36 + 3.6 = 39.6

I. Can we compute root with absolute accuracy $< 10^{-6}$ by bisection method

No, we can't. Because $128 = (10000000)_2 = (1.0000000)_2 \times 2^7$ takes up 8 digits within 24 digits. As a result, the minimum number can be denoted is $min = 2^{7-23} = 2^{-16} \approx 15.26 \times 10^{-6}$ so that $\epsilon_u = \frac{min}{2} = 7.63 \times 10^{-6} > 10^{-6}$. In other word, we can't compute roots with absolute accuracy $< 10^{-6}$.

5

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

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

$cond_f(x) = |\frac{\alpha x (x-1)^{\alpha-1}}{(x-1)^\alpha}| = |\frac{\alpha x}{x-1}|$, when $x \to 1$, $cond_f(x) \to \infty$

b. $\ln x$,

$cond_f(x) = |\frac{x \cdot \frac{1}{x}}{\ln x}| = |\frac{1}{\ln x}|$, when $x \to 1$, $cond_f(x) \to \infty$

8

c. e^x ,

$cond_f(x) = |\frac{xe^x}{e^x}| = |x|$, when $x \to \infty$, $cond_f(x) \to \infty$

d. $\arccos x$

$cond_f(x) = |\frac{x}{\sqrt{1-x^2}\arccos x}|$, when $x \to 1$ or $x \to -1$, $cond_f(x) \to \infty$

III. Repeat Example 1.25 for $f(x) = \frac{\sin x}{1+\cos x}$ on $(0, \frac{\pi}{2})$

We can know that

$$f'(x) = \frac{1}{1+\cos x} \tag{1}$$

$$cond_f(x) = |\frac{x}{\sin x}| \tag{2}$$

By theorem 1.29,

$$f_A(x) = \frac{(1+\delta_1)\sin x}{(1+\delta_3)(1+(1+\delta_2)\cos x)}(1+\delta_4) \tag{3}$$

$$= \frac{\sin x}{1+\cos x}(1+\delta_1+\delta_4-\delta_3-\frac{\delta_2\cos x}{1+(1+\delta_2)\cos x}) \tag{4}$$

5

where $|\delta_i| < \epsilon_u$, consequently, $\varphi(x) = 3 + \frac{\cos x}{1+cosx}$. Finally,

$$cond_A(x) \le \frac{\sin x}{x}(3+\frac{\cos x}{1+\cos x}) \tag{5}$$

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

a. Show that $cond_f(x) \le 1$ for $x \in [0,1]$

Because $cond_f(x) = |\frac{xe^{-x}}{1-e^{-x}}| = \frac{x}{e^x-1}$ on $(0,1]$ and $cond_f(0) = 1$ by L'Hospital theorem. Besides, we know that $e^x > 1+x$, therefore, $cond_f(x) < 1$ on $(0,1]$ and $cond_f(x) \le 1$ on $[0,1]$. Hence proved.

b. Estimate $cond_A(x)$ for $x \in [0,1]$

We can know that

$$f_A(x) = (1+\delta_2)(1-(1+\delta_1)e^{-x}) \tag{6}$$

$$= (1-e^{-x})(1+\delta_1+\delta_2+\delta_1\delta_2-\frac{\delta_1(1+\delta_2)}{1-e^{-x}}) \tag{7}$$

10

where $|\delta_i| < \epsilon_u$, consequently, $\varphi(x) = 2 + \frac{1}{1-e^{-x}}$. Therefore, for $x \in (0, 1]$

$$cond_A(x) \leq \frac{e^x - 1}{x}(2 + \frac{1}{1 - e^{-x}}) \tag{8}$$

不太对 $\frac{e^x}{x}$

c. Plot $cond_f(x)$ and $cond_A(x)$ as function of x on $[0, 1]$

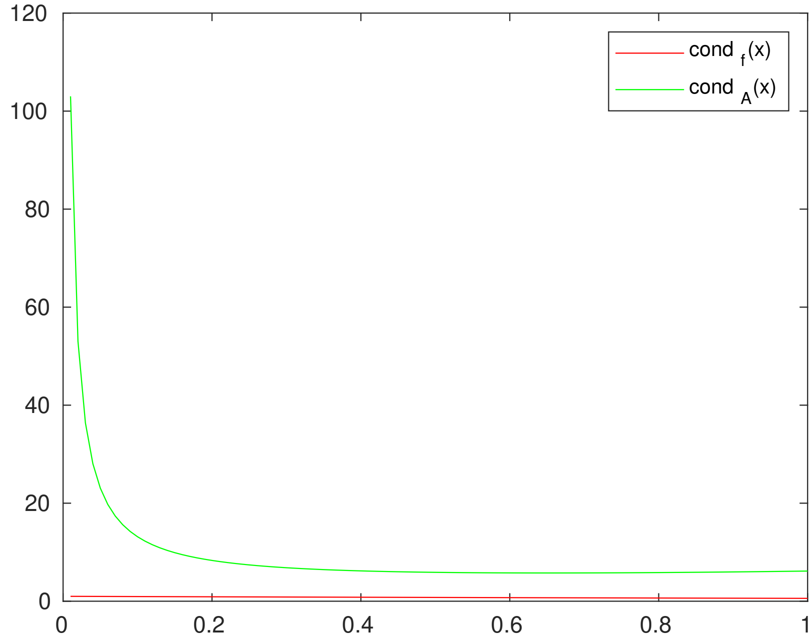

Figure 1: comparison between $cond_f(x)$ and $cond_A(x)$

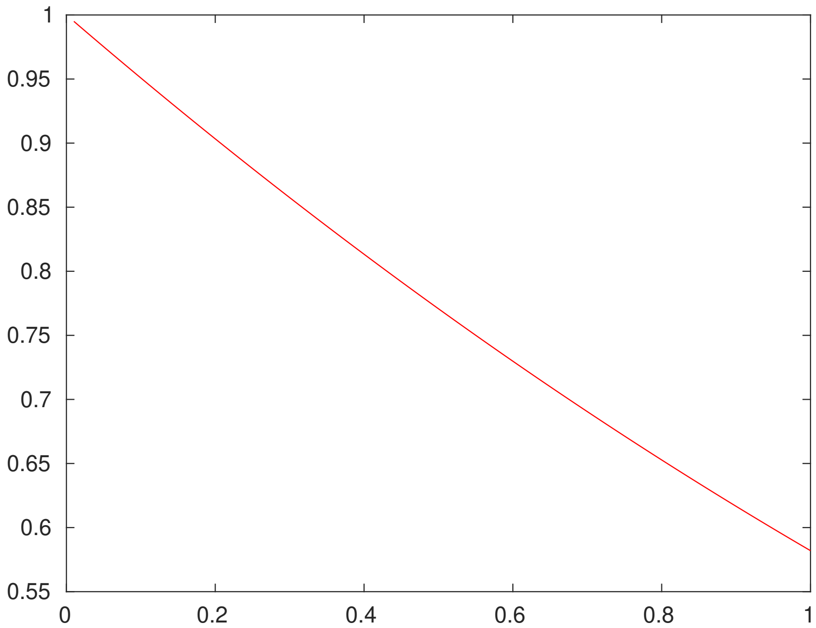

Figure 2: $cond_f(x)$

These two figures differ from each other totally. I think one of reasons is that $cond_A$ figure just shows a upper bound rather than explicit expression. Besides, relative error of $f_A \to \infty$ when $x \to 0^+$, so $cond_A \to \infty$ as result.

V. For Wilkinson example, compute condition number and compare result

Assume $q(r) = 0$, namely,

$$q(r) = a_0 + a_1 r + \cdots + a_{n-1} r^{n-1} + r^n = 0 \tag{9}$$

We can take equation as a function f of a_i, denoted by

$$r = f(a_0, a_1, \cdots, a_{n-1}) \tag{10}$$

Also, let $F(a_0, a_1, \cdots, a_{n-1}, r) = f(a_0, a_1, \cdots, a_{n-1}) - r$. Consequently, we can get

$$\frac{\partial r}{\partial a_i} = -\frac{r^i}{a_1 + 2a_2 r + \cdots + (n-1)a_{n-1} r^{n-2} + n r^{n-1}} \tag{11}$$

8

by

$$\frac{\partial F}{\partial r} = a_1 + 2a_2 r + \cdots + (n-1)a_{n-1} r^{n-2} + n r^{n-1} \tag{12}$$

$$\frac{\partial F}{\partial a_i} = x^i, \forall i = 0, 1, \cdots, n-1 \tag{13}$$

$$\frac{\partial r}{\partial a_i} = -\frac{\frac{\partial F}{\partial a_i}}{\frac{\partial F}{\partial r}} \tag{14}$$

According to definition 1.45, let $cond_f(\overrightarrow{x}) = ||A(\overrightarrow{x})||$, where $A = [a_{1i}(\overrightarrow{x})]$, $\overrightarrow{x} = (a_0, a_1, \cdots, a_{n-1})$. To be specific,

$$a_{1i}(\overrightarrow{x}) = |\frac{x_i \frac{\partial f}{\partial x_i}}{f(\overrightarrow{x})}| = |\frac{a_i}{r}\frac{\partial r}{\partial a_i}| = |\frac{a_i r^{i-1}}{a_1 + 2a_2 r + \cdots + (n-1)a_{n-1} r^{n-2} + n r^{n-1}}| \tag{15}$$

Therefore,

$$cond_f(\overrightarrow{x}) = ||A(\overrightarrow{x})|| = \max_{0 \le i \le n-1} |\frac{a_i r^{i-1}}{a_1 + 2a_2 r + \cdots + (n-1)a_{n-1} r^{n-2} + n r^{n-1}}| \tag{16}$$

$$= \max_{0 \le i \le n-1} \{|a_i r^{i-1}|\} |\frac{1}{a_1 + 2a_2 r + \cdots + (n-1)a_{n-1} r^{n-2} + n r^{n-1}}| \tag{17}$$

As for Wilkinson Example, namely , take $f(x) = \prod_{k=1}^{p}(x-k)$ into (17) and we can know that

$$cond_f(\overrightarrow{x}) = |\frac{p^{n-2}}{a_1 + 2a_2 p + \cdots + (n-1)a_{n-1} p^{n-2} + n p^{n-1}}| \tag{18}$$

The comparison reveals that conponentwise condition number of vector function is a better way to measure how is a function sensitive to variation than condition number.