Numerical Methods

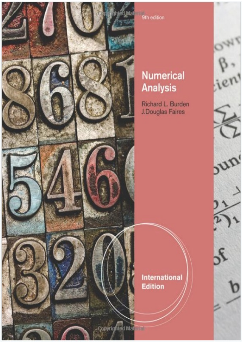
DS288 and UMC201

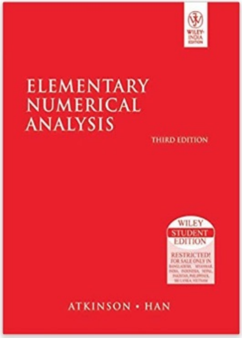
Ratikanta Behera

Department of computational and Data Sciences,
Indian Institute of Science Bangalore

August-December 2025

Books (Numerical Methods)





- Numerical Analysis (Richard L. Burden & J. Douglas Faires)
- Elementary Numerical Analysis (Kendall Atkinson & Weimin Han).

Exam and Grading Policy

Grading Policy: The grade will be calculated as follows:

Assessment	Course Weight	Due date
Three Assignments	$30\ \% = 3 \times 10\%$	21-08-2025 16-09-2025 16-10-2025
Midterm	20%	23-09-2025
Final Project	20 %	21-11-2025
Final Exam	30 %	-

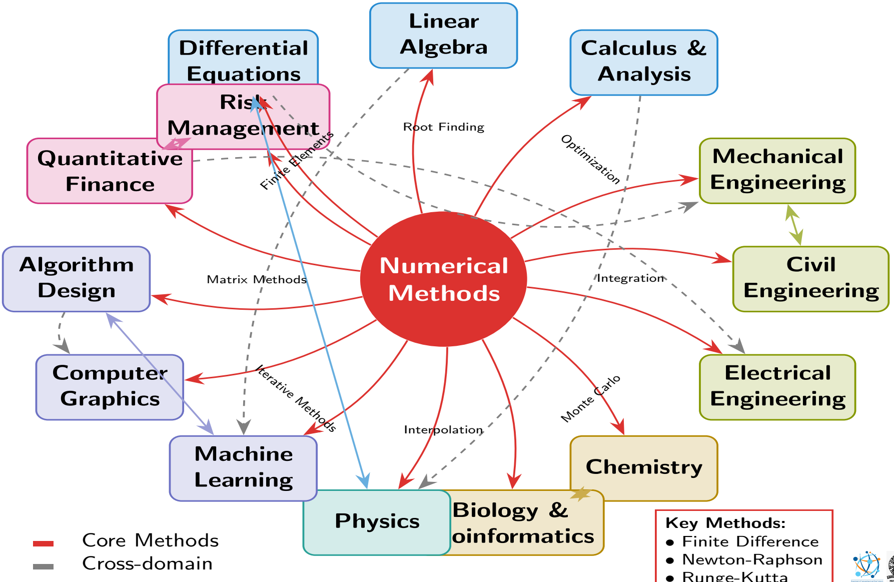
What are Numerical Methods?

- Numerical methods are mathematical techniques used to solve problems that cannot be solved analytically or where analytical solutions are impractical.
- Numerical methods provide approximate solutions using computational algorithms.

Why Study Numerical Methods?

1. Real-world problems often lack closed-form solutions
2. Complex systems require computational approaches
3. Engineering applications rely heavily on numerical solutions
4. Scientific computing is fundamental to modern research

Numerical Methods



Key Methods:
- Finite Difference
- Newton-Raphson
- Runge-Kutta
- Interpolation

Scientific Computing

- Scientific Computing is the collection of tools, techniques, and theories required to solve mathematical models of problems in Science and Engineering on a computer.

Scientific Computing

- Scientific Computing is the collection of tools, techniques, and theories required to solve mathematical models of problems in Science and Engineering on a computer.
- ☞ This set of mathematical theories and techniques is called Numerical Mathematics or Numerical Methods.

Scientific Computing

- Scientific Computing is the collection of tools, techniques, and theories required to solve mathematical models of problems in Science and Engineering on a computer.
- This set of mathematical theories and techniques is called Numerical Mathematics or Numerical Methods.
- Numerical methods are procedures that allow for efficient solution of a mathematically formulated problem in a finite number of steps to within an arbitrary precision.

Scientific Computing

- Scientific Computing is the collection of tools, techniques, and theories required to solve mathematical models of problems in Science and Engineering on a computer.
- This set of mathematical theories and techniques is called Numerical Mathematics or Numerical Methods.
- Numerical methods are procedures that allow for efficient solution of a mathematically formulated problem in a finite number of steps to within an arbitrary precision.

Numerical Calculation vs Symbolic Calculation

- Numerical Calculation:

Numerical Calculation vs Symbolic Calculation

- Numerical Calculation: (involve numbers directly) $\rightarrow$ Manipulate numbers to produce a numerical result

➢ Example:

$$\frac{(17.36)^2 - 1}{17.36 + 1} = 16.36 \tag{1}$$

- Symbolic Calculation:

Numerical Calculation vs Symbolic Calculation

- Numerical Calculation: (involve numbers directly) $\rightarrow$ Manipulate numbers to produce a numerical result

➢ Example:

$$\frac{(17.36)^2 - 1}{17.36 + 1} = 16.36 \tag{1}$$

- Symbolic Calculation: (symbols represent numbers) $\rightarrow$ Manipulate symbols according to mathematical rules to produce a symbolic result

➢ Example:

$$\frac{x^2 - 1}{x + 1} = x - 1 \tag{2}$$

Analytic Solution Vs Numerical Solution

- Analytic Solution:

Analytic Solution Vs Numerical Solution

- Analytic Solution: The exact numerical/symbolic representation of the solution. It may use special characters such as

$$\frac{1}{4},\ \frac{1}{5},\ 7\pi,\quad e,\quad or\quad tan(83) \tag{3}$$

- Numerical Solution:

Analytic Solution Vs Numerical Solution

- Analytic Solution: The exact numerical/symbolic representation of the solution. It may use special characters such as

$$\frac{1}{4}, \ \frac{1}{5}, \ 7\pi, \quad e, \quad or \quad tan(83) \tag{3}$$

- Numerical Solution: The computational representation of the solution. It is entirely numerical
- Example:

$$.25, \ 0.33333\cdots, \ 3.14159\cdots, \quad 0.88472\cdots$$

Floating point representation

A non-zero real number x to be stored in the computer of the form

$$x = \pm(0.a_1 a_2 a_3 \cdots a_k) \times 10^n \tag{4}$$

where $1 \leq a_1 \leq 9$, and $0 \leq a_i \leq 9, \quad i = 2, 3, \cdots, k$

- $10 \longrightarrow$ base number.
- $n \longrightarrow$ an integer called exponent (positive or negative or zero)
- $0.a_1 a_2 a_3 \cdots a_k \longrightarrow$ mantissa

Here the Eq. (4) is called normalized k decimal floating point form.

Example

Normalized decimal representation the following numbers

$$12345.67 \longrightarrow 0.1234567 \times 10^5 \tag{5}$$

$$0.00123 \longrightarrow 0.123 \times 10^{-2} \tag{6}$$

Significant Figures

Significant figures are the number of digits in a value that are known with some degree of confidence, often a measurement, that contribute to the degree of accuracy of the value.

- All non zero digits are significant.
  - 549 $\rightarrow$ has three significant figures
  - 1.892 $\rightarrow$ has four significant figures

Significant Figures

Significant figures are the number of digits in a value that are known with some degree of confidence, often a measurement, that contribute to the degree of accuracy of the value.

- All non zero digits are significant.
  - 549 $\rightarrow$ has three significant figures
  - 1.892 $\rightarrow$ has four significant figures
- Zeros between non zero digits are significant.
  - 4023 $\rightarrow$ has four significant figures
  - 50014 $\rightarrow$ has five significant figures

Significant Figures

Significant figures are the number of digits in a value that are known with some degree of confidence, often a measurement, that contribute to the degree of accuracy of the value.

- All non zero digits are significant.
  - 549 $\rightarrow$ has three significant figures
  - 1.892 $\rightarrow$ has four significant figures
- Zeros between non zero digits are significant.
  - 4023 $\rightarrow$ has four significant figures
  - 50014 $\rightarrow$ has five significant figures
- Zeros to the left of the first non zero digit are not significant.
  - 0.000034 $\rightarrow$ has only two significant figures. (as it is $3.4x10^{-5}$)
  - 0.001111 $\rightarrow$ has four significant figures.

Approximation of numbers

- Most real numbers x cannot be represented exactly by the normalized decimal floating-point form.
- We approximate by a nearby number to represent in a computer.

➥ Translating a real number x into k-digit floating point number

Suppose we want to round off the number (Rounding or Chopping)

$$x = \pm(0.a_1 a_2 a_3 \cdots a_k a_{k+1} \cdots) \times 10^n \quad \text{with} \quad a_1 \neq 0$$

➢ let $fl(x)$ denote its normalized decimal floating-point form. Then

$$fl(x) = \pm(0.a_1 a_2 a_3 \cdots a_k{}^*) \times 10^n \text{ where}$$

$$\begin{aligned} a_k{}^* &= a_k && \textit{if} \quad a_{k+1} < 5 \quad \text{and} \\ a_k{}^* &= a_k + 1 && \textit{if} \quad 5 \leq a_{k+1} \end{aligned}$$

ERROR

The following types of error will come for any numerical computation.

- Inherent error:

ERROR

The following types of error will come for any numerical computation.

- Inherent error: Errors which are already present in the statement of the problem before its solution.
- Rounding error:

ERROR

The following types of error will come for any numerical computation.

- Inherent error: Errors which are already present in the statement of the problem before its solution.
- Rounding error: Arise from the process of rounding off the numbers during the computations.
- Truncation error:

ERROR

The following types of error will come for any numerical computation.

- Inherent error: Errors which are already present in the statement of the problem before its solution.
- Rounding error: Arise from the process of rounding off the numbers during the computations.
- Truncation error: This errors are caused by using approximate results or replacing an infinite process by a finite one. (type of algorithm error)
- Absolute, Relative and Percentage errors

Absolute, Relative and Percentage errors

- Consider $X \longrightarrow$ true value of the quantity and
- $X' \longrightarrow$ approximate value
  - Absolute error, $E_a = |X - X'|$
  - Relative error, $E_r = \frac{|X-X'|}{|X|}$
  - Percentage error, $E_p = 100E_r = 100\frac{|X-X'|}{|X|}$

Remark

- The relative and percentage error are independent of unit.
- Absolute error is expressed in terms of these unit.

Chapter-2 (Solving Nonlinear Equations)

Introduction

- In scientific and engineering studies, a frequently occurring problem is to find the roots or zeros of equations of the form $f(x) = 0$. (Root Finding Problem)
- $f(x)$ may be algebraic or transcendental or a combination of both.
- Algebraic functions of the form $P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$, are called polynomials.
- A non-algebraic function is called a transcendental function.

Chapter-2 (Solving Nonlinear Equations)

Introduction

- In scientific and engineering studies, a frequently occurring problem is to find the roots or zeros of equations of the form $f(x) = 0$. (Root Finding Problem)
- $f(x)$ may be algebraic or transcendental or a combination of both.
- Algebraic functions of the form $P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, are called polynomials.
- A non-algebraic function is called a transcendental function.

Numerical Methods

Polynomials up to degree 4 can be solved exactly. Finding the root of $f(x) = 0$ analytically is not always possible.

- Bisection Method, Regula Falsi Method
- Newton-Raphson, Secant, Fixed-Point Method.

Convergence and order of convergence

Convergence: The method is said to be convergent if the sequence $x_0, x_1, x_2, \cdots, x_n, \cdots$ convergences to the true solution, i.e., for a given $\epsilon > 0$ there exists a positive integer n_0 such that

$$\lim_{n\to\infty} x_n = \alpha \quad or \quad |x_n - \alpha| < \epsilon \quad for \quad n \geq n_0 \tag{7}$$

Convergence and order of convergence

Convergence: The method is said to be convergent if the sequence $x_0, x_1, x_2, \cdots, x_n, \cdots$ convergences to the true solution, i.e., for a given $\epsilon > 0$ there exists a positive integer n_0 such that

$$\lim_{n\to\infty} x_n = \alpha \quad or \quad |x_n - \alpha| < \epsilon \quad for \quad n \geq n_0 \tag{7}$$

Rate of convergence: Let x_n be a sequence that converges to α. If there exists a sequence $\{\beta_n\}$ that converges to zero and a positive constant c, independent of n, such that

$$|x_n - \alpha| \leq c|\beta_n| \quad \text{for sufficiently large } n. \tag{8}$$

Then $\{x_n\}$ is said to converge to α with rate of convergence $O(\beta_n)$.

Convergence and order of convergence

Convergence: The method is said to be convergent if the sequence $x_0, x_1, x_2, \cdots, x_n, \cdots$ convergences to the true solution, i.e., for a given $\epsilon > 0$ there exists a positive integer n_0 such that

$$\lim_{n\to\infty} x_n = \alpha \quad or \quad |x_n - \alpha| < \epsilon \quad for \quad n \geq n_0 \tag{7}$$

Rate of convergence: Let x_n be a sequence that converges to α. If there exists a sequence $\{\beta_n\}$ that converges to zero and a positive constant c, independent of n, such that

$$|x_n - \alpha| \leq c|\beta_n| \quad \text{for sufficiently large } n. \tag{8}$$

Then $\{x_n\}$ is said to converge to α with rate of convergence $O(\beta_n)$.
Order of convergence: The sequences $\{x_n\}$ converges to α with an order of convergence $p \geq 1$ if

$$|x_{n+1} - \alpha| \leq c|x_n - \alpha|^p \quad for \quad n \geq 0 \tag{9}$$

where $c \geq 0$ is the asymptotic error constant.

Convergence and rate of convergence

Let $x_0, x_1, \cdots$, be the values of the, α of an equation at the $0th, 1st, 2nd, \cdots$ iterations, while its actual value is 3.5567.

Root	1st Method	2nd Method	3rd Method
x_0	5	5	5
x_1	5.6	3.8527	4
x_2	6.4	3.5693	3.8327
x_3	8.3	3.5589	3.5683
x_4	9.7	3.5578	3.5567
x_5	10.6	3.5577	
x_6	11.9	3.5567	

Convergence and rate of convergence

Let $x_0, x_1, \cdots$, be the values of the, α of an equation at the $0th, 1st, 2nd, \cdots$ iterations, while its actual value is 3.5567.

Root	1st Method	2nd Method	3rd Method
x_0	5	5	5
x_1	5.6	3.8527	4
x_2	6.4	3.5693	3.8327
x_3	8.3	3.5589	3.5683
x_4	9.7	3.5578	3.5567
x_5	10.6	3.5577	
x_6	11.9	3.5567	

➢ Rate of convergence:

Convergence and rate of convergence

Let $x_0, x_1, \cdots,$ be the values of the, α of an equation at the $0th, 1st, 2nd, \cdots$ iterations, while its actual value is 3.5567.

Root	1st Method	2nd Method	3rd Method
x_0	5	5	5
x_1	5.6	3.8527	4
x_2	6.4	3.5693	3.8327
x_3	8.3	3.5589	3.5683
x_4	9.7	3.5578	3.5567
x_5	10.6	3.5577	
x_6	11.9	3.5567	

- Rate of convergence: How fast the method converge
- Linear convergent

Convergence and rate of convergence

Let $x_0, x_1, \cdots$, be the values of the, α of an equation at the $0th, 1st, 2nd, \cdots$ iterations, while its actual value is 3.5567.

Root	1st Method	2nd Method	3rd Method
x_0	5	5	5
x_1	5.6	3.8527	4
x_2	6.4	3.5693	3.8327
x_3	8.3	3.5589	3.5683
x_4	9.7	3.5578	3.5567
x_5	10.6	3.5577	
x_6	11.9	3.5567	

- Rate of convergence: How fast the method converge
- Linear convergent $\rightarrow |e_{n+1}| \leq c|e_n|$, where $c < 1$.
- Order p convergent

Convergence and rate of convergence

Let $x_0, x_1, \cdots$, be the values of the, α of an equation at the $0th, 1st, 2nd, \cdots$ iterations, while its actual value is 3.5567.

Root	1st Method	2nd Method	3rd Method
x_0	5	5	5
x_1	5.6	3.8527	4
x_2	6.4	3.5693	3.8327
x_3	8.3	3.5589	3.5683
x_4	9.7	3.5578	3.5567
x_5	10.6	3.5577	
x_6	11.9	3.5567	

- Rate of convergence: How fast the method converge
- Linear convergent $\rightarrow |e_{n+1}| \leq c|e_n|$, where $c < 1$.
- Order p convergent $\rightarrow |e_{n+1}| \leq c|e_n|^p$, where c is not necessarily less than 1.

Detail convergence analysis

- $p = 1, p = 2$ and $p = 3 \longrightarrow$ linear convergence, quadratic convergence, and cubic convergence.
- Solve a problem with three methods ($p = 1, p = 2$ and $p = 3$).
- Consider the asymptotic error constant $c = 0.5$ and $e_0 = 1$

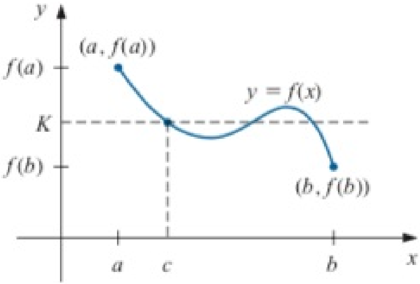
	Linear $\lvert e_{n+1}\rvert \approx 0.5\lvert e_n\rvert$	Quadratic $\lvert e_{n+1}\rvert \approx 0.5\lvert e_n\rvert^2$	Cubic $\lvert e_{n+1}\rvert \approx 0.5\lvert e_n\rvert^3$
e_1	0.5	0.5	0.5
e_2	0.25	0.125	0.0625
e_3	0.125	$7.8125 * 10^{-3}$	$1.2207 * 10^{-4}$
e_4	0.0625	$3.0518 * 10^{-5}$	$9.0949 * 10^{-13}$
e_5	0.03125	$4.6566 * 10^{-10}$	$3.7616 * 10^{-37}$
e_6	0.015625	$1.0842 * 10^{-19}$	
e_7	$7.8125 * 10^{-3}$	$5.8775 * 10^{-39}$	

The Existence of Roots

First finding an interval which is guaranteed to contain a root and then systematically shrinking the size of that interval.

Theorem-1(Intermediate Value Theorem (IVT)):

Let $f(x)$ be continuous function in $[a, b]$, where $a < b$ and let k be any number between $f(a)$ and $f(b)$. Then, there exists a number c in (a, b) such that $f(c) = k$.



The Existence of Roots

Corollary-1 (Application of IVT)

If f is a continuous function on the closed interval $[a, b]$ where $f(a) < 0 < f(b)$ or $f(a) > 0 > f(b)$, then f contains at least one root on the interval $[a, b]$.

The Existence of Roots

Corollary-1 (Application of IVT)

If f is a continuous function on the closed interval $[a, b]$ where $f(a) < 0 < f(b)$ or $f(a) > 0 > f(b)$, then f contains at least one root on the interval $[a, b]$.

➢ Hence, an equation $f(x) = 0$, where $f(x)$ is a real continuous function, has at least one root between a and b if $f(a)f(b) < 0$.

The Existence of Roots

Corollary-1 (Application of IVT)

If f is a continuous function on the closed interval $[a, b]$ where $f(a) < 0 < f(b)$ or $f(a) > 0 > f(b)$, then f contains at least one root on the interval $[a, b]$.

➢ Hence, an equation $f(x) = 0$, where $f(x)$ is a real continuous function, has at least one root between a and b if $f(a)f(b) < 0$.

Example

Show that $x^5 - 2x^3 + 3x^2 - 1 = 0$ has a solution in the interval $[0, 1]$.

$$f(0) = -1 < 0 \quad \text{and} \quad f(1) = 1 > 0$$

Note: This method may produce a false root if $f(x)$ is discontinuous $[a, b]$.

Bisection Method

The Bisection Method (Algorithm)

Algorithm for Bisection Method

- Suppose $f(x)$ is continuous and we have two numbers a_1 and b_1 such that $f(a_1)f(b_1) < 0,$ then by Intermediate Value Theorem we know that there is a root for $f(x)$ between a_1 and $b_1.$
- Then divide the interval into two parts and take the middle point suppose that is $x_1,$
  - if $f(x_1) = 0$ then we got the root,
- if not so then by IVT the root lies within a_1 and x_1 or x_1 and $b,$ depending on whether $f(a_1)f(x_1) < 0$ or $f(x_1)f(b_1) < 0.$

The Bisection Method (Example)

Example

Find an approximation to $\sqrt{3}$ correct to two decimal places

Here $f(x) = x^2 - 3$ and $f(1) < 0$ and $f(2) > 0$ root lies in $[1, 2]$.

- $x_1 = (a + b)/2 = 1.5$ and $f(x_1) = -7.5 < 0$

$$f(1.5) < 0 \quad \text{and} \quad f(2) > 0 \text{ root lies in } [1.5, 2]$$

- $x_2 = (x_1 + b)/2 = 1.75$ and $f(x_2) = 0.0625 > 0$

$$f(1.5) < 0 \quad \text{and} \quad f(x_2) > 0 \text{ root lies in } [1.5, 1.75]$$

- $x_3 = (x_1 + x_2)/2 = 1.625$ and $f(x_3) = -0.359375 < 0$
- $x_4 = 1.6875, \quad x_5 = 1.71875, \quad x_6 = 1.734375$

Number of Iterations for Bisection Method

- ➢ Each new interval contains the root.
- ➢ Each interval is the half the length of the previous interval.
- ➥ Thus, the interval width is reduced by the factor of $\frac{1}{2}$ at each time.
- ➢ At the end of the nth step, the interval length $\frac{(b-a)}{2^n}$
- ➢ Suppose the nth interval is less than ϵ, i.e., $\frac{(b-a)}{2^n} \leq \epsilon$.

$$n \geq \frac{log(b-a) - log\epsilon}{log2}$$

Example

A root of the equation $f(x) = x^3 + x - 4 = 0$ lies in the interval $(1, 4)$. Find the number of iterations necessary to obtain an approximation to the root with an error less than $\epsilon = 10^{-3}$.

Number of Iterations for Bisection Method

- Each new interval contains the root.
- Each interval is the half the length of the previous interval.
- Thus, the interval width is reduced by the factor of $\frac{1}{2}$ at each time.
- At the end of the nth step, the interval length $\frac{(b-a)}{2^n}$
- Suppose the nth interval is less than ϵ, i.e., $\frac{(b-a)}{2^n} \leq \epsilon$.

$$n \geq \frac{log(b-a) - log\epsilon}{log2}$$

Example

A root of the equation $f(x) = x^3 + x - 4 = 0$ lies in the interval $(1, 4)$. Find the number of iterations necessary to obtain an approximation to the root with an error less than $\epsilon = 10^{-3}$.

- Here $a = 1, b = 4$. Thus $\frac{log(3) - log(10^{-3})}{log(2)} \approx 11.5$, i.e. n=12


ANY
QUESTIONS?