Numerical Solution of ODEs

$v''(t) + v(t) = 0$ 2nd order differential equation

general solution: $v(t) = A\cos t + B\sin t$

where A, B are constants of integration.
This two-parameter family of functions are called integral curves. In this case, the curves oscillate with period 2π. The constants select the amplitude and phase of the oscillation.

In order to obtain a unique solution, need to specify initial or boundary data.

Suppose $v(0) = \alpha$, $v'(0) = \beta$ (initial data)

$\Rightarrow A = \alpha$, $B = \beta$

unique solution: $v(t) = \alpha\cos(t) + \beta\sin(t)$

Or suppose instead $v(0) = \alpha$, $v(b) = \beta$ (boundary data)

solution is: $v(t) = \alpha\cos t + \left(\frac{\beta - \alpha\cos b}{\sin b}\right)\sin t$

unique solution if $\sin b \neq 0$, otherwise there are NO solutions or an INFINITE number of solutions.

IVP is a problem consisting of an ODE and initial data. "Solution evolves locally"

BVP is a problem consisting of an ODE and boundary data (usually specified at endpoints of the ~~domain~~ interval of interest.)

BVP are typically more difficult to handle either analytically or numerically, but the solution set is often more interesting.

Initial Value Problems

Focus on IVP's in the form of a system of 1st order eqns.

$$\underline{y}'(t) = \underline{f}(t, \underline{y}) \quad , \quad 0 \le t \le b \quad , \quad \underline{y}(0) = \underline{c}$$

(constant vector)

where $\underline{y}'(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix}$, $\underline{f}(t, \underline{y}) = \begin{bmatrix} f_1(t, \underline{y}) \\ f_2(t, \underline{y}) \\ \vdots \\ f_m(t, \underline{y}) \end{bmatrix}$, $\underline{c} \in \mathbb{R}^m$

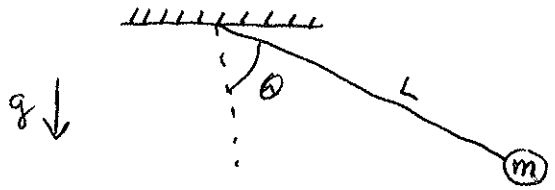
Assume that $\underline{f}(t, \underline{y})$, the "slope function" is a smooth function of $t, \underline{y}$. Note that if the interval is $[a, b]$, then the change of variables $\tau = a + \frac{b-a}{b} t$

transforms to $[0, b]$.

Examples

(1) Pendulum (dynamical system, governed by autonomous ODEs)

Consider a mass suspended on a rigid arm attached to a pivot:



$\theta(t)$ = angular displacement (radians)
m = mass
L = length
g = acceleration due to gravity

Newton's Second Law: (vertical direction

$$mL\frac{d^2\theta}{dt^2} = -mg\sin\theta \quad \Rightarrow \quad \frac{d^2\theta}{dt^2} = -\frac{g}{L}\sin\theta$$

Initial Conditions $\quad \theta(0) = \theta_0, \quad \frac{d\theta}{dt}(0) = \omega_0$

First-order system

$$y_1 = \theta(t)$$
$$y_2 = \frac{d}{dt}\theta(t)$$

$$\Rightarrow \quad \boxed{y_1' = y_2, \quad y_2' = -\frac{g}{L}\sin(y_1)}$$

$$\boxed{y_1(0) = \theta_0, \quad y_2(0) = \omega_0}$$

We expect solutions to oscillate.

IF $|\theta|$ is small then $\sin\theta \sim \theta \Rightarrow \theta'' + \frac{g}{L}\theta = 0$

The general solution is $\theta(t) = A\cos\sqrt{\frac{g}{L}}t + B\sin\sqrt{\frac{g}{L}}t$

$A \uparrow \theta_0$ $\qquad$ $B \uparrow \omega_0\sqrt{\frac{L}{g}}$

What about finite θ?

Consider an integral of the eqn: $\theta'' + \frac{g}{L}\sin\theta = 0$

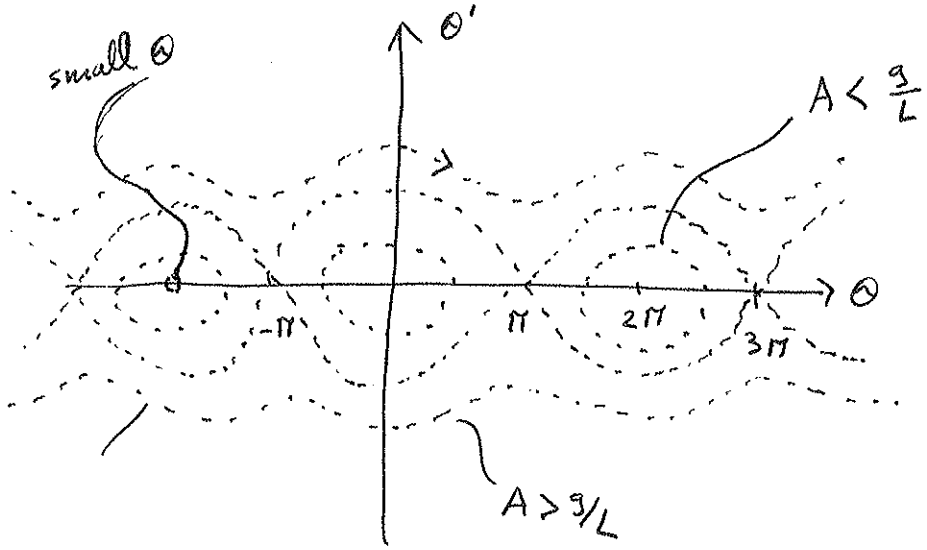
Multiply through by θ':

$$\theta''\theta' + \frac{g}{L}\sin\theta\cdot\theta' = 0$$

$$\Rightarrow \frac{1}{2}\theta'^2 - \frac{g}{L}\cos\theta = A \quad , \quad A = \text{constant}$$

the initial conditions specify: $A = \frac{1}{2}\omega_0^2 - \frac{g}{L}\cos\theta_0$

Plot the curves in the phase plane:



(2) Predator - Prey (dynamical systems

Dynamical system that arises in population dynamics.
Consider the interaction of coupled populations:

$y_1(t)$ = population of a prey species
$y_2(t)$ = population of a predator species

In isolation, assume

$$\left.\begin{aligned} \frac{dy_1}{dt} &= \alpha y_1 \\ \frac{dy_2}{dt} &= \gamma y_2 \end{aligned}\right\}$$ change in population is proportional to the population

Generally assume that $\alpha > 0$, $\gamma < 0$. Assume that the interaction is proportional to the product of the populations, and that the interaction inhibits the growth of prey and promotes the growth of predator.

With interaction:

$$\frac{dy_1}{dt} = \alpha y_1 - \beta y_1 y_2$$

$$\frac{dy_2}{dt} = \gamma y_2 + \delta y_1 y_2$$

where $\beta, \delta > 0$. Include initial conditions

$y_1(0) = A$, $y_2(0) = B$.

Solutions oscillate. Consider phase plane,

equilibrium points: $(0,0)$ or $\left(-\frac{\gamma}{\delta}, \frac{\alpha}{\beta}\right)$

Consider a linearization about $\left(-\frac{\gamma}{\delta}, \frac{\alpha}{\beta}\right)$:

$$y_1 = -\frac{\gamma}{\delta} + \tilde{y}_1 \quad , \quad y_2 = \frac{\alpha}{\beta} + \tilde{y}_2$$

Find that:

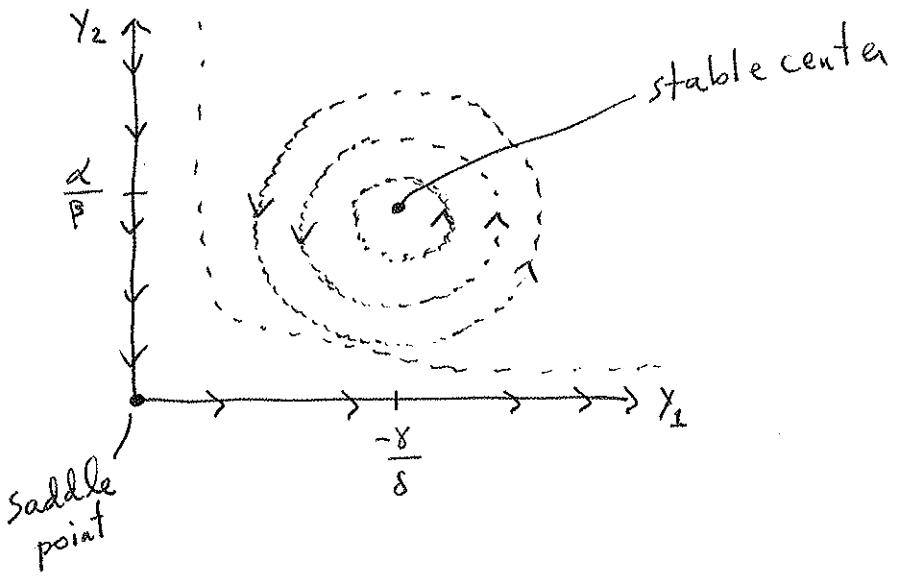
$$\tilde{y}_j'' + \sigma_j \tilde{y}_j = 0 \quad , \quad \sigma_j > 0 \text{ is constant}$$

Integrate:

$$\frac{dy_2}{dy_1} = \frac{\delta y_2 + \delta y_1 y_2}{\alpha y_1 - \beta y_1 y_2} = \frac{y_2(\gamma + \delta y_1)}{y_1(\alpha - \beta y_2)}$$

The equation is separable (Boyce Di Prima)

$$\Rightarrow \underbrace{\alpha \ln y_2 - \beta y_2 - \gamma \ln y_1 - \delta y_1}_{\text{const. of motion}} = \text{constant}$$



(3) Biochemical kinetics

4/4

2-step biochemical reaction involving a sub_strate $s(t)$, an enzyme $e(t)$, a "complex" $c(t)$, and a product $p(t)$.

$$s + e \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} c \xrightarrow{k_2} p + e$$

k_1, k_{-1}, k_2 are rate constants.

Rate laws (laws of mass action):

$*$ - dimensional quantities

(1) $$\frac{ds^*}{dt^*} = -k_1 s^* e^* + k_{-1} c^*$$

(2) $$\frac{de^*}{dt^*} = -k_1 s^* e^* + k_{-1} c^* + k_2 c^*$$

(3) $$\frac{dc^*}{dt} = k_1 s^* e^* - k_{-1} c^* - k_2 c^*$$

(4) $$\frac{dp^*}{dt} = k_2 c^*$$

Initial conditions: $s^*(0) = \bar{s}$, $e^*(0) = \bar{e}$, $c^*(0) = p^*(0) = 0$

1/17/03

notice, add eqns (2) and (3)

$$\Rightarrow \frac{d}{dt^*}\left(e^* + c^*\right) = 0 \quad \rightarrow \quad e^* + c^* = \bar{e}$$

Add (1), (3) and (4), again right-hand side vanishes

$$\frac{d}{dt^*}\left(s^* + c^* + p^*\right) = 0 \quad \rightarrow \quad s^* + c^* + p^* = \bar{s}$$

In effect we've integrated the system twice, so we are left with two ODEs and two algebraic relations.

Use these 2 algebraic relations to eliminate e^* and p^*.

$$\frac{ds^*}{dt^*} = -k_1 \bar{e} s^* + (k_1 s^* + k_{-1}) c^* \quad , \quad s^*(0) = \bar{s}$$

$$\frac{de^*}{dt^*} = k_1 \bar{e} s^* - (k_1 s^* + k_{-1} + k_2) c^* \quad , \quad c^*(0) = 0$$

Choose scales for t^*, s^*, c^* (make dimensionless)

let $s = \frac{s^*}{\bar{s}}$, $c = \frac{c^*}{\bar{e}}$, $t = \frac{t^*}{1/k_1 \bar{e}}$

$$\Rightarrow \boxed{\begin{aligned} s' &= -s + (s + K - \lambda) c \quad , \quad s(0) = 1 \\ \varepsilon c' &= s - (s + K) c \quad , \quad c(0) = 0 \end{aligned}}$$

where $\varepsilon = \frac{\bar{e}}{\bar{s}}$, $K = \frac{k_{-1} + k_2}{k_1 \bar{s}}$, $\lambda = \frac{k_2}{k_1 \bar{s}}$

Often, $\bar{e} \ll \bar{s} \Rightarrow 0 < \varepsilon \ll 1$

$K, \lambda = O(1)$

Analyze the behaviour of the solution.

set $\varepsilon = 0$:

$$s' = -s + (s + K - \lambda) c$$
$$0 = s - (s + K) c \quad \Rightarrow \quad c = \frac{s}{s + K}$$

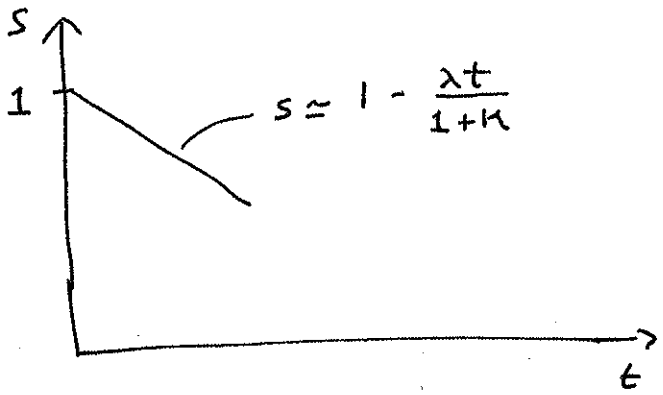
$$\Rightarrow \quad s' = \frac{-\lambda s}{s + K} \quad , \quad s(0) = 1$$

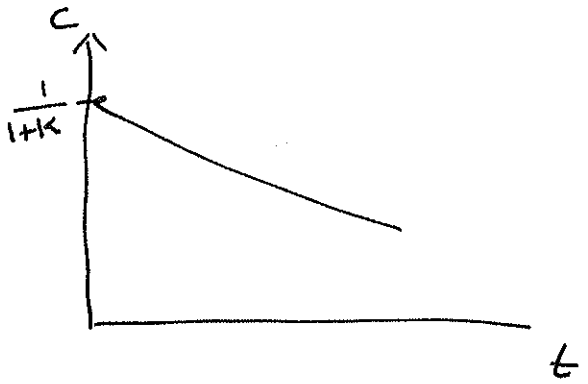
separable equation, integrate

$$\Rightarrow \boxed{s + K \ln s = -\lambda t + 1 \quad , \text{ for } \varepsilon \to 0}$$

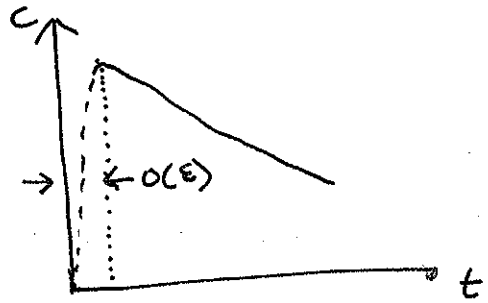
$s + K \ln s = -\lambda t + 1$

notice, when near $s = 1$, linear





Problem with this solution is that c does not satisfy initial condition, $c(0) = 0$.



Consider an "initial layer" solution. This is a small region near $t=0$ where c' is large so that

$\varepsilon c' = O(1)$

Set $\tau = \frac{t}{\varepsilon}$ $\Rightarrow$ when $t = O(\varepsilon)$, $\tau = O(1)$
(exact change of variables)

$$\Rightarrow \quad \frac{ds}{d\tau} = \varepsilon \left[-s + (s + K - \lambda) c \right]$$

$$\frac{dc}{d\tau} = s - (s+K)c$$

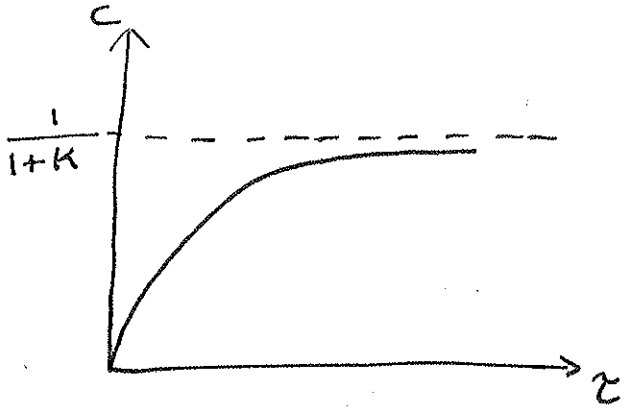
since ε is small, First equation goes to zero so that easily integrated $\Rightarrow$ $s = 1$

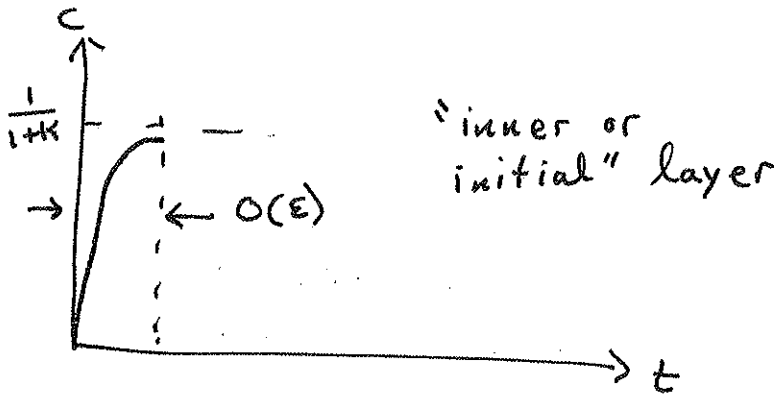
$$\Rightarrow \quad \frac{dc}{d\tau} = 1 - (1+K)c \quad , \quad c(0) = 0$$

Integrate: $\quad c = \frac{1}{1+K} \left(1 - e^{-(1+K)\tau} \right)$

$$c = \frac{1}{1+K}\left(1 - e^{-(1+K)\tau}\right)$$

$$\Rightarrow c = \frac{1}{1+K}\left(1 - e^{-(1+K)t/\varepsilon}\right)$$





The point is, s is a "slow" variable whereas c is a "Fast" variable.

Numerically, we refer to this rapid behavior as "stiffness" in the problem.

EXAMPLE: Diffusion Problem

ODEs often arise through some reduction (analytical or numerical) of a PDE

$$u_t = (p(x)u_x)_x + g(x,u) \quad , \quad 0<x<1 \, , \quad t>0$$

where $u(x,t)$ = temperature
$p(x)$ = diffusivity
$g(x,u)$ = heat source

Boundary conditions: $u(0,t) = \alpha(t)$, $u(1,t) = \beta(t)$

Initial conditions: $u(x,0) = g(x)$

use Method of Lines

introduce grid: $x_j = jh$, $h = \frac{1}{m+1}$, $1 \le j \le m$

$v_j(t) \approx v(x_j, t)$

Replace derivatives in x with Finite difference approximations

$$\frac{d}{dt} v_j = \frac{1}{h}\left[p(x_j + \tfrac{h}{2})\left(\frac{v_{j+1} - v_j}{h}\right) - p(x_j - \tfrac{h}{2})\left(\frac{v_j - v_{j-1}}{h}\right)\right]$$

pv_x at $x_j + \frac{h}{2}$; pv_x at $x_j - \frac{h}{2}$

set of ODEs, $1 \le j \le m$, $v_0(t) = \alpha(t)$

$v_{m+1}(t) = \beta(t)$

$v_j(0) = g(x_j)$

Remarks

1) Unlike previous cases, the number of ODEs is typically large. Might use only a lower order method and match the order of accuracy of the spatial discretization. Can be expensive computationally.

2) Equations are stiff

A Regularity Result for IVPs

IVP $\quad \underline{y}' = \underline{f}(t, \underline{y}) \quad , \quad \underline{y}(0) = \underline{c} \quad , \quad 0 \le t \le b$

Theorem Let $\underline{f}(t, \underline{y})$ be continuous for all $(t, \underline{y})$ in a region

$$D = \left\{ 0 \le t \le b \ , -\infty < \|y\| < \infty \right\} .$$

Moreover, assume that $\underline{f}$ is Lipschitz in the variable $\underline{y}$, ie a constant L exists such that

$$\| \underline{f}(t, \underline{y}) - \underline{f}(t, \hat{\underline{y}}) \| \le L \| \underline{y} - \hat{\underline{y}} \|$$

for all $(t, \underline{y})$, $(t, \hat{\underline{y}})$ in D.

Then:

(1) For any $\underline{c} \in \mathbb{R}^m$ there exists a unique solution $\underline{y}(t)$ throughout the interval $[0, b]$. The solution is differentiable.

(2) The solution depends continuously on the initial data, ie if $\hat{y}(t)$ also satisfies the ODE, then

$$\| y(t) - \hat{y}(t) \| \le e^{Lt} \| y(0) - \hat{y}(0) \|$$

(3) IF $\hat{y}(t)$ satisfies, more generally, a perturbed ODE

$$\hat{y}' = \underline{f}(t, \hat{\underline{y}}) + \underline{r}(t, \hat{\underline{y}}) \quad \text{where } \underline{r} \text{ is continuous and}$$

bounded in D, $\|\underline{r}\| < M$. Then

$$\| y(t) - \hat{y}(t) \| \le e^{Lt} \| y(0) - \hat{y}(0) \| + \frac{M}{L} \left(e^{Lt} - 1 \right)$$

Remarks

1) If $\underline{f}(t,\underline{y})$ is smooth then $\underline{F}$ is Lipschitz if

$$\left\| \frac{\partial \underline{f}}{\partial \underline{y}}(t,\underline{y}) \right\| \leq L \quad \forall \ (t,\underline{y}) \text{ in } D.$$

2) Often $\underline{F}$ cannot be shown to be Lipschitz for D if $-\infty < \|y\| < \infty$. A restricted result:

Suppose $\underline{F}$ satisfies

$$\| \underline{F}(t,\hat{\underline{y}}) - \underline{F}(t,\underline{y}) \| \leq L \| \hat{\underline{y}} - \underline{y} \| \quad \text{for}$$

$D = \{ 0 \leq t \leq b, \ \| \underline{y} - c \| \leq \gamma \}$. Suppose also that $\| \underline{F} \| \leq M$ on D. Then the results above hold but for $0 \leq t \leq \min\{ b, \gamma/M \}$

Boundary Value Problems

A BVP involves an ODE defined on some interval and data supplied at more than one point on the interval. Typically, the data is supplied on the boundaries.

2-point BVP

For example: $\underline{y}' = \underline{F}(t, \underline{y})$, $0 \le t \le b$

$$g(y(0), y(b)) = 0$$

$\underline{y}(t)$ is a vector of m scalar functions, $\underline{F}$ is a given "slope" function and g is a set of m algebraic equations

Consider the solution of an associated initial value problem, $\underline{y}' = \underline{F}(t, \underline{y})$, $\underline{y}(0) = \underline{s}$, $0 \le t \le b$ call the solution $\underline{y}(t; \underline{s})$. Then to solve the BVP,

$g(\underline{s}, y(b, \underline{s})) = 0$ - a set of m nonlinear algebraic equations for $\underline{s}$.

Even if the IVP is well behaved, there is no guarantee that the solution of the BVP exists and if it exists, whether it is unique.

The solution behavior of a BVP is non-local.

Example

1) A diffusion problem

$$u_t = (p(x) u_x)_x + q(x, u) \quad , \quad 0 \le x \le 1 \quad , \quad t \ge 0$$

BCs $u(0, t) = \alpha(t)$, $u(1, t) = \beta(t)$

ICs $u(x, 0) = g(x)$

Suppose that α, β approach constant values for t sufficiently large, so that as $t \to \infty$ a steady state is reached.

The steady state solves for $v = v(x)$

$$0 = (p(x)v_x) + g(x,v) \quad , \quad 0 \le x \le 1$$

$$v(0) = \alpha \quad , \quad v(1) = \beta$$

write as a first-order system:

set $y_1(x) = v$ (temperature)

and $y_2(x) = p(x)v'$ (heat flux)

ODEs:

$$y_1' = \frac{y_2}{p(x)} \quad , \quad y_2' = -g(x,v) \quad , \quad 0 \le x \le 1$$

$$y_1(0) = \alpha \quad , \quad y_1(1) = \beta$$

no conditions given for y_2

Often $p(x) \neq 0$, in fact, $p(x) > 0$. There are cases when $p(x) = 0$ (typically at a boundary) and then some special treatment is required.

2) Singular Perturbation Problem

Consider the second-order, linear BVP:

$$\varepsilon v'' + a(x) v' + b(x) v = 0, \qquad 0 \le x \le 1$$

$$v(0) = \alpha, \quad v(1) = \beta$$

Suppose $a(x)$ and $b(x)$ are smooth and that $\varepsilon > 0$ and small in magnitude. We want to determine the solution behavior in the limit $\varepsilon \to 0$.

Consider the case $\varepsilon = 0$.

$$a(x) v' + b(x) v = 0$$

$$\rightarrow \quad \frac{v'}{v} = -\frac{b(x)}{a(x)} \qquad \text{assume } a(x) \neq 0 \text{ for } 0 \le x \le 1$$

Integrate:

$$\ln v = -\int \frac{b(x)}{a(x)} dx + c$$

$$v = c \exp\left\{ -\int \frac{b(x)}{a(x)} dx \right\}$$

Choose c such that v satisfies the boundary condition:

eg, if $v(0) = \alpha$ then

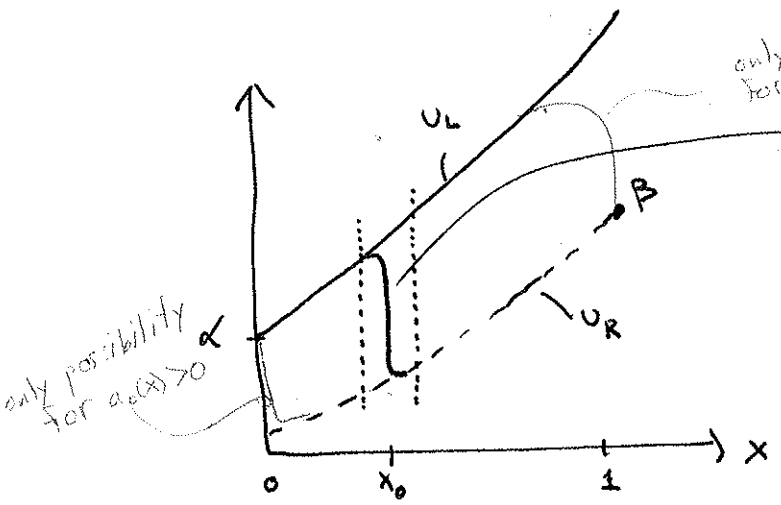
$$v_L = \alpha \exp\left\{ -\int_0^x \frac{b(s)}{a(s)} ds \right\}$$

notice the negative sign

or if $v(1) = \beta$

$$v_R = \beta \exp\left\{ \int_x^1 \frac{b(s)}{a(s)} ds \right\}$$

In general, $v_L \neq v_R$ unless $\beta = \alpha \exp\left\{ -\int_0^1 \frac{b(s)}{a(s)} ds \right\}$



layer solution where u'' is large

Consider "layer-type" solutions

$$\xi = \frac{x - x_0}{\varepsilon} \quad \longrightarrow \quad x = x_0 + \xi\varepsilon$$

$$\frac{d}{dx} = \frac{d}{d\xi}\cdot\frac{d\xi}{dx} = \frac{1}{\varepsilon}\frac{d}{d\xi}$$

$$\frac{d^2}{dx^2} = \frac{1}{\varepsilon^2}\frac{d^2}{d\xi^2}$$

$$\Rightarrow \varepsilon\left(\frac{1}{\varepsilon^2}\frac{d^2}{d\xi^2}u\right) + a(x_0+\xi\varepsilon)\frac{1}{\varepsilon}\frac{d}{d\xi}u + b(x_0+\xi\varepsilon)u = 0$$

third term is small compared to first two terms

Consider $\varepsilon = 0$: $\quad u'' + a(x_0)u' = 0$

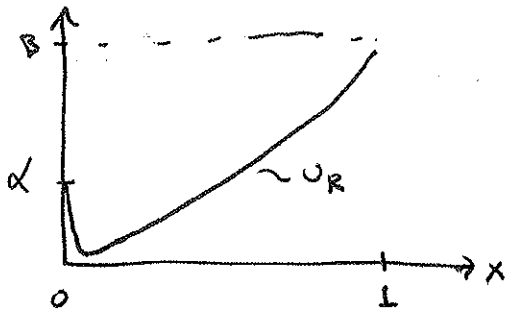
general solution: $\quad u = A + Be^{-a(x_0)\xi}$

solution behavior depends on value of $a(x_0)$

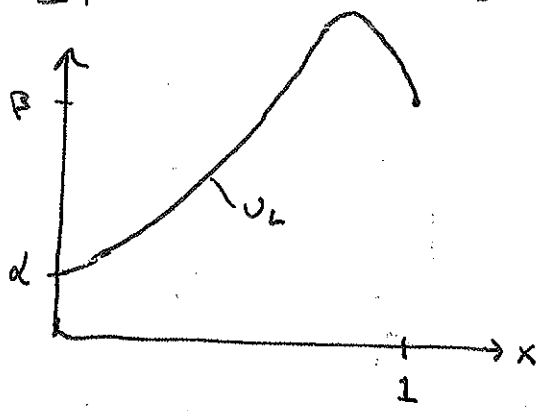
$a(x_0) > 0 \Rightarrow$ decay exponentially as ξ increases

$a(x_0) < 0 \Rightarrow$ decay exponentially as ξ decreases

If $a > 0$ then $x_0 = 0$



If $a < 0$ then $x_0 = 1$



No internal layers are possible

Suppose $a_0(x_0) > 0$

$$v = A + B e^{-a\xi}$$

at $\xi = 0$ $\quad v = A + B = \alpha$

as $\xi \to \infty$ $\quad v = A = v_R(0) = \beta \exp\left\{ \int_0^{\cdot} \frac{b(s)}{a(s)} ds \right\}$

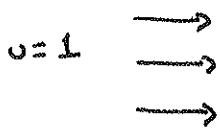
then $B = \alpha - A$

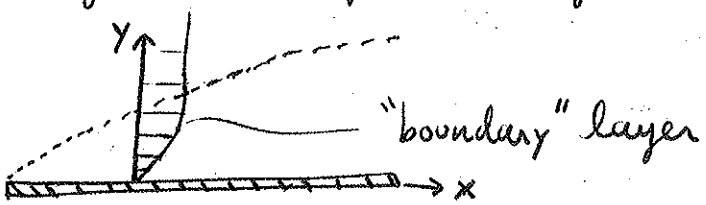
This is the solution, to 1st order of accuracy.

---

3) Blasius Problem (BOUNDARY LAYER)

Consider the flow of an incompressible fluid over a flat plate.





Eqns for velocity: (u, v)

$$u_x + v_y = 0 \quad \text{(conservation of mass)}$$

$$uu_x + vu_y = \nu u_{yy} \quad \text{(balance of momentum in x-direction)}$$

Boundary conditions

$$u = v = 0 \text{ at } y = 0$$

$$u = 1 \text{ at } y \to \infty$$

Reduce the eqns for $u(x,y)$ and $v(x,y)$ to an ODE for a single variable ψ.

Define $\psi(x,y)$ such that

$$u = \psi_y \,, \quad v = -\psi_x$$

with this choice, the conservation of mass eqn is satisfied identically. The momentum balance eqn becomes

$$\psi_y \psi_{xy} - \psi_x \psi_{yy} = \nu \psi_{yyy}$$

BCs $\quad \psi_x = \psi_y = 0 \quad$ on $y = 0 \quad \Rightarrow \quad \psi = 0 \quad$ on $y = 0$

$$\psi_y \to 1 \quad \text{as} \quad y \to \infty$$

Solution for ψ has the form

$$\psi(x,y) = \sqrt{\nu x}\, f(\eta) \,, \quad \eta = \frac{y}{\sqrt{\nu x}}$$ similarity solution

where η = similarity variable

Need to find a differential equation for $f(\eta)$

$$\psi_x = \frac{1}{2}\sqrt{\frac{\nu}{x}}\, f(\eta) + \sqrt{\nu x}\, f'(\eta)\left(\frac{-y x^{-3/2}}{\sqrt{\nu}}\right)$$

$$\psi_x = \frac{\sqrt{\nu}}{2} f(\eta)\left(x^{-1/2}\right) - \frac{y}{2x} f'(\eta) \qquad y = \eta\sqrt{2x}$$

$$\psi_x = \frac{1}{2}\sqrt{\frac{\nu}{x}}\, f(\eta) - \frac{1}{2x}\,\eta\sqrt{\nu x}\, f'(\eta)$$

$$\bullet \quad \psi_x = \frac{1}{2}\sqrt{\frac{\nu}{x}}\left(f(\eta) - \eta f'(\eta)\right)$$

$$\psi_y = \sqrt{\nu x}\, f'(\eta)\frac{1}{\sqrt{\nu x}} \quad \Rightarrow \quad \bullet \quad \psi_y = f'(\eta)$$

$$\psi_{yx} = \psi_{xy} = f''(\eta)\left(-\frac{1}{2}\frac{y}{\sqrt{\nu}\, x^{3/2}}\right) = f''(\eta)\left(\frac{-\eta\sqrt{\nu x}}{2\sqrt{x^3\nu}}\right) = \frac{-\eta}{2x} f''(\eta)$$

$$\psi_{yy} = \frac{1}{\sqrt{\nu x}} f''(\eta)$$

$$\psi_{yyy} = \frac{1}{\nu x} f'''(\eta)$$

Substitute partial derivatives into

$$\psi_y \psi_{xy} - \psi_x \psi_{yy} = \nu \psi_{yyy}$$

$$\Rightarrow \frac{1}{2}\left((f'(\eta))^2 - f(\eta) f''(\eta) \right) = f'''(\eta)$$

Boundary conditions

$\psi_y = 0$ on $y = 0$ $\Rightarrow$ $f'(0) = 0$

$\psi_x = 0$ on $y = 0$ $\Rightarrow$ $f(0) = 0$

$\psi_y \to 1$ as $y \to \infty$ $\Rightarrow$ $\lim_{\eta \to \infty} f'(\eta) = 1$

Let $f''(0) = s$, shear stress

pick a value for s, integrate out to infinity, compare $f'(\eta)$ to $\lim_{\eta \to \infty} f'(\eta) = 1$. Adjust s as necessary. This is the shooting method.

Differential Algebraic Equation

$y' = F(t, y)$ explicit ODE

$F(t, y, y') = 0$ implicit ODE

IF $\frac{\partial F}{\partial y'}$, Jacobian matrix, is non singular, then the implicit equation can be "solved" for y'.

A DAE is one for which $\frac{\partial F}{\partial y'}$ is singular. Then the eqn involves both "differential" variables and "algebraic" variables.

Some cases this is made explicit:

$$\text{DAE} \begin{cases} y' = f(t, y, z) \\ 0 = g(t, y, z) \end{cases}$$

$z = z(t)$

In this case, $y(t)$ = differential variables, z = algebraic variables.

One issue is what boundary or initial data is allowed.

eg: $y' = z$, $0 = y - t \rightarrow y = t$ $\Rightarrow$ $1 = z$

Here, no data was needed.

IF the sequations are simple enough, solve for z interms of t, y and substitute into y' for ODE.

1/24/03

Example

(1) Mechanics: Let some mechanical system be described by a set of n generalized coordinates.

q_j, $j = 1, \ldots, n$ = generalized coordinates

Typically, q_j = positions or possibly angles, areas, etc...

Then define generalized velocities as q_j', $j = 1, \ldots, n$.

Let $g_i(t, \underline{q}) = 0$ be a set of m constraints on the motion. $(i = 1, \ldots, m)$

Suppose $v = v(\underline{q})$ is the potential energy of the system and $T = T(\underline{q}, \underline{q}')$ is the kinetic energy of the system.

1/24/03

A Lagrangian for the system is

$$L = -U + T - \sum_{i=1}^{m} \lambda_i g_i \quad , \quad L = L(t)$$

where λ are the ~~Lagrangian~~ Lagrange multipliers.

Hamilton's Principle - the motion of the system from $t=0$ to $t=b$ is one that minimizes the integral $I_H = \int_0^b L(t)\,dt$ $\Rightarrow$ Calculus of Variations

$\Rightarrow$ Euler Lagrange equations.

$$\frac{\partial}{\partial q_j} L - \frac{d}{dt}\left(\frac{\partial L}{\partial q_j'}\right) = 0 \quad , \quad j = 1, \ldots, n$$

$\Rightarrow$ system of n 2nd order ODEs for q_j

let $\underline{v} = \underline{q}'$ be the generalized velocities

$\Rightarrow$

$$\boxed{\begin{aligned} \underline{\underline{M}}(t,\underline{q})\,\underline{v}' &= \underline{f}(t,\underline{q},\underline{v}) - \underline{\underline{G}}^T(t,\underline{q})\,\lambda \\ \underline{v} &= \underline{q}' \\ 0 &= \underline{g}(t,\underline{q}) \end{aligned}}$$

where $\underline{v}$ - generalized velocities

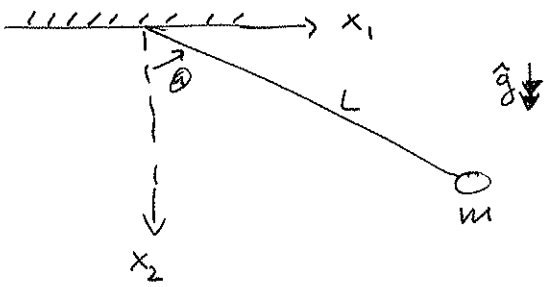
$\underline{\underline{M}}$ - "mass matrix," (positive definite)

$\underline{f}$ - applied forces

$\underline{\underline{G}}$ - $\dfrac{\partial \underline{g}}{\partial \underline{q}}$

Specific

Now consider a ~~2nd~~ case:



(x_1, x_2) - Cartesian coordinates

$\hat{g}$ - gravity

let $q_j = x_j$, $j = 1, 2$

potential energy, $U = -m\hat{g}x_2$ (PE goes to 0 when $x_2 \to 0$)

kinetic energy, $T = \frac{m}{2}\left(x_1'^2 + x_2'^2\right)$

constraint: $g = x_1^2 + x_2^2 - L^2 = 0$

Lagrangian: $L = -U + T - \lambda g$

$$L = \frac{m}{2}\left(x_1'^2 + x_2'^2\right) + m\hat{g}x_2 - \lambda\left(x_1^2 + x_2^2 - L^2\right) = 0$$

Euler-Lagrange eqns:

$$\frac{\partial L}{\partial x_1} - \frac{d}{dt}\left(\frac{\partial L}{\partial x_1'}\right) = 0 \Rightarrow \boxed{-2\lambda x_1 - m x_1'' = 0}$$

$$\frac{\partial L}{\partial x_2} - \frac{d}{dt}\left(\frac{\partial L}{\partial x_2'}\right) = 0 \Rightarrow \boxed{m\hat{g} - 2\lambda x_2 - m x_2'' = 0}$$

let $x_1' = v_1$ and $x_2' = v_2$

$\Rightarrow$

$$\boxed{\begin{aligned} m v_1' &= -2\lambda x_1 \\ m v_2' &= m\hat{g} - 2\lambda x_2 \\ x_1^2 + x_2^2 - L^2 &= 0 \end{aligned}}$$

Note, that if you set

$x_1 = L\sin\theta$
$x_2 = L\cos\theta$ $\Rightarrow mL\ddot{\theta} + mg\sin\theta = 0$

Initial Value Problems

$$\underline{y}' = \underline{F}(t, \underline{y}) \quad , \quad t \geq 0 \quad , \quad \underline{y}(0) = \underline{c}$$

"Problem Stability"

- idea: given a solution $\underline{y}(t)$ of the IVP, the stability refers to the sensitivity of the solution (output) to perturbations in the data (input)

caution on "stability"

- "stability" is overused. Here stability might better be termed "conditioning," and "stability" refering to a numerical method,

Unfortunately there are many versions of "stability" defined for the numerics.

Test Equation (scalar equation)

$$y' = \lambda y, \quad t \geq 0 \quad , \quad y(0) = c$$

Here, λ is a constant, possibly complex. Often λ is an eigenvalue.

solution: $y(t) = y(0)\, e^{\lambda t}$, $\lambda = a + ib$

$$y(t) = y(0)\, e^{at + ibt}$$

$$= y(0)\, e^{at}(\cos bt + i \sin bt)$$

imaginary part of λ represents oscillatory behavior and the real part of λ represents exponential growth or decay.

$\Rightarrow$ in terms of stability, we're only interested in the $\mathrm{Re}\,\lambda$

Suppose $\hat{y}(t)$ solves the test eqn but with $\hat{y}(0)$ as the initial condition.
Let $y(t)$ be "the" solution to the test equation and $\hat{y}(t)$ the perturbed solution, then

$$|\hat{y}(t) - y(t)| = |y(\hat{0})e^{\lambda t} - y(0)e^{\lambda t}|$$
$$= e^{(\mathrm{Re}\lambda)t}|y(\hat{0}) - y(0)|$$

IF $\mathrm{Re}\lambda \le 0$, then $|\hat{y}(t) - y(t)|$ remains bounded, ie stable.

IF $\mathrm{Re}\lambda < 0$, then $|\hat{y}(t) - y(t)|$ approaches zero $\Rightarrow$ asymptotically stable.

IF $\mathrm{Re}\lambda > 0$, unstable (perturbations grow).

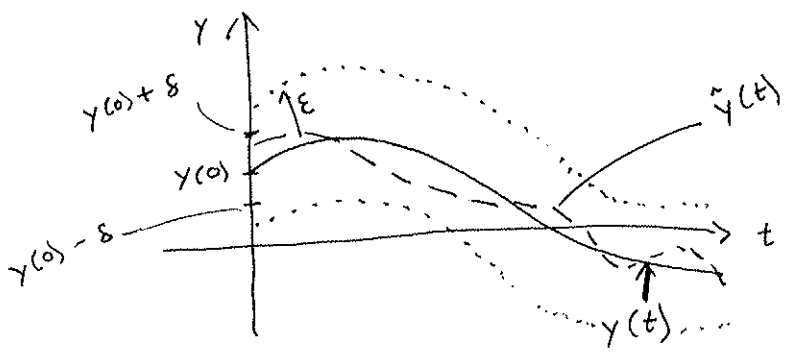
Full Equation

$$\underline{y}' = \underline{f}(t, \underline{y}), \quad \underline{y}(0) = \underline{c}, \quad t \ge 0$$

The idea is similar but the details a little harder.
Suppose $\underline{y}(t)$ is the solution of the ODE for $t \ge 0$.
The solution is <u>stable</u> if given any $\varepsilon > 0$ there is a $\delta > 0$ st any other solution $\hat{y}(t)$ satisfying the ODE and

$$\|\hat{\underline{y}}(0) - \underline{y}(0)\| < \delta$$

also satisfies $\|\hat{\underline{y}}(t) - \underline{y}(t)\| < \varepsilon$ for all $t \ge 0$.

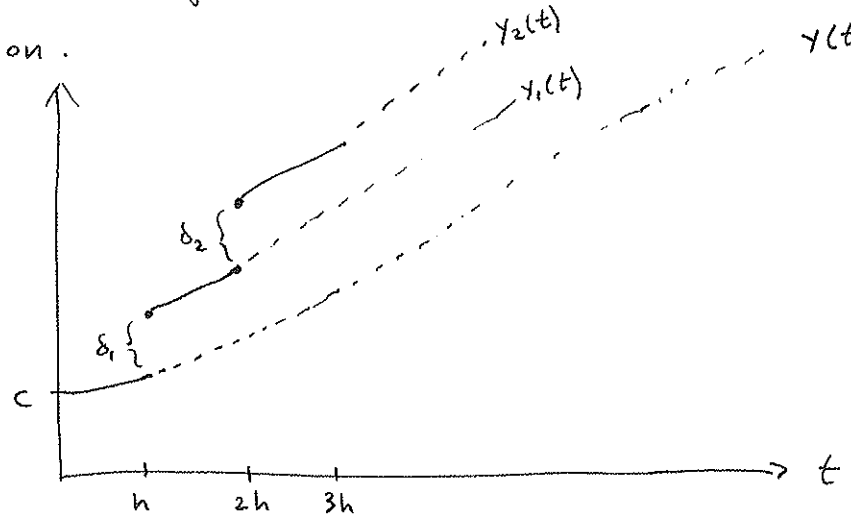


The solution is <u>asymptotically stable</u> if in addition to being stable
$\|\underline{y}(t) - \hat{\underline{y}}(t)\| \to 0$ as $t \to \infty$.

Example : Consider the effect of a sequence of perturbations on the solution of the test equation.

We have $y' = \lambda y$, $y(0) = c$, $t \geq 0$

Suppose the solution is perturbed at $t = h$ by an amount δ_1. The perturbed solution is then itself perturbed at $t = 2h$ by δ_2 and so on.



Model for a sequence of discretization errors.

$$y(t) = c e^{\lambda t}$$

$$y_1(t) = \left(y(h) + \delta_1\right) e^{\lambda(t-h)}$$

$$= \left(c e^{\lambda h} + \delta_1\right) e^{\lambda(t-h)} = c e^{\lambda t} + \delta_1 e^{\lambda(t-h)}$$

$$y_2(t) = \left(y_1(2h) + \delta_2\right) e^{\lambda(t-2h)} = \left(c e^{\lambda 2h} + \delta_1 e^{\lambda h} + \delta_2\right) e^{\lambda(t-2h)}$$

$$= c e^{\lambda t} + \delta_1 e^{\lambda(t-h)} + \delta_2 e^{\lambda(t-2h)}$$

$$\vdots$$

$$\boxed{y_j(t) = c e^{\lambda t} + \sum_{k=1}^{j} \delta_k e^{\lambda(t-kh)}}$$ perturbed solution

Difference: $$y_j(t) - y(t) = \sum_{k=1}^{j} \delta_k e^{\lambda(t-kh)}$$

Behavior depends on λ :

If $\mathrm{Re}\,\lambda \le 0$, then the perturbations only accumulate linearly with no exponential growth in time (at most).

If $\mathrm{Re}\,\lambda < 0 \Rightarrow$ asymptotically stable.

Linear Systems

1) Constant coefficient, homogeneous problems,

$$\underline{y}' = \underline{\underline{A}}\,\underline{y}, \quad t \ge 0, \quad A \text{ is } m \times m \text{ matrix (real)}.$$

$$\boxed{\underline{y}(t) = e^{\underline{\underline{A}}t}\,\underline{y}(0)}$$

notice, $e^{\underline{\underline{A}}t}$ is a matrix

$e^{\underline{\underline{A}}t}$ is defined via power series:

$$e^{\underline{\underline{A}}t} = I + \underline{\underline{A}}t + \frac{1}{2}\underline{\underline{A}}^2 t^2 + \frac{1}{3!}\underline{\underline{A}}^3 t^3 + \cdots$$

Note, if $\underline{\underline{A}}$ is diagonalizable then $\underline{\underline{T}}^{-1}\underline{\underline{A}}\,\underline{\underline{T}} = \underline{\underline{\Lambda}} = \mathrm{diag}(\lambda_i)$ $i = 1, \ldots, m$

and column vectors of $\underline{\underline{T}}$ are the (right) eigenvectors of $\underline{\underline{A}}$.

If $\underline{\underline{T}}$ s.t. $\underline{\underline{T}}^{-1}\underline{\underline{A}}\,\underline{\underline{T}} = \underline{\underline{\Lambda}} \Rightarrow \underline{\underline{A}} = \underline{\underline{T}}\,\underline{\underline{\Lambda}}\,\underline{\underline{T}}^{-1}$

$$\underline{\underline{A}}^2 = \left(\underline{\underline{T}}\,\underline{\underline{\Lambda}}\,\underline{\underline{T}}^{-1}\right)\left(\underline{\underline{T}}\,\underline{\underline{\Lambda}}\,\underline{\underline{T}}^{-1}\right) = \underline{\underline{T}}\,\underline{\underline{\Lambda}}^2\,\underline{\underline{T}}^{-1}$$

$$\Rightarrow \quad e^{\underline{\underline{A}}t} = \underline{\underline{T}}\left(\underline{\underline{I}} + \underline{\underline{\Lambda}}t + \frac{1}{2}\underline{\underline{\Lambda}}^2 t^2 + \cdots\right)\underline{\underline{T}}^{-1}$$

$$e^{\underline{\underline{A}}t} = \underline{\underline{T}}\begin{bmatrix} e^{\lambda_1 t} & & & \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ & & & e^{\lambda_m t} \end{bmatrix}\underline{\underline{T}}^{-1}$$

$$\underline{y}(t) = \underline{\underline{T}} \begin{pmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_m t} \end{pmatrix} \underline{\underline{T}}^{-1} \underline{y}(0)$$

$$\Rightarrow \quad \underline{\underline{T}}^{-1} \underline{y}(t) = \begin{pmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_m t} \end{pmatrix} \underline{\underline{T}}^{-1} \underline{y}(0)$$

IF $\underline{w}(t) = \underline{\underline{T}}^{-1} \underline{y}(t)$ then

$$\underline{w}(t) = \begin{pmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_m t} \end{pmatrix} \underline{w}(0) \quad \rightarrow \quad w_j(t) = w_j(0)\, e^{\lambda_j(t)}$$

For diagonalizable $\underline{\underline{A}}$, if $\mathrm{Re}\,\lambda_j \leq 0$ $\forall j \Rightarrow$ problem stability.

if $\mathrm{Re}\,\lambda_j < 0$ $\forall j \Rightarrow$ asymptotic stability.

1/28/03

IF the difference between $\hat{y}$ and y remains bounded, then the IVP is stable.

IF the difference is bounded and approaches zero as $t \to \infty$, then the IVP is asymptotically stable.

The IVP is unstable if the difference is unbounded.

Suppose $\underline{\underline{A}}$ is not diagonalizable.

Then $\exists$ T matrix nonsingular, satisfying

$$\underline{\underline{T}}^{-1}\,\underline{\underline{A}}\,\underline{\underline{T}} = \begin{bmatrix} \underline{\underline{\Lambda}}_1 & & \\ & \underline{\underline{\Lambda}}_2 & \\ & & \ddots \\ & & & \underline{\underline{\Lambda}}_n \end{bmatrix},$$ being in Jordan Canonical form

where $\Lambda_j = \begin{bmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{bmatrix}$ — Jordan block

For a diagonalizable matrix Λ_j is one by one

$$\underline{y}(t) = \underline{\underline{T}} \begin{bmatrix} \underline{\underline{\Lambda}}_1 & & \\ & \underline{\underline{\Lambda}}_2 & \\ & & \ddots \\ & & & \underline{\underline{\Lambda}}_n \end{bmatrix} \underline{\underline{T}}^{-1} y(0)$$

The question of stability depends on behaviour of term $e^{\underline{\underline{\Lambda}}_j t}$. The corresponding ODE is

$$\begin{aligned} y_1' &= \lambda y_1 + y_2 \\ y_2' &= \lambda y_2 + y_3 \\ &\vdots \\ y_{n-1}' &= \lambda y_{n-1} + y_n \\ y_n' &= \lambda y_n \end{aligned}$$

For example, suppose $n=2$ then

$$y_1' = \lambda y_1 + y_2, \quad y_1(0) = \alpha, \quad y_2(0) = \beta$$

$$y_2' = \lambda y_2$$

$$\searrow\ y_2 = \beta e^{\lambda t} \longrightarrow y_1' = \lambda y_1 + \beta e^{\lambda t}$$

$$\downarrow$$

$$y_1 = (\alpha + \beta t)\, e^{\lambda t}$$

$$y_1 = (\alpha + \beta t) e^{\lambda t}$$

$$y_2 = \beta e^{\lambda t}$$

1) Stable if all eigenvalues of $\underline{\underline{A}}$ satisfy either

$\operatorname{Re}\lambda < 0$ OR $\operatorname{Re}\lambda = 0$ and λ is simple

(λ is simple - belongs to 1×1 Jordan block)

2) asymptotic stability if $\operatorname{Re}\lambda < 0$ for all eigenvalues of A.

Variable Coefficient, Non-homogeneous

$$\underline{y}'(t) = \underline{\underline{A}}(t)\,\underline{y}(t) + \underline{q}(t), \quad t \geq 0$$

where $\underline{\underline{A}}$ is an $m \times m$ real matrix function and $\underline{q}$ is an m-vector, real function, both smooth.

The exact solution is found in terms of the Fundamental solution matrix: $\underline{\underline{Y}}' = \underline{\underline{A}}(t)\underline{\underline{Y}}$, $t \geq 0$ where $\underline{\underline{Y}}(0) = I$. We assume that $\underline{\underline{Y}}(t)$ is known.

Set $\underline{y}(t) = \underline{\underline{Y}}(t)\,\underline{g}(t)$ — vector function to be determined

This approach is variation of parameters.

$$\rightarrow \quad \underline{y}'(t) = \underline{\underline{Y}}'(t)\,\underline{g}(t) + \underline{\underline{Y}}(t)\,\underline{g}'(t) = \underline{\underline{A}}\,\underline{\underline{Y}}\,\underline{g} + \underline{\underline{Y}}\,\underline{g}' = \underline{\underline{A}}\,\underline{y} + \underline{\underline{Y}}\,\underline{g}'$$

$$\Rightarrow \underline{\underline{A}}\,\underline{y} + \underline{\underline{Y}}\,\underline{g}' = \underline{\underline{A}}\,\underline{y} + \underline{q} \quad \Rightarrow \quad \underline{\underline{Y}}\,\underline{g}' = \underline{q}$$

$$\underline{g}' = \underline{\underline{Y}}^{-1}\,\underline{g}$$

$$\Rightarrow \underline{g}(t) = \int_0^t \underline{\underline{Y}}^{-1}(s)\,\underline{q}(s)\,ds + \underline{c}$$

$$\boxed{\underline{y}(t) = \underline{\underline{Y}}(t)\left[\underline{c} + \int_0^t \underline{\underline{Y}}^{-1}(s)\,\underline{q}(s)\,ds\right]}$$

IF a perturbed solution $\hat{\underline{y}}(t)$ solves the same ODE but with initial conditions $\hat{\underline{y}}(0) = \hat{\underline{c}}$, then the difference is bounded if $\underline{\underline{Y}}(t)$ is bounded.

IF a perturbed solution $\hat{\underline{y}}(t)$ solves a perturbed ODE as well, then require $\underline{\underline{Y}}(t)\,\underline{\underline{Y}}^{-1}(t)$

Asymptotic stability if $\underline{\underline{Y}} \to 0$ as $t \to \infty$.

Note that in the variable coefficient case, a solution may grow over certain intervals of time but still be stable or asymptotically stable.

$$\frac{dy}{dt} = \cos t\, y(t) \quad \to \quad \underline{y}(t) = e^{\sin t}$$

The problem is stable (clearly $e^{\sin t}$ is a bounded function) but NOT asymptotically stable.

Nonlinear Case

$$\underline{y}' = \underline{f}(t, \underline{y}) \quad , \quad t \geq 0 \quad , \quad \underline{y}(0) = \underline{c}$$

If $\underline{y}(t)$ is a known trajectory, then is it stable?

One approach is to linearize about $\underline{y}(t)$. Consider a perturbed solution $\hat{y}(t)$ solving

$$\hat{\underline{y}}' = \underline{f}(t, \underline{y}) \quad , \quad \hat{\underline{y}}(0) = \hat{\underline{c}}$$

Expand about $\underline{y}(t)$: (Jacobian matrix)

$$\hat{\underline{y}}' = \underbrace{\underline{f}(t, \underline{y})}_{\underline{y}'} + \underbrace{\frac{\partial \underline{f}}{\partial \underline{y}}(t, \underline{y})}_{\underline{\underline{A}}(t)} (\hat{\underline{y}} - \underline{y}) + \cdots \text{ higher order terms}$$

If $\underline{w}(t) = \hat{\underline{y}} - \underline{y}$, then $\underline{w}(t)$ solves

$$\underline{w}'(t) = \underline{\underline{A}}(t)\, \underline{w} + \cdots$$

If $\underline{w}$ is small then $\underline{w}$ solves

$$\underline{w}'(t) = \underline{\underline{A}}(t)\, \underline{w}(t)$$

to a first approximation.

Can perform an analysis of the linearized problem and this could provide necessary conditions for the nonlinear problem.

Hamiltonian Systems

A Hamiltonian system is a set of ODEs of the form $q_i' = \frac{\partial H}{\partial p_i}$, $p_i' = -\frac{\partial H}{\partial q_i}$, $i = 1, 2, \ldots, l$

This is a system of $2l$ equations for the variables (q_i, p_i), $i = 1, \ldots, l$, depending on time t. H is a scalar function, $H = H(q, p)$, and is a smooth function called the Hamiltonian.

In a mechanical system, q = position of a set of particles and p = momentum of particles, and H is the sum of potential and kinetic energies. (The total energy of the system).

$$\frac{dH}{dt} = \frac{\partial H}{\partial \underline{p}} \cdot \frac{d\underline{p}}{dt} + \frac{\partial H}{\partial \underline{q}} \cdot \frac{d\underline{q}}{dt}$$

$$= \frac{\partial H}{\partial \underline{p}} \cdot \left(-\frac{\partial H}{\partial \underline{q}}\right) + \frac{\partial H}{\partial \underline{q}} \cdot \left(\frac{\partial H}{\partial \underline{p}}\right)$$

$$= 0 \quad \Rightarrow \quad H = \text{constant} \Rightarrow \text{total energy is conserved}$$

A geometric implication of this is that the flow conserves area.

Consider a 2-vector problem

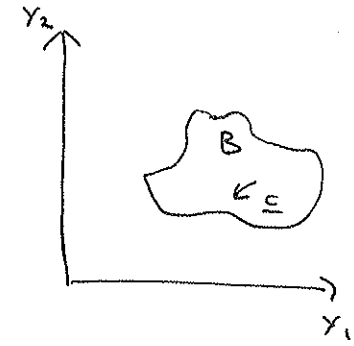
(autonomous system)

$$\underline{Y}'(t) = \underline{F}(\underline{Y})$$

$$\underline{Y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \underline{Y}(0) = \underline{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Assume a solution is given by $\underline{y}(t;\underline{c})$.

Consider the set B and the flow of B under the trajectories $\underline{y}(t;\underline{c})$ where, $\underline{c} \in B$.



Define $S(t)B = \{\underline{y}(t;\underline{c}),\ \underline{c} \in B\}$

If $\underline{y}$ is governed by a constant coefficient linear system then the behavior of the region depends on the stability of the equations. If the problem is asymptotically stable, for example, then the region shrinks to zero. If the $\nabla \cdot \underline{f} = \frac{\partial f_1}{\partial y_1} + \frac{\partial f_2}{\partial y_2} = 0$

then the area remains fixed. (The area of $S(t)B$ remains fixed).

For the Hamiltonian system with $\ell = 1$,

$$f_1 = \frac{\partial H}{\partial p}, \quad f_2 = -\frac{\partial H}{\partial q}, \quad \underline{y} = \begin{bmatrix} q \\ p \end{bmatrix}$$

$$\nabla \cdot \underline{f} = \frac{\partial}{\partial q}\left(\frac{\partial H}{\partial p}\right) + \frac{\partial}{\partial p}\left(-\frac{\partial H}{\partial q}\right) = 0$$

$\Rightarrow$ the area of the flow is preserved

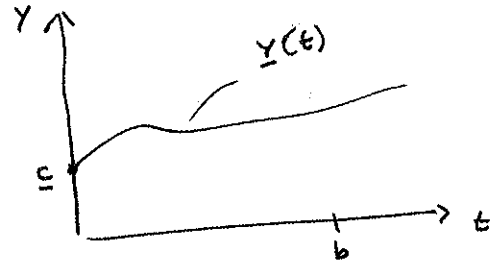
For $\ell = 1$ it turns out that the area is preserved in every (q_i, p_i) plane. The flow is said to be symplectic.

The implication for problem stability is that a Hamiltonian system cannot be asymptotically stable or unstable. A numerical solution may want to mimick this area preserving property $\Rightarrow$ simplectic numerical methods.

Numerical Methods for IVPs

$\underline{y}' = \underline{F}(t, \underline{y})$, $t \geq 0$, $\underline{y}(t) = \underline{c}$

the solution is a trajectory $\underline{y}(t)$:



Approximate the solution on a mesh or grid.

$0 = t_0 < t_1 < \cdots < t_N = b$

mesh spacing: $h_n = t_n - t_{n-1} > 0$

The solution is approximated by a set of vectors $\underline{y}_n \in \mathbb{R}^m$ st

$\underline{y}_n \simeq \underline{y}(t_n)$ for $n = 0, \ldots, N$.

Basic Concepts (consistency, stability, convergence)

Illustrate the basic concepts for Euler's method:

$$\underline{y}(t_n) = \underline{y}(t_{n-1}) + \underline{y}'(t_{n-1})(t_n - t_{n-1}) + \frac{1}{2}\underline{y}''(t_{n-1})(t_n - t_{n-1})^2 + \cdots$$

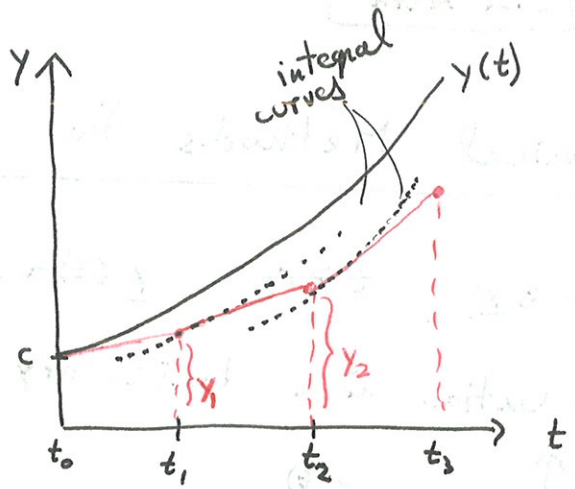
$$= \underline{y}(t_{n-1}) + \underline{F}\left(t_{n-1}, \underline{y}(t_{n-1})\right) h_n + O(h_n^2) + \cdots$$

Euler's method ignores $O(h_n^2)$ terms (and higher) to write

$$\underline{y}_n = \underline{y}_{n-1} + h_n \underline{F}(t_{n-1}, \underline{y}_{n-1}), \quad \underline{y}_0 = \underline{c},$$

Geometry

Move along successive tangents to the integral curves.



the smaller the mesh spacing, the more accurate the solution

BASIC CONCEPTS

1/31/03

Euler's Method: $Y_n = Y_{n-1} + h\,F(t_{n-1}, Y_{n-1})$

$$\frac{Y_n - Y_{n-1}}{h} = F(t_{n-1}, Y_{n-1})$$

Define the difference operator N_h

$$N_h\, v(t_n) \equiv \frac{v(t_n) - v(t_{n-1})}{h_n} - F(t_{n-1}, Y_{n-1})$$

N_h acts on a grid function $v(t_n)$

N_h is an approximation of N, a differential operator, where $N\,y(t) = y' - f(t, y)$

The local truncation error measures the amount by which the exact solution of the differential equation fails to satisfy the difference eqn.

Define d_n as the local truncation error and set $d_n = N_h\, y(t_n)$. A numerical method is consistent if $d_n \to 0$ as $h_n \to 0$.

A numerical method is consistent of order p if $|d_n| = O(h_n^p)$ as $h_n \to 0$

(The numerical method is p^{th}-order accurate.)

$$|d_n| = O(h_n^p) \text{ as } h_n \to 0 \Rightarrow \lim_{h_n \to 0} \frac{|d_n|}{h_n^p} = \text{const.}$$

For Euler's Method

$$N_h y(t_n) = \frac{y(t_n) - y(t_{n-1})}{h_n} - f(t_{n-1}, y(t_{n-1}))$$

$$N_h y(t_n) = \frac{\left(y(t_{n-1}) + h_n y'(t_{n-1}) + \frac{h_n^2}{2} y''(t_{n-1}) + \cdots\right) - y(t_{n-1})}{h_n} - f(t_{n-1}, y(t_{n-1}))$$

$$N_h y(t_n) = \underbrace{\frac{h_n}{2} y''(t_{n-1})}_{\text{dominant behavior of } d_n} + \cdots$$

Hence, Euler's method is first-order accurate

If a numerical method is consistent then $N_h \to N$ as $h \to 0$, but what does this say about the the error, $e_n = y_n - y(t_n)$?

Define error, $e_n = y_n - y(t_n)$

A numerical method is convergent if

$$\max_{0 \le n \le N} |e_n| \to 0 \text{ as } h \to 0 \quad \text{where } h = \max_n h_n.$$

A numerical method is convergent of order p if

$$\max_{0 \le n \le N} |e_n| = O(h^p) \text{ as } h \to 0$$

Typically, if a numerical method is consistent of order p then it is convergent of order p. To ensure this you must show that the method is stable.

The difference method is "0-stable" (or just "stable") if $h_0 > 0$ and $K > 0$ exist st

$$|x_n - z_n| \le K\left[|x_0 - z_0| + \max_{1\le j\le N} |N_h x_j - N_h z_j|\right] \quad \text{for } 0 \le n \le N$$

for any mesh functions x_n, z_n and $h \le h_0$

(Recall our discussion of the regularity of the differential equation - this is similar,

Thm: If a difference method is consistent of order p with local truncation error d_n and if it is 0-stable with constant K, then it is convergent of order p st $|e_n| \le K \max_{1\le j\le N} |d_j| = O(h^p)$

whenever you talk of numerical methods for differential eqns, the usual theorem is

consistency + stability $\Rightarrow$ convergence

Proof If a method is 0-stable then

$$|x_n - z_n| \le K\left[|x_0 - z_0| + \max_{1\le j\le N} |N_h x_j - N_h z_j|\right]$$

set $x_n = y_n$, set $z_n = y(t_n)$

$$|e_n| \le K\left[\underbrace{|e_0|}_{0} + \max_{1\le j\le N} |0 - d_j|\right]$$

$$\Rightarrow |e_n| \le K \max_{1\le j\le N} |d_j|$$

Examine 0-Stability for Euler

Set $s_n = x_n - z_n$, $B = \max_j |N_h x_j - N_h z_j|$

use $N_h x_j = \frac{x_j - x_{j-1}}{h_j} - f(t_{j-1}, x_{j-1})$

$N_h z_j = \frac{z_j - z_{j-1}}{h_j} - f(t_{j-1}, z_{j-1})$

so $B \geq |N_h x_{j_n} - N_h z_{j_n}|$ for $1 \leq n \leq N$

$$\Rightarrow B \geq \left| \frac{s_n - s_{n-1}}{h_n} + f(t_{j_n-1}, x_{j_n-1}) - f(t_{j_n-1}, z_{n-1}) \right|$$

$$\geq \left| \frac{s_n}{h_n} - \left(\frac{s_{n-1}}{h_n} + f(t_{n-1}, x_{n-1}) - f(t_{n-1}, z_{n-1}) \right) \right|$$

$$\geq \left| \frac{s_n}{h_n} \right| - \left| \frac{s_{n-1}}{h_n} + f(t_{n-1}, x_{n-1}) - f(t_{n-1}, z_{n-1}) \right|$$

$$\Rightarrow \frac{|s_n|}{h_n} \leq \left| \frac{s_{n-1}}{h_n} + f(t_{n-1}, x_{n-1}) - f(t_{n-1}, z_{n-1}) \right| + B$$

multiply by h_n, use triangle inequality

$$\Rightarrow |s_n| \leq |s_{n-1}| + h_n \left| f(t_{n-1}, x_{n-1}) - f(t_{n-1}, z_{n-1}) \right| + h_n B$$

Key step: assume f satisfies Lipschitz condition

$$|f(t_{n-1}, x_{n-1}) - f(t_{n-1}, z_{n-1})| \leq L |s_{n-1}|$$

$$\Rightarrow |s_n| \leq (1 + h_n L) |s_{n-1}| + h_n B$$

$$\leq (1 + h_n L) \left[(1 + h_{n-1} L) |s_{n-2}| + h_{n-1} B \right] + h_n B$$

$$= (1 + h_n L)(1 + h_{n-1} L) |s_{n-2}| + \left(h_n + (1 + h_n L) h_{n-1} \right) B$$

remember Triangle inequality

$$||a| - |b|| \leq |a+b| \leq |a| + |b|$$

continue with this process:

$$|s_n| \le (1+h_nL)(1+h_{n-1}L)\cdots(1+h_1L)|s_0| + \left[h_n + (1+h_nL)h_{n-1} + (1+h_nL)(1+h_{n-1}L)h_{n-2} + \cdots + (1+h_nL)(1+h_{n-1}L)\cdots(1+h_2L)h_1\right]B$$

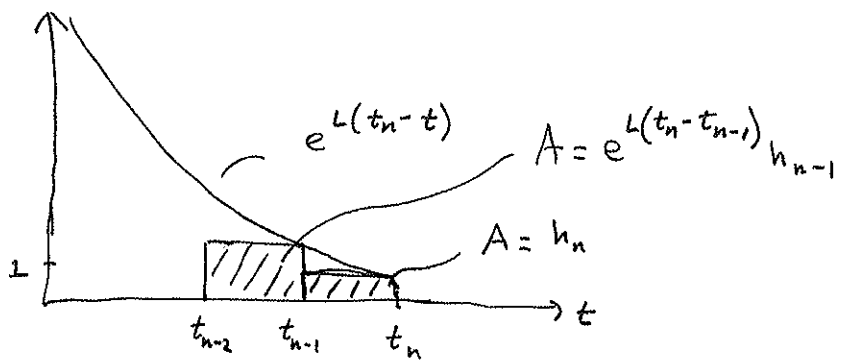
Use $1+hL \le e^{Lh}$

Note $(1+h_nL)(1+h_{n-1}L)\cdots(1+\underbrace{h_1}_{1})L \le e^{L(h_1+\cdots+h_n)} = e^{Lt_n}$

$$(1+h_nL)(1+h_{n-1}L)\cdots(1+\underbrace{h_{j+1}}_{j+1})L \le e^{L(h_{j+1}+\cdots+h_n)} = e^{L(t_n-t_j)}$$

$$\Rightarrow |s_n| \le e^{Lt_n}|s_0| + \left[h_n + e^{L(t_n-t_{n-1})}h_{n-1} + e^{L(t_n-t_{n-2})}h_{n-2} + \cdots + e^{L(t_n-t_1)}h_1\right]B$$

Notice:



$$\Rightarrow \left[h_n + e^{L(t_n-t_{n-1})}h_{n-1} + e^{L(t_n-t_{n-2})}h_{n-2} + \cdots + e^{L(t_n-t_1)}h_1\right] \le \int_0^{t_n} e^{L(t_n-t)}\,dt$$

$$= \frac{1}{L}\left(e^{Lt_n}-1\right)$$

$$\Rightarrow |s_n| \le e^{Lt_n}|s_0| + \frac{B}{L}\left(e^{Lt_n}-1\right)$$

We want to show that $|s_n| \le K(|s_0|+B)$

$$|s_n| \le e^{Lt_n}|s_0| + \frac{B}{L}\left(e^{Lt_n}-1\right)$$

Let $K = \max\left\{e^{Lb}, \frac{1}{L}\left(e^{Lb}-1\right)\right\}$

then $|s_n| \le K\left(|s_0| + B\right)$

Remarks:

1) Recall the error bound $|e_n| \le K \max_{1\le j\le N} |d_j|$

where $e_n = y_n - y(t_n)$, $d_j = N_h y(t_j)$ = truncation error

For Euler's method,

$$d_j = \frac{h_j}{2} y''(t_{j-1}) + O(h_j^2)$$

$$\Rightarrow |e_n| \le \frac{K}{2} \max_j |h_j y''(\tau_j)| \quad , \quad t_{j-1} \le \tau_j \le t_j$$

where $K = \max\left(e^{Lb}, \frac{1}{L}(e^{Lb}-1)\right)$

A direct analysis of the error gives the bound that $|e_n| \le \frac{Mh}{2L}\left(e^{Lb}-1\right)$ where $M = \max_{0\le t\le b} |y''(t)|$, $h = \max_{1\le j\le N} h_j$

2) The above bound is a very poor estimate for the error. One reason is that the bound produces exponential growth in the error, which is not necessarily the case. There are other ways to estimate the error for finite values of h. The purpose of the bound above is to establish the behavior as $h \to 0$.

(3) The stability analysis for Euler's method may be extended to explicit single-step methods of the form

$$N_h y_n = \frac{y_n - y_{n-1}}{h_n} - \phi(t_{n-1}, y_{n-1}, h_n) = 0$$

where ϕ is the stepping function.
(this includes Runge Kutta methods).

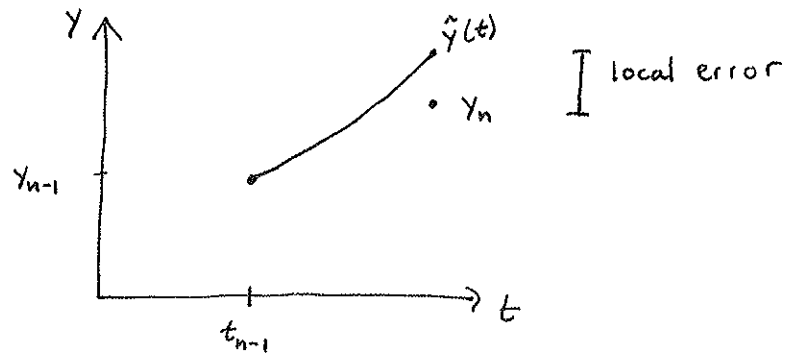
Require that $\phi(t, y, h)$ satisfy a Lipschitz condition in the variable y.

Local Error

This is a measure of discretization error and is NOT the local truncation, though it is related. The local error is a measure of the error committed by one step of the difference formula.

Consider the IVP: $\hat{y}' = f(t, \hat{y})$

$$\hat{y}(t_{n-1}) = y_{n-1}$$



The local error is related to the local truncation error.

For Euler's method

$$\ell_n = \hat{y}(t_n) - \left(y_{n-1} + h_n F(t_{n-1}, y_{n-1})\right)$$

$$= \hat{y}(t_n) - \left[\hat{y}(t_{n-1}) + h_n F(t_{n-1}, \hat{y}(t_{n-1}))\right]$$

$$= h_n \left[\frac{\hat{y}(t_n) - \hat{y}(t_{n-1})}{h_n} - F(t_{n-1}, \hat{y}(t_{n-1}))\right]$$

$$= h_n \underbrace{N_n \hat{y}(t_n)}_{\text{almost } d_n}$$

It can be shown that $\ell_n = h_n d_n \left(1 + O(h_n)\right)$, $h = \max h_n$

IF the numerical method is consistent of order p then $|\ell_n| = O(h^{p+1})$ as $h \to 0$

2/4/03

Absolute Stability

So far we have introduced concepts that concern behavior of a numerical method in the limit $h \to 0$,

e.g. a method is <u>consistent of order p</u> if

$$d_n = N_n y(t_n) = \text{truncation error} = O(h^p) \text{ as } h \to 0$$

and if the method is <u>0-stable</u> then it is <u>convergent of order p</u>, ie

$$e_n = y_n - y(t_n) = O(h^p) \text{ as } h \to 0$$

IF we make h sufficiently small then we can achieve a desired accuracy. The question is, how small does h have to be to get an acceptable numerical solution.

Consider the test equation: $y' = \lambda y$, $t \geq 0$, $y(0) = c$ where λ - complex constant. Apply Euler's method:

$$y_n = y_{n-1} + h_n f(t_{n-1}, y_{n-1}) \quad , \quad y_0 = c$$

$$y_n = y_{n-1} + h_n \lambda y_{n-1} \quad \Rightarrow \quad y_n = (1 + h_n \lambda) y_{n-1}$$

Suppose h is fixed, then

$$y_n = (1 + h\lambda) y_{n-1}$$

$$\Rightarrow \boxed{\begin{aligned} y_n &= (1+h\lambda)^n c \quad \text{- numerical solution} \\ y(t) &= c e^{\lambda t} \quad \text{- exact solution} \end{aligned}}$$

How does the numerical and exact solutions compare for a fixed h?

Cases:

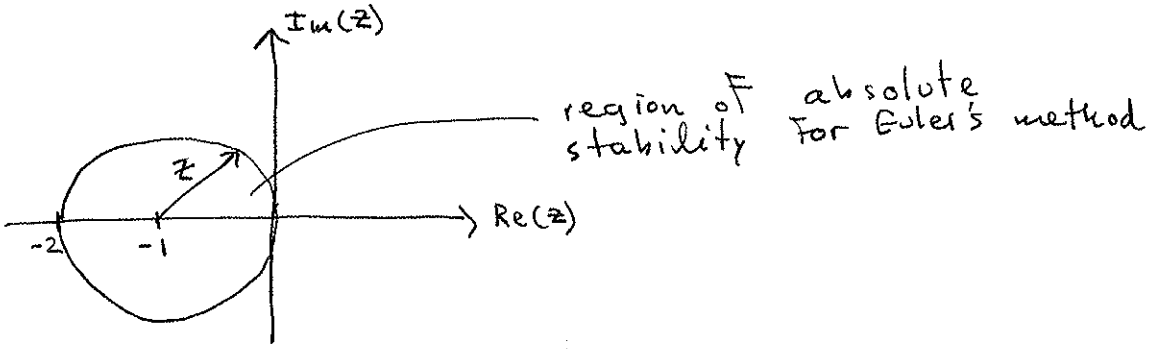
1) $\mathrm{Re}\,\lambda > 0$. For the unstable case, the exact solution grows exponentially and $(1+h\lambda)^n$ grows with n. Qualitatively, the solutions have the same growth behavior. Absolute error grows exponentially but the relative error might be ok.

2) $\mathrm{Re}\,\lambda \leq 0$. The exact solution is bounded and decays exponentially if $\mathrm{Re}\,\lambda < 0$. The quantity $(1+h\lambda)$ only decays for a certain h. Therefore we require

$$|1 + h\lambda| \leq 1$$

For a bounded numerical solution. This requirement implies a region of absolute stability.

Let $z = h\lambda$ = complex. Describe the region in the z-plane such that $|1 + z| \leq 1$



Suppose that λ is real and negative then we're sitting somewhere on negative real axis, not necessarily restricted to region of absolute stability. For a decaying numerical solution, we require

$$-2 \leq \lambda h \leq 0 \quad \Rightarrow \quad -2 < \lambda h < 0 \quad \Rightarrow \quad h < \frac{2}{-\lambda}$$

this implies that as $-\lambda$ becomes large, h must become small so that the decay is captured.

Extend the analysis to constant coefficient systems.

Constant Coefficient systems

$$\underline{y}' = \underline{\underline{A}}\,\underline{y}, \quad t \geq 0, \quad \underline{y}(0) = \underline{c}$$

Apply Euler's method with fixed h.

$$\underline{y}_n = \underline{y}_{n-1} + h\,\underline{\underline{A}}\,\underline{y}_{n-1} \quad \Rightarrow \quad \underline{y}_n = \left(\underline{\underline{I}} + h\underline{\underline{A}}\right)\underline{y}_{n-1}$$

If $\underline{\underline{A}}$ is diagonalizable, then set $\underline{w}_n = \underline{\underline{T}}^{-1}\underline{y}_n$

$$\Rightarrow \underline{\underline{T}}^{-1}\underline{y}_n = \underline{\underline{T}}^{-1}\left(\underline{\underline{I}} + h\underline{\underline{A}}\right)\underline{\underline{T}}\,\underline{\underline{T}}^{-1}\underline{y}_{n-1}$$

$$\Rightarrow \underline{w}_n = \left(\underline{\underline{I}} + h\,\underline{\underline{T}}^{-1}\underline{\underline{A}}\,\underline{\underline{T}}\right)\underline{w}_n$$

$$\boxed{\underline{w}_n = \left(\underline{\underline{I}} + h\underline{\underline{\Lambda}}\right)\underline{w}_{n-1}}$$

$$\underline{w}_n = \left(\underline{\underline{I}} + h\underline{\underline{A}}\right)\underline{w}_{n-1}$$

For a bounded solution, require $|1 + h\lambda_j| \leq 1$

For $j = 1, \ldots, m$ where λ_j are the eigenvalues of $\underline{\underline{A}}$.

Example: spring equation: $mv'' + \nu v' + kv = 0$, $t \geq 0$

m - mass, ν - damping coeff, $v(0) = \alpha$, $v'(0) = \beta$

k - spring constant : $m, \nu, k > 0$

The general solution is: $v(t) = Ae^{\lambda_1 t} + Be^{\lambda_2 t}$

where λ_1, λ_2 are roots of the characteristic polynomial: $m\lambda^2 + \nu\lambda + k = 0$

$$\lambda = \frac{-\nu \pm \sqrt{\nu^2 - 4mk}}{2m}$$

Since m, ν, k are positive, the real parts of both eigenvalues λ_1, λ_2 are negative $\Rightarrow$ the problem is asymptotically stable.

Consider the behavior of Euler's method for two cases:

1) Overdamped: $\nu^2 - 4mk > 0$ $\Rightarrow$ both roots λ_1, λ_2 are real

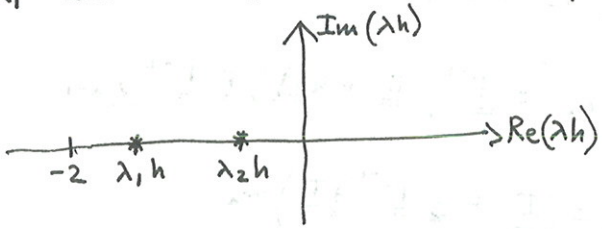
the solution is $v = Ae^{\lambda_1 t} + Be^{\lambda_2 t}$

solution decays - real eigenvalues

where $\lambda_1 = \frac{-\nu - \sqrt{\nu^2 - 4mk}}{2m}$, $\lambda_2 = \frac{-\nu + \sqrt{\nu^2 - 4mk}}{2m}$ $(\lambda_1 < \lambda_2)$

$$A = \frac{\alpha\lambda_2 - \beta}{\lambda_2 - \lambda_1}, \quad B = \frac{\beta - \alpha\lambda_1}{\lambda_2 - \lambda_1}$$

We require that $h\lambda_1$ and $h\lambda_2$ lie in the region of absolute stability



$\Rightarrow h < \frac{2}{-\lambda_1}$ we're only worried about $\lambda_1 h$

$m u'' + \nu u' + k u = 0, \quad t \geq 0, \quad u(0) = \alpha, \quad u'(0) = \beta$

$\lambda_1 = \frac{-\nu - \sqrt{\nu^2 - 4mk}}{2m}, \quad \lambda_2 = \frac{-\nu + \sqrt{\nu^2 - 4mk}}{2m}$

$m = 0.5, \quad k = 0.5, \quad \nu = 0.2$

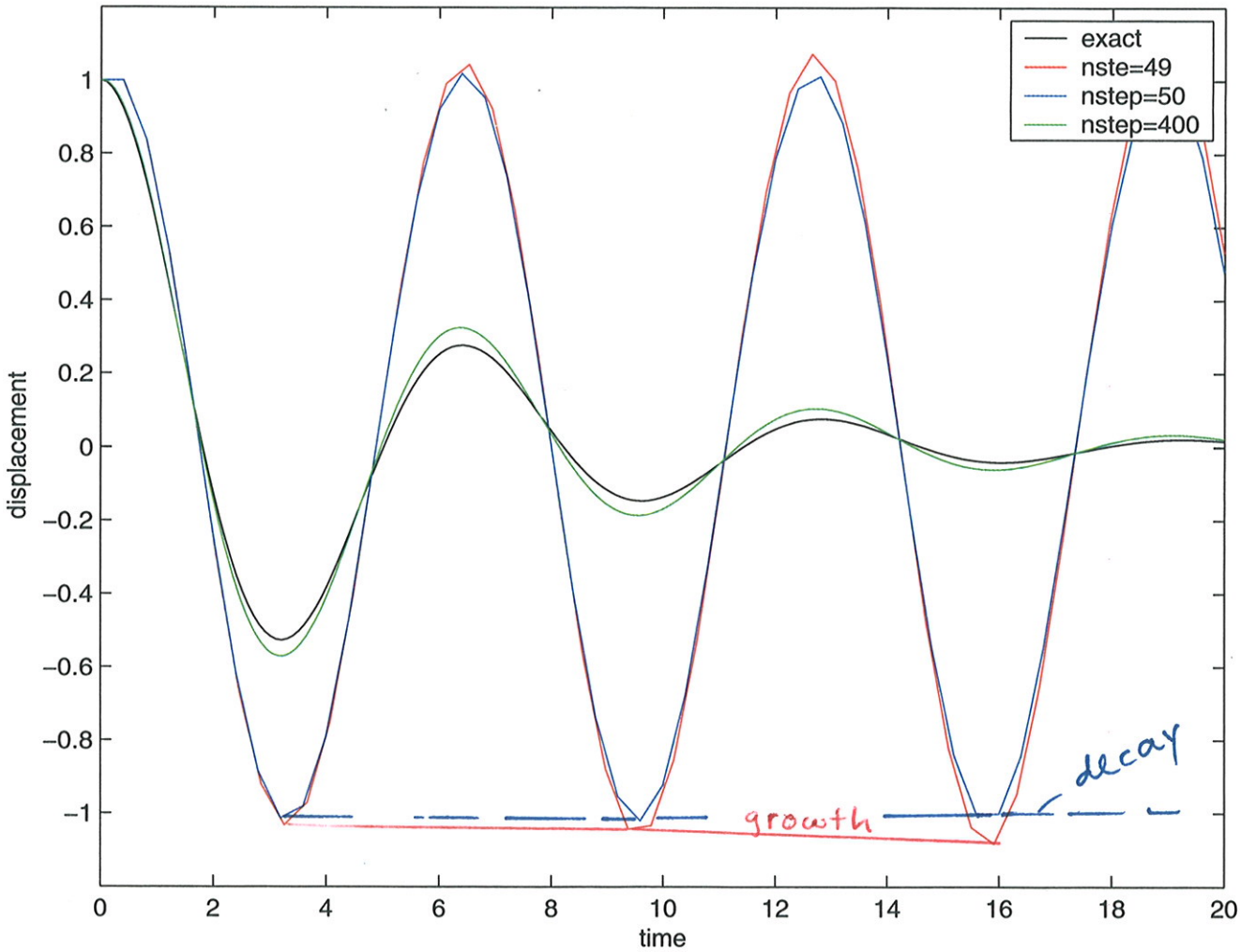
~~$\lambda_1 = -0.2 - i\,0.98\ldots$~~
~~$\lambda_2 = -0.2 + i\,0.98\ldots$~~
~~UNDERDAMPED~~

$\lambda = \frac{-\nu}{2m} = -0.2$

Using Euler's method, absolute stability requires that ~~$h < \frac{2}{-\lambda_1}$~~

$h \leq \frac{\nu}{k} = 0.4$

Region of Absolute Stability



when nstep = 49, $h = \frac{20}{49} = 0.408$, does not obey region, so grows

nstep = 50, $h = \frac{20}{50} = 0.4$, on boundary of disk of stability

Suppose $m = k = \frac{1}{2}$, $\nu = 2$

$\lambda_1 = -2 - \sqrt{3}$

$\Rightarrow \quad h < \frac{2}{2+\sqrt{3}} \simeq 0.536$

---

2) Underdamped: $\nu^2 - 4mk < 0$

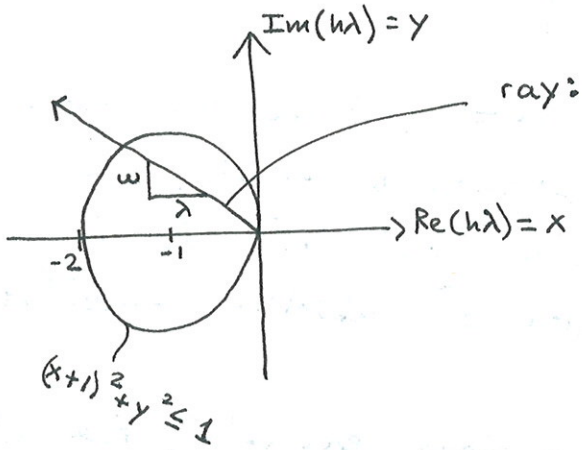
Here λ_1, λ_2 are complex conjugates

The solution is: $u = e^{\lambda t}\left(A\cos(\omega t) + B\sin(\omega t)\right)$

solution is oscillatory (imaginary eigenvalues)

where $\lambda = \frac{-\nu}{2m}$, $\omega = \frac{1}{2m}\sqrt{4mk - \nu^2}$, $A = \alpha$, $B = \frac{\beta - \lambda\alpha}{\omega}$

Absolute stability



$(x+1)^2 + y^2 = (\lambda h + 1)^2 + (\omega h)^2 \le 1 \Rightarrow \lambda^2 h^2 + 2\lambda h + 1 + \omega^2 h^2 \le 1$

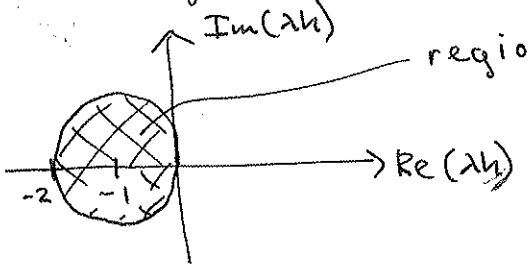
$\rightarrow h(h\lambda^2 + 2\lambda + \omega^2 h) \le 0 \rightarrow \lambda^2 h + 2\lambda + \omega^2 h \le 0$

$\rightarrow h \le \frac{-2\lambda}{\lambda^2 + \omega^2} = \frac{\nu}{k}$

If $m = k = \frac{1}{2}$, $\nu = \frac{1}{5}$ then $h \le \frac{2}{5} = 0.4$

Stiffness & Implicit Methods

Region of absolute stability - apply a numerical method to the test equation, $y' = \lambda y$. Plot the region in the complex λh-plane for which the numerical solution is bounded. For Euler's method, the region is a disk of unit radius:



IF $\mathrm{Re}\,\lambda \le 0$, then the exact solution of the test equation is stable and the numerical solution is bounded only if h is chosen st $h\lambda$ is in the region of absolute stability.

In the previous example, the choice of h was determined by accuracy and not stability $\Rightarrow$ "non-stiff".

In other cases, the situation could be reversed and a small h is dictated by stability $\Rightarrow$ "stiff."

Example: spring with a small mass

$$\varepsilon v'' + 2v' + v = 0 \quad , \quad 0 \le t \le O(1)$$

$$v(0) = 0 \quad , \quad \varepsilon v'(0) = 1$$

where ε is a positive constant

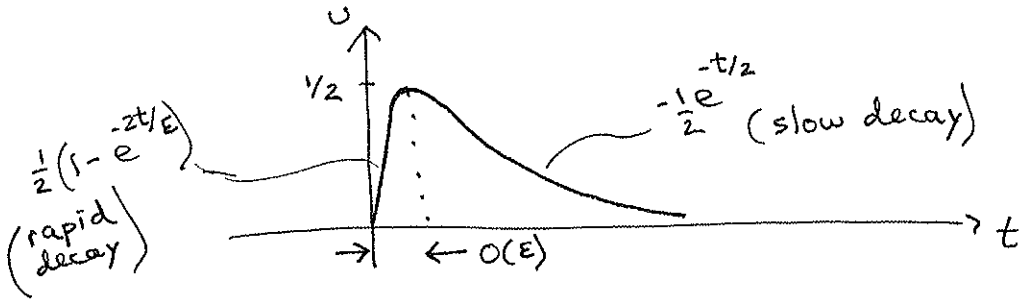
exact solution: $v(t) = A(e^{r_1 t} - e^{r_2 t})$

where $r_1 = \frac{-1}{\varepsilon} + \frac{\sqrt{1-\varepsilon}}{\varepsilon}$, $r_2 = \frac{-1}{\varepsilon} - \frac{\sqrt{1-\varepsilon}}{\varepsilon}$, $A = \frac{1}{2\sqrt{1-\varepsilon}}$

in the limit as $\varepsilon \to 0$, $r_1 \to \frac{-1}{2}$, $r_2 \to \frac{-2}{\varepsilon}$, $A \to \frac{1}{2}$

and $v \to \frac{1}{2}\left(e^{-t/2} - e^{-2t/\varepsilon}\right)$

this is a problem of multiple time scales



Accuracy $\Rightarrow$ $h \sim O(\varepsilon)$ for $t \sim O(\varepsilon)$ to resolve rapid decay

and $h \sim O(1)$ for $t \sim O(1)$ to resolve slow decay

Suppose we want to apply Euler's method, then take the equation and write it as a system:

$x_1 = v$, $x_2 = v'$ $\Rightarrow$

$$x_1' = x_2$$
$$x_2' = -\frac{1}{\varepsilon}(x_1 + 2x_2)$$

this is of the form $y' = Ay$, where $A = \begin{pmatrix} 0 & 1 \\ -1/\varepsilon & -2/\varepsilon \end{pmatrix}$

We'll have: $y_n = y_{n-1} + hAy_{n-1}$ } choose h

$y_0 =$ given

Concerning absolute stability, $h \leq \frac{2}{-\lambda_j}$, $j = 1, 2$

What are the eigenvalues? $\lambda_1 \sim -1/2$, $\lambda_2 = -2/\varepsilon$

λ_2 is the 'controlling' eigenvalue as its magnitude is greater than that of λ_1 $\Rightarrow$ $\boxed{h \leq \frac{2}{-(-2/\varepsilon)} = \varepsilon}$ this is true for all time

The problem of stiffness is essentially an issue of multiple scales, ie, the solution evolves on different time scales.

eg, $\underline{y}' = \underline{\underline{A}}\,\underline{y}$, $0 \le t \le b$

A has eigenvalues λ_j, $j = 1, \ldots, m$

We require all eigenvalues with $\mathrm{Re}\,\lambda_j \le 0$ to lie in region of abs stability

The problem is stiff if $b(-\mathrm{Re}\,\lambda_j) \gg 1$.

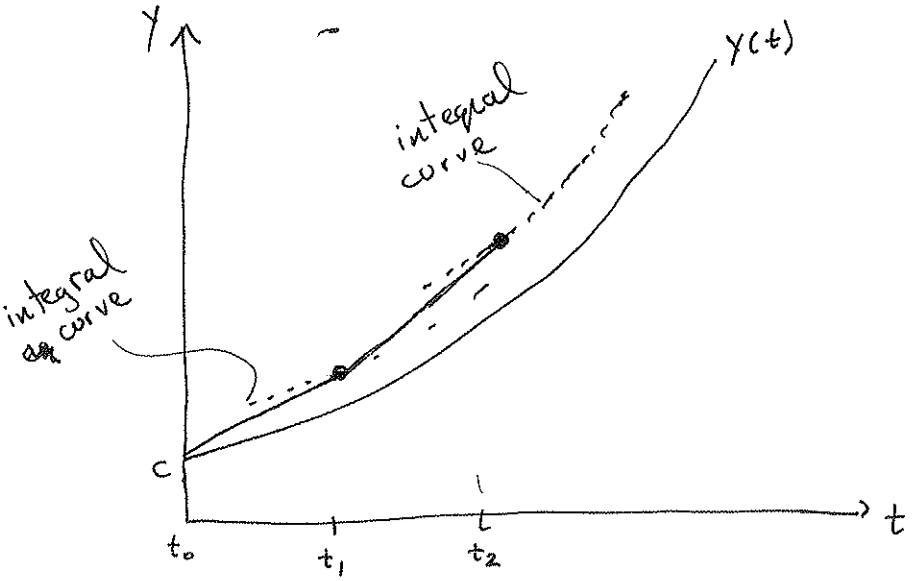
Most methods for stiff equations are implicit.

Standard case: backward Euler method

$$y' = f(t,y), \quad t \ge 0, \quad y(0) = c$$

backward Euler: $y_n = y_{n-1} + h f(t_n, y_n)$

The derivation involves a Taylor series centered at (t_n, y_n).



Solve

y_n is defined implicitly $\Rightarrow$ need to solve a set of (nonlinear) algebraic eqns to find y_n at each step.

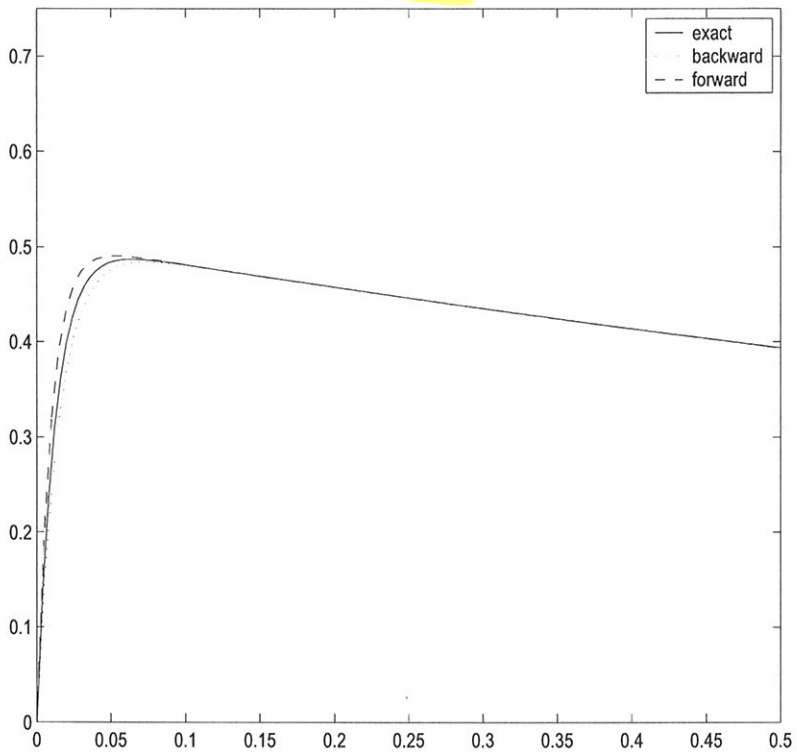
25a

$$\varepsilon u'' + 2u' + u = 0$$

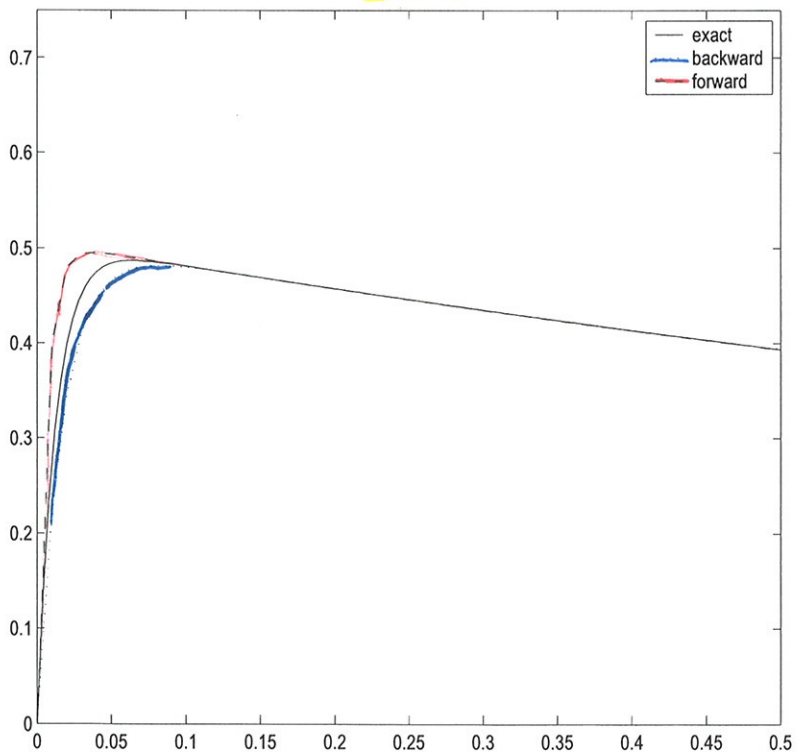
$$u(0) = 0$$

$$\varepsilon u'(0) = 1$$

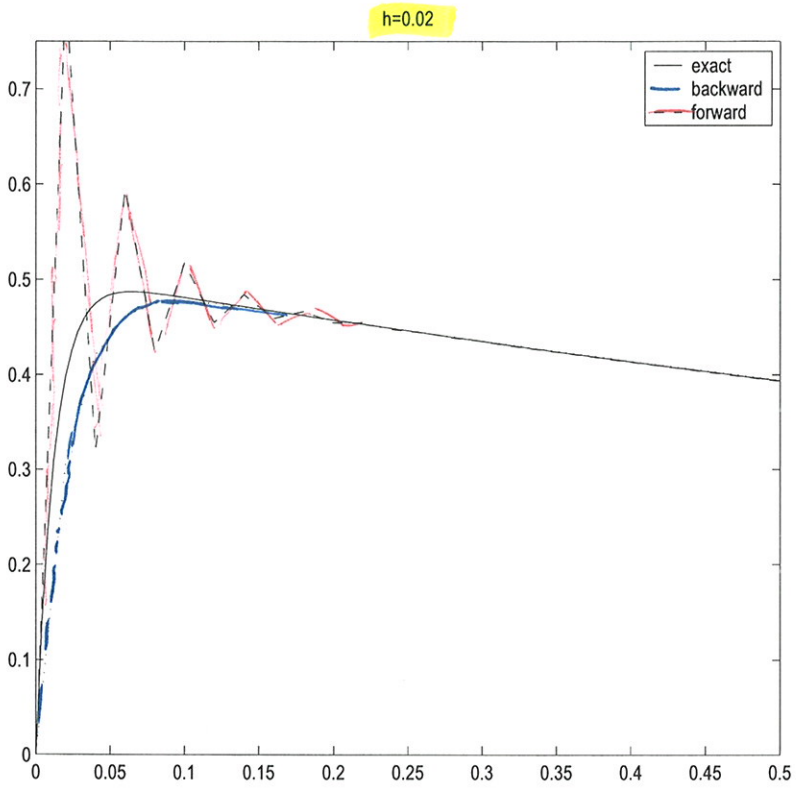
h=0.005



h=0.01



2.5b


h=0.02
exact
backward
forward
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

Apply backward Euler to the test equation:

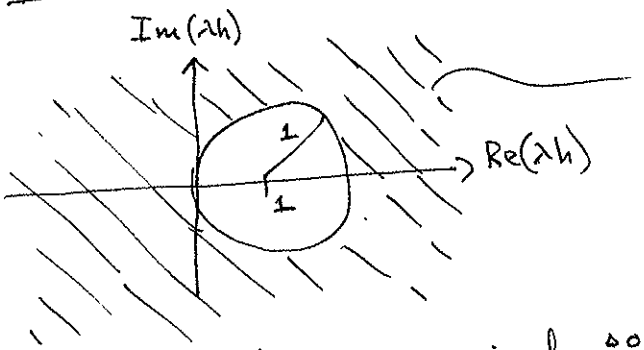
$$y' = \lambda y \Rightarrow y_n = y_{n-1} + h f(t_n, y_n)$$

$$y_n = y_{n-1} + h\lambda y_n$$

$$(1 - h\lambda) y_n = y_{n-1}$$

$$y_n = \frac{y_{n-1}}{1 - h\lambda}$$

absolute stability: $\left|\frac{1}{1-h\lambda}\right| \leq 1 \rightarrow 1 \leq |1 - h\lambda|$



region of absolute stability for backwards Euler

Note that a numerical solution of $y' = \lambda y$ using backwards Euler decays whenever $Re\lambda < 0$ independent of h. The region of absolute stability covers the whole left half of the λh-plane. Such methods are called A-stable.

Example: small-mass, spring problem

$$\varepsilon v'' + 2v' + v = 0, \quad 0 \leq t \leq 2, \quad v(0) = 0, \quad \varepsilon v'(0) = 1$$

solve using backward Euler

$$\underline{Y}' = \underline{\underline{A}}\, Y, \quad A = \begin{pmatrix} 0 & 1 \\ -1/\varepsilon & -2/\varepsilon \end{pmatrix}, \quad y(0) = \begin{pmatrix} 0 \\ 1/\varepsilon \end{pmatrix}, \quad Y_1 = v, \quad Y_2 = v'$$

$$Y_n = Y_{n-1} + hAY_n \rightarrow (I - hA) Y_n = Y_{n-1}$$

$$\rightarrow Y_n = B Y_{n-1}, \text{ where } B = (I - hA)^{-1} = \frac{1}{1 + \frac{h}{\varepsilon}(2+h)} \begin{bmatrix} 1 + 2h/\varepsilon & h \\ -h/\varepsilon & 1 \end{bmatrix}$$

nonlinear case

$y' = f(t,y)$, $0 \le t \le b$

Now we measure stiffness locally about a given trajectay $\hat{y}(t)$. If we linearize about $\hat{y}(t)$, set $y(t) = \hat{y}(t) + \tilde{y}(t)$, where $\tilde{y}(t)$ is small.

$$\Rightarrow \quad \hat{y}' + \tilde{y}' = f(t, \hat{y}+\tilde{y})$$
$$= f(t,\hat{y}) + \frac{\partial f}{\partial y}(t,\hat{y})\,\tilde{y} + \dots$$

Suppose $\hat{y}' = f(t,\hat{y})$, then $\tilde{y}$ solves $\tilde{y}' = \frac{\partial f}{\partial y}(t,\hat{y})\,\tilde{y}$ to a first approximation. Set $\underline{\underline{A}}_0 = \frac{\partial f}{\partial y}(t_0, \hat{y}(t_0))$

Stiffness depends on the eigenvalues of $\underline{\underline{A}}_0$

stiffness $\Rightarrow$ $b(-\mathrm{Re}\lambda_j) >> 1$.

If the problem were stiff then you might use backward Euler $\Rightarrow$ $y_n = y_{n-1} + h f(t_n, y_n)$

Need to solve for y_n given y_{n-1}, h, using some Functional iteration: choose a function $g(y)$ and iterate: $y^\nu = g(y^{\nu-1})$, $\nu = 1, 2, \dots$ such that $y^\nu \to y_n$ as $\nu \to \infty$.

Problem becomes the choice of $g(y)$ and the starting value y^0.

One choice (poor) is $g(y) = y_{n-1} + h\, f(t_n, y)$

clearly y_n is a fixed-point of the iteration (ie, $y_n = g(y_n)$).

The question is does this Functional iteration converge?

Compute the Jacobian: $g_y = h\, f_y(t_n, y)$

Require that $\rho(g_y) < 1$, where ρ - spectral radius, the max of the magnitudes of the eigenvalues.

$$\rho(g_y) = \max_j |h\lambda_j| < 1$$

we can always Find h small enough so that the iteration converges. However if the problem is stiff then h must be be chosen to be "too" small for this to happen. This is not a good iteration scheme.

Newton's method:

$$\underline{g}(\underline{Y}) = \underline{Y} - \underline{\underline{A}}(\underline{Y})\left(\underline{Y}_{n-1} - \underline{Y} + h\,\underline{F}(t_n, \underline{Y})\right)$$

where $\underline{\underline{A}}(\underline{Y}) = \left(-\underline{\underline{I}} + h\,\underline{\underline{F}}_{\underline{Y}}(t_n, \underline{Y})\right)^{-1}$

Jacobian: $\underline{g}_{\underline{Y}}(\underline{Y}) = \underline{\underline{I}} - \underbrace{\underline{\underline{A}}(\underline{Y})\left(-\underline{\underline{I}} + h\,\underline{\underline{F}}_{\underline{Y}}(t_n, \underline{Y})\right)}_{\underline{\underline{I}}} - \underline{\underline{A}}_{\underline{Y}}(\underline{Y})\left(\underline{Y}_{n-1} - \underline{Y} + h\,\underline{F}(t_n, \underline{Y})\right)$

$$\rightarrow \quad \underline{g}_{\underline{Y}}(\underline{Y}) = -\underline{\underline{A}}_{\underline{Y}}(\underline{Y})\left(\underline{Y}_{n-1} - \underline{Y} + h\,\underline{F}(t_n, \underline{Y})\right)$$

$\Rightarrow \underline{g}_{\underline{Y}}(\underline{Y}_n) = 0 \Rightarrow$ quadratic convergence (at least), independent of choice of h.

The asymptotic rate of convergence is quadratic and independent of h but the choice of Y^0 for convergence does depend on h. Evidently, $Y^0 = Y_{n-1}$ is generally good enough.

Algorithm

1. select $Y^0 = Y_{n-1}$
2. For $\nu = 1, 2, \ldots$ do 3 → 7
3. $\underline{X}^\nu = \underline{Y}^\nu - \underline{Y}_{n-1} - h\,\underline{F}(t_n, \underline{Y}^\nu)$
4. $\underline{\underline{J}}^\nu = \underline{\underline{I}} - h\,\underline{\underline{F}}_{\underline{Y}}(t_n, \underline{Y}^\nu)$ (calculate Jacobian)
5. solve $\underline{\underline{J}}^\nu \underline{\Delta Y} = \underline{X}^\nu$ (solve linear system)
6. $\underline{Y}^{\nu+1} = \underline{Y}^\nu - \underline{\Delta Y}$ (update step)
7. stop if $\|\underline{\Delta Y}\| \leq$ tol

Implicit Methods and Newtons Method

Remember, solving the initial value problem $\underline{y}' = \underline{f}(t, \underline{y})$, $\underline{y}(0) = \underline{c}$, $t \geq 0$

Numerical solution using backward Euler

$$\underline{y}_n = \underline{y}_{n-1} + h_n \underline{f}(t_n, \underline{y}_n)$$

This turns into a root-finding problem for $\underline{y}_n$.

Define: $\underline{F}(\underline{y}) \equiv \underline{y} - \underline{y}_{n-1} - h_n \underline{f}(t_n, \underline{y})$

Clearly $\underline{F}(\underline{y}_n) = 0$

Apply Newton's method to $\underline{F}(\underline{y})$:

1) $\underline{y}^0 = \underline{y}_{n-1}$, initial guess

2) for $\nu = 0, 1, \cdots,$

3) compute $\underline{F}^\nu = \underline{F}(\underline{y}^\nu)$

4) compute $\underline{\underline{J}}^\nu = \underline{\underline{F}}_{\underline{y}}(\underline{y}^\nu)$ (Jacobian)

Notice, $\underline{\underline{J}}(\underline{y}) = \underline{\underline{I}} - h_n \underline{\underline{f}}_{\underline{y}}(t_n, \underline{y})$

5) solve $\underline{\underline{J}}^\nu \Delta \underline{y} = \underline{F}^\nu$ ($m \times m$ linear system)

6) update, $\underline{y}^{\nu+1} = \underline{y}^\nu - \Delta \underline{y}$

7) check convergence, $\| \Delta \underline{y} \| \leq tol$

Remarks:

1) The expensive step is (5), solving the linear system. Typically use Gaussian elimination, partial pivoting. $\Rightarrow$ cost $\frac{2}{3} m^3$ (LU factorization of $\underline{\underline{J}}^\nu$). May want to freeze the Jacobian and its LU-factorization after one or two steps.

2) Don't need to set tolerance, tol $= \| \ \| \varepsilon_{machine}$, instead only need tol $\ll$ local error $= \| \underline{\ell}_n \|$

3) Can avoid the analytic calculation of $\underline{F}_{\underline{Y}}$ by using finite differences.

set $\underline{a}_j = \frac{1}{n}\left(\underline{F}(t_n, \underline{Y}^\nu + n\underline{e}_j) - \underline{F}(t_n, \underline{Y}^\nu)\right)$

($\underline{e}_j$: j^{th} column of $m \times m$ identity)

$n \sim \sqrt{\varepsilon_{machine}} \approx 10^{-8}$

Finite diff approximation of j^{th} column of $\underline{F}_{\underline{Y}}(\underline{Y}^\nu)$

Trapezoidal Method

- one parameter family of methods

$$Y_n = Y_{n-1} + h_n\left[\Theta F(t_n, Y_n) + (1-\Theta) F(t_{n-1}, Y_{n-1})\right]$$

where Θ is a parameter, $\Theta \in [0,1]$

if $\Theta = 0 \Rightarrow$ Euler

$\Theta = 1 \Rightarrow$ backward Euler

$\Theta > 0 \Rightarrow$ implicit method

Consider the order of accuracy:

$$N_n Y_n = \frac{Y_n - Y_{n-1}}{h_n} - \left[\Theta F(t_n, Y_n) + (1-\Theta) F(t_{n-1}, Y_{n-1})\right]$$

local truncation error is: $d_n = N_n Y(t_n) = O(h_n)$ if $\Theta \neq \frac{1}{2}$

$= O(h_n^2)$ if $\Theta = \frac{1}{2}$

If $\Theta = 1/2$

$$Y_n = Y_{n-1} + \frac{h_n}{2}\left[F(t_n, Y_n) + F(t_{n-1}, Y_{n-1})\right]$$

$\rightarrow$ Trapezoidal Method

You can consider the integral of the ODE:

$$\int_{t_{n-1}}^{t_n} y'(t)\,dt = \int_{t_{n-1}}^{t_n} F(t, y(t))\,dt$$

$$y(t_n) - y(t_{n-1}) = \int_{t_{n-1}}^{t_n} f(t, y(t))\,dt \rightarrow \text{approximate using trapezoidal rule}$$

$$\approx \frac{h_n}{2}\left[f(t_n, y(t_n)) + f(t_{n-1}, y(t_{n-1}))\right]$$

This leads to the Trapezoidal method

Consider the stability properties of the Trapezoidal method

$$y' = \lambda y$$

$$y_n = y_{n-1} + \frac{\lambda h}{2}(y_{n-1} + y_n) \rightarrow \left(1 - \frac{\lambda h}{2}\right) y_n = \left(1 + \frac{\lambda h}{2}\right) y_{n-1}$$

$$\rightarrow y_n = \left(\frac{1 + \frac{\lambda h}{2}}{1 - \frac{\lambda h}{2}}\right) y_{n-1}$$

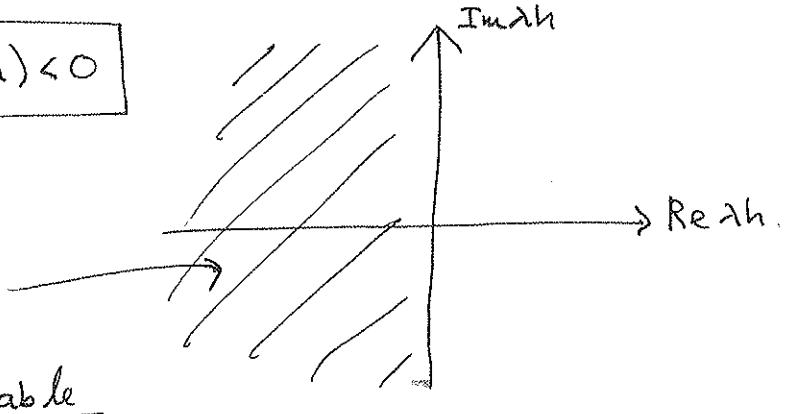
"growth" factor

Find λh complex st $\left|\frac{1 + \frac{\lambda h}{2}}{1 - \frac{\lambda h}{2}}\right| \leq 1$

set $\lambda h = z = a + ib \rightarrow \left|1 + \frac{a+ib}{2}\right|^2 \leq \left|1 - \frac{a+ib}{2}\right|^2$

$$\rightarrow \left(1 + \frac{a}{2}\right)^2 + \frac{b^2}{4} \leq \left(1 - \frac{a}{2}\right)^2 + \frac{b^2}{4} \rightarrow 1 + a + \frac{a^2}{4} \leq 1 - a + \frac{a^2}{4}$$

$$\rightarrow \boxed{a \leq 0} \Rightarrow \boxed{\mathrm{Re}(\lambda h) < 0}$$



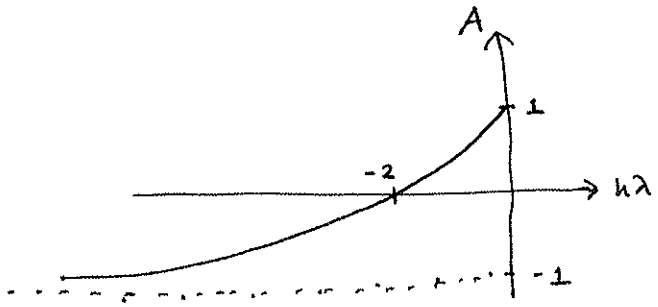
region of absolute stability for trap method

hence, A-stable

so the trapezoidal method is A-stable

=> if the exact solution of the test equation decays, then the numerical ~~method~~ solution decays independent of h, and if the exact solution grows the numerical soln grows (independent of h)

Consider real λ with $\lambda < 0$, let $A(h\lambda) \equiv \dfrac{1 + \frac{\lambda h}{2}}{1 - \frac{\lambda h}{2}}$



In the very stiff limit, $-\mathrm{Re}(\lambda h) >> 1$, then

$$A(h\lambda) = \frac{1 + {}^{2}/_{\lambda h}}{-1 + {}^{2}/_{\lambda h}} \sim -1 + \frac{4}{(-\lambda h)} + \cdots = -1 + \delta$$

so, $y_n \sim -y_{n-1}$, decay, but very slowly. The solution has a ± 1 oscillation.

Remember, backward Euler, $A = \frac{1}{1-\lambda h} \to 0$ as $-\mathrm{Re}(\lambda h) \to \infty$, which is referred to as "stiff" decay.

Example: $\varepsilon v'' + 2v' + v = 0 \quad , t \geq 0$

$v(0) = 0 \quad , \quad \varepsilon v'(0) = 1 \quad , \quad 0 < \varepsilon << 1$

$$\Rightarrow \underline{Y}' = \underline{\underline{A}}\,\underline{Y} \quad , \quad \underline{\underline{A}} = \begin{bmatrix} 0 & 1 \\ -1/\varepsilon & -2/\varepsilon \end{bmatrix} \quad , \quad \underline{Y} = \begin{pmatrix} v(t) \\ v'(t) \end{pmatrix}$$

Apply Trapezoidal method:

$$Y_n = Y_{n-1} + \frac{h}{2}\left(A Y_{n-1} + A Y_n\right) \quad \rightarrow \quad \left(I - \tfrac{h}{2}A\right) Y_n = \left(I + \tfrac{h}{2}A\right) Y_{n-1}$$

$$Y_n = \underbrace{\left(I - \tfrac{h}{2}A\right)^{-1}\left(I + \tfrac{h}{2}A\right)}_{B} Y_{n-1} \quad \Rightarrow \quad Y_n = B\, Y_{n-1}$$

$\rightarrow$ numerical solution $Y_n = B^n Y_0 \quad , \quad Y_0 = \begin{pmatrix} 0 \\ Y_\varepsilon \end{pmatrix}$

and $$B = \begin{bmatrix} 1 + \frac{h}{\varepsilon} - \frac{h^2}{4\varepsilon} & h \\ -\frac{h}{\varepsilon} & 1 - \frac{h}{\varepsilon} - \frac{h^2}{4\varepsilon} \end{bmatrix} \frac{1}{\left(1 + \frac{h}{\varepsilon} + \frac{h^2}{4\varepsilon}\right)}$$

Single-Step Methods

General form of an explicit single-step method is

$$\underline{Y}_n = \underline{Y}_{n-1} + h_n \underline{\Phi}\left(t_{n-1}, \underline{Y}_{n-1}, h_n\right)$$

The function $\phi(t, y, h)$ is called the stepping function.

For example, $\phi(t, y, h) = f(t, y) \Rightarrow$ Euler

Implicit single-step methods have the form

$$\underline{Y}_n = \underline{Y}_{n-1} + h_n \hat{\underline{\Phi}}\left(t_{n-1}, \underline{Y}_{n-1}, \underline{Y}_n, h_n\right)$$

For example $\hat{\phi}(t, y, w, h) = f(t+h, w) \Rightarrow$ backwards Euler

$$\hat{\phi}(t, y, w, h) = \frac{1}{2}\left(f(t, y) + f(t+h, w)\right) \Rightarrow \text{trapezoidal method}$$

Main problem is to construct $\phi(t,y,h)$ to obtain higher-order accuracy. Simplest approach is via Taylor expansions.

$$y(t_n) = y(t_{n-1} + h)$$
$$= y(t_{n-1}) + h y'(t_{n-1}) + \frac{h^2}{2} y''(t_{n-1}) + \cdots + \frac{h^p}{p!} y^{(p)}(t_{n-1}) + \cdots$$

we know that $y(t)$ solves the ODE $y'(t) = f(t, y(t))$

$$\Rightarrow \quad y'(t_{n-1}) = f(t_{n-1}, y(t_{n-1}))$$

$$y''(t_{n-1}) = \frac{d}{dt} f(t, y(t))\Big|_{t=t_{n-1}} = \left[\frac{\partial f}{\partial t} + \frac{\partial f}{\partial y} f\right]_{t=t_{n-1}}$$

$$y'''(t_{n-1}) = \left[f_{tt} + f_{ty} f + f_{yt} f + f_{yy} f^2 + f_y f_t + f_y^2 f\right]_{t=t_{n-1}}$$

$\vdots$

p^{th} order Taylor method

$$y_n = y_{n-1} + h f\Big|_{(t_{n-1}, y_{n-1})} + \frac{h^2}{2} \frac{df}{dt}\Big|_{(t_{n-1}, y_{n-1})} + \cdots + \frac{h^p}{p!} \frac{d^{p-1} f}{dt^{p-1}}\Big|_{(t_{n-1}, y_{n-1})}$$

Here, $\phi = f\Big|_{(t_{n-1}, y_{n-1})} + \frac{h}{2} \frac{df}{dt}\Big|_{(t_{n-1}, y_{n-1})} + \cdots + \frac{h^{p-1}}{p!} \frac{d^{p-1} f}{dt^{p-1}}\Big|_{(t_{n-1}, y_{n-1})}$

Eg, 2nd order Taylor method

$$y_n = y_{n-1} + h\left[f(t_{n-1}, y_{n-1}) + \frac{h}{2}\left(f_t(t_{n-1}, y_{n-1}) + f_y(t_{n-1}, y_{n-1}) f(t_{n-1}, y_{n-1})\right)\right]$$

- this is kind of like Euler's method but with a correction
- not hard to show this is 2nd order accurate
- ~~in principle you can get arbitrarily high order of accuracy~~

$$d_n = \frac{y_n - y_{n-1}}{h} - \left\{f(t_{n-1}, y_{n-1}) + \frac{h}{2}\left(f_t(t_{n-1}, y_{n-1}) + f_y(t_{n-1}, y_{n-1}) f(t_{n-1}, y_{n-1})\right)\right\}$$

$$d_n = y'(t_{n-1}) + \frac{h}{2} y''(t_{n-1}) + \frac{h^2}{6} y'''(t_{n-1}) - \left\{y'(t_{n-1}) + \frac{h}{2} y''(t_{n-1})\right\}$$

$$= \frac{h^2}{6} y'''(t_{n-1}) + O(h^3) \quad \Rightarrow \quad \boxed{d_n = O(h^2) \text{ as } h \to 0}$$

Remarks on Taylor method

- In principle you can get arbitrarily high order of accuracy
- There is an explosion of terms involved in the analytic calculation of the derivatives. This can be avoided by imbedding functions evaluations $\Rightarrow$ Runge-Kutta methods.

2/14/03

Runge Kutta Methods

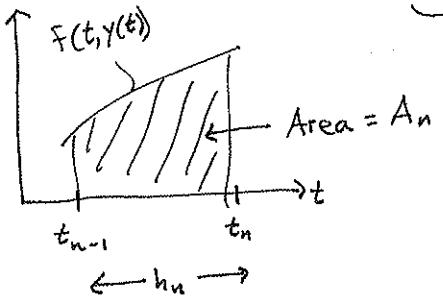
ODE: $y' = f(t,y)$

Main idea: Suppose we know y at t_{n-1} (approximately perhaps) we want to find an approximation for y at $t_n = t_{n-1} + h$

Integrate the ODE:

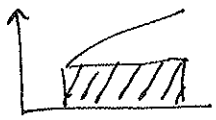
$$y(t_n) = y(t_{n-1}) + \int_{t_{n-1}}^{t_n} f(t, y(t))\, dt$$

area



Approximate A_n

eg: $A_n \simeq h f(t_{n-1}, y(t_{n-1}))$ $\Rightarrow$ Euler



(EXPLICIT)

$A_n \simeq h f(t_n, y(t_n))$ $\Rightarrow$ Backward Euler

(IMPLICIT)

$A_n \simeq \frac{h}{2}\left(f(t_{n-1}, y(t_{n-1})) + f(t_n, y(t_n))\right)$

$\Rightarrow$ Trapezoidal method

(IMPLICIT)

EXPLICIT MIDPOINT METHOD

Consider the implicit midpoint method:

$$y_n = y_{n-1} + h f\left(t_{n-1/2}, \frac{y_{n-1} + y_n}{2}\right)$$

approximate $y(t_{n-1/2})$ by forward Euler then substitute:

$$\left.\begin{aligned} \hat{y}_{n-1/2} &= y_{n-1} + \tfrac{h}{2} f(t_{n-1}, y_{n-1}) \\ y_n &= y_{n-1} + h f\left(t_{n-1/2}, \hat{y}_{n-1/2}\right) \end{aligned}\right\} \text{ 2-stage RK method}$$

local truncation error

$$d_n = \frac{y(t_n) - y(t_{n-1})}{h} - f\left(t_{n-1/2}, y(t_{n-1}) + \tfrac{h}{2} f(t_{n-1}, y(t_{n-1}))\right)$$

Expand in a Taylor series: (centered about $t_{n-1}, y(t_{n-1})$)

1st term:

$$\frac{y(t_n) - y(t_{n-1})}{h} = \frac{y(t_{n-1}) + h y'(t_{n-1}) + \frac{h^2}{2} y''(t_{n-1}) + \frac{h^3}{6} y'''(t_{n-1}) + \cdots - y(t_{n-1})}{h}$$

$$= y'(t_{n-1}) + \frac{h}{2} y''(t_{n-1}) + \frac{h^2}{6} y'''(t_{n-1}) + \cdots$$

2nd term

$$f\left(t_{n-1/2}, y(t_{n-1}) + \tfrac{h}{2} f(t_{n-1}, y(t_{n-1}))\right) = f\left(t_{n-1} + \tfrac{h}{2}, y(t_{n-1}) + \Delta y\right)$$

$$= f + \frac{h}{2} f_t + \Delta y\, f_y + \frac{1}{2!}\left(\left(\frac{h}{2}\right)^2 f_{tt} + 2 f_{ty}\left(\frac{h}{2}\right)(\Delta y) + f_{yy}(\Delta y)^2\right) + \cdots$$

$$= f + \frac{h}{2} f_t + \frac{h}{2} f \cdot f_y + \frac{1}{2!}\left(\frac{h^2}{4} f_{tt} + 2 f_{ty} \frac{h}{2} \cdot \frac{h}{2} f + f_{yy}\left(\frac{h}{2} y\right)^2\right) + \cdots$$

$$= f + \frac{h}{2}\left(f_t + f f_y\right) + \frac{h^2}{8}\left(f_{tt} + 2 f_{ty} f + f_{yy} f^2\right) + \cdots$$

$$\Rightarrow d_n = y' + \frac{h}{2} y'' + \frac{h^2}{6} y''' - \left(f + \frac{h}{2}(f_t + f f_y) + \frac{h^2}{8}\left(f_{tt} + 2 f_{ty} f + f_{yy} f^2\right)\right) + O(h^3)$$

Notice: $y' = f(t, y)$ so $y'' = \frac{\partial f}{\partial t}\frac{\partial t}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} = f_t + f_y f$

$$\Rightarrow \frac{h}{2} y'' - \frac{h}{2}(f_t + f f_y) = 0$$

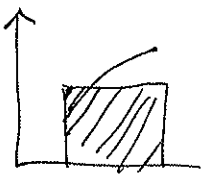
$$\Rightarrow d_n = \frac{h^2}{6} y''' - \frac{h^2}{8}\left(f_{tt} + 2 f_{ty} f + f_{yy} f^2\right) + O(h^3)$$

$$\Rightarrow \lim_{h \to 0} \frac{|d_n|}{h^2} = C \Rightarrow$$

the method is consistent of order 2

$$A_n \simeq h\, f\left(t_{n-1/2}, \frac{y(t_{n-1}) + y(t_n)}{2}\right)$$

$\Rightarrow$ Midpoint rule



(IMPLICIT)

Convert the implicit methods to explicit methods using a suitable Euler step:

eg: $\hat{y} = y_{n-1} + \frac{h}{2} f(t_{n-1}, y_{n-1}) \simeq y(t_{n-1/2})$ $\Rightarrow$ Euler half step

use this in Midpoint rule

$$y_n = y_{n-1} + h\, f(t_{n-1/2}, \hat{y})$$

} "two stage" Runge Kutta (RK) Method - explicit, midpoint method

---

Consider:

$$\tilde{y} = y_{n-1} + h f(t_{n-1}, y_{n-1}) \simeq y(t_n)$$

substitute into trapezoidal method

$$y_n = y_{n-1} + \frac{h}{2}\left(f(t_{n-1}, y_{n-1}) + f(t_n, \tilde{y})\right)$$

Classical 4th order Runge Kutta (RK4)

$$Y_1 = y_{n-1}$$

$$Y_2 = y_{n-1} + \frac{h}{2} f(t_{n-1}, Y_1)$$

$$Y_3 = y_{n-1} + \frac{h}{2} f(t_{n-1/2}, Y_2)$$

$$Y_4 = y_{n-1} + h f(t_{n-1/2}, Y_3)$$

$$y_n = y_{n-1} + \frac{h}{6}\left[f(t_{n-1}, Y_1) + 2f(t_{n-1/2}, Y_2) + 2f(t_{n-1/2}, Y_3) + f(t_n, Y_4)\right]$$

Example: $y' = e^{-t} - y$, $y(0) = 0$

Calculate the solution for $0 \le t \le 4$ using various values for h and various RK methods (all explicit) and compare to the exact solution , $y(t) = te^{-t}$.

error : p^{th} order method

$$e(h) \simeq c h^p \quad \rightarrow \text{global error}$$

$$e\left(\tfrac{h}{2}\right) \simeq c\left(\tfrac{h}{2}\right)^p$$

$$\rightarrow \frac{e(h/2)}{e(h)} = 2^{-p} \rightarrow \log\left(\frac{e(h/2)}{e(h)}\right) = -p \log 2$$

$$\rightarrow -p \approx \frac{\log\left(e(h/2)/e(h)\right)}{\log 2} \rightarrow p \approx \frac{\log\left(e(h)/e(h/2)\right)}{\log 2}$$

General forms

General s-stage RK methods

$$Y_i = y_{n-1} + h \sum_{j=1}^{s} a_{ij} f(t_{n-1} + c_j h , Y_j) \quad , \quad i = 1, \dots, s$$

Then

$$y_n = y_{n-1} + h \underbrace{\sum_{i=1}^{s} b_i f(t_{n-1} + c_j h , Y_i)}_{\text{quadrature rule}}$$

Note that Y_i is an approximation for y at $t_{n-1} + c_j h$, $0 \le c_i \le 1$.

the constants $\{a_{ij}\}$, $\{c_j\}$, $\{b_i\}$ are to choose.

Consistency will place some constraints on the constants

Let us consider consistancy:

Expand for h small

$$y_{n-1} + h\sum_{j=1}^{s} a_{ij} f(t_{n-1} + c_i h, Y_j)$$

$$= y_{n-1} + h\sum_{j=1}^{s} a_{ij} f(t_{n-1}, y_{n-1}) + O(h^2)$$

$$= y_{n-1} + h\left(\sum_{j=1}^{s} a_{ij}\right) f(t_{n-1}, y_{n-1}) + O(h^2) \approx y(t_{n-1} + c_i h)$$

↳ like Euler's method, so we require $\boxed{\sum_{j=1}^{s} a_{ij} = c_i \quad , i = 1, \ldots s}$

Also $y_{n-1} + h\sum_{i=1}^{s} b_i f(t_{n-1} + c_i h, Y_i)$

$$= y_{n-1} + h\left(\sum_{i=1}^{s} b_i\right) f(t_{n-1}, y_{n-1}) + O(h^2) \approx y(t_n)$$

like Euler's method, so require $\boxed{\sum_{i=1}^{s} b_i = 1}$

Convenient way to express the coefficients

$$\begin{array}{c|ccc} c_1 & a_{11} & \cdots & a_{1s} \\ c_2 & a_{21} & \cdots & a_{2s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & & a_{ss} \\ \hline & b_1 & \cdots & b_s \end{array}$$

← sums to c_i ; ← sums to 1

} most general form which includes implicit methods

If you restrict to explicit methods, the matrix becomes triangular

$a_{ij} = 0$ for $j > i$

$$\begin{array}{c|ccccc} 0 & 0 & & & & \\ c_2 & a_{21} & 0 & & & \\ c_3 & a_{31} & a_{32} & 0 & & \\ \vdots & \vdots & & \ddots & \ddots & \\ c_s & a_{s1} & a_{s2} & \cdots & a_{s,s-1} & 0 \\ \hline & b_1 & b_2 & \cdots & & b_s \end{array}$$

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}$$

Example

- Euler: $\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$

backward Euler: $\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}$

- Midpoint (explicit) $\begin{array}{c|cc} 0 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ \hline & 0 & 1 \end{array}$

Midpoint (implicit) $\begin{array}{c|c} 1/2 & 1/2 \\ \hline & 1 \end{array}$

$s = 1$
$p = 2$

- Trapezoidal Rule $\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array}$

- RK4 $\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline & 1/6 & 2/6 & 2/6 & 1/6 \end{array}$

Order of Accuracy

For an s-stage RK method we have many parameters to choose

Impose: $\sum_{i=1}^{s} b_j = 1$, $\sum_{j=1}^{s} a_{ij} = c_i$, $i = 1, \dots, s$ For consistency.

Beyond that, choose coeffs to:

(1) explicit vs implicit
(2) order of accuracy (want high order methods)
(3) ~~order of accuracy~~ Favorable stability properties
(4) error control

Focus on the choice for higher order methods. (accuracy)
For single-step methods, obtaining a higher order method becomes choosing parameters st

truncation error $\sim O(h^p)$
For p^{th} order method

All RK methods are 0-stable if f satisfies a Lipschitz condition.

$\Rightarrow |e_n| \le k \max_j |d_j| = O(h^p)$

↳ truncation error

So, $d_n = \frac{y(t_n) - y(t_{n-1})}{h} - h\sum_{i=1}^{s} b_i f(t_{n-1} + c_i h, Y_i)$

$$Y_i = y(t_{n-1}) + h\sum_{j=1}^{s} a_{ij} f(t_{n-1} + c_j h, Y_j)$$

$$\frac{y(t_n) - y(t_{n-1})}{h} = \frac{y(t_{n-1}) + hy'(t_{n-1}) + \frac{h^2}{2}y''(t_{n-1}) + \cdots - y(t_{n-1})}{h}$$

$$\Rightarrow d_n = y'(t_{n-1}) + \frac{h}{2}y''(t_{n-1}) + \cdots - h\sum_{i=1}^{s} b_i f(t_{n-1} + c_i h, Y_i) = O(h^p)$$

For this to $O(h^p)$ the sum from $i = 1 \ldots s$ has to cancel the first p terms of the expansion

$$\Rightarrow \underbrace{\sum_{i=1}^{s} b_i f(t_{n-1} + c_i h, Y_i)}_{} = y'(t_{n-1}) + \frac{h}{2}y''(t_{n-1}) + \cdots + \frac{h^{p-1}}{p!}y^{(p)}(t_{n-1}) + O(h^p)$$

Expand in Taylor series - this is an "enormous task"
- Taylor series within Taylor series $\Rightarrow$ "an explosion of terms"

- Instead, develop certain necessary conditions for p-order accuracy by applying the previous formula for "smart" choices of $f(t,y)$.

Choose: $f(t,y) = y + t^{l-1}$ where l is a positive integer
(linear in y, polynomial in t)

set $t_{n-1} = 0$, $y_{n-1} = 0$

we're only interested in one step:

$$Y_i = h\sum_{j=1}^{s} a_{ij} f(t_{n-1} + c_j h, Y_i) = h\sum_{j=1}^{s} a_{ij}\left(Y_j + (c_j h)^{l-1}\right)$$

set $\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_s \end{bmatrix}$, $\underline{\underline{A}} = \begin{bmatrix} a_{ij} \end{bmatrix}$, $\underline{z} = \begin{bmatrix} c_1^{l-1} \\ c_2^{l-1} \\ \vdots \\ c_s^{l-1} \end{bmatrix}$

$$\underline{\underline{Y}} = h\,\underline{\underline{A}}\left(\underline{\underline{Y}} + h^{l-1}\underline{\underline{z}}\right) = h\,\underline{\underline{A}}\,\underline{\underline{Y}} + h^{l}\underline{\underline{A}}\,\underline{\underline{z}}$$

$$\Rightarrow \left(I - h\,\underline{\underline{A}}\right)\underline{\underline{Y}} = h^{l}\underline{\underline{A}}\underline{\underline{z}} \quad \Rightarrow \quad \underline{\underline{Y}} = h^{l}\left(I - h\,\underline{\underline{A}}\right)^{-1}\underline{\underline{A}}\,\underline{\underline{z}}$$

Then: $Y_n = Y_{n-1} + h\sum_{i=1}^{s} b_i\, f\left(t_{n-1} + c_i h,\ \underline{\underline{Y}}_i\right)$

(Y_{n-1}: 0; t_{n-1}: 0; $\underline{\underline{Y}}_i + (c_i R)^{l-1}$)

let $\underline{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_s \end{pmatrix}$

$$\rightarrow Y_n = h\,\underline{b}^{T}\underbrace{\left(\underline{\underline{Y}} + h^{l-1}\underline{\underline{z}}\right)}_{h^{-1}\underline{\underline{A}}^{-1}\underline{\underline{Y}}}$$

($h^{-1}\underline{\underline{A}}^{-1}\underline{\underline{Y}}$ (From top line of page))

NOTE, Y_n is a scalar

$$\rightarrow Y_n = b^{T}A^{-1}\underline{\underline{Y}}$$

$$\boxed{Y_n = h^{l}b^{T}A^{-1}(I - hA)^{-1}A\,z} \quad (1)$$

this is an approximation for $y(t_n)$ where $y' = t^{l-1} + y$ with $y(t_{n-1}) = 0$, $t_{n-1} = 0$

The exact solution: $y(t) = \int_0^t e^{t-\tau}\,\tau^{l-1}\,d\tau$

So $\boxed{y(t_n) = y(h) = \int_0^h e^{h-\tau}\,\tau^{l-1}\,d\tau}$ (2)

General s-stage Runge Kutta method

$$Y_i = Y_{n-1} + h\sum_{j=1}^{s} a_{ij} F(t_j, Y_j) \quad , i = 1,\dots,s$$

then $$Y_n = Y_{n-1} + h\sum_{i=1}^{s} b_i F(t_i, Y_i)$$

nodes: $t_i = t_{n-1} + c_i h$, $0 \le c_i \le 1$

$Y_i \simeq y(t_i)$

constraints that enforce consistency: $\sum_{i=1}^{s} b_i = 1$, $\sum_{j=1}^{s} a_{ij} = c_i$, $i = 1,\dots,s$

Matrix form:

c_i	a_{ij}
	b_j

explicit then $a_{ij} = 0$ if $j \ge i$

constraints that enforce order of accuracy:
- Find constraints on the a's, b's, c's st the method is p^{th} order accurate.

choose a specific ODE ⇒ necessary order conditions

let $y' = y + t^{l-1}$, $l = 1, 2, \dots$

$y(t_{n-1}) = Y_{n-1}$

without loss you can take $t_{n-1} = Y_{n-1} < 0$

apply an s-stage RK to the IVP

(1) ⇒ $$\boxed{Y_n = h^l \underline{b}^T \underline{\underline{A}}^{-1} \left(\underline{\underline{I}} - h\underline{\underline{A}}\right)^{-1} \underline{\underline{A}}\, \underline{z}}$$, where $\underline{z} = \begin{bmatrix} c_1^{l-1} \\ \vdots \\ c_s^{l-1} \end{bmatrix}$

notice, Y_n is a scalar

the exact solution is $y(t) = \int_0^t e^{t-\tau} \tau^{l-1} d\tau$

(2) ⇒ $$\boxed{y(t_n) = y(h) = \int_0^h e^{h-\tau} \tau^{l-1} d\tau}$$

Expand (1) and (2) in Taylor series about $h = 0$, then match terms up to and including h^p.

⇒ local error = $O(h^{p+1})$

⇒ truncation error = $O(h^p)$

⇒ p^{th} order method

Note, $(\underline{\underline{I}} - h\underline{\underline{A}})^{-1}$ can be expanded as geometric series

$$(I - h\underline{\underline{A}})^{-1} = \underline{\underline{I}} + h\underline{\underline{A}} + h^2\underline{\underline{A}}^2 + h^3\underline{\underline{A}}^3 + \cdots$$

$$A^{-1}(I - hA^{-1})A = \underline{\underline{I}} + h\underline{\underline{A}} + h^2\underline{\underline{A}}^2 + h^3\underline{\underline{A}}^3 + \cdots$$

$$\Rightarrow y_n = h^\ell \underline{b}^T (\underline{\underline{I}} + h\underline{\underline{A}} + h^2\underline{\underline{A}}^2 + \cdots)\underline{z} \qquad (1e)$$

$$y(h) = y(0) + hy'(0) + \frac{h^2}{2}y''(0) + \cdots$$

$$= (\ell-1)!\left[\frac{h^\ell}{\ell!} + \frac{h^{\ell+1}}{(\ell+1)!} + \cdots\right] \qquad (2e)$$

Math terms in (1e) and (2e)

coeffs of h^ℓ: $\dfrac{(\ell-1)!}{\ell!} = \underline{b}^T\underline{z} \Rightarrow \boxed{\underline{b}^T\underline{z} = \frac{1}{\ell}}$ $\quad \ell = 1, 2, \ldots, p$

$h^{\ell+1}$: $\dfrac{(\ell-1)!}{(\ell+1)!} = \underline{b}^T\underline{\underline{A}}\,\underline{z} \Rightarrow \boxed{\underline{b}^T\underline{\underline{A}}\,\underline{z} = \frac{1}{\ell(\ell+1)}}$, $\ell = 1, 2, \ldots, p-1$

$\vdots$

$h^{\ell+k}$: $\boxed{\underline{b}^T\underline{\underline{A}}^k\underline{z} = \frac{1}{\ell(\ell+1)\cdots(\ell+k)}}$, $\ell = 1, \ldots, p-k$, $\quad 0 \le k \le p-1$

These are necessary conditions for a p^{th} order method

SPECIAL CASES

1) $\underline{b}^T\underline{z} = \frac{1}{\ell}$, $\ell = 1, 2, \cdots, p$ $\Rightarrow \boxed{\sum_{i=1}^{s} b_i c_i^{\ell-1} = \frac{1}{\ell}} \Rightarrow$ "pure" quadrature order condition

This condition obtained by considering the ODE $y' = t^{\ell-1}$

$y(h) = y(0) + \underbrace{\int_0^h \tau^{\ell-1}\,d\tau}$ approximating this quadrature

$\sum_{i=1}^{s} b_i c_i^{\ell-1} = \frac{1}{\ell} \Rightarrow$ the integral is exact for polynomials of degree p or less

SPECIAL CASES

(2) Set $l=1$, replace k by $k-1$ for general case

$$\Rightarrow \quad \underline{b}^T \underline{\underline{A}}^k \underline{e} = \frac{1}{l(l+1)\cdots(l+k)}$$

$$\Rightarrow \quad \underline{b}^T \underline{\underline{A}}^{k-1} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{k!}$$

This is the order condition for $y'=y$

Notice: if $k=1 \Rightarrow \sum_{i=1}^{s} b_i = 1$ (already established this constraint)

if $k=2 \Rightarrow \underline{b}^T \underline{\underline{A}} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \underline{b}^T \underline{c} = \frac{1}{2}$

this overlaps with the other special case (1) when $l=2$

IF you set $l=1$, the leading term in the truncation error is $h^p\left(\underline{b}^T \underline{\underline{A}}^p \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} - \frac{1}{(p+1)!}\right)$ IF the method is explicit, then $\underline{\underline{A}}^i = 0$ if $i \geq s$. What this implies is that the maximal order of an explicit s-stage method is $p=s$.

Example: Consider the classical RK4 method

0	0	0	0	0
1/2	1/2	0	0	0
1/2	0	1/2	0	0
1	0	0	1	0
	1/6	2/6	2/6	1/6

check order conditions:

(1) $\sum_{i=1}^{s} b_i c_i^{l-1} = \frac{1}{l}$, $l = 1, \cdots, p$

$$\Rightarrow \sum_{i=1}^{4} b_i c_i^{l-1} = \frac{1}{l}, \quad l = 1, \cdots, 4$$

examine $l=1 \Rightarrow \sum_{i=1}^{4} b_i = 1$ (satisfied)

$l=2 \Rightarrow \sum_{i=1}^{4} b_i c_i = \underbrace{\frac{2}{6}\left(\frac{1}{2}\right)}_{i=2} + \underbrace{\frac{2}{6}\left(\frac{1}{2}\right)}_{i=3} + \underbrace{\frac{1}{6}(1)}_{i=4} = \frac{1}{2}$ (satisfied)

$l=3 \Rightarrow \sum_{i=1}^{4} b_i c_i^2 = \frac{2}{6}\left(\frac{1}{2}\right)^2 + \frac{2}{6}\left(\frac{1}{2}\right)^2 + \frac{1}{6}(1)^2 = \frac{1}{3}$ (satisfied)

$l=4 \Rightarrow \sum_{i=1}^{4} b_i c_i^3 = \frac{2}{6}\left(\frac{1}{2}\right)^3 + \frac{2}{6}\left(\frac{1}{2}\right)^3 + \frac{1}{6}(1)^3 = \frac{1}{4}$ (satisfied)

(2) check: $\underline{b}^T \underline{\underline{A}}^{k-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{k!}$, $k = 3, 4$

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad A^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \end{bmatrix}$$

$$A^2 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1/4 \\ 1/2 \end{bmatrix}, \quad b^T A^2 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{4} b_3 + \frac{1}{2} b_4 = \frac{1}{4}\left(\frac{1}{2}\right) + \frac{1}{2}\left(\frac{1}{4}\right) = \frac{1}{6} = \frac{1}{3!} \text{ (satisfied)}$$

$$A^3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 \end{bmatrix}, \quad b^T A^3 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{4} b_4 = \frac{1}{4} \cdot \frac{1}{6} = \frac{1}{24} = \frac{1}{4!} \text{ (satisfied)}$$

necessary but not sufficient
(RK4 is a candidate for 4th order, but it isn't necessarily, ie we haven't shown this yet)

REMARKS

(1) The order conditions given by the two special cases are necessary and sufficient for $s = p \leq 3$.

(2) For $p = s = 4$ also require

$$\underline{b}^T \underline{\underline{A}} \begin{bmatrix} c_1^2 \\ \vdots \\ c_s^2 \end{bmatrix} = \frac{1}{12} \quad \text{(covered by general case)}$$

and $\left[b_1 c_1 \;\; b_2 c_2 \cdots b_s c_s\right] \underline{\underline{A}} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_s \end{bmatrix} = \frac{2}{4!}$ (not covered by the general case)

IVP must consider more general ~~case~~ to derive this condition

(3) See text for $p = 5$

(4)

s	1	2	3	4	5	6	7	8	9	10
p_{max}	1	2	3	4	4	5	6	6	7	7

Region of Absolute Stability

Test equation: $y' = \lambda y$

Apply an s-stage RK method:

$$Y_i = y_{n-1} + h\lambda \sum_{j=1}^{s} a_{ij} Y_j \quad , i = 1, \dots, s$$

$$y_n = y_{n-1} + h\lambda \sum_{i=1}^{s} b_i Y_i = y_{n-1} + h\lambda \, \underline{b}^T \underline{Y}$$

$$\underline{Y} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} y_{n-1} + h\lambda \underline{\underline{A}} \underline{Y} \rightarrow \left[\underline{\underline{I}} - h\lambda \underline{\underline{A}}\right] \underline{Y} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} y_{n-1}$$

$$\rightarrow \underline{Y} = \left[\underline{\underline{I}} - h\lambda \underline{\underline{A}}\right]^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} y_{n-1}$$

$$\Rightarrow y_n = y_{n-1} + h\lambda \underline{b}^T \left[\underline{\underline{I}} - h\lambda \underline{\underline{A}}\right]^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} y_{n-1}$$

$$y_n = \left(1 + h\lambda \underline{b}^T \left(\underline{\underline{I}} - h\lambda \underline{\underline{A}}\right)^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right) y_{n-1}$$

use $\left(\underline{\underline{I}} - h\lambda \underline{\underline{A}}\right)^{-1} = I + h\lambda A + (h\lambda)^2 A^2 + \dots$

remember, $\underline{b}^T \underline{\underline{A}}^{k-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \frac{1}{k!}$, $k = 1, \dots, p$

$$\Rightarrow y_n = \left[1 + \lambda h + \frac{(\lambda h)^2}{2!} + \dots + \frac{(\lambda h)^p}{p!} + \sum_{j>p} (\lambda h)^j \underline{b}^T \underline{\underline{A}}^{j-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\right] y_{n-1}$$

if explicit, then $A^{j-1} = 0$ for $j > s$

let $\lambda h = z$

$$\Rightarrow y_n = \left[\underbrace{1 + z + \frac{z^2}{2!} + \dots + \frac{z^p}{p!}}_{G} + \underbrace{\sum_{j=p+1}^{s} z^j \underline{b}^T \underline{\underline{A}}^{j-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{\text{only if } s > p, \text{ typically, for } p > 4}\right] y_{n-1}$$

If $s = p \le 4$ then $G = 1 + z + \frac{z^2}{2!} + \dots + \frac{z^p}{p!}$

then z st $|G| \le 1 \Rightarrow$ region of absolute stability

to plot region of stability, select θ, Find z st $G(z) = e^{i\theta}$
OR create grid in z-plane, compute $|G(z)|$, plot contour where $|G(z)| = 1$

NO EXPLICIT RUNGE-KUTTA METHOD IS A-STABLE

2/24/03

Error Control

Choice of the step size is done by some error control scheme. The main task is to estimate the error and pick h st local error is less than some user defined tolerance.

Consider a pair of single-step methods

$y_n = y_{n-1} + h\phi(t_{n-1}, y_{n-1}, h)$, p^{th} order method

and

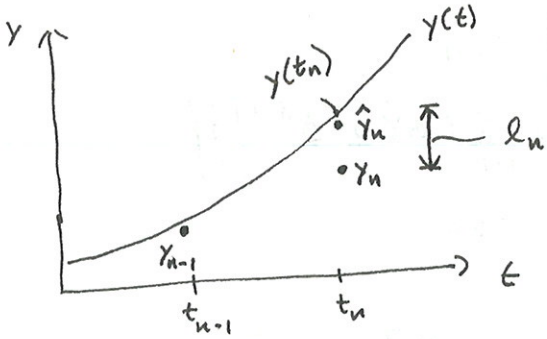
$\hat{y}_n = y_{n-1} + h\hat{\phi}(t_{n-1}, y_{n-1}, h)$, $p+1$ order method

Then the local errors are $|y_n - y(t_n)| \cong$ [illegible] $= O(h^{p+1})$

$|\hat{y}_n - y(t_n)| = O(h^{p+2})$

see pg 22 of notes - if numerical method consistent of order p then $|l_n| = O(h^{p+1})$ as $h \to 0$

where $y(t)$ solves the ODE $y' = f(t,y)$, $y(t_{n-1}) = y_{n-1}$



use $|l_n| \approx |y_n - \hat{y}_n|$

If $|y_n - \hat{y}_n| \leq \delta$ (some chosen tolerance, or $\delta|y_n|$ - relative tol) then the step to t_n would be accepted. If not then y_n is rejected and a new h (smaller) is needed.

$$|y_n - \hat{y}_n| = ch^{p+1} + \cdots$$

Want $\tilde{h}$ st $|y_n - \hat{y}_n| = \delta = c\tilde{h}^{p+1}$

Eliminate c $\Rightarrow \left(\frac{\tilde{h}}{h}\right)^{p+1} = \frac{\delta}{|y_n - \hat{y}_n|} \Rightarrow \tilde{h} = h\left(\frac{\delta}{|y_n - \hat{y}_n|}\right)^{1/p+1}$

Remarks

1) May include a "Fudge" Factor Θ to be conservative. Also include limits on q. $\tilde{h} = h\left(\frac{\Theta\,\delta}{|y_n - \hat{y}_n|}\right)^{\frac{1}{p+1}}$

2) Even if the step is accepted, calculate $\tilde{h}$ and use it for the next step.

For Runge Kutta Methods, the computational work to compute y_n and $\hat{y}_n$ can be reduced if ϕ, $\hat{\phi}$ share stage calculations.

For ϕ, might have

$\underline{c}$	$\underline{\underline{A}}$
	$\underline{b}^T$

so ϕ and $\hat{\phi}$ share $\underline{c}$ and $\underline{\underline{A}}$

For $\hat{\phi}$, want

$\underline{c}$	$\underline{\underline{A}}$
	$\underline{\hat{b}}^T$

Combine to write :

$\underline{c}$	$\underline{\underline{A}}$
	$\underline{b}^T$
	$\underline{\hat{b}}^T$

Simplest Example:

$y_n = y_{n-1} + h f(t_{n-1}, y_{n-1})$ Euler $p = 1$

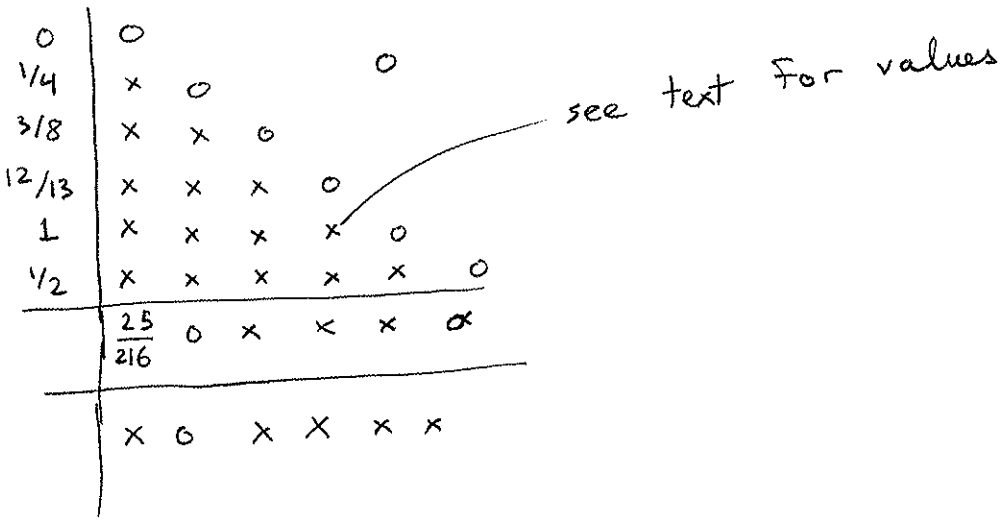
$\hat{y}_n = y_{n-1} + \frac{h}{2}\left(f(t_{n-1}, y_{n-1}) + f(t_n, y_n)\right)$, Modified Euler

0	0	0
1	1	0
	1	0
	1/2	1/2

Two - stage order 1/2 pair.

Very popular example is the

Runge-Kutta Fehlberg 4/5 pair. (



Implicit Runge Kutta Methods

implicit methods are use for stiff equations. Many are based on quadrature formulas. $y' = f(t,y)$

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

Families of Methods

- Gaussian Methods: select e's for a given s to max p, $p = 2s$

  eg, midpoint method, $s=1$, $p=2$

- Radau methods: select c's for a given s with $c_s = 1$ to maximize p. $p = 2s-1$

  eg, backward euler, $s=1$, $p=1$

- Lebatto methods: select c's for a given s with $c_1 = 0$, $c_s = 1$ to maximize p. $p = 2s-p$

  eg, trapezoidal method, $s=2$, $p=2$

These Families are collocation methods.

Pick a set of collocation points

$$0 \le c_1 < c_2 < \cdots < c_s \le 1$$

Construct a polynomial $\psi(t)$ of degree s or less that "collocates" the ODE. ie,

$$\left.\begin{aligned} \psi(t_{n-1}) &= y_{n-1} \\ \psi'(t_i) &= f(t_i, \psi(t_i)) \quad , i = 1, \ldots, s \end{aligned}\right\} \; s+1 \text{ constraints on } \psi$$

where $t_i = t_{n-1} + c_i h$

These constraints define ψ uniquely, then $y_n = \psi(t_n)$.

Show that this collocation method leads to implicit Runge K methods.

$\psi'(t)$ = polynomial of degree $s-1$ or less

$$\psi'(t) = \sum_{j=1}^{s} \psi'(t_j)\, L_j(t) \quad \leftarrow \text{polynomial in Lagrange interpolating form}$$

(note, the choices for c_i determine L)

$$L_j(t) = \prod_{\substack{k=1 \\ k \ne j}}^{s} \frac{t - t_k}{t_j - t_k}$$

Now integrate: $\psi(t) = \underbrace{\psi(t_{n-1})}_{y_{n-1}} + \int_{t_{n-1}}^{t} \psi'(\tau)\, d\tau$

$$= y_{n-1} + \int_{t_{n-1}}^{t} \sum_{j=1}^{s} \psi'(t_j)\, L_j(\tau)\, d\tau$$

$$= y_{n-1} + \sum_{j=1}^{s} \psi'(t_j) \int_{t_{n-1}}^{t} L_j(\tau)\, d\tau$$

$$\Rightarrow y_n = y_{n-1} + h \sum_{j=1}^{s} f(t_j, \psi(t_j))\, b_j \quad , \text{ where } b_j \equiv \frac{1}{h} \int_{t_{n-1}}^{t_n} L_j(\tau)\, d\tau$$

Set $Y_i = \psi(t_i) \Rightarrow Y_i = y_{n-1} + h \sum_{j=1}^{s} f(t_j, Y_j)\, a_{ij}$

$$a_{ij} = \frac{1}{h} \int_{t_{n-1}}^{t_i} L_j(\tau)\, d\tau$$

Examples: if $c_1 = \frac{1}{2} - \frac{\sqrt{3}}{6}$, $c_2 = \frac{1}{2} + \frac{\sqrt{3}}{6}$

(these are the Gauss nodes for the interval $[0,1]$)

Find:

c_1	$1/4$	$\frac{3-2\sqrt{3}}{12}$
c_2	$\frac{3+2\sqrt{3}}{12}$	$1/4$
	$1/2$	$1/2$

} 2-stage Gauss method, $p = 4$

Behavior for stiff Equations

Test equation, $y' = \lambda y$

apply implicit RK method

Generally, $\begin{array}{c|c} \underline{c} & \underline{\underline{A}} \\ \hline & \underline{b}^T \end{array}$

we had, $y_n = R(z)\, y_{n-1}$, $z = \lambda h$, where $R(z) = 1 + z\underline{b}^T \left[\underline{\underline{I}} - z\underline{\underline{A}}\right]^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

For an explicit method then R is a polynomial in z.

This implies that no explicit method is A-stable, because regions of absolute stability have to be bounded. For implicit methods, R = rational function of z, $R(z) = \frac{P(z)}{Q(z)}$

$\Rightarrow$ the region of absolute stability may be unbounded.

It turns out that the families of collocation methods (Gauss, Radau, Lebatto) are all A-stable. Additionally, if the degree of Q is greater than the degree of P, then $R \to 0$ as $z \to \infty$ $\Rightarrow$ method has stiff decay, (like backward Euler). Radau methods have stiff decay.

One way to obtain stiff decay is if $\underline{\underline{A}}$ is nonsingular and the last row of $\underline{\underline{A}}$ is equal to $\underline{b}^T$, then this will have stiff decay.

$$A = \begin{bmatrix} \cdots \\ \cdots \\ \cdots \\ \underline{b}^T \end{bmatrix}, \quad A^{-1} = \left[\begin{array}{ccc|c} & & & \frac{\underline{b}}{\|\underline{b}\|^2} \\ & & & \\ & & & \end{array}\right]$$

$s-1$ columns orthogonal to $\underline{b}$

$$R = I + z b^T (I - z\underline{\underline{A}})^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\to 1 - z b^T \left(\tfrac{1}{z}\underline{\underline{A}}^{-1}\right)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1 - \underline{b}^T \underline{\underline{A}}^{-1}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad (\text{assuming } z\underline{\underline{A}} \gg I)$$

(this is in the limit as $z \to \infty$)

Notice, $\underline{b}^T \underline{\underline{A}}^{-1}$ is a row vector, $[0, 0, \ldots, 1]$

$$\Rightarrow \lim_{z\to\infty} = 1 - \underline{b}^T \underline{\underline{A}}^{-1}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1 - (0, 0, \ldots, 0, 1)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 0$$

---

backward Euler: $\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}$ $(s = 1)$

some Radau method $\begin{array}{c|cc} 1/3 & 5/12 & -1/12 \\ 1 & 3/4 & 1/4 \\ \hline & 3/4 & 1/4 \end{array}$ $(s = 2,\ p = 3)$

3/4/03

Implicit Runge-Kutta Methods

Families of Implicit Runge-Kutta methods, derived using polynomial collocation:

Gauss: $p = 2s$ (symmetric) mid.
Radau: $p = 2s - 1$ (stiff decay) b. euler
Lobatto: $p = 2s - 2$ (symmetric) trap
} (a-stable)

Best for stiff eqns.

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}$$

- test eqn, $y' = \lambda y$

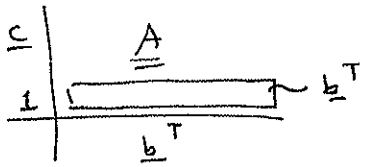
  Apply RK method: $y_n = R(z) y_{n-1}$, $z = \lambda h$

  $$R(z) = 1 + z\, \underline{b}^T (\underline{\underline{I}} - z\underline{\underline{A}})^{-1}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

  explicit → R = polynomial in z
  implicit → R = rational fct of z
  ⇒ possible to have an unbounded region of absolute stability

  All members of the families above (Gauss, Radau, Lobatto) are A stable.

Also may have stiff decay if last row of A is $\underline{b}^T$



Diagonally Implicit RK Methods (DIRK)

main difficulty - must solve a system of nonlinear algebraic equations at each time step.

General Form:

$$Y_i = y_{n-1} + h \sum_{j=1}^{s} a_{ij} f(t_j, Y_j) \quad , i = 1, \cdots, s$$

where $t_j = t_{n-1} + c_j h$

then

$$y_n = y_{n-1} + h \sum_{i=1}^{s} b_i f(t_i, Y_i)$$

Need to solve for the Y_i's. Consider some linearization: $Y_i = \hat{Y}_i + \delta_i$ ← correct to Final

↑ current guess

$$\hat{Y}_i + \delta_i = y_{n-1} + h \sum_{j=1}^{s} a_{ij} f(t_j, Y_j + \delta_j)$$

$$= y_{n-1} + h \sum_{j=1}^{s} a_{ij} \left[f(t_j, \hat{Y}_j) + f_y(t_j, \hat{Y}_j)\, \delta_j + \cdots \right]$$

Ignore dots, solve for δ_i:

$$\delta_i - h \sum_{j=1}^{s} a_{ij} f_y(t_j, \hat{Y}_j)\, \delta_j = y_{n-1} - \hat{Y}_i + h \sum_{j=1}^{s} a_{ij} f(t_j, Y_j)$$

linear system for δ

$$\begin{bmatrix} \underline{\underline{I}} - a_{11} h f_y(t_1, \hat{Y}_1) & -a_{12} h f_y(t_2, \hat{Y}_2) & \cdots \\ -a_{21} h f_y(t_1, \hat{Y}_1) & \underline{\underline{I}} - a_{22} h f_y(t_2, Y_2) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} \underline{\delta}_1 \\ \underline{\delta}_2 \\ \vdots \\ \underline{\delta}_s \end{bmatrix}$$

$m \times 1$ vector

each element an $m \times m$ matrix

so $ms \times ms$ system

Consider various approaches to solving for the s's.

1) Freeze calculation and LU-factorization of the Jacobian
2) Design its method and its a_{ij}'s s.t. the Jacobian has a simpler structure.

One choice: $a_{ij} = 0$ for $j > i$
$a_{ii} = \alpha$ for $j = i$

$\Rightarrow$ diagonally ~~impl~~ implicit runge kutta methods (DIRK)

$$Y_i = Y_{n-1} + h\alpha f(t_i, Y_i) + \underbrace{h \sum_{j=1}^{i-1} a_{ij} f(t_j, Y_j)}_{\text{explicit terms}}$$

set $Y_i = \hat{Y}_i + \delta_i$

$$\hat{Y}_i + \delta_i = Y_{n-1} + h\alpha \left(f(t_i, \hat{Y}_i) + f_y(t_i, \hat{Y}_i)\delta_i + \cdots + \right) + h\sum_{j=1}^{i-1} a_{ij} f(t_j, Y_j)$$

$$\Rightarrow \underbrace{\left(\underline{\underline{I}} - \alpha h f_y(t_i, \hat{Y}_i)\right)\delta_i}_{\text{some matrix}} = \text{known} \quad \text{(some "structure" for each } i\text{)}$$

use $\left(\underline{\underline{I}} - \alpha h f_y(t_{n-1}, Y_{n-1})\right)\delta_i = \text{known}$, for $i = 1, \ldots, s$

minimize updates and recalcs of LU

DIRK Methods: $p = s + 1$

$$\begin{array}{c|cc} 1/2 & 1/2 & \\ \hline & 1 & \end{array}$$ → midpoint, $s = 1$, $p = 2$

$$\begin{array}{c|cc} \gamma & \gamma & \\ 1-\gamma & 1-2\gamma & \gamma \\ \hline & 1/2 & 1/2 \end{array}$$ $s = 2$, $p = 3$, $\gamma = \frac{3+\sqrt{3}}{6}$ → A-stable

Stiff Decay, $p = s$

eg $\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}$ $s = p = 1$, b. euler

$$\begin{array}{c|cc} \gamma & \gamma & \\ 1 & 1-\gamma & \gamma \\ \hline & 1-\gamma & \gamma \end{array}$$ $s = p = 2$, $\gamma = \frac{2-\sqrt{2}}{2}$ (notice last row of A and b^T)

Midterm: Friday, in class, 2hrs
open notes
no book
covers everything through RK

Topics:

1) Intro
   a) IVP, BVP, BAE
   b) examples - phase plane, initial layers
   c) regularity of IVPs, Lipschitz
   d) BVP: boundary layer problem

2) IVP
   a) problem stability
      - scalar test
      - linear coeff (constant) system
      - variable coeff sys
      - non linear
   b) Hamiltonian systems
   c) Basic numerical methods and concepts
      - illustrated with Eulers method
        - local truncation error
        - consistancy
        - order of accuracy
        - global error
        - convergence
        - 0-stability

      Thm: consistent + 0-stability ⇒ convergent

        - local error vs local truncation error
        - absolute stability ($y' = \lambda y$) → region of absolute stability in the complex λh plane

   d) stiffness, implicit methods
      - backward Euler
      - A stability,
      - stiff decay
      - solving non-linear eqns
        - Newton's method

Trapezoidal rule - A-stable

e) single-step methods

$y_n = y_{n-1} + h\phi(t_{n-1}, y_{n-1}, h)$

explicit methods

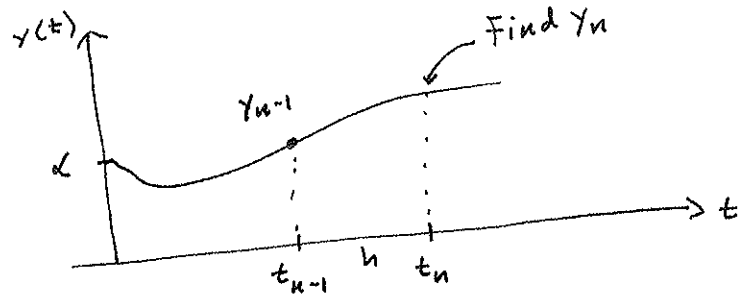
- Taylor's method
- Runge Kutta Methods
  - s-stage method $\Rightarrow$ $\begin{array}{c|c} c & A \\ \hline & b^T \end{array}$
  - consistency $\Rightarrow$ $\sum_{j=1}^{s} a_{ij} = c_i$, $\sum_{i=1}^{s} b_i = 1$
  - order conditions $\Leftarrow$ how obtained, special forms
  - regions of absolute stability
- Error control - For explicit methods
  - local error estimates
  - step size selection
- Implicit Methods
  - Gauss, Radau, Labatto
  - collocation methods
- DIRK Methods

POST SPRING BREAK

3/18/03

Linear Multistep Methods

IVP: $y' = f(t,y)$, $0 \le t \le b$, $y(0) = \alpha$



single step method uses data at y_{n-1} to calculate y_n

whereas multi-step method might use data at y_{n-2}, y_{n-3}, etc . . .

$\Rightarrow$ A k-step multistep method uses previous k solutions values to compute y_n.

General form

$$\sum_{j=0}^{k} \alpha_j y_{n-j} = h \sum_{j=0}^{k} \beta_j f_{n-j}$$

"linear" k-step multistep method

$\{\alpha_j\}$, $\{\beta_j\}$ coefficients for the method

set $\alpha_0 = 1 \Rightarrow y_n = \text{stuff}$

Assume $|\alpha_j| + |\beta_j| \neq 0$

define $f_j \equiv f(t_j, y_j)$

Observations

1) "linear" merely implies that y's, f's appear as a linear combination

2) Fixed h (traditionally). Typically, methods are derived using fixed h so that α's, β's independent of h. (Though we will discuss varying h for error control.)

3) The method is explicit if $\beta_0 = 0$ and implicit otherwise.

Adams Family

The Adams family is popular for non-stiff ODEs.

Form: ADAMS-BASHFORTH ($j=1 \dots k$)

$$y_n = y_{n-1} + h \sum_{j=0}^{k} \beta_j f_{n-j}, \quad \alpha_0 = -\alpha_1 = 1, \quad \alpha_j = 0 \text{ for } j \geq 2$$

The β_j's are chosen for accuracy.

Amongst this family of methods there are 2 subclasses:

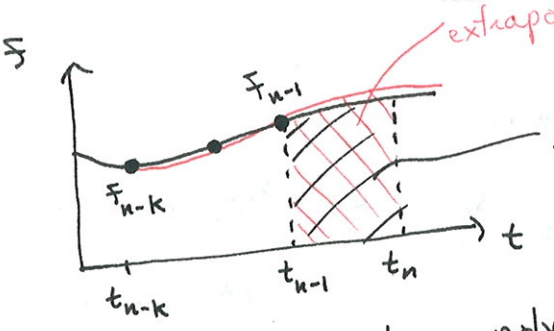
1) Adams-Bashforth - explicit methods ($\beta_0 = 0$)
   order of accuracy $\equiv p = k$

2) Adams-Moulton - implicit methods ($\beta_0 \neq 0$)
   order of accuracy $\equiv p = k+1$

derivation

$y' = f(t,y)$, integrate: $y(t_n) = y(t_{n-1}) + \underbrace{\int_{t_{n-1}}^{t_n} f(t, y(t))\, dt}_{A_n}$

Need to approximate the integral



interpolate data and extrapolate for t_{n-1} to t_n

$\Rightarrow$ Fit an interpolating polynomial to the data (t_{n-j}, f_{n-j}), $j = 1, \dots, k$

Interpolant: $\tilde{f}(t) = \sum_{j=1}^{k} f_{n-k}\, \tilde{L}_j(t)$ — Lagrange polynomials

where $\tilde{L}_j(t) = \prod_{\substack{i=1 \\ i\neq j}}^{k} \frac{(t - t_{n-i})}{(t_{n-j} - t_{n-i})}$ ← $(k-1)$ degree polynomial

$$A_n \approx \int_{t_{n-1}}^{t_n} \tilde{f}(t)\, dt = \sum_{j=1}^{k} f_{n-k} \underbrace{\int_{t_{n-1}}^{t_n} \tilde{L}_j(t)\, dt}_{h\beta_j}$$

$$\Rightarrow \quad \beta_j = \frac{1}{h} \int_{t_{n-1}}^{t_n} \tilde{L}_j(t)\, dt$$

For example (remember $\beta_0 = 0$)

AB1: , $\beta_1 = 1$ (this is just Euler's method)

AB2: , $\beta_1 = \frac{3}{2}$, $\beta_2 = -\frac{1}{2}$

AB3: $\beta_1 = \frac{23}{12}$, $\beta_2 = -\frac{16}{12}$, $\beta_3 = \frac{5}{12}$

For example, the AB3 method becomes

$$y_n = y_{n-1} + \frac{h}{12}\left[23 f_{n-1} - 16 f_{n-2} + 5 f_{n-3}\right]$$

Accuracy (more later)

$$d_n = C_{p+1} h^p y^{(p+1)}(t_n) + O(h^{p+1})$$

AB: $p = k$ so $C_2 = 1/2$, $C_3 = 5/12$, $C_4 = 3/8$, ...

↑ 3rd order

Adams-Moulton Fits a k^{th} degree polynomial to the data (t_{n-j}, f_{n-j}), $j = 0, \ldots, k$

$$\hat{f}(t) = \sum_{j=0}^{k} f_{n-k}\, \hat{L}_j(t)$$

$$\hat{L}_j(t) = \prod_{\substack{i=0 \\ i \neq j}}^{k} \frac{(t - t_{n-i})}{(t_{n-j} - t_{n-i})}$$

so the AM β_j's are $\beta_j = \frac{1}{h} \int_{t_{n-1}}^{t_n} \hat{L}_j(t)\, dt$

For example:

AM1: $\beta_0 = 1/2$, $\beta_1 = 1/2$ (trapezoidal rule)

AM2: $\beta_0 = 5/12$, $\beta_1 = 8/12$, $\beta_2 = -1/12$

(see text for others)

Remember, Adams Family For nonstiff eqns...

Accuracy $d_n = C_{p+1} h^p y^{(p+1)}(t_n) + \cdots$

$p = k + 1$

$C_3 = -1/12$, $C_4 = -1/24$ ↑ 3rd order

Backward Difference Formulas

3/21/03

Linear Multistep Methods

$y' = f(t,y)$, $y(0) = c$, $0 \le t \le b$

form: $\sum_{j=0}^{k} \alpha_j y_{n-j} = h \sum_{j=0}^{k} \beta_j f_{n-j}$, where $f_{n-j} = f(t_{n-j}, y_{n-j})$

for the Adams Method: $\alpha_0 = 1$, $\alpha_1 = -1$, $\alpha_j = 0$, $j \ge 2$

$$\Rightarrow \quad y_n = y_{n-1} + h \sum_{j=0}^{k} \beta_j f_{n-j}$$

if $\beta_0 = 0 \Rightarrow$ Adams-Bashforth (explicit) $p = k$ (order)

if $\beta_0 \neq 0 \Rightarrow$ Adams-Moulton (implicit) $p = k+1$

BDF Methods - Backwards Difference Formulas (stiff problems)

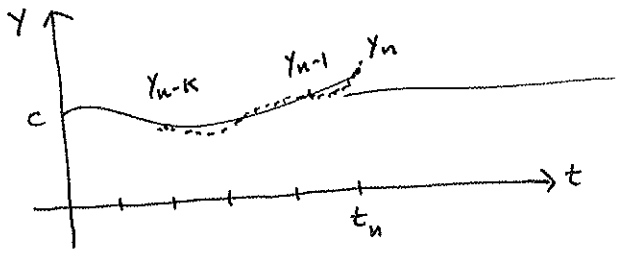
form: $\sum_{j=0}^{k} \alpha_j y_{n-j} = h \beta_0 f_n$, $\alpha_0 = 1$

start with differential eqn evaluated at t_n

$$y'(t_n) = f(t_n, y(t_n))$$

now replace $y'(t_n)$ with a backward difference formula
(this naturally derives an implicit method)

Derivation: use the data (t_{n-j}, y_{n-j}), $j = 0, \ldots, k$



fit a polynomial interpolant
(previously we approximated f, now we approximate y)

$$\Rightarrow \quad y(t) \cong \sum_{j=0}^{k} y_{n-j} \hat{L}_j(t) \quad , \quad \hat{L}_j(t) = \prod_{\substack{i=0 \\ i \neq j}}^{k} \frac{t - t_{n-i}}{t_{n-j} - t_{n-i}}$$

differentiate: $y'(t_n) \cong \sum_{j=0}^{k} y_{n-j} \hat{L}_j'(t_n)$

$$\Rightarrow \quad \sum_{j=0}^{k} y_{n-j} \hat{L}_j'(t_n) = f(t_n, y_n) \quad , \text{ normalize by } \hat{L}_0'(t_n)$$

$$\Rightarrow \alpha_j = \frac{\hat{L}_j'(t_n)}{\hat{L}_0'(t_n)}, \quad \beta_0 = \frac{1}{h\,\hat{L}_0'(t_n)}$$

Suppose $k = 2$

$$\hat{L}_0(t) = \frac{t-t_{n-1}}{t_n - t_{n-1}} \cdot \frac{t-t_{n-2}}{t_n - t_{n-2}} = \frac{1}{2h^2}(t-t_{n-1})(t-t_{n-2})$$

$$\hat{L}_0'(t) = \frac{1}{2h^2}(2t - t_{n-1} - t_{n-2})$$

$$\hat{L}_0'(t_n) = \frac{1}{2h^2}(2t_n - t_{n-1} - t_{n-2}) = \frac{1}{2h^2}(t_n - t_{n-1} + t_n - t_{n-2}) = \frac{3}{2h}$$

$$\hat{L}_1(t) = \frac{t-t_n}{t_{n-1}-t_n} \cdot \frac{t-t_{n-2}}{t_{n-1}-t_{n-2}} \rightarrow \hat{L}_1'(t_n) = -\frac{2}{h}$$

$$\hat{L}_2(t) = \frac{t-t_n}{t_{n-2}-t_n} \cdot \frac{t-t_{n-1}}{t_{n-2}-t_{n-1}} \rightarrow \hat{L}_2'(t_n) = \frac{1}{2h}$$

$$\Rightarrow \alpha_1 = -4/3, \quad \alpha_2 = 1/3, \quad \beta_0 = 2/3$$

$$\Rightarrow \text{BDF2}: \quad y_n = \frac{4}{3}y_{n-1} - \frac{1}{3}y_{n-2} + \frac{2}{3}hf(t_n, y_n)$$

Notice, this is an implicit method

see text for others...

- Order of accuracy is $p = k$

Example:

Consider the initial value problem

$y' = e^{-t} - y, \quad y(0) = 0, \quad 0 \le t \le 4$

Compare solutions of linear multistep methods for this IVP. The exact solution is $y(t) = te^{-t}$.

starting values: (this is "cheating")

$y(0) = 0, \quad y_j = t_j e^{-t_j}, \quad j = 1, \ldots, k-1$

Adams-Moulton $y_n = y_{n-1} + h\sum_{j=0}^{k} \beta_j f_{n-j}$

$$\rightarrow y_n = y_{n-1} + h\beta_0 f_n + h\sum_{j=1}^{k} \beta_j f_{n-j}$$

$$f_n = e^{-t_n} - y_n$$

$$\rightarrow (1+h\beta_0)y_n = y_{n-1} + h\beta_0 e^{-t_n} + h\sum_{j=1}^{k} \beta_j f_{n-j}$$

$$\rightarrow y_n = \frac{1}{1+h\beta_0}\left[y_{n-1} + h\beta_0 e^{-t_n} + h\sum_{j=1}^{k} \beta_j f_{n-j}\right]$$

another "cheat" to avoid implicit calculations - turn into explicit using exact solution

Observations

1) Adams-Moulton order p is generally more accurate than Adams-Bashford for the same h.

2) Compare back to RK methods.

AB4, $h = 0.00625$, error $= 8.8 \times 10^{-10}$

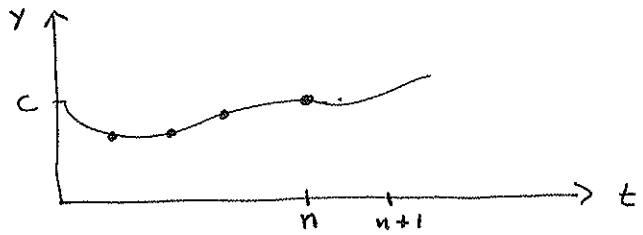
AM3, $h = 0.00625$, error $= 6.6 \times 10^{-11}$

RK4, $h = 0.00625$, error $= 6.8 \times 10^{-12}$

(For this particular IVP)

Starting Values

$$\sum_{j=0}^{k} \alpha_j y_{n-j} = h \sum_{j=0}^{k} \beta_j F_{n-j} \quad , \quad n \geq k$$



Require values for y_j, $j = 1, \ldots, k-1$ to get started

2 Basic Approaches

1) use a single step method of correct order

2) Bootstrap using multistep methods of the same family but with lower order. Need some form of error extrapolation.

Consider the Adams-Bashforth Family

AB1 $\Rightarrow$ Euler

$$y(h) = \underbrace{y(0) + hF(0)}_{y_{AB1}(h)} + ch^2 + \cdots \quad (1)$$

$$y(h) = y_{AB1}(h/2) + c(h/2)^2 + \cdots \quad (2)$$

$4(2) - (1) \Rightarrow$

$$3y(h) = 4y_{AB1}(h/2) - y_{AB1}(h) + O(h^3)$$

$$y(h) = \underbrace{\frac{4y_{AB1}(h/2) - y_{AB1}(h)}{3}}_{O(h^2) \text{ accurate approximation}} + O(h^3)$$

Order of Accuracy

Define $N_h y_n = \frac{1}{h}\sum_{j=0}^{k} \alpha_j y_{n-j} - \sum_{j=0}^{k} \beta_j f_{n-j}$

Local truncation error

$$d_n = N_h y(t_n) = \frac{1}{h}\sum_{j=0}^{k} \alpha_j y(t_{n-j}) - \sum_{j=0}^{k} \beta_j \underbrace{f(t_{n-j}, y(t_{n-j}))}_{y'(t_{n-j})}$$

Expand using Taylor series.

$$y(t_{n-j}) = y(t_n - jh) = \sum_{m=0}^{\infty} \frac{(-jh)^m}{m!} y^{(m)}(t_n)$$

$$y'(t_{n-j}) = \sum_{m=0}^{\infty} \frac{(-jh)^m}{m!} y^{(m+1)}(t_n)$$

$$\Rightarrow d_n = \frac{1}{h}\sum_{j=0}^{k} \alpha_j \sum_{m=0}^{\infty} \frac{(-jh)^m}{m!} y^{(m)}(t_n) - \sum_{j=0}^{k} \beta_j \sum_{m=0}^{\infty} \frac{(-jh)^m}{m!} y^{(m+1)}(t_n)$$

$$= (\;)h^{-1} + (\;)h^{0} + (\;)h^{1} + \cdots$$

h^{-1}: 0 — h^0: 0 — for consistency; h^1: 0 ← for order of accuracy →

$$\rightarrow d_n = \sum_{m=0}^{\infty}\left(\sum_{j=0}^{k} \alpha_j (-j)^m\right)\frac{h^{m-1}}{m!} y^{(m)}(t_n) - \sum_{m=0}^{\infty}\left(\sum_{j=0}^{k} \beta_j (-j)^m\right)\frac{h^m}{m!} y^{(m+1)}(t_n)$$

shift of m-index

$$d_n = \sum_{m=-1}^{\infty}\left(\sum_{j=0}^{k} \alpha_j (-j)^{m+1}\right)\frac{h^m}{(m+1)!} y^{(m+1)}(t_n) - \sum_{m=0}^{\infty}\left(\sum_{j=0}^{k} \beta_j (-j)^m\right)\frac{h^m}{m!} y^{(m+1)}(t_n)$$

h^{-1}: $\sum_{j=0}^{k} \alpha_j y(t_n) = 0 \Rightarrow$ **$\sum_{j=0}^{k} \alpha_j = 0$, For consistency** (don't forget to consider h^0 term)

h^m: $\sum_{j=0}^{k} \alpha_j (-j)^{m+1} \frac{y^{(m+1)}(t_n)}{(m+1)!} - \sum_{j=0}^{k} \beta_j (-j)^m \frac{y^{(m+1)}(t_n)}{m!} = 0$ For p^{th} order method for $m = 0, \ldots, p-1$

$$\Rightarrow \boxed{\sum_{j=0}^{k} \alpha_j (-j)^{m+1} = \sum_{j=0}^{k} \beta_j (-j)^m (m+1), \text{ For } m = 0, 1, \ldots, p-1}$$

In summary:

For consistency:

$$\sum_{j=0}^{k} \alpha_j = 0 \quad \text{AND} \quad -\sum_{j=0}^{k} j\alpha_j = \sum_{j=0}^{k} \beta_j$$

Truncation Error:

$$d_n = \underbrace{\left| \frac{1}{(p+1)!}\left(\sum_{j=0}^{k} \alpha_j (-j)^{p+1}\right) - \frac{1}{p!}\sum_{j=0}^{k} \beta_j (-j)^p \right|}_{C_{p+1}} h^p y^{(p+1)}(t_n) + \cdots$$

Example

use the order conditions to ~~derive~~ obtain the α_j, β_j in a linear multistep method.

$$y_n = y_{n-1} + h\sum_{j=0}^{2} \beta_j f_{n-j} \qquad (AB\,2)$$

in this form, we've already chosen

$$\alpha_0 = 1, \quad \alpha_1 = -1, \quad \alpha_2 = 0 \qquad (\text{Notice, } \textstyle\sum \alpha_j = 0)$$

Problem is to choose $\beta_0, \beta_1, \beta_2$ st the method is 3rd order.

$$\rightarrow \quad \sum_{j=0}^{2} (-j)^{m+1}\alpha_j = (m+1)\sum_{j=0}^{2} (-j)^m \beta_j \quad , \quad m = 0, 1, 2$$

$m=0$: $\quad 1 = \beta_0 + \beta_1 + \beta_2$

$m=1$: $\quad -1 = 2(-\beta_1 - 2\beta_2)$

$m=2$: $\quad 1 = 3(\beta_1 + 4\beta_2)$

$$\Rightarrow \beta_0 = \frac{5}{12}, \quad \beta_1 = \frac{8}{12}, \quad \beta_2 = \frac{-1}{12}$$

Root Condition $\quad \alpha_0 \xi^k + \alpha_1 \xi^{k-1} + \cdots + \alpha_{k-1}\xi + \alpha_k = 0$

stability condition involving roots of this characteristic polynomial. what are the magnitudes of the roots?

0-Stability & The Root Condition For Linear Multistep Methods

General k-step method gives

$$N_h y_n = \frac{1}{h}\sum_{j=0}^{k} \alpha_j y_{n-j} - \sum_{j=0}^{k} \beta_j F_{n-j}$$

0-stability requires that

$$|x_m - z_m| \le K\left[\sum_{j=0}^{k-1} |x_j - z_j| + \max_{k\le n\le N} |N_h x_n - N_h z_n|\right]$$

For $0 \le m \le N$, $h \le h_0$

$$\Rightarrow |x_m - z_m| \le K\left[\sum_{j=0}^{k-1} |x_j - z_j| + \max_{k\le n\le N} \left|\underbrace{\frac{1}{h}\sum_{j=0}^{k} \alpha_j (x_{n-j} - z_{n-j})}_{\text{dominant term in this expression}} - \sum_{j=0}^{k} \beta_j \left(f(x_{n-j}) - f(z_{n-j})\right)\right|\right]$$

Consider the behavior of solutions to the difference eqn

$$\sum_{j=0}^{k} \alpha_j v_{n-j} = 0$$

$$\Rightarrow \alpha_0 v_n + \alpha_1 v_{n-1} + \cdots + \alpha_{k-1} v_{n-k+1} + \alpha_k v_{n-k} = 0$$

Constant coeff linear difference equation.

Set $v_n = \xi^n$, where ξ = (complex) constant

or equivalently, $v_n = e^{n \ln \xi} = e^{t_n\left(\frac{\ln \xi}{h}\right)}$, $t_n = nh$

→ exponential solution for linear constant coeff difference eqn

$$\rightarrow \alpha_0 \xi^n + \alpha_1 \xi^{n-1} + \cdots + \alpha_k \xi^{n-k} = 0$$

$$\xi^{n-k}\underbrace{\left(\alpha_0 \xi^k + \alpha_1 \xi^{k-1} + \cdots + \alpha_k\right)}_{p(\xi)} = 0$$

$p(\xi)$ = characteristic polynomial of the difference eqn. There are k roots of $p(\xi) = 0$, $\xi_1, \xi_2, \ldots, \xi_k$.

general solution: $v_n = \sum_{m=1}^{k} c_m \xi_m^n$, c_m determined by initial conditions

Remarks

1) If we assume that multistep method is consistent

then $\sum_{j=0}^{k} \alpha_j = 0$ and $\rho(1) = \sum_{j=0}^{k} \alpha_j = 0$

$\Rightarrow \xi = 1$ is always a root of the characteristic polynomial for a consistent multistep method

$\Rightarrow v_n = c_1 + \sum_{m=2}^{k} c_m \xi_m^n$

2) If the roots are repeated then the form changes slightly.

eg, suppose $\xi_2 = \xi_3 = \hat{\xi}$

then $v_n = c_1 + (c_2 + nc_3)\hat{\xi}^n + \sum_{m=4}^{k} c_m \xi_m^n$

triple root $\Rightarrow \hat{\xi}^n, n\hat{\xi}^n, n^2\hat{\xi}^n$, and so on...

3) terminology:

$\xi_1 = 1 \Rightarrow$ principal root

ξ_m, $m = 2, \ldots, k \Rightarrow$ extraneous roots

The extraneous roots determine stability

Root Condition

Require that $|\xi_m| \leq 1$ for all m For stability, (and if $|\xi_m| = 1$, ξ_m must be a simple root, ie non-repeatable).

"strongly stable" $\Rightarrow$ $|\xi_m| < 1$ for $m \geq 2$

"weakly stable" $\Rightarrow$ 0-stable, but not strongly stable

Examples

① Adams Methods all have

$\alpha_0 = 1$, $\alpha_1 = -1$, $\alpha_j = 0$ for $j \geq 2$

$\Rightarrow p(\xi) = \xi^k - \xi^{k-1}$

$= \xi^{k-1}(\xi - 1)$

$\Rightarrow \xi_1 = 0$, $\xi_m = 0$, $m \geq 2$ $\Rightarrow$ **strongly stable**

② "leap frog" - 2 step method (2 roots to char poly)

$y_n = y_{n-2} + 2h f_{n-1}$

$p(\xi) = \xi^2 - 1 \Rightarrow \xi_1 = 1, \xi_2 = -1$ → extraneous root sitting on boundary

$\Downarrow$

weakly stable

③ $y' = f(t, y)$

take $y'(t_{n-1}) = f(t_{n-1}, y(t_{n-1}))$

and now approximate $y'(t_{n-1})$ using some finite difference

$\Rightarrow \frac{1}{14h}(5y_n + 6y_{n-1} - 13y_{n-2} + 2y_{n-3}) = f_{n-1}$

this is a 3-step method

$5y_n + 6y_{n-1} - 13y_{n-2} + 2y_{n-3} = 14h f_{n-1}$

$\Rightarrow p(\xi) = 5\xi^3 + 6\xi^2 - 13\xi + 2$

$= (\xi - 1)(5\xi^2 + 11\xi - 2)$

$\xi_1 = 1$; extraneous roots: $\xi_2, \xi_3 = \frac{-11 \pm \sqrt{161}}{10}$

$= 0.169, -2.37$

↑

UNSTABLE

Test these using $y' = e^{-t} - y$, $y(0) = 0$

see plots on website

code

Thm: consistent method + 0-stability $\Rightarrow$ convergent

↓ easy to show for linear multistep method

↓ not too hard to show

Absolute stability

Apply a general k-step method to test equation, $y' = \lambda y$

Have $\sum_{j=0}^{k} \alpha_j y_{n-j} = h\lambda \sum_{j=0}^{k} \beta_j y_{n-j}$

let $z = h\lambda$

$$\Rightarrow \sum_{j=0}^{k} \alpha_j y_{n-j} = z \sum_{j=0}^{k} \beta_j y_{n-j}$$

set $y_n = \xi^n \Rightarrow \sum_{j=0}^{k} \alpha_j \xi^{n-j} = z \sum_{j=0}^{k} \beta_j \xi^{n-j}$

$$\Rightarrow \underbrace{\sum_{j=0}^{k} \alpha_j \xi^{k-j}}_{\rho(\xi)} = z \underbrace{\sum_{j=0}^{k} \beta_j \xi^{k-j}}_{\sigma(\xi)}$$

generally, $\rho(\xi)$ and $\sigma(\xi)$ are k-degree polynomials
(if explicit, then (k-1) degree polynomials)

roots $\xi_1, \ldots, \xi_k$ of the polynomial

$$\rho(\xi) - z\,\sigma(\xi) = 0$$

For a given complex value for z

general solution: $y_n = \sum_{m=1}^{k} c_m \xi_m^n$

we want decay $\Rightarrow$

Region of absolute stability: $\{ z \text{ st } |\xi_m| \leq 1 \}$

To find region, set $\xi = e^{i\theta}$, $-\pi \le \theta \le \pi$

then compute $z = \dfrac{\rho(e^{i\theta})}{\sigma(e^{i\theta})}$ = boundary of region of absolute stability

Implementation

(1) predictor/corrector methods

- used for non-stiff problems, Adams-Bashforth and Adams-Moulton methods are often used in a predictor-corrector way.

predictor-corrector scheme "PECE"

predictor step: AB, explicit

example: 3-step

$$\tilde{y}_n = y_{n-1} + \frac{h}{12}\left[23 f_{n-1} - 16 f_{n-2} + 5 f_{n-3}\right]$$

$\tilde{y}_n$: predicted value

corrector step: AM, "implicit" (2-step)

$$y_n = y_{n-1} + \frac{h}{12}\left[5\tilde{f}_n + 8 f_{n-1} - f_{n-2}\right]$$

$\tilde{f}_n = f(t_n, \tilde{y}_n)$

cost = 2 function evaluations per step

(2) Error estimates

For a predictor-corrector scheme, an error estimate is natural and comes for free.

predictor step (AB): $\tilde{y}_n - y(t_n)$ = local error $= h d_n + \cdots$

where $d_n = \tilde{c}_{p+1} h^p y^{(p+1)}(t_n) + \cdots$

$$\Rightarrow \quad \tilde{y}_n - y(t_n) = \tilde{c}_{p+1} h^{p+1} y^{(p+1)}(t_n) + \cdots$$

corrector step (AM): $y_n - y(t_n) = \underbrace{c_{p+1} h^{p+1} y^{(p+1)}(t_n) + \cdots}$

assume this even though y_n is obtained using $\tilde{f}_n$

subtract: $\tilde{y}_n - y_n = \left(\tilde{c}_{p+1} - c_{p+1}\right) h^{p+1} y^{(p+1)}(t_n) + \cdots$

$$\rightarrow \quad h^{p+1} y^{(p+1)}(t_n) = \frac{\tilde{y} - y_n}{\tilde{c}_{p+1} - c_{p+1}} + \cdots$$

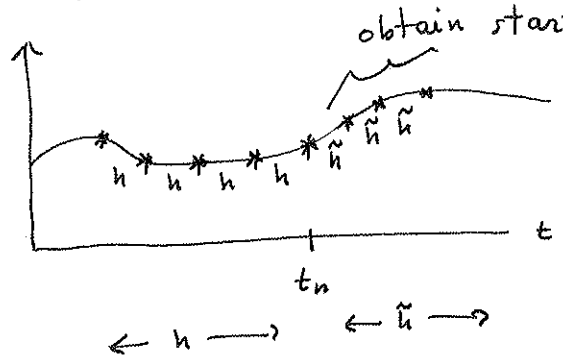
substitute this error into local error of corrector step

$$\rightarrow \quad y_n - y(t_n) \approx \frac{c_{p+1}}{\tilde{c}_{p+1} - c_{p+1}} \left(\tilde{y}_n - y_n\right), \quad \text{Milne estimate}$$

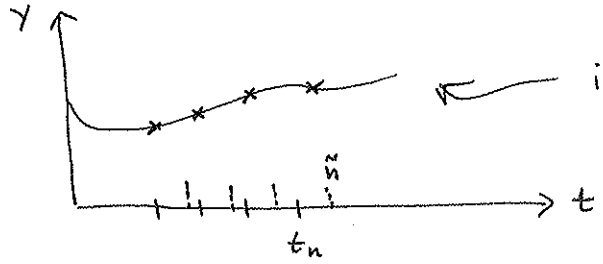
(3) Variable step size

In an error control algorithm you may need to change h according to some error estimate. Choices to accomodate the change in h.

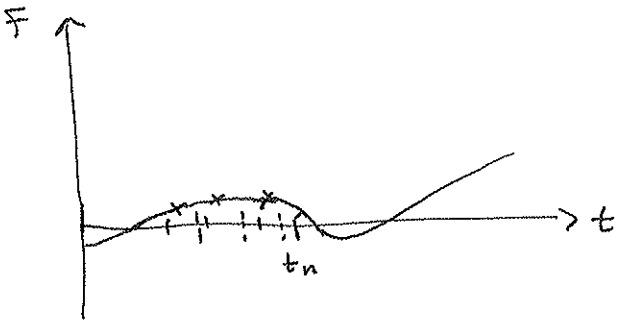
[a] restart multi-step method with new h using a single step method.



[b] Interpolate y or F.



interpolate curves then grab values of at intervals of $\tilde{h}$ to continue computations at $\tilde{h}$



[C] Variable step multistep method

had $$\sum_{j=0}^{k} \alpha_j y_{n-j} = h \sum_{j=0}^{k} \beta_j f_{n-j}$$

where α_j, β_j are found assuming uniform h

Instead find α_j, β_j assuming arbitrary set of nodes t_{n-j}. The formulas are more complicated and involve h_n.

Boundary Value Problems

General system of m ODEs.

$$\underline{y}' = \underline{f}(t, \underline{y}) \quad , \quad a \le t \le b$$

with m boundary conditions, $\underline{g}(\underline{y}(a), \underline{y}(b)) = 0$.

Solution of the BVP is not local, ie for IVP we can write solution locally using a Taylor series but this is not possible for a BVP in general.

This makes the question of existence and uniqueness of solutions more difficult.

Consider a corresponding IVP

$$\underline{y}' = \underline{f}(t, \underline{y}) \quad , a \le t \le b \, , \quad \underline{y}(a) = \underline{s}$$

Suppose the IVP is well-posed. The solution exists uniquely and it depends continuously on f, c.

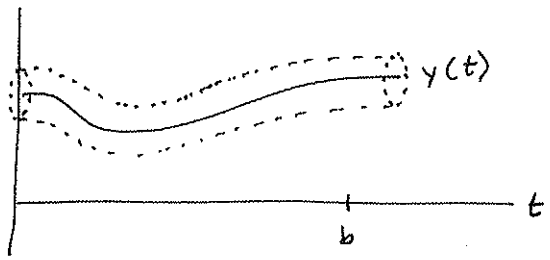
Solution of the IVP is $\underline{y} = \underline{y}(t, \underline{s})$. For the BVP, $\underline{s}$ is chosen st $g(\underline{s}, \underline{y}(b, \underline{s})) = 0$ ← m algebraic equations for the m-components of $\underline{s}$. This eqn may have none, one, or many solutions.

For certain cases, you can show that a solution exists.

For example, $u'' = F(t, u, u')$, $a \le t \le b$, $u(a) = \alpha$, $u(b) = \beta$ ★

If $F(t, u, v)$ is continuous and Lipschitz in the variables u, v for a domain $a \le t \le b$ and u, v bounded, then ★ has a unique solution for b sufficiently small (see Ascher, Mattheij, Russel).

For the general case, suppose a solution exists. Consider now the local uniqueness of the solution $\Rightarrow$ whether the solution is isolated.



A solution is isolated if there are no other solutions in a "tube" about it.

Variational problem

Linearize the general boundary value problem about a solution $\underline{y}(t)$:

$$\underline{z}'(t) = \underline{\underline{A}}(t, \underline{y}(t))\,\underline{z}(t), \quad 0 \le t \le b$$
$$\underline{\underline{B}}_0\,\underline{z}(0) + \underline{\underline{B}}_1\,\underline{z}(b) = \underline{0}$$

} linear, homogeneous problem

where $\underline{\underline{A}} = \dfrac{\partial \underline{f}}{\partial \underline{y}}(t, \underline{y}(t))$, Jacobian of $\underline{f}$ evaluated at t and $\underline{y}(t)$

$$\underline{\underline{B}}_0 = \frac{\partial \underline{g}}{\partial \underline{u}}(\underline{y}(0), \underline{y}(b)) \quad \text{where } \underline{g}(\underline{y}(0), \underline{y}(b)) = \underline{g}(\underline{u}, \underline{v}) = \underline{0}$$

$$\underline{\underline{B}}_1 = \frac{\partial \underline{g}}{\partial \underline{v}}(\underline{y}(0), \underline{y}(b))$$

the solution $\underline{y}(t)$ is isolated if the variational problem has only the trivial solution, $\underline{z} = 0$.

Linear BVPs

Consider the BVP, $\underline{y}' = \underline{\underline{A}}(t)\,\underline{y} + \underline{q}(t)$, $0 \le t \le b$

with boundary conditions: $\underline{\underline{B}}_0\,\underline{y}(0) + \underline{\underline{B}}_1\,\underline{y}(b) = \underline{\gamma}$

} non-homogeneous linear problem

Corresponding IVP:

$\underline{y}' = \underline{\underline{A}}(t)\,\underline{y} + \underline{q}$ (same ODE), $0 \le t \le b$

with initial condition $\underline{y}(0) = \underline{c}$

If $\underline{\underline{A}}(t)$ is continuous on $[0, b]$, then there is a fundamental solution matrix $\underline{\underline{X}}(t)$ s.t.

$$\underline{\underline{X}}'(t) = \underline{\underline{A}}(t)\,\underline{\underline{X}}, \quad 0 \le t \le b$$

$$\underline{\underline{X}}(0) = \underline{\underline{I}}$$

The solution of the IVP is

$$\underline{y}(t) = \underline{\underline{X}}(t)\left[\underline{c} + \int_0^t \underline{\underline{X}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau\right]$$

The solution of the linear BVP occurs for some choice of $\underline{c}$.

BC $\Rightarrow$ $\underline{\underline{B}}_0\,\underline{y}(0) + \underline{\underline{B}}_1\,\underline{y}(b) = \underline{\gamma}$

$\Rightarrow$ ~~$\underline{\underline{B}}_0\underline{c} + \underline{\underline{B}}_1 \underline{\underline{X}}(b)\left[\underline{c} + \int_0^b \underline{\underline{X}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau\right]$~~

$$\Rightarrow \underline{\underline{B}}_0\,\underline{c} + \underline{\underline{B}}_1\,\underline{\underline{X}}(b)\left[\underline{c} + \int_0^b \underline{\underline{X}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau\right] = \underline{\gamma}$$

$$\rightarrow \underbrace{\left[\underline{\underline{B}}_0 + \underline{\underline{B}}_1\,\underline{\underline{X}}(b)\right]}_{\underline{\underline{Q}}}\underline{c} = \underline{\gamma} - \underline{\underline{B}}_1\,\underline{\underline{X}}(b)\int_0^b \underline{\underline{X}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau$$

A unique solution of the linear boundary value problem exists if $\underline{\underline{Q}}$ is <u>nonsingular</u>.

<u>Example</u> $\quad u'' + \lambda u = 0, \quad 0 \le t \le b$

$u(0) = \alpha, \quad u'(b) = \beta$

Dirichlet condition ↙ ; ↳ Neumann condition

Rewrite as a system: $\underline{y} = \begin{pmatrix} u \\ u' \end{pmatrix}$

$$\Rightarrow \underline{y}' = \begin{pmatrix} 0 & 1 \\ -\lambda & 0 \end{pmatrix}\underline{y}, \quad 0 \le t \le b$$

boundary conditions to be written as $\underline{\underline{B}}_0\begin{pmatrix} u(0) \\ u'(0) \end{pmatrix} + \underline{\underline{B}}_1\begin{pmatrix} u(b) \\ u'(b) \end{pmatrix} = ?$

$$\rightarrow \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}}_{\underline{\underline{B}}_0}\begin{pmatrix} u(0) \\ u'(0) \end{pmatrix} + \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}}_{\underline{\underline{B}}_1}\begin{pmatrix} u(b) \\ u'(b) \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$ ↖ separated BCs

Want to analyze $\underline{\underline{Q}}$, so calculate Fundamental solution $\underline{\underline{X}}$

The general soln of $u'' + \lambda u = 0$ is $u(t) = c_1 \cos\sqrt{\lambda}t + c_2 \sin\sqrt{\lambda}t$

$\rightarrow u'(t) = -\sqrt{\lambda}c_1 \sin\sqrt{\lambda}t + \sqrt{\lambda}c_2 \cos\sqrt{\lambda}t$

$\rightarrow \underline{Y}(t) = \begin{pmatrix} u \\ u' \end{pmatrix} = c_1 \begin{pmatrix} \cos\sqrt{\lambda}t \\ -\sqrt{\lambda}\sin\sqrt{\lambda}t \end{pmatrix} + c_2 \begin{pmatrix} \sin\sqrt{\lambda}t \\ \sqrt{\lambda}\cos\sqrt{\lambda}t \end{pmatrix}$

$\underline{\underline{X}}(t) = \begin{bmatrix} | & | \\ x_1 & x_2 \\ | & | \end{bmatrix}$, $\underline{x}_1(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\underline{x}_2(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Find $\underline{\underline{X}}(t) = \begin{bmatrix} \cos\sqrt{\lambda}t & \frac{1}{\sqrt{\lambda}}\sin\sqrt{\lambda}t \\ -\sqrt{\lambda}\sin\sqrt{\lambda}t & \cos\sqrt{\lambda}t \end{bmatrix}$

$\underline{\underline{Q}} = \underline{\underline{B}}_0 + \underline{\underline{B}}_1 \underline{\underline{X}}(b) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\sqrt{\lambda}b & \frac{1}{\sqrt{\lambda}}\sin\sqrt{\lambda}b \\ -\sqrt{\lambda}\sin\sqrt{\lambda}b & \cos\sqrt{\lambda}b \end{bmatrix}$

$= \begin{bmatrix} 1 & 0 \\ -\sqrt{\lambda}\sin\sqrt{\lambda}b & \cos\sqrt{\lambda}b \end{bmatrix}$

$\det(Q) = \cos\sqrt{\lambda}b = 0 \Rightarrow Q$ is singular

$\Rightarrow \sqrt{\lambda}b \neq \pi\left(k - \frac{1}{2}\right)$, $k = 1, 2, \ldots$ Then unique soln exists

Green's Functions

Had the following solution for the BVP

$$(\star)\quad \underline{y}(t) = \underline{\underline{X}}(t)\left[\underline{c} + \int_0^t \underline{\underline{X}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau\right]$$

where $\underline{c}$ solves $\underline{\underline{Q}}\,\underline{c} = \underline{\gamma} - \underline{\underline{B}}_1 \underline{\underline{X}}(b) \int_0^b \underline{\underline{X}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau$

Object is to rearrange this solution to the form

$$\underline{y}(t) = \underline{\underline{\Phi}}(t)\,\underline{\gamma} + \int_0^b \underline{\underline{G}}(t,\tau)\,\underline{q}(\tau)\,d\tau$$

$\underline{\underline{\Phi}}(t)$: scaled Fundamental solution matrix

$\underline{\underline{G}}(t,\tau)$: kernel of integral - Greens Function

substitute $\underline{c}$ into $(\star)$:

$$\underline{y}(t) = \underline{\underline{X}}(t)\,\underline{\underline{Q}}^{-1}\,\underline{\gamma} - \underline{\underline{X}}(t)\,\underline{\underline{Q}}^{-1}\,\underline{\underline{B}}_1\,\underline{\underline{X}}(b)\int_0^b \underline{\underline{X}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau + \underline{\underline{X}}(t)\int_0^t \underline{\underline{X}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau$$

Let $\underline{\underline{\Phi}}(t) = \underline{\underline{X}}(t)\,\underline{\underline{Q}}^{-1}$ $\rightarrow$ $\underline{\underline{X}}(t) = \underline{\underline{\Phi}}(t)\,\underline{\underline{Q}}$ and $\underline{\underline{\Phi}}^{-1}(t) = \underline{\underline{Q}}\,\underline{\underline{X}}(t)^{-1}$

$$\underline{y}(t) = \underline{\underline{\Phi}}(t)\,\underline{\gamma} - \underline{\underline{\Phi}}(t)\,\underline{\underline{B}}_1\,\underline{\underline{X}}(b)\,\underline{\underline{Q}}^{-1}\underline{\underline{Q}}\int_0^b \underline{\underline{X}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau + \underline{\underline{X}}(t)\,\underline{\underline{Q}}^{-1}\underline{\underline{Q}}\int_0^t \underline{\underline{X}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau$$

$$\rightarrow\ \underline{y}(t) = \underline{\underline{\Phi}}(t)\,\underline{\gamma} \ominus \underline{\underline{\Phi}}(t)\,\underline{\underline{B}}_1\,\underline{\underline{\Phi}}(b)\int_0^b \underline{\underline{\Phi}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau + \underline{\underline{\Phi}}(t)\int_0^t \underline{\underline{\Phi}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau$$

Note: $\underline{\underline{Q}} = \underline{\underline{B}}_0 + \underline{\underline{B}}_1\,\underline{\underline{X}}(b)$

$= \underline{\underline{B}}_0\,\underline{\underline{X}}(0) + \underline{\underline{B}}_1\,\underline{\underline{X}}(b)$, because $\underline{\underline{X}}(0) = \underline{\underline{I}}$

$= \underline{\underline{B}}_0\,\underline{\underline{\Phi}}(0)\,\underline{\underline{Q}} + \underline{\underline{B}}_1\,\underline{\underline{\Phi}}(b)\,\underline{\underline{Q}}$

$\Rightarrow\ \underline{\underline{I}} = \underline{\underline{B}}_0\,\underline{\underline{\Phi}}(0) + \underline{\underline{B}}_1\,\underline{\underline{\Phi}}(b)$

then $\underline{\underline{B}}_1\,\underline{\underline{\Phi}}(b)\int_0^b \underline{\underline{\Phi}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau = \underline{\underline{B}}_1\,\underline{\underline{\Phi}}(b)\int_0^t \underline{\underline{\Phi}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau + \underline{\underline{B}}_1\,\underline{\underline{\Phi}}(b)\int_t^b \underline{\underline{\Phi}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau$

$$= \left(\underline{\underline{I}} - \underline{\underline{B}}_0\,\underline{\underline{\Phi}}(0)\right)\int_0^t \underline{\underline{\Phi}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau + \underline{\underline{B}}_1\,\underline{\underline{\Phi}}(b)\int_t^b \underline{\underline{\Phi}}^{-1}(\tau)\,\underline{q}(\tau)\,d\tau$$

$$\Rightarrow \quad \underline{y}(t) = \underline{\underline{\Phi}}(t)\underline{\gamma} - \underline{\underline{\Phi}}(t)\Big((\underline{\underline{I}} - \underline{\underline{B}}_0\underline{\underline{\Phi}}(0)) \int_0^t \underline{\underline{\Phi}}^{-1}(\tau)\underline{q}(\tau)\,d\tau + \underline{\underline{B}}_1\underline{\underline{\Phi}}(b)\int_t^b \underline{\underline{\Phi}}^{-1}(\tau)\underline{q}(\tau)\,d\tau \Big)$$

$$+ \underline{\underline{\Phi}}(t)\int_0^t \underline{\underline{\Phi}}^{-1}(\tau)\underline{q}(\tau)\,d\tau$$

$$\rightarrow \quad \underline{y}(t) = \underline{\underline{\Phi}}(t)\underline{\gamma} + \underline{\underline{\Phi}}(t)\underline{\underline{B}}_0\underline{\underline{\Phi}}(0)\int_0^t \underline{\underline{\Phi}}^{-1}(\tau)\underline{q}(\tau)\,d\tau - \underline{\underline{\Phi}}(t)\underline{\underline{B}}_1\underline{\underline{\Phi}}(b)\int_t^b \underline{\underline{\Phi}}^{-1}(\tau)\underline{q}(\tau)\,d\tau$$

$$\Rightarrow \boxed{\underline{y}(t) = \underline{\underline{\Phi}}(t)\underline{\gamma} + \int_0^b \underline{\underline{G}}(t,\tau)\underline{q}(\tau)\,d\tau}$$

where $$\underline{\underline{G}}(t,\tau) = \begin{cases} \underline{\underline{\Phi}}(t)\underline{\underline{B}}_0\underline{\underline{\Phi}}(0)\underline{\underline{\Phi}}^{-1}(\tau), & 0 \le \tau \le t \\ -\underline{\underline{\Phi}}(t)\underline{\underline{B}}_1\underline{\underline{\Phi}}(b)\underline{\underline{\Phi}}^{-1}(\tau), & t \le \tau \le b \end{cases}$$

GREENS FUNCTION

Problem Stability

Consider the linear boundary value problem

$$\underline{y}' = \underline{\underline{A}}(t)\underline{y} + \underline{q}(t), \quad 0 \le t \le b, \quad \underline{\underline{B}}_0\underline{y}(0) + \underline{\underline{B}}_1\underline{y}(b) = \underline{\gamma}$$

solution is

$$\underline{y}(t) = \underline{\underline{\Phi}}(t)\underline{\gamma} + \int_0^b \underline{\underline{G}}(t,\tau)\underline{q}(\tau)\,d\tau$$

The behavior of the solution depends on $\underline{\underline{\Phi}}(t)$ and $\underline{\underline{G}}(t,\tau)$.

Consider a perturbed problem:

$$\tilde{\underline{y}}' = \underline{\underline{A}}(t)\tilde{\underline{y}} + \underline{q}(t) + \underline{\delta}(t), \quad \underline{\underline{B}}_0\tilde{\underline{y}}(0) + \underline{\underline{B}}_1\tilde{\underline{y}}(b) = \underline{\gamma} + \underline{\epsilon}$$

Consider the difference $\underline{w}(t) = \tilde{\underline{y}}(t) - \underline{y}(t)$

$\underline{w}(t)$ solves $\underline{w}' = \underline{\underline{A}}(t)\underline{w} + \underline{\delta}(t)$, $\underline{\underline{B}}_0\underline{w}(0) + \underline{\underline{B}}_1\underline{w}(b) = \underline{\epsilon}$

$$\Rightarrow \quad \underline{w}(t) = \underline{\underline{\Phi}}(t)\underline{\epsilon} + \int_0^t \underline{\underline{G}}(t,\tau)\underline{\delta}(\tau)\,d\tau$$

If $$K = \max_{0 \le t,\tau \le b}\Big(\|\underline{\underline{G}}(t,\tau)\|_\infty, \|\underline{\underline{\Phi}}(t)\|_\infty \Big)$$

then $$\|\underline{w}(t)\| = \max_{0 \le t \le b} \|\underline{w}(t)\| \le K\left(\|\underline{\epsilon}\|_\infty + \int_0^b \|\underline{\delta}(t)\|_\infty\,d\tau \right)$$

↳ K is an absolute condition number

Problem Stability for Linear BVPs

BVP: $\underline{y}' = \underline{\underline{A}}(t)\,\underline{y} + \underline{q}(t)$, $0 \le t \le b$

$$\underline{\underline{B}}_0\,\underline{y}(0) + \underline{\underline{B}}_1\,\underline{y}(b) = \underline{\gamma}$$

Recall: Fundamental solution matrix for the IVP

$$\underline{\underline{X}}' = \underline{\underline{A}}(t)\,\underline{\underline{X}} \quad , \quad \underline{\underline{X}}(0) = \underline{\underline{I}}$$

Define $\underline{\underline{Q}} = \underline{\underline{B}}_0 + \underline{\underline{B}}_1\,\underline{\underline{X}}(b)$, then a solution of the BVP exists if $\underline{\underline{Q}}$ is nonsingular.

Suppose $\underline{\underline{Q}}$ is nonsingular. Define $\underline{\underline{\Phi}}(t) = \underline{\underline{X}}(t)\,\underline{\underline{Q}}^{-1}$, then the solution of BVP is:

$$\underline{y}(t) = \underline{\underline{\Phi}}(t)\,\underline{\gamma} + \int_0^b \underline{\underline{G}}(t,\tau)\,\underline{q}(\tau)\,d\tau ,$$

$$\underline{\underline{G}}(t,\tau) = \begin{cases} \underline{\underline{\Phi}}(t)\,\underline{\underline{B}}_0\,\underline{\underline{\Phi}}(0)\,\underline{\underline{\Phi}}^{-1}(\tau) & t > \tau \\ -\underline{\underline{\Phi}}(t)\,\underline{\underline{B}}_1\,\underline{\underline{\Phi}}(b)\,\underline{\underline{\Phi}}^{-1}(\tau) & t < \tau \end{cases}$$

Problem stability depends on the behavior of $\underline{\underline{\Phi}}(t)$, $\underline{\underline{G}}(t,\tau)$.

- Assume the boundary conditions are **separated**:

$$\Rightarrow \underline{\underline{B}}_0 = \begin{bmatrix} \times & \cdots & \times & \cdots & \times & \cdots & \times \\ \vdots & & \vdots & & \vdots & & \vdots \\ \times & \cdots & \times & \cdots & \times & \cdots & \times \\ 0 & & & & & & 0 \end{bmatrix} \begin{matrix} \} k \\ \} m \end{matrix} \quad , \quad \underline{\underline{B}}_1 = \begin{bmatrix} 0 & & & & 0 \\ \times & \cdots & \times & \cdots & \times \\ \vdots & & \vdots & & \vdots \\ \times & \cdots & \times & \cdots & \times \end{bmatrix} \begin{matrix} \} k \\ \} m-k \end{matrix}$$

(m columns)

"k modes of the solution are forced from the left whereas $m-k$ modes are forced from the right"

If $\underline{\underline{\Phi}}(t)$ has first k columns that decay as t increases and the last $m-k$ columns that decay as t decreases, then the BVP is **stable**. This is called an (exponential) **dichotomy**.

return to boundary conditions

$$\underline{\underline{B}}_0 \underline{\underline{\Phi}}(0) + \underline{\underline{B}}_1 \underline{\underline{\Phi}}(b) = \underline{\underline{I}} \quad , \text{ because } \underline{\underline{\Phi}}(t) = \underline{\underline{X}}(t)\, Q^{-1}$$

if $\underline{\underline{B}}_0 = \begin{bmatrix} \text{////} \\ \hline 0 \end{bmatrix}$ and $\underline{\underline{B}}_1 = \begin{bmatrix} 0 \\ \hline \text{////} \end{bmatrix}$

then $\underline{\underline{B}}_0 \underline{\underline{\Phi}}(0) = \left[\begin{array}{c|c} \underline{\underline{I}} & 0 \\ \hline 0 & 0 \end{array}\right] = \underline{\underline{P}}$

$\leftarrow k \rightarrow \leftarrow m-k \rightarrow$

and $\underline{\underline{B}}_1 \underline{\underline{\Phi}}(b) = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \underline{\underline{I}} \end{array}\right] = \underline{\underline{I}} - \underline{\underline{P}}$

$k \quad m-k$

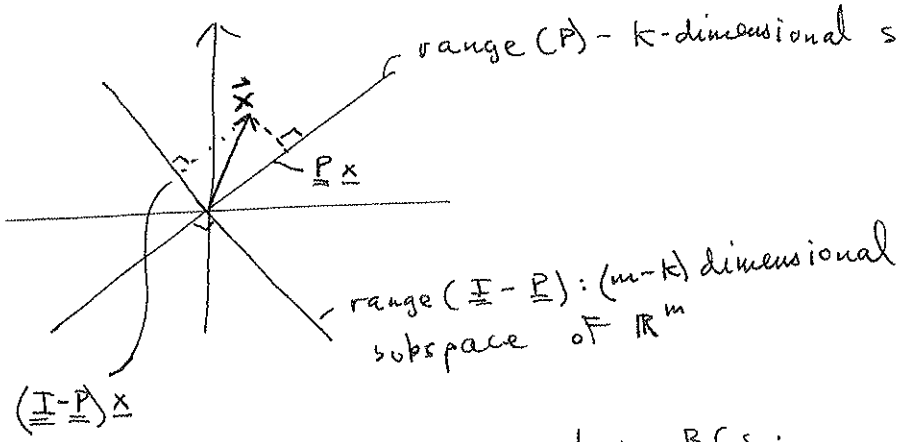
The matrix $\underline{\underline{P}}$ is an orthogonal projector:

$\Rightarrow \underline{\underline{P}}^2 = P$ (projection matrix)

$\Rightarrow \underline{\underline{P}}^T = \underline{\underline{P}}$ (orthogonal)

The matrix $\underline{\underline{P}}$ projects $\mathbb{R}^m$ into a k-dimensional subspace.

The matrix $\underline{\underline{I}} - \underline{\underline{P}}$ is also an orthogonal projector. It projects $\mathbb{R}^m$ into a $(m-k)$-dimensional subspace orthogonal to that of $\underline{\underline{P}}$.



So for the case of <u>separated BCs</u>:

$$\underline{\underline{G}}(t,\tau) = \begin{cases} \underline{\underline{\Phi}}(t)\, \underline{\underline{P}}\, \underline{\underline{\Phi}}^{-1}(\tau) & t > \tau \\ -\underline{\underline{\Phi}}(t)(\underline{\underline{I}} - \underline{\underline{P}})\, \underline{\underline{\Phi}}^{-1}(\tau) & t < \tau \end{cases}$$

For dichotomy, we want $\underline{\underline{P}}$ to decay as t- increases and $\underline{\underline{I}} - \underline{\underline{P}}$ to decay as t decreases.

If $\| \underline{\underline{\Phi}}(t)\, \underline{\underline{P}}\, \underline{\underline{\Phi}}^{-1}(\tau) \| \leq K$ for $\tau < t$ and

if $\| \underline{\underline{\Phi}}(t)(\underline{\underline{I}} - \underline{\underline{P}})\, \underline{\underline{\Phi}}^{-1}(\tau) \| < K$ for $\tau > t$ then

the boundary value problem possesses a dichotomy and is (stable).

IF $\| \underline{\underline{\Phi}}(t)\,\underline{\underline{P}}\,\underline{\underline{\Phi}}^{-1}(\tau) \| \le K e^{\alpha(\tau - t)}$, $\tau < t$

and $\| \underline{\underline{\Phi}}(t)(\underline{\underline{I}} - \underline{\underline{P}})\,\underline{\underline{\Phi}}^{-1}(\tau) \| \le K e^{\beta(t-\tau)}$, $t < \tau$

for some $K > 0$, $\alpha, \beta > 0$ then the BVP possesses an exponential dichotomy (asymptotic stability).

Example

$v'' - v = 0$, $0 \le t \le b$, $v(0) = \alpha$, $v(b) = \beta$

write as a first order system, $\underline{Y}' = \underline{\underline{A}}\,\underline{Y}$,

where $\underline{Y} = \begin{pmatrix} v \\ v' \end{pmatrix}$ and $\underline{\underline{A}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

Find fundamental solution matrix

general solution of original problem, $v(t) = c_1 \cosh t + c_2 \sinh t$

$v'(t) = c_1 \sinh t + c_2 \cosh t$

$\Rightarrow \underline{Y}(t) = c_1 \begin{pmatrix} \cosh t \\ \sinh t \end{pmatrix} + c_2 \begin{pmatrix} \sinh t \\ \cosh t \end{pmatrix}$

Pick c_1 and c_2 to solve for $\underline{\underline{X}}$ $\therefore \underline{\underline{X}} = \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix}$

satisfies $\underline{\underline{X}}(0) = \underline{\underline{I}}$

Note that $\underline{\underline{X}}(t)$ grows as t increases, which suggests the IVP is unstable, though it turns out the BVP is stable.

From BCs: $v(0) = \alpha$, $v(b) = \beta$

write as $\underline{\underline{B}}_0 \underline{Y}(0) + \underline{\underline{B}}_1 \underline{Y}(b) = \underline{\gamma}$

$$\underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}}_{\underline{\underline{B}}_0} \begin{pmatrix} v(0) \\ v'(0) \end{pmatrix} + \underbrace{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}}_{\underline{\underline{B}}_1} \begin{pmatrix} v(0) \\ v'(0) \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

$$\underline{\underline{Q}} = \underline{\underline{B}}_0 + \underline{\underline{B}}_1 \underline{\underline{X}}(b) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cosh b & \sinh b \\ \sinh b & \cosh b \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ \cosh b & \sinh b \end{pmatrix}$$

Notice $\det(\underline{\underline{Q}}) = \sinh(b) \Rightarrow \underline{\underline{Q}}$ is nonsingular.

$$\underline{\underline{\Phi}}(t) = \underline{\underline{X}}(t)\,\underline{\underline{Q}}^{-1} = \frac{1}{\sinh b}\begin{bmatrix} \sinh(b-t) & \sinh(t) \\ -\cosh(b-t) & \cosh(t) \end{bmatrix}$$

↑ this column decays as t increases

↑ this column decays as t decreases from b

Calculate $\underline{\underline{\Phi}}(t)\,\underline{\underline{P}}\,\underline{\underline{\Phi}}^{-1}(\tau) = \underline{\underline{L}}$

$$\underline{\underline{P}}\,\underline{\underline{\Phi}}^{-1}(\tau) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \cosh\tau & -\sinh\tau \\ \cosh(b-\tau) & \sinh(b-\tau) \end{bmatrix} = \begin{bmatrix} \cosh\tau & -\sinh\tau \\ 0 & 0 \end{bmatrix}$$

$$\Rightarrow \underline{\underline{L}} = \frac{1}{\sinh b}\begin{bmatrix} \sinh(b-t) & \sinh(t) \\ \cosh(b-t) & \cosh(t) \end{bmatrix}\begin{bmatrix} \cosh\tau & -\sinh\tau \\ 0 & 0 \end{bmatrix}$$

$$\underline{\underline{L}} = \frac{1}{\sinh b}\begin{bmatrix} \cosh\tau\sinh(b-t) & -\sinh\tau\sinh(b-t) \\ -\cosh\tau\cosh(b-t) & \sinh\tau\cosh(b-t) \end{bmatrix}$$

Pick a norm: Frobenius norm: $\|L\|_F^2 = \text{trace}\left(\underline{\underline{L}}\,\underline{\underline{L}}^T\right)$

$$\Rightarrow \|L\|_F^2 = \frac{\cosh 2\tau \cdot \cosh 2(b-t)}{\sinh^2 b}$$

Notice $\tau \geq 0$, $b - t \geq 0$ then $\cosh 2t = \frac{1}{2}\left(e^{2t} + e^{-2t}\right)$

$\rightarrow$ ~~$\cosh 2t = \frac{e^{2t}}{2}$~~

$$\rightarrow \cosh 2t = \frac{e^{2t}}{2}\left(1 + e^{-4t}\right),$$

$$\cosh 2(b-t) = \frac{1}{2}e^{2(b-t)}\left(1 + e^{-4(b-t)}\right),$$

$$\sinh 2b = \frac{1}{2}e^{b}\left(1 - e^{-2b}\right)$$

$$\Rightarrow \|L\|_F^2 = e^{2\tau + 2(b-t) - 2b} = e^{2(\tau - t)}.$$

$$\Rightarrow \boxed{\|L\|_F = e^{\tau - t}\sqrt{\frac{(\quad)}{(\quad)}}}$$

the $\alpha = 1$ in theory

$$K = \frac{2}{(1 - e^{-1})}$$

You also Find $\|R\|_F \leq e^{t-\tau}K$

↑ where $K = \frac{2}{(1-e^{-1})}$

Stiff BVP

For an asymptotically stable BVP (ie one that possesses an exponential dichotomy) solution modes decay in from the boundaries depending on the number of BCs on each side.

If the eqn has constant coeffs then the rate of decay is determined by the eigenvalues of the coefficient matrix.

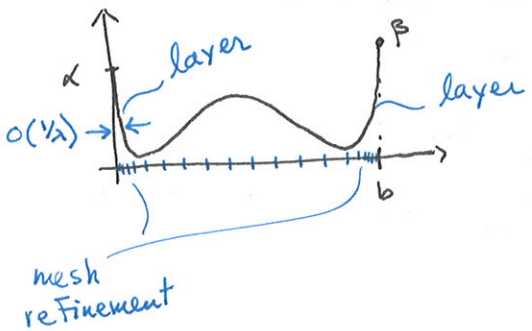
Suppose λ_j, $j = 1, \dots, m$ are these eigenvalues. The BVP would be stiff if the real part of λ_j times b is large For any j:

$$|b \cdot \mathrm{Re}(\lambda_j)| \gg 1 \quad \text{for any } j$$

Previously, for an IVP integrating for t increasing, the problem is stiff if $-(\mathrm{Re}\,\lambda_j)b \gg 1$. This differs from a BVP where both signs of $\mathrm{Re}\,\lambda_j$ must be considered.

For IVPs, the preferred methods had stiff decay (eg backward Euler, BDF, ...). There are no known methods for BVPs akin to this "stiff" decay because modes decay in both directions.

For BVPs, stiffness is handled numerically by resolving layers using some form of grid adaptation.



shooting methods

BVP: $\underline{y}' = \underline{f}(t, \underline{y})$, $0 \le t \le b$, $\underline{g}(\underline{y}(0), \underline{y}(b)) = \underline{0}$

construct solution by solving associated IVP
(the IVP has same diff eqn)

IVP: $\underline{y}' = \underline{f}(t, \underline{y})$
$\underline{y}(0) = \underline{c}$ $\rightarrow$ $\underline{y}(t; \underline{c})$

Need $\underline{c}$ s.t. $\underline{g}(\underline{c}, \underline{y}(b; \underline{c})) = 0$ $\Rightarrow$ root-finding problem

use Newton's method

$\underline{c}^{NEW} = \underline{c}^{OLD} - \underline{\Delta c}$ ← correction

how to find correction?

4/8/03

Shooting Methods for BVPs

General 2-point BVP: $\underline{y}' = \underline{f}(t, \underline{y})$, $0 \le t \le b$

with boundary conditions: $\underline{g}(\underline{y}(0), \underline{y}(b)) = 0$

The idea behind shooting is to solve an associated initial value problem for some chosen initial condition and then iterate on the choice of the initial condition until BCs are satisfied.

Associated IVP: (same diff eq)

$$\underline{y}' = \underline{f}(t, \underline{y}) \quad , 0 \le t \le b \quad , \quad \underline{y}(0) = \underline{c}$$

The solution of the IVP is $\underline{y}(t, \underline{c})$. Want to find $\underline{c}$ st $\underline{g}(\underline{c}, \underline{y}(b, \underline{c})) = 0$

$\Rightarrow$ Root-finding problem for $\underline{c}$

solve the Root-finding problem using Newton:

define $\underline{F}(\underline{c}) \equiv \underline{g}(\underline{c}, \underline{y}(b, \underline{c})) = 0$

We need Jacobian matrix of $\underline{F}$

(shooting)

For $\underline{g} = \underline{g}(\underline{v}, \underline{\nu})$

$$\frac{\partial \underline{F}}{\partial \underline{c}} = \frac{\partial \underline{g}}{\partial \underline{v}} \frac{\partial \underline{v}}{\partial \underline{c}} + \frac{\partial \underline{g}}{\partial \underline{\nu}} \frac{\partial \underline{\nu}}{\partial \underline{c}}$$

let $\underline{v} = \underline{c}$, $\underline{\nu} = \underline{y}(b, \underline{c})$

$$= \frac{\partial \underline{g}}{\partial \underline{v}} \underline{\underline{I}} + \frac{\partial \underline{g}}{\partial \underline{\nu}} \frac{\partial \underline{y}}{\partial \underline{c}}$$

need to compute this derivative, which becomes a variational problem

to compute $\frac{\partial \underline{y}}{\partial \underline{c}}$ we have

$$\frac{\partial \underline{y}}{\partial t} = \underline{f}(t, \underline{y}(t, \underline{c}))$$

Differentiate wrt $\underline{c}$

$$\frac{\partial}{\partial \underline{c}}\left(\frac{\partial \underline{y}}{\partial t}\right) = \frac{\partial}{\partial \underline{c}}\left(\underline{f}(t, \underline{y}(t, \underline{c}))\right)$$

$$\rightarrow \frac{\partial}{\partial t}\left(\frac{\partial \underline{y}}{\partial \underline{c}}\right) = \frac{\partial \underline{f}}{\partial \underline{y}}(t, \underline{y}) \cdot \frac{\partial \underline{y}}{\partial \underline{c}}$$

$\frac{\partial \underline{y}}{\partial \underline{c}} \rightarrow \underline{\underline{z}}$, $\frac{\partial \underline{y}}{\partial \underline{c}} \rightarrow \underline{\underline{z}}$

$$\rightarrow \boxed{\underline{\underline{z}}' = \underline{\underline{A}}(t, \underline{y})\,\underline{\underline{z}}}, \quad \underline{\underline{A}} = \frac{\partial \underline{f}}{\partial \underline{y}}$$

with IC: $\underline{y}(0, \underline{c}) = \underline{c}$

$$\downarrow$$

$$\frac{\partial \underline{y}}{\partial \underline{c}}(0, \underline{c}) = \underline{\underline{I}}$$

$$\downarrow$$

$$\underline{\underline{z}}(0) = \underline{\underline{I}}$$

$\Rightarrow$ Variational Problem

$$\underline{\underline{z}}' = \underline{\underline{A}}\,\underline{\underline{z}}, \quad \underline{\underline{z}}(0) = \underline{\underline{I}}, \quad \underline{\underline{A}} = \frac{\partial \underline{f}}{\partial \underline{y}},$$

Algorithm

1) Choose $\underline{c}^{(0)}$

2) for $k = 0, 1, \ldots$

3) Solve the IVPs

$\underline{A}$ is computed analytically

$$y' = F(t, y), \quad y(0) = c^{(k)}$$

$$\underline{z}_j' = \underline{\underline{A}}(t, y)\,\underline{z}_j, \quad \underline{z}_j(0) = \underline{e}_j, \quad j = 1, \ldots, m$$

$m+1$ IVPs, solve numerically

where $\underline{\underline{Z}} = \begin{bmatrix} | & | & & | \\ \underline{z}_1 & \underline{z}_2 & \cdots & \underline{z}_m \\ | & | & & | \end{bmatrix}$

4) Compute $\underline{F}^{(k)} = \underline{g}(\underline{c}^{(k)}, \underline{y}(b))$

5) compute

$$\underline{\underline{J}}^{(k)} = \frac{\partial \underline{g}}{\partial \underline{u}}(\underline{c}^{(k)}, \underline{y}(b)) + \frac{\partial \underline{g}}{\partial \underline{v}}(\underline{c}^{(k)}, \underline{y}(b))\,\underline{\underline{Z}}(b)$$

6) solve the $m \times m$ linear system

$$\underline{\underline{J}}^{(k)}\,\Delta \underline{c}^{(k)} = \underline{F}^{(k)}$$

7) update $\underline{c}^{(k+1)} = \underline{c}^{(k)} - \Delta \underline{c}^{(k)}$

8) stop if $\| \Delta \underline{c}^{(k)} \| \leq tol$

Consider a specific case:

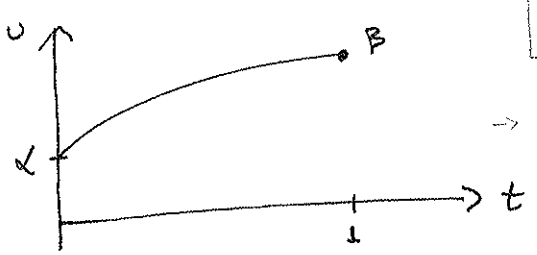
$u'' - G(u) = 0, \quad 0 \leq t \leq 1, \quad u(0) = \alpha, \quad u(1) = \beta$

Formulate and solve using shooting!

Associated IVP for a system of 1st order ODEs:

$$\underline{y} = \begin{pmatrix} u \\ u' \end{pmatrix} \rightarrow \underline{y}' = \begin{pmatrix} y_2 \\ G(y_1) \end{pmatrix}, \quad 0 \leq t \leq 1$$

pick initial condition: $\underline{y}(0) = \begin{pmatrix} \alpha \\ c \end{pmatrix}$



$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{pmatrix} u(0) \\ u'(0) \end{pmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{pmatrix} u(1) \\ v(1) \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

$$\rightarrow g(\underline{y}(0), \underline{y}(1)) = \begin{pmatrix} u(0) - \alpha \\ v(1) - \beta \end{pmatrix}$$

there's only one parameter, so it turns into a scalar problem:

$$F(c) = Y_1(1, c) - \beta = 0$$

to calculate, integrate out to $t = 1$ to obtain $Y_1(1, c)$, then subtract Beta

$$\text{Newton}: \quad c^{(k+1)} = c^{(k)} - \frac{F(c^{(k)})}{F'(c^{(k)})}$$

Now calculate $F'(c)$

we have $y' = \begin{pmatrix} y_2 \\ G(y_1) \end{pmatrix}$, $y(0) = \begin{pmatrix} \alpha \\ c \end{pmatrix}$

$\underline{z} \equiv \frac{\partial \underline{y}}{\partial c}$, Differentiate wrt c

$$\rightarrow \quad \frac{\partial \underline{z}}{\partial t} = \begin{pmatrix} z_2 \\ G'(y_1)\, z_1 \end{pmatrix}, \quad \underline{z}(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

} solve IVP numerically

$$\Rightarrow \quad c^{(k+1)} = c^{(k)} - \frac{Y_1\left(1, c^{(k)}\right) - \beta}{z_1\left(1, c^{(k)}\right)}$$

Example: $\quad u'' - 9u = 0, \quad 0 \le t \le 1$

$$u(0) = 0, \quad u(1) = \sinh 3$$

this is a linear BVP whose solution is $u(t) = \sinh(3t)$

Solve using shooting:

set $\quad \underline{w} = \begin{pmatrix} Y_1 \\ Y_2 \\ z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} u \\ u' \\ \partial u/\partial c \\ \partial u'/\partial c \end{pmatrix}$

$$\rightarrow \quad \underline{w}' = \begin{pmatrix} w_2 \\ 9w_1 \\ w_4 \\ 9w_3 \end{pmatrix}, \quad \underline{w}(0) = \begin{pmatrix} \alpha \\ c \\ 0 \\ 1 \end{pmatrix}$$

$$c^{NEW} = c - \frac{w_1(1) - \sinh 3}{w_3(1)}$$

Example

$$u'' - e^{u+1} = 0, \quad 0 \le t \le 1, \quad u(0) = u(1) = 0$$

"Bratu" problem

Here $G(u) = e^{u+1}$

$$\underline{w}' = \begin{pmatrix} w_2 \\ e^{w_1+1} \\ w_4 \\ w_3 e^{w_1+1} \end{pmatrix}, \quad \underline{w}(0) = \begin{pmatrix} 0 \\ c \\ 0 \\ 1 \end{pmatrix}$$

Difficulties

The main difficulty with shooting involves growing "modes" present in a stable BVP. If the modes grow too quickly then the linear system to determine the correction $\underline{\Delta c}$ become ill-conditioned.

BVP: $\underline{y}' = \underline{f}(t, \underline{y}), \quad 0 \le t \le b, \quad \underline{g}(\underline{y}(0), \underline{y}(b)) = 0$

For shooting, we solve some IVPs and a variational problem

$$\underline{y}' = \underline{f}(t, \underline{y}), \quad 0 \le t \le b, \quad \underline{y}(0) = \underline{c}$$

$$\underline{\underline{Z}}' = \underline{\underline{A}}\,\underline{\underline{Z}}, \quad \underline{\underline{Z}}(0) = \underline{\underline{I}}, \quad \underline{\underline{A}} = \frac{\partial \underline{f}}{\partial \underline{y}}$$

Solve a linear system of the form $\underline{\underline{Q}}\,\underline{\Delta c} = \underline{g}$

where $\underline{\Delta c}$ is the correction to $\underline{c}$ and $\underline{\underline{Q}} = \frac{\partial \underline{g}}{\partial \underline{v}} + \frac{\partial \underline{g}}{\partial \underline{w}} \cdot \underline{\underline{Z}}(b)$

If a solution mode (one of the columns of $\underline{\underline{Z}}$) grows too quickly as t increases, then $\underline{\underline{Q}}$ becomes ill-conditioned.

Example (pg 179 text)

$$u''' - 2\lambda u'' - \lambda^2 u' + 2\lambda^3 u = g(t), \quad 0 \le t \le 1$$

linear constant coeff ODE with nonhomogeneous term on right,

with BCs: $u(0) = \alpha, \quad u(1) = \beta, \quad u'(1) = \gamma$

First examine the homogeneous problem

substitute $v = e^{rt}$

$\Rightarrow r^3 - 2\lambda r^2 - \lambda^2 r + 2\lambda^3 = 0$ (char polynomial for homogeneous ODE)

$\rightarrow (r+\lambda)(r-\lambda)(r-2\lambda) = 0$

general solution: $v_h(t) = c_1 e^{-\lambda t} + c_2 e^{\lambda t} + c_3 e^{2\lambda t}$, homogeneous solution

$\rightarrow v(t) = v_h(t) + v_p(t)$

$v(t) = c_1 e^{-\lambda t} + c_2 e^{\lambda t} + c_3 e^{2\lambda t} + v_p(t)$

one mode that ~~decreases~~ decays for increasing t and two modes that grow as t increases.

$\Rightarrow$ this suggests that IVP is unstable but the BVP would be nicely stable. If λ gets too big then shooting will become difficult.

The text chooses $g(t) = (\lambda^2 + \pi^2)(\pi \sin \pi t + 2\lambda \cos \pi t)$

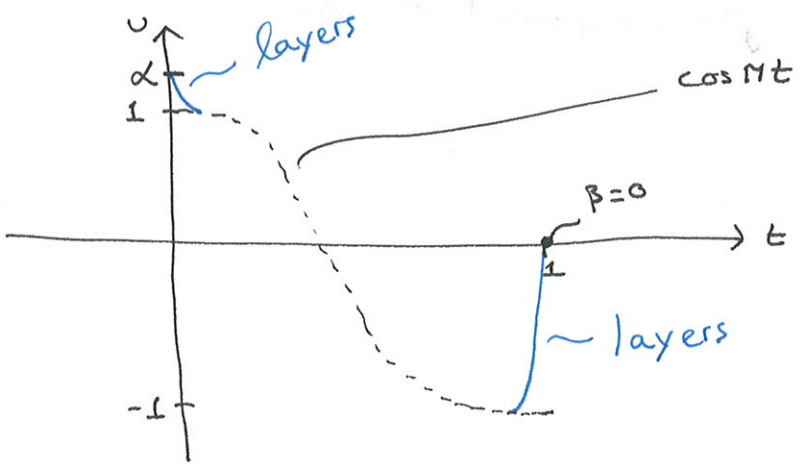
$$\alpha = \frac{3 + 2e^{-\lambda} + e^{-2\lambda}}{2 + e^{-\lambda}}$$

$$\beta = 0$$

$$\gamma = \lambda\left(\frac{3 - e^{-\lambda}}{2 + e^{-\lambda}}\right)$$

For any λ, with $g(t)$, β, α, γ as specified above, the exact solution is:

$$v(t) = \frac{1}{2+e^{-\lambda}}\left(e^{-\lambda t} + e^{\lambda(t-1)} + e^{2\lambda(t-1)}\right) + \cos \pi t$$



the BCs force layers, depending on λ

Solve using shooting

IVP: $\underline{y}' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2\lambda^3 & \lambda^2 & 2\lambda \end{bmatrix} \underline{y} + \begin{pmatrix} 0 \\ 0 \\ g(t) \end{pmatrix} \equiv \underline{\underline{A}}$, where $\underline{y} = \begin{pmatrix} v \\ v' \\ v'' \end{pmatrix}$

with $\underline{y}(0) = \begin{pmatrix} \alpha \\ c_1 \\ c_2 \end{pmatrix}$ where c_1 and c_2 are shooting parameters

variational problems: $\underline{z}_1 = \frac{\partial \underline{y}}{\partial c_1}$, $\underline{z}_2 = \frac{\partial \underline{y}}{\partial c_2}$

$\rightarrow \underline{z}_j' = \underline{\underline{A}}\,\underline{z}_j$, $j = 1, 2$

$\rightarrow \underline{z}_j(0) = \begin{cases} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} & \text{if } j = 1 \\ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} & \text{if } j = 2 \end{cases}$

So there are 3 ODEs to solve, one ODE for y, one for z_1 and for z_2.

to solve, create "super-system"

$\underline{w} = \begin{pmatrix} \underline{y} \\ \underline{z}_1 \\ \underline{z}_2 \end{pmatrix} \Rightarrow \underline{w}' = \begin{bmatrix} \underline{\underline{A}} & & \\ & \underline{\underline{A}} & \\ & & \underline{\underline{A}} \end{bmatrix} \underline{w}$, $\underline{w}(0) = \begin{pmatrix} \alpha \\ c_1 \\ c_2 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

(don't have to create "super-system" cause ODEs are decoupled - just a matter of style)

Eqns to satisfy are

$y_1(1) - \beta = 0$, $y_2(1) - \gamma = 0$ (these depend on c_1, c_2)

Linear system

$$\begin{bmatrix} z_{11}(1) & z_{21}(1) \\ z_{12}(1) & z_{22}(1) \end{bmatrix} \begin{pmatrix} \Delta c_1 \\ \Delta c_2 \end{pmatrix} = \begin{pmatrix} y_1(1) - \beta \\ y_2(1) - \gamma \end{pmatrix}$$

where $z_{ij}(1)$ is j^{th} component of $\underline{z}_i$ at $t = 1$

BVP: $\underline{y}' = \underline{f}(t, \underline{y})$, $0 \le t \le b$, $g(\underline{y}(0), \underline{y}(b)) = 0$

associated IVP (same diff eq, same interval)

$\underline{Y}' = \underline{f}(t, \underline{Y})$, $0 \le t \le b$, $\underline{Y}(0) = \underline{c}$

variational problem $\underline{\underline{Z}}' = \underline{\underline{A}}(t)\,\underline{\underline{Z}}$, $\underline{\underline{Z}}(0) = \underline{\underline{I}}$, $\underline{\underline{A}}(t) = \frac{\partial \underline{f}}{\partial \underline{y}}(t, \underline{y}(t))$

solve $\underline{\underline{Q}}\,\Delta\underline{c} = \underline{g}$, where $\underline{g}(\underline{c}, \underline{Y}(b))$ $\qquad \underline{\underline{Z}}(t) = \frac{\partial \underline{Y}}{\partial \underline{c}}(t)$

where $\underline{\underline{Q}} = \underline{g}_{\underline{v}}\,\underline{\underline{Z}}(b)$

then $\underline{c}^{NEW} = \underline{c}^{OLD} - \Delta\underline{c}$

Main Difficulty: occurs if this linear system is not solved accurately

---

Example

BVP $v''' - 2\lambda v'' - \lambda^2 v' + 2\lambda^3 v = g(t)$, $0 \le t \le 1$

$v(0) = \alpha$, $v(1) = \beta$, $v'(1) = \gamma$

Pick $\alpha, \beta, \gamma, g(t)$ such that exact solution is

$$v(t) = \frac{e^{-\lambda t} + e^{\lambda(t-1)} + e^{2\lambda(t-1)}}{2 + e^{-\lambda}} + \cos \pi t$$

Notice, two growing modes in solution

associated IVP

$\underline{Y}' = \underline{\underline{A}}\,\underline{Y} + \underline{q}(t)$ where $\underline{Y}(t) = \begin{pmatrix} v \\ v' \\ v'' \end{pmatrix}$, $\underline{\underline{A}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2\lambda^3 & \lambda^2 & 2\lambda \end{bmatrix}$, $\underline{q}(t) = \begin{pmatrix} 0 \\ 0 \\ g(t) \end{pmatrix}$

$\underline{Y}(0) = \begin{pmatrix} \alpha \\ c_1 \\ c_2 \end{pmatrix}$

ASIDE

Suppose $v'' - v = 0$, $v(0) = \alpha$, $v(1) = \beta$

then $\underline{Y}' = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \underline{Y}$, $\underline{g}(\underline{Y}(0), \underline{Y}(1)) = \begin{pmatrix} Y_1(0) - \alpha \\ Y_1(1) - \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\frac{\partial \underline{g}}{\partial \underline{v}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\frac{\partial \underline{g}}{\partial \underline{v}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$

Require c_1 and c_2 s.t.

$y_1(1) = \beta$, $y_2(1) = \gamma$

Iterate on (c_1, c_2) using Newton's method

The exact solution tells us there are two growing modes. In this case it is better to shoot in the reverse direction.

$$\underline{y}' = \underline{\underline{A}}\,\underline{y} + \underline{q}(t), \quad 1 \geq t \geq 0$$

$$\underline{y}(1) = \begin{pmatrix} \beta \\ \gamma \\ c \end{pmatrix}$$

solve for some choice of c:

Let $\underline{z}(t) = \frac{\partial \underline{y}}{\partial c} \rightarrow \underline{z}' = \underline{\underline{A}}\,\underline{z}, \ \underline{z}(1) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \ 1 \geq t \geq 0$

want c s.t. $y_1(0) = \alpha$

scalar root finding problem: $F(c) = y_1(0, c) - \alpha = 0$

$F'(c) = z_1(0, c)$

$$\rightarrow \quad c^{NEW} = c^{OLD} - \frac{F(c^{OLD})}{F'(c^{OLD})}$$

Multiple Shooting

"Simple" shooting becomes ineffective if b is too big or $\mathrm{Re}(\text{eigenvalues of } \underline{\underline{J}}_{\underline{y}})$ is too large. Rough estimate of problem is $b\|\underline{\underline{J}}_{\underline{y}}\|$ is too big. (Note, $b\|\underline{\underline{J}}_{\underline{y}}\|$ is dimensionless.)
A fix is to use multiple shooting.

Consider a linear BVP:

$$\underline{y}' = \underline{\underline{A}}(t)\,\underline{y} + \underline{q}(t), \quad 0 \leq t \leq b, \quad \underline{\underline{B}}_0\,\underline{y}(0) + \underline{\underline{B}}_1\,\underline{y}(b) = \underline{\gamma}$$

Introduce a partition of the interval:

$$0 = t_0 < t_1 < t_2 < \cdots < t_N = b$$

Let's assume the partition is known.

Idea: solve an IVP for each subinterval.

For $t_{n-1} \le t \le t_n$, define

$$\underline{\underline{Y}}_n'(t) = \underline{\underline{A}}(t)\,\underline{\underline{Y}}_n(t), \quad \underline{\underline{Y}}_n(t_{n-1}) = \underline{\underline{I}}$$ } homogeneous solution

this is a Fundamental solution matrix for IVP defined on subinterval

with

$$\underline{v}_n'(t) = \underline{\underline{A}}(t)\,\underline{v}_n(t) + \underline{q}(t), \quad \underline{v}_n(t_{n-1}) = 0$$ } particular solution

The general solution for the subinterval $t_{n-1} \le t \le t_n$ is

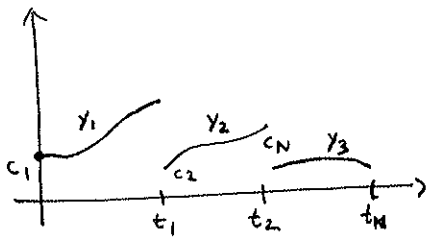
$$\underline{y}_n(t) = \underbrace{\underline{\underline{Y}}_n(t)\,\underline{c}_n}_{\text{homogeneous}} + \underbrace{\underline{v}_n(t)}_{\text{particular}}$$

Overall

$$\underline{y}(t) = \begin{cases} y_1(t), & t_0 \le t \le t_1 \\ y_2(t), & t_1 \le t \le t_2 \\ \vdots \\ y_N(t), & t_{N-1} \le t \le t_N \end{cases}$$

this is a piecewise smooth solution, not necessarily continuous

For instance



Choose $\underline{c}$'s so that $\underline{y}(t)$ is continuous and satisfies the boundary conditions.

For continuity: $\underline{y}_n(t_n) = \underline{y}_{n+1}(t_n)$ for $n = 1, 2, \ldots, N-1$ (interior nodes)

$$\to \underline{y}_n(t_n) = \underline{\underline{Y}}_n(t_n)\,\underline{c}_n + \underline{v}_n(t_n) = \underline{\underline{Y}}_{n+1}(t_n)\,\underline{c}_{n+1}$$

$$= \underline{y}_{n+1}(t_n) = \underline{c}_{n+1}$$

$$\Rightarrow \boxed{-\underline{\underline{Y}}_n(t_n)\,\underline{c}_n + \underline{c}_{n+1} = \underline{v}_n(t_n)}$$

Boundary Conditions:

$$\underline{\underline{B}}_0 \underline{Y}_1(0) + \underline{\underline{B}}_1 \underline{Y}(b) = \underline{\gamma}$$

$$\underline{Y}(b) = \underline{\underline{Y}}_N(b)\underline{c}_N + \underline{Y}_N(b)$$

$$\rightarrow \boxed{\underline{\underline{B}}_0 \underline{c}_1 + \underline{\underline{B}}_1 \underline{\underline{Y}}_N(b)\underline{c}_N = \underline{\gamma} - \underline{\underline{B}}_1 \underline{Y}_N(b)} \quad m$$

unknowns: N $\underline{c}$'s $\Rightarrow$ Nm unknowns

equations: $(N-1)m$ (From continuity)
m (From BCs)

total: Nm equations, Nm unknowns

Collect the Nm linear eqns for $\underline{c}$'s

$$\underline{c} = \begin{bmatrix} \underline{c}_1 \\ \underline{c}_2 \\ \vdots \\ \underline{c}_N \end{bmatrix} \in \mathbb{R}^{mN} \Rightarrow \begin{bmatrix} -\underline{\underline{Y}}_1(t_1) & \underline{\underline{I}} & & \\ & -\underline{\underline{Y}}_2(t_2) & \underline{\underline{I}} & \\ & & \ddots & \ddots \\ & & & -\underline{\underline{Y}}_{N-1}(t_{N-1}) \quad \underline{\underline{I}} \\ \underline{\underline{B}}_0 & & & \underline{\underline{B}}_1 \underline{\underline{Y}}_N(b) \end{bmatrix} \underline{c} = \begin{bmatrix} \underline{v}_1(t_1) \\ \underline{v}_2(t_2) \\ \vdots \\ \underline{v}_{N-1}(t_{N-2}) \\ \underline{\gamma} - \underline{\underline{B}}_1 \underline{v}_N(b) \end{bmatrix}$$

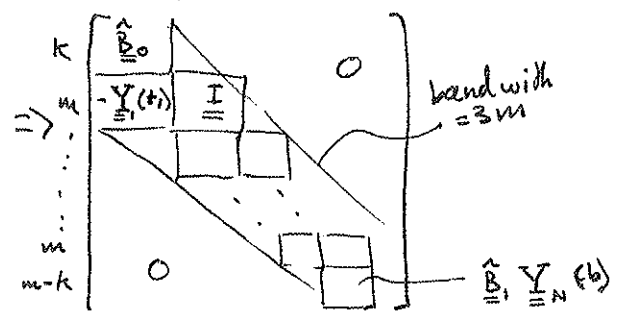
Solve the linear system for $\underline{c}$, operation count $= O(m^3N^3)$. The system is sparse so that the operation count can be reduced.

For example, if the BCs are separated then

$$\underline{\underline{B}}_0 = \begin{bmatrix} \hat{\underline{\underline{B}}}_0 \\ \hline 0 \end{bmatrix} \begin{matrix} \} k \\ \} m-k \end{matrix}, \quad \underline{\underline{B}}_1 = \begin{bmatrix} 0 \\ \hline \hat{\underline{\underline{B}}}_1 \end{bmatrix} \begin{matrix} \} k \\ \} m-k \end{matrix}$$

Move the k eqns associated with k BCs at $t=0$ to the top of the linear system.

$\hat{\underline{\underline{Q}}} \equiv$



Now genuinely banded so solve using Gaussian elimination for a banded matrix (with pivoting)

operation count $\approx O(mN \cdot m^2)$ — size of system times sea band with squared

(omit 9 due to order of mag $(\text{bw})^2 = (3m)^2 = 9m^2$)

Remarks

1) For a given partition, each initial value problem is independent, so that these could be solved in parallel. ⇒ parallel shooting
2) May want to choose the nodes adaptively based on the growth of the solutions ⇒ IVP are no longer independent
3) The matrix $\hat{\underline{\underline{Q}}}$ is better conditioned than $\underline{\underline{Q}}$ if the partition is fine enough.
4) IF the problem is nonlinear then a linearization leads to a system similar to $\hat{\underline{\underline{Q}}}$.

Finite Difference Methods

4/14/03

BVP: $\underline{y}' = \underline{f}(t, \underline{y})$, $0 \le t \le b$, $\underline{g}(\underline{y}(0), \underline{y}(b)) = 0$

Introduce a mesh or grid, $0 = t_0 < t_1 < \cdots < t_n = b$

Set $h_n = t_n - t_{n-1} > 0$, $n = 1, \ldots, N$

- Choose some approximation
  - Natural choice is a one-step method, symmetric, such as midpoint method or its higher order extensions (Gauss method); or trapezoidal method or its higher order extensions (Labato methods).
  - There is no essential distinction between explicit versus implicit methods here because the plan is to solve for $\underline{y}$ of all grid points together.

4/14/03

Midpoint Method

set $\underline{y}_n = \underline{y}(t_n)$, $n = 0, \ldots, N$

Let $\underline{y}_n$ solve

$$\begin{cases} \underline{y}_n = \underline{y}_{n-1} + h_n \underline{f}\left(t_{n-1/2}, \frac{1}{2}(\underline{y}_n + \underline{y}_{n-1})\right) & , n = 1, 2, \ldots, N \\ \text{subject to} \quad \underline{g}(\underline{y}_0, \underline{y}_N) = 0 \end{cases}$$

$\Rightarrow$ set of $m(N+1)$ equations for $m(N+1)$ unknowns, $\{\underline{y}_n\}_{n=0}^{N}$

• Consider the linear case

$$\underline{f}(t, \underline{y}) = \underline{\underline{A}}(t)\,\underline{y} + \underline{q}(t)$$

$$\underline{g}(\underline{u}, \underline{v}) = \underline{\underline{B}}_0 \underline{u} + \underline{\underline{B}}_1 \underline{v} - \underline{\gamma}$$

Have $\begin{cases} \underline{y}_n = \underline{y}_{n-1} + h_n \left\{ \underline{\underline{A}}(t_{n-1/2}) \frac{1}{2}(\underline{y}_{n-1} + \underline{y}_n) + \underline{q}(t_{n-1/2}) \right\} & , n = 1, 2, \ldots, N \\ \underline{\underline{B}}_0 \underline{y}_0 + \underline{\underline{B}}_1 \underline{y}_N = \underline{\gamma} \end{cases}$

($\underline{\underline{A}}(t_{n-1/2}) \to \underline{\underline{A}}_{n-1/2}$, $\underline{q}(t_{n-1/2}) \to \underline{q}_{n-1/2}$)

collect as a linear system, rewrite as:

$$\left(\underline{\underline{I}} - \frac{h_n}{2}\underline{\underline{A}}_{n-1/2}\right)\underline{y}_n - \left(\underline{\underline{I}} + \frac{h_n}{2}\underline{\underline{A}}_{n-1/2}\right)\underline{y}_{n-1} = h_n \underline{q}_{n-1/2}$$

Set $\underline{\underline{D}}_n = -\underline{\underline{I}} - \frac{h_n}{2}\underline{\underline{A}}_{n-1/2}$, $\underline{\underline{U}}_n = \underline{\underline{I}} - \frac{h_n}{2}\underline{\underline{A}}_{n-1/2}$

$\to$ $\underline{\underline{U}}_n \underline{y}_n + \underline{\underline{D}}_n \underline{y}_{n-1} = h_n \underline{q}_{n-1/2}$

$$\begin{bmatrix} \underline{\underline{D}}_1 & \underline{\underline{U}}_1 & & & \\ & \underline{\underline{D}}_2 & \underline{\underline{U}}_2 & & \\ & & & & \\ & & & \underline{\underline{D}}_N & \underline{\underline{U}}_N \\ \underline{\underline{B}}_0 & & & & \underline{\underline{B}}_1 \end{bmatrix} \begin{pmatrix} \underline{y}_0 \\ \underline{y}_1 \\ \vdots \\ \underline{y}_{N-1} \\ \underline{y}_N \end{pmatrix} = \begin{pmatrix} h_1 \underline{q}_{1/2} \\ h_2 \underline{q}_{3/2} \\ \vdots \\ h_n \underline{q}_{N-1/2} \\ \underline{\gamma} \end{pmatrix}$$

If the boundary conditions are separated then

$$\underline{\underline{B}}_0 = \begin{bmatrix} \hat{\underline{\underline{B}}}_0 \\ \hline 0 \end{bmatrix} \begin{matrix} \} k \\ \} m-k \end{matrix}, \quad \underline{\underline{B}}_1 = \begin{bmatrix} 0 \\ \hline \hat{\underline{\underline{B}}}_1 \end{bmatrix} \begin{matrix} \} k \\ \} m-k \end{matrix}$$

perform manipulation to obtain banded matrix.

Now have

$$\begin{bmatrix} \hat{\underline{\underline{B}}}_0 & & & & \\ \underline{\underline{D}}_1 & \underline{\underline{U}}_1 & & & \\ & \underline{\underline{D}}_2 & \underline{\underline{U}}_2 & & \\ & & \ddots & \ddots & \\ & & & \underline{\underline{D}}_N & \underline{\underline{U}}_N \\ & & & & \hat{\underline{\underline{B}}}_1 \end{bmatrix} \cdot \begin{pmatrix} \underline{Y}_0 \\ \underline{Y}_1 \\ \underline{Y}_2 \\ \vdots \\ \underline{Y}_N \end{pmatrix} = \begin{pmatrix} \underline{\gamma}_0 \\ h_1 \underline{g}_{1/2} \\ h_2 \underline{g}_{3/2} \\ \vdots \\ h_N \underline{g}_{N-1/2} \\ \underline{\gamma}_1 \end{pmatrix}$$

BANDED MATRIX

(same structure as that for multishoot method, pg 60, back)

- Using direct banded solver, the operation count is
  $O(\text{size of system} \times \text{bandwidth}^2) \Rightarrow O(Nm^3)$
- tends to be a larger system than multishoot but don't have to integrate to get individual blocks

Example

$u'' - qu = 0$, $0 \le t \le 1$, $u(0) = 0$, $u(1) = \sinh 3$, $u_{exact}(t) = \sinh(3t)$

let $\underline{Y} = \begin{pmatrix} u \\ u' \end{pmatrix} \Rightarrow \underline{Y}' = \underline{\underline{A}}\,\underline{Y}$, $\underline{\underline{A}} = \begin{bmatrix} 0 & 1 \\ q & 0 \end{bmatrix}$

$\hat{\underline{\underline{B}}}_0 = \hat{\underline{\underline{B}}}_1 = [1\ 0]$, $\underline{Y}_0 = 0$, $\underline{\gamma}_1 = \sinh(3)$

→ see Matlab code, my linear BVP.m

if have $\underline{\underline{B}}_0 = \begin{bmatrix} \hat{\underline{\underline{B}}}_0 \\ \hline 0 \end{bmatrix}$ and $\underline{\underline{B}}_1 = \begin{bmatrix} 0 \\ \hline \hat{\underline{\underline{B}}}_1 \end{bmatrix}$

then define $\underline{\underline{B}} = \begin{bmatrix} \hat{\underline{\underline{B}}}_0 \\ \hline \hat{\underline{\underline{B}}}_1 \end{bmatrix}$ for storage

use reshape function in Matlab

Example- stiffness

$$v''' - 2\lambda v'' - \lambda^2 v' + 2\lambda^3 v = g(t), \quad 0 \le t \le 1$$

$$v(0) = \alpha, \quad v(1) = \beta, \quad v'(1) = \gamma$$

with $g(t) = (\lambda^2 + \pi^2)(\pi \sin \pi t + 2\lambda \cos \pi t)$

$$\alpha = \frac{3 + 2e^{-\lambda} + e^{-2\lambda}}{2 + e^{-\lambda}}, \quad \beta = 0, \quad \gamma = \frac{\lambda(3 - e^{-\lambda})}{2 + e^{-\lambda}}$$

The exact solution is $v_{exact}(t) = \dfrac{e^{-\lambda t} + e^{\lambda(t-1)} + e^{2\lambda(t-1)}}{2 + e^{-\lambda}} + \cos(\pi t)$

The system becomes, $\underline{y}'(t) = \underline{\underline{A}}\,\underline{y} + \underline{q}(t)$

$$\underline{y} = \begin{pmatrix} v \\ v' \\ v'' \end{pmatrix}, \quad \underline{\underline{A}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2\lambda^3 & \lambda^2 & 2\lambda \end{bmatrix}, \quad \underline{q}(t) = \begin{bmatrix} 0 \\ 0 \\ g(t) \end{bmatrix}, \quad \underline{\underline{B}} = \begin{bmatrix} \hat{\underline{\underline{B}}}_0 \\ \hat{\underline{\underline{B}}}_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\underline{\gamma} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}$$

Recall we encountered problems shooting in either direction

can play the "mesh game"

Nonlinear BVPs

consider nonlinear BVP, but with linear separated BCs

$$\underline{y}' = \underline{F}(t, \underline{y}), \quad a \le t \le b$$

$$\begin{bmatrix} \hat{\underline{\underline{B}}}_0 \\ 0 \end{bmatrix} \underline{y}(a) + \begin{bmatrix} 0 \\ \hat{\underline{\underline{B}}}_1 \end{bmatrix} \underline{y}(b) = \begin{bmatrix} \underline{\gamma}_0 \\ \underline{\gamma}_1 \end{bmatrix}$$

with mesh: $a = t_0 < t_1 < \cdots < t_N = b$

using midpoint method

$$\underline{y}_n = \underline{y}_{n-1} + h_n \underline{F}\left(t_{n-1/2}, \frac{1}{2}(\underline{y}_n + \underline{y}_{n-1})\right), \quad n = 1, \cdots, N$$

$$\hat{\underline{\underline{B}}}_0 \underline{y}_0 = \underline{\gamma}_0, \quad \hat{\underline{\underline{B}}}_1 \underline{y}_N = \underline{\gamma}_1$$

Set $\quad \underline{y}_{n-1} = \tilde{\underline{y}}_{n-1} + \Delta \underline{y}_{n-1}$

$\underline{y}_n = \tilde{\underline{y}}_n + \Delta \underline{y}_n$

where $\tilde{\underline{y}}_j$ are approximations to discrete solution, $\Delta \underline{y}_j$ are corrections

substitute into equations

$$\Rightarrow \quad \tilde{\underline{y}}_n + \Delta \underline{y}_n = \tilde{\underline{y}}_{n-1} + \Delta \underline{y}_{n-1} + h_n \underline{F}\left(t_{n-1/2}, \frac{1}{2}\left(\tilde{\underline{y}}_{n-1} + \tilde{\underline{y}}_n\right) + \frac{1}{2}\left(\Delta \underline{y}_{n-1} + \Delta \underline{y}_n\right)\right)$$

Expand for small Δy s:

$$\tilde{\underline{y}}_n + \Delta \underline{y}_n = \tilde{\underline{y}}_{n-1} + \Delta \underline{y}_{n-1} + h_n \underline{F}\left(t_{n-1/2}, \frac{1}{2}\left(\tilde{\underline{y}}_{n-1} + \tilde{\underline{y}}_n\right)\right)$$

$$+ h_n \underline{F}_{\underline{y}}\left(t_{n-1/2}, \frac{1}{2}\left(\tilde{\underline{y}}_{n-1} + \tilde{\underline{y}}_n\right)\right) \frac{\Delta \underline{y}_{n-1} + \Delta \underline{y}_n}{2} + \cdots$$

Notation:

let $\quad \underline{F}\left(t_{n-1/2}, \frac{1}{2}\left(\tilde{\underline{y}}_{n-1} + \tilde{\underline{y}}_n\right)\right) = \underline{F}_{n-1/2}$

and $\quad \underline{F}_{\underline{y}}\left(t_{n-1/2}, \frac{1}{2}\left(\tilde{\underline{y}}_{n-1} + \tilde{\underline{y}}_n\right)\right) = \underline{\underline{A}}_{n-1/2}$

$$\Rightarrow \quad \tilde{\underline{y}}_n + \Delta \underline{y}_n = \tilde{\underline{y}}_{n-1} + \Delta \underline{y}_{n-1} + h_n \underline{F}_{n-1/2} + \frac{h_n}{2} \underline{\underline{A}}_{n-1/2}\left(\Delta \underline{y}_{n-1} + \Delta \underline{y}_n\right) + \cdots$$

$$\Rightarrow \quad \left(\underline{\underline{I}} - \frac{h_n}{2} \underline{\underline{A}}_{n-1/2}\right) \Delta \underline{y}_n - \left(\underline{\underline{I}} + \frac{h_n}{2} \underline{\underline{A}}_{n-1/2}\right) \Delta \underline{y}_{n-1} = +\tilde{\underline{y}}_{n-1} - \tilde{\underline{y}}_n + h_n \underline{F}_{n-1/2}$$

$n = 1, \ldots, N$

BCs: $\quad \hat{\underline{\underline{B}}}_0 \left(\tilde{\underline{y}}_0 + \Delta \underline{y}_0\right) = \underline{\gamma}_0 \quad \rightarrow \quad \hat{\underline{\underline{B}}}_0 \Delta \underline{y}_0 = \underline{\gamma}_0 - \hat{\underline{\underline{B}}}_0 \tilde{\underline{y}}_0$

and similarly ..... $\quad \hat{\underline{\underline{B}}}_1 \Delta \underline{y}_N = \underline{\gamma}_1 - \hat{\underline{\underline{B}}}_1 \tilde{\underline{y}}_N$

$$\begin{bmatrix} & \text{as} \\ & \text{before} \\ & \end{bmatrix} \begin{pmatrix} \Delta \underline{y}_0 \\ \Delta \underline{y}_1 \\ \vdots \\ \Delta \underline{y}_N \end{pmatrix} = \text{RHS} \rightarrow \text{drive to zero}$$

Finite Difference Methods for Nonlinear BVPs

BVP: $\underline{y}' = \underline{f}(t, \underline{y})$, $0 \le t \le b$

with separated linear BCs: $\hat{\underline{\underline{B}}}_0 \underline{y}(0) = \underline{\gamma}_0$, $\hat{\underline{\underline{B}}}_1 \underline{y}(b) = \underline{\gamma}_1$

(k BCs on left & $m-k$ BCs on right)

Notice - $\hat{\underline{\underline{B}}}_0$, $\hat{\underline{\underline{B}}}_1$ not square matrices

Consider the midpoint method:

mesh: $0 = t_0 < t_1 < \dots < t_N = b$, $h_n = t_n - t_{n-1} > 0$

Let $\underline{y}_n \simeq \underline{y}(t_n)$ and set $\underline{y}_n = \underline{y}_{n-1} + h_n \underline{f}\left(t_{n-1/2}, \frac{1}{2}(\underline{y}_{n-1} + \underline{y}_n)\right)$, $n = 1, \dots, N$

~~$\hat{\underline{\underline{B}}}_0 \underline{y}_0 = \underline{\gamma}_0$, $\hat{\underline{\underline{B}}}_1 \underline{y}_N = \underline{\gamma}_1$~~

$\hat{\underline{\underline{B}}}_0 \underline{y}_0 = \underline{\gamma}_0$, $\hat{\underline{\underline{B}}}_1 \underline{y}_N = \underline{\gamma}_1$

k eqns; $m-k$ eqns

number of eqns & unknowns - $m(N+1)$

- this is generally a set of nonlinear algebraic eqns, generally so employ iterative scheme, like Newton's method.

set $\underline{y}_n = \tilde{\underline{y}}_n + \Delta \underline{y}_n$, $n = 0, \dots, N$

substitute into diff eqns, linearize for small ~~Δ~~ $\Delta \underline{y}_n$

$$\Rightarrow \quad \left(-\underline{\underline{I}} + \frac{h_n}{2} \underline{\underline{A}}_n\right) \Delta \underline{y}_{n-1} + \left(\underline{\underline{I}} + \frac{h_n}{2} \underline{\underline{A}}_n\right) \Delta \underline{y}_n = \tilde{\underline{y}}_n - \tilde{\underline{y}}_{n-1} - h_n \underline{f}_n$$

where $\underline{f}_n = \underline{f}\left(t_{n-1/2}, \frac{1}{2}(\tilde{\underline{y}}_{n-1} + \tilde{\underline{y}}_n)\right)$

$$\underline{\underline{A}}_n = \frac{\partial \underline{f}}{\partial \underline{y}}\left(t_{n-1/2}, \frac{1}{2}(\tilde{\underline{y}}_{n-1} + \tilde{\underline{y}}_n)\right), \quad m \times m \text{ matrix}$$

and $\hat{\underline{\underline{B}}}_0 \Delta \underline{y}_0 = \underline{\gamma}_0 - \hat{\underline{\underline{B}}}_0 \tilde{\underline{y}}_0$, $\hat{\underline{\underline{B}}}_1 \Delta \underline{y}_N = \underline{\gamma}_1 - \hat{\underline{\underline{B}}}_1 \tilde{\underline{y}}_N$

Linear system for Δy's:

$$\begin{bmatrix} \hat{\underline{\underline{B}}}_0 & 0 & 0 & & \\ \underline{\underline{I}} + \frac{h_1}{2}\underline{\underline{A}}_1 & -\underline{\underline{I}} + \frac{h_1}{2}\underline{\underline{A}}_1 & 0 & & \\ 0 & \underline{\underline{I}} + \frac{h_2}{2}\underline{\underline{A}}_2 & -\underline{\underline{I}} + \frac{h_2}{2}\underline{\underline{A}}_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \underline{\underline{I}} + \frac{h_N}{2}\underline{\underline{A}}_N & \underline{\underline{I}} - \frac{h_N}{2}\underline{\underline{A}}_N & \\ & & 0 & \hat{\underline{\underline{B}}}_1 & \end{bmatrix} \begin{pmatrix} \Delta y_0 \\ \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_{N-1} \\ \Delta y_N \end{pmatrix} = \begin{bmatrix} \gamma_0 - \hat{\underline{\underline{B}}}_0 \tilde{y}_0 \\ \tilde{y}_1 - \tilde{y}_0 - h_1 F_1 \\ \tilde{y}_2 - \tilde{y}_1 - h_2 F_2 \\ \vdots \\ \tilde{y}_N - \tilde{y}_{N-1} - h_N F_N \\ \gamma_1 - \hat{\underline{\underline{B}}}_0 \tilde{y}_N \end{bmatrix}$$

For a given $\tilde{y}_n$'s, you solve the linear system for Δy_n's

then update:

$$y_n^{new} = \tilde{y}_n + \Delta y_n$$

then repeat with $\tilde{y}_n = y_n^{new}$ until, $\max_n \|\Delta y_n\| < tol$

Example:

$$u'' - G(u) = 0 \quad , \quad 0 \le t \le 1$$

$$u(0) = \alpha \, , \quad u(1) = \beta$$

set $y = \begin{pmatrix} u \\ u' \end{pmatrix} \rightarrow y' = \begin{pmatrix} y_2 \\ G(y_1) \end{pmatrix}$, $\hat{\underline{\underline{B}}}_0 = \hat{\underline{\underline{B}}}_1 = [1 \;\; 0]$

$\gamma_0 = \alpha$, $\gamma_1 = \beta$

Midpoint

$$y_n = y_{n-1} + h_n \begin{pmatrix} \frac{1}{2}(y_{2n-1} + y_{2n}) \\ G\left(\frac{1}{2}(y_{1n-1} + y_{1n})\right) \end{pmatrix} \quad , \quad y_{10} = \alpha \, , \; y_{1N} = \beta$$

(y_{2n})

so $F_n = \begin{pmatrix} \frac{1}{2}(\tilde{y}_{2n-1} + \tilde{y}_{2n}) \\ G\left(\frac{1}{2}(\tilde{y}_{1n-1} + \tilde{y}_{1n})\right) \end{pmatrix}$

remember, $F = \begin{pmatrix} y_2 \\ G(y_1) \end{pmatrix}$ $\quad \underline{\underline{F}}_y = \begin{pmatrix} 0 & 1 \\ G'(y_1) & 0 \end{pmatrix} \rightarrow \underline{\underline{A}}_n = \begin{bmatrix} 0 & 1 \\ G'\left(\frac{1}{2}(\tilde{y}_{1n-1} + \tilde{y}_{1n})\right) & 0 \end{bmatrix}$

Case ① : $G(v) = 9v$, $\alpha = 0$, $\beta = \sinh 3$

exact soln: $v(t) = \sinh(3t)$

Case ② $G(v) = -e^{v+1}$, $\alpha = \beta = 0$, "Bratu" problem

Aside

given $\tilde{v}_n = \tilde{y}_{1n}$ and need $\tilde{y}_{2n}$, where $y_2 = v'$

then use $\tilde{y}_{2n} = \dfrac{\tilde{y}_{1n} - \tilde{y}_{1n-1}}{h_n}$

Consistency, 0-stability, and convergence for finite difference methods.

For midpoint, for example, have

$$N_h \underline{y}_n \equiv \frac{\underline{y}_n - \underline{y}_{n-1}}{h_n} - \underline{F}\left(t, \tfrac{1}{2}(\underline{y}_n + \underline{y}_{n-1})\right)_{t_{n-1/2}} \quad , \quad (Ny = y' - F)$$

$n = 1, \ldots, N$

$N_h \underline{y}_n = 0$, if $\underline{y}_n$ solution of diff. eqn.

Local truncation error: $\underline{d}_n = N_h \underline{y}(t_n)$, $n = 1, \ldots, N$

For midpoint, can show (using Taylor series) that $d_n = O(h^2)$ for $h = \max_n h_n$

Since $d_n \to 0$ as $h \to 0$ $\Rightarrow$ midpoint method is consistent

Would like to show that the finite difference approximation is convergent.

global error $\underline{e}_n = \underline{y}_n - \underline{y}(t_n)$

$\Rightarrow \max_n \|\underline{e}_n\| \to 0$ as $h \to 0$

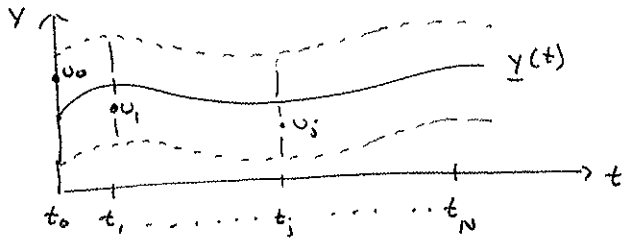
Specifically, $\max_n \|e_n\| = O(h^p)$, where p is order of accuracy (For midpoint, $p = 2$)

The step From consistency to convergence requires 0-stability.

0-stability

define a "discrete tube" about a chosen trajectory $\underline{y}(t)$:

$$S_\rho(\underline{y}) = \left\{ \underline{v}_n \text{ st } \|\underline{v}_n - \underline{y}(t_n)\| < \rho \quad \forall n \right\}$$



The positive value ρ defines the radius of the tube.
A difference method is 0-stable if ρ, h_0 and K (all positive) exist such that For any mesh with $h \le h_0$ and any mesh functions $\underline{x}_n$ and $\underline{z}_n$ in $S_\rho(\underline{y})$, we have

$$(*) \qquad \|\underline{x}_n - \underline{z}_n\| \le K \left\{ \|\underline{g}(\underline{x}_0, \underline{x}_N) - \underline{g}(\underline{z}_0, \underline{z}_N)\| + \max_{1 \le j \le N} \|N_h x_j - N_h z_j\| \right\}$$

For all n.

IF $\underline{x}_n = \underline{y}_n$, $\underline{z}_n = \underline{y}(t_n)$ then (*) tells us

$$\|e_n\| \le K \max_{1 \le j \le N} \|d_j\| \qquad \forall n = 1, \ldots, N$$

$\uparrow$ $O(h^2)$

$$\Rightarrow \quad \|e_n\| = O(h^2)$$

0-stability for BVPs

BVP $\underline{y}' = \underline{f}(t, \underline{y})$, $0 \le t \le b$

$\underline{g}(\underline{y}(0), \underline{y}(b)) = 0$

Finite difference method:

$N_h \underline{y}_n = 0$, $n = 1, \cdots, N$

$\underline{g}(\underline{y}_0, \underline{y}_N) = 0$

For Midpoint: $N_h \underline{y}_n = \dfrac{\underline{y}_n - \underline{y}_{n-1}}{h_n} - \underline{f}\left(t_{n-1/2}, \dfrac{\underline{y}_{n-1} + \underline{y}_n}{2}\right)$

~~$N_h \underline{y}_n = 0$~~

Define: a discrete tube about a solution $\underline{y}(t)$:

$$S_\rho(\underline{y}) = \left\{ \underline{v}_n \text{ st } \|\underline{v}_n - \underline{y}(t_n)\| < \rho \ \forall n \right\}$$

A difference method is 0-stable if ρ, K, h_0 exist st for any mesh with $h \le h_0$ and any mesh functions $\underline{x}_n$, $\underline{z}_n$ in $S_\rho(\underline{y})$, we have

$$\|\underline{x}_n - \underline{z}_n\| \le K \left\{ \|g(\underline{x}_0, \underline{x}_N) - g(\underline{z}_0, \underline{z}_N)\| + \max_{1 \le j \le N} \|N_h \underline{x}_j - N_h \underline{z}_j\| \right\} \forall n$$

In order to establish convergence we need to show consistency (easy) and 0-stability (harder).

If the BVP is nonlinear, then you would linearize about a solution $y(t)$ and try to show 0-stability for the linearized problem. This result would hold for ρ sufficiently small.

Consider the linear case: $\underline{f}(t, \underline{y}) = \underline{\underline{A}}(t)\,\underline{y} + \underline{q}(t)$

$$\underline{g}(\underline{v}, \underline{w}) = \underline{\underline{B}}_0 \underline{v} + \underline{\underline{B}}_1 \underline{w} - \underline{\gamma}$$

Calculate:

$$g(\underline{x}_0, \underline{x}_N) - g(\underline{z}_0, \underline{z}_N) = \left(\underline{\underline{B}}_0 \underline{x}_0 + \underline{\underline{B}}_1 \underline{x}_N - \underline{\gamma}\right) - \left(\underline{\underline{B}}_0 \underline{z}_0 + \underline{\underline{B}}_1 \underline{z}_N - \underline{\gamma}\right)$$

$$\boxed{g(\underline{x}_0, \underline{x}_N) - g(\underline{z}_0, \underline{z}_N) = \underline{\underline{B}}_0 \underline{w}_0 + \underline{\underline{B}}_1 \underline{w}_N}$$

where $\underline{w}_0 = \underline{x}_0 - \underline{z}_0$, $\underline{w}_N = \underline{x}_N - \underline{z}_N$

$$N_h \underline{x}_j = \frac{\underline{x}_j - \underline{x}_{j-1}}{h_j} - \underline{\underline{A}}\left(t_{j-1/2}\right) \frac{1}{2}\left(\underline{x}_j + \underline{x}_{j-1}\right) \overset{\ominus}{+} \underline{q}(t_{j-1/2})$$

$$N_h \underline{x}_j = \left(\frac{1}{h_j}\underline{\underline{I}} - \frac{1}{2}\underline{\underline{A}}_{j-1/2}\right)\underline{x}_j - \left(\frac{1}{h_j}\underline{\underline{I}} + \frac{1}{2}\underline{\underline{A}}_{j-1/2}\right)\underline{x}_{j-1} \overset{\ominus}{+} \underline{q}(t_{j-1/2})$$

where $\underline{\underline{A}}_{j-1/2} = \underline{\underline{A}}(t_{j-1/2})$

and similarly,

$$N_h \underline{z}_j = \left(\frac{1}{h_j}\underline{\underline{I}} - \frac{1}{2}\underline{\underline{A}}_{j-1/2}\right)\underline{z}_j - \left(\frac{1}{h_j}\underline{\underline{I}} + \frac{1}{2}\underline{\underline{A}}_{j-1/2}\right)\underline{z}_{j-1} - \underline{q}(t_{j-1/2})$$

$$\Rightarrow \boxed{N_h \underline{x}_j - N_h \underline{z}_j = \underline{\underline{R}}_j \underline{w}_j + \underline{\underline{S}}_j \underline{w}_{j-1}}$$

where $\underline{\underline{R}}_j = \frac{1}{h_j}\underline{\underline{I}} - \frac{1}{2}\underline{\underline{A}}_{j-1/2}$, $\underline{\underline{S}}_j = \left(\frac{1}{h_j}\underline{\underline{I}} + \frac{1}{2}\underline{\underline{A}}_{j-1/2}\right)$

Define: $\underline{\delta}_{j-1} \equiv \underline{\underline{S}}_j \underline{w}_{j-1} + \underline{\underline{R}}_j \underline{w}_j \quad , j = 1, \ldots, N$

$$\underline{\delta}_N \equiv \underline{\underline{B}}_0 \underline{w}_0 + \underline{\underline{B}}_1 \underline{w}_N$$

$$\Rightarrow \begin{bmatrix} \underline{\underline{S}}_1 & \underline{\underline{R}}_1 & & & \\ & \underline{\underline{S}}_2 & \underline{\underline{R}}_2 & & \\ & & \ddots & \ddots & \\ & & & \underline{\underline{S}}_N & \underline{\underline{R}}_N \\ \underline{\underline{B}}_0 & & & & \underline{\underline{B}}_1 \end{bmatrix} \begin{pmatrix} \underline{w}_0 \\ \underline{w}_1 \\ \vdots \\ \underline{w}_{N-1} \\ \underline{w}_N \end{pmatrix} = \begin{pmatrix} \underline{\delta}_0 \\ \underline{\delta}_1 \\ \vdots \\ \underline{\delta}_{N-1} \\ \underline{\delta}_N \end{pmatrix}$$

$\underline{\underline{J}}$

$$\Rightarrow \underline{\underline{J}}\,\underline{w} = \underline{\delta}$$

$\underline{\underline{J}}$ nonsingular $\Rightarrow$ $\underline{w} = \underline{\underline{J}}^{-1}\underline{\delta}$

For 0- stability, must show that $\|\underline{\underline{J}}^{-1}\| \le K$

- Consider the inverse of $\underline{\underline{J}}$:

First look at calculation:

$$\underline{\underline{R}}_n^{-1}\underline{\underline{S}}_n = -\left(\frac{1}{h_n}\underline{\underline{I}} - \frac{1}{2}\underline{\underline{A}}_{n-1/2}\right)^{-1}\left(\frac{1}{h_n}\underline{\underline{I}} + \frac{1}{2}\underline{\underline{A}}_{n-1/2}\right)$$

$$= -\left(\underline{\underline{I}} - \frac{h_n}{2}\underline{\underline{A}}_{n-1/2}\right)^{-1}\left(\underline{\underline{I}} + \frac{h_n}{2}\underline{\underline{A}}_{n-1/2}\right)$$

$$= -\left(\underline{\underline{I}} + \frac{h_n}{2}\underline{\underline{A}}_{n-1/2} + \cdots\right)\left(\underline{\underline{I}} + \frac{h_n}{2}\underline{\underline{A}}_{n-1/2}\right)$$

$$= -\left(\underbrace{\underline{\underline{I}} + h_n\underline{\underline{A}}_{n-1/2} + O(h_n^2)}\right)$$

approximation of Fundamental solution matrix $\underline{\underline{Y}}_n(t_n) + \cdots$

where $\underline{\underline{Y}}_n' = \underline{\underline{A}}(t)\,\underline{\underline{Y}}_n$, $\underline{\underline{Y}}_n(t_{n-1}) = \underline{\underline{I}}$

why? consider one step of midpoint/Euler

$$\underline{\underline{Y}}_n(t_n) = \underbrace{\underline{\underline{Y}}_n(t_{n-1})}_{\underline{\underline{I}}} + h_n\underline{\underline{A}}(t_{n-1/2})\underbrace{\underline{\underline{Y}}_n(t_{n-1})}_{\underline{\underline{I}}} + \cdots$$

Now set $\underline{\underline{D}}\,\underline{\underline{S}} = \underline{\underline{M}} + \underline{\underline{E}}$

where $\underline{\underline{D}} = \begin{bmatrix} \underline{\underline{R}}_1^{-1} & & & \\ & \underline{\underline{R}}_2^{-1} & & \\ & & \ddots & \\ & & & \underline{\underline{R}}_N^{-1} \\ & & & & \underline{\underline{I}} \end{bmatrix}$, where $\underline{\underline{D}}$ is block diagonal

$$\Rightarrow \underline{\underline{M}} = \begin{bmatrix} -\underline{\underline{Y}}_1(t_1) & \underline{\underline{I}} & & & \\ & -\underline{\underline{Y}}_2(t_2) & \underline{\underline{I}} & & \\ & & -\underline{\underline{Y}}_3(t_3) & \underline{\underline{I}} & \\ & & & \ddots & \ddots \\ & & & & -\underline{\underline{Y}}_N(t_N) \quad \underline{\underline{I}} \\ \underline{\underline{B}}_0 & & & & \underline{\underline{B}}_1 \end{bmatrix}$$

$\underline{\underline{E}}$ is a matrix of size $O(h^2)$ with the same zero structure as

$$\underline{\underline{\Sigma}}^{-1} = \left(\underline{\underline{D}}^{-1} \left(\underline{\underline{M}} + \underline{\underline{E}} \right) \right)^{-1} \rightarrow \underline{\underline{\Sigma}}^{-1} = \left(\underline{\underline{M}} + \underline{\underline{E}} \right)^{-1} \underline{\underline{D}}$$

Let $\left(\underline{\underline{M}} + \underline{\underline{E}} \right)^{-1} = \underline{\underline{M}}^{-1} \left(\underline{\underline{I}} - \underline{\underline{X}} \right)$, where $\underline{\underline{X}}$ is a small quantity

$$\rightarrow \underline{\underline{M}} + \underline{\underline{E}} = \left(\underline{\underline{I}} - \underline{\underline{X}} \right)^{-1} \underline{\underline{M}}$$

$$= \left(\underline{\underline{I}} + \underline{\underline{X}} + \cdots \right) \underline{\underline{M}}$$

$$= \underline{\underline{M}} + \underline{\underline{X}}\,\underline{\underline{M}} + \cdots \rightarrow \underline{\underline{E}} = \underline{\underline{X}}\,\underline{\underline{M}} \rightarrow \underline{\underline{X}} = \underline{\underline{E}}\,\underline{\underline{M}}^{-1}$$

$$\Rightarrow \left(\underline{\underline{M}} + \underline{\underline{E}} \right)^{-1} = \underline{\underline{M}}^{-1} \left(\underline{\underline{I}} - \underline{\underline{E}}\,\underline{\underline{M}}^{-1} + \cdots \right)$$

$$\rightarrow \underline{\underline{\Sigma}}^{-1} = \underline{\underline{M}}^{-1} \left(\underline{\underline{I}} - \underline{\underline{E}}\,\underline{\underline{M}}^{-1} + \cdots \right) \underline{\underline{D}}$$

Green's Functions

$$\underline{\underline{M}}^{-1} = \begin{bmatrix} \underline{\underline{G}}(t_0, t_1) & \underline{\underline{G}}(t_0, t_2) & \underline{\underline{G}}(t_0, t_3) & \cdots & \underline{\underline{G}}(t_0, t_N) & \underline{\underline{\Phi}}(t_0) \\ \underline{\underline{G}}(t_1, t_1) & \underline{\underline{G}}(t_1, t_2) & & & \underline{\underline{G}}(t_1, t_N) & \underline{\underline{\Phi}}(t_1) \\ \vdots & \vdots & & & & \\ \underline{\underline{G}}(t_N, t_1) & \underline{\underline{G}}(t_N, t_2) & & & \underline{\underline{G}}(t_N, t_N) & \underline{\underline{\Phi}}(t_N) \end{bmatrix}$$

IF $\| \underline{\underline{G}}(t, \tau) \| \leq K$ and $\| \underline{\underline{\Phi}} \| \leq K$

then $\| M^{-1} \| \leq NK = \frac{\tilde{K}}{h}$

and also, $\| \underline{\underline{D}} \| = O(h)$, $\| \underline{\underline{E}} \| = O(h^2)$,

$\Rightarrow \| \underline{\underline{\Sigma}}^{-1} \| \leq \tilde{K} (1 + O(h))$ $\Rightarrow$ as $h \rightarrow 0$, $\| \underline{\underline{\Sigma}}^{-1} \|$ is bounded

Higher Order Methods

1) Collocation

The idea with collocation is to assume that the soln can be described by a chosen function (a polynomial, for instance) and then choose the coeffs st the ODEs are solved exactly.

⇒ Families: Radau, Gauss, Lebatta

↑ B. Euler ↑ Midpoint ↑ Trapezoidal

symmetric methods - well suited for BVPs

General Form:

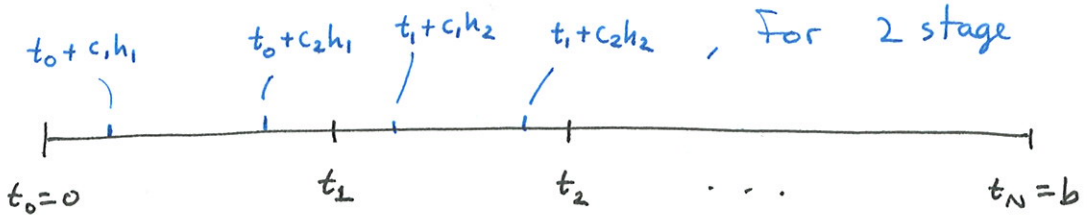
$$\underline{Y}_n = \underline{Y}_{n-1} + h_n \sum_{i=1}^{s} b_i \underline{F}(t_{n-1} + c_i h_n, \underline{Y}_i)$$

$$\underline{Y}_i = \underline{Y}_{n-1} + h_n \sum_{j=1}^{s} a_{ij} \underline{F}(t_{n-1} + c_j h_n, \underline{Y}_j) \quad , i = 1, \ldots, s$$

eg, Gauss, $s = 2$ $(p = 4)$

$$\begin{array}{c|cc} c_1 = \frac{3-\sqrt{3}}{6} & \frac{1}{4} & \frac{3-2\sqrt{3}}{12} \\ c_2 = \frac{3+\sqrt{3}}{6} & \frac{3+2\sqrt{3}}{12} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \\ & \uparrow b_1 & \uparrow b_2 \end{array} \quad \left.\right\} a_{ij}$$

Define a mesh: $0 = t_0 < t_1 < \cdots < t_N = b$



, For 2 stage

so you have discretization points and collocation points

Define $\underline{Y}_n$'s on discretization points

and $\underline{Y}_{ni}$'s on collocation points

Some Details for the Linear case

$$\underline{y}'(t) = \underline{\underline{A}}\,\underline{y}(t) + \underline{q}(t)$$

discretize using Gauss with $s=2$, (4^{th} order method)

$$\underline{y}_n = \underline{y}_{n-1} + h_n \Big[b_1 \Big(\underline{\underline{A}}(t_{n-1} + c_1 h_n)\,\underline{Y}_{n1} + \underline{q}(t_{n-1} + c_1 h_n) \Big) + b_2 \Big(\underline{\underline{A}}(t_{n-1} + c_2 h_n)\,\underline{Y}_{n2} + \underline{q}(t_{n-1} + c_2 h_n) \Big) \Big]$$

For notation, define

$$\underline{\underline{A}}_{n1} = \underline{\underline{A}}(t_{n-1} + c_1 h_n)$$

$$\underline{\underline{A}}_{n2} = \underline{\underline{A}}(t_{n-1} + c_2 h_n)$$

$$\underline{q}_{n1} = \underline{q}(t_{n-1} + c_1 h_n)$$

$$\underline{q}_{n2} = \underline{q}(t_{n-1} + c_2 h_n)$$

$$\rightarrow \quad \underline{y}_n = \underline{y}_{n-1} + h_n \Big[b_1 \big(\underline{\underline{A}}_{n1}\,\underline{Y}_{n1} + \underline{q}_{n1} \big) + b_2 \big(\underline{\underline{A}}_{n2}\,\underline{Y}_{n2} + \underline{q}_{n2} \big) \Big], \quad n = 1, \ldots, N$$

Collocation Pts:

$$\underline{Y}_{n1} = \underline{y}_{n-1} + h_n \Big(a_{11} \big(\underline{\underline{A}}_{n1}\,\underline{Y}_{n1} + \underline{q}_{n1} \big) + a_{12} \big(\underline{\underline{A}}_{n2}\,\underline{Y}_{n2} + \underline{q}_{n2} \big) \Big)$$

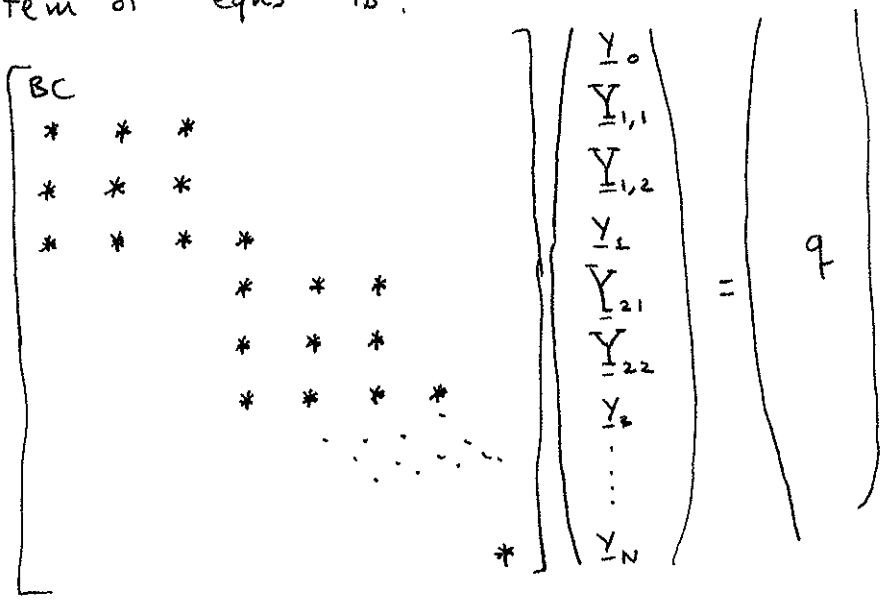
$$\underline{Y}_{n2} = \underline{y}_{n-1} + h_n \Big(a_{21} \big(\underline{\underline{A}}_{n1}\,\underline{Y}_{n1} + \underline{q}_{n1} \big) + a_{22} \big(\underline{\underline{A}}_{n2}\,\underline{Y}_{n2} + \underline{q}_{n2} \big) \Big)$$

BCs

$$\hat{\underline{\underline{B}}}_0\,\underline{y}_0 = \underline{\gamma}_0, \qquad \hat{\underline{\underline{B}}}_1\,\underline{y}_N = \underline{\gamma}_1$$

Number of Eqns: $(3N+1)\,m$ = Number of unknowns

The basic structure of the resulting system of eqns is:



4/

Higher Order Methods

BVP: $\underline{y}' = \underline{f}(t, \underline{y})$, $0 \le t \le b$, $\underline{g}(\underline{y}(0), \underline{y}(b)) = 0$

Various approaches to obtain higher order ($p > 2$) finite difference methods.

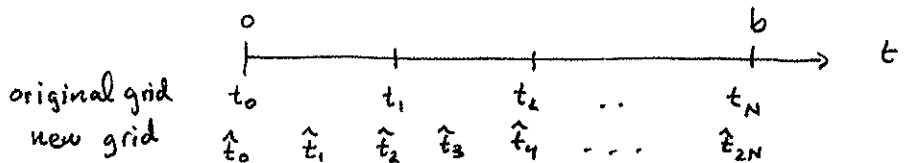
1) collocation methods
   eg midpoint → Gauss methods
   trapezoidal → Lobatto methods (sym RK schemes)

2) Richardson Extrapolation
   idea is to use linear combinations of approximate solutions of lower order methods to construct higher order approximations

   e.g. approximate solution using midpoint on grid with uniform spacing h.

   (1) $\underline{y}_n - \underline{y}(t_n) = \underline{e}(t_n) h^2 + O(h^4)$, $n = 0, 1, \ldots, N$

   Now consider the approximate solution using midpoint on grid with uniform spacing $\hat{h} = \frac{h}{2}$.



Have for the solution on the finer grid

② $\hat{\underline{y}}_n - \underline{y}(\hat{t}_n) = \underline{\varepsilon}(\hat{t}_n)\hat{h}^2 + O(\hat{h}^4)$, $n = 0, \ldots, 2N$

where $\underline{\varepsilon}(t)$ is the same function for both grids - $\underline{\varepsilon}(t)$ only depends on the method and is independent of grid size.

Consider the finer solution at every other grid point, or where ① and ② have common points:

$$\hat{\underline{y}}_{2n} = \hat{\underline{y}}(\hat{t}_{2n}) = \underline{\varepsilon}(\hat{t}_{2n})\hat{h}^2 + O(\hat{h}^4) \quad , n = 0, \ldots, N$$

use $\hat{t}_{2n} = t_n$, $\hat{h} = \frac{h}{2}$

③ → $\hat{\underline{y}}_{2n} - \underline{y}(t_n) = \underline{\varepsilon}(t_n)\frac{h^2}{4} + O(h^4)$

Eliminate 2nd order term in error from ① and ③

ie $4 \times ③ - ①$

$$\Rightarrow \quad 4\hat{\underline{y}}_{2n} - \underline{y}_n - 3\underline{y}(t_n) = O(h^4)$$

$$\Rightarrow \quad \boxed{\frac{4\hat{\underline{y}}_{2n} - \underline{y}_n}{3} - \underline{y}(t_n) = O(h^4)}$$

this is a 4th order accurate approximation of $\underline{y}(t_n)$

Can also obtain an error estimate: eqn ① - eqn ③

③ $\hat{\underline{y}}_{2n} - \underline{y}(t_n) = \underline{\varepsilon}(t_n)\frac{h^2}{4} + O(h^4)$

① $\underline{y}_n - \underline{y}(t_n) = \underline{\varepsilon}(t_n)h^2 + O(h^4)$

$$\Rightarrow \quad \underline{y}_n - \hat{\underline{y}}_{2n} = \frac{3}{4}\underline{\varepsilon}(t_n)h^2 + O(h^4)$$

$$\rightarrow \quad \underline{\varepsilon}(t_n)h^2 = \underbrace{\frac{4}{3}(\underline{y}_n - \hat{\underline{y}}_{2n})}_{\text{error estimate}} + O(h^4)$$

Continuation Methods

A discretization of a BVP (either by shooting, finite differences, or some other method) leads to a set of nonlinear algebraic eqns to solve for the discrete solution.

Let us refer to the collection of discrete unknowns as $\underline{v}$ and the set of equations that define them as $\underline{G}$. We want to solve $\underline{G}(\underline{v}) = \underline{0}$. (In the case of shooting, $\underline{v}$ is the shooting parameter(s). In the case of finite differences, $\underline{v}$ is the solution on the grid...).

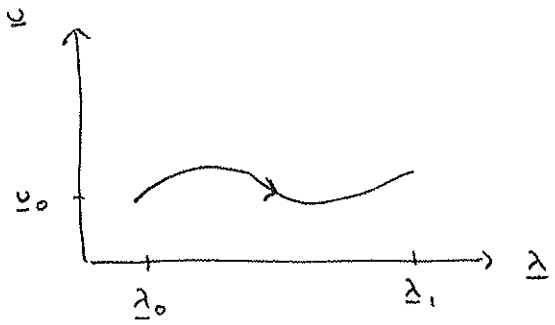
$\underline{G}(\underline{v}) = \underline{0}$ is a standard root-finding problem. The usual approach is some iteration, ie Newton's method or variation of it, and we need some initial guess to get started.

Continuation is an approach that can be used to obtain initial guesses in a rational way.

Typically, the equations (exact or discrete) involve a set of parameters, $\underline{\lambda}$ say.

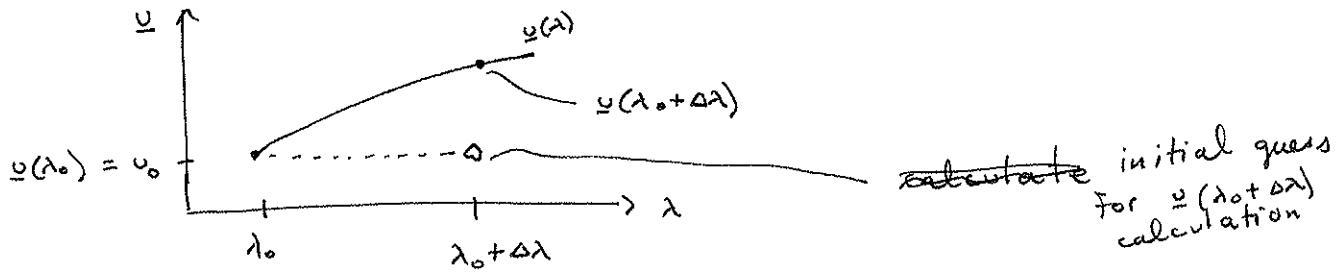
$$\Rightarrow \quad \underline{G}(\underline{v}, \underline{\lambda}) = 0 \quad \Rightarrow \quad \underline{v} \text{ depends on } \underline{\lambda}$$

Exploit this dependence to obtain a good initial guess. More generally, might want to determine how v depends on $\underline{\lambda}$.

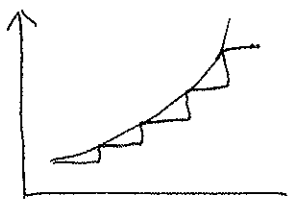


Suppose want to compute $\underline{u}$ at $\underline{\lambda} = \underline{\lambda}_1$ (hard) and we start at $\underline{\lambda} = \underline{\lambda}_0$ where the solution is easy (perhaps linear for instance). The question is, how can we move in parameter space in a systematic way $\Rightarrow$ continuation.

Consider a step in parameter space: (with λ a scalar)



"easy" continuation ~~part~~ - use $\underline{u}_0$ as initial guess for Newton at $\lambda = \lambda_0 + \Delta\lambda$ and then converge (hopefully) to $\underline{u}(\lambda_0+\Delta\lambda)$ using Newton, with λ fixed



this often works but there are more sophisticated ideas.

Instead of stair-casing along, why not find tangents ...

Euler-Newton "continuation": Regard $\underline{u}$ as a differentiable function of scalar parameter λ.

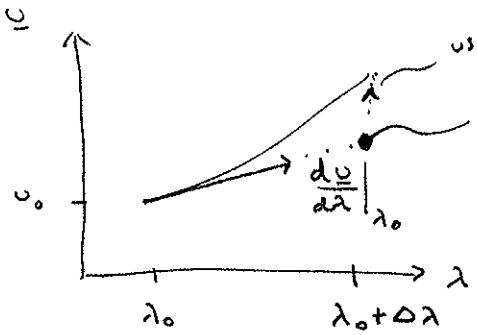
$$\Rightarrow \quad \underline{G}(\underline{u}(\lambda), \lambda) = 0$$

Take a derivative with respect to λ:

$$\underline{G}_{\underline{u}}(\underline{u}, \lambda)\frac{\partial \underline{u}}{\partial \lambda} + \underline{G}_{\lambda}(\underline{u}, \lambda) = 0 \qquad \text{(linear system for } \frac{d\underline{u}}{d\lambda}\text{)}$$

Apply at u_0, λ_0. Define $G_u(u_0,\lambda_0) = G_u^0$, $G_\lambda(u_0,\lambda_0) = G_\lambda^0$

Solve $G_u^0 \left.\frac{du}{d\lambda}\right|_{\lambda=\lambda_0} = -G_\lambda^0$ (*) For $\frac{du}{d\lambda}$



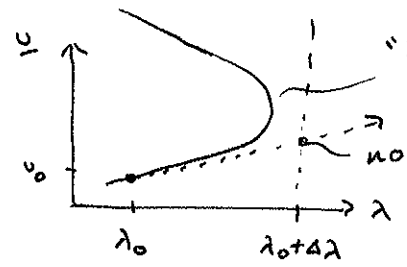
set $\tilde{u} = u_0 + \Delta\lambda \left.\frac{du}{d\lambda}\right|_{\lambda=\lambda_0}$

↑ initial guess for Newton at $\lambda_0 + \Delta\lambda$

Almost all of this information is available as is, in the code. Certainly the Jacobian is known. The solution of the linear system (*) comes for free

Arclength Continuation

A difficulty occurs for either the "trivial" or "Euler-Newton" continuation near "Folds" or "turning points".



} multiple solutions exist

Regard u and λ depending on a parameter s:

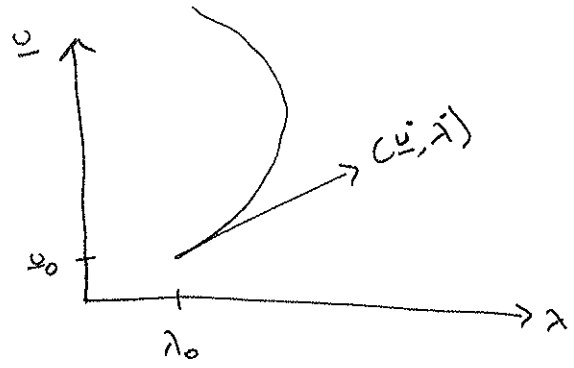
$\Rightarrow G(u(s), \lambda(s)) = 0$

Pick s: For example, if s is arclength then

$\|\dot{u}(s)\|^2 + \dot{\lambda}(s)^2 = 1$, where $\dot{} \equiv \frac{d}{ds}$

stems from $ds^2 = dx^2 + dy^2$ $\Rightarrow 1 = \dot{x}^2 + \dot{y}^2$

Find $\dot{u}, \dot{\lambda}$ at (λ_0, u_0) (ie the tangent at λ_0, u_0)



To find tangent, differentiate wrt s

$\underline{G}(\underline{u}(s), \lambda(s)) = 0$

$\Rightarrow \underline{G}_{\underline{u}}(\underline{u}_0, \lambda_0)\,\dot{\underline{u}} + \underline{G}_{\lambda}(u_0, \lambda_0)\,\dot{\lambda} = 0$ (★)

matrix / vector / vector / scalar

this represents n equations

set $\dot{\underline{u}} = \alpha \underline{\phi}$, where α = scalar, $\underline{\phi}$ is a vector that solves

$\underline{G}_{\underline{u}}(\underline{u}_0, \lambda_0)\,\underline{\phi} = \underline{G}_{\lambda}(u_0, \lambda_0)$

Substitute $\dot{\underline{u}} = \alpha \underline{\phi}$ into (★): $\underline{G}_{\underline{u}}(\underline{u}_0, \lambda_0)\,\alpha\underline{\phi} + \underline{G}_{\lambda}(u_0, \lambda_0)\,\dot{\lambda} = 0$

$\rightarrow \alpha\,\underline{G}_{\lambda}(u_0, \lambda_0) + \underline{G}_{\lambda}(u_0, \lambda_0)\,\dot{\lambda} \rightarrow (\alpha + \dot{\lambda})\,\underline{G}_{\lambda} = 0 \rightarrow \boxed{\alpha = -\dot{\lambda}}$

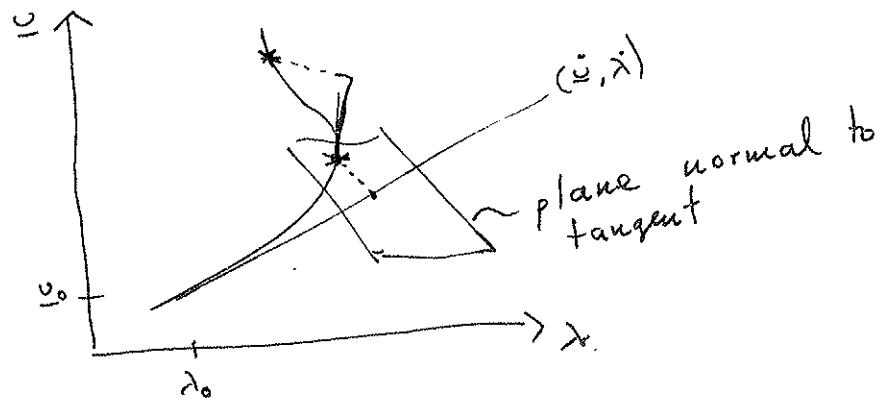
$\Rightarrow \dot{\underline{u}} = -\dot{\lambda}\underline{\phi}$

therefore the equation $\|\dot{u}(s)\|^2 + \dot{\lambda}(s)^2 = 1$ becomes

$\|-\dot{\lambda}\underline{\phi}\|^2 + \dot{\lambda}^2 = 1 \rightarrow \dot{\lambda}^2 \|\underline{\phi}\|^2 + \dot{\lambda}^2 = 1 \rightarrow \boxed{\dot{\lambda} = \pm \frac{1}{\sqrt{1 + \|\phi\|^2}}}$

+ value for ↗ direction, − value for ↖

Now that we have calculated $(\dot{\underline{u}}, \dot{\lambda})$ at (u_0, λ_0). Now want to move back to curve, but not at fixed λ, which gave us problems with this fold. Consider a plane normal to the tangent. Use Newton's method to move along the plane back to the curve.



Eqn for the curve: $\underline{G}(\underline{v}, \lambda) = 0$

Eqn for the plane: $N(v,\lambda) = \underline{\dot{v}}^T(\underline{v} - \underline{v}_0) + \dot{\lambda}(\lambda - \lambda_0) - \Delta s = 0$ (scalar eqn)

Perform Newton on $\underline{G} = 0$, $N = 0$ to find v, λ for a given Δs. Need to calculate a Jacobian:

$\underline{v} = \tilde{\underline{v}} + \underline{\Delta v}$, $\lambda = \tilde{\lambda} + \Delta\lambda$, where $\tilde{\underline{v}}, \tilde{\lambda}$ are initial guess

$\underline{G}(\tilde{\underline{v}} + \underline{\Delta v}, \tilde{\lambda} + \Delta\lambda) = 0$

$\rightarrow \underline{G}(\tilde{\underline{v}}, \tilde{\lambda}) + \underline{G}_{\underline{v}}(\tilde{\underline{v}}, \tilde{\lambda})\underline{\Delta v} + \underline{G}_{\lambda}(\tilde{\underline{v}}, \tilde{\lambda})\Delta\lambda + \cdots = 0$ (neglect higher order terms)

① $\rightarrow \underline{G}_{\underline{v}}(\tilde{\underline{v}}, \tilde{\lambda})\underline{\Delta v} + \underline{G}_{\lambda}(\tilde{\underline{v}}, \tilde{\lambda})\Delta\lambda = -\underline{G}(\tilde{\underline{v}}, \tilde{\lambda})$ — linear equation for $\underline{v}$, $\tilde{\lambda}$

and

$N(\tilde{\underline{v}} + \underline{\Delta v}, \tilde{\lambda} + \Delta\lambda) = 0$ similarly:

$\rightarrow N(\tilde{\underline{v}}, \tilde{\lambda}) + \underbrace{N_{\underline{v}}(\tilde{\underline{v}}, \tilde{\lambda})}_{\underline{\dot{v}}^T}\underline{\Delta v} + \underbrace{N_{\lambda}(\tilde{\underline{v}}, \tilde{\lambda})}_{\dot{\lambda}}\Delta\lambda = 0$

② $\rightarrow \underline{\dot{v}}^T\underline{\Delta v} + \dot{\lambda}\Delta\lambda = -N(\tilde{\underline{v}}, \tilde{\lambda})$

combine ① and ② into a linear system

$$\begin{bmatrix} \underline{G}_{\underline{v}} & \vdots & \underline{G}_{\lambda} \\ \cdots & \cdots & \cdots \\ \underline{\dot{v}}^T & \vdots & \dot{\lambda} \end{bmatrix} \begin{pmatrix} \underline{\Delta v} \\ \cdots \\ \Delta\lambda \end{pmatrix} = \begin{pmatrix} -\underline{G} \\ \cdots \\ -N \end{pmatrix}$$

this becomes the problem to solve for the corrections $\underline{\Delta v}$ and $\Delta\lambda$

Must update both solutions

$\underline{v}^{NEW} = \tilde{\underline{v}} + \underline{\Delta v}$

$\lambda^{NEW} = \tilde{\lambda} + \Delta\lambda$

typically solve using efficient algorithms

Review - since Midterm

1) linear multi-step methods (IVPs)

a) Adams Family For non-stiff problems
   explicit ($B_0 = 0$) vs implicit ($B_0 \neq 0$)

b) Backwards Difference Formula (BDF)
   - For stiff problems

c) starting values

d) order of accuracy → order conditions

e) 0-stability, root-condition

f) absolute stability

g) implementation issues
   - predictor/corrector scheme
   - error estimates (Milne estimate)
   - variable mesh size

2) BVP - theory

a) general problem, locally isolated solutions, linearization

b) general solution of linear BVP ⇒ Greens Function

c) problem stability → dichotomy, exponential dichotomy

3) BVP - Shooting Methods

a) general framework, Newton's method

b) difficulties: if BVP is stable, IVP is unstable

c) multiple shooting (for linear case)

4) BVP - Finite difference Methods

a) linear case, midpoint method → linear system

b) nonlinear case, midpoint method → Newton → linear system

c) consistency, 0-stability, convergence

d) higher order methods
   - collocation and extrapolation

e) continuation methods

Take-Home Final Exam Rules

- email to get the final
- 1st available time, saturday morning (May 3rd), to return monday morning
- 48 hrs to complete the exam
- All exams to be completed by 5pm Wed May 7th
- NO COLLABORATION or TEXTBOOKS