Numerical Analysis – FMN011

Carmen Arévalo
Lund University
carmen@maths.lth.se

Lecture 1

What is Numerical Analysis/Scientific Computing?

A few problems can be solved exactly:

Linear systems

$$\begin{aligned} x + y &= 1 \\ x - y &= 0 \Rightarrow x = y = 1/2 \end{aligned}$$

Integrals of polynomials

$$I = \int_0^1 x^3 dx \Rightarrow I = 1/4$$

Most problems must be solved approximately

Some problems cannot be solved exactly:

Non-linear equations

$$x^5 + 3x^4 - 7x^3 + x^2 + 2x - 2 = 0 \qquad \Rightarrow x = ?$$

Differential equations

$$\begin{aligned} \dot{x} &= e^{t^2} \\ x(1) &= 0 \qquad \Rightarrow x(t) = ? \end{aligned}$$

Categories of Mathematical Problems

Category	linear	non-linear
algebra	computable	not computable
analysis	not computable	not computable

All categories can be computed with numerical methods!

"Numerical analysis is the area of mathematics and computer science that creates, analyzes, and implements algorithms for solving numerically the problems of continuous mathematics."

Our goal: to construct and analyze mathematical software.

Error analysis

A “numerical solution” is different from a “mathematical (closed) solution.”

Numerical solutions are approximations to the exact solution.
How “good” a numerical solution is depends on how close it is to the exact solution and what is the error we are willing to tolerate.

If $\hat{p}$ is an approximation to p,

- the absolute error is $E_p = |p - \hat{p}|$
- The relative error is $R_p = \frac{|p-\hat{p}|}{|p|}$. It may be expressed as a percentage.

Notation and Accuracy

Floating point form of a number:

$$fl(p) = \pm(\frac{d_0}{\beta^0} + \frac{d_1}{\beta^1} + \frac{d_2}{\beta^2} + \cdots + \frac{d_{k-1}}{\beta^{k-1}}) \times \beta^E$$

d_i is an integer ($0 \leq d_i \leq \beta - 1$), $\beta > 1$ is the base, k is the precision (number of digits).

Normalized decimal form of a number: $p = \pm d_1.d_2d_3 \cdots d_kd_{k+1} \cdots \times 10^n$

Chopping off: $fl_{chop}(p) = \pm d_0.d_1d_2 \cdots d_{k-1} \times 10^n$

Rounding off: $fl_{round}(p) = \pm d_0.d_1d_2 \cdots r_{k-1} \times 10^n$ where $d_{k-1}d_k \cdots$ is rounded to the nearest integer. In case of a tie, round to nearest even number.

Machine epsilon

Also called rounding unit,

$$\eta = \frac{1}{2}\beta^{1-k}$$

gives the bound on the relative error when we represent a number in a floating point system. In the standard floating point system,

$$\frac{|x - \mathrm{fl}(x)|}{|x|} \leq \eta \approx 1.1e^{-16}$$

Correct significant digits

We approximate

$$x = \pi \approx 3.14159\ldots$$

by

$$\hat{x} = 3.14$$

$\hat{x}$ approximates π to 3 correct (significant) digits.

$$\hat{x} = 3.141$$

also approximates π to 3 correct (significant) digits

Types of errors

- **Truncation error:** occurs when an exact formula is replaced by another to make it easier (or possible) to solve numerically.

- **Round-off error:** occurs because computers cannot represent all real numbers exactly, as computer numbers have a limited number of digits.

- **Noise:** is the error in data. There can also be errors in the mathematical model. The accuracy of the numerical result must take these errors into consideration.

Loss of Significance or Cancelation Error

Loss of accuracy can occur when two very similar numbers are subtracted.

Two numbers with 7-digit accuracy are subtracted

$$\begin{array}{r} 123.4567 \\ -123.4566 \\ \hline 000.0001 \end{array}$$

the result has only 1-digit accuracy.

Taylor's polynomial

Many functions can be approximated in $[x_0 - \delta, x_0 + \delta]$ as

$$f(x) = P_n(x) + R_n(x)$$

where P_n is the **Taylor polynomial of degree** n,

$$P_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

and the **remainder** is

$$R_n(x) = \frac{f^{(n+1)}(\zeta)}{(n+1)!}(x - x_0)^{n+1}$$

with ζ between x and x_0

A Taylor approximation

The Taylor polynomial of degree 3 for $f(x) = x^{-2}$ about $x = 1$ is

$$p(x) = 1 - 2(x-1) + 3(x-1)^2 - 4(x-1)^3$$

We can approximate $1/(0.9)^2$ as

$$p(0.9) = 1 + (.9-1)\cdot(-2 + (.9-1)\cdot(3 + (.9-1)\cdot(-4))) = 1.2340$$

The remainder is $0.0005/0.9^5 < 0.00085$.

The exact value is 1.2346.

The error is 0.0006.