

## Numerical Methods ..

8/1/2020 - 10:00 am

- Solving problems where theory does not go/apply.
- Seek solutions to practical problems  
(good enough)
  - Efficiency
  - Accuracy
  - Precision  
[level of detail]
- Different from other branches of mathematics.  
One needs to execute a computation & find an answer.

## Examples :- (That we have seen)

- Roots of polynomials / function.

Solve :-  $f(x) = 0$   
for  $x \in \mathbb{R}$ .

- Solutions to linear system of equations.

$$Ax = b$$

$$x \in \mathbb{R}^n, A_{n \times n}, b_{n \times 1}$$

.  $\{(x_i, y_i) : 1 \leq i \leq n\}$  - Data

Relationship :  $y_i = f(x_i)$

+  $\mathbb{R}^m \ni y_i; x_i \in \mathbb{R}^d$

f - linear, polynomial

.

$$\int_a^b f(x) dx$$

$f: \mathbb{R} \rightarrow \mathbb{R}$   
bounded  
continuous

.

$$y: [0, T] \rightarrow \mathbb{R}$$

$f: \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$

$$\frac{dy}{dt} = f(t, y)$$

For each of the above, we :-

- Explore where theory begins / ends.
- Reach for cases where manual computation is not possible.
- Find the answer.

Cannot , then get as close to it  
as possible.

Question:- What are the limitations. ?

Machines can only store / work with numbers that can be represented by finitely many digits. [number of bits allocated]

$\Rightarrow$  There is a largest number & a smallest number

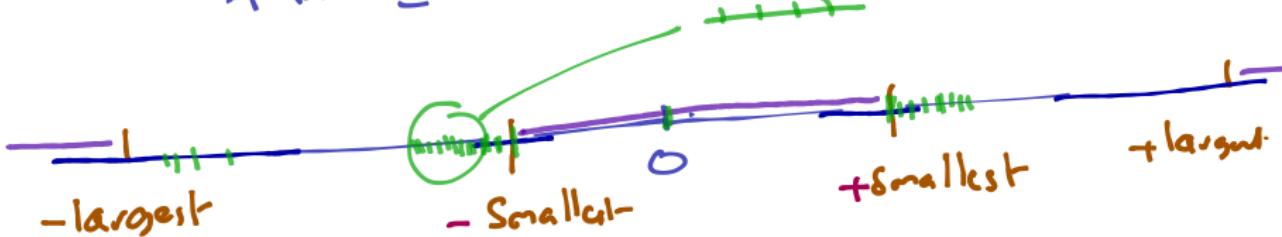
We are used to real line  $\mathbb{R}$ .

$a, b \in \mathbb{R}$  -  $\#\{x \in \mathbb{R} \mid a \leq x \leq b\}$  is infinite & we have spent classes understanding intuitively.

For a machine the real line does not exist. It works with floating point line. - fR

- There  
 $\hat{a}, \hat{b}$  are floating point numbers  
 $= \# \{ x \in fR \mid \hat{a} \leq x \leq \hat{b} \}$   
is finite (could be zero).

$$fR \subseteq \mathbb{R}$$



Range is limited.

⇒ Values created in any computations  
(add, subtracting, multiplying, dividing)  
will also have limited range.

Further, this will introduce errors

E.g.  $(\mathbb{R}, +, \times, -)$  is well understood

object.  
 $\mathbb{fR}$  is not "closed" under  
arithmetic operations. In the sense addition  
of floating point values cannot be  
represented always as another floating pt.  
value.

Introduces:

[for arithmetic]

- Round off error
- Catastrophic cancellation
- Truncation error

. keep track of each computation

&

also

Number of computations.

Size of Example is  $\wedge$   
theory A — results in computation  $a_n$   
 $b_n$   
B

We must know how to compare  $a_n < b_n$ .

## Stability issues of a Theory :-

Say we know a method to solve  
 $Ax = b$        $A_{n \times n}$        $b_{n \times 1}$   
 $x \in \mathbb{R}^n$

Suppose change  $A$   
 $\tilde{a}_{ij} = a_{ij} + \varepsilon$   $\leftarrow$  small #

$\tilde{x}$  solves  $\tilde{A}\tilde{x} = b$   
Is  $\tilde{x}$  close to  $x$ ?

# Some Computations : Jan 10 - 2pm

We wish to evaluate the polynomial

$$f(x) = a_0 + a_1 x + \dots + a_n x^n$$

To evaluate  $f(c)$

## Algorithm 1

$$\text{CoefH} = c(a_0, a_1, \dots, a_n)$$

$$x = c;$$

$$\text{Eval} = 0$$

$$\text{for } i = 1, \dots, n+1 \\ \quad \text{Eval} = \text{Eval} + \text{CoefH}[i] * x^{i-1}$$

end

In each loop the algorithm executes one addition and (i<sup>th</sup> loop) 'i' multiplications.

$\Rightarrow$  Total number of multiplication

$$\text{& addition, } = n + \sum_{i=0}^n i = n + \frac{n^2+n}{2}$$

Algorithm 2:

$$\text{CoefH} = c(a_0, a_1, \dots, a_n)$$

$$x = c$$

$$\text{Eval} = 0$$

for  $i = 1, \dots, n$

$$\text{Eval} = \text{Eval} + \text{CoefH}[i] \cdot C[x]$$

$$C[x] = C[x] \times x.$$

end  
return -Eval)

The above algorithm reduces multiplications  
to  $2n$ .

Horner's Method :- Rewrite

$$\begin{aligned}f(x) &= a_0 + a_1 x + \dots + a_n x^n \\&= a_0 + a_1 x + \dots + x^{n-1}(a_{n-1} + x a_n) \\&\vdots a_0 + x(a_1 + \dots + x(a_{n-1} + x a_n))\end{aligned}$$

Algorithm Horner  
 $\text{CoefH} = c(a_0, a_1, \dots, a_n)$   
 $x = c ;$   
 $\text{Eval} = 0$   
 for  $i = 1, \dots, n+1$   
 $\quad \text{Eval} = \text{Eval} * x + \text{CoefH}[i]$   
 end  
 $\text{return}(\text{Eval})$

In this method in each loop  
 there is one addition and one  
 multiplication.

- One must try to minimize the number of computations for any given situation.

