Numerical Analysis/Loss of Significance

Loss of significance occurs in numerical calculations when too many significant digits cancel. As a result of the floating point arithmetic used by computers, when a number is subtracted from another number that is almost exactly the same, catastrophic cancellation may occur and an erroneous value returned. As an example, consider the behavior of

$$f(x) = \sqrt{x^2+1} - 1$$

as x approaches 0. Evaluating this function at $x = 1.89 \times 10^{-9}$ using Matlab incorrectly returns the answer 0, which shows that too many significant digits have cancelled.

If x and y are positive, normalized floating point binary numbers such that $x > y$ and

$$2^{-q} \leq 1 - \frac{y}{x} \leq 2^{-p}\,, \qquad \textbf{(bound)}$$

then at most q and at least p significant binary bits are lost in the subtraction $x - y$.

On this page we will consider several exercises/examples of using this formula and show how sometimes we can rearrange the calculation to reduce loss of significance. Please try the exercise yourself before revealing the solution.

Exercises

Exercise 1

Use $2^{-q} \leq 1 - \frac{y}{x} \leq 2^{-p}$ to find a lower bound on the input x if one desires to lose no more than 1 significant binary bit in the calculation of $f(x) = \sqrt{x^2+1} - 1$.

Solution:

In (**bound**), "x" is $\sqrt{x^2+1}$ and "y" is 1; "q" = 1. We have $2^{-1} \leq 1 - \frac{1}{\sqrt{x^2+1}}$. Solving for x, we have $x^2 \geq 3$, which is our lower bound on x.

Exercise 2

Use $2^{-q} \leq 1 - \frac{y}{x} \leq 2^{-p}$ to find a lower bound on the input x if one desires to lose at most 3 significant binary bits in the calculation of $f(x) = \sqrt{x^2+1} - 1$.

Solution:

In (**bound**), "x" is $\sqrt{x^2+1}$ and "y" is 1; "q" = 3. We have $2^{-3} \leq 1 - \frac{1}{\sqrt{x^2+1}}$. Solving for x, we have $x^2 \geq \frac{15}{49} \approx .306122449$, which is our lower bound on x.

Exercise 3

Consider

$$f(x) = \sqrt{x^2+1} - x\,.$$

As x gets very large, loss of significance can occur. What is the bound on x if we want to lose no more than one binary digit?

Solution:

We have

$$2^{-1} \leq 1 - \frac{x}{\sqrt{x^2+1}} \Rightarrow x^2 \leq \frac{1}{3}\,.$$

Exercise 4

Now consider the function

$$f(x) = \log(x+1) - \log(x)\,.$$

For large values of x, loss of significance may occur.

Use (**bound**) to find a bound on the input so that at most 1 significant binary bit will be lost in the calculation.

Solution:

In (**bound**), "x" is $\log(x+1)$, "y" is $\log(x)$, and "q" is 1. We have $2^{-1} \leq 1 - \frac{\log(x)}{\log(x+1)}$. Solving for x gives the interval $0 < x \leq \frac{1+\sqrt{5}}{2} \approx 1.618033989$.

Exercise 5

In w:Numerical differentiation, the w:Finite difference

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

is often used to approximate derivatives. w:Truncation error can be reduced by decreasing "h", the step size, but if h becomes too small, loss of significance can become a factor. For $\frac{f(x+h)-f(x)}{h}$, find a bound on h such that at most 1 binary bit will be lost in the calculation.

Solution:

We have

$$2^{-1} \leq 1 - \frac{f(x)}{f(x+h)} \Rightarrow \frac{f(x)}{f(x+h)} \leq \frac{1}{2} \Rightarrow 2f(x) \leq f(x+h)\,.$$

Exercise 6

For $\frac{f(x+h)-f(x)}{h}$, find a bound on h such that at most 1 binary bit will be lost in the calculation, if $f(x) = x^2$.

Solution:

We have

$$2x^2 \leq (x+h)^2 = x^2 + 2xh + h^2 \Rightarrow h^2 + 2xh - x^2 \geq 0\,.$$

Comparing this equation to the general form,

$$ax^2 + bx + c = 0$$

and using the w:quadratic formula

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

we see that the variable in our equation is "h", a = 1, b = 2x, and c = x^2. We use the quadratic formula to solve for "h".

The quadratic formula itself can be a cause of w:loss of significance if the quantity "4ac" is very small. This can be remedied by not subtracting.

If "b" (in this case, "2x") is positive, subtraction can be avoided by using

$$x_1 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \text{ or,in this case } h_1 = \frac{-2x - \sqrt{4x^2 + 4x^2}}{2} = -(1 + \sqrt{2})x\,.$$

Unfortunately, this gives a value for "h" that is always negative, which is unacceptable. Using one of w:Vieta's formulas,

$$h_2 = \frac{c}{ah_1} = \frac{x}{1 + \sqrt{2}}$$

which gives positive values for h. If "b" (here, "2x") is negative, we use

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \text{ or, in this case, } h_1 = \frac{-2x + \sqrt{4x^2 + 4x^2}}{2} = |x| + |x|\sqrt{2} = (\sqrt{2} + 1)|x|$$

which, is positive. Therefore, there's no need to find the other solution.

The bounds on h are then:

$$\begin{aligned} &h \geq \frac{x}{1 + \sqrt{2}} \quad \text{for } x > 0 \quad \text{and} \\ &h \geq |x|(1 + \sqrt{2}) \quad \text{for } x < 0 \end{aligned}$$

Exercise 7

Rewrite $f(x) = \sqrt{x^2 + 1} - 1$ so that there is no loss of significance, then evaluate it at $x = 1.89 \times 10^{-9}$.

Solution:

Rationalize the expression:

$$\frac{\sqrt{x^2 + 1} - 1}{1} \times \frac{\sqrt{x^2 + 1} + 1}{\sqrt{x^2 + 1} + 1} = \frac{x^2 + 1 - 1}{\sqrt{x^2 + 1} + 1} = \frac{x^2}{\sqrt{x^2 + 1} + 1}.$$

Evaluating this expression at $x = 1.89 \times 10^{-9}$ gives an answer of 1.78605×10^{-18}.

Exercise 8

Rewrite $f(x) = \log(x + 1) - \log(x)$ so that loss of significance will be minimized.

Solution:

Use the Quotient Property of logarithms to rewrite $f(x) = \log(x + 1) - \log(x)$ as $\log\left(\frac{x + 1}{x}\right)$.

Exercise 9

For very small values of x, loss of significance can occur in $f(x) = \frac{1 - x}{1 + x} - \frac{1}{3x + 1}$. Rewrite this function in a way that will minimize loss of significance.

Solution:

Giving both terms a common denominator and combining them into a single rational expression reduces loss of significance. We then have

$$\frac{1-x}{1+x}-\frac{1}{3x+1}=\frac{(1-x)(3x+1)-(1+x)}{(1+x)(3x+1)}=\frac{x-3x^2}{3x^2+4x+1}\,.$$

Exercise 10

As x gets close to zero, loss of significance can occur in $f(x)=\frac{1-\cos(x)}{\sin(x)}$. Rewrite this function in a way that will minimize loss of significance.

Solution:

The loss of significance occurs in the numerator, so rewrite the numerator using a trigonometric identity to get

$$\frac{1-\cos(x)}{\sin(x)}=\frac{2\sin^2(x/2)}{\sin(x)}\,.$$
