1. Introduction and terminology

Numerical analysis is the application of computing techniques to approximate solutions to analytical and algebraic problems. In calculus, you learn about differentiation and integration:

1. differentiation determines the rate of change, while
2. integration determines the rate of accumulation.

For example, what is the rate of change of the electromagnetic force at a point in time versus what is the cumulative effect of a force on a point over a period of time? In linear algebra, you learned about setting up and solving systems of linear equations. For example, what are the currents flowing through a specific circuit with given voltages?

In engineering, we are not usually looking for an exact solution: exact solutions do not exist in mathematics, for there is not even a general formula for the roots of a quintic polynomial $a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0$. Additionally, engineers are not interested in good approximations, but rather sufficiently good approximations for the engineering problem at hand. Estimating where a self-driving car will be in the next second should likely be accurate up to a centimeter or two, but landing an aircraft on autopilot will likely see a touchdown that need only be a few meters from the optimal point.

1.1 Error, absolute error and relative error

Thus, we must first define the accuracy of an approximation:

If we are attempting to approximate an exact value *x* through an approximation *a*, the *error* will be defined as

$$|x-a|,$$

although often we will use the *absolute error*, defined as

$$|x-a|,$$

while the *relative error* will be defined as

$$\frac{|x-a|}{|x|}.$$

That is, how large is the error relative to the quantity we are attempting to approximate. If we multiply the relative error by 100%, we get the percent relative error:

$$\frac{|x-a|}{|x|}\cdot 100\%.$$

Here are three examples:

1. A resistor is graded at 120 Ω, but a careful measurement reveals that the actual resistance is 126.38 Ω, then the absolute error of the stated resistance of 120 Ω is 6.38 Ω while the relative error is 0.0504 or 5.04 %.
2. For example, how good is $\frac{355}{113}$ an approximation of π? The absolute error is approximately 2.67 × 10^{-7} while the relative error 8.49 × 10^{-8}.
3. Erwin Hubble estimated the distance to the Andromeda galaxy to be 1.5 billion light years, while modern estimates put the distance closer to 2.54 million light years. The absolute error of Hubble's estimate is 1.05 million light years, while the relative error is 0.413 or 41.3 %.

Note that the absolute error retains the units: an absolute error of 6.38 Ω is also an absolute error of 6380000 μΩ as well as 0.00638 kΩ. The relative error is, however, calculated as being proportional to the actual value.

1.2 Significant digits

You may suggest that 3.14 is, as an approximation to π, correct to three significant digits. On the other hand, consider 1.995 and 2.09 as approximations to 2. You may suggest that the first has no significant digits, while the second has two; however, the percent relative error of the first approximation is 0.25 %, while the percent relative error of the second is 4.5 %, so the second is a much worse approximation than the first.

Instead, we will use the phrase that a approximates x to n significant digits if the relative error is no greater than ½ 10^{-n}. Thus, 1.995 is correct to 3 significant digits as an approximation to 2, while 2.09 is only correct to one significant digit as an approximation to 2.

Please note, *significant digits* will only be used colloquially in this course. If a approximates x to 3 significant digits, the relative error is less than 0.0005, while if a approximates x to 12 significant digits, the relative error is less than

$$0.\underbrace{000000000000}_{\text{12 zeros}}5\ .$$

1.3 Precision versus accuracy

Another two concepts that we will use to describe numerical algorithms are accuracy and precision.

A method or technique is precise if it is capable of generating approximations that have very small absolute error; however, a method or technique is accurate if it gives an unbiased approximation; that is, an approximation that does not have some form of systematic error.

For example, as an approximation to the mean (average) height of 1000 students, taking the average of eleven randomly chosen students will give a reasonably good approximation as to the global mean. Another approximation of the global mean is the median, but the median is not as precise a tool as the average.

As an alternate example, suppose you want to find the maximum and minimum heights of 1000 students, and so you take a sample of 11 students. If you were to use the tallest and shortest of these students to approximate the maximum and minimum heights, you would always find that your approximation of the maximum height to be underestimated, and the approximation of the minimum height to be overestimated. Instead, if you assume that the heights are uniformly distributed, a better approximation would be to find the tallest and shortest heights, call them *b* and *a*, respectively, and then approximate the maximum height with $b+\frac{b-a}{n}$ and the minimum height by $a-\frac{b-a}{n}$. These two estimators are more accurate, even if their precision is unchanged.

To demonstrate this, suppose we take 10 random samples from a uniform distribution of 20 to 30:

23.957, 21.931, 20.224, 28.002, 24.276, 28.426, 24.123, 29.964, 23.864, 26.946
27.730, 27.306, 21.065, 23.964, 29.449, 22.109, 27.501, 24.547, 27.366, 23.298
26.157, 28.470, 24.711, 23.926, 28.364, 24.743, 22.241, 20.786, 27.207, 25.597
27.597, 24.301, 28.676, 28.668, 29.669, 24.972, 25.562, 23.636, 29.884, 27.353

The minimum and maximum of each of these is:

20.224, 29.964
21.065, 29.449
20.786, 28.470
23.636, 29.884

so each of the first numbers is an over-estimate of the lower bound of 20, and the second number is an under-estimate of the upper bound of 30; however, if you calculate the spread and subtract or and one-tenth of each spread from the numbers, respectively, you get reasonably bester estimators of the lower and upper bounds:

19.2500, 30.9380
20.2266, 30.2874
20.0176, 29.2384
23.0112, 30.5088

Note that these four samples were generated randomly and the third case demonstrates that samples are not always perfect at estimating global properties.

As another example, if you use simple elementary school ruler to measure the width of a wire, you may get an answer that has an absolute error of at most 0.2 mm. If, however, you were to use a micrometer, the answer will be likely accurate to the nearest micrometre. If the micrometer is not calibrated correctly, however, it may give inaccurate measurements, and similarly, if the wooden ruler from school was, for example, dumped in water, it may warp and thus may similarly become less accurate.

As another example, the .308 calibre Lee Enfield rifle is not as precise as the 0.338 calibre C14 Timberwolf, the current sniper rifle adopted by the Canadian Forces. If the sites are correctly aligned, each firearm may be accurate, however, the Timberwolf will always be more precise.

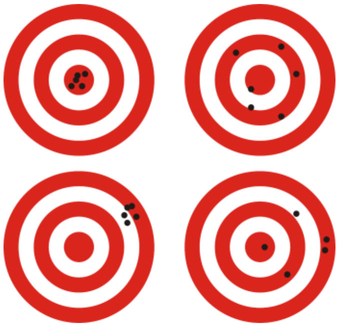

Figure 1. Accurate and inaccurate groupings of a precision firearm on the left, with accurate and inaccurate groupings of a less precise firearm on the right.

As a final example, the `double` type used in C++ can store an approximation that is correct up to 16 significant digits, while the `float` type can only store 7 significant digits. Consequently, the double-precision floating-point representation is significantly more precise than the single-precision floating-point representation.