LECTURE 20

SOLUTION TO SINGLE 1ST ORDER INITIAL VALUE PROBLEMS (IVP's)

- Solve

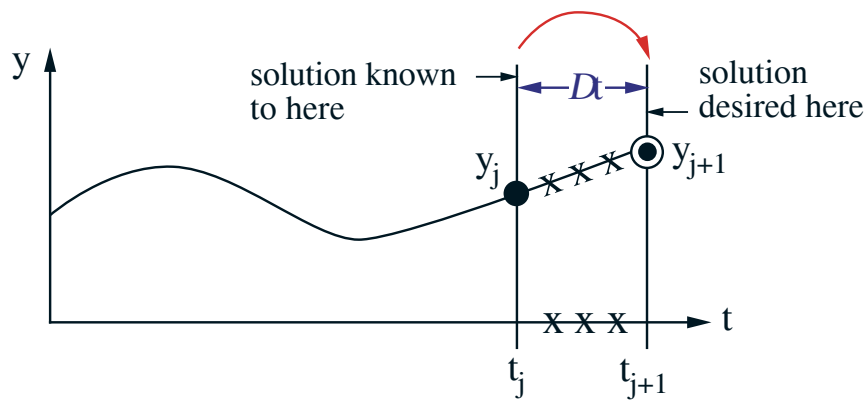
$$\frac{dy}{dt} = f(y, t) \qquad \text{i.c. } y(t_o) = y_o$$

- Consider two classes of methods:
  - Runge-Kutta type formulae
    - single step methods
    - very simple to program
    - self starting (only need i.c.'s)
  - Multi-step formulae
    - Multi-step methods are much more efficient than single step methods (for the same accuracy)
    - Multi-step methods are not self starting → use single step method to start up and then go over to multi-step

Runge-Kutta type formulas → Single Step Methods

- Solution y_{j+1} is obtained in terms of y_j, $f(y_j, t_j)$ and $f(y, t)$ evaluated for various values of y between t_j and t_{j+1} ⇒ ***self starting***.

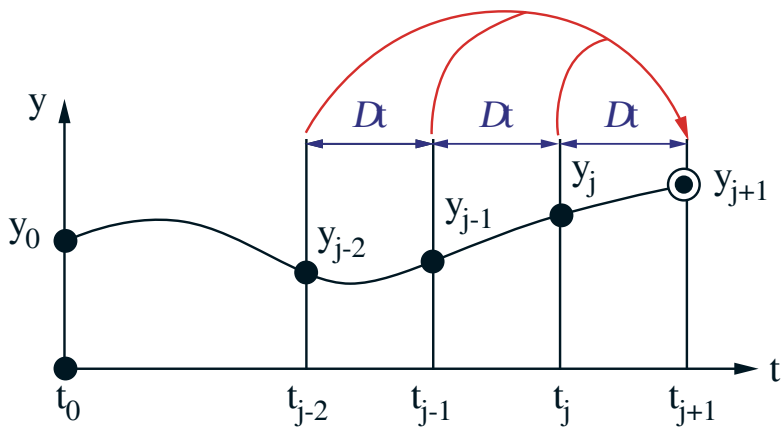
- Self starting since solution involves only information between $t_j < t < t_{j+1}$ ⇒ therefore all information required is available at the 1st step (i.e. the response function of the previous step only).



- Various orders of accuracy are available:
  - 1st order - Euler
  - 2nd order - Improved Euler, Modified Euler
  - 4th order - Runge-Kutta

Multi-step Methods

- Require information for $t < t_j$ in order to predict the value at t_{j+1}



- Multi-step methods are dependent on several previous conditions
- Use F.D.'s in the development of the multi-step formulae
  - Adams open formula
  - Adams closed formula
  - Predictor - Corrector Methods (combination of the above 2 methods)
- Multi-step methods are much more efficient than single step methods (for the same accuracy)
- Multi-step methods are not self starting → use single step method to start up

Runge-Kutta type methods

- Solve $\frac{dy}{dt} = f(t, y) \qquad y(0) = y_o$

- Recursive relationship for all Runge-Kutta methods:

$$y_{j+1} = y_j + \Delta t \Phi(y_j, t_j, \Delta t)$$

$$\Phi \equiv a_1 g_1 + a_2 g_2 + a_3 g_3 + \ldots + a_n g_n$$

where

$$g_1 \cong f(t, y) \qquad (g_1 \text{ is therefore the slope as per the definition of } f)$$

$$g_2 \cong f(t + p_1 \Delta t, y + p_2 \Delta t g_1)$$

$$g_3 \cong f(t + p_3 \Delta t, y + p_4 \Delta t g_2)$$

$$\vdots$$

- We must select a_i's and p_i's
  - Select these coefficients such that you minimize errors.
  - Compare Taylor Series expansion of y_{j+1} and select the recursive relationship coefficients such that you eliminate the appropriate error terms.

Euler Method

- 1st order method

Derivation 1

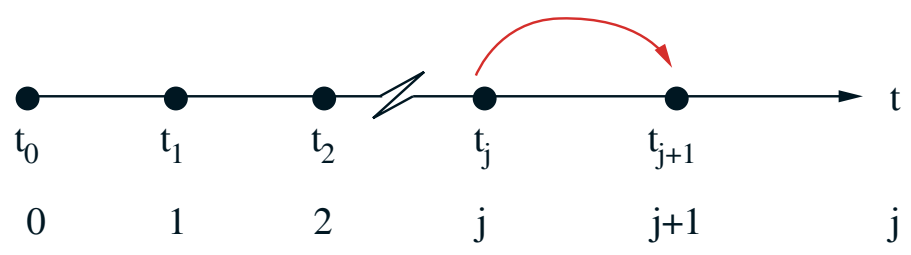
- Use forward difference approximation for $\frac{dy}{dt}$

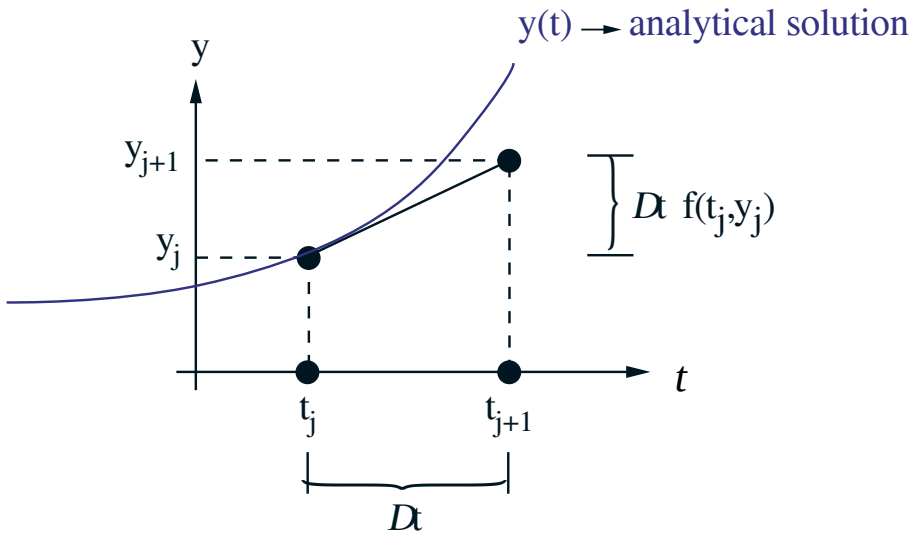
$$\frac{dy}{dt} = f(y, t) \quad \Rightarrow$$

$$\frac{y_{j+1} - y_j}{\Delta t} = f(y_j, t_j) \quad \Rightarrow$$

$$y_{j+1} = y_j + \Delta t\, f(y_j, t_j)$$

- Simply "march" forward in time from $t = 0$





- $f(t_j, y_j)$ slope at t_j
- Simply add $\Delta t\ f(t_j, y_j)$ to y_j

Derivation 2

- Cast into generic Runge-Kutta form with Φ expanded to only 1 term

$$y_{j+1} = y_j + \Delta t \; a_1 \; g_1 - E_L$$

where $E_L \equiv$ the local truncation error per time step

$$y_{j+1} = y_j + \Delta t a_1 \; f(t_j, y_j) - E_L \qquad (1)$$

- Develop Taylor Series expansion for y_{j+1} about t_j

$$y_{j+1} = y_j + \Delta t \left.\frac{dy}{dt}\right|_j + \frac{(\Delta t)^2}{2} \left.\frac{d^2y}{dt^2}\right|_j + O(\Delta t)^3$$

- Note that

$$\left.\frac{dy}{dt}\right|_j = f(t_j, y_j)$$

$$\left.\frac{d^2y}{dt^2}\right|_j = \dot{f}(t_j, y_j)$$

- Therefore

$$y_{j+1} = y_j + \Delta t\, f(t_j, y_j) + \frac{(\Delta t)^2}{2}\dot{f}(t_j, y_j) + O(\Delta t)^3 \qquad (2)$$

- Now compare Equations (1) and (2)

$$y_j + \Delta t\, a_1\, f(t_j, y_j) - E_L = \quad y_j + \Delta t\, f(t_j, y_j) + \frac{(\Delta t)^2}{2}\dot{f}(t_j, y_j) + O(\Delta t)^3$$

- Comparing terms we note that

$$a_1 = 1$$

$$E_L = -\frac{(\Delta t)^2}{2}\dot{f}(t_j, y_j)$$

- Thus the Euler Method (substituting for a_1 into the formula)

$$\boldsymbol{y_{j+1} = y_j + \Delta t\, f(t_j, y_j)} \qquad \textbf{(3)}$$

- We also note that the ***local*** truncation error (i.e. per time step), E_L, is second order! However this error builds up as we time step and will in fact become first order.

Detailed analysis of truncation error for the Euler Method

- Neglect roundoff error for this analysis
- Let the exact solution be denoted Y
  - Assume that at some starting time t_j we know the exact solution Y_j
- First time step taken:

$$y_{j+1} = Y_j + \Delta t\, f(t_j, Y_j) \tag{4}$$

- Now let's apply Taylor Series to find Y_{j+1}

$$Y_{j+1} = Y_j + \Delta t \left.\frac{dY}{dt}\right|_{t_j} + \frac{(\Delta t)^2}{2}\left.\frac{d^2Y}{dt^2}\right|_{t_j} + \frac{(\Delta t)^3}{3!}\left.\frac{d^3Y}{dt^3}\right|_{t_j} + \ldots$$

- However the o.d.e. states that

$$\frac{dY}{dt} = f(t, Y) \quad \Rightarrow \quad \left.\frac{dY}{dt}\right|_{t_j} = f(t_j, Y_j) \quad \Rightarrow \quad \left.\frac{d^2Y}{dt^2}\right|_{t_j} = \dot{f}(t_j, Y_j) \quad \text{etc.}$$

- Thus

$$Y_{j+1} = Y_j + \Delta t\, f(t_j, Y_j) + \frac{(\Delta t)^2}{2}\dot{f}(t_j, Y_j) + \frac{(\Delta t)^3}{3!}\ddot{f}(t_j, Y_j) + \ldots \tag{5}$$

- Taking the difference between Equations (4) and (5) defines the local truncation error

$$E_{j+1} \equiv y_{j+1} - Y_{j+1} = -\frac{(\Delta t)^2}{2}\dot{f}(t_j, Y_j) + O(\Delta t)^3 \tag{6}$$

  - E_{j+1} = the truncation error of the Euler formula ***per time step*** = $O(\Delta t)^2$
  - This is consistent with the $O(\Delta t)$ error in the forward difference approximation used to evaluate $\frac{dy}{dt}$.
  - This equation assumes that Y_j is exact

- The total solution error at t_{j+1} is due to the truncation error at every time step (local error) ***plus*** the error that has accumulated in all previous steps.
  - Only for the first step when using the i.c. $y_j = Y_j$
  - For other steps the solution y_j carries an accumulated error!

- In general the total solution has an error which can be defined by examining the difference between the Euler formula, Equation (3) and our Taylor Series expansion, Equation (5).

$$y_{j+1} - Y_{j+1} = y_j - Y_j + \Delta t(f(t_j, y_j) - f(t_j, Y_j)) - \frac{(\Delta t)^2}{2}\dot{f}(t_j, Y_j) + O(\Delta t)^3 \tag{7}$$

- The total solution error:

$$\varepsilon_{j+1} \equiv y_{j+1} - Y_{j+1} \tag{8}$$

$$\varepsilon_j \equiv y_j - Y_j \tag{9}$$

- Also it can be shown that (by one of the mean value theorems):

$$\frac{f(t_j, y_j) - f(t_j, Y_j)}{y_j - Y_j} = \frac{\partial f}{\partial y}(t_j, \xi_j) \qquad y_j < \xi_j < Y_j \tag{10}$$

- Substituting Equations (6), (8), (9), (10) into Equation (7):

$$\varepsilon_{j+1} = \varepsilon_j + \Delta t\ \varepsilon_j \frac{\partial f}{\partial y}(t_j, \xi_j) + E_{j+1}$$

  - where
    - ε_{j+1} = the total new error
    - ε_j = the total old error
    - $\Delta t\ \varepsilon_j \frac{\partial f}{\partial y}(t_j, \xi_j)$ = the added error due to not evaluating the slope at quite the right point (we evaluate the slope at (t_j, y_j) instead of at (t_j, Y_j)
    - E_{j+1} = local truncation error due to applying a constant slope over the interval $t_j \rightarrow t_{j+1}$

- Letting $p_j \equiv \frac{\partial f}{\partial y}(\xi_j, t_j)$

$$\varepsilon_{j+1} = \varepsilon_j + \Delta t\ \varepsilon_j\ p_j + E_{j+1} \quad \Rightarrow$$

$$\boldsymbol{\varepsilon_{j+1} = \varepsilon_j(1 + \Delta t\ p_j) + E_{j+1}} \tag{11}$$

- Let's apply Equation (11) to a simplified scenario
  - For real problems, p_j and E_{j+1} change every time step
  - ***Estimate ε_{j+1} by assuming that p and E are constants over the interval of interest***

$$\varepsilon_{j+1} = \varepsilon_j(1 + \Delta t\ p) + E$$

  - Starting with the i.c.'s at $j = 0$

$$\varepsilon_o = 0$$

$$\varepsilon_1 = E$$

$$\varepsilon_2 = E(1 + \Delta t\ p) + E \quad \Rightarrow$$

$$\varepsilon_2 = E(2 + \Delta t\ p)$$

$$\varepsilon_3 = E(2 + \Delta t\ p)(1 + \Delta t\ p) + E \quad \Rightarrow$$

$$\varepsilon_3 = E(3 + 3\Delta t\ p + (\Delta t)^2 p^2)$$

$$\varepsilon_4 = E(3 + 3\Delta t\ p + (\Delta t)^2 p^2)(1 + \Delta t p) + E \quad \Rightarrow$$

$$\varepsilon_4 = 4E + 6p\ \Delta t\ E + 4p^2(\Delta t)^2 E + (\Delta t)^3 p^3 E$$

- However for Euler $E = O(\Delta t)^2$

$$\varepsilon_4 = 4 \times O(\Delta t)^2 + 6pO(\Delta t)^3 + 4p^2 O(\Delta t)^4 + p^3 O(\Delta t)^5$$

  - The leading $4 \times O(\Delta t)^2$ term > all other terms

- Thus

$$\varepsilon_4 = 4E + O(\Delta t)^3$$

$$\vdots$$

$$\varepsilon_{j+1} = (j+1)E + O(\Delta t)^3$$

- We note that

$$j + 1 = \frac{t_{j+1}}{\Delta t}$$

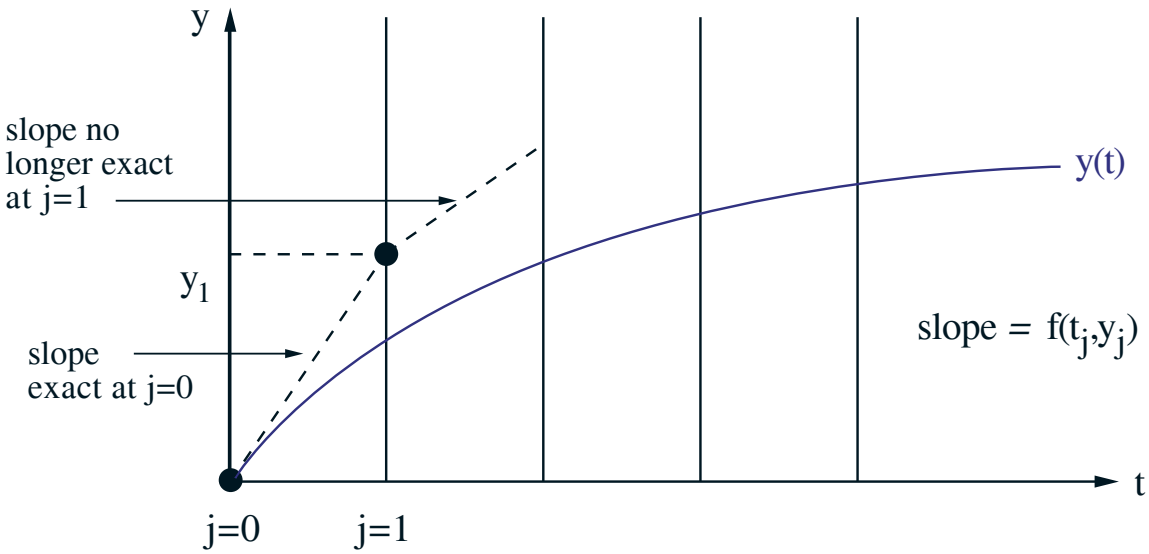
$$E = O(\Delta t)^2$$

- Substituting

$$\varepsilon_{j+1} = \frac{t_{j+1}}{\Delta t} O(\Delta t)^2$$

$$\varepsilon_{j+1} = O(\Delta t)$$

- The Euler method is first order!
- Total numerical error due to truncation is $O(\Delta t)$, one order less than the truncation error per time step.
- ***For higher order methods, a similar relationship holds: The total solution error is one order less than truncation error per time step.***


y
slope no
longer exact
at j=1
y(t)
y_1
slope = $f(t_j,y_j)$
slope
exact at j=0
t
j=0
j=1

- <u>Convergence</u>: A numerical method is convergent if (assuming no round off) the numerical solution approaches the exact solution as $\Delta t \downarrow 0$.

- <u>Stability</u>: Deals with the artificial amplification of components of the numerical solution.
  - Under certain circumstances, components of the discrete solution (often the short wavelength components) experience artificial (i.e. not physical) sustained growth from time step to time step which ultimately leads to numerical overflow (i.e. the computer can not hold the numbers anymore) $\rightarrow$ unstable solution.
  - Stability is a property of both the differential equation and the numerical method $\rightarrow$ i.e. the difference equations determine stable/unstable behavior.
  - Discrete solution can be
    - unconditionally unstable
    - conditionally stable $\rightarrow$ restrictions on time step Δt
    - unconditionally stable

- An unstable scheme is ***always*** inaccurate. A stable scheme may be inaccurate (depends on the truncation error).