MA 7007: Numerical Solution of Differential Equations I
Initial Value Problem for ODEs

Suh-Yuh Yang (楊肅煜)

Department of Mathematics, National Central University
Jhongli District, Taoyuan City 32001, Taiwan

E-mail: syyang@math.ncu.edu.tw
Website: http://www.math.ncu.eud.tw/∼syyang/

Introduction

- Initial-value problem (IVP): find $u(t)$ such that

$$u'(t) = f(u(t), t) \quad \text{for } t > t_0 \tag{*}$$

  with some initial data $u(t_0) = \eta$. We will often assume $t_0 = 0$ for simplicity.
- In general, (*) may represent a system of ODEs, i.e., for $t > t_0$,

$$\begin{aligned} u_1'(t) &= f_1(u_1, u_2, \cdots, u_s, t), \\ u_2'(t) &= f_2(u_1, u_2, \cdots, u_s, t), \\ &\vdots \\ u_s'(t) &= f_s(u_1, u_2, \cdots, u_s, t). \end{aligned}$$

- A higher-order ODE can be converted to a first-order system.

Example

Consider the IVP for the ODE,

$$v'''(t) = v'(t)v(t) - 2t(v''(t))^2 \qquad \text{for } t > 0.$$

with three initial values $v(0) = \eta_1$, $v'(0) = \eta_2$, and $v''(0) = \eta_3$.

Let $u_1(t) := v(t)$, $u_2(t) := v'(t)$, and $u_3(t) := v''(t)$. Then we have

$$\begin{aligned} u_1'(t) &= u_2(t), \\ u_2'(t) &= u_3(t), \qquad\qquad t > 0, \\ u_3'(t) &= u_1(t)u_2(t) - 2tu_3^2(t), \end{aligned}$$

with the initial conditions $u_1(0) = \eta_1$, $u_2(0) = \eta_2$, and $u_3(0) = \eta_3$.

Example: An autonomous first-order system

We also may define $u_4(t) = t$ such that $u_4'(t) = 1$ and $u_4(0) = 0$. Then the system takes the form

$$u'(t) = f(u(t)), \quad t > 0,$$

with

$$f(u) = \begin{bmatrix} u_2 \\ u_3 \\ u_1 u_2 - 2u_4 u_3^2 \\ 1 \end{bmatrix} \quad \text{and} \quad u(0) = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ 0 \end{bmatrix}.$$

The equation is said to be autonomous since the RHS f does not depend explicitly on time t.

Linear ordinary differential equations

- The system of ODEs (*) is linear if

$$f(u,t) = A(t)u + g(t),$$

where $A(t) \in \mathbb{R}^{s\times s}$ and $g(t) \in \mathbb{R}^{s}$.

- An important special case is the constant coefficient linear system

$$u'(t) = Au(t) + g(t),$$

where $A \in \mathbb{R}^{s\times s}$ is a constant matrix. If $g(t) \equiv 0$, then the equation is homogeneous. The solution to the homogeneous system $u'(t) = Au(t)$ with $u(t_0) = \eta$ is

$$u(t) = e^{A(t-t_0)}\eta,$$

where the matrix exponential is defined as in Appendix D.

Duhamel's principle

- If $g(t) \not\equiv 0$, then the solution to the constant coefficient system can be written as

$$u(t) = e^{A(t-t_0)}\eta + \int_{t_0}^{t} e^{A(t-\tau)} g(\tau) d\tau.$$

This is known as *Duhamel's principle*.

- In particular, if A is constant and so is $g(t) \equiv g \in \mathbb{R}^s$. Then the solution reduces to

$$u(t) = e^{A(t-t_0)}\eta + \left(\int_{t_0}^{t} e^{A(t-\tau)} d\tau \right) g.$$

This integral can be computed by expressing $e^{A(t-\tau)}$ as a Taylor series and then integrating term by term. We obtain

$$\int_{t_0}^{t} e^{A(t-\tau)} d\tau = A^{-1}(e^{A(t-t_0)} - I),$$

and so

$$u(t) = e^{A(t-t_0)}\eta + A^{-1}\left(e^{A(t-t_0)} - I\right) g.$$

Existence and uniqueness of solutions

We consider the following IVP ($s = 1$),

$$u'(t) = f(u(t), t), \quad t > t_0 \quad \text{and} \quad u(t_0) = \eta.$$

- We say that $f(u,t)$ is Lipschitz continuous in u over the domain

$$\mathcal{D} := \{(u,t) : |u - \eta| \le a, t_0 \le t \le t_1\}$$

  if there exists some constant $L \ge 0$ so that

$$|f(u,t) - f(u^*,t)| \le L|u - u^*|$$

  for all (u,t) and (u^*,t) in $\mathcal{D}$.
- If f is Lipschitz continuous over the region $\mathcal{D}$ then there is a unique solution to the IVP at least up to time $T^* = \min\{t_1, t_0 + a/S\}$, where

$$S = \max_{(u,t)\in\mathcal{D}} |f(u,t)|.$$

Example

Consider the following IVP

$$u'(t) = (u(t))^2, \qquad u(0) = \eta > 0.$$

- Let $f(u) = u^2$. By MVT, $\exists\ \xi$ between u_1 and u_2 such that

$$f(u_1) - f(u_2) = \frac{\partial f(\xi)}{\partial u}(u_1 - u_2) \Longrightarrow |u_1^2 - u_2^2| \le 2(\eta + a)|u_1 - u_2|.$$

  Then $f(u)$ is Lipschitz continuous in u over any finite interval $|u - \eta| \le a$ with $L = 2(\eta + a)$.
- Since $|f(u)| \le (\eta + a)^2 := S$. Then the theorem guarantees that a unique solution exists at least up to time $a/(\eta + a)^2$. Since a is arbitrary, we can set $a = \eta$ and so there is a solution at least up to time $1/(4\eta)$.
- In fact, the unique solution is given by

$$u(t) = \frac{1}{\eta^{-1} - t}.$$

  Note that $u(t) \to \infty$ as $t \to 1/\eta$ and there is no solution beyond time $1/\eta$.

Non-uniqueness

If f is continuous but not Lipschitz continuous, we may have more than one solution, e.g.,

$$u'(t) = \sqrt{u(t)}, \qquad u(0) = 0.$$

The function $f(u) = \sqrt{u}$ is not Lipschitz continuous near $u = 0$ since $f'(u) = 1/(2\sqrt{u}) \to \infty$ as $u \to 0^+$.

Note that this IVP does not have a unique solution. In fact it has two distinct solutions:

$$u(t) \equiv 0 \quad \text{and } u(t) = \frac{1}{4}t^2.$$

Systems of equations: $s > 1$

- We say that function $f(u,t)$ is Lipschitz continuous in u in some norm $\|\cdot\|$ if there exists some constant $L \geq 0$ such that

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

  for all (u,t) and (u^*,t) in some domain

$$\mathcal{D} := \{(u,t) : \|u - \eta\| \leq a, t_0 \leq t \leq t_1\}.$$

- By the equivalence of finite-dimensional norms, if f is Lipschitz continuous in one norm then it is Lipschitz continuous in any other norm, although the Lipschitz constant may depend on the norm chosen.
- The theorems on existence and uniqueness carry over to systems of equations.

Example

Consider the pendulum problem

$$\theta''(t) = -\sin(\theta(t)).$$

Let $v(t) := \theta'(t)$, then it can be rewritten as a first-order system of two equations: $u = (\theta, v)^\top, f(u) = (v, -\sin(\theta))^\top$ and

$$u'(t) = f(u).$$

Consider the max-norm. In view of

$$\begin{aligned} |v - v^*| &\le \|u - u^*\|_\infty, \\ |\sin(\theta) - \sin(\theta^*)| \le |\theta - \theta^*| &\le \|u - u^*\|_\infty, \end{aligned}$$

we have the Lipschitz continuity of f with $L = 1$,

$$\|f(u) - f(u^*)\|_\infty \le \|u - u^*\|_\infty.$$

Basic numerical methods

For simplicity, we consider the following autonomous IVP:

$$u'(t) = f(u(t)) \quad \text{for } t > 0 \text{ and } u(0) = \eta.$$

Let k be the time step, so $t_n = nk$ for $n \geq 0$. Let U^n represent an approximation to $u(t_n)$ and $U^0 := \eta$.

- Euler's method (also called forward Euler):

$$\frac{U^{n+1} - U^n}{k} = f(U^n), \quad n = 0, 1, \cdots,$$

  or $U^{n+1} = U^n + kf(U^n)$, which is a time-marching method.

- Backward Euler method:

$$\frac{U^{n+1} - U^n}{k} = f(U^{n+1}), \quad n = 0, 1, \cdots$$

- Trapezoidal method: averaging the forward and backward Euler methods

$$\frac{U^{n+1} - U^n}{k} = \frac{1}{2}\left(f(U^n) + f(U^{n+1})\right), \quad n = 0, 1, \cdots$$

Remarks

- The backward Euler and trapezoidal methods give an equation that must be solved for U^{n+1}, they are called implicit methods and can be solved by using e.g., Newton's method.
- The forward Euler method is an explicit method.
- The trapezoidal method is second order accurate, whereas the Euler methods are only first order accurate.
- The above methods are all one-step methods, meaning that U^{n+1} is determined from U^n alone and previous values are not needed.

Multistep methods

One way to get higher order accuracy is to use a multistep method that involves other previous values.

- Midpoint method (the leapfrog method): Using the approximation

$$\frac{u(t+k)-u(t-k)}{2k} = u'(t) + \frac{1}{6}k^2 u'''(t) + O(k^3)$$

yields a second order accurate explicit 2-step method

$$\frac{U^{n+1}-U^{n-1}}{2k} = f(U^n)$$

or

$$U^{n+1} = U^{n-1} + 2kf(U^n).$$

- Backward differentiation formula (BDF) methods: Using the approximation

$$\frac{3u(t+k)-4u(t)+u(t-k)}{2k} = u'(t+k) - \frac{1}{3}k^2 u'''(t+k) + \cdots$$

yields a second order accurate implicit 2-step method

$$\frac{3U^{n+1}-4U^n+U^{n-1}}{2k} = f(U^{n+1}).$$

Local truncation error

As usual, the local truncation error (LTE) of, for example, the midpoint method

$$\frac{U^{n+1} - U^{n-1}}{2k} = f(U^n)$$

is defined by

$$\begin{aligned}\tau^n &= \frac{u(t_{n+1}) - u(t_{n-1})}{2k} - f(u(t_n)) \\ &= \left(u'(t_n) + \frac{1}{6}k^2 u'''(t_n) + O(k^3)\right) - u'(t_n) \\ &= \frac{1}{6}k^2 u'''(t_n) + O(k^3).\end{aligned}$$

So, the local truncation error is $O(k^2)$ and we say that methods is second order accurate, although it is not yet clear that the global error will have this behavior (we need some form of *stability* to guarantee it).

One-step errors

Based on the form $U^{n+1} = U^{n-1} + 2kf(U^n)$, we define

$$\mathcal{L}^n := u(t_{n+1}) - u(t_{n-1}) - 2kf(u(t_n)) = \frac{1}{3}k^3 u'''(t_n) + O(k^4) = 2k\tau^n.$$

- Using this alternative definition, many standard results in ODE theory sat that a pth order accurate method should have an LTE that is $O(k^{p+1})$.
- But here we will call $\mathcal{L}^n$ the one-step error, since this can be viewed as the error that would be introduced in one time step if the past values $U^n, U^{n-1}, \cdots$ were all taken to be the exact values from $u(t)$.
- For example, in the midpoint method, we suppose that $U^n = u(t_n)$ and $U^{n-1} = u(t_{n-1})$, then

$$U^{n+1} = u(t_{n-1}) + 2kf(u(t_n))$$
$$\Longrightarrow u(t_{n+1}) - U^{n+1} = u(t_{n+1}) - u(t_{n-1}) - 2kf(u(t_n)) = \mathcal{L}^n = O(k^3).$$

  We see that in one step the error introduced is $O(k^3)$.
- If we want to compute an approximation to the true solution $u(T)$ at some fixed time T, we need to take T/k time steps. Then a rough estimate of the global error at time T is taking sum of all one-step errors,

$$\text{GE} \approx O(k^3) \times T/k = O(k^2).$$

Taylor series methods

Consider the IVP: $u'(t) = f(u(t), t)$ for $t > 0$ with the initial condition $u(0) = \eta$.

- By the Taylor series expansion, we have

$$u(t_{n+1}) = u(t_n) + ku'(t_n) + \frac{1}{2}k^2 u''(t_n) + \cdots .$$

  If we drop all terms of order k^2 and higher, we obtain

$$u(t_{n+1}) = u(t_n) + kf(u(t_n), t_n) + O(k^2).$$

  This suggests the forward Euler method

$$U^{n+1} = U^n + kf(U^n, t_n).$$

  and the *one-step error* is $O(k^2)$.

- A Taylor series method of higher accuracy can be derived by keeping more terms in the Taylor series. If we keep the first $p+1$ terms of the Taylor series expansion

$$u(t_{n+1}) \approx u(t_n) + ku'(t_n) + \frac{1}{2}k^2 u''(t_n) + \cdots + \frac{1}{p!}k^p u^{(p)}(t_n),$$

  then we obtain a pth order accurate method.

Taylor series methods (continued)

The problem is that we are given only

$$u'(t) = f(u(t), t)$$

and we must compute the higher derivatives by repeated differentiation of this function. For example, we can compute

$$\begin{aligned} u''(t) &= f_u(u(t), t)u'(t) + f_t(u(t), t) \\ &= f_u(u(t), t)f(u(t), t) + f_t(u(t), t). \end{aligned}$$

Example. Consider the equation $u'(t) = t^2 \sin(u(t))$. Then we can compute

$$\begin{aligned} u''(t) &= 2t \sin(u(t)) + t^2 \cos(u(t))u'(t) \\ &= 2t \sin(u(t)) + t^4 \cos(u(t)) \sin(u(t)). \end{aligned}$$

Hence, a second order method is given by

$$U^{n+1} = U^n + kt_n^2 \sin(U^n) + \frac{1}{2}k^2 \Big(2t_n \sin(U^n) + t_n^4 \cos(U^n) \sin(U^n)\Big).$$

Advantages and disadvantages of Taylor-series methods

- Disadvantages:
  - The method depends on repeated differentiation of the differential equation, unless we intend to use only the method of order 1. Thus, $f(u,t)$ must have partial derivatives of sufficient high order in the region where are solving the problem. Such an assumption is not necessary for the existence of a solution.
  - The various derivatives formula need to be programmed.
- Advantages:
  - Conceptual simplicity.
  - Potential for high precision: If we get e.g. 10 derivatives of $u(t)$, then the method is order 10 (i.e., terms up to and including the one involving k^{10}).

Basic concepts of Runge-Kutta methods

We wish to approximate the following nonautonomous IVP:

$$\begin{cases} u'(t) & = & f(u(t),t), \\ u(0) & = & \eta. \end{cases}$$

- From the Taylor theorem, we have

$$u(t+k) = u(t) + ku'(t) + \frac{k^2}{2!}u''(t) + O(k^3).$$

- By the chain rule, we obtain

$$\begin{cases} u''(t) & = & f_t + f_u u' = f_t + f_u f, \\ u'''(t) & = & f_{tt} + f_{tu} f + (f_t + f_u f) f_u + f(f_{ut} + f_{uu} f). \end{cases}$$

Basic concepts of Runge-Kutta methods (continued)

- In the Taylor expansion, we have

$$\begin{aligned} u(t+k) &= u(t) + kf(u,t) + \frac{k^2}{2}(f_t(u,t) + f_u(u,t)f(u,t)) + O(k^3) \\ &= u(t) + \frac{k}{2}f(u,t) + \frac{k}{2}\left[f(u,t) + kf_t(u,t) + kf_u(u,t)f(u,t))\right] + O(k^3) \\ &= u(t) + \frac{k}{2}f(u,t) + \frac{k}{2}f(u + kf(u,t), t+k) + O(k^3). \end{aligned}$$

- Note that the term in the square blankets above can be obtained by the Taylor expansion in two variables

$$f(u + kf(t,u), t+k) = f(u,t) + kf_t(u,t) + kf(u,t)f_u(u,t) + O(k^2).$$

A second-order Runge-Kutta method

- Therefore, we have

$$u(t+k) = u(t) + \frac{k}{2} f(u,t) + \frac{k}{2} f(u + kf(u,t), t+k) + O(k^3).$$

Then a 2nd-order Runge-Kutta (RK) method is given by

$$U^{n+1} = U^n + \frac{k}{2}\Big(F_1 + F_2\Big),$$

where

$$F_1 = f(U^n, t_n) \quad \text{and} \quad F_2 = f(U^n + kF_1, t_n + k).$$

- It is a multistage (two-stage) explicit method.
- It is known as Heun's method.

The general second-order Runge-Kutta method

- In general, the 2nd order RK method needs

$$\begin{aligned} u(t+k) &= u(t) + \omega_1 kf + \omega_2 kf(u + \beta kf, t + \alpha k) + O(k^3), \\ &= u(t) + \omega_1 kf + \omega_2 k\Big(f + \alpha k f_t + \beta k f f_u\Big) + O(k^3). \end{aligned}$$

- Compare with

$$u(t+k) = u(t) + kf + \frac{k^2}{2}\Big(f_t + f_u f\Big) + O(k^3),$$

we have

$$\begin{aligned} \omega_1 + \omega_2 &= 1, \\ \omega_2 \alpha &= 1/2, \\ \omega_2 \beta &= 1/2. \end{aligned}$$

The modified Euler method

- The previous method (Heun's method) is obtained by setting

$$\begin{cases} \omega_1 = \omega_2 = 1/2, \\ \alpha = \beta = 1. \end{cases}$$

- Setting

$$\begin{cases} \omega_1 = 0, \\ \omega_2 = 1, \\ \alpha = \beta = 1/2, \end{cases}$$

  we have

$$u(t+k) = u(t) + kf\left(u + \frac{k}{2}f(u,t), t + \frac{1}{2}k\right) + O(k^3).$$

  Then we obtain the following modified Euler method:

$$U^{n+1} = U^n + kF_2,$$

  where $F_1 = f(U^n, t_n)$ and $F_2 = f(U^n + \frac{k}{2}F_1, t_n + \frac{1}{2}k)$.

Fourth-order RK methods

- The derivations of higher order RK methods are tedious. However, the formulas are rather elegant and easily programmed once they have been derived.
- The most popular 4th order RK is:

$$U^{n+1} = U^n + \frac{k}{6}\Big(F_1 + 2F_2 + 2F_3 + F_4\Big),$$

where

$$\begin{cases} F_1 &= f(U^n, t_n), \\ F_2 &= f\Big(U^n + \frac{k}{2}F_1, t_n + \frac{k}{2}\Big), \\ F_3 &= f\Big(U^n + \frac{k}{2}F_2, t_n + \frac{k}{2}\Big), \\ F_4 &= f(U^n + kF_3, t_n + k). \end{cases}$$