تحليلات عددية

أ.م.عبدالله إبراهيم عبدالله السبعاوي

كلية هندسة الالكترونيات
قسم هندسة النظم والسيطرة

Contains

Chapter 1: Introduction

Types and sources of errors

Chapter2: Numerical solutions for nonlinear equations

- Fixed point Method
- Newton-Raphson
- Secant Method

Chapter 3: Numerical solutions of linear systems

- Direct methods
- Gauss Elimination Method
- Gauss-Jordan Method
- Indirect methods
- Gauss – Sidel algorithm

Chapter 4: Numerical Solutions of First Order Ordinary Differential Equations

- Euler Method
- Modified Euler Method
- Runge-Kutta Method of order2
- Runge-Kutta Method of order4

Chapter 5: Numerical integration

- Trapezoidal method
- Composite Trapezoidal method
- Simpson Method
- Composite Simpson method

Chapter One
Introduction

Numerical Analysis is a field of mathematics that concerned with the study of approximate solutions of mathematical problems, where it is difficult or impossible to find the exact solutions for these problems.

For instance, we can't find the exact value of the following integral

, by using known integration methods, $\int_0^1 e^{x^2}\,dx$

while we can find an approximate value for this integral, using numerical methods.

The other example, if we aim to find the solution of a linear system $Ax = b$, where $A \in R^{n\times n}$, when n is too large , it so difficult to find the exact solutions for this system by hand using known methods, so in this case, it is easier to think about how to find the approximate solution using a suitable algorithm and computer programs.

The importance of Numerical Analysis

To interpret any real phenomena, we need to formulate it, in a mathematical form. To give a realistic meaning for these phenomena we have to choose complicated mathematical models, but the problem is, it so difficult to find explicit formulas to find the exact solutions for these complicated models. Therefore, it might be better to find the numerical solutions for complicated form rather than finding the exact solutions for easier forms that can't describe these phenomena in realistic way.

The nature of Numerical analysis

Since for any numerical Algorithm (the steps of the numerical method), we have lots of mathematical calculations, we need to choose a suitable computer language such as Matlab or Mable and write the algorithm processes in programing steps.

In fact, the accuracy of numerical solutions, for any problem, is controlled by three criteria:

1- The type of algorithm,

2- The type of computer language and programs,

3- The advancement of the computers which are used.

Types and sources of Errors

When we compute the numerical solutions of mathematical model that we use to describe a real phenomenon, we get some errors; therefore we should study the types and sources of these errors.

We can point out the most important types and sources of these errors as follows:

1- *Rounding errors*: this type of errors can be got, because of the rounding of numbers in computer programming languages.

Example:- 5.99…9 round to 6, and 3.0001 round to 3.

2- *Truncation Errors:* Since most of numerical algorithms depend on writing the functions as infinite series, and since it's impossible to take more than few terms of these series when we formulate the algorithm, therefore, we get errors, called truncation errors.

Example :-

$$\mathbf{f(x) = e^{x^2} = 1 + x + x^2/2! + x^3/3! + \cdots}$$

If we compute $f(5)$, with taking 4 terms, we get more errors than with taking 10 terms.

3-*Total errors*:- Since any numerical algorithm is about iterative presses, the solution in step n depends on the solution in step $(n-1)$. Therefore, for larger number of steps we get more errors, and those errors are the total of all previous types of errors.

Let $\bar{x}$ is the approximate value of x, there are two methods can be used to measure the errors:-

1- Absolute Error:- $\boldsymbol{E_x = |x - \bar{x}|}$,

2- Relative error:- $\boldsymbol{R_x = \frac{E_x}{|x|} = \frac{|x-\bar{x}|}{|x|}}$

Example:- Let $\bar{x} = 3.14, \quad x = 3.141592.$ Find the Absolute and Relative errors

Solution

$E_x = |x - \bar{x}| = 0.001592$,

$R_x = \frac{0.001592}{3.141592} = 0.000507$

Remark: Clearly, $R_x < E_x$

Questions

Q1:

i- What is numerical analysis concerned with?, and what is the importance of studying this subject ?,

ii- What are the most important types of errors that arise from using a numerical method to compute the numerical solution of a mathematical problem?

iii- What are the criteria that control the accurse of numerical solutions?

Q2: Let E_x, R_x be the absolute and relative errors, respectively, in an approximate value of x . Show that $R_x \leq E_x$, if $|x| \geq 1$.

Numerical Analysis

2.Roots of Single Equations

2.1 Fixed-Point Iteration Method:

The method requires one initial guess only.

►Procedure:

Consider the equation f(x) = 0

1- Re-write the equation as x = g(x)

2- Assume an initial guess for the root = x_0 and calculate the first estimate of the root x_1 from: $x_1=g(x_0)$

3- Repeat step 2 several times until convergence is achieved, i.e.

$$x_{i+1}=g(x_i) \text{ until } \epsilon_a < \epsilon_s$$

Convergence Condition: |g(x)| < 1 in the region of interest.

Example1: Use the fixed-point iteration method to estimate the root of $f(x)=e^{-x}-x$ with an accuracy of ϵ_t= 5%. (The exact root is 0.56714329).

Solution:

The above iterative formula becomes

$$x_{i+1} = e^{-x_i}$$

Performing the iterations, we get:

Iter#	x_{i+1} (i=0,1,2,……) **Let** $x_0=0$
1	$e^{-0} = 1$
2	$e^{-1} = 0.367879$
3	$e^{-0.367879} = 0.692201$
4	$e^{-0.692201} = 0.500473$
5	$e^{-0.500473} = 0.606244$
.	.
10	$e^{-0.571143} = 0.564879$

Example2: Use the fixed-point iteration method to estimate one of the roots of $f(x)=x^2-2x-3$ with an accuracy of ϵ_t= 5%. (The exact roots are: x=-1 and x=3).

Solution:

Alternative (1): Re-write the equation in the form: $x=\sqrt{2x+3}$, so that $g(x)=\sqrt{2x+3}$. Let x_0=0. and use the above iterative formula to generate the following results:

Iter #	x
1	1.73205
2	2.54246
.	.
.	.
6	2.99413

*Note that the method here, **converges** to the root **x=3**.*

Alternative (2): Re-write the equation in the form: $x=\frac{3}{x-2}$, so that $g(x)=\frac{3}{x-2}$

For x_0=0, we get:

Iter #	x
1	-1.50000
2	-0.85714
.	.
.	.
5	-1.00549

*So that the method **converges** to the root **x = -1.***

Alternative (3): Re-write the equation in the form: $x = x^2 - x - 3$, so that $g(x) = x^2 - x - 3$. Starting with $x_0 = 0$, we get:

Iter #	x
1	-3.0000
2	9.00000
3	69.0000
4	4689.00

It is obvious that the method does **not converge** for the above choice of g(x).

Let us repeat the procedure with a different initial guess, say x_0=2.9 which is very close to one of the root. We get:

Iter #	x
1	2.51000
2	0.79010
.	….
.	….
5	90.6150
6	8117.47

It does **not converge**! Even if you try x_0 = -1.1 which is very close to the other root, you will find out that the method **diverges again**!! The reason will be clarified later.

Alternative (4): Re-write the equation in the form: $x = \frac{x^2 - 3}{2}$, so that $g(x) = \frac{x^2 - 3}{2}$. Starting with $x_0 = 0$, we get:

Iter #	x
1	-1.50000
2	-0.37500
.	….
14	-0.66217
15	-1.28076

It is obvious that the method **diverges** for this choice of g(x). The same thing will happen, even if we start with x_0 very close to one of the roots!!

2.2The Newton-Raphson Metod :

► The method is based on the first order Taylor expansion, i.e.:

$$f(x_{i+1}) \approx f(x_i) + f'(x_i)(x_{i+1} - x_i)$$

When we hit the root, $f(x_{i+1}) = 0$, then:

$$x_{i+1} \approx x_i - \frac{f(x_i)}{f'(x_i)}$$

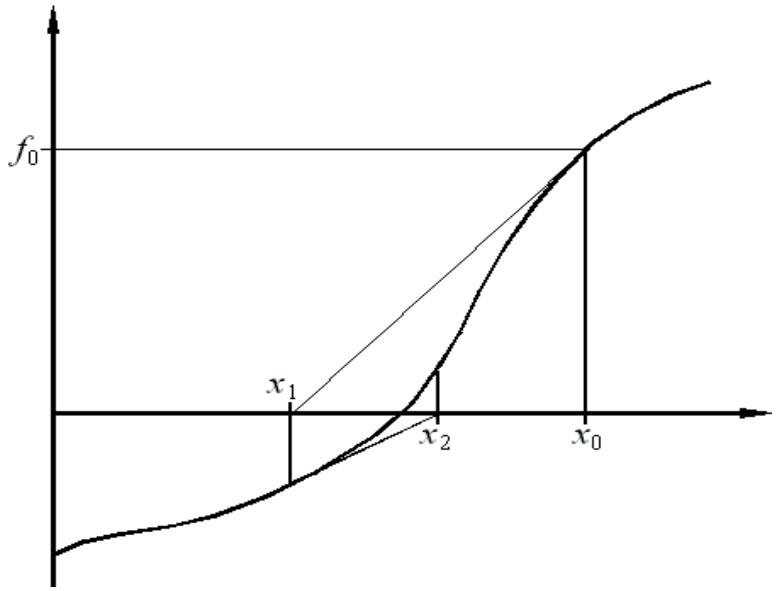
►Procedure:

4- Assume an initial guess for the root = x_0 and calculate the first estimate of the root x_1 from: $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$

5- Repeat step 1 several times until convergence is achieved, i.e.

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \text{ until } \epsilon_a < \epsilon_s$$

The procedure is illustrated by the figure shown below.



Example1: Use Newton-Raphson method to estimate the root of $f(x)=e^{-x}-x$. Show all details of the iterations. Hint: the root is located between 0 and 1.

Iter	x_i	$f(x_i)$	$f'(x_i)=-e^{-x}-1$	$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}$
1	0.	1.	-2.	0.5
2	0.5	0.1065	-1.6065	0.5663
3	0.5663	0.0013	-1.5676	0.5671
4	0.5671	0.0000	-1.5676	0.5671

Example2: Repeat Example1 starting with x0 = 5

Iter	x_i	$f(x_i)$	$f'(x_i)=-e^{-x}-1$	$x_{i+1}=x_n-\frac{f(x_i)}{f'(x_i)}$
1	5.	-4.9933	-1.0067	0.04016
2	0.04016	0.92048	-1.9606	0.5096
3	0.5096	0.0911	-1.6007	0.5665
4	0.5665	0.0010	-1.5675	0.5671

Example3:

Use the Newton- Raphson method ,with 1.5 as starting point ,to find solution of $f(x)= x-2\sin x$

$f(x)=x-2\sin x$, $x_0=1.5$

$f'(x) = 1-2\cos x$

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}=x_n-\frac{x_n-2\sin x_n}{1-2\cos x_n}$$

$$=\frac{2(\sin x_n-x_n\cos x_n)}{1-2\cos x_n}$$

x1	**2.0765582**
x2	1.9105066
x3	1.8956220
x4	1.89549427
x5	1.895494267033
x6	1.895494267033

2.3 Secant Method:

The method is used instead of Newton-Raphson method when the derivative of the function is difficult to obtain or if the slope of the function = 0 near the root. The formula of the method is obtained by replacing the exact derivative in Newton-Raphson's formula by $\frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}$ so that

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

►Procedure:

6- Assume two initial guess x_0 and x_1 to calculate the first estimate of the root x_2 from: $x_2 = x_1 - \frac{x_1 - x_0}{f(x_1) - f(x_0)} f(x_1)$

7- Repeat step 1 several times until convergence is achieved, i.e.

$$x_{i+1} = x_i - \frac{x_i - x_{i-1}}{f(x_i) - f(x_{i-1})} f(x_i) \quad \text{until } \epsilon_a < \epsilon_s$$

►Example3: Repeat Examples 1 using the secant method with the two initial guesses $x_0 = 0.$ and $x_1 = 1.$

Solution:

Iter	X_{i-1}	X_i	X_{i+1}
1	1.	2.	0.6127
2	2.	0.6127	0.563838
3	0.48714	0.563838	0.56717

►Example4: Repeat Examples 3 using the secant method with the two initial guesses $x_0 = 2.$ and $x_1 = 3.$

Solution:

Iter	X_{i-1}	X_i	X_{i+1}
1	2.	3.	0.2823
2	3.	0.2823	0.6570
.	.	.	.
5	0.5719	0.5671	0.5671

Chapter 3

The Numerical Solutions of Linear Systems

It is well known from the linear algebra that, that there are many methods used to find the exact solutions of linear systems, $Ax = b$, where $A \in R^{n\times n}$, $b \in R^{n\times 1}$, such as Gauss elimination or, Gauss-Gordan, or Kramer's method. But using these methods becomes so difficult when the dimension n, of the matrix A, is large. Therefore, we need to compute the solutions numerically by using computers.

In general, there are two types of numerical methods, which can be used to find the numerical solutions of linear systems: **direct methods** and **indirect methods**.

Before starting to study these methods, let's revision some equivalent algebraic facts of the linear system: $Ax = b$, $A \in R^{n\times n}$, $b \in R^{n\times 1}$

1-The homogenous system $Ax = 0$, has only zero solution, iff $|A| \neq 0$, which means A is nonsingular matrix.

2-The linear system $Ax = b$, has a unique solution , iff $|A| \neq 0$, which means A is nonsingular matrix.

From above, before to think about finding a numerical solution of a linear system, we need make sure that the exact solution exists, which means we should check whether $|A| \neq 0$, or A is nonsingular.

Direct methods

These methods can be convergent very fast, but when the dimension of A is large, it is not recommended to use these methods because, we need to compute lots of mathematical operations, which means, the errors become bigger.

Next, we will study some of direct methods.

Gauss Elimination Algorithm

The idea of this method, is to convert A in the linear system $Ax = b$, to upper U or lower matrix L. Thus, we get a lower or upper triangular system:

$$Ux = b_1 \text{ or } Lx = b_2$$

It is clear that, for solving lower triangular system we use **Forward substitutions** and for solving upper triangular system we use **Backward substitutions**.

In fact, solving lower (upper) triangular system is easier than solving the original system.

Steps of Gauss elimination algorithm:

1- Write the system $Ax = b$, in the matrix form [A: b].

2- Convert [A: b] to appear triangular form $[L: b_1]$ or lower triangular from $[U: b_2]$

3-Solve the lower (upper) triangular system, $Lx = b_1$ $(Ux = b_2)$, by using forward (backward) substitutions.

Example:- Solve the following linear system by using Gauss algorithm

$$5x_1 + 2x_2 = 3$$
$$4x_1 + 3x_2 = 1$$

Solution

Firstly, we write the system in matrix form $[A: b]$, as follows

$$\begin{bmatrix} 5 & 2 & : & 3 \\ 4 & 3 & : & 1 \end{bmatrix}$$

$$(L_1)/5 \ , \ -4(L_1/5)+L_2$$

$$\begin{bmatrix} 1 & 2/5 & : & 3/5 \\ 0 & 7/5 & : & -7/5 \end{bmatrix}$$

Thus, we get the following **upper triangular system**

$$x_1 + (2/5)x_2 = 3/5$$
$$(7/5)x_2 = -7/5$$

Finally, we solve the last system by using the backward substitutions, to get

$$x_2 = (-7/5)/(7/5) = -1$$
$$x_1 = 3/5 + 2/5 = 5/5 = 1$$

Thus, the solution is

$$x=\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

H.W. In the last example, try to change the system $Ax = b$, to a lower triangular system and find its solution by using the forward substitutions.

Matlab Code for Gauss Method

Write a Matlab program, which can be used to find the solution of the following system by using Gauss method with back ward substitutions

$$\begin{aligned} 4x_1 - 9x_2 + 2x_3 &= 5 \\ 2x_1 - 4x_2 + 6x_3 &= 3 \\ x_1 - x_2 + 3x_3 &= 4 \end{aligned}$$

```
A=[4,-9,2;2,-4,6;1,-1,3];
b=[5;3;4];
n=3;
for k=1:n-1;
    for i=k+1:n
        m(i,k)=A(i,k)/A(k,k);
        for j=k:n
            A(i,j)=A(i,j)-m(i,k)*A(k,j);
        end
        b(i)=b(i)-m(i,k)*b(k);
    end
end
x(n)=b(n)/A(n,n);
for i=n-1:-1:1
    s=0;
    for j=i+1:n
        s=s+A(i,j)*x(j);
    end
    x(i)=(b(i)-s)/A(i,i);
end
```

```
disp(x);


    6.9500    2.5000   -0.1500
```

Gauss- Gordan algorithm

This algorithm is considered, as a modified to the Gauss elimination algorithm. We convert the linear system, $Ax = b$, to $I_n x = b_1$, where I_n is the identity matrix of order n. Thus to solve the last system we use the direct substitutions.

$$Ax=b \longrightarrow I_n x = b_1$$

Steps of Gauss- Gordan algorithm

1- Write the system $Ax = b$, in the matrix form [A: b].

2- Do some mathematical operations to convert [A: b] to the diagonal form $[I_n : b_1]$.

3- Find x by solving the system $I_n x = b_1$ using direct substitutions.

Example :- Find the solution of the following system, by using Gauss-Gordan Method

$$x_1 + 2x_2 = 1$$
$$2x_1 + x_2 = 3$$

Solution

Firstly, we write the system in matrix form $[A : b]$, as follows:

$$\begin{bmatrix} 1 & 2 & : & 1 \\ 2 & 1 & : & 3 \end{bmatrix} \ldots\ldots(1) \qquad -2(L_1)+L_2 \begin{bmatrix} 1 & 2 & : & 1 \\ 0 & -3 & : & 1 \end{bmatrix} \ldots\ldots..(2)$$

$$-(L_2/3) \begin{bmatrix} 1 & 2 & : & 1 \\ 0 & 1 & : & -1/3 \end{bmatrix} \ldots\ldots(3)$$

$$-2(L_2)+L_1 \begin{bmatrix} 1 & 0 & : & 5/3 \\ 0 & 1 & : & -1/3 \end{bmatrix} \ldots\ldots(4)$$

Thus, we get

$$I_n x = b_1$$
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5/3 \\ -1/3 \end{bmatrix}$$

LU Algorithm *:*

This method is called by LU, because the matrix A in the linear system $Ax = b$, decomposes into the multiplication of two matrices : lower L and upper U, and this decomposition works for any vector b**,** which means $A = LU$.
Thus, we get lower triangular and upper triangular systems.
In order to get the solution of the system $Ax = b$, we need to solve these two systems.

Steps of LU Method

1- Decompose A, in the form $A = LU$, where U is an upper matrix and L is a lower matrix.
2- Set Ux=y, which leads to Ly=b
3-Solve first the lower triangular system, Ly=b, using the *forward substitutions* to get y, and then solve the lower triangular system, Ux=y, using the *backward substitutions* to get *x* .

Remarks:-

1-This method can be considered better than Gauss and Gauss-Gordan methods and that because the decomposition of the matrix A works for any vector b, while in Gauss and Gauss - Gordan, the mathematical operations, which we have to do on [A:b], should be redone again when we choose another vector b.

2-In fact, not any matrix A can be decomposed to LU, unless the following condition (**The diagonal control condition),** is satisfied.

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \quad i=1,2,..n$$

Example:- Can we decompose the matrix A to LU ?, where

$$A=\begin{pmatrix}3 & 1 & 1\\ 5 & 6 & 0\\ 0 & 1 & 7\end{pmatrix}$$

Solution

$$2 = |a_{12}| + |a_{13}| \le |a_{11}| = 3$$
$$5 = |a_{21}| + |a_{23}| \le |a_{22}| = 6$$
$$1 = |a_{31}| + |a_{32}| \le |a_{33}| = 7$$

Since, A satisfies the diagonal control condition
Therefore, it follows that A can be decomposed to LU.

The following example shows how to use LU algorithm to solve a linear system.

Example:- Use LU Algorithm to find the solution of the following system.

$$x_1 + 2x_2 = 3$$
$$3x_1 + x_2 = 5$$

Solution

Set

$$A=\begin{bmatrix}1 & 2\\ 3 & 1\end{bmatrix}=\begin{bmatrix}1 & 0\\ 3 & 1\end{bmatrix}\begin{bmatrix}u_{11} & u_{12}\\ 0 & u_{22}\end{bmatrix}$$

L U

$u_{11} = 1, u_{12} = 2,\ 3u_{12} + u_{22} = 1 \qquad u_{22} = 1 - (3)(2) = -5$

It follows that

$$L=\begin{bmatrix}1 & 0\\ 3 & 1\end{bmatrix},\quad u=\begin{bmatrix}1 & 2\\ 0 & -5\end{bmatrix}$$

Set $Ux = y, Ly = b$

We need to solve first the system $[Ly = b]$ by using Forward substitutions

$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$, thus $y_1 = 3$

$y_2 = 5 - (3)(3) = -4$ y= $\begin{bmatrix} 3 \\ -4 \end{bmatrix}$

Secondly, we solve the system $[Ux = y]$, by using Backward substitutions

$\begin{bmatrix} 1 & 2 \\ 0 & -5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$ $x_2 = -4/(-5) = 4/5$

$x_1 = 3 - 2(4/5) = 3 - 8/5 = 7/5$

Thus $x = \begin{bmatrix} 7/5 \\ 4/5 \end{bmatrix}$

H.W. Consider that, we have the following linear system

$$\begin{aligned} x_1 - x_2 + x_3 &= 2 \\ 3x_1 + 3x_3 &= 0 \\ 2x_1 + 5x_2 + x_3 &= 1 \end{aligned}$$

Solve the system by using,

1- Gauss elimination (with Backward or Forward substitutions).
2- LU algorithm.

Indirect methods

In these methods, we don't need to do lots of matrix operations as in the direct methods, but it is known that indirect methods are slower than direct methods in convergence. Moreover, the main different between direct and indirect methods that indirect methods needs an initial solution, $x^0 = (x_1^0, x_2^0, \dots x_n^0)$, in order to start and depending on this initial condition, we can get $x^1, x^2, \dots \dots$

Therefore, indirect methods are also called the *iterative methods*.

We will study two algorithms of indirect methods: *Jacobi & Gauss-Sidel*

Jacobi iterative algorithm

Consider that, we have the following the linear system:

$$A_{n\times n}x = b_{n\times 1}$$

which can be written as follows:

$a_{11}x_1+a_{12}x_2+\ldots\ldots\ldots+a_{1n}x_n=b_1$

$a_{21}x_1+a_{22}x_2+\ldots\ldots\ldots\ldots+a_{21}x_n=b_2$

$a_{n1}x_1+a_{n2}x_2+\ldots\ldots\ldots+a_{nn}x_{n=b_n}$

where $a_{\mathrm{ii}} \neq 0, \quad i = 1,2,\ldots.n$

Remarks:

1- In case of $a_{\mathrm{ii}} = 0,$ for some i, we replace the equation number i, with another equation in order to get $a_{\mathrm{ii}} \neq 0, \quad i = 1,2,\ldots.n$

2- On the other hand, in order to get faster convergence, we should make sure that, **The diagonal control condition** is satisfied

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|$$

In case of this condition is not satisfied, we can also switch places between the equations, until we get this condition is satisfied.

Steps of Jacobi algorithm

1-Make sure that the system has a unique solution, $|A| \neq 0$.

2-Choose the initial solution $x^0 = (x_1^0, x_2^0, \ldots\ldots x_{\mathrm{n}}^0) \in R^n$

3-Make sure that $a_{ii} \neq 0, and\ |a_{ii}| \geq \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|$

and in case of one of these condition is not satisfied, switch places of the equations, until we get the two conditions are satisfied.

4- Write the system in iterative form: $x_i^{k+1} = (b_i - \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij}\ x_j^k)/a_{ij},$ for

$$i = 1,2,\ldots.n$$

5-Compute the solution iterative iteratively, (for $k = 0,1,2, ...$), until we get the following the stop condition is satisfied:

$\left\|x^{(k+1)} - x^{(k)}\right\| < \epsilon$, where

$$E_k = \left\|x^{(k+1)} - x^{(k)}\right\| = \max_{1 \le i \le n} \left|x_i^{k+1} - x_i^k\right|$$

Example:- Use Jacobi algorithm to solve the following linear system, for two iterative step and find the iterative error at each step.

Note: The exact solution for this system is x=(1; 2; 3)

$$\begin{aligned} 4x_1 + 2x_2 + x_3 &= 11 \\ 2x_1 + x_2 + 4x_3 &= 16 \\ -x_1 + 2x_2 &= 3 \end{aligned}$$

Solution

Since $a_{33} = 0$, we need to switch the places of equation 2 and 3:

$$\begin{aligned} 4x_1 + 2x_2 + x_3 &= 11 \\ -x_1 + 2x_2 &= 3 \\ 2x_1 + x_2 + 4x_3 &= 16 \end{aligned}$$

It is clear that, the last system satisfies the **diagonal control condition.**

We can rewrite the last system as follows:

$x_1 = \frac{11}{4} - \frac{1}{2}x_2 - \frac{1}{4}x_3$

$x_2 = \frac{3}{2} + \frac{1}{2}x_1$

$x_3 = 4 - \frac{1}{2}x_1 - \frac{1}{2}x_2$

We write the system in the iterative form as follows:

$x_1^{k+1} = \frac{11}{4} - \frac{1}{2}x_2^k - \frac{1}{4}x_3^k$

$x_2^{k+1} = \frac{3}{2} + \frac{1}{2}\, x_1^k$ k=0,1,…….

$x_3^{k+1} = 4 - \frac{1}{2}x_1^k - \frac{1}{4}x_2^k$

Let x^0=(x_1^0,x_2^0,x_3^0)=(0.8, 1.8, 2.8)

Set k=0

$x_1^{(1)}$= $\frac{11}{4}$ - $\frac{1}{2}$(1.8) - $\frac{1}{4}$(2.8) = 1.15

$x_2^{(1)}$= $\frac{3}{2}$ + $\frac{1}{2}$(0.8) = 1.9

$x_3^{(1)}$ = 4 - $\frac{1}{2}$(0.8) - $\frac{1}{4}$(1.8) = 3.15

$x^{(1)} = (1.15, 1.9, 3.15)$

$$E_1 = \max\{\,|1.15 - 0.8|, |1.9 - 1.8|, |3.15 - 2.8|\} = 0.35$$

set $k = 1$, by using the same way obtain:

$x_1^{(2)} = \frac{11}{4} - \frac{1}{2}(1.9) - \frac{1}{4}\,(3.15) = 1.0125$,

$x_2^{(2)} = \frac{3}{2} + \frac{1}{2}\,(1.15) = 2.0750$,

$x_3^{(2)} = 4 - \frac{1}{2}(1.15) - \frac{1}{4}(1.9) = 2.95$

$x^{(2)} = (1.0125, 2.075, 2.95)$

$$E_2 = \max\{|1.0125 - 1.15|, |2.075 - 1.9|, |2.95 - 3.15|\} = 0.175$$

We continue iteratively, until the convergent condition can be satisfied:

$||x^{(k+1)} - x^k|| < \in$

Gauss – Sidel algorithm

The main difference between this method and Jacobi method, is for any iterative step ,k, the new approximate values of the x_j, $j = 1, \ldots i - 1$, are used directly, to

compute the approximate values of x_i, while in Jacobi we don't use the new approximate values x_j until, we consider the next iterative step, $k+1$.

Remark: The steps of Gauss-Sidel algorithm are the same as the step of Jacobi method expect for in step (4), we use the Gauss-Sidel iterative system:

$$x_i^{k+1} = (\text{bi} - \sum_{\substack{\text{j=1} \\ \text{j}\neq\text{i}}}^{i-1} \text{a}_{ij}\ x_j^{k+1} - \sum_{\substack{\text{j=i} \\ \text{j}\neq\text{i}}}^{n} \text{a}_{ij}\, x_j^k)\ /a_{ii}$$

for $i = 1,2,3, \dots n, k = 0,1,2, \dots ..$

Example:- Use Gauss-Sidel algorithm to find the approximate solution to the following system for two iterative steps, with x^0=(0.8, 1.8, 2.8), and find the iterative errors at each step.

$$\begin{aligned} 4x_1 + 2x_2 + x_3 &= 11 \\ -x_1 + 2x_2 &= 3 \\ 2x_1 + x_2 + 4x_3 &= 16 \end{aligned}$$

It is clear that, $a_{ii} \neq 0$ *and* the last system satisfies the **diagonal control condition.**

We can rewrite the last system as follows:

$x_1 = \frac{11}{4} - \frac{1}{2}x_2 - \frac{1}{4}x_3$

$x_2 = \frac{3}{2} + \frac{1}{2}x_1$

$x_3 = 4 - \frac{1}{2}x_1 - \frac{1}{2}x_2$

We write the system in the iterative form as follows

$x_1^{k+1} = \frac{11}{4} - \frac{1}{2}x_2^k - \frac{1}{4}x_3^k$

$x_2^{k+1} = \frac{3}{2} + \frac{1}{2}x_1^{k+1}$ $\qquad k = 0,1, \dots ....$

$x_3^{k+1} = 4 - \frac{1}{2}x_1^{k+1} - \frac{1}{4}x_2^{k+1}$

Set k=0, we get

$x_1^{(1)} = \frac{11}{4} - \frac{1}{2}(1.8) - \frac{1}{4}(2.8) = 1.15$

$x_2^{(1)} = \frac{3}{2} + \frac{1}{2}(1.15) = 2.075$

$x_3^{(1)} = 4 - \frac{1}{2}(1.15) - \frac{1}{4}(2.075) = 2.9063$

$x^{(1)} = (1.15, 2.075, 2.9063)$

$$E_1 = \max\{|1.15 - 0.8|, |2.075 - 1.8|, |2.9063 - 2.8|\} = 0.35$$

Set k=1, we can, in the same way, compute

$x_1^{(2)} = \frac{11}{4} - \frac{1}{2}(2.075) - \frac{1}{4}(2.9063) = 0.9859$

$x_2^{(2)} = \frac{3}{2} + \frac{1}{2}(0.9859) = 1.993$

$= 4 - \frac{1}{2}(0.9859) - \frac{1}{4}(1.9930) = 3.0088$

$$x^{(2)} = (0.9859, 0.9859, \ 3.0088)$$

$$E_2 = \max\{|0.9859 - 1.15|, |1.993 - 2.075|, |3.0088 - 2.9063|\} = 0.1641$$

We continue iteratively, until the convergent condition can be satisfied:

$\|x^{(k+1)} - x^k\| < \in$ or $\left\|b - A * x^{(k)}\right\| < \in$

Remark: From last example, we see that, the approximate results that we get by using Gauss-Sidel are more accurate and closer to the exact solution, compared with the results that we got by using Jacobi method. Moreover, for $k \geq 2$,the iterative errors those arise from using Gauss-Sidel method are less than the iterative errors those arise from using Gauss-Sidel. Which means Gauss-Sidel method is faster than Jacobi method in convergence.

Matlab Code

Write a matlab program, which can be used to find the approximate solution of the following system by using

1- Jacobi Method

2-Gauss-Sidel

$$\begin{aligned} 9x_1 - 4x_2 + 2x_3 &= 5 \\ 2x_1 - 4x_2 + x_3 &= 3 \\ x_1 - x_2 + 3x_3 &= 4 \end{aligned}$$

with $x^0 = (0,0,0)$

Jacobi

```
A=[9,-4,2;2,-4,1;1,-1,3];
b=[5;3;4];
n=3;
x0=[0;0;0];
r=norm(b-A*x0);
k=0;
while r> 0.01
    k=k+1;
    for i=1:n
        s=0;
        for j=1:i-1
            s=s+A(i,j)*x0(j);
        end
        for j=i+1:n
            s=s+A(i,j)*x0(j);
        end
       x(i)=(b(i)-s)/A(i,i);
    end
    x0=x' ;
    r=norm(b-A*x0);
end
disp(x);
disp(k);
```

```
0.1205   -0.4005    1.1605      k=19
```

Gass-Sidel

```
    A=[9,-4,2;2,-4,1;1,-1,3];
b=[5;3;4];
n=3;
x0=[0;0;0]; x=x0';
r=norm(b-A*x0);
k=0;
while r> 0.01
    k=k+1;
    for i=1:n
        s=0;
        for j=1:i-1
            s=s+A(i,j)*x(j);
        end
        for j=i+1:n
            s=s+A(i,j)*x0(j);
        end
       x(i)=(b(i)-s)/A(i,i);
    end
    x0=x' ;
    r=norm(b-A*x0);
end
disp(x);
disp(k);
```

```
 0.1188   -0.4004    1.1603     k=5
```

Exercises

Q1: Consider that, we have the following linear system 4x4

$$\begin{aligned} 2x_1 + x_2 - x_3 + 5x_4 &= 0 \\ x_1 + 2x_3 - x_4 &= 2 \\ x_1 + 4x_2 + x_3 + x_4 &= 1 \\ 3x_1 + x_2 + x_3 - x_4 &= 3 \end{aligned}$$

a. Make sure that, the diagonal control condition is satisfied,
b. Use **Gauss-Sidel** method to find the approximate solution for two iterative step ($x^{(1)}, x^{(2)}$), with considering $x^{(0)} = (\ 0\ ,\ -1\ ,\ \frac{1}{2}\ ,\ 2\ )$.
c. Compute the iterative errors, at each step.
d. What is the stop condition ?

Q2: Consider that, we have the following linear system 3x3

$$\begin{aligned} ax_1 + bx_2 + c\,x_3 &= 2 \\ dx_1 + ex_3 &= 8 \\ +f\,x_2 + g\,x_3 &= 1 \end{aligned}$$

1- Under which condition the above system has a unique solution, in terms of the elements of A ?
2- Use **Gauss** Method, with Forward substitution, to find the solution of this system in terms of the elements of A.

Q3: Consider that, we have the following linear system

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 2 \\ 3x_1 + 3x_3 &= 0 \\ 2x_1 + 5x_2 + x_3 &= 1 \end{aligned}$$

a. Does the system have a unique solution ? why ?
b. Make sure that the diagonal control condition is satisfied.
c. Solve the system by using **LU** method.

Chapter 4

Numerical integration

In this chapter, we study some methods; used to find the approximate value for the definite following integral

$$\int_a^b f(x)dx \qquad (\ \forall x \in [a,b]\ \ f \in C[a,b]\ ),$$

when it's difficult to find the exact value by using known methods (integration methods), such as:

$$\int_a^b e^{x^2} dx$$

The general idea of the integration methods is to divide the interval $[a,b]$, into n of subintervals :

$$[a,b] = [x_0,x_1]\ \cup [x_1,x_2] \dots \dots \cup [x_{n-1},x_n].$$

It is not needed to be the distances between the points $\{x_i\}$ are equal.

Next, we consider the polynomial $p_n(x)$ as an approximate form for f.

$f(x) \cong p_n(x),\ \ \forall\, x \in [a,b]$

which means the problem becomes

$$\int_a^b f(x)dx \cong \int_a^b p_n(x)\,dx$$

From the last form, we note that, the formula of numerical integration depends on the way of choosing the polynomial p_n.

The general formula of integration methods takes the form:

$$\int_a^b f(x)dx \cong \int_a^b p_n(x)\,dx = \sum_{i=0}^{n} a_i\, f(xi) \dots\dots\dots (*)$$

where,

$\{a_i\}_{i=0}^n$ are called the coefficients

$\{x_i\}_{i=0}^n$ are called the nodes.

If we used Lagrange polynomial, we could very easy get $\{a_i\}_{i=0}^n$, $\{x_i\}_{i=0}^n$ as follows:

We divide the interval $[a, b]$, to the n of subintervals

$$[a, b] = [x_0, x_1] \cup [x_1, x_2] \dots\dots \cup [x_{n-1}, x_n],$$

where the distances between the nodes $\{\text{xi}\}_{i=0}^n$ are equal

i.e. $x_{i+1} - x_i = h$

Thus $f(x) = p_n(x) + T(x)$, $\forall\, x \in [a, b]$,

where p_n is the Lagrange polynomial

$\int_a^b f(x)dx = \int_{x_0}^{x_n} p_n(x)\,dx = \int_{x_0}^{x_n} \sum_{i=0}^{n} L_i(x) f(x_i)dx + \int_{x_0}^{x_n} \frac{f^{(n+1)}(\delta(x))}{(n+1)!} \prod_{i=0}^{n}(x - x_i)dx,$

where $\int_{x_0}^{x_n} \frac{f^{(n+1)}(\delta(x))}{(n+1)!} \prod_{i=0}^{n}(x - x_i)dx,$ is the truncation error formula

So, $\int_a^b f(x)dx \cong \sum_{i=0}^{n}(\int_{x_0}^{x_n} L_i(x)dx) f(x_i))$ ……(1),

which means

$a_i = (\int_{x_0}^{x_n} L_i(x)dx)$

In the last formula,

For n=1, method is called (**Trapezoidal method**),

For n=2, method is called (**Simpson method**).

Trapezoidal method

From the general form of integration, with choosing Lagrange polynomial, and n=1, we get

$h = x_1 - x_0 \quad , b = x_1 \quad , a = x_0$

$$\int_a^b f(x)dx = \sum_{i=0}^{n} \left(\int_{x_0}^{x_n} L_i(x)dx \right) f(x_i)dx + \int_{x_0}^{x_1} \frac{f''(\delta(x))}{2}(x - x_0)(x - x_1)dx$$

$$\text{Set , } \mathrm{a}_0 = \int_{x_0}^{x_1} L_0(x)dx = \int_{x_0}^{x_1} \frac{(x-x_1)}{(x_0-x_1)}dx = \frac{(x-x_1)^2}{2(x_0-x_1)} = |_{x_0}^{x_1} = \frac{h}{2}$$

$$\mathrm{a}_1 = \int_{x_0}^{x_1} L_1(x)dx = \int_{x_0}^{x_1} \frac{(x - x_0)}{(x_1 - x_0)}dx = \frac{(x - x_0)^2}{2(x_1 - x_0)} = |_{x_0}^{x_1} = \frac{h}{2}$$

Substitute a_0, a_1 in equation (1), we get

$$\int_a^b f(x)dx \cong \sum_{i=0}^{1} a_i\, f(x_i) = \frac{h}{2} f(x_0) + \frac{h}{2} f(x_1),$$

which is the **Trapezoidal formula.**

where the truncation error formula takes the form

$$T(h) = \int_{x_0}^{x_1} \frac{f''(\delta(x))}{2}(x - x_0)(x - x_1)\mathrm{dx} = -\frac{h^3}{12} f''(\delta),$$

Remark:- We note that, if f is polynomial of order less than or equal one, then the truncation error equal zero, which means:

$$\int_a^b f(x)dx = \sum_{i=0}^{1} a_i\, f(x_i) = \frac{h}{2} f(x_0) + \frac{h}{2} f(x_1),$$

Example:- Use the Trapezoidal method to find the approximate value of the following integral

$$\int_a^b (x^3 + 1)\, dx$$

Solution

$$n = 1\ ,\ x_0 = 0\ ,\ x_1 = 1\ ,\ h = \frac{x_1 - x_0}{1} = 1$$

$$\int_a^b f(x)dx \cong \frac{1}{2}\big(f(x_0) + f(x_1)\big)$$

$$I_n = \int_a^b (x^3 + 1)\, dx \cong \frac{1}{2}\big((x^3 + 1)|_0 + (x^3 + 1)|_1\big) = \frac{1}{2}(1 + 2) = \frac{3}{2} = 1.5$$

While, we can find the exact value as follows:

$$I_e = \int_a^b (x^3 + 1)\, dx = \left(\frac{x^4}{4} + x\right)|_0^1 = \frac{1}{4} + 1 = 1.25$$

Thus, the absolute error is

$$E = |I_n - I_e| = |1.5 - 1.25| = 0.25$$

In order to get more accurate value to the integration, we use the composite Trapezoidal methods.

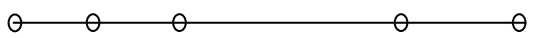
Composite Trapezoidal methods

The idea is, we divide the interval [a, b] into *n* of subintervals

$$[a, b] = [x_0, x_1] \cup [x_1, x_2] \ldots \ldots \cup [x_{n-1}, x_n],$$

Since the summation of all the integrals on the subintervals is equal the integral on the whole interval $[a, b]$, we can apply the Trapezoidal formula, on each of the integrals as follows:

$x_0 \quad x_1 \quad x_2 \qquad\qquad x_{n-1} \quad x_n$



$$I_N = I_1 + I_2 + \cdots I_n$$

$$\int_a^b f(x)dx = \int_{x_0}^{x_1} f(x)dx + \int_{x_1}^{x_2} f(x)dx + \cdots + \int_{x_{n-1}}^{x_n} f(x)dx$$

$$= [\frac{h}{2}\big(f(x_0) + f(x_1)\big)] + [\frac{h}{2}\big(f(x_1) + f(x_2)\big)] + \cdots + [\frac{h}{2}\big(f(x_{n-1}) + f(x_n)\big)$$

$$- \sum_{i=1}^{n} \frac{h^3}{12} f''(\delta_i),$$

where $x_{i-1} \leq \delta_i \leq x_i$

$$\therefore \int_a^b f(x)dx \cong \left[\frac{h}{2}\big(f(x_0) + f(x_n)\big)\right] + h\sum_{i=2}^{n} f(x_{i-1})$$

The last equation is called, the **Composite Trapezoidal Formula.**

Steps of composite Trapezoidal algorithm

1-Input a, b

2- Input the number of partitions n

3-Define the function $f(x)$

4-Find $h = \frac{b-a}{n}$

5- Use composite Trapezoidal formula

$$\int_a^b f(x)dx \cong \left[\frac{h}{2}\big(f(x_0) + f(x_n)\big)\right] + h\sum_{i=2}^{n} f(x_{i-1})$$

Example: For the last example, find the approximate value of the integral, using composite Trapezoidal methods with taking $n = 2$

$$\int_0^1 (x^3 + 1)\mathrm{dx}$$

Solution

$$h = \frac{b-a}{n} = \frac{1}{2}$$

$$[0,1] = [0,0.5] \cup [0.5,1]$$

$$\int_0^1 (x^3 + 1)\mathrm{dx} = \int_0^{1/2} (x^3 + 1)\mathrm{dx} + \int_{1/2}^1 (x^3 + 1)\mathrm{dx}$$

$$I_n \cong \frac{h}{2}[(x^3 + 1)dx|_0 + (x^3 + 1)dx|_{0.5}] + \frac{h}{2}[(x^3 + 1)dx|_{0.5} + [(x^3 + 1)dx|_1]$$

=1/4(1+1/8+1+1/8+1+2)= 1/4[21/4]=[21/16]=1.3125

$$E=|I_n\text{-}I_e|=|1.31\text{-}1.25|=0.06$$

Remark

In last example, it is clear that, from the absolute errors, the result of the composite Trapezoidal method is more accurate than the result that we get by using normal Trapezoidal method**.**

Matlab Code for composite Trapezoidal method

Next, we write dawn the Matlab Code for last example with $n = 40$

```
a=0; b=1; n=40;
h=(b-a)/n; g=0;
x=sym('x');
f=x^3+1;
m=subs(f,x,a)+subs(f,x,b);
for i=1:n-1
```

```
    d=a+i*h;
    g=g+2*subs(f,x,d);
end
T=(h/2)*(m+g);
fprintf('I=%f',T);
```

```
I=1.250156
```

Simpson method

Here, we set $n = 2$, which means

$$[a,b] = [x_0, x_1] \cup [x_1, x_2], \qquad h = \frac{x_2 - x_0}{2}$$

We use Lagrange polynomial, of order 2, to find an approximate form for the function f.

$$\mathrm{f(x)} \cong P_2(x), \qquad \forall x \in [a,b]$$

$$\int_a^b f(x)dx = \sum_{i=0}^{n} \left(\int_{x_0}^{x_n} L_i(x)dx \right) f(x_i) + \int_{x_0}^{x_1} \frac{f'''(\delta(x))}{6}(x - x_0)(x - x_1)(x - x_2)dx$$

we can show that:

$$a_0 = \int_{x_0}^{x_2} \mathrm{L_0(x)dx} = \frac{\mathrm{h}}{3}$$
$$a_1 = \int_{x_0}^{x_2} \mathrm{L_1(x)dx} = \frac{4\mathrm{h}}{3}$$
$$a_2 = \int_{x_0}^{x_2} \mathrm{L_2(x)dx} = \frac{\mathrm{h}}{3}$$

$$T(h) = \int_{x_0}^{x_1} \frac{f'''(\delta(x))}{6}(x - x_0)(x - x_1)(x - x_2)dx = -\frac{h^5}{90} f^{(4)}(\delta) \dots\dots(2)$$

Thus, we get

$$\therefore \int_a^b f(x)dx \cong \frac{h}{3}\mathrm{f}(\mathrm{x}_0) + \frac{4h}{3}\mathrm{f}(\mathrm{x}_1) + \frac{h}{3}\mathrm{f}(\mathrm{x}_2)$$

which is the **Simpson integral formula**

Remark:- We note that, if f is polynomial of order less than or equal 3, then the truncation error equal zero, which means:

$$\int_a^b f(x)dx = \sum_{i=0}^{2} a_i\, f(x_i) = \frac{h}{3}f(x_0) + \frac{4h}{3}f(x_1) + \frac{h}{3}f(x_2),$$

Example: find the approximate value of the following integral , using Simpson method

$$\int_1^2 e^{x^2}\, dx$$

Solution:

$$h = \frac{b-a}{n} = \frac{2-1}{2} = \frac{1}{2}$$

$$[1,2] = [1,1.5]U[1.5,2]$$

$$I_n = \int_1^2 e^{x^2}\, dx = \frac{1}{6}[e^{x^2}\Big|_1 + 4e^{x^2}\Big|_{1.5} + e^{x^2}|_2]$$

$$= \frac{1}{6}[2.71 + 4(9.48) + 54.5] = 15.855$$

H.W. Find the following integral, by using Simpson method.

$$\int_0^1 (x^3 + 1)\mathrm{dx}$$

(answer 1.25, $I_n = I_e$)

Composite Simpson method

We divide, the interval $[a,b]$, into k, pairs of subintervals, where $n = 2k$ (n should be odd), as follows:

$[a,b] = ([x_0,x_1] \cup [x_1,x_2]) \cup ([x_2,x_3] \cup [x_3,x_4]) \dots\dots \cup ([x_{n-2},x_{n-1}] \cup [x_{n-1},x_n])$

Apply Simpson formula for each of the pairs of subintervals

$$\int_a^b f(x)dx = \int_{x_0}^{x_2} f(x)dx + \int_{x_2}^{x_4} f(x)dx + \dots\dots \int_{x_{n-2}}^{x_n} f(x)dx$$

$$= \frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)] + \frac{h}{3}[f(x_2) + 4f(x_3) + f(x_4)] \dots\dots + \frac{h}{3}[f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)] - \sum_{i=1}^{k} \frac{h^5}{90} f^{(4)}(\delta_i)$$

where $x_{i-1} \le \delta_i \le x_i$

$$\int_a^b f(x)dx \cong \frac{h}{3}[f(x_0) + f(x_n)] + \frac{4h}{3}\sum_{i=1}^{k} f(x_{2i-1}) + \frac{2h}{3}\sum_{i=1}^{k-1} f(x_{2i})$$

which is the **'Composite Simpson Integral formula '**

Steps of composite Simpson algorithm

1-Input a,b

2-Input the numbers of the pairs of subintervals, k, and set $n = 2k$

3-Define the function f

4- $h = \frac{b-a}{n}$

5-Use the composite Simpson integral formula,

$$\int_a^b f(x)dx \cong \frac{h}{3}[f(x_0)+f(x_n)] + \frac{4h}{3}\sum_{i=1}^{k} f(x_{2i-1}) + \frac{2h}{3}\sum_{i=1}^{k-1} f(x_{2i})$$

Example:- Use Composite Simpson integral formula to find the value of the following integral, consider n=4

$$\int_1^2 e^{x^2} dx$$

solution

$$h = \frac{2-1}{4} = \frac{1}{4}$$

$[1,2] = ([1,1.25] \cup [1.25,1.5]) \cup ([1.5,1.75] \cup [1.75,2])$

$$\int_1^2 e^{x^2}\, dx = \int_1^{1.5} e^{x^2} dx + \int_{1.5}^2 e^{x^2}\, dx$$

Apply Simpson formula for each integral, we get that

$$\int_1^2 e^{x^2} dx \cong \frac{h}{3}(e^{x^2}|_1 + 4e^{x^2}|_{1.25} + e^{x^2}|_{1.5}) + \frac{h}{3}(e^{x^2}|_{1.5} + 4e^{x^2}|_{1.75} + e^{x^2}|_2)$$

$= \frac{1}{12}(2.71 + 4(4.77) + 9.48) + \frac{1}{12}(9.48 + 4(21.3) + 54.5) = 15.0375$

Remark: If we increase *n*, $(n = 6, or\ n = 8)$, we can get more accurate results to the integration.

H.W. Compare between the two absolute errors, those can arise from using *Simpson* method with n=2 and 4, respectively, to find the approximate value for the following integral:

$$\int_0^1 (x^4 + 1)dx$$

Matlab Code for composite Simpson method

Next, we write dawn the Matlab Code for which can be used to find the following integer $\int_0^1 (x^3 + 1)$dx, using composite Simpson method, with $n = 40$.

```
a=0; b=1; n=40;
h=(b-a)/n; g=0;
x=sym('x');
f=x^3+1;
m=subs(f,x,a)+subs(f,x,b);
r=0; g=0;
for i=1:n-1
d=a+i*h;
if rem(i,2)==0
    r=r+2*subs(f,x,d);
else
    g=g+4*subs(f,x,d);
end
end
T=(h/3)*(m+r+g);
fprintf('I=%f',T);

I=1.25
```

Note that
$I_N = I_e = 1.25$, and that because f is a polynomial of order three.

Exercises

Q1: Use both of Trapezoidal and composite Simpson formulas to derive a new composite formula, with n=5.

a. What is the truncation error's form of this new formula?
b. What should be the degree of f to guarantee that there is no absolute error.
c. Use the new formula to find the approximate value of the following integral, and find the absolute error.

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

Q2: Use **Trapezoidal** method, with **n=1**, and **n=3** to find the approximate value of the following integral, and find the absolute error in each case.

$$\int_1^2 (x^2 + 2x + 1) dx$$

Chapter 5

Numerical Solutions of First Order Ordinary Differential Equations

It is well known that, an ordinary differential equation (O.D.E.), is an equation has unknown function *y*, of one variable *x*, and some of its derivatives. For instance

$$\frac{dy}{dx} = y \sin x$$

while, partial differential equation, has unknown function of two or more variables and some of its partial derivatives. For instance

$$\frac{\partial y}{\partial \mathrm{x}} = \frac{\partial^2 y}{\partial x^2}$$

The order of the differential equation, is the highest derivative appears in the differential equation.

In this chapter, we will study, the numerical solutions for first order ordinary differential equations, which takes the general form:

$$y' = f(x, y)$$

The solution of the differential equation above, is an differentiable function defined on an interval $[a, b]$, and satisfies the differential equation. Each solution has a constant *c*, which can be determined, if *y* is known for at least one point $x_0 \in [a, b]$. ($y(x_0) = y_0$ *initial condition*)

Remark : The problem of the differential equation with an initial condition is called **Initial Value Problem:**

$$y' = f(x, y), \quad y(x_0) = y_0$$

Example:

$$y = -\sin x$$
$$y(0) = 1$$

While, if y are given at more than one point, then the problem is called
Boundary values problem:

Example

$$y = xy$$
$$\mathrm{y}(0) = 1\,, \qquad \mathrm{y}(1) = 2$$

In this chapter, we will only study the numerical solutions of first order initial vale problems.
Our aim is find the approximate solution, for this problem at certain points: $\{x_i\}_{i=1}^{n} \in [a, b]$, which means, we only need to find $\{y_i\}_{i=1}^{n}$

Next, we will study some important methods that can be used to find the numerical solutions of initial value problems

Euler Algorithm
The general idea of this method, is to divide the interval $[a, b]$, into *n* of subintervals, as follows:

$$x_0 = a\,, x_1 = x_0 + h \ldots\ldots \quad x_n = x_{n-1} + h = b$$

$$h = \frac{(b-a)}{n}$$

In order to find the approximate solution of the initial value problem at the point $x_{i+1}, i = 0,1,2, \ldots n-1$, we consider the definition of $y'(x_i)$, as follows:

$$y'(x_i) = \lim_{x \to x_i} \left(\frac{y(x_{i+1}) - y(x_i)}{x_{i+1} - x_i} \right) \cong \frac{y(x_{i+1}) - y(x_i)}{x_{i+1} - x_i}$$

since, $y' = f(x, y)$, we obtain

$$\frac{y(x_{i+1}) - y(x_i)}{x_{i+1} - x_i} \cong f(x_i, y_i), \qquad i = 0,1,2, \ldots\ldots n$$

$$\therefore \quad y(x_{i+1}) \cong y(x_i) + hf\big(x_i, y(x_i)\big)$$

Assume that, y_i is the approximate value of $y(x_i)$, we get

$$y_{i+1} = y_i + hf(x_i, y_i) \qquad i = 0,1,2, \dots \dots \dots, n-1$$

The last equation is called **Euler formula.**

Remark:-

Euler formula, can only be used only for finding the numerical solutions at the points $x_i)_0^n$, while if we would like to find the approximate value of $y(x^*)$, where $x^* \neq x_i)_0^n$, $x^* \in [a, b]$, then in this case, we can use an interpolation method.

Steps of Euler algorithm

1-Input a, b

2-Input the points, x_1^*, x_2^*, ...

3-Define $f(x, y)$

4- Input, n, $h = \frac{(b-a)}{n}$

5- Set $x_{i+1} = x_i + h$, $i = 0,1,2, \dots . n-1$

6-Compute the approximate values of $y(x_i), i = 1,2, \dots n$, using Euler formula:

$$y_{i+1} = y_i + hf(x_i, y_i) \qquad i = 0,1,2, \dots \dots \dots, n-1$$

6- If $x_k^* \in [a, b]$, $x_k^* \neq x_i, \forall i$, then use a Newton's finite difference formula, to find the approximate value of the initial value problem at this point.

Example: - Consider that, we have the following initial value problem

$$y' = -xy, \qquad y(0) = 0.5$$
$$x \in [0,0.2],$$

Use Euler method to find the approximate values of y(0.1),y(0.15), y(0.2) and compare these values with the exact solution (find the absolute errors)

Solution :-

$$h=\frac{b-a}{2} = \frac{0.2-0}{2} = 0.1, \quad x_0 = 0\ , x_1 = 0.1\ , x_2 = 0.2$$

by using Euler formula

$$y_1 = y_0 + h\, f(x_0, y_0)$$

$$= 0.5 + (0.1)(-(0)(0.5)) = 0.5$$

$$y_2 = y_1 + h\, f(x_1, y_1)$$

$$= 0.5 + (0.1)(-(0.1)(0.5)) = 0.495$$

Thus, we get the following database

x	0	0.1	0.2
y	0.5	0.5	0.495

We can show that, by using separation of variables, the exact solution of this problem takes the form: $y = \frac{1}{2} e^{\frac{-x^2}{2}}$

which leads to $y(0.1) = 0.4975$, $y(0.2) = 0.4901$

Thus, the absolute errors can be computed as follows:

$$E_1 = |y(0.1) - y_1| = |0.4975 - 0.5| = 0.0025$$

$$E_2 = |y(0.2) - y_1| = |0.4901 - 0.495| = 0.0049$$

From the above database, we can compute the approximate value of $y(0.15)$, by using Backward Newton's formula studied in Chapter 4. (**H.W**.)

Modified Euler method

In order to get more accurate results than Euler method, we define the modified Euler method, which has the following general formula

$$y_{i+1} = y_i + \frac{h}{2}\left(f(x_i, y_i) + f(x_{i+1}, y^*_{i+1})\right), \qquad \textbf{Modified Euler}$$

where

$$y^*_{i+1} = y_i + h\ f(x_i, y_i) \qquad \textbf{Normal Euler}$$

$$i = 0,1,2, \ldots \ldots . n-1$$

In fact, the first equation depends on the second equation, and this way is called (**Estimation–Correction**), because from the first equation, we get estimate value for $y(x^*)$, and then in the second equation, we get a correction for this value.

The steps of Modified Euler algorithm are the same as the steps of normal Euler algorithm and we only have to add one step more after step 5, which is:

6- Correct the estimated value y_{i+1}^*, that we get from using normal Euler formula, by using the modified Euler formula.

Example:- Use Modified Euler method for the last example.
Solution

$h = \frac{b-a}{2} = \frac{0.2-0}{2} = 0.1, \quad n=2$

Estimation: $y_1^* = y_0 + h\ f(x_0, y_0)$

$$= 0.5 + (0.1)(0) = 0.5$$

Correction: $y_1 = y_0 + \frac{h}{2}(f(x_0, y_0) + f(x_1, y_1^*))$

$$= 0.5 + \frac{0.1}{2}\big(-(0.1)(0.5)\big) = 0.4975$$

Estimation: $y_2^* = y_1 + h\ f(x_1, y_1)$

$$= 0.4975 + (0.1)\big(-(0.1)(0.4975)\big) = 0.4925$$

Correction: $y_2 = y_1 + \frac{h}{2}(f(x_1, y_1) + f(x_2, y_2^*))$

$$= 0.4975 + \frac{(0.1)}{2}(-(0.1)(0.4975) - (0.2)(0.4925)) = 0.4901$$

Thus, we get the following data base

x	0	0.1	0.2
y	0.5	0.4975	0.4901

Next, we compute the absolute errors:

$$E_1 = |y(0.1) - y_1| = |0.4975 - 0.4975| = 0$$

$$E_2 = |y(0.2) - y_1| = |0.4901 - 0.4901| = 0$$

It is clear that, this database is much different from that we have got from using normal Euler method. Moreover, it is more accurate.

Matlab Codes of Euler methods

For the last example, write Matlab codes, that can be used to find $y(0.1), y(0.2)$, by using Euler & modified Euler methods

```
Euler
x0=0; y0=0.5;h=0.1; X(1)=x0+h; X(2)=X(1)+h;
x=sym('x');
y=sym('y');
f=-x*y;
for i=1:2
    Y(i)=y0+h*subs(f,{x,y},{x0,y0});
    x0=X(i); y0=Y(i);
end
disp(Y)
```

Answer: 0.5 0.4950

Modified Euler

```
x0=0; y0=0.5;h=0.1; X(1)=x0+h; X(2)=X(1)+h;
x=sym('x');
y=sym('y');
f=-x*y;
```

```
for i=1:2
    Ye(i)=y0+h*subs(f,{x,y},{x0,y0});

Yc(i)=y0+(h/2)*(subs(f,{x,y},{x0,y0})+subs(f,{x,y},{X(i
),Ye(i)}))
    x0=X(i); y0=Yc(i);
end
disp(Yc)
```

Answer: 0.4975 0.4901

Explicit & Implicit Methods

Normal Euler method, belongs to group of methods called **explicit methods**, and that because of , computing y_{i+1} depends only on x_i ,while modified Euler methods belongs to **implicit methods**, and that because of , computing y_{i+1} depends on both of x_i & x_{i+1}

Runge-Kutta Methods

Since modified Euler method needs two steps to get the solutions, it is considered a two-steps method. Moreover, Euler methods need to approximate the derivatives by using special forms. Thus, we will use *Runge-Kutta* method , which is a one-step method and it can be used to avoid determining higher order derivatives.

Runge-Kutta Method of order 2:

Set $k_1 = hf(x_i, y_i)$

$k_2 = hf(x_i + h, y_i + k_1)$

Recall modified Euler formula

$$y_{i+1} = y_i + \frac{1}{2}\left(hf(x_i, y_i) + hf(x_{i+1}, y_{i+1}^*)\right),$$

where

$y_{i+1}^* = y_i + h\ f(x_i, y_i)$

This equation can rewrite as follows:

$$y_{i+1} = y_i + \frac{1}{2}(k_1 + k_2) \ldots\ldots \textbf{Ruga} - \textbf{kutta formula of order2}$$

Runge-Kutta Method of order 4:

Set $k_1 = hf(x_i, y_i)$,

$$k_2 = hf\left(x_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1\right),$$

$$k_3 = hf\left(x_i + \frac{1}{2}h, y_i + \frac{1}{2}k_2\right),$$

$$k_4 = hf(x_i + h, y_i + k_3)$$

$$y_{i+1} = y_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

which is **Ruga-Kutta** formula of order4.

Example:- Consider that, we have the following intial value problem:

$$y' = x + y, \qquad y(0) = 1$$

Use Runga-Kutta method of order 2 and order 4 to compute the approximate values of $y(0.25), y(0.5)$

Solution

$$h = 0.25, \qquad x_0 = 0, \qquad x_1 = 0.25, \quad x_2 = 0.5$$

$$i = 0, \qquad k_1 = hf(x_0, y_0) = 0.25(0 + 1) = 0.25$$

$$k_2 = hf(x_0 + h, y_0 + k_1) = 0.25(0.25 + 1 + 0.25) = 0.3750$$

$$y_1 = y_0 + \frac{1}{2}(k_1 + k_2) = 1 + 0.5(0.25 + 0.3750) = 1.3125$$

$$i = 1, \qquad k_1 = hf(x_1, y_1) = 0.25(0.25 + 1.3125) = 0.3906$$

$$k_2 = hf(x_1 + h, y_1 + k_1) = 0.25(0.25 + 0.25 + 1.3125 + 0.3906) = 0.5508$$

$$y_2 = y_1 + \frac{1}{2}(k_1 + k_2) = 1.3125 + 0.5(0.3906 + 0.5508) = 1.7832$$

Next, we compute these values by using Euler method of order4

$$i = 0, \qquad k_1 = hf(x_0, y_0) = 0.25(0 + 1) = 0.25$$

$$k_2 = hf\left(x_0 + \frac{1}{2}h, y_0 + \frac{1}{2}k_1\right) = 0.25\left(\frac{1}{8} + 1 + \frac{1}{8}\right) = 0.3125$$

$$k_3 = hf\left(x_0 + \frac{1}{2}h, y_0 + \frac{1}{2}k_2\right) = 0.25\left(\frac{1}{8} + 1 + \frac{0.3125}{2}\right) = 0.3203$$

$$k_4 = hf(x_0 + h, y_0 + k_3) = 0.25(0.25 + 1 + 0.3203) = 0.3926$$

$$y_1 = y_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

$$= 1 + \frac{1}{6}(0.25 + 2(0.3125) + 2(0.3203) + 0.3926) = 1.3180$$

$$i = 1, \qquad k_1 = hf(x_1, y_1) = 0.25(0.25 + 1.3180) = 0.3920$$

$$k_2 = hf\left(x_1 + \frac{1}{2}h, y_1 + \frac{1}{2}k_1\right) = 0.25\left(\frac{1}{4} + \frac{1}{8} + 1.3180 + \frac{0.3920}{2}\right)$$

$$= 0.4723$$

$$k_3 = hf\left(x_1 + \frac{1}{2}h, y_1 + \frac{1}{2}k_2\right) = 0.25\left(\frac{1}{4} + \frac{1}{8} + 1.3180 + \frac{0.4723}{2}\right)$$

$$= 0.4823$$

$$k_4 = hf(x_1 + h, y_1 + k_3) = 0.25(0.25 + 0.25 + 1.3180 + 0.4823)$$

$$= 0.5751$$

$$y_2 = y_1 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

$$= 1.3180 + \frac{1}{6}(0.3920 + 2(0.4723) + 2(0.4823) + 0.5751)$$

$$= 1.7974$$

Since $y' = x + y$ is linear equation, we can show that, by using integration factor, the exact solution of problem in the last example takes the form:

$$y = 2e^x - x - 1$$

Thus $y(0.25) = 1.3181,\ y(0.5) = 1.7974$

Thus, the absolute errors, obtained by Ruge-Kutta of order 2 are as follows:

$$E_1 = |y(0.25) - y_1| = |1.3181 - 1.3125| = 0.0056$$

$$E_2 = |y(0.5) - y_1| = |1.7974 - 1.7832| = 0.0142$$

while, the absolute errors, obtained by Ruge-Kutta of order 4 are as follows:

$$E_1 = |y(0.25) - y_1| = |1.3181 - 1.3180| = 0.0001$$

$$E_2 = |y(0.5) - y_1| = |1.7974 - 1.7974| = 0$$

It is clear that, the results of Ruga-Kutta method of order 4 are much accurate than the results of Ruga-Kutta method of order 2.

Matlab Codes for Runge-Kutta Methods

We can write matlab codes that can be used to solve the last example by using Ruga-Kutta method of order 2 and 4, as follows:

Order2

```
x0=0; y0=1; h=0.25;
X(1)=0.25; X(2)=0.5;
x=sym('x');
y=sym('y');
f=x+y;
for i=1:2
    k1=subs(f,{x,y},{x0,y0});
    k2=subs(f,{x,y},{x0+h,y0+h*k1});
    Y(i)=y0+(h/2)*(k1+k2);
    x0=X(i);
    y0=Y(i);
end
    disp(Y)
```

```
    1.3125    1.7832
```

Order4

```
   x0=0; y0=1; h=0.25;
X(1)=0.25; X(2)=0.5;
x=sym('x');
y=sym('y');
f=x+y;
for i=1:2
    k1=subs(f,{x,y},{x0,y0});
    k2=subs(f,{x,y},{x0+h/2,y0+h/2*k1});
    k3=subs(f,{x,y},{x0+h/2,y0+h/2*k2});
    k4=subs(f,{x,y},{x0+h,y0+h*k3});
    Y(i)=y0+(h/6)*(k1+2*k2+2*k3+k4);
    x0=X(i);
    y0=Y(i);
end
    disp(Y)
```

```
  1.3180    1.7974
```

Exercises

Q1: Consider the following initial value problem:

$$y' = x/y \ , \quad y(0) = 2 \ .$$

a. Use **Runge-Kutta** method of order 2 to find the approximate values of $y(0.1),\ y(0.2), y(0.3)$ and $y'(0.2)$.

b. Depending on the results of a. , find the approximate value of $y(0.15)$

c. Find the absolute error at each point in a.

Q2: Consider the following initial value problem:

$$y' - 2xy = 0 \ , \quad y(0) = 1$$

$$x \in [0, 0.4]$$

a. Use **Modified Euler** method to find the approximate values for $\{x_i\}_{i=1}^{2}$, with $h = 0.2$
b. Find the absolute error at each point.