Approximations in Scientific Computing

Jamie M. Foster
http://www.math.mcmaster.ca/∼jmfoster

Sources of Error

- Approximations **before** computation begins:

  1. **Modelling error**: Physics may be neglected or omitted during model formulation, *e.g.*, neglecting air resistance in trajectory of a cannonball.
  2. **Previous computations/experimental data**: Approximations in input data that were computed previously. Experimental data is imperfect since lab equipment has finite precision.

- Approximations **during** computations:

  1. **Truncation or discretisation error**: Some features of the mathematical model are approximated, *e.g.*, derivatives are replaced by finite difference approximations, or an infinite series may be approximated by a finite one.
  2. **Rounding error**: Representation of real numbers and arithmetic operations on them is limited to some finite precision, *i.e.*, the calculation is not really exact!

Example: Compute the surface area of the Earth

Approximation 1: Earth's shape is idealized as sphere (modeling error). Thus, the surface area, $S = 4\pi r^2$.

Approximation 2: $r \approx 6370$ km, a value from empirical measurements and previous calculations.

Approximation 3: Value of $\pi \approx 3.142$ (truncation error).

Approximation 4: Calculated value of $4\pi r^2$ is finally rounded (rounding error).

Error analysis

Studying the effects of such approximations on the accuracy and the stability of numerical algorithms is called error analysis.

Absolute error:

$$\text{Absolute error} = \text{Approximate value} - \text{True value} \tag{1}$$

Note that some others define the absolute error as

$$\text{Absolute error} = |\text{Approximate value} - \text{True value}| \tag{2}$$

Comment: Absolute error is of little practical use. Example: An absolute error of 1 in the count of population of a country is not a large error. However, an absolute error of 1 in the count of the count of a population of a household is large.

Relative error:

$$\text{Relative error} = \frac{\text{Approximate value} - \text{True value}}{\text{True value}}. \tag{3}$$

$$= \frac{\text{Absolute error}}{\text{True value}}. \tag{4}$$

Comment: Relative error is more meaningful because it 'scales' with the 'size' of the problem.

Precision vs. accuracy

Precision: the number of digits with which a number is represented.

Accuracy: the number of **correct** significant digits in approximating the quantity. Comment: Computing a quantity with a high precision does not guarantee high accuracy.

Data error and computational error

Error in the final calculation is due to a combination of error in the input data and/or the computations using the data.

Example: Consider the following 1-dimensional problem $f : \Re \to \Re$, and denote

$$\begin{aligned} x &\quad \to \text{True value of input} \\ f(x) &\quad \to \text{True value of output} \\ \hat{x} &\quad \to \text{Inexact input} \\ \hat{f} &\quad \to \text{Approximate output} \end{aligned}$$

$$\begin{aligned} \text{Total error} &= \hat{f}(\hat{x}) - f(x) \\ &= \left(\hat{f}(\hat{x}) - f(\hat{x})\right) + (f(\hat{x}) - f(x)) \\ &= \text{computational error} + \text{data error} \end{aligned} \tag{5}$$

Example: Calculate $\sin(\pi/8)$ without using a calculator
Approximate solution: $\pi \approx 22/7 \approx 3$, and $\sin(x) \approx x$ (for small x), $\sin(\pi/8) \approx \sin(3/8) \approx 3/8 = 0.3750$. Thus $\hat{f}(\hat{x}) = 0.3750$.
True solution: $\sin(\pi/8) = 0.3827$

$$\text{Total error} = \hat{f}(\hat{x}) - f(\hat{x}) \approx 0.3750 - 0.3827. \tag{6}$$

$$\text{Propagated data error} = f(\hat{x}) - f(x) = \sin(3/8) - \sin(\pi/8) \tag{7}$$

$$\approx 0.3663 - 0.3827 = -0.0164. \tag{8}$$

1. Data error and computational errors can have opposite signs.
2. One of them can dominate.
3. To decrease the total error: (a.) Use more accurate value of input data. (b.) Use more accurate mathematical representation of f.

Computational errors

Truncation error: Difference between true result and the result produced by the given algorithm. Sources: truncating infinite series, replacing derivatives by finite differences and terminating iterative sequence before convergence. Dominant error in problems involving integrals and derivatives and non-linearities.

Rounding errors: Difference between the results produced by the given algorithm using exact arithmetic and the results produced by the same algorithm using finite precision, rounded arithmetic. Dominant error in purely algebraic problems with finite solution algorithms.

Sensitivity and conditioning

Problem formulation itself might be very sensitive to input data!

Sensitivity: is qualitative notion of how sensitive an output is to the input.

Conditioning: is the quantitative measure of sensitivity. Together they define the problem as *well-conditioned* or *ill-conditioned*.

The condition number:

$$C = \frac{\left|\left(f(\hat{x}) - f(x)\right)/f(x)\right|}{\left|\left(\hat{x} - x\right)/x\right|}. \tag{9}$$

If $C \gg 1$ then the problem is ill-conditioned. If $C \ll 1$ the problem is well-conditioned.