CS206p Assignment 2
Jingcheng Lu (42326170)

---

1. Question 1

   (a) 1.A

$$\begin{aligned}
\varepsilon_r(f(x_0)) &= \frac{|f(x_0) - f(t)|}{f(t)} \\
&= \frac{|f(t \times (1+\delta_0)) - t^2|}{t^2} \\
&= \frac{\left|t^2 \cdot (1 + 2\cdot\delta_0 + {\delta_0}^2) - t^2\right|}{t^2} \\
&= 2\cdot\delta_0 + {\delta_0}^2 \\
&\approx 2\cdot\delta_0
\end{aligned} \tag{1}$$

   (b) 1.B

$$\begin{aligned}
x_k &= t^{2^k} \cdot (1+\delta_k) \\
&= (1+\delta_0)\cdot f(x_{k-1}) \\
&= (1+\delta_0)\cdot {x_{k-1}}^2 \\
&= (1+\delta_0)\cdot t^{2^k} \cdot (1+\delta_{k-1})^2
\end{aligned} \tag{2}$$

Because of (2), we can get:

$$\begin{aligned}
&(1+\delta_k) = (1+\delta_0)\cdot(1+\delta_{k-1})^2 \\
&\rightarrow 2\cdot\delta_{k-1} + \delta_0 + \varepsilon = \delta_k \\
&\rightarrow 2\cdot\delta_{k-1} + E = \delta_k + E - \delta_0 - \varepsilon \\
&\rightarrow |\delta_k| \le |2\cdot\delta_{k-1}| + E
\end{aligned} \tag{3}$$

Based on (3), for n=1,

$$|\delta_1| \le 2\cdot|\delta_0| + E$$

. If for n=k-1:

$$|\delta_{k-1}| \le 2^{k-1}\cdot|\delta_0| + (2^{k-1}-1)\cdot E$$

So for n=k, we have:

$$
\begin{aligned}
&|\delta_k| \leq 2 \cdot |\delta_{k-1}| + E \\
&= 2 \cdot (2^{k-1} \cdot |\delta_0| + (2^{k-1} - 1) \cdot E) + E \\
&= 2^k \cdot |\delta_0| + (2^k - 1) \cdot E
\end{aligned}
\tag{4}
$$

2. Question 2

(a) 2.A

$$
\begin{aligned}
\varepsilon_r(f(x_0)) &= \frac{|f(x_0) - f(t)|}{f(t)} \\
&\approx \frac{|f'(t)|}{f(t)} \cdot (x_0 - t) \\
&= \frac{\frac{1}{2} \cdot t^{\frac{1}{2}-1} \cdot t \cdot \delta_0}{t^{\frac{1}{2}}} \\
&= \frac{1}{2} \cdot \delta_0
\end{aligned} \tag{5}
$$

(b) 2.B

Because of:

$$
\begin{aligned}
X_k &= \sqrt{t^{2^{1-k}} \cdot (1 + \delta_{k-1})} \\
&= t^{2^{-k}} \cdot \sqrt{(1 + \delta_{k-1})} \\
&\approx t^{2^{-k}} \cdot \sqrt{(1 + \delta_{k-1} + \frac{1}{4}{\delta_{k-1}}^2)} \\
&= t^{2^{-k}} \cdot (1 + \frac{1}{2}\delta_{k-1}) \\
&= t^{2^{-k}} \cdot (1 + \delta_k)
\end{aligned} \tag{6}
$$

We can get:

$$
|\delta_k| \leq \left| \frac{1}{2}\delta_{k-1} \right| + E \tag{7}
$$

Based on (8), for n=1,

$$
|\delta_1| \leq \frac{1}{2^1} \cdot |\delta_0| + E
$$

If for n=k-1:

$$
|\delta_{k-1}| \leq \frac{1}{2^{k-1}} \cdot |\delta_0| + (2 - \frac{1}{2^{k-2}}) \cdot E
$$

So for n=k, we have:

$$
\begin{aligned}
&|\delta_k| \leq \frac{1}{2} \cdot |\delta_{k-1}| + E \\
&\leq \frac{1}{2} \cdot (\frac{1}{2^{k-1}} \cdot |\delta_0| + (2 - \frac{1}{2^{k-2}}) \cdot E) + E \\
&= \frac{1}{2^k} \cdot |\delta_0| + (1 - \frac{1}{2^{k-1}}) \cdot E + E
\end{aligned} \tag{8}
$$

3. Question 3

(a) 3.A

The function exp1 is for the experiment 1, the other exp2 is for the experiment 2.

```
import numpy as np
import math

alpha = np.float64(2.37e-7);

def computeX(x0,n):
    xn = x0;
    for n in range(0, n):
        xn = math.pow(xn,2);
    return xn;

def computeY(y0,n):
    yn = y0;
    for n in range(0, n):
        yn = math.sqrt(yn);
    return yn;

def exp1():
    x0 = 1 + alpha;
    t = 1 + alpha;
    for n in range(1, 31):
        xn = computeX(x0,n);
        errx = xn - t;
        y0 = xn;
        yn = computeY(y0,n);
        erry = yn - t;
        print("n: ",n,"; errx: ",errx,"; erry: ",erry);
    return 0;

def exp2():
    y0 = 1 + alpha;
    t = 1 + alpha;
    for n in range(1, 31):
        yn = computeY(y0, n);
        erry = yn - t;
        x0 = yn;
        xn = computeX(x0,n);
        errx = xn - t;
        print("n: ",n,"; errx: ",errx,"; erry: ",erry);
    return 0;
```

The result in experiment1 is
('n: ', 1, '; errx: ', 2.3700005624682774e-07, '; erry: ', 0.0)
('n: ', 2, '; errx: ', 7.1100033727233836e-07, '; erry: ', 0.0)
('n: ', 3, '; errx: ', 1.6590015732287355e-06, '; erry: ', 0.0)
('n: ', 4, '; errx: ', 3.5550067414291675e-06, '; erry: ', 0.0)
('n: ', 5, '; errx: ', 7.347027862314448e-06, '; erry: ', 0.0)
('n: ', 6, '; errx: ', 1.493111324224472e-05, '; erry: ', 0.0)
('n: ', 7, '; errx: ', 3.0099456556298421e-05, '; erry: ', 0.0)
('n: ', 8, '; errx: ', 6.0436833413168856e-05, '; erry: ', 0.0)
('n: ', 9, '; errx: ', 0.00012111434814054967, '; erry: ', 0.0)
('n: ', 10, '; errx: ', 0.00024248042243080192, '; erry: ', 0.0)
('n: ', 11, '; errx: ', 0.00048525675660893164, '; erry: ', 0.0)
('n: ', 12, '; errx: ', 0.00097098621740565605, '; erry: ', 0.0)
('n: ', 13, '; errx: ', 0.0019431527093494161, '; erry: ', 0.0)
('n: ', 14, '; errx: ', 0.0038903191822612371, '; erry: ', 0.0)
('n: ', 15, '; errx: ', 0.0077960117919297911, '; erry: ', 0.0)
('n: ', 16, '; errx: ', 0.015653042079085244, '; erry: ', 0.0)
('n: ', 17, '; errx: ', 0.031551346304098393, '; erry: ', 0.0)
('n: ', 18, '; errx: ', 0.064098432017192231, '; erry: ', 0.0)
('n: ', 19, '; errx: ', 0.1323057404041601, '; erry: ', 0.0)
('n: ', 20, '; errx: ', 0.28211658946519047, '; erry: ', 0.0)
('n: ', 21, '; errx: ', 0.64382331970517148, '; erry: ', 0.0)
('n: ', 22, '; errx: ', 1.7021556485788403, '; erry: ', 0.0)
('n: ', 23, '; errx: ', 6.3016461929683665, '; erry: ', 0.0)
('n: ', 24, '; errx: ', 52.314040351269789, '; erry: ', 0.0)
('n: ', 25, '; errx: ', 2841.3869236106784, '; erry: ', 0.0)
('n: ', 26, '; errx: ', 8079162.4248600332, '; erry: ', 0.0)
('n: ', 27, '; errx: ', 65272881645598.922, '; erry: ', 0.0)
('n: ', 28, '; errx: ', 4.2605490783204951e+27, '; erry: ', 0.0)
('n: ', 29, '; errx: ', 1.815227844877762e+55, '; erry: ', 0.0)
('n: ', 30, '; errx: ', 3.2950521288195639e+110, '; erry: ', 0.0)
The result in experiment2 is
('n: ', 1, '; errx: ', -2.2204460492503131e-16, '; erry: ', -1.185000071401987e-07)
('n: ', 2, '; errx: ', -2.2204460492503131e-16, '; erry: ', -1.7775000538122754e-07)
('n: ', 3, '; errx: ', -2.2204460492503131e-16, '; erry: ', -2.0737500316947433e-07)
('n: ', 4, '; errx: ', -2.2204460492503131e-16, '; erry: ', -2.2218750173053081e-07)
('n: ', 5, '; errx: ', -2.2204460492503131e-16, '; erry: ', -2.2959375090003675e-07)
('n: ', 6, '; errx: ', -7.3274719625260332e-15, '; erry: ', -2.3329687559581203e-07)

('n: ', 7, '; errx: ', -7.3274719625260332e-15, '; erry: ', -2.3514843783267736e-07)
('n: ', 8, '; errx: ', -3.5749181392930041e-14, '; erry: ', -2.3607421906213233e-07)
('n: ', 9, '; errx: ', -9.2592600253738055e-14, '; erry: ', -2.3653710967685981e-07)
('n: ', 10, '; errx: ', -2.0627943797535409e-13, '; erry: ', -2.3676855498422356e-07)
('n: ', 11, '; errx: ', -2.0627943797535409e-13, '; erry: ', -2.3688427752688312e-07)
('n: ', 12, '; errx: ', -2.0627943797535409e-13, '; erry: ', -2.3694213879821291e-07)
('n: ', 13, '; errx: ', -2.0627943797535409e-13, '; erry: ', -2.369710694338778e-07)
('n: ', 14, '; errx: ', -2.0627943797535409e-13, '; erry: ', -2.3698553475171025e-07)
('n: ', 15, '; errx: ', -2.0627943797535409e-13, '; erry: ', -2.3699276741062647e-07)
('n: ', 16, '; errx: ', -7.48223705215878e-12, '; erry: ', -2.3699638385110688e-07)
('n: ', 17, '; errx: ', -7.48223705215878e-12, '; erry: ', -2.3699819196032479e-07)
('n: ', 18, '; errx: ', -3.6586067508892484e-11, '; erry: ', -2.3699909612595604e-07)
('n: ', 19, '; errx: ', -9.4793728422359891e-11, '; erry: ', -2.3699954820877167e-07)
('n: ', 20, '; errx: ', -2.1120905024929471e-10, '; erry: ', -2.3699977425017948e-07)
('n: ', 21, '; errx: ', -4.4403969390316433e-10, '; erry: ', -2.3699988727088339e-07)
('n: ', 22, '; errx: ', -4.4403969390316433e-10, '; erry: ', -2.3699994367021304e-07)
('n: ', 23, '; errx: ', -4.4403969390316433e-10, '; erry: ', -2.3699997186987787e-07)
('n: ', 24, '; errx: ', -2.3066850651787263e-09, '; erry: ', -2.3699998608073258e-07)
('n: ', 25, '; errx: ', -6.0319762518190601e-09, '; erry: ', -2.3699999318615994e-07)
('n: ', 26, '; errx: ', -1.3482558181010518e-08, '; erry: ', -2.3699999673887362e-07)
('n: ', 27, '; errx: ', -2.8383722705527248e-08, '; erry: ', -2.3699999851523046e-07)
('n: ', 28, '; errx: ', -5.8186050200248474e-08, '; erry: ', -2.3699999940340888e-07)
('n: ', 29, '; errx: ', -1.1779070430151251e-07, '; erry: ', -2.3699999984749809e-07)
('n: ', 30, '; errx: ', -2.3700000006954269e-07, '; erry: ', -2.3700000006954269e-07)

(b) 3.B

Based on the result of question 1, we can get:

$$\begin{aligned} x_{n-1} = {}^1\delta_{n-1} \leq 2^{k-1}\left|\delta_0\right| + (2^{k-1}-1)\cdot E \\ y_{n-1} = {}^2\delta_{n-1} \leq \tfrac{1}{2^{k-1}}\left|\delta'_0\right| + (2-\tfrac{1}{2^{k-2}})\cdot E \end{aligned}$$

In experiment 1,

$$\delta'_0 = {}^1\delta_{n-1}$$

so

$$\begin{aligned} & y_{n-1} = {}^2\delta_{n-1} \leq \tfrac{1}{2^{k-1}}\left|\delta'_0\right| + (2-\tfrac{1}{2^{k-2}})\cdot E \\ & \leq \tfrac{1}{2^{k-1}}(2^{k-1}\left|\delta_0\right| + (2^{k-1}-1)\cdot E) + (2-\tfrac{1}{2^{k-2}})\cdot E \\ & = \left|\delta_0\right| + 3\cdot(1-\tfrac{1}{2^{k-1}})\cdot E \end{aligned}$$

In experiment 2, with same logic we can get

$$|x_{n-1}| \leq |\delta_0| + 3 \cdot (2^{k-1} - 1) \cdot E$$