- Numerical methods — replace DE by <u>difference equations</u>

<u>Simplest example</u> - "Forward Euler"

Approximate DE $\mathring{x} = f(t, x)$ by difference equation $D^+ y = f(t, y)$ or

$$\frac{y(t+h) - y(t)}{h} = f(t, y(t)).$$

Given $y_0 \equiv y(0)$, this formula recursively generates $y_1 \equiv y(h)$, $y_2 \equiv y(2h)$, ... $y_n \equiv y(nh)$ is supposed to approximate solution $x(t)$ of DE at $t = nh$.

<u>Errors</u>

(i) <u>Local truncation error</u> - Substitute an <u>exact solution</u> of DE into $D^+ x - f(t, x)$:

$$D^+ x - f(t, x) =$$

$$\dot{x}(t) + \frac{h}{2}\ddot{x}(\xi) - f(t, x(t)) = \underbrace{\frac{h}{2}\ddot{x}(\xi)}$$

$t < \xi < t+h$

Local truncation error - ie - error in one step due to replacing $\frac{d}{dt}$ by D^+.

(ii) <u>Global truncation error</u> - due to accumulation of local truncation errors after many steps.

Consider DE $\dot{x} = f(t, x)$ with $f(t, x)$ continuously differentiable in t, x with $|f_x| < L$ for all x. There is a unique solution $x(t)$ of the DE in an interval I about $t=0$ which satisfies IC $x(0) = x_0$.

Take $T > 0$ in interval I, and define $h \equiv \frac{T}{N}$,

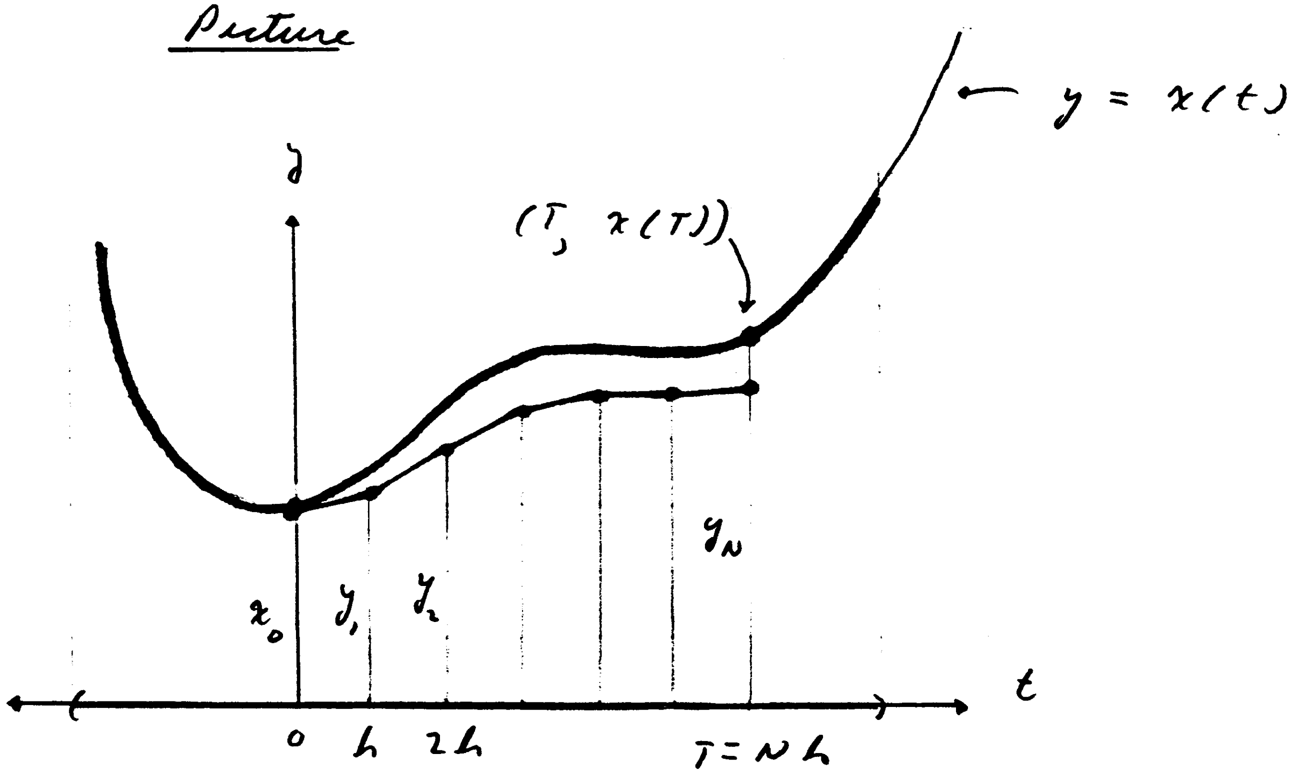
N = integer. Let $\{y_n\}$ be solution of difference equation

$$y_{n+1} - y_n = h\, f(nh, y_n), \quad n = 0, 1, 2, \ldots$$

$$y_0 = x_0.$$

We seek an upper bound on $|y_N - x(T)|$.

<u>Picture</u>



(i) Let $x_n \equiv x(nh)$ and $\varepsilon_n \equiv y_n - x_n$ = difference between numerical and exact solution at $t = nh$.

(i) We calculate

$$\varepsilon_{n+1} - \varepsilon_n = y_{n+1} - x_{n+1} - (y_n - x_n) =$$

$$y_{n+1} - y_n - (x_{n+1} - x_n) =$$

$$h f(nh, y_n) - (x_{n+1} - x_n).$$

(ii) Recall $x(t+h) - x(t) =$

$$h f(t, x(t)) + \frac{h^2}{2} \ddot{x}(\xi),$$

where $t < \xi < t+h$. Set $t = nh \Rightarrow$

$$x_{n+1} - x_n = h f(nh, x_n) + \frac{h^2}{2} \ddot{x}(\xi_n),$$

where $nh < \xi_n < (n+1)h$. Substitute this result into (i) $\Rightarrow$

$$\varepsilon_{n+1} - \varepsilon_n =$$

$$h\{f(nh, y_n) - f(nh, x_n)\} - \frac{h^2}{2} \ddot{x}(\xi_n).$$

(iii) $|f_x| < L \Rightarrow |f(t, y) - f(t, x)| < L|x - y|$ so

$$f(nh, y_n) - f(nh, x_n) = l_n (x_n - y_n) = \varepsilon_n l_n,$$

where $|l_n| < L$. Substitute this result into (ii) $\Rightarrow$

$$\varepsilon_{n+1} - \varepsilon_n = h\, l_n\, \varepsilon_n - \frac{h^2}{2}\ddot{x}(\xi_n) \quad \text{or}$$

$$\varepsilon_{n+1} = (1 + h\, l_n)\, \varepsilon_n - \frac{h^2}{2}\ddot{x}(\xi_n).$$

(iv) $|\varepsilon_{n+1}| \leqq (1 + hL)|\varepsilon_n| + h^2 M,$

where M is an upper bound on $\frac{1}{2}\ddot{x}(t)$ on $[0, T]$. Multiply inequality by $(1 + hL)^{-(n+1)}$ $\Rightarrow$

$$(1+hL)^{-(n+1)}|\varepsilon_{n+1}| - (1+hL)^{-n}|\varepsilon_n| \leqq h^2 M (1+hL)^{-(n+1)}.$$

Define $a_n \equiv (1+hL)^{-n}|\varepsilon_n|$. Get

$$a_{n+1} - a_n = h^2 M (1 + hL)^{-(n+1)}.$$

Since $a_0 = 0$, we have

$$a_N = (a_N - a_{N-1}) + (a_{N-1} - a_{N-2}) + \cdots (a_1 - a_0) =$$

$$\leqq h^2 M \sum_{1}^{N} (1+hL)^{-n} \leqq h^2 M \sum_{0}^{\infty} (1+hL)^{-n} =$$

$$h^2 M \frac{1-\left(\frac{1}{1+hL}\right)^{N+1}}{1-\frac{1}{1+hL}} \leq \frac{M}{L}(1+hL)\,h.$$

In summary,

$$(1+hL)^{-N}|\varepsilon_N| \leq \frac{M}{L}(1+hL)\,h \quad \text{or}$$

$$|y_N - x(T)| \leq \frac{M}{L}(1+hL)^{N+1}\,h.$$

Recall $h = \frac{T}{N}$, so

$$|y_N - x(T)| \leq \frac{M}{L}\underbrace{\left(1+\frac{TL}{N}\right)^{N+1}}\frac{T}{N}.$$

increasing with N

converges to e^{TL}

$$|y_N - x(T)| \leq \left(\frac{MT}{L}e^{TL}\right)\frac{1}{N}.$$

- RHS converges to <u>zero</u> as $N \to \infty$, so numerical solution $\{y_n\}$ converges to exact solution of IVP as stepsize $\to 0$.

Higher order methods

Given $x(t)$ how do we generate approximations to $x(t+h)$ with small local truncation error

- 2d order methods

$$\mathring{x} = f(t, x) \Rightarrow$$

$$\int_t^{t+h} \mathring{x}(\sigma)\, d\sigma = \int_t^{t+h} f(\sigma, x(\sigma))\, d\sigma$$

$$LHS = x(t+h) - x(t)$$

Approximate RHS by trapezoid rule

$$\int_t^{t+h} f(\sigma, x(\sigma))\, d\sigma =$$

$$\frac{h}{2}\left\{ f(t, x(t)) + f(t+h, x(t+h)) \right\} + O(h^3).$$

now $x(t+h)$ is the unknown. We substitute

$$x(t+h) = x(t) + h\,\mathring{x}(t) + O(h^2) =$$

$$x(t) + h\, f(t, x(t)) + O(h^2)$$

into this RHS:

$$f(t+h,\ x(t+h)) =$$

$$f(t+h,\ x(t) + h f(t,\ x(t))) + O(h^2) =$$

$$f(t+h,\ x(t) + h F(t)) + O(h^2),$$

$F(t) \equiv f(t,\ x(t))$. Hence

$$\int_t^{t+h} f(\sigma,\ x(\sigma))\, d\sigma =$$

$$\frac{h}{2}\left\{ F(t) + f(t+h,\ x(t) + h F(t)) \right\} + O(h^3).$$

And the difference scheme is

$$x(t+h) =$$

$$x(t) + \frac{h}{2}\left\{ F(t) + f(t+h,\ x(t) + h F(t)) \right\} + O(h^3).$$

- This scheme is a special case of <u>Rung-Kutta</u> <u>methods</u>, which read

$$x(t+h) - x(t) =$$

$$w_1 h F(t) + w_2 h f(t+\alpha h,\ x(t) + \beta h F(t)) + O(h^3)$$

We <u>choose</u> w_1, w_2, α, β so truncation error <u>is</u> $O(h^3)$. We compute

$$f(t + \alpha h,\ x + \beta h F(t)) =$$

$$F(t) + h\{\alpha h_t + \beta F h_x\} + O(h^2) \quad \text{so}$$

$$x(t+h) - x(t) =$$

$$h(w_1 + w_2)F(t) +$$

$$h^2(w_2 \alpha f_t + w_2 \beta F f_x) + O(h^3).$$

Compare RHS with Taylor series of LHS,

$$x(t+h) - x(t) = \dot{x}(t) h + \frac{\ddot{x}(t)}{2} h^2 + O(h^3)$$

We get

$$(w_1 + w_2) F(t) = \dot{x}(t)$$

$$w_2 \alpha f_t + w_2 \beta F f_x = \tfrac{1}{2} \ddot{x}(t)$$

Now $\dot{x}(t) = f(t, x(t)) = F(t)$ and

$\ddot{x}(t) = f_t + f_x \dot{x} = f_t + F f_x$ so

these equations become

$$\left. \begin{aligned} &(w_1 + w_2) F = F, \\ &w_2 \alpha f_t + w_2 \beta F f_x = \tfrac{1}{2}(f_t + F f_x) \end{aligned} \right\} \Rightarrow$$

$$w_1 + w_2 = 1, \qquad w_2 \alpha = \frac{1}{2}, \qquad w_2 \beta = \frac{1}{2}.$$

This is an <u>underdetermined</u> system of 3 equations in 4 unknowns. One solution: $w_1 = \frac{1}{2}$, $w_2 = \frac{1}{2}$, $\alpha = \beta = 1 \Leftrightarrow$ the scheme we already examined. Another solution: $w_1 = 0$, $w_2 = 1$, $\alpha = \beta = \frac{1}{2}$. The scheme is

$$x(t+h) - x(t) = h f\left(t + \frac{h}{2}, x + \frac{h}{2} F\right)$$

(Modified Euler method)

- 4th order Runge-Kutta method - numerical solution satisfies difference equation

$$y(t+h) - y(t) = \frac{h}{6}\{F_1 + 2F_2 + 2F_3 + F_4\},$$

$$F_1 \equiv f(t, y(t)),$$

$$F_2 \equiv f(t + \frac{h}{2}, y + \frac{h}{2} F_1),$$

$$F_3 \equiv f(t + \frac{h}{2}, y + \frac{h}{2} F_2),$$

$$F_4 = f(t+h, y + h F_3).$$

Local truncation error - let $x(t)$ be an exact solution of DE $\dot{x} = f(t, x)$. Evaluate $F_1 \to F_4$ with $y(t)$ replaced by $x(t)$. An unpleasant calculation shows

$$x(t+h) - x(t) - \frac{h}{6}\{F_1 + 2F_2 + 2F_3 + F_4\} = O(h^5).$$

Generalization to <u>Autonomous system</u>

$\dot{\underline{x}} = \underline{f}(\underline{x})$ (RHS has no explicit t dependence)

$$\underline{y}(t+h) - \underline{y}(t) = \frac{h}{6}\left\{ \underline{F}_1 + 2\underline{F}_2 + 2\underline{F}_3 + \underline{F}_4 \right\},$$

$$\underline{F}_1 \equiv \underline{f}(\underline{y}),$$

$$\underline{F}_2 \equiv \underline{f}(\underline{x} + \frac{h}{2}\underline{F}_1),$$

$$\underline{F}_3 \equiv \underline{f}(\underline{x} + \frac{h}{2}\underline{F}_2),$$

$$\underline{F}_4 \equiv \underline{f}(\underline{x} + h\underline{F}_3).$$