Introduction to Numerical Analysis I
Handout 3

1 The Error

Let $\varepsilon = 0.0.....01$ a (smallest) difference between two successive real numbers on computer and let

$$r < \alpha < r + \varepsilon.$$

What should represent α on computer r or $r + \varepsilon$? There is two possible options:

- Chopping/Truncation of the lest significant digits - fast, but less accurate.
- Rounding to the closest number

Different computer system may use either of the option, but we, usually, won't distinct between chopping and rounding and use the term *round-off error*.

Definition 1.1 (Error). Denote x an approximate solution and denote $\bar{x}$ the exact solution. One says that $\bar{x} = x + \text{Error}$, thus one defines

$$\text{Error} = \delta x = \bar{x} - x$$

however we often interesting in the magnitude of the error only, thus we define an absolute error.

Definition 1.2 (Absolute Error). If x approximates $\bar{x}$, then

$$\text{Absolute Error} = |\delta x| = |\bar{x} - x|$$

Definition 1.3 (Relative Error). If x approximates $\bar{x}$, then

$$\text{Relative Error} = \frac{|\delta x|}{|\bar{x}|} = \frac{|\bar{x} - x|}{|\bar{x}|}$$

Example 1.4. If r represents $\bar{r}$ on computer with m-bits mantissa, then

The absolute error is bounded by

$$|\delta r| = |\bar{r} - r| \le 2^{e-(m+1)}$$

The relative errors is bounded by

$$\frac{|\delta r|}{|\bar{r}|} = \frac{|\bar{r} - r|}{|r|} \le \frac{2^{e-(m+1)}}{2^e \cdot 1/2} = 2^{-m} = 2^{-52} \approx 2{\cdot}10^{-16}$$

2 Numerical Stability

A numerical problem may often be solved using more then one algorithm. Consider only the algorithms that yield the same accuracy. Some algorithms have sensitivity to a small computational errors. These computational errors are approximation errors that arise from round-off error and propagate through prior computations.

Definition 2.1. An algorithm considered numerically stable if it can be proven that it is not magnify approximation error.

2.1 Loss of Significant Digits

A key mechanism for amplification of error is due to the loss of significant digits, which is purely computer related phenomena which occurs because numbers are represented by finite number of digits. See the following example.

$$a - b = 0.a_1a_2...a_n - 0.b_1b_2...b_n = 0.0..0c_kc_{k+1}...c_n$$

$$= 1.c_{k+1}c_{k+2}...c_n\,\overbrace{c_{n+1}...c_{n+k}}^{\text{artifacts}}$$

The only solution is to avoid subtraction between close numbers, however it could be very tricky and it is easy to fall apart.

2.2 Condition number

Condition number is a property of a problem that is quantitatively describe how the problem is "bad" or difficult.

Definition 2.2. For a problem described by equation $F(d, x) = 0$ the condition number is defined as the relation between change in input d and the output $x = x(d)$. The absolute condition number defined as

$$\mathbf{K}(d) = \sup_{\delta d \neq 0} \left\{ \frac{||\delta x||}{||\delta d||} \right\} \approx x'(d) = -\frac{F_d}{F_x}$$

The relative condition number defined as $(x, d \neq 0)$

$$\mathbf{K}(d) = \sup_{\delta d \neq 0} \left\{ \frac{||\delta x||/||x||}{||\delta d||/||d||} \right\}$$