Introduction to the Numerical Solution of IVPs for ODEs

Well-Posed Problems

$$\text{IVP} \qquad x' = f(t,x), \quad \text{with} \quad x(a) = x_a$$

Assume $f(t,x)$ is continuous in t, x and uniformly Lipschitz in x (with Lipschitz constant L) on $I \times \mathbb{R}^n$ with $I = [a,b]$.

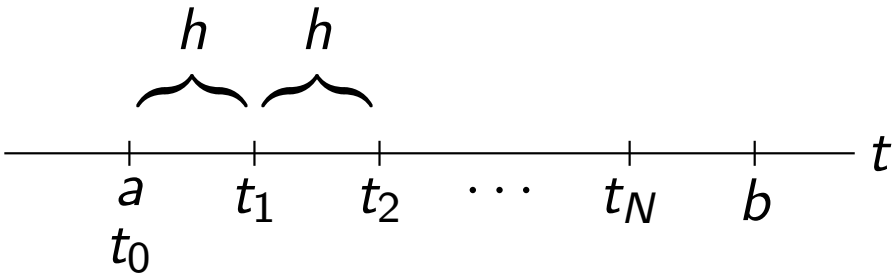
(1) **Existence.** There exists a solution of the IVP on $[a,b]$.

(2) **Uniqueness.** The solution, for each given x_a, is unique.

(3) **Continuous Dependence.** The solution depends continuously on the data.

The map $x_a \mapsto x(t,x_a)$ is continuous from $\mathbb{R}^n$ into $(C([a,b]), \|\cdot\|_\infty)$.

Grids



Choose a mesh width h (with $0 < h \leq b - a$), and let $N = \left[\frac{b-a}{h}\right]$ (greatest integer $\leq (b-a)/h$). Let

$$t_i = a + ih \qquad (i = 0, 1, \ldots, N)$$

be the grid points in t (note: $t_0 = a$), and let x_i denote the approximation to $x(t_i)$. Note that t_i and x_i depend on h, but we will usually suppress this dependence in our notation.

Explicit One-Step Methods

Start with $x_0 \approx x_a$.

Recursively compute $x_1, \ldots, x_N$ by

$$x_{i+1} = x_i + h\psi(h, t_i, x_i), \qquad i = 0, \ldots, N-1.$$

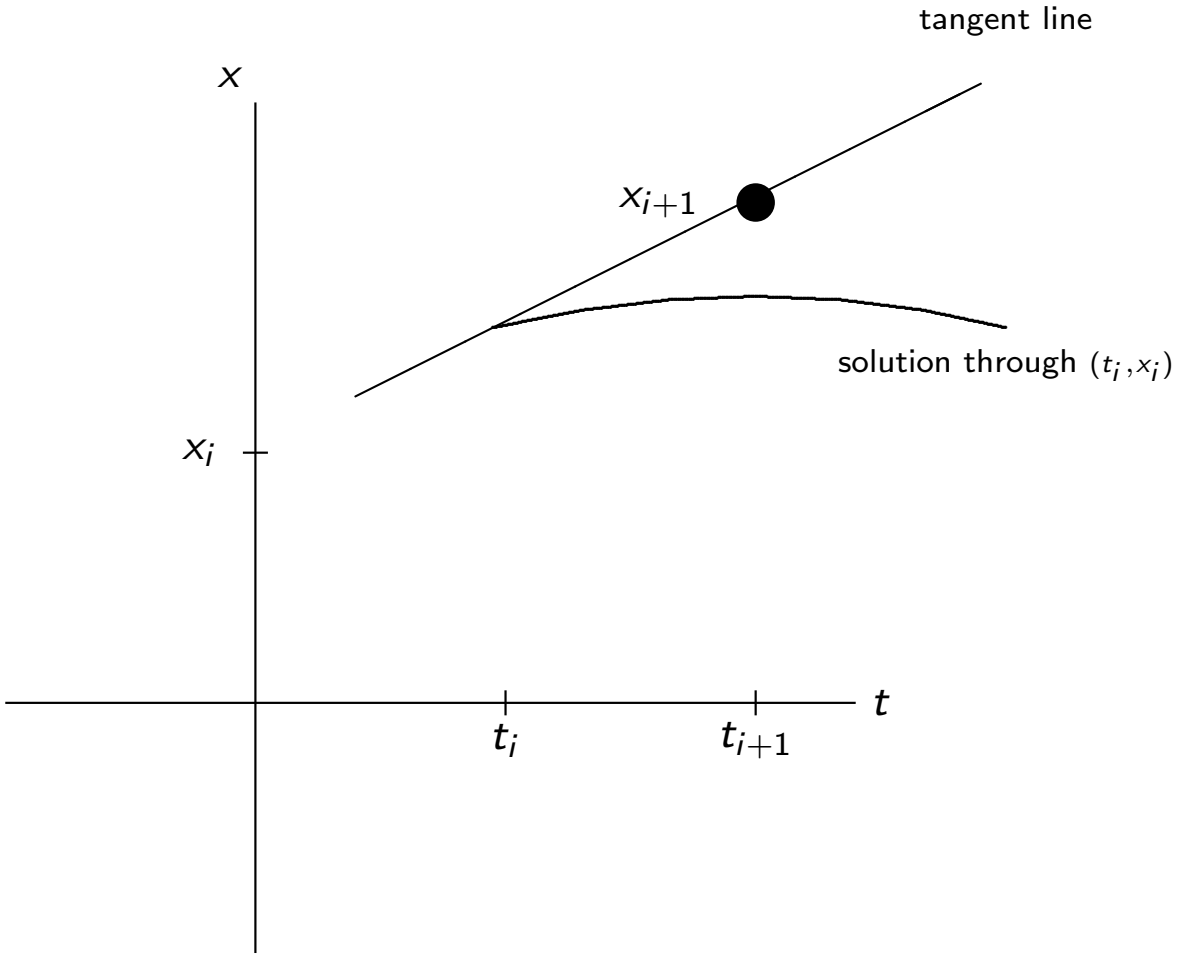
$\psi(h, t, x)$ is a function defined for

$$0 \leq h \leq b - a, \quad a \leq t \leq b, \quad x \in \mathbb{R}^n$$

which depends on $f(t, x)$.

Euler's Method: $\psi(h, t, x) := f(t, x)$

$$x_{i+1} = x_i + hf(t_i, x_i)$$



Taylor Methods

Let p be an integer ≥ 1 and consider the pth-order Taylor expansion of a C^{p+1} solution $x(t)$ of $x' = f(t,x)$:

$$x(t+h) = x(t) + hx'(t) + \cdots + \frac{h^p}{p!}x^{(p)}(t) + \underbrace{O(h^{p+1})}_{\text{remainder term}}$$

where $O(h^{p+1})$ is the Taylor's Theorem remainder term. Replace $x'(t), x''(t), \ldots$ by expressions involving f and its derivatives:

$$\begin{aligned}
x'(t) &= f(t,x(t)) \\
x''(t) &= \frac{d}{dt}(f(t,x(t))) = \overset{(n\times 1)}{D_t f}\Big|_{(t,x(t))} + \overset{(n\times n)}{D_x f}\Big|_{(t,x(t))} \overset{(n\times 1)}{\frac{dx}{dt}} \\
&= (D_t f + (D_x f)f)\Big|_{(t,x(t))} \qquad (\text{for } n=1, \text{ this is } f_t + f_x f),
\end{aligned}$$

and so forth.

Taylor Methods

This yields

$$p = 1: \quad x_{i+1} = x_i + hf(t_i, x_i) \quad \text{(Euler's method, } \psi(h,t,x) = f(t,x)\text{)}$$

$$p = 2: \quad x_{i+1} = x_i + hf(t_i, x_i) + \frac{h^2}{2}\left(D_t f + (D_x f) f\right)\Big|_{(t_i, x_i)}$$

For the case $p = 2$, we have

$$\psi(h,t,x) = T_2(h,t,x) \equiv \left(f + \frac{h}{2}\left(D_t f + (D_x f) f\right)\right)\Big|_{(t,x)}.$$

We will use the notation $T_p(h,t,x)$ to denote the $\psi(h,t,x)$ function for the Taylor method of order p.

Modified Euler's Method

$$\begin{aligned} x_{i+1} &= x_i + hf\left(t_i + \tfrac{h}{2}, x_i + \tfrac{h}{2} f(t_i, x_i)\right) \\ (\text{so } \psi(h,t,x) &= f\left(t + \tfrac{h}{2}, x + \tfrac{h}{2} f(t,x)\right)). \end{aligned}$$

Here $\psi(h,t,x)$ tries to approximate

$$x'\left(t + \tfrac{h}{2}\right) = f\left(t + \tfrac{h}{2}, x\left(t + \tfrac{h}{2}\right)\right),$$

using the Euler approximation to $x\left(t + \frac{h}{2}\right)$ $\left(\approx x(t) + \frac{h}{2} f(t, x(t))\right)$.

Improved Euler's Method (or Heun's Method)

$$x_{i+1} = x_i + \frac{h}{2}\left(f(t_i, x_i) + f\left(t_{i+1}, x_i + hf(t_i, x_i)\right)\right)$$

$$\left(\text{so } \psi(h, t, x) = \frac{1}{2}\left(f(t, x) + f\left(t + h, x + hf(t, x)\right)\right)\right).$$

Here again $\psi(h, t, x)$ tries to approximate

$$x'\left(t + \tfrac{h}{2}\right) \approx \tfrac{1}{2}(x'(t) + x'(t + h)).$$

Or $\psi(h, t, x)$ can be viewed as an approximation to the trapezoid rule applied to

$$\frac{1}{h}\left(x(t + h) - x(t)\right) = \frac{1}{h}\int_t^{t+h} x' \approx \tfrac{1}{2}x'(t) + \tfrac{1}{2}x'(t + h).$$

Consistency

Modified Euler and Improved Euler are examples of 2^{nd} order two-stage Runge-Kutta methods. Notice that no derivatives of f need be evaluated, but f needs to be evaluated *twice* in each step (from x_i to x_{i+1}).

It is evident from the above discussion that $\psi(h, t, x(t))$ should approximate $x'(t)$ as $h \to 0$ if $x(t)$ is a solution of the differential equation. Since $x' = f(t, x)$ for a solution, one expects that any useful method will satisfy the following condition.

Definition. A one-step method is called *consistent* if

$$\psi(0, t, x) = f(t, x).$$

All the methods described above are consistent.

Local Truncation Error

Let $x_{i+1} = x_i + h\psi(h, t_i, x_i)$ be a one-step method, and let $x(t)$ be a solution of the DE $x' = f(t, x)$. The *local truncation error* (LTE) for $x(t)$ is defined to be

$$l(h, t) \equiv x(t + h) - (x(t) + h\psi(h, t, x(t))) = \text{actual} - \text{predicted}.$$

$l(h, t)$ is defined for $0 < h \leq b - a$ and $a \leq t \leq b - h$.

Define

$$\tau(h, t) = \frac{l(h, t)}{h} \quad \text{and} \quad \tau(h) = \max_{a \leq t \leq b-h} |\tau(h, t)|.$$

Set $\tau_i(h) = \tau(h, t_i)$.

Characterizing Consistency

Proposition. Consider the one-step method

$$x_{i+1} = x_i + h\psi(h, t_i, x_i),$$

where $\psi(h, t, x)$ is continuous for $0 \leq h \leq h_0$, $a \leq t \leq b$, $x \in \mathbb{R}^n$ for some $h_0 \in (0, b - a]$.

This method is consistent with the DE $x' = f(t, x)$ if and only if

$$\tau(h) \to 0 \quad \text{as} \quad h \to 0^+ \quad \text{for all } C^1 \text{ solutions } x(t),$$

or, equivalently,

$$l(h, t) = o(h) \quad \text{as} \quad h \to 0^+ \quad \text{for all } C^1 \text{ solutions } x(t).$$

Proof

$\Rightarrow$: Fix a solution $x(t)$. For $0 < h \le h_0$, let

$$Z(h) = \max_{a \le s,t \le b,\ |s-t| \le h} |\psi(0,s,x(s)) - \psi(h,t,x(t))|.$$

By uniform continuity, $Z(h) \to 0$ as $h \to 0^+$. Now

$$\begin{aligned} I(h,t) &= x(t+h) - x(t) - h\psi(h,t,x(t)) \\ &= \int_t^{t+h} \left[x'(s) - \psi(h,t,x(t))\right] ds \\ &= \int_t^{t+h} \left[f(s,x(s)) - \psi(h,t,x(t))\right] ds \\ &= \int_t^{t+h} \left[\psi(0,s,x(s)) - \psi(h,t,x(t))\right] ds, \end{aligned}$$

so $|I(h,t)| \le hZ(h)$. Therefore $\tau(h) \le Z(h) \to 0$.

Proof

$\Leftarrow$: Conversely, suppose $\tau(h) \to 0$. For any $t \in [a, b)$ and any $h \in (0, b - t]$,

$$\frac{x(t+h) - x(t)}{h} = \psi(h, t, x(t)) + \tau(h, t).$$

Taking the limit as $h \downarrow 0$ gives $f(t, x(t)) = x'(t) = \psi(0, t, x(t))$.

Accurate of Order p

The Proposition states that consistency is equivalent to the condition that $l(h,t) = o(h)$ as $h \to 0^+$. For most useful methods, $l(h,t)$ actually goes to zero much more rapidly.

Definition. A one-step method is called *accurate of order* p (for a positive integer p) if for any solution $x(t)$ of the DE $x' = f(t,x)$, where f is C^p, we have $l(h,t) = O(h^{p+1})$.

Consistency is a minimal version of accuracy. It can be thought of as the correct notion of accuracy of order 0.

Example: Taylor method of order p

If $f \in C^p$, then $x \in C^{p+1}$, and

$$\begin{aligned} I(h,t) &= x(t+h) - \left(x(t) + hx'(t) + \cdots + \frac{h^p}{p!} x^{(p)}(t) \right) \\ &= \frac{1}{p!} \int_t^{t+h} (t+h-s)^p x^{(p+1)}(s) ds. \end{aligned}$$

So

$$|I(h,t)| \leq M_{p+1} \frac{h^{p+1}}{(p+1)!} \quad \text{where} \quad M_{p+1} = \max_{a \leq t \leq b} |x^{(p+1)}(t)|.$$

Characterization of p Order Accuracy

A one-step method $x_{i+1} = x_i + h\psi(h, t_i, x_i)$ is accurate of order p if and only if

$$\psi(h, t, x) = T_p(h, t, x) + O(h^p),$$

where T_p is the "ψ" for the Taylor method of order p.

Proof.

Since $x(t+h) - x(t) = hT_p(h, t, x(t)) + O(h^{p+1})$,
we have for any given one-step method that

$$\begin{aligned} l(h,t) &= x(t+h) - x(t) - h\psi(h, t, x(t)) \\ &= hT_p(h, t, x(t)) + O(h^{p+1}) - h\psi(h, t, x(t)) \\ &= h(T_p(h, t, x(t)) - \psi(h, t, x(t))) + O(h^{p+1}). \end{aligned}$$

So $l(h, t) = O(h^{p+1})$ iff $h(T_p - \psi) = O(h^{p+1})$ iff $\psi = T_p + O(h^p)$.

□

Convergence of One-Step Methods

Theorem Suppose $f(t,x)$ is continuous in t,x and uniformly Lipschitz in x on $[a,b]\times\mathbb{R}^n$. Suppose that satisfies

1. (*Stability*) $\psi(h,t,x)$ is continuous in h,t,x and uniformly Lipschitz in x (with Lipschitz constant K) on $0\le h\le h_0$, $a\le t\le b$, $x\in\mathbb{R}^n$ for some $h_0>0$ with $h_0\le b-a$, and
2. (*Consistency*) $\psi(0,t,x)=f(t,x)$.

Let $x(t)$ be the solution of the IVP $x'=f(t,x)$, $x(a)=x_a$ on $[a,b]$. Let $e_i(h)=x(t_i(h))-x_i(h)$, where $x_i(h)$ is obtained from the one-step method $x_{i+1}(h)=x_i(h)+h\psi(h,t_i(h),x_i(h))$, and set $e_0(h)=x_a-x_0(h)$ (the error in the initial value $x_0(h)$). Then

$$|e_i(h)|\le e^{K(t_i(h)-a)}|e_0(h)|+\tau(h)\left(\frac{e^{K(t_i(h)-a)}-1}{K}\right),\ \text{so}$$

$$|e_i(h)|\le e^{K(b-a)}|e_0(h)|+\frac{e^{K(b-a)}-1}{K}\tau(h).$$

Moreover, $\tau(h)\to 0$ as $h\to 0$. Therefore, if $e_0(h)\to 0$ as $h\to 0$, then

$$\max\left\{|e_i(h)|\ :\ 0\le i\le h^{-1}(b-a)\right\}\ \to 0\quad\text{as}\quad h\to 0,$$

that is, the approximations converge uniformly on the grid to the solution.

Proof

Hold $h > 0$ fixed, and ignore rounding error. Subtracting

$$x_{i+1} = x_i + h\psi(h, t_i, x_i)$$

from

$$x(t_{i+1}) = x(t_i) + h\psi(h, t_i, x(t_i)) + h\tau_i, \qquad (\tau_i := \tau(h, t_i))$$

gives

$$\begin{aligned} |e_{i+1}| &\leq |e_i| + h|\psi(h, t_i, x(t_i)) - \psi(h, t_i, x_i)| + h|\tau_i| \\ &\leq |e_i| + hK|e_i| + h\tau(h) \\ &= (1 + hK)|e_i| + h\tau(h). \end{aligned}$$

So

$$\begin{aligned} |e_1| &\leq (1 + hK)|e_0| + h\tau(h), \quad \text{and} \\ |e_2| &\leq (1 + hK)|e_1| + h\tau(h) \\ &\leq (1 + hK)^2|e_0| + h\tau(h)(1 + (1 + hK)). \end{aligned}$$

Proof

By induction,

$$\begin{aligned} |e_i| &\leq (1+hK)^i|e_0| + h\tau(h)(1+(1+hK)+(1+hK)^2+\cdots+(1+hK)^{i-1}) \\ &= (1+hK)^i|e_0| + h\tau(h)\frac{(1+hK)^i-1}{(1+hK)-1} \\ &= (1+hK)^i|e_0| + \tau(h)\frac{(1+hK)^i-1}{K} \end{aligned}$$

Since $(1+hK)^{\frac{1}{h}} \uparrow e^K$ as $h \to 0^+$ (for $K > 0$), and $i = \frac{t_i - a}{h}$, we have

$$(1+hK)^i = (1+hK)^{\frac{t_i-a}{h}} \leq e^{K(t_i-a)}.$$

Thus

$$|e_i| \leq e^{K(t_i-a)}|e_0| + \tau(h)\frac{e^{K(t_i-a)}-1}{K}.$$

The preceding proposition shows $\tau(h) \to 0$, and the theorem follows.

pth order convergence

If $f \in C^p$, then $x(t) \in C^{p+1}$, so the theorem implies that if a one-step method is accurate of order p and stable [i.e. ψ is Lipschitz in x], then

$$I(h,t) = O(h^{p+1}) \quad \text{and thus} \quad \tau(h) = O(h^p).$$

If, in addition, $e_0(h) = O(h^p)$, then

$$\max_i |e_i(h)| = O(h^p),$$

i.e. we have p^{th} order convergence of the numerical approximations to the solution uniformly on $[a, b]$.

Explicit Runge-Kutta Methods

A problem with Taylor methods is the need to evaluate higher derivatives of f. Runge-Kutta (RK) methods only require the evaluation of f when going from x_i to x_{i+1}.

An m-stage (explicit) RK method is of the form

$$x_{i+1} = x_i + h\psi(h, t_i, x_i),$$

with

$$\psi(h, t, x) = \sum_{j=1}^{m} a_j k_j(h, t, x),$$

where $a_1, \ldots, a_m$ are given, $k_1(h, t, x) = f(t, x)$, and for $2 \leq j \leq m$,

$$k_j(h, t, x) = f\left(t + \alpha_j h, x + h \sum_{r=1}^{j-1} \beta_{jr} k_r(h, t, x)\right)$$

with $\alpha_2, \ldots, \alpha_m$ and β_{jr} $(1 \leq r < j \leq m)$ given constants.

Explicit Runge-Kutta Methods

Since $k_j(0, t, x) = f(t, x)$, the method is consistent if and only if $\sum_{j=1}^{m} a_j = 1$, so this condition will always be imposed.

We usually choose $0 < \alpha_j \leq 1$, and for accuracy reasons we choose

$$\alpha_j = \sum_{r=1}^{j-1} \beta_{jr} \qquad (2 \leq j \leq m). \qquad (*)$$

Examples: $m = 2$

$$x_{i+1} = x_i + h(a_1 k_1(h, t_i, x_i) + a_2 k_2(h, t_i, x_i))$$

where

$$\begin{aligned} k_1(h, t_i, x_i) &= f(t_i, x_i) \\ k_2(h, t_i, x_i) &= f(t_i + \alpha_2 h, x_i + h\beta_{21} k_1(h, t_i, x_i)). \end{aligned}$$

For simplicity, write α for α_2 and β for β_{21}. Expanding in h,

$$\begin{aligned} k_2(h, t, x) &= f(t + \alpha h, x + h\beta f(t, x)) \\ &= f(t, x) + \alpha h D_t f(t, x) + (D_x f(t, x))(h\beta f(t, x)) + O(h^2) \\ &= [f + h(\alpha D_t f + \beta (D_x f) f)](t, x) + O(h^2). \end{aligned}$$

So

$$\psi(h, t, x) = (a_1 + a_2) f + h(a_2 \alpha D_t f + a_2 \beta (D_x f) f) + O(h^2).$$

Examples: $m = 2$

Recalling that

$$T_2 = f + \frac{h}{2}(D_t f + (D_x f)f),$$

and that the method is accurate of order two if and only if

$$\psi = T_2 + O(h^2),$$

we obtain the following necessary and sufficient conditions on a two-stage (explicit) RK method to be accurate of order two:

$$a_1 + a_2 = 1, \quad a_2\alpha = \tfrac{1}{2}, \quad \text{and} \quad a_2\beta = \tfrac{1}{2}.$$

We require $\alpha = \beta$ as in (*) (we now see why this condition needs to be imposed), whereupon these conditions become:

$$\boxed{a_1 + a_2 = 1, \qquad a_2\alpha = \tfrac{1}{2}}.$$

Therefore, there is a one-parameter family (e.g., parameterized by α) of 2^{nd} order, two-stage ($m = 2$) explicit RK methods.

Examples: $m = 2$

Two instances of this one parameter family are $\alpha = \frac{1}{2}, 1$.

(1) Setting $\alpha = \frac{1}{2}$ gives $a_2 = 1$, $a_1 = 0$, yielding the Modified Euler method.

(2) Choosing $\alpha = 1$ gives $a_2 = \frac{1}{2}$, $a_1 = \frac{1}{2}$, yielding the Improved Euler method, or Heun's method.

The Popular 4^{th} Order Four-Stage RK Method

$$x_{i+1} = x_i + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

where

$$\begin{aligned} k_1 &= f(t_i, x_i) \\ k_2 &= f\left(t_i + \tfrac{h}{2}, x_i + \tfrac{h}{2}k_1\right) \\ k_3 &= f\left(t_i + \tfrac{h}{2}, x_i + \tfrac{h}{2}k_2\right) \\ k_4 &= f(t_i + h, x_i + hk_3). \end{aligned}$$