1. INTRODUCTION TO NUMERICAL METHODS

Analytical methods are used extensively to solve many mathematical and engineering problems. These methods give results in terms of mathematical functions whose behaviour and properties are often apparent. However, many practical engineering problems are so complex that analytical solutions cannot be obtained or the cost or effort of performing an analytical solution would be prohibitive. For example, the length of the curve

$$y = \sin x$$

from $x = 0$ to $x = \pi$ can be found analytically by solving the definite integral

$$s = \int_0^{\pi} \sqrt{1 + \cos^2 x}\, dx.$$

But this integral is "not easy" to evaluate.

Numerical methods have most of the following characteristics:

1. The solution procedure is iterative, with the accuracy of the estimated solution improving with each iteration.
2. The solution procedure provides only the approximation to the true, but unknown, solution.
3. An initial estimate of the solution may be required.
4. The solution procedure is conceptually simple, with algorithms representing the solution procedure that can be easily programmed on a digital computer.
5. The solution procedure may occasionally diverge from rather than converge to the true solution.

The next example illustrates the characteristics of a numerical method.

Example 1. 1.
Suppose we wish to estimate $\sqrt{x}$, where x is a positive real number. Let $x_0 + h_0 = \sqrt{x}$ and assume that h_0 is "small". Then

$$x = (x_0 + h_0)^2 = x_0^2 + 2x_0h_0 + h_0^2.$$

Since h_0 is small, h_0^2 is much smaller so that

$$h_0 \approx \frac{x - x_0^2}{2x_0}. \tag{1.1.1}$$

Hence, we have the following approximations:

$$x_1 = x_0 + h_0 \ldots\ldots\ldots\ldots\ldots\ldots 1^{st} \text{ iteration}$$

$$x_2 = x_1 + h_1 \ldots\ldots\ldots\ldots\ldots\ldots 2^{nd} \text{ iteration}$$

$$\vdots$$

$$x_n = x_{n-1} + h_{n-1} \ldots\ldots\ldots\ldots\ldots\ldots n^{th} \text{ iteration}$$

where each h_i, $i = 1, 2, \ldots, n-1$, is approximated by using (1.1.1).

For example, if $x = 150$, then we can choose $x_0 = 12$ so that

$$h_0 \approx \frac{x - x_0^2}{2x_0} = \frac{150 - 12^2}{2(12)} = \frac{150 - 144}{24} = 0.25$$

$$\Rightarrow x_1 = x_0 + h_0 = 12 + 0.25 = 12.25$$

$$h_1 \approx \frac{150 - (12.25)^2}{2(12.25)} = -0.00255$$

$$\Rightarrow x_2 = x_1 + h_1 = 12.25 - 0.00255 = 12.24745$$

$$h_2 \approx \frac{150 - (12.24745)^2}{2(12.24745)} = -0.000001286$$

$$\Rightarrow x_3 = x_2 + h_2 = 12.24745 - 0.000001286 = 12.24744871.$$

$$\vdots$$

Using a calculator, the actual value is 12.24744871.

Although numerical methods have many advantages, they also have disadvantages. The main disadvantages are:

- They are solved iteratively and, thus, more computation time is required.
- An exact solution may not be found.
- Initial estimates of the solution are often required.

Thus, any numerical method is judged by how reliable it is, how fast it is and the cost of doing one iteration. Since the solution to a problem by numerical methods is not exact, error analysis and error estimation are often necessary.

1.1 Analysis of Numerical Errors

An error in estimating or determining a quantity of interest can be defined as a deviation from its known true value. In numerical analysis, we usually encounter the following types of errors:

(i) **Inherent Error:** This error is caused by using data which are approximate or due to limitations of the computing aid such as a calculator or computer.

(ii) **Truncation Error:** This error is caused by using approximate formulae in computation. For example, when a function $f(x)$ is evaluated from an infinite series for x after 'truncating' it at a certain stage, a truncation error does arise.

(iii) **Round-off Error:** This is a type of error of inherent error which arises because the arithmetic performed in a machine (computer or calculator) involves numbers with only a finite number of digits resulting in many calculations being performed with an approximate representation of the actual numbers.

(iv) **Propagated Error:** This is the error which arises because there was an error in the proceeding steps of a process (cumulative).

If an error stays at one point in a process and does not aggregate further as the calculation continues, then it is considered a numerically stable error. This happens when the error causes only a very small variation in the formula result. If the opposite occurs and the error propagates bigger as the calculation continues, then it is considered numerically unstable.

The following definitions describe methods of measuring approximate error:

Definition 0.1.
Suppose that P^* is an approximation to P. The absolute error is given by

$$|P - P^*|,$$

the relative error is given by

$$\frac{|P - P^*|}{|P|}$$

and the percentage error is given by

$$\frac{|P - P^*|}{|P|} \times 100\%.$$

In real world applications, true value is generally not known in which case the value is replaced by the best available estimate of the true value, i.e.

$$\text{Relative} = \frac{|Present\ Approximation - Previous\ Aproximation|}{|Present\ Approximation|}.$$

NOTE: The relative error is generally a better measure of accuracy than the absolute error because it takes into consideration the size of the number being approximated.

Example 1. 2.

1. *Determine the absolute error, relative error and percentage error when approximating P by P^* when*

   (a) $P = 0.3000 \times 10^1$ *and* $P^* = 0.3100 \times 10^1$

   (b) $P = 0.3000 \times 10^{-3}$ *and* $P^* = 0.3100 \times 10^{-3}$

   (c) $P = 0.3000 \times 10^4$ *and* $P^* = 0.3100 \times 10^4$

2. *Three approximate values of* $P = \frac{1}{3}$ *are given as* $0.30,\ 0.33$ *and* 0.34. *Which of these is the best approximation?*

Solutions

1. **(a)** *Absolute Error* $= |P - P^*| = \left|0.3000 \times 10^1 - 0.3100 \times 10^1\right| = 0.1$

   $$\textit{Relative Error} = \frac{|P - P^*|}{|P|} = \frac{\left|0.3000 \times 10^1 - 0.3100 \times 10^1\right|}{|0.3000|} = 0.\overline{3} \times 10^{-1}$$

   and

   $$\textit{Percentage Error} = \frac{|P - P^*|}{|P|} \times 100\% = 0.\overline{3} \times 10^{-1} \times 100\% = 3.\overline{3}\%$$

   (b) *Absolute Error* $= |P - P^*| = \left|0.3000 \times 10^{-3} - 0.3100 \times 10^{-3}\right| = 0.00001$

Relative Error $= \dfrac{|P - P^*|}{|P|} = \dfrac{0.00001}{0.3000 \times 10^{-3}} = 0.0\overline{3}$

and

Percentage Error $= \dfrac{|P - P^*|}{|P|} \times 100\% = 0.0\overline{3} \times 100\% = 3.\overline{3}\%$

(c) *Exercise*

2. *Using relative error, we get*

$$\frac{|P - P^*|}{|P|} = \frac{\left|\frac{1}{3} - 0.30\right|}{\left|\frac{1}{3}\right|} = 0.1$$

$$\frac{|P - P^*|}{|P|} = \frac{\left|\frac{1}{3} - 0.33\right|}{\left|\frac{1}{3}\right|} = 0.01$$

and

$$\frac{|P - P^*|}{|P|} = \frac{\left|\frac{1}{3} - 0.34\right|}{\left|\frac{1}{3}\right|} = 0.02$$

Therefore, we conclude that 0.33 *with the smallest relative error is the best approximation.*

Definition 0.2.

The number P^* is said to approximate P to t significant digits (or figures) if t is the largest non-negative integer for which

$$\frac{|P - P^*|}{|P|} \leq 5 \times 10^{-t}.$$

Example 1. 3.

1. *The following numbers are accurate to t significant figures. Find t in each case.*

   (a) 1036.52 ± 0.01

   (b) 9.321 ± 0.1

   (c) 0.05 ± 0.01

2. *Estimate* $e^{0.5}$ *to* 3 *significant figures using*

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}$$

Solutions:

1. **(a)** *The absolute error is* 0.01 *implying that*

$$\frac{|P - P^*|}{|P|} = \frac{0.01}{1036.52} = 0.000009647667194 < 0.00005$$

$$\therefore t = 5$$

(b) $\frac{|P - P^*|}{|P|} = \frac{0.1}{9.321} = 0.010728462 < 0.05$

$$\therefore t = 2$$

(c) *Exercise*

2. *We require* P_n *such that*

$$Relative\ Error \leq 5 \times 10^{-3} = 0.005$$

$P_0 = 1,\ P_1 = 1 + 0.5 = 1.5$

$$\therefore Relative\ Error = \frac{|Present\ Approximation - Previous\ Aproximation|}{|Present\ Approximation|}$$

$$= \frac{|1.5 - 1|}{|1.5|} = 0.33 > 0.005.$$

$P_2 = 1 + 0.5 + \frac{(0.5)^2}{2} = 1.625$

$$\Rightarrow Relative\ Error = \frac{|1.625 - 1.5|}{|1.625|} = 0.083 > 0.005.$$

$P_3 = 1 + 0.5 + \frac{(0.5)^2}{2} + \frac{(0.5)^3}{3!} = 1.6458$

$$\Rightarrow Relative\ Error = \frac{|1.6458 - 1.625|}{|1.6458|} = 0.012 > 0.005.$$

$$P_4 = 1 + 0.5 + \frac{(0.5)^2}{2} + \frac{(0.5)^3}{3!} + \frac{(0.5)^4}{4!} = 1.64844$$

$$\Rightarrow \textit{Relative Error} = \frac{|1.64844 - 1.6458|}{|1.64844|} = 0.0015 < 0.005.$$

$$e^{0.5} \approx P_4 = 1.64844$$

to 3 significant figures.

1.2 Taylor Series Expansion

A Taylor series is commonly used as a basis of approximation in numerical analysis.

Theorem 1. 1.
Suppose that $f \in C^n[a,b]$, that $f^{(n+1)}$ exists on $[a,b]$ and $x_0 \in [a,b]$. For every $x_0 \in [a,b]$, there exists a number $\xi(x)$ between x_0 and x with

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

where

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

$$= \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$$

and

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

In Theorem 1.1, $P_n(x)$ is called the n^{th} Taylor polynomial for f at x_0 and $R_n(x)$ is called the remainder term associated with $P_n(x)$. The number $\xi(x)$ depends on the value of x at which the polynomial is being evaluated. It cannot be evaluated explicitly but it lies between x and x_0. The infinite series obtained by taking the limit of $P_n(x)$ as $n \to \infty$ is called the Taylor series of f at x_0. In a case where $x_0 = 0$, the series is called Maclaurin series.

Example 1. 4.

(a) *Find the second Taylor polynomial for* $f(x) = e^x \cos x$ *about* $x_0 = 0$

(b) *Use part (a) to approximate*

(i) $f(0.5)$ ***(ii)*** $\int_0^1 f(x)\,dx$

(c) *Find upper bounds of the errors in part (b).*

Solutions

(a) $f(x) = e^x \cos x \Rightarrow f(0) = 1$

$f'(x) = e^x \cos x - e^x \sin x \ \Rightarrow f'(0) = 1$

$f''(x) = e^x \cos x - e^x \sin x - (e^x \sin x + e^x \cos x) = -2e^x \sin x \ \Rightarrow f''(0) = 0$

$f'''(x) = -2(e^x \sin x + e^x \cos x)$

$\therefore f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + R_2(x)$

$= f(0) + f'(0)(x - 0) + \frac{f''(0)}{2!}(x - 0)^2 + R_2(x)$

$= 1 + 1(x) + \frac{0}{2}(x)^2 + R_2(x),$

where

$$R_2(x) = \frac{-2e^{\xi(x)}(\sin \xi(x) + \cos \xi(x))}{3!}x^3, \quad 0 < \xi(x) < x.$$

$$\therefore P_2(x) = 1 + x$$

(b) **(i)** $f(0.5) \approx P_2(0.5) = 1 + 0.5 = 1.5$

(ii) $$\int_0^1 f(x)\,dx = \int_0^1 (1 + x)\,dx = \left(x + \frac{x^2}{2}\right)\Big|_0^1 = 1.5$$

(c) **(i)** *The truncation error associated with* $P_n(x)$ *is*

$$R_n(x) = f(x) - P_n(x) = \frac{-2e^{\xi(x)}(\sin \xi(x) + \cos \xi(x))}{3!}x^3.$$

Therefore, an upper bound, for $0 < \xi(x) < 0.5$, *is*

$$|R_n(x)| = \left|\frac{-2e^{\xi(x)}(\sin \xi(x) + \cos \xi(x))}{3!}x^3\right| \leq \frac{1}{3}(0.5)^3 \max_{\xi(x)\in[0,0.5]} |e^{\xi(x)}(\sin \xi(x) + \cos \xi(x))|.$$

To maximise $g(x) = e^{\xi(x)}(\sin \xi(x) + \cos \xi(x))$, *we notice that*

$$g' = e^{\xi(x)}(\sin \xi(x) + \cos \xi(x)) + e^{\xi(x)}(\cos \xi(x) - \sin \xi(x))$$

$$= 2e^{\xi(x)} \cos \xi(x) > 0, \text{ for } \xi(x) \in [0, 0.5].$$

Thus, $g(0) = 1$ *and* $g(0.5) = e^{0.5}(\sin 0.5 + \cos 0.5) \approx 2.24$

$$\therefore |R_n(x)| \leq \frac{1}{3}(0.5)^3(2.24) \approx 0.0932.$$

(ii) *An upper bound of the error is*

$$\left| \int_0^1 R_2(x)\,dx \right| = \left| \int_0^1 f(x)\,dx - \int_0^1 P_2(x)\,dx \right|$$

$$\leq \int_0^1 |R_2(x)|\,dx$$

$$\leq \int_0^1 \frac{1}{3}e^1(\sin 1 + \cos 1)\,x^3\,dx, \text{ since } 0 < \xi(x) < 1$$

$$\leq 1.252 \int_0^1 x^3\,dx = (1.252)\left(\frac{x^4}{4}\right)\Big|_0^1 = 0.313$$

THE END!