1 $477 = 111011101$
$= (1.11011101)_2 \times 2^8$

1	477
0	238
1	119
1	59
1	29
0	14
1	7
1	3
1	1

2 $\frac{3}{5} = \frac{6}{10} = 0.6$
$0.6 = (0.100110011001\ldots)_2$
$= (1.00110011\ldots)_2 \times 2^{-1}$

	0.6
1	1.2
0	0.4
0	0.8
1	1.6
1	1.2
0	0.4
0	0.8
1	1.6

3 $x = \beta^e,\ e \in \mathbb{Z},\ L < e < U,\quad x_R - x = \beta(x - x_L)$

Let $x = (1.000\ldots00)_\beta \times \beta^e$
$x_L = [(\beta-1).(\beta-1)(\beta-1)\cdots(\beta-1)]_\beta \times \beta^{e-1}$
$x_R = (1.000\ldots01)_\beta \times \beta^e$
$x_R - x = (0.000\ldots01)_\beta \times \beta^e$
$= \beta^{e-p+1}$
$x - x_L = (0.000\ldots01)_\beta \times \beta^{e-1}$
$= \beta^{e-p}$
$\therefore x_R - x = \beta \times (x - x_L)$

4 $x = (1.00110011\ldots)_2 \times 2^{-1}$
$x_L = (1.00110011001100110011 00)_2 \times 2^{-1}$
$x_R = (1.00110011001100110011 01)_2 \times 2^{-1}$
$x - x_L = (1.1001100\ldots)_2 \times 2^{-23} = \frac{8}{5} \times 2^{-23}$
$x_R - x_L = 2^{-22}$
$x_R - x = (x_R - x_L) - (x - x_L)$
$= 2^{-22} - \frac{8}{5} \times 2^{-23}$
$= \frac{2}{5} \times 2^{-23}$
$\therefore x - x_L > x_R - x,\quad fl(x) = x_R$
Relative roundoff error is $\frac{|x_R - x|}{|x|} = \frac{2}{3} \times 2^{-23}$

5. from $\varepsilon_M = \beta^{1-p}$ and IEEE 754 single-precision protocols

$\beta = 2,\ p = 24,\ \varepsilon_M = \beta^{-23}$

$\therefore \varepsilon_u = \varepsilon_M = 2^{-23}$

6 from Theorem 4.49, $1 > \cos x,\ x = \frac{1}{4}$

$$\beta^{-t} \le 1 - \frac{\cos x}{1} \le \beta^{-s}$$

$$2^{-6} \le 0.0310875\ldots \le 2^{-5}$$

$\therefore$ When $\beta = 2$, $1-\cos x$ lost at most 6 and at least 5 significant digits.

7 Two ways compute $1-\cos x$ to avoid catastrophic cancellation

(1) Taylor's expansion:

$$1-\cos x = 1-\left(1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots\right)$$

$$= \frac{x^2}{2!}-\frac{x^4}{4!}+\frac{x^6}{6!}-\cdots$$

(2) Trigonometric identity

$$1-\cos x = 2\sin^2\frac{x}{2}$$

8 (1) $(x-1)^\alpha$

$\alpha \ne 0$, $C_f(x) = \left|\frac{x f'(x)}{f(x)}\right|$

$$= \left|\frac{x\alpha(x-1)^{\alpha-1}}{(x-1)^\alpha}\right|$$

$$= \left|\frac{x\alpha}{x-1}\right|$$

$\therefore$ When $x \to 1$, $C_f(x) \to +\infty$

$\alpha = 0$, $C_f(x) = 0$, never be large

(2) $\ln x$

$C_f(x) = \left|\frac{1}{\ln x}\right|$

$\therefore x \to 0^+$, $C_f(x) \to +\infty$

(3) e^x

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

$\therefore x \to +\infty$, $C_f(x) \to +\infty$

(4) $\arccos x$

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

$\therefore x = \pm 1,\ C_f(x) \to +\infty$

9(1) $f(x) = 1 - e^{-x},\quad x \in [0,1]$

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

$$= \left| \frac{x}{e^{x}-1} \right|$$

$\lim_{x \to 0} C_f(x) = 1$, $C_f(x)$ decrease in $x \in [0,1]$

$\therefore C_f(x) \leq 1$

(2) $|f(x_A) - f(x)| = |f'(\xi)|\,|x - x_A| < \varepsilon_u$, ξ between x_A and x

$$|x - x_A| < \frac{\varepsilon_u}{|f'(\xi)|} = \frac{\varepsilon_u}{e^{-\xi}} \leq e\,\varepsilon_u$$

$$\mathrm{Cond}_A(x) = \frac{1}{\varepsilon_u} \min_{\{x_A\}} \frac{|x - x_A|}{x}$$

$$< \frac{e}{x}$$

(3)

10) $q(x) = \sum_{i=0}^{n} a_i x^i \quad a_n = 1,\ a_0 \neq 0,\ a_i \in \mathbb{R}$

$$q(r) = \sum_{i=0}^{n} a_i r^i$$

$$r^j + \sum_{i=0}^{n} i a_i r^{i-1} \frac{\partial r}{\partial a_j} = 0, \quad j = 1, 2, \dots, n-1$$

$$\nabla r = -\frac{(1, r, \dots, r^{n-1})}{q'(r)}$$

$$\mathrm{cond}_f(a) = \frac{\|a\|_1 \|\nabla r\|_1}{|r|}$$

$$= \frac{\sum_{i=0}^{n-1} |a_i| \sum_{i=0}^{n-1} |r|^i}{\left|\sum_{i=0}^{n} i a_i r^i\right|}$$

In Wilkinson example, $q(x) = \prod_{i=1}^{n} (x - i)$

$$\sum_{i=0}^{n-1} |a_n r^i| \geq -\sum_{i=0}^{n-1} a_i r^i = r^n = n^n$$

$$\left|\sum_{i=0}^{n} i a_i r^i\right| = |r|\,|q'(r)|$$

$$= n\,|q'(n)|$$

$$< n^2\, n!$$

Date:

No.:

11 $\beta = 2,\ p = 2,\ L = -1,\ U = 1$

$a = (1.0)_2 \times 2^0$, $b = (1.1)_2 \times 2^0$

$\frac{a}{b} = (0.101010\cdots)_2$

$fl(\frac{a}{b}) = fl[(0.101)_2]$

$= (0.10)_2 \times 2^0$

$= 0.5$

Relative error $: \left|\frac{0.5}{\frac{2}{3}} - 1\right| = 0.25$

Machine precision: $\varepsilon_M = 2^{-2} = 0.25$

12 $\varepsilon_M = 2^{-23}$

$2^7 \varepsilon_M = 2^{-16} \approx 1.5259 \times 10^{-5} > 2 \times 10^{-6}$

$\therefore$ We can't compute the root with absolute accuracy $< 10^{-6}$

13 calculate $s(x) = ax^3 + bx^2 + cx + d$ on $[x_i, x_{i+1}]$ by $s(x_i)$, $s(x_{i+1})$, $s'(x_i)$, $s'(x_{i+1})$.

Coefficient matrix : $\begin{pmatrix} x_i^3 & x_i^2 & x_i & 1 \\ x_{i+1}^3 & x_{i+1}^2 & x_{i+1} & 1 \\ 3x_i^2 & 2x_i & 1 & 0 \\ 3x_{i+1}^2 & 2x_{i+1} & 1 & 0 \end{pmatrix}$

It has large condition number when x_i and x_{i+1} are close enough, so the result will be very inaccurate.