A1

September 25, 2022 1:28 PM

Problem 1 [2 points] (p. 32, exercise 25) For small x, the approximation $\sin x \approx x$ is often used. For what range of x is this good to a relative accuracy of $\frac{1}{2}10^{-14}$?

$\sin(x) \approx x$

$\delta = 5 \times 10^{-15}$

$$\delta = \left| \frac{\sin(x) - x}{\sin(x)} \right|$$

TS for $\sin(x) \approx x - \frac{x^3}{3!} - \frac{x^5}{5!} - \frac{x^7}{7!} \ldots$

for $|x| << 1$

$\sin(x) \approx x - x^3/3!$

$\sin(x) - x \approx -x^3/3!$

substitute $\sin(x) - x \approx -x^3/3!$ into δ

$$\delta = \left| \frac{-x^3/3!}{x - x^3/3!} \right|$$

$$\delta = \left| \frac{-x^2/6}{1 - x^2/6} \right|$$

$$|1 \quad 1 - x^2/6 \quad| \approx 1$$

$$\therefore \; |x| \ll 1, \; 1 - x^2/6 \approx 1$$

$$\delta = \left| \frac{-x^2/6}{1} \right| = \left| -\frac{x^2}{6} \right| = \frac{x^2}{6}$$

$$\delta = \frac{x^2}{6} > 5 \times 10^{-15}$$

$$x^2 > 30 \times 10^{-15}$$

$$x > \sqrt{3 \times 10^{-14}}$$

$$-1.73 \times 10^{-7} < x < 1.732 \times 10^{-7}$$

Problem 2 [4 points] (p. 33, exercise 41) Write the Taylor series for

a. e^{x+2h}

b. $\sin(x - 3h)$

$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)(x-a)^n}{n!}$ $\quad f(x) = e^{x+2h} \quad f'(x) = e^{x+2h} \ldots$

a)

$$n = 0 \rightarrow \frac{f(0)\,(x-0)^0}{0!} = \frac{e^{2h}(1)}{1} = e^{2h}$$

$$n = 1 \rightarrow \frac{f'(0)\,(x-0)^1}{1!} = \frac{e^{2h}x}{1}$$

$$n = 2 \rightarrow \frac{f''(0)(x-0)^2}{2!} = \frac{e^{2h}x^2}{2}$$

$n=0 \rightarrow \frac{f^{(n)}(a)(x-a)^n}{n!} = \frac{e^{2h}x^n}{2}$

$$T(x) = \sum_{n=0}^{\infty} \frac{e^{2h}x^n}{n!}$$

b) $f(x) = \sin(x-3h)$ $\quad f'''(x) = -\sin(x-3h)$

$f'(x) = \cos(x-3h)$ $\quad f''''(x) = -\cos(x-3h)$

$n=0 \rightarrow \frac{f(0)(x-0)^0}{0!} = \frac{\sin(-3h)(1)}{1} = \frac{-\sin(3h)x^0}{0!}$

$n=1 \rightarrow \frac{f'(0)(x-0)^1}{1!} = \frac{\cos(-3h)(x)}{1!} = \frac{\cos(3h)x^1}{1!}$

$n=2 \rightarrow \frac{f''(0)(x-0)^2}{2!} = \frac{-\sin(-3h)x^2}{2} = \frac{\sin(3h)x^2}{2!}$

$n=3 \rightarrow \frac{f'''(0)(x-0)^3}{3!} = \frac{-\cos(-3h)x^3}{6} = \frac{-\cos(3h)x^3}{3!}$

$$T(x) = -\sin(3h) + \cos(3h)x + \frac{\sin(3h)x^2}{2} - \frac{\cos(3h)x^3}{6} \ldots$$

Problem 3 [3 points] Suppose you approximate e^x by its truncated Taylor series. For given $x = 0.5$, derive how many terms of the series are needed to achieve accuracy of 10^{-10}.

$$\text{Error} = |R_n| = \left| \frac{f^{(n+1)}(a)}{(n+1)!} x^{n+1} \right|$$

For e^x, $f^{(n+1)}(a) = e^x$

For e^x, $f^{(n+1)}(a) = e^x$

$|R_n| = 10^{-10} \geq \left| \frac{e^x}{(n+1)!} x^{n+1} \right|$

let $x = 0.5$

$$10^{-10} \geq \left| \frac{e^{0.5}}{(n+1)!} 0.5^{n+1} \right| = \frac{e^{0.5}}{(n+1)!} 0.5^{n+1}$$

@ $n = 8$

$$\frac{e^{0.5}}{9!} 0.5^9 = 8.8739 \times 10^{-9} \not\leq 10^{-10}$$

@ $n = 9$

$$\frac{e^{0.5} \cdot 0.5^{10}}{10!} = 4.4369 \times 10^{-9} \not\leq 10^{-10}$$

@ $n = 10$

$$\frac{e^{0.5} \cdot 0.5^{11}}{11!} = 2.0168 \times 10^{-11} < 10^{-10}$$

∴ 10 terms of the series are needed to achieve accuracy of 10^{-10}

Problem 4 [2 points] Consider the expression $(1 - a)(1 + a)$. In double precision, for what values of a does this expression evaluate to 1?

$(1-a)(1+a) >$

greatest # less than 1 in binary double

$2^{-1}\left[\left(\frac{1}{2}\right)(1) + \left(\frac{1}{2}\right)^2(1) + \ldots + \left(\frac{1}{2}\right)^{52}(1)\right]$

$$2^{-1}\left[\left(\tfrac{1}{2}\right)(1) + \left(\tfrac{1}{2}\right)^2(1) + \dots + \underbrace{\left(\tfrac{1}{2}\right)^{52}}_{\varepsilon_{mach}}(1)\right]$$

$$= 1 - \underbrace{\left(\tfrac{1}{2}\right)^{52}}_{\varepsilon_{mach}}$$

Smallest # that rounds up to 1

$$1 - \varepsilon_{mach} < (1+a)(1-a)$$

$$1 - \varepsilon_{mach} < 1 - a^2$$

$$\varepsilon_{mach} > a^2$$

$$\sqrt{\varepsilon_{mach}} > a$$

$$\sqrt{\varepsilon_{mach}} > a > -\sqrt{\varepsilon_{mach}}$$

$$(1/2)^{26} > a > -(1/2)^{26}$$

```
>> (1+0.5^27)*(1-0.5^27)==1

ans =

  logical

   1

>> (1+0.5^26)*(1-0.5^26)==1

ans =

  logical

   0
```

Problem 5 [2 points] Give an example in base-10 computer arithmetic when

a. $(a+b)+c \neq a+(b+c)$

b. $(a*b)*c \neq a*(b*c)$

a) assume $t = 2$, $\varepsilon_{mach} = 0.1 \rightarrow U = 0.05$

$a = 1$ $\qquad (a+b) = \emptyset + c = 0.05$

$b = -1$ $\qquad 1 + (-1 + 0.05) = \emptyset$

$c = 0.05$

b) assume $t = 3$

$a = 1.23$
$b = 2.82$
$c = 1.01$

$(a \cdot b) = 1.23 \times 2.82 \doteq 3.47$
$(a \cdot b) \cdot c = 3.47 \cdot 1.01 = 3.50$

$(b \cdot c) = 2.82 \times 1.01 \doteq 2.85$
$a \cdot (b \cdot c) = 1.23 \times 2.85 \doteq 3.51$

Problem 6 [8 points] Suppose you need to generate $n+1$ equally spaced points in the interval $[a, b]$ with spacing $h = (b-a)/n$, $n > 1$. You can use either

$$x_0 = a, \quad x_i = x_{i-1} + h, \quad i = 1, \ldots, n \quad \text{or} \tag{1}$$
$$x_i = a + ih, \quad i = 0, \ldots, n. \tag{2}$$

Denote by $\tilde{x}_i$ the computed value in (1) and by $\hat{x}_i$ the computed value in (2).

a. [2 points] Which of $|x_i - \tilde{x}_i|$ and $|x_i - \hat{x}_i|$ is more accurate? Explain why.

b. [2 points] Write a MATLAB program that implements both methods and illustrates the difference between them.

a) $\tilde{x}$ propagates the error of each calculation because of its recursive nature

For Example: let $a = 1$, $b = 1 + \varepsilon_{mach}$, $n = 10$

$\tilde{x}$ $\tilde{x}$ $\hat{x}$: 1

For example let $a = 1$, $b = 1 + \varepsilon_{mach}$, $n = 10$

$\tilde{x}_i = \tilde{x}_{i-1} + h$, $x_0 = a$ $\qquad$ $\hat{x}_i = a + ih$

Using FP arithmetic:

$\tilde{x}_0 = 1$

$\tilde{x}_1 = 1 + \varepsilon_{mach}/10 \approx 1$

$\tilde{x}_2 = 1 + \varepsilon_{mach}/10 \approx 1$

...

$\tilde{x}_9 = 1 + \varepsilon_{mach}/10 \approx 1$

$\tilde{x}_{10} = 1 + \varepsilon_{mach}/10 \approx 1$

$\therefore \tilde{x}_{10} = 1$

$$h = \frac{((1 + \varepsilon_{mach}) - 1)}{10}$$

@ $i = 10$,

$\hat{x}_1 = 1 + 10h$

...

$\hat{x}_{10} = 1 + 10 \dfrac{((1+\varepsilon_{mach})-1)}{10}$

$\hat{x}_{10} = 1 + \varepsilon_{mach}$

$\hat{x}_{10} = 1 + \varepsilon_{mach}$

Using manual arithmetic

$x_0 = 1$

$x_1 = 1 + \varepsilon_{mach}/10$

$x_2 = (1 + \varepsilon_{mach}/10) + \varepsilon_{mach}/10$

$x_3 = (1 + 2\varepsilon_{mach}/10) + \varepsilon_{mach}/10$

...

$x_4 = (1 + 8\varepsilon_{mach}/10) + \varepsilon_{mach}/10$

$x_{10} = 1 + 10\varepsilon_{mach}/10 = \underbrace{1 + \varepsilon_{mach}}$

$\therefore |x_{10} - \hat{x}_{10}| = 0$ however $|x_{10} - \tilde{x}_{10}| = \varepsilon_{mach}$

$\therefore \hat{x}$ is more accurate in this case

```
global n;
global a;
global b;
n = 10;
a = 1;
b = 1+2*eps;

for j = 0:n
    fprintf("****i=%i****\n", j)
    fprintf("x1: %f\n  ", x1(j))
    x1(j) == 1
    fprintf("x2: %f\n  ", x2(j))
    x2(j) == 1
end

function x = x1(i)
    global n;
    global a;
    global b;
    h = (b-a)/n;

    if i == 0
        x = a;
    else
        x = x1(i-1)+h;
    end
end

function x = x2(i)
    global n;
    global a;
    global b;
    h = (b-a)/n;

    x = a + i*h;
end
```

$\tilde{x}$

$\hat{x}$

```
****i=0****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   1

****i=1****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   1

****i=2****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   1

****i=3****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   0

****i=4****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   0

****i=5****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   0

****i=6****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   0

****i=7****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   0

****i=8****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   0

****i=9****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   0

****i=10****
x1: 1.000000
ans =
  logical
   1
x2: 1.000000
ans =
  logical
   0
```

Since $\hat{x}_3$ and all further $\hat{x}$ terms are not rounded to 1 & all $\tilde{x}_3$ terms are rounded to 1, therefore the analysis from part (a) is proven & $\hat{x}$ is more accurate

Problem 7 [6 points] Consider the approximation, $h > 0$

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

Assume that $f'''(x)$ in continuous on $[x-h, x+h]$.

a. [2 points] If we approximate $f'(x)$ by $(f(x+h) - f(x-h))/(2h)$, what is the truncation error of this approximation?

$$f(x+h) = f(x) + \frac{f'(x)h}{1!} + \frac{f''(x)h^2}{2!} + \frac{f'''(\xi)h^3}{3!}$$

$$f(x-h) = f(x) - \frac{f'(x)h}{1!} + \frac{f''(x)h^2}{2!} - \frac{f'''(\xi)h^3}{3!}$$

* ξ between $x-h$ & $x+h$

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

$$= f'(x)2h + \frac{f'''(\xi)h^3}{3!}$$

$$= f'(x)2h + \frac{f''''(\xi)h^2}{3!}2h$$

$$= 2h\left(f'(x) + \frac{f'''(\xi)h^2}{3!}\right)$$

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

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

$\therefore$ truncation error is $-\frac{f'''(\xi)h^2}{3!}$

b. [2 points] When evaluated on a computer, for what value of h the error of this approximation is the smallest?

$$f_1 = f(x+h) + \delta_1$$
$$f_2 = f(x-h) + \delta_2$$

$$\frac{f_1 - f_2}{2h} = \frac{f(x+h) - f(x-h)}{2h} + \frac{\delta_1 - \delta_2}{2h}$$

$$f'(x) - \frac{f_1 - f_2}{2h} = \frac{f(x+h) - f(x-h)}{2h} - \frac{1}{6} f'''(\xi) h^2 - \frac{f(x+h) - f(x-h)}{2h} - \frac{\delta_1 - \delta_2}{2h}$$

$$f'(x) - \frac{f_1 - f_2}{2h} = -\frac{1}{6} f'''(\xi) h^2 - \frac{\delta_1 - \delta_2}{2h}$$

$$\left| f'(x) - \frac{f_1 - f_2}{2h} \right| \leq \left| \frac{1}{6} f'''(\xi) h^2 \right| + \left| \frac{\delta_1 - \delta_2}{2h} \right| \quad \text{* } |\delta_1|, |\delta_2| \leq \varepsilon_{mach}$$

$$\left| f'(x) - \frac{f_1 - f_2}{2h} \right| \leq \frac{1}{6} M h^2 + \frac{2\varepsilon_{mach}}{2h}$$

$$g(h) = \frac{1}{6} M h^2 + \varepsilon_{mach}/h$$

$$g'(h) = \frac{1}{3} M h - \varepsilon_{mach}/h^2$$

$$0 = \frac{1}{3} M h - \varepsilon_{mach}/h^2$$

$$\frac{1}{3} M h = \varepsilon_{mach}/h^2$$

$$h^3 = 3\varepsilon_{mach}/M$$

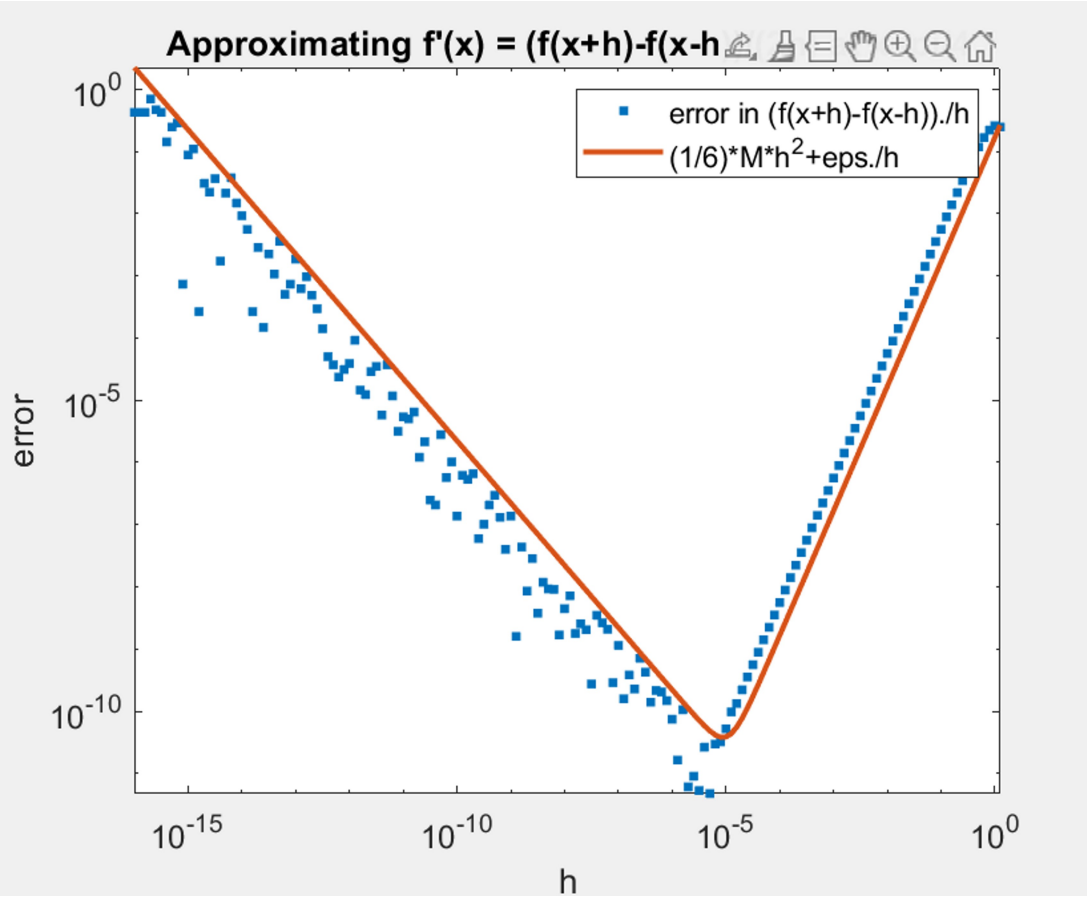
$$h = \sqrt[3]{3\varepsilon_{mach}/M}$$

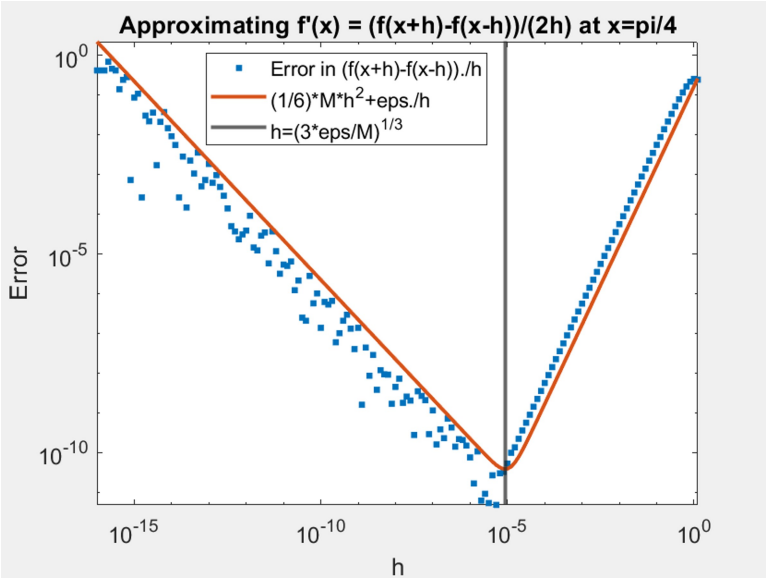
$\therefore$ $g(h)$ is smallest @ $h = \sqrt[3]{3\varepsilon_{mach}/M}$

c. [2 points] For the function $f(x) = \sin x e^{\cos x}$ plot the error

$$\left| f'(x) - \frac{f(x+h) - f(x-h)}{2h} \right|$$

versus h for appropriate values of h. Plot on a `loglog` scale. Submit your plot in the hardcopy. How does the error match your derivation in the previous part?





As demonstrated in the graph to the left, the error as represented on the graph matches with the derivation from the previous part. The gray line represents the value derived in part b, and it can be observed that the line intersects the minimum of the red error line from part c. Therefore, the minimum error occurs at the value derived in part b.

Problem 8 [4 points] Consider

$$f(x) = \frac{e^x - x - 1}{x^2}.$$

When evaluated with $|x| < 1$, the relative error can be large.

a. [1 point] Explain why this error can be large.

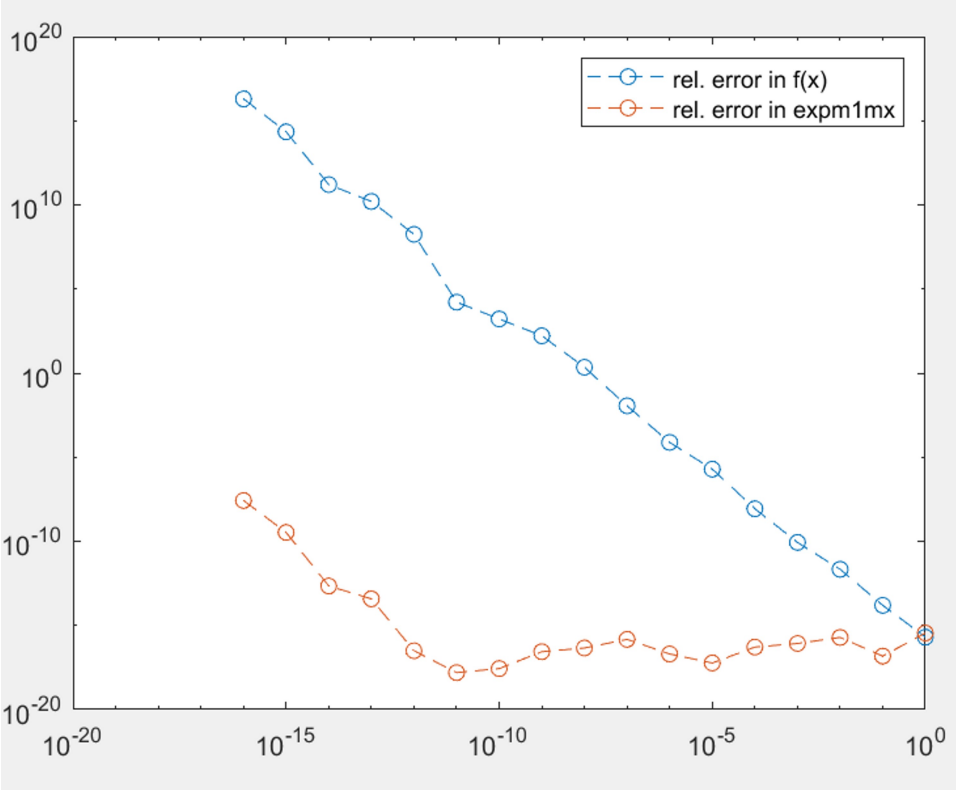
For very small values of x, e^x-x can be rounded to 1. This can be demonstrated by the taylor series of e^x:

$$e^x \approx 1 + x + x^2/2! + x^3/3! + \ldots$$

So, if the third term of the series is evaluates to being smaller than machine epsilon, then e^x is rounded to 1 + x. This can cause issues in the formula above: with a sufficiently small value for x, the numerator will be equal to zero due to rounding error. Thereby making the entire formula equal to zero and due to large rounding error for a sufficiently small x.

b. [3 points] Write a MATLAB function

```
function y = expm1mx(x)
% Evaluates (exp(x) - 1 - x)/x^2 accurately for |x| < 1.
```



```
function y = expm1mx1(x)
    i = 2; y = 0;
    while i<=18
        %factorial calculator
        fact = 1;
        if i ~= 0
            for n = 1:i
                fact = fact*n;
            end
        end
        %function
        y = y + (x.^(i-2))./fact;
        i = i + 1;
    end
end
```

Problem 9 (7 points) The following MATLAB script

```
g = @(x) (exp(x)-1-x)./x.^2;
h = @(x) (exp(x)-x-1)./x.^2;
x = 1e-10;
fprintf('x=%.16e\ng(x)=%.16e\nh(x)=%.16e\n', x, g(x), h(x))
x = 2^(-33);
fprintf('x=%.16e\ng(x)=%.16e\nh(x)=%.16e\n', x, g(x), h(x))
```

produces (on my machine)

```
x=1.0000000000000000e-10
g(x)=8.2740370962658164e+02
h(x)=0.0000000000000000e+00
x=1.1641532182693481e-10
g(x)=0.0000000000000000e+00
h(x)=0.0000000000000000e+00
```

(3, 2, 1,1 points) Explain the values for each of the `g(x)` and `h(x)`.

Taylor series of e^x is 1 + x + x^2/2…
If x = 1e-10, x^2/2 = 5e-21 < emach, therefore e^x = 1+ x if e^x is being added to or subtracted from 1
If x = 2^(-33), x^2/2 = 6.7763e-21 < emach, therefore e^x = 1+x **if** e^x is being added to or subtracted from 1

$g(1\times10^{-10})$

$$e^x \doteq 1 + 10^{-10} + 10^{-20}/2! + 10^{-30}/3! + \dots$$

$$e^x - 1 = (1 + 10^{-10} + 10^{-20}/2! + 10^{-30}/3! + \dots) - 1$$

$$e^x - 1 = 10^{-10} + 10^{-20}/2! + 10^{-30}/3! + \dots$$

$$e^x - 1 - x = 10^{-20}/2! + 10^{-30}/3! + \dots$$

$h(1\times10^{-10})$

$$e^x \doteq 1 + 10^{-10} + \frac{(10^{-10})^2}{2!} + \frac{(10^{-10})^3}{3!} + \dots$$

$$(e^x) - x = (1 + 10^{-10} + (10^{-10})^2/2! + (10^{-10})^3/3! + \dots) - 10^{-10}$$

$$= 1 + (10^{-10})^2/2! + (10^{-10})^3/3! + \dots$$

$$(e^x - x) - 1 = (1 + (10^{-10})^2/2! + (10^{-10})^3/3! + \dots) - 1$$

- $(10^{-10})^2/2! < \varepsilon_{mach}$, $(10^{-10})^3/3! < \varepsilon_{mach}$ etc
- $(10^{-10})^2/2! - 1 = -1$, and so on ...

$$(e^x - x) - 1 = 1 - 1 = \emptyset$$

$$h(1\times10^{-10}) = \emptyset$$

$g(2^{-33})$

$$e^x \doteq 1 + 2^{-33} + 2^{-66}/2! + 2^{-99}/3! \dots$$

∵ every term in the taylor series after term 2 is $< \varepsilon_{mach}$,

∴ $e^x - 1 = 1 + 2^{-33} - 1 = 2^{-33}$

$$(e^x) - 1 = 2^{-33}$$

$$(e^x - 1) - x = (2^{-33}) - 2^{-33} = \emptyset$$

$$g(2^{-33}) = \emptyset$$

$h(2^{-33})$

$$e^x \doteq 1 + 2^{-33} + 2^{-66}/2! + 2^{-99}/3! \dots$$

$$(e^x) - x = (1 + 2^{-33} + 2^{-66}/2! + 2^{-99}/3! \dots) - 2^{-33}$$

$$= (1 + 2^{-66}/2! + 2^{-99}/3! \dots)$$

$$(e^x - x) - 1 = (1 + 2^{-66}/2! + 2^{-99}/3! + \dots) - 1 = \emptyset$$

$$h(2^{-33}) = \emptyset$$