I. Can we compute the root with absolute accuracy $< 10^{-6}$?

We start the bisection method with the interval $[128, 129]$, and the interval will be shorten by half after one iteration. Assume that the interval after the n-th iteration is $[a_n, b_n]$, $a_0 = 128 = 1 \times 2^7$ $b_0 = 129 = \left(1 + 2^{-7}\right) \times 2^7$, $n \leq$ the iteration times. We can easily find that $a_n \in \mathcal{F}$, $b_n \in \mathcal{F}$ until a_n's or b_n's mantissa contains 2^{-23} and the iteration will end. So we know the iteration will end when $b_n - a_n = 2^{23} \times 2^7 = \epsilon_M \times 2^7$.

Thus we have $\max E_a bs = \epsilon_u \times 2^7 = 2^{-17} \approx 7.629 \times_{10}^{-6} > 10^{-6}$. So the absolute accuracy may be bigger than 10^{-6}.

II. What are the condition numbers of the following functions?Where are they large?

II-a $(x-1)^\alpha$.

$$C_f(x) = \left| \frac{x\alpha (x-1)^{\alpha-1}}{(x-1)^\alpha} \right| = \left| \frac{\alpha x}{x-1} \right| \to \infty \quad \text{when} \quad x \to 1.$$

II-b $\ln x$.

$$C_f(x) = \left| \frac{x \frac{1}{x}}{\ln x} \right| = \left| \frac{1}{\ln x} \right| \to \infty \quad \text{when} \quad x \to 1.$$

II-c e^x.

$$C_f(x) = \left| \frac{xe^x}{e^x} \right| = |x| .$$

The $C_f(x)$ will be large when x is large.

II-d $\arccos x$.

$$C_f(x) = \left| \frac{x \frac{-1}{\sqrt{1-x^2}}}{\arccos x} \right| = \left| \frac{x}{\arccos x \sqrt{1-x^2}} \right| .$$

Since $\sqrt{1-x^2} = o(x)$ and $\arccos x$ is bouded, we have $C_f(x) \to \infty$ when $x \to \pm 1$.

III. The last Exercise in Section 1.3.5 in the notes.

$$\begin{aligned}
\text{cond}_f(x) &= \left| \frac{xf'(x)}{f(x)} \right| = \frac{x}{\sin x} \\
f_A(x) &= \frac{\sin x (1+\delta_2)}{[1+\cos x (1+\delta_1)](1+\delta_3)} (1+\delta_4) \qquad |\delta_i| < \epsilon_u \\
&= \frac{\sin x}{1+\cos x} \left(1 - \frac{\cos x \delta_1}{1+\cos x + \cos x \delta_1}\right) \frac{(1+\delta_2)(1+\delta_4)}{(1+\delta_3)} \\
&\approx \frac{\sin x}{1+\cos x} \left(1 + \delta_2 + \delta_4 - \delta_3 - \frac{\cos x}{1+\cos x}\delta_1\right) \\
\Longrightarrow &\varphi(x) = 3 + \frac{\cos x}{1+\cos x} \\
\therefore \text{cond}_A(x) &\leq \frac{\sin x}{x} \left(3 + \frac{\cos x}{1+\cos x}\right).
\end{aligned}$$

IV. Consider the function $f(x) = 1 - e^{-x}$ for $x \in [0, 1]$.

IV-a Show that $\text{cond}_f(x) \leq 1$ for $x \in [0, 1]$.

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

Because $e^x - 1 = o(x)$ when $x \to 0$, so $\text{cond}_f(x) \to 1$ when $x \to 0$.

Assume $g_1(x) = x$, $g_2(x) = e^x - 1$. Because $g_1(0) = g_2(0) = 0$, $\forall x \in [0,1]$ $g_1'(x) = 1 \leq e^x = g_2'(x)$, we have $\forall x \in [0,1]$ $g_1(x) \leq g_2(x)$. Thus $\text{cond}_f(x) \leq 1$.

IV-b Estimate $\text{cond}_A(x)$ for $x \in [0,1]$.

$$\begin{aligned} A &= \left(1 - e^{-x}(1+\delta_1)\right)(1+\delta_2) \qquad |\,\delta_i\,| < \epsilon_u \\ &= \left(1-e^{-x}\right)\left(1+\delta_2 - \frac{1}{e^x-1}\delta_1(\delta_2+1)\right) \\ &\approx \left(1-e^{-x}\right)\left(1+\delta_2 - \frac{1}{e^x-1}\delta_1\right) \\ \Longrightarrow \varphi(x) &= 1 + \left|\frac{1}{e^x-1}\right| = \frac{e^x}{e^x-1} \\ \therefore \text{cond}_A(x) &\le \frac{\varphi(x)}{\text{cond}_f(x)} = \frac{e^x}{x} \to \infty \quad \text{when} \quad x \to 0. \end{aligned}$$

IV-c Use c++ to plot $\text{cond}_f(x)$ and $\text{cond}_A(x)$ as a function of x on $[0,1]$.

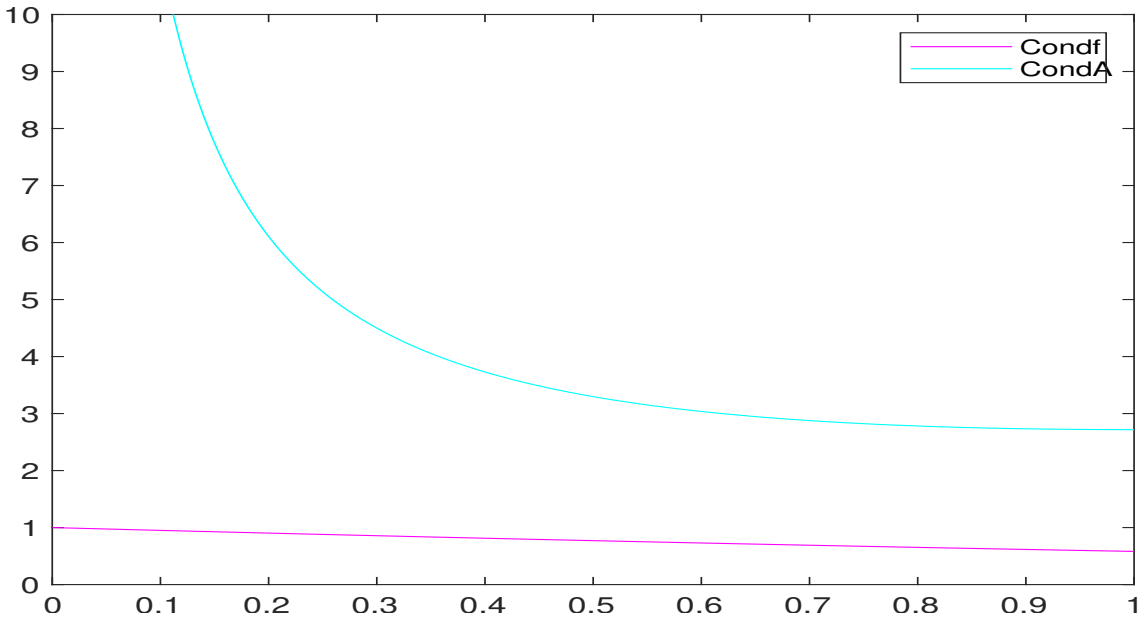

Figure 1: $\text{cond}_f(x)$ and $\text{cond}_A(x)$.

We find that $\text{cond}_A(x)$ is far larger than $\text{cond}_f(x)$ especially when x is close to zero. $\text{cond}_A(x) \to \infty$ when $x \to 0$, which satisfies the theorem that the subtraction in $1 - e^{-x}$ is not accurate , especially when $x \to 0$, $1 - e^{-x} \to 0$.

V. Derive the componentwise condition number of f based on the 1-norm. And compare the result with that in the Wilkinson Example.

$r = f(a_0, a_1, \ldots, a_{n-1})$, assume that $A = (a_0, a_1, \ldots, a_{n-1})$.

We have $B = (b_j)$, where $b_j = \left|\frac{a_j \frac{\partial f}{\partial a_j}}{f(A)}\right|$, $j = 0, 1 \ldots (n-1)$.

As f finds the root of polynomial $q(x) = 0$, the $\frac{\partial f}{\partial a_j}$ can be view as the change of the root when slightly change a_j. So we assume $g(x) = x^j$, the new function is $F = q + \epsilon g$ and the root of F is $r + h$, by corollary 1.43 $h \approx -\epsilon \frac{g(r)}{q'(r)}$.So we have

$$\frac{\partial f}{\partial a_j} = \lim_{\epsilon \to 0} \frac{r - r + h}{\epsilon} \approx -\frac{g(r)}{q'(r)} = -\frac{r^j}{\sum_{i=0}^{n-1}(i+1)\,a_{i+1}r^i}.$$

So the componentwise condition number of $f(x)$ is

$$\text{cond}_f(x) = \|\,B\,\|_1 = \sum_{j=0}^{n-1} \frac{|\,a_j\,|\,r^j}{|\,q'(r)\,|} = \sum_{j=0}^{n-1} \frac{|\,a_j\,|\,r^j}{\sum_{i=0}^{n-1}(i+1)\,a_{i+1}r^i}.$$

In the Wilkinson Example, $q(x) = \prod_{k=1}^{p}(x-k) = \sum_{i=0}^{p} a_i x^i$, so

$$\mathrm{cond}_f(x) = \frac{\sum_{j=0}^{p-1} \mid a_j \mid p^j}{(p-1)!}.$$

Because $x^p = -\sum_{i=0}^{p-1} a_i x^i$, so

$$\frac{p^p}{(p-1)!} = \frac{-\sum_{i=0}^{p-1} a_i p^i}{(p-1)!} \leq \mathrm{cond}_f(x).$$

Because in the Wilkinson Example, we calculate the change in the root for a small change in the highest item x^p. The change equals to that fix a_p and slightly change $a_i \quad (i = 0, 1 \ldots, p-1)$ uniformly. While the componentwise condition number of f means the change in the root for small changes in the coefficient of $q(x)$, where $a_i \quad (i = 0, 1, \ldots, p-1)$ can be changed seperately. So obviously the change in the latter one includes the change in the formal one. So the result in the Wilkinson Example in the notes is smaller than the condition number which we calculated in the problem.