## Contents

7 Numerical Integration of Ordinary Differential Equations 2

7.1 Euler step ........................................... 3
7.2 Heun step ........................................... 8
7.3 Runge Kutta 4 step .................................. 8
7.4 A method for second order differential equations .......... 9
7.5 Global Error Estimate Revisited ........................ 10
7.6 Error estimate in practise ............................. 11
7.7 Implicit Methods ..................................... 14
7.8 Linear Multi-step Methods ............................ 15
Chapter 7

Numerical Integration of Ordinary Differential Equations

Up until now solved for \( \ddot{q} \) from

\[
M \ddot{q} = \dot{Q}(q, \dot{q}, t)
\]

(7.1)

or

\[
\begin{pmatrix}
M & D^T \\
D & 0 \\
\end{pmatrix}
\begin{pmatrix}
\ddot{x} \\
\lambda \\
\end{pmatrix}
= \begin{pmatrix}
f \\
-D^2 \dddot{x} \\
\end{pmatrix}
\]

(7.2)

But what we really want to know is \( q(t) \).

So given \( t_0, q_0 \) and \( \dot{q}_0 \) and \( \ddot{q}_0 \) we want to know \( q_1 \) and \( \dot{q}_1 \) for time \( t_1 = t_0 + h \).

A natural, but naive, way would be,

\[
q_1 = q_0 + h \dot{q}_0, \\
\dot{q}_1 = \dot{q}_0 + \frac{1}{2} h^2 \dddot{q}_0.
\]

Which finds its origin in the Taylor expansion of a function \( f(x) \).

\[
f(x + h) = f(x) + hf'(x) + \frac{1}{2} h^2 f''(x) + \cdots + \frac{1}{n!} h^n f^n(x) + O(h^{n+1})
\]

(7.3)

Knowing this you could raise the question if it were not better to use,

\[
q_1 = q_0 + h \dot{q}_0 + \frac{1}{2} h^2 \dddot{q}_0 ?
\]

An ordinary differential equation (ODE) is usually expressed as, \( y' = f(x, y) \) with initial values \( y(x_0) = y_0 \). We have, \( \dot{y} = f(t, y) \) with the initial values at \( t = t_0, y(t_0) = y_0 \). Well actually we have a set of \( n \) second order differential equations,

\[
M \ddot{q} = \dot{Q}(t, q, \dot{q}),
\]

(7.4)
with, \(q(t_0) = q_0\) and \(\dot{q}(t_0) = \dot{q}_0\). Which we can transform into first order differential equations by substitution of \(u = \dot{q}\) into (7.4),

\[
Mu = \dot{Q}(t, q, u).
\]

This lead to a set of 2n first order differential equations,

\[
\begin{align*}
\dot{q} &= u \\
u &= M^{-1}\dot{Q}(t, q, u) \\
qu(0) &= q_0 \\
u(0) &= u_0,
\end{align*}
\]

which we can write in the standard first order form

\[
\dot{y} = f(t, y) \quad y(t_0) = y_0 \quad \text{with} \quad y = \begin{pmatrix} q \\ u \end{pmatrix}.
\] (7.5)

Let’s look at a simple example, a mass-spring-damper system. The equation of motion for this system with linear spring-damper is,

\[
m\ddot{x} = mg - F_s - F_d = mg - kx - cx.
\]

Substitute \(\dot{x} = u\) in the previous equation results in,

\[
\dot{u} = g - \frac{k}{m}x - \frac{c}{m}u.
\]

And rewritten in the standard first order form

\[
\begin{pmatrix} \dot{x} \\ \dot{u} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{c}{m} \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} + \begin{pmatrix} 0 \\ g \end{pmatrix}.
\] (7.7)

And finally with initial conditions for example starting from rest, \(x = 0\) and \(\dot{x} = 0\), or \(y_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\).

### 7.1 Euler step

The first method, \(y_{n+1} = y_n + hf(t_n, y_n)\), is known as an Euler step. A somewhat more refined method is,

\[
y^*_{n+1} = y_n + hf(t_n, y_n)
\]

\[
y_{n+1} = y_n + \frac{h}{2} \left( f(t_n, y_n) + f(t_{n+1}, y^*_{n+1}) \right)
\]
or Heun’s method. This method takes a sort of average of \( f(y') \) into account. Note that predictor \( y_{n+1}^* \) is a normal Euler step. What do we mean by more refined? ⇒ More

![Figure 7.2: Sketch of the Heun integration method](image)

accurate results! So let’s look at the accuracy of \( y \) after one step, the so-called truncation error. The truncation error \( \epsilon \) (after one step) for the two methods follows from the Taylor expansion and are,

\[
\begin{align*}
\text{Euler} & \quad \epsilon = O\left(h^2\right) \\
\text{Heun} & \quad \epsilon = O\left(h^3\right)
\end{align*}
\]

Far more important is the error in final result, or the so-called global error. This global truncation error \( E \) is the result after integration over the time-span \( t = 0 \cdots T \), which is the sum of \( (T/h) \) local true errors,

\[
E = \left(\frac{T}{h}\right) \epsilon
\]

So the global truncation error for the two methods are,

\[
\begin{align*}
\text{Euler} & \quad E = O\left(h\right) \\
\text{Heun} & \quad E = O\left(h^2\right)
\end{align*}
\]

These are all order of magnitude things. In practice we usually do not know the value of this error and sometimes even not know the order of magnitude. But we can estimate the global error. Let’s look at a simple one-dimensional problem. We assume that the approximate value is

\[
y = \hat{y} + C_1 h^p, \quad (7.8)
\]

with the true answer (which we do not know) \( \hat{y} \) and the truncation error \( C_1 h^p \). For Euler \( p = 1 \) and for Heun \( p = 2 \). This equation has two unknowns, the true value \( \hat{y} \) and the coefficients \( C_1 \). We can determine these values by integration the system twice with two different stepsizes. We first integrate from \( t = 0 \) to \( t = T \) with step size \( h \), call this result \( y_h \). Next, we integrate again from \( t = 0 \)
to $t = T$, but now with half the stepsize $h/2$ and we call this $y_{h/2}$. This results in the following two equations,

$$
\begin{align*}
    y_h &= \hat{y} + C_1 h^p \\
    y_{h/2} &= \hat{y} + C_1 \left(\frac{h}{2}\right)^p.
\end{align*}
$$

From which we can solve $\hat{y}$ and $C_1$. But since these are error bounds, plus or minus, we can only give a global truncation error on our best result $y_{h/2}$.

$$
E = |\hat{y} - y_{h/2}| = \frac{1}{2^p - 1} |(y_{h/2} - y_h)|
$$

So this was accuracy. Far more important is stability, or how do previous made errors propagate? To investigate this we need a test equation. We will again look at a simple one-dimensional case and use the most generic and archetypal differential equation,

$$
\dot{y} = \lambda y.
$$

The general solution to this equation is $y = ce^{\lambda t}$ and with complex $\lambda = a + bi$ and complex $c$ we have the solutions of the form,

$$
y = ce^{(a+bi)t} = ce^{at}e^{bit}.
$$

Which we transform with the Euler identities $e^{i\phi} = \cos \phi + i \sin \phi$ into sines and cosines, as

$$
y = ce^{at}(\cos (bt) + i \sin (bt))
$$

We see that $e^{at}$ gives rise to growing or decaying solutions and that $\cos (bt) + i \sin (bt)$ is the oscillating part.

How does this connect to our mechanical systems? Lets look at the mass-spring-damper system again. The equation of motion is:

$$
m\ddot{x} + c\dot{x} + kx = f
$$

Look at the reduced equations for the homogenous solution $m\ddot{x} + c\dot{x} + kx = 0$ and introduce the eigenfrequency $\omega_0$, from $\omega_0^2 = k/m$, the relative damping $\beta$, from $2\beta\omega_0 = c/m$, which results in

$$
\ddot{x} + 2\beta\omega_0 \dot{x} + \omega_0^2 x = 0.
$$

Transform this second order ODE into 2 first order equations with $v = \dot{x}$

$$
\begin{pmatrix}
    \dot{x} \\
    \dot{v}
\end{pmatrix} = \begin{pmatrix}
    0 & 1 \\
    -\omega_0^2 & -2\beta\omega_0
\end{pmatrix} \begin{pmatrix}
    x \\
    v
\end{pmatrix}
$$

(7.11)

$$
\dot{y} = Ay
$$

assume solution of the form $y = y_0 e^{\lambda t}$

$$
\lambda y_0 e^{\lambda t} = Ay_0 e^{\lambda t} \text{ holds for any } t
$$

$$(A - \lambda I)y_0 = 0 \Rightarrow |A - \lambda I| = 0
$$
\[
\begin{vmatrix}
-\lambda & 1 \\
-\omega_0^2 & -\lambda - 2\beta\omega_0
\end{vmatrix}
= \lambda^2 + 2\beta\omega_0\lambda + \omega_0^2 = 0 \quad (7.12)
\]

\[
\lambda = -\beta_0\omega_0 \pm \sqrt{1 - \beta^2}i\omega_0
\]

Re(\(\lambda\)) \rightarrow "Damping" \quad Im(\(\lambda\)) \rightarrow "Oscillation"

So now we have an idea what type of motions are described by the test equation \(\dot{\mathbf{y}} = \lambda \mathbf{y}\).

After one numerical integration step we have,

\[
\mathbf{y} = \hat{\mathbf{y}} + \epsilon,
\]

with \(\hat{\mathbf{y}}\) as the true solution and \(\epsilon\) as the truncation error of the method.

\[
\dot{\hat{\mathbf{y}}} + \dot{\epsilon} = \lambda (\hat{\mathbf{y}} + \epsilon)
\]

So the true solution \(\dot{\hat{\mathbf{y}}} = \lambda \hat{\mathbf{y}}\) leads that the error can propagate:

\[
\dot{\epsilon} = \lambda \epsilon
\]

Re(\(\lambda\)) > 0: Stable only in a fixed interval.
Re(\(\lambda\)) < 0: Unconditionally stable.

Back to the test equation and the Method. Let’s with the Euler method,

\[
\mathbf{y}_{n+1} = \mathbf{y}_n + hf(t_n, \mathbf{y}_n) \quad \text{with} \quad f(t_n, \mathbf{y}_n) = \lambda \mathbf{y}_n
\]

\[
\begin{align*}
\mathbf{y}_{n+1} &= \mathbf{y}_n + h\lambda \mathbf{y}_n \\
\mathbf{y}_{n+1} &= (1 + h\lambda) \mathbf{y}_n \\
\mathbf{y}_{n+1} &= C(h\lambda) \mathbf{y}_n
\end{align*}
\]
where $C(h\lambda)$ is called the amplification factor. For the Euler method we have $C(h\lambda) = (1 + h\lambda)$.

For absolute stability $C(h\lambda) < 1$, where $\lambda$ can be complex, like in $h\lambda = a + bi$. So for stability we have to look at $|C(h\lambda)| < 1$, which for the Euler method results in $|1 + a + bi| < 1$. Definition stable: local errors do not propagate!

![Euler stability area in the complex coordinates](image)

Note that $h\lambda = bi$ is not in the stability region. Euler’s method is unstable for pure oscillating systems. For example the solution of an undamped pendulum will become infinity over a period of time.

![Pendulum solution using Euler](image)
7.2 Heun step

Let’s look at Heun:

\[ y_{n+1}^* = y_n + h\lambda y_n \]
\[ y_{n+1} = y_n + \frac{h}{2} \left( \lambda y_n + \lambda (y_n + h\lambda y_n) \right) \]
\[ y_{n+1} = y_n + \frac{h\lambda}{2} y_n + \frac{h\xi}{2} - y_n \]
\[ y_{n+1} = \left( 1 + (h\lambda) + \frac{1}{2} (h\lambda)^2 \right) y_n \]

where \( C (h\lambda) = \left( 1 + (h\lambda) + \frac{1}{2} (h\lambda)^2 \right) \).

\( z = h\lambda \) the condition becomes: \(|1 + z + 1/2z^2| < 1\).

From the figure it is shown that Heun is also unstable for pure oscillating systems.

![Figure 7.6: Heun stability area in the complex coordinates](image)

7.3 Runge Kutta 4 step

A very accurate and efficient method is the so-called Runge Kutta method,

\[
\begin{align*}
  k_1 &= f(t_n, y_n) \\
  k_2 &= f(t_n + \frac{h}{2}, y_n + \frac{h}{2} k_1) \\
  k_3 &= f(t_n + \frac{h}{2}, y_n + \frac{h}{2} k_2) \\
  k_4 &= f(t_n + h, y_n + h k_3) \\
  y_{n+1} &= y_n + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4)
\end{align*}
\]

The local truncation error is \( \epsilon = O(h^5) \) and the amplification factor for this method is: \( C (h\lambda) = 1 + (h\lambda) + 1/2 (h\lambda)^2 + 1/6 (h\lambda)^3 + 1/24 (h\lambda)^4 \).

This method is stable for pure oscillating systems.

For stability:
\( h\lambda < 2.8 \) and with \( \lambda = \omega_0 \)
Figure 7.7: Runge Kutta stability area in the complex coordinates

$h < 2.8/\omega_0$ or with $T_0 = 2\pi/\omega_0 \rightarrow h < (2.8/2\pi)T_0 \rightarrow h < 0.4T_0$

These were all 1-step methods for systems first order differential equations.

Figure 7.8: Runge Kutta maximum time step to guarantee stability

7.4 A method for second order differential equations

Now if you look at our system of differential equations half of them are pretty simple.

\[
\begin{align*}
\dot{q} &= z \\
\dot{z} &= f(t, q, z)
\end{align*}
\]

We can write this explicit in a very simple one (Euler 2)

\[
\begin{align*}
k_1 &= f(t_n + \frac{h}{2}, q_n + \frac{h}{2} z_n, z_n) \\
q_{n+1} &= q_n + h z_n + \frac{h^2}{2} k_1 \\
z_{n+1} &= z_n + h k_1
\end{align*}
\]

with $t_n + h/2$ is the midpoint and $z_n$ not interesting, just pick $z_n$. Note that only 1 function ($k_1$) calls to the differential equation at the midpoint.
The local truncation error is:

\[ \epsilon = O(h^2) \]

but if \( f \) is weak in \( z \) (the velocities)

\[ \epsilon = O(h^3) \]

To investigate the stability we use as a test equation the simplest second order equation that of a linear damped oscillatory system.

\[ \ddot{q} + 2\beta\omega_0 \dot{q} + \omega_0^2 q = 0 \]

and with \( \dot{q} = z \):

\[
\begin{pmatrix}
\dot{q} \\
z
\end{pmatrix} =
\begin{pmatrix}
0 & 1 \\
-\omega_0^2 & -2\beta\omega_0
\end{pmatrix}
\begin{pmatrix}
q \\
z
\end{pmatrix}
\]

which can be seen as a first order differential equation: \( \dot{y} = \lambda y \).

\[ \lambda_{1,2} = -\beta\omega_0 \pm \sqrt{1 - \beta^2 \omega_0^2}i \]

The stability comes out as:

Stable for pure oscillating systems.

Figure 7.9: Runge Kutta stability detail

An "arcade" scheme, suitable for the entertainment industry, moderate accuracy but fast only 1 function call per step. Note that this function call is at the time increment \( t_n + h/2 \). So in the end we know the position \( q \) and the velocity \( z \) at the discrete points but the acceleration \( \ddot{q} \) and all the forces etc. are only known at the midpoints.

### 7.5 Global Error Estimate Revisited

To get more accurate results you just decrease the step size \( h \), but \( \cdots \) we only looked at local truncation errors of the type \( \epsilon = O(h^{p+1}) \) or \( E = O(h^p) \). Since we work with finite word lengths in the computer to represent our numbers we also have round-off errors which re of the type \( \epsilon = C_2 \leftarrow \) a very small number but since we take the \( n = T/h \) steps the global or accumulated round-off error is:

\[ E = \frac{T}{h}C_2 = O\left(\frac{1}{h}\right) \]
In refining our step size we would hit this. So the fatal global error is like:

\[ E = C_1 h^p + \frac{C_2}{h} \]

For example Heun’s method we have \( E = C_1 h^2 + C_2/h \) which is shown in Figure (7.10), or better on a log-log scale in Figure (7.11).

![Figure 7.10: global error versus step size on linear scale](image)

![Figure 7.11: global error versus step size on logarithmic scale](image)

### 7.6 Error estimate in practice

The global error estimate has the term:

\[ y = \hat{y} + C_1 h^p + C_2 \frac{1}{h} \]
Now we expect to be either in the second term or in the third term to be dominant for the error estimate.

Dominant truncation: \( y = \hat{y} + C_1 h^p \)

Dominant round-off: \( y = \hat{y} + C_2/h \)

Take a number of successive steps with the step sizes \( h = T/2^n \) and calculate \( T = \text{timeperiod} \) (usually start with \( n = 6 \) or \( 8 \)) the differences at two successive solutions:

\[ D_n = |y_n - y_{n-1}| \]

where \( y_n \) is \( \int_0^T \) at step size \( h = T/2^n \) and plot this on a log-log scale.

Figure 7.12: global error of the truncation and round-off effect versus step size on logarithmic scale

Figure 7.13: global error of the truncation and round-off effect versus step size on logarithmic scale
Figure 7.14: the truncation and round-off error for different integration methods versus step size on logarithmic scale
7.7 Implicit Methods

These were all explicit methods, recall Forward Euler \( y_{n+1} = y_n + hf(t_n, y_n) \) in contrast to the Backward Euler \( y_{n+1} = y_n + hf(t_{n+1}, y_{n+1}) \).

We call this method implicit, in order to know \( y_{n+1} \) we need to evaluate \( f(t_{n+1}, y_{n+1}) \), so in general we need to solve these usually non-linear equations.

\[
y_{n+1} = y_n + hf(t_{n+1}, y_{n+1})
\]

Why better? Accuracy and stability.

Accuracy: The local truncation error \( \epsilon = O(h^2) \) is the same as for the forward Euler.

Stability: with the same test equation:

\[
\dot{y} = \lambda y
\]

We came up with:

\[
y_{n+1} = y_n + h\lambda y_{n+1}
\]

And for the error:

\[
\frac{\epsilon_{n+1}}{\epsilon_n} = \frac{\epsilon_n + h\lambda \epsilon_{n+1}}{1 - h\lambda \epsilon_n}
\]

The amplification factor \( C(h\lambda) = \frac{1}{1 - h\lambda} \). Stability is expressed by

\[
|C(h\lambda)| < 1.
\]

And with \( \lambda \) complex we get:

Even high frequency signals are not blown up! There are higher order implicit solutions which show the same stability behaviour.
7.8 Linear Multi-step Methods

Even earlier than Runge-Kutta they were devised by J.C. Adams to solve a problem of F. Bashfort. The origin of the method has been dated back to 1855 when F. Bashfort made an application to the Royal Society for assistance from the Government grant. There he wrote: "... but I am indebted to Mr. Adams for a method of rewriting the differential equation ... which gives the theoretical form of the drop with an accuracy exceeding that of the most refined measurements". Eventually the methods were published by Bashfort in 1883.

What happened in 1855 in the United states of America? In 1855 Whitman published at his own expense a volume of 12 poems, Leaves of Grass. Smith & Wesson invents the revolver. It was 6 years before the civil war would started.

\[
y' = f(t, y)
\]

This is the idea:

\[
y_1 = y_0 + \int_{t_0}^{t_1} f(t, y) \, dt
\]

Describe \( f \) in a Langrange polynomial as a function of the past and \( \int \) one step.
in the fictive with the interpolated function value. For instance the Lagrange polynomial for 3 \( f \)’s.

\[
p(t) = f_0 \frac{(t + 2h)(t + h)}{(h)(-h)} + f_1 \frac{(t + 2h)(t + h)}{(2h)(h)} + f_2 \frac{(t + h)(t - h)}{(-2h)(-h)}
\]

Approximate \( \int_{t_0}^{t_1} f(t, y) \, dt \) by \( \int_{0}^{h} p(t) \, dt \) which is \( h \left( \frac{5}{12}f_{-2} - 16/12f_{-1} + 23/12f_0 \right) \). Thus we get \( y_{n+1} = y_n + h/12 (5f_{n+1} - 16f_n - 1 + 5f_n - 2f_n) \) which is the explicit Adams-Bashfort method. Now you could make an implicit scheme by adding \( f_{n+1} \) in the polynomial resulting in:

\[
y_{n+1} = y_n + \frac{h}{12} (5f_{n+1} + 8f_n - f_{n-1})
\]

The implicit Adams-Moulton method.

We do not like implicit schemes because of the necessary iterations.

\[
y_{n+1} = y_n + h/24 (9f_{n+1} + 19f_n - 5f_{n-1} + f_{n-2})
\]

Now a very popular scheme is the following PECE (Shampine 1975)

Predict with \( A-B \) order \( k \) \( y_{n+1}^p \)

Evaluate the differential equation \( f_{n+1} (t_{n+1}, y_{n+1}^p) \)

Correct with \( A-M \) order \( k + 1 \) \( y_{n+1} \)

Evaluate the differential equations \( f_{n+1} (t_{n+1}, y_{n+1}) \)

This scheme uses only 2 future evaluations/step. This is the ode113 solver of Matlab!

Pro’s:

- Only 2 Function Evaluations/step independent of the order
- Interpolated values of \( y \) for “free”

Con’s:

- Startup is difficult
- Order \( k > 0 \) stability decreases

Change of step size \( h \) and order \( k \)?

Compare this with a carrace in reverse.
Finally

There are BDF-methods Backward Differentiating made popular by Gear 1971 subscribing DIFSUB. Again we build a Lagrange polynomial but now on $y_{-3}, \ldots, y_{-1}, y_0$ and predict $y_1$. But where is the differential equation? Differentiate the polynomial and set the result being $f$. For instance for a BDF with constant step size $h$ and order 2 has the form of:

$$
\frac{3}{2}y_1 - 2y_0 + \frac{1}{2}f_{-1} = hf(t_1y_n)
$$

$$
y_1 = -\frac{1}{3}y_{-1} + \frac{4}{3}y_0 + \frac{2}{3}hf(t_1y_n)
$$

Again, thus is an implicit scheme, and think of the startup problem. The methods are unconditionally stable up to order 6. This is ode15s with option BDF.

Before we start let’s introduce Auckland New Zealand 1987 the J.C. Butcher for a s-stage Runge-Kutta method:

\[ y_{n+1} = y_n + h \sum_{i=1}^{s} b_i k_i \]

with: \( k_i = f \left( t_n + c_i h, y_n + h \sum_{j=1}^{s} a_{ij} k_j \right) \).

The method is characterised by \( s \), the stages and the coefficients \( a_{ij} \), \( b_i \) and \( c_i \).

These coefficients are usually written in the following scheme: Butcher arrays.

\[
\begin{array}{c|ccc}
   & a_{11} \cdots a_{1s} \\
n & \vdots & \ddots & \\
   & a_{s1} \cdots a_{ss} \\
\hline
b_1 \cdots b_s
\end{array}
\]

or

\[
\begin{array}{c|ccc}
c_1 & a_1 & \cdots & a_s \\
\vdots & \ddots & \ddots & \ddots \\
c_s & a_s & \cdots & a_s
\end{array}
\]

So RK4 looks like

\[
\begin{align*}
k_1 &= f \left( t_n, y_n \right) \\
k_2 &= f \left( t_n + \frac{h}{2}, y_n + \frac{h}{2} k_1 \right) \\
k_3 &= f \left( t_n + \frac{h}{2}, y_n + \frac{h}{2} k_2 \right) \\
k_4 &= f \left( t_n + h, y_n + h k_3 \right)
\end{align*}
\]

\[
y_{n+1} = y_n + \frac{h}{6} \left( k_1 + 2 k_2 + 2 k_3 + k_4 \right)
\]

Now if \( a_{ij} = 0 \) \( j \geq i \) → Explicit Method

and \( a_{ij} = 0 \) \( j > i \) → Diagonal Implicit

else → Implicit sensotritio

( The number \( c_i \) have to fulfill by: \( c_i = \Sigma_{j=1}^{s} a_{ij} \) in order the integrated time dependent system which we make arbcenucus)

Methods with one stage, \( s = 1 \)

\[
\begin{array}{c|cc}
0 & 0 & 1 \\
1 & 0 & 0 \\
\end{array} \quad \text{Forward Euler}
\]

\[
\begin{array}{c|cc}
1 & 0 & 1 \\
0 & 1 & 0 \\
\end{array} \quad \text{Backward Euler}
\]

Explicit Methods with second stages, \( s = 2 \)

\[
\begin{array}{c|ccc}
0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 \\
\end{array} \quad \text{Heun} \quad \begin{pmatrix}
0 & 0 & 0 \\
\frac{1}{2} & 0 & 0 \\
1 - \alpha & \alpha & 0 \\
\end{pmatrix}
\]

\( \epsilon = O \left( h^3 \right) \)

Basis of the RKs-stage methods

ODE \( y' = f \left( y \right) \)

Taylor expansion solution is:

\[
y \left( t + h \right) = y \left( t \right) + h y' + \frac{h^2}{2} y'' + \cdots + \frac{h^p}{p!} y^p + O \left( h^{p+1} \right)
\]

Runge-Kutta 2 stage method explicit (hope for \( \epsilon = O \left( h^3 \right) \))

\[
\begin{array}{c|ccc}
0 & 0 & 0 & 0 \\
\frac{b_2}{a_{21}} & a_{21} & 0 & 0 \\
\frac{1}{c_1} & \frac{1}{c_2} & 0 & 0 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
k_1 &= f \left( y_n \right) \\
k_2 &= f \left( y_n + ah k_1 \right) \\
y_{n+1} &= y_n + h \left( c_1 k_1 + c_2 k_2 \right)
\end{array}
\]

\[
\begin{pmatrix}
0 & 0 & 0 \\
\frac{1}{2} & 0 & 0 \\
1 - \alpha & \alpha & 0 \\
\end{pmatrix}
\]

18
Taylor expansion:

\[
k_2 = f(y_n) + ahk_1f'(y_n) + \frac{(ahk_1)^2}{2}f''(y_n) + \cdots
\]

\[
y_{n+1} = y_n + h_{c1}f(y_n) + h_{c2}f(y_n) + h_{c2}ahk_1f'(y_n) + O(h^3)
\]

\[
y_{n+1} = y_n + h(c_1 + c_2)f(y_n) + h^2c_2ahk_1f'(y_n) + O(h^3)
\]

Taylor expansion solution:

\[
y_{n+1} = y_n + h\frac{\partial f}{\partial y} + \frac{h^2}{2}\frac{\partial^2 f}{\partial y^2} + \frac{h^3}{6}\frac{\partial^3 f}{\partial y^3} + O(h^4)
\]

\[
y'' = \frac{\partial f}{\partial y} = \frac{\partial f}{\partial t} = f' = f
\]

\[
h(c_1 + c_2) = h \quad \left\{ \begin{array}{l} c_1 + c_2 = 1 \\
 c_2a = \frac{1}{2}
\end{array} \right.
\]

2 equations but 3 constants \((c_1, c_2, a)\) so 1 equation missing. Is coming from:

\[
\begin{array}{cccc}
0 & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 \\
1 - \alpha & \alpha & 0 & 0
\end{array}
\]

Now let's try and connect this to ODE23 (print font 10 batch)

Explicit 3-stage methods

\[
\begin{array}{cccc}
b_1 + b_2 + b_3 = 1 & 0 & 0 & 0 \\
b_2c_2 + b_3c_3 = \frac{1}{2} & x & x & 0 \\
b_2c_2^2 + b_3c_3^2 = \frac{1}{4} & x & x & 0 \\
b_3a_32a_{21} = \frac{1}{6} & x & x & x
\end{array}
\]

6 coefficients and 4 equations \(\rightarrow 2\) free parameters a \(3^{rd}\) order method. Max # coefficients = 0.

\[
\begin{array}{cccc}
0 & 0 & 0 & 0 \\
\frac{1}{4} & \frac{1}{3} & 0 & 0 \\
\frac{1}{4} & 0 & \frac{1}{2} & 0 \\
\frac{1}{4} & 0 & 0 & \frac{1}{2}
\end{array}
\]

RK3

\[
\begin{array}{cccc}
0 & 0 & 0 & 0 \\
\frac{1}{3} & \frac{1}{3} & 0 & 0 \\
\frac{1}{3} & 0 & \frac{1}{3} & 0 \\
\frac{1}{3} & 0 & 0 & \frac{1}{3}
\end{array}
\]

"ODE23"

But when you read closely you see the following Butcher arrays.

\[
A = \begin{pmatrix}
\frac{1}{4} & 0 \\
\frac{1}{4} & \frac{1}{3} \\
\frac{1}{4} & 0 \\
\frac{1}{4} & \frac{1}{3}
\end{pmatrix} \quad \left(\begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & \frac{1}{7} & 0 & 0 \\
0 & \frac{2}{7} & \frac{3}{8} & 0 \\
0 & \frac{3}{7} & \frac{3}{8} & 0
\end{pmatrix}\right) = B^T \\
\begin{pmatrix}
\frac{1}{4} & \frac{1}{8} \\
\frac{1}{4} & \frac{3}{8} \\
\frac{1}{4} & \frac{5}{8} \\
\frac{1}{4} & \frac{7}{8}
\end{pmatrix} = E^T
\]

\[
hB = h \ast B
\]

\[
y_{n+1} = y_n + f \ast hB (\cdot, 3)
\]

\[f (\cdot, y) = f (y_{n+1}) \text{ this } f_y \text{ is used for local error estimate.}\]
function varargout = ode(ode,tspan,y0,options,varargin)
% ODE23 Solve non-stiff differential equations, low order method.
% [T,Y] = ODE23(ODEFUN,TSPAN,Y0) with TSPAN = [T0 TFINAL] inte-
grates the
% ... cut ...
% Initialize method parameters.
pow = 1/3;
A = [1/2; 3/4; 1];
B = [
  1/2 0 2/9
  0 3/4 1/3
  0 0 4/9
  0 0 0];
E = [-5/72; 1/12; 1/9; -1/8];
f = zeros(neq,4);
hmin = 16*eps*abs(t);
% ... cut ...
% LOOP FOR ADVANCING ONE STEP.
nofailed = true; % no failed attempts
while true
  hB = h * B;
  f(:,2) = feval(ode,t+h*A(1),y+f*hB(:,1),args:);
  f(:,3) = feval(ode,t+h*A(2),y+f*hB(:,2),args:);
  tnew = t + h*A(3);
  if done
    tnew = tfinal; % Hit end point exactly.
  end
  h = tnew - t; % Purify h.
  ynew = y + f*hB(:,3);
  f(:,4) = feval(ode,tnew,ynew,args:);
  stats.nfevals = stats.nfevals + 3;
  % Estimate the error.
  if normcontrol
    normynew = norm(ynew);
    err = absh * (norm(f * E) / max(max(normy,normynew),threshold));
  else
    err = absh * (norm(f * E) ./ max(max(abs(y),abs(ynew)),threshold),inf);
  end
  % Accept the solution only if the weighted error is no more than the toler-
  ance rtol. Estimate an h that will yield an error of rtol on the next step or
  the next try at taking this step, as the case may be, and use 0.8 of this value to
  avoid failures.
  if err > rtol % Failed step
    ... cut ...
  function yinterp = ntrp23(tinterp,t,y,tnew,ynew,h,f)
  %NTRP23 Interpolation helper function for ODE23.
\[ BI = \begin{bmatrix} 1 & -4/3 & 5/9 \\ 0 & 1 & -2/3 \\ 0 & 4/3 & -8/9 \\ 0 & -1 & 1 \end{bmatrix}; \]

\[ s = ((tinterp - t) / h)'; \] % may be a row vector

\[ yinterp = y(:,ones(length(tinterp),1)) + f^*(h*BI)*cumprod(s(ones(3,1),:)); \]

RK3 (“ODE23”)

\[
\begin{align*}
  k_1 &= f(t_n, y_n) \\
  k_2 &= f(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_1) \\
  k_3 &= f(t_n + \frac{3h}{4}, y_n + \frac{3h}{4}k_2) \\
  y_{n+1} &= y_n + \frac{h}{9} (2k_1 + k_2 + 4k_3)
\end{align*}
\]

\[ \epsilon = abs(h \cdot norm(f \cdot E)) \] abs: back iterative norm: multi distinctive

Local truncation error estimate:

\[ \epsilon = h \left( -\frac{5}{72}k_1 + \frac{1}{12}k_2 + \frac{1}{9}k_3 - \frac{1}{8}k_4 \right) \]

If the step is accepted then: ...

\[ f(:,1) = f(:,4) \] FSAL (F irst S ame A s L ast)

So apparently 4 stage but actually 3-stage method.

Finally, for dense output or event interpolation we want to interpolate accurately between \( y_n \) and \( y_{n+1} \) without extra function evaluations to \( f \).

\[
\begin{align*}
  y_{n+\alpha} &= y_n + h(k_1 \left( \alpha - \frac{4}{9}\alpha^2 + \frac{5}{9}\alpha^3 \right) + \\
  &\quad h(k_2 \left( \alpha^2 - \frac{2}{3}\alpha^3 \right) + \\
  &\quad h(k_3 \left( \frac{4}{9}\alpha^2 - \frac{8}{9}\alpha^3 \right) + \\
  &\quad h(k_4 \left( -\alpha^2 + \alpha^3 \right) \right) \right) \] (+O (h^4)??)
\]

So we have two extra things here:

- Local truncation error estimate and step size adaption!
- Accurate interpolation results.

Now look for yourself at ode45

% Initialize method parameters.
\[
\begin{align*}
  pow &= 1/5; \\
  A &= [1/5; 3/10; 4/5; 8/9; 1; 1]; \\
  B &= [1/5; 3/10; 4/5; 8/9; 1; 1]; \\
\end{align*}
\]
1/5 3/40 44/45 19372/6561 9017/3168 35/384
0 9/40 −56/15 −25360/2187 −355/33 0
0 0 32/9 64448/6561 46732/5247 500/1113
0 0 0 −212/729 49/176 125/192
0 0 0 0 −5103/18656 −2187/6784
0 0 0 0 0 11/84
0 0 0 0 0 0;
E = [71/57600; 0; −71/16695; 71/1920; −17253/339200; 22/52 5; −1/40];
f = zeros(neq,7);
hmin = 16*eps*abs(t);
% ... cut ...
% LOOP FOR ADVANCING ONE STEP.
nofailed = true; % no failed attemps
while true
hA = h * A;
hB = h * B;
f(:,2) = feval(ode,t+hA(1),y+f*hB(:,1),args:);
f(:,3) = feval(ode,t+hA(2),y+f*hB(:,2),args:);
f(:,4) = feval(ode,t+hA(3),y+f*hB(:,3),args:);
f(:,5) = feval(ode,t+hA(4),y+f*hB(:,4),args:);
f(:,6) = feval(ode,t+hA(5),y+f*hB(:,5),args:);
tnew = t + hA(6);
if done
tnew = tfinal; % Hit end point exactly.
end
h = tnew - t; % Purify h.
ynew = y + f*hB(:,6);
f(:,7) = feval(ode,tnew,ynew,args:);
stats.nfevals = stats.nfevals + 6;

% Estimate the error.
if normcontrol
normynew = norm(ynew)
err = absh * (norm(f * E) / max(max(normy,normynew),threshold));
else
err = absh * norm((f * E) ./ max(max(abs(y),abs(ynew)),threshold),inf); end

% Accept the solution only if the weighted error is no more than the
% tolerance rtol. Estimate an h that will yield an error of rtol on
% the next step or the next try at taking this step, as the case may be,
% and use 0.8 of this value to avoid failures.
if err > rtol % Failed step
% ... cut ...
function yinterp = ntrp45(tinterp,t,y,tnew,ynew,h,f)
%NTRP45 Interpolation helper function for ODE45.
BI = []
1  -183/64  37/12  -145/128
0   0        0        0
0  1500/371  -1000/159  -1000/371
0  -125/32  125/12  -375/64
0  9477/3392  -729/106  25515/6784
0  -11/7  11/3  -55/28
0  3/2  -4  5/2;
s = ((tinterp - t) / h)';  % may be a row vector
yinterp = y(:,ones(length(tinterp),1)) + f*(h*BI)*cumprod(s(ones(4,1),1:));
Bibliography


