4. $f(x) = 1 - e^{-x}$; $x \in [0,1]$

(a) $(\text{cond} f)(x) = \left\| \frac{x f'(x)}{f(x)} \right\| = \left\| \frac{x e^{-x}}{1-e^{-x}} \right\| = \left\| \frac{x}{e^{x}-1} \right\|$

Now; $e^{x} - 1 = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$

so; $(\text{cond} f)(x) = \left\| \frac{x}{\cancel{1}+x+\frac{x^2}{2!}+\cdots - \cancel{1}} \right\| = \left\| \frac{1}{1+\frac{x}{2!}+\cdots} \right\|$ (for $x \neq 0$)

Now; $1 + \frac{x}{2!} + \frac{x^2}{3!} + \cdots \geq 1$ for $x \in [0,1]$

and so $\left\| \frac{1}{1+\frac{x}{2!}+\cdots \frac{x^n}{(n+1)!}+\cdots} \right\| \leq 1$ for $x \in [0,1]$

However; in the previous step; we have cancelled x in numerator and denominator

$\Rightarrow x \in (0,1]$

$\Rightarrow \left\| \frac{1}{1+\frac{x}{2!}+\cdots} \right\| < 1$ (always less than 1 in our domain of interest).

(b) ~~$fl(1-e^{-x})$~~ $fl(1 - fl(\exp(-fl(x))))$

$= fl(1 - fl(\exp(-x)))$ (as x is a machine # and negation of x does not produce any error)

$= fl(1 - \exp(-x)(1+\epsilon_e))$

$= \{1 - \exp(-x)(1+\epsilon_e\}(1+\epsilon_s)$

$\Rightarrow (1 - e^{-x}(1+\epsilon_e))(1+\epsilon_s)$

$\Rightarrow 1 - e^{-x} - \epsilon_e e^{-x} + \epsilon_s - \epsilon_s e^{-x} + O(\epsilon^2)$

$\Rightarrow (1-e^{-x})\left\{1 + \frac{\epsilon_s - e^{-x}(\epsilon_e+\epsilon_s)}{1-e^{-x}}\right\}$

ϵ_e = due to exponential

ϵ_s = due to subtraction

So; $f_A(x) = (1-e^{-x})(1+\epsilon_A)$

where $\boxed{\epsilon_A = \dfrac{\epsilon_s - e^{-x}(\epsilon_s+\epsilon_e)}{1-e^{-x}}}$

and; $f(x_A) = 1-e^{-x_A}$

Now; $f(x_A) = f_A(x)$ (the equaling to get x_A).

$\therefore\ 1-e^{-x_A} = (1-e^{-x})(1+\epsilon_A)$.

$\Rightarrow\ \cancel{1}+e^{-x_A} = \cancel{1}+e^{-x} - \epsilon_A + \epsilon_A e^{-x}$

$\Rightarrow\ e^{-x_A} = e^{-x} + \epsilon_A e^{-x} - \epsilon_A$

$\Rightarrow\ e^{-(x_A-x)} = 1 + \epsilon_A - \epsilon_A e^{x}$

$\Rightarrow\ -(x_A-x) = \ln(1+\epsilon_A - \epsilon_A e^{x})$

$\Rightarrow\ \cancel{x - x_A}\ \ x_A - x = -\ln(1+\epsilon_A-\epsilon_A e^{x})$.

So; $x_A - x = -\{\epsilon_A - \epsilon_A e^{x}\}$ (1st order in ϵ_A).

$\left[\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \cdots\right]$

So; $x_A - x = \epsilon_A e^{x} - \epsilon_A$

$\Rightarrow\ x_A - x = \epsilon_A\left(\cancel{1} + x + \frac{x^2}{2!} + \cdots - \cancel{1}\right)$

So; $\boxed{\dfrac{x_A - x}{x} = \epsilon_A\left(1 + \frac{x}{2!} + \cdots\right)}$

$\therefore\ \cancel{(\mathrm{cond}\,A)(x) = \left\|\frac{x_A-x}{x}\right\| / \epsilon}$

$\Rightarrow\ \cancel{\left\|\frac{\epsilon_A}{\epsilon}\left(1+\frac{x}{2!}+\cdots\right)\right\|}$

Now: $\epsilon_A = \dfrac{\epsilon_s - e^{-x}(\epsilon_s+\epsilon_e)}{1-e^{-x}} = \cancel{\dfrac{\epsilon_s - e^{-x}\epsilon_e}{1-e^{-x}}}$.

So, $\left\| \frac{x_A - x}{x} \right\| / \epsilon = (\text{cond}A)(x) = \left\| \frac{\epsilon_A}{\epsilon} \left(1 + \frac{x}{2!} + \cdots \right) \right\|$

Now, $\left(1 + \frac{x}{2!} + \frac{x^2}{3!} + \cdots \right) \geq 1$ for $x \in [0,1]$

and $$\epsilon_A = \frac{\epsilon_s - e^{-x}(\epsilon_s + \epsilon_e)}{1 - e^{-x}}$$

Now; ϵ_s = error propagation due to subtraction
$= \frac{x \epsilon_x - y \epsilon_y}{x - y}$ and here; $x \equiv 1$ and $y \equiv e^{-x}$.

In our domain; $1 - e^{-x}$ very small.

~~So, it is~~ Also, we are computing e^{-x} using in-built routine $\Rightarrow$ very small error ϵ_e. So; it is fair to assume that in our domain; $\epsilon_s \gg \epsilon_e$.

So, $\epsilon_A \approx \epsilon_s$ and if ϵ is machine precision; then $\epsilon_s = \epsilon_A \sim \frac{1}{1 - e^{-x}}$ which is large in $x \in [0,1] \Rightarrow \frac{\epsilon_A}{\epsilon} > 1$ (another fair assumption)

So, $(\text{cond}A)(x) = \left\| \underbrace{\frac{\epsilon_A}{\epsilon}}_{>1} \underbrace{\left(1 + \frac{x}{2!} + \cdots \right)}_{>1} \right\| > 1 \Rightarrow$ is greater than 1 everywhere in our domain

4. (c) So, the root cause of $(\text{cond}A)(x) > 1$ in our domain is ~~that~~ two fold :-

- We have <u>subtraction operations</u> that cause ϵ_A to be large ~~at~~ in our domain due to a factor of $(1 - e^{-x})$ in the denominator.
- We can set an upper bound to $\left(1 + \frac{x}{2!} + \cdots \right)$ by approximating it as :- $\frac{e^x - 1}{x}$ whose upper bound in $x \in [0,1]$ is $\frac{e - 1}{x}$ and so; as $x \to 0$ the upper bound of $(\text{cond}A)(x)$ increases drastically.

(d) $f(x) = 1 - e^{-x}$. in the operation $(x-y)$

so, the loss of significant bits can be between b and a when $\cancel{2^{-a} < |x - y/x| < 2^{-b}}$ $2^{-a} \leq 1 - y/x \leq 2^{-b}$.

Now, $x \equiv 1$ and $y \equiv e^{-x}$ in our problem

so, $2^{-a} \leq 1 - e^{-x} \leq 2^{-b}$

Now; $1 - e^{-x} \geq 2^{-a}$

$\Rightarrow -e^{-x} \geq 2^{-a} - 1 \Rightarrow e^{-x} \leq 1 - 2^{-a}$

so, $-x \leq \ln(1-2^{-a})$

or; $\boxed{x \geq -\ln(1-2^{-a})}$

so, to ensure atmost 1 bit of significance is lost;

$a = 1$.

so; $x \geq -\ln(1-\frac{1}{2}) \Rightarrow \boxed{x \geq \ln 2}$

so; least value of $x = \underline{\underline{0.6931}}$

- for $a=2 \longrightarrow x^2_{min} = -\ln(1-\frac{1}{4}) = \ln\left(\frac{4}{3}\right) = \underline{\underline{0.2876}}$
- for $a=3 \longrightarrow x^3_{min} = -\ln(1-\frac{1}{8}) = \ln\left(\frac{8}{7}\right) = \underline{0.1335}$
- for $a=4 \longrightarrow x^4_{min} = -\ln(1-\frac{1}{16}) = \ln\left(\frac{16}{15}\right)$

$= \underline{\underline{0.0645}}$

(e) so, we need to compute:-

$$(\text{cond} f)(x) = \left| \frac{1}{1 + \frac{x}{2!} + \cdots} \right|$$

so putting ~~upper~~ lower bounds in x would mean putting upper bounds in $(\text{cond} f)(x)$.

we can also write:- $(\text{cond} f)(x) \leq \left| \frac{x}{e^x - 1} \right|_{x_{lower}}$

So; $(\text{cond } f)(x_{\text{lower bound}} = \ln 2) = \frac{\ln 2}{2-1} = \ln 2$

$= \underline{\underline{0.693}}$

$(\text{cond} f)\left(x_{\text{lower bound}} = \ln \frac{4}{3}\right) = \frac{\ln(4/3)}{\frac{4}{3}-1} = 3\ln\left(\frac{4}{3}\right)$

$= \underline{\underline{0.8628}}$

$(\text{cond} f)\left(x_{\text{lower bound}} = \ln\left(\frac{8}{7}\right)\right) = \frac{\ln(8/7)}{\frac{8}{7}-1} = 7\ln\left(\frac{8}{7}\right)$

$= \underline{\underline{0.9345}}$

$(\text{cond} f)\left(x_{\text{lower bound}} = \ln\left(\frac{16}{15}\right)\right) = \frac{\ln(16/15)}{\frac{16}{15}-1} = 15\ln\left(\frac{16}{15}\right)$

$= \underline{\underline{0.9675}}$

So; indeed although we end up losing more and more significant bits; the (cond f) is <u>not</u> ill-posed and so; the underlying math problem is well posed. So; the problem ~~might~~ ~~be~~ of ill-conditioned algorithm might be solved by choosing a better algorithm.

(f) The major problem in the algorithm was the subtraction operation and the fact that $\epsilon_{x-y} \sim \frac{x\epsilon_x - y\epsilon_y}{x-y}$. Here; x and y became ~~to~~ very similar as $x \to 0$.

So; changing the algorithm to get rid of subtraction operations would be the key :-

$$1 - e^{-x} = \frac{e^x - 1}{e^x} = \boxed{\frac{x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots}{e^x} = f_A(x)}$$

This algorithm involves only additions and division. Neither of these operations ~~suf~~ suffer from the error blowing up as $x \to 0$. So; this can be an algorithm to mitigate error propagation and make the algorithm well conditioned.