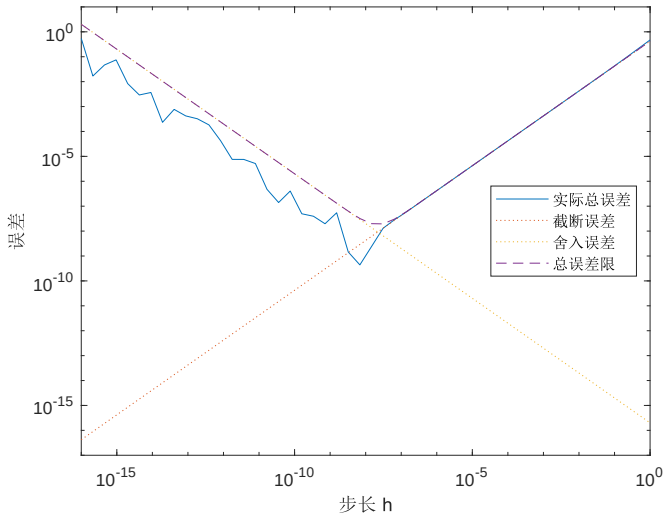
数值计算导论-上机作业

张晨，2017011307

1.1

【复现结果】



【关键代码】

```
h = logspace(-16, 0);
err = abs(cos(1) - (sin(1+h) - sin(1)) ./ h);
trunc = sin(1)*h/2;
round = 2 * 1e-16 ./ h;
```

【分析】使用差商$f'(x) \approx \frac{f(x+h)-f(x)}{h}$近似一阶导数，主要引起两类误差。由泰勒展式知，

$$f(x+h) \approx f(x) + h * f'(x) + \frac{h^2}{2} * f''(x)$$

故差商引起的截断误差为

$$e_T \approx hf''(\xi)/2$$

设M是$f''(\xi)$的上界，则截断误差的误差限为

$$\epsilon_T = Mh/2$$

设计算$f(x)$的误差限为ϵ，则

$$\epsilon_R = 2\epsilon/h$$

故有

$$\epsilon_{tot} = \frac{Mh}{2} + \frac{2\epsilon}{h}$$

对于$f(x) = \sin(x), x = 1$，可视$M = sin(1)$，为计算方便，取$M = 1$。使用matlab双精度浮点数进行计算，可认为$\epsilon = 10^{-16}$，由此可解得当$h \approx 2 * 10^{-8}$时，差商法的计算误差最小。

在图里，开始时误差抖动较为剧烈，这是因为在步长h较小时，误差主要由计算$f(x)$的不精确造成，这种不精确具有一定的随机性。函数上升段较为平稳，是因为h较大时，误差的主要来源是差商法的舍入误差，其值为$h * \sin(1)/2$，这个误差是确定的。

1.3

(1) 当$n = 2097152$时，值不再变动，此时的和为15.403683。

【分析】IEEE单精度浮点数的ϵ_{mach}约为$6 * 10^{-8}$。定义s_n为使用单精度浮点数计算$\sum_{i=1}^{n} \frac{1}{i}$的结果。由课本定理1.6知，若

$$\left| \frac{1/n}{s_{n-1}} \right| < \frac{1}{2}\epsilon_{mach} \qquad (*)$$

则计算值不再变动。n较大时,

$$\sum_{i=1}^{n} \frac{1}{i} \approx \ln(n) \qquad (**)$$

。将$(**)$式代入$(*)$得，当计算值开始不变时，

$$\left| \frac{1/n}{\ln(n-1)} \right| \approx \frac{1}{2}\epsilon_{mach} \qquad (***)$$

用二分法解之得

$$n \approx 2.1 * 10^6$$

(2) 用双精度浮点数计算，得到$\sum_{x=1}^{2097151} 1/x = 15.13307$。故单精度浮点数的计算误差为0.270376,相对误差为1.7866%.

【分析】由于前$n-1$项的和与$1/n$值相差较大，每次加法都会引起一定的舍入误差，误差累计后便出现明显的计算误差.

(3) 双精度浮点数的ϵ_{mach}约为$6 * 10^{-8}$，将其代入（***）式，并用二分法求解，可得n约为10^{14}量级。在当前做实验的计算机上用双精度浮点数计算该级数的前10^{10}项，花费了12.616秒，故估计运行至求和结果不再变化，需要花费$12.616 * 10^4$秒，大概为35小时。

【关键代码】

```
fprintf("----------question 1----------\n");
sum = single(0);
n = single(0);
while (1)
    n = n + 1;
    cur = sum + single(1/n);
    if (cur == sum)
        break
    end
    sum = cur;
end
fprintf("end at %d : %f\n", n, sum)
fprintf("----------question 2----------\n");
sum_acc = 0;
n = double(n);
for i = 1:n
    sum_acc = sum_acc + 1/i;
end
fprintf("sum %f\nerror %f\nrelative error %f\n", sum_acc, ...
    sum-sum_acc, (sum-sum_acc)/sum_acc)
fprintf("----------question 3----------\n");
tic
n = 1e10;
sum_acc = 0;
for i = 1 : n
    sum_acc = sum_acc + 1/i;
end
t = toc;
fprintf('time used: %fs\n', t)
```

【程序输出】

```
----------question 1----------
end at 2097152 : 15.403683
----------question 2----------
sum 15.133307
error 0.270376
relative error 0.017866
----------question 3----------
time used: 12.960664s
```