CSC336: Assignment 1

Junjie Cheng, 1002770539

October 12 2017

1. (a) Absolute error$= A - T = 2.72 - 2.71828182845905 = 1.72 * 10^{-3}$
   Relative error$= \frac{A-T}{T} = 6.32 * 10^{-4}$
   (b) Absolute error:$= A - T = 2.718 - 2.71828182845905 = -2.82 * 10^{-4}$
   Relative error:$\frac{A-T}{T} = -1.04 * 10^{-4}$
   (c) Absolute error:$= A - T = 2.71828183 - 2.71828182845905 = 1.54 * 10^{-9}$
   Relative error:$= \frac{A-T}{T} = 5.67 * 10^{-10}$
2. (a) $4.21 * 10^0 + 5.47 * 10^{-2} = 4.2647 * 10^0$. It will be rounded to $4.26 * 10^0$
   (b) $6.52 * 10^1 - 7.27 * 10^10^{-1} = 6.4473 * 10^1$. It will be rounded to $6.45 * 10^1$
   (c) $5.61 * 10^1 + 6.67 * 10^{-4} = 5.6100667 * 10^1$. It will be rounded to $5.61 * 10^1$
   (d) $4.52 * 10^4 - 3.82 * 10^6 = -3.7748 * 10^6$. It will be rounded to $-3.77 * 10^6$
   (e) $7.51 * 10^12 - 5.25 * 10^5$ will be rounded to $7.51 * 10^{12}$
   (f) $3.82 * 10^1 + 8.42 * 10^2 = 8.803 * 10^2$. It will be rounded to $8.80 * 10^2$
   (g) $4.47 * 10^{10} * 5.81 * 10^{15} = 2.59707 * 10^{-4}$. It will be rounded to $2.60 * 10^{-4}$
   (h) $2.41 * 10^{10} * 4.81 * 10^{12} = 1.15921 * 10^{-21} = 0.115921 * 10^{-20}$. It will be rounded to $0.12 * 10^{-20}$, a subnormal floating point.
   (i) $6.37 * 10^{10} * 5.28 * 10 * 15 = -3.36336 * 10^{-24}$. It will be rounded to 0 because of underflow.
   (j) $6.27 * 10^{10}/(2.72 * 10^{15}) = -2.305 * 10^{25}$. It will be rounded to -Inf because of overflow.
3. (a) Let Δx denote the change of input x (i.e. $\Delta x = \hat{x} - x$).
   For very small Δx, we can write $f(\hat{x}) - f(x) = \Delta x f'(x)$
   $RelativeForwardError$
   $= \frac{f(\hat{x})-f(x)}{f(x)}$
   $= \frac{f(x+\Delta h)-f(x)}{f(x)} = \frac{f'(x)(\hat{x}-x)}{f(x)} = \frac{xf'(\hat{x})}{f(x)} \times \frac{\hat{x}-x}{x} = \frac{1}{\log_e(x)} \times \frac{\hat{x}-x}{x}$
   When x is close to 1, the condition number $|\frac{1}{\log_e(x)}|$ approaches Inf. The function is ill-conditioned in a relative sense with respect to small relative changes in the value of the input argument x for x close to 1.
   When x is close to 10, the condition number $|\frac{1}{\log_e(x)}| \approx \frac{1}{\log_e 10} \approx 0.4343$. The function is well-conditioned in a relative sense with respect to small relative changes in the value of the input argument x for x close to 10.
   (b) See attachments for code and result.
   In part (a), we calculated that the condition number can (almost) be expressed as $|\frac{1}{\log_e(x)}|$. In the computational result of part (b), the condition numbers are very close for the same

x. Also, the condition numbers are very large when $x = 1$ and small when $x = 10$, this also agrees with the calculation we have done in part (a).

4. (a) $RelativeError$

$= \frac{(\frac{1}{1-x}-\frac{1}{1+x})-(\frac{1}{(1-x)(1+\sigma_1)}(1+\sigma_2)-\frac{1}{(1+x)(1+\sigma_3)}(1+\sigma_4))(1+\sigma_5)}{\frac{1}{1-x}-\frac{1}{1+x}}$

$= \frac{\frac{2x(1+\sigma_1)(1+\sigma_3)+(1-x)(1+\sigma_1)(1+\sigma_4)(1+\sigma_5)-(1+x)(1+\sigma_3)(1+\sigma_2)(1+\sigma_5)}{(1-x^2)(1+\sigma_1)(1+\sigma_3)}}{\frac{2x}{1-x^2}}$

$= 1 + \frac{(1+\sigma_2)(1+\sigma_3)(1+\sigma_5)-(1+\sigma_1)(1+\sigma_4)(1+\sigma_5)}{2} - \frac{(1+\sigma_2)(1+\sigma_3)(1+\sigma_5)+(1+\sigma_1)(1+\sigma_4)(1+\sigma_5)}{2x(1+\sigma_1)(1+\sigma_3)}$

Note that the last part of the relative error is inverse proportional to x. That is, when x is extremely small, the absolute value of relative error can be very large.

(b) We choose $\frac{2x}{(1-x)(1+x)} = \frac{1}{1-x} - \frac{1}{1+x}$.

$RelativeError$

$= \frac{\frac{2x}{(1-x)(1+x)} - \frac{(2x)(1+\sigma_1)}{((1-x)(1+\sigma_2)(1+x)(1+\sigma_3))(1+\sigma_5)}(1+\sigma_4)}{\frac{2x}{(1-x)(1+x)}}$

$= 1 - \frac{\frac{(2x)(1+\sigma_1)}{((1-x)(1+\sigma_2)(1+x)(1+\sigma_3))(1+\sigma_5)}(1+\sigma_4)}{\frac{2x}{(1-x)(1+x)}}$

$= 1 - \frac{(1+\sigma_1)(1+\sigma_4)}{(1+\sigma_2)(1+\sigma_3)(1+\sigma_5)}$

$= \frac{\sigma_2+\sigma_3+\sigma_5+\sigma_2\sigma_3+\sigma_2\sigma_5+\sigma_3\sigma_5+\sigma_2\sigma_3\sigma_5-\sigma_1-\sigma_4-\sigma_1\sigma_4}{1+\sigma_2+\sigma_3+\sigma_5+\sigma_2\sigma_3+\sigma_2\sigma_5+\sigma_3\sigma_5+\sigma_2\sigma_3\sigma_5}$

Note that $|\sigma_i| < \epsilon_{machine}$, so in the worst case, we have

$$|RelativeError| = \frac{5\epsilon_{machine} + 4\epsilon_{machine}^2 + \epsilon_{machine}^3}{1 - 3\epsilon_{machine} - 3\epsilon_{machine}^2 - 3\epsilon_{machine}^3}$$

Since $\epsilon_{machine} << 1$, $|RelativeError| < 6\epsilon_{machine}$ in the worst case. So, we can conclude that $\frac{2x}{(1-x)(1+x)}$ has a very small relative error for all values of x, provided that there is no overflow or underflow.

5. (a) See attachments for code and output.

(b) For $x > -18$, the function produce very accurate approximations, (when $x = -18$ the approximation is also kind of accurate as the relative error is around only 0.05).

For $x < -19$, the relative error gets much larger. Of course rounding error contributes at lot on the poor approximations, the main reason that the function performs well on $x > 19$ but badly on $x < -19$ is that, $exp(x)$ gets too small for $x < -19$. When denominator gets too small, the whole fraction denoting the relative number gets large.

On the other hand, when $|x|$ gets larger, it will require a higher i for $\frac{x^i}{i!}$ to be insignificant so that the accumulated sum stops changing. Much more operations are needed to compute each $\frac{x^i}{i!}$ with a very high i, resulting a higher error for $\frac{x^i}{i!}$. This can make the relative error grows even faster for more negative x. However, the denominator $exp(x)$ gets too small is the main reason for the poor approximation for $x < -19$.

(c) See attachments for code and output.

Junjie Cheng
1002770539
chenjunj

Question 3
b)

MatLab Program:
```
function Question3b(~)
    function printConditionNumber(x, diff_x)
        fprintf('Choose x = %f, f(%f) = %f\n', x, x, log(x))
        fprintf('f(%f) = %f, conditionNumber = %f\n', x-diff_x, log(x-
diff_x), cNum(x,x-diff_x))
        fprintf('f(%f) = %f, conditionNumber = %f\n', x+diff_x,
log(x+diff_x), cNum(x,x+diff_x))
    end

    function conditionNumber = cNum(x, new_x)
        relY = log(new_x) / log(x) - 1;
        relX = new_x / x - 1;
        conditionNumber = abs(relY/relX);
    end

    printConditionNumber(0.99990, 0.000005)
    printConditionNumber(1.00010, 0.000005)
    printConditionNumber(9.99990, 0.000005)
    printConditionNumber(10.00010, 0.000005)

end
```

Output:
```
>> Question3b
Choose x = 0.999900, f(0.999900) = -0.000100
f(0.999895) = -0.000105, conditionNumber = 9999.524993
f(0.999905) = -0.000095, conditionNumber = 9999.474991
Choose x = 1.000100, f(1.000100) = 0.000100
f(1.000095) = 0.000095, conditionNumber = 10000.524990
f(1.000105) = 0.000105, conditionNumber = 10000.474993
Choose x = 9.999900, f(9.999900) = 2.302575
f(9.999895) = 2.302575, conditionNumber = 0.434296
f(9.999905) = 2.302576, conditionNumber = 0.434296
Choose x = 10.000100, f(10.000100) = 2.302595
f(10.000095) = 2.302595, conditionNumber = 0.434293
f(10.000105) = 2.302596, conditionNumber = 0.434292
```

Question 5
a)

MatLab Program:
```
function result = exp1(x)
    exp_result = 0;
    new_result = 1;
    i = 1;
    while exp_result ~= new_result;
        exp_result = new_result;
        new_result = new_result + (power(x, i)/factorial(i));
        i = i + 1;
    end
    result = exp_result;
```

```
end

function Question5a(~)
         x_array = -25:1:25;
    for j = x_array(1:end);
        fprintf('exp(%d)=%f, relative error = %f\n', j, exp1(j), rErr(j))
    end

    function relativeError = rErr(x)
        relativeError = (exp1(x) - exp(x))/exp(x);
    end
end
```

Output:

```
>> Question5a
exp(-25)=0.000001, relative error = 58226.187024
exp(-24)=0.000000, relative error = 9966.350698
exp(-23)=0.000000, relative error = 66.228996
exp(-22)=-0.000000, relative error = -115.073655
exp(-21)=0.000000, relative error = 35.386515
exp(-20)=0.000000, relative error = 1.024904
exp(-19)=0.000000, relative error = -0.544198
exp(-18)=0.000000, relative error = 0.049480
exp(-17)=0.000000, relative error = 0.001020
exp(-16)=0.000000, relative error = 0.000285
exp(-15)=0.000000, relative error = 0.000010
exp(-14)=0.000001, relative error = -0.000009
exp(-13)=0.000002, relative error = -0.000001
exp(-12)=0.000006, relative error = 0.000000
exp(-11)=0.000017, relative error = 0.000000
exp(-10)=0.000045, relative error = -0.000000
exp(-9)=0.000123, relative error = -0.000000
exp(-8)=0.000335, relative error = -0.000000
exp(-7)=0.000912, relative error = 0.000000
exp(-6)=0.002479, relative error = -0.000000
exp(-5)=0.006738, relative error = 0.000000
exp(-4)=0.018316, relative error = 0.000000
exp(-3)=0.049787, relative error = 0.000000
exp(-2)=0.135335, relative error = 0.000000
exp(-1)=0.367879, relative error = 0.000000
exp(0)=1.000000, relative error = 0.000000
exp(1)=2.718282, relative error = 0.000000
exp(2)=7.389056, relative error = -0.000000
exp(3)=20.085537, relative error = -0.000000
exp(4)=54.598150, relative error = 0.000000
exp(5)=148.413159, relative error = -0.000000
exp(6)=403.428793, relative error = 0.000000
exp(7)=1096.633158, relative error = -0.000000
exp(8)=2980.957987, relative error = -0.000000
exp(9)=8103.083928, relative error = 0.000000
exp(10)=22026.465795, relative error = -0.000000
exp(11)=59874.141715, relative error = 0.000000
exp(12)=162754.791419, relative error = -0.000000
exp(13)=442413.392009, relative error = -0.000000
exp(14)=1202604.284165, relative error = 0.000000
exp(15)=3269017.372472, relative error = 0.000000
exp(16)=8886110.520508, relative error = 0.000000
exp(17)=24154952.753575, relative error = 0.000000
exp(18)=65659969.137331, relative error = 0.000000
exp(19)=178482300.963187, relative error = -0.000000
```

```
exp(20)=485165195.409790, relative error = -0.000000
exp(21)=1318815734.483214, relative error = -0.000000
exp(22)=3584912846.131593, relative error = 0.000000
exp(23)=9744803446.248905, relative error = 0.000000
exp(24)=26489122129.843472, relative error = 0.000000
exp(25)=72004899337.385880, relative error = 0.000000
```

c)

MatLab Program:

```
function result = exp2(x)
    if x >= 0;
        result = exp1(x);
    else;
        result = 1/exp(-x);
end

function Question5c(~)
    x_array = -25:1:25;
    for j = x_array(1:end);
        fprintf('exp(%d)=%f, relative error = %f\n', j, exp2(j), rErr(j))
    end

    function relativeError = rErr(x)
        relativeError = (exp2(x) - exp(x))/exp(x);
    end
end
```

Output:

```
>> Question5c
exp(-25)=0.000000, relative error = -0.000000
exp(-24)=0.000000, relative error = 0.000000
exp(-23)=0.000000, relative error = 0.000000
exp(-22)=0.000000, relative error = 0.000000
exp(-21)=0.000000, relative error = 0.000000
exp(-20)=0.000000, relative error = 0.000000
exp(-19)=0.000000, relative error = -0.000000
exp(-18)=0.000000, relative error = 0.000000
exp(-17)=0.000000, relative error = 0.000000
exp(-16)=0.000000, relative error = 0.000000
exp(-15)=0.000000, relative error = 0.000000
exp(-14)=0.000001, relative error = 0.000000
exp(-13)=0.000002, relative error = 0.000000
exp(-12)=0.000006, relative error = 0.000000
exp(-11)=0.000017, relative error = 0.000000
exp(-10)=0.000045, relative error = 0.000000
exp(-9)=0.000123, relative error = -0.000000
exp(-8)=0.000335, relative error = 0.000000
exp(-7)=0.000912, relative error = 0.000000
exp(-6)=0.002479, relative error = 0.000000
exp(-5)=0.006738, relative error = 0.000000
exp(-4)=0.018316, relative error = 0.000000
exp(-3)=0.049787, relative error = 0.000000
exp(-2)=0.135335, relative error = 0.000000
exp(-1)=0.367879, relative error = -0.000000
exp(0)=1.000000, relative error = 0.000000
exp(1)=2.718282, relative error = 0.000000
exp(2)=7.389056, relative error = -0.000000
exp(3)=20.085537, relative error = -0.000000
```

```
exp(4)=54.598150, relative error = 0.000000
exp(5)=148.413159, relative error = -0.000000
exp(6)=403.428793, relative error = 0.000000
exp(7)=1096.633158, relative error = -0.000000
exp(8)=2980.957987, relative error = -0.000000
exp(9)=8103.083928, relative error = 0.000000
exp(10)=22026.465795, relative error = -0.000000
exp(11)=59874.141715, relative error = 0.000000
exp(12)=162754.791419, relative error = -0.000000
exp(13)=442413.392009, relative error = -0.000000
exp(14)=1202604.284165, relative error = 0.000000
exp(15)=3269017.372472, relative error = 0.000000
exp(16)=8886110.520508, relative error = 0.000000
exp(17)=24154952.753575, relative error = 0.000000
exp(18)=65659969.137331, relative error = 0.000000
exp(19)=178482300.963187, relative error = -0.000000
exp(20)=485165195.409790, relative error = -0.000000
exp(21)=1318815734.483214, relative error = -0.000000
exp(22)=3584912846.131593, relative error = 0.000000
exp(23)=9744803446.248905, relative error = 0.000000
exp(24)=26489122129.843472, relative error = 0.000000
exp(25)=72004899337.385880, relative error = 0.000000
```