MODULE 3

UNIT 1 REVIEW OF CALCULUS

CONTENTS

NUMERICAL ANALYSIS

Mathematical modeling of physical/biological problems generally gives rise to ordinary or partial differential equations or an integral equation or in terms of a set of such equation. A number of these problems can be solved exactly by mathematical analysis but most of them cannot be solved exactly. Thus, a need arises to devise numerical methods to solve these problems. These methods for solution of mathematical methods may give rise to a system of algebraic equations or a non-linear equation or system of non-linear equations. The numerical solution of these systems of equations is quantitative in nature but when interpreted give qualitative results and are very useful. Numerical analysis deals with the development and analysis of the numerical methods. We are offering this course of numerical analysis to students entering the Bachelor's Degree Programme as an elective subject.

It was in the year 1624 that the English mathematician, Henry Briggs used a numerical procedure to construct his celebrated table of logarithms. The interpolation problem was first taken up by Briggs but was solved by the 17th century mathematicians and physicists, Sir Isaac Newton and James Gregory. Later on, other problems were considered and solved by more and more efficient methods. In recent years the invention and development of electronic calculators/computers have strongly influenced the development of numerical analysis.

This course assumes the knowledge of the course MTH 112, MTH 122. They are prerequisite for this course. Number of results from linear algebra are also used in this course. These results have been stated wherever required. For details of these results our linear algebra course MTH 121 may be referred. This course is divided into 4 blocks. The first block, deals with the problem of finding approximate roots of a non-linear equation in one unknown. We have started the block with a recall of four important theorems from calculus which are referred to throughout the course. After introducing the concept of 'error' that arise due to approximations, we have discussed two basic approximation methods namely, bisection and fixed point iteration methods and two commonly used methods, namely. secant and Newton-Raphson methods. In Block 2, we have considered the problem of finding the solution of system of linear equations. We have discussed both direct and iterative methods of solving system of linear equations.

Block 3 deals with the theory of interpolation. Here, we are concerned only with polynomial interpolation. The existence and uniqueness of interpolating polynomials are discussed. Several form of interpolating polynomials like Lagrange's and Newton's divided difference forms with error terms are discussed. This block concludes with a discussion on Newton's forward and backward difference form.

In Block 4, using interpolating polynomials we have obtained numerical differentiation and integration formulae together with their error terms. After a brief introduction to difference equations the numerical solution of the first order ordinary differential equation is dealt with. More precisely, Taylor series, Euler's and second order Runge Kutta methods are derived with error terms for the solution of differential equations.

Each block consists 4 units. All the concepts given in the units are followed by a number of examples well as exercises. These will help you get a better grasp of the techniques discussed in this course. We have used a scientific calculator for doing computations throughout the course. While attempting the exercises given in the units, you would also need a calculator which is available at your study centre. The solutions/answers to the exercises in a unit are given at the end of the unit. We suggest that you look at them only after attempting the exercises. A list of symbols and notations are also given in for your reference.

You ma like to look up some more books on the subject and try to solve some exercises given in them. This will help you get a better grasp of the techniques discussed in this course. We are giving you a list of titles which will be available in your study centre for reference purposes.

Some useful books

Wrede, R.C. and Spegel M. (2002). Schaum's and Problems of Advanced Calculus. McGraw – Hill N.Y.

Keisler, H.J. (2005). Elementary Calculus. An Infinitesimal Approach. 559 Nathan Abbott, Stanford, California, USA

NOTATION AND SYMBOLS

Ì	belong to
É	contains
< (£)	less than (less than or equal to)
>(³)	greater than (greater than or equal to)
R	set of real numbers
C	set of complex numbers
n!	n(n-1) ... 3. 2. 1 (n factorial)
[ ]	closed interval
] [	open interval
|x|	absolute value of a number x
i.e.	that is
$\overset{n}{\underset{j=1}{\text{å}}}\ a_i$	$a_1 + a_2 + ... + a_n$
x® a	x tends to a
$\underset{x\text{®}a}{\lim} f(x)$	limit of f(x) as x tends to a
$P_n(x)$	nth degree polynomial
f'(x)	derivative of f(x) with respect to x
»	approximately equal to
a	alpha
b	beta
g	gama
e	epilson
p	pi
å	capital sigma
z	zeta

BLOCK INTRODUCTION

This is the first of the four blocks which you will be studying in the Numerical Analysis course. In this block we shall be dealing with the problem of finding approximate roots of a non-linear equation in one unknown. In the Elementary Algebra course you have studied some methods for solving polynomial equations of degree up to and including four. In this block we shall introduce you to some numerical methods for finding solutions of equation. These methods are applicable to polynomial and transcendental equations.

This block consists of four units. In Unit 1, we begin with a recall of our important theorems from calculus which are referred to throughout the course. We then introduce you to the concept of 'error' that arise due to approximation. In Unit 2, we shall discuss two types of errors that are common in numerical approximation methods, namely, bisection method and fixed point iteration method. Each of these methods involve a process that is repeated until an answer or required accuracy is achieved. These methods are known as iteration methods. We shall also discuss two accurate methods, namely, secant and Newton-Raphson methods in Unit 3. Unit 4, which is the last unit of this block, deals with the solutions of the most well-known class of equations, the polynomial equations. For finding the roots of polynomial equations we shall discuss Birge-Vieta and Graeffe's root squaring methods.

As already mentioned in the course introduction, we shall be using a scientific calculator for doing computations throughout the block. While attempting the exercises given in this block, you would also need a calculator which is available at your centre. We therefore suggest you to go through the instructions manual, supplied with the calculator, before using it.

Lastly we remind you to through the solved examples carefully, and to attempt all exercises in each unit. This will help you to gain some practice over various methods discussed in this block.

1.0 INTRODUCTION

The study of numerical analysis involves concepts from various branches of mathematics including calculus. In this unit, we shall briefly review certain important theorems in calculus which are essential for the development and understanding of numerical methods. You are already familiar with some fundamental theorems about continuous functions from your calculus course. Here we shall review three theorems given in that course, namely, intermediate value theorem, Rolle's Theorem and Lagrange's mean value theorem. Then we state another important theorem in calculus due to B. Taylor and illustrate the theorem through various examples.

Most of the numerical methods give answers that are approximation to the desired solutions. In this situation, it is important to measure the accuracy of the approximate solution compared to the actual solution. To find the accuracy we must have an idea of the possible errors that can arise in computational procedures. In this unit we shall introduce you to different forms of errors which are common in numerical computations.

The basic ideas and result that we have illustrated in this unit will be used often throughout this course. So we suggest you go through this unit very carefully.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- apply
  - Intermediate value theorem
  - Rolle's Theorem
  - Lagrange's mean value theorem
  - Taylor's theorem
- define the term 'error' in approximation
- distinguish between rounded-off error and truncation error and calculate these errors as the situation demands.

3.0 MAIN CONTENT

3.1 Three Fundamental Theorems

In this section we shall discuss three fundamental theorems, namely, intermediate value theorem, Rolle's Theorem and Lagrange's mean value theorem. All these theorems give properties of continuous functions defined on a closed interval [a, b]. we shall not prove them here, but we shall illustrate their utility with various examples. Let us take up these theorems one by one.

3.1.1 Intermediate Value Theorem

The intermediate value theorem says that a function that is continuous on a closed interval [a, b] takes on every intermediate value i.e., every value lying between f(a) and f(b) if $f(a) \neq f(b)$.

Formally, we can state the theorem a follows:

Theorem 1: let f be a function defined on a closed interval [a, b]. let c be a number lying between f(a) and f(b) (i.e. $f(a) < c < f(b)$ if $f(a) < f(b)$ or $f(b) < c < f(a)$ if $f(b) < f(a)$). Then there exists at least one point $x_0 \in [a, b]$ such that $f(x_0) = c$.

The following figure (Fig. 1) may help you to visualise the theorem more easily. It gives the graph of a function f.

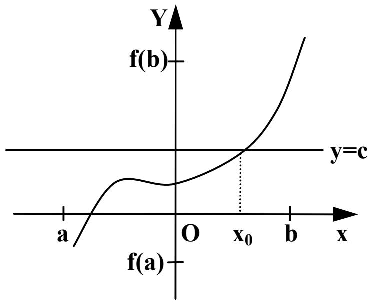

Fig. 1

In this figure f(a) < f(b). the condition f(a) < c < f(b) implies that the points (a, f(a)) and (b, f(b)) lie on opposite sides of the line y = c. This, together with the fact that f is continuous, implies that the graph crosses the line y = c at some point. In Fig. 1 you see that the graph crosses the line y = c at (x_0, c).

The importance of this theorem is as follows: If we have a continuous function f defined on a closed interval [a, b], then the theorem guarantees the existence of a solution of the equation f(x) = c, where c is as in Theorem 1. However, it does not say what the solution is. We shall illustrate this point with an example.

Example 1: Find the value of x in 0 £ x £ $\frac{p}{2}$ for which sin (x) = $\frac{1}{2}$.

Solution: You know that the function f(x) = sin x is continuous on $\left[0, \frac{P}{2}\right]$. Since f(0) = 0 and $f\left(\frac{p}{2}\right)$ = 1, we have f(0) < $\frac{1}{2}$ < $f\left(\frac{p}{2}\right)$. Thus, f satisfies all the conditions of Theorem 1. Therefore, there exists at least one value of x, say x_0 such that sin (x_0) = $\frac{1}{2}$, that is, the theorem guarantees that there exists a point x_0 such that sin (x_0) = $\frac{1}{2}$. Let us try to find this point from the graph of sin x in (see Fig. 2).

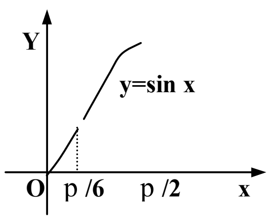

Fig. 2

From the figure, you can see that the line $x = \frac{1}{2}$ cuts the graph at the point $\left(\frac{p}{6}, \frac{1}{2}\right)$. Hence there exists a point $x_0 = \frac{p}{6}$ in $\left[0, \frac{P}{2}\right]$ such that sin $((x_0) = \frac{1}{2}$.

Let us consider another example.

Example 2: Show that the equation $2x^3 + x^2 - x + 1 = 5$ has a solution in the interval [1, 2].

Solution: Let $f(x) = 2x^3 + x^2 - x + 1$. Since f is a polynomial in x, f is continuous in [1, 2]. Also f(1) = 3, f(2) = 19 and 15 lies between f(1) and f(2). Thus f satisfied all conditions of Theorem 1. Therefore, there exists a number x_0 between 1 and 2 such that $f(x_0) = 5$. That is, the equation $2x^3 + x^2 - x + 1 = 5$ has solution in the interval [1, 2].

Thus we saw that the theorem enables us in establishing the existence of the solutions of certain equations of the type f(x) = 0 without actually solving them. In other words, if you want to find an interval in which a solution (or root) of f(x) = 0 exists, then find two numbers a, b such that f(a) f(b) < 0. Theorem 1, then states that the solution lies in]a, b[. We shall need some other numerical methods for finding the actual solution. We shall study the problem of finding solution of the equation f(x) = 0 more elaborately in Unit 2.

Let us now discuss another important theorem in calculus.

3.1.2 Rolle's Theorem

In this section we shall review the Rolle's Theorem. The theorem is named after the seventeenth century French mathematician Michel Rolle (1652 – 1719).

Theorem 2 (Rolle's Theorem): Let f be a continuous function defined on [a, b] and differentiable on]a, b[. If f(a) = f(b), then there exists a number x_0 in]a, b[such that $f'(x_0) = 0$.
Geometrically, we can interpret the theorem easily. You know that since f is continuous, the graph of f is a smooth curve (see Fig. 3).

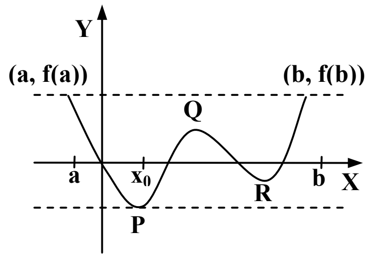

Fig. 3

You have already seen in your calculus course that the derivative $f'(x_0)$ at some point x_0 gives the slope of the tangent at $(x_0, f(x_0))$ to the curve $y = f(x)$. Therefore the theorem states that if the end values f(a) and f(b) are equal, then there exists a point x_0 in]a, b[such that the slope of the tangent at the point $P(x_0, f(x_0))$ is zero, that is, the tangent is parallel to x-axis at the point (see Fig. 3). In fact we can have more than one point at which $f'(x) = 0$ as shown in Fig. 3. This shows that the number x_0 in Theorem 2 may not be unique.

The following example give an application of Rolle's Theorem.

Example 3:Use Rolle's Theorem to show that there is a solution of the equation cot x = x in $[0, \frac{\pi}{2}]$.

Solution: Here we have to solve the equation $\cot x - x = 0$. We rewrite $\cot x - x$ as $\frac{\cos x \quad x \sin x}{\sin x}$. Solving the equation $\frac{\cos x \quad x \sin x}{\sin x} = 0$ in $[0, \frac{\pi}{2}]$ is sme as solving the equation $\cos x - x \sin x = 0$. now we shall see whether we can find a function f which satisfies the conditions of Rolle's Theorem and for which $f'(x) = \cos x - x \sin x$. Our experience in differentiation suggests that we try $f(x) = x \cos x$. this function f is continuous in $]0, \frac{\pi}{2}[$, differentiable in $[0, \frac{\pi}{2}]$ and the derivative $f'(x) = \cos x - x \sin x$. Also $f(0) = 0 = f(\frac{\pi}{2})$. Thus f satisfies all the requirements of Rolle's Theorem. Hence, there exists a point x_0 in [a, b] such that $f'(x_0) = \cos x_0 - x_0 \sin x_0 = 0$. This shows that a solution to the equation $\cot x - x = 0$ exists in $[0, \frac{\pi}{2}]$.

Now, let us look at Fig. 3 carefully. We see that the line joining (a, f(a)) and (b, f(b)) is parallel to the tangent at $(x_0, f(x_0))$. Does this property hold when $f(a) \neq f(b)$ also? In other words, does there exists a point x_0 in]a, b[such that the tangent at $(x_0, f(x_0))$ is parallel to the line joining (a, f(a)) and (a, f(b))? The answer to this question is the content of the well-known theorem. "Lagrange's mean value theorem", which we discuss next.

3.1.3 Lagrange's Mean Value Theorem

This theorem was first proved by the French mathematician Count Joseph Louis Lagrange (1736 – 1813).

Theorem 3: Let f be a continuous function defined on [a, b] and differentiable in]a, b[. Then there exists a number x_0 in [a, b] such that

$$f'(x_0) = \frac{f(b) - f(a)}{b - a} \qquad (1)$$

geometrically we can interpret this theorem as given in Fig. 4.

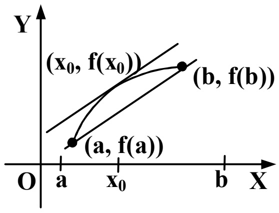

Fig. 4

In this figure you can see that the straight line connecting the end points (a, f(a)) and (b, f(b)) of the graph is parallel to some tangent to the curve at an intermediate point.

You may be wondering why this theorem is called 'mean value theorem'. This is because of the following physical interpretation.

Suppose f(t) denotes the position of an object at time t. Then the average (mean) velocity during the interval [a, b] is given by

$$\frac{f(b) - f(a)}{b - a}$$

Now Theorem 3 states that this mean velocity during an interval [a, b] is equal to the velocity $f'(x_0)$ at some instant x_0 in [a, b].

We shall illustrate the theorem with an example.

Example 4: Apply the mean value theorem to the function $f(x) = \sqrt{x}$ in [0, 2] (see Fig. 5).

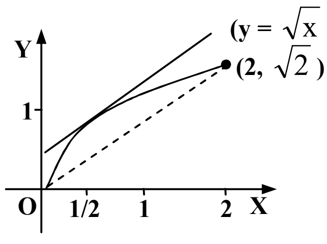

Fig. 5: Graph of $f(x) = \sqrt{x}$

Solution: We first note that the function $f(x) = \sqrt{x}$ is continuous on [0, 2] and differentiable in [0, 2] and $f'(x) = \frac{1}{2\sqrt{x}}$.

Therefore by Theorem 3, there exists a point x_0 in]0, 2[such the f(2) = $\sqrt{2}$ and f(0) = f'(x_0) (2 – 0)

Now f(2) = $\sqrt{2}$ and f(0) = 0 and f'(x_0) = $\frac{1}{2x_0}$.

Therefore we have

$\sqrt{2} = \frac{1}{x_0}$

i.e. $\frac{1}{x_0} = \frac{1}{\sqrt{2}}$ and $x_0 = \frac{1}{2}$.

Thus we get that the line joing the end points (0, 0) and (2, $\sqrt{2}$) of the graph of f is parallel to the tangent to the curve at the point ($\frac{1}{2}$, $\frac{1}{\sqrt{2}}$).

We shall consider one more example.

Example 5: Consider the function f(x) = (x – 1) (x – 2) (x – 3) in [0, 4]. Find a point x_0 in]0, 4[such that

$$f'(x_0) = \frac{f(4) - f(0)}{4 - 0}.$$

Solution: We rewrite the function f(x) as

$$f(x) = (x - 1)(x - 2)(x - 3) = x^3 - 6x^2 + 11x - 6$$

we know that f(x) is continuous on [0, 4], since f is a polynomial in x. Also the derivative

$$f'(x) = 3x^2 - 12x = 11$$

exists in]0, 4[. Thus f satisfies all conditions of the mean value theorem. Therefore, there exists a point x_0 in [0, 4] such that

$$f'(x_0) = \frac{f(4) - f(0)}{4 - 0}$$

i.e., $3x_0^2 - 12x_0 + 11 = \frac{6 + 6}{4 - 0} = 3$

i.e., $3x_0^2 - 12x_0 + 8 = 0$

This is a quadratic equation in x_0. The roots of this equation are

$\frac{6+2\sqrt{3}}{8}$ and $\frac{6-2\sqrt{3}}{8}$

Taking $\sqrt{3}$ = 1.732, we see that there are twp values for x_0 lying in the interval [0, 4].

The above example shows that the number x_0 in Theorem 3 may not be unique. Again, as we mentioned in the case of theorems 1 and 2, the mean value theorem guarantees the existence of a point only.

So far we have used the mean value theorem to show the existence of appoint satisfying Eqn. 1. Next we shall consider an example which shows another application of mean value theorem.

Example 6: Find an approximate value of $\sqrt[3]{26}$ using the mean value theorem.

Solution: Consider the function $f(x) = x^{1/3}$. Then $f(26) = \sqrt[3]{26}$. The number nearest to 26 for which the cube root is known is 27, i.e., $f(27) = \sqrt[3]{27} = 3$. Now we shall apply the mean value theorem to the function $f(x) = x^{1/3}$ in the interval [26, 27]. The function f is continuous in [26, 27] and the derivative is

$$f'(x) = \frac{1}{3x^{2/3}}$$

Therefore, there exists a point x_0 between 26 and 27 such that

$$\sqrt[3]{27} - \sqrt[3]{26} = \frac{1}{3x_0^{2/3}}(27-26)$$

$$\text{i.e., } \sqrt[3]{26} = 3 - \frac{1}{3x_0^{2/3}} \qquad (2)$$

Since x_0 is close to 27, we approximate $\frac{1}{3x_0^{2/3}}$ by $\frac{1}{3(27)^{2/3}}$, i.e.;

$$\frac{1}{3x_0^{2/3}} \approx \frac{1}{27}$$

Substituting this value in Eqn. (2) we get

$$\sqrt[3]{26} = 3 - \frac{1}{27} = 2.963.$$

Note that in writing the value of we have rounded off the number after three decimal places. Using the calculator we find that the exact value of $\sqrt[3]{26}$ is 2.9624961.

We have given this example just to illustrate the usefulness of the theorem. The mean value theorem has got many other applications which you will come across in later units.

Now we shall discuss another theorem in calculus.

3.2 Taylor's Theorem

You are already familiar with the name of the English mathematician Brook Taylor (1685 – 1731) from your calculus course. In this section we shall introduce you to a well-known theorem due to B. Taylor. Here we shall state the theorem without proof and discuss some of its applications.

You are familiar with polynomial equations of the form $f(x) = a_0 + a_1 x + ... + a_n x^n$ where $a_0, a_1 ....., a_n$ are real numbers. We can easily compute the value of a polynomial at any point x = a by using the four basic operation of addition, multiplication, subtraction and division. On the other hand there are function like e^x, cos x. In x etc. which occur frequently in all branches of mathematics which cannot be evaluated in the same manner. For example, evaluating the function f(x) = cos x at 0.524 is not so simple. Now, to evaluate such functions we try to approximate them by polynomials which are easier to evaluate. Taylor's theorem gives us a simple method for approximating functions f(x) by polynomials.
Let f(x) be a real-valued function defined on R which is n-times differentiable. Consider the function

$P_1(x) = f(x_0) + (x - x_0) f'(x_0)$
where x_0 is any given real number.

Now $P_1(x)$ is a polynomial in x of degree 1 and $P_1(x_0) = f(x_0)$ and $P'_1(x_0) = f'(x_0)$. The polynomial $P_1(x)$ is called the first Taylor polynomial of f(x) at x_0. Now consider another function

$$P_2(x_0) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!} f''(x_0).$$

Then $P_2(x)$ is a polynomial in x of degree 2 and $P_2(x_0) = f(x_0)$, $P'_2(x_0) = f'(x_0)$ and $P''_2(x_0) = f''(x_0)$. $P_2(x)$ is called the second Taylor polynomial of f(x) at x_0.

Similarly, we can define the rth Taylor polynomial of f(x) at x_0 where $1 \leq r \leq n$. The rth Taylor polynomial at x_0 is given by

$$P_r(x) = f(x_0) + (x - x_0) f'(x_0) + \dots + \frac{f^{(n)}(x_0)}{r!} (x - x_0)^r. \qquad (3)$$

You can check that $P_r(x_0) = f(x_0)$, $P'_r(x_0) = f'(x_0)$, ….

$P_r^{(r)}(x_0) = f^{(r)}(x_0)$ (see E6)
Let us consider an example.

Example 7: Find the fourth Taylor polynomial of f(x) = In x about x_0=1.

Solution: The fourth Taylor polynomial of f(x) is given by

$$P_4(x) = f(1) + (x - 1)f'(1) + \frac{(x-1)^2}{2!} f''(1) + \frac{(x-1)^3}{3!} f^{(3)}(1) + \frac{(x-1)^4}{4!} f^{(4)}(1).$$

Now, f(1) = In1 = 0

$f'(x) = \frac{1}{x}$; f'(1) = 1

$f''(x) = (-\frac{1}{x^2})$; f''(1) = -1

$f^{(3)}(x) = \frac{2}{x^3}$; $f^{(3)}(1) = 2$

$f^{(4)}(x) = \frac{-6}{x^4}$; $f^{(4)}(1) = -6$

Therefore, $P_4(x) = (x - 1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4}$

We are now ready to state the Taylor's theorem.

Theorem 4 (Taylor's Theorem): Let f be a real valued function having (n + 1) continuous derivatives on]a, b[for some n $\geq$ 0. Let x_0 be any point in the interval]a, b[. Then for any x $\in$]a, b[, we have

$$f(x) = f(x_0) + \frac{(x - x_0)}{1!} f'(x_0) + \frac{(x - x_0)}{2!} f^{(2)}(x_0) + \ldots$$
$$+ \ldots + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0) + \frac{(x - x_0)^{n+1}}{n+1!} f^{n+1}(c). \qquad (4)$$

where c is point between x_0 and x.

The series given in Eqn. (4) is called the nth Taylor's expansion of f(x) at x_0.

We rewrite Eqn. (4) in the form

$f(x) = P_n(x) + R_{n+1}(x)$

where $P_n(x)$ is the nth Taylor polynomial of f(x) about x_0 and

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

$R_{n+1}(x)$ depends on x, x_0 and n. $R_{n+1}(x)$ is called the remainder (or error) of the nth Taylor's expansion after n + 1 terms.

Suppose we put $x_0 = a$ and $x = a + h$ where $h > 0$, in Eqn (4). Then any point between a and a + h will be of the form $a +/\theta h$, $0 < \theta < 1$.

Therefore, Eqn (4) can be written as

$$f(a+h) = f(a)+h\, f'(a)+\frac{h^2}{2!}\, f''(a)+\ldots+\frac{h^n}{n!}\, f^{(n)}(a)+\frac{h^{n+1}}{n+1!}\, f^{(n+1)}(a+\theta h) \qquad (5)$$

Let us now make some remarks on the Taylor's theorem.

Remark 1: Suppose that the function f(x) in Theorem 4 is a polynomial of degree m. Then $f^{(r)}(x) = 0$ for all r > m. Therefore $R_{n+1}(x) = 0$ for all $n \geq m$. Thus, in this case, the mth Taylor's expansion of f(x) about x_0 will be

$$f(x) = f(x_0) + \frac{(x - x_0)}{1!} f'(x_0) + \ldots + \frac{(x - x_0)^m}{m!} f^{(m)}(x_0).$$

Note that the right hand side of the above equation is simply a polynomial in $(x - x_0)$.

Therefore, finding Taylor's expansion of a polynomial function f(x) about x_0 is the same as expressing f(x) as a polynomial in $(x - x_0)$ with coefficients from R.

Remark 2: Suppose we put $x_0 = a$, $x = b$ and $n = 0$ in Eqn. (4). Then Eqn (4) becomes

$$f(b) = f(a) + f'(c)(b - a)$$

or equivalently

$$f(b) - f(a) = f'(c)\,(b - a)$$

which is the Lagrange's mean value theorem. Therefore we can consider the mean value theorem as a special case of Taylor's theorem.

Let us consider some examples.

Example 8: Expand $f(x) = x^4 - 5x^3 + 5x^2 + x + 2$ in powers of $(x - 2)$.

Solution: The function f(x) is a polynomial in x of degree 4. Hence, derivatives of all orders exists and are continuous. Therefore by Taylor's theorem, the 4^{th} Taylor expansion of f(x) about 2 is given by

$$f(x) = f(2) + \frac{(x-2)}{1!} f'(2) + \frac{(x-2)^2}{2!} f''(2) + \frac{(x-2)^3}{3!} f^{(3)}(2) + \frac{(x-2)^4}{4!} f^{(4)}(2).$$

Here f(2) = 0

$f'(x) = 4x^3 – 15x^2 + 10x + 1$, $f'(2) = -7$

$f''(x) = 12x^2 – 30x + 10$, $f''(2) = -2$

$f^{(3)}(x) = 24x – 30$, $f^{(3)}(2) = 18$

$f^{(4)}(x) = 24$, $f^{(4)}(2) = 24$

Hence the expansion is

$$f(x) = -7(x – 2) - \frac{2(x-2)^2}{2!} + \frac{18(x-2)^3}{3!} + \frac{24(x-2)^4}{4!}$$

$$= -7(x – 2) – (x – 2)^2 + 3(x – 2)^3 + (x – 2)^4$$

Example 9: Find the nth Taylor expansion of 1n (1 + x) about x = 0 for $x \in [-1, 1]$.

Solution: We first note that the point x = 0 lies in the given interval. Further; the function f(x) = 1n (1 + x) has continuous derivatives of all orders. The derivatives are given by

$f'(x) = \frac{1}{1+x}$, $f'(0) = 1$

$f''(x) = \frac{-1}{(1+x)^2}$, $f''(0) = -1$

$f^{(3)}(x) = \frac{(-1)^2 2!}{(1+x)^3}$, $f^{(3)}(0) = 2$

.. .

.. .

.. .

$f^{(n)})x) = \frac{(-1)^{n-1}(n-1)!}{(1+x)^n}$, $f^{(n)}(0) = (-1)^{n-1}(n – 1)!$

Therefore by applying Taylor's theorem we get that for any $x \in [-1, 1]$

$$1n\,(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots + \frac{(-1)^{n-1}x^n}{n} + \frac{(-1)^{n-1}n!\,x^{n+1}}{(n+1)!(1+c)^{n+1}}$$

where c is a point lying between 0 and x.

Now, let us consider the behaviour of the remainder in a small interval, say, [0, 0.5]. then for x in [0, 0.5], we have

$$|R_{n+1}(x)| = \left|\frac{(-1)^n n!\,x^{n+1}}{(n+1)!(1+c)^{n+1}}\right|$$

where 0 < c < x.

Since $|x| < 1$, $|x|^{n+1} < 1$ for any positive integer n.

Also since $c > 0$, $\frac{1}{(1+c)^{n+1}} < 1$. Therefore we have

$|R_{n+1}(x)| < \frac{1}{n+1}$

Now $\frac{1}{n+1}$ can be made as small as we like by choosing n sufficiently large i.e. $\lim_{n\to\infty} \frac{1}{n+1} = 0$. This shows that $\lim_{n\to\infty} |R_{n+1}(x)| = 0$.

The above example shows that if n is sufficient large, the value of the nth Taylor polynomial $P_n(x)$ at any x_0 will be approximately equal to the value of the given function $f(x_0)$. In fact, the remainder $R_{n+1}(x)$ tell(s) us how close the value $P_n(x_0)$ is to $f(x_0)$.

Now we shall make some general observations about the remainder $R_{n+1}(x)$ in the Taylor's expansion of a function f(x).

Remark 3: Consider the nth Taylor expansion of f about x_0 given by
$f(x) = P_n(x) + R_{n+1}(x)$.

Then $R_{n+1}(x) = f(x) - P_n(x)$. If $\lim_{n\to\infty} R_{n+1}(x) = 0$ for some x, then for that x we say that we can approximate f(x) by $P_n(x)$ and we write f(x) as the infinite series.

$$f(x) = f_0(x) + f'(x)(x-x_0) + \frac{f^{(2)}(x_0)}{2!}(x - x_0)^2 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \ldots$$
$$= \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} x^n$$

is called Maclaurin's series.

Remark 4: If the remainder $R_{n+1}(x)$ satisfies the condition that $|R_{n+1}(x)| < M$ for some n at some fixed point x = a, then M is called the bound of the error at x = a.

In this case we have

$|R_{n+1}(x)| = |f(x) - | < M$

That is, f(x) lies in the interval $[P_n(x) - M, P_n(x) + M]$.

Now if M is considerably small for some n, then this interval becomes very small. In this case we say that f(x) is approximately equal to the value of the nth Taylor polynomial with error M. Thus the remainder is used to determine a bound for the accuracy of the approximation.

We shall explain these concepts with an example.

Example 10: Find the 2nd Taylor's expansion of $f(x) = \sqrt{1+x}$ in [-1, 1] about x = 0. find the bound of the error at x = 0.2.

Solution: Since $f(x) = \sqrt{1+x}$, we have

$f(0) = 1$

$f'(x) = \frac{1}{2\sqrt{1+x}}$, $f(0) = \frac{1}{2}$

$f''(x) = -\frac{1}{4}(1+x)^{-3/2}$, $f'(0) = -\frac{1}{4}$

$f^{(3)}(x) = \frac{3}{8}(1+x)^{-5/2}$,

Applying Taylor's theorem to f(x), we get

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3(1+c)^{-5/2}$$

where c is a point lying between 0 and x.

The error is given by $R_3(x) = \frac{x^3}{16}(1+c)^{-5/2}$.

When x = 0.2, we have

$$R_3(0.2) = \frac{(0.2)^3}{16(1+c)^{5/2}}$$

Where 0 < c < 0.2. Since c > 0 we have

$$\left|\frac{1}{(1+c)^{5/2}}\right| < 1.$$

Hence,

$$|R_3(0.2)| \leq \frac{(0.2)^3}{16} = (0.5)\ 10^{-3}$$

Hence the bound of the error for n = 2 at x = 0.2 is $(0.5)\ 10^{-3}$.

There are some functions whose Taylor's expansion is used very often. We shall list their expansion here.

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \frac{x^{n+1}}{(n+1)!}e^c \ldots \qquad (7)$$

$$\text{Sin } x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots + \frac{(-1)^{n-1}x^{2n-1}}{(2n-1)!} + \frac{(-1)^n x^{2n-1}}{(2n+1)!}\cos(c) \quad (8)$$

$$\text{Cos } x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots + \frac{(-1)^n x^{2n}}{(2n)!} + \frac{(-1)^{n+1}x^{2n+2}}{(2n+2)!}\cos(c). \quad (9)$$

$$\frac{1}{1-x} = 1 + x + x^2 + \ldots + x^n + \frac{x^{n+1}}{(1-c)^{n+2}} \quad (10)$$

where c, in each expansion, is as given in Taylor's theorem.

Now, let us consider some examples that illustrate the use of finding approximate values of some functions at certain points using truncated Taylor series.

Example 11: Using Taylor's expansion for sin x about x = 0, find the approximate value of sin 10° with error less than 10^{-7}.

Solution: The nth Taylor's expansion for sin x given in Eqn. (9) is

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots + \frac{(-1)^{n-1}x^{2n-1}}{(2n-1)!} + \frac{(-1)^n x^{2n+1}}{(2n+1)!}\cos(c). \quad (11)$$

where x is the angle measured in radians.

Now, in radian measure , we have

$$10^\circ = \frac{\pi}{18} \text{ radians.}$$

Therefore, by putting $x = \frac{\pi}{18}$ in Eqn. (11) we get

$$\sin\frac{\pi}{18} = \frac{\pi}{18} - \frac{1}{3!}\left(\frac{\pi}{18}\right)^3 + \frac{1}{5!}\left(\frac{\pi}{18}\right)^5 + \ldots + R_{n+1}\left(\frac{\pi}{18}\right)$$

where $R_{n+1}(\frac{\pi}{18})$ is the remainder after (n + 1) terms.

Now

$$R_{n+1}\left(\frac{\pi}{18}\right) = \frac{(-1)^n}{(2n+1)!}\left(\frac{\pi}{18}\right)^{2n+1}\cos c.$$

If we approximate sin $\frac{\pi}{18}$ by $P_n(\frac{\pi}{18})$, then the error introduced will be less than 10^{-7} if

$$\left|\sin\left(\frac{\pi}{18}\right) - P_n\left(\frac{\pi}{18}\right)\right| = \left|R_{n+1}\left(\frac{\pi}{18}\right)\right| = \left|\frac{(-1)^n}{(2n+1)!}\left(\frac{\pi}{18}\right)^{2n+1}\cos c\right| < 10^{-7}.$$

Maximizing cos c, we require that

$$\frac{1}{(2n+1)!}\left(\frac{\pi}{18}\right)^{2n+1} < 10^{-7} \qquad (12)$$

Using the calculator, we find that the value of left hand side of Eqn. (12) for various n is

n	1	2	3
Left hand side	89×10^{-3}	13×10^{-5}	99×10^{-9}

From the table we find that the inequality in (12) is satisfied for n = 3. Hence the required approximation is

$$\sin\left(\frac{\pi}{18}\right) \approx \frac{\pi}{18} - \frac{1}{3!}\left(\frac{\pi}{18}\right)^3 + \frac{1}{5!}\left(\frac{\pi}{18}\right)^5 = 0.1745445$$

with error less than 1.0×10^{-7}.

Let us now find the approximate value of e using Taylor's theorem.

Example 12: Using Maclaurin's series for e^x, show that $e \approx 2.71806$ with error less than 0.001. (Assume that e < 3).

Solution: The Maclaurin's series for e^x is

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

Putting x = 1 in the above series, we get

$$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \ldots$$

Now we have to find n for which

$$|e - P_n(1)| = |R_{n+1}(1)| < 0.001.$$

Now $|R_{n+1}(1)| \le e^c \frac{1}{(n+1)!}$

Since we have chosen $x_0 = 0$ and x = 1, the value c lies between 0 and 1 i.e. 0 < c < 1. Since $e^c < c < 3$, we get

$$|R_{n+1}(1)| \le e^c \frac{3}{(n+1)!}$$

The bound for $R_{n+1}(1)$ for different n is given in the following table.

n	1	2	3	4	5	6
Bounds for R_{n+1}	1.5	.5	.1	.125	.004	.0006

From this table, we see that

$R_{n+1} < .001$ if $n = 6$

Thus $P_6(1)$ is the desired approximation to e. i.e.

$$e \approx 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \frac{1}{720} + \frac{1957}{720} \approx 2.71806$$

In numerical analysis we are concerned with developing a sequence of calculations that will give a satisfactory answer to a problem. Since this process involves a lot of computations, there is a chance for the presence of some errors in these computations. In the next section we shall introduce you to the concept of 'errors' that arise in numerical computations.

3.3 Errors

In this section we shall discuss the concept of an 'error'. We consider two types of errors that are commonly encountered in numerical computations.

You are already familiar with the rounding off a number which has non-terminal decimal expansion from your school arithmetic. For example we use 3.1425 for 22/7. These rounded off numbers are approximations of the actual values. In any computational procedure we make use of these approximate values instead of the true values. Let x_T denote the true value and x_A denote the approximate value. How do we measure the goodness of an approximation x_A to x_T? The simplest measure which naturally comes to our mind is the difference between x_T and x_A. This measure is called the 'error'. Formally, we define error as a quantity which satisfies the identity.

True value x_T = Approximate value x_A + error.

Now if an 'error' in approximation is considered small (according to some criterion), then we say that 'x_A is a good approximation to x'.

Let us consider an example.

Example 13: The true value of π is 3.14159265 … In some mensuration problems the value 22/7 is commonly used as an approximation to π. What is the error in this approximation?

Solution: The true value of π is

$\pi = 3.14159265$ (13)

Now, we convert 22/7 to decimal form, so that we can find the difference between the approximate value and true value. Then the approximate value of π is

$$\frac{22}{7} = 3.14285714 \qquad (14)$$

Therefore,

error = True value – approximate value = -0.00126449 (15)

Note that in this case the error is negative. Error can be positive or negative. We shall in general be interested in absolute value of the error which is defined as

|error| = |True value – approximate value|

For example, the absolute Error in Example 13 is

|error | = |-0.00126449…| = 0.00126…

Sometimes, when the true value is very small we prefer to study the error by comparing it with the value. This is known as Relative error and we define this error as

$$|\text{Relative error}| = \left|\frac{\text{True value - approximate value}}{\text{True value}}\right|$$

In the case of Example 13,

$$|\text{Relative error}| = \frac{0.00126449...}{3.14159265...} = 0.00040249966\ldots$$

But note that in certain computations, the true value may to be available. In that case we replace the true value by the computed approximate value by the computed approximate value in the definition of relative error.

In numerical calculations, you will encounter mainly two types of errors: round-off error and truncation error. We shall discuss these errors in the next two subsections 1.4.1 and 1.4.2 respectively.

3.3.1 Round-off Error

Let us look at Example 13 again. You can see that the numbers appearing in Eqn. (13), (14) and (15) consists of 8 digits after the decimal point followed by dots. The line of dots indicates that the digits continue and we are not able to write all of them. That is, these numbers cannot be represented exactly by a terminating decimal expansion. Whenever we use much numbers in calculations we have to decide how many digits we are going to take into account. For example, consider again the approximate value of π. If we approximate π using 2 digits after the decimal point (say), chopping off the other digits, then we have

$\pi = 3.14$

The error in this approximation is

error = 0.00159265 (16)

If we use 3 digits after the decimal point, then using chopping we have
$\pi \approx 3.141$

In this case the error is given by

error = -0.00059265 (17)

Now suppose we consider the approximate value rounded-off to three decimal places. You already know how to round off a number which has non-terminal decimal expansion . Then the value of π rounded-off to 3 digits is 3.142. The error in this case is

error = -0.00040734…

which is smaller, in absolute value than 0.00059265…given in Eqn. (17). Therefore in general whenever we want to use only a certain number of digits after the decimal point, then it is always better to use the value rounded-off to that many digits because in this case the error is usually small. The error involved in a process where we use rounding-off method is called round-off error.

We now discuss the concept of floating point arithmetic.

In scientific computations a real number x is usually represented in the form

$x = \pm(.\, d_1\, d_2 \ldots d_n)\, 10^m$

where $d_1\, d_2 \ldots d_n$ are natural numbers between 0 and 9 and m is an integer called exponent. Writing a number in this form is known as floating point representation. We denote this representation by fl(x). Such a floating point number is said to be normalized if $d_1 \neq 0$. To translate a number into floating point representation we adopt any of the two methods – rounding and chopping. For example, suppose we want to represent the number 537 in the normalized floating point representation with n = 1, then we get

$$\begin{aligned} fl\,(537) &= .5 \times 10^3 \text{ chopped} \\ &= .5 \times 10^3 \text{ rounded} \end{aligned}$$

In this case we are getting the same representation in rounding and chopping. Now if we take n = 2, then we get

$$\begin{aligned} fl\,(537) &= .53 \times 10^3 \text{ chopped} \\ &= .54 \times 10^3 \text{ rounded} \end{aligned}$$

In this case, the representations are different.

Now if we take n = 3, then we get

$$\begin{aligned} fl\,(537) &= .537 \times 10^3 \text{ chopped} \\ &= .537 \times 10^3 \text{ rounded} \end{aligned}$$

The number n in the floating point representation is called precision.

The difference between the true value of a number x and rounded fl(x) is called round-off error. From the earlier discussion it is clear that the round-off error decreases when precision increases.

Mathematically, we define these concepts as follows:

Definition 2: Let x be a real number and x* be a real number having non-terminal decimal expansion, then we say that x* represents x rounded to k decimal places if

$|x - x^*| \leq \frac{1}{2}\, 10^{-k}$, where k > 0 is a positive integer.

Next definition gives us a measure by which we can conclude that the round-off error occurring in an approximation process is negligible or not.

Definition 3: Let x be a real number and x* be an approximation to x. Then we say that x* is accurate to k decimal places if

$$\frac{1}{2}10^{-(k+1)} \leq |x - x^*| \leq \frac{1}{2}10^{-k} \qquad (18)$$

Let us consider an example.

Example 14: Find out to how many decimal places the value of 22/7 obtained in Example 13 is accurate as an approximation to $\pi = 3.14159265$?

Solution: We have already seen in Example 13 that

$$\left|\pi - \frac{22}{7}\right| = 0.00126449\ldots$$

Now $.0005 < .00126\ldots < 0.005$
or $\frac{1}{2}10^{-3} < .00126\ldots < \frac{1}{2}10^{-2}$

Therefore the inequality (18) is satisfied for k = 2.

Hence, by Definition 3, we conclude that the approximation is accurate to 2 decimal places.

Now we make an important remark.

Remark 5: Round-off errors can create serious difficulties in lengthy computations. Suppose we have a problem which involves a long calculation. In the course of these computations many rounding errors (some positive, and some negative) may occur in a number of ways. At the end of the calculations these errors will get accumulated and we don't know the magnitude of this error. Theoretically it can be large. But, in reality some of these errors (between positive and negative errors) may get cancelled so that the accumulated error will be much smaller.

Let us now define another type of error called Truncation error.

3.3.2 Truncation Error

We shall first illustrate this error with a simple example. In Sec. 1.3. we have already discussed how to find approximate value of a certain function f(x) for a given value of x using Taylor's series expression. Let

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n$$

denote the Taylor's series of f(x) about x_0. In practical situations, we cannot, in general, find the sum of an infinite number of terms. So we must stop after a finite number of terms, say N.

This means that we are taking

$$f(x) = \sum_{n=0}^{N} (x - x_0)^n$$

and ignoring the rest of the terms, that is, $\sum_{n=N+1}^{\infty} a_n (x - x_0)^n$

There is an error involved in this truncating process which arises from the terms which we exclude. This error is called the 'truncation error'. We denote this error by T E. Thus we have

$$T\ E = f(x) - \sum_{n=0}^{N} a_n (x - x_0)^n \quad \sum_{n=N+1}^{\infty} a_n (x - x_0)^n$$

You already know how to calculate this error from Sec. 1.3. There we saw that using Taylor's theorem we can estimate the error (or remainder) involved in a truncation process in some cases.

Let's see what happen if we apply Taylor's theorem to the function f(x) about the point $x_0 = 0$. We assume that f satisfies all conditions of Taylor's theorem. Then we have

$$f(x) = \sum_{n=0}^{N} a_n x^n + \frac{x^{N+1}}{N+1!} f^{N+1}(c) \qquad (19)$$

where $a_n = \dfrac{f^{(n)}(0)}{n!}$ and $0 < c < x$.

now, suppose that we want to approximate f(x) by $\sum_{n=0}^{N} a_n x^n$.

Then Eqn. (19) tells us that the truncation error in approximating f(x) by $\sum_{n=0}^{N} a_n x^n$ is given by

$$T\ E = R_{N+1}(x) = \frac{x^{N+1}}{N+1!} f^{N+1}(c) \qquad (20)$$

Theoretically we can use this formula for truncation error for any sufficiently differentiable function. But practically it is not easy to calculate the nth derivative of many functions. Because of the complexity in differentiation of such functions, it is better to obtain indirectly their Taylor polynomials by using one of the standard expansions we have listed in Sec. 1.3.

For example consider the function $f(x) = e^{x^2}$. It is difficult to calculate the nth derivative of this function. Therefore, for convenience, we obtain Taylor's expansion of e^{x^2} using Taylor's expansion of e^y by putting $y = x^2$. We shall illustrate this in the following example.

Example 15: Calculate a bound for the truncation error in approximation e^{x^2} by

$$e^{x^2} \approx 1 + x^2 + \frac{x^4}{2!} + \frac{x^6}{3!} + \frac{x^8}{4!} \text{ for } x \in]-1, 1[.$$

Solution: Put $u = x^2$. Then $e^{x^2} = e^u$. Now we apply the Taylor's theorem to function $f(u) = e^u$ about $u = 0$. Then, we have

$$e^u = 1 + u + \frac{u^2}{2!} + \frac{u^3}{3!} + \frac{u^4}{4!} + R_5(u) \text{ where}$$

$$R_5(u) = \frac{e^c u^5}{5!}$$

And $0 < c < u$. Since $|x| < 1$, $u = x^2 < 1$ i.e. $c < 1$. Therefore, $e^c < e < 3$. Thus

$$|R_5(u)| \leq \left|\frac{3x^{10}}{5!}\right| < \frac{3}{5!} = \frac{1}{40} = .025$$

Hence the truncation error in approximating e^{x^2} by the above expression is less than 25×10^{-1}.

If the absolute value of the TE is less, then we say that he approximation is good.

Now, in practical situations we should be able to find out the value of n for which the summation $\Sigma a_n x^n$ gives a good approximation to f(x). For this we always specify the accuracy (or error bound) required in advance. Then we find n using formula (20) such that the absolute error $|R_{n+1}(x)|$ is less than the specified accuracy. This gives the approximation within the prescribed accuracy.

Let us consider an example.

Example 16: Find an approximate value of the integral

$$\int_0^1 e^{x^2} dx$$

with an error less than 0.025

Solution: In Example 15 we observed that

$$e^{x^2} \approx 1 + \frac{x^2}{1!} + \frac{x^4}{2!} + \frac{x^6}{3!} + \frac{x^8}{4!}$$

with TE = $\frac{e^{x^2} x^{10}}{5!}$ dx.

Now we use this approximation to calculate the integral. We have

$$\int_0^1 e^{x^2}\,dx \approx \int_0^1 (1 + x^2 + \frac{x^4}{2!} + \frac{x^6}{3!} + \frac{x^8}{4!})dx \qquad (20)$$

with the truncation error

$$TE = \int_0^1 \frac{e^{x^2} x^{10}}{5!}\,dx.$$

We have

$$|TE| \int_0^1 \frac{e^{x^2}\,|x|^{10}}{5!} \le \frac{3}{5!} = .25 \times 10^{-1}$$

Integrating the right hand side of (21), we get

$$\int_0^1 e^{x^2} \approx \int_0^1 (1 + x^2 + \frac{x^4}{2!} + \frac{x^6}{3!} + \frac{x^8}{4!})dx = \left(x + \frac{x^3}{3} + \frac{x^5}{5\times 2!} + \frac{x^7}{7\times 3!} + \frac{x^9}{9\times 4!}\right)_0^1$$

$$= \left(x + \frac{x^3}{3} + \frac{x^5}{10} + \frac{x^7}{42} + \frac{x^9}{216}\right)_0^1$$

$$= 1 + \frac{1}{2} + \frac{1}{10} + \frac{1}{40} + \frac{1}{216}$$

$= 0.0048$

Here is an important remark.

Remark: The magnitude of the truncation error could be reduced within any prescribed accuracy by retaining sufficient large number of terms.
Likewise the magnitude of the round-off error could be reduced by retaining additional digits.

You can now try the following self assessment exercises.

SELF ASSESSMENT EXERCISE

i Calculate a bound for the truncation error in approximation f(x) = sin x by

$$\sin x \approx 1 - \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} \text{ where } -1 \le x \le 1.$$

ii Using the approximation in (a), calculate an approximate value of the integral

$$\int_0^1 \frac{\sin x}{x}\,dx$$

with an error 10^{-4}.

SELF ASSESSMENT EXERCISE

i Calculate the truncation error in approximating e^{-x^2} by $1 - x^2 + \frac{x^4}{2}$, $-1 \le x \le 1$.

ii Using the approximation in (a) calculate an approximate value of $\int_0^1 e^{-x^2} dx$ within an error bound of 10^{-7}.

4.0 CONCLUSION

We end this unit by summarizing what we have learnt in this unit.

5.0 SUMMARY

In this unit we have:

- recalled three important theorems in calculus, namely

  i) Intermediate value theorem
  ii) Rolle's Theorem
  iii) Lagrange's mean value theorem
- State Taylor's theorem and demonstrated it with the help of examples. The nth Taylor's expansion:

$$f(x) = f(x_0) + \frac{(x \quad x_0)}{1!} f'(x_0) + \frac{(x \quad x_0)^2}{2!} f^{(2)}(x_0) + \ldots$$

$$\ldots + \frac{(x \quad x_0)^n}{n!} f^{(n)}(x_0) + \frac{(x \quad x_0)^{n+1}}{(n+1)!} f^{(n+1)}(c)$$

- Defined the term 'error' occurring in numerical computations.
- Discussed two types of errors namely
  i) Round-off error: Error occurring in computations where we use rounding off method to represent a number is called round-off error.
  ii) Truncation error: Error occurring in computations where we use truncation process to represent the sum of an infinite number of terms.
- Explained how Taylor's theorem is used to calculate the truncation error.

6.0 TUTOR-MARKED ASSIGNMENT

i Show that the following equations have a solution in the interval given alongside.

ii Using Rolle's Theorem show that there is a solution to the equation tan x – 1 + x = 0 in]0, 1[.

iii Let $f(x) = \frac{1}{3}x^3 + 2x$. Find a number x_0 in]0, 3[such that

$$f'(x_0) = \frac{f(3) - f(0)}{3 - 0}$$

iv Find all numbers x_0 in the interval]-2, 1[for which the tangent to the graph of $f(x) = x^3 + 4$ is parallel to the line joining the end points (-2, f(-2)) and (1, f(1)).

v. Show that Rolle's Theorem is a special case of mean value theorem.

vi. If P_r denotes the rth Taylor polynomial as given y Eqn (3), then show that $P_r(x_0) = f(x_0)$, $P'_r(x_0) = f'(x_0)$, ... $P_r^{(r)}(x_0) = f^{(r)}(x_0)$.

vii. Obtain the third Taylor polynomial of $f(x) = e^x$ about x = 0.

viii. Obtain the nth Taylor expansion of the function $f(x) = \frac{1}{1+x}$ in $]-\frac{1}{2}, 1[$ about $x_0 = 0$.

ix. Does $f(x) = \sqrt{x}$ have a Taylor series expansion about x = 0? Justify your answer.

x. Obtain the 8th Taylor expansion of the function f(x) = cos x in $[-\frac{\pi}{4}, \frac{\pi}{4}]$ about $x_0 = 0$. Obtain a bound for the error $R_9(x)$.

xi. Using Maclaurin's expansion for cos x, find the approximate value of cos $\frac{\pi}{4}$ with the error bound 10^{-5}.

xii. How large should n be chosen in Maclaurin's expansion for e^x to have $|e^x - P_n(x)| \leq 10^{-5}$, $-1 \leq x \leq 1$.

xiii. In some approximation problems where graphic methods are used, the value $\frac{355}{133}$ is used as an approximation to

$\pi = 3.14159265$….To how many decimal places the value $\frac{355}{133}$ is accurate as an approximation to π?

7.0 REFERENCES/FURTHER READINGS

Wrede, R.C. and Spegel M. (2002). Schaum's and Problems of Advanced Calculus. McGraw – Hill N.Y.

Keisler, H.J. (2005). Elementary Calculus. An Infinitesimal Approach. 559 Nathan Abbott, Stanford, California, USA

UNIT 2 ITERATION METHODS FOR LOCATING ROOT

CONTENTS

1.0 INTRODUCTION

We often come across equation of the form $x^4 + 3x^2 + 2x + 1 = 0$ or $e^x = x - 2$ or tanh x = x etc. Finding one or more values of x which satisfy these equations is one of the important problems in Mathematics. From your elementary algebra course, you are already familiar with some methods of solving equations of degrees 1, 2, 3 and 4. equations of degree 1, 2, 3 and 4 are called linear, quadratic, cubic and biquadratic respectively. There you might have realized that it is very difficult to use the methods available for solving cubic and biquadratic equations. In fact no formula exists for solving equations of degree $n \geq 5$. In these cases we take recourse to approximate methods for the determination of the solution of equations of the form.

$$f(x) = 0 \qquad (1)$$

The problem of finding approximate values of roots of polynomial equations of higher degree was initiated by Chinese mathematicians. The methods of solution in various forms appeared in the 13th century work che' in kiu-shoo. The first noteworthy work in this direction was done in Euope by the English mathematician Fibonacci. Later in the year 1600 Vieta and Isaac Newton made significant contribution to the theory.

In this unit as well as in the next two units we shall discuss some numerical methods which gives an approximate solution of an equation f(x) = 0. We can classify the methods of solution into two types namely (i) Direct methods and (ii) Iteration methods.
Direct methods produce solution by in finite number of steps whereas iteration methods give an approximate solution by repeated application of a numerical process. You will find later that for using iteration methods we have to start with an

approximate solution. Iteration methods improve this approximate solution. We shall begin this unit by first discussing methods which enable us to determine an initial approximate solution and then discuss iteration methods to refine this approximate solution.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- find an initial approximation of the root using (1) tabulation method (2) graphical method
- use bisection method for finding approximate roots
- use fixed point iteration method for finding approximate roots.

3.0 MAIN BODY

3.1 Initial Approximation to a Root

You know that in many problems of engineering and physical sciences you come across equations in one variable of the form f(x) = 0.

For example, in Physical, the pressure-volume-temperature relationship of real gases can be described by the equation

$$PV = RT + \frac{\beta}{V} + \frac{r}{V^2} + \frac{s}{V^3} \qquad (2)$$

where P, V, T are pressure, volume and temperature respectively. R, β, r, s are constants. We can rewrite Eqn. (2) as

$$PV^4 - RTV^3 - \beta V^3 - rV - s = 0 \qquad (3)$$

Therefore the problem of finding the specific volume of a gas at a given temperature and pressure reduces to solving the biquadratic equation Eqn. (3) for the unknown variable V.

Consider another example in life sciences, the study of genetic problem of recombination of chromosomes can be described in the form

$$p(1 - p) = p^2 - p + k - 0,$$

where p stands for the recombination fraction with the limitation $0 \leq p \leq \frac{1}{2}$ and (1 – p) stands for the non-recombination fraction. The problem of finding the recombination fraction of a gene reduces to the problem of finding roots of the quadratic equation $p^2 - p + k = 0$.

In these problems we are concerned with finding value (or values) of the unknown variable x that satisfies the equation f(x) = 0. the function f(x) may be a polynomial of the form

$$f(x) = a_0 + a_1 x + ... + a_n x_n$$

or it may be a combination of polynomials, trigonometric, exponential or logarithmic functions. By a root of this equation we mean a number x_0 such that $f(x_0) = 0$. The root is also called a zero of f(x).

If f(x) is linear, then Eqn. (1) is of the form ax + b = 0, a $\neq 0$ and it has only one root given by $x = -\frac{b}{a}$. Any equation which is not linear is called a non-equation. In this unit we shall discuss some methods for finding roots of the equation f(x) = 0 where f(x) is a non linear function. You are already familiar with various methods for calculating roots of quadratic, cubic and biquadratic equations. But there is no such formula for solving polynomial equations of degree more than 4 or even for a simple equation like

$$x - \cos x = 0$$

Here we shall discuss some of the numerical approximation methods. These methods involve two steps:

- Step 1: To find an initial approximation of a root.
- Step 2: To improve this approximation to get a more accurate value.

We first consider step 1. Finding an initial approximation to a root means locating (or estimating) a root of an equation approximately. There are two ways for achieving this-tabulation mehod and graphical method.

Let us start with Tabulation method.

3.1.1 Tabulation Method

This method is based on the intermediate value theorem (IV Theorem), (see Theorem 1, Unit 1). Let us try to understand the various steps involved in the method through an example.

Suppose we want to find a root of the equation

$$2x - \log_{10} x = 7$$

We first compute value of $f(x) = 2x - \log_{10}x - 7$ for different value of x, say x = 1, 2, 3 and 4.

When x = 1, we have $f(1) = 2 - \log_{10}1 - 7 = -5$

Similarly, we have

$f(2) = 4 - \log_{10}2 - 7 = -3.301$

(Note that $\log_{10}2$ is computed using a scientific calculator.)

$f(3) = 6 - \log_{10}2 - 7 = -1.477$
$f(4) = 8 - \log_{10}4 - 7 = -0.3977$

These values are given in the following table:

Table 1

x	1	2	3	4
f(x)	-5	-3.301	-1.477	0.397

We find that f(3) is negative and f(4) is positive. Now we apply IV Theorem to the function $f(x) = 2x - \log_{10}x - 7$ in the interval $I_1 = [3, 4]$. Since f(3) and f(4) are of opposite signs, by IV theorem there exists a number x_0 lying between 3 and 4 such that $f(x_0) = 0$. That is, a root of the function lies in the interval]3, 4[. Note that this root is positive.

Let us now repeat the above computations for some values of x lying in]3, 4[say x = 3.5, 3.7 and 3.8. In the following table we report the values of f(x).

Table 2

x	3.5	3.7	3.8
f(x)	-0.544	-0.168	0.0202

We find that f(3.7) are of opposite signs. By applying IV theorem again to f(x) in the interval $I_2 = [3.7, 3.8]$, we find that the root of f(x) lies in the interval]3.7, 3.8[. Note that this interval is smaller than the previous interval. We call this interval a refinement of the previous interval. Let us repeat the above procedure once again for the interval I_2. In Table 3 we give the values of f(x) for some x between 3.7 and 3.8.

Table 3

x	3.75	3.78	3.79
f(x)	-0.074	-0.017	-0.00137

Table 3 shows that the root lies within the interval]3.78, 3.79[and this interval is much smaller compared to the original interval]3, 4[. The procedure is terminated by taking any value of x between 3. 78 and 3.79 as an approximate value of the root of the equation $f(x) = 2x - \log_{10}x - 7 = 0$.

The method illustrated above is known as Tabulation method. Let us write the steps involved in the method.

Step 1: Select some numbers $x_1, x_2, ..., x_n$ and calculate $f(x_1)$ and $f(x_2), ..., f(x_n)$. If $f(x_i) = 0$ for some i, then x_i is a root of the equation. If none of the x_is are zero, then proceed to step 2.

Step 2: Find values x_i and x_{i+1} such that $f(x_i)\, f(x_{i+1}) < 0$. Rename $x_i = a_1$ and $x_{i+1} = b_1$. Then by the IV Theorem a root lies in between a_1 and b_1. Test for all values of $f(x_j)$, j = 1, 2, ..., n and determine other intervals, if any, in which some more roots may lie.

Step 3: Repeat Step 1 by taking some numbers between a_1 and b_1. Again, if $f(x_j) = 0$ for some x_j between a_1 then we have found the root x_j. Otherwise, continue step 2.

Continue the step 1, 2, 3 till we get a sufficiently small interval]a, b[in which the root lies. Then any value between]a, b[can be chosen as an initial approximation to the root. You may have noticed that the test values x_j, j = 1, 2, ..., n chosen are dependent on the nature of the function f(x).

We can always gather some information regarding the root either from the physical problem in which the equation f(x) = 0 occur, or it is specified in the problem. For example, we may ask for the smallest positive root or a root closest to a given number etc.

For a better understanding of the method let us consider one more example.
Example 1: Find the approximate value of the real root of the equation

$2x - 3 \sin x - 5 = 0$.

Solution: Let $f(x) = 2x - 3 \sin x - 5$.

Since $f(-x) = -2x + 3 \sin x - 5 < 0$ for $x > 0$, the function f(x) is negative for all negative real numbers x. Therefore the function has no negative real root. Hence the roots of this equation must lie in $[0, \infty[$. Now following step 1, we compute values of f(x), for x = 0, 1, 2, 3, 4, ...

We have

$f(0) = -5.0$,
$f(1) = 2 - 3 \sin 1 - 5 = 5.5224$

using the calculator. Note that x is in radians. The values f(0), f(1), f(2) and f(3) are given in Table 4.

Table 4

x	0	1	2	3
f(x)	-5.0	-5.51224	-3.7278	0.5766

Now we follow step 2. From the table we find that f(2) and f(3) are of opposite signs. Therefore a root lies between 2 and 3. Now, to get a more refined interval, we evaluate f(x) for some values between 2 and 3. The values are given in Table 5.

Table 5

x	2	2.5	2.8	2.9
f(x)	-3.7278	-1.7954	-0.4049	0.0822

This table of values shows that f(2.8) and f(2.9) are of opposite signs and hence the root lies between 2.8 and 2.9. We repeat the process once again for the interval [2.8, 2.9] by taking some values as given in Table 6.

Table 6

x	2.8	2.85	2.88	2.89
f(x)	-0.4049	-1.1624	-0.0159	0.0232

From Table 6 we find that the root lies between 2.88 and 2.89. This interval is small, therefore we take any value between 2.88 and 2.89 as an initial approximation of the root. Since f(2.88) is near to zero than f(2.89), we can take any number near to 2.88 as an initial approximation to the root.

You might have realized that the tabulation method is a lengthy process for finding an initial approximation of a root. However, since only a rough approximation to the root is required, we normally use only one application of the tabulation method. In the next sub-section we shall discuss the graphical method.

3.1.2 Graphical Method

In this method, we draw the approximate graph of y = f(x). The points where the curve cuts the x-axis are taken as the required approximate values of the roots of the equation f(x) = 0. Let us consider an example.

Example 2: Find an approximate value of a root of the biquadractic equation

$x^4 + 4x^3 + 4x^2 – 2 = 0$
using graphical method.

Solution: We first sketch the fourth degree polynomial $f(x) = x^4 + 4x^3 + 4x^2 - 2$. This graph is given in Fig. 1.

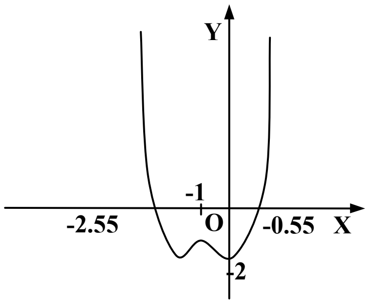

Fig. 1: Graph of f(x) = + $4x^3 + 4x^2 - 2$

The figure shows that the graph cuts the x-axis at two points -2.55 and 0.55, approximately. Hence -2.55 and 0.55 are taken as the approximate roots of the equation

$$x^4 + 4x^3 + 4x^2 - 2 = 0$$

Now go back for a moment to Unit 1 and see Example 1 in Sec. 1.2. There we applied graphical method to find the roots of the equation $\sin x = \frac{1}{2}$.

Let us consider another example.

Example 3: Find the approximate value of a root of

$$x^2 - e^x = 0$$

using graphical method.

Solution: First thing to do is to draw the graph of the function $f(x) = x^2 - e^x$. It is not easy to graph this function. Now if we split the function as

$$f(x) = f_1(x) - f_2(x)$$

where $f_1(x) = x^2$ and $f_2(x) = e^x$, then we can easily draw the graphs of the functions $f_1(x)$ and $f_2(x)$. The graphs are given in fig. 2.

The figure shows that the two curves $y = x^2$ and $y = e^x$ intersect at some point P. From the figure, we find that the approximate point of intersection of the two curves is -0.7. Thus we

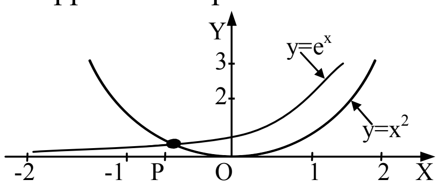

Fig. 2: Graphs of $f_1(x) = x^2$ and $f_2(x) = e^x$.

have $f_1(-0.7) - f_2(-0.7)$, and therefore $f(-0.7) = f_1(-0.7) - f_2(-0.7) \approx 0$. Hence -0.7 is an approximate value of the root of the equation f(x) = 0.

From the above example we observe the following: Suppose we want to apply the graphic method for finding an approximate root of f(x) = 0. Then we may try to simply the method by splitting the equation as

$$f(x) = f_1(x) - f_2(x) = 0 \qquad (4)$$

where the graphs of $f_1(x)$ and $f_2(x)$ are easy to draw. From Eqn. (4), we have $f_1(x) = f_2(x)$. The x-coordinate of the point at which the two curves $y_1 = f_1(x)$ and $y_2 = f_2(x)$ intersect gives an approximate value of the root of the equation f(x) = 0. Note that we are interested only in the x-coordinate, we don't have to worry about the point of intersection of the curves.

Often we can split the function f(x) in the form (4) in a number of ways. But we should choose that form which involves minimum calculations and the graphs of $f_1(x)$ and $f_2(x)$ are easy to draw. We illustrate this point in the following example.

Example 4: Find an approximate value of the positive real root of 3x – cos x – 1 = 0 using graphic method.
Solution: Since it is easy to plot 3x – 1 and cosx, we rewrite the equation as 3x – 1 = cos x. The graphs of $y = f_1(x) = 3x - 1$ and $y = f_2(x) = \cos x$ are given in Figure 3.

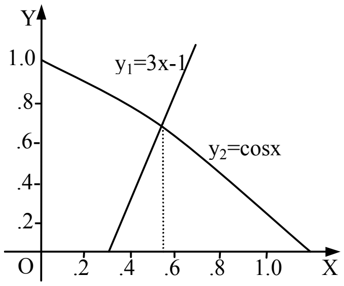

Fig. 3: Graphs of $f_1(x) = 3x - 1$ and $f_2(x)$ cos x

It is clear from the figure that the x-coordinate of the point of intersection is approximately 0.6. Hence x = 0.6 is an approximate value of the root of the equation 3x – cos x – 1 = 0.

We now make a remark.

Remark 1: You should take some care while choosing the scale for graphing. A magnification of the scale may improve the accuracy of the approximate value.

We have discussed two methods, namely, tabulation method and graphical method which help us in finding an initial approximation to a root. But these two methods give only a rough approximation to a root. Now to obtain more accurate results, we need to improve these crude approximations. In the tabulation method we found that one way of improving the process is refining the intervals within which a root lies. A modification of this method is known as bisection method. In the next section we discuss this method.

3.2 Bisection Method

In the beginning of the previous section we have mentioned that there are two steps involved in finding an approximate solution. The first step has already been discussed. In this section we consider the second step which deals with refining an initial approximation to a root.

Once we know an interval in which a root lies, there are several procedures to refine it. The bisection method is one of the basic methods among them. We repeat the steps 1, 2, 3 of the tabulation method given in subsection 3.3.1 in a modified form. For convenience we write the method as an algorithm.
Suppose that we are given a continuous function f(x) defined on [a, b] and we want to find the roots of the equation f(x) = 0 by bisection method. We described the procedure in the following steps:

Step 1: Find points x_1, x_2 in the interval [a, b] such that $f(x_1). f(x_2) < 0$. That is, those points x_1 and x_2 for which $f(x_1)$ and $f(x_2)$ are of opposite signs-(see Step 1 subsection 3.3.1). This process is called "finding an initial bisecting interval". Then IV theorem a root lies in the interval $]x_1, x_2[$.

Step 2: Find the middle point c of the interval $]x_1, x_2[$ i.e., $c = \frac{x_1 + x_2}{2}$. If f(c) = 0, then c is the required root of the equation and we can stop the procedure. Otherwise we go to Step 3.

Step 3: Find out if

$f(x_1) f(c) < 0$

If it holds, then the root lies in $]x_1, c[$. Otherwise the root lies in $]c, x_2[$ (see Fig 4). Thus in either case we have found an interval half as wide as the original interval that contains the root.

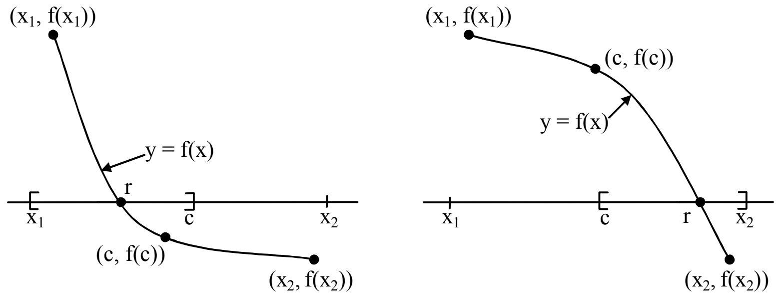

Fig. 4: The decision process for the bisection method

Step 4: Repeat Step 2 and 3 with the new interval. This process either gives you the root or an interval having width ¼ of the original interval $]x_1, x_2[$ which contains the required root.

Step 5: Repeat this procedure until the interval width is as small as we desire. Each bisection halves the length of the preceding interval. After N steps, the original interval length will be reduced by a factor $1/2^N$.

Now we shall see how this method helps in refining the initial intervals in some of the problems we have done in subsection 2.2.1.

Example 5: Consider the equation $2x - \log_{10}x - 7$ lies in [3.78, 3.79]. Apply bisection method to find an approximate root of the equation correct to three decimal places.

Solution: Let $f(x) = 2x - \log_{10}x - 7$. From Table 2 in subsection 3.3.1, we find that f(3.78) = -0.01749 and f(3.79) = 0.00136. Thus s root lies in the interval]3.78, 3.79[.

Then we find the middle point of the interval]3.78, 3.79[. The middle point is c = (3.78 + 3.79)/2 = 3.785 and f(c) = f3.785) = -0.0806 $\neq$ 0. Now, we check the condition in Step 3. Since f(3.78) f(3.785) > 0, the root does not lie in the interval [3.78. 3.78]. Hence the root lies in the interval [3.785, 3.9]. We have to refine this interval further to get better approximation. Further bisection are shown in the following Table.

Table 7

Number of Bisection	Bisected value x_i	$f(x_i)$	Improved Interval
1	3.785	-0.00806	]3.785, 3.79[
2	3.7875	-3.3525×10^{-3}	]3.7875, 3.79[
3	3.78875	9.9594×10^{-4}	]3.78875, 3.79[
4	3.789375	1.824×10^{-4}	]3.78875, 3.789375[
5	3.7890625	-4.068×10^{-4}	]3.78906, 3.7989375[

The table shows that the improved interval after 5 bisections is [3.78906, 3.789375]. The width of this interval in 3.789375 - 3.78906 = 0.000315. If we stop further bisections, the maximum absolute error would be 0.000315. The approximate root can therefore be taken as (3.78906 - 3.789375)/2 = 3.789218. Hence the desired approximate value of the root rounded off to three decimal places is 3.789.

Example 6: Apply bisection method to find an approximation to the positive root of the equation.

$2x - 3 \sin x - 5 = 0$

rounded off to three decimal places.

Solution: Let f(x) = 2x – 3 sin x – 5.

In Example 1, we had shown that a positive root lies in the interval [2.8, 2.9]. Now we apply bisection method to this interval. The results are given in the following table.

Table 8

Number of Bisection	Bisected value x_i	$f(x_i)$	Improved Interval
1	2.85	-0.1624	]2.85, 2.79[
2	2.875	-0.0403	]2.875, 2.79[
3	2.8875	0.02089	]2.875, 2.8875[
4	2.88125	-9.735×10^{-3}	]2.88125, 2.8875[
5	2.884375	5.57781×10^{-3}	]2.88125, 2.884375[
6	2.8828125	-2.0795×10^{-3}	]2.8828125, 2.884375[
7	2.8835938	1.7489×10^{-3}	]2.8828125, 2.8835938[
8	2.8832031	-1.6539×10^{-4}	]2.8832031, 2.8835938[

After we bisection the width of the interval is 2.8835938 - 2.8832031 = 0.0003907. Hence, the maximum possible absolute error to the root is 0.0003907. Therefore the required approximation to the root is 2.883.

Now let us make some remarks.

Remark 2: While applying bisection method we must be careful to check that f(x) is continuous. For example, we may come across functions like f(x) = $\frac{1}{x-1}$. If we consider the interval [.5, 1.5], then f(.5) f(1.5) < 0. In this case we may be tempted to use bisection method. But we cannot use the method here because f(x) is not defined at the middle point x = 1. We can overcome these difficulties by taking f(x) to be continuous throughout the initial bisecting interval. (Note that if f(x) is continuous by IV theorem f(x) assumes all values between the intervals.)

Therefore you should always examine the continuity of the function in the initial interval before attempting the bisection method.

Remark 3: It may happen that a function has more than one root in an interval. The bisection method helps us in determining one root only. We can determine the other roots by properly choosing the initial intervals.

While applying bisection method we repeatedly apply steps 2, 3, 4 and 5. You recall that in the introduction we classified such a method as an Iteration method. As we mentioned in the beginning of Sec. 3.1, a numerical process starts with an initial approximation and iteration improves this approximation until we get the desired accurate value of the root.

Let us consider another iteration method now.

3.3 Fixed Point Iteration Method

The bisection method we have described earlier depends on our ability to find an interval in which the root lies. The task of finding such intervals is difficult in certain situations. In such cases we try an alternate method called Fixed Point Iteration Method. We shall discuss the advantage of this method later.

The first step in this method is to rewrite the equation f(x) = 0 as

$$x = g(x) \qquad (5)$$

For example consider the equation $x^2 - 2x - 8 = 0$. We can write it as

$$x = \sqrt{2x+8} \qquad (6)$$

$$x = \frac{2x+8}{x} \qquad (7)$$

$$x = \frac{x^2 - 8}{2} \qquad (8)$$

We can choose the form (5) in several ways. Since f(x) = 0 is the same s x = g(x), finding a root of f(x) = 0 is the same as finding a root of x = g(x) i.e., a fixed point of g(x). Each such g(x) given in (6), (7) or (8) is called an iteration function for solving f(x) = 0.

Once an iteration function is chosen, our next step is to take a point x_0 close to the root as the initial approximation of the root.

Starting with x_0, we find the first approximation x_1 as

$$x_1 = g(x_0)$$

Then we find the next approximation as

$$x_2 = g(x_1)$$

Similarly we find the successive approximation x_2, x_3, x_4 ... as

$$x_3 = g(x_2)$$
$$x_4 = g(x_3)$$
. .
. .
. .
$$x_{n+1} = g(x_n)$$

Each computation of the type $x_{n+1} = g(x_n)$ is called an iteration. Now, two questions arise (i) when do we stop these iterations? (ii) Does this procedure always give the required solution?

To ensure this we make the following assumptions on g(x):

Assumption*
The derivative g'(x) of g(x) exists g'(x) is continuous and satisfies $|g'(x)| < 1$ in an interval containing x_0. (That would mean that we require $|g'(x)| < 1$ at all iterates x_i.)

The iteration is usually stopped whenever $|x_{i+1}|$ is less than the accuracy required.

In Unit 3 you will prove that if g(x) satisfies the above conditions, then there exists a unique point α such that $g(\alpha) = \alpha$ and the sequence of iterates approach α, provided that the initial approximation is close to the point α.

Now we shall illustrate this method with the following example.

Example 7: Find an approximate root of the equation

$x^2 - 2x - 8 = 0$
using fixed point iteration method, starting with $x_0 = 5$. Stop the iteration whenever

$|x_{i+1} - x_i| < 0.001$.

Solution: Let $f(x) = x^2 - 2x - 8$. We saw that the equation f(x) = 0 can be written in three forms (6), (7) and (8). We shall take up the three forms one by one.

Case 1: Suppose we consider form (5). In this form the equation is written as

$x = (2x + 8)^{1/2}$

Here $g(x) = (2x + 8)^{1/2}$. Let's see whether Assumption (*) is satisfied for this g(x). We have

$$g'(x) = \frac{1}{(2x+8)^{1/2}}$$

Then $|g'(x)| < 1$ whenever $(2x + 8)^{1/2} > 1$. For any positive real number x, we see that the inequality $(2x + 8)^{1/2} > 1$ is satisfied. Therefore, we consider any interval on the positive side of x-axis. Since the starting point is $x_0 = 5$, we may consider the interval at I = [3, 6]. This contains the point 5. Now, g(x) satisfies the condition that g'(x) exists on I, g'(x) is continuous on I and $|g'(x)| < 1$ for every x in the interval [3, 6]. Now we apply fixed point iteration method to g(x).

We get
$x_1 = g(5) = \sqrt{18} = 4.243$
$x_2 = g(4.243) = 4.060$
$x_3 = 4.015$
$x_4 = 4.004$
$x_5 = 4.001$
$x_6 = 4.000.$

Since $|x_6 - x_5| = |-0.001| = 0.001$, we conclude that an approximate value of a root of $f(x) = 0$ is 4.

Case 2: Let us consider the second form,

$$x = \frac{2x+8}{x}$$

Here $g(x) = \frac{2x+8}{x}$ and $g'(x) = \frac{-8}{x^2}$. The $|g'(x)| < 1$ for any real number $x \geq 3$. Hence g(x) satisfies Assumption (*) in the interval [3, 6]. Now we leave is as an exercise for you to complete the computations (See TMA 6).

Case 3: Here we have $x = \frac{x^2 - 8}{2}$. Then $g(x) = \frac{x^2 - 8}{2}$ and $g'(x) = x$. In this case $|g'(x)| < 1$ only if $|x| < 1$ i.e. if x lies in the interval]-1, 1[. But this interval does not contain 5. Therefore g(x) does not satisfy the Assumption (*) in any interval containing the initial approximation. Hence, the iteration method cannot provide approximation to the desired root.

Note: This example may appear artificial to you. You are right because in this case we have got a formula for calculating the root. This example is taken to illustrate the method in a simple way.

Let us consider another example.

Example 8: Use fixed point iteration procedure to find an approximate root of 2x = 3 sin x – 5 = 0 starting with the point $x_0 = 2.8$. Stop the iteration whenever $|x_{i+1} + x_i| < 10^{-5}$.

Solution: We can rewrite the equation in the form,

$$x = \frac{3}{2}\sin x + \frac{5}{2}.$$

Here $g(x) = \frac{3}{2}\sin x + \frac{5}{2}$ and $g'(x) = \frac{3}{2}\cos x$.

Now at $x_0 = 2.8$, we have

$|g'(2.8)| = 1.413$
which is greater than 1. Thus g(x) does not satisfy Assumption (*) and therefore in this form the iteration method fails.

Let us now rewrite the equation in another form. We write

$$x = x - \frac{2x - 3\sin x - 5}{2 - 3\cos x}$$

Then $g(x) = x - \frac{2x - 3\sin x - 5}{2 - 3\cos x}$

You may wonder how did we get this form. Note that here g(x) is of the form $g(x) = x - \frac{f(x)}{f'(x)}$. You will find later that the above equation is the iterated formula for another popular iteration method.

Then $g'(x) = 1 - \left(\frac{(2 - 3\cos x)(2 - 3\cos x) - (2x - 3\sin x + 5)3\sin x}{(2 - 3\cos x)^2}\right)$

$= \frac{2x - 3\sin x + 5}{(2 - 3\cos x)^2} 3\sin x$

At $x_0 = 2.8$ $|g'(x_0)| = 0.0669315$ (or 0.02174691) < 1

Therefore g(x) satisfies the Assumption (*). Using the initial approximation as $x_0 = 2.8$, we get the successive approximation as

$x_1 = 2.8839015$
$x_2 = 2.8832369$
$x_3 = 2.8832369$

Since $|x_2 - x_3| < 10^{-5}$ we stop the iteration here and conclude that 2.88323 is an approximate value of the root.

Next we shall use another form

$$x = \sin^{-1}\left(\frac{2x - 5}{3}\right)$$

Here $g(x) = \sin^{-1}\left(\frac{2x - 5}{3}\right)$ and $g'(x) = \frac{2}{\sqrt{9 - (2x - 5)^2}}$

At $x_0 = 2.8$, $g'(x_0) = 0.6804 < 1$. In fact, we can check that in any small interval containing 2.8 $|g'(x)| < 1$. Thus g(x) satisfies the Assumption (*). Applying the iteration method, we have

$$x_1 = \sin^{-1}\left(\frac{2(2.8) - 5}{3}\right) = 0.201358$$

We find that there are two values which satisfy the above equation. One value is 0.201358 and the other is π - 0.201358 = 2.940235. In situations, we take a value close to the initial approximation. In this case the value close to the initial approximation is 2.940235. Therefore we take this value as the starting point of the next approximation.

$x_1 = 2.940235$

Next we calculate

$$x_2 = \sin^{-1}\left(\frac{2(2.940235) - 5}{3}\right)$$
$= 0.297876$ or 2.843717

Continuing like this, it needed 17 iteration to obtain the value $x_{17} = 2.88323$, which we got from the previous form. This means that in this form the convergence is very slow.

From examples 7 and 8, we learn that if we choose the form x = g(x) properly, then we can get the approximate root provided that the initial approximation is sufficiently close to the root. The initial approximation is usually given in the problem or we can find using the IV theorem.

Now we shall make a remark here

Remark: The Assumption (*) we have given for an iteration function, is a stronger assumption. In actual practice there are a variety of assumptions which the iteration function g(x) must satisfy to ensure that the iterations approach the root. But, to use those assumptions you would require a lot of practice in the application of techniques in mathematical analysis. In this course, we will be restricting ourselves to functions that satisfies Assumption (*). If you would like to know about the other assumptions, you may refer to 'Elementary Numerical Analysis' by Samuel D Conte and Carl de Boor.

4.0 CONCLUSION

Let us now briefly recall what we have done in this unit.

5.0 SUMMARY

In this unit we have covered the following points:

- We have seen that the methods for finding an approximate solution of an equation involve two steps:
  i) Find an initial approximation to a root.
  ii) Improve the initial approximation to get a more accurate value of the root.
- We have described the following iteration methods for improving an initial approximation of a root.
  i) Bisection method
  ii) Fixed point iteration method.

6.0 TUTOR-MARKED ASSIGNMENT (TMA)

i. Find an initial approximation to a root of the equation $3x - \sqrt{1+\sin x} = 0$ using tabulation method.

ii. Find a initial approximation to a positive root of the equation $2x - \tan x = 0$ using tabulation method.

iii. Find the approximate location of the roots of the following equations in the regions given using graphic method.
   a. $f(x) = e^{-x} - x = 0$, in $0 \le x \le 1$
   b. $f(x) = e^{-0.4x} - 0.4x - 9 = 0$, in $0 < x \le 7$

iv. Starting with the interval $[a_0, b_0]$, apply bisection method to be the following equations and find an interval of width 0.05 that contains a solution of the equations
   a. $e^x - 2 - x = 0$, $[a_0, b_0] = [1.0, 1.8]$
   b. $1n\ x - 5 + x = 0$, $[a_0, b_0] = [3.2, 4.0]$

v. Using bisection method find an approximate root of the equation $x^3 - x - 4 = 0$ in the interval]1, 2[to two places of decimal.

vi. Apply fixed point iteration method to the form $x = \frac{2x+8}{x}$ starting with $x_0 = 5$ to obtain a root of $x^2 - 2x - 8 = 0$.

vii. a) Apply fixed point iteration method to the following equations with the initial approximation given alongside. In each case find an approximate root rounded off to 4 decimal places.

   6.1 $x = -45 + \frac{2}{x}$ $x_0 = 20$.

   6.2 $x = \frac{1}{2} + \sin x$, $x_0 = 1$.

   b) Compute the exact roots of the equation $x^2 + 45x - 2 = 0$ using quadratic formula and compare with the approximate root obtained in (a) (i).

7.0 REFERENCES/FURTHER READINGS

Wrede, R.C. and Spegel M. (2002). Schaum's and Problems of Advanced Calculus. McGraw – Hill N.Y.

Keisler, H.J. (2005). Elementary Calculus. An Infinitesimal Approach. 559 Nathan Abbott, Stanford, California, USA

UNIT 3 CHORD METHOD FOR FINDING ROOTS

CONTENTS

1.0 INTRODUCTION

In the last unit we introduced you to two iteration methods for finding roots of an equation f(x) = 0. There we have shown that a root of the equation f(x) = 0 can be obtained by writing the equation in the form x = g(x). Using this form we generate a sequence of approximations $x_{i+1} = g(x_i)$ for i = 0, 1, 2, ... We had also mentioned there that the success of the iteration methods depends upon the form of g(x) and the initial approximation x_0. In this unit, we shall discuss two iteration methods: regula-falsi and Newton-Raphson methods. These methods produce results faster than bisection method. The first two sections of this unit deal with derivations and the use of these two methods. You will be able to appreciate these iteration methods better if you can compare the efficiency of these methods. With this in view we introduce the concept of convergence criterion which helps us to check the efficiency of each method. Sec. 3.3 is devoted to the study of rate of convergence of different iterative methods.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- apply regula-falsi and secant methods for finding roots
- apply Newton-Raphson method for finding roots
- define ‘order of convergence’ of an iterative scheme
- obtain the order of convergence of the following four methods:
  - bisection method
  - fixed point iteration method
  - secant method
  - Newton-Raphson method

3.0 MAIN BODY

3.1 Regula-Falsi Method (or Method of False Position)

In this section we shall discuss the 'regula-falsi method'. The Latin word 'Regula Falsi' means rule of falsehood. It does not mean that rule is a false statement. But it conveys that the roots that we get according to the rule are approximate roots and not necessarily exact roots. The method is also known as the method of false position. This method is similar to the bisection method you have learnt in Unit 3.

The bisection method for finding approximate roots has a draw back that it makes use of only the signs of f(a) and f(b). It does not use the values f(a), f(b) in the computations. For example, if f(a) = 700 and f(b) = -0.1, then by the bisection method the first approximate value of a root of f(x) is the mid value x_0 of the interval]a, b[. But at x_0, $f(x_0)$ is nowhere near 0. Therefore in this case it makes more sense to take a value near to -0.1 than the middle value as the approximation to the root. This drawback is to some extent overcome by the regula-falsi method. We shall first describe the method geometrically.

Suppose we want to find a root of the equation f(x) = 0 where f(x) is a continuous function. As in the bisection method, we first find an interval]a, b[such that f(a) f(b) < 0. Let us look at the graph of f(x) given in Fig. 1.

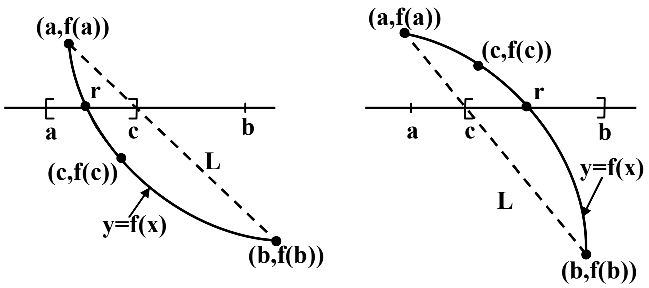

Fig 1: Regula-Falsi

The condition f(a) f(b) < 0 means that the points (a, f(a)) and (b, f(b)) lie on the opposite sides of the x-axis. Let bus consider the line joining (a, f(a)) and (b, f(b)). This line crosses the x-axis at some point (c, 0) [see Fig. 1]. Then we take the x-coordinate of that point as the first approximation. If f(c) = 0, then x = c is the required root. If f(a) f(c) < 0, then the root lies in]a, c[(see Fig. 1 (a)). In this case the graph of y = f(x) is concave near the root r). Otherwise, if f(a) f(c) > 0, the root lies in]c, b[(see Fig. 1 (b)). In this case the graph of y = f(x) is convex near the root. Having fixed the interval in which the roots lies, we repeat the above procedure.

Let us now write the above procedure in the mathematical form. Recall the formula for the line joining two points in the Cartesian plane. The line joining (a, f(a)) and (b, f(b)) is given by

$$y - f(a) = \frac{f(b) - f(a)}{b - a}(x - a)$$

We can rewrite this in the form

$$\frac{y - f(a)}{f(b) - f(a)} = \frac{x - a}{b - a} \qquad (1)$$

Since the straight line intersects the x-axis at (c, 0), he point (c, 0) lies on the straight line. Putting x = c, y = 0 in Eqn. (1), we get

$$\frac{- f(a)}{f(b) - f(a)} = \frac{c - a}{b - a}$$

i.e. $\frac{c}{b - a} - \frac{a}{b - a} = \frac{- f(a)}{f(b) - f(a)}$

Thus $c = a \frac{f(a)}{f(b) - f(a)} (b – a)$. (2)

This expression for c gives an approximate value of a root of f(x). Simplifying (2), we can also write as

$$\frac{af(b) - bf(a)}{f(b) - f(a)}$$

Now, examine the sign of f(c) and decide in which interval]a, c[or]c, b[, the root lies. We thus obtain a new interval such that f(x) is of opposite signs at the end points of this interval. By repeating this process, we get a sequence of intervals]a, b[,]a, a_1[,]a, a_2[, ... as shown in Fig. 2.

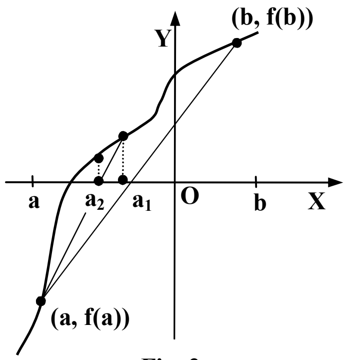

Fig. 2

We stop the process when either of the following holds.

i) The interval containing the zero of f(x) is of sufficiently small length or
ii) The difference between two successive approximation is negligible.

In the iteration format, the method is usually written as

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$$

where $]x_0, x_1[$ is the interval in which the root lies.

We now summarise this method in the algorithm form. This will enable you to solve problems easily.

- Step 1: Find numbers x_0 and x_1 such that $f(x_0)\, f(x_1) < 0$, using the tabulation method.
- Step 2: Set $x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$. This gives the first approximation.
- Step 3: If $f(x_2) = 0$ then x_2 is the required root. If $f(x_2) \neq 0$ and $f(x_0)\, f(x_2) < 0$, then the next approximation lies in $]x_0, x_2[$. Otherwise it lies in $]x_2, x_1[$.
- Step 4: Repeat the process till the magnitude of the difference between two successive iterated values x_i and x_{i+1} is less than the accuracy required. (Note that $|x_{i+1} – x_i|$ gives the error after ith iteration).

Let us now understand these steps through an example.

Example 1: It is known that the equation $x^3 + 7x^2 + 9 = 0$ has a root between -8 and -7. Use the regula-falsi method to obtain the root rounded off to 3 decimal places. Stop the iteration when $|x_{i+1} – x_i| < 10^{-4}$.

Solution: For convenience we rewrite the given function f(x) as

$$f(x) = x^3 + 7x^2 + 9$$
$$= x^2(x + 7) + 9$$

Since we are given that $x_0 = -8$ and $x_1 = -7$, we do not have to use Step 1. Now to get the first approximation, we apply the formula in Step 2.

Since, $f(x_0) = f(-8) = -55$ and $f(x_1) = f(-7) = 9$ we obtain

$$x_2 = \frac{(-8)9 - (-7)(-5)}{9 + 55} = -7.1406$$

Therefore our first approximation is -7.1406.

To find the next approximation we calculate $f(x_2)$ with the signs of $f(x_0)$ and $f(x_1)$. We can see that $f(x_0)$ and $f(x_2)$ are of opposite signs. Therefore a root lies in the interval]-8, -7.1406[. We apply the formula again by renaming the end points of the interval as $x_1 = -8$, $x_2 = -7.1406$. Then we get the second approximation as

$$x_3 = \frac{-8\ f(-7.1406) + 7.1406\ f(-8)}{1.862856 + 55} = -7.168174.$$

We repeat this process using Step 2 and 3 given above. The iterated values are given in the following table.

Table 1

Number of iterations	Interval	Iterated Values x_i	The function value $f(x_i)$
1	]-8,-7[	-7.1406	1.862856
2	]-8,-7.1406[	-7.168174	0.3587607
3	]-8,-7.168174[	-7.1735649	0.0683443
4	]-8,-7.1735649[	-7.1745906	0.012994
5	]-8,-7.1745906[	-7.1747855	0.00246959
6	]-8, -7.1747855[	-7.1748226	0.00046978

From the able, we see that the absolute value of the difference between the 5th and 6th iterated values is |7.1748226 – 7.1747855| = .0000371. Therefore we stop the iteration here. Further, the values of f(x) at 6th iterated value is $.00046978 = 4.6978 \times 10^{-4}$ which is close to zero. Hence we conclude that -7.175 is an approximate root of
$x^3 + 7x^2 + 9 = 0$
Rounded off to three decimal places.

You note that in regula-falsi method, at each stage we find an interval $]x_0, x_1[$ which contains a root and then apply iteration formula (3). This procedure has a disadvantage. To overcome this, regula-falsi method is modified. The modified method is known as secant method. In this method we choose x_0 and x_1 as any two approximations of the root. The Interval $]x_0, x_1[$ need not contain the root. Then we supply formula (3) with x_0, x_1, $f(x_0)$ and $f(x_1)$.

The iterations are now defined as:

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x)}$$

$$x_3 = \frac{x_1 f(x_2) - x_2 f(x_1)}{f(x_2) - f(x_1)}$$

....

....

$$x_{n+1} = \frac{x_{n-1} f(x_n) - x_n f(x_{n-1})}{f(x_n) - f(x_{n-1})} \qquad (4)$$

Note: Geometrically, in secant Method, we replace the graph of f(x) in the interval $]x_n, x_{n+1}[$ by a straight line joining two points $(x_n, f(x_{n+1}), (x_{n+1}), f(x_{n+1}))$ on the curve and take the point of intersection with x-axis as the approximate value of the root. Any line joining two points on the curve is called a secant line. That is why this method is known as secant method. (see Fig. 3).

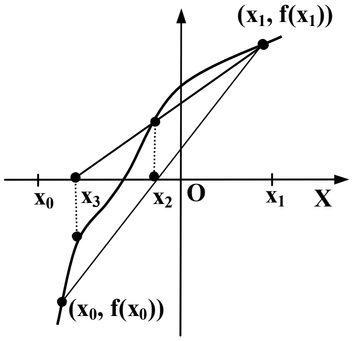

Fig. 3

Let us solve an example.

Example 2: Determine an approximate root of the equation

$\cos x - x e^x = 0$

using

i) secant method starting with the two initial approximations as $x_0 = 1$ and $x_1 = 1$ and

ii) regula-falsi method.
(This example was considered in the book 'Numerical methods for scientific and engineering computation' by M. K. Jain, S. R. K. Iyengar and R. K. Jain).

Solution: Let $f(x) = \cos x - x e^x$.

Then $f(0) = 1$ and $f(1) = \cos 1 - e = -2.177979523$. Now we apply formula (4) with $x_0 = 0$ and $x_1 = 1$. Then

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)} = \frac{0(-2177979523 + (-1)1}{-2.177979523 - 1}$$

$$= \frac{-1}{-2.177979523 - 1} = \frac{1}{3.177979523} = 0.3146653378.$$

Therefore the first iterated value is 0.3146653378. to get the 2nd iterated value, we apply formula (4) with $x_1 = 1$, $x_2 = 0.3144653378$. Now $f(1) = -2.177979523$ and f(0. 3144653378) = 0.519871175.

Therefore

$$x_3 = \frac{x_1 f(x_2) - x_2 f(x_1)}{f(x_2) - f(x_1)}$$

$$= \frac{1(0.519871175) - 0.3146653378(-2.177979523)}{0.519871175 + 2.177979523}$$

= 0.4467281466

We continue this process. The iterated values are tabulated in the following table.

Table 2: Secant Method

Number of iterations	Iterated Values x_i	$f(x_i)$
1	0.3146653378	0.519871
2	0.4467281466	0.203545
3	0.5317058606	-0.0429311
4	0.5169044676	.00259276
5	0.5177474653	0.00003011
6	0.5177573708	-0.215132×10^{-7}
7	0.5177573637	0.178663×10^{-12}
8	0.5177573637	0.222045×10^{-15}

From the table we find that the iterated values for 7th and 8th iterations are the same. Also the value of the function at the 8th iteration is closed to zero. Therefore we conclude that 0.5177573637 is an approximate root of the equation.

ii) To apply regula-falsi method, let us first note that f(0) f(1) < 0. Therefore a root lies in the interval [0, 1]. Now we apply formula (3) with $x_0 = 0$ and $x_1 = 1$. then the first approximation is

$$x_2 = \frac{0(-2177979523 + (-1)1}{-2.177979523 - 1}$$

$$= 0.3146653378$$

You may have noticed that we have already calculated the expression on the right hand side of the above equation in part (i).

Now $f(x_2) = 0.51987 > 0$. This shows that the root lies in the interval]0.3146653378, 1[. To get the second approximation, we compute

$$x_3 = \frac{0.3146653378\ f(1)\ -1f(0.3146653378)}{f(1) - f(0.3146653378)} = 0.4467281446$$

which is same as x_3 obtained in (i). We find $f(x_2) = 0.203545 > 0$. Hence the root lies in [0.4467281446, 1]. To get the third approximation, we calculate

$$x_4 = \frac{0.4467281446\ f(1)\ -1f(0.4467281446)}{f(1) - f(0.4467281446)}$$

The above expression on the right hand side is different from the expression for x_4 in part (i). This is because when we use regula-falsi method, at each stage, we have to check the condition $f(x_1)\ f(x_{i-1}) < 0$.

The computed values of the rest of the approximations are given in Table 3.

Table 3: Regula-Falsi Method

No.	Interval	Iterated value x_i	$f(x_i)$
1	[0, 1[	0.3146653378	0.519871
2	].04467281446, 1[	0.4467281446	0.203545
3	]0.4940153366, 1[	0.4940153366	0.708023×10^{-1}
4	]0.5099461404, 1[	0.5099461404	0.236077×10^{-1}
5	]0.5152010099, 1[	0.5152010099	0.776011×10^{-2}
6	]0.5176683450, 1[	0.5177478783	0.288554×10^{-4}
7	]0.5177478783, 1[	0.5177573636	0.396288×10^{-9}

From the table, we observe that we have to perform 20 iterations using regula-falsi method to get the approximate value of the root 0.5177573637 which we obtained by secant method after 8 iterations. Note that the end point 1 is fixed in all iterations given in the table.

Next we shall discuss another iteration method.

3.2 Newton-Raphson Method

This method is one of the most useful methods for finding roots of an algebraic equation.

Suppose that we want to find an approximate root of the equation $f(x) = 0$. If $f(x)$ is continuous, then we can apply either bisection method or regula-falsi method to find approximate roots. Now if $f(x)$ and $f'(x)$ are continuous, then we can use a new iteration method called Newton-Raphson method. You will learn that this method gives the result more faster than the bisection or regula-falsi methods. The underlying idea of the method is due to mathematician Isac Newton. But the method as now used is due to the mathematician Raphson.

Let us begin with an equation $f(x) = 0$ where $f(x)$ and $f'(x)$ and are continuous. Let x_0 be an initial approximation and assume that x_0 is close to the exact root α and $f'(x) \neq 0$. Let $\alpha = x_0 + h$ where h is a small quantity in magnitude. Hence $f(\alpha) = f(x_0 + h) = 0$.

Now we expand $f(x_0 + h)$ using Taylor's theorem. Note that $f(x)$ satisfies all the requirements of Taylor's theorem. Therefore, we get

$$f(x_0 + h) = f(x_0) + hf'(x_0) + ... = 0$$

Neglecting the terms containing h^2 and higher powers we get

$f(x_0) + hf'(x_0) = 0.$

Then, $h = \frac{-f(x_0)}{f'(x_0)}$

This gives a new approximation to α as

$x_1 = x_0 + h = x_0 - \frac{-f(x_0)}{f'(x_0)}$

Now the iteration can be defined by

$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$

$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} \qquad (5)$$

Eqn. (5) is called the Newton-Raphson formula. Before solving some examples we shall explain this method geometrically.

Geometrical Interpretation of Newton-Raphson Method

Let the graph of the function $y = f(x)$ be as shown in Fig. 4.

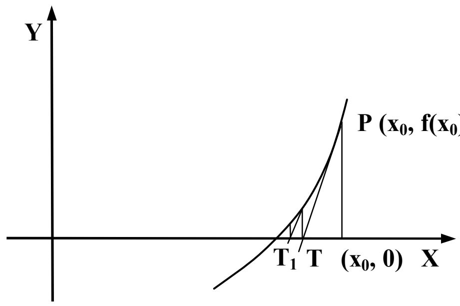

Fig. 4 Newton-Raphson Method

If x_0 is an initial approximation to the root, then the corresponding point on the graph is $P(x_0, f(x_0))$. We draw a tangent to the curve at P. Let it intersect the x-axis at T. (see Fig. 4). Let x_1 be the x-coordinate of T. Let $S(\alpha, 0)$ denote the point on the x-axis where the curve cuts the x-axis. We know that α is a root of the equation $f(x) = 0$. We take x_1 as the new approximation which may be closer to α than x_0. Now let us find the tangent at $P(x_0, f(x_0))$. The slope of the tangent at $P(x_0, f(x_0))$ is given by $f'(x_0)$. Therefore by the point-slope form of the expression for a tangent to a curve, we can write

$y - f(x_0) = f'(x_0)(x_1 - x_0)$

This tangent passes through the point $T(x_1, 0)$ (see fig. 4). Therefore we get

$0 – f(x_0) = f'(x_0)\ (x_1 – x_0)$

i.e. $x_1\ f'(x_0) = x_0 f'(x_0) – f(x_0)$

i.e. $x_1 = x_0 – \frac{f(x_0)}{f'(x_0)}$

This is the first iterated value. To get the second iterated value we again consider a tangent at a point $P(x_1, f(x_1))$ on the curve (see Fig. 4) and repeat the process. Then we get a point $T_1(x_2, 0)$ on the x-axis. From the figure, we observe that T_1 is more closer to $S(\alpha, 0)$ than T. therefore after each iteration the approximation is coming closet and closer to the actual root. In practice we do not know the actual root of a given function.

Let us now take up some examples.

Example 3: Find the smallest positive root of

$2x – \tan x = 0$

by Newton-Raphson method, correct to 5 decimal places.

Solution: Let $f(x) = 2x – \tan x$. Then $f(x)$ is a continuous function and $f'(x) = 2 – \sec^2 x$ is also a continuous function. Recall that the given equation has already appeared in an exercise in Unit 2 (see TMA in Unit 2). From that exercise we know that an initial approximation to the positive root of the equations is $x = 1$. Now we apply the Newton-Raphson iterated formula.

$x_1 = x_{i-1} - \frac{f(x_i)}{f'(x_i)}$, $i = 1, 2, 3 ....$

Here $x_0 = 1$. Then $f(x_0) = f(1) = 2 – \tan 1 = 0.4425922$

$$f'(x_0) = f'(1) = 2 – \sec^2 1 = 2 – (1 + \tan^2 1)$$
$$= 1 - \tan^2 1$$
$$= -1.425519$$

Therefore $x_1 = 1 - \frac{0.4425922}{-1.425519}$

$= 1.31048$

For $i = 2$, we get
$x_3 = 1.17605$

$x_4 = 1.165926$
$x_5 = 1.165562$
$x_6 = 1.165561$

Now x_5 and x_6 are correct to five decimal places. Hence we stop the iteration process here. The root correct to 5 decimal places is 1.16556.

Next we shall consider an application of Newton-Raphson formula. We know that finding the square root of a number is not easy unless we use a calculator. Calculators use some algorithm to obtain such an algorithm for calculating square roots. Let's consider an example.

Example 4: Find an approximate value of $\sqrt{2}$ using the Newton-Raphson formula.

Solution: Let $x = \sqrt{2}$. Then we have $x^2 = 2$ i.e. $x^2 – 2 = 0$. Hence we need to find the positive root of the equation $x^2 – 2 = 0$. Let

$f(x) = x^2 – 2$.

Then f(x) satisfies all the conditions for applying Newton-Raphson method. We choose $x_0 = 1$ as the initial approximation to the root. This is because we know that $\sqrt{2}$ lies between $\sqrt{1}$ and $\sqrt{4}$ and therefore we can assume that the root will be close to 1.

Now we compute the iterated values.

The iteration formula is

$$x_i = x_{i-1} - \frac{x_{i-1}^2 - 2}{2x_{i-1}}$$

$$= \frac{1}{2}\left(x_{i-1} + \frac{2}{x_{i-1}}\right)$$

Putting i = 1, 2, 3 ….. we get

$$x_1 = \frac{1}{2}\left(x_0 + \frac{2}{x_0}\right) = 1.5$$

$$x_2 = \frac{1}{2}\left(1.5 + \frac{2}{1.5}\right) = 1.4166667$$

$$x_3 = \frac{1}{2}\left(1.4166667 + \frac{2}{1.416667}\right)$$

$= 1.41242157$

Similarly

$x_4 = 1.4142136$
$x_5 = 1.4142136$

Thus the value of $\sqrt{2}$ correct to seven decimal places is 1.4142136. Now you can check this value with the calculator.

Note 1: The method used in the above example is applicable for finding square root of nay positive real number. For example suppose we want to find an approximate value of $\sqrt{A}$ where A is a positive real number. Then we consider the equation $x^2 - A = 0$. The iterated formula in this case is

$$x_i = \left(\frac{1}{2}\ x_{i-1} + \frac{A}{x_{i-1}}\right)$$

This formula involves only the basic arithmetic operations +, -, × and ÷.

Note 2: From examples (3) and (4), we find that Newton-Raphson method gives the root very fast. One reason for this is that the derivative $|f'(x)|$ is large compared to $|f(x)|$ for any $x = x_i$. The quantity $\left|\frac{f(x)}{f'(x)}\right|$ which is the difference between two iterated values is small in this case. In general we can say that if $|f'(x_i)|$ is large compared to $|f(x_i)|$, then we can obtain the desired root very fast by this method.

The Newton-Raphson method has some limitations. In the following remarks we mention some of the difficulties.

Remark 1: Suppose $f'(x_i)$ is zero in a neighbourhood of the root, then it may happen that $f'(x_i) = 0$ for some x_i. In this case we cannot apply Newton-Raphson formula, since division by zero is not allowed.

Remark 2: Another difficulty is that it may happen that $f'(x)$ is zero only at the roots. This happens in either of the situations.

i) f(x) has multiple root at α. Recall that a polynomial function f(x) has a multiple root α of order N if we can write

$$f(x) = (x - \alpha)^N h(x)$$

where h(x) is a function such that $h(\alpha) \neq 0$. For a general function f(x), this means $f(\alpha) = 0 = f'(\alpha) = ... = f^{N-1}(\alpha)$ and $f^N(\alpha) \neq 0$.

ii) f(x) has a stationary point (point of maximum of minimum) point at the root [recall from your calculus course that if $f'(x) = 0$ at some point x then x is called a stationary point].

In such cases some modifications to the Newton-Raphson method are necessary to get an accurate result. We shall not discuss the modifications here as they are beyond the scope of this course.

You can try some exercise now. Whenever needed, should use a calculator for computation.

In the next section we shall discuss a criterion using which we can check the efficiency of an iteration process.

3.3 Convergence Criterion

In this section we shall introduce a new concept called 'convergence criterion' related to an iteration process. This criterion gives us an idea of how much successive iteration has to be carried out to obtain the root to the desired accuracy. We begin with a definition.

Definition 1: Let x_0, x_1 …..x_n …. be the successive approximation of an iteration process. We denote the sequence of these approximation as $\{x_n\}_{n=0}^{\infty}$. We say that $\{x_n\}_{n=0}^{\infty}$ converges to a root α with order p $\geq$ 1 if

$$|x_{n+1} - \alpha| \leq \lambda |x_n - \alpha|^P \qquad (6)$$

for some number $\lambda > 0$. p is called the order of convergence and λ is called the asymptotic error constant.

For each i. we denote by $\varepsilon_i = x_i - \alpha$. Then the above inequality be written as

$$|\varepsilon_{i+1}| \leq \lambda |\varepsilon_i|^P \qquad (7)$$

This inequality shows the relationship between the errors in successive approximations. For example, suppose p = 2 and $|\varepsilon_i| \approx 10^{-2}$ for some i. then we can expect that $|\varepsilon_{i+1}| \approx \lambda 10^{-4}$. Thus if p is large, the iteration converges rapidly. When p takes the integer values 1, 2, 3 then we say that the convergences are linear, quadratic and cubic respectively. In the case of linear convergence (i.e. p =1). Then we require that $\lambda < 1$. In this case we can write (6) as

$$|x_{n+1} - \alpha| \leq \lambda |x_n - \alpha| \text{ for all } n \geq 0 \qquad (8)$$

In this condition is satisfied for an iteration process then we say that the iteration process converges linearly.

Setting n = 0 in the inequality (8), we get

$$|x_1 - \alpha| \leq \lambda |x_0 - \alpha|$$

For n = 1, we get

$|x_2 - \alpha| \leq \lambda |x_1 - \alpha| \leq \lambda^2 |x_0 - \alpha|$

Similarly for n = 2, we get

$|x_3 - \alpha| \leq \lambda |x_2 - \alpha| \leq \lambda^2 |x_1 - \alpha| \leq \lambda^3 |x_0 - \alpha|$

Using induction on n, we get that

$$|x_n - \alpha| \leq \lambda^n |x_0 - \alpha| \text{ for } n \geq 0 \qquad (9)$$

If either of the inequality (8) or (9) is satisfied, then we conclude that $\{x_n\}_{n=0}^{\infty}$ converges to the root.

Now we shall find the order of convergence of the iteration methods which you have studied so far.

Let us first consider bisection method.

Convergence of bisection method

Suppose that we apply the bisection method on the interval $[a_0, b_0]$ for the equation $f(x) = 0$. In this method you have seen that we construct intervals $[a_0, b_0] \supseteq [a_1, b_1] \supseteq [a_2, b_2] \supseteq \ldots$ each of which contains the required root of the given equation.

Recall that in each step the interval width is reduced by $\frac{1}{2}$ i.e.

$$b_1 - a_1 = \frac{b_0 - a_0}{2}$$

$$b_2 = a_2 = \frac{b_1 - a_1}{2} = \frac{b_0 - a_0}{2^2}$$

$$\vdots \qquad \vdots$$

$$\text{and } b_n - a_n = \frac{b_0 - a_0}{2^n} \qquad (10)$$

We know that the equation $f(x) = 0$ has a root in $[a_0, b_0]$. Let α be the root of the equation. Then α lies in all intervals $[a_i, b_i]$, i = 0, 1, 2, …. For any n, let $c_n = \frac{a_n - b_n}{2}$ denote the middle point of the interval $[a_n, b_n]$. Then c_0, c_1, c_2, … are taken as successive approximations to the root α. Let's check the inequality (8) for $\{c_n\}_{n=0}^{\infty}$ converges to the root α. Hence we can say the bisection method always converges.

For practical purposes, we should be able to decide at what stage we can stop the iteration to have an acceptably good approximate value of α. The number of iterations required to achieve a given accuracy for the bisection method can be obtained. Suppose that we want an approximate solution within an error bound of 10^{-M} (Recall that you have studied error bounds in Unit 1, Sec. 3.4). Taking logarithms on both sides of Eqn. (10), we find that the number of iteration required, say n, is approximately given by

$$n = \text{int}\left[\frac{\text{In}(b_0 - a_0) - \text{In}10^{-M}}{\text{In}2}\right] \qquad (11)$$

where the symbol 'int' stands for the integral part of the number in the bracket and $]a_0, b_0[$ is the initial interval in which a root lies.

Let us work out an example.

Example 5: Suppose that the bisection method is used to find a zero of f(x) in the interval [0, 1]. How many times this interval be bisected to guarantee that we have an approximate root with absolute error less than or equal to 10^{-5}.

Solution: Let n denote the required number. To calculate n, we apply the formula in Eqn. (11) with $b_0 = 1$, $a_0 = 0$ and M = 5.

Then

$$n = \text{int}\left[\frac{\text{In}1 - \text{In}10^{-5}}{\text{In}2}\right]$$

Using a calculator, we find

$$n = \text{int}\left[\frac{11.51292547}{0.69314718}\right]$$
$$= \text{int}\ [16.60964047] = 17$$

The following table gives the minimum number of iterations required to find an approximate root in the interval]0, 1[for various acceptable errors.

E	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
n	7	10	14	17	20	24

This table shows that for getting an approximate value with an absolute error bounded by 10^{-5}, we have to perform 17 iterations. Thus even though the bisection method is simple to use, it requires a large number of iterations to obtain a reasonably good approximate root. This is one of the disadvantages of thee bisection method.

Note: The formula given in Eqn. (11) shows that, given an acceptable error, the number of iterations depends upon the initial interval and thereby depends upon the initial approximation of the root and not directly on the values of f(x) at these approximations.

Next we shall obtain the convergence criteria for the secant method.

Convergence criteria for Secant Method

Let f(x) = 0 be the given equation. Let α denote a simple root of the equation f(x) = 0. Then we have $f'(\alpha) \neq 0$. The iteration scheme for the secant method is

$$x_{i+1} = x_i - \frac{x_i - x_{i-1}}{f(x_i) - f(x_{i-1})} \qquad (12)$$

For each i, set $\varepsilon_i = x_i - \alpha$. Then $x_i + \alpha$. Substituting in Eqn. (12) we get

$$\varepsilon_{i+1} + \alpha = \varepsilon_i + \alpha - \frac{\varepsilon_i - \varepsilon_{i-1}}{f(\varepsilon_i + \alpha) - f(\varepsilon_{i+1} + \alpha)} f(\varepsilon_i + \alpha)$$

$$\varepsilon_{i+1} = \varepsilon_i - \frac{\varepsilon_i - \varepsilon_{i-1}}{f(\varepsilon_i + \alpha) - f(\varepsilon_{i-1} + \alpha)} f(\varepsilon_i + \alpha) \qquad (13)$$

Now we expand $f(\varepsilon_i + \alpha)$ and $f(\varepsilon_i - \alpha)$ using Taylor's theorem about the point $x = \alpha$.

We get $f(\varepsilon_i + \alpha) = f(\alpha) + \frac{f'(\alpha)}{1}\varepsilon_i + \frac{f''(\alpha)}{2}\varepsilon_i^2 + ...$

i.e. $f(\varepsilon_i + \alpha) = f'(\alpha)\left[\varepsilon_i + \frac{f''(\alpha)}{2f'(\alpha)}\varepsilon_i^2 + ...\right]$ (14)

since $f'(\alpha) = 0$.

Similarly,

$$f(\varepsilon_{i-1} + \alpha) = f'(\alpha)\left[\varepsilon_{i-1} + \frac{f''(\alpha)}{2f'(\alpha)}\varepsilon_{i-1}^2 + ...\right] \qquad (15)$$

Therefore $f(\varepsilon_i + \alpha) - f(\varepsilon_{i-1} + \alpha) = f'(\alpha)\left[\varepsilon_i - \varepsilon_{i-1} + (\varepsilon_i^2 - \varepsilon_i^2)\frac{f''(\alpha)}{2f'(\alpha)} + ...\right]$

$$= f'(\alpha)(\varepsilon_i - \varepsilon_{i-1})\left[1 + (\varepsilon_i + \varepsilon_{i-1})\frac{f''(\alpha)}{2f'(\alpha)} + ..\right] \qquad (16)$$

Substituting Eqn. (14) and Eqn. (13), we get

$$\varepsilon_{i+1} = \varepsilon_i - \left[\varepsilon_i + \frac{1}{2}\varepsilon_i^2\frac{f''(\alpha)}{f'(\alpha)} + ...\right]\left[1 + \frac{1}{2}(\varepsilon_i + \varepsilon_{i-1})\frac{f''(\alpha)}{f'(\alpha)} + ...\right]^{-1}$$

$$= \varepsilon_i - \left[\varepsilon_i + \frac{1}{2}\varepsilon_i^2 \frac{f''(\alpha)}{f'(\alpha)} + ...\right]\left[1 - \frac{1}{2}(\varepsilon_i + \varepsilon_{i-1})\frac{f''(\alpha)}{f'(\alpha)} + ...\right]$$

$$= \varepsilon_i - \left[\varepsilon_i + \frac{1}{2}\frac{f''(\alpha)}{f'(\alpha)}(\varepsilon_i^2 - \varepsilon_i^2 - \varepsilon_i\,\varepsilon_{i-1}) + ...\right]$$

By neglecting the terms involving $\varepsilon_i\,\varepsilon_{i-1}^2 + \varepsilon_i^2\,\varepsilon'_{i-1}$ the above expression, we get

$$\varepsilon_{i+1} \approx \varepsilon_i\,\varepsilon_{i-1}\left[\frac{f''(\alpha)}{2f'(\alpha)}\right] \qquad (17)$$

This relationship between the errors is called the error equation. Note that this relationship holds only if α is a simple root. Now using Eqn. (17) we will find a number p and λ such that

$$\varepsilon_{i+1} = \lambda\,\varepsilon_i^p \quad i = 0, 1, 2, ... \qquad (18)$$

Setting i = j – 1, we obtain

$$\varepsilon_j = \lambda\,\varepsilon_{j-1}^p$$

or

$$\varepsilon_i = \lambda\,\varepsilon_{i-1}^p$$

Taking pth root on both sides, we get

$$\varepsilon_i^{1/p} = \lambda^{1/p}\,\varepsilon_{i-1}$$

$$\text{i.e. } \varepsilon_{i-1} = \lambda^{-1/p}\,\varepsilon_i^{1/p} \qquad (19)$$

Combining Eqns. (17) and (18). We get

$$\lambda\,\varepsilon_i^p = \varepsilon_i\,\varepsilon_{i-1}\frac{f''(\alpha)}{2f'(\alpha)}$$

Substituting the expression for ε_{i-1} from Eqn. (19) in the above expression we get

$$\lambda\,\varepsilon_i^p \approx \frac{f''(\alpha)}{2f'(\alpha)}\,\varepsilon_i\,\lambda^{-1/p}\,\varepsilon_i^{1/p}$$

$$\text{i.e. } \lambda\,\varepsilon_i^p \approx \frac{f''(\alpha)}{2f'(\alpha)}\,\lambda^{-1/p}\,\varepsilon_i^{1+1/p} \qquad (20)$$

Equating the powers of ε_i on both sides of Eqn. (20) we get

$p = 1 + \frac{1}{p}$ or $p^2 - p - 1 = 0$.

This is a quadratic equation in p. The roots are given by

$$p = \frac{1+\sqrt{5}}{2} \approx 1.618.$$

Now, to get the number λ, we equate the constant terms on both sides of Eqn. (20). Then we get

$$\lambda = \left[\frac{f''(\alpha)}{2f'(\alpha)}\right]^{P/1+p}$$

Hence the order of convergence of the secant method is p = 1.62 and the asymptotic error constant is $\frac{f''(\alpha)}{2f'(\alpha)}^{P/1+p} \left[\quad\right]$

Example 6: The following are the five successive iterations obtained by secant method to find the root α = -2 of the equation $x^3 - 3x + 2 = 0$.

$x_1 = -2.6$, $x_2 = -2.4$, $x_3 = -2.106598985$.
$x_4 = -2.022641412$ and $x_5 = -2.000022537$.

Compute the asymptotic error constant and show that $\varepsilon_5 \approx \frac{2}{3}\varepsilon_4$.

Solution: Let $f(x) = x^3 - 3x + 2$

Then
$f'(x) = 3x^2 - 3$, $f'(-2) = 9$
$f''(x) = 6x$, $f(-2) = -12$

Therefore $\lambda = \left[-\frac{12}{18}\right]^{.618}$

$= \left[-\frac{2}{3}\right]^{.618} = -0.778351205$

Now
$\varepsilon_5 = | x_5 - \alpha | = | -2.000022537 + 2$
$= 0.000022537$
and
$\varepsilon_4 = | -2.022641412 + 2 | = 0.022641412$.

Then $\lambda\, \varepsilon_4 = 0.778351205 \times 2.022641412$
$= 0.000021246$
≈ 0.00002253

Hence we get that $\lambda\, \varepsilon_4 \approx \varepsilon_5$

Convergence criterion for fixed point iteration method
Recall that in this method we write the equation in the form

$x = g(x)$

Let α denote a root of the equation. Let x_0 be an initial approximation to the root. The iteration formula is

$$x_{i+1} = g(x_i),\ i = 0, 1, 2, \ldots \qquad (21)$$

We assume that g'(x) exists and is continuous and $|\, g'(x)\, | < 1$ in an interval containing the root α. We also assume that $x_0, x_1, \ldots.$ lie in this interval.

Since g'(x) is continuous near the root and $|\, g'(x)\, | < 1$, there exists an interval $]\,\alpha - h, \alpha + h[$, where $h > 0$, such that $|\, g'(x)\, | \leq k$ for some k, where $0 < k < 1$.

Since α is a root of the equation, we have

$$\alpha = g(\alpha). \qquad (22)$$

Subtracting (22) from (21) we get

$x_{i+1} - \alpha = g(x_i) = g(\alpha)$

Now the function g(x) is continuous in the interval $]x_i, \alpha[$ and g'(x) exists in this interval. Hence g(x) satisfies all the conditions of the mean value theorem [see Unit 1]. Then, by the mean value theorem there exists a ξ between x_i and α such that

$|\, x_{i+1} - \alpha| \leq |\, g(x_i) - g(\alpha)\, | \leq |\, g'(\xi)\, |\, |\, (x_i - \alpha|$

Note that ξ lies in $]\alpha - h, \alpha + h[$ and therefore $|\, g'(\xi)\, | < k$ and hence
$|\, x_{i+1} - \alpha| \leq |\, x_i - \alpha|$

Setting i = 0, 1, 2, ..., n we get
$|\, x_1 - \alpha| \leq k\, |\, x_0 - \alpha|$
$|\, x_2 - \alpha| \leq k\, |\, x_1 - \alpha| \leq k^2\, |\, x_0 - \alpha|$
. .
. .

. .
$| x_n - \alpha| \leq k^n | x_0 - \alpha|$

This shows that the sequence of approximation $| x_i |$ converges to α provided that the initial approximation is close to the root.

We summarise the result obtained for this iteration process in the following Theorem.

Theorem 1: If g(x) and g'(x) are continuous in an interval about a root α of the equation x = g(x), and if | g'(x) | < 1 for all x in the interval, then the successive approximations $x_1, x_2, ...$ given by

$x_i = g(x_{i-1}), i = 1, 2, 3, ...$
converges to the root α provided that the initial approximation x_0 is chosen in the above interval.

We shall now discuss the order of convergence of this method. From the previous discussions we have the result.

$| x_{i+1} - \alpha| \leq g'(\xi) | (x_i - \alpha) |$

Note that ξ is dependent on each x_i. Now we wish to determine the constant λ and p independent of x_i such that

$| x_{i+1} - \alpha| \leq c | (x_i - \alpha) |^P$

Note that as the approximations x_i get closer to the root α, $g'(\xi)$ approaches a constant value $g'(\alpha)$. Therefore, in the limiting case, as $i \to \infty$,the approximation satisfy the relation

$| x_{i+1} - \alpha| \leq g'(\alpha) | (x_i - \alpha) |$

Therefore, we conclude that if $g'(\alpha) \neq 0$, then the convergence of the method is linear.

If $g'(\alpha) = 0$, then we have

$_{i+1} - \alpha = g(x_i) - \alpha$
$= g(x_i - \alpha) + \alpha - \alpha$
$= g(\alpha) + (x_i - \alpha) g'(\alpha) + \frac{(x_i \alpha)^2}{2} g''(\xi) - \alpha$
$= \frac{(x_i \alpha)^2}{2} g''(\xi)$
since $g(\alpha) = \alpha$ and $g'(\alpha) = 0$ and ξ lies between x_i and α.

Therefore, in the limiting case we have

$$| x_{i+1} - \alpha | \le \frac{1}{2} | g''(\alpha) | \, | (x_i - \alpha) |^2$$

Hence, if $f'(\alpha) = 0$ and $g'(\alpha) \neq 0$, then this iteration method is of order 2.

Example 7: Suppose α and β are the roots of the equation $x^2 + ax + b = 0$. Consider a rearrangement of this equation as

$$x = -\frac{(ax + b)}{x}$$

Show that the iteration $x_{i+1} = -\frac{(ax_i + b)}{x_i}$ will converge near $x = \alpha$ when $|\alpha| > |\beta|$

Solution: The iteration are given by

$$x_{i+1} = g(x_i) = -\frac{(ax_i + b)}{x_i}, \; i = 0, 1, 2, ...$$

By Theorem 1, these iterations converge to α if $|g'(x)| < 1$ near α i.e. if $|g'(x)| = \left| -\frac{b}{x^2} \right| < 1$. Note that $g'(x)$ is continuous near α. If the iterations converge to $x = \alpha$, then we require $|g'(x)| = \left| -\frac{b}{\alpha^2} \right| < 1$.

Thus $| b | < | \alpha |^2$

i.e. $| \alpha |^2 > | b |$. (23)

Now you recall from your elementary algebra course that if α and β are the roots, then

$\alpha + \beta = -a$ and $\alpha\beta = b$

Therefore $| b | = |\alpha| \, |\beta|$. Substituting in Eqn. (23), we get

$|\alpha|^2 > | b | = |\alpha| \, | \beta|$.

Hence $|\alpha| > |\beta|$

Finally, we shall discuss the convergence of the Newton-Raphson method.

Convergence of Newton-Raphson Method

Newton-Raphson iteration formula is given by

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \qquad (24)$$

To obtain the order of the method we proceed as in the secant method. We assume that α is a simple root of f(x) = 0. Let

$x_i - \alpha = \varepsilon_i$, i = 0, 1, 2,...

Then we have

$$\varepsilon_{i+1} + \alpha = \varepsilon_i + \alpha - \frac{f(\varepsilon_i + \alpha)}{f'(\varepsilon_i + \alpha)}$$

$$\text{i.e. } \varepsilon_{i+1} = \frac{\varepsilon_i f'(\varepsilon_i + \alpha) - f(\varepsilon_i + \alpha)}{f'(\varepsilon_i + \alpha)}$$

Now we expand $f(\varepsilon_i + \alpha)$ and $f'(\varepsilon_i + \alpha)$, using Taylor's theorem about the point α. We have

$$\varepsilon_{i+1} = \frac{\left[\varepsilon_i \left\{f'(\alpha + \varepsilon_i f''(\alpha) + \frac{\varepsilon_i^2}{2} f'''(\alpha) + \ldots\right\} - \left\{f(\alpha)\varepsilon_i f'(\alpha) + \frac{\varepsilon_i^2}{2} f''(\alpha) + \ldots\right\}\right]}{f'(\alpha) + \varepsilon_i f''(\alpha) + \varepsilon_1^2 f''(\alpha) + \ldots}$$

But $f(\alpha) = 0$ and $f'(\alpha) \neq 0$. Therefore

$$\varepsilon_{i+1} = \left[\frac{\varepsilon_i^2}{2} f''\alpha + \ldots\right] \frac{1}{f''(\alpha)} \left[1 + \frac{\varepsilon_i f''(\alpha)}{f'(\alpha)} + \ldots\right]^{-1}$$

$$= \frac{1}{f'(\alpha)} \left[\frac{\varepsilon_i^2}{2} f''\alpha + \ldots\right]\left[1 - \frac{\varepsilon_i f''(\alpha)}{f'(\alpha)} + \ldots\right]$$

Hence, by neglecting higher powers of ε_i, we get

$$\varepsilon_{i+1} \approx \frac{f''(\alpha)}{2f'(\alpha)} \varepsilon_i^2$$

This shows that the errors satisfy Eqn. (6) with p = and $\lambda = \frac{f''(\alpha)}{2f'(\alpha)}$. Hence, Newton-Raphson method is of order 2. That is at each step, the error is proportional to the square of the previous error.

Now, we shall discuss an alternate method for showing that the order is 2. Note that we can write (24) in the form x = g(x) where

$$g(x) = x \frac{f(x)}{f'(x)}$$

$$g'(x) = \frac{d}{dx}\left[x - \frac{f(x)}{f(x)}\right] = 1 - \frac{[f'(x)]^2 - f(x)f''(x)}{[f'(x)]^2}$$

$$= \frac{f(x)f''(x)}{[f'(x)]^2}$$

Now, $g'(\alpha) = \frac{f(\alpha)f''(\alpha)}{[f'(\alpha)]^2} = 0$, since $f(\alpha) = 0$ and $f'(\alpha) \neq 0$.

Hence by the conclusion drawn just above Example 7, the method is of order 2. Note that this is true only if α is a simple root. If α is a multiple root i.e. if $g'(\alpha) = 0$, then the convergence is not quadratic, but only linear. We shall not prove this result, but we shall illustrate this with an example.

Let us consider an example.

Example 8: Let $f(x) = (x - 2)^4 = 0$. Starting with the initial approximation $x_0 = 2.1$, compute the iterations x_1, x_2, x_3 and x_4 using Newton-Raphson method. Is the sequence conveying quadratically or linearly?

Solution: The given function has multiple roots at x = 2 and is of order 4.
Newton-Raphson iteration formula for the given equation is

$$x_{i+1} = x_i - \frac{(x_i - 2)^4}{4(x_i - 2)^3} = x_i - \frac{1}{4}(x_i - 2)$$

$$= \frac{1}{4}(3x_i - 2) \qquad (25)$$

Starting with $x_0 = 2.1$, the iteration are given by

$$x_1 = \frac{1}{4}(6.3 + 2) = \frac{8.3}{2} = 2.075$$

Similarly,
$x_2 = 2.05625$
$x_3 = 2.0421875$
$x_4 = 2.031640625$

Now $\varepsilon_0 = x_0 - 2 = 0.1$, $\varepsilon_i = x_1 - 2 = 0.075$, $\varepsilon_2 = 0.05625$, $\varepsilon_3 = 0.0421875$, $\varepsilon_4 = 0.031640625$.

Then

$$\varepsilon_i = .075 = \frac{3}{4} \times 0.1 = \frac{3}{4}\varepsilon_0$$

and

$$\varepsilon_2 = \frac{3}{4}\varepsilon_i$$

$$\varepsilon_3 = \frac{3}{4}\varepsilon_2$$

$$\varepsilon_4 = \frac{3}{4}\varepsilon_3$$

Thus the convergence is linear in this case. The error is reduced by a factor of $\frac{3}{4}$ with each iteration. This result can also be obtained directly from Eqn. (25).

4.0 CONCLUSION

5.0 SUMMARY

In this unit we have

- described the following methods for finding a root of an equation $f(x) = 0$
  - i) Regula-Falsi method:
    The formula is
    $$c = \frac{a\,f(b) - b\,f(a)}{f(b) - f(a)}$$
    where]a, b[is an interval such that $f(a)\,f(b) < 0$.
  - ii) Secant method:
    The iteration formula is
    $$x_{i+1} = \frac{x_{i-1}f(x_i) - x_i f(x_{i-1})}{f(x_i) - f(x_{i-1})} \quad i = 0, 1, 2, ...$$
    where x_0 and x_1 are any two given approximation of the root.
  - iii) Newton-Raphson method:
    The iteration formula is
    $$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)},\ i = 0, 1, 2, ...$$
    where x_0 is an initial approximation to the root.
- introduced the concept called convergence criterion of an iteration process.
- discussed the convergence of the following iterative methods
  - i) Bisection method.
  - ii) Fixed point iteration method.
  - iii) Secant method.
  - iv) Newton-Raphson method.

6.0 TUTOR-MARKED ASSIGNMENT (TMA)

i. Obtain an approximate root for the following equations rounded off to three decimal places, using regula-falsi method
   a. $x \log_{10} x - 1.2 = 0$
   b. $x \sin x - 1 = 0$

ii. Use secant method to find an approximate root to the equation $x^2 - 2x + 1 = 0$, rounded off to 5 decimal places, starting with $x_0 = 2.6$ and $x_1 = 2.5$. Compare the result with the exact root $1 + \sqrt{2}$.

iii. Find an approximate root of the cubic equation $x^3 + x^2 + 3x - 3 = 0$ using
   a. i) regula-falsi method, correct to three decimal places.

      ii) secant method starting with a = 1, b = 2, rounded-off to three decimal places.

   b. compare the results obtained by (i) and (ii) in part (a).

iv. Starting with $x_0 = 0$ find an approximate root of the equation $x^3 - 4x + 1 = 0$, rounded off to five decimal places using Newton-Raphson method.

v. The motion of a planet in the orbit is governed by an equation of the form $y = x - e \sin x$ where e stands for the eccentricity. Let y = 1 and $e = \frac{1}{2}$. Then find a approximate root of $2x - 2 - \sin x = 0$ in the interval $[0, \pi]$ with error less than 10^{-5}. Start with $x_0 = 1.5$.

vi. Using Newton-Raphson square root algorithm, find the following roots within an accuracy of 10^{-4}.
   i) $8^{1/2}$ starting with $x_0 = 3$
   ii) $91^{1/2}$ starting with $x_0 = 10$

vii. Can Newton-Raphson iteration method be used to solve the equation $x^{1/3} = 0$? Give reasons for your answer.

viii. For the problem given in Example 5, Unit 2, find the number n of bisection required to have an approximate root with absolute error less than or equal to 10^{-7}.

ix. For the equation given in Example 7, show that the iteration $x_{i+1} = \frac{b}{x_i + a}$ will converge to the root $x = \alpha$, when $|\alpha| < |\beta|$.

7.0 REFERENCES/FURTHER READINGS

Wrede, R.C. and Spegel M. (2002). Schaum's and Problems of Advanced Calculus. McGraw – Hill N.Y.

Keisler, H.J. (2005). Elementary Calculus. An Infinitesimal Approach. 559 Nathan Abbott, Stanford, California, USA

UNIT 4 APPROXIMATE ROOTS OF POLYNOMIAL EQUATION

CONTENTS

1.0 INTRODUCTION

In the last two units we discussed methods for finding approximate roots of the equation f(x) = 0. In this unit we restrict our attention to polynomial equations. Recall that a polynomial equation is an equation of the form f(x) = 0 where f(x) is a polynomial in x. Polynomial equation arise very frequently in all branches of science especially in physical applications. For example, the stability of electrical of mechanical systems is related to the real part of one of the complex roots of a certain polynomial equation. Thus there is a need to find all roots, real and complex, of a polynomial equation. The four iteration methods, we have discussed so far, applies to polynomial equations also. But you have seen that all those methods are time consuming. Thus it is necessary to find some efficient methods for obtaining roots of polynomial equations.

The sixteenth century French mathematician Francois Vieta was the pioneer to develop methods for finding approximate roots of polynomial equations. Later, several other methods were developed for solving polynomial equations. In this unit we shall discuss two simple methods: Birge-Vieta's and Graeffe's root squaring methods. To apply these methods we should have some prior knowledge of location and nature of roots of a polynomial equation. You are already familiar with some results regarding location and nature of roots from the elementary algebra course. We shall begin this unit by listing some of the important result about the roots of polynomial equations.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- apply the following methods for finding approximate roots of polynomial equations
  - Birge-Vieta method

- Graeffe's root squaring method
- list the advantages of the above methods over the methods discussed in the earlier units.

3.0 MAIN BODY

3.1 Some Results on Roots of Polynomial Equations

The main contribution in the study of polynomial equations due to the French mathematician Rene Descarte's. The results appeared in the third part of his famous paper 'La geometric' which means 'The geometry'.

Consider a polynomial equation of degree n

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + ... +a_1 x + a_0 \qquad (1)$$

where a_0, a_1, ... a_n are real numbers and $a_n \neq 0$. You know that the roots of a polynomial equation need not be real numbers, it can be complex numbers, that is numbers of the form z = a + ib where a and b are real numbers. The following results are basic to the study of roots of polynomial equations.

Theorem 1: (Fundamental Theorem of Algebra): Let p(x) be a polynomial of degree n ≥ 1 given by Eqn. (1). Then p(x) = 0 has at least one root: that is there exists a number $\alpha \in C$ such that $p(\alpha) = 0$. In fact p(x) has n complex roots which may not be distinct.

Theorem 2: Let p(x) be a polynomial of degree n and α is a real number. Then

$$p(x) = (x - \alpha)\, q_0(x) + r_0 \qquad (2)$$

for some polynomial $q_0(x)$ of degree n – 1 and some constant number r_0 . $q_0(x)$ and r_0 are called the quotient polynomial and the remainder respectively.

In particular, if α is a root of the equation p(x) = 0, then $r_0 = 0$: that is $(x - \alpha)$ divides p(x).

Then we get

$$p(x) = (x - \alpha)\, q_0(x)$$

How do we determine $q_0(x)$ and r_0? We can find them by the method of synthetic division of a polynomial p(x). Let us now discuss the synthetic division procedure.

Consider the polynomial p(x) as given in Eqn. (1)

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + ... +a_1 x + a_0$$

Dividing p(x) by $x - \alpha$ we get

$$p(x) = q_0(x)(x - \alpha) + r_0 \qquad (3)$$

where $q_0(x)$ is a polynomial of degree $n - 1$ and r_0 is a constant.

Let $q_0(x)$ be represented as

$$q_0(x) = b_n x^{n-1} + b_{n-1}x^{n-2} + ... + b_2 x + b_1$$

(Note that for convenience we are denoting the coefficient by $b_1, ..., b_n$ instead of b_0, $b_1, ... b_{n-1}$). Set $b_0 = r_0$. Substituting the expressions for $q_0(x)$ and r_0 in Eqn. (3) we get

$$p(x) = (x - \alpha)(b_n x^{n-1} + b_{n-1}x^{n-2} + ... + b_2 x + b_1) + b_0 \qquad (4)$$

Now, to find $b_0, b_1 ..... b_n$ we simplify the right hand side of Eqn. (4) and compare the coefficients of x^i, $i = 0, 1, ... n$ on both sides. Note that $p(\alpha) = b_0$. Comparing the coefficient we get

Coefficient of x^n : $a_n = b_n \qquad b_n = a_n$
Coefficient of x^{n-1} : $a_{n-1} = b_{n-1} - \alpha b_n, \qquad b_{n-1} = a_{n-1} + \alpha b_n$
.
.
.
Coefficient of x^k : $a_k - b_k - \alpha b_{k+1}, \qquad b_k = a_k + \alpha b_{k+1}$
.
.
.
Coefficient of x^0 : $a_0 = b_0 - \alpha, \qquad b_0 = a_0 + \alpha b_1$

It is easy to perform the calculations if we write the coefficient of p(x) on a line and perform the calculation $b_k = a_k + \alpha b_{k+1}$ below a_k as given in the table below.

Table 1: Horner's table for synthetic division procedure

α	a_n	a_{n-1}	a_{n-2}	...	a_k	...	a_2	a_1	a_0
		αb_n	αb_{n-1}	...	αb_{k+1}	...	αb_3	αb_2	αb_1
	b_n	b_{n-1}	b_{n-2}		b_k		b_2	b_1	$b_0 = p_0(\alpha)$

We shall illustrate this procedure with an example.

Example 1: Divide the polynomial

$p(x) = x^5 - 6x^4 + 8x^3 + 8x^2 + 4x - 40$
by $x - 3$ by the synthetic division method and find the remainder.

Solution: Here p(x) is a polynomial of degree 5. If a_5, a_4, a_3, a_2, a_1, a_0 are the coefficients of p(x), then the Horner's table in this case is

Table 2

a_5	a_4	a_3	a_2	a_1	a_0
1	-6	8	8	4	-40
	3	-9	-3	15	57
1	-3	-1	5	19	17
b_5	b_4	b_3	b_2	b_1	b_1

Hence the quotient polynomial $q_0(x)$ is

$q_0(x) = x^4 - 3x^3 - x^2 + 5x + 19$
and the remainder is $r_0 = b_0 = 17$. thus we have p(3) = b =17.

Theorem 3: Suppose that z = a + ib is a root of the polynomial equation p(x) = 0. Then the conjugate of z, namely $\overline{z}$, = a – ib is also a root of the equation p(x) = 0, i.e. complex roots occur in pairs.

We denote by p(-x) the polynomial obtained by replacing x by –x in p(x). We next give an important Theorem due to Rene Descarte.

Theorem 4: (Descarte's Rule of signs): A polynomial equation p(x) = 0 cannot have more positive roots than the number of changes in sign of its coefficients. Similarly p(x) = 0 cannot have more negative roots than the number of changes in sign of the coefficients of p(-x).

For example, let us consider the polynomial equation

$p(x) = x^4 - 15x^2 + 7x - 11 = 0$
$= 1x^4 - 15x^2 + 7x - 11 = 0$

We count the changes in the sign of the coefficients. Going from left to right there are changes between 1 and -15, between -15 and 7 and between 7 and -11. The total number of changes is 3 and hence it can have at most 3 positive roots. Now we consider

$p(-x) = (-x)^4 - 15(-x)^2 + 7(-x) - 11 = 0$
$= x^4 - 15x^2 - 7x - 11$

Here there is only one change between 1 and -15 and hence the equation cannot have more than one negative root.

We now give another theorem which helps us in locating the real roots.
Theorem 5: Let p(x) = 0 be a polynomial equation of degree n $\geq$ 1. Let a and b be two real numbers with a < b. Suppose further that p(a) $\neq$ 0 and p(b) $\neq$ 0. Then,

i) if p(a) and p(b) have opposite signs, the equation p(x) = 0 has an odd number of roots between a and b.

ii) if p(a) and p(b) have like signs, then p(x) = 0 either has no root or an even number of roots between a and b.

Note: In this theorem multiplicity of the root is taken into consideration i.e. if a is a root of multiplicity k it has to be counted k times.

As a corollary of Theorem 5, we have the following results.

a. Corollary 1: An equation of odd degree with real coefficients has at least one real root whose sign is opposite to that of the last term.

b. Corollary 2: An equation of even degree whose constant term has the sign opposite to that of the leading coefficient, has at least two real roots one positive and the other negative.

c. Corollary 3: the result given in Theorem 5(i) is the generalization of the Intermediate value theorem.

The relationship between roots and coefficients of a polynomial equation is given below.

Theorem 6: Let $\alpha_1, \alpha_2, ..., \alpha_n$ be a roots ($n \geq 1$) of the polynomial equation

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0 = 0$$

Then $\alpha_1 + \alpha_2 + ... + \alpha_n = \frac{-a_{n-1}}{a_n}$

$$\alpha_1\alpha_2 + \alpha_2\alpha_3 + ... + \alpha_{n-1}\alpha_n = \frac{a_{n-2}}{a_n}$$

....

....

$$\alpha_1\alpha_2...\alpha_n = (-1)^n \frac{a_0}{a_n}$$

In the next section we shall discuss one of the simple methods for solving polynomial equations.

3.2 Birge-Vieta Method

We shall now discuss the Birge-Vieta method for finding the real roots of a polynomial equation. This method is based on an original method due to two English mathematicians Birge-Vieta. This method is a modified form of Newton-Raphson method.

Consider now, a polynomial equation of degree n, say
$p_n(x) = a_n x^n + ... + a_1 x + a_0 = 0.$ (5)

Let x_0 be an initial approximation to the root α. The Newton-Raphson iterated formula for improving this approximation is

$$x_i = x_{i-1} - \frac{p_n(x_{i-1})}{p'_n(x_{i-1})}, \quad i = 1, 2, ... \qquad (6)$$

To apply this formula we should be able to evaluate both $p_n(x)$ and $p'_n(x_i)$ at any x_i. The most natural way is to evaluate

$$p_n(x_i) = a_n x_i^n + a_{n-1} x^{n-1} + ... + a_2 x_i^2 + a_1 x_i + a_0$$
$$p'_n(x_i) = n\, a_n x_i^{n-1} + (n - 1) a_{n-1} x_i^{n-2} + ... + 2a_2 x_i + a_1$$

However, this is the most inefficient way of polynomial because of the amount of computations involved and also due to the possible growth of round off errors. Thus there is a need to look for some efficient method for evaluating $p_n(x_i)$ and $p'_n(x_i)$.

Let us consider the evaluation of $p_n(x_i)$ and $p'_n(x_i)$ at x_0 using Horner's method as discussed in the previous section.

We have

$$p_n(x_i) = (x - x_0)\, q_{n-1}(x) + r_0 \qquad (7)$$

where
$q_{n-1}(x) = b_n x^{n-1} + b_{n-2} x^{n-2} + \ldots + b_2 x + b_1$

and $b_0 = p_n(x_0) = r_0$ (8)

We have already discussed in the previous section how to find b_1, I = 1, 2, …, n.

Next we shall find the derivative $p'_n(x_0)$ using Horner's method. We divide $q_{n-1}(x)$ by $(x - x_0)$ using Horner's method. That is, we write

$$q_{n-1}(x) = (x - x_0)\, q_{n-2}(x) + r_1$$
$$q_{n-1}(x) = c_n x^{n-2} + c_{n-1} x^{n-3} + \ldots + c_3 x + c_2.$$

Comparing the coefficients, we get c_i as given in the following table

Table 3

	b_n	b_{n-1}	...	b_k	...	b_2	b_1
x_0		$x_0 c_n$	...	$x_0 c_{k+1}$	...	$x_0 c_3$	$x_0 c_2$
	$c_n = b_n$	c_{n-1}		c_k		c_2	c_1

As observed in Sec. 1, we have

$c_1 = q_{n-1}(x_0)$ (9)

Now, from Eqn. (7) and (8), we have

$p_n(x) = (x – x_0)\, q_{n-1}(x) + p_n(x_0)$ (10)

Differentiating both sides of Eqn. (10) w.r.t.x, we get

$p'_n(x) = q_{n-1}(x) + (x – x_0)\, q'_{n-1}(x)$ (11)

Putting $x = x_0$ in Eqn. (11), we get

$p'_n(x_0) = q_{n-1}(x_0)$ (12)

Comparing (9) and (12), we get

$p'_n(x_0) = q_{n-1}(x_0) = c_1$

Hence the Newton-Raphson method (Eqn. (6)) simplies to

$$x_i = x_{i-1} - \frac{b_0}{c_1} \qquad (13)$$

We summarise the evaluation of b_i and c_i in the following table

Table 4

	a_n	a_{n-1}	...	a_k	...	a_2	a_1	a_0
x_0		x_0b_n	...	x_0b_{k+1}	...	x_0b_3	x_0b_2	x_0b_1
	$a_n=b_n$	b_{n-1}		b_k		b_2	b_1	$b_0=p_n(x_0)$
x_0		x_0c_n	...	x_0c_{k+1}	...	x_0c_3	x_0c_2	
	$c_n=b_n$	c_{n-1}		c_k		c_2		$c_1=p'_n(x_0)$

Let us consider an example.

Example 2: Evaluate p'(3) for the polynomial

$p(x) = x^5 – 6x^4 + 8x^3 + 8x^2 + 4x – 40.$

Solution: Here the coefficients are $a_0 = -40$, $a_1 = 4$, $a_2 = 8$, $a_3 = 8$, $a_4 = -6$ and $a_5 = 1$. To compute b_0, we form the following table.

Table 5

3	1	-6	8	8	4	-40
		3	-9	-3	15	57
3	1	-3	-1	5	19	$17=p(3)=b_0$
		3	0	-3	6	
	1	0	-1	2	$25 = p'(3)=c_1$	

Therefore p'(3) = 25

Now we shall illustrate why this method is more efficient than the direct method. Let us consider an example. Suppose we want to evaluate the polynomial

$$p(x) = -8x^5 + 7x^4 - 6x^3 + 5x^2 - 4x + 3$$

for any given x.

When we evaluate by direct method, we compute each power of x by multiplying with x the preceding power of x as

$x^3 = x(x^2)$, $x^4 = x(x^3)$ etc.

Thus each term c^k takes two multiplications for $k > 1$. Then the total number of multiplications involved in the evaluation pf p(x) is 1 + 2 + 2 + 2 + 2 = 9.

When we use Horner's method the total number of multiplications in 5. The number of additions in both cases are the same. This shows that less computation is involved while using Horner's method and therapy reduces the error in computation.

Let us now solve some problems using Birge-Vieta method.

Example 3: Use Birge-Vieta method to find all the positive real roots, rounded off to three decimal places of the equation

$$x^4 + 7x^3 + 24x^2 + x - 15 = 0$$

Stop the iteration whenever $| x_{i+1} - x_i | < 0.0001$

Solution: We first note that the given equation

$p4(x) = x^4 + 7x^3 + 24x^2 + x - 15 = 0$

is of degree 4. Therefore, by Theorem 1, this equation has 4 roots. Since there is only one change of sign in the coefficients of this equation, Descarte's rule of signs (see Theorem 4), states that the equation can have at most one positive real root.

Now let us examine whether the equation has a positive real root.

Since p4(0) = -15 and p4(1) = 19, by Intermediate value theorem, the equation has a root lying in]0, 1[.

We take $x_0 = 0.5$ as the initial approximation to the root. The first iteration is given by

$$x_1 = x_0 - \frac{p_4(x_0)}{p'_4(x_0)}$$

$$= 0.5 - \frac{p_4(0.5)}{p'_4(0.5)}$$

Now we evaluate $p_4(0.5)$ and $p'_4(0.5)$ using Horner's method. The results are given in the following table.

Table 6

	1	7	24	1	-15
0.5		0.5	3.75	13.875	7.4375
	1	7.5	27.75	14.875	$-75625 = p_4(0.5)$
0.5		0.5	4.00	15.875	
	1	8.0	31.75	$30.750 = p'_4(0.5)$	

Therefore $x_1 = 0.5 - \frac{-7.5625}{30.75} = 0.7459$

The second iteration is given by

$$x_2 = x_1 - \frac{p_4(x_1)}{p'_4(x_1)} = 0.7459 - \frac{p_4(0.7459)}{p'_4(0.7459)}$$

Uisng synthetic division, we form the following table of values

Table 7

	1	7	24	1	-15
0.7459		0.7459	5.7777	22.2119	17.3138
	1	7.7459	29.7777	23.2119	2.3138
0.7459		0.7459	6.3340336		26.935717
	1	8.4918	36.111701		50.146879

Therefore $x_2 = 0.7459 - \frac{2.3132}{50.1469} = 0.6998$

Third iteration is given by

$$x_3 = x_2 - \frac{p_4(0.6998)}{p'_4(0.6998)}$$

Table 8

	1	7	24	1	-15
0.6998		0.6998	5.3881	20.5649	15.0905
	1	7.6998	29.3881	21.5649	0.0905
0.6998		.6998	5.8778	24.6780	
	1	8.3996	35.2659	46.2429	

$$x_3 = 0.6998 - \frac{0.0905}{46.2429} = 0.6978$$

For the fourth iteration we have

$$x_4 = x_3 - \frac{p_4(0.6978)}{p'_4(0.6978)}$$

Table 9

	1	7	24	1	-15
0.6978		0.6978	5.3715248	20.495459	14.999525
	1	7.6978	29.3715248	21.495459	0.0905
0.6978		.6978	5.8584497	24.583476	
	1	8.3956	35.229975	46.078926	

$x_4 = 0.6978 - \dfrac{0.0005}{46.0789} = 0.6978$

Since x_3 and x_4 are the same, we get $| x_4 - x_3 | < 0.0001$ and therefore we stop the iteration here. Hence the approximate value of the root rounded off to three decimal places is 0.698.
Next we shall illustrate how Birge-Vieta's method helps us to find all real roots of a polynomial equation.

Consider Eqn. (4)

$p(x) = (x - \alpha) (b_n x^{n-1} + b_{n-1} x^{n-2} + ... + b_2 x + b_1) + b_0$

If α is a root of the equation $p(x) = 0$, then $p(x)$ is exactly divisible by $x - \alpha$, that is, $b_0 = 0$. In finding the approximations to the root by the Birge-Vieta method, we find that b_0 approaches zero ($b_0 \to 0$) as x_i approaches α ($x_i \to \alpha$). Hence, if x_n is taken as the final approximation, to the root satisfying the criterion $| x_n - x_{n-1} | < \varepsilon$, then to this approximation, the required quotient is

$q_{n-1}(x) = b_n x^{n-1} + b_{n-1} x^{n-2} + ... + b_1$

where b'_1 are obtained by using x_n and the Horner's method. This polynomial is called the deflated polynomial or reduced polynomial. The next root is now obtained using $q_{n-1}(x)$ and not $p_n(x)$. Continuing this process, we can successively reduce the degree of the polynomial and find one real root at a time.

Let us consider an example.

Example 4: Find all the roots of the polynomial equation $p_3(x) = x^3 + x - 3 = 0$ rounded off to three decimal places. Stop the iteration whenever $|x_{i+1} - x_i| < 0.0001$.

Solution: The equation $p_3(x) = 0$ has three root. Since there is only one change in the sign of the coefficients, by Decarts' rule of signs the equation can have at most one positive real root. The equation has no negative real root since $p_3(-x) = 0$ has no change of sign of coefficients. Since $p_3(x) = 0$ is of odd degree it has at least one real root. Hence the given equation $x^3 + x - 3 = 0$ has one positive real root and a complex pair. Since $p(1) = -1$ and $p(2) = 7$, by intermediate value theorem the equation has a real root lying in the interval]1, 2[. Let us find the real root using Birge-Vieta Method. Let the initial approximation be 1.1.

First iteration

Table 10

	1	0	14	-3
1.1		1.1	1.21	2.431
	1	1.1	2.21	0.0905
1.1		1.1	2.42	
	1	2.2	4.63	

Therefore $x_1 = 1.1 - \frac{-0.569}{4.63} = 1.22289$

Similarly, we obtain

$x_2 = 1.21347$
$x_3 = 1.21341$

Since $| x_2 – x_3 | < 0.0001$, we stop the iteration here. Hence the required value of the root is 1.213, rounded off to three decimal places. Next let us obtain the deflated polynomial of $p_3(x)$. To get the deflated polynomial of, we have to find the polynomial $q_2(x)$ by using the final approximation $x_3 = 1.213$ (see Table 11).

Table 11

	1	0	1	-3
1.213		1.213	1.4714	2.9978
	1	1.213	2.4714	-0.0022

Note that $p_3(1.213) = -0.0022$. That is, the magnitude of the error in satisfying $p_3(x_3) = 0$ is 0.0022.
We find $q_2(x) = x^2 + 1.213x + 2.4714 = 0$

This is a quadratic equation and its roots are given by

$$x = \frac{-1.213 \pm \sqrt{(1.213)^2 - 4 \times 2.4714}}{2}$$

$$= \frac{-1.213 \pm 2.9009i}{2}$$

$$= 0.\ 6065 \pm 1.4505\ i$$

Hence the three roots of the equation rounded off to three deciml places are 1.213, 0.6065 + 1.4505 i and -0.6065 – 1.4505 i.

Remark: We now know that we can determine all the real roots of a polynomial equation using deflated polynomials. This procedure reduces the amount of computations also. But this method has certain limitations. The computations using deflated polynomial can cause unexpected errors. If the roots are determined only approximately, the coefficients of the deflated polynomials will contain some errors due to rounding off. Therefore we can expect loss of accuracy in the remaining roots. There are some ways of minimizing this error. We shall not be going into the details of these refinements.

3.3 Graeffe's Root Squaring Method

In the last section we have discussed a method for finding real roots of polynomial equations. Here we shall discuss a direct method for solving polynomial equations. This method was developed independently by three mathematicians Dandelin, Lobachesky and Graeffe. But Graeffe's name is usually associated with this method. The advantage of this method is that it finds all roots of a polynomial equation simultaneously: the roots may be real and distinct, real and equal (multiple) or complex roots.

The underlying idea of the method is based on the following fact: Suppose β_1, β_2, ..., β_n are the n real and distinct roots of a polynomial equation of degree n such that they are widely separated, that is,

$$|\beta_1| >> |\beta_2| >> |\beta_3| >> ... >> |\beta_n|$$

where >> stands for 'much greater than'. Then we can obtain the roots approximately from the coefficients of the polynomial equation as follows:

Let the polynomial equation whose roots are β_1, β_2, ..., β_n be

$$a_0 + a_1x + a_2x^2 + ... + a_nx^n = 0, a_n \neq 0.$$

Using the relations between the roots and the coefficients of the polynomial as given in Sec. 4.2, we get

$$\left.\begin{aligned}
&\beta_1 + \beta_2 + \ldots + \beta_n = -\frac{a_{n-1}}{a_n} \\
&\beta_1 , \beta_2 + \beta_1\beta_3 + \ldots + \beta_{n-1}\beta_n = \frac{a_{n-2}}{a_n} \\
&\beta_1\beta_2\beta_3 + \ldots + \beta_{n-2}\beta_{n-1}\beta_n = -\frac{a_{n-3}}{a_n} \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\beta_1\beta_2 \ldots\beta_n = (-1)^n \frac{a_0}{a_n}
\end{aligned}\right\} \qquad (14)$$

Since $|\beta_1| >> |\beta_2| >> |\beta_3| >> \ldots >> |\beta_n|$, we have from (14) the approximations

$$\left.\begin{aligned}
&\beta_1 \approx -\frac{a_{n-1}}{a_n} \\
&\beta_1\beta_2 \approx \frac{a_{n-2}}{a_n} \\
&\beta_1\beta_2\beta_3 \approx -\frac{a_{n-3}}{a_n} \\
&\ldots \\
&\ldots \\
&\beta_1\beta_2 \ldots\beta_n \approx (-1)^n \frac{a_0}{a_n}
\end{aligned}\right\} \qquad (15)$$

These approximations can be simplified as

$$\left.\begin{aligned}
&|\beta_1| \approx -\frac{a_{n-1}}{a_n} \\
&|\beta_2| \approx \frac{a_{n-2}}{a_n}\frac{a_n}{a_{n-1}} \approx \frac{a_{n-2}}{a_{n-1}} \\
&|\beta_3| \approx -\frac{a_{n-3}}{a_n}\frac{a_{n-1}}{a_{n-2}}\frac{a_n}{a_{n-1}} = \frac{a_{n-3}}{a_{n-2}} \\
&. \\
&. \\
&. \\
&|\beta_n| \approx \frac{a_0}{a_1}
\end{aligned}\right\} \qquad (16)$$

So the problem now is to find from the given polynomial equation, a polynomial equation whose roots are widely separated. This can be done by the method which we shall describe now.

In the present course we shall discuss the application of the method to a polynomial equation whose roots are real and distinct.

Let α_1, α_2, ..., α_n be the n real and distinct roots of the polynomial equation of degree n given by

$$a_0 + a_1x + a_2x^2 + ... + a_nx^n = 0. \qquad (17)$$

where a_0, a_1, a_2, ..., a_{n-1}, a_n are real numbers and $a_n \neq 0$. We rewrite Eqn. (17) by collecting all even terms on one side and all odd terms on the other side, i.e.

$$a_0 + a_2x^2 + a_4x^4 + ... = -(a_1x + a_3x^3 + a_5x^5 + ...) \qquad (18)$$

Squaring both sides of Eqn. (18), we get

$$(a_0 + a_2x^2 + a_4x^4 + ...)^2 = (a_1x + a_3x^3 + a_5x^5 + ...)^2$$

Now we expand both the right and left sides and simplify by collecting the coefficients. We get

$$a_0^2 - (a_1^2 - 2a_0a_2)x^2 + (a_2^2 - 2a_1a_3 + 2a_0a_4)x^4 -$$
$$(a_3^2 - 2a_2a_4 + 2a_1a_5 - 2a_0a_6)x^6 + ... + (-1)^n a_n^2 x^{2n} = 0 \qquad (19)$$

Putting $x^2 = -y$ in Eqn. (19), we obtain a new equation gien by

$$b_0 + b_1y + b_2y^2 + ... + b_n = 0 \qquad (20)$$

where

$b_0 = a_0^2$

$b_1 = a_1^2 - 2a_0a_2$

$b_2 = a_2^2 - 2a_1a_3 + 2a_0a_4$

$b_n = a_n^2$

The following table helps us to compute the coefficients b_0, b_1, ..., b_n of Eqn. (20) directly from Eqn. (17).

Table 12

a_0	a_1	a_2	a_3...	a_n
a_0^2	a_1^2	a_2^2	a_3^2	a_n^2
0	$-2a_0a_2$	$-2a_1a_3$	$-2a_2a_4$	0
0	0	$-2a_0a_4$	$-2a_1a_5$	0
0	0	0	$-2a_0a_6$	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
b_0	b_1	b_2	b_3...	b_n

To form Table 12 we first write the coefficients a_0, a_1, a_2, ..., a_n as the first row. Then we form (n + 1) columns as follows.

The terms in each column alternate in sign starting with a positive sign. The first term in each column is the square of the coefficients a_k, k = 0, 1, 2, ..., n. The second term in each column is twice the product of the nearest neghbouring coefficients, if there are nay with negative sign: otherwise put it as zero. For example, the second term in the first column is zero and second term in the second column is $-2a_0 a_2$. Likewise the second term of the (k + 1)th column is $2a_{k-1} a_{k+1}$. The third term in the (k + 1)th column is twice the product of the next neighbouring coefficients a_{k-2} and a_{k+2}, if there are nay, otherwise put it as zero. This procedure is continued until there are no coefficients available to form the cross products. Then we add all the term in each column. The sum gives the coefficients b_k for k = 0, 1, 2, ..., n which are listed as the last term in each column. Since the substitution $x^2 = -y$ is used, it is easy to see that if $\alpha_1, \alpha_2, ..., \alpha_n$ are the n roots of Eqn. (17), then $-\alpha_1^2, \alpha_2^2, ..., \alpha_n^2$ are the roots of Eqn. (20).

Thus, starting with a given polynomial equation, we obtained another polynomial equation whose roots are the squares of the roots of the original equation with negative sign.

We repeat the procedure for Eqn. (20) and obtain another equation

$c_0 + c_1x + ... + c_nx^n = 0.$

Whose roots are the squares of the roots of Eqn. (20) with a negative sign i.e. they are fourth powers of the roots of the original equation with a negative sign. Let this procedure be repeated n times. Then, we obtain an equation

$q_0 + q_1x + ... + q_nx^n = 0$ (21)

whose roots $\gamma_1, \gamma_2, ..., \gamma_n$ are given by

γ_i =, i = 0, 1, 2, ..., n. (22)

Now, since all the roots of Eqn. (17) are real and distinct, we have

$|\alpha_1| > |\alpha_2| > \ldots\ldots\ldots > |\alpha_n|$

Hence $|\gamma_1| = |\alpha_1^{2^m}| = \left|\dfrac{q_{n-1}}{q_n}\right|$

$|\gamma_2| = |\alpha_2^{2^m}| = \left|\dfrac{q_{n-2}}{q_{n-1}}\right|$

. . .

. . .

. . .

$|\gamma_n| = |\alpha_n^{2^m}| = \left|\dfrac{q_0}{q_1}\right|$

The magnitude of the roots of the original equations are therefore given by

$|\alpha_1| = \sqrt[2^m]{\dfrac{q_{n-1}}{q_n}}$

$|\alpha_2| = \sqrt[2^m]{\dfrac{q_{n-2}}{q_{n-1}}}$

.

.

.

$|\alpha_n| = \sqrt[2^m]{\dfrac{q_0}{q_1}}$

This gives the magnitude of the roots. To determine the sign of the roots, we substitute these approximations in the original equation and verify whether positive or negative value satisfies it.

We shall now illustrate this method with an example.

Example 5: Find the roots of the cubic equation $x^3 – 15x^2 + 62x – 72 = 0$ by Graeffe's method using three squaring.

Solution: Let $P_3(x) = x^3 – 15x^2 + 62x – 72 = 0$.

The equation has no negative real roots. Let us now apply the root squaring method successively. The get the following results:

First Squaring

Table 13

a_0	a_1	a_2	a_3
-72	62	-15	1
a_0^2 =5184	a_1^2=3844	a_2^2=225	a_3^2=1
0	$-2a_0a_2$=-2160	$-2a_1a_3$=-124	0
5184	1684	101	1
b_0	b_1	b_2	b_3

Therefore the new equation is

$x^3 + 101x^2 + 168x + 5184 = 0.$

Applying the squaring method to the new equation we get the following results.

Second Squaring

Table 14

5184	1684	101	1
26873856	2835856	10201	1
0	-1047168	-3368	0
26873856	1788688	6833	1

Thus the new equation is
$x^3 + 6833x^2 + 1788688x + 26873856 = 0.$

For the third squaring, we have the following results.

Third Squaring

Table 15

26873856	1788688	6833	1
7.2220414×10^{14}	3.1994048×10^{12}	46689889	1
0	-3672581×10^{12}	-3577376	0
7.2220414×10^{14}	2.83214×10^{12}	43112513	1
q_0	q_1	q_2	q_3

Hence the new equation is

$x^3 + 43112513x^2 + (2.83214 \times 10^{12})x + (7.2220414 \times 10^{14}) = 0$

After three squaring, the roots γ_1, γ_2, and γ_3 of this equation are given by

$$|\gamma_1| = \left|\frac{q_2}{q_3}\right| = 43112513$$

$$|\gamma_2| = \left|\frac{q_1}{q_2}\right| = \frac{2.83214 \times 10^{12}}{43112513}$$

$$|\gamma_3| = \left|\frac{q_0}{q_1}\right| = \frac{7.22204 \times 10^{14}}{2.83214 \times 10^{12}}$$

Hence, the roots

$$|\alpha_1| = \sqrt[8]{443112513} = 9.0017$$

$$|\alpha_2| = \sqrt[8]{\frac{2.83214 \times 10^{12}}{43112513}} = 4.0011$$

$$|\alpha_3| = \sqrt[8]{\frac{7.22204 \times 10^{14}}{2.83214 \times 10^{12}}} = 1.9990$$

Since the equation has no negative real roots, all the roots are positive. Hence the roots can be taken as 9.0017, 4.0011 and 1.9990. If the approximations are rounded to 2 decimal places, we have the roots as 9, 4 and 2. Alternately, we can substitute the approximate roots in the given equation and find their sign.

4.0 CONCLUSION

We have seen that Graeffe's root squaring method obtain all real roots simultaneously. There is considerable saving in time also. The method can be extended to find multiple and complex roots also. However the method is not efficient to find these roots. We shall not discuss these extensions.

We shall end this block by summarizing what we have covered in this unit.

5.0 SUMMARY

In this unit we have:

- discussed the following methods for finding approximate roots of polynomial equations.
  - i) Birge-Vieta method.
  - ii) Graeffe's root squaring method.
- Mentioned the advantage and disadvantages of the above methods.

6.0 TUTOR-MARKED ASSIGMENT (TMA)

i Find the quotient and the remainder when $2x^3 - 5x^2 + 3x - 1$ is divided by $x - 2$.

ii. Using synthetic division check whether $\alpha_0 = 3$ is a root of the polynomial equation $x^4 + x^3 - 13x^2 - x + 12 = 0$ and find the quotient polynomial.

iii. How many negative roots does the equation $3x^7 + 5x^5 + 4x^3 + 10x - 6 = 0$ have? Also determine the number of positive roots, if any.

iv. Show that the biquadratic equation
$p(x) = x^4 + x^3 - 2x^2 + 4x - 24 = 0$ has at least two real roots one positive and the other negative.

v. Using synthetic division, show that 2 is a simple root of the equation
$p(x) = x^4 - 2x^3 - 7x^2 + 8x + 12 = 0$.

vi. Evaluate p(0.5) and p'(0.5) for
$p(x) = -8x^5 + 7x^4 - 6x^3 + 5x^2 - 4x + 3$

vii. Find an approximation to one of the roots of the equation
$p(x) = 2x^4 - 3x^2 + 3x - 4 = 0$
using Birge-Vieta method starting with the initial approximation $x_0 = -2$. Stop the iteration whenever $| x_{i+1} - x_i | < 0.4 \times 10^{-2}$.

viii. Find all the roots of the equation $x^3 - 2x - 5 = 0$ using Birge-Vieta method.

ix. Find the real root rounded off to two decimal places of the equation $x^4 - 4x^3 - 3x + 23 = 0$ lying in the interval]2, 3[by Birge-Vieta method.

x. Determine all roots of the following equations by Graeffe's root squaring method using three squaring.
i) $x^3 + 6x^2 - 36x + 40 = 0$
ii) $x^3 - 2x^2 - 5x + 6 = 0$
iii) $x^3 - 5x^2 - 17x + 20 = 0$

7.0 REFERENCES/FURTHER READINGS

Wrede, R.C. and Spegel M. (2002). Schaum's and Problems of Advanced Calculus. McGraw – Hill N.Y.

Keisler, H.J. (2005). Elementary Calculus. An Infinitesimal Approach. 559 Nathan Abbott, Stanford, California, USA