

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

a)  $f'(x) = e^{-x}$ . Thus,

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

on  $[0, 1]$ ,  $e^x \geq 1$  and  $e^x \leq e$

As  $x \rightarrow 0$   $\frac{x}{e^x - 1} \rightarrow 1 \leq 1$  and as  $x \rightarrow 1$ ,  
 $\frac{x}{e^x - 1} \rightarrow \frac{1}{e - 1} \leq 1$

It seems that, as  $(\text{cond } f)(x)$  has no local minima or maxima, it goes monotonically from 1 to  $1/(e-1)$  on this interval.

b) The condition of  $A$  is:

$$(\text{cond } A)(\bar{x}) = \frac{|\bar{x}_A - \bar{x}|}{|\bar{x}|}$$

find first what  $f_A(x)$  yields. Then find

$$x_A \equiv f^{-1}(f_A(x))$$

$f_A(x)$ : first,  $x \rightarrow -x(1 + \epsilon_x)$

then,  $e^{-x(1+\epsilon_x)}(1+\epsilon) = e^{-x}(1 - x\epsilon_x + \epsilon)$

and,  $1.0 - e^{-x}(1 - x\epsilon_x + \epsilon) = f_A(x)$

Assume the 1.0 has no error. So any error only comes from the  $e^{-x}$ . Let's try to move the error to the exponent for ease of finding  $x_A$ .

$$e^{-x}(1 + \epsilon - x\epsilon_x) \approx e^{-x} e^{\epsilon - x\epsilon_x} \approx e^{-x + \epsilon - x\epsilon_x}$$

$$\Rightarrow x_A = x + \epsilon + x\epsilon_x$$

$$\Rightarrow \text{Cond A}(x) = \frac{|x + x\varepsilon_x + \varepsilon - x|}{|x|} \varepsilon^{-1}$$

$$= |\varepsilon_x + \varepsilon/x| \varepsilon^{-1}$$

Assume that  $\varepsilon = \varepsilon_x$ . Then

$$\text{Cond A}(x) = |1 + 1/x| > 1 \text{ on } 0 < 1.$$

In fact as  $x \rightarrow 0$ ,  $\text{Cond A}(x)$  diverges.

(C) See plot attached.

The root of the problem is that the error from exponentiation,  $\varepsilon$ , gets comparable to the magnitude of  $x$  as  $|x|$  gets small.

d) Suppose we have a  $p$  bit mantissa.

Any number on  $[0, 1]$  can be written as

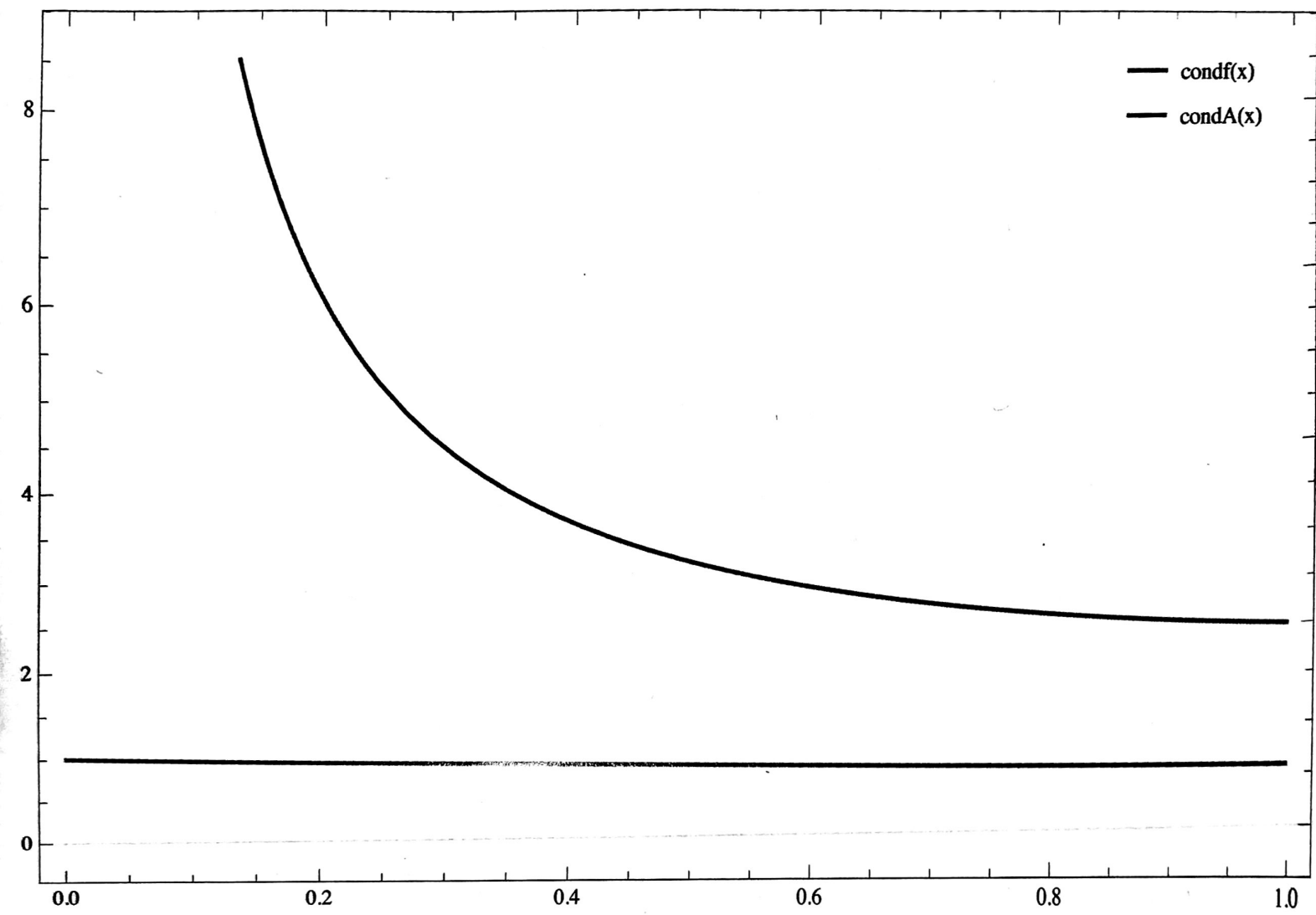
$$x = \sum_{n=1}^p b_n 2^{-n} \text{ (rounded)}$$

The last bit of significance is  $b_p 2^{-p}$ .

This is true even when  $x$  is small. Let's at least

write 
$$x = 2^e \left( \sum_{n=1}^p b_n 2^{-n} \right)$$

c.



It's not clear to me if it's meant  
 $n$  bits in  $x_A$  or in  $f$ .

$f$  isn't too sensitive to errors, so I'll  
 assume it's in  $x$ .

If  $\epsilon = 2^{-p-1}$ , then  $2\epsilon$  is losing one digit  
 of significance, and  $2^n \epsilon$  is losing  $n$  bits of  
 significance.

When does  $1 + 1/x = 2^n$ ?

$$(2^n - 1)^{-1} = x_{\min}$$

We require  $x > x_{\min}$  to lose at most  
 $n$  bits of significance. When  $n=1$ ,

$x > x_{\min} = 1$ . We always lose one bit of  
 precision.

What if  $n=2, 3, 4$ ?

$n$	$x_{\min}$
2	$1/3$
3	$1/7$
4	$1/15$

e) We know the relative error is bounded by  
 $\text{cond}(x) (\epsilon + \epsilon(\text{cond } A)x) = \text{err}(x)$

So:

$x$	$\text{err}(x)/\epsilon$
1	1.746
$1/3$	1.966
$1/7$	1.993
$1/15$	1.999

f) When  $x$  is small, we could use an alternative method. Namely,

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

When  $x$  is small, this is well behaved.

We see that  $f(x) \rightarrow \frac{1 + 2x}{1 - x}$  for small  $x$ .

This is fine.