MA 214 - Introduction to Numerical Analysis

Vishal Neeli

1 Interpolation Theory

1.1 Introduction

- Given finite set of points, reconstructing the original curve is interpolation.
- There will be obviously infinitely many curve.
- Interpolation problem
  Given n+1 real distinct points: $x_0, x_1, \ldots, x_n$ and real numbers: $y_0, y_1, \ldots, y_n$
  Find a function $f : \mathbb{R} \to \mathbb{R}$ such that
  $$f(x_i) = y_i \qquad \text{for } i = 0, 1, \ldots, n$$
  Such a function is called **interpolant** and points x_i are called **interpolation points**.
  We attempt to rebuild original function using polynomial functions. This is called polynomial interpolation and function is polynomial interpolant.
- A polynomial is function of the form
  $$p(x) = a_0 + a_1 x + \ldots + a_n x^n$$
- $\mathbb{P}_n$ is the set of polynomials consisting of all polynomials of degree $\leq n$

1.2 Polynomial Interpolation

Theorem (Joseph-Louis Lagrange Theorem). *Given $n + 1$ data points with unique x_is, then there exists a unique polynomial $p_n \in \mathbb{P}_n$ such that*
$$p(x_i) = y_i \qquad \textit{for } i = 0, 1, \ldots, n$$

Proof. (1) This can be shown by linear algebra. In a n degree polynomial, we substitute the points and get $n+1$ equations in n+1 variables (coeff) and all the rows are unique (since $x_0, x_1, \ldots, x_n$ are unique), hence in $AX = b$, $|A| \neq 0$. □

Proof. (2) Part 1: Uniqueness : If there is an interpolant, then the interpolant is unique
Let there be 2 interpolants, p_n and q_n and let $r(x) = p(x) - q(n)$,
$$r(x) = 0 \text{ for } i = 0, 1, \ldots, n$$
This contradicts the fundamental theorem of Algebra. (A polynomial of degree n can have at most n real roots). Therefore
$$r(x) = 0 \quad \forall x \in \mathbb{R}$$
$$p(x) = q(x) \quad \forall x \in \mathbb{R}$$
Part 2: Existence (construction):
Given n+1 data points, build n+1 Langrange polynomials
$$L_k^n(x) = \begin{cases} 0 & \text{for } i \neq k \\ 1 & \text{for } i = k \end{cases}$$
$$L_k^n(x) = \frac{(x - x_0)\ldots(x - x_{k-1})(x - x_{k+1})\ldots(x - x_n)}{(x_k - x_0)\ldots(x_k - x_{k-1})(x_k - x_{k+1})\ldots(x_k - x_n)}$$
$$p(x) = \sum_{k=0}^{n} y_k L_k^n(x)$$
□

1.3 Closeness between functions

Given two continuous functions $f, g : [a, b] \to \mathbb{R}$, to evaluate how close the functions are consider the following

$$max_{x\in[a,b]}|f(x) - g(x)|$$

1.4 Set of continuous Functions

$C[a, b]$ is the set of all continuous functions on [a,b]
$C[a, b]$ is a infinite dimensional vector space

$$f, g \in C \implies f + g \in C \text{ and } \lambda f \in C$$

We define norm on $C[a, b]$ as

$$||f|| = max_{x\in[a,b]}|f(x)|$$

$C^k[a, b]$ denotes the set of all functions which are continuously k-times differentiable

1.5 Polynomial Approximation and Error

Theorem (Weierstrass approximation Theorem)**.** *Given a function $f \in C[a, b]$ and given $\epsilon > 0$, there exists a polynomial $p(x)$ such that,*

$$||f(x) - p|| < \epsilon$$

Using Langrange's recipe to approximate
Take $n+1$ interpolation points in the [a,b] and collect the function values at all the points. We have $n + 1$ data points. Using Lagrange polynomials, find the interpolant

Theorem (Error equation)**.** *Let $f \in C^k[a, b]$, $x_0, x_1, \ldots, x_n \in [a, b]$ and $p \in \mathbb{P}_n$ be the interpolant using these points, then for all x, there exists a $\zeta = \zeta(x) \in (a, b)$ such that*

$$\boxed{f(x) - p(x) = \frac{1}{(n+1)!} f^{(n+1)}(\zeta) \prod_{k=0}^{n} (x - x_k)}$$

Note: *Here ζ is dependent on the x, i.e, for every x you choose, ζ generally changes.*

Proof. Consider the function,

$$\psi(t) = (f(t) - p(t)) \prod_{k=0}^{n} (x - x_k) - (f(x) - p(x)) \prod_{k=0}^{n} (t - x_k)$$

This $n + 2$ roots (n+1 data points and x), applying rolle theorem's gives us that $\phi^{(1)}(t)$ has at least n+1 roots. Applying like this repeatedly on its derivatives, we get that $f^{(n+1)}$ has at least 1 root in $[a, b]$. Assuming the root to be ζ. We have,

$$f(x) - p(x) = \frac{1}{(n+1)!} f^{(n+1)}(\zeta) \prod_{k=0}^{n} (x - x_k)$$

□

Approximating the error:
Taking norm on both the sides of error equation, we have,

$$\max_{x\in[a,b]} |f(x) - p(x)| = \frac{1}{(n+1)!} ||f^{(n+1)}(\zeta) \prod_{k=0}^{n} (x - x_k)|| \tag{1}$$

$$\max_{x\in[a,b]} |f(x) - p(x)| \leq \frac{1}{(n+1)!} ||f^{(n+1)}|| \; max_{x\in[a,b]} \prod_{k=0}^{n} (x - x_k) \tag{2}$$

Chebyshef interpolation points:

$$x_k = \frac{a+b}{2} + \frac{b-a}{2} cos\left(\frac{j\pi}{n}\right)$$

These points minimise $max_{x\in[a,b]} \prod_{k=0}^{n}(x - x_k)$ and therefore prefered over equally spaced points on real line. These points can be visualised as projections of equally spaced points on the arc of the semicircle with $\frac{a+b}{2}$ as center and $\frac{(b-a)}{2}$ as radius.

1.6 Some more methods for calculating interpolant

- This is similar to the linear algebra method (given as proof(1) to Joseph-Louis Lagrange Theorem) for finding the interpolant.
  Consider the polynomial
  $$p(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \ldots + a_n(x - x_0)...(x - x_{n-1})$$
  Find the coefficients $a_0, a_1, \ldots, a_n$ by substituting the data points. On substituting x_0, we get a_0, again on substituting x_1 and using a_0, we get a_1 and so on.
- Interpolant $p(x)$ of $x_0, x_1, \ldots, x_n$ can be calculated using interpolants $a(x)$ and $b(x)$ of $x_1, x_2, \ldots, x_n$ and $x_0, x_1, \ldots, x_{n-1}$ respectively as
  $$p(x) = \frac{(x - x_0)a(x) - (x - x_n)b(x)}{x_n - x_0}$$

1.7 Divided difference - recursion relation

- **Divided difference** : It is the coefficient of x_n in the interpolant $p \in \mathbb{P}_n$ and denoted by $f[x_0, x_1 \ldots, x_n]$.
  Using Langrange polynomials, we have
  $$p(x) = \sum_{k=0}^{n} f(x_k) \prod_{j=0, j\neq k}^{n} \frac{x - x_j}{x_k - x_j}$$
  So the divided difference is
  $$f[x_0, x_1, \ldots, x_n] = \sum_{k=0}^{n} f(x_k) \prod_{j=0, j\neq k}^{n} \frac{1}{x_k - x_j}$$

Theorem (Divided difference recursion theorem)**.**

$$\boxed{f[x_0, x_1, \ldots, x_{m+1}] = \frac{[f[x_1, x_2, \ldots, x_{m+1}] - f[x_0, x_1, \ldots, x_m]}{x_{m+1} - x_0}}$$

Proof. Let p(x) be the interpolant for $x_0, x_1, \ldots, x_m$ and q(x) be the interpolant for $x_1, x_2 \ldots, x_{m+1}$. Then,

$$L(x) = \frac{(x - x_0)q(x) + (x_{m+1} - x)p(x)}{x_{m+1} - x_0}$$

is an interpolant. Since, interpolant is unique, considering coeff of x_m we have,

$$f[x_0, x_1, \ldots, x_m] = \frac{f[x_1, x_2, \ldots, x_{m+1}] - f[x_0, x_1, \ldots, x_m]}{x_{m+1} - x_0}$$

□

Theorem (Interpolant using divided differences)**.** *Suppose $x_0, x_1, \ldots, x_n$ be the data points. Then interpolant $p \in \mathbb{P}_n$ is*

$$\boxed{p(x) = f[x_0] + f[x_0, x_1](x - x_0) + \ldots + f[x_0, x_1, \ldots, x_n] \prod_{j=0}^{n-1}(x - x_j)}$$

Proof. We prove this by induction. Base case $n = 0$ is trivially satisfied. Assume that this is satisfied for p_k,

$$p_k(x) = f[x_0] + f[x_0, x_1](x - x_0) + \ldots + f[x_0, x_1, \ldots, x_k] \prod_{j=0}^{k-1}(x - x_j)$$

Consider the polynomial $p_{k+1}(x) - p_k(x) \in \mathbb{P}_{k+1}$ which has $x_0, x_1, \ldots, x_k$ as roots. Hence,

$$p_{k+1}(x) - p_k(x) = c \prod_{j=0}^{k}(x - x_j)$$

Comparing leading coefficient on both sides, we have $c = f[x_0, x_1, \ldots, x_k]$. Hence,

$$p_{k+1}(x) = f[x_0] + f[x_0, x_1](x - x_0) + \ldots + f[x_0, x_1, \ldots, x_{k+1}] \prod_{j=0}^{k}(x - x_j)$$

By PMI,

$$p(x) = f[x_0] + f[x_0, x_1](x - x_0) + \ldots + f[x_0, x_1, \ldots, x_n] \prod_{j=0}^{n-1}(x - x_j)$$

□

1.8 Weierstrass theorem consequences

- In Weierstrass approximation Theorem, take $\epsilon_n = \frac{1}{n}$. Then weierstrass theorem proves the existence of sequence of polynomials $p^{(1)}, p^{(2)}, \ldots$ such that

$$\lim_{n\to\infty} ||f - p^{(n)}|| = 0$$

- If **f in not a polynomial**, then

$$\lim_{n\to\infty} \text{degree of } p(n) = 0$$

1.9 Spline Interpolation

- **Piece wise polynomial**: $\phi \in C[a,b]$ is a piecewise polynomial function, if there exists $a = x_0 < x_1 < \ldots < x_n = b$ such that $\phi \in \mathbb{P}_m$ when $x \in [x_i, x_{i+1}]$ for all $i = 0, 1, \ldots, n$ and some $m > 0$.
- Piece wise polynomial ϕ need not be polynomial in whole domain.
- Splines interpolation for $f \in C[a,b]$
  - Pick some data points $x_0, x_1, \ldots, x_n$ such that $a = x_0 < x_1 < \ldots < x_n = b$
  - Fix $m \leq n$
  - Build ϕ in each subinterval $[x_i, x_{i+1}]$ using the following conditions:

$$\phi(x_i) = f_i \qquad \text{for } i = 0, 1, \ldots, n$$

$$\lim_{h\to 0+} \phi(x_i - h) = \lim_{h\to 0+} \phi(x_i + h) \qquad \text{for } i = 1, 2, \ldots, n-1$$

$$\lim_{h\to 0+} \frac{\mathrm{d}\phi(x_i - h)}{\mathrm{d}x} = \lim_{h\to 0+} \frac{\mathrm{d}\phi(x_i + h)}{\mathrm{d}x} \qquad \text{for } i = 1, 2, \ldots, n-1$$

$$\lim_{h\to 0+} \frac{\mathrm{d}^2\phi(x_i - h)}{\mathrm{d}x^2} = \lim_{h\to 0+} \frac{\mathrm{d}^2\phi(x_i + h)}{\mathrm{d}x^2} \qquad \text{for } i = 1, 2, \ldots, n-1$$

$$\vdots$$

$$\lim_{h\to 0+} \frac{\mathrm{d}^{m-1}\phi(x_i - h)}{\mathrm{d}x^{m-1}} = \lim_{h\to 0+} \frac{\mathrm{d}^{m-1}\phi(x_i + h)}{\mathrm{d}x^{m-1}} \qquad \text{for } i = 1, 2, \ldots, n-1$$

  - We have $(n+1) + m(n-1) = n(m+1) - (m-1)$ conditions. We need $m-1$ more conditions.

1.10 Linear Splines

Constructing linear splines:
Since the degree is only 1, we can constructs the splines using the equation of straight lines between the knots (data points).

$$s_0 = \frac{f_1 - f_0}{x_1 - x_0}x + \frac{x_1 f_0 - x_0 f_1}{x_1 - x_0} \qquad \text{for } x \in [x_0, x_1]$$

$$s_1 = \frac{f_2 - f_1}{x_2 - x_1}x + \frac{x_2 f_1 - x_1 f_2}{x_2 - x_1} \qquad \text{for } x \in [x_1, x_2]$$

$$\vdots$$

$$s_{n-1} = \frac{f_n - f_{n-1}}{x_n - x_{n-1}}x + \frac{x_n f_{n-1} - x_{n-1} f_n}{x_n - x_{n-1}} \qquad \text{for } x \in [x_{n-1}, x_n]$$

Theorem (Linear Splines error)**.** *Let $f \in C^2[a,b]$ and $s_L(x)$ be the interpolating **linear spline** at (n+1) knots $a = x_0 < x_1 < \ldots < x_n = b$ and let h be the maximum subinterval length, then*

$$\boxed{||f - s_L|| \leq \frac{h^2}{8}||f''||}$$

Proof. Consider the interval $[x_i, x_{i+1}]$, then $s_L(x)$ is the interpolating polynomial in this interval. Using the error equation for interpolating polynomials,

$$f(x) - s_L(x) = \frac{1}{2} f''(\zeta)(x - x_i)(x - x_{i+1})$$

Taking absolute value on both the sides,

$$\begin{aligned} |f(x) - s_L(x)| &= \frac{1}{2}|f''(\zeta)||(x - x_i)(x - x_{i+1})| \\ &\leq \frac{1}{2}||f''||\frac{h_i^2}{4} \qquad \text{where } h_i = \frac{x_{i+1} - x_i}{2} \\ &\leq \frac{h_i^2}{8}||f''|| \end{aligned}$$

Considering $h = max(h_i)$, then for $x \in [a, b]$

$$max_{x\in[a,b]}|f(x) - s_L(x)| \leq \frac{1}{8}h^2||f''||$$

□

1.11 Cubic Splines

Constructing cubic splines:

$$s_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i$$

Using the conditions given in the section 1.9, we have 4n-2 equations to for 4n coefficients. We choose the other two conditions as

$$s_0''(x_0) = s_{n-1}''(x_n) = 0$$

We have 4n variables (coefficients) and 4n equations, i.e, we have a $4n \times 4n$ matrix which can be solved to get the coefficients of the spline.

We can simplify this by choosing the form of spline as

$$s_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \qquad \text{for } i = 0, 1, \ldots, n-1$$

We will work only with **equally spaced knots**, i.e, $x_{i+1} - x_i = const$ for $i = 0, 1, \ldots, n-1$. We also define

$$\sigma_i = s''(x_i) \qquad \text{for } i = 0, 1, \ldots, n$$

After doing a lot of simplification, we get

$$\boxed{a_i = \frac{\sigma_{i+1} - \sigma_i}{6h}}$$

$$\boxed{b_i = \frac{\sigma_i}{2}}$$

$$\boxed{c_i = \frac{f_{i+1} - f_i}{h} - \frac{h}{6}(2\sigma_i + \sigma_{i+1})}$$

$$\boxed{d_i = f_i}$$

σ_i s can be obtained by solving these equations

$$\boxed{\sigma_{i-1} + 4\sigma_i + \sigma_{i+1} = \frac{6}{h^2}(f_{i-1} - 2f_i + f_{i+1})}$$

This can be put in matrix form as

$$\begin{bmatrix} 4 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 1 & 4 & 1 & \ldots & 0 & 0 & 0 \\ 0 & 1 & 4 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 4 & 1 & 0 \\ 0 & 0 & 0 & \ldots & 1 & 4 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \vdots \\ \sigma_{n-3} \\ \sigma_{n-2} \\ \sigma_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} f_0 - 2f_1 + f_2 \\ f_1 - 2f_2 + f_3 \\ f_2 - 2f_3 + f_4 \\ \vdots \\ f_{n-4} - 2f_{n-3} + f_{n-2} \\ f_{n-3} - 2f_{n-2} + f_{n-1} \\ f_{n-2} - 2f_{n-1} + f_n \end{bmatrix}$$

Theorem (Error - Cubic Splines (equispaced knots))**.** *Let $f \in C^4[a,b]$ and $s(x) \in C^2[a,b]$ be the interpolating* ***natural cubic spline*** *at (n+1)* ***equispaced knots*** *$a = x_0 < x_1 < \ldots < x_n = b$ and let h be the subinterval length ($h = x_{i+1} - x_i$), then*

$$\boxed{||f - s|| \leq C||f^{(4)}||h^4 \qquad \textit{for some } C > 0}$$

Proof. Consider the function g which is defined as $g = f - s_i$ on each subinterval $[x_i, x_{i+1}]$. Then $g(x_i) = 0$ for $i = 0, 1, \ldots, n-1$.
We can see that the zero polynomial is the linear spline of $g(x)$ and using the theorem Linear Splines error, we have

$$\begin{aligned}||g - 0|| &\leq \frac{h^2}{8}||g''|| \\ ||f - s|| &\leq \frac{h^2}{8}||f'' - s''||\end{aligned}$$

Assuming that $||f'' - s''|| \leq Ch^2||f^{(4)}||$, we have

$$||f - s|| \leq Ch^4||f^{(4)}|| \qquad \text{for some } C > 0$$

□

2 Numerical Integration

2.1 Introduction

- Given f be a real valued function, we want to evaluate

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

- If we can find its antiderivative $F(x)$, then we can use the fundamental theorem of calculus and evaluate it. But finding antiderivate is not so straightforward.

$$\int_a^b f(x)dx = F(b) - F(a)$$

2.2 Newton-Cotes Formula

Let $f(x) : [a,b] \to \mathbb{R}$ and $p(x) \in \mathbb{P}_n$ be the polynomial interpolant using the data points $a = x_1 < x_2 < \ldots < x_n = b$, then the definite integral $\int_a^b f(x)dx$ can be approximated as

$$\begin{aligned}\int_a^b f(x)dx &\approx \int_a^b p(x)dx \\ &= \int_a^b \sum_{i=0}^{n} f(x_i)L_i(x)dx \\ &= \sum_{i=0}^{n} f(x_i) \int_a^b L_i(x)dx\end{aligned}$$

Let $x_i = a + ih$ for $i = 0, 1, \ldots, n$ and $x = a + th$ for $t \in [0, n]$, then we have,

$$\int_a^b L_i(x)dx = \int_a^b \prod_{k=0,k\neq i}^{n} \frac{x - x_k}{x_i - x_k}dx = \int_0^n \prod_{k=0,k\neq i}^{n} \frac{t-k}{i-k}hdt = h\int_0^n \varphi_i(t)dt = hw_i$$

$$\text{where } w_i = \int_0^n \varphi_i(t)dt \text{ and } \varphi_i(t) = \prod_{k=0,k\neq i}^{n} \frac{t-k}{i-k}$$

Then,

$$\boxed{\int_a^b f(x)dx \approx h\sum_{i=0}^{n} w_i f(x_i)}$$

Note: w_is are independent of f, end points a and b and h. w_is are dependent on only dependent on n. **Note**: w_is are symmetric i.e,

$$w_k = w_{n-k}$$

Trapezium rule $(n = 1)$:

$$\int_a^b f(x)dx \approx \frac{b-a}{2}(f(a)+f(b))$$

Simpson's rule $(n = 2)$:

$$\int_a^b f(x)dx \approx \frac{h}{3}\left(f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right)$$

2.3 Newton-Cones formula error

We use the Error equation, for interpolation polynomial to calculate the error in Newton-Cotes formula.

$$\begin{aligned}
|I_f - I_P| &= \left|\int_a^b f(x) - p(x)dx\right| \\
&\leq \int_a^b |f(x) - p(x)|dx
\end{aligned}$$

$$\boxed{|I_f - I_P| \leq \frac{1}{(n+1)!}||f^{(n+1)}|| \int_a^b \prod_{i=0}^{n} |x - x_i|dx}$$

For trapezium rule (n=1)

$$\begin{aligned}
|I_f - I_{P_1}| &\leq \frac{1}{2}||f''|| \int_a^b |(x-a)(x-b)|dx \\
&= \frac{1}{12}||f''||(b-a)^3
\end{aligned}$$

For simpson's rule (n=2)

$$\begin{aligned}
|I_f - I_{P_2}| &\leq \frac{1}{6}||f'''|| \int_a^b |(x-a)(x-\frac{a+b}{2})(x-b)|dx \\
&= \frac{1}{192}||f''||(b-a)^4
\end{aligned}$$

2.4 Convergence of the approximation

- The difference $|I_f - I_{P_n}|$ **does not converge** to 0 as we increase n. This is (crudely) because the weights sometimes takes negative values.
- A similar approximation which converges as n increases is Gaussian quadratures.

  $$G_n(f) = \sum_{i=0}^{n} W_i f(x_i)$$

  where weights W_i are

  $$W_i = \int_a^b (L_i(x))^2 dx = \int_a^b \left[\prod_{k=0,k\neq i}^{n} \frac{x - x_k}{x_i - x_k}\right]^2 dx$$

  Here, the points x_ks are **not** equally spaces.
  x_ks are the roots of Legendre Polynomials

2.5 Composites Rule

Similar to the splines case, we approximate the integral using Newton-Cotes formula in each subinterval.

- **Composite Trapezoidal Rule**: We divide the interval $[a, b]$ into m subintervals and apply trapezoidal rule in each subinterval.

$$\begin{aligned}
\int_a^b f(x)dx &\approx \int_a^b C_{p_1} dx \\
&= h(\frac{1}{2}f(a) + \frac{1}{2}f(a+h)) + h(\frac{1}{2}f(a+h) + \frac{1}{2}f(a+2h)) + \ldots + h(\frac{1}{2}f(a+(m-1)h) + \frac{1}{2}f(b)) \\
&= h(\frac{1}{2}f(a) + f(a+h) + f(a+2h) + \ldots + f(a+(m-1)h) + \frac{1}{2}f(b))
\end{aligned}$$

- **Composite Simpson's Rule**: We divide the interval $[a, b]$ into 2m subintervals and apply simpons' rule.

$$\begin{aligned}\int_a^b f(x)dx &\approx \int_a^b C_{p_2} dx \\ &= h(\frac{1}{3}f(a) + \frac{4}{3}f(a+h) + \frac{1}{3}f(a+2h)) + h(\frac{1}{3}f(a+2h) + \frac{4}{3}f(a+3h) + \frac{1}{3}f(a+4h)) \\ &+ \ldots + h(\frac{1}{3}f(a+(2m-2)h) + \frac{4}{3}f(a+(2m-1)h) + \frac{1}{3}f(b)) \\ &= h(\frac{1}{3}f(a) + \frac{4}{3}f(a+h) + \frac{2}{3}f(a+2h) + \frac{4}{3}f(a+3h) + \frac{2}{3}f(a+4h) + \ldots \\ &+ \frac{2}{3}f(a+(2m-2)h) + \frac{4}{3}f(a+(2m-1)h) + \frac{1}{3}f(b))\end{aligned}$$

2.6 Error - Composites Rule

- To get an estimate of the error $|I_f - C_{p_n}|$, we sum the error in each subinterval.
- For composite trapezoid rule,

$$\begin{aligned}|I_f - C_{p_1}| &\leq \sum_{i=0}^{m} \frac{1}{12}||f''||h^3 \\ &= \frac{1}{12}||f''||h^2(b-a)\end{aligned}$$

3 Ordinary Differential Equation

3.1 Uniqueness and Existence Theorem

Theorem (Cauchy-Lipschitz-Picard). *Consider the initial value problem*

$$y' = f(t, y) \qquad y(t_0) = y_0$$

1. *If the function f is continuous for $t_0 \leq t \leq T$ and $y_0 - C \leq y \leq y_0 + C$ for some constants T, $C > 0$.*
2. *If the function f satisfies lipschitz condition.*
   ***Lipschitz condition** : There exists a constant $L > 0$ such that*

   $$|f(t, u) - f(t, v)| \leq L|u - v|$$

   for all $t \in [t_0, T]$ and $u, v \in [y_0 - C, y_0 + C]$

Then there exists a unique solution $y(t) \in C^1[t_0, T]$ such that

$$y' = f(t, y) \qquad y(t_0) = y_0$$

Lipschitz condition imposes the condition that the slope of the function f (with t treated as constant) is bounded.

3.2 Integral Formula

Consider the initial value problem

$$y' = f(t, y) \qquad y(t_0) = y_0$$

Interating with y from y_0 to y and t from t_0 to t

$$y = y_0 + \int_{t_0}^{t} f(s, y(s))ds$$

3.3 Order of Numerical Method

Consider a numerical method given by the recurrence relation

$$y_{i+1} = F(t, f, y_0, y_1, \ldots, y_i, y_{i+1})$$

Order of the numerical method is said to be p if

$$y(t_{i+1}) - F(t, f, y(t_0), y(t_1), \ldots, y(t_i), y(t_{i+1})) = O(h^{p+1})$$

This in some sense represents order of error that is included while approximating from y_i to y_{i+1} since we use exact values while computing y_{i+1} and get the difference between $y(t_{i+1})$ and t_{i+1}. Order can also be intepreted as the highest degree of polynomial which is exactly recovered after approximating the solution using the numerical method. (Here, order is the degree of the polynomial and not function f)

3.4 Global error

In an interval [0,T] for a given $h > 0$, there are $\left[\frac{T}{h}\right] + 1$ are equally spaced mesh points. Let

$$e_{n,h} = y_{n,h} - y(t_n)$$

3.5 Convergence of Numerical Method

A numerical method is said to be convergent if

$$\lim_{h \to 0^+} max_{i=0,1,\ldots,[\frac{T}{h}]} |e_{i,h}| = 0$$

3.6 Euler Method

To approximate the solution, we divide the interval [0,T] into n mesh points, i.e, $t_i = ih$, where $h = \frac{T}{n}$ At the mesh point t_i, we find approximations y_i for exact solution $y(t_i) = y(t_{i-1}) + \int_{t_{i-1}}^{t_i} f(s, y(s))ds$ using

$$y_i = y_{i-1} + hf(t_{i-1}, y_{i-1}) \qquad \text{for i} = 1, 2, \ldots, n$$

Order:

$$\begin{aligned}
&y(t_{i+1}) - F(t, f, y(t_0), y(t_1), \ldots, y(t_n), y(t_{n+1})) \\
&= y(t_{i+1}) - (y(t_i) + hf(t_i, y(t_i))) \\
&= y(t_i) + hy'(t_i) + \frac{h^2}{2} y''(\zeta) - (y(t_i) + hf(t_i, y(t_i))) \qquad \text{using taylor's expansion} \\
&= \frac{y''(\zeta)}{2} h^2 \\
&= O(h^2)
\end{aligned}$$

Hence, the order of Euler's method is 1.

Using the above result and lipschitz condition, we can prove that

$$|e_{i,h}| \leq \frac{Ch}{\lambda}((1 + \lambda h)^i - 1) \qquad \text{for i } = 0, 1, \ldots, [\frac{T}{h}]$$

Since, $|e_{i,h}| = O(h)$, so $\lim_{h\to} |e_{i,h}| = 0$ and hence euler's method is convergent. Since the global error converges to zero at the rate of h^1, the order of convergence of Euler method is said to be 1.

3.7 Trapezoidal Rule

The recurrence relation for trapezoidal rule is implicit.

$$y_i = y_{i-1} + \frac{h}{2}(f(t_{i-1}, y_{i-1}) + f(t_i, y_i))$$

Order:

$$\begin{aligned}
&y(t_{i+1}) - F(t, f, y(t_0), y(t_1), \ldots, y(t_n), y(t_{n+1})) \\
&= y(t_{i+1}) - (y(t_i) + \frac{h}{2}(y'(t_i) + y'(t_{i+1}))) \\
&= y(t_i) + hy'(t_i) + \frac{h^2}{2} y''(t_i) + \frac{h^3}{6} y'''(\zeta_1) - (y(t_i) + \frac{h}{2}(y'(t_i) + y'(t_i) + hy''(t_i) + \frac{h^2}{2} y'''(\zeta_2))) \\
&= (\frac{y'''(\zeta_1)}{6} - \frac{y'''(\zeta_2)}{4})h^3 \\
&= O(h^3)
\end{aligned}$$

Hence, the order of Trapezoidal Rule is 2.

Using the above result and lipschitz condition, we can prove that

$$|e_{i,h}| \leq \frac{Ch^2}{\lambda} \left(\left(\frac{1 + \frac{\lambda h}{2}}{1 - \frac{\lambda h}{2}} \right)^i - 1 \right) \qquad \text{for i } = 0, 1, \ldots, [\frac{T}{h}]$$

Since, $|e_{i,h}| = O(h^2)$, so $\lim_{h\to} |e_{i,h}| = 0$ and hence trapezoidal method is convergent. Since the global error converges to zero at the rate of h^2, the order of convergence of trapezoidal method is said to be 2.

3.8 Multistep Methods

- **Simpson's Rule** We have

$$y(t_{i+1}) - y(t_{i-1}) = \int_{t_{i-1}}^{t_{i+1}} f(t,y)dt$$

We use the simpsons rule to approximate the integral,

$$\boxed{y_{t_{i+1}} = y_{t_{i-1}} + \frac{h}{3}\left(f(t_{i-1}, y_{i-1} + 4f(t_i, y_i) + f(t_{i+1}, y_{i+1})\right)}$$

Note: Here, $h = t_{i+2} - t_i$

- **Adam Bashfort Method:** (Explicit 4 step method)

$$y_{i+4} = y_{i+3} + \frac{h}{24}(55f(t_{i+3}, y_{i+3}) - 59f(t_{i+2}, y_{i+2} + 37f(t_{i+1}, y_{i+1} - 9f(t_i, y_i))))$$

- **Adam Moulton Method:** (Implicit 3 step method)

$$y_{i+3} = y_{i+2} + \frac{h}{24}(9f(t_{i+3}, y_{i+3}) + 19f(t_{i+2}, y_{i+2} - 5f(t_{i+1}, y_{i+1} - 9f(t_i, y_i))))$$

3.9 General multistep method

Any s step method is of the form

$$\sum_{m=0}^{s} a_m y_{i+m} = h \sum_{m=0}^{s} b_m f(t_{i+m}, y_{i+m}) \qquad \text{for i} = 0, 1, 2, \ldots$$

Here, a_m, b_m, s are constants (independent of h,i and f)
For any s step method we need s starting points $(y_0, y_1, y_2, \ldots, y_{s-1})$

3.10 Constructing s step method

Adams method of constructing s step method:

1. Take the s points $t_i, t_{i+1}, \ldots, t_{i+s-1}$ and corresponding values (approximated by using y_i instead of $y(t_i)$) $f(t_i, y_i), f(t_{i+1}, y_{i+1}), \ldots, f(t_{i+s-1}, y_{i+s-1})$ and find the interpolating polynomial
2. Integrate the obtained polynomial between t_{i+s-1} to t_{i+s} for obtaining $y_{i+s} - y_{i+s-1}$, i.e,

$$y_{i+s} - y_{i+s-1} = \int_{t_{i+s-1}}^{t_{i+s}} \sum_{m=0}^{s-1} f(t_{i+m}, y_{i+m}) \prod_{p=0, p\neq m}^{s-1} \left(\frac{t - t_{i+p}}{t_{i+m} - t_{i+p}}\right) dt$$

On simplifying we have

$$y_{i+s} - y_{i+s-1} = h \sum_{m=0}^{s-1} b_m f(t_{i+m}, y_{i+m})$$

where $b_m = \int_{s-1}^{s} \prod_{p=0, p\neq m}^{s-1} \left(\frac{x-p}{m-p}\right) dx$

3.11 Order of s step method

Theorem. *The order of s step method is p if*

1. $$\sum_{m=0}^{s} a_m = 0$$
2. $$\sum_{m=0}^{s} a_m m^k = k \sum_{m=0}^{s} b_m m^{k-1} \qquad \text{for } k = 1, 2, \ldots, p$$
3. $$\sum_{m=0}^{s} a_m m^{p+1} \neq (p+1) \sum_{m=0}^{s} b_m m^p$$

So to compute order, we keep on checking if $\sum_{m=0}^{s} a_m m^k = k \sum_{m=0}^{s} b_m m^{k-1}$ while increasing, if this condition is not satisfied for k (first time), then $k-1$ is the required order.

3.12 Convergence

Consider the general s step method

$$\sum_{m=0}^{s} a_m y_{i+m} = h \sum_{m=0}^{s} b_m f(t_{i+m}, y_{i+m}) \qquad \text{for i} = 0, 1, 2, \ldots$$

The associated characteristics polynomials are

$$\rho(z) = \sum_{m=0}^{s} a_m z^m \qquad\qquad \sigma(z) = \sum_{m=0}^{s} b_m z^m$$

Theorem (Dahlquist equivalence theorem)**.** *A s-step method is convergent iff*

1. *The method is of order* $p \geq 0$
2. *The roots of the characteristic polynomial* $\rho(z)$ *lies in the closed unit disc in the complex plane, with any roots that lie on the unit circle being simple.*