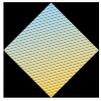


Lies My Calculator and Computer Told Me





Lies My Calculator and Computer Told Me



▲ See Section 1.4 for a discussion of graphing calculators and computers with graphing software.

A wide variety of pocket-size calculating devices are currently marketed. Some can run programs prepared by the user; some have pre-programmed packages for frequently used calculus procedures, including the display of graphs. All have certain limitations in common: a limited range of magnitude (usually less than 10^{100} for calculators) and a bound on accuracy (typically eight to thirteen digits).

A calculator usually comes with an owner's manual. Read it! The manual will tell you about further limitations (for example, for angles when entering trigonometric functions) and perhaps how to overcome them.

Program packages for microcomputers (even the most fundamental ones, which realize arithmetical operations and elementary functions) often suffer from hidden flaws. You will be made aware of some of them in the following examples, and you are encouraged to experiment using the ideas presented here.

▲ Preliminary Experiments with your Calculator or Computer

To have a first look at the limitations and quality of your calculator, make it compute $2 \div 3$. Of course, the answer is not a terminating decimal so it can't be represented exactly on your calculator. If the last displayed digit is 6 rather than 7, then your calculator approximates $\frac{2}{3}$ by truncating instead of rounding, so be prepared for slightly greater loss of accuracy in longer calculations.

Now multiply the result by 3; that is, calculate $(2 \div 3) \times 3$. If the answer is 2, then subtract 2 from the result, thereby calculating $(2 \div 3) \times 3 - 2$. Instead of obtaining 0 as the answer, you might obtain a small negative number, which depends on the construction of the circuits. (The calculator keeps, in this case, a few "spare" digits that are remembered but not shown.) This is all right because, as previously mentioned, the finite number of digits makes it impossible to represent $2 \div 3$ exactly.

A similar situation occurs when you calculate $(\sqrt{6})^2 - 6$. If you do not obtain 0, the order of magnitude of the result will tell you how many digits the calculator uses internally.

Next, try to compute $(-1)^5$ using the y^x key. Many calculators will indicate an error because they are built to attempt $e^{5 \ln(-1)}$. One way to overcome this is to use the fact that $(-1)^k = \cos k\pi$ whenever k is an integer.

Calculators are usually constructed to operate in the decimal number system. In contrast, some microcomputer packages of arithmetical programs operate in a number system with base other than 10 (typically 2 or 16). Here the list of unwelcome tricks your device can play on you is even larger, since not all terminating decimal numbers are represented exactly. A recent implementation of the BASIC language shows (in double precision) examples of incorrect conversion from one number system into another, for example,

$$8 \times 0.1 \stackrel{?}{=} 0.79999\ 99999\ 99999\ 9$$

whereas $19 \times 0.1 \stackrel{?}{=} 1.90000\ 00000\ 00001$

Yet another implementation, apparently free of the preceding anomalies, will not calculate standard functions in double precision. For example, the number $\pi = 4 \times \tan^{-1}1$, whose representation with sixteen decimal digits should be 3.14159 26535 89793, appears as 3.14159 29794 31152; this is off by more than

3×10^{-7} . What is worse, the cosine function is programmed so badly that its “cos” $0 = 1 + 2^{-23}$. (Can you invent a situation when this could ruin your calculations?) These or similar defects exist in other programming languages too.

The Perils of Subtraction

You might have observed that subtraction of two numbers that are close to each other is a tricky operation. The difficulty is similar to this thought exercise: Imagine that you walk blindfolded 100 steps forward and then turn around and walk 99 steps. Are you sure that you end up exactly one step from where you started?

The name of this phenomenon is “loss of significant digits.” To illustrate, let’s calculate

$$8721\sqrt{3} - 10,681\sqrt{2}$$

The approximations from my calculator are

$$8721\sqrt{3} \approx 15105.21509 \quad \text{and} \quad 10,681\sqrt{2} \approx 15105.21506$$

and so we get $8721\sqrt{3} - 10,681\sqrt{2} \approx 0.00003$. Even with three spare digits exposed, the difference comes out as 0.00003306. As you can see, the two ten-digit numbers agree in nine digits that, after subtraction, become zeros before the first nonzero digit. To make things worse, the formerly small errors in the square roots become more visible. In this particular example we can use rationalization to write

$$8721\sqrt{3} - 10,681\sqrt{2} = \frac{1}{8721\sqrt{3} + 10,681\sqrt{2}}$$

(work out the details!) and now the loss of significant digits doesn’t occur:

$$\frac{1}{8721\sqrt{3} + 10,681\sqrt{2}} \approx 0.00003310115 \quad \text{to seven digits}$$

(It would take too much space to explain why all seven digits are reliable; the subject *numerical analysis* deals with these and similar situations.) See Exercise 7 for another instance of restoring lost digits.

Now you can see why in Exercises 2.2 your guess at the limit of $(\tan x - x)/x^3$ was bound to go wrong: $\tan x$ becomes so close to x that the values will eventually agree in all digits that the calculator is capable of carrying. Similarly, if you start with just about any continuous function f and try to guess the value of

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

long enough using a calculator, you will end up with a zero, despite all the rules in Chapter 3!

Where Calculus is More Powerful than Calculators and Computers

One of the secrets of success of calculus in overcoming the difficulties connected with subtraction is symbolic manipulation. For instance, $(a + b) - a$ is always b , although the calculated value may be different. Try it with $a = 10^7$ and $b = \sqrt{2} \times 10^{-5}$. Another powerful tool is the use of inequalities; a good example is the Squeeze Theo-

rem as demonstrated in Section 2.3. Yet another method for avoiding computational difficulties is provided by the Mean Value Theorem and its consequences, such as l'Hospital's Rule (which helps solve the aforementioned exercise and others) and Taylor's Inequality.

The limitations of calculators and computers are further illustrated by infinite series. A common misconception is that a series can be summed by adding terms until there is "practically nothing to add" and "the error is less than the first neglected term." The latter statement is true for certain alternating series (see the Alternating Series Estimation Theorem) but not in general; a modified version is true for another class of series (Exercise 10). As an example to refute these misconceptions, let's consider the series

$$\sum_{n=1}^{\infty} \frac{1}{n^{1.001}}$$

which is a convergent p -series ($p = 1.001 > 1$). Suppose we were to try to sum this series, correct to eight decimal places, by adding terms until they are less than 5 in the ninth decimal place. In other words, we would stop when

$$\frac{1}{n^{1.001}} < 0.000000005$$

that is, when $n = N = 196,217,284$. (This would require a high-speed computer and increased precision.) After going to all this trouble, we would end up with the approximating partial sum

$$S_N = \sum_{n=1}^N \frac{1}{n^{1.001}} < 19.5$$

But, from the proof of the Integral Test, we have

$$\sum_{n=1}^{\infty} \frac{1}{n^{1.001}} > \int_1^{\infty} \frac{dx}{x^{1.001}} = 1000$$

Thus, the machine result represents less than 2% of the correct answer!

Suppose that we then wanted to add a huge number of terms of this series, say, 10^{100} terms, in order to approximate the infinite sum more closely. (This number 10^{100} , called a *googol*, is outside the range of pocket calculators and is much larger than the number of elementary particles in our solar system.) If we were to add 10^{100} terms of the above series (only in theory; a million years is less than 10^{26} microseconds), we would still obtain a sum of less than 207 compared with the true sum of more than 1000. (This estimate of 207 is obtained by using a more precise form of the Integral Test, known as the Euler-Maclaurin Formula, and only then using a calculator. The formula provides a way to accelerate the convergence of this and other series.)

If the two preceding approaches didn't give the right information about the accuracy of the partial sums, what does? A suitable inequality satisfied by the remainder of the series, as you can see from Exercise 6.

Computers and calculators are not replacements for mathematical thought. They are just replacements for some kinds of mathematical labor, either numerical or symbolic. There are, and always will be, mathematical problems that can't be solved by a calculator or computer, regardless of its size and speed. A calculator or computer does stretch the human capacity for handling numbers and symbols, but there is still considerable scope and necessity for "thinking before doing."



Exercises

A [Click here for answers.](#)

1. Guess the value of

$$\lim_{x \rightarrow 0} \left(\frac{1}{\sin^2 x} - \frac{1}{x^2} \right)$$

and determine when to stop guessing before the loss of significant digits destroys your result. (The answer will depend on your calculator.) Then find the precise answer using an appropriate calculus method.

2. Guess the value of

$$\lim_{h \rightarrow 0} \frac{\ln(1+h)}{h}$$

and determine when to stop guessing before the loss of significant digits destroys your result. This time the detrimental subtraction takes place inside the machine; explain how (assuming that the Taylor series with center $a = 1$ is used to approximate $\ln x$). Then find the precise answer using an appropriate calculus method.

3. Even innocent-looking calculus problems can lead to numbers beyond the calculator range. Show that the maximum value of the function

$$f(x) = \frac{x^2}{(1.0001)^x}$$

is greater than 10^{124} . [Hint: Use logarithms.] What is the limit of $f(x)$ as $x \rightarrow \infty$?

4. What is a numerically reliable expression to replace $\sqrt{1 - \cos x}$, especially when x is a small number? You will need to use trigonometric identities. (Recall that some computer packages would signal an unnecessary error condition, or even switch to complex arithmetic, when $x = 0$.)
5. Try to evaluate

$$D = \ln \ln(10^9 + 1) - \ln \ln(10^9)$$

on your calculator. These numbers are so close together that you will likely obtain 0 or just a few digits of accuracy. However, we can use the Mean Value Theorem to achieve much greater accuracy.

- (a) Let $f(x) = \ln \ln x$, $a = 10^9$, and $b = 10^9 + 1$. Then the Mean Value Theorem gives

$$f(b) - f(a) = f'(c)(b - a) = f'(c)$$

where $a < c < b$. Since f' is decreasing, we have $f'(a) > f'(c) > f'(b)$. Use this to estimate the value of D .

- (b) Use the Mean Value Theorem a second time to discover why the quantities $f'(a)$ and $f'(b)$ in part (a) are so close to each other.

6. For the series $\sum_{n=1}^{\infty} n^{-1.001}$, studied in the text, exactly how many terms do we need (in theory) to make the error less than 5 in the ninth decimal place? You can use the inequalities from the proof of the Integral Test:

$$\int_{N+1}^{\infty} f(x) dx < \sum_{n=N+1}^{\infty} f(n) < \int_N^{\infty} f(x) dx$$

7. Archimedes found an approximation to 2π by considering the perimeter p of a regular 96-gon inscribed in a circle of radius 1. His formula, in modern notation, is

$$p = 96\sqrt{2 - \sqrt{2 + \sqrt{2 + \sqrt{2 + \sqrt{3}}}}}$$

- (a) Carry out the calculations and compare with the value of p from more accurate sources, say $p = 192 \sin(\pi/96)$. How many digits did you lose?
- (b) Perform rationalization to avoid subtraction of approximate numbers and count the exact digits again.
8. This exercise is related to Exercise 2. Suppose that your computing device has an excellent program for the exponential function $\exp(x) = e^x$ but a poor program for $\ln x$. Use the identity

$$\ln a = b + \ln \left(1 + \frac{a - e^b}{e^b} \right)$$

and Taylor's Inequality to improve the accuracy of $\ln x$.

9. The cubic equation

$$x^3 + px + q = 0$$

where we assume for simplicity that $p > 0$, has a classical solution formula for the real root, called Cardano's formula:

$$x = \frac{1}{3} \left[\left(\frac{27q + \sqrt{729q^2 + 108p^3}}{2} \right)^{1/3} + \left(\frac{27q - \sqrt{729q^2 + 108p^3}}{2} \right)^{1/3} \right]$$

For a user of a pocket-size calculator, as well as for an inexperienced programmer, the solution presents several stumbling blocks. First, the second radicand is negative and the fractional power key or routine may not handle it. Next, even if we fix the negative radical problem, when q is small in magnitude and p is of moderate size, the small number x is the difference of two numbers close to $\sqrt[3]{p/3}$.

- (a) Show that all these troubles are avoided by the formula

$$x = \frac{-9q}{a^{2/3} + 3p + 9p^2 a^{-2/3}}$$

where $a = \frac{27|q| + \sqrt{729q^2 + 108p^3}}{2}$

Hint: Use the factorization formula

$$A + B = \frac{A^3 + B^3}{A^2 - AB + B^2}$$

(b) Evaluate

$$u = \frac{4}{(2 + \sqrt{5})^{2/3} + 1 + (2 + \sqrt{5})^{-2/3}}$$

If the result is simple, relate it to part (a), that is, restore the cubic equation whose root is u written in this form.

10. (a) Consider the power series

$$f(x) = \sum_{n=1}^{\infty} \frac{x^n}{100^n + 1}$$

It is easy to show that its radius of convergence is $r = 100$. The series will converge rather slowly at $x = 99$: find out how many terms will make the error less than 5×10^{-7} .

(b) We can speed up the convergence of the series in part (a). Show that

$$f(x) = \frac{x}{100 - x} - f\left(\frac{x}{100}\right)$$

and find the number of the terms of this transformed series that leads to an error less than 5×10^{-7} .

[*Hint:* Compare with the series $\sum_{n=1}^{\infty} (x^n/100^n)$, whose sum you know.]

11. The positive numbers

$$a_n = \int_0^1 e^{1-x} x^n dx$$

can, in theory, be calculated from a reduction formula obtained by integration by parts: $a_0 = e - 1$, $a_n = na_{n-1} - 1$. Prove, using $1 \leq e^{1-x} \leq e$ and the Squeeze Theorem, that $\lim_{n \rightarrow \infty} a_n = 0$. Then try to calculate a_{20} from the reduction formula using your calculator. What went wrong?

The initial term $a_0 = e - 1$ can't be represented exactly in a calculator. Let's call c the approximation of $e - 1$ that we can enter. Verify from the reduction formula (by observing the pattern after a few steps) that

$$a_n = \left[c - \left(\frac{1}{1!} + \frac{1}{2!} + \dots + \frac{1}{n!} \right) \right] n!$$

and recall from our study of Taylor and Maclaurin series that

$$\frac{1}{1!} + \frac{1}{2!} + \dots + \frac{1}{n}$$

converges to $e - 1$ as $n \rightarrow \infty$. The expression in square brackets converges to $c - (e - 1)$, a nonzero number, which gets multiplied by a fast-growing factor $n!$. We conclude that even if all further calculations (after entering a_0) were performed without errors, the initial inaccuracy would cause the computed sequence $\{a_n\}$ to diverge.

12. (a) A consolation after the catastrophic outcome of Exercise 11: If we rewrite the reduction formula to read

$$a_{n-1} = \frac{1 + a_n}{n}$$

we can use the inequality used in the squeeze argument to obtain improvements of the approximations of a_n . Try a_{20} again using this reverse approach.

(b) We used the reversed reduction formula to calculate quantities for which we have elementary formulas. To see that the idea is even more powerful, develop it for the integrals

$$\int_0^1 x^{n-\theta} e^{1-x} dx$$

where θ is a constant, $0 < \theta < 1$, and $n = 0, 1, \dots$. For such θ the integrals are no longer elementary (not solvable in "finite terms"), but the numbers can be calculated quickly. Find the integrals for the particular choice $\theta = \frac{1}{3}$ and $n = 0, 1, \dots, 5$ to five digits of accuracy.

13. An advanced calculator has a key for a peculiar function:

$$E(x) = \begin{cases} 1 & \text{if } x = 0 \\ \frac{e^x - 1}{x} & \text{if } x \neq 0 \end{cases}$$

After so many warnings about the subtraction of close numbers, you may appreciate that the definition $\sinh x = \frac{1}{2}(e^x - e^{-x})$ gives inaccurate results for small x , where $\sinh x$ is close to x . Show that the use of the accurately evaluated function $E(x)$ helps restore the accuracy of $\sinh x$ for small x .



Answers

1. $\frac{1}{3}$ 3. 0 5. (a) $4.82549424 \times 10^{-11}$ 9. (b) $1, x^3 + 3x - 4 = 0$