Introduction

Numerical analysis, area of mathematics and computer science that creates, analyzes, and implements algorithms for obtaining numerical solutions to problems involving continuous variables. Such problems arise throughout the natural sciences, social sciences, engineering, medicine, and business. Since the mid 20th century, the growth in power and availability of digital computers has led to an increasing use of realistic mathematical models in science and engineering.

The goal of this course of lectures is to introduce some of the most important and basic numerical algorithms that are used in practical computations. Beyond merely learning the basic techniques, it is crucial that an informed practitioner develop a thorough understanding of how the algorithms are constructed, why they work, and what their limitations are.

Numerical methods are mathematical techniques used for solving mathematical problems that cannot be solved or are difficult to solve. The numerical solution is an approximate numerical value for the solution. Although numerical solutions are an approximation, they can be very accurate.

Absolute and relative errors.

Since numerical solutions are an approximation, and since the computer program that executes the numerical method might have errors, a numerical solution needs to be examined closely.

Definition 1: The error is the difference between the true value, x, and approximate value, x_a.

$$e = \Delta x = x - x_a$$

Definition 2: The absolute error, δ, is the magnitude of the error

$$\delta = |x - x_a|.$$

Definition 3: The relative error, ε, is the absolute error divided by the magnitude of the value.

$$\varepsilon = \varepsilon_t = \left|\frac{x - x_a}{x}\right|$$

Example 1. Suppose we measure the length of a table to be $x_a = 59.7$ in, but the true length is $x = 59.5$ in. Find the error, absolute error and relative error?

Solution:

$$e = x - x_a = 59.5 - 59.7 = -0.2\ in$$

$$\delta = |x - x_a| = |-0.2| = 0.2\ in.$$

$$\varepsilon_t = \left|\frac{x - x_a}{x}\right| = \left|\frac{0.2}{59.5}\right| = 0.003361 \ldots. \approx 0.3361..\%$$

Example 2. Suppose the measured value of the temperature is $T_a = 146.2$, but the true temperature is $T = 145.9$. What are the error, the absolute error, and the relative error?

Solution:

The error is $\Delta T = T - T_a = 145.9 - 146.2 = -0.3$.

The absolute error is $\delta = |T\text{-}T_a| = |\text{-}0.3| = 0.3$.

The relative error is $\varepsilon_t = \left|\frac{T-T_a}{T}\right| = \left|\frac{0.3}{145.9}\right| = 0.002056 \approx 0.2\%$.

Errors in numerical solutions

Error in solving an engineering or science problem can arise due to several factors. First, the error may be in the modeling technique. A mathematical model may be based on using assumptions that are not acceptable. For example, one may assume that the drag force on a car is proportional to the velocity of the car, but actually it is proportional to the square of the velocity of the car. This itself can create huge errors in determining the performance of the car, no matter how accurate the numerical methods you may use are. Second, errors may arise from mistakes in programs themselves or in the measurement of physical quantities. But, in applications of numerical methods itself, the two errors we need to focus on are.

1. Round off error
   Round-off error occurs because of the computing device's inability to deal with certain numbers. Such numbers need to be rounded off to some near approximation which is dependent on the word size used to represent numbers of the device. For instance:
   $\frac{1}{9} = 0.1111 \ldots \approx 0.111$ .Then the round off error is $\geq 1.10^{-4}$
2. Truncation error
   The word 'Truncate' means 'to shorten'. Truncation error refers to an error in a method, which occurs because some number/series of steps (finite or infinite) is truncated (shortened) to a fewer number. Such errors are essentially algorithmic errors and we can predict the extent of the error that will occur in the method. For example, the Maclaurin series for e^x is given as
   $e^x = 1 + x + \frac{x^2}{2!} + \ldots\ldots + \frac{x^n}{n!} + \ldots.$
   This series has an infinite number of terms but when using this series to calculate e^x , only a finite number of terms can be used. If the formula is used to calculate e 0.3 we get
   $e^{0.3} = 1 + 0.3 + \frac{0.3^2}{2!} + \ldots\ldots + \frac{0.3^n}{n!} + \ldots.$

Where do we stop the calculation? How many terms do we include? Theoretically the calculation will never stop. There are always more terms to add on. If we do stop after a finite number of terms, we will not get the exact answer. For example if we do take the first four terms as the approximation we get

$r = e^{0,3} \approx 1 + 0.3 + \frac{0.3^2}{2!} + \frac{0.3^3}{3!}$.

Then the Truncation error $= e^x - r = \frac{0.3^4}{4!} + \frac{0.3^5}{5!} + \ldots$.

Note that the errors defined above cannot be determined in problems that require numerical methods for their solution, this is because the exact solution is not known. These error quantities are useful for evaluating the accuracy of different numerical methods when the exact solution is known (problem solved analytically).

Since the true errors cannot, in most cases, be calculated, other means are used for estimating the accuracy of a numerical solution

Chapter 1

Solutions Nonlinear Equation in One Variable

Introduction

In this chapter we consider one of the most basic problems of numerical approximation, the **root-finding problem**. This process involves finding a **root**, or solution, of an equation of the form $f(x) = 0$, for a given function f. A root of this equation is also called a **zero** of the function f, where f is a nonlinear function with respect to x, and x is the only independent variable.

1- The Bisection Method

The first technique, based on the Intermediate Value Theorem, is called the Bisection, or Binary-search, method.
Suppose f is a continuous function defined on the interval $[a, b]$, with $f\ (a)$ and $f\ (b)$ of opposite sign. The Intermediate Value Theorem implies that a number c exists in (a, b) with $f\ (p) = 0$. Although the procedure will work when there is more than one root in the interval (a, b), we assume for simplicity that the root in this interval is unique. The method calls for a repeated halving (or bisecting) of subintervals of $[a, b]$ and, at each step, locating the half containing p.

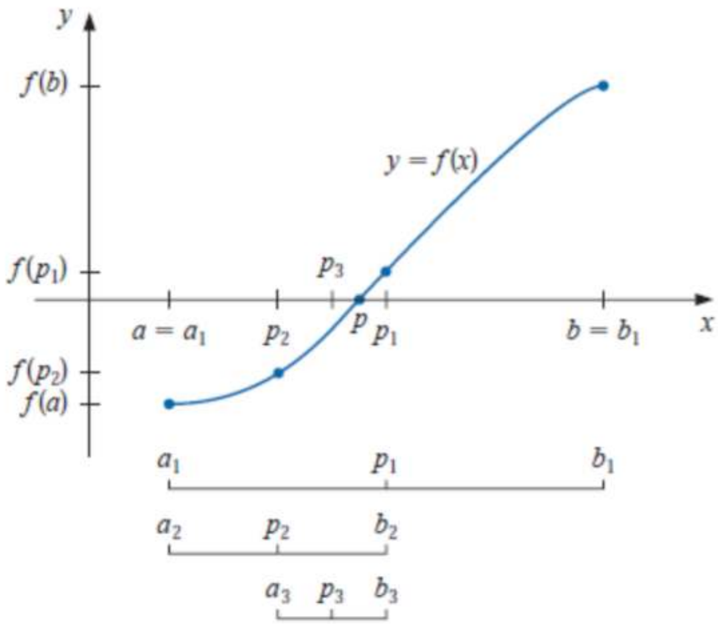
To begin, set $a_1 = a$ and $b_1 = b$, and let p_1 be the midpoint of $[a, b]$; that is,

$$p_1 = a_1 + \frac{b_1 - a_1}{2} = \frac{a_1 + b_1}{2}$$

- If $f\ (p_1) = 0$, then $p = p_1$, and we are done.
- If $f(p_1) \neq 0$, then $f(p_1)$ has the same sign as either $f(a_1)$ or $f(b_1)$.
- If $f(p_1)$ and $f(a_1)$ have the same sign, $p \in (p_1, b_1)$. Set $a_2 = p_1$ and $b_2 = b_1$.

· If $f(p_1)$ and $f(a_1)$ have opposite signs, $p \in (a1, p1)$. Set $a_2 = a_1$ and $b_2 = p_1$.

Then reapply the process to the interval $[a_2, b_2]$. This produces the method described in Algorithm 1



Algorithm 1:

To find a solution to $f(x) = 0$ given the continuous function f on the interval $[a, b]$, where $f(a)$ and $f(b)$ have opposite signs:

INPUT endpoints a, b; tolerance TOL; maximum number of iterations N_0.

OUTPUT approximate solution p or message of failure.

Step 1 Set $i = 1$;

$FA = f(a)$.

Step 2 While $i \leq N_0$ do Steps 3–6.

Step 3 Set $p = a + (b - a)/2$; (Compute p_i.)

$FP = f(p)$.

Step 4 If $FP = 0$ or $(b - a)/2 < TOL$ then

OUTPUT (p); (Procedure completed successfully.)

STOP.

Step 5 Set $i = i + 1$.

Step 6 If $FA \cdot FP > 0$ then set $a = p$; (Compute a_i, b_i.)

$FA = FP$
else set $b = p$. (FA is unchanged.)
Step 7 OUTPUT ('Method failed after N_0 iterations, $N_0 =$ ', N_0);
(The procedure was unsuccessful.)
STOP.
Other stopping procedures can be applied in Step 4 of Algorithm 1 or in any of the iterative techniques in this chapter. For example, we can select a tolerance $\varepsilon > 0$ and generate $p_1, \ldots, p_N$ until one of the following conditions is met:

$$|p_N - p_{N-1}| < \varepsilon, \qquad (1)$$

$$\frac{|p_N - p_{N-1}|}{|p_N|} < \varepsilon, \quad |p_N| \neq 0 \qquad (2)$$

$$|f(p_N)| < \varepsilon. \qquad (3)$$

Example:
By use the Bisection method find the root of the function
$f(x) = x^3 - 4x - 9 = 0.$
Solution:

Example:

Show that $f(x) = x^3 + 4x^2 - 10 = 0$ has a root in $[1, 2]$, and use the Bisection method to determine an approximation to the root that is accurate to at least within 10^{-4}.

1- Find the approximated value of x till 6 iterations for $x^3-4x+9=0$ using Bisection Method. Take a = -3 and b = -2.

Answer

Iteration table is given as follows:

No.	a	b	c	f(a)*f(c)
1	-3.0	-2.0	-2.5	(-)
2	-3.0	-2.5	-2.75	(-)
3	-2.75	-2.5	-2.625	(-)
4	-2.75	-2.625	-2.6875	(-)
5	-2.75	-2.6875	-2.71875	(+)
6	-2.71875	-2.6875	-2.703125	(-)

Hence we stop the iterations after 6. Therefore the approximated value of x is -2.703125.

2- A function is defined as $f(x) = x^3 – x – 11 = 0$. Between the interval [2,3] find the root of the function by Bisection Method upto 8 iteration?

Answer

The Iteration Table is given as follows:

No.	a	b	c	f(a)*f(c)
1	2.0	3.0	2.5	(-)
2	2.0	2.5	2.25	(+)
3	2.25	2.5	2.375	(+)
4	2.375	2.5	2.4375	(-)
5	2.375	2.4375	2.40625	(-)
6	2.375	2.40625	2.390625	(-)
7	2.375	2.390625	2.3828125	(-)
8	2.375	2.3828125	2.37890625	(-)

Thus, with the 9th iteration, we note that the final value. Hence 2.37890625 is our approximation of the root.

3- Find the root of xe^{-x} -0.3 = 0 using Bisection Method in the interval [1,5].

Answer

Iteration table is given as follows:

No.	a	b	c	f(a)*f(c)
1	1.0	5.0	3.0	(-)
2	1.0	3.0	2.0	(-)
3	1.0	2.0	1.5	(+)
4	1.5	2.0	1.75	(+)
5	1.75	2.0	1.875	(-)
6	1.75	1.875	1.8125	(-)
7	1.75	1.8125	1.78125	(+)
8	1.78125	1.8125	1.796875	(-)
9	1.78125	1.796875	1.7890625	(-)

After 9 iterations the difference between the last 2 values is less than 0.01. Thus the approximated value of x is 1.7890625.