Roundoff error propagation

If roundoff error caused by u stayed as big as $u \rightarrow$ no problem! Does it?

Example

Multiplication:

$$(x \otimes y) \otimes z \equiv fl(fl(x \times y) \times z) = [xy(1+\delta_1)] \times z(1+\delta_2)$$

where $| \delta_i | < u$

$$(x \otimes y) \otimes z = xyz(1+\delta_1)(1+\delta_2)$$

$$\left| \frac{(x \otimes y) \otimes z - xyz}{xyz} \right| = | (1+\delta_1)(1+\delta_2) - 1 | \leq (1+u)^2 - 1 \approx 2u$$

no problem with floating point multiplication!

How about addition?

If x, y are machine numbers (contain no roundoff error themselves)

$$x \oplus y \equiv fl(x+y) = [x+y](1+\delta_1)$$

$$\left| \frac{x \oplus y - (x+y)}{x+y} \right| = |\delta_1| \leq u$$

so no problem.
But what if they have roundoff errors from previous computations?

Floating point addition

$$fl(x) \oplus fl(y) = [x(1+\delta_1) + y(1+\delta_2)](1+\delta_3)$$

$$\begin{aligned}
\left| \frac{fl(x) \oplus fl(y) - (x+y)}{x+y} \right| &\leq \frac{|x|}{|x+y|} \left| (1+\delta_1)(1+\delta_3) - 1 \right| \\
&\quad + \frac{|y|}{|x+y|} \left| (1+\delta_2)(1+\delta_3) - 1 \right| \\
&\leq \frac{|x| + |y|}{|x+y|}((1+u)^2 - 1) \\
&\sim 2u \frac{|x| + |y|}{|x+y|}
\end{aligned}$$

but $\frac{|x|+|y|}{|x+y|}$ can be large if x, y are nearly equal and opposite!

If 2 nearly equal numbers (with error) are subtracted, the relative error can be greatly magnified!

Severe cancellation or subtractive cancellation

can greatly magnify the relative error so lose lots of precision in final answer

Example

Estimating the budget surplus/deficit

Remedy? $\rightarrow$ change formula/algorithm to avoid subtraction

Example

quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

what if $b^2 \gg 4ac$?

Example

sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Alternative form:

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

$\Longrightarrow$ try to avoid calculations that rely on cancellation ...
or use higher precision

Example

quadruple precision (not in MATLAB)

Example

in a symbolic environment

(not possible for big problems)

Stability of Algorithms

Since some error (roundoff) is usually present, we must have algorithms where the error doesn't grow too fast

Some growth is inevitable

Example

$S_n = \sum_{k=1}^{n} a_k$ typically has error $\sim n^{1/2} u$ and is guaranteed to have (absolute) error $< (n-1)u \sum_k \mid a_k \mid + O(u^2)$

but exponential growth (error $\sim K^n u, K > 1$) is disastrous!

An unstable recurrence relation ...

To compute the integral $I = \int_0^1 \frac{x^{100}}{x+2}\, dx$, we can derive the recurrence relation for $I_n = \int_0^1 \frac{x^n}{x+2}\, dx$

$$I_n = \frac{1}{n} - 2I_{n-1}$$

and run for $n = 1..100$, starting from $I_0 = \log(3/2)$

- **Demo** `BadRecurrence.m`

Remedy?

run recurrence **backwards!**

- **Demo** `GoodRecurrence.m`

Roundoff vs. truncation error

Sometimes there is a tradeoff between truncation error and roundoff

Example

forward difference approximation to the derivative

Approximate $f'(x)$ by $\frac{f(x+h)-f(x)}{h}$

- **Demo** `ForwardDifference.m`

Explanation

In the absence of roundoff error, there is still a truncation error. Use Taylor series for $f(x+h)$, assuming $f \in C^2$
$\rightarrow$ truncation error $\leq K_1 h$ (K_1 depends on f and x)

In the absence of roundoff error, approximation becomes exact as $h \rightarrow 0$

Using our model for roundoff error, get additional error $\leq K_2 u/h + K_3 u$ (K_2, K_3 depend on f and x)

Use bounds (worst case analysis) $\rightarrow$ minimum total (absolute) error at an optimal $h \approx u^{1/2}$, ignoring constants
$\implies$ no point using h smaller than this!

Roundoff error sets a lower bound to achievable accuracy

Summary: main effects of roundoff error

after *Afternotes on Numerical Analysis (Stewart)*
Roundoff error

1. can accumulate over long computations : inevitable

Example

sums

2. can reveal other errors by cancellation $\rightarrow$ try to do the problem another way
3. can grow so fast it swamps the actual answer $\rightarrow$ try to do the problem another way

Example

recurrence relations

End of Topic

End of Lecture 8