MINOR-1

MINOR-1

Numerical Methods and Computation

M1 25
M2 25
Q 10
M 40

---

x_i	f_i	f_i'
x_0	f_0	f_0'
.		
.		
.		
x_n	f_n	f_n'

x_i's are distinct

What is the polynomial $H(x)$ such that

$$H(x_i) = f(x_i)$$
$$H'(x_i) = f'(x_i)$$

] Hermite interpolation

This process is called INTERPOLATION, more precisely Polynomial Interpolation.

Approximation

Given $\varepsilon > 0$, $\exists$ a polynomial of degree n such that $|f(x_i) - P(x_i)| < \varepsilon \quad \forall \quad x \in [a, b]$

$$f(x) \simeq P(x_i \; c_0, c_1 \cdots c_n)$$

$$= c_0 P_0(x) + c_1 P_1(x) \cdots c_n P_n(x)$$

where P_i s are linearly independent.

Problem is to determine c_i s such that error/ norm $E(f_i, x)$ is small

$$E(f_i, x) = \| f(x_i) - [c_0 P_0(x_i) \cdots + c_n P_n(x_i)] \|$$

→ Least Squares Approximation
→ Uniform Approximation

Numerical Differentiation

x_i	f_i
x_0	f_0
⋮	⋮
x_n	f_n

→ $n+1$ nodal points and their values.

$f'(x_i)$

$f''(x_i)$

[error appears in -ve powers of x.]

Numerical Integration

[error is in +ve powers of x]

$$\int_a^b f(x)\,dx \qquad \int_0^\infty e^{-x^2}\,dx.$$

→ Newton-cotes formulas

→ gauss-Quadrature formulas.

Linear algebra problems

1. Solution of $AX = b$
2. Eigenvalues and eigenvectors of A.

We don't use cramer's rule to solve $Ax = b$ because it is computationally not practical

20×20 matrix [or 20 eqn 20 var] will take more than a million years.

Direct methods

- Gauss elimination
- LU decomposition
- LL^T Cholesky fac.

Iterative methods

→ Improve after each iteration

- Gauss Jacobi
- Gauss Seidel
- SOR Relocation Method.

→ For order greater than 5, find roots by iterative methods.

Transcendental → other than polynomial
(e^x, $\sin x$, $\cos hx$ etc]

Methods for transcedental can be used for polynomial as well.

→ Bisection Method
Regular Falsi Method
Secant Method
Newton Raphson
One point iteration

For ODEs

→ Taylor's series method
→ Euler's / Improved / Modified / Backward Euler's method
→ Range - Kutta method.
→ 4th order classical R-K

↓

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

$y(x_0) = y_0$.

Computer Representation of Numbers and Floating point arithmatic.

All real numbers cannot be represented in computers → only a subset of them can be represented

General representation of $x \in R$ in base β is given by –

$$x = \pm \left(a_n \beta^N \cdots + a_1 \beta + a_0 + a_1' \beta^{-1} \right.$$
$$\left. \cdots a_p' \beta^{-p} \right)$$
$$= \pm \left(a_N a_{N-1} \cdots a_1 a_0 \cdots a_p' \right)_\beta$$

where $0 \le N < \infty$, $1 \le p < \infty$

$a_N \neq 0$ is the most significant digit

$a, a_i' \in S_\beta$

$S_{10} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

$S_2 = \{0, 1\}$

$S_5 = \{0, 1, 2, 3, 4\}$

The number is characterised by -

① its sign $\pm$

② its integral part

$$E(x) = \sum_{i=1}^{N} a_i \beta^i$$

③ its fractional part

$$F(x) = \sum_{i=1}^{p} a_i' \beta^{-i}$$

leading to general expression of x

$$x = \pm \left(E(x) + F(x) \right)$$

$$= \pm \left(E(x) \cdot F(x) \right)$$

$\downarrow$

(decimal)

in case $p = \infty$, the fractional part is said to be infinite

Representing a real no in decimal form $\rightarrow$

37.2	$=$	$.372 \times 10^2$
$.0031$	$=$	$.31 \times 10^{-2}$
29.33	$=$	$.2933 \times 10^2$

floating point scientific

Jan-6 Class 3

This is called Normalized Scientific Notation.
Also called as Normalized Floating Point Representation.

NOTE : Unless it is the 0 number, first digit should not be 0

$$x = \pm 0.d_1 d_2 d_3 \ldots \times 10^n$$

where $d_1 \neq 0$ and
n is an integer.

The real number x, if different from 0, can be represented in normalised floating decimal

$$x = \pm r \times 10^n$$

$$\frac{1}{10} \leq r < 1 \qquad r \in \left[\frac{1}{10}, 1\right)$$

$r \rightarrow$ normalised mantissa
$n \rightarrow$ exponent

//

Any real number x can be represented as

$$x = \pm (.d_1 d_2 \cdots d_n d_{n+1} \cdots)_\beta \times \beta^e$$

with $d_1 \neq 0$ or $d_1 = d_2 \cdots = d_n = 0$.

or

$$x = (-1)^s \times (.d_1 d_2 \cdots d_n d_{n+1} \cdots)_\beta \times \beta^e$$

where $s = 1$ or 0.

or

$$x = \sigma \tilde{x} \beta^e$$

e: exponent

$\sigma = +1$ or -1 $\quad \tilde{x} = .d_1 d_2 \cdots$ (decimal point)

(Notation in Atkinson)

β-fraction

or

or

$$x = \pm m \times \beta^e$$

m-digit mantissa or significand.

upto which we want to retain

Formal def of Floating Point :

Let x be a non zero real number. An n-digit floating point number in base β has the form

$$fl(x) = (-1)^s \times (.d_1 d_2 \ldots d_n)_\beta \times \beta^e$$

(floating)

where

$$(.d_1 d_2 \ldots d_n)_\beta = \frac{d_1}{\beta} + \frac{d_2}{\beta^2} + \ldots \frac{d_n}{\beta^n}$$

is a β-fraction called mantissa or significand, $s = 1$ or 0

e is an integer called exponent

The number β is also called radix

eg $\beta = 2$, floating point representation is called the binary floating point representation

$\beta = 10$ — decimal —.

Normalization

A floating point number is said to be normalised if either $d_1 \neq 0$ or

$$d_1 = d_2 = \ldots = d_n = 0$$

→ exponent e is limited To a range

$$m \leq e \leq M$$

During calculation, if some computed number has exponent $e > M$, we say memory overflow and if $e < m$, we say memory underflow.

$\beta, t, m, M.$

example of overflow.

Let $\beta = 10 \quad t = 3 \quad L = -3 \quad , \quad U = 3$

$a = .111 \times 10^3 \qquad b = .12 \times 10^3$

t → precision # of digits in decimal place.

$c = ab = .133 \times 10^5$ ← results in overflow

∴ exponent 5 is too large.

example of underflow

Let $\beta = 10 \quad t = 3 \quad L = -2 \quad U = 3$

$a = .1 \times 10^{-1} \qquad b = .2 \times 10^{-1}$

$ab = .2 \times 10^{-3}$ ↑ results in underflow.

→ Underflow is not a big issue. Computer treats it as 0 and carried the calculation forward

→ Overflow results in NAN → Not A Number error.

$0 . \infty \rightarrow$ Also results in NAN

→ [will only consider t digits during calc]

The number n is called precision or length of floating point representation.

$$x = (-1)^s \times (.d_1 d_2 \cdots d_n d_{n+1} \cdots)_\beta \times \beta^e$$

After n, the number is chopped or rounded

The chopped approximation is given by -

$$fl(x) = (-1)^s \times (.d_1 d_2 \cdots d_n)_\beta \times \beta^e$$

The rounded approximation is given by -

$$fl(x) = (-1)^s \times (.d_1 d_2 \cdots d_n)_\beta \times \beta^e$$

$$, 0 \leq d_{n+1} < \frac{\beta}{2}$$

$$fl(x) = (-1)^s \times (.d_1 d_2 \cdots (d_{n+1}))_\beta \times \beta^e$$

$$, \frac{\beta}{2} \leq d_{n+1} < \beta$$

Jan 10 Classy

Floating point representation of a binary number is given by

$$f(x) = (-1)^s \times (1.a_1 a_2 \cdots a_n) \times 2^e \quad - (*)$$

a_1 to a_n are 0 or 1

This is bound to be 1 if $x \neq 0$ ⇒ no storage req.

→ The IEEE 754 standard always uses binary operations

→ IEEE single precision floating point format uses 4 bytes (32 bits) to store a number.

→ Out of 32 bits, 24 are allocated for storing mantissa, 1 for sign (s) and 8 for exponent e

[→ One binary digit needs 1 bit storage space]

The storage scheme is given by –

[Single Precision]

Sign b_1 | exponent $b_2 \cdots b_9$ | mantissa $b_{10} \cdots b_{32}$

$b_1 = 0$ $\sigma = 1$

$b_1 = 1$ $\sigma = -1$

1	23	8
sign	mantissa	expo.

⇒ There are only 23 bits used for mantissa because the digit 1 before the binary point in (*) is not stored in the memory and will be inserted at the time of calculation.

[So one bit "1" is not stored]

Instead of the exponent e, we store the non negative integer $E = (b_1 \cdots b_9)_2$ and define $E = e + 127$. If all b_i's $(i = 2, 3, \ldots 9)$ are zero, then $E = (0)_{10}$ and if all b_i's are 1 then $E = (255)_{10}$. [see table on next page]

In addition to this one space corresponding to $e = 128$ (or $E = 255$) is ∞ or NaN depending on whether $b_{10} = \cdots b_{32} = 0$ or otherwise. Then in IEEE754 we have $-126 \leq e \leq 127$ (Note that it is not from -127 but -126, because the number is reserved as those numbers not represented otherwise and one memory space for NaN.)

(*) The decimal number 0 needs a special representation, which is stored as $E = 0$ (i.e. $b_2 = \cdots = b_9 = 0$), $b_1 = 0$ and $b_{10} = \cdots b_{32} = 0$)

$E = e + 127$

$E = (c_1 \dots c_8)_2$	x
$(00000000)_2 = (0)_{10}$	$\pm (0.a_1a_2 \dots a_{23})_2 \cdot 2^{-126}$
$(00000001)_2 = (1)_{10}$	$\pm (1.a_1a_2 \dots a_{23})_2 \cdot 2^{-126}$
$(00000010)_2 = (2)_{10}$	$\pm (1.a_1a_2 \dots a_{23})_2 \cdot 2^{-125}$
$\vdots$	$\vdots$
$(01111111)_2 = (127)_{10}$	$\pm (1.a_1a_2 \dots a_{23})_2 \cdot 2^{0}$
$(10000000)_2 = (128)_{10}$	$\pm (1.a_1a_2 \dots a_{23})_2 \cdot 2^{1}$
$\vdots$	$\vdots$
$(11111101)_2 = (253)_{10}$	$\pm (1.a_1a_2 \dots a_{23})_2 \cdot 2^{126}$
$(11111110)_2 = (254)_{10}$	$\pm (1.a_1a_2 \dots a_{23})_2 \cdot 2^{127}$
$(11111111)_2 = (255)_{10}$	$\pm\infty$ if $a_1 = \dots = a_{23} = 0$; *NaN* otherwise

Advantage of binary over other base systems
We do not need to store the "1" in memory.

In IEEE single precision storage system the overflow occurs for real number $|x| > x_{max}$

$$x_{max} = 1.1111\ldots1 \times 2^{127}$$

$$\approx 3.40 \times 10^{38}$$

The IEEE double precision floating point representation of a number has a precision of 53 binary digits and the exponent e is limited by

$$-1022 \le e \le 1023$$

Note

$$x = \sigma \quad \bar{x} \cdot \beta^e$$

The allowable number of binary digits in $\bar{x}$ is called the precision of the binary floating point representation of x.

→ IEEE single precision floating-point representation of x has a precision of 24 binary digits and exponent e is limited by $-126 \leq e \leq 127$.

$$x = \sigma \cdot (1.a_1 a_2 \ldots a_{23}) \cdot 2^e$$

in binary,

$$-(1111111\,0)_2 \leq e \leq (1111111)_2$$

2.1.1 Accuracy of Floating-Point Representation

Consider how accurately a number can be stored in the floating-point representation. This is measured in various ways, with the *machine epsilon* being the most popular. The machine epsilon for any particular floating-point format is the difference between 1 and the next larger number that can be stored in that format. In single precision IEEE format, the next larger binary number is

$$1.00000000000000000000001 \tag{2.8}$$

with the final binary digit 1 in position 23 to the right of the binary point. Thus, the *machine epsilon in single precision IEEE format* is 2^{-23}. As an example, it follows that the number $1 + 2^{-24}$ cannot be stored exactly in IEEE single precision format. From

$$2^{-23} \doteq 1.19 \times 10^{-7} \tag{2.9}$$

we say that IEEE single precision format can be used to store approximately 7 decimal digits of a number x when it is written in decimal format. In a similar fashion, the *machine epsilon in double precision IEEE format* is $2^{-52} = 2.22 \times 10^{-16}$; IEEE double precision format can be used to store approximately 16 decimal digits of a number x. In MATLAB, the machine epsilon is available as the constant named eps.

Any positive integer $\leq M$ can be represented exactly in this floating-point representation. In the IEEE single precision format,

$$M = 2^{24} = 16777216$$

Different Types of Error

Error = True value − Approximate value.

Absolute error is the absolute value of error defined above.

Relative error is the measure of the error in relation to the size of the True value as given by

$$\text{Relative error} = \frac{\text{error}}{\text{True value}}$$

$$\% \text{ error} = \frac{\text{error}}{\text{True value}} \times 100$$

$$= \text{Relative error} \times 100.$$

Truncation error is used to denote the error resulting from approximating smooth function by truncating its Taylor series representation as a finite number of terms

Let x_A be the approximation of real number x. Then

$$E(x_A) = \text{Error}(x_A) = x - x_A$$

$$E_a(x_A) = \text{Absolute error}(x_A) = |E(x_A)|$$

$$E_r(x_A) = \text{Relative error}(x_A) = \frac{E(x_A)}{x}$$

Significant Digit

If x_A is an approximation to x then we say x_A approximates x to r significant β-digits if

$$|x - x_A| \leq \frac{1}{2}\beta^{s-r+1}$$ ← smallest. → least upper bound

where s the largest integer such that

$$\beta^s \leq |x|$$

eg –

$$x = \frac{1}{3}$$

Approx no $= 0.333$

has 3 significant digits, since

$$|x - x_A| = .00033 < .0005 = .5 \times 10^{-3}$$

but $10^{-1} < .333\ldots = x$.

Therefore, in this case $s = -1$

and hence $r = 3$

eg For $x = .02138$, the approximate number $x_A = 0.02144$ has absolute error $|x - x_A| = .000006 < .0005 = .5 \times 10^{-3}$

But $10^{-2} < .02138 = x$

Therefore in this case, $s = -2$ and therefore the number x_A has only 2 significant digits, but not 3, with respect to x.

Jan 11 Class5

NOTE

⇒ In a very simple way, the number of leading non-zero digits of x_A that are correct relative to the corresponding digits in the true value of x is called the NUMBER OF SIGNIFICANT DIGITS in x_A

Propagation of Error

x_A & y_A denote the numbers used in calculations

x_T & y_T be the corresponding true values

1. Propagated error in addition & subtraction

$$x_T = x_A + \varepsilon \qquad \varepsilon, \eta \rightarrow \text{errors}$$
$$y_T = y_A + \eta$$

Let these be +ve numbers.

Relative error $E_R(x_A \pm y_A)$ is given by -

$$E_R(x_A \pm y_A) = \frac{(x_T \pm y_T) - (x_A \pm y_A)}{x_T \pm y_T}$$

$$= \frac{x_T \pm y_T - (x_T - \varepsilon \pm y_T - \eta)}{x_T \pm y_T}$$

$$= \frac{\varepsilon \pm \eta}{x_T \pm y_T}$$

This shows that relative error propagates slowly with addition ($\because$ large denominator)

whereas, this amplifies drastically with subtraction when $x_T \approx y_T$ ($\because$ denominator $\to 0$ and $E_r \to \infty$.

2. Propagated error in multiplication

The relative error $E_r(x_A \times y_A)$

$$E_r(x_A \times y_A) = \frac{(x_T \times y_T) - (x_A \times y_A)}{x_T \times y_T}$$

$$= \frac{(x_T \times y_T) - ((x_T - \varepsilon) \times (y_T - \eta))}{x_T \times y_T}$$

$$= \frac{\eta x_T + \varepsilon y_T - \varepsilon\eta}{x_T \times y_T}$$

$$= \boxed{\frac{\varepsilon}{x_T} + \frac{\eta}{y_T} - \frac{\varepsilon}{x_T}\frac{\eta}{y_T}}$$

$$= E_r(x_A) + E_r(y_A) - E_r(x_A)\,E_r(y_A)$$

This shows that relative error propagates slowly with multiplication.

3. Propagated error in division

The relative error $E_r\left(\frac{x_A}{y_A}\right)$ is given by

$$E_r\left(\frac{x_A}{y_A}\right) = \frac{\frac{x_I}{y_T} - \frac{x_A}{y_A}}{\frac{x_T}{y_T}}$$

$$= \frac{\frac{x_I}{y_T} - \frac{x_T - \xi}{y_T - \eta}}{\frac{x_T}{y_T}}$$

$$= \frac{x_T(y_T - \eta) - y_T(x_T - \xi)}{x_T(y_T - \eta)}.$$

$$= \frac{y_T\xi - x_T\eta}{x_T(y_T - \eta)}$$

$$= \frac{y_T}{y_T - \eta}\left(\frac{\varepsilon}{x_T} - \frac{\eta}{y_T}\right)$$

$$\boxed{= \frac{1}{1 - \frac{\eta}{y_T}}\left(\frac{\varepsilon}{x_T} - \frac{\eta}{y_T}\right)}$$

$$= \frac{1}{1 - Er(y_A)}\left(Er(x_A) - Er(y_A)\right)$$

This shows that relative error in division propagates very slowly unless $Er(y_A) \simeq 1$. But $1 - Er(y_A) \simeq 0$ is very unlikely because we always expect error to be very small i.e. very close to 0, in which case, the expression becomes approximately equal to

$$Er\left(\frac{x_A}{y_A}\right) = Er(x_A) - Er(y_A).$$

4. Total Calculation Error

$$w : +, -, \times, \div$$

when using floating point arithmatic on a computer, the calculation of $x_A \, w \, y_A$ involves additional rounding or chopping error

The computed value $x_A w y_A$ will involve the propagated error

To be more precise, let $\hat{w}$ denote the complete operation as carried out on a computer involving any rounding or chopping. Then the total error is given by -

$$(x_T \, w \, y_T) - (x_A \, \hat{w} \, y_A)$$

$$= \underbrace{[(x_T w y_T) - (x_A w y_A)]}_{\text{propagation error}} + \underbrace{[(x_A w y_A) - (x_A \hat{w} y_A)]}_{\text{rounding or chopping error.}}$$

eg Propagation error in function evaluation

Q Consider evaluating f using approximate value x_A rather than x. How well $f(x_A)$ approximate $f(x)$?

Using mean value theorem, we get

$$f(x) - f(x_A) = f'(\xi_1)(x - x_A)$$

ξ is an unknown btw (x & x_A)

The relative error of $f(x_A)$ w.r.t. $f(x_A)$ is given by –

divide & multiply by x.

$$Er(f(x_A)) = \frac{f'(\xi)}{f(x)}(x - x_A) = \frac{f'(\xi)}{f(x)} \times Er(x)$$

Since x and x_A are assumed to be very close to each other and ξ lies between x and x_A, we make the approximation $\xi \to x$ and hence

$$\boxed{Er(f(x_A)) = \frac{f'(x)}{f(x)} \times Er(x)}$$

Defn (Condition Number of a function)

The condition number of a function f at a point $x = c$ is given by

$$\left| \frac{f'(c)}{f(c)} \right| c$$

f must be differentiable at c

Condition(x) [no. of x]

$$c(x) = \left| \frac{f'(x)}{f(x)} x \right|$$

Condition number tells us how much larger the relative perturbation in y is compared to relative perturbation in x.

derivation: Let $y = f(x)$

$$\Delta y = f(x+\Delta x) - f(x)$$

$$\frac{\Delta y}{y} = \frac{f(x+\Delta x)}{f(x)} - 1 = \left[\frac{f(x) + f'(x)\Delta x + \frac{f''(x)}{2!}\Delta x^2 \ldots}{f(x)}\right]$$

$$= 1 + \frac{f'(x)}{f(x)}\Delta x - 1$$

$$= \frac{f'(x)}{f(x)}\Delta x = \left(\frac{x f'(x)}{f(x)}\right)\frac{\Delta x}{x}$$

↓

condition number.

Jan 13 Class 6

eg $f(x) = e^x$

$$C(x) = \left| \frac{e^x}{e^x} x \right| = |x|,$$ which is large for large x.

is ill-conditioned when x is large.

eg $f(x) = \sqrt{x}$

$$C(x) = \left| \frac{\frac{1}{2\sqrt{x}}}{\sqrt{x}} x \right| = \frac{1}{2}$$

i.e. Taking square root is a well conditioned process

eg $f(x) = \frac{10}{1-x^2}$

$$f'(x) = \frac{-10}{(1-x^2)^2}(-2x) = \frac{20x}{(1-x^2)^2}$$

$$C(x) = \left| \frac{f'(x)}{f(x)} x \right| = \left| \frac{\frac{20x}{(1-x^2)^2}}{\frac{10}{1-x^2}} x \right| = \left| \frac{2x^2}{1-x^2} \right|$$

This no. could be very large near 1 or -1

Suppose for x is a vector, the condition number for a vector can be defined in a same way by using norms instead of absolute value.

The condition number of function of several variables (or a vector) can be defined by replacing $f'(x)$ by ∇f

$$c(x) = \frac{\|x\| \quad \|\nabla f\|}{\|f\|}$$

Eg $x = (x_1, x_2)^T$ a vector and problem is to obtain the scalar $f(x) = x_1 - x_2$

$$\nabla f = (1, -1)^T$$

The condition no of f (w.r.t. infinite norm)

$$c(x) = \frac{\|x\| \quad \|\nabla f\|_\infty}{\|f\|_\infty} = \frac{\max(|x_1|, |x_2|)}{\|x_1 - x_2\|}$$

which means that the problem is ill conditioned if $x_1 \approx x_2$

(1) One Norm $\|\vec{v}\|_1$

The one-norm (also known as the L_1-norm, ℓ_1 norm, or mean norm) of a vector $\vec{v}$ is denoted $\|\vec{v}\|_1$ and is defined as the sum of the absolute values of its components:

$$\|\vec{v}\|_1 = \sum_{i=1}^{n} |v_i| \tag{1}$$

for example, given the vector $\vec{v} = (1, -4, 5)$, we calculate the one-norm:

$$\|(1, -4, 5)\|_1 = |1| + |-4| + |5| = 10$$

(2) Two Norm $\|\vec{v}\|_2$

The two-norm (also known as the L_2-norm, ℓ_2-norm, mean-square norm, or least-squares norm) of a vector $\vec{v}$ is denoted $\|\vec{v}\|_2$ and is defined as the square root of the sum of the squares of the absolute values of its components:

$$\|\vec{v}\|_2 = \sqrt{\sum_{i=1}^{n} |v_i|^2} \tag{2}$$

for example, given the vector $\vec{v} = (1, -4, 5)$, we calculate the two-norm:

$$\|(1, -4, 5)\|_2 = \sqrt{|1|^2 + |-4|^2 + |5|^2} = \sqrt{42}$$

(3) Infinity Norm $\|\vec{v}\|_\infty$

The infinity norm (also known as the L_∞-norm, ℓ_∞-norm, max norm, or uniform norm) of a vector $\vec{v}$ is denoted $\|\vec{v}\|_\infty$ and is defined as the maximum of the absolute values of its components:

$$\|\vec{v}\|_\infty = \max\{|v_i| : i = 1, 2, \ldots, n\} \tag{3}$$

for example, given the vector $\vec{v} = (1, -4, 5)$, we calculate the infinity-norm:

$$\|(1, -4, 5)\|_\infty = \max\{|1|, |-4|, |5|\} = 5$$

(4) p-Norm $\|\vec{v}\|_p$

In general, the p-norm (also known as L_p-norm or ℓ_p-norm) for $p \in \mathbb{N}$, $1 \leq p < \infty$ of a vector $\vec{v}$ is denoted $\|\vec{v}\|_p$ and is defined as:

$$\|\vec{v}\|_p = \sqrt[p]{\sum_{i=1}^{n} |v_i|^p} \tag{4}$$

Condition number of a matrix vector produce

Suppose $A \in R^{m \times n}$, x is a vector

The condition no. of Ax (w.r.t. perturbation of x) is given by

$$K = \frac{\|A\| \quad \|x\|}{\|Ax\|}$$

where matrix norm to subordinate to vector norm

If A is square and invertible, $K = \|A\| \, \|A^{-1}\|$

$$\text{General condition no} = \frac{\text{relative error in solution}}{\text{relative perturbation in data}}$$

$$\lim_{\varepsilon \to 0} \sup \left\{ \frac{\|f(x) - f(y)\|}{\|f(x)\|} \;\middle|\; \|x - y\| < \varepsilon \right\}$$

Norm	Vector	Matrix
One-norm	$\|x\|_1 = \sum_i \|x_i\|$	$\|A\|_1 = \max_j \sum_i \|a_{ij}\|$
Two-norm	$\|x\|_2 = (\sum_i \|x_i\|^2)^{1/2}$	$\|A\|_2 = \max_{x \neq 0} \|Ax\|_2 / \|x\|_2$
Frobenius norm	$\|x\|_F = \|x\|_2$	$\|A\|_F = (\sum_{ij} \|a_{ij}\|^2)^{1/2}$
Infinity-norm	$\|x\|_\infty = \max_i \|x_i\|$	$\|A\|_\infty = \max_i \sum_j \|a_{ij}\|$

Stability and Instability in evaluating a function :

If any one step is ill conditioned while evaluating the function then the total process is said to have instability

Stability → only when all steps are well-conditioned.

eg : Consider $f(x) = \sqrt{x+1} - \sqrt{x}$

for sufficiently large x, the condition no is

$$C(x) = \left| \frac{\frac{1}{2\sqrt{x+1}} - \frac{1}{2\sqrt{x}}}{\sqrt{x+1} - \sqrt{x}} \, x \right|$$

$$= \frac{1}{2} \left| \frac{\sqrt{x} - \sqrt{x+1}}{(\sqrt{x+1} - \sqrt{x})\sqrt{x}\sqrt{x+1}} \right|$$

$$\frac{1}{2} \left| \frac{x}{x\sqrt{1+\frac{1}{x}}} \right| \approx \frac{1}{2}$$

which is quite good.

$f(12345)$ in six digit rounding arithmetic

$$f(12345) = \sqrt{12346} - \sqrt{12345}$$
$$= 111.113 - 111.108 = 0.005$$

while actually $f(12345) = .00450032$
The answer has 10% error

Let us evaluate computational errors. It consists of following four steps –

$$x_0 = 12345$$
$$x_1 = x_0 + 1$$
$$x_2 = \sqrt{x_1}$$
$$x_3 = \sqrt{x_0}$$
$$x_4 = x_2 - x_3$$

Now consider the last 2 steps when we already computed x_2 and now going to compute x_3 and finally evaluate the function

$$f_3(t) = x_2 - t$$

At this step, the condition no. for f_3 is given by

$$\left| \frac{f'(t)}{f(t)} t \right| = \left| \frac{t}{x_2 - t} \right|$$

Thus f is ill conditioned when t approaches x_2 for an instance $t \simeq 111.11$ while $x_2 - t \simeq .005$, The condition no for f_3 to t approximately 22.222 or more than 40.000 times as big as

the condition number of f itself. Therefore the above process of evaluating the function $f(x)$ is UNSTABLE

Let us rewrite $f(x)$ as

$$f(x) = \frac{1}{\sqrt{x+1} + \sqrt{x}}.$$

In six-digit rounding arithmatic, this gives

$$f(12345) = .0045002$$

Which is in error by only .003%. The computational process –

$x_0 = 12345$ → constant, not a computed value

$x_1 = x_0 + 1$ → const + 1 (not $x+1$) (rather $x+1$!)

$x_2 = \sqrt{x_1}$

$x_3 = \sqrt{x_0}$

$x_4 = x_2 + x_3$

$x_5 = \frac{1}{x_4}$

It is easy to verify that the condition number of each of the above step is well conditioned. For instance, last step defines

$f_3 = \frac{1}{x_2 + t}$ and condition no of this

function is approximately

$$\left| \frac{f_3'(x)}{f_3(x)} x \right| = \left| \frac{t}{x_2 + t} \right| = \frac{1}{2}$$

for t sufficiently close to x_2

Therefore the process of evaluating $f(x)$ is STABLE.

Class 7 Jan 17

Theorem: Let $fl(x)$ denote the floating point representation of a real number x, then

$$\frac{|fl(x) - x|}{|x|} \leq u = \begin{cases} \beta^{1-t} & \text{for chopping} \\ \frac{1}{2}\beta^{1-t} & \text{for rounding} \end{cases} \quad -(1)$$

Proof

We establish the $\leq$ bound for rounding and leave the other part as H.W.

Let x be written as

$$x = (.d_1 d_2 \cdots d_t d_{t+1} \cdots) \times \beta^e$$

where $d_1 \neq 0 \quad 0 \leq d_i < \beta$

When we round off x we obtain one of the following floating point numbers -

$$x' = (.d_1 d_2 \cdots d_t) \times \beta^e$$

$$x'' = [(.d_1 d_2 \cdots d_t) + \beta^{-t}] \times \beta^e$$

Let $x \in (x', x'')$. Assume without loss of generality, that x is closer to x'. Then we have

$$|x - x'| \leq \frac{1}{2} |x' - x''| = \frac{1}{2} \beta^{e-t}$$

Thus the relative error

$$\frac{|x - x'|}{x} \leq \frac{1}{2} \left(\frac{\beta^{-t}}{.d_1 d_2 \cdots d_t} \right)$$

$$\leq \frac{1}{2} \frac{\beta^{-t}}{\frac{1}{\beta}} \quad (\because d_1 < \beta)$$

$$= \frac{1}{2} \beta^{1-t}$$

The number u in (1) is called machine precision or unit round off error. It is the smallest positive floating point number such that

$$fl(1+u) > 1$$

eg consider 3-digit representation of the decimal number $x = .2346$ $(\beta = 10, t = 3)$

in case of rounding, $x' = fl(x) = .235$

Relative error $= = .001705 < \frac{1}{2} 10^{-2}$

in case of chopping $x'' = fl(x) = .234$

Relative error $= .0025575 < 10^{-2}$

NOTE

(1) $$\frac{|fl(x) - x|}{|x|} \leq u = \begin{cases} \beta^{1-t} & \text{for chopping} \\ \frac{1}{2}\beta^{1-t} & \text{for rounding} \end{cases}$$

can be written in the form $fl(x) = x(1+\delta)$, where $|\delta| \leq u$

Laws of floating point arithmetic

[Assuming IEEE standards hold,]

Let x and y be two floating point numbers, and let $fl(x+y)$, $fl(x-y)$, $fl(xy)$, $fl\left(\frac{x}{y}\right)$ denote the computed sum, difference, product and division. Then

1. $fl(x+y) = (x \pm y)(1+\delta_1)$, $|\delta_1| \leq u$
2. $fl(xy) = xy(1+\delta_2)$, $|\delta_2| \leq u$
3. $fl\left(\frac{x}{y}\right) = \frac{x}{y}(1+\delta_3)$, $|\delta_3| \leq u$

The above results 1-3 of above theorem along with (*) the FUNDAMENTAL LAWS OF FLOATING POINT ARITHMATIC.

The fundamental laws form the basis for establishing bounds for relative errors in other floating point representation.

On the computers that do not use IEEE standard, the following floating point addition may hold

$$fl(x+y) = x(1+\delta_1) + y(1+\delta_2)$$

when $|\delta_1| \leq u$

$|\delta_2| \leq u$

To summarize,

Laws of Floating Point Arithmetic

$$fl(x \odot y) = (x \odot y)(1+\delta) \qquad |\delta| \leq u$$

where $\odot$ indicates any of the 4 basic operations $+ \; - \; \times \; \div$

$x, y \in F_s = F(\beta, t, L, U)$ (e_{min}, e_{max})

↳ single precision

set of all real numbers written as normalised floating point form.

Obviously $F \subset \mathbb{R}$

Algebraic Properties in Floating Point operations

Since F is a proper subset of R, elementary algebraic operations on floating point numbers do NOT satisfy all properties of algebraic operations in R.

1. Floating point addition is commutative in F.

$$x \oplus y = fl(x+y) = fl(y+x) = y \oplus x$$

2. Floating point multiplication is commutative in F

$$x \otimes y = fl(xy) = fl(yx) = y \otimes x$$

3. Floating point addition is NOT associative in F.

$$(x \oplus y) \oplus z \neq x \oplus (y \oplus z)$$

4. Floating point multiplication is NOT associative in F

$$(x \otimes y) \otimes z \neq x \otimes (y \otimes z)$$

5. Floating point multiplication is NOT distributive w.r.t. addition in F

$$x \otimes (y \oplus z) \neq (x \otimes y) + (x \otimes z)$$

[where 3, 4, 5 are true for real numbers]

Addition of n Floating Point numbers

consider adding n floating point numbers $x_1, x_2 \ldots x_n$.

Define $S_2 = fl(x_1 + x_2) = (x_1 + x_2)(1 + \delta_2)$, $|\delta_2| \leq u$

$$S_2 - (x_1 + x_2) = \delta_2 (x_1 + x_2)$$

Define $S_3, S_4 \ldots S_n$ recursively by

$$S_{i+1} = fl(S_i + x_{i+1}) \quad i = 2, 3 \ldots n-1$$

$$S_3 = fl(S_2 + x_3)$$

$$= (S_2 + x_3)(1 + \delta_3)$$

$$= x_1(1 + \delta_2)(1 + \delta_3) + x_2(1 + \delta_2)(1 + \delta_3) + x_3(1 + \delta_3)$$

$$S_3 - (x_1 + x_2 + x_3) = (x_1 + x_2)\delta_2 +$$
$$(x_1 + x_2)(1+\delta_2)\delta_3 +$$
$$x_3 \delta_3$$

(neglecting the terms $\delta_2 \delta_3$, which is very small and so on)

Thus by induction, we can see that

$$S_n - (x_1 + x_2 \cdots + x_n) \approx (x_1 + x_2)\delta_2$$
$$+ (x_1 + x_2 + x_3)\delta_3$$
$$\vdots$$
$$+ (x_1 + x_2 \cdots + x_n)\delta_n$$

— (*)

(neglecting $\delta_i \delta_j$ terms which are very small)

(*) can be written as

$$S_n - (x_1 + x_2 \cdots + x_n) = x_1(\delta_2 + \delta_3 \cdots + \delta_n)$$
$$+ x_2(\delta_2 + \delta_3 + \delta_4 \cdots + \delta_n)$$
$$\vdots$$
$$+ x_n \delta_n$$

where each $|\delta_i| \leq \frac{1}{2}\beta^{1-t} = u$ → in rounding

Defining $\delta_1 = 0$, we write the following theorem.

Theorem: (Rounding error in floating point addition)

Let $x_1, x_2 \cdots x_n$ be n floating point numbers. Then

$$Er\left[fl(x_1 + x_2 \cdots + x_n)\right] \approx x_1(\delta_1 + \delta_2 \cdots + \delta_n) + x_2(\delta_2 + \delta_3 \cdots + \delta_n) + \vdots + x_n \delta_n$$

where $|\delta_i| \leq u = \frac{1}{2}\beta^{1-t}$

and $|x_1| < |x_2| \cdots < |x_n|$

Class 8 Jan 18

Roots of non linear equations of one variable

$f(x) = 0$ — (1)

- algebraic
- transcedental

→
- finite roots
- infinite roots
- no roots

To find solution of (1) i.e. to find x^* such that $f(x^*) \simeq 0$

Assumptions:

1. f is a continuously differentiable and real valued function of f.
2. The equation (1) has only one isolated root. i.e. for each root of (1) there is a neighbourhood which does not contain any other root of the equation

[We will find only one root at a time]

Zero of a function - A number ξ is a solution of $f(x) = 0$ if $f(\xi) = 0$ then ξ is called a zero / root of the function.

Geometrically, a root of $f(x) = 0$ is value of x where $y = f(x)$ intersects x-axis

Multiple Root - If $f(x) = (x - \xi)^m g(x) = 0$ where $g(x)$ is bounded and $g(\xi) \neq 0$, then ξ is called a multiple root of $f(x) = 0$ with multiplicity 'm'

for $m = 1$, ξ is called a simple root

Key ideas in approximating an isolated root

1. Initial guess. The smallest interval $[a, b]$ that contains the root. Take a point $x_0 \in [a, b]$ as an initial guess

2. Improve value of root

The process of improving value of root is called iteration process and the methods are called iteration methods.

A general form of iterative method may be written as -

If $x_k, x_{k-1}, \cdots x_{k-m+1}$ are m approximations to the root, then a multipoint iteration method is defined as

$$x_{k+1} = \phi(x_k, x_{k-1}, \cdots, x_{k-m+1})$$

This is called m-point iteration method since m initial guesses are required

ϕ is called multi point iteration function.

If $m=1$, we get one point iteration method

$$x_{k+1} = \phi(x_k)$$

* In iteration process, we get a sequence of x $\{x_n\}$ and the sequence should converge to the root of (1)

- In practice, except in rare cases, it is not possible to find ξ which satisfies the given equation exactly. So therefore we attempt to find an approximation root ξ^* such that either

$$|f(\xi^*)| < \varepsilon$$

or

$$|x_{k+1} - x_k| < \varepsilon$$

where x_{k+1} and x_k are 2 successive iterates and ε is the error or tolerance

- If 1st 2nd 3rd 4th decimal accuracy is required, we choose ε to be

  1st — .05
  2nd — .005
  3rd — .0005
  4th — .00005

- **Convergence of a method**

If the sequence of approximations $x_0, x_1, \ldots x_n \ldots$ for the root ξ of $f(x)=0$ are such that $\lim_{n\to\infty} x_n = \xi$. We say that the method is convergent

[or $\lim_{n\to\infty} |x_n - \xi| = 0$]

- **Order of Convergence / Rate of Convergence**

Assume $\{x_n\}$ converges to ξ. Let $\varepsilon_n = x_n - \xi$, $n \geq 0$

If 2 constants $c \neq 0$ and $p > 0$ exist and

$$\lim_{n\to\infty} \frac{|x_{n+1} - \xi|}{|x_n - \xi|^p} = \lim_{n\to\infty} \frac{|\varepsilon_{n+1}|}{|\varepsilon_n|^p} = c \quad \text{— (*)}$$

The sequence is said to converge to ξ with order of convergence 'p'

If $p=1$, the convergence of sequence $\{x_n\}$ is called linear

$p=2$, — quadratic

Some sequence converge at a rate not equal to integer i.e. p need not be an integer

- $$\lim_{n\to\infty} \frac{|\varepsilon_{n+1}|}{|\varepsilon_n|^p} = C$$

For large values of n, we have an approximation

$$|\varepsilon_{n+1}| = C\,|\varepsilon_n|^p \quad \because \text{[used in proofs]}$$

→ A sequence is said to converges with order p to the root 'ξ' if

$$|x_{n+1} - \xi| \le C\,|x_n - \xi|^p, \quad n \ge 0$$

for some constant $C > 0$ and p is a positive real number.

BISECTION METHOD

This method is based on repeated application of Intermediate Value Theorem.

If f is a continuous function on some interval $[a, b]$ and $f(a) \cdot f(b) < 0$ ($\Leftrightarrow$ f(a) and f(b) have opposite signs) then $f(x) = 0$ has at least one real root and odd number of roots in (a, b).

Class 9, Jan 20

If we know that a root of $f(x)=0$ lies in the interval $I_0 = (a_0, b_0)$, we bisect I_0 at the point

$$m_1 = \frac{1}{2}(a_0 + b_0)$$

We denote $I_1 = (a_0, m_1)$ or (m_1, b_0) acc to:

- If $f(a_0) \times f(m_1) < 0 \rightarrow$ we bisect (a_0, m_1)
- If $f(m_1) \times f(b) < 0 \rightarrow$ we bisect (m_1, b_0)

Therefore I_1 contains the root. We bisect I_1 and get a sub-interval I_2 and continue the process

We thus obtain a sequence of nested intervals $I_0 > I_1 > I_2 \ldots$ such that each sub interval contains the root

After repeating the process q times, we either find the root or find interval I_q of length $\frac{b_0 - a_0}{2^q}$ which contains the root. We take mid-point of last interval as desired approximation of the root

$\xi^* =$ mid point of I_q

This root has error not greater than half of length of interval of which it is a mid-point.

$$|\xi^* - \xi| = \varepsilon \leq \frac{1}{2} \text{len}(I_q) = \frac{1}{2}\left(\frac{b_0 - a_0}{2^q}\right)$$

Then we have

$$m_{k+1} = a_k + \frac{1}{2}(b_k - a_k), \quad k = 0, 1, 2..$$

$$(a_{k+1}, b_{k+1}) = \begin{cases} (a_k, m_{k+1}) & \text{if } f(a_k) f(m_{k+1}) < 0 \\ (m_{k+1}, b_k) & \text{if } f(m_{k+1}) f(b_k) < 0 \end{cases}$$

If the bisection method is applied to a continuous function f on an interval $[a, b]$, where $f(a)f(b) < 0$, then after n steps an approximate root will have been computed with error of ε.

$$\frac{1}{2}\frac{(b_0 - a_0)}{2^n}$$

Convergence of bisection method.

Suppose α is the exact root

$$b_{k+1} - a_{k+1} = \frac{1}{2}(b_k - a_k), \quad b \geq 0$$

$$b_k - a_k = \frac{1}{2^k}(b_0 - a_0)$$

α lies in the interval (a_k, x_{k+1}) or (x_{k+1}, b_k) where $x_{k+1} = a_k + \frac{1}{2}(b_k - a_k)$

$$|\alpha - x_{k+1}| \leq x_{k+1} - a_k = b_k - x_{k+1}$$

$$= \frac{1}{2}(b_k - a_k)$$

[x_{k+1} is mid point of (a_k, b_k)]

$$|\alpha - x_{k+1}| \leq \frac{1}{2^{k+1}} (b_0 - a_0) \quad \text{— (*)}$$

This shows that x_{k+1} converges to α as $k \to \infty$

Alternate Condition for LINEAR CONVERGENCE:

→ A sequence $\{x_n\}$ exhibits a linear convergence to a limit x if there is a constant c in the interval $[0, 1)$ such that

$$|x_{n+1} - x| \leq c\,|x_n - x|$$

NOTE ✓ There is no lt $n \to \infty$ when $|c| < 1$

If this inequality is true for all n, then

$$\begin{aligned} |x_{n+1} - x| &\leq c\,|x_n - x| \\ &\leq c^2 |x_{n-1} - x| \\ &\leq c^3 |x_{n-2} - x| \\ &\vdots \\ &\leq c^n |x_1 - x_0| \end{aligned}$$

Thus it is a sequence of linear convergence that

$$|x_{n+1} - x| \leq A\,c^n \quad (0 \leq c < 1)$$

— (*)

The sequence produced by bisection method obeys inequality (*)

Note :

If we know beforehand how much tolerance we need, then no of steps n can be given as -

$$\frac{b-a}{2^n} \leq \varepsilon$$

$$\ln 2^n \geq \ln\left(\frac{b-a}{\varepsilon}\right)$$

$$n \geq \frac{\ln \frac{b-a}{\varepsilon}}{\ln 2}$$

$$n = \text{ceil}\left(\frac{\ln\left(\frac{b-a}{\varepsilon}\right)}{\ln 2}\right)$$

* Often the bisection method is used to get close to the root before we apply a faster method

Crude approximation:

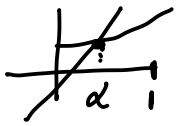
When solving a non linear equation $f(x)=0$, first we have to find a crude approximation x_0 the desired root. There are 3 ways to do this -

1. From graphical representation of a function
2. By tabulating function
3. By use of bisection method.

→ Sometimes it helps to rearrange the equation

$2^x - 5x + 2 = 0$

$2^x = 5x - 2$



⇒ It has exactly one root α between 0 & 1

Iteration methods based on first degree equations

If $f(x)$ is a first degree equation in x, then it can readily be solved ($ax+b=0 \Rightarrow x=-\frac{b}{a}$). We will study the iterative methods which will produce exact results whenever $f(x)=0$ is a first degree equation

If we approximate $f(x)$ by a first degree equation in the neighbourhood of the root, we may write

$$f(x) = a_1 x + a_0 = 0 \quad -(1)$$

solution of (1): $x = -\frac{a_0}{a_1}$, $a_1 \neq 0$ ——— (2)

a_1 and a_0 are to be determined by prescribing two approximate conditions on $f(x)$ or its derivatives

1. Secant method / Chord method
2. Regula Falsi method
3. Newton Raphson method.

Secant Method / Chord Method.

If x_{k-1}, x_k are 2 approximations to the root, then we determine a_0, a_1 using the conditions:

$$f_{k-1} = a_1 x_{k-1} + a_0$$
$$f_k = a_1 x_k + a_0$$

Solve for a_1 and a_0. $f_k - f_{k-1} = a_1 (x_k - x_{k-1})$

$$\Rightarrow \quad a_1 = \frac{f_k - f_{k-1}}{x_k - x_{k-1}} \quad \Big] -(3)$$

Solving for a_0 gives $\quad a_0 = \dfrac{x_k f_{k-1} - x_{k-1} f_k}{x_k - x_{k-1}}$

The next approximation for root is given by

$$x_{k+1} = \frac{-a_0}{a_1}$$

$$x_{k+1} = \frac{f_k x_{k-1} - f_{k-1} x_k}{f_k - f_{k-1}} \qquad -(4)$$

add and sub.

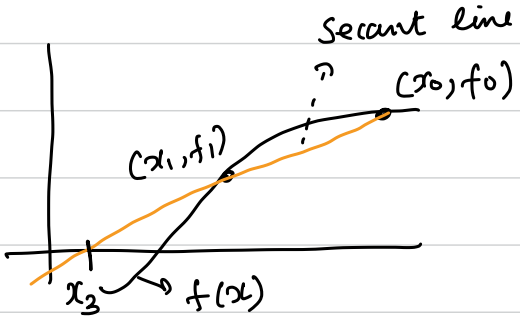
$$x_{k+1} = \frac{-x_k f_{k-1} + x_{k-1} f_k + x_k f_k - x_k f_k}{f_k - f_{k-1}}$$

$$= \frac{x_k (f_k - f_{k-1}) + x_{k-1} f_k - x_k f_k}{f_k - f_{k-1}}$$

$$\boxed{x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f_k - f_{k-1}} f_k} \quad -(5)$$

This is called Secant method or Chord method. (4 or 5)

Geometrically,



Geometrically, we replace the function $f(x)$ by a st. line or chord passing through the points (x_k, f_k) and (x_{k-1}, f_{k-1}) and take the point of intersection of st. line with x-axis as the next approximation to the root.

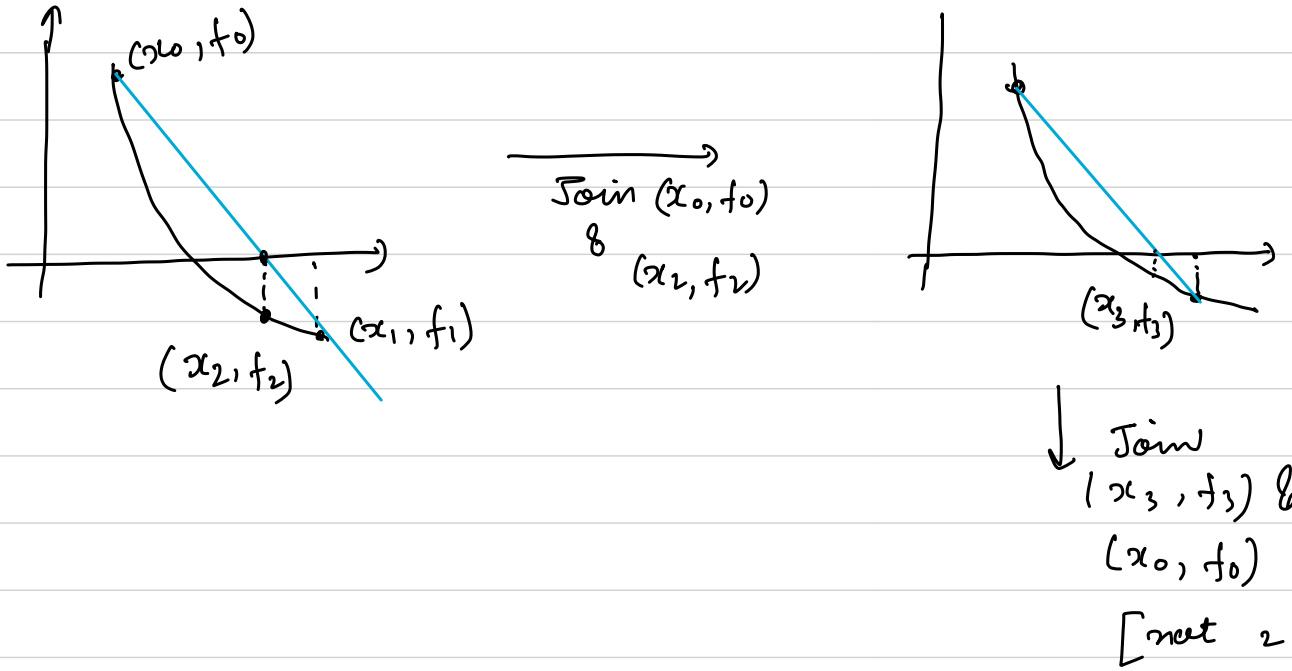
f

Regula Falsi Method.

If the approximations are such that $f_k f_{k+1} < 0$, then (4) or (5) is known as Regula falsi Method.

We can also devise regula falsi method or method of false position independently.

[Join f_1 and f_3, not $f_2 f_3$ if the sign of $f_1 f_3 < 0$]
This method always converges.



Newton Raphson Method

We determine a_1 and a_0 in (1) using the condition

$$f_k = a_1 x_k + a_0$$
$$f'_k = a_1$$

$$\therefore a_1 = f_k - f_k$$

Next approximation to the root is given by

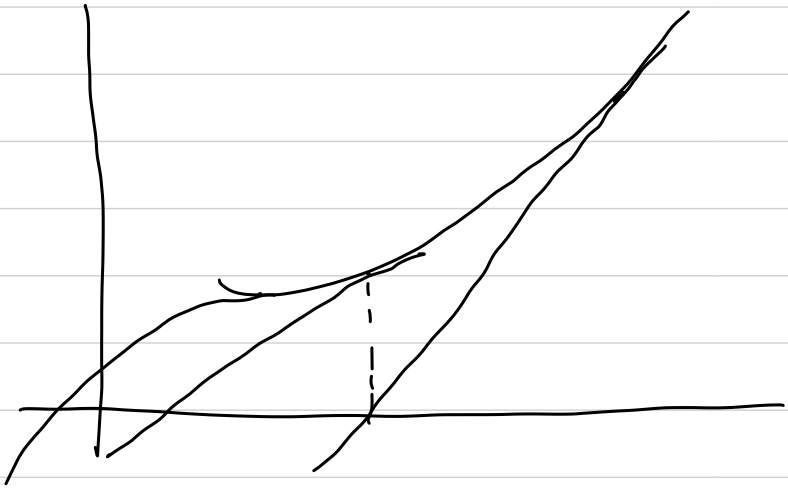
$$x_{k+1} = \frac{-a_0}{a_1} = -\frac{f_k - f_k' x_k}{f_k'}$$

$$\boxed{x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}} \quad \text{— (6)}$$

This is called the Newton Raphson Method.

Jan 24, Class 10

In $\lim x_{k+1} \longrightarrow x_k$, the chord joining the points (x_k, f_k) and (x_{k+1}, f_{k+1}) becomes tangent. Thus in this case, problem of finding root of equation $f(x) = 0$ is equivalent to finding the intersection of tangent to the curve $y = f(x)$ at point (x_k, f_k) with x-axis



→ Tangent method/ Newton Raphson.

Deriving Newton Raphson method using Taylor series

Let x_0 be an initial approximation to root ξ of $f(x) = 0$

$$\xi = x_0 + h$$

Since $f(\xi) = 0$

$$f(x_0+h) = 0$$

$$\Rightarrow f(x_0) + h f'(x_0) + \frac{h^2}{2!} f''(x_0) + \frac{h^3}{3!} f'''(x_0)$$

$$\cdots \quad \frac{h^n}{n!} f^n(x_0) = 0$$

Assume that h is small such that h^2 and higher powers of h can be neglected

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

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

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

x_1 is the first approximation to ξ

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

$$\vdots$$

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

→ This is called Newton Raphson method.

Error Analysis of Newton Raphson method.

Assume: ① f is at least twice continuously differentiable for all x in some interval about the root 'ξ'

② $f'(\xi) \neq 0$ i.e. simple root

This says that the graph of $y = f(x)$ is NOT tangent to the x-axis when the graph intersects it at $x = \xi$.

The case in which $f'(x) = 0$ will be treated separately.

Combining $f'(\xi) \neq 0$ with the continuity of $f'(x)$ implies that $f'(x) \neq 0$ for all x near to ξ

$$\varepsilon_n = x_n - \xi \quad : \text{error in } n\text{th iteration}$$

or

$$\xi = x_n - \varepsilon_n$$

$$f(\xi) = f(x_n - \varepsilon_n) = f(x_n) - \varepsilon_n f'(x_n) + \frac{\varepsilon_n^2}{2!} f''(x_n) - \frac{\varepsilon_n^3}{3!} f'''(x_n) \cdots$$

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

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

$$0 = x_n - x_{n+1} - \varepsilon_n + \frac{\varepsilon_n^2}{2!} \frac{f''(x_n)}{f'(x_n)}$$

$$0 = x_n - x_{n+1} - (x_n - \xi) + \frac{\varepsilon_n^2}{2!} \frac{f''(x_n)}{f'(x_n)}$$

$$0 = \xi - x_{n+1} + \frac{\varepsilon_n^2}{2!} \frac{f''(x_n)}{f'(x_n)}$$

$\xi - x_{n+1}$ → error in $(n+1)$ th iteration

$$\varepsilon_{n+1} = -\frac{1}{2}\frac{f''(x_n)}{f'(x_n)}\varepsilon_n^2$$

$$\frac{\varepsilon_{n+1}}{\varepsilon_n^2} = c \text{ (constant)}$$

⇒ This says that error at x_{n+1} is proportional to the square of error at x_n

⇒ Newton Raphson method has second order / Quadratic convergence.

Must Do Proof

Suppose $|f'(x)| \geq m$ and $|f''(x)| \leq M$ $\forall\ x \in I = [a,b]$

let $k = \frac{M}{2m}$

then $|x_{n+1} - \xi| \leq k\,|x_n - \xi|^2$ for $n \in N$

Q: Find the root of the equation $x^3 - 5x^2 + 6x - 1 = 0$ that lies between 3 and 4. by Newton-Raphson correct to fourth decimal place

Sol

$$f(x) = x^3 - 5x^2 + 6x - 1$$
$$f'(x) = 3x^2 - 10x + 6$$

$f(0) = -1$
$f(1) = 1$
] ⇒ one root lies btw 0 & 1

This is hit and trial when it is not given in question

$f(3) = 27 - 45 + 18 - 1 = -1$

$f(4) = 64 - 80 + 24 - 1 = 7$

$\Big]$ → one root lies btw. $(3,4)$

$\alpha \in (3,4)$

Let $x_0 = 3$

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

$$= 3 - \frac{(-1)}{3} = \frac{10}{3} = 3.33333$$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 3.33333 - \frac{f(3.33333)}{f'(3.33333)}$$

$$= 3.25309$$

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = 3.25309 - \frac{f(3.25309)}{f'(3.25309)}$$

$$= 3.24701$$

$$x_4 = x_3 - \frac{f(x_3)}{f'(x_3)} = 3.24698$$

$|x_4 - x_3| = .00003 < .00005$

∴ The fourth decimal place accuracy is reached.

$\alpha = 3.24698$ can be taken as the root.

At least 4 digits after decimal.

H.W.

Find a non zero root of $x^2 + 4\sin x = 0$ correct upto 4 decimal places.

ONE-POINT ITERATION METHOD.
or
Fixed point iteration

Let $f(x) = 0$ — ①

From ①, we can derive $x = g(x)$ — ②

Each solution of ② is a solution of ①. Each solution of ② is called a <u>fixed point of g</u>.

If $\phi = x - \frac{f(x)}{f'(x)}$ it becomes newton method

NA is specid case of o.p.i.

$$f(x) = x^2 - x + 2 = 0$$

ⓐ $x = x^2 + 2 = g(x)$ $\quad x_{k+1} = x_k^2 + 2$

or

$x^2 = x - 2$

ⓑ $x = \sqrt{x-2} = g(x)$ $\quad x_{k+1} = \sqrt{x_k - 2}$

ⓒ $x = 1 - \frac{2}{x} = g(x)$ $\quad x_{k+1} = 1 - \frac{2}{x_k}$

$$x_{k+1} = \phi(x_k)$$

This is one point iteration method. $\phi \rightarrow$ iteration func.

Which $g(x)$ to choose?

<u>Theorem</u>:

Assume that the iteration function ϕ has a real fixed point ξ and that $|\phi'(x)| \leq m < 1 \;\forall\; x \in I$ where I is an interval around ξ, $I = \{x : |x - \xi| \leq s\}$ for some s. Then —

ⓐ $x_k \in I$, $k = 0, 1, 2 \ldots$ [I is the interval, not integers]

ⓑ $\lim_{k \to \infty} x_k = \xi$

ⓒ ξ is the only root in I of the equation $x = \phi(x)$

↪ [see p.]

Class 11 Jan 25

Proof of theorem: (a) Prove it by induction

Assume $x_{k-1} \in I$. Mean value theorem gives

$$x_k - \xi = \phi(x_{k-1}) - \phi(\xi) = \phi'(\xi_k)(x_{k-1} - \xi)$$

and since ξ_k lies between x_{k-1} and $\xi \Rightarrow \xi_k \in I$.

$$\therefore |x_k - \xi| \leq m |x_{k-1} - \xi| \leq m\delta < \delta$$

which means $x_k \in I$.

$\Rightarrow$ Proved.

(b) From the above argument, we have

$$|x_k - \xi| \leq m |x_{k-1} - \xi|$$

$$\leq m\,m |x_{k-2} - \xi|$$

$$\vdots$$

$$\leq m^k |x_0 - \xi|$$

$m^k \to 0$ as $k \to \infty$ $\quad \lim_{k\to\infty} m^k = 0$

$$\Rightarrow \lim_{k\to\infty} x_k \to \xi$$

Proved.

(c) Uniqueness is proved by contradiction.

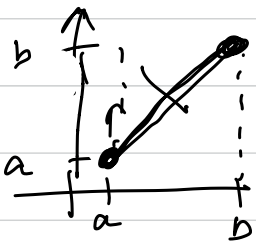
Assume that there is another root $\tilde{\xi}$,

$$\phi(\tilde{\xi}) = \tilde{\xi}, \quad \tilde{\xi} \neq \xi$$

The mean value theorem gives $\xi < \alpha < \tilde{\xi}$

$$|\xi - \tilde{\xi}| = |\phi(\xi) - \phi(\tilde{\xi})| = \phi'(\alpha)\,|\xi - \tilde{\xi}| \leq m\,|\xi - \tilde{\xi}|$$

which is a contradiction
(because we get $m \geq 1$)



If $\phi \in C[a, b]$ and $a \leq \phi(x) \leq b$ for all $x \in [a, b]$, then there is a fixed point x^* in $[a, b]$
If $\phi'(x)$ exists & $|\phi'(x)| \leq m < 1$, then fixed point is unique.

Proof: Suppose there are 2 values a and b such that $\phi(a) \geq a$ and $\phi(b) \leq b$. If $\phi(a) = a$ or $\phi(b) = b$, then a fixed point is found; so assume $g(a) > a$ and $g(b) < b$

Then condition for which $g(x) = \phi(x) - x$
We have $g(a) > 0$ and $g(b) < 0$. Hence by intermediate value theorem, there is a root x^*, $a < x^* < b$ such that $g(x^*) = 0$

Then $\phi(x^*) = x^*$
So x^* is a fixed point

(Uniqueness – same as above)

Defn Order of one point iteration method.

The iteration method $x_{n+1} = \phi(x_n)$, $n = 0, 1, 2 \cdots$ is said to be of order p if

$$\phi'(\xi) = \phi''(\xi) \cdots = \phi^{p-1}(\xi) = 0 \quad \text{and} \quad \phi^{(p)}(\xi) \neq 0$$

where ξ is solution of $x = \phi(x)$

→ Consider Newton – Raphson method $x_{n+1} = x_n - \dfrac{f(x_n)}{f'(x_n)}$

can be thought as one point iteration method

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

$$\begin{cases} x_{k+1} = g(x_k) \\ g(x_k) = x_k - \dfrac{f(x_k)}{f'(x_k)} \end{cases}$$

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

$$g'(x) = \frac{f(x)\, f''(x)}{[f'(x)]^2} \quad \text{— ✓ ①}$$

$$g'(\xi) = \frac{f(\xi)\, f''(\xi)}{[f'(\xi)]^2} = 0 \qquad \because f(\xi) = 0$$

but $f'(\xi) \neq 0$ for this.

$$g''(\xi) =$$

$$g''(\xi) = \frac{f'(\xi)\, f''(\xi)}{[f'(\xi)]^2}$$

$$g''(\xi) \neq 0$$

∴ Newton Raphson has quadratic convergence.

Convergence acceleration for fixed point iteration method or Aitkin's Δ^2 method :

Suppose the fixed point iteration method has at least first order convergence.

The linear convergence of iteration method can be improved with Aitkin's Δ^2 process.

Suppose x_n, x_{n+1}, x_{n+2} be three successive approximations to the root ξ of $x = \phi(x)$.

Error in 2 successive approximations may be written as -

$$\varepsilon_{n+1} = c\,\varepsilon_n \quad \text{— (1)}$$

$$\varepsilon_{n+2} = c\,\varepsilon_{n+1} \quad \text{— (2)}$$

$$c = \max\{c_1, c_2\}$$

$$\frac{\varepsilon_{n+1}}{\varepsilon_{n+2}} = \frac{\varepsilon_n}{\varepsilon_{n+1}}$$

$$(\varepsilon_{n+1})^2 = \varepsilon_n\,\varepsilon_{n+2}$$

$$(x_{n+1} - \xi)^2 = (x_n - \xi)(x_{n+2} - \xi)$$

$$x_{n+1}^2 + \xi^2 - 2x_{n+1}\xi = x_n x_{n+2} + \xi^2 - \xi x_{n+2} - \xi x_n$$

$$x_{n+1}^2 - 2x_{n+1}\xi = x_n x_{n+2} - \xi(x_{n+2} + x_n)$$

$$\xi = \frac{x_n x_{n+2} - x_{n+1}^2}{x_{n+2} + x_n - 2x_{n+1}}.$$

$$= \frac{x_n x_{n+2} - x_{n+1}^2 + (x_n^2 - 2x_n x_{n+1}) - (x_n^2 - 2x_n x_{n+1})}{x_{n+2} - 2x_{n+1} + x_n}$$

(Add, sub)

$$\text{Eq} = \frac{x_k\left[x_{k+2} + x_k - 2x_{k+1}\right] - x_k^2 + 2x_k x_{k+1} - x_{k+1}^2}{x_{k+2} - 2x_{k+1} + x_k}$$

$$\text{Eq} = x_k - \frac{(x_{k+1} - x_k)^2}{x_{k+2} - 2x_{k+1} + x_k}$$

Define $\Delta x_k = x_{k+1} - x_k$

$$\underline{\Delta^2 x_k} = \Delta(\Delta x_k)$$
$$= \Delta x_{k+1} - \Delta x_k$$
$$= x_{k+2} - x_{k+1} - (x_{k+1} - x_k)$$
$$= x_{k+2} - 2x_{k+1} + x_k$$

$$\boxed{\text{Eq} = x_k - \frac{(\Delta x_k)^2}{\Delta^2 x_k}}$$

Call $x^*_{k+2} = \text{Eq}$

$$x^*_{k+2} = x_k - \frac{(\Delta x_k)^2}{\Delta^2 x_k}$$

Eq or x^*_{k+2} gives an improved value of approximation of x_{k+2}

This has second order convergence.

→ $x = e^{-x}$ has a root (unique) in $[0, 1]$

$x = \phi(x)$ type

$\phi(x_k) = e^{-x_k}$

(condition for unique root & convergence $|\phi'(x)| \leq m < 1$

$\phi'(x) = -e^{-x}$

$|-e^{-x}| = 1$ ← contradiction

∵ This is only sufficient condition, not a necessary condition

Contraction mapping

A function g is said to be a contraction mapping on an interval $[a, b]$ if

① $x \in [a, b] \Rightarrow g(x) \in [a, b]$

② g satisfies a lipschitz condition with lipschitz constant $L < 1$ Strictly less

Then $x \in [a, b] \Rightarrow |g(x) - g(y)| \leq L|x - y|$, $L < 1$

Condition ② is sometimes referred to as "closure condition"

We can now give less restrictive result than above

Theorem :

If there is an interval $[a, b]$ on which g is a contraction mapping, then —

i) the equation $x = g(x)$ has a unique root (say α) in $[a, b]$

ii) for any $x_0 \in [a, b]$, the sequence defined by

$$x_{k+1} = g(x_k), \quad k = 0, 1, 2 \ldots$$

converges to α.

Class 12, Jan 27

Proof:

For any $x \in [a, b]$, $g(x) \in [a, b]$. In particular,

for $x = a$ $\quad a - g(a) \leq 0$ $\quad -(1)$

$x = b$ $\quad b - g(b) \geq 0$ $\quad -(2)$

The Lipschitz condition implies that g is continuous on $[a, b]$. Thus $x - g(x)$ is continuous on $[a, b]$ and from inequalities (1) & (2), we can say that $x - g(x) = 0$ has at least one real root in $[a, b]$

Uniqueness:

Assume that there is more than one root and α, β be $(a \leq \alpha, \beta \leq b)$, denote 2 distinct roots.

Therefore $\alpha = g(\alpha)$ $\quad \beta = g(\beta)$

$$\alpha - \beta = g(\alpha) - g(\beta)$$

From the contraction mapping, we deduce that

$$|\alpha - \beta| = |g(\alpha) - g(\beta)| \leq L|\alpha - \beta|$$

This gives $L = 1$, which is a contradiction $(L < 1)$ $\because L$ should be less than 1

So $\underline{\alpha = \beta}$ and there is only a unique root.

Convergence :

Note that, since g is a contraction mapping on $[a,b]$

$x_k \in [a,b] \Rightarrow g(x_k) \in [a,b]$, which means that $x_{k+1} \in [a,b]$

By induction, every $x_k \in [a,b]$ if $x_0 \in [a,b]$
From lipschitz condition, it follows that

$$|x_{k+1} - \alpha| = |g(x_k) - g(\alpha)| \leq L|x_k - \alpha|$$
$$\leq L^2 |x_{k-1} - \alpha|$$
$$\leq L^3 |x_{k-2} - \alpha|$$
$$\vdots$$
$$\leq L^{k+1} |x_0 - \alpha|$$

Since $0 \in L < 1$, we conclude that sequence $\{x_k\}$ converges

Newton Raphson Method for Repeated Roots

Let 'ξ' be a root of multiplicity 'm' of $f(x) = 0$
i.e. $f(x) = (x-\xi)^m p(x)$, where $p(\xi) \neq 0$

$$f'(x) = m(x-\xi)^{m-1} p(x) + (x-\xi)^m p'(x)$$

$$\therefore \frac{f(x)}{f'(x)} = \frac{(x-\xi)^m p(x)}{m(x-\xi)^{m-1} p(x) + (x-\xi)^m p'(x)}$$

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

NR - Method.

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

compare $x_{n+1} = g(x)$

$g(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}$

$$= x - \frac{(x-\xi)\, p(x)}{m\, p(x) + (x-\xi)\, p'(x)}$$

$$g'(x) = 1 - \frac{d}{dx}\left[\frac{(x-\xi)\, p(x)}{m\, p(x) + (x-\xi)\, p'(x)}\right]$$

On simplification, it gives

Check it.

$$g'(x) = 1 - \frac{m\, p^2(x) + (x-\xi)^2\left[p'(x^2) - (x-\xi)^2 p(x)\, p''(x)\right]}{m^2 p^2(x)\left[1 + \frac{(x-\xi)\, p'(x)}{m\, p(x)}\right]^2}$$

Divide numerator and denominator by $m^2 p^2(x)$.
On simplification, we get

$$g'(x) = \frac{\left(1 - \frac{1}{m}\right) + \frac{2(x-\xi)\, p'(x)}{m\, p(x)} + \frac{(x-\xi)^2\, p''(x)}{m^2 p(x)}}{\left[1 + \frac{(x-\xi)\, p'(x)}{m\, p(x)}\right]^2} \quad \text{— (*)}$$

For the values of x sufficiently close to 'ξ'. $|g'(x)| < 1$ so that method converges. Since $g'(\xi) \neq 0$, it is no longer a second order method.

$\left[\because g'(\xi) = 0 \text{ and } g''(\xi) \neq 0 \text{ for second order}\right]$

Hence, it is now a first order method.

It is second order only when $m=1$ (i.e. simple root)
For $m>1$, the method converges like a first order method and loses its efficiency.
However, if we know or guess the multiplicity m of the root, it is possible to modify N-R method such that the modified method has second order convergence

$$f(x) = (x-\xi)^m \, p(x), \quad m>1, \quad p(\xi) \neq 0$$

consider $x_{k+1} = x_k - m \dfrac{f(x_k)}{f'(x_k)}$

So we have $g(x) = x - \dfrac{m f(x)}{f'(x)}$

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

$$= 1 - m + m \frac{f f''}{(f')^2}$$ → substitute from *

$$= 1 - m + m \left[\frac{\left(1-\frac{1}{m}\right) + \frac{2(x-\xi)p'(x)}{m\,p(x)} + \frac{(x-\xi)^2 p''(x)}{m^2 p(x)}}{\left[1 + \frac{(x-\xi)p'(x)}{m\,p(x)}\right]^2} \right]$$

$g'(\xi) = 0$ and $g''(\xi) \neq 0$
This shows method is second order

Problems
(1) We do not know multiplicity
(2) We do not know root is a multiple root.

$f(x) = 0$ has root 'ξ' of multiplicity m. $f'(x) = 0$ has the same root 'ξ' of g with multiplicity $m-1$

Hence let $g(x) = \frac{f(x)}{f'(x)}$ has a simple root

We can now apply N-R on $g(x)$

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}$$

to find approximate value of multiple root ξ

Simplifying we have

$$x_{n+1} = x_n - \frac{f_n f'_n}{f'^2_n - f_n f''_n}$$

Secant method can also be generalised to find a multiple root. We apply secant method for $g(x) = \frac{f(x)}{f'(x)}$, which has a single root.

In this case, secant method becomes -

$$x_{k+1} = \frac{x_{k-1} g_k - x_k g_{k-1}}{g_k - g_{k-1}} = \frac{x_{k-1} f_k f'_{k-1} - x_k f_{k-1} f'_k}{f_k f'_{k-1} - f_{k-1} f'_k}$$

Jan 31, Class 12

Iteration methods based on second degree approx.

Chebychev's Method

$$f(x) = 0$$

$$f(x) = a_0 x^2 + a_1 x + a_2 = 0$$

where a_0, a_1 and a_2 are three arbitrary constants to be determined by prescribing three appropriate conditions on $f(x)$ or its derivatives

$$f_k = a_0 x_k^2 + a_1 x_k + a_2$$

$$f'_k = 2a_0 x_k + a_1$$

$$f''_k = 2a_0$$

$$a_0 = \frac{1}{2} f''_k$$

$$a_1 = f'_k - 2a_0 x_k$$

$$= f'_k - f''_k x_k$$

$$a_2 = f_k - a_0 x_k^2 - a_1 x_k$$

$$= f_k - \frac{1}{2} f''_k x_k^2 - (f'_k - f''_k x_k) x_k$$

Substituting a_0 a_1 a_2 in (1), we get

$$\underbrace{\frac{1}{2} f''_k}_{a_0} x^2 + \underbrace{(f'_k - f''_k x_k)}_{a_1} x + \underbrace{f_k - \frac{1}{2} f''_k x_k^2 - (f'_k - f''_k x_k) x_k}_{a_2} = 0$$

$$f_k + (x - x_k) f'_k + \left(\frac{1}{2} x^2 - x x_k - \frac{1}{2} x_k^2 + x_k^2\right) f''_k = 0$$

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

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

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

$$(x - x_n) = -\frac{f_n}{f_n'} - \frac{1}{2}(x - x_n)^2 \frac{f_n''}{f_n'}$$

$$x = x_n - \frac{f_n}{f_n'} - \frac{1}{2}(x - x_n)^2 \frac{f_n''}{f_n'}$$

$\therefore$ Iteration method is

$$x_{n+1} = x_n - \frac{f_n}{f_n'} - \frac{1}{2}(x_{n+1} - x_n)^2 \frac{f_n''}{f_n'}$$

replace $x_{n+1} - x_n \longrightarrow -\frac{f_n}{f_n'}$ (by Newton Raphson)

$$x_{n+1} = x_n - \frac{f_n}{f_n'} - \frac{1}{2}\left(-\frac{f_n}{f_n'}\right)^2 \frac{f_n''}{f_n'}$$

$$\boxed{x_{n+1} = x_n - \frac{f_n}{f_n'} - \frac{1}{2}\frac{f_n^2 f_n''}{(f_n')^3}}$$

This is called Chebychev Method of order '3'
We can derive higher order Chebychev methods

Polynomial Equations.

To determine a real number 'p' such that $(x - p)$ is a factor of polynomial equation

Birge - Vieta Method.

This is same as Chebychev's method where $f(x_0)$, $f'(x_0)$, $f''(x_0)$ ... are determined by successive synthetic

division of $f(x)$ by $(x - x_0)$

Let $f(x) = a_0 x^n + a_1 x^{n-1} \ldots + a_{n-1} x + a_n$

Synthetic division of $f(x)$ by $(x - p)$ where $f(x)$ is a polynomial of degree n

p	a_0	a_1	a_2	$\ldots$	a_{n-3}	a_{n-2}	a_{n-1}	a_n
		pb_0	pb_1	$\ldots$	pb_{n-4}	pb_{n-3}	pb_{n-2}	pb_{n-1}
	b_0	b_1	b_2	$\ldots$				$b_n \Rightarrow R = f(p)$ (Remainder)
		pc_0	pc_1	$\ldots$			pc_{n-2}	pc_{n-1}
	c_0	c_1	c_2	$\ldots$		c_{n-2}	c_{n-1}	$= f'(p)$
		pd_0	pd_1	$\ldots$		pd_{n-3}		
	d_0	d_1	d_2	$\ldots$	d_{n-3}	d_{n-2}		$= \frac{1}{2!} f''(p)$

If the divisions of successive quotient polynomial by $(x-p)$ are continued, the remainders in successive divisions respectively represent $-\frac{1}{3!} f'''(p)$, $\frac{1}{4!} f''''(p)$

Let $f(x) = a_0 x^n + a_1 x^n \ldots + a_{n-1} x + a_n = 0$ be the equation and x_0 be an initial approximation to the root

N-R method $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$

In particular $x_{n+1} = x_0 - \frac{f(x_0)}{f'(x_0)}$

If b_n, c_{n-1}, d_{n-2} are the remainders in the successive divisions of $f(x)$ by $(x - x_0)$, then we see that

$$f(x_0) = b_n \qquad f'(x_0) = c_{n-1} \qquad \frac{f''(x_0)}{2!} = d_{n-2}$$

NR = second order (Birge-Vieta method)

Third order Birge-Vieta method

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} - \frac{(f(x_0))^2 f''(x_0)}{2! (f'(x_0))^3}$$

$$\therefore \quad x_1 = x_0 - \frac{b_n}{c_{n-1}} - \frac{b_n^2 d_{n-2}}{c_{n-1}^3}$$

Similarly the successive approximations $x_2, x_3 \ldots$ can be calculated.

Q: Use third order Birge-Vieta method to find the real root of $x^3 - x^2 - x - 1 = 0$ near 2

$x_0 = 2$

Birge-Vieta method $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} - \frac{f_n^2 f_n''}{2(f_n')^3}$

or

$$x_{n+1} = x_n - \frac{b_n}{c_{n-1}} - \frac{b_n^2 d_{n-2}}{c_{n-1}^3}$$

If b_n, c_{n-1}, d_{n-2} are the remainders in the successive synthetic division of $f(x) = x^3 - x^2 - x - 1$ by

$(x - x_0)$, then third order Birge - Vieta method is given by-

$$x_1 = x_0 - \frac{b_n}{c_{n-1}} - \frac{b_n^2 d_{n-2}}{c_{n-1}^3}$$

Illy x_2, x_3 -- can be determined.

$x_0 = 2$

2	1	−1	−1	−1
		2	2	2
	1	1	1	1 = b_n
2		2	6	
	1	3	7 = c_{n-1}	
		2		
	1	5 = d_{n-2}		

Improved value $x_1 = 1.8426$

1.8426	1	−1	−1	−1
		1.8426	1.5526	1.0182
	1	.8426	.5526	.0182 = b_n
		1.8426	4.9477	
	1	2.6852	5.603 = c_{n-1}	
		1.8426		
	1	4.5278 = d_{n-2}		

$$x_2 = 1.8426 - \frac{.0182}{5.5003} - \frac{(.0182)^2 \times 4.5278}{(5.5003)^3}$$

$$= 1.8393$$

Do one more iteration, it will be found that $x_3 = 1.8394$ → correct to fourth decimal.

Solution of System of Non-Linear equations by Newton Raphson method.

$$f(x,y) = 0$$
$$g(x,y) = 0$$

(x_0, y_0) is initial guess.

Assume that (α, β) is the exact solution.

Let $\alpha = x_0 + h$

$\beta = y_0 + k$

Since $f(\alpha, \beta) = 0$, $g(\alpha, \beta) = 0$

$f(x_0+h, y_0+k) = 0$ $\quad g(x_0+h, y_0+k) = 0$

Taylor series about (x_0, y_0)

$$f(x_0, y_0) + \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right) f\Big|_{(x_0,y_0)} + \frac{1}{2!}\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^2 f\Big|_{(x_0,y_0)} + \cdots$$

$$g(x_0, y_0) + \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right) g\Big|_{(x_0,y_0)} + \frac{1}{2!}\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^2 g\Big|_{(x_0,y_0)} + \cdots$$

Neglecting terms beyond second order derivative.

$$f(x_0, y_0) + h\left(\frac{\partial f}{\partial x}\right)_{(x_0,y_0)} + k\left(\frac{\partial f}{\partial y}\right)_{(x_0,y_0)} = 0$$

$$f_0 + h(f_x)_0 + k(f_y)_0 = 0 \quad \text{— (1)}$$

Ily $$g_0 + h(g_x)_0 + k(g_y)_0 = 0 \quad \text{— (2)}$$

Solving, we get

Feb 1, Class 13

$$\frac{h}{(f_y)_0 g_0 - (g_y)_0 f_0} = \frac{k}{f_0 (g_x)_0 - g_0 (f_x)_0} = \frac{1}{(f_x)_0 (g_y)_0 - (g_x)_0 (f_y)_0}$$

Jacobian $= \begin{vmatrix} f_x & f_y \\ g_x & g_y \end{vmatrix}_{(x_0, y_0)}$

$$h = \frac{(f_y)_0 g_0 - (g_y)_0 f_0}{(f_x)_0 (g_y)_0 - (g_x)_0 (f_y)_0}$$

$$k = \frac{f_0 (g_x)_0 - g_0 (f_x)_0}{(f_x)_0 (g_y)_0 - (g_x)_0 (f_y)_0}$$

Next approximation to (α, β) is given by (x_1, y_1).

$$x_1 = x_0 + h \; ; \; y_1 = y_0 + k$$

Repeat the procedure with (x_1, y_1) now and compute (h, k) and we get (x_2, y_2) $\quad x_2 = x_1 + h, \; y_2 = y_2 + k$.

Repeat till desired accuracy is reached.

i.e. till $|x_{k+1} - x_k| < \varepsilon$ and $|y_{k+1} - y_k| < \eta$ for some ε, η.

Equations (1) and (2) can also be written as

$$\begin{bmatrix} (f_x)_0 & (f_y)_0 \\ (g_x)_0 & (g_y)_0 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix} = - \begin{bmatrix} f_0 \\ g_0 \end{bmatrix}$$

or

$$\begin{bmatrix} h \\ k \end{bmatrix} = -\begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}^{-1}_{at\,(x_0, y_0)} \begin{bmatrix} f \\ g \end{bmatrix}_{at\,(x_0, y_0)}$$

$$= -J_0^{-1} \begin{bmatrix} f(x_0, y_0) \\ g(x_0, y_0) \end{bmatrix}$$

where $J = \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}$ and $J_0 = \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{at\,(x_0, y_0)}$

↓ Jacobian matrix of f and g.

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_0 + h \\ y_0 + k \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} h \\ k \end{bmatrix}$$

$$= \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} - J_0^{-1} \begin{bmatrix} f(x_0, y_0) \\ g(x_0, y_0) \end{bmatrix}$$

or

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} - J_0^{-1} F(x_0, y_0), \text{ where } F = \begin{bmatrix} f \\ g \end{bmatrix}$$

↓

$$X^{(1)} = X^{(0)} - J_0^{-1} F(X^{(0)}), \text{ where } X^{(0)} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

Thus we can write Newton Raphson method for the system

$$\boxed{X^{(k+1)} = X^{(k)} - J_k^{-1} F(X^{(k)})}$$

where $X^k = \begin{bmatrix} x^{(k)} \\ y^{(k)} \end{bmatrix}$, $F(X^{(k)}) = \begin{bmatrix} f(x_k, y_k) \\ g(x_k, y_k) \end{bmatrix}$

$$J_k^{-1} = \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(x^k, y^k)} = \frac{1}{f_x g_y - g_x f_y} \begin{bmatrix} g_y & -f_y \\ -g_x & f_x \end{bmatrix}_{(x^k, y^k)}$$

Newton Raphson method for complex roots

$f(z) = 0$ $\quad z$ is (x_0, y_0)

$f(z) = u(x,y) + i\, v(x,y) = 0 + 0i$

$$\Rightarrow \quad \left.\begin{aligned} u(x,y) &= 0 \\ v(x,y) &= 0 \end{aligned}\right\} \rightarrow (*)$$

Thus, the problem of finding the roots of the single equation $f(z) = 0$ is analogous to determining the roots of two simultaneous equations (*)

We apply N-R method for the system.

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - J^{-1} \begin{bmatrix} u(x_k, y_k) \\ v(x_k, y_k) \end{bmatrix}$$

For n equations

$$\left.\begin{aligned} f_1(x_1, x_2 \ \ldots \ x_n) &= 0 \\ f_2(x_1, x_2 \ \ldots \ x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2 \ \ldots \ x_n) &= 0 \end{aligned}\right\} - (1)$$

in the n unknowns $x_1, x_2 \ldots x_n$.

Assume the partial derivatives of f_i $(i = 1, 2, \ldots n)$ are first order continuous.

Let $x = (x_1, x_2 \cdots x_n)$

$F = (f_1, f_2 \cdots f_n)^T$

(1) can be written as $F(X) = 0$. Assume $X^{(0)}$ is the initial approximation is known.

N-R for (1) is

$$X^{(k+1)} = X^{(k)} - [J(X^{(k)})]^{-1} F(X^{(k)})$$

$$J(X^{(k)}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}_{(x_1^{(k)}, x_2^{(k)}, \ldots x_n^{(k)})}$$

Convergence in case of n equations solved using N-R method.

If $X^{(0)}$ is sufficiently close to 'ξ', and second order partial derivatives are continuous in the neighbourhood of ξ and $|J(\xi)| \neq 0$, then N-R method has Quadratic Convergence.

This means that there exists a constant c, such that

$$\| x^{(k+1)} - \xi \|_2 \simeq c \| x^{(k)} - \xi \|_2^2$$

→ Euclidean/quadratic norm.

Only Newton raphson and fixed point method are extendable to n equations. That's why they are so popular and effective

One-point iteration method for the non-linear system:

$$F(X) = 0 \Rightarrow \begin{aligned} f_1(x_1, x_2 \dots x_n) &= 0 \\ f_2(x_1, x_2 \dots x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2 \dots x_n) &= 0 \end{aligned}$$

$$\Downarrow$$

$$\begin{aligned} x_1 &= \phi_1(x_1, x_2 \dots x_n) \\ x_2 &= \phi_2(x_1, x_2 \dots x_n) \\ &\vdots \\ x_n &= \phi_n(x_1, x_2 \dots x_n) \end{aligned}$$

$$X = \underline{\Phi}(X) = [\phi_1(X) \quad \phi_2(X) \dots \phi_n(X)]^T$$

Let $X^{(0)}$ be initial approximation. Let the sequence of approximations $X^{(1)}, X^{(2)} \dots$ is computed by

$$X^{(k+1)} = \underline{\Phi}(X^{(k)})$$

The convergence criteria can be generalised

Assume $\xi = \underline{\Phi}(\xi)$ and the partial derivatives

$$d_{ij} = \frac{\partial \Phi_i}{\partial x_j}(X) \qquad i,j = 1, 2, \ldots n.$$

exist for $X \in I = \{X : \| X - \xi \| < \delta\}$

Let $D(X)$ be a $n \times n$ matrix with elements $d_{ij}(X)$

A sufficient condition for the fixed point iteration to converge for any $X^{(0)} \in I$ is that

$$\| D(X) \| \leq m < 1, \quad X \in I \quad \text{where}$$

$\| \cdot \| \rightarrow$ induced matrix norm.

If the condition is satisfied, we have linear convergence

$$\| X^{(k+1)} - \xi \| \leq m \| X^{(k)} - \xi \|$$

Two easily computable p- norms are -

Maximum row sum norm = Row norm = $\max_{1 \leq i \leq n} \left\{ \sum_{j=1}^{n} |a_{ij}| \right.$

or

Matrix infinity norm. $= \| A \|_\infty$

Maximum column sum norm = Column norm = $\max_{1 \leq j \leq n} \left| \sum_{i=1}^{n} |a_{ij}| \right|$

$= \| A \|_1$

Contraction mapping

[For System of non linear equations]

A mapping 'g' which maps a real n-dimensional vectors into real n-dimensional vectors is said to be a contraction mapping with respect to a norm $\|\cdot\|$ on a closed region R if

(i) $X \in R \Rightarrow g(X) \in R$

(ii) $\|g(X) - g(Y)\| \leq L \|X - Y\|$ with $0 \leq L < 1$ $\forall X, Y \in \mathbb{R}^n$

We have extended the closure and lipschitz conditions to n-variables and then extended the notion of contraction mapping.

Class 14, Feb 3.

Def:

Closed Region.

A region R in n-dim space is said to be closed if every convergent sequence $\{X_n\}$ with each $X_n \in R$ such that its limit $\alpha \in R$.

Theorem: If there is a closed region R on which 'g' is a contraction mapping w.r.t. some norm, $\|\cdot\|$, Then

i) the equation $X = g(X)$ has unique solution (say α) belonging to R

ii) for any $X_0 \in R$, the sequence $\{X_k\}$ defined by

$$X_{k+1} = g(X_k) \quad k = 0, 1, 2.$$

converges to α.

In contraction mapping, differentiability is not required, but is req. in fixed-point.

Q: The methods for finding $\sqrt{a}$ are given by -

$$x_{n+1} = \frac{x_n}{2}\left(1 + \frac{a}{x_n^2}\right) \quad \text{and} \quad x_{n+1} = \frac{x_n}{2}\left(3 - \frac{x_n^2}{a}\right)$$

Find order of convergence of above methods. Hence show that

$$x_{n+1} = \frac{1}{8} x_n \left(6 + \frac{3a}{x_n^2} - \frac{x_n^2}{a}\right) \text{ gives}$$

a sequence with third order convergence

Assume method converges

$$\lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} \frac{x_n}{2}\left(1 + \frac{a}{x_n^2}\right)$$

$$x = \frac{x}{2}\left(\frac{x^2 + a}{x^2}\right)$$

$$2x^2 = x^2 + a$$

$$x^2 = a$$

Regula-falsi always converges. → super-linear

Secant - may / may not

Sol $x_{n+1} = \frac{x_n}{2}\left(1 + \frac{a}{x_n^2}\right)$

Let $\varepsilon_n = x_n - \xi \Rightarrow x_n = \varepsilon_n + \xi$

$\varepsilon_{n+1} = x_{n+1} - \xi \Rightarrow x_{n+1} = \varepsilon_{n+1} + \xi$ ξ is root $\xi^2 = a$

Iteration scheme,

$$(\varepsilon_{n+1} + \xi) = \frac{(\varepsilon_n + \xi)}{2}\left(1 + \frac{a}{(\varepsilon_n + \xi)^2}\right)$$

$$= \frac{(\varepsilon_n + \xi)}{2}\left[1 + \frac{1}{\left(1 + \frac{\varepsilon_n}{\xi}\right)^2}\right]$$

$$= \frac{1}{2}(\varepsilon_n + \xi)\left[1 + \left[1 + \frac{\varepsilon_n}{\xi}\right]^{-2}\right]$$

Assume $\left|\frac{\varepsilon_n}{\xi}\right| < 1$

Taylor series,

$$= \frac{1}{2}(\varepsilon_n + \xi)\left[1 + 1 - 2\frac{\varepsilon_n}{\xi} + 2\frac{3}{2}\left(\frac{\varepsilon_n}{\xi}\right)^2 \cdots\right]$$

$$= \frac{1}{2}(\varepsilon_n + \xi)\left[2 - 2\frac{\varepsilon_n}{\xi} + 2\frac{3}{2}\left(\frac{\varepsilon_n}{\xi}\right)^2 - \cdots\right]$$

$$= (\varepsilon_n + \xi)\left[1 - \frac{\varepsilon_n}{\xi} + \frac{3}{2}\left(\frac{\varepsilon_n}{\xi}\right)^2 \cdots\right]$$

$$\varepsilon_{n+1} + \xi = (\varepsilon_n + \xi) - \frac{\varepsilon_n^2}{\xi} - \varepsilon_n + \frac{3}{2}\frac{\varepsilon_n^3}{\xi^2} + \frac{3}{2}\frac{\varepsilon_n^2}{\xi} \cdots$$

$$\boxed{\varepsilon_{n+1} = \frac{1}{2\xi}\varepsilon_n^2} + \underbrace{O\,\varepsilon_n^3 + \text{Higher order neglected.}}_{-(1)}$$

$\Rightarrow$ This is a second order method. with error constant $\frac{1}{2\xi}$

Now consider the iteration scheme.

$$x_{n+1} = \frac{x_n}{2}\left(3 - \frac{x_n^2}{a}\right)$$

$$\varepsilon_{n+1} + \xi = \frac{(\varepsilon_n + \xi)}{2}\left(3 - \frac{(\varepsilon_n+\xi)^2}{\xi^2}\right)$$

$$= \left(\frac{\varepsilon_n + \xi}{2}\right)\left(3 - \left(1 + \frac{\varepsilon_n}{\xi}\right)^2\right)$$

Assume $\left|\frac{\varepsilon_n}{\xi}\right| < 1$

$$\left(\frac{\varepsilon_n+\xi}{2}\right)\left(3 - \left(1 - \frac{\varepsilon_n}{\xi} + \left(\frac{\varepsilon_n}{\xi}\right)^2\right.\right.$$

$$\vdots$$

$$\boxed{\varepsilon_{n+1} = -\frac{3}{2}\frac{\varepsilon_n^2}{\xi} + O\,\varepsilon_n^3} \quad \text{— (2)}$$

The method has second order convergence with error constant $c = \frac{-3}{2\xi}$

Observe (1) and (2)

Error in (1) is about $\frac{1}{3}$rd of (2) in magnitude

$3\times$ (1) + (2) $\Rightarrow$ $\varepsilon_{n+1} = 0\,\varepsilon_n^3$

$$3x_{n+1} = \frac{3}{2}(x_n)\left(1 + \frac{a}{x_n^2}\right)$$

$$+ \quad x_{n+1} = \frac{1}{2}x_n\left(3 - \frac{x_n^2}{a}\right)$$

$$4x_{n+1} = x_n\left[\frac{3}{2} + \frac{3}{2}\frac{a}{x_n^2} + \frac{3}{2} - \frac{x_n^2}{2a}\right]$$

$$x_{n+1} = \frac{1}{4}x_n\left(3 + \frac{3}{2}\frac{a}{x_n^2} - \frac{x_n^2}{2a}\right]$$

$$x_{n+1} = \frac{1}{8}x_n\left[6 + \frac{3a}{x_n^2} - \frac{x_n^2}{a}\right]$$

Q: The iteration formula $x_{k+1} = \frac{1}{n}\left[(n-1)x_k + \frac{a}{x_{n-1}}\right]$ is used to find a certain quantity. Determine the quantity.

As $k \to \infty$ $\quad x_{k+1} = x_k = x$.

$$x = \frac{1}{n}\left[(n-1)x + \frac{a}{x_{n-1}}\right]$$

$$nx = nx - x + \frac{a}{x_{n-1}}$$

$$x = \frac{a}{x_{n-1}}$$

So we are trying to find root of $f(x) = 0$ where $f(x) = x - \frac{a}{x_{n-1}}$

MINOR-2

Interpolation

Suppose values of x and $f(x)$ are given in a tabular form.

x	x_0	x_1	-	-	x_n
$f(x)$	f_0	f_1			f_n

The form of the function may / may not be known. We want to find a polynomial of degree at most n which agrees with the values of $f(x)$ at $x = x_0, x_1 \ldots x_n$

In other words, we want to find a polynomial $P_n(x)$ such that

$$P_n(x_j) = f(x_j) \quad , \quad j = 0, 1, 2 \ldots$$

Then $P_n(x)$ is called INTERPOLATING POLYNOMIAL of $f(x)$.

Assume all x_i are different. x_i s are called NODAL POINTS.

Class 15, Feb 14

Forward Difference Operator 'Δ'

Let equally spaced points $x_0, x_1, x_2 \ldots x_n$

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

or

$$x_i = x_0 + ih \qquad i = 1, 2, 3 \ldots$$

First order forward difference of f at x_k is denoted by Δf_k and is defined as

$$\Delta f_k = f_{k+1} - f_k \qquad f_k = f(x_k)$$

$$\Delta f_k = \Delta f(x_k) = f(x_{k+1}) - f(x_k) = f(x_k + h) - f(x_k)$$

eg $\Delta f_0 = f_1 - f_0$

$\Delta f_1 = f_2 - f_1$

Projection : i) $\Delta(cu_n) = c\,\Delta u_n$

ii) $\Delta(u_m + u_n) = \Delta u_m + \Delta u_n$

iii) $\Delta^m(\Delta^n u_k) = \Delta^{m+n} u_k = \Delta^n(\Delta^m u_k)$

iv) $\Delta(u_m v_m) = \Delta u_m\, v_{m+1} + u_{m+1}\,\Delta v_m$

] → They imply Δ is a linear op.

All these can be verified from definitions

Second order forward difference 'Δ^2' $\Delta^2 f_k$

Defined as $\Delta(\Delta f_k)$

$= \Delta f_{k+1} - \Delta f_k$

$= f_{k+2} - f_{k+1} - (f_{k+1} - f_k)$

$= f_{k+2} - 2f_{k+1} + f_k$

eg $\Delta^2 f_3 = \Delta(\Delta f_3)$

$= \Delta(f_4 - f_3) = \Delta f_4 - \Delta f_3$

$= f_5 - f_4 - (f_4 - f_3)$

$= f_5 - 2f_4 + f_3$

Third order forward difference Δ^3

$\Delta^3 f_k = \Delta(\Delta^2 f_k)$

$= \Delta(f_{k+1} - 2f_{k+1} + f_k)$

$= f_{k+3} - 3f_{k+2} + 3f_{k+1} - f_k$

Similarly, Higher order forward differences can be defined

We can see that Δ^n has coefficients of expansion of $(1-x)^n$

eg. $\Delta^4 f_k = f_{k+4} - 4f_{k+3} + 6f_{k+2} - 4f_{k+1} + f_k$

$$\boxed{\Delta^m f_k = \sum_{r=0}^{m} (-1)^r \, {}^mC_r \, f_{k+m-r}}$$

Coeff. are coeff. in binomial expansion of $(1-x)^m$.

Shift operator 'E'

$$E(f_k) = f_{k+1}$$
$$E^2(f_k) = E(f_{k+1}) = f_{k+2}$$
$$\vdots$$

$$\boxed{E^r(f_k) = f_{k+r}}$$

Relation btw E & Δ $\rightarrow$ $\boxed{E = 1 + \Delta}$

1 is the identity operator.

Proof: $\Delta f_n = f_{n+1} - f_n = E(f_n) - f_n$

$\Delta = E - 1$ or $E = 1 + \Delta$

All these are linear operators.

$$\Delta^2 f_n = (E-1)^2 f_n$$

$$= (E^2 - 2E + 1) f_n$$

$$= f_{n+2} - 2f_{n+1} + f_n$$

Illy $\Delta^3 f_n = (E-1)^3 f_n$

Forward Difference Table.

x_k	f_k	Δf	$\Delta^2 f$	$\Delta^3 f$	$\Delta^4 f$
x_0	f_0	Δf_0	$\Delta^2 f_0$	$\Delta^3 f_0$	$\Delta^4 f_0$
x_1	f_1	Δf_1	$\Delta^2 f_1$	$\Delta^3 f_1$	
x_2	f_2	Δf_2	$\Delta^2 f_2$		
x_3	f_3	Δf_3			
x_4	f_4				

Backward Difference Operator '∇'

The first order backward difference of f at x_k is denoted by ∇f_k, is defined as

$$\nabla f_k = f_k - f_{k-1}$$

eg $\nabla f_1 = f_1 - f_0$

$\nabla f_2 = f_2 - f_1$

Second order $\nabla^2 f_k = \nabla(\nabla f_k) = \nabla(f_k - f_{k-1})$

$$= \nabla f_k - \nabla f_{k-1}$$

$$= (f_k - f_{k-1}) - (f_{k-1} - f_{k-2})$$

$$= f_k - 2f_{k-1} + f_{k-2}$$

Third order

$$\nabla^3 f_x = \nabla^2 f_x - \nabla^2 f_{x-1}$$

$$= (f_x - 2f_{x-1} + f_{x-2}) - (f_{x-1} - 2f_{x-2} + f_{x-3})$$

$$= f_x - 3f_{x-1} + 3f_{x-2} - f_{x-3}$$

$$\nabla f_x = f_x - f_{x-1}$$
$$\nabla^2 f_x = f_x - 2f_{x-1} + f_{x-2}$$
$$\nabla^3 f_x = f_x - 3f_{x-1} + 3f_{x-2} - f_{x-3}$$

Note that the coeff. in the expansion of $\nabla^m f_x$ are the coeff in expansion of $(1-x)^m$

$$\nabla^4 f_x = f_x - 4f_{x-1} + 6f_{x-2} - 4f_{x-3} + f_{x-4}$$

More generally

$$\nabla^m f_x = \sum_{r=0}^{m} (-1)^r \, {}^mC_r \, f_{x-r}$$

$$\Delta^m f_x = \nabla^m f_m$$

or

$$\nabla^m f_m = \Delta^m f_{m-m}$$

$$\boxed{\nabla = 1 - E^{-1}}$$

// -ve powers of E not defined for Δ

Relation btw ∇ and E

$$\nabla f_k = f_k - f_{k-1} = f_k - E^{-1} f_k$$

$$\nabla = 1 - E^{-1}$$

$$E f_k = f_{k+1}$$
$$E^{-1} f_k = f_{k-1}$$

$$E^{-2} f_k = f_{k-2}$$

$$\nabla^2 f_k = (1 - E^{-1})^2 = (1 - 2E^{-1} + E^{-2}) f_k$$

$$= f_k - 2 f_{k-1} + f_{k-2}$$

Backward difference table

x_k	f_k	∇f_k	$\nabla^2 f_k$	$\nabla^3 f_k$	$\nabla^4 f_k$
x_0	f_0				
		∇f_1			
x_1	f_1		$\nabla^2 f_2$		
		∇f_2		$\nabla^3 f_3$	
x_2	f_2		$\nabla^2 f_3$		$\nabla^4 f_4$
		∇f_3		$\nabla^3 f_4$	
x_3	f_3		$\nabla^2 f_4$		
		∇f_4			
x_4	f_4				

Feb 15, Class 16

Divided Differences

In this case the abscissas $x_0, x_1, x_2 \ldots$ (nodal points or nodes or arguments) need NOT be equally spaced

Suppose f_i is value of $f(x)$ at x_i.

The divided difference [DD] of zeroth order of f with argument x_k is denoted by $f[x_k]$ and is defined as $f[x_k] = f(x_k) = f_k$

The divided difference of f of first order with arguments x_0, x_1 is denoted by $f[x_0, x_1]$ and is defined as

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

Similarly $f[x_3, x_4] = \frac{f(x_4) - f(x_3)}{x_4 - x_3}$

The second order divided difference of f with arguments x_0, x_1, x_3 is denoted by $f[x_0, x_1, x_2]$ and is defined as

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$$

- The third order d.d. with arguments x_0, x_1, x_2, x_3 is denoted by $f[x_0, x_1, x_2, x_3]$ and is defined as

$$f[x_0, x_1, x_2, x_3] = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0}$$

Kth order

Similarly the Kth order d.d is defined as -

$$f[x_0, x_1, \ldots x_k] = \frac{f[x_1, x_2 \cdots x_k] - f[x_0, x_1, \ldots x_{k-1}]}{x_k - x_0}$$

Order of arguments does NOT matter

Symmetric form of divided difference

D.D. are symmetric w.r.t. the arguments

Consider $f[x_0, x_1] = \frac{f_1 - f_0}{x_1 - x_0} = \frac{f_1}{x_1 - x_0} - \frac{f_0}{x_1 - x_0}$

$$= \frac{f_1}{x_1 - x_0} + \frac{f_0}{x_0 - x_1}$$

$$f[x_1, x_0] = \frac{f_0 - f_1}{x_0 - x_1} = \frac{f_0}{x_0 - x_1} - \frac{f_1}{x_0 - x_1}$$

$$= \frac{f_0}{x_0 - x_1} + \frac{f_1}{x_1 - x_0}$$

D.D. remain unchanged if 2 arguments are interchanged.

$$f[x_0, x_1, x_2] = f[x_1, x_0, x_2] = f[x_2, x_1, x_0]$$

Lemma: $f[x_0, x_1, \ldots, x_n]$ is symmetric w.r.t. its arguments i.e. to show

$$f[x_0, x_1, \ldots, x_n] = \sum_{k=0}^{n} \frac{f(x_k)}{\prod_{\substack{j=0 \\ j \neq k}}^{n} (x_k - x_j)}$$

Proof: By induction.

$\Rightarrow$ D.D. are invariant to permutations of $x_0, x_1 \ldots x_n$

Newton's Divided Difference Interpolation formula.

We have $f[x_0, x] = \dfrac{f(x) - f(x_0)}{x - x_0}$

$\therefore \quad f(x) = f(x_0) + (x - x_0) f[x_0, x] \quad -(1)$

Consider $f[x_1, x_0, x] = \dfrac{f[x_0, x] - f[x_1, x_0]}{x - x_1}$

or

$$f[x_0, x_1, x] = \frac{f[x_0, x] - f[x_0, x_1]}{x - x_1}$$

$$f[x_0, x] = f[x_0, x_1] + (x - x_1) f[x_0, x_1, x]$$

Put this value in (1)

$$f(x) = f(x_0) + (x - x_0)\left[f[x_0, x_1] + (x - x_1) f[x_0, x_1, x]\right]$$

$$f(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x] \quad -(2)$$

Consider

$$f[x_2, x_1, x_0, x] = \frac{f[x_1, x_0, x] - f[x_2, x_1, x_0]}{x - x_2}$$

(dd are symm.)

$$\| \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$f[x_0, x_1, x_2, x] = \frac{f[x_0, x_1, x] - f[x_0, x_1, x_2]}{x - x_2}$$

$$f[x_0, x_1, x] = f[x_0, x_1 x_2] + (x - x_2) f[x_0, x_1, x_2, x]$$

Put this in (2)

$$f(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1 x_2] + (x - x_0)(x - x_1)(x - x_2) f[x_0, x_1, x_2, x]$$

Similarly, if (n+1) nodal points or arguments are given $x_0, x_1, x_2 \ldots x_n$ where values of f are known

$$f(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2] + \ldots + (x - x_0)(x - x_1) \ldots (x - x_{n-1}) f[x_0, x_1 \ldots, x_n] + R$$

$$R = (x - x_0) \ldots (x - x_n) f[x_0, x_1 \ldots x_n, x]$$

If we ignore R, we get

$$f(x) \simeq P_n(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2] + \ldots + (x - x_0) \ldots (x - x_{n-1}) f[x_0, x_1 \ldots, x_n] \quad -(3)$$

which is n degree polynomial in x.

(3) is called Newton's Fundamental Interpolation formula. or Newton's General / Newton's D.D. Interpolation formula.

$$f(x_i) = P_n(x_i) \qquad i = 0, 1, 2 \ldots$$

Divided Difference Table.

x_k	$f(x_k)$	d.d.	d.d 2nd	d.d 3rd	d.d with
x_0	f_0	$f[x_0 x_1]$	$f[x_0, x_1 x_2]$	$f[x_0 x_1 x_2 x_3]$	$f[x_0 x_1 x_2 x_3 x_4]$
x_1	f_1	$f[x_1 x_2]$	$f[x_1, x_2 x_3]$	$f[x_1 x_2 x_3 x_4]$	
x_2	f_2	$f[x_2, x_3]$	$f[x_2 x_3 x_4]$		
x_3	f_3	$f[x_3, x_4$			
x_4	f_4				

Feb 17, Class 17

Q Given values of x and $f(x)$ as shown in the following table, use Newton's D.D. formula to obtain the interpolating polynomial of $f(x)$

x	0	2	3	5	6
$f(x)$	1	9	28	126	217

Sol

x	$f(x)$	Δ	Δ^2	Δ^3	Δ^4
0	1				
		4			
2	9		5		
		19		1	
3	28		10		0
		49		1	
5	126		14		
		91			
6	217				

$\frac{19-4}{3-0}$, $\frac{49-19}{5-2}$, $\frac{91-49}{6-3}$

$\frac{10-5}{5-0}$, $\frac{14-10}{6-2}$

The formula is

$$f(x) = f(x_0) + (x-x_0) f[x_0, x_1] + (x-x_0)(x-x_1) f[x_0, x_1, x_2] + (x-x_0)(x-x_1)(x-x_2) f[x_0, x_1, x_2, x_3]$$

Substituting these values,

$$f(x) = 1 + (x-0)\,4 + (x-0)(x-2)\,5 + (x-0)(x-2)(x-3)\,1 + 0$$

$$= 1 + 4x + 5(x^2 - 2x) + (x^3 - 5x^2 + 6x)$$

$$= \underline{x^3 + 1}$$

Newton's Forward Difference Formula from general interpolation formula –

i.e. for equally spaced points

$$f(x) \simeq P_n(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2]$$
$$+ \ldots \ldots + (x - x_0) \ldots (x - x_{n-1}) f[x_0, x_1 \ldots, x_n] \quad -(1)$$

If abscissas are equally spaced then $x_k - x_{k-1} = h$

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f_1 - f_0}{h} = \frac{\Delta f_0}{h}$$

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = \frac{\frac{\Delta f_1}{h} - \frac{\Delta f_0}{h}}{2h}$$

$$= \frac{\Delta(f_1 - f_0)}{2h^2} = \frac{\Delta^2 f_0}{2!\, h^2}$$

$$f[x_0, x_1, x_2, x_3] = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0}$$

$$= \frac{\frac{\Delta^2 f_1}{2h^2} - \frac{\Delta^2 f_0}{2h^2}}{3h} = \frac{\Delta^2(f_1 - f_0)}{2!\, 3\, h^3}$$

$$= \frac{\Delta^3 f_0}{3!\, h^3}$$

etc.

So ① takes the form

$$f(x) \simeq P_n(x) = f_0 + (x-x_0)\frac{\Delta f_0}{h} + (x-x_0)(x-x_1)\frac{\Delta^2 f_0}{2!\,h^2} + (x-x_0)(x-x_1)(x-x_2)\frac{\Delta^3 f_0}{3!\,h^3} + \cdots + \cdots + (x-x_0)(x-x_1)\cdots(x-x_{n-1})\frac{\Delta^n f_0}{n!\,h^n} \quad (*)$$

which is called Newton's Forward Difference interpolating formula for $f(x)$

If value of $f(x)$ is required at a non tabular value of $f(x)$, we proceed as follows:

In (*), put $\frac{x-x_0}{h} = s$, then

$$\frac{x-x_i}{h} = s-i$$

$$\left[\frac{x-x_i}{h} = \frac{(x-x_0)-(x_i-x_0)}{h} = \frac{sh-ih}{h} = s-i\right]$$

then (*) takes the form

$$f_s = f_0 + s\Delta f_0 + \frac{s(s-1)}{2!}\Delta^2 f_0 + \frac{s(s-1)(s-2)}{3!}\Delta^3 f_0 + \cdots + \frac{s(s-1)\cdots(s-(n-1))}{n!}\Delta^n f_0$$

For $k > 0$, $f[x_0, x_1, \ldots x_n] = \frac{1}{k!\, h^k} \Delta^k f_0$ — (1)

This is the relation between divided difference and forward difference when the nodal points are equally spaced

Proof:

For $k = 0$, $f[x_0] = \frac{1}{0!\, h^0} \Delta^0 f_0 = f_0$

The result is trivially true

for $k = 1$, $f[x_0, x_1] = \frac{f_1 - f_0}{x_1 - x_0} = \frac{1}{h} \Delta f_0$

Assume that the result is true for $k \leq r$, then for $k = r+1$,

$$f[x_0, x_1 \ldots x_r, x_{r+1}] = \frac{f[x_1, x_2 \ldots x_{r+1}] - f[x_0, x_1 \ldots x_r]}{x_{r+1} - x_0}$$

$$= \frac{\frac{1}{r!\, h^r} \Delta^r f_1 - \frac{1}{r!\, h^r} \Delta^r f_0}{(r+1)\, h}$$

$$= \frac{\Delta^r (f_1 - f_0)}{r!\, (r+1)\, h^r\, h}$$

$$= \frac{\Delta^{r+1} f_0}{(r+1)!\, h^{r+1}}$$

Similarly we can write for backward difference

Lemma : For $k \geq 0$

$$f[x_n, \ldots, x_k] = \frac{1}{(n-k)!\, h^{n-k}} \nabla^{n-k} f_n \quad -(2)$$

Proof : By induction (H.W.)

//

From (1) and (2) we can also write

$$f[x_0, \ldots x_n] = \frac{\Delta^n f_0}{n!\, h^n}$$

and $$f[x_n, \ldots x_0] = \frac{\nabla^n f_n}{n!\, h^n}$$

[**Proof :** H.W.]

Derivation of Newton's Backward Difference Interpolation Formula :

We introduce the abscissas in the order

$x_n, x_{n-1}, \ldots x_1, x_0 \rightarrow$ i.e. in reverse order, the Newton's General interpolation formula can be written as

$$f(x) \simeq P_n(x) = f(x_n) + (x - x_n) f[x_n, x_{n-1}] + (x - x_n)(x - x_{n-1}) f[x_n, x_{n-1}, x_{n-2}] + \ldots + (x - x_n) \ldots (x - x_1) f[x_n, \ldots x_1, x_0] \quad -(1)$$

$$f[x_n, x_{n-1}] = \frac{f_{n-1} - f_n}{x_{n-1} - x_n} = \frac{f_n - f_{n-1}}{x_n - x_{n-1}}$$

$$= \frac{\nabla f_n}{h}$$

$$f[x_n, x_{n-1}, x_{n-2}] = \frac{f[x_n, x_{n-1}] - f[x_{n-1}, x_{n-2}]}{x_n - x_{n-2}}$$

$$= \frac{\frac{\nabla f_n}{h} - \frac{\nabla f_{n-1}}{h}}{2h} = \frac{\nabla(f_n - f_{n-1})}{2h^2}$$

$$= \frac{\nabla^2 f_n}{2!\, h^2}$$

Similarly, we can show

$$f[x_n, x_{n-1}, x_{n-3}] = \frac{\nabla^3 f_n}{3!\, h^3}$$

$$f[x_n, x_{n-1}, \ldots x_1, x_0] = \frac{\nabla^3 f_n}{n!\, h^n}$$

[can be proved by induction as done in prev. lemma]

① can be written as

$$f(x) = P_n(x) = f(x_n) + (x - x_n)\frac{\nabla f_n}{h} + (x - x_n)(x - x_{n-1})\frac{\nabla^2 f_n}{2!\, h^2}$$

$$+ \ldots + (x - x_n)\ldots(x - x_1)\frac{\nabla^n f_n}{n!\, h^n}$$

— (*)

This is called Newton's Backward Difference interpolating polynomial

Put $\frac{x - x_n}{h} = s$, then $\frac{x_n - x_{n-i}}{h} = s + i$

$$\left[\frac{x_n - x_{n-i}}{h} = \frac{(x - x_n) + (x_n - x_{n-i})}{h} = \frac{sh + ih}{h} = s + i \right]$$

$\therefore$ (*) becomes

$$f(x) \simeq P_n(x) = F_s = f_n + s \nabla f_n + \frac{s(s+1)}{2!} \nabla^2 f_n$$

$$+ \frac{s(s+1)(s+2)}{3!} \nabla^3 f_n + \cdots \cdots +$$

$$\frac{s(s+1) \cdots (s+(n-1))}{n!} \nabla^n f_n$$

Feb 21, Class 18

Central Difference Operator 'δ'

$x_0, x_1, x_2 \cdots x_n$ → abscissas | nodal points | nodes

These are equispaced, $x_i - x_{i-1} = h$

First order central difference of f at x_k is denoted by δf_k and is defined as

$$\delta f_k = f\left(x_k + \frac{h}{2}\right) - f\left(x_k - \frac{h}{2}\right)$$

$$= f_{k+\frac{1}{2}} - f_{k-\frac{1}{2}}$$

$$\left[f_{k+\frac{1}{2}} = f\left(x_k + \frac{h}{2}\right) \right]$$

Second order central difference $\delta^2 f_k$

$$\delta^2 f_k = \delta(\delta f_k) = \delta\left(f_{k+\frac{1}{2}} - f_{k-\frac{1}{2}}\right)$$

$$= \delta f_{k+\frac{1}{2}} - \delta f_{k-\frac{1}{2}}$$

$$= f_{k+1} - f_k - \left[f_k - f_{k-1} \right]$$

$$= f_{k+1} - 2f_k + f_{k-1}$$

Third order central difference $\delta^3 f_k$

$$\delta^2(\delta f_k) = \delta^2\left(f_{k+\frac{1}{2}} - f_{k-\frac{1}{2}}\right)$$

$$= \delta^2 f_{k+\frac{1}{2}} - \delta^2 f_{k-\frac{1}{2}}$$

$$= \left[f_{k+\frac{3}{2}} - 2f_{k+\frac{1}{2}} - f_{k-\frac{1}{2}} \right] - \left[f_{k+\frac{1}{2}} - 2f_{k-\frac{1}{2}} + f_{k-\frac{3}{2}} \right]$$

$$= f_{k+\frac{3}{2}} - 3f_{k+\frac{1}{2}} + 3f_{k-\frac{1}{2}} - f_{k-\frac{3}{2}}$$

// We see that

$$\delta f_k = f_{k+\frac{1}{2}} - f_{k-\frac{1}{2}}$$

$$\delta^2 f_k = f_{k+1} - 2f_k + f_{k-1}$$

$$\delta^3 f_k = f_{k+\frac{3}{2}} - 3f_{k+\frac{1}{2}} + 3f_{k-\frac{1}{2}} - f_{k-\frac{3}{2}}$$

$$\delta^4 f_k = f_{k+2} - 4f_{k+1} + 6f_k - 4f_{k-1} + f_{k-2}$$

The coefficients in the expansion of $(1-x)^2$

Relation between central difference operator 'δ' and shift operator 'E' :

$$\boxed{\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}}}$$

$$\delta f_k = f_{k+\frac{1}{2}} - f_{k-\frac{1}{2}} = E^{\frac{1}{2}} f_k - E^{-\frac{1}{2}} f_k$$

$$= (E^{\frac{1}{2}} - E^{-\frac{1}{2}}) f_k$$

$$\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}}$$

$$\delta^2 f_k = (E^{\frac{1}{2}} - E^{-\frac{1}{2}})^2 f_k = (E - 2 + E^{-1}) f_k$$

$$= f_{k+1} - 2f_k + f_{k-1}$$

Central differences at non nodal points $x_{k+\frac{1}{2}}$

$$\delta f_{k+\frac{1}{2}} = f_{k+\frac{1}{2}+\frac{1}{2}} - f_{k+\frac{1}{2}-\frac{1}{2}}$$

$$= f_{k+1} - f_k$$

$$\delta^2 f_{k+\frac{1}{2}} = f_{k+\frac{3}{2}} - 2f_{k+\frac{1}{2}} + f_{k-\frac{1}{2}}$$

$$\delta^3 f_{k+\frac{1}{2}} = f_{k+2} - 3f_{k+1} + 3f_k - f_{k-1}$$

$$\delta^4 f_{(k+\frac{1}{2})} = f_{k+\frac{5}{2}} - 4f_{k+\frac{3}{2}} + 6f_{k+\frac{1}{2}} - 4f_{k-\frac{1}{2}} + f_{k-\frac{3}{2}}$$

Central Difference Table

x	$f(x)$	δf	$\delta^2 f$	$\delta^3 f$	$\delta^4 f$	$\delta^5 f$	$\delta^6 f$
x_{-3}	f_{-3}						
		$\delta f_{-\frac{5}{2}}$					
x_{-2}	f_{-2}		$\delta^2 f_{-2}$				
		$\delta f_{-\frac{3}{2}}$		$\delta^3 f_{-\frac{3}{2}}$			
x_{-1}	f_{-1}		$\delta^2 f_{-1}$		$\delta^4 f_{-1}$		
		$\delta f_{-\frac{1}{2}}$		$\delta^3 f_{-\frac{1}{2}}$		$\delta^5 f_{-\frac{1}{2}}$	
x_0	f_0		$\delta^2 f_0$		$\delta^4 f_0$		$\delta^6 f_0$
		$\delta f_{\frac{1}{2}}$		$\delta^3 f_{\frac{1}{2}}$		$\delta^5 f_{\frac{1}{2}}$	
x_1	f_1		$\delta^2 f_1$		$\delta^4 f_1$		
		$\delta f_{\frac{3}{2}}$		$\delta^3 f_{\frac{3}{2}}$			
x_2	f_2		$\delta^2 f_2$				
		$\delta f_{\frac{5}{2}}$					
x_3	f_3						

Averaging operator 'μ'

$$\mu f(x) = \frac{f\left(x+\frac{1}{2}\right) + f\left(x-\frac{1}{2}\right)}{2}$$

$$\mu f(x_n) = \frac{f_{n+\frac{1}{2}} + f_{n-\frac{1}{2}}}{2}$$

//

$$\mu \delta f_0 = \frac{\delta f_{\frac{1}{2}} + \delta f_{-\frac{1}{2}}}{2}$$

$$\left[= \delta \mu f_0\right]$$

Lagrange's Interpolation formula

[Existence : Since non homogeneous system of eqns has a unique solution]

$\Delta = \prod (x_j - x_i)$ $\quad \forall\, j > i$

Suppose f at $x_0, x_1, \ldots, x_n$ are given as $f_1, f_2, \ldots, f_n$. The abscissas need NOT be equispaced, but disjoint (i.e. distinct).

The Lagrange's interpolating polynomial $P_n(x)$ of degree $\leq n$ such that

$$P_n(x) = \sum_{k=0}^{n} L_k(x) f_k$$

where

$$L_k(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_k)\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}$$

$$= \frac{\Pi(x)}{(x-x_k)\,\Pi'(x_k)}$$

where $\Pi(x) = (x-x_0)(x-x_1)\cdots(x-x_n)$

$$L_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^{n} \frac{(x-x_j)}{(x_k-x_j)}$$

Proof

Assume $P_n(x) = \sum_{k=0}^{n} L_k(x) f_k$ where

$L_k(x)$ is a polynomial of degree $\leq n$

We have to choose the polynomial $L_k(x)$ such that $P_n(x_k) = f_k$

$k = 0, 1, 2 \ldots n$

$$\begin{cases} L_k(x_k) = 1 \\ L_k(x_j) = 0, \; j \neq k \\ L_k(x_j) = \delta_{kj} \end{cases}$$

$$P_n(x) = L_0(x) f_0 + L_1(x) f_1 + \ldots + L_{k-1}(x) f_{k-1}$$

$$+ L_k(x) f_k + L_{k+1}(x) f_{k+1} \ldots L_n(x) f_n$$

[nothing is deleted]

$$P_n(x_k) = f_k \Rightarrow L_k(x_k) = 1$$

and

$$L_j(x_k) = 0 \quad \text{for } j \neq k$$

The polynomial $L_j(x)$ should be constructed such that

$$L_j(x_j) = 1$$
$$L_j(x_k) = 0 \quad j \neq k$$

Since $L_j(x) = 0$ for $x = x_0, x_1, \ldots x_{j-1}, x_{j+1}, \ldots x_n$

$\Rightarrow$

$(x - x_0), (x - x_1), \ldots (x - x_{j-1}), (x - x_{j+1}) \ldots, (x - x_n)$

are factors of $L_j(x)$

Since $L_j(x)$ is a polynomial of degree n.

$$L_j(x) = A_j (x-x_0)(x-x_1) \dots (x-x_{j-1})(x-x_{j+1}) \dots (x-x_n)$$

$A_j \rightarrow$ constant

$$L_j(x_j) = 1 \Rightarrow 1 = A_j (x_j-x_0)(x_j-x_1) \dots (x_j-x_{j-1})(x_j-x_{j+1}) \dots (x_j-x_n)$$

$$A_j = \frac{1}{(x_j-x_0)(x_j-x_1)\dots(x_j-x_{j-1})(x_j-x_{j+1})\dots(x_j-x_n)}$$

$$(1) \Rightarrow L_j(x) = \frac{(x-x_0)(x-x_1)\dots(x-x_{j-1})(x-x_{j+1})\dots(x-x_n)}{(x_j-x_0)\dots(x_j-x_{j-1})(x_j-x_{j+1})\dots(x_j-x_n)}$$

$$\therefore P_n(x) = \sum_{j=0}^{n} L_j(x) f_j$$

\# $\Pi(x) = (x-x_0)(x-x_1)\dots(x-x_n)$

$= (x-x_j)\,\phi_n(x)$, where $\phi_n(x)$ is a polynomial of degree n

and $\phi_n(x) = x - x_0$

// The exact forms of $L_0(x)$ $L_1(x)$ . . . .

$$L_0(x) = \frac{(x-x_1)(x-x_2) \cdots (x-x_n)}{(x_0-x_1)(x_0-x_2) \cdots (x-x_n)}$$

$$L_1(x) = \frac{(x-x_0)(x-x_2) \cdots (x-x_n)}{(x_1-x_0)(x_1-x_2) \cdots (x_1-x_n)}$$

$\vdots$

$L_n(x) =$

$$f(x) \simeq P_n(x) = \sum_{j=0}^{n} L_j(x) f_j$$

Feb 22, Class 19

Q Find the Lagrange's interpolating polynomial which fits the following data

x	-1	0	1	2
$f(x)$	1	1	1	-5

Sol

Lagrange's interpolating polynomial

$$P_n(x) = \frac{(x-0)(x-1)(x-2)}{(-1-0)(-1-1)(-1-2)} \times 1$$

$$+ \frac{(x-(-1))(x-1)(x-2)}{(0-(-1))(0-1)(0-2)} \times 1$$

$$+ \frac{(x-(-1))(x-0)(x-2)}{(1-(-1))(1-0)(1-2)} \times 1$$

$$+ \frac{(x-(-1))(x-0)(x-1)}{(2-(-1))(2-0)(2-1)} \times 5$$

$$\left[P_n(x) = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_1)(x_0-x_3)} \times f_0 + \dots \right]$$

$$= \frac{x(x^2-3x+2)}{(-1)(-2)(-3)} + \frac{(x^2-1)(x-2)}{(1)(-1)(-2)} + \frac{x(x^2-x-2)}{2(1)(-1)}$$

$$+ \frac{x(x^2-1)}{3(2)(1)} \times -5$$

$$= -\frac{1}{6}(x^3 - 3x^2 + 2x) + \frac{1}{2}(x^3 - 2x^2 - x + 2)$$
$$- \frac{1}{2}(x^3 - x^2 - 2x) - \frac{5}{6}(x^3 - x)$$

$$= (\ )x^3 + (\ )x^2 + (\ )x + (\ )$$

$$= -x^3 + x + 1$$

⇒ Interpolating polynomial is unique.

$P_n(x)$: interpolating polynomial
$Q_n(x)$: " "

$f(x) \simeq P_n(x)$: $P_n(x_i) = f(x_i)$
$f(x) \simeq Q_n(x)$: $Q_n(x_i) = f(x_i)$

The polynomial $P_n(x) - Q_n(x)$ vanishes at all x_i
$(f(x_i) - f(x_i) = 0)$

and $(x - x_0)(x - x_1) \ldots (x - x_n)$ are factors of $P_n - Q_n$ [(n+1) factors]

Contradicts fundamental theorem of algebra.

[Every non-zero polynomial of degree n has n factors]

$\therefore P_n(x) - Q_n(x) = 0$

$\Rightarrow P_n(x) = Q_n(x)$

Error associated with interpolating polynomial

Error:

$$E(x) = f(x) - P_n(x)$$

$$= \frac{\Pi(x) f^{n+1}(\xi)}{(n+1)!}$$

where $\Pi(x) = (x-x_0)(x-x_1) \cdots (x-x_n)$

ξ is in smallest interval containing $x_0, x_1 \cdots x_n$

$\xi \in (x_0, x_n)$ given $x_0 < x_1 < \cdots < x_n$ $\rightarrow H\{x_0, x_1 \cdots x_n\}$

Denote $$K(x) = \frac{f(x) - P_n(x)}{(x-x_0) \cdots (x-x_n)}$$ — (1) $\left[K = \frac{\text{error}}{\Pi(x)}\right]$

Define the function $F(t)$ as

$$F(t) = f(t) - P_n(t) - (t-x_0)(t-x_1) \cdots (t-x_n) K(x)$$

where we consider x to be a fixed value

$F(t)$ vanishes for $t = x_0, x_1 \cdots, x_n$

(because of interpolation conditions $P_n(x_i) = f_i$

$i = 0, 1, 2 \cdots n$)

see $F(x_0) = f(x_0) - P_n(x_0) - 0 (x_0 - x_1) \cdots K(x_0)$

$= 0$

$\because$ at x_i $f(x_i) = P_n(x_i)$

Also, $F(x) = 0$

i.e. $F(t)$ vanishes at $t = x$

$\therefore$ $F(t)$ vanishes at $t = x, x_0, x_1 \ldots x_n$, which are all distinct $[n+2$ points$]$

We assume that $f^{(n+1)}(t)$ is continuous on the interval $H\{x, x_0, \ldots x_n\}$

[smallest interval containing $\{x, x_0 \ldots x_n\}$
$x_0 < x_1 \ldots < x_n$]

$F(t)$ is continuous and $F(t)$ vanishes at $(n+2)$ points
By Rolle's Theorem, $F'(t) = 0$ at at least $(n+1)$ values of t other than $x_0, x_1, \ldots x_n$

By Rolles Theorem, $F''(t)$ vanishes at at least n points

$F'''(t)$ " " $(n-1)$

Rolle's theorem states that if a function f is continuous on the closed interval [a, b] and differentiable on the open interval (a, b) such that f(a) = f(b), then f′(x) = 0 for some x with a ≤ x ≤ b.

$F^{(n+1)}(t)$ vanishes once for $t = \xi$ (say)

$$F^{(n+1)}(t) = f^{(n+1)}(t) - 0 - (n+1)!\,k(x)$$

$\because P_n$ is of degree n

fixed as we are differentiating w.r.t. t

$\therefore$ t^{n+1} degree polynomial

$$\therefore \quad k(x) = \frac{f^{(n+1)}(t) - F^{(n+1)}(t)}{(n+1)!}$$

when $t = \xi$

$$k(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

Substitute $k(x)$ in ①,

$$\frac{f^{(n+1)}(\xi)}{(n+1)!} = \frac{f(x) - P_n(x)}{(x-x_0)(x-x_1)\cdots(x-x_n)} = \pi(x)$$

$$\therefore E(x) = f(x) - P_n(x) = \frac{\pi(x)\, f^{n+1}(\xi)}{(n+1)!}$$

$$P_n(x) = a_0 + a_1 x + a_2 x^2 \cdots a_n x^n$$

$$f(x_i) = P(x_i)$$

$$\begin{aligned}
a_0 + a_1 x_0 \cdots &+ a_n x_0^n = f_0 \\
a_0 + a_1 x_1 \cdots &+ a_n x_1^n = f_1 \\
&\vdots \\
a_0 + a_1 x_n \cdots &+ a_n x_n^n = f_n
\end{aligned}$$

$$\begin{bmatrix} 1 & x_0 & x_0^2 \cdots & x_0^n \\ 1 & x_1 & x_1^2 \cdots & x_1^n \\ 1 & x_2 & x_2^2 \cdots & x_2^n \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \cdots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$$

$$A \quad X \quad = B$$

Non singular if $x_0, x_1 \cdots x_n$ are diff.

[Vander Monde Matrix]

$$\det A = \prod_{i>j}^{n} (x_i - x_j)$$

Error in Newton forward difference interpolation formula.

$$E(x) = \frac{\Pi(x)\, f^{n+1}(\xi)}{(n+1)!}$$

$$\Pi(x) = (x-x_0)(x-x_1)\cdots(x-x_n)$$

Put $\frac{x-x_0}{h} = s$, $\frac{x-x_i}{h} = s-i$

$\left[\frac{x-x_1}{h} = \frac{x-(x_0+h)}{h} = \frac{x-x_0-h}{h} = \frac{x-x_0}{h} - 1 = s-1 \quad x-x_1 = h(s-1)\right]$

$$E_s = sh(s-1)h \cdots (x-n)h \frac{f^{n+1}(\xi)}{(n+1)!}$$

$$= h^{n+1} s(s-1)\cdots(s-n) \frac{f^{n+1}(\xi)}{(n+1)!}$$

Error associated with Newton Backward Difference Interpolation formula.

$$E(x) = (x-x_n)(x-x_{n-1}) \cdots (x-x_1)(x-x_0) \frac{f^{n+1}(\xi)}{(n+1)!}$$

Put $\frac{x-x_n}{h} = s$ then

$$\frac{x-x_{n-i}}{h} = s+i$$

$$\therefore E_s = sh(s+1)h \cdots (s+n)h \frac{f^{n+1}(\xi)}{(n+1)!}$$

$$= h^{n+1} s(s+1)\cdots(s+n) \frac{f^{n+1}(\xi)}{(n+1)!}$$

Error in linear interpolation

$$E_n(x) = \frac{\Pi(x)\, f^{n+1}(\xi)}{(n+1)!}$$

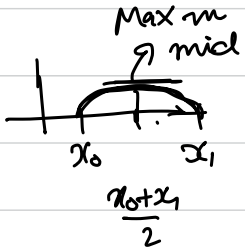
x	x_0	x_1
$f(x)$	f_0	f_1

$n = 1$ $\quad E_1(x) = \frac{1}{2}(x-x_0)(x-x_1) f''(\xi_1)$

Let $|f''(x)| \le M_2 \quad \forall\ x \in [x_0, x_1]$

$$\therefore \quad |E_1(x)| \le \frac{1}{2}(x-x_0)(x-x_1) M_2$$

$$= \frac{1}{2} |(x-x_0)(x-x_1)| M_2$$



$$\le \frac{1}{2} \left| \frac{1}{2}(x_0+x_1) - x_0 \right| \left| \frac{1}{2}(x_0+x_1) - x_1 \right| M_2$$

$$\le \frac{1}{2} \left| \frac{1}{2}(x_1-x_0) \right| \left| \frac{1}{2}(x_0-x_1) \right| M_2$$

$$\le \frac{1}{8}(x_1-x_0)^2 M_2$$

if x_i are equispaced, $x_1 - x_0 = h$

$$\le \frac{1}{8} h^2 M_2.$$

$$|E_1(x)| \le \frac{1}{8} h^2 \operatorname{Max}_{x_0 < x < x_n} |f''(x)|$$

Prove the results :

Q. $E_1(x) \le \frac{h^2}{8} M_2$ for $x \in [x_0, x_1]$ → Error in linear int.

$E_2(x) \le \frac{h^3 M_3}{9\sqrt{3}}$ for $x \in [x_0, x_2]$ → quadratic interpolation

$E_3(x) \le \frac{h^4 M_4}{24}$ for $x \in [x_0, x_3]$ → error in cubic interpolation

Q: Show that the magnitude of error in the linear interpolation of the error function

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

between x_0 and x_1 cannot exceed $\frac{(x_1 - x_0)^2}{2\sqrt{2\pi e}}$

Sol

$$\frac{d}{dx}\left(\int_{u(x)}^{v(x)} f(t)\, dt \right) = f(v(x)) \frac{dv}{dx} - f(u(x)) \frac{d(u(x))}{dx}$$

$$\frac{d}{dx}[erf(x)] = \frac{2}{\sqrt{\pi}}\left[e^{-x^2} \cdot 1 - 0\right] = \frac{2}{\sqrt{\pi}} e^{-x^2}$$

$$\frac{d^2}{dx^2}[\quad] = \frac{2}{\sqrt{\pi}} e^{-x^2} \times (-2x) = \frac{-4x}{\sqrt{\pi}} e^{-x^2}$$

To find $\underset{x_0 \le x \le x_1}{Max} \left| \frac{d^2}{dx^2} erf(x) \right|$

Find third derivative and equate to zero

We get the point where $f''(x)$ takes the max value

$$\frac{d^3}{dx^3}(\text{erf}(x)) = \frac{-4}{\sqrt{\pi}}\left(e^{-x^2} + x \cdot e^{-x^2}(-2x)\right) = 0$$

$$1 - 2x^2 = 0$$

$$x = \pm \frac{1}{\sqrt{2}}$$

$\therefore$ Max $\frac{d^2}{dx^2}(\text{erf}(x))$ occurs at $x = \frac{1}{\sqrt{2}}$ (check)

Error in linear interpolation —

$$E_1(x) = \frac{1}{2}(x - x_0)(x - x_1) f''(\zeta)$$

$$E_1(x) = \frac{1}{2}\left((x - x_0)(x - x_1)\right)\left(\frac{-4}{\sqrt{\pi}} x e^{-x^2}\right)$$

$$\leq \left|\frac{1}{2}\left(\frac{(x_1 - x_0)}{2}\frac{(x_1 - x_0)}{2}\right)\right| \left|\left(\frac{-4}{\sqrt{\pi}}\left(\frac{1}{\sqrt{2}}\right) e^{-\frac{1}{2}}\right)\right| \quad \downarrow \text{at } x = \frac{1}{\sqrt{2}}$$

$$\leq \frac{1}{8}(x_1 - x_0)^2 \frac{4}{\sqrt{2}\sqrt{\pi}\sqrt{e}}$$

$$\leq \frac{1}{2\sqrt{2\pi e}}(x_1 - x_0)^2$$

$\downarrow$

Max error in linear interpolation.

Feb 28, Class 20

Inverse Interpolation

f_i	x_i
f_0	x_0
⋮	⋮
f_n	x_n

is called inverse interpolation

$$x = \frac{(f-f_1)(f-f_2)\cdots(f-f_n)}{(f_0-f_1)(f_0-f_2)\cdots(f_0-f_n)}x_0 + \frac{(f-f_0)(f-f_2)\cdots(f-f_n)}{(f_1-f_0)(f_1-f_2)\cdots(f_1-f_n)}x_1$$

$$\cdots + \frac{(f-f_0)\cdots(f-f_{n-1})}{(f_n-f_0)(f_n-f_1)\cdots(f_n-f_{n-1})}x_n$$

$$x = \sum_{k=0}^{n} L_k(f)\, x_k$$

where $L_k(f) = \dfrac{(f-f_0)\cdots(f-f_{k-1})(f-f_{k+1})\cdots(f-f_n)}{(f_k-f_0)(f_k-f_1)\cdots(f_k-f_{k-1})(f_k-f_{k+1})\cdots(f_k-f_n)}$

$$L_k(f_n) = 1$$
$$L_k(f_j) = 0 \qquad k \neq j$$

All f_i's are distinct.

Hermite Interpolation Polynomial

x_i	x_0	x_1	..	x_n
f_i	f_0	f_1		f_n
f_i'	f_0'	f_1'		f_n'

$$H(x_k) = f_k \quad \left.\begin{matrix} \\ \end{matrix}\right\} \; k = 0, 1, 2 \ldots n \quad \rightarrow \text{Interpolation conditions}$$
$$H'(x_k) = f_k'$$

$\Rightarrow H(x)$ is a polynomial of degree $\leq 2n+1$ - ? HOW

$$H(x) = \sum_{j=0}^{n} h_j(x) f_j + \sum_{j=0}^{n} \bar{h}_j(x) f_j'$$

where $h_j(x)$ and $\bar{h}_j(x)$ are polynomial of degree $\leq 2n+1$ in x and

$$h_j(x) = [1 - 2(x - x_j) l_j'(x_j)] \, l_j^2(x)$$

$$\bar{h}_j(x) = (x - x_j) \, l_j^2(x)$$

$$l_j(x) = \frac{(x-x_0) \ldots (x-x_{j-1})(x-x_{j+1}) \ldots (x-x_n)}{(x_j-x_0) \ldots (x_j-x_{j-1})(x_j-x_{j+1}) \ldots (x_j-x_n)}$$

Error in Hermite Interpolation

$f \in C^{2n+2}[x_0, x_n]$ — continuous and differentiable (upto 2n+2 derivatives exist) on the given interval

$x_0 < x_1 \ldots < x_n$

$$E(x) = f(x) - H(x) = [\Pi(x)]^2 \frac{f^{2n+2}(\xi)}{(2n+2)!}$$

where $\Pi(x) = (x-x_0)(x-x_1) \ldots (x-x_n)$

[By repititive application of Rolle's Theorem]

General Hermite Interpolation

Let $x_0 < x_1 \cdots < x_n$ and integers $m_0, m_1 \cdots m_n$ are all greater than zero.
We find a unique polynomial of degree

$$N = m_0 + m_1 \cdots + m_n - 1$$

Solving the interpolation problem

$$p^{j}(x_i) = f^{j}(x_i)$$

$$j = 0, 1, 2 \cdots, m_i - 1$$

$$i = 0, 1, 2 \cdots n$$

$p(x)$ is called the General Hermite Interpolation polynomial.

Usual Hermite Interpolating Polynomial

$$x_0 < x_1 \cdots < x_n$$

$$m_0 = m_1 \cdots = m_n = 2$$

Error in General Hermite Interpolation Polynomial.

If $f \in C^{n+1}[a, b]$, then for $x \in [a, b]$, there exists a ξ in (a, b) such that all nodal points are placed in (a, b) $\quad a \leq x_0 < x_1 \cdots < x_n \leq b$

$$f(x) - P(x) = \frac{f^{(N+1)}(\xi)}{(N+1)!} (x - x_0)^{m_0} (x - x_1)^{m_1} \cdots (x - x_n)^{m_n}$$

where $N = m_0 + m_1 + m_2 \cdots + m_n - 1$

$$= \left(\sum_{i=0}^{n} m_i \right) - 1$$

Taylor Polynomial regarded as interpolating polynomial

Let $f(x)$ has $n+1$ continuous derivatives on $[a, b]$ for $n \geq 0$

Let x, $x_0 \in [a, b]$. Then

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

where $P_n(x) = f(x_0) + \frac{(x-x_0)}{1!} f'(x_0) + \frac{(x-x_0)^2}{2!} f''(x_0) + \cdots$

$$\cdots + \frac{(x-x_0)^n}{n!} f^n(x) \quad \text{— (1)}$$

$R_{(n+1)}(x)$ → remainder after n terms

$$R_{n+1} = \frac{1}{n!} \int_{x_0}^{x} (x-t)^n f^n(t)\, dt \rightarrow \text{integral form of remainder}$$

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

$P_n(x)$ may be regarded as interpolating polynomial of degree 'n' satisfying the conditions

$$P_n^k(x_0) = f^k(x_0) \qquad k = 0, 1, \cdots n$$

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

$x_0 < \xi < x$

Error Bound

The number of terms to be retained in (1) may be determined by acceptable error. If the error is $\varepsilon > 0$ and the series is truncated at the term $f^n(x_0)$, then

$$\frac{1}{(n+1)!}(x-x_0)^{n+1}\left|f^{n+1}(\xi)\right| \leq \varepsilon \qquad \varepsilon \sim 10^{-3}, 10^{-5}$$

$$\boxed{\frac{(x-x_0)^{n+1}}{(n+1)!} M_{n+1} \leq \varepsilon} \quad — (2)$$

where $M_{n+1} = \max_{a \leq x \leq b} \left|f^{n+1}(x)\right|$

ε, x, n

Assume that value of M_{n+1} or its estimate is available

ε, x, n

For a given ε and x, (2) will determine 'n'.

- If n, x are given, (2) will determine 'ε'
- When both n, ε are given, (2) will give an upper bound on $(x-x_0)$

⇓

It will give an interval about x_0 in which this Taylor Polynomial approximates $f(x)$ to the prescribed accuracy.

It is not necessary that more is the degree of approximated function, more is the accuracy.

March 1, class 21

Bivariate Interpolation

$$(x_i, y_i) \quad i = 0, 1, \cdots m$$
$$j = 0, 1 \cdots n$$

Denote $f(x_i, y_j) = f_{ij}$

We want to approximate $f(x, y)$ by a polynomial of degree at most m in x and at most n in y such that

$$P(x_i, y_j) = f_{ij}$$

$$P_{m,n}(x, y) = \sum_{i=0} \sum_{j=0} X_{m,i}(x)\, Y_{n,j}(y)\, f_{ij}$$

where $X_{m,i} = \dfrac{w(x)}{(x - x_i)\, w'(x_i)} \quad i = 0, 1, 2.$

$$Y_{n,j} = \frac{w^*(y)}{y - y_j \; w^{*\prime}(y_j)} \quad j = 0, 1, 2$$

where $w(x) = (x - x_0)(x - x_1) \cdots (x - x_m)$

$w^*(y) = (y - y_0)(y - y_1) \cdots (y - y_n)$

$X_{m,i}(x)$ and $Y_{n,j}(y)$ are polynomials of degree m in x and n in y resp. These polynomials satisfy -

$$X_{m,i}(x_k) = \delta_{ik}$$

$$Y_{n,j}(y_k) = \delta_{jk}$$

Error

$$E(x) = f(x) - P_n(x) = \frac{\Pi(x) f^{n+1}(\xi)}{(n+1)!}$$

$$\Pi(x) = (x-x_0)(x-x_1) \cdots (x-x_n)$$

eg $f(x) = \frac{1}{1+x^2}$, $-5 \le x \le 5$

Let $n > 0$ be an even integer, define $h = \frac{10}{n}$

$$x_j = -5 + jh, \quad j = 0, 1, 3, \ldots$$

$-5, -5+h, -5+2h \ldots, -5+10$

$P_{10}(x)$ (in red)

$[x_i, x_{i+1}]$ approximate $f(x)$ by a polynomial of lower degree

Piecewise Polynomial Interpolation

Cubic Splines

$$a = x_0 < x_1 < x_1 \quad \cdots \quad < x_n = b$$

$x_0, x_1, x_2 \ldots x_n$ are called knots in cubic spline.
A function S is said to be a cubic spline on interval $[a, b]$ if —

(i) S, S' and S'' are continuous on $[a, b]$
(ii) S is a polynomial of degree ≤ 3 in each knot interval / sub interval $[x_{i-1}, x_i]$

The spline function S is composed of 'n' cubic polynomials. (n intervals, n+1 knots)

$$S = S_i(x), \quad x_{i-1} \leq x \leq x_i$$

Each S_i has 4 coefficients

$$S \Big|_{[x_{i-1}, x_i]} S_i = a_i x^3 + b_i x^2 + c_i x + d_i$$

n+1 knots
⇓
n intervals
⇓
(n-1) interval functions

But the continuity requirement gives $3(n-1)$ conditions —

- $S_i(x_i) = S_{i+1}(x_i)$
- $S_i'(x_i) = S_{i+1}'(x_i)$

To force continuity

- $S_i''(x_i) = S_{i+1}''(x_i)$

n intervals. Each has 4 unknowns a, b, c, d. ∴ 4n unknowns

There remain $4n - 3(n-1) = n+3$ degrees of freedom

If we choose to make the spline interpolate in the knots $S_i(x_i) = f(x_i)$, then we have $(n+1)$ conditions. Common choices for the remaining conditions are given by the following theorem

Then $n+3-(n+1) = 2$ degrees of freedom remain

Theorem : Let $\{x_i\}_{i=0}^{n}$ be given points with

$$a = x_0 < x_1 < x_2 - - - < x_n = b$$

A cubic spline S with the knots $x_0, x_1 \ldots, x_n$ is uniquely determined by the interpolation condition $S(x_i) = f(x_i)$, $i = 0, 1, 2 \ldots, n$.

Splines with any of the following choices of two extra conditions —

Natural spline : $S_1''(x_0) = S_n''(x_n) = 0$

[second derivative is zero at end points]

Correct boundary condition : $S_1'(x_0) = f'(x_0)$,

$S_n'(x_n) = f'(x_n)$

Not-a-Knot : $S_1^{(3)}(x_1) = S_2^{(3)}(x_1)$,

$S_{n-1}^{(3)}(x_{n-1}) = S_n^{(3)}(x_{n-1})$

Periodic Boundary condition : $S_1^{(r)}(x_0) = S_n^{(r)}(x_n)$ $r = 1, 2$

Theorem gives choices for the remaining 2 degrees of freedom

APPROXIMATION

Weierstrass Approximation Theorem

If the function f is continuous on a closed interval $[a,b]$, then given $\varepsilon > 0$, there exists an $n : n(\varepsilon)$ and a polynomial of degree n such that –

$$|f(x) - P(x)| < \varepsilon \quad \forall \; x \in [a, b]$$

Given a $f \in C[a,b]$ and $\varepsilon > 0$, there is a $p \in P_n =$ set of all polynomials such that

$$\|f - p\| < \varepsilon$$

Best approximation

Let $p_n \in P_n \longrightarrow$ degree $\leq n$

$$\|f - p_n\| \leq \|f - q_n\| \quad \forall \quad q_n \in P_n$$

Then p_n is called the best approximation

$$f(x) \simeq a_0 \phi_0(x) + a_1 \phi_1(x) + \cdots + a_n \phi_n(x)$$

where $\phi_0, \phi_1, \ldots \phi_n$ are approximately chosen linearly independent function

$a_0, a_1, \ldots, a_n$ are the parameters to be determined

ϕ_i are called coordinate functions.

$\phi_i = x^i \quad i = 0, 1, 2 \ldots n$ for polynomial approx.

Error of approximation

$$E = \| f(x) - (a_0\phi_0(x) + a_1\phi_1(x) \ldots + a_n\phi_n(x)) \| \quad \text{— (*)}$$

$\| \cdot \|$ → well defined norm

The problem of approximation is to determine $a_0, a_1 \ldots a_n$ such that the error is as small as possible in some sense

- We get least square approximation if we use Euclidean norm
- We get uniform approximation if we use uniform norm or maximum norm or Chebychev norm.

March 3, Class 22

Most commonly used norms.

- L^p - norm $\quad \|x\|_p = \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p}, \quad p \geq 1$

where $x = (x_1, x_2 \ldots, x_n)^T$

- Euclidean Norm : $\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$

which is a particular case of above norm where $p=2$

- Uniform norm $= \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$

which is a particular case of above norm where $p=2$

- For function defined on a grid $\{x_i\}_{i=1}^{m}$, the corresponding p norm are defined as

$$\|f\|_p = \left(\sum_{i=1}^{n} w(x_i) \, |f(x_i)|^p \right)^{1/p}$$

where. $w(x)$ is a weight function.

- When continuous data

If function f is continuous on $[a, b]$ and $|f(x)|^p$ is integrable on $[a, b]$, then

$$\|f\|_p = \left(\int_a^b w(x) \, |f(x)|^p \right)^{\frac{1}{p}} \quad , p \geq 1 \qquad -(2)$$

- Euclidean norm : for $p = 2$ in (2), we get euclidean norm or square norm

$$\|f\|_2 = \left(\int_a^b w(x) \, |f(x)|^2 \right)^{1/2}$$

- Uniform norm : for $p = \infty$ in (2), we get uniform norm

$$\|f\|_\infty = \max_{a \leq x \leq b} |f(x)|$$

NOTE

- When we use the Euclidean norm, we obtain least square approximation
- When we use Uniform norm, we obtain uniform approximation

Functions which are continuous on $[a, b]$ and given explicitly

$$S = \int_a^b w(x) \left[f(x) - \sum_{i=0}^{n} a_i \phi_i \right]^2 dx \quad - (A)$$

$= S(a_0, a_1 \ldots a_n) = \min$

↓

Square of norm of error w.r.t weight factor $w(x)$

The necessary condition for S to have minimum value is that

$$\frac{\partial S}{\partial a_j} = 0 \quad , \; j = 0, 1, \ldots, n$$

This gives a system of $(n+1)$ linear equations in $(n+1)$ unknowns $a_0, a_1 \ldots, a_n$
These are called NORMAL EQUATIONS

The normal equations are –

$$\int_a^b w(x) \left[f(x) - \sum_{i=0}^{n} a_i \phi_i(x) \right] \phi_j(x) \, dx = 0 \quad - (B)$$

$j = 0, 1, \ldots, n$

Coordinate functions ϕ_i are chosen as $\phi_i(x) = x_i$, $i = 0, 1, \ldots n$ for polynomial approximation
Also, sometimes we choose $w(x) = 1$

Example Find linear approximation to the function $f(x) = x^3$ on $[0,1]$ by method of least square

[Here $w(x) = 1$. If nothing is mentioned about weight function, choose $w(x) = 1$]

Solution

Let $p(x) = a_0 + a_1 x$, $w(x) = 1$

where a_0, a_1 are constants to be determined

$$S(a_0, a_1) = \int_0^1 (x^3 - a_0 - a_1 x)^2 \, dx$$

$$\frac{\partial S}{\partial a_0} = 0 \quad \text{and} \quad \frac{\partial S}{\partial a_1} = 0$$

$$\frac{\partial S}{\partial a_0} = 0 \Rightarrow 2\left(\int_0^1 (x^3 - a_0 - a_1 x)\, dx\right) \times (-1) = 0$$

$$\Rightarrow -2\left(\left[\frac{x^4}{4} - a_0 x - \frac{a_1 x^2}{2}\right]_0^1\right) = 0$$

$$\frac{1}{4} - a_0 - \frac{a_1}{2} = 0$$

$$1 - 4a_0 - 2a_1 = 0$$

$$4a_0 + 2a_1 = 1 \qquad -(1)$$

$$\frac{\partial S}{\partial a_1} = 0 \Rightarrow 2 \times \left(\int_0^1 (x^3 - a_0 - a_1 x)\, dx\right) \times (-x) = 0$$

$$\Rightarrow \int_0^1 (x^4 - a_0 x - a_1 x^2)\, dx = 0$$

$$\left[\frac{x^5}{5} - \frac{a_0 x^2}{2} - \frac{a_1 x^3}{3}\right]_0^1 = 0$$

$$\frac{1}{5} - \frac{a_0}{2} - \frac{a_1}{3} = 0.$$

$$\frac{a_0}{2} + \frac{a_1}{3} = \frac{1}{5} \qquad -(2)$$

Solving (1) and (2) we get

$$a_0 = \frac{-1}{5}, \quad a_1 = \frac{9}{10}$$

$$\therefore \quad P(x) = \frac{-1}{5} + \frac{9}{10}x.$$

Functions whose values are given at N+1 points

$x_0, x_1, \ldots, x_N$

$$S = \sum_{k=0}^{N} w(x_k) \left[f(x_k) - \sum_{i=0}^{n} a_i \phi_i(x_k) \right]^2 \qquad -(C)$$

$= S(a_0, a_1 \ldots a_n) = \min$

square of norm of error w.r.t. weight factor $w(x_k)$

Necessary conditions for S to have minimum value -

$$\frac{\partial S}{\partial a_j} = 0 \qquad j = 0, 1, 2 \ldots$$

This gives a system of (n+1) linear equations in (n+1) unknowns $a_0, a_1 \ldots, a_n$

These are called NORMAL EQUATIONS

The normal equations are -

$$\sum_{k=0}^{N} w(x_k)\left[f(x_k) - \sum_{i=0}^{n} a_i \phi_i(x_k)\right]\phi_j(x_k) = 0 \quad \text{(D)}$$

In the polynomial approximation, we choose $\phi_i = x^i$

Example Find the least squares approximation of first degree for the discrete set

	x_0	x_1	x_2	x_3	x_4
x	-2	-1	0	1	2
$f(x)$	15	1	1	3	19

Solution Let $p(x) = a_0 + a_1 x$

Here $w(x) = 1$ (since nothing is mentioned)

$$S = \sum_{k=0}^{4} w(x_k)\left[f(x_k) - (a_0 + a_1 x_k)\right]^2$$

$$\frac{\partial S}{\partial a_0} = 0 \Rightarrow \sum_{k=0}^{4} 1 \cdot 2\left[f(x_k) - (a_0 + a_1 x_k)\right] \times (-1) = 0$$

$$\frac{\partial S}{\partial a_1} = 0 \Rightarrow \sum_{k=0}^{4} 1 \cdot 2\left[f(x_k) - (a_0 + a_1 x_k)\right] \times (-x_k) = 0$$

$$\sum_{k=0}^{4} f(x_k) - \sum_{k=0}^{4} a_0 - a_1 \sum_{k=0}^{4} x_k = 0$$

$$(15+1+1+3+19) - 5a_0 - a_1(-2-1+0+1+2) = 0$$

$$5a_0 = 39$$

$$a_0 = \frac{39}{5}$$

$$\sum_{k=0}^{4} x_k f_k - a_0 \sum_{k=0}^{4} x_k - a_1 \sum_{k=0}^{4} x_k^2 = 0$$

$$a_1 = 1$$

$$\Rightarrow \boxed{p(x) = \frac{39}{5} + x}$$

March 14, Class 23

Theorem (Least Squares Approximation)

Let U be a subspace of a normed linear space L.
Let $f \in L$.
Assume $\phi_1, \phi_2 \dots, \phi_n$ are L.I. Then there is a unique element $f^* \in U$, where

$$f^* = \sum_{j=0}^{n} c_j^* \phi_j \quad \text{such that}$$

$$\|f - f^*\|_2 \leq \|f - g\|_2 \quad \text{for all } g = \sum_{j=0}^{n} L_j \phi_j$$

↓

[Euclidean norm / Square norm]

f^* is characterized by

$$\langle f - f^*, \phi_k \rangle = 0 \quad , k = 0, 1, 2 \dots, n$$

$$\langle f, \phi_k \rangle - \langle f^*, \phi_k \rangle = 0$$

$$\langle f, \phi_k \rangle = \langle f^*, \phi_k \rangle$$

$$\left\langle \sum_{j=0}^{n} c_j^* \phi_j , \phi_k \right\rangle = \langle f, \phi_k \rangle$$

$$\Rightarrow \boxed{\sum_{j=0}^{n} \langle \phi_j, \phi_k \rangle c_j^* = \langle f, \phi_k \rangle}$$

$k = 0, 1, \dots, n$

System of $(n+1)$

$\{c_j\}$ are obtained by solving $(n+1)$ system

x_i	x_0	x_1	$\cdots$	x_N
f_i	f_0	f_1	$\cdots$	f_N

1. • Linear approximation $f \simeq f^* = a + bx$

   • Normal equations

$$\begin{cases} a(N+1) + b\sum_{k=0}^{N} x_k = \sum_{k=0}^{N} f(x_k) \\ a\sum_{k=0}^{N} x_k + b\sum_{k=0}^{N} x_k^2 = \sum_{k=0}^{N} x_k f(x_k) \end{cases}$$

2. • Quadratic Approximation $f \approx f^* = a + bx + cx^2$

   • Normal equations

$$\begin{cases} a(N+1) + b\sum_{k=0}^{N} x_k + c\sum_{k=0}^{N} x_k^2 = \sum_{k=0}^{N} f(x_k) \\ a\sum_{k=0}^{N} x_k + b\sum_{k=0}^{N} x_k^2 + c\sum_{k=0}^{N} x_k^3 = \sum_{k=0}^{N} x_k f(x_k) \\ a\sum_{k=0}^{N} x_k^2 + b\sum x_k^3 + c\sum x_k^4 = \sum x_k^2 f(x_k) \end{cases}$$

Why the solution of the least squares problem is unsatisfactory ?

because $$\int_a^b w(x) f(x) \phi_j(x)\, dx = \int_a^b w(x) \left(\sum_{i=0}^{n} a_i \phi_i(x) \right) \phi_j(x)\, dx$$

Consider the special case

$w(x) = 1$

$[a, b] = [0, 1]$

$\phi_i(x) = x^i$

$\phi_j(x) = x^j$

Then $$\int_0^1 f(x)\, x^j\, dx = \int_0^1 \left(\sum_{i=0}^{n} a_i x^i \right) x^j\, dx$$

$$= \sum_{i=0}^{n} \int_0^1 a_i\, x^{i+j}\, dx$$

$$\int_0^1 f(x)\, x^j\, dx = \sum_{j=0}^{n} \frac{a_i}{i+j+1}$$

or $$\sum_{j=0}^{n} \frac{a_i}{i+j+1} = \int_0^1 f(x)\, x^j\, dx \qquad (*)$$

The coeff. matrix is a Hilbert matrix of order $(n+1)$

Hilbert Matrix

It is a $n \times n$ non singular matrix

$$H = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+1} \\ \vdots & & & & \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n-1} \end{bmatrix}$$

$$H = [k_{ij}], \quad \text{where } k_{ij} = \frac{1}{i+j-1}$$

The solution of linear system (*) is extremely sensitive to small changes in the coeff. matrix on right hand side constant.

Thus, this is NOT a good way to approach least square solution

ORTHOGONAL FUNCTIONS

Def: A set of function $\{\phi_i(x)\}$ is said to be orthogonal on interval $[a,b]$ w.r.t. weight function $w(x)$ if

(Notation: $\langle \phi_i(x), \phi_j(x) \rangle = 0$)

$$\int_a^b w(x)\,\phi_i(x)\,\phi_j(x)\,dx = 0 \qquad i \neq j$$

Def: A set of function $\{\phi_i(x)\}$ is said to be orthogonal over a set of points $\{x_i\}_{i=0}^N$ w.r.t. weight function $w(x)$ if

$$\sum_{k=0}^{N} w(x_k)\,\phi_i(x_k)\,\phi_j(x_k) = 0$$

Recall NORMAL EQUATIONS

eqn (B) : $\int_a^b w(x) \left[f(x) - \sum_{i=0}^{n} a_i \phi_i(x) \right] \phi_j(x) \, dx = 0$

eqn (D) : $\sum_{k=0}^{N} w(x_k) \left[f(x_k) - \sum_{i=0}^{n} a_i \phi_i(x_k) \right] \phi_j(x_k) = 0$

- For continuous

If functions $\{\phi_i(x)\}$ $i = 0, 1 \ldots n$ are orthogonal, then from (B),

$$\int_a^b w(x) f(x) \phi_j(x) \, dx = \int_a^b w(x) \left(\sum_{i=0}^{n} a_i \phi_i(x) \right) \phi_j(x) \, dx$$

$$= \sum_{i=0}^{n} a_i \int_a^b w(x) \phi_i(x) \phi_j(x) \, dx$$

$$= a_j \int_a^b w(x) \phi_j^2(x) \, dx$$

or

$$a_j = \frac{\int_a^b w(x) f(x) \phi_j(x) \, dx}{\int_a^b w(x) \phi_j^2(x) \, dx}$$

$$= \frac{\int_a^b w(x) f(x) \phi_j(x) \, dx}{\| \phi_j \|^2}$$

- For discrete,

If function $\{\phi_i(x)\}$, $i = 0, 1, 2 \ldots n$ are orthogonal,

then (D) can be written as

$$\sum_{k=0}^{N} w(x_k) \left[f(x_k) - \sum_{i=0}^{n} a_i \phi_i(x_k) \right] \phi_j(x_k) = 0$$

$$\sum_{k=0}^{N} w(x_k) \phi_j^2(x_k) a_j = \sum_{k=0}^{N} w(x_k) \phi_j(x_k) f(x_k)$$

or

$$\boxed{a_j = \frac{\sum_{k=0}^{N} w(x_k) \phi_j(x_k) f(x_k)}{\sum_{k=0}^{N} w(x_k) \phi_j^2(x_k)}}$$

Gram - Schmidt Orthogonalisation Process.

Given the polynomials $\phi_i(x)$ of degree 'i', the polynomials $\phi_i^*(x)$ of degree 'i' which are orthogonal over $[a, b]$ w.r.t. weight factor $w(x)$ can be generated recursively from the relation

$$\phi_i^*(x) = \phi_i - \sum_{j=0}^{i-1} a_{ij} \phi_j^*(x), \quad i = 1, 2 \ldots, n$$

where $a_{ij} = \dfrac{\int_a^b w(x) \phi_i(x) \phi_j^*(x) \, dx}{\int_a^b w(x) \phi_j^{2*}(x) \, dx}$ and $\phi_0^*(x) = 1$

or $a_{ij} = \dfrac{\langle \phi_i(x), \phi_j^*(x) \rangle}{\langle \phi_j^*(x), \phi_j^*(x) \rangle}$ and $\phi_0^*(x) = 1$

$$\phi_i^*(x) = \phi_i(x) - \sum_{j=0}^{i-1} \frac{\langle \phi_i(x), \phi_j^*(x) \rangle}{\langle \phi_j^*(x), \phi_j^*(x) \rangle} \phi_j^*(x)$$

and

$\phi_0^*(x) = 1$

NOTE

On a discrete set of points, the integral is replaced by summation.

March 15, Class 24

Q. Using Gram - Schmidt orthogonalisation process, compute first 3 orthogonal polynomials $P_0(x)$ $P_1(x)$ $P_2(x)$ which are orthogonal on $[0,1]$ w.r.t. weight function $w(x) = 1$

$$\{1, x, x^2\} \xrightarrow[\text{orthogonalisation process.}]{\text{Gram - Schmidt}} \{P_0(x)\ P_1(x)\ P_2(x)\}$$

$$P_0(x) = 1 = \phi_0^*(x)$$

$$P_1(x) = \phi_1^*(x) = x - a_{10}\,\phi_0^*(x)$$

where $$a_{10} = \frac{\int_0^1 1 \cdot x \cdot 1\, dx}{\int_0^1 1 \cdot 1^2 dx} = \frac{\left[\frac{x^2}{2}\right]_0^1}{1} = \frac{1}{2}$$

$$P_1(x) = x - \frac{1}{2} = \phi_1^*(x)$$

$$P_2(x) = \phi_2^*(x) = x^2 - a_{20}\,\phi_0^*(x) - a_{21}\,\phi_{21}^*(x)$$

where $$a_{20} = \frac{\int_0^1 1 \cdot x^2 \cdot 1\, dx}{\int_0^1 1 \cdot 1^2} = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$$

$$a_{21} = \frac{\int_0^1 1 \cdot x^2 \left(x - \frac{1}{2}\right) dx}{\int_0^1 \left(x^2 - x + \frac{1}{4}\right) dx}$$

$$= \frac{\left[\frac{x^4}{4} - \frac{x^3}{6}\right]_0^1}{\left[\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right]_0^1} = \frac{\frac{1}{4} - \frac{1}{6}}{\frac{1}{3} - \frac{1}{2} + \frac{1}{4}} = 1$$

$$P_2(x) = x^2 - \frac{1}{3} \cdot 1 - 1\left(x - \frac{1}{2}\right)$$

$$= x^2 - x + \frac{1}{6}$$

Legendre Polynomials

$$(1-x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + n(n+1)y = 0$$

$x \in (-1, 1)$

n is a non-negative integer

$$y = A P_n(x) + B Q_n(x)$$

A, B are arbitrary constants.

$P_n(1) = 1$ is Legendre Polynomial.

$P_n(x)$ is called Legendre Polynomial of first Kind

$Q_n(x)$ is called Legendre Series of first Kind.

Properties

(1) Legendre Polynomials satisfy the recurrance relations

$$(n+1)P_{n+1}(x) = (2n+1)x P_n(x) - n P_{n-1}(x)$$

(2) $P_n(x)$ is an even degree polynomial if n is even and odd if n is odd

(3) $P_n(-x) = (-1)^n P_n(x)$

(4) $P_n(x)$ defined on $[-1, 1]$ are orthogonal and satisfy

$$\int_{-1}^{1} P_m(x) P_n(x)\, dx = 0$$

$$= \begin{cases} 0 & \text{if } m \neq n \\ \frac{2}{2n+1} & \text{if } m = n \end{cases}$$

Few Legendre Polynomials

$$P_0(x) = 1$$
$$P_1(x) = 2$$
$$P_2(x) = \frac{1}{2}(3x^2 - 1)$$
$$P_3(x) = \frac{1}{3}(5x^3 - 3x)$$

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$$

Rodrigues Formula

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}[(x^2-1)^n]$$

Chebychev Polynomials

Chebychev D.E. : $(1-x^2)\frac{d^2y}{dx^2} - x\frac{dy}{dx} + n^2 y = 0$

General Solution - $y = A T_n(x) + B\sqrt{1-x^2}\, U_{n-1}(x)$, $n = 1, 2 \ldots$

$= A + B\sin^{-1}(x)$, $n = 0$

$T_n(x)$: Chebychev Polynomial of first kind

$$T_n(x) = \cos(n\cos^{-1}x) \quad \text{or} \quad \cos n\theta \quad \text{where } \theta = \cos^{-1}x$$

$$U_n(x) = \sin[(n+1)\cos^{-1}x]$$

(1) Chebychev polynomials $T_n(x)$ satisfy the recurrence relation

$$T_{n+1}(x) = 2x\,T_n(x) - T_{n-1}(x)$$

(2) $T_n(x) = 2^{n-1}x^n + \text{term of lower degree}$

Few chebychev polynomials

$$T_0(x) = 1$$
$$T_1(x) = x$$
$$T_2(x) = 2x^2 - 1$$
$$T_3(x) = 4x^3 - 3x$$
$$T_4(x) = 8x^4 - 8x^2 + 1$$

$$1 = T_0(x)$$
$$x = T_1(x)$$
$$x^2 = \frac{1}{2}\left[T_2(x) + T_0(x)\right]$$
$$x^3 = \frac{1}{4}\left[T_4(x) + 3T_1(x)\right]$$

Various powers of x can be written in terms of Chebychev polynomials

Properties

(1) $T_n(x)$ is a polynomial of degree n, $T_n(x)$ is an even degree polynomia and if n is odd, $T_n(x)$ is an odd polynomial.

(2) $T_n(x)$ has n simple zeros

$$x_k = \cos\left(\frac{2k-1}{2n}\pi\right) \qquad k = 1, 2 \ldots n$$

or

$$x_k = \cos\left(\frac{(2k+1)\pi}{2n}\right) \qquad k = 0, 1 \ldots n-1$$

(3) $T_n(x)$ assumes extreme values (max / min) at $(n+1)$ points

$$x_k = \cos\left(\frac{k\pi}{n}\right) \qquad k = 0, 1, 2 \ldots n$$

(These are the points where $T_n'(x) = 0$)
and extreme values at x_k is $(-1)^k$

(4) $|T_n(x)| \leq 1 \;, \qquad x \in [-1, 1]$

(5) If $P_n(x)$ is any polynomial of degree n with leading coeff. unity (monic polynomial) and

$$\tilde{T}_n(x) = \frac{T_n(x)}{2^{n-1}}$$ is monic chebychev polynomial.

then $$\underset{-1 \leq x \leq 1}{\text{Max}} \left|\tilde{T}_n(x)\right| \leq \underset{-1 \leq x \leq 1}{\text{Max}} \left|P_n(x)\right|$$

This property is called Mini Max Property

(6) $T_n(x)$ are orthogonal w.r.t. the weight function

$$w(x) = \frac{1}{\sqrt{1-x^2}} \quad \text{on } [-1,1]$$

$$\int_{-1}^{1} \frac{T_m(x)\, T_n(x)}{\sqrt{1-x^2}}\, dx = \begin{cases} 0 & m \neq n \\ \frac{\pi}{2} & m = n \neq 0 \\ \pi & m = n = 0 \end{cases}$$

$$\int_{-1}^{1} \frac{T_m(x)\, T_n(x)}{\sqrt{1-x^2}}\, dx = \int_{0}^{\pi} \cos m\theta \cos n\theta \, d\theta = 0, \quad m \neq n$$

- Any function which is continuous on $[-1,1]$ may be expressed uniquely in terms of a series of Chebyshev Polynomials.

Least Squares Approximation in terms of Legendre Polynomials.

In this $a = -1$, $b = 1$, $w(x) = 1$

$$\phi_k(x) = P_k(x) \quad \text{and} \quad \|P_k\|^2 = \frac{2}{2k+1}$$

The nth degree polynomial approximation of given function $f(x)$ in terms of Legendre polynomials is given by

$$f(x) \approx \sum_{k=0}^{n} a_k P_k(x) \quad \text{where}$$

$$a_k = \frac{2k+1}{2} \int_{-1}^{1} f(x) P_k(x)\, dx$$

$$a_k = \frac{\int_{-1}^{1} 1 \cdot f(x) \cdot P_k(x)\, dx}{\int_{-1}^{1} P_k^2(x)\, dx = \frac{2}{2k+1}}$$

Least Squares Approximation in terms of Chebychev Polynomials.

$a=-1 \quad b=1, \quad w(x) = \frac{1}{\sqrt{1-x^2}}, \quad \phi_k(x) = T_k(x),$

$$\| \phi_k(x) \|^2 = \| T_k(x) \|^2 = \begin{cases} \pi & \text{when } k=0 \\ \frac{\pi}{2} & \text{when } k \neq 0 \end{cases}$$

Approximation of $f(x)$ in terms of Chebychev polynomials of degree $\leq n$ is given by

$$f(x) = \sum_{k=0}^{n} a_k T_k, \quad \text{where} \quad a_0 = \frac{1}{\pi} \int_{-1}^{1} \frac{f(x_0)}{\sqrt{1-x^2}}\, dx.$$

$$a_k = \frac{2}{\pi} \int_{-1}^{1} \frac{f(x_a) \cdot T_k(x)}{\sqrt{1-x^2}}\, dx \qquad k = 1, 2 \ldots$$

How to generate

$[a,b] \longrightarrow [-1,1]$

Put $x = \frac{1}{2}\left[(a+b) + (b-a)\,t\right]$ } $f(x)$

$f(x) = F(t)$

March 17, Class 25

Problem Find the least square approximation to $f(x) = x^2$ on the interval $[0, 1]$ by a straight line (or linear / first degree polynomial) using Legendre Polynomials

Sol Put $x = \frac{1}{2}[(1+0) + (1-0)t]$

$[0,1] \to [-1,1]$
when $x=0$ $t=-1$
$x=1$ $t=1$

$$x = \frac{1+t}{2}$$

$$f(x) = x^2 = \left(\frac{1+t}{2}\right)^2 = F(t)$$

Recall
$P_0(x) = 1$
$P_1(x) = x$
$P_2(x) = \frac{1}{2}(3x^2-1)$
$P_3(x) = \frac{1}{3}(5x^3-3x)$

Let $f(t) \simeq a_0 P_0(t) + a_1 P_1(t)$

$$a_k = \frac{2k+1}{2} \int_{-1}^{1} F(t) P_k(t)\, dt \qquad t=0,1\ldots$$

$$a_0 = \frac{1}{2} \int_{-1}^{1} \left(\frac{1+t}{2}\right)^2 \cdot 1 \, dt = \frac{1}{8} \int_{-1}^{1} (1+2t+t^2)\, dt$$

$$= \frac{2}{8} \int_{0}^{1} (t^2+1)\, dt = \frac{1}{4}\left[\frac{t^3}{3} + t\right]_0^1 = \frac{1}{3}$$

$$a_1 = \frac{3}{2} \int_{-1}^{1} \left(\frac{1+t}{2}\right)^2 \cdot t \, dt = \frac{3}{8} \int_{-1}^{1} (t+2t^2+3t^3)\, dt$$

$$= \frac{3 \cdot 2}{8} \int_{0}^{1} 2t^2\, dt = \frac{3}{2}\left[\frac{t^3}{3}\right]_0^1 = \frac{1}{2}$$

$$F(t) = a_0 + a_1 t$$

$$= \frac{1}{3} + \frac{1}{2} t$$

But $t = 2x - 1$ $\quad$ (from (1) : $x = \frac{1+t}{2}$)

$$\therefore f(x) = \frac{1}{3} + \frac{1}{2}(2x-1)$$

$$= x - \frac{1}{6}$$

H.W. Find least squares approximation of second degree (or parabola or quadratic) using Legendre Polynomials.

Problem Find the least squares straight line (or linear or first degree) approximation to $f(x) = x^2$ on $[0,1]$ using Chebychev polynomials

Sol $\quad [0,1] \xrightarrow{?} [-1,1]$

Put $\quad x = \frac{1}{2}\left[(b+a) + t(b-a)\right] = \frac{1+t}{2}$

$$f(x) = \left(\frac{1+t}{2}\right)^2 = F(t)$$

$$F(t) = a_0 T_0(t) + a_1 T_1(t)$$

$a_0 = ?$ $\quad a_1 = ?$

$$a_0 = \frac{1}{\pi} \int_{-1}^{1} \frac{F(t)}{\sqrt{1-t^2}}\, dt \qquad \left[\begin{array}{l} T_0 = 1 \\ T_1(t) = t \end{array}\right]$$

$$a_1 = \frac{2}{\pi} \int_{-1}^{1} \frac{F(t)\, T_1(t)}{\sqrt{1-t^2}}\, dt$$

$$a_0 = \frac{1}{\pi}\int_{-1}^{1} \frac{1}{\sqrt{1-t^2}}\left(\frac{1+t}{2}\right)^2 dt$$

$$= \frac{1}{4\pi}\int_{-1}^{1} \frac{1+2t+t^2}{\sqrt{1-t^2}}\, dt$$

$$= \frac{1}{4\pi}\int_{-1}^{1} \frac{2t}{\sqrt{1-t^2}}\, dt + \frac{1}{4\pi}\int_{-1}^{1} \frac{1+t^2}{\sqrt{1-t^2}}\, dt$$

$\because$ odd func.

$$= \frac{1}{4\pi}\int_{-1}^{1} \frac{1+t^2}{\sqrt{1-t^2}}\, dt = \frac{1}{2\pi}\int_{0}^{1} \frac{1+t^2}{\sqrt{1-t^2}}\, dt$$

Put $t = \sin\theta$

$dt = \cos\theta\, d\theta$

$$= \frac{1}{2\pi}\int_{0}^{\frac{\pi}{2}} \frac{(1+\sin^2\theta)}{\sqrt{1-\sin^2\theta}} \cos\theta\, d\theta$$

$$= \frac{1}{2\pi}\int_{0}^{\frac{\pi}{2}} (1+\sin^2\theta)\, d\theta$$

$$= \frac{1}{2\pi}\int_{0}^{\frac{\pi}{2}} 1 + \left(\frac{1-\cos 2\theta}{2}\right) d\theta$$

$$= \frac{1}{4\pi}\int_{0}^{\frac{\pi}{2}} (3 - \cos 2\theta)\, d\theta$$

$$= \frac{1}{4\pi}\left[3\theta - \frac{\sin 2\theta}{2}\right]_0^{\frac{\pi}{2}}$$

$$= \frac{1}{4\pi}\left(\left[3\frac{\pi}{2}\right] - 0\right) = \boxed{\frac{3}{8}}$$

$$a_1 = \frac{2}{\pi} \int_{-1}^{1} \frac{1}{\sqrt{1-t^2}} \left(\frac{1+t}{2}\right)^2 t \, dt$$

$$= \frac{2}{4\pi} \int_{-1}^{1} \frac{t + 2t^2 + t^3}{\sqrt{1-t^2}} \, dt \quad \text{(odd)}$$

$$= \frac{4}{4\pi} \int_{-1}^{1} \frac{t^2}{\sqrt{1-t^2}} \, dt = \frac{2}{\pi} \int_{0}^{1} \frac{t^2}{\sqrt{1-t^2}} \, dt$$

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

$$F(t) = \frac{3}{8} + \frac{1}{2} T_1(t) = \frac{3}{8} + \frac{1}{2} t$$

Put $t = 2x - 1$

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

$$= x - \frac{1}{8}$$

[Earlier we got $\left(x - \frac{1}{6}\right) = f(x)$]

H.W. Find the least square approximation of second degree to above problem using Chebychev Polynomials.

Equi - oscillation

A continuous function E is said to equi-oscillate on n points of $[a, b]$ if there exist n points x_i, with

$$a \le x_1 \le x_2 \dots \le x_n \le b \quad \text{such that}$$

$$|E(x_i)| = \max_{a \le x \le b} |E(x)| \qquad i = 1, 2 \dots, n$$

and

$$E(x_i) = -E(x_{i+1}) \qquad i = 1, 2 \dots, n-1$$

- Minimax Approximation / Uniform approximation / Chebyshev Approximation

$$\text{Error} = E_r(f) = \inf_{p \in P_n} \max_{a \le x \le b} |f(x) - P_n(x)|$$

inf = infimum means greatest lower bound
(Infimum may not belong to the set)

→ If f is continuous on (a, b)
(Maximum of absolute error is small)

Theorem: $f \in C[a, b]$, $p_n \in P_n$ = { set of all algebraic polynomials of degree $\le n$ }

is such that

$$E_n = \max |f(x) - p_n(x)| \le \max |f(x) - q_n(x)|$$

$\forall\ p_n(x), q_n(x) \in P_n$

or

$$E_n = \|f - P_n\|_\infty \le \|f - q_n\|_\infty$$

i.e. P_n is the best uniform approximation
(∵ error in this is least)

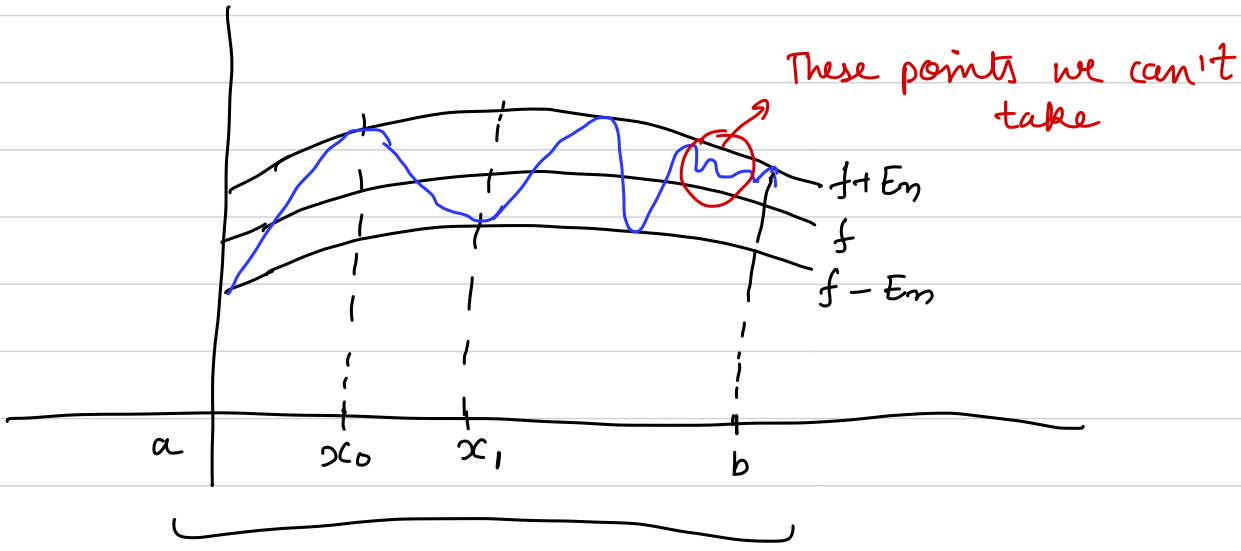
iff there exist at least $(n+2)$ points

$$a \le x_0 < x_1 \ \ldots \ < x_{n+1} \le b$$

such that

$$f(x_i) - P_n(x_i) = (-1)^i \sigma \, E_n \qquad i = 0, 1, \ldots, n+1$$

$\sigma = 1$ or -1 depending on f and n.



Equioscillation of $P_n(x)$

$$\mathcal{E}(x_i) = \max_{a \le x \le b} |\mathcal{E}_n(x)| = \pm E_n$$

$$\mathcal{E}(x_i) = -\mathcal{E}(x_{i+1})$$

Apply Rolle's theorem

$$\mathcal{E}'(\xi_i) = 0 \qquad (\because \ x_i \text{ and } x_{i+2} \text{ have same value})$$

Problem Obtain Uniform / Minimax / Chebychev linear approximation to the function $f(x) = x^3$ on $[0,1]$

Sol Let $P_1(x) = a + bx$

$x_0 = 0 \qquad x_1 = \alpha \text{ (say)} \qquad x_2 = 1$

[$\because n+2$ points are required]

We have $\varepsilon_1(x) = f(x) - P_1(x)$

$= x^3 - a - bx$

Using the property $\varepsilon_n(x_i) = -\varepsilon_n(x_{i+1})$

$\varepsilon_1(0) = -\varepsilon_1(\alpha) \Rightarrow \varepsilon_1(0) + \varepsilon_1(\alpha) = 0$ — (1)

$\varepsilon_1(\alpha) = -\varepsilon_1(1) \Rightarrow \varepsilon_1(\alpha) + \varepsilon_1(1) = 0$ — (2)

and $\varepsilon_1'(\alpha) = 0$ (by Rolle's theorem) — (3)

i.e. $3x^2 - b = 0$ at $x = \alpha$

$3\alpha^2 - b = 0$

$\alpha = \pm\sqrt{\frac{b}{3}}$ discard -ve value $\because \alpha \in [0,1]$

$\alpha = \sqrt{\frac{b}{3}}$

From (1) $-a + \alpha^3 - a - b\alpha = 0$

(2) $\alpha^3 - a - b - \alpha + 1 - a - b = 0$

(3) $3\alpha^2 - b = 0$ or $b = 3\alpha^2$

System of non linear eqns.

⇓

$$\alpha^3 - b\alpha - 2a = 0$$
$$\alpha^3 - b(\alpha+1) - 2a + 1 = 0$$

i.e. $3\alpha^2 - b = 0$.

Substitute in (1) $\Rightarrow \alpha^3 - 3\alpha^2 - 2a = 0$ i.e. $-2\alpha^3 - 2a = 0$

$$a = -\alpha^3$$

Substitute a and b in (2)

$$\alpha^3 - (\alpha+1)b - 2a + 1 = 0$$
$$\alpha^3 - \alpha b - b - 2a + 1 = 0$$
$$\alpha^3 - \alpha - 3\alpha^2 - 3\alpha^2 + 2\alpha^3 + 1 = 0$$

$$3\alpha^2 = 1$$

$$\alpha = \pm \frac{1}{\sqrt{3}}$$

$$\Rightarrow \alpha = \frac{1}{\sqrt{3}}$$

$$\therefore a = -\alpha^3 = -\left(\frac{1}{\sqrt{3}}\right)^3 = -\frac{\sqrt{3}}{9}$$

$$b = 3\alpha^2 = 3 \times \frac{1}{3} = 1$$

$\therefore$ Chebyshev linear approximation is

$$= -\frac{\sqrt{3}}{9} + x$$

March 21, Class 26

f is continuous on $[-1, 1]$, then f is uniquely expressed in terms of Chebychev series

$$f(x) = a_0 T_0 + a_1 T_1 + \cdots + a_n T_n + \cdots$$
$$= \sum a_i T_i$$

$$\int_{-1}^{1} \frac{f(x)\, T_i(x)}{\sqrt{1-x^2}}\, dx = \int_{-1}^{1} \frac{a_i\, T_i^2(x)}{\sqrt{1-x^2}}\, dx$$

(due to orthogonality of T_i)

$$= a_i \frac{\pi}{2} \quad \text{if } i \neq 0$$

or $$a_i = \frac{2}{\pi} \int_{-1}^{1} \frac{f(x)\, T_i(x)}{\sqrt{1-x^2}}\, dx$$

and $$a_0 = \frac{1}{\pi} \int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}}\, dx$$

then $f(x) = \sum_{i=0}^{\infty} a_i T_i$ is Chebychev series expansion $-$ (*)

Near Minimax Approximation

Suppose $P_n(x) = \frac{a_0}{2} + \sum_{i=1}^{n} a_i T_i(x)$

Let $f(x) = \frac{a_0}{2} + \sum_{i=1}^{\infty} a_i T_i(x)$ where $f(x) \in [-1, 1]$

be Chebychev series expansion

$f(x) - P_n(x)$

Also $f(x) - \left(\sum_{i=1}^{n} a_i T_i(x) + \frac{a_0}{2} \right) = \sum_{i=n+1}^{\infty} a_i T_i(x)$

$\simeq a_{n+1} T_{n+1}(x)$ — (A)

If $a_{n+1} \neq 0$ and if other coefficients a_j, $(j > n+1)$ rapidly converge to zero.

If $a_{n+1} = 0$, the first non-zero term in the error will still oscillate on $(n+2)$ points

From the decomposition of $T_{n+1}(x)$,

$$|T_{n+1}(x)| \leq 1 \quad \text{— (B)}$$

Also the points $x_j = \cos\left(\frac{\pi j}{n+1}\right)$, $j = 0, 1, 2 \ldots, n+1$

We have $T_{n+1}(x) = (-1)^j$

The bound (B) is attained at exactly $(n+2)$ points, the minimum possible times. Applying this to (A), the term $a_{n+1} T_{n+1}$ has exactly $(n+2)$ relative maxima and minima, all are equal in magnitude

From Chebyshev equioscillation theorem, we would therefore expect $P_n(x)$ to be nearly equal to minima approximation

Moral: The partial sum $P_n(x) = \frac{a_0}{2} + \sum_{i=1}^{n} a_i T_i(x)$ is very nearly the solution of the problem

$$\min \left(\max_{-1 \leq x \leq 1} |f(x) - P_n(x)| \right) \quad \text{i.e.}$$

the partial sum $P_n(x)$ is nearly the least Uniform approximation

Lancoz Economization

Suppose $f(x) = \sum_{i=1}^{n} a_i T_i(x) + \frac{a_0}{2}$ is Chebychev Series expansion for a continuous function $f(x)$ on $[-1, 1]$. Then partial sum $P_n(x) \approx \frac{a_0}{2} + \sum_{i=1}^{n} a_i T_i(x)$ is a good uniform approximation to $f(x)$ in the sense -

$$\max_{-1 \le x \le 1} |f(x) - P_n(x)| \le |a_{n+1}| + |a_{n+2}| + \dots$$
$$\le \varepsilon$$

Given ε, it is possible to find the number of terms that should be retained in (*)
This process is known as Lancoz Economization
Replacing each $T_i(x)$ by its polynomial function and rearranging the terms, we get the economized polynomial approximation.

Problem (Using Chebychev Polynomials) Find nearly best uniform approximation of degree 3 or less to x^4 on $[-1,1]$
i.e. near minimax approximation

Expressing x^4 in terms of Chebychev Polynomials, we get

$$x^4 = \frac{1}{8}[T_4 + 4T_2 + 3T_0]$$

$$x^4 = \frac{3}{8}T_0 + \frac{1}{2}T_2 + \frac{1}{8}T_4$$

Since T_4 is a polynomial of degree 4, we approximate

$$f(x) = x^4 \text{ by } \frac{3}{8}T_0 + \frac{1}{2}T_2$$

Therefore we have, $x^4 - \left(\frac{3}{8}T_0 + \frac{1}{2}T_2\right) = \frac{1}{8}T_4$

The uniform polynomial approximation of degree 3 or less to x^4 is

$$\frac{3}{8}T_0 + \frac{1}{2}T_2 = x^2 - \frac{1}{8}$$

$$\left[\frac{1}{2}T_2 = \frac{1}{2}(2x^2-1) = x^2 - \frac{1}{2}\right]$$

$$T_0 = 1$$

and the error of approximation on $[-1, 1]$ is

$$\max_{-1\le x\le 1} \left| x^4 - \left(\frac{3}{8}T_0 + \frac{1}{2}T_1\right)\right| = \max_{-1\le x\le 1} \left|\frac{1}{8}T_4(x)\right| = \frac{1}{8}$$

NOTE : In the problem, we have approximated a polynomial by a polynomial of lower degree polynomial. For a general function, we first express the function as a power series in x and write each term in terms of Chebychev polynomials.

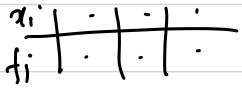
Post Minor 2

Numerical Differentiation

Numerical differentiation methods can be obtained based on –

1. Interpolation
2. Finite difference operator
3. Undetermined coeff.

Methods based on interpolation



$$f(x) \simeq P_n(x)$$

$$f^{(r)}(x) = P_n^{(r)}(x) \rightarrow r^{th} \text{ derivative of } f(x)$$

Error

$E^{(r)}(x) = f^{(r)}(x) - P_n^{(r)}(x)$ is called the error in approximation of r^{th} derivative at any point.

Order

If $|E^{(r)}(x)| \leq C h^p$

where C is constant independent of h, then the numerical differentiation is said to be of order 'p'

Uniform nodal points

Derivative using Newton's Forward Difference formula.

$$f(x) = f(s) = f_0 + s\Delta f_0 + \frac{s(s-1)}{2!}\Delta^2 f_0 + \cdots$$

where $s = \frac{x - x_0}{h} \Rightarrow ds = \frac{dx}{h}$ or $\frac{ds}{dx} = \frac{1}{h}$

$$f'(x) = f'(s)\frac{ds}{dx}$$

$$f'(x) \approx \frac{d}{ds}(f(s)) \frac{ds}{dx}$$

$$= \frac{d}{ds}\left[f_0 + s\Delta f_0 + \frac{s(s-1)}{2}\Delta^2 f_0 + \cdots \right] \times \frac{1}{h}$$

$$= \frac{1}{h}\left[\Delta f_0 + \frac{2s-1}{2}\Delta^2 f_0 + \cdots \right]$$

Put $x = x_0$ then $s = 0$

$$f'(x_0) = \frac{1}{h}\left(\Delta f_0 - \frac{1}{2}\Delta f_0^2 + \frac{1}{3}\Delta f_0^3 \cdots \right]$$

Similarly, approximation of second order derivative $f''(x)$ at x_0 can be obtained as -

$$f''(x_0) = \frac{1}{h^2}\left[\Delta^2 f_0 - \Delta^3 f_0 + \frac{11}{12}\Delta^4 f_0 \cdots \right]$$

March 22, Class 27

Derivative using Newton's Backward Difference formula.

$$f(x) = F_s = f_n + s\nabla f_n + \frac{s(s+1)}{2}\nabla^2 f_n + \cdots$$

when $\frac{x - x_n}{h} = s \qquad \frac{ds}{dx} = \frac{1}{h}$

$$f'(x) = \frac{d}{ds}\left[f_n + s\nabla f_n + \frac{s(s+1)}{2}\nabla^2 f_n + \cdots \right] \frac{ds}{dx}$$

$$= \frac{1}{h}\left[\nabla f_n + \frac{2s+1}{2}\nabla^2 f_n + \frac{3s^2+6s+2}{6}\nabla^3 f_n + \cdots \right]$$

At nodal points,

Put $x = x_n$, then $s = 0$

$$f'(x_n) \simeq \frac{1}{h}\left[\nabla f_n + \frac{1}{2}\nabla^2 f_n + \frac{1}{3}\nabla^3 f_n + \cdots\right]$$

$$f''(x_n) \simeq \frac{1}{h^2}\left[\nabla^2 f_n + \nabla^3 f_n + \frac{11}{12}\nabla^4 f_n + \cdots\right]$$

Non uniform Nodal Points

x_i	x_0	-	-	x_n
f_i	f_0	-	-	f_n

$$P_n(x) = \sum_{k=0}^{n} l_k(x) f_k \quad -(1)$$

$l_k(x)$ is Lagrange's Polynomial

$$l_k(x) = \frac{\Pi(x)}{(x - x_k)\,\Pi'(x_k)}$$

$$f_k = f(x_k)$$

$$\Pi(x) = (x - x_0)(x - x_1) \cdots (x - x_n)$$

Error of approximation

$$E(x) = f(x) - P_n(x) = \frac{\Pi(x)}{(n+1)!} f^{n+1}(\xi) \quad -(2)$$

Differentiate (1) and (2) w.r.t. x.

$$P_n'(x) = \sum_{k=0}^{n} l_k'(x) f_k$$

and $$E'(x) = \frac{\Pi'(x) f^{n+1}(\xi)}{(n+1)!} + \frac{\Pi(x)}{(n+1)!}\frac{d}{dx}\{f^{n+1}(\xi)\}$$

[ξ depends on x]

$[\pi(x_k) = 0]$

$$E_n'(x_k) = \frac{\pi'(x_k)}{(n+1)!} f^{n+1}(\xi) \quad \text{provided } \frac{d}{dx}\left(f^{n+1}(\xi)\right) \text{ is bounded}$$

$[E(x) = 0$ at nodal points$]$

Difference Approximation to derivatives

First order derivative

Forward difference quotient

$$f(x+h) = f(x) + h f'(x) + \frac{h^2}{2!} f''(\xi) \qquad x < \xi < x+h$$

$$\Rightarrow f'(x) = \frac{1}{h}\left[f(x+h) - f(x)\right] - \frac{h}{2} f''(\xi)$$

$$f'(x) \simeq \frac{f(x+h) - f(x)}{h} \qquad \Rightarrow \text{quotient} \quad -(*)$$

with truncation error (or error of approximation) $\boxed{\frac{-h}{2} f''(\xi)}$ $\Rightarrow$ Remainder

$(*)$ is called Forward difference approximation

Backward Difference Quotient

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

$$\Rightarrow f'(x) = \frac{f(x) - f(x-h)}{h} + \frac{h}{2} f''(\xi)$$

$$f'(x) \simeq \frac{f(x) - f(x-h)}{h} \qquad -(**) \qquad \text{or } f'(x_k) = \frac{f(x_k) - f(x_{k-1})}{h}$$

$\nearrow f(x-h)$

$(**)$ is called Backward difference approximation

Central Difference Quotient

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \frac{h^3}{3!}f'''(\xi_1)$$

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2!}f''(x) - \frac{h^3}{3!}f'''(\xi_2)$$

$$f(x+h) - f(x-h) = 2hf'(x) + \frac{h^3}{3!}\left(f'''(\xi_1) + f'''(\xi_2)\right)$$

[suppose f''' is continuous]

$$f'''(\xi) = \frac{f'''(\xi_1) + f'''(\xi_2)}{2}$$

for some $\xi \in (\xi_1, \xi_2)$

$$f(x+h) - f(x-h) = 2hf'(x) + \frac{h^3}{3!}\,2f'''(\xi)$$

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \frac{h^2}{3!}f'''(\xi)$$

error

Error is of order h^2

We can also derive, further higher order derivatives using more points

$$f'(x) \simeq \frac{-f(x+2h) - 8f(x+h) - 8(x-h) + f(x+2h)}{12h}$$

Fourth order approximation

Higher order derivatives

Second order derivatives

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \frac{h^3}{3!}f'''(x) + \frac{h^4}{4!}f''''(\xi_1)$$

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2!}f''(x) - \frac{h^3}{3!}f'''(x) + \frac{h^4}{4}f''''(\xi_2)$$

(+)

$$f(x+h) + f(x-h) = 2f(x) + h^2 f''(x) + \frac{h^4}{4!}\left(f^{(4)}(\xi_1) + f^4(\xi_2)\right)$$

$$f^4(\xi) = \frac{f^4(\xi_1) + f^4(\xi_2)}{2}$$

$$\Rightarrow f''(x) \simeq \frac{f(x+h) + f(x-h) - f(x)}{h^2} + \frac{h^2}{12}f^{(4)}(\xi)$$

$$f''(x) \simeq \frac{f(x+h) - f(x) + f(x-h)}{h^2}$$

With the Truncation error $\frac{h^2}{12}f^{(4)}(\xi)$

Error is of order h^2

METHODS BASED ON UNDETERMINED COEFF.

To approximate $f''(x)$

$$f''(x_k) \approx A f(x_k+h) + B f(x_k) + C f(x_k-h)$$

with A, B, C unspecified constants [only 3 points]

Replace $f(x_k+h)$ and $f(x_k-h)$ by Taylor Polynomial approximation

$$C \times f(x_k-h) \approx f(x_k) - h f'(x_k) + \frac{h^2}{2!} f''(x_k) - \frac{h^3}{6} f'''(x_k) + \frac{h^4}{24} f''''(\xi_1)$$

$$A \times f(x_k+h) = f(x_k) + h f'(x_k) + \frac{h^2}{2!} f''(x_k) + \frac{h^3}{6} f'''(x_k) + \frac{h^4}{24} f''''(\xi_2)$$

(*)

Substitute these eqns in formula and collect the common powers of h, we get

$$f''(x_k) = (A+B+C) f(x_k) + h(A-C) f'(x_k) + \frac{h^2}{2}(A+C) f''(x_k) + \frac{h^3}{6}(A-C) f'''(x_k) + \frac{h^4}{24}\left(A f''''(\xi_1) + C f''''(\xi_2)\right)$$

It is necessary to require —

$$A+B+C=0 \quad \text{(1)} : \text{coeff of } f(x_n)$$

$$h(A-C)=0 \quad \text{(2)} : \text{coeff of } f'(x_n)$$

$$\frac{h^2}{2}(A+C)=1 \quad \text{(3)} : \text{coeff of } f''(x_n)$$

(2) $\Rightarrow A = C$

(3) $\Rightarrow A = \frac{1}{h^2} = C$

(1) $\Rightarrow B = -\frac{2}{h^2}$

This determines, $f''(x_n) \approx \frac{f(x_n+h) - 2f(x_n) + f(x_n-h)}{h^2}$

$$f''(x_n) = \frac{f(x_n+h) - 2f(x_n) + f(x_n-h)}{h^2} + \frac{h^4}{24} \cdot \frac{1}{h^2} \cdot 2f''''(\xi)$$

$$\text{Error} = \frac{h^2}{12} f''''(\xi)$$

$$f(x_i) = f_i + \xi_i \quad (\xi_i \rightarrow \text{error})$$

$f(x_i)$ ↓ Exact value

f_i ↓ Tabulated value

x_i	x_1	-	-
f_i	f_1	-	-

'h' → optimum step size.

i) $|RE| = |TE|$

ii) $|RE| + |TE|$: Minimum

RE : Rounding error

TE : Truncation error

March 28, Class

Problem

For the method $f'(x_0) = \frac{f(x_1) - f(x_0)}{h} - \frac{h}{2} f''(\xi), \quad x_0 < \xi < x_1$

find the optimal value of h using the criteria –

i) $|RE| = |TE|$

ii) $|RE| + |TE|$ = Minimum.

Sol If $\varepsilon_0 \to$ Rounding error in f_0

$\varepsilon_1 \to$ Rounding error in f_1

then we have

$$f'(x_0) = \frac{f_1 - f_0}{h} + \underbrace{\frac{\varepsilon_1 - \varepsilon_0}{h}}_{RE} - \underbrace{\frac{h}{2} f''(\xi)}_{TE}$$

(Round off error) (Truncation error)

If we take $\varepsilon = \max\{|\varepsilon_0|, |\varepsilon_1|\}$

$$M_2 = \max_{x_0 \le x \le x_1} f''(x)$$

then we get $|RE| = \frac{2\varepsilon}{h}$, $|TE| \le \frac{h}{2} M_2$

For the criteria $|RE| = |TE|$

$$\frac{2\varepsilon}{h} = \frac{h}{2} M_2 \Rightarrow h^2 = \frac{4\varepsilon}{M_2}$$

$h_{optimum}$ or $h_{opt} = 2\sqrt{\frac{\varepsilon}{M_2}}$

For this value of h_{opt},

$$|RE| = |TE| = \frac{2\varepsilon}{h} \qquad \frac{2\varepsilon}{2\sqrt{\frac{\varepsilon}{M_2}}} = \sqrt{\varepsilon M_2}$$

For the criteria $|RE| + |TE| = \text{Minimum}$

$$= \frac{2\varepsilon}{h} + \frac{h M_2}{2} = \text{Min}$$

Using the necessary cond.,

$$g'(h) = 0$$

$$-\frac{2\varepsilon}{h^2} + \frac{M}{2} = 0$$

$$h^2 = \frac{4\varepsilon}{M_2}$$

$$h_{opt} = 2\sqrt{\frac{\varepsilon}{M_2}}$$

For this h_{opt},

$$|RE| + |TE| = \frac{2\varepsilon}{h} + \frac{h}{2} M_2$$

$$= \frac{2\varepsilon}{2\sqrt{\frac{\varepsilon}{M_2}}} + \frac{2}{2}\sqrt{\frac{\varepsilon}{M_2}}\, M_2$$

$$= \sqrt{\varepsilon M_2} + \sqrt{\varepsilon M_2}$$

$$= 2\sqrt{\varepsilon M_2}$$

Note : h_{opt} is same in both cases is a matter of chance. In general, h_{opt} may be different for different criteria

✱ The notation $O(h^{k+1})$ is conventionally used to stand for "a sum of terms of order h^{k+1} and higher".

Extrapolation Methods.

The technique of combining 2 computed values obtained by using the same formula with two different step sizes, to obtain a higher order method is called EXTRAPOLATION METHOD or

RICHARDSON'S EXTRAPOLATION

Let $g(x)$ denotes the approximate value of 'g' obtained using a method of order 'p', with step size 'h' and $g(qh)$ denotes the value of g obtained by using the same method of order p using step size 'qh'

$$q^p \times \quad g = g(h) + ch^p + O(h^{p+1})$$

$$\ominus \quad g = g(qh) + c q^p h + O(h^{p+1})$$

$$g(q^p - 1) = q^p g(h) - g(qh) + O(h^{p+1})$$

or

$$g = \frac{q^p g(h) - g(qh)}{q^p - 1} + O(h^{p+1})$$

$$g = g^{(1)}(h) + O(h^{p+1})$$

$$\boxed{g^{(1)}(h) = \frac{q^p g(h) - g(qh)}{q^p - 1}}$$

Thus we obtain $\boxed{g^{(1)}(h) = g + O(h^{p+1})}$

which is of order $(p+1)$

[The $O(h^{p+1})$, O, takes care of sign anomalies.]

- If truncation error associated with the method known as power series in h, then by repeating the extrapolation procedure a number of times, we can obtain the method of any arbitrary order.

- The application of this procedure becomes simplified when the step length form a geometric sequence. For simplicity, we take $q = \frac{1}{2}$

For example, take the method

$$f'(x_0) \approx \frac{f(x_0+h) - f(x_0-h)}{2h} \quad \text{(Central difference approximation for first derivative)}$$

The truncation error of the above method is obtained

$$E(x) = c_1 h^2 + c_2 h^4 + c_3 h^6 + \ldots$$

where $c_1, c_2, c_3 \ldots$ are constants independent of h.

Let $g = f(x_0)$ be the quantity which is to be obtained and $g\left(\frac{h}{2^r}\right)$ denotes the approximate value of g obtained using $(*)$ with step length $\frac{h}{2^r}$, $r = 0, 1, 2 \ldots$

Then we have

$$g(h) = g(x) + c_1 h^2 + c_2 h^4 + c_3 h^6 + \cdots \quad -(1)$$

$$g\left(\frac{h}{2}\right) = g(x) + \frac{c_1 h^2}{4} + \frac{c_2 h^4}{16} + \frac{c_3 h^6}{64} + \cdots \quad -(2)$$

$$g\left(\frac{h}{2^2}\right) = g(x) + \frac{c_1 h^2}{16} + \frac{c_2 h^4}{256} + \frac{c_3 h^6}{4096} + \cdots \quad -(3)$$

[Using the same procedure twice as compared to once, as did earlier]

Eliminating c_1 from (1) and (2): we obtain

$4 \times (2) - (1) \Rightarrow$

$$g^{(1)}(h) = \frac{4 g\left(\frac{h}{2}\right) - g(h)}{4 - 1} = \frac{1}{3}\left[\begin{array}{l} (4-1)\, g(x) \\ -\frac{3}{4} c_2 h^4 \\ -\frac{15}{16} c_3 h^6 + \cdots \end{array} \right]$$

$$= g(x) - \frac{1}{4} c_2 h^4 - \frac{5}{16} c_3 h^6 \quad -(4)$$

[We have made it seem like $g^{(1)}(h)$ has been obtained from method of order 4 $O(h^4)$]

Eliminating c_1 from (2) and (3), we obtain

$$g^{(1)}\left(\frac{h}{2}\right) = \frac{4 g\left(\frac{h}{4}\right) - g\left(\frac{h}{2}\right)}{3}$$

$$= g(x) - \frac{1}{64} c_2 h^4 - \frac{5 c_3 h^6}{1024} + \cdots \quad -(5)$$

Then $g^{(1)}(h)$, $g^{(1)}(\frac{h}{2})$, .. are of $O(h^4)$ approximation to $g(x)$

Eliminating c_2 from (4) and (5), we obtain

$$g^{(2)}(h) = \frac{4^2 g^{(1)}(\frac{h}{2}) - g^{(1)}(h)}{4^2 - 1}$$

$$g^{(2)}(h) = g(x) + \frac{1}{64} c_3 h^6 + \cdots$$

(which is of order $O(h^6)$ approximation)

Thus the successive higher order results can be obtained from the formula.

$$g^{(m)}(h) = \frac{a^m g^{(m-1)}(\frac{h}{2}) - g^{(m-1)}(h)}{4^m - 1}, \quad m = 1, 2 \ldots$$

where $g^{(0)}(h) = g(h)$

Step size h	Second $O(h^2)$	Fourth $O(h^4)$	Sixth $O(h^6)$	$O(h^8)$
h	$g(h)$	$g^{(1)}(h)$		
$\frac{h}{2}$	$g(\frac{h}{2})$	$g^{(1)}(\frac{h}{2})$	$g^{(2)}(h)$	$g^{(3)}(h)$
$\frac{h}{2^2}$	$g(\frac{h}{2^2})$	$g^{(1)}(\frac{h}{2^2})$	$g^{(2)}(\frac{h}{2})$	
$\frac{h}{2^3}$	$g(\frac{h}{2^3})$			

$$\left| g^{(k)}(h) - g^{(k+1)}\left(\frac{h}{2}\right) \right| < \epsilon$$

March 31

Numerical Integration
(Numerical Quadrature)

x_i	x_0 - -	x_n
f_i	f_0	f_n

$\rightarrow \quad x_0 < x_1 < - - - < x_n$

$$\int_{x_0}^{x_n} f(x)\,dx = ?$$

More generally, $\int_a^b w(x)\, f(x)\, dx = ?$

Suppose if x_i are equispaced $x_i - x_{i-1} = h$

(Newton's forward)

$$f(x) = f_s = f_0 + s\,\Delta f_0 + \frac{s(s-1)}{2!}\Delta^2 f_0 + \frac{s(s-1)(s-2)}{3!}\Delta^3 f_0$$

$$s = \frac{x - x_0}{h} \quad \Rightarrow \quad dx = h\,ds$$

$$\left[\begin{array}{l} \text{at } x = x_0, \; s = 0 \\ \text{at } x = x_n, \; s = n \end{array}\right]$$

$$\int_{x_0}^{x_n} f(x)\,dx = \int_0^n f_s\, h\, ds$$

$$= \int_0^n \left[f_0 + s\Delta f_0 + \frac{s(s-1)}{2!}\Delta^2 f_0 + \cdots \right] h\, ds$$

$$\int_{x_0}^{x_n} f(x)\,dx = h\left[\Delta f_0 + \frac{\Delta^2}{2}\Delta f_0 + \frac{1}{2}\left(\frac{\Delta^3}{3} - \frac{\Delta^2}{2}\right)\Delta^2 f_0 + \cdots\right]_{\Delta=0}^{n}$$

Trapezoidal Rule

for $n = 1$

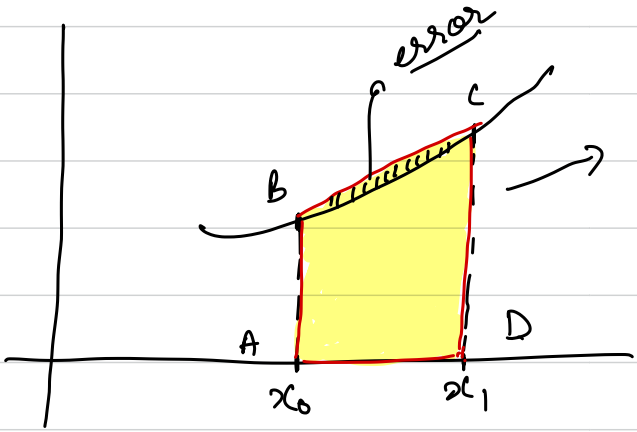
$$\int_{x_0}^{x_1} f(x)\,dx = h\left[\Delta f_0 + \frac{\Delta^2}{2}\Delta f_0\right]_{\Delta=0}^{1}$$

$\Delta^2 f_0$, $\Delta^3 f_0$ ... etc all are zero.

$$= h\left[f_0 + \frac{1}{2}\Delta f_0\right] = h\left[f_0 + \frac{1}{2}(f_1 - f_0)\right]$$

$$= \frac{h}{2}\left[f_0 + f_1\right]$$

$$\int_{x_0}^{x_1} f(x)\,dx = \frac{h}{2}\left[f_0 + f_1\right]$$ → Trapezoidal Rule.



→ Area under the curve.

Simpson's $\frac{1}{3}$rd Rule.

Let $n = 2$ (n = no. of intervals)

(i) takes the form.

$$\int_{x_0}^{x_2} f(x)\,dx = h\left[s f_0 + \frac{s^2}{2}\Delta f_0 + \frac{1}{2}\left(\frac{s^3}{3} - \frac{s^2}{2}\right)\Delta^2 f_0\right]_{s=0}^{s=2}$$

$$= h\left[2f_0 + \frac{4}{2}\Delta f_0 + \frac{1}{2}\left(\frac{8}{3} - \frac{4}{2}\right)\Delta^2 f_0\right]$$

$$= h\left[2f_0 + 2(f_1 - f_0) + \frac{1}{3}(f_2 - 2f_1 + f_0)\right]$$

$$= h\left[\frac{1}{3}f_0 + \left(2 - \frac{2}{3}\right)f_1 + \frac{1}{3}f_2\right]$$

$$= \frac{h}{3}\left[f_0 + 4f_1 + f_2\right]$$

$$\boxed{\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{3}\left[f_0 + 4f_1 + f_2\right]}$$

Composite Trapezoidal Rule.

$$\int_a^b f(x)\,dx = ?$$

Divide $[a, b]$ into n sub-intervals $a = x_0 < x_1 < \cdots < x_n = b$

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

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

$$= \frac{h}{2}[f_0 + f_1] + \frac{h}{2}[f_1 + f_0] - \cdots + \frac{h}{2}[f_{n-1} + f_n]$$

$$= \frac{h}{2}\left[(f_0 + f_n) + 2(f_1 + f_2 \cdots + f_{n-1})\right]$$

$$= \frac{h}{2}\left[\text{sum of extreme ordinates} + 2\left(\text{sum of intermediate ordinates}\right)\right]$$

Composite Simpson's $\frac{1}{3}$rd Rule

To evaluate $\int_a^b f(x)\,dx$.

Divide $[a, b]$ into $n = 2m$ (even) subintervals

$$\int_a^b f(x)\,dx = \int_{x_0}^{x_2} f(x)\,dx + \int_{x_2}^{x_4} f(x)\,dx + \cdots + \int_{x_{2m-2}}^{x_{2m}} f(x)\,dx.$$

Apply Simpson's $\frac{1}{3}$rd rule.

$$\int_a^b f(x)\,dx = \frac{h}{3}\left[f_0 + 4f_1 + f_2\right] + \frac{h}{3}\left(f_2 + 4f_3 + f_4\right]$$

$$\cdots + \frac{h}{3}\left[f_{2m-2} + 4f_{2m-1} + f_{2m}\right]$$

$$= \frac{h}{3}\left[(f_0 + f_{2m}) + 4(f_1 + f_3 \cdots + f_{2m-1}) + 2(f_2 + f_4 + \cdots + f_{2m-2})\right]$$

$h = \frac{b-a}{2m}$.

$$= \frac{h}{3}\left[\left(\begin{array}{c}\text{sum of first}\\ \text{and last}\\ \text{ordinates}\end{array}\right) + 4\left(\begin{array}{c}\text{sum of even}\\ \text{ordinates}\end{array}\right) + 2\left(\begin{array}{c}\text{sum of remaining}\\ \text{ordinates}\end{array}\right)\right]$$

#	
1	$f_0 \rightarrow$ odd
2	$f_1 \rightarrow$ even
3	$f_2 \rightarrow$ odd

Error in Trapezoidal Rule.

$$\int_{x_0}^{x_1} f(x)\,dx \simeq \frac{h}{2}\left[f(x_0) + f(x_1)\right] \quad \text{where } h = x_1 - x_0$$

Error associated with the formula is $\int_{x_0}^{x_1} f(x)\,dx - \frac{h}{2}\left[f_0 + f_1\right]$

Idea: Write Taylor series of $f(x)$ in powers of $(x - x_0)$ i.e. about x_0 and use $h = x_1 - x_0$

Rough

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

$$\int_{x_0}^{x_1} P_N(x)\,dx = \left[x f_0 + \frac{(x-x_0)^2}{2!} f_0' + \frac{f_0''}{2!}\frac{(x-x_0)^3}{3} + \cdots \right]_{x_0}^{x_1}$$

$$= (x_1 - x_0) f_0 + \frac{h^2}{2!} f_0' + \frac{f_0''}{2!}\frac{h^3}{3} + \cdots$$

$$\int_{x_0}^{x_1} f(x)\,dx \simeq h f_0 + \frac{h^2}{2!} f_0' + \frac{h^3}{3!} f_0'' + \cdots \quad -(1)$$

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

$$f(x_0) + f'(x_0)(x_1 - x_0) + \frac{f''(x_0)}{2!}(x_1 - x_0)^2 + \cdots$$

Also, $\frac{h}{2}\left[f(x_0) + f(x_1)\right]$, when expanded about x_0,

we get $= h f_0 + \frac{h^2}{2!} f_0' + \frac{h^3}{4} f_0'' + \cdots \quad -(2)$

Subtracting (2) from (1)

$$\int_{x_0}^{x_1} f(x)\,dx - \frac{h}{2}(f_0 + f_1) = \left(\frac{1}{6} - \frac{1}{2}\right) h^3 f_0'' + \text{terms containing higher powers of } h.$$

The principal error term $= \left(\frac{1}{6} - \frac{1}{4}\right) h^3 f_0'' - \frac{h^3}{12} f_0''$

$= O(h^3)$

Error in composite Trapezoidal Rule

$$\int_a^b f(x)\,dx \simeq \frac{h}{2}\left[(f_0 + f_n) + 2(f_1 + f_2 \cdots + f_{n-1}) \right]$$

$$\text{Error} = -\frac{h^3}{12} f_0'' - \frac{h^3}{12} f_1'' - \cdots - \frac{h^3}{12} f_{n-1}''$$

+ terms containing higher powers of h

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

$$\text{Error}\left(\int_{x_0}^{x_n} f(x)dx\right) = \text{Error}\left(\int_{x_0}^{x_1}\right) + \text{Error}\left(\int_{x_1}^{x_2}\right)$$

$$+ \cdots + \text{Error}\left(\int_{x_{n-1}}^{x_n} f(x)\,dx\right)$$

$$\text{Error} = -\frac{h^3}{3}\left[f_0'' + f_1'' \cdots + f_{n-1}''\right] + \text{terms containing higher powers of } h.$$

If f'' is continuous on $[a,b]$, then there exists a 'ξ' such that

$$f''(\xi) = \frac{f_0'' + f_1'' \cdots + f_{n-1}''}{n}$$

$\therefore$ Error of composite Trap. Rule $= -\frac{h^3}{12} n f''(\xi)$ $\quad n = \frac{b-a}{h}$

$$= \frac{-h^3}{12} \frac{(b-a)}{h} f''(\xi)$$

$$= \frac{-h^2}{12} (b-a) f''(\xi)$$

$O(h^2)$

Error in Simpson's $\frac{1}{3}$rd rule.

$$\int_{x_{n-1}}^{x_{n+1}} f(x)\,dx = \frac{h}{3}\left(f_{n-1} + 4f_n + f_{n+1}\right)$$

$$\text{Error} = \frac{-1}{90} h^5 f_n + \ldots \text{(higher order)}$$

$O(h^5)$

H.W.

Error in Composite Simpson's $\frac{1}{3}$rd Rule

$$= \frac{-1}{180} h^4 (b-a) f^{(4)}(\xi) \cdot O(h^4)$$

or $$-\frac{h^5}{90} \frac{N}{2} f^4(\xi)$$

	Method	Error Order.
1	Trapezoidal	$O(h^3)$
2	Composite Trapezoidal	$O(h^2)$
3	Simpson's	$O(h^5)$
4	Composite Simpson's	$O(h^4)$

April 1.

Problem Using Simpson's $\frac{1}{3}$rd Rule, evaluate $\int_0^1 \frac{dx}{1+x}$ dividing the interval into 10 equal parts

Sol $f(x) = \frac{1}{1+x}$, $a=0$, $b=1$ $\quad h = \frac{b-a}{n} = .1$

x	0.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1
$f(x)$	1	.9091	.8333	.7892	.7143	.6667	.6750	.5882	.5556	.5263	.5

S_1 = Sum of extreme ordinates : $1 + .5 = 1.5$

S_2 = Sum of even ordinates $= .9091 + .7692 + .6667 + .5882 + .5263$
$= 3.4595$

S_3 = Sum of odd ordinates $= .8333 + .7143 + .6750 + .5556 = 2.7282$

$$\int_0^1 f(x)\,dx = \int_0^1 \frac{dx}{1+x} = \frac{h}{3}\left[1.5 + 4(3.4545) + 2(2.7282)\right]$$

where $(h = .1)$ =

Romberg Integration

Richardson's extrapolation procedure applied to integration methods is called Romberg Integration

First, we find the power series expansion of the error terms in the integration method. Then by

eliminating the leading terms in the error expansion by using the computed results, we obtain new methods (results like) which are of higher order than the previous method

Consider the integral $I = \int_a^b f(x)\,dx$

Error in composite trapezoidal rule $= -\frac{(b-a)}{12} h^2 f''(\xi)$ can be obtained as

$$I = I_T + c_1 h^2 + c_2 h^4 + c_3 h^6 + \cdots$$

where 'c's are constants independent of h.

The extrapolation technique for the trapezoidal rule is given by –

$$g^{(m)}(h) = \frac{4^m g^{(m-1)}\left(\frac{h}{2}\right) - g^{(m-1)}(h)}{4^m - 1} \qquad m = 1, 2, 3 \ldots$$

where $g^{(0)}(h) = g(h)$

$$I_T^{(m)}(h) = \frac{4^m I_T^{(m-1)}\left(\frac{h}{2}\right) - I_T^{(m-1)}(h)}{4^m - 1} \qquad m = 1, 2 \ldots$$

↓ order $2m+2$

[not derived in the course]

Error in Composite Simpson's $\frac{1}{3}$rd Rule can be obtained as –

$$I = I_S + d_1 h^4 + d_2 h^6 + d_3 h^8 + \ldots$$

where 'd's are independent of h

The extrapolation procedure for Simpson's $\frac{1}{3}$rd Rule becomes –

$$g^{(m)}(h) = \frac{4^m g^{(m-1)}\left(\frac{h}{2}\right) - g^{(m-1)}(h)}{4^m - 1} \qquad m = 1, 2, 3$$

with $g^{(0)}(h) = g(h)$

becomes –

$$I_S^{(m)}(h) = \frac{4^{m+1} I_S^{(m-1)}\left(\frac{h}{2}\right) - I_S^{(m-1)}(h)}{4^{m+1} - 1} \qquad m = 1, 2, 3 -$$

↓

of order $= 2m+4$

[not derived in the course]

NEWTON – COTES FORMULA

x_i	x_0	. .	x_n
f_i	f_0		f_n

$x_0 < x_1 < \ldots < x_n$

Objective : To find approximate value of the integral $\int_{x_0}^{x_n} f(x)\,dx$

Lagrange Interpolation formula.

$$f(x) = \sum_{k=0}^{n} l_k(x) f_k \quad \text{where } l_k(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}$$

$$\int_{x_0}^{x_n} f(x)\,dx = \int_{x_0}^{x_n} \sum_{k=0}^{n} l_k(x) f_k \,dx$$

$$= \sum_{k=0}^{n} \int_{x_0}^{x_n} l_k(x) f_k \,dx \qquad -(1)$$

Consider $$\int_{x_0}^{x_n} l_k(x)\,dx = \int_{x_0}^{x_1} \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}\,dx$$

Suppose x_i are equispaced, i.e. $x_k - x_{k-1} = h$

Put $\frac{x-x_0}{h} = s$. then $dx = h\,ds$

$$\frac{x-x_i}{h} = s-i$$

Also when $x = x_0 \quad s = 0$

$x = x_n \quad s = n$

$$\int_{x_0}^{x_n} l_k(x)\,dx = \int_{s=0}^{n} \frac{s h\ (s-1)h \cdots [s-(k-1)]h\ [s-(k+1)]h \cdots (s-n)h}{k h\ (k-1)h \cdots (k-(k-1))h\ (k-(k+1))h \cdots (k-n)h}\, h\,ds$$

$$= h \int_{s=0}^{n} \frac{s(s-1)\cdots[s-(k-1)]\,[s-(k+1)]\cdots(s-n)}{k(k-1)\cdots(k-(k-1))\,(k-(k+1))\cdots(k-n)}\,ds$$

$= \lambda_k$ (say)

$$= h\,\lambda_k \quad \text{(say)}$$

Now (1) $\Rightarrow$ $\displaystyle\int_{x_0}^{x_n} f(x)\,dx \simeq \sum_{k=0}^{n} (h\lambda_k) f_k$

$$\int_{x_0}^{x_n} f(x)\,dx = h \sum_{k=0}^{n} \lambda_k f_k$$

where $\displaystyle\lambda_k = \int_{s=0}^{n} \frac{s(s-1)\cdots(s-(k-1))\,(s-(k+1))\cdots(s-n)}{k(k-1)\cdots(k-(k-1))\,(k-(k+1))\cdots(k-n)}\,ds$ — (2)

(1) NEWTON-COTES CLOSED TYPE $(n+1)$ POINT FORMULA

When the end values x_0 and x_n are included in the interpolation, the resulting integration, quadrature formula is called Newton Cotes closed type formula

When the end values x_0 and x_n are NOT included in the interpolation, the resulting integration, quadrature formula is called Newton Cotes open type formula

In the quadrature formula, $h\lambda_0, h\lambda_1, \ldots, h\lambda_n$ are called the weights of the quadrature formula.

Trapezoidal Rule (Two point formula, $n=1$)

(n → no of intervals)

Put $n=1$

$$\int_{x_0}^{x_1} f(x)\,dx = h\left[\lambda_0 f_0 + \lambda_1 f_1\right]$$

where $\lambda_0 = \int_0^1 \frac{s-1}{0-1}\,ds = -\int_0^1 s-1 = \frac{1}{2}$

$$\lambda_1 = \int_0^1 \frac{s}{1}\,ds = \frac{1}{2}$$

$$\int_{x_0}^{x_1} f(x)\,dx = h\left[\frac{1}{2}f_0 + \frac{1}{2}f_1\right]$$

$$\boxed{= \frac{h}{2}\left[f_0 + f_1\right]}$$

Simpson's $\frac{1}{3}$rd Rule, $n=2$

Put $n=2$

$$\int_{x_0}^{x_2} f(x)\,dx \simeq h\sum_{k=0}^{2}\lambda_k f_k$$

$$= h\left[\lambda_1 f_1 + \lambda_2 f_2 + \lambda_3 f_3\right]$$

$$\lambda_0 = \int_0^2 \frac{(s-1)(s-2)}{(0-1)(0-2)} = \int_0^2 \frac{1}{2}\left(s^2 - 3s + 2\right)ds$$

$$= \frac{1}{3}$$

$$\lambda_1 = \int_0^2 \frac{(s-0)(s-2)}{(1-0)(1-2)}\,ds = -\int_0^2 (s^2 - 2s)\,ds$$

$$= -\left[\frac{s^3}{3} - \frac{2s^2}{2}\right]_0^2 = \frac{4}{3}$$

$$\lambda_2 = \int_0^2 \frac{(s-0)(s-1)}{(2-0)(2-1)}\,ds = \frac{1}{2}\int_0^2 (s^2 - s)\,ds$$

$$= \frac{1}{2}\left[\frac{s^3}{3} - \frac{s^2}{2}\right]_0^2 = \frac{1}{3}$$

$$\int_{x_0}^{x_2} f(x)\,dx = h\left[\frac{1}{3}f_0 + \frac{4}{3}f_1 + \frac{1}{3}f_2\right]$$

$$= \frac{h}{3}\left[f_0 + 4f_1 + f_2\right]$$

More generally we can write:

$$\int_{x_{n-1}}^{x_{n+1}} f(x)\,dx = \frac{h}{3}\left[f_{n-1} + 4f_n + f_{n+1}\right]$$

Simpson's $\frac{3}{8}$ th Rule, Four point formula.

Put $n=3$ $\quad (n=3)$

$$\int_{x_0}^{x_3} f(x)\,dx = h\left[\lambda_0 f_0 + \lambda_1 f_1 + \lambda_2 f_2 + \lambda_3 f_3\right]$$

$$\lambda_0 = \int_0^3 \frac{(s-1)(s-2)(s-3)}{(0-1)(0-2)(0-3)}\,ds = \frac{-1}{6}\int_0^3 (s^3 - 6s^2 + 11s - 6)\,ds$$

$$= \frac{-1}{6}\left[\frac{s^4}{4} - \frac{6s^3}{3} + \frac{11}{2}s^2 - 6s\right]_0^3 = \frac{-1}{6}\left[\frac{81}{4} - 54 + \left(\frac{11}{2}\cdot 9\right) - 18\right]$$

$$= \frac{-1}{6}\left[\frac{81}{4} + \frac{99}{2} - 72\right] = \frac{3}{8}$$

$$\lambda_1 = \int_0^3 \frac{s(s-2)(s-3)}{(1-0)(1-2)(1-3)}\,ds = \frac{9}{8}$$ HW.

$$\lambda_2 = \int_0^3 \frac{(s-0)(s-1)(s-3)}{(2-0)(2-1)(2-3)}\,ds = \frac{9}{8}$$ H.W.

$$\lambda_3 = \int_0^3 \frac{(s-0)(s-1)(s-2)}{(3-0)(3-1)(3-2)}\,ds = \frac{3}{8}$$ H.W.

$$\int_0^3 f(x)\,dx = \frac{3h}{8}\left[f_0 + 3f_1 + 3f_2 + f_3\right]$$

Composite Trapezoidal Rule

Composite Simpson's $\frac{1}{3}$rd Rule

Composite Simpson's $\frac{3}{8}$th Rule

April 5

Error in Newton Cotes closed-type formula

x_i	x_0	- -
t_i	t_0	- -

$$E_n = f(x) - P_n(x)$$

$$= \Pi(x) \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

$$\int_{x_0}^{x_n} (f(x) - P_n(x))\, dx = \int_{x_0}^{x_n} E(x)\, dx = \int_{x_0}^{x_n} \Pi(x) \frac{f^{(n+1)}(\xi)}{(n+1)!}\, dx$$

Error in Newton-Cotes integration formula.

$$= \int_{x_1}^{x_n} (x-x_0)(x-x_1) \cdots (x-x_n) \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

Assume $x_i - x_{i-1} = h$

Put $\frac{x-x_0}{h} = s$ Then $dx = h\,ds$

and $\frac{x-x_i}{h} = s-i$

Error term, $\int_{s=0}^{n} sh\,(s-1)h \cdots (s-n)h \frac{f^{(n+1)}(\xi)}{(n+1)!} h\,ds$

$$= \frac{h^{n+2}}{(n+1)!} \int_{s=0}^{n} s(s-1) \cdots (s-n) f^{(n+1)}(\xi)\, ds$$

$n \rightarrow$ no of intervals

It is proved that the error =

error = 0 ← when degree $\leq n$

$$\frac{h^{n+2}}{(n+1)!} f^{(n+1)}(\xi_1) \int_0^n s\,(s-1)\,\ldots\,(s-n)\,ds$$

when n = odd

error = 0 ← when degree $\leq (n+1)$

$$\frac{h^{n+3}}{(n+2)!} f^{(n+2)}(\xi_1) \int_0^n \left(s - \frac{n}{2}\right) s\,(s-1)\,\ldots\,(s-n)\,ds$$

when n = even

i.e. no of intervals are taken to be even.

If $f(x)$ and $g(x)$ are continuous in $[a,b]$ and $g(x)$ does not change sign in $[a,b]$ then $\int_a^b f(x)\,g(x)\,dx = f(\xi) \int_a^b g(x)\,dx$ for some ξ

Def Precision of a Quadrature formula or degree/order of an integration formula.

A quadrature formula is said to have a precision 'n' if it is exact for all polynomials of degree $\leq n$

[So Cotes formulas have precision of at most $(n+1)$]

Error in Trapezoidal Rule. : Two point formula $n=1$

$$\frac{h^{n+2}}{(n+1)!} f^{(n+1)}(\xi) \int_0^n s(s-1)\,\ldots\,(s-n)\,ds$$

$$E = \frac{h^3}{2!} f''(\xi) \int_0^1 s(s-1)\,ds$$

$$= \frac{h^3}{2} f''(\xi) \int_0^1 (s^2 - s)\,ds$$

$$= \frac{h^3}{2} f''(\xi) \left[\frac{1}{3} - \frac{1}{2}\right] = \frac{-h^3}{12} f''(\xi)$$

Error in Simpson's $\frac{1}{3}$rd Rule. $n=2$ (even)

$$E = \frac{h^{n+3}}{(n+2)!} f^{(n+2)}(\xi) \int_0^n (s-\tfrac{n}{2})\, s(s-1)(s-2)\cdots(s-n)\, ds$$

$$= \frac{h^5}{4!} f^{(4)}(\xi) \int_0^2 (s-1)\, s(s-1)(s-2)\, ds$$

$$= \frac{h^5}{24} f^{(4)}(\xi) \int_0^2 (s^4 - 4s^3 + 5s^2 - 2s)\, ds$$

$$= \frac{h^5}{24} f^{(4)}(\xi) \left[\frac{s^5}{5} - s^4 + \frac{5}{3}s^3 - s^2 \right]_0^2$$

$$= \frac{h^5}{24} f^{(4)}(\xi) \left[\frac{32}{5} - 16 + \frac{5}{3} \times 8 - 4 \right]$$

$$= \frac{h^5}{24} f^{(4)}(\xi) \left[\frac{96 + 200 - 300}{15} \right]$$

$$= \frac{h^5}{24} f^{(4)}(\xi) \left[\frac{-4}{15} \right]$$

$$= \frac{-h^5 f^{(4)}(\xi)}{90}$$

$$\begin{array}{r} 32 \\ \times 3 \\ \hline 96 \\ 40 \\ -100+96 \end{array}$$

Error in Simpson's $\frac{3}{8}$th rule. $n=3$ [4 point formula]

$$\text{Error} = \frac{h^5}{4!} f^{(4)}(\xi) \int_0^3 s(s-1)(s-2)(s-3)\, ds$$

(n = odd)

$$= \frac{-3}{80} h^5 f^{(4)}(\xi)$$

Verify H.W.

Newton Composite - add error in all intervals and add.

Newton - Cotes (N-C) Open type Quadrature Formula.

x_i	x_0	x_1	. .	x_n
f_i	f_0	f_1	. .	f_n

Consider the interpolating polynomial of $f(x)$ interpolating the value of f at (n-1) intermediate points

$\downarrow$

$x_1, x_2 \cdots x_{n-1}$

(We will get polynomial of degree $(n-1)-1 = n-2$)

$$P_{n-2}(x) = \sum_{k=1}^{n-1} l_k(x) f_k$$

where $$l_k(x) = \frac{(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1}) - - (x-x_{n-1})}{(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_{n-1})}$$

$$\therefore \int_{x_0}^{x_n} f(x)\,dx = \int_{x_0}^{x_n} P_{n-2}(x)\,dx$$

$$\therefore \int_{x_0}^{x_n} \sum_{k=1}^{n-1} l_k(x) f_k = \sum_{k=1}^{n-1} \int_{x_0}^{x_n} l_k(x)\,dx\, f_k$$

Consider $\int_{x_0}^{x_n} l_k(x)\,dx = \int_{x_0}^{x_n} \frac{(x-x_1)\ldots(x-x_{k-1})(x-x_{k+1})\ldots(x-x_{n-1})}{(x_k-x_1)\ldots(x_k-x_{k-1})(x_k-x_{k+1})\ldots(x_k-x_{n-1})}\,dx$

Suppose $x_i - x_{i-1} = h \quad , i = 1, 2 \ldots, n$

Put $\frac{x-x_0}{h} = s \Rightarrow dx = h\,ds$

$$\int_{x_0}^{x_n} l_k(x)\,dx = \int_{s=0}^{n} \frac{[(s-1)h\ (s-2)h \ldots (s-(k-1))h\ (s-(k+1))h \ldots (s-(n-1))h]\,h\,ds}{[(k-1)h \ldots (k-(k-1))h\ (k-(k+1))h \ldots (k-(n-1))h]}$$

$$= h\int_{s=0}^{n} \frac{(s-1)(s-2)\ldots(s-(k-1))\ (s-(k+1))\ldots(s-(n-1))\,ds}{(k-1)(k-2)\ldots(k-(k-1))\ (k-(k+1))\ldots(k-(n-1))}$$

$$= \lambda_k \text{ (say)}$$

$$\int_{x_0}^{x_n} f(x)\,dx = \sum_{k=1}^{n-1} h\lambda_k f_k = h\sum_{k=1}^{n-1} \lambda_k f_k$$

The $(n+1)$ point N-C open type formula is given by -

$$\int_{x_0}^{x_1} f(x)\,dx \simeq h\sum_{k=1}^{n-1} \lambda_k f_k$$

where $\lambda_k = \int_{s=0}^{n} \frac{(s-1)(s-2)\ldots(s-(k-1))\ (s-(k+1))\ldots(s-(n-1))\,ds}{(k-1)(k-2)\ldots(k-(k-1))\ (k-(k+1))\ldots(k-(n-1))}$

$\frac{s}{k}$ ← Present in closed type → $\frac{s-n}{k-n}$

Error in N-C open type

$$\left[\int_{x_0}^{x_n} \pi(x)\, \frac{f^{(n+1)}(\xi)}{(n+1)!}\, dx\right]$$

$$\int_{x_0}^{x_n} (x-x_1)(x-x_2)\cdots(x-x_{n-1})\, \frac{f^{(n-1)}(\xi)}{(n-1)!}\, dx$$

Suppose $x_i - x_{i-1} = h$

Put $\frac{x-x_0}{h} = s \Rightarrow dx = h\,ds$

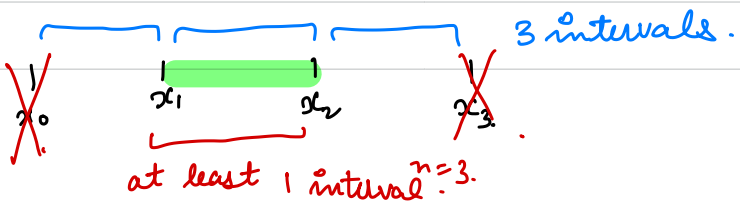
$$\text{Error} = h^n \int_0^n (s-1)(s-2)\cdots(s-(n-1))\, \frac{f^{(n-1)}(\xi)}{(n-1)!}\, ds$$

This is shown -

$$\text{Error} = \begin{cases} h^n \dfrac{f^{(n-1)}(\xi_1)}{(n-1)!} \displaystyle\int_0^n (s-1)(s-2)\cdots(s-(n-1))\, ds & \text{when } n \to \text{odd} \\ h^{n+1} \dfrac{f^{n}(\xi_2)}{n!} \displaystyle\int_0^n \left(s-\frac{n-2}{2}\right)(s-1)\cdots(s-(n-1))\, ds & n \to \text{even} \end{cases}$$

N-C open

Minimum value of $n = 3$ (no of sub-intervals) to derive N-C open type formula



For $n=3$ $\int_{x_0}^{x_3} f(x)\,dx \simeq h[\lambda_1 f_1 + \lambda_2 f_2]$

where $\lambda_1 = \int_0^3 \frac{s-2}{1-2}\,ds = \frac{3}{2}$

$\lambda_2 = \int_0^3 \frac{s-1}{2-1}\,ds = \frac{3}{2}$

∴ Four point N-C open type formula is

$$\int_{x_0}^{x_3} f(x)\,dx = h\left[\frac{3}{2}f_1 + \frac{3}{2}f_2\right]$$

$$= \frac{3h}{2}[f_1 + f_2]$$

Error associated with this :

$$= \frac{h^3 f''(\xi)}{(3-1)!}\int_0^3 (s-1)(s-2)\,ds$$

$$= \frac{h^3}{2} f''(\xi)\int_0^3 (s^2 - 3s + 2)\,ds$$

$$= \frac{h^3}{3} f''(\xi)\left[\frac{s^3}{3} - \frac{3}{2}s^2 + 2s\right]_0^3$$

$$= \frac{3}{4} h^3 f''(\xi)$$

Q Find Newton-Cotes Open-Type 5-point formula and error associated with it

April 11

Gauss Quadrature Formula.

N-C Formula $\int_a^b f(x)\,dx \simeq h \sum_{k=1}^{n} \lambda_k f(x_k)$

$\lambda_k \rightarrow$ weights

More generally, $\int_a^b w(x) f(x)\,dx$ — (1)

$w(x) \rightarrow$ weight function
$\rightarrow$ +ve continuous f^n [Utility: To ensure norm is a finite quantity]

Now, how to increase precision of a quadrature formula?

Make λ_k's arbitrary $\lambda_1, \ldots, \lambda_n$
$\Rightarrow$ $2n$ unknowns. $x_1, \ldots, x_n$

(1) is exact for $f(x) = 1, x, x^2 \ldots, x^{2n-1}$

i.e. error is zero.

$\Rightarrow$ System of $2n$ non-linear equations [not easy to solve]

$\rightarrow$ x_k's are zeros of orthogonal polynomial.

Gauss Jacobi Theorem : Let $x_1, x_2 \ldots x_n$ be zeros of orthogonal polynomial $P_n(x)$ (of degree n) w.r.t. weight function $w(x)$ on $[a, b]$. Then

$$\int_a^b w(x) f(x)\, dx \simeq \sum_{k=1}^{n} \lambda_k f(x_k)$$

$$\lambda_k = n\lambda_k$$

is exact whenever $f(x)$ is a polynomial of degree $\leq 2n-1$ where λ_k = is given by. –

$$\lambda_k = \frac{1}{P_n'(x)} \int_a^b \frac{w(x)\, P_n(x)}{(x - x_k)}\, dx$$

Further the weights λ_k are positive.

2 point Gauss Legendre Quadrature Formula (Method of undetermined coefficients)

$$\int_{-1}^{1} f(x)\, dx \simeq \lambda_1 f(x_1) + \lambda_2 f(x_2)$$

$\int_a^b w(x) f(x)\, dx$

$a = -1$
$b = 1$
$w(x) = 1$

4 unknowns λ_1, λ_2
x_1, x_2

So it is exact for polynomials of degree ≤ 3

$$f(x) = x^i \qquad i = 0, 1, 2, 3$$

$$[1, x, x^2, x^3]$$

$$\int_{-1}^{1} 1 \cdot dx = \lambda_1 \cdot 1 + \lambda_2 \cdot 1 \qquad [\because f(x) = 1]$$

$$= \lambda_1 + \lambda_2 = 2 \qquad \text{— (1)}$$

$$\int_{-1}^{1} x\,dx = \lambda_1 x_1 + \lambda_2 x_2 \qquad [f(x) = x]$$

(odd) = 0

$$\lambda_1 x_1 + \lambda_2 x_2 = 0 \qquad \text{— (2)}$$

$$\int_{-1}^{1} x^2\,dx = \lambda_1 x_1^2 + \lambda_2 x_2^2$$

$$\lambda_1 x_1^2 + \lambda_2 x_1^2 = \frac{2}{3} \qquad \text{— (3)}$$

$$\int_{-1}^{1} x^3\,dx = \lambda_1 x_1^3 + \lambda_2 x_2^3$$

(∵ odd) $= 0$

$$\lambda_1 x_1^3 + \lambda_2 x_2^3 = 0 \qquad \text{— (4)}$$

Suppose x_1 and x_2 are roots of quadratic equation

$$(x - x_1)(x - x_2) = c_0 + c_1 x + c_2 x^2$$

$c_2 = 1$

Also,

$$\left.\begin{aligned} c_0 + c_1 x_1 + c_2 x_1^2 &= 0 \\ c_0 + c_1 x_2 + c_2 x_2^2 &= 0 \end{aligned}\right\} \text{ as } x_1 \text{ and } x_2 \text{ are roots of } (x - x_1)(x - x_2)$$

(1) c_1 + (2) c_2 + (3) c_3

$$\Rightarrow \lambda_1 (c_0 + c_1 x_1 + c_2 x_2^2) + \lambda_2 (c_0 + c_1 x_2 + c_2 x_2^2)$$

($c_1 x_1 + c_2 x_2^2 = 0$, $c_1 x_2 + c_2 x_2^2 = 0$)

$$= 2c_0 + \frac{2}{3} c_2$$

$$\Rightarrow 2c_0 + \frac{2}{3} c_2 = 0$$

$$c_0 = -\frac{1}{3} c_2$$

$$\text{or} \quad c_0 = -\frac{1}{3}.$$

(2) c_0 + (3) c_1 + (4) c_2

$$\Rightarrow \lambda_1 x_1 (c_0 + c_1 x_1 + c_2 x_1^2) + \lambda_2 x_2 (c_0 + c_1 x_2 + c_2 x_2^2)$$

($c_1 x_1$ and $c_1 x_2$ terms $\to 0$; $c_0 \to 0$)

$$= \frac{2}{3} c_1$$

$$\Rightarrow \quad c_1 = 0$$

$$\therefore \quad c_0 = -\frac{1}{3} \qquad c_1 = 0 \qquad c_2 = 1$$

$$(x - x_1)(x - x_2) = c_0 + c_1 x + c_2 x^2$$

$$\therefore \quad x^2 - \frac{1}{3} = 0$$

$$x = \pm \sqrt{\frac{1}{3}}$$

Let $x_1 = -\frac{1}{\sqrt{3}} \qquad x_2 = +\frac{1}{\sqrt{3}}$

$$\lambda_1 + \lambda_2 = 2$$

$$-\frac{\lambda_1}{\sqrt{3}} + \frac{\lambda_2}{\sqrt{3}} = 0$$

$$\Rightarrow \quad \lambda_1 + \lambda_2 = 2$$

$$\lambda_1 - \lambda_2 = 0$$

$$\Rightarrow \lambda_1 = \lambda_2 = 1$$

$$\therefore \quad \lambda_1 = 1 \qquad \lambda_2 = 1$$

$$x_1 = -\frac{1}{\sqrt{3}} \qquad x_2 = +\frac{1}{\sqrt{3}}.$$

$$\int_{-1}^{1} f(x)\,dx \simeq 1 \cdot f\left(\frac{-1}{\sqrt{3}}\right) + 1 \cdot f\left(\frac{+1}{\sqrt{3}}\right)$$

$$\simeq f\left(\frac{-1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right)$$

$P_2(x)$ = 2nd degree Legendre polynomial.

$$P_2(x) = \frac{1}{2}(3x^2 - 1)$$

$$\Rightarrow \quad x = \pm \frac{1}{\sqrt{3}}$$

$$\int_{-1}^{1} f(x)\,dx) = \lambda_1 f\left(\frac{-1}{\sqrt{3}}\right) + \lambda_2 f\left(\frac{1}{\sqrt{3}}\right)$$

$f(x) = 1$ $\qquad \int_{-1}^{1} 1\,dx = \lambda_1 + \lambda_2$

$$\boxed{\lambda_1 + \lambda_2 = 2}$$

$f(x) = x$ $\qquad \int_{-1}^{1} x\,dx = \lambda_1\left(\frac{-1}{\sqrt{3}}\right) + \lambda_2 \frac{1}{\sqrt{3}}$

$$\Rightarrow \frac{-\lambda_1}{\sqrt{3}} + \frac{\lambda_2}{\sqrt{3}} = 0$$

$$\boxed{\lambda_1 - \lambda_2 = 0}$$

$$\lambda_1 = 1$$

$$\lambda_2 = 1$$

Error Term

Suppose the integration method is exact for all polynomials of degree $\leq n$ i.e.

$$R_n = 0 \quad \text{when } f(x) = x^i, \quad i = 0, 1 \ldots n$$

$\downarrow$

We can write error term in the form

$$\boxed{R_n = c \, \frac{f^{(n+1)}(\xi)}{(n+1)!}}$$

This follows as corollary of peano - kernel theorem

where $c = \int_a^b w(x) \, x^{n+1} \, dx - \sum_{k=1}^{n} \lambda_k \, x_k^{n+1}$

c is called error constant

If $c = 0$ for $f(x) = x^{n+1}$, then take next term $f(x) = x^{n+2}$

Error in 2-point Gauss - Legendre Quadrature formula.

$$c = \int_{-1}^{1} x^3 \, dx - (\lambda_1 x_1^3 + \lambda_2 x_2^3)$$

odd $= 0$ $\qquad \lambda_1 = 1 \quad \lambda_2 = 1 \qquad x_1 = \frac{-1}{\sqrt{3}} \quad x_2 = \frac{1}{\sqrt{3}}$.

$$= 0 - \left(\left(\frac{-1}{\sqrt{3}}\right)^3 + \left(\frac{1}{\sqrt{3}}\right)^3 \right)$$

$$= 0$$

$$\text{Error} = \frac{c \, f^{(4)}(\xi)}{4!}$$

$$C = \int_{-1}^{1} x^4\,dx - (\lambda_1 x_1^4 + \lambda_2 x_2^4)$$

$$= \frac{2}{5} - \left(\frac{1}{9} + \frac{1}{9}\right) = \frac{8}{45}$$

$$\text{Error} = \frac{8}{45} \times \frac{f^{(4)}}{4!} = \frac{f^{(4)}}{135}$$

April 12

3-point Gauss Legendre Quadrature Formula.

$$\int_{-1}^{1} f(x)\,dx \simeq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \lambda_3 f(x_3)$$

x_1, x_2, x_3 are zeros of $P_3(x)$ (Legendre polynomial of degree 3)

$$P_3(x) = \frac{1}{2}(5x^3 - 3x)$$

Zeroes of $P_3(x) = \frac{1}{2}(5x^3 - 3x) = 0$

$$\Rightarrow \quad x = 0, \sqrt{\frac{3}{5}}, -\sqrt{\frac{3}{5}}$$

Let $x_1 = -\sqrt{\frac{3}{5}}$ $\quad x_2 = 0$ $\quad x_3 = \sqrt{\frac{3}{5}}$

$f(x) = 1$ $\quad \int_{-1}^{1} 1\,dx = \lambda_1 + \lambda_2 + \lambda_3$

$$\lambda_1 + \lambda_2 + \lambda_3 = 2 \quad -(1)$$

$$f(x) = x \qquad \int_{-1}^{1} x\,dx = -\sqrt{\frac{3}{5}}\lambda_1 + 0 + \sqrt{\frac{3}{5}}\lambda_3$$

$$-\sqrt{\frac{3}{5}}\lambda_1 + \sqrt{\frac{3}{5}}\lambda_3 = 0$$

$$\lambda_1 - \lambda_2 = 0 \qquad -(2)$$

$$\Rightarrow \lambda_1 = \lambda_3$$

$$f(x) = x^2 \qquad \int_{-1}^{1} x^2\,dx = \frac{3}{5}\lambda_1 + 0 + \frac{3}{5}\lambda_3$$

$$(\lambda_1 + \lambda_3) = \frac{10}{9} \qquad -(3)$$

$$\Rightarrow \qquad \lambda_1 = \frac{5}{9} \qquad \lambda_3 = \frac{5}{9} \qquad \lambda_2 = 2 - \frac{10}{9} = \frac{8}{9}$$

3-point Gauss Legendre Quadrature Formula is -

$$\int_{-1}^{1} f(x)\,dx = \frac{5}{9} f\left(-\sqrt{\frac{3}{5}}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\sqrt{\frac{3}{5}}\right)$$

Error

$$E = \frac{c\, f^{(6)}(\xi)}{6!}$$

where $$c = \int_{-1}^{1} x^6\,dx - \sum_{k=1}^{3} \lambda_k f_k = \frac{8}{175} \qquad \text{(Verify)}$$

$$E = \frac{8}{175} \times \frac{f^{(6)}(\xi)}{6!}$$

$$= \frac{1}{15750} f^{(6)}(\xi)$$

$$\left[\int_{-1}^{1} x^6 \, dx - \left(\frac{5}{9} f\left(-\sqrt{\frac{3}{5}}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\sqrt{\frac{3}{5}}\right) \right) \right]$$

$f(0) \to 0$

$$\frac{2}{7} - \left(0 + 2 \times \frac{5}{9} \left(\sqrt{\frac{3}{5}}\right)^6 \right)$$

$$\frac{2}{7} - \frac{6}{25} = \frac{8}{175}$$

Q Find w_1, w_2, x_1, x_2 in

$$\int_0^1 \sqrt{x} f(x) \, dx \simeq w_1 f(x_1) + w_2 f(x_2)$$

Also find the error.

Q. Evaluate $\int_{-4}^{4} \frac{dx}{1+x^2}$ using 3-point Gauss Legendre Quadrature Formula.

Sol $[-4, 4] \longrightarrow [-1, 1]$

$$x = \frac{1}{2}\left[(b+a) + (b-a)\, t \right]$$

$$x = 4t$$
$$dx = 4\,dt$$

$$I = \int_{-4}^{4} \frac{dx}{1+x^2} = \int_{-1}^{1} \frac{4\,dt}{1+16t^2} = \int_{-1}^{1} f(t)\, dt$$

where $f(t) = \frac{4}{1+16t^2}$

To this apply 3-point Gauss Legendre Quadrature Formula.

$$I = \frac{5}{9} f\left(-\sqrt{\frac{3}{5}}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\sqrt{\frac{3}{5}}\right)$$

solve rest part in H.W.

Gauss Chebyshev Quadrature Formula.

$$\int_a^b w(x) f(x)\, dx \simeq \sum_{k=1}^{n} \lambda_k f(x_k)$$

$$a = -1 \qquad b = 1 \qquad w(x) = \frac{1}{\sqrt{1-x^2}}$$

where x_k are zeros of Chebyshev Polynomials.

$$T_n(x) = \cos(n \cos^{-1} x)$$

$$x_k = \cos\left(\frac{2k-1}{2n}\right)\pi \qquad k = 1, 2 \ldots n$$

on $[-1, 1]$

3-point Gauss - Chebychev Q.F:

$n=3$ in $\quad n$-point $- \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\,dx \simeq \sum_{k=1}^{n} \lambda_k f(x_k)$

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\,dx \simeq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \lambda_3 f(x_3)$$

$$x_k = \cos\left(\frac{2k-1}{2n}\right)\pi \qquad k=1,2,3$$

$$x_1 = \cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}$$

$$x_2 = \cos\frac{3\pi}{6} = 0$$

$$x_3 = \cos\frac{5\pi}{6} = -\frac{\sqrt{3}}{2}$$

[Zeroes of $T_3(x)$: $4x^3 - 3x = 0$

$x = 0, -\frac{\sqrt{3}}{2}, +\frac{\sqrt{3}}{2}$]

$$x_1 = -\frac{\sqrt{3}}{2} \qquad x_2 = 0, \qquad x_3 = \frac{\sqrt{3}}{2}$$

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\,dx = \lambda_1 f\left(-\frac{\sqrt{3}}{2}\right) + \lambda_2 f(0) + \lambda_3 f\left(\frac{\sqrt{3}}{2}\right)$$

$f(x)=1$ $\quad \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} \cdot 1\,dx = \lambda_1 + \lambda_2 + \lambda_3$

Put $x = \sin\theta$

$dx = \cos\theta\, d\theta$

$$2\int_0^{\frac{\pi}{2}} \frac{\cos\theta}{\cos\theta}\, d\theta = \pi$$

$$\lambda_1 + \lambda_2 + \lambda_3 = \pi \quad \text{—(1)}$$

$f(x) = x$ $\quad \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\, x\, dx = \lambda_1\left(\frac{-\sqrt{3}}{2}\right) + \lambda_2(0) + \lambda_3\left(\frac{\sqrt{3}}{2}\right)$

$= 0$ (odd func.) $\quad \lambda_1 - \lambda_3 = 0 \quad \text{—(2)}$

$\lambda_1 = \lambda_3$

$f(x) = x^2$ $\quad \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\, x^2\, dx = \frac{3}{4}\lambda_1 + \lambda_2(0) + \frac{3}{4}\lambda_3$

$= \frac{3}{4}(\lambda_1 + \lambda_3)$

Put $x = \sin\theta$

$$2\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta\, \cos\theta\, d\theta}{\cos\theta} = 2\int_0^{\frac{\pi}{2}} \left(\frac{1-\cos 2\theta}{2}\right) d\theta$$

$$= \frac{\pi}{2} - \left[\sin 2\theta\right]_0^{\frac{\pi}{2}}$$

$$= \frac{\pi}{2}$$

$$\Rightarrow \quad \lambda_1 + \lambda_3 = \frac{2\pi}{3} \quad \text{—(3)}$$

$$\lambda_1 = \frac{\pi}{3} \qquad \lambda_3 = \frac{\pi}{3} \qquad \lambda_2 = \pi - 2\frac{\pi}{3}$$

$$\lambda_2 = \frac{\pi}{3}$$

$$\lambda_1 = \lambda_2 = \lambda_3 = \frac{\pi}{3}$$

[In general, in n-point Chebychev, all weights are equal.]

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\,dx \simeq \frac{\pi}{3} f\left(-\frac{\sqrt{3}}{2}\right) + \frac{\pi}{3} f(0) + \frac{\pi}{3} f\left(\frac{\sqrt{3}}{2}\right)$$

Error :

$$c \frac{f^{(6)}(\xi)}{6!}$$

$$= \frac{\pi}{23040} f^{(6)}(\xi) \qquad \text{(verify)}$$

2-point Gauss Chebychev Quadrature Formula.

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\,dx \simeq \lambda_1 f(x_1) + \lambda_2 f(x_2)$$

x_1 and x_2 are zeros of $T_2(x)$

$$x_k = \cos\left(\frac{2k-1}{2n}\right) \qquad k = 1, 2 \ldots n$$

$$T_2(x) = 2x^2 - 1$$

$$x = \pm \frac{1}{\sqrt{2}}$$

$$x_1 = \frac{-1}{\sqrt{2}} \qquad x_2 = \frac{1}{\sqrt{2}}$$

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\, dx = \lambda_1 f\left(\frac{-1}{\sqrt{2}}\right) + \lambda_2 f\left(\frac{1}{\sqrt{2}}\right)$$

$$f(x) = 1 \qquad \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\, dx = \lambda_1 + \lambda_2$$

$$\lambda_1 + \lambda_2 = \pi$$

$$f(x) = x \qquad \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\, x\, dx = \frac{-1}{\sqrt{2}}\lambda_1 + \frac{1}{\sqrt{2}}\lambda_2$$

$$\lambda_1 - \lambda_2 = 0$$

$$\Rightarrow \lambda_1 = \lambda_2 = \frac{\pi}{2}$$

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\, dx = \frac{\pi}{2} f\left(\frac{-1}{\sqrt{2}}\right) + \frac{\pi}{2} f\left(\frac{1}{\sqrt{2}}\right)$$

4 - point Gauss - Chebychev Q.f.

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\, dx \approx \frac{\pi}{4}\left[f\left(\cos\frac{\pi}{8}\right) + f\left(\cos\frac{3\pi}{8}\right) + f\left(\cos\frac{5\pi}{8}\right) + f\left(\cos\frac{7\pi}{8}\right)\right]$$

$$T_4(x) = 8x^4 - 8x^2 + 1$$

$$x_k = \cos\frac{(2k-1)}{2n}\pi \qquad k = 1, 2, 3, 4.$$

$$x_1 = \cos\frac{\pi}{8} \qquad x_2 = \cos\frac{3\pi}{8} \qquad x_3 = \cos\frac{5\pi}{8} \qquad x_4 = \cos\frac{7\pi}{8}$$

H.W.

$$\int_0^\infty e^{-x} f(x) \simeq \lambda_1 f(x_1) + \lambda_2 f(x_2)$$

→ Solve

NOTE.

$$\int_0^\infty \quad , \quad \int_{-\infty}^{\infty}$$

Gauss - Lagrange

$$\int_0^\infty w(x) f(x)\, dx$$

$$w(x) = e^{-x}$$

Gauss - Hermite

$$\int_{-\infty}^{\infty} w(x) f(x)\, dx$$

$$w(x) = e^{-x^2}$$

April 12 5-7pm Extra class

SYSTEM OF LINEAR ALGEBRAIC EQUATIONS

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 \cdots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m
\end{aligned}
$$

$$\Leftrightarrow \quad Ax = b$$

where $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & & & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & & & a_{mn} \end{bmatrix}_{m \times n}$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n \times 1}, \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}_{m \times 1}$$

Methods to solve $Ax = b$

Direct

Methods which produce exact solution to the problem in the absence of round-off errors

Iterative

Methods which produce sequence of approximations to the solution $\{x^k\}$,

$\{x^k\} \longrightarrow \{x\}$ as $k \to \infty$

or $\|x - x^k\| < \varepsilon$

Problem: Given A, b find x such that $Ax = b$

If $m = n$: no of equations is same as no. of unknowns

$m > n$: Over-determined system

$m < n$: Under-determined system → Always consistent

→ $Ax = b$ has unique solution iff

$$\text{rank}(A|B) = \text{rank}(A) = n = \text{no. of unknowns}$$

→ Assume $m = n$

$Ax = b$, $\text{rank}(A) = n$ i.e. $\det(A) \neq 0$.

→ $Ax = b$, $|A| \neq 0$ $\quad x = A^{-1}b$

This finding $A^{-1}b$ is NOT preferred numerically

Finding A^{-1} is 2.5 times costlier than solving $Ax = b$.

→ $\frac{2}{3}n^3$ flops (floating point operations) for LU decomposition

→ $2n^3$ flops needed to compute A^{-1}

(*) Finding A^{-1} is solving n linear systems

$$A \cdot B = I$$

$$\downarrow$$

$$(A^{-1})$$

→ Cramer's rule is also NOT preferred. It involves $(n+1)!$ additions $(n+2)!$ multiplications.

$$A \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix} = I = \begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix}$$

$\downarrow$

columns of B

$\rightarrow$ e_i are i^{th} coordinate vector in R^n

$$Ab_i = e_i \quad , \quad 1 \le i \le n$$

ith e_i = $\begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$ $\rightarrow$ at ith

$\rightarrow$ Cramer's rule is NOT recommended practically

$\downarrow$

$$x_i = \frac{\det B_i}{\det A} \quad , \quad B_i = \begin{bmatrix} & \begin{bmatrix} \\ \\ \end{bmatrix} & \end{bmatrix}$$

$\downarrow$ ith column of A.

$\rightarrow$ $(n+1)!$ additions

$(n+2)!$ multiplications

n divisions

For $n=10$

G.E. $\frac{2}{3}n^3$.

Some special cases

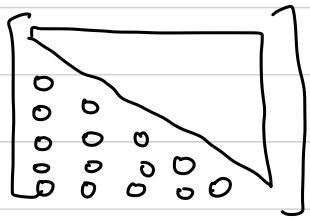
① If $A = D$ = diagonal matrix

i.e. $a_{ij} = 0$ when $i \neq j$

Solution of $Ax = b$ is immediate $\begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$

$$x_i = \frac{b_i}{d_i}, \quad i = 1, 2, \ldots n. \qquad d_i \neq 0$$

② $A = U =$ Upper triangular matrix



$A = [a_{ij}]$

$a_{ij} = 0$ if $i > j$

(strictly upper triangular)

Unit upper triangular : Upper triangular matrix in which diagonal entries are 1.

Upper triangular system

$$\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Solve using back substitution

$$u_{nn} x_n = b_n \qquad - \quad (1)$$

$$x_n = \frac{b_n}{u_{nn}}$$

$$u_{n-1,n-1} x_{n-1} + u_{n-1,n} x_n = b_{n-1} \qquad - (2)$$

$$x_{n-1} = \frac{b_{n-1} - u_{n-1,n} x_n}{u_{n-1,n-1}}$$

$$\vdots$$

$$x_1 = \frac{b_1 - \sum_{j=2}^{n} u_{ij} x_j}{u_{11}}$$

Subtraction is also considered addition

Total number of additions $= 0 + 1 + 2 \ldots + n-1$

$$= \frac{(n-1)(n-1+1)}{2}$$

$$= \frac{n(n-1)}{2}$$

Total number of multiplications $= 0 + 1 + 2 \ldots + n-1$

$$= \frac{n(n-1)}{2}$$

Total number of divisions $= n$.

(3) Lower Triangular Matrix

$A = [a_{ij}]$ is a lower triangular if $a_{ij} = 0$ when $i < j$.

Lower Triangular System

$$\begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & & \ddots & \\ l_{n1} & \cdots & & l_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Solve using Forward Elimination

$$l_{11} x_1 = b_1$$

$$x_1 = \frac{b_1}{l_{11}}$$

$$l_{21} x_1 + l_{22} x_2 = b_2$$

$$x_2 = \frac{b_2 - l_{21} x_1}{l_{22}}$$

$$\vdots$$

$$x_n = \frac{b_n - \sum_{j=1}^{n-1} l_{nj} x_j}{l_{nn}}$$

→ Operations count for Forward Elimination are same as those of Backward Substitution

$$Ax = b$$

$$[a_1 \; a_2 \; \cdots \; a_n]$$

↓

columns of A

$$Ax = \sum_{i=1}^{n} a_i x_i$$

↓

These are columns

→ Product / Inverse of two upper / lower triangular matrices is an upper / lower triangular matrix

$$U_1^{-1} = U_2 \qquad L_1^{-1} = L_2$$

$$U_1 U_2 \cdots U_n = U \qquad L_1 L_2 \cdots L_n = L$$

Gauss - Elimination

$$(A|b) \xrightarrow[\text{row - operations}]{\text{Elementary}} (U|C)$$

↓

upper triangular

Why not columns operations?

We have to simultaneously change x variables accordingly as well.

Column → Post multiplication

Row → Pre - multiplication

Gauss Elimination Method

For $k = 1, 2 \ldots , n-1$, carry out the following elimination step

$$\left[A^{(k)} \mid b^{(k)}\right] = \left[\begin{array}{cccc|c} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} & b_1^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} & b_2^{(2)} \\ & & \ddots & & \vdots \\ & & a_{kk}^{(k)} & \cdots a_{kn}^{(k)} & b_k^{(k)} \\ & & \vdots & & \vdots \\ & & a_{nk}^{(k)} & \cdots a_{nn}^{(k)} & b_n^{(k)} \end{array}\right] \rightarrow \text{in } k\text{th step}$$

$a_{22}^{(2)} \rightarrow$ means a_{22} is unchanged after second step

(Assume $a_{kk} \neq 0$)

Define the multipliers –

$$l_{ik} = \boxed{m_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}}, \quad i = k+1, \ldots n.$$

Perform the following operations to move

$$\left[A^{(k)} \mid b^{(k)}\right] \longrightarrow \left[A^{(k+1)} \mid b^{(k+1)}\right]$$

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik}\, a_{kj}^{(k)}, \quad i, j = k+1, \ldots n$$

$$b_i = b_i^{(k)} - m_{ik}\, b_k^{(k)}, \quad i = k+1, \ldots, n$$

- When $(n-1)$ th step is completed, the linear system will be in upper triangular form $Ux = c$
- Then solve by back substitution.

1. $A \longrightarrow U$
2. $B \longrightarrow C$
3. Back substitution $\left(\frac{n(n-1)}{2}\right) \rightarrow O(n^2)$

Operations Count

① $A \longrightarrow U$

No. of divisions at k^{th} step : $n - (k+1) + 1$

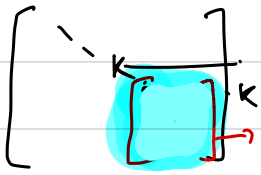
$= n - k$

∴ Total no of divisions in whole procedure

$= \sum_{k=1}^{n-1} (n-k)$

$= (n-1) + (n-2) \cdots + 2 + 1 = \frac{n(n-1)}{2}$

No. of multiplications in k^{th} step $= (n-k)^2$



→ $(n-k) \times (n-k)$ matrix.

→ All get updated by multiplication by m_{ik}

Total no. of multiplications in whole procedure

$= \sum_{k=1}^{n-1} (n-k)^2 = (n-1)^2 + (n-2)^2 \cdots + 2^2 + 1^2$

$= \frac{n(n-1)(2n-1)}{6}$

No. of additions at k^{th} step $= (n-k)^2$

Total no. of additions in whole process $= \frac{n(n-1)(2n-1)}{6}$

(2) B ⟶ C

No. of additions at k^{th} step $= n-k$

Total no. of additions in whole procedure

$$= \sum_{k=1}^{n-1} (n-k) = \frac{n(n-1)}{2}$$

$\begin{bmatrix} b_k \\ \int \end{bmatrix} \to (n-k)$

No. of multiplications in k^{th} step $= n-k$

Total no. of multiplications in whole procedure

$$= \sum_{k=1}^{n-1} (n-k) = \frac{n(n-1)}{2}$$

NO divisions because the division is in the multipliers are counted in A ⟶ U.

(3) Operation Count for Back Substitution

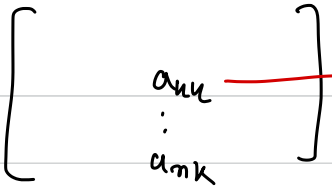
division $= n$

Multiplication $= \frac{n(n-1)}{2}$

Addition $= \frac{n(n-1)}{2}$

(1) + (2) + (3) gives total operation count in Gauss elimination

Partial Pivoting



→ inter-change a_{kk} with numerically largest element below it

So that we have numerical stability

'm' multiplier becomes small.

Pivoting Strategy

REQUIREMENT → for numerical stability

Partial pivoting → $|a_{pk}^{(k)}| = \max_{k \le i \le n} |a_{ik}^{(k)}|$

Interchange p^{th} and k^{th} row.

Complete pivoting. $|a_{pq}^{(k)}| = \max_{k \le i,j \le n} |a_{ij}^{(k)}|$

Interchange p^{th} and k^{th} row
q^{th} and k^{th} column. → NOTE x_k and x_q will also inter change

---

An example that shows the rounding errors with disastrous effects arise as a result of division by pivots which are too small.

Let us assume that floating point arithmatic is used with a three-digit mantissa, and in decimal (in order to fix ideas, and above all to facilitate calculation

In other words, the data and intermediate results of all calculations are rounded to 3 significant digits.

Consider the exact system : $10^{-4}x_1 + x_2 = 1$ —①

$x_1 + x_2 = 2$ —②

with exact solution $x_1 = 1.00010\ldots \simeq 1$

$x_2 = .99990\ldots \simeq 1$

Consider $a_{11} = 10^{-4}$ as pivot, as it is not zero, leading to the following procedure of elimination

$R_2 - 10^4 R_1$ $\qquad 10^{-4}x_1 + x_2 = 1$

$\qquad\qquad - 9999\,x_2 = -9999$

Since the number $-10^4 + 1 = -9999$ and $-10^4 + 2 = -9998$

So the solution found this way

$x_2 = 1$, $x_1 = 0$.

is very far from the true solution.

On the other hand, if we begin by inter-changing the two equations, that is to say if the pivot is the element $a_{21} = 1$, then we are led to the following calculations.

$x_1 + x_2 = 2$ —①

$10^{-4}x_1 + x_2 = 1$ —②

$R_2 - 10^{-4}R_1$ $\qquad x_1 + x_2 = 2$

$\qquad\qquad .999\,x_2 = .999$

Since the numbers $10^{-4} + 1 = .9999$ and $10^{-4} + 2 = .9998$

are both rounded to the same number .999 (chopping)
The current solution is $x_2 = 1$, $x_1 = 1$
is now very satisfactory.

This is why in practice we follow one of the two pivoting strategies

Diagonally Dominant Matrix

$$A = [a_{ij}]$$

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

Strictly diagonally dominant:

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \quad 1 \leq i \leq n$$

April 18

Q: When does LU factorisation exist?

Ans: When all leading principal submatrices are non-singular.

$A =$ 

are all non singular

Uniqueness

Idea: Suppose

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & & \\ l_{21} & l_{22} & \\ \vdots & & \ddots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} u_{11} & \cdots & u_{1n} \\ & u_{22} & \vdots \\ & & \ddots \\ & & u_{nn} \end{bmatrix}$$

The elements of L and U are calculated by comparing the elements of A with corresponding elements in the product

A has n^2 elements

LU has n^2+n unknowns $\left[\frac{n^2+n}{2} + \frac{n^2+n}{2}\right]$

If we reduce no. of unknowns to n^2 by fixing n unknowns, we can have a unique solution.

LU Decomposition of a matrix.

Let $A = [a_{ij}]$ be a square matrix of order n such that the leading principal submatrices

$$D_k = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ a_{21} & & a_{2k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}, \quad 1 \le k \le n$$

are invertible, then there exists a lower triangular matrix $L = [l_{ij}]$ with $l_{ii} = 1$, $1 \le i \le n$ and upper triangular matrix U such that $A = LU$.

Moreover, the factorisation is unique

$A = LU$ → usual LU factorisation.

↓

unit lower triangular matrix

This factorisation is called Doolittle LU factorisation

$A = L'U'$

↓

unit upper triangular matrix

This factorisation is called Crout's LU factorisation

How to get Crout's from Doolittle ?

Let $D = \text{diag}(u_{11}, \ldots u_{nn}) = \begin{bmatrix} u_{11} & & 0 \\ & u_{22} & \\ 0 & & u_{nn} \end{bmatrix}$

$$A = LU = L D D^{-1} U$$
$$= L' U'$$

$$\left[A^{(k)} \mid b^{(k)}\right] = \left[\begin{array}{cccc|c} a_{11}^{(1)} & \cdots & \cdots & a_{1n}^{(1)} & b_1^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} & b_2^{(2)} \\ & & \ddots & & \\ & & a_{kk}^{(k)} \cdots & a_{kn}^{(k)} & b_k^{(k)} \\ & & \vdots & & \\ & & a_{nk}^{(k)} \cdots & a_{nn}^{(k)} & b_n^{(k)} \end{array}\right]$$

$$m_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, \quad i = k+1, \ldots, n$$

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)}, \quad i, j = k+1, \ldots n$$

$$A^{(k)} \longrightarrow A^{(k+1)}$$

In matrix notation

[Gauss Transform]

$$G^{(k)} = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & -\frac{a_{k+1,k}^{(k)}}{a_{kk}^{(k)}} & \ddots & \\ & & \vdots & & \\ & & -\frac{a_{nk}^{(k)}}{a_{kk}^{(k)}} & & 1 \end{bmatrix}$$

diagonal entries are one

only entries below kth diagonal element are non-zero.

$$(G_1^{(k)})^{-1} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & \frac{a_{k+1,k}^{(k)}}{a_{kk}^{(k)}} & 1 & & \\ & & \vdots & & \ddots & \\ & & \frac{a_{nk}^{(k)}}{a_{kk}^{(k)}} & & & 1 \end{bmatrix}$$

negative sign disappears.

$$A^{(k+1)} = G_1^{(k)} A^{(k)} = \begin{bmatrix} a_{11}^{(1)} & - & - & - & a_{1n}^{(1)} \\ & a_{22}^{(2)} & & & a_{2n}^{(2)} \\ & & \ddots & & \\ & & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} \\ & & 0 & a_{k+1,k+1}^{(k+1)} \cdots & a_{k+1,n}^{(k+1)} \\ & & \vdots & \vdots & \\ & & 0 & a_{n,k+1}^{(k+1)} & a_{nn}^{(k+1)} \end{bmatrix}$$

Similarly, $b^{(k+1)} = G_1^{(k)} b^{(k)}$

$$A^{(n)} = G_1^{(n-1)} A^{(n-1)}$$

$$\| \quad = G_1^{(n-1)} G_1^{(n-2)} \cdots G_1^{(2)} G_1^{(1)} A$$

$$\cup$$

$$A = \left[G_1^{(n-1)} G_1^{(n-2)} \cdots G_1^{(2)} G_1^{(1)} \right]^{-1} U$$

$$\| \quad L$$

L unit lower triangular matrix.

[Product] inverse of a lower triangular matrix is a lower triangular matrix

where $L = \begin{bmatrix} 1 & & & & \\ l_{21} & 1 & & & \\ l_{31} & & 1 & & \\ \vdots & & & 1 & \\ l_{n1} & \cdots & & & l_{n,n-1}\ 1 \end{bmatrix}$

Gauss Jordan Method for Solution of $Ax = b$

$$[A \mid b] \longrightarrow [I \mid d]$$

$$[(\tilde{A}^{(k)} \mid b) \longrightarrow (\tilde{A}^{(k+1)} \mid \tilde{b}^{(k+1)})]$$

$$\tilde{A}^{(k+1)} = \begin{bmatrix} a_{11}^{(1)} & \cdots & a_{1k}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & & a_{2n}^{(2)} \\ & & \ddots & & \vdots \\ & & & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} \\ & & & \vdots & & \vdots \\ & & & a_{nk}^{(k)} & & a_{nn}^{(k)} \end{bmatrix}$$

kth column.

$$\tilde{G}^{(k)} = \begin{bmatrix} 1 & & 0 & -\frac{a_{1k}^{(k)}}{a_{kk}^{(k)}} & & \\ & 1 & & \vdots & & \\ & & 1 & & & \\ & & & \ddots & & \\ 0 & & & -\frac{a_{k+1,k}^{(k)}}{a_{kk}^{(k)}} & & \\ & & & \vdots & 1 & \\ & & & -\frac{a_{nk}^{(k)}}{a_{kk}^{(k)}} & & 1 \end{bmatrix}$$

$$\tilde{A}^{(k+1)} = \tilde{G}^{(k)} \tilde{A}^{(k)}$$

i.e. appropriate multiples are subtracted from the kth row to all other rows $(1, 2 \ldots k-1, k+1, \ldots n)$ so that all elements in the kth column of the $\tilde{A}^{(k)}$ become zero except k^{th} element

CHOLESKY FACTORIZATION

$$A = LL^T$$

$\downarrow$

lower triangular matrix

$$(A|I) \longrightarrow [I|B]$$

$$\Rightarrow B = A^{-1}$$

This is possible if and only if A is Symmetric Positive Definite

Symmetric Positive Definite (SPD)

$$\forall \; x \neq 0 \quad x^T A x > 0$$

NOTE: In $A = LL^T$, L need not be unit lower triangular matrix

April 19.

$$L = \begin{bmatrix} l_{11} & & \\ l_{21} & l_{22} & \\ \vdots & \ddots & \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \qquad L^T = \begin{bmatrix} l_{11} & \cdots & l_{n1} \\ & l_{22} & l_{n2} \\ & & \vdots \\ & & l_{nn} \end{bmatrix}$$

By matrix multiplication and equating coeff, proceeding top to bottom from left to right, we get successively the equation containing the unknown
On simplification -

Step 1 $\quad l_{11} = \sqrt{a_{11}}$

Step 2 $\quad l_{i1} = \dfrac{a_{i1}}{l_{11}}$

Step 3 $\quad l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2}$, $i = 1, 2, \ldots n$

Step 4 $\quad l_{ij} = \dfrac{1}{l_{ii}} \left[a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk} \right]$, $i = j+1, \ldots n$

Theorem (Cholesky Factorisation) If A is a symmetric positive definite matrix, then there exists at least one real lower triangular matrix B (L) such that $A = BB^T$ (LL^T)

Moreover, it is possible to require that the diagonal element of the matrix B should be positive, the corresponding factorisation in $A = BB^T$ is unique.

Problem Find Cholesky Factorisation of the matrix -

$$\begin{bmatrix} 2 & -1 & 2 \\ -1 & 1 & -1 \\ 2 & -1 & 3 \end{bmatrix}$$

Sol

$$\text{Let } A = LL^T = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

$$\begin{bmatrix} l_{11}^2 & l_{11}\,l_{21} & l_{11}\,l_{31} \\ l_{21}\,l_{11} & l_{21}^2 + l_{22}^2 & l_{21}\,l_{31} + l_{22}\,l_{32} \\ l_{31}\,l_{11} & l_{31}\,l_{21} + l_{32}\,l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 2 \\ -1 & 1 & -1 \\ 2 & -1 & 3 \end{bmatrix}$$

Row 1

$$l_{11} = \sqrt{2}$$

$$l_{21} = \frac{-1}{\sqrt{2}}$$

$$l_{31} = \sqrt{2}$$

Row 2

$$l_{21}^2 + l_{22}^2 = 1$$

$$\frac{1}{2} + l_{22}^2 = 1$$

$$l_{22} = \frac{1}{\sqrt{2}}$$

$$l_{21}\,l_{31} + l_{22}\,l_{32} = -1$$

$$\frac{-1}{\sqrt{2}}\sqrt{2} + \frac{1}{\sqrt{2}}\,l_{32} = -1$$

$$l_{32} = 0$$

Row 3

$$l_{31}^2 + l_{32}^2 + l_{33}^2 = 3$$

$$2 + 0 + l_{33}^2 = 3$$

$$l_{33} = 1$$

$$A = \begin{bmatrix} 2 & -1 & 2 \\ -2 & 1 & -1 \\ 2 & -1 & 3 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \sqrt{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & -\frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

If we want to find A^{-1}, $A^{-1} = (LL^T)^{-1} = (L^T)^{-1} L^{-1}$

$$= (L^{-1})^T (L^{-1})$$

Consider the linear system

$$\left.\begin{aligned} 5x_1 + 7x_2 &= .7 \\ 7x_1 + 10x_2 &= 1 \end{aligned}\right\} \circledast$$

has the solution $x_1 = 0 \quad x_2 = .1$

The perturbed system –

$$\begin{aligned} 5\hat{x_1} + 7\hat{x_2} &= .69 \\ 7\hat{x_1} + 10\hat{x_2} &= 1.01 \end{aligned}$$

This has the solution, $\hat{x_1} = -.17 \quad \hat{x_2} = 0.22$

Moral : A relatively small change in the right hand side of ⊛ has led to relatively large change in the solution

$$\begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 22 \\ 23 \\ 33 \\ 31 \end{bmatrix}$$

Solution $= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

Keep b as such and perturb $A \to \tilde{A}$, $\tilde{A}\tilde{x} = b$

$$\tilde{A} \begin{bmatrix} 10 & 7 & 8.1 & 7.2 \\ 7.08 & 5.04 & 6 & 5 \\ 8 & 5.98 & 9.89 & 9 \\ 6.99 & 4.99 & 9 & 9.98 \end{bmatrix} \qquad b = \begin{bmatrix} 22 \\ 23 \\ 33 \\ 31 \end{bmatrix}$$

Then $\tilde{x} = \begin{bmatrix} -81 \\ 137 \\ -34 \\ 22 \end{bmatrix}$

Hilbert Matrix

The Hilbert matrices -

$$H_n = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & & \frac{1}{n+1} \\ \vdots & & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & & \frac{1}{2n-1} \end{bmatrix}$$

$n = 1, 2, 3 \ldots$ are notoriously ill conditioned and

cond $(H_n) \longrightarrow \infty$ very rapidly as $n \to \infty$

$[K_1(H_n)]$

eg $H_3 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$, $H_3^{-1} = \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}$

$\|H_3\|_1 = \|H_3\|_\infty = \frac{11}{6}$

$\|H_3^{-1}\|_1 = \|H_3^{-1}\|_\infty = 408$

$K_1(H_3) = K_\infty(H_3) = 748$

$K_A = \|A\| \ \|A^{-1}\|$

↓

cond (A) Condition no.

For n as large as 6, the ill conditioning is extremely bad

$K_1(H_6) = K_\infty(H_6) \simeq 29 \times 10^6$!!

Vector norm

$\|\cdot\| \quad R^2 \longrightarrow R^2$

1. $\|x\| \geq 0$, $\|x\| = 0 \hookrightarrow x = 0$
2. $\|\alpha x\| = |\alpha| \ \|x\|$
3. $\|x+y\| \leq \|x\| + \|y\|$

Matrix norm

$\|\cdot\| \quad R^{m \times n} \longrightarrow R^1$

1. $\|A\| \geq 0$ $\|A\| = 0 \iff A = 0$
2. $\|\alpha A\| = |\alpha| \|A\|$
3. $\|A+B\| \leq \|A\| + \|B\|$

Subordinate matrix norm.

Given a matrix A and a vector norm $|\cdot|$ a non negative number defined by -

$$\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

$$\|Ax\| \leq \|A\| \, \|x\|$$

Two easily computable p-norm

$$\|A\|_1 = \max_{1 \leq j \leq n} \left(\sum_{i=1}^{n} |a_{ij}| \right)$$ max column sum norm

$$\|A\|_\infty = \max_{1 \leq i \leq n} \left(\sum_{j=1}^{n} |a_{ij}| \right)$$ max row sum norm.

Another useful p-norm is called Spectral Norm, denoted by $\|A\|_2$

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

It can be shown that $\|A\|_2 = \sqrt{\text{maximum eigen value of } (A^T A)}$

→ Eigen values of $A^T A$ are real and non-singular.

Matrix norm is consistent $\|AB\| \leq \|A\| \, \|B\|$

→ Matrix norm has submultiplicative property

Frobenius Norm $F(A) = \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2}$

Def. Spectral Radius $\rho(A) = \text{Max } |\lambda_i|$

λ_i are eigen values of 'A'

$$Ax = \lambda x$$
$$\lambda x = Ax$$
$$\|\lambda x\| = \|Ax\|$$
$$|\lambda| \, \|x\| \leq \|A\| \, \|x\| \qquad x \neq 0$$
$$|\lambda| \leq \|A\|$$

$$\Rightarrow \text{Max } |\lambda| \leq \|A\|$$
$$\boxed{\rho(A) \leq \|A\|}$$

Theorem:

Let A be any square matrix of order n. Then A^m converges to Zero matrix as $m \to \infty$ if and only if $\rho(A) < 1$.

//

$$Ax = b$$

perturb b
perturb A
perturb both A & b

→ then $\frac{\|x\|}{\|\delta x\|} \leq ?$

① Perturb b

$$A(x + \delta x) = b + \delta b$$
$$Ax + A\delta x = b + \delta b \qquad (\because Ax = b)$$
$$A\delta x = \delta b$$
$$\delta x = A^{-1} \delta b$$

$$\Rightarrow \quad \|\delta x\| = \|A^{-1}\,\delta b\|$$

$$\Rightarrow \quad \|\delta x\| \leq \|A^{-1}\| \;\|\delta b\| \qquad -(1)$$

Consider $b = Ax$

$$\|b\| = \|Ax\| \leq \|A\| \;\|x\|$$

$$\Rightarrow \quad \frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|} \qquad -(2)$$

From (1) and (2), we have

$$\boxed{\frac{\|\delta x\|}{\|x\|} \leq \{\|A\|\;\|A^{-1}\|\}\;\frac{\|\delta b\|}{\|b\|}}$$

i.e. Relative error in the solution is bounded by relative error in the data. So $\|A\|\;\|A^{-1}\|$ is the bound for magnifying factor for the relative error in the solution. This number is called the condition number of A and is denoted by $K(A)$

$$K(A) = \|A\|\;\|A^{-1}\|$$

Sometimes we represent Condition Number by cond(A)

NOTE : # For different norms, we get different condition numbers.

The matrix is called well-conditioned if the condition number is NOT far away from one. Otherwise A is ill conditioned

② Perturb A

Now we perturb A and compare the exact solution x, $x+\delta x$ of the system

$$Ax = b$$

$$(A+\delta A)(x+\delta x) = b$$

$$A(x+\delta x) + \delta A(x+\delta x) = b$$

$$A\delta x + \delta A(x+\delta x) = 0$$

$$A\delta x = -\delta A(x+\delta x)$$

$$\delta x = -A^{-1}\delta A(x+\delta x)$$

$$\Rightarrow \|\delta x\| \leq \|A^{-1}\| \|\delta A\| \|x+\delta x\|$$

$$\Rightarrow \frac{\|x\|}{\|x+\delta x\|} \leq \|A^{-1}\| \|\delta A\|$$

$$\Rightarrow \boxed{\frac{\|x\|}{\|x+\delta x\|} \leq \{\|A\| \|A^{-1}\|\} \frac{\|\delta A\|}{\|A\|}}$$

$$\boxed{\frac{\|\delta x\|}{\|x+\delta x\|} \leq K(A) \frac{\|\delta A\|}{\|A\|}}$$

April 21

Other Result

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\| \, \|A^{-1}\| \, \frac{\|\delta A\|}{\|A\|} \left(1 + O(\|\delta A\|)\right)$$

→ Let F be any square matrix. If for some operator norm, $\|F\| < 1$, then

$$\|(I+F)^{-1}\| \leq \frac{1}{1-\|F\|}$$

[Also $\|(I-F)^{-1}\| \leq \frac{1}{1-\|F\|}$]

Proof : Invertibility of $(I+F)$

Consider $(I+F)x = 0$ (Homogeneous system)

$x = 0$ guarantees invertibility of $I+F$.

(By contradiction)

Suppose $x \neq 0$

$$x + Fx = 0$$

$$Fx = -1x$$

$\Rightarrow -1$ is an eigen value of F

Since $\rho(F) \leq \|F\| < 1$

$\therefore$ -1 cannot be an eigenvalue of F

$\therefore$ The assumption $x \neq 0$ is wrong

$\Rightarrow (I+F)$ is invertible

$$(I+F)(I+F)^{-1} = I$$

$$I(I+F)^{-1} + F(I+F)^{-1} = I$$

$$(I+F)^{-1} = I - F(I+F)^{-1}$$

$$\|(I+F)^{-1}\| \leq 1 + \|F\| \, \|(I+F)^{-1}\|$$

$$\|(I+F)^{-1}\| \left(1 - \|F\|\right) \leq 1$$

$$\| (I+F)^{-1} \| \leq \frac{1}{1 - \|F\|}$$

Hence Proved.

[$\|I\| \geq 1$ for any norm
$\|I\| = 1$ for operator norm]

Theorem: Let A and B be square matrices of same order. Assume A is non-singular and suppose $\|A - B\| \leq \frac{1}{\|A^{-1}\|}$. Then B is also non-singular

[$\|A^{-1}\| \|A-B\| \leq 1$]

and $$\|B^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \|A - B\|}$$

Proof: $B = A - (A-B) = A(I - A^{-1}(A-B))$

The matrix $A[I - A^{-1}(A-B)]$ is non singular

Because of the result that if $\|F\| < 1$, then $(I-F)^{-1}$ exists and $\|(I-F)^{-1}\| \leq \frac{1}{1 - \|F\|}$

(Proved above)

[Also, given $\|A-B\| \leq \frac{1}{\|A^{-1}\|}$

$\Leftrightarrow \|A\|^{-1} \|A-B\| \leq 1$]

$$\|A^{-1}(A-B)\| \leq \|A^{-1}\| \|A-B\| \leq 1$$

$$\Rightarrow \|A^{-1}(A-B)\| \leq 1$$

B is non singular as it is product of 2 non-singular matrices

Also $B^{-1} = \left[A\left(I - A^{-1}(A-B)\right)\right]^{-1}$

$= \left(I - A^{-1}(A-B)\right)^{-1} A^{-1}$

$\Rightarrow \|B^{-1}\| \leq \|I - A^{-1}(A-B)\| \, \|A^{-1}\|$

$\leq \dfrac{\|A^{-1}\|}{1 - \|A^{-1}(A-B)\|}$

$\leq \dfrac{\|A^{-1}\|}{1 - \|A^{-1}\| \, \|A-B\|}$?

$\|A^{-1}(A-B)\| \leq \|A^{-1}\| \, \|A-B\|$

$-\|A^{-1}(A-B)\| \geq -\|A^{-1}\| \, \|A-B\|$

$1 - \|A^{-1}(A-B)\| \geq 1 - \|A^{-1}\| \, \|A-B\|$

$\Rightarrow \dfrac{1}{1 - \|A^{-1}(A-B)\|} \leq \dfrac{1}{1 - \|A^{-1}\| \, \|A-B\|}$

General Perturbation Theorem.

Consider the system $Ax = b$. Let δA and δb be perturbations of A and b and assume $\|\delta A\| \leq \dfrac{1}{\|A^{-1}\|}$

Then $A + \delta A$ is non singular and if we define δx implicitly by

$$(A + \delta A)(x + \delta x) = b + \delta b$$

Then $$\frac{\|\delta x\|}{\|x\|} \leq \frac{\text{cond}(A)}{1 - \text{cond}(A)\frac{\|\delta A\|}{\|A\|}} \left\{ \frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right\}$$

Proof:

Let A and B be square matrices of same order. Assume A is non singular and suppose that $\|A - B\| \leq \frac{1}{\|A^{-1}\|}$, then B is also non singular

and $$\|B^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \|A - B\|}$$

In this theorem, $B = A + \delta A$ $\therefore A - B = -\delta A$

$\therefore (A + \delta A)$ is non singular and

$$\|(A + \delta A)^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \|\delta A\|} \quad \text{—} \ (*)$$

Consider $(A + \delta A)(x + \delta x) = (b + \delta b)$

$$(A + \delta A)x + (A + \delta A)\delta x = b + \delta b$$

$$(\delta A)x + (A + \delta A)\delta x = \delta b$$

$$(A + \delta A)\delta x = \delta b - (\delta A)x$$

$$\|\delta x\| \leq \|(A + \delta A)^{-1}\| \, \|\delta b - (\delta A)x\|$$

$$\|\delta x\| \leq \|(A + \delta A)^{-1}\| \{ \|\delta b\| + \|\delta A\| \|x\| \}$$

$$\leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\,\|\delta A\|} \left\{ \|\delta b\| + \|\delta A\|\,\|x\| \right\}$$

(using ⊛)

$$\leq \frac{\|A\|\,\|A^{-1}\|}{1 - \|A\|\,\|A^{-1}\|\,\frac{\|\delta A\|}{\|A\|}} \left\{ \frac{\|\delta b\|}{\|A\|} + \frac{\|\delta A\|}{\|A\|}\|x\| \right\}$$

[Dividing num. & den. by $\|A\|$]

$$\|\delta x\| \leq \frac{\|A\|\,\|A^{-1}\|}{1 - \|A\|\,\|A^{-1}\|\,\frac{\|\delta A\|}{\|A\|}} \left\{ \frac{\|\delta b\|}{\|b\|}\|x\| + \frac{\|\delta A\|}{\|A\|}\|x\| \right\}$$

$\frac{\|\delta b\|}{\|b\|}$?

because $b = Ax$

$$\|b\| \leq \|A\|\,\|x\|$$

$$\frac{1}{\|A\|} \leq \frac{\|x\|}{\|b\|}.$$

$$\therefore \frac{\|\delta x\|}{\|x\|} \leq \frac{\text{cond}(A)}{1 - \text{cond}(A)\,\frac{\|\delta A\|}{\|A\|}} \left\{ \frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right\}$$

Problem Given the system $Ax = b$, where

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \end{bmatrix}$$

The vector b consists of 3 quantities measured with an error bounded by ε. Derive error bounds for –

(a) The components of x

(b) The sum of the components of $y = x_1 + x_2 + x_3$

$$A(x + \delta x) = b + \delta b$$

$= \hat{x}$ (say)

$$Ax + A\,\delta x = b + \delta b$$

$$\delta x = A^{-1}\,\delta b$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \to \begin{bmatrix} x_1+\delta x_1 \\ x_2+\delta x_2 \\ x_3+\delta x_3 \end{bmatrix}$$

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \to \begin{bmatrix} b_1+\delta b_1 \\ b_2+\delta b_2 \\ b_3+\delta b_3 \end{bmatrix}$$

$$A^{-1} = 12 \begin{bmatrix} 6 & -20 & 15 \\ -20 & 75 & -60 \\ 15 & -60 & 50 \end{bmatrix} \quad \text{(Verify !)}$$

$$\delta x = A^{-1}\,\delta b$$

$$\begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \delta x_3 \end{bmatrix} = A^{-1} \begin{bmatrix} \delta b_1 \\ \delta b_2 \\ \delta b_3 \end{bmatrix} = 12 \begin{bmatrix} 6 & -20 & 15 \\ -20 & 75 & -60 \\ 15 & -60 & 50 \end{bmatrix} \begin{bmatrix} \delta b_1 \\ \delta b_2 \\ \delta b_3 \end{bmatrix}$$

$$\delta x_1 = 12\,[6\,\delta b_1 - 20\,\delta b_2 + 15\,\delta b_3] \Rightarrow |\delta x_1| \le 12(6+20+15)\,\varepsilon = 492\,\varepsilon$$

$$\delta x_2 = 12\,[-20\,\delta b_1 + 75\,\delta b_2 - 60\,\delta b_3] \Rightarrow |\delta x_2| \le 12(155)\,\varepsilon = 1860\,\varepsilon$$

$$\delta x_3 = 12\,[15\,\delta b_1 - 60\,\delta b_2 + 50\,\delta b_3] \Rightarrow |\delta x_3| \le 12(125)\,\varepsilon = 1500\,\varepsilon$$

$$\sum_{i=1}^{3} \delta x_i = 12\,[\delta b_1 - 5\,\delta b_2 + 5\,\delta b_3]$$

∴ The error term for sum of components of $y = x_1 + x_2 + x_3$ is given by –

$$\delta y = \delta x_1 + \delta x_2 + \delta x_3 = 12\,(\delta b_1 - 5\,\delta b_2 + 5\,\delta b_3)$$

$$|\delta y| \le 12\,(1+5+5)\,\varepsilon$$

$$= 132\,\varepsilon$$

$$\Rightarrow |\delta y| \le 132\,\varepsilon$$

April 26

Iteration Methods or Splitting Method

$$Ax = b$$

Suppose $A = M - N$, when $|M| \neq 0$

then $Ax = b$ becomes $(M - N)x = b$

$$Mx = Nx + b$$

$$x = M^{-1}Nx + M^{-1}b$$

Now we can write the iterative method

$$\boxed{x^{k+1} = \underbrace{(M^{-1}N)}_{= H \text{ (say)}} x^k + \underbrace{M^{-1}b}_{= c}} \quad \text{— (A)}$$

$(M^{-1}N)$ ↓ iterative matrix

OR In general, we represent iteration method for the solution of $Ax = b$ is -

$$x^{(k+1)} = Hx^{(k)} + C \quad \text{— (*)}$$

If H is constant matrix, the iteration method (*) is known as stationary iteration method.

Some results :

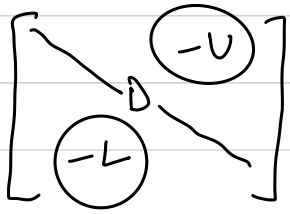
- The iteration method (*) for the solution of $Ax = b$ converges for any initial vector $x^{(0)}$ if $\|H\| < 1$.
- The iteration method (*) for the solution of $Ax = b$ converges for any initial vector $x^{(0)}$ if and only if $\rho(H) < 1$. (necessary)
- Rate of convergence of iterative method (*) is given by

$$R = -\log(\rho(H))$$

3 standard iteration methods –

① Jacobi or Gauss – Jacobi Iterative method
② Gauss Seidel Iteration Method
③ Successive Relaxation method.

Split $A = D - L - U$ ie. $A =$



Jacobi Method

Take $M = D$, then (A) becomes.
$N = L + U$.

$$x^{(k+1)} = D^{-1}(L+U)\,x^k + D^{-1}b \qquad -①$$

The iteration matrix $D^{-1}(L+U)$ is known as Jacobi Matrix and is denoted by J

① may be written as

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij}\, x_j^{(k)}\right) \quad i = 1, 2, \ldots n$$

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{\substack{j=1 \\ j<i}}^{i-1} a_{ij}\, x_j^{(k)} - \sum_{\substack{j=i+1 \\ j>i}}^{n} a_{ij}\, x_j^{(k)}\right)$$

$i = 1, 2 \ldots n$
$k = 0, 1 \ldots$

Gauss Seidel Iteration Method.

In $A = M - N$, if we take $M = D - L$ and $N = U$ then (A) becomes. $\left[\, x^{(k+1)} = (M^{-1}N)x^{(k)} - M^{-1}b \right.$

$$x^{(k+1)} = (D-L)^{-1}U\,x^k - (D-L)^{-1}b \qquad -②$$

The iteration matrix $(D-L)^{-1}U$ is denoted by L_1 and is known as Gauss Seidel Matrix

OR ② may be represented as -

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j<i} a_{ij} x_j^{(k+1)} - \sum_{j>i} a_{ij} x_j^{(k)} \right)$$

$i = 1, 2 \ldots n$

$k = 0, 1 \ldots$

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k)} \right)$$

useful for computation

✱ In Gauss Seidel Iteration method, latest values of x_i s are used

Relaxation Method or Successive Relaxation Method.

In the splitting $A = M - N$, if we choose, $M = \left(\frac{D}{\omega} - L\right)$ and $N = \left(\frac{1-\omega}{\omega} D + U\right)$, $\omega \neq 0$.

a real parameter, i.e. part of D matrix is shunted over to the matrix N.

which leads to the iteration method -

$$\Rightarrow \quad x^{(k+1)} = \left(\frac{D}{\omega} - L\right)^{-1} \left(\frac{1-\omega}{\omega} D + U\right) x^{(k)} + \left(\frac{D}{\omega} - L\right)^{-1} b. \qquad -③$$

The iteration matrix $L_\omega = \left(\frac{D}{\omega} - L\right)^{-1}\left(\frac{1-\omega}{\omega} D + U\right)$

$= (D - \omega L)^{-1}\left((1-\omega) D + \omega U\right)$

L_ω is called Relaxation matrix

→ It reduces to Gauss Seidel matrix for $\omega = 1$

☆ A necessary condition for relaxation method ③ To converge is that $0 < \omega < 2$. $\omega \in (0, 2)$

If the iteration method with $1 < \omega < 2$ or $0 < \omega < 1$ is called. → over relaxation, under relaxation.

In general a choice of ω in this range i.e. $\omega \in (0, 2)$ will NOT give the convergence. but in the important case that the coeff. matrix A is semi positive definite (spd), we have the following theorem

Theorem 1 : If A is spd, then for any $\omega \in (0,2)$ and any starting vector $x^{(0)}$, ③ converges to the solution of $Ax = b$

Theorem 2 : The Jacobi iterates converge for any $x^{(0)}$ if and only if the matrix $M + N$ is also positive definite

Theorem 3 : Assume that $A = M - N$ is spd (semi positive definite), then Gauss Seidel iterates converge to unique solution of $Ax = b$ for any starting vector $x^{(0)}$

Apr 26

Component form of ③ : Relaxation Method.

Given the current approximation $x^{(k)}$, we first compute the Gauss-Siedel iterate -

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j<i} a_{ij} x_j^{(k+1)} - \sum_{j>i} a_{ij} x_j^{(k)} \right) \quad (*)$$

as intermediate value, then take the final value of new approximation the i^{th} component to be -

$$x_i^{(k+1)} = x_i^{(k)} + \omega \left(\hat{x}_i^{(k+1)} - x_i^{(k)} \right)$$

Here ω is a parameter that has been introduced to accelerate the convergence

$$x_i^{(k+1)} = (1-\omega)\, x_i^{(k)} + \omega\, \hat{x}_i^{(k+1)} \quad - (**)$$

Combining (*) and (**), we get

$$x_i^{(k+1)} = \frac{\omega}{a_{ii}} \left(b_i - \sum_{\substack{j<i \\ \Downarrow \\ j=1}}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{\substack{j>i \\ \Downarrow \\ j=i+1}}^{n} a_{ij} x_j^{(k)} \right) - (\omega-1)\, x_i^{(k)}$$

$$x_i^{(k+1)} = \omega \, (\text{RHS of Gauss Seidel Method}) - (\omega-1)\, x_i^{(k)}$$

Problem Find the solution of the following system using Jacobi iteration method

$$2x_1 + x_2 + x_3 = 5$$
$$3x_1 + 5x_2 + 2x_3 = 15$$
$$2x_1 + x_2 + 4x_3 = 8$$

Jacobi. $$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j^{(k)} \right)$$

$$x_1^{(k+1)} = \frac{1}{2} \left(5 - x_2^{(k)} - x_3^{(k)} \right)$$

$$x_2^{(k+1)} = \frac{1}{5} \left(15 - 3x_1^{(k)} - 2x_3^{(k)} \right)$$

$$x_3^{(k+1)} = \frac{1}{4} \left(8 - 2x_1^{(k)} - x_2^{(k)} \right)$$

Assume initial value of x_1, x_2, x_3 to be zero.
i.e. $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 0$.

Ist iteration

$$x_1^{(1)} = \frac{1}{2} (5 - 0 - 0) = \frac{5}{2}$$

$$x_2^{(1)} = \frac{1}{5} (15 - 0 - 0) = 3$$

$$x_3^{(1)} = \frac{1}{4} (8 - 0 - 0) = 2$$

2nd iteration

$$x_1^{(2)} = \frac{1}{2} (5 - 3 - 2) = 0$$

$$x_2^{(2)} = \frac{1}{5} \left(15 - 3 \times \frac{5}{2} - 2 \times 2\right) = \frac{7}{10}$$

$$x_3^{(2)} = \frac{1}{4} \left(8 - 2 \times \frac{5}{2} - 3\right) = 0$$

3rd iteration

Problem Find the solution of the following system using Gauss-Seidel Iteration method

$$2x_1 + x_2 + x_3 = 5$$
$$3x_1 + 5x_2 + 2x_3 = 15$$
$$2x_1 + x_2 + 4x_3 = 8$$

$$x_i^{(k+1)} = \left(b_i - \sum_{j<i} a_{ij} \cdot x_j^{(k+1)} - \sum_{j>i} a_{ij} \cdot x_j^{(k)} \right)$$

Sol

$$x_1^{(k+1)} = \frac{1}{2}\left(5 - x_2^{(k)} - x_3^{(k)}\right)$$

$$x_2^{(k+1)} = \frac{1}{5}\left(15 - 3x_1^{(k+1)} - 2x_3^{(k)}\right)$$

$$x_3^{(k+1)} = \frac{1}{4}\left(8 - 2x_1^{(k+1)} - x_2^{(k+1)}\right)$$

Assume initial $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 0$

1st iteration.

$$x_1^{(1)} = \frac{1}{2}(5-0-0) = \frac{5}{2} = 1.5$$

$$x_2^{(1)} = \frac{1}{5}\left(15 - 3\times\frac{5}{2} - 0\right) = \frac{15}{10} = \frac{3}{2}$$

$$x_3^{(1)} = \frac{1}{4}\left(8 - 3\times\frac{5}{2} - \frac{3}{2}\right) = \frac{3}{8}$$

2nd iteration

H.W.

$Ax = b$

eg. ① $A = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{bmatrix}$ $\qquad \rho(J) < 1 < \rho(\mathcal{L}_1)$

Since $\rho(J) < 1$, Jacobi method works (converges)
Since $\rho(\mathcal{L}_1) > 1$, Gauss seidel method does not work (converge)

② $A = \begin{bmatrix} 2 & -1 & 1 \\ 2 & 2 & 2 \\ -1 & -1 & 2 \end{bmatrix}$ $\qquad \rho(\mathcal{L}_1) < 1 < \rho(J)$

Since $\rho(\mathcal{L}_1) > 1$, Gauss seidel method works (converges)
Since $\rho(J) < 1$, Jacobi method does not work (converge)

③ Assume that matrix A is strictly diagonally dominant

i.e.

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

[diagonally dominant → need not be non-singular
strictly dd → is non singular]

Then the Jacobi and Gauss-Seidel iteration converge to the solution of $Ax = b$ for any (initial) starting vector $x^{(0)}$

Proof Jacobi Iterative Method. Convergence

$\rho(J) \leq \|J\|$

$J = D^{-1}(L+U)$

$= D^{-1}(D - A)$

$[A = D - L - U \quad L + U = D - A]$

$$J = [J_{ij}]$$

$$J_{\substack{ij \\ i \neq j}} = \left(\frac{-a_{ij}}{a_{ii}} \right)$$

So $\|J\|_\infty = \max_{1 \le i \le n} \left(\sum_{j=1}^{n} |J_{ij}| \right)$ [max row sum]

$$= \max \left(\frac{1}{|a_{ii}|} \sum_{j=1}^{n} |a_{ij}| \right) \forall i$$

?

Because of strict diagonal dominance, each row sum in absolute value of J is < 1 $\Rightarrow \|J\|_\infty < 1$

$$\rho(J) \le \|J\|_\infty < 1$$

$\therefore$ $\rho(J) < 1$ $\Rightarrow$ Convergence of Jacobi method.

Gauss Seidel Convergence

$$L_1 = (D-L)^{-1} U$$

Let λ be an eigen value of L_1 and v be the corresponding eigen vector.

$$\lambda v = L_1 v = (D-L)^{-1} U v$$

$$\lambda (D-L) v = U v \quad -(1)$$

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Let $|v_k| = \max \{ |v_i| , i = 1, 2 \dots n \}$ $-(2)$

The kth equation of (1) is

$$\lambda \left(a_{kk} v_k + \sum_{j<k} a_{kj} v_j \right) = - \sum_{j>k} a_{kj} v_j \quad -(3)$$

Let $\alpha = \sum_{j<k} \frac{a_{kj} v_j}{a_{kk} v_k}$, $\beta = \sum_{j>k} \frac{a_{kj} v_j}{a_{kk} v_k}$

Then ③ becomes, [dividing ③ by $a_{kk} v_k$]

$$\lambda (1+\alpha) = -\beta$$

$$\lambda = \frac{-\beta}{1+\lambda} = \frac{-\beta}{1-(-\alpha)}$$

$$|\lambda| = \frac{|\beta|}{|1-(-\alpha)|} \leq \frac{|\beta|}{1-|\alpha|}$$

$$|1-(-\alpha)| \geq |1-|\alpha|| \geq 1-|\alpha|$$

as $|\alpha| < 1$

$$|a_{kk}| > \sum_{j<k} |a_{kj}| + \sum_{j>k} |a_{kj}|$$

$$|a_{kk}||v_k| \geq \sum_{j<k} |a_{kj} v_j| + \sum_{j>k} |a_{kj} v_j|$$

$$|a_{kk} v_k| - \sum_{j<k} |a_{kj} v_j| \geq \sum_{j>k} |a_{kj} v_j|$$

$$1 > \frac{\sum_{j>k} |a_{kj} v_j|}{|a_{kk} v_k| - \sum_{j<k} |a_{kj} v_j|}$$

$$1 > \frac{\sum_{j>k} |a_{kj} v_j|}{|a_{kk} v_k|}$$

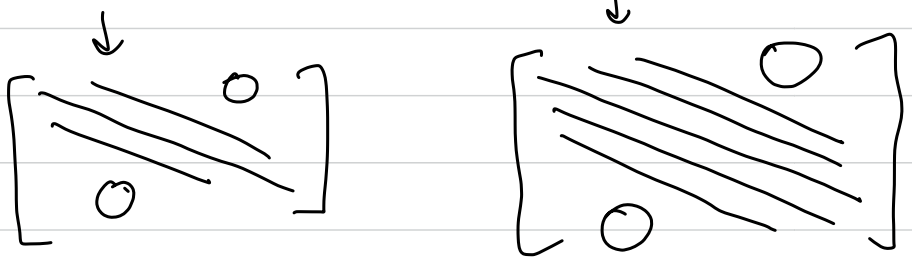
$\therefore$ $|v_k > v_j|$

Def A - Property

A matrix B is said to have A property if there exists a permutation matrix P such that

$$P B P^T = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$ where B_{11} B_{22} are diagonal matrices.

eg - Tridiagonal, Pentadiagonal matrices have A property



In $Ax = b$, if A is symmetric and having "A property", then W_{opt} is given by

$$W_{opt} = \frac{2}{1 + \sqrt{1 - (\rho(J))^2}}$$

where $\rho(J)$ is the spectral radius of Jacobi matrix

SOR → Successive Over Relaxation method.

$W \in (0, 2)$ $\quad$ $W_{opt} \in (1, 2)$ in SOR

April 28.

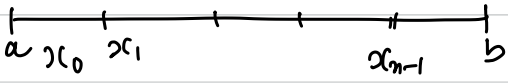
Numerical Solution of ODE

$$y' = f(x, y) \qquad y(x_0)$$

Analytical Solution Vs Numerical Solution.

Numerical solution of first order ODE -

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



$x_k - x_{k-1} = h$ [equi spaced]

To find numerically the solution at the points

Numerical methods

Single step methods
Method which requires solution at the abscissa at x_n to find solution at x_{n+1}

Multi-step methods.
Method which requires solution at more than one preceding point say x_n, x_{n-1}

k-step method. - The method requires the solution at k points $x_n, x_{n-1} \ldots x_{n-k+1}$ to find solution at x_{n+1}

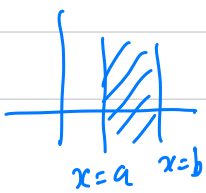
Initial Value Problem (IVP)

Theorem : A unique solution of IVP $y' = f(x, y)$, $y(x_0) = y_0$ exists in $[a, b]$ when -

i) $f(x, y)$ is continuous and bounded in the domain $a = x_0 \leq x \leq b$, $-\infty < y < \infty$

ii) There exists a constant L (Lipschitz constant) such that for $x \in [a, b]$ any $y, y^* \rightarrow$

$$|f(x, y) - f(x, y^*)| \leq L |y - y^*|$$



Single-step method.

Taylor-series method.

Suppose we know y at x_n is $y(x_n)$ is known;
To find $y(x_{n+1})$

$$y(x_{n+1}) = y(x_n + h), \quad \text{expanding by Taylor series}$$

$$= y(x_n) + h\,y'(x_n) + \frac{h^2}{2!}y''(x_n) \cdots + \frac{h^p}{p!}y^{(p)}(x_n) + \frac{h^{p+1}}{(p+1)!}y^{(p+1)}(\xi)$$

where $y' = f(x, y)$ [some implicit function of x, y] $\quad x_n < \xi < x_{n+1}$

Neglecting the last term -

$$y(x_{n+1}) \approx y(x_n) + h\,y'(x_n) + \cdots + \frac{h^p}{p!}y^{(p)}(x_n)$$

The neglected term $\frac{h^{p+1}}{(p+1)!}f^{(p+1)}(\xi)$ is called local truncation error.

Denoting the approximation for $y(x_n)$ by y_n, we get

$$y(x_{n+1}) = y_{n+1} \approx y_n + h\,y_n' + \cdots + \frac{h^p}{p!}y_n^{(p)}$$

where y_n', y_n'' .. are approximation of $y'(x_n)$, $y''(x_n)$ ...

∴ The approximation y_{n+1} for $y(x_{n+1})$ is given by -

$$y_{n+1} = y_n + h y_n' + \frac{h^2}{2!} y_n'' + \cdots + \frac{h^p}{p!} y_n^{(p)} \qquad \text{— (1)}$$

This is called **pth order Taylor Series Method**.

It may be noted that $y_0 = y(x_0)$ and $y_n \simeq y(x_n)$ for $x \geq 1$

LTE (Local Truncation Error)

$$y(x_{n+1}) = y(x_n) + h\, y'(x_n) + \frac{h^2}{2!} y''(x_n) \cdots$$

$$\cdots + \frac{h^p}{p!} y^{(p)}(x_n) + \frac{h^{p+1}}{(p+1)!} y^{(p+1)}(\zeta)$$

The neglected term $\frac{h^{p+1}}{(p+1)!} f^{(p+1)}(\zeta)$ is called local truncation error.

TTE [Total Truncation Error]

$y(x_{n+1}) - y_{n+1}$ where y_{n+1} is given by (1)

TTE at x_{n+1} is denoted by e_{n+1}

$$e_{n+1} = y(x_{n+1}) - y_{n+1}$$

$$TTE > LTE$$

Order of a Numerical Method.

If the LTE associated with a numerical method is of order $p+1$, the method is called a pth order method.

In Taylor Series method, if $p=1$, it is called Euler method

Therefore, Euler method is given by -

$$y_{n+1} = y_n + h y_n'$$

i.e. $$y_{n+1} = y_n + h f(x_n, y_n) \qquad n = 0, 1, 2 \ldots$$

$$\begin{cases} y_1 = y_0 + h f(x_0, y_0) \\ y_2 = y_1 + h f(x_1, y_1) \\ y_3 = y_2 + h f(x_2, y_2) \end{cases}$$

Convergence of a Numerical Method.

Let y_n be an approximate solution to the exact value of $y(x_n)$ of y at x_n, where $x_n = x_0 + nh$

If $\lim_{h \to 0} y_n = y(x_n)$, then the method by which y_n is computed is said to be convergent

i.e. $$\lim_{h \to 0} [y(x_n) - y_n] \longrightarrow 0$$

$$\lim_{n \to 0} e_n = 0$$

Lemma

If w_0, w_1, w_2 is a sequence of real numbers such that $w_{k+1} \leq (1+a) w_k + B$, where a and B are +ve number, then

$$w_n \leq B \left(\frac{e^{na} - 1}{a} \right) \quad \text{where } w_0 = 0$$

-(1)

Proof: This is proved by induction

Since $w_0 = 0$, the inequality $w_n \leq B\left(\frac{e^{na}-1}{a}\right)$

is satisfied. $\left(w_0 = 0 \leq \frac{B(e^0-1)}{a} = 0 \right.$

$\left. 0 \leq 0 \right)$

We assume it is true for $n = k-1$ and prove that it is true for $n = k$

We assume that $w_{k-1} \leq \frac{B\left(e^{(k-1)a}-1\right)}{a}$

From the hypothesis $w_k \leq (1+a)\, w_{k-1} + B$

$$\therefore w_k \leq (1+a)\, B\left(\frac{e^{(k-1)a}-1}{a}\right) + B$$

$$\leq B\left[(1+a)\left(\frac{e^{(k-1)a}-1}{a}\right) + 1\right.$$

$$\leq B\left[\frac{(1+a)\, e^{(k-1)a} - 1 - a + a}{a}\right]$$

$$\leq B\left[\frac{(1+a)\, e^{(k-1)a} - 1}{a}\right]$$

Since a, B are +ve, $1+a < e^a$

$$w_k \leq B\left[\frac{e^a\, e^{(k-1)a} - 1}{a}\right]$$

$$w_k \leq B\left[\frac{e^{ka}-1}{a}\right]$$

Therefore, the inequality (1) is true for $n=k$.

Convergence of Taylor Series Method.

Theorem: Let $h\phi(x_n, y(x_n), h) \equiv h y'(x_n) + \frac{h^2}{2!} y''(x_n) + \cdots + \frac{h^p}{p!} y^{(p)}(x_n)$

If –

i) $f(x,y)$ and its derivatives upto p^{th} order are continuous

ii) There exists a constant L such that

$$|\phi(x, y(x), h) - \phi(x, y^*(x), h)| \leq L|y(x) - y^*(x)|$$

iii) $|y^{(p+1)}(x)| \leq M_{p+1}$ for $x \in [a,b]$

Then –

$$|e_n| = |y(x_n) - y_n| \leq \left[\frac{e^{L(x_n-a)} - 1}{L}\right] M_{p+1} \frac{h^p}{(p+1)!}$$

$\longrightarrow 0.$

as fast as $h^p \to 0$ as $h \to 0$

Proof: We have

$$y(x_{n+1}) = y(x_n) + h\phi(x_n, y(x_n), h) + \frac{h^{p+1}}{(p+1)!} y^{(p+1)}(\xi) \quad \text{— (1)}$$

$x_n < \xi < x_{n+1}$

The p^{th} order Taylor Series method is –

$$y_{n+1} = y_n + h y_n' + \frac{h^2}{2!} y_n'' + \cdots + \frac{h^p}{p!} y_n^{(p)}$$

$$y_{n+1} = y_n + h\,\phi(x_n, y_n, h) \qquad -②$$

Subtracting ② from ①

$$y(x_{n+1}) - y_{n+1} = [y(x_n) - y_n] + h\left[\phi(x_n, y(x_n), h) - \phi(x_n, y_n, h)\right] + \frac{h^{p+1}}{(p+1)!} y^{p+1}(\zeta)$$

$$\Rightarrow |y(x_{n+1}) - y_{n+1}| \leq |y(x_n) - y_n| + h\left|\phi(x_n, y(x_n), h) - \phi(x_n, y_n, h)\right| + \frac{h^{p+1}}{(p+1)!} y^{p+1}(\zeta)$$

By hypothesis, $\left|\phi(x_n, y(x_n), h) - \phi(x_n, y_n, h)\right| \leq L\,|y(x_n) - y_n|$

$$|y(x_{n+1}) - y_{n+1}| \leq |y(x_n - y_n| + hL|y(x_n) - y_n| + \frac{h^{p+1}}{(p+1)!} y^{p+1}(\zeta)$$

$$|e_{n+1}| \leq |e_n| + hL|e_n| + \frac{h^{p+1}}{(p+1)!} M_{p+1}$$

$$|e_{n+1}| \leq |e_n|(1 + \underbrace{hL}_{a}) + \underbrace{\frac{h^{p+1}}{(p+1)!} M_{p+1}}_{B} \quad \text{—(3)}$$

The sequence of real numbers $|e_0|, |e_1| \ldots$ satisfy the inequality where $e_0 = y(x_0) - y_0 = 0$

Using the lemma, it follows that —

$$|e_n| \leq \frac{h^{p+1}}{(p+1)!} M_{p+1} \left[\frac{e^{nhL} - 1}{hL}\right]$$

$$w_n \leq B\left[\frac{e^{na} - 1}{a}\right]$$

$$\leq \frac{h^p}{(p+1)!} M_{p+1} \left[\frac{e^{(x_n - x_0)L} - 1}{L}\right]$$

$[\because x_0 = a]$

$$\leq \frac{h^p}{(p+1)!} M_{p+1} \left[\frac{e^{(x_n - a)L} - 1}{L}\right]$$

$\to 0$ as $h \to 0$

$|e_n| \to 0$ as $h \to 0$

$\therefore e_n \to 0$ as $h \to 0$

$\therefore$ The method is convergent

Apr 29

Problem Find the solution of IVP $y' = -xy$, $y(0) = 1$ at $x = .1, .2$ using Taylor's second order method

Taylor's Second order method: $y_{n+1} = y_n + hy_n' + \frac{h^2}{2!} y_n''$

Here $x_0 = 0$, $y_0 = 1$, $h = .1$

Put $n = 0$ $\quad y_1 = y_0 + hy_0' + \frac{h^2}{2!} y_0''$

$y' = f(x, y) = -xy$ | $y_0' = -0 \times 1 = 0$

$y'' = -(y + xy')$ | $y_0'' = -(y_0 + x_0 y_0') = -y_0 = -1$

$\therefore y_1 = 1 + \frac{(.1)^2}{2}(-1) = .995$

$y(.1) \simeq .995$

Put $n = 1$

$y_2 = y_1 + hy_1' + \frac{h^2}{2!} y_1''$

$x_1 = .1$ $\quad y_1 = .995$, $\quad h = .1$

$y_1' = -.1 \times .995 = -0.0995$

$y_1'' = -(.995 + .1 \times (-.0995)) = -.9851$

$y_2 = .995 + .1(-.0995) + \frac{(.01)^2}{2!}(-.9851)$

$y_2 = .9801$

$y(.2) \simeq .9801$

MVT f is continuous in $[a, b]$, differentiable in (a, b), then $\exists$ a $\xi \in (a, b)$ such that

$$f(b) = f(a) + (b-a) f(\xi) \quad , a < \xi < b$$

Proof: $y' = f(x, y)$, if y is a solution

$$y(x_{n+1}) = y(x_n) + (x_{n+1} - x_n)\, y'(\xi)$$

$$x_n < \xi < x_{n+1}$$

If we take x_n, x_{n+1} as 2 consecutive nodal points, then $x_{n+1} - x_n \approx h$

$$\therefore y(x_{n+1}) = y(x_n) + h f(\xi, y(\xi))$$

Approximations to $y(x_{n+1})$ can be obtained by approximating the value $f(\xi, y(\xi))$

Approx 1 Approximate $f(\xi, y(\xi)) = \frac{1}{2}\left[f(x_n, y(x_n)) + f(x_{n+1}, y(x_{n+1}))\right]$

we get $y(x_{n+1}) \approx y(x_n) + \frac{h}{2}\left[f(x_n, y(x_n)) + f(x_{n+1}, y(x_{n+1}))\right]$ — ①

On RHS of ①, an approximate value of $y(x_{n+1})$ is obtained by the Euler's method,
When it is used, we get a better approximation

Euler's method : $y_{n+1} = y_n + h f(x_n, y_n)$

Substitute this on RHS of ①

$$y_{n+1} = y_n + \frac{h}{2}\left[f(x_n, y_n) + f(x_{n+1}, y_n + h f(x_n, y_n)) \right]$$

Using approximation, we get

$$y_{n+1} = y_n + \frac{h}{2}\left[f(x_n, y_n) + f(x_{n+1}, y_{n+1}) \right] \quad \text{--- (2)}$$

where on KUT, $y_{n+1} = y_n + h f(x_n, y_n)$ --- (3)

(2) is called Improved Euler's Method
or Trapezium / Trapezoidal method
or Heun's method

Implicit method → eg $y_{n+1} = y_n + \frac{h}{2}[f(x_n, y_n) + f(x_{n+1}, y_{n+1})]$

Explicit method → eg $y_{n+1} = y_n + h f(x_n, y_n)$

→ This is a special case of second order Range-Kutta Method.

Define
$$\left.\begin{aligned} k_1 &= h f(x_n, y_n) \\ k_2 &= h f(x_n + h, y_n + k_1) \end{aligned}\right\} \quad (4)$$

Now (2) can be written as

$$y_{n+1} = y_n + \frac{1}{2}\left[h f(x_n, y_n) + h f(x_n + h, y_n + k_1) \right]$$

Substitute (4), we get

$$\boxed{y_{n+1} = y_n + \frac{1}{2}(k_1 + k_2)} \quad -(5)$$

(5) Represents a second order Range Kutta method.

The improved Euler's method can be represented as a second order R-K method as

$$k_1 = h f(x_n, y_n)$$
$$k_2 = h f(x_n + h, y_n + k_1)$$
$$y_{n+1} = y_n + \frac{1}{2}k_1 + \frac{1}{2}k_2 = y_n + \frac{1}{2}(k_1 + k_2)$$

2nd order R-K method.

[= h * slope evaluated at some point]

$$k_1 = h f(x_n, y_n)$$
$$k_2 = h\,(x_n + \alpha_2 h\ ,\ y_n + \beta_2 k_1)$$

$$y_{n+1} = y_n + \frac{1}{2}(k_1 + k_2)$$

= previous point + weighted average of slopes

Approximation 2. From MVT, we have

$$y(x_{n+1}) = y(x_n) + h f(\xi, y(\xi))$$

Suppose we consider ξ to be the mid point of x_n and x_{n+1}

i.e. we assume $\xi = x_n + \frac{h}{2}$

$$\therefore\ y(x_{n+1}) \approx y(x_n) + h f\left((x_n + \tfrac{h}{2}), y(x_n + \tfrac{h}{2})\right) \quad -(1)$$

where $y(x_n + \frac{h}{2})$ is calculated by Euler's Method.

i.e. $y\left(x_n + \frac{h}{2}\right) \simeq y_n + \frac{h}{2} f(x_n, y_n)$ — (2)

From (1) and (2), we get the approximation y_{n+1} for $y(x_{n+1})$ by the formula.

$$y_{n+1} = y_n + h f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2} f(x_n, y_n)\right)$$

This is called Modified Euler's Method.

Put $k_1 = h f(x_n, y_n)$

$k_2 = h f\left(x_n + \frac{h}{2}, y_n + \frac{1}{2} k_2\right)$

$$y_{n+1} = y_n + h f\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right)$$

$$y_{n+1} = y_n + k_2$$

which is a special case of 2nd order R-K method.

Classical Range - Kutta Method (4th Order)

$k_1 = h f(x_n, y_n)$

$k_2 = h f\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right)$

$k_3 = h f\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right)$

$k_4 = h f\left(x_n + \frac{h}{2}, y_n + k_3\right)$

$$y_{n+1} \simeq y_n + \frac{1}{6} k_1 + \frac{2}{6} k_2 + \frac{2}{6} k_3 + \frac{1}{6} k_4$$

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

H.W. Find the solution of D.E. (IVP) $y' = 1 + xy$, $y(0) = 1$ by 4th order R-K method at $x = 0.1$ using step size 0.1.

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

Forward D.D. $\dfrac{y(x_{n+1}) - y(x_n)}{h} \simeq f(x_n, y(x_n))$

$$\Rightarrow y(x_{n+1}) = y(x_n) + h f(x_n, y(x_n))$$

Replace $y(x_n)$ by y_n
$y(x_{n+1})$ by y_{n+1}

We get $\boxed{y_{n+1} = y_n + h f(x_n, y_n)}$ → Euler's method.

Backward D.D. $\dfrac{y(x_n) - y(x_{n-1})}{h} \simeq f(x_n, y(x_n))$

or $y(x_n) \simeq y(x_{n-1}) + h f(x_n, y(x_n))$

or $y(x_{n+1}) \simeq y(x_n) + h f(x_{n+1}, y(x_{n+1}))$

$\boxed{y_{n+1} = y_n + h f(x_{n+1}, y_{n+1})}$ → More popular.

unconditionally stable

Backward Euler's Method.

Central Difference $\dfrac{y(x_{n+1}) - y(x_{n-1})}{2h} \simeq f(x_n, y(x_n))$

$$y(x_{n+1}) = y(x_{n-1}) + 2h f(x_n, y(x_n))$$

$$\boxed{y_{n+1} = y_{n-1} + 2h f(x_n, y_n)}$$

Mid point method.

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

$$\int_{x_n}^{x_{n+1}} y' \, dx = y(x_{n+1}) - y(x_n)$$

RHS cannot be computed directly since $y(x)$ is unknown

Using trapezoidal rule, we get

$$\int_{x_n}^{x_{n+1}} f(x, y) \, dx \approx \frac{1}{2} h \left[f(x_n, y(x_n)) + f(x_{n+1}, y(x_{n+1})) \right]$$

$$\Rightarrow \boxed{y_{n+1} = y_n + \frac{1}{2} h \left[f(x_n, y_n) + f(x_{n+1}, y_{n+1}) \right]}$$

Trapezoidal Method.

Numerical method for solution of $y' = f(x, y)$, $y(x_0) = y_0$

Q)

(i) How well the difference equations approximate the D. E. (Truncation error)

(2) What happens when step size h tends to zero. (convergence)

(3) How sensitive the difference equation to the perturbations in data (Stability)

$$f(x,y) = \lambda y$$

$$y' = \lambda y$$

$h < \frac{2}{|\lambda|}$

λ be -ve real with.

Euler's method

$y_{n+1} = y_n + h f(x_n, y_n)$ becomes

$$y_{n+1} = y_n + h\lambda y_n$$

$$= (1 + \lambda h)\, y_n$$

The solution is decreasing if the condition $|1 + \lambda h| < 1$ is satisfied

$$\Rightarrow \quad -2 < \lambda h < 1$$

Since $\lambda < 0$, we must choose h so small that

$$\boxed{h < \frac{2}{|\lambda|}}$$

If $|\lambda|$ is very large, we must choose a very small step length 'h' in the Euler's method to get a decreasing solution