Lecture 1

MA 214 Numerical Analysis

Instructor: Prof. Tony J. Puthenpurakal

email: tputhen@math.iitb.ac.in

Textbooks

1. Elementary Numerical Analysis by S.D. Conte and C. deBoor

2. Numerical Analysis by Richard L. Burden and J.D. Faires

80% Attendence Required

2 Quizes (10% each)	20%
Midsem	30%
Endsem	40%

Passing = min{40, 40% of max score}

Note last time around 30 students got above 90%. So passing marks ~ 36%.

Calculators are a must for this course
You need it in tutorials, quizes, midsem
and endsem.

All calculations will be done to
4 sig digit accuracy.
(SCI-4 in Casio calculators)

Note Programmable calculators,
tablet P.C's are not allowed in
exams.

Note: All angles will be in Radians.
All logarithms will be to base e

Introduction

What is Numerical Analysis?

I will explain it through examples

1. $\int_a^b f(x)\,dx$ exists for example when $f : [a,b] \to \mathbb{R}$ is continuous.

However in most cases it is impossible to compute it

Examples

1. $I_1 = \int_0^1 \sin(x^2)\,dx$

2. $I_2 = \int_0^1 e^{-x^2}\,dx$

3. $I_3 = \int_0^{1/4} \frac{1}{\sqrt{1-x^3}}\,dx$

Not only do we have to approximate the integrals, we also have to do it with pre-assigned accuracy

For example

approximate I_1 upto 10^{-6}

i.e $|I_1 - \text{approx}| \leq 10^{-6}$.

(2) Interpolation

Suppose you are given a function

$f: [0,2] \to \mathbb{R}$ at some values

$x_0, x_1, \ldots x_n$.

Problem Approximate $f(t)$ at

$t \in [0,2] \setminus \{x_0, x_1, \ldots, x_n\}$

Graphically



Goal

To find a curve passing through these points

There are two standard methods to do this job

1. Lagrange Interpolation

2. Piecewise Methods
   a) Piecewise Linear
   b) Cubic Spine Interpolation

(3) Intial value differential equations

$$\begin{cases} \frac{dy}{dx} = f(x,y) \\ y(x_0) = y_0 \end{cases}$$

In general not possible to find y exactly

Example $\frac{dy}{dx} = \sin(x+y^2)$

$y(0) = 1.$

However for many applications approximate value is enough.

Suppose in the previous example you have to approximate $y(1)$

All methods will first approximate inbetween point and then find $y(1)$

for example $y(0.1)$ is approximated first.

Then using this $y(0.2)$ is approximated

$y(0.3)$ is approximated using $y(0.2)$

$y(0.4)$ " " " $y(0.3)$

⋮

$y(1)$ " " " $y(0.9)$

This creates an additional issue and that is of error propagation.

(4) Eigenvalue and Eigenvectors

Let A be a $n \times n$ matrix.

Recall $\lambda \in \mathbb{R}$ is said to be an eigenvalue of A if there exists $\bar{x} \neq 0$ such that

$$A\bar{x} = \lambda \bar{x}$$

$\bar{x}$ is called eigen-vector corresponding to λ.

Question How do you find eigenvalues and eigenvectors ?

In applications the size of the matrix is large

$n \geq 10,000$ is common

$n \geq 1$ million for significant % of cases

So usual method of finding

$P(t) = |tI - A|$

and then finding roots of $P(t)$ is not feasible.

In practise A will have a dominant eigen value i.e $\exists\ \lambda_0$ s.t

$|\lambda_0| > |\lambda_i|$ for all other eigen values λ_i

and it is enough for applications to find

λ_0

Floating point Arithmetic

n-digit floating point number in base β has the form

$$n = \pm (.d_1 d_2 \cdots d_n)_\beta \beta^e$$

$(0.d_1 d_2 \cdots d_n)_\beta \longrightarrow$ mantissa

$e \longrightarrow$ exponent

for most computers $\beta = 2$

For calculators $\beta = 10$

$x \rightarrow$ real number

$fl(x) \leftarrow$ floating pt. representation of x

example

$\sqrt{0.5} = 7.071\ E-1$ in 4 sig digits

Suppose x^* is an approximation to x

$|x - x^*|$ = Absolute error

$\frac{|x - x^*|}{|x|}$:= relative error (provided $x \neq 0$)

Problem We do not know x
So how to find relative error?

$$\alpha = \frac{x - x^*}{x}$$

$$\frac{x - x^*}{x^*} = \frac{\alpha}{1 - \alpha} \approx \alpha \quad \text{if } \alpha \text{ is small}$$

Definition :- x^* is said to approximate x to t significant digits if

$$\left| \frac{x - x^*}{x^*} \right| \leq 5 \times 10^{-t}$$

(we are assuming $x \neq 0$)

Loss of significant digits

It is not true that if

1) $x \sim x^*$, $y \sim y^*$ } in 4 sig digits

$$\Rightarrow \begin{cases} x \pm y \sim x^* \pm y^* & \text{in 4 sig digits} \\ xy \sim x^* y^* & \text{" " "} \\ \frac{x}{y} \sim \frac{x^*}{y^*} & \text{" " "} \end{cases}$$

Things which create loss of significant digits

1) Subtraction of nearly equal quantities

2) division by a number which is close to zero

Examples

1. $f(x) = 1 - \cos x$

$\cos 0.01 = 1$ in 4 sig digits

$f(0.01) = 1 - 1 = 0$

Actual answer is $5\ E{-5}$ (i.e $5\ 10^{-5}$)

Loss of sig digits arises since $\cos(0.01) = 1$ in 4 sig digits

To avoid this

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

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

$$= \frac{\sin^2 x}{1 + \cos x}$$

$$f(0.01) = \frac{1\ E^{-4}}{1+1} = 5\ E^{-5}$$

Example 2

$$x^2 + 111.11x + 1.2121 = 0$$

$b^2 = 1.235\ E4$

$b^2 - 4ac = 1.234\ E4$

$\sqrt{b^2-4ac} = 1.111\ E2$

$= b$ in 4 sig digits

$$x_1 = \frac{-b + \sqrt{b^2-4ac}}{2a}$$

$= 0$

Again loss of sig digits occur because we are subtracting nearly equal quantities

$$x_1 = \frac{-b + \sqrt{b^2-4ac}}{2a} \times \frac{-b - \sqrt{b^2-4ac}}{-b - \sqrt{b^2-4ac}}$$

$$= \frac{-2c}{b + \sqrt{b^2-4ac}}$$

$$x_1 = \frac{1.212}{222.1} = 1.091\ E-2$$

correct upto 4 sig digits

Example 3

$$f(x) = \frac{x - \sin x}{\tan x}$$

$$f(0.01) = \frac{0.01 - 1E\text{-}2}{1E\text{-}2} = 0$$

Rewrite $$f(x) = \frac{x - \sin x}{\tan x} \cdot \frac{x + \sin x}{x + \sin x}$$

$$= \frac{x^2 - \sin^2 x}{\tan x (x + \sin x)}$$

again $$f(0.01) = \frac{1E\text{-}4 - 1E\text{-}4}{1E\text{-}2\,(*)} = 0$$

Actual value $f(0.01) \approx 1.667\,E\text{-}5$ 4 sig digits

So one has to use Taylor series

$$\sin x \approx x - \frac{x^3}{6} \qquad \tan x \approx x - \frac{x^3}{3}$$

$$f(x) \approx \frac{x^3/6}{x - x^3/3} \approx \frac{x^2/6}{1 - x^2/3}$$

$$f(0.01) = 1.667\ E\text{-}5$$

Error propagation

Once an error is commited it contaminates subsequent results

Error propagation is studied in terms of two related concepts

1. condition
2. instability

1. condition

condition $\Leftrightarrow$ sensitivity of $f(x)$ to changes in x

$$= \max \left\{ \frac{\frac{|f(x) - f(x^*)|}{|f(x)|}}{\frac{|x - x^*|}{|x|}} \;\; \forall : |x - x^*| \text{ small} \right\}$$

$$\approx \left| \frac{f'(x)\, x}{f(x)} \right|$$

The larger the condition, the more ill-conditioned the function is said to be

Examples

Example 1

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

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

$$\left| \frac{f'(x)\, x}{f(x)} \right| = \left| \frac{\frac{1}{2\sqrt{x}}\, x}{\sqrt{x}} \right| = 1/2$$

So taking square-roots is well-conditioned since it actually reduces the relative error

~~[illegible]~~

Example 2

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

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

large when $|x|$ is close to 1

What to do

Put $x = 1 - y$ $\quad y \sim 0$ if $x \sim 1$

$$f(x) = \frac{10}{1-(1-y)^2} \cong \frac{10}{2y-y^2} \underset{\uparrow}{\approx} \frac{10}{2y} = \frac{5}{y} = g(y)$$

$y \sim 0$

$g(y) = 5/y$

$$\left| \frac{g'(y)\, y}{g(y)} \right| = 1$$

Example of instability

$$\begin{cases} \dfrac{du_1}{dt} = 9u_1 + 24u_2 + 5\cos t - \frac{1}{3}\sin t \\ \dfrac{du_2}{dt} = -24u_1 - 51u_2 - 9\cos t + \frac{1}{3}\sin t \end{cases}$$

$$u_1(0) = \frac{4}{3} \qquad u_2(0) = 2/3$$

exact solution

$$u_1(t) = 2e^{-3t} - e^{-39t} + \frac{1}{3}\cos t$$

$$u_2(t) = -e^{-3t} + 2e^{-39t} - \frac{1}{3}\cos t$$

RK method of order 4 stepsize 0.1

$\tilde{u}_1$ approx to u_1

$\tilde{u}_2$ " " u_2

t	$u_1(t)$	$\tilde{u}_1(t)$
0.1	1.793061	−2.645169
0.2	1.423901	−18.45158
⋮	⋮	⋮
0.9	0.3416143	−695332
1.0	~~0.2799~~ 0.2796748	−3099671

t	$u_2(t)$	$\tilde{u}_2(t)$
0.1	−1.032001	7.844527
0.2	−0.8746809	38.8763)
⋮	⋮	⋮
0.9	−0.2744088	1390664
1.0	−0.2298877	6199352

Instability = sensitivity of the numerical process for the calculation of $f(x)$ from x to the inevitable rounding error committed in a calculator or a computer.

Example

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

$$\text{condition} = \left| \frac{f'(x)\,x}{f(x)} \right| = \frac{1}{2} \frac{x}{\sqrt{x+1}\,\sqrt{x}} \simeq \frac{1}{2}$$

So conditioning is good

$$f(12345) = \cancel{\text{11113}}\ 1.111\,E2 - 1.111\,E2$$

$$= 0.0$$

Actual value $= 4.5\ E-3$

So we analyze what ~~happens~~ goes wrong

$$x_0 = 12345$$

$$x_1 = f_1(x_0) = x_0 + 1 = 12346$$

$$x_2 = f_2(x_1) = \sqrt{x_1}$$

$$x_3 = f_3(x_0) = \sqrt{x_0}$$

$$f_4(t) = x_2 - t$$

$$x_4 = f_4(x_3) = f(x)$$

Condition of f_4 is

$$\left| \frac{f_4'(t)\, t}{f_4(t)} \right| = \left| \frac{t}{x_2 - t} \right|$$

f_4 is well conditioned except when $t = x_2$

In our example $x_2 - x_3 \approx 0.005$

$x_3 - t \approx 111.11$

So condition of $f_4 \approx 22,222$

$\approx 40,000$ times condn of f.

what to do

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

$$f(12345) = \frac{1}{1.111\,E+2 + 1.111\,E+2}$$

$\approx 4.502\ E-3$

correct upto 3 sig digits

Note It is possible to estimate the effects of instability by considering the rounding errors one at a time

Mathematical Preliminaries

1. Intermediate-value theorem for continuous function's

Let $f:[a,b] \longrightarrow \mathbb{R}$ be a continuous function,

$x_1, x_2 \in [a,b]$ and say

$$f(x_1) < \alpha < f(x_2)$$

Then $\exists\ c \in [a,b]$ such that $f(c) = \alpha$.

Corollary 1

Let $f:[a,b] \longrightarrow \mathbb{R}$ be a continuous function.

Let $x_1, x_2, \cdots, x_n \in [a,b]$ and let

$g_1, g_2, \cdots, g_n$ be real numbers all of one sign. Then

$$\sum_{i=1}^{n} f(x_i) g_i = f(\xi) \sum_{i=1}^{n} g_i$$

for some $\xi \in [a,b]$

Proof (sketch)

We may assume $g_i \geq 0$.

Without loss of generality. assume

$$f(x_1) = \min \{ f(x_i) : i = 1, \dots, n \}$$

$$f(x_n) = \max \{ f(x_{ni}) : i = 1, \dots, n \}$$

Then

$$f(x_1) \sum_{i=1}^{n} g_i \leq \sum_{i=1}^{n} f(x_i) g_i \leq f(x_n) \sum_{i=1}^{n} g_i$$

$$h(x) = f(x) \sum_{i=1}^{n} g_i \quad \text{is continuous}$$

$$h(x_1) \leq \sum_{i=1}^{n} f(x_i) g_i \leq h(x_2)$$

So by intermediate value theorem $\exists \xi$

s.t $h(\xi) = \sum_{i=1}^{n} f(x_i) g_i$.

Note $h(\xi) = f(\xi) \sum_{i=1}^{n} g_i$

Similarly one has the following:-

Cor 2 Let $g : [a,b] \longrightarrow \mathbb{R}$ be a non-negative (or non-positive) integrable function.

Let $f : [a,b] \longrightarrow \mathbb{R}$ be continuous

Then

$$\int_a^b f(x)\, g(x)\, dx = f(\xi) \int_a^b g(x)\, dx$$

for some $\xi \in [a,b]$.

Remark :- The assumption $g(x)$ is of one sign is essential.

Example $f(x) = g(x) = x$ $\quad x \in [-1,1]$

$$\int_{-1}^{1} f g\, dx = \int_{-1}^{1} x^2\, dx = \frac{2}{3}.$$

However $\int_{-1}^{1} g(x)\, dx = 0$

Existence of maximum and minimum of continuous functions.

Theorem :- Let $f: [a,b] \longrightarrow \mathbb{R}$ be continuous.

Then $\exists\ \alpha, \beta \in [a,b]$ such that

$$f(\alpha) \leq f(x) \leq f(\beta) \quad \forall\ x \in [a,b]$$

Let us recall

Theorem (Rolle's theorem)

Let $f: [a,b] \longrightarrow \mathbb{R}$ be a continuous function and assume $f: (a,b) \longrightarrow \mathbb{R}$ is differentiable.

If $f(a) = f(b)$ then $\exists\ \xi \in (a,b)$ such that $f'(\xi) = 0$

Rolle's theorem implies the famous mean value theorem.

Theorem (M.V.T)

Let $f: [a,b] \to \mathbb{R}$ be a continuous function.

Assume $f: (a,b) \to \mathbb{R}$ is differentiable

Then $$\frac{f(b)-f(a)}{b-a} = f'(\xi)$$

for some $\xi \in (a,b)$

Proof Apply Rolle's theorem to

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

Theorem (Taylor's formula with integral remainder)

If $f(x)$ has $(n+1)$ continuous derivatives on $[a,b]$ and c is some point in $[a,b]$.

Then for all $x \in [a,b]$

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2}(x-c)^2 + \cdots + \frac{f^{(n)}(c)}{n!}(x-c)^n + R_{n+1}(x)$$

where $$R_{n+1}(x) = \frac{1}{n!}\int_c^x (x-s)^n f^{(n+1)}(s)\, ds$$

Fundamental Theorem of Algebra

Let $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ $\quad a_i \in \mathbb{C}$

$a_n \neq 0 \quad n \geq 1$

It is easy to see that $p(x)$ has almost n roots.

However the following is non-trivial to prove

Theorem (FTA)

$p(x)$ has a root, i.e., $\exists\, \xi \in \mathbb{C}$ such that $p(\xi) = 0$

Measuring how fast sequences converge

Note $\frac{1}{n^p} \to 0$ for any $p > 0$

Intuitively $\frac{1}{n} \to 0$ more slowly

than $\frac{1}{n^2} \to 0$ and so-on

Definition :- Let $\{\alpha_n\}_{n \geq 1}$ and $\{\beta_n\}$ be two sequences. We say α_n is big-oh of β_n and we write

$$\alpha_n = O(\beta_n)$$

if $|\alpha_n| \leq K|\beta_n|$ for some constant K and for all $n \geq 0$.

Examples $\{\frac{1}{n}\}$, $\{\frac{1000}{n}\}$, $\{\frac{10}{n} - \frac{40}{n^2} + e^{-n}\}$

$\{\frac{1}{n^p}\}$ (here $p > 1$

are all $O(\frac{1}{n})$

Definition

Let $\{\alpha_n\}$, $\{\beta_n\}$ be sequences.

We say $\{\alpha_n\}$ is little-oh of $\{\beta_n\}$ and write it as $\alpha_n = o(\beta_n)$ if

$$\lim_{n\to\infty} \frac{\alpha_n}{\beta_n} = 0$$

Example

1) $\{1/n^2\}$
2) $\left\{\frac{1}{n \ln n}\right\}$

$\Big\} = o\left(\frac{1}{n}\right)$

Important Remark

A convergence order rate of $1/n$ is much too slow to be useful in calculations.

Example

$$\frac{\pi}{4} = \sum_{i=0}^{\infty} \frac{(-1)^i}{2i+1} = 1 - \sum_{j=1}^{\infty} \frac{2}{16j^2-1}$$

Set $\alpha_n = 1 - \sum_{j=1}^{n} \frac{2}{16j^2-1}$

The sequence $\{\alpha_n\}$ is monotonically decreasing.

Moreover

$$0 \le \alpha_n - \frac{\pi}{4} \le \frac{1}{4n+3} \qquad n = 1, 2, \cdots$$

To calculate $\pi/4$ correctly to within 10^{-6} we would need $10^6 \le 4n+3$ or roughly $n = 250{,}000$ calculations.

However round of error's in calculating $\alpha_{250,000}$ will be usually greater than 10^{-6}.

Here $\alpha_n = \pi/4 + O(\frac{1}{n})$

$\alpha_n \neq \pi/4 + o(\frac{1}{n})$