MATH 125 Lecture 1

In calculus, linear algebra and differential equations, we have studied various problem. A few of these problems are

i) finding roots
ii) Differentiation
iii) Integration
iv) Solving linear systems
v) Eigenvalue problems
vi) Differential equations
vii) optimization

Numerical analysis is the subject that involves the study, development and analysis of algorithms for obtaining numerical solutions to various mathematical problems

One might wonder or ask the reason for numerical analysis.

1. In most cases, we can not find exact solutions to mathematical problems
2. Finite precision computer arithmetic
   $\Rightarrow$ How do round off errors propogate in an algorithm
3. Existing analytical methods for some problems may not be scalable

Sources of error

- Mathematical formulation of problem
- Numerical solution

Mathematical

i) The model or mathematical statement is only an approximation to physical situation

Example classical mechanics → ignores relativistic effects

Black-Scholes model → constant volatility

Navier-Stokes → incompressible fluid

ii) Inaccuracies in physical data
- Due to error in empirical measurements
- difficult to model as there is typically random behaviour

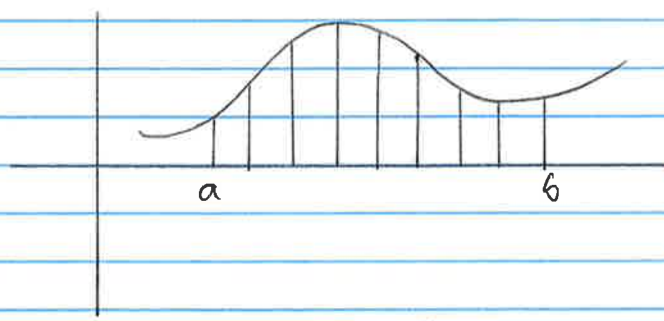
Computational

i) Error by programmer or user error

ii) Approximation error

Example $\sin(x) \approx x - \frac{x^3}{3!} + \frac{x^5}{5!}$

Example



Approximating area using numerical integration

Example Iterative methods
- They converge in the limit as number of iterations go to infinity
- Typically, few/minimal iterations but always finite

* We refer to these errors as truncation errors.

iii) Arithmetic calculations can never be done in complete accuracy
$\Rightarrow$ This error is called roundoff error

<u>classifying error</u>

Two measurements 2 ± 0.01 which measurement
$10^6 \pm 0.01$ do you "trust" more?

$$\underline{\text{Absolute error}} \equiv \text{Approximate value} - \text{True value}$$

$$\underline{\text{Relative error}} \quad \frac{\text{absolute error}}{\text{true value}}$$

Measurement 1 : Absolute error $= 0.01$
Relative error $= \dfrac{0.01}{2} = 0.005$

Measurement 2 : Absolute error $= 0.01$
Relative error $= \dfrac{0.01}{10^6} = 10^{-8}$

<u>Problem</u> compute $\sqrt{2}$

<u>solution</u> Find a number x such that $x^2 = 2$
Equivalently, $x = \dfrac{2}{x}$

<u>Initial guess</u> $x = 1$. Does it work? No
It is too low.
$\dfrac{2}{x} = 2$

<u>Average</u> $x_1 = \dfrac{1 + \frac{2}{1}}{2} = \dfrac{3}{2}$

$x_2 = \dfrac{\frac{3}{2} + \frac{4}{3}}{2} = \dfrac{17}{12}$

$\left(\frac{3}{2}\right)^2 > 2$

$\dfrac{2}{3/2} = 4/3$

$\Rightarrow \left(\frac{4}{3}\right)^2 < 2$

Next step $x_3 = \dfrac{\frac{17}{12} + \frac{24}{17}}{2} = \dfrac{577}{408}$

$x_3 \approx 1.414215686$

- YBC7289 Tablet dating 1750 B.C. was discovered near Baghdad in 1962

$\Rightarrow$ It is speculated that the above calculation was used by the Babylonians to compute $\sqrt{2}$

Interesting pattern

i) $3^2 = 2 \cdot 2^2 + 1$
$17^2 = 2 \cdot 12^2 + 1$ $\Longrightarrow$ $p^2 = 2q^2 + 1$
$577^2 = 2 \cdot 408^2 + 1$

i) $1 + \frac{1}{2} = \frac{3}{2}$

$$1 + \frac{1}{2+\frac{1}{2}} = 1 + \frac{1}{5/2} = \frac{7}{5}$$

$$1 + \frac{1}{2+\frac{1}{2+1/2}} = 1 + \frac{1}{2+\frac{1}{5/2}} = 1 + \frac{1}{2+\frac{2}{5}} = 1 + \frac{5}{12} = \frac{17}{12}$$

(Pell's numbers)

What is the underlying algorithm?

$$X_{n+1} = \frac{X_n + \frac{2}{X_n}}{2} \qquad n = 0, 1, 2, \ldots$$

This is called a fixed point iteration. One of the topics we study in this course is to understand when fixed point iterations converge

Loss of significance

consider evaluating $\frac{1-\cos(x)}{\sin^2(x)}$ as $x \to 0$

Infinite precision

$$T_1 = \frac{1-\cos(x)}{\sin^2(x)} \qquad T_2 = \frac{1-\cos(x)}{\sin^2(x)} \cdot \frac{1+\cos(x)}{1+\cos(x)}$$

$$= \frac{1-\cos^2(x)}{\sin^2(x)\cdot[1+\cos(x)]}$$

$$T_2 = \frac{\sin^2(x)}{\sin^2(x)\cdot[1+\cos(x)]} = \frac{1}{1+\cos(x)}$$

$T_1 \equiv T_2$ in infinite precision.

However, in finite precision aritametic, T_1 and T_2 differ as $x \to 0$

Main reason T_1 is prone to "concellation error"

Note that as $x \to 0$ $\cos(x) \to 1$
The numerator in T_1 is prone to loss of significance
This leads to getting 0 an as answer for T_1 as $x \to 0$

$\Delta \Longrightarrow$ However, T_2 avoids the concellation error.
As $x \to 0$, we obtain the correct limiting value of 0.5

---

computational complexity

In this class, each elementory operation $+ - \times \div$ costs 1 floating point operation (flop)

Exomple What is the cost of $x^T y$ where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$?

Answer $x^T y = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$

Addition $\equiv n-1$
Multiplication $\equiv n$

computation complexity $\equiv$ Total cost $= n + n - 1 = 2n - 1$

In practice, we are interested in the "scale" of the cost. For that, we rely on the Big O notation.

Let x_n and y_n be two different sequences
$x_n = O(y_n)$ if there are constants c and n_0 such that $|x_n| \le c|\alpha_n|$ when $n \ge n_0$

Let x_n and y_n be two different sequences
$x_n = o(y_n)$ if for any $c > 0$ there exists an integer n_0 such that $|x_n| < c|\alpha_n|$ when $n \ge n_0$

Examples

$$\frac{n+1}{n^2} = O\left(\frac{1}{n}\right)$$

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

Exercise Let $x_n = 2n+5$ and $y_n = n$. Is $x_n = o(y_n)$?

Solution choose any c
we have to find n_0 for which
$x_n < c\, y_n$ holds

Let $c = 3$ Can we find n_0 such that
$2n+5 < 3n$ works? Yes! $n_0 = 6$

Let $c = \frac{1}{10}$ can we find n_0 such that
$2n+5 < \frac{1}{10} n$ works? No positive n_0

Therefore, $x_n \ne o(y_n)$

Summary Asymptotic behaviour
- $x_n = O(y_n)$ The asymptotic growth of x_n is no faster than y_n
- $x_n = o(y_n)$ The asymptotic growth of x_n is strictly slower than y_n

Therefore, the cost of computing $x^T y$ for $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ is $O(n)$.

Evaluating a polynomial

What is the best way to evaluate
$$P(x) = 2x^4 + 3x^3 - 3x^2 + 5x - 1$$
at $x = 1/2$?

Method 1

$$P\left(\tfrac{1}{2}\right) = 2\left(\tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{1}{2}\right) + 3\left(\tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{1}{2}\right)$$
$$-3\left(\tfrac{1}{2}\cdot\tfrac{1}{2}\right) + 5\left(\tfrac{1}{2}\right) - 1$$

Multiplications $\equiv$ 10
Additions $\equiv$ 4
Total $\equiv$ 14

Method 2

Let's compute the following

$$\left(\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{2}\right) = \left(\tfrac{1}{2}\right)^2$$
$$\left(\tfrac{1}{2}\right)^2\cdot\left(\tfrac{1}{2}\right) = \left(\tfrac{1}{2}\right)^3$$
$$\left(\tfrac{1}{2}\right)^3\cdot\left(\tfrac{1}{2}\right) = \left(\tfrac{1}{2}\right)^4$$

store this

$$P\left(\tfrac{1}{2}\right) = 2\left(\tfrac{1}{2}\right)^4 + 3\left(\tfrac{1}{2}\right)^3 - 3\left(\tfrac{1}{2}\right)^2 + 5\left(\tfrac{1}{2}\right) - 1$$

Multiplications $\equiv$ 7
Additions $\equiv$ 4
Total $\equiv$ 11

Could we do even better?

Method 3

$$P(x) = -1 + x\,(5 - 3x + 3x^2 + 2x^3)$$
$$= -1 + x\,(5 + x\,(-3 + 3x + 2x^2))$$
$$= -1 + x\,(5 + x\,(-3 + x\,(3 + 2x)))$$
$$= -1 + \underbrace{x\,(5 + \underbrace{x\,(-3 + \underbrace{x\,(\underbrace{3 + 2\cdot x}_{①})}_{②})}_{③})}_{④}$$

Evaluate inside out

(1) costs 2
(2) costs 2 $\Rightarrow$ 4 multiplications $\equiv 8$
(3) costs 2 4 additions
(4) costs 2

This is Horner's method. Although attributed to Horner, this method was known in ancient persia and china

* For generic polynomials, Horner's method is optimal

---

order of convergence

$$\lim_{n\to\infty} x_n = L$$

For each $\varepsilon > 0$, there is a real number r such that $|x_n - L| < \varepsilon$ whenever $n > r$

Example $\displaystyle \lim_{n\to\infty} \frac{n+1}{n} = 1$

proof $\displaystyle \left| \frac{n+1}{n} - 1 \right| = \left| \frac{1}{n} \right| < \varepsilon$ whenever $n > \varepsilon^{-1}$

In practice, we are interested to "quantify" the rapidity of convergence.

Linear convergence $|x_{n+1} - x^*| \le c\,|x_n - x^*| \quad (n \ge N)$

$c < 1$

superlinear convergence $|x_{n+1} - x^*| \le \varepsilon_n\,|x_n - x^*| \quad (n \ge N)$

A sequence ε_n tending to zero

Quadratic convergence $|x_{n+1} - x^*| \le c\,|x_n - x^*|^2 \quad (n \ge N)$

Note c is positive but not necessarily less than 1

In general, $|x_{n+1} - x^*| \le c\,|x_n - x^*|^{\alpha} \quad (n \ge N)$

convergence of order α