Numerical Methods for Initial Value Problems of Ordinary Differential Equations

Defn. An Ordinary Differential Equation (ODE) is an equation that involves one or more derivatives of an unknown function which depends on only one independent variable with domain $[a, b]$.

Defn A solution of an ODE is a differentiable function that satisfies the equation $[a, b]$.

Ex. A first-order ODE:

$y' - y = e^t$, $a < t \leq b$. Solution: $y(t) = te^t + c_1 e^t$ for any c_1

Ex. A second-order ODE

$y'' + 9y = 0$, $0 < t \leq \pi$. Solution: $y(t) = c_1 \cos 3t + c_2 \sin 3t$. for any c_1 and c_2.

Note the existence of the arbitrary constants c_1 and c_1, c_2 respectively. To specify these, i.e. to have a unique solution, extra conditions must be imposed. Typically, as many conditions as the order of the equations must be imposed.

Defn An Initial Value problem for an ODE is one where all the conditions are imposed at $t = a$.

Ex.
$$\begin{cases} y'' + 9y = 0, & 0 < t \leq \pi \\ y(0) = 1 \\ y'(0) = 0 \end{cases}$$

Then unique solution is $y(t) = \cos t$.

There is no guarantee that a given ODE will have a solution. Indeed the first-order ODE

$(y')^2 + y^2 = -1$ cannot possibly have a real solution.

On the other hand, we have the following fundamental result.

Theorem Consider the IVP for a first-order ODE

$$\begin{cases} y'(t) = f(t,y), & a < t \le T \\ y(a) = y_0. \end{cases}$$

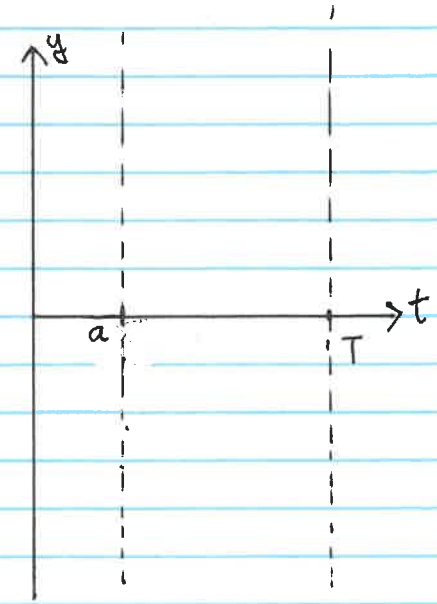
Assume that

(a) $f(t,y)$ is defined and continuous in the strip $a \le t \le T$, $-\infty < y < \infty$ with a and T finite

(b) f is Lipschitz continuous in y i.e. there exists L such that

$$|f(t,y) - f(t,z)| \le L|y-z|$$

$\forall\, t \in [a,T]$ and $\forall\, y, z \in (-\infty, \infty)$.



Then, there exists a unique solution y of the problem, i.e. there exists a function y which is differentiable on $[a,T]$ and satisfies the equation and the initial condition. Moreover there can be only one such solution.

Remarks (i) Condition (a) guarantees the existence of a local solution, i.e. one that exists only on an interval $[a, a+\delta]$ for some positive δ. Also, uniqueness is not guaranteed. Indeed, consider the problem

$$\begin{cases} y' = y^2 \\ y(0) = 0. \end{cases}$$

There is a solution $y(t) = 0$, which is obviously of no interest. On the other hand, there exists a positive solution given by $y(t) = \tan(t)$ (verify!) which becomes singular at $t = \frac{\pi}{2}$.

With the intent here being not to give chase to such theoretical oddities, we shall always assume that there exists a unique solution which is sufficiently regular

(ii) Although many (analytical) methods "of integration" exist for obtaining solutions in "closed form", many problems, in fact most, are resistant to such techniques. Hence, the need to resort to numerical approximation techniques.

(iii) As far as numerical methods are concerned, we shall restrict ourselves to techniques for first-order equations or systems. Indeed, second or higher order ODE's can always be cast as first-order systems.

Ex. $\begin{cases} y'' + 9y = e^t. \quad \text{let } y_1 = y, \; y_2 = y_1' = y' \\ y(0) = 1, \; y'(0) = 0 \end{cases}$

This problem can be expressed as:

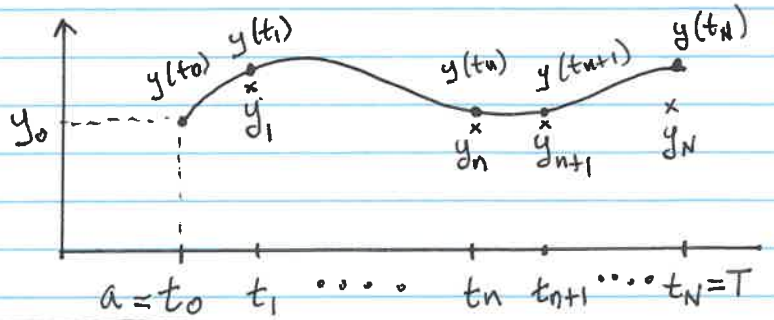
$$\begin{bmatrix} y_1' \\ y_2' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -9 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} 0 \\ e^t \end{bmatrix}.$$

or

$$Y' = AY + b, \qquad A = \begin{bmatrix} 0 & 1 \\ -9 & 0 \end{bmatrix}, \; b = \begin{bmatrix} 0 \\ e^t \end{bmatrix}.$$

The Numerical methods we shall study involve a process of <u>time stepping</u>. It begins with a partition of the interval $[a, T]$: $a = t_0 < t_1 < t_2 < \cdots < t_N = T$

we let $\boxed{h_n = t_{n+1} - t_n}$



The basic idea is this:

(i) Develop a sequence of discrete approximations $y_0, y_1, \ldots, y_N$ with $\boxed{y_n \approx y(t_n), \; n = 0, \ldots, N}$.

Note that we have already used the fact that $y(t_0)$ is known.

(ii) A particular numerical method is one that specifies how to calculate y_{n+1} given that $y_0, \ldots, y_n$ have already been calculated.

Ex. Euler's method

$$\begin{cases} y_{n+1} = y_n + h_n f(t_n, y_n), & n = 0, 1, \ldots, N-1. \\ y_0 \text{ given } (= y(t_0)). \end{cases}$$

we apply it to the problem

$$\begin{cases} y' = -y + t + \frac{1}{2}, & 0 < t < 1 \\ y(0) = 1, \end{cases}$$

with solution $y(t) = t + \frac{3}{2} e^{-t} - \frac{1}{2}$.
we take $N = 10$ and use a uniform partition $\Rightarrow h = .1$.
This results in the procedure

$$\begin{cases} y_{n+1} = y_n + h\left[-y_n + t_n + \frac{1}{2}\right], & n = 0, 1, \ldots, 9, \\ y_0 = 1 \end{cases}$$

or

$$y_{n+1} = y_n + \frac{1}{100}\left[-y_n + \frac{n}{10} + \frac{1}{2}\right].$$

t_n	y_n	$y(t_n)$
0.0	1	1
0.1	0.95	0.95725613
0.2	0.915	0.92809613
0.3	0.8935	0.91122733
0.4	0.88415	0.90548007
0.5	0.885735	0.90979599
0.6	0.8971615	0.92321745
0.7	0.91744535	0.94487796
0.8	0.94570081 5	0.97399346
0.9	0.98113073 35	1.00985449
1.0	1.023017660 15	1.05181916

we measure the error as follows

$$E = E(h) = \max_{0 \le n \le N} |y_n - y(t_n)|$$
$$= .0288$$

in this case.

Some other well-known methods are:

Backward or Implicit Euler

$$y_{n+1} = y_n + h f(t_{n+1}, y_{n+1}).$$

Trapezoidal rule

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

Midpoint rule (Implicit)

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

Midpoint rule (Explicit)

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

The manner in which a numerical method is "built" can be as instructive as the study of its numerical properties. So let us show a way to construct several of these methods.

From the fundamental Theorem of Calculus, the solution of the IVP satisfies

$$y(t_{n+1}) - y(t_n) = \int_{t_n}^{t_{n+1}} f(s, y(s))\, ds.$$

It is now a matter of replacing the integral by one of many approximation seen earlier.

1) Using $\int_a^b f(s)ds \approx (b-a)f(a)$, we obtain

$y(t_{n+1}) - y(t_n) \approx h_n f(t_n, y(t_n))$. This motivates defining the numerical approximations by

$$y_{n+1} = y_n + h_n f(t_n, y_n)$$

which is precisely Euler's method.

2) Using $\int_a^b f(s)\,ds \approx (b-a) f(b)$ yields the Backward Euler method

3) Using $\int_a^b f(s)\,ds \approx \frac{b-a}{2}[f(a)+f(b)]$ results in the Trapezoidal rule for IVP's.

4) Using $\int_a^b f(s)\,ds \approx (b-a) f\left(\frac{a+b}{2}\right)$, we obtain

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

Further using the approximation $y\left(\frac{t_n+t_{n+1}}{2}\right) \approx \frac{y(t_n)+y(t_{n+1})}{2}$ can be seen as the motivation behind the implicit midpoint method.

whereas, using $y\left(\frac{t_n+t_{n+1}}{2}\right) = y\left(t_n + \frac{t_{n+1}-t_n}{2}\right)$

$$\approx y(t_n) + h_n y'(t_n) = y(t_n) + h_n f(t_n, y(t_n))$$

yields

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

which motivates the explicit version of the midpoint method.

Classification of Numerical Methods for ODE's

There are at this point a vast array of numerical methods for IVP's of ODE's differing in terms of their computational (implementational) as well as analytical properties.

To begin, we shall divide them in 4 categories. In order to do so, it will help to assume that $y_0, \ldots, y_n$ have been computed and the method is characterized by how it defines y_{n+1}.

Single-step vs. Multistep: A method is called single-step (or one-step) if y_{n+1} does not depend explicitly on any subset of $\{y_0, \ldots, y_{n-1}\}$. We define a method as multistep if it is not single-step.

Explicit vs. Implicit A method is called implicit if there is no explicit dependence on y_{n+1} through f. Otherwise it is classified as Explicit.

Ex. Euler's method is single-step and explicit.
The Backward differentiation method

$$y_{n+1} - \frac{4}{3} y_n + \frac{1}{3} y_{n-1} = \frac{2}{3} f(t_{n+1}, y_{n+1})$$

is implicit and 2-step, multistep.

The distinctions shown above refer to type rather than property, although sometimes the two can be related, e.g. the true statement "No Explicit method can be A-Stable".

General considerations in the numerical approximation of IVP's.

As in the case of any area of numerical analysis, the numerical approximation of solution of initial value problems of ordinary differential equations entails a study of fundamental issues such as the convergence and the accuracy of the approximations as well as the efficiency of the scheme involved.

A time-stepping scheme, involves the interplay between three things:

(i) A numerical method

(ii) A given initial value problem $y' = f(t,y), \quad a < t \leq T, \quad y(a) = y_0,$

(iii) A partition $\mathrm{P} : \mathrm{a} = \mathrm{t}_0 < \mathrm{t}_1 < \ldots, \mathrm{t}_N = \mathrm{T},$.

resulting in a sequence $\{y_n^{(\mathrm{P})}\}_{n=0}^N$ of approximations $y_n^{(\mathrm{P})} \approx y(t_n),\ n = 0, \ldots, N.$

We shall characterize the *fineness* of the partition P by

$$h = \max_{0 \leq n \leq N} h_n, \quad h_n = t_{n+1} - t_n,\ n = 0, \ldots, N-1$$

and to simplify the notation we shall henceforth omit the superscript P.

One way of studying the behavior of a particular numerical method is to fix the first two items and then to study the behavior of the set of sequences $\{y_n\}_{n=0}^N$ generated over a sequence of finer and finer partitions. Later, we will be able to make more general statements concerning the behavior of a given numerical method over a class of IVP's.

Definition 1 *A numerical scheme as described above is convergent if*

(i) There exists $h_0 > 0$ such that the sequences generated over all partitions with fineness $h \leq h_0$ are well-defined.

(ii) For all partitions with fineness $h \leq h_0$.

$$\lim_{h \to 0} \max_{n \leq 0 \leq N} |y_n - y(t_n)| = 0.$$

At this point, we can also define the *order* or *rate* of convergence of the scheme.

Definition 2 *We say that a scheme is convergent at the rate m for some integer $m \geq 1$ if there exists a constant c such that for $h \leq h_0$*

$$\max_{n \leq 0 \leq N} |y_n - y(t_n)| \leq ch^m.$$

The restriction of the rate to integer values is typical of this area. Moreover, if a scheme is convergent with order $m > 0$, then it is convergent according to the previous and more general sense.

It turns out that the rate of convergence of a scheme is related to an important property of the numerical method which we now introduce. At this stage, it is proper to restrict the definition to the case of a uniform partition, i.e. one for which $h = h_n$.

Definition 3 *Let y^*_{n+1} be the approximation obtained when the scheme is applied to the exact solution, i.e. using $y(t_n)$ instead of y_n in the case of a single-step method and $y(t_n), \ldots, y(t_{n-k+1})$ in the case of a multistep method. Then the Local Truncation Error (LTE) of the scheme is the quantity*

$$\frac{1}{h}|y(t_{n+1}) - y^*_{n+1}|.$$

Remark 1 *Essentially, the information provided by the* LTE is the amount of error introduced in just one step. Moreover, the reason for having the factor $\frac{1}{h}$ is the fact that when a scheme is applied over a total of $N \approx O(\frac{1}{h})$ steps, the LTE, as defined, turns out to be a more accurate predictor of the convergence rate.

Ex. Calculate the LTE of the Euler method.

We have upon applying Taylor's Theorem and using the equation $y'(t) = f(t, y(t))$:

$$y^*_{n+1} = y(t_n)+hf(t_n, y(t_n)) = y(t_n)+hy'(t_n) = y(t_{n+1})-\frac{1}{2}f''(\zeta_n)h^2,\ \zeta_n \in (t_n, t_{n+1}).$$

From this it follows that $\frac{1}{h}|y^*_{n+1} - y(t_{n+1})| = \frac{1}{2}h|f''(\zeta_n)|$ which is the LTE of Euler's method. Since the LTE is $O(h)$, we say that Euler's method is a first-order accurate method in the sense that we expect that under certain conditions the order or rate of convergence will be one.

Ex. calculate the LTE of the two-step method $y_{n+1} - \frac{4}{3}y_n + \frac{1}{3}y_{n-1} = \frac{2}{3}hf(t_{n+1}, y_{n+}$

Again, an application of Taylor's Theorem shows that for $\zeta_n^{(1)} \in (t_{n-1}, t_n), \zeta_n^{(2)}, \zeta_n^{(3)}$ (t_n, t_{n+1}),

$$
\begin{aligned}
y^*_{n+1} &= \frac{4}{3}y(t_n) - \frac{1}{3}y(t_{n-1}) + \frac{2}{3}hf(t_{n+1}, y(t_{n+1}) \\
&= \frac{4}{3}y(t_n) - \frac{1}{3}y(t_{n-1}) + \frac{2}{3}hy'(t_{n+1}) \\
&= \frac{4}{3}y(t_n) - \frac{1}{3}\left[y(t_n) - hy'(t_n) + \frac{1}{2}h^2y''(t_n) + \frac{1}{6}h^3y'''(\zeta_n^{(1)})\right] \\
&\quad + \frac{2}{3}h\left[y'(t_n) + hy''(t_n) + \frac{1}{2}h^2y'''(\zeta_n^{(2)})\right], \\
&= y(t_n) + hy'(t_n) + \frac{1}{2}h^2y''(t_n) - \frac{1}{18}h^3y'''(\zeta_n^{(1)}) + \frac{1}{3}h^3y'''(\zeta_n^{(2)}) \\
&= y(t_{n+1}) - \frac{1}{6}h^3y'''(\zeta_n^{(3)}) - \frac{1}{18}h^3y'''(\zeta_n^{(1)}) + \frac{1}{3}h^3y'''(\zeta_n^{(2)}).
\end{aligned}
$$

Hence, the LTE of this method is

$$
\frac{1}{h}|y^*_{n+1} - y(t_{n+1})| = h^2\left|\frac{1}{6}h^3y'''(\zeta_n^{(3)}) - \frac{1}{18}h^3y'''(\zeta_n^{(1)}) + \frac{1}{3}h^3y'''(\zeta_n^{(2)})\right| \le \frac{5}{9}h^2 \max_{t_{n-1}\le t\le t_{n+1}} |y'''(t)|
$$

which shows that the method is second-order accurate.

Stability

The preceding discussion centered on what can be described as the accuracy of the numerical approximations to solutions of IVP's. There is yet another concept that is crucial to the study of the behavior of the approximations, namely that of *stability*. We all have intuitive notions what a stable *system* is. For instance, a pencil lying on its side is considered as stable in that small perturbations do not cause it to "fly apart" whereas the same pencil standing on its tip, yes this can be done, is considered as unstable. While completely adhoc, such intuitive ideas are the starting points of some very powerful and rigorously developed stability concepts and results.

The very nature of the time-stepping schemes generates a compounding effect of the errors with possibly disastrous results. As an illustration of this compounding effect, let us consider the application of Euler's method to the IVP $y' = -100y$, $y(0) = 1$ with solution $y(t) = e^{-100t}$. We have $y_{n+1} = y_n + h(-100)y_n = (1-100h)y_n$ with general solution $y_n = (1-100h)^n$. If we use the stepsize $h = .1$, then $y_n = (-9)^n$. Besides the absence of any

pretense of closeness between y_n and $y(t_n) = e^{-100hn} = e^{-10n}$, there is a vast discrepancy between their qualitative behaviors: One is positive and decaying while the other is growing and with alternating signs. On the other hand if we reduce the step size to .001, then we obtain $y_n = (.9)^n$ and $y(t_n) = e^{-.1n}$, an obvious improvement. We would be justified in calling these two occurrences as manifestations of instability and stability, respectively.

With the aim of developing an understanding of the phenomena underlying these and other occurrences, we introduce an approach to the issue of stability that will eventually lead to a body of knowledge known as *linear stability theory.* It starts by asking the following question: How does a given numerical method behave when applied to the very simple but nontrivial initial value problem

$$y' = \lambda y,\ y(0) = 1,$$

where λ is a complex parameter. The solution is given by $y(t) = e^{\lambda t}$. The thinking here is that if the scheme does not behave "satisfactorily" for this simple problem, then could it be expected to handle more general problems? More specifically, we would like to see the degree to which the behavior of the numerical approximations matches that of the exact solution and in particular its decay

$$\lim_{t\to\infty} e^{\lambda t} = 0 \iff \mathrm{Re}\lambda < 0.$$

Now $y(t_n) = e^{\lambda hn} \to 0$ as $n \to \infty \iff \mathrm{Re}\lambda < 0$. We argue that if y_n is to be an approximation of $y(t_n)$, then we should have $\lim_{n\to\infty} y_n = 0$. These considerations are the motivation behind some important concepts of *linear stability theory.* First and in order to put things on a rigorous footing, let us agree that we have a numerical method which is applied to the model problem using the uniform partition $t_n = nh,\ n = 0, \ldots$ generating a sequence $\{y_n\},\ n = 0, \ldots$. Furthermore, it will turn out that it is preferable to deal with the complex number $z = \lambda h$ rather than with λ and h separately.

Definition 4 *We say that the numerical scheme is absolutely stable for $z = \lambda h$ if the resulting sequence satifies $\lim_{n\to\infty} y_n = 0$.*

Definition 5 *The region of absolute stability of a numerical method is the collection of complex numbers z for which the numerical scheme is absolutely stable.*

Definition 6 *A numerical method is A_0-stable if it region of absolute stability contains the half-line* $(-\infty, 0)$.

Definition 7 *A numerical method is said to be A-stable if its region of absolute stability contains the negative half-plane* $\{\zeta \in \mathbb{C},\ \mathrm{Re}\zeta < 0\}$.

As an example, let us apply Euler's method to the model problem above. We have

$$y_{n+1} = y_n + h\lambda y_n = (1+\lambda h)y_n \iff y_n = (1+\lambda h)^n y_0 = (1+\lambda h)^n,\ n = 0, \ldots, N.$$

With $z = \lambda h$, we see that $\lim_{n\to\infty} y_n = 0 \iff |1+z| < 1$. In other words the region of absolute stability of Euler's method is the open disk $\mathcal{D}$ in the complex plane with center at $(-1, 0)$ and radius 1. In the light of this we can offer an explanation for the behavior seen earlier: With $\lambda = -100$ and $h = .1$, $z = -10$ a value that did not belong to $\mathcal{D}$ whereas with $h = .001$ $z = -.1$ does. In general, for a given λ, the values of h for which the behaviour of the sequence $\{y_n\}$ generated by Euler's method will match that of the solution are precisely those for which λh belong to $\mathcal{D}$.

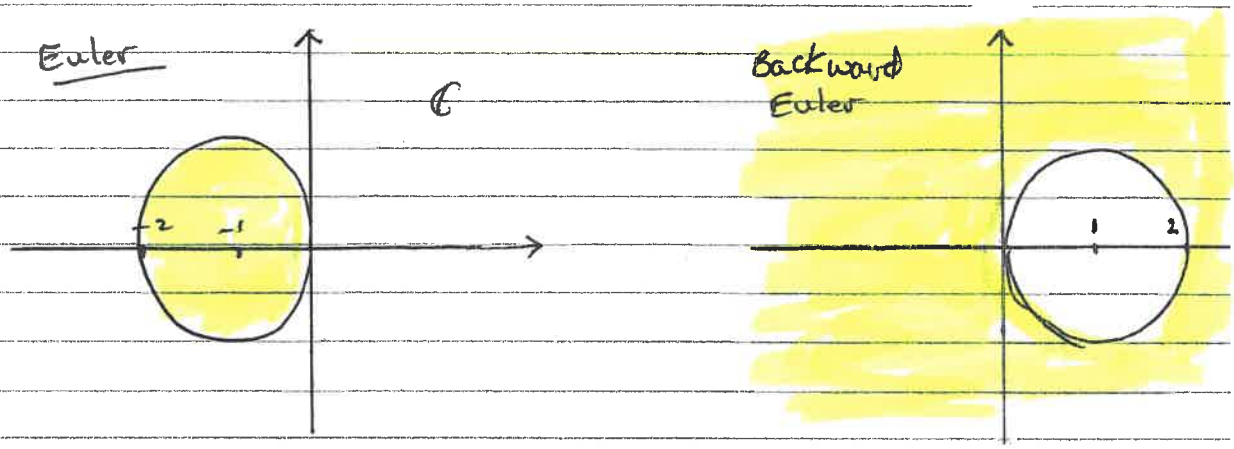
Remark 2 *The region of absolute stability of Euler's method is bounded. Indeed, this is the case of all explicit methods. For large values of λ, it is necessary to restrict the size of h in order to have $z \in \mathcal{D}$. This is an example of conditional stability.*

Let us now consider the backward Euler method. In this case, we have

$$y_{n+1} = y_n + h\lambda y_{n+1} \implies y_{n+1} = \frac{1}{1-z} y_n \implies y_n = \frac{1}{(1-z)^n},\ n = 0, 1, \ldots.$$

Points z for which the method is absolutely stable are those for which $|1-z| > 1$. With $z = x + \mathrm{i}y$ simple calculations show that this is the case iff $(1-x)^2 + y^2 > 1$. Thus the region of absolute stability of the Backward Euler method is the complement in the complex plane of the closed disk centered at $(1, 0)$ and radius 1.

Remark 3 *Since the region includes the open left-half plane, the Backward Euler method is A-stable. Furthermore,* $\lim_{n\to\infty} y_n = 0$ *as long as* $\mathrm{Re}\, z < 0$, *irrespective of the size of h. This is an instance of unconditional stability.*



Regions of absolute stability

Methods based on Taylor's Theorem

These are also known as Taylor series methods and form a family of explicit, single-step methods.

Let $y(t)$ be the solution of the IVP $y'(t) = f(t, y(t))$ assumed to be sufficiently regular. We have

$$(1) \quad y(t_{n+1}) = y(t_n) + h y'(t_n) + \frac{h^2}{2} y''(t_n) + \cdots + \frac{h^m}{m!} y^{(m)}(t_n) + O(h^{m+1}).$$

The idea here is to express time derivatives of y in terms of f and its partial derivatives using the equation itself which we write as $y' = f$ for simplicity

$$y' = f$$

$$y'' = f_t + f_y y' = f_t + f_y f$$

$$y''' = f_{tt} + f_{ty} y' + (f_{yt} + f_{yy} y') f + f_y (f_t + f_y y')$$

$$= f_{tt} + 2 f_{ty} f + f_{yy} f^2 + f_y f_t + f_y^2 f,$$

and so on ....

A Taylor series method is obtained by choosing a value for m and replacing time derivatives of y in (1) by the corresponding expressions in terms of f and its derivatives.

Ex. $m = 1$.

$$y(t_{n+1}) \doteq y(t_n) + h f(t_n, y(t_n)) + O(h^2)$$

which forms the motivation behind the method

$$y_{n+1} = y_n + h f(t_n, y_n)$$

which we recognize as Euler's method.

$\underline{m=2}$ Arguing as above, we arrive at the method

$$y_{n+1} = y_n + hf + \frac{h^2}{2}[f_t + f_y f]$$

with the understanding that f, f_t, f_y are evaluated at (t_n, y_n).

$\underline{m=3}$

$$y_{n+1} = y_n + hf + \frac{h^2}{2}[f_t + f_y f] + \frac{h^3}{6}[f_{tt} + 2f_{ty}f + f_{yy}f^2 + f_y f_t + f_y^2 f].$$

It is clear from the construction process that the method of order m has order of accuracy m, i.e.

$$LTE(m) = O(h^m).$$

Remarks (i) For higher values of m the expressions become quite unwieldy, which limits the efficiency and thus the applicability of Taylor series methods. Indeed more efficient alternatives, e.g. explicit Runge-Kutta methods, are available.

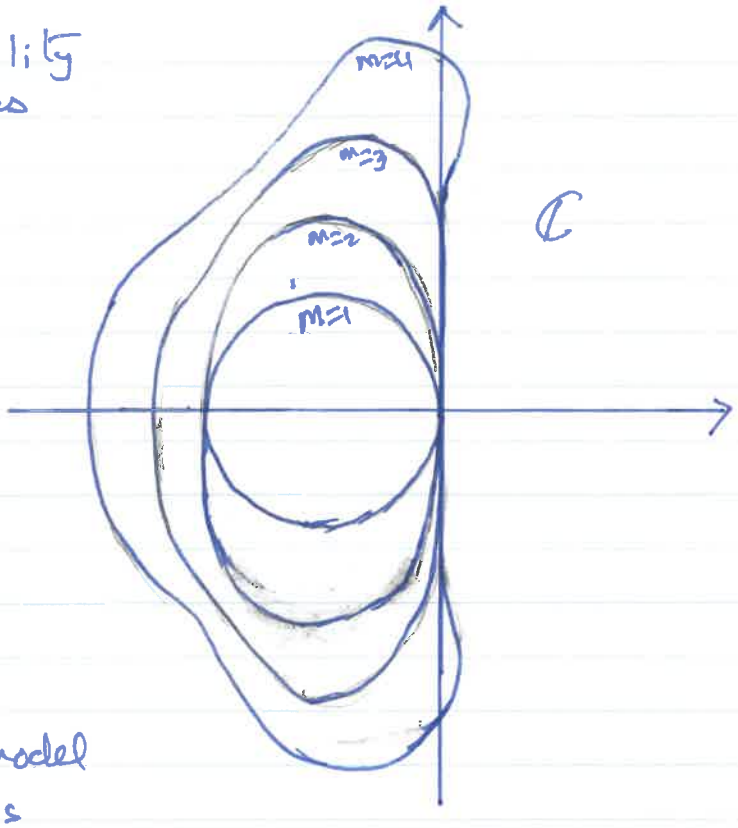
(ii) It is easy to see that the Taylors series method of order m when applied to the model equation $y' = \lambda y$ yields

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

where $T_m(z) = 1 + z + \cdots + \frac{z^m}{m!}$ is the Taylor polynomial of degree m of e^z. This opens the way to the study of the linear stability properties of these methods. Also, since $y_n = (T_m(z))^n y_0$, $n = 0, 1, \ldots$,

the region of absolute stability is given by $\{z \in \mathbb{C}, |T_m(z)| < 1\}$.

Regions of absolute stability of the first 4 Taylor series methods.



It can be shown that any explicit single-step method whose local truncation error is $O(h^m)$, can be expressed as

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

when applied to the model problem $y' = \lambda y$. Thus all single-step, explicit methods of order of accuracy m share the same linear stability properties, including the regions of absolute stability.

(RK)

Runge-Kutta methods offer a wide gamut of single-step methods and are quite popular in view of their properties. They can be presented in compact form as a tableau of coefficients.

What we have here is an s-stage RK method. Given y_n, y_{n+1} is defined by

$$\begin{array}{c|c} A & \tau \\ \hline w & \end{array} = \begin{array}{c|c} \begin{matrix} a_{11} & \cdots & a_{1s} \\ \vdots & & \vdots \\ a_{s1} & \cdots & a_{ss} \end{matrix} & \begin{matrix} \tau_1 \\ \vdots \\ \tau_s \end{matrix} \\ \hline \begin{matrix} w_1 & \cdots & w_s \end{matrix} & \end{array}$$

$$y_{n+1} = y_n + h \sum_{i=1}^{s} w_i K_i$$

where the K_i's are given by

$$K_i = f\left(t_n + h\,\tau_i,\ y_n + h \sum_{j=1}^{s} a_{ij} K_j\right), \quad i = 1, \ldots, s.$$

RK methods can be of explicit or implicit type, with an RK method being explicit if and only if there is a permutation of the matrix A in its tableau which is strictly lower triangular.

Ex. The tableau

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \hline & 0 & 1 \end{array}$$

leads to $K_1 = f(t_n, y_n)$

$$K_2 = f\left(t_n + \frac{h}{2},\ y_n + \frac{h}{2} K_1\right)$$

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

$$y_{n+1} = y_n + h K_2 = y_n + h\, f\left(\frac{t_n + t_{n+1}}{2},\ y_n + \frac{h}{2} f(t_n, y_n)\right).$$

we recognize this as the Explicit mid point method.

Ex. The tableau

0	0
1	

can be easily seen to correspond to Euler's method.

Ex.

0	0	0
1	0	1
$\frac{1}{2}$	$\frac{1}{2}$	

translates into the method

$$y_{n+1} = y_n + \frac{h}{2} f(t_n, y_n) + \frac{h}{2} f(t_{n+1}, y_n + h f(t_n, y_n))$$

known as the <u>modified Euler method.</u>

Ex. The Tableau

0	0	0
$\frac{2}{3}$	0	$\frac{2}{3}$
$\frac{1}{4}$	$\frac{3}{4}$	

yields the method

$$y_{n+1} = y_n + \frac{h}{4} f(t_n, y_n) + \frac{3h}{4} f(t_n + \frac{2}{3}h, y_n + \frac{2}{3} h f(t_n, y_n))$$

known as <u>Heun's method.</u>

A widely used 4-stage explicit RK method is given by the tableau

0	0	0	0	0
$\frac{1}{2}$	0	0	0	$\frac{1}{2}$
0	$\frac{1}{2}$	0	0	$\frac{1}{2}$
0	0	1	0	1
$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	

$$K_1 = f(t_n, y_n)$$

$$K_2 = f(t_n + \frac{h}{2}, y_n + \frac{h}{2} K_1)$$

$$K_3 = f(t_n + \frac{h}{2}, y_n + \frac{h}{2} K_2)$$

$$K_4 = f(t_n + h, y_n + h K_3)$$

$$y_{n+1} = y_n + \frac{h}{6}[K_1 + 2K_2 + 2K_3 + K_4].$$

Some Implicit Runge-Kutta (IRK) methods

Ex.

1	1
	1

$$K_1 = f(t_n + h,\ y_n + h K_1)$$

$$y_{n+1} = y_n + h K_1 = y_n + h f(t_{n+1},\ \underbrace{y_n + h K_1}_{y_{n+1}})$$

$$= y_n + h f(t_{n+1},\ y_{n+1})$$

which is Backward Euler method seen earlier.

Ex.

$\frac{1}{2}$	$\frac{1}{2}$
1	

results in the implicit Midpoint method.

Ex. The tableau

0	0	0
$\frac{1}{2}$	$\frac{1}{2}$	1
$\frac{1}{2}$	$\frac{1}{2}$	

can be seen to lead to the trapezoidal method.

Ex. A two-stage IRK method is given by the tableau. Together with the implicit midpoint method, they are the first two members of IRK methods of Gauss-Legendre type.

$\frac{1}{4}$	$\frac{1}{4} - \frac{1}{2\sqrt{3}}$	$\frac{1}{2} - \frac{1}{2\sqrt{3}}$
$\frac{1}{4} + \frac{1}{2\sqrt{3}}$	$\frac{1}{4}$	$\frac{1}{2} + \frac{1}{2\sqrt{3}}$
$\frac{1}{2}$	$\frac{1}{2}$	

Indeed, for arbitrary $s \geq 1$, $\tau_1, \ldots, \tau_s$ are the roots of the Legendre polynomial of degree s but mapped to the interval $[0,1]$. The weights $w_1, \ldots, w_s$ are precisely the weights of the associated Gauss-Legendre quadrature rule.

There are also IRK methods corresponding to the Radau and Lobatto quadrature rules. (cf. J. Butcher, Dekker & Verwer)

Explicit RK methods are relatively inexpensive to use in the sense that each k_i can be evaluated explicitly in terms of known quantities. The implicit RK methods are more expensive in terms of work since the k_i's are now implicitely defined through f. Moreover, different k_i's may be coupled requiring the solution of larger systems.

In this latter context, it is worth mentioning that there are so-called Diagonally Implicit Runge-Kutta (DIRK) methods which are implicit, yet the $s \times s$ matrix is lower triangular with identical diagonal elements.

β	0	β
$1-2\beta$	β	$1-\beta$
$\frac{1}{2}$	$\frac{1}{2}$	

2-stage, Third-order DIRK

$$\beta = \frac{1}{2} + \frac{1}{2\sqrt{3}}$$

γ	0	0	γ
$\frac{1}{2}-\gamma$	γ	0	$\frac{1}{2}$
2γ	$1-4\gamma$	γ	$1-\gamma$
$\frac{1}{24(\frac{1}{2}-\gamma)^2}$	$1-\frac{1}{12(\frac{1}{2}-\gamma)^2}$	$\frac{1}{24(\frac{1}{2}-\gamma)^2}$	

3-stage, fourth-order DIRK

$$\gamma = \frac{1}{2} + \frac{1}{\sqrt{3}} \cos\frac{\pi}{18} \simeq 1.068579021\,3$$

The advantage of DIRK methods over other IRK methods is that the different k_i's are now decoupled. Moreover, instead of the system of ODE's $y' =$, the k_i's are the given as the solution of a linear system of the form

$$(I - h\beta M)\, k_i = r_i, \qquad i = 1, \ldots, s,$$

i.e. one has to deal with the same matrix for each k_i, and if the stepsize is not changed, over the entire span of integration.

Order of accuracy of Runge-Kutta methods

Let us recall that the order of accuracy of a time-stepping scheme is the index m appearing in the Local Truncation Error of the scheme, i.e. $LTE = O(h^m)$, when the scheme is applied to the ODE $y' = f(t, y(t))$. Here of course we assume that f and the solution y are sufficiently regular.

The determination of the order of accuracy of Runge-Kutta methods turns out to be nontrivial as the next example shows

Ex. Determine the LTE and thus the order of accuracy of the modified Euler method

$$y_{n+1} = y_n + \frac{1}{2} h f(t_n, y_n) + \frac{1}{2} h f(t_n + h, y_n + h f(t_n, y_n)).$$

we have

$$LTE = \frac{1}{h} \Big| y(t_{n+1}) - \Big[y(t_n) + \frac{1}{2} h f(t_n, y(t_n)) + \frac{1}{2} h f(t_n + h, y(t_n) + h f(t_n, y(t_n))) \Big] \Big|.$$

Note that $f(t_n, y(t_n)) = y'(t_n)$. Hence

$$LTE = \frac{1}{h} \Big| y(t_{n+1}) - \Big[y(t_n) + \frac{h}{2} y'(t_n) + \frac{h}{2} f(t_n + h, y(t_n) + h y'(t_n)) \Big] \Big|$$

To handle the last term, we use Taylor's Theorem in two variables

$$f(x + \Delta x, y + \Delta y) = f(x, y) + f_x \Delta x + f_y \Delta y + \frac{1}{2} (\Delta x, \Delta y) \begin{bmatrix} f_{xx} & f_{yx} \\ f_{xy} & f_{yy} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + O(|\Delta x|^3 + |\Delta y|^3).$$

Applying this with $x = t_n$, $\Delta x = h$, $y = y(t_n)$, $\Delta y = h y'(t_n)$,

$$LTE = \frac{1}{h}\left| y(t_{n+1}) - \left[y(t_n) + \frac{h}{2} y'(t_n) + \frac{h}{2}\left\{ \overbrace{f}^{=y'(t_n)} + h\left[f_t + f_y \overbrace{y'}^{=y''(t_n)} \right] + \frac{h^2}{2}\left(f_{tt} + 2 f_{ty} y' + f_{yy} (y')^2 \right) + O(h^3) \right\} \right] \right|$$

with $y' = y'(t_n)$ and f and its partials are evaluated at $(t_n, y(t_n))$. Hence

$$LTE = \frac{1}{h}\left| \overbrace{y(t_{n+1}) - \left[y(t_n) + h y'(t_n) + \frac{h^2}{2} y''(t_n) \right]}^{= y(t_{n+1}) - \frac{y'''(t_n)}{6} h^3} + \frac{h^3}{4}\left(f_{tt} + 2 f_{ty} f + f_{yy} f^2 \right) + O\left(h^3 + h^3 |y'(t_n)|^3\right) \right|$$

$$= h^2 \left| \frac{1}{6} y'''(t_n) - \frac{1}{4}\left(f_{tt} + 2 f_{ty} f + f_{yy} f^2 \right) + O\left(h^2 + h^2 |y'(t_n)|^3\right) \right|$$

which shows that the method is 2nd-order accurate for the general ODE $y'(t) = f(t, y(t))$.

One can easily imagine the difficulties encountered in doing such calculations for general Runge-Kutta methods and especially so for the implicit ones. The method of rooted trees pioneered by J. C. Butcher in the 1960's offers a systematic way of handling the various partial derivatives of f.

On the other hand, one could restrict attention to the model problem $y' = \lambda y$ and thus get a value for m for this restricted class of problems. Of course the value thus obtained may not be the order of accuracy for the general problem, but only an upper bound.

Indeed, the calculations in this case turn out to be quite simple. So consider the s-stage RK method, explicit or otherwise. Applying it to the model problem $y' = \lambda y$ gives

$$y_{n+1} = y_n + h(w_1 k_1 + \cdots + w_s k_s) \quad \text{where } k_1, \ldots, k_s \text{ are given by}$$

$$k_1 = \lambda\left(y_n + h(a_{11} k_1 + \cdots + a_{1s} k_s)\right.$$

$$\vdots$$

$$k_s = \lambda\left(y_n + h(a_{s1} k_1 + \cdots + a_{ss} k_s)\right.$$

Let $\mathbb{1}$ denote the s-vector with components equal to 1. we have

$$\begin{bmatrix} k_1 \\ \vdots \\ k_s \end{bmatrix} = \lambda y_n \mathbb{1} + \lambda h A \begin{bmatrix} k_1 \\ \vdots \\ k_s \end{bmatrix} \Rightarrow \begin{bmatrix} k_1 \\ \vdots \\ k_s \end{bmatrix} = (I - h\lambda A)^{-1}(\lambda y_n \mathbb{1})$$

Thus with $w = (w_1, \ldots, w_s) \in \mathbb{R}^s$,

$$y_{n+1} = y_n + h\lambda y_n w^T (I - h\lambda A)^{-1} \mathbb{1}$$

$$\Rightarrow \boxed{y_{n+1} = \left(1 + z\, w^T (I - zA)^{-1} \mathbb{1}\right) y_n, \quad z = h\lambda}$$

we will show that the function $\phi(z) \equiv 1 + z w^T (I - zA)^{-1} \mathbb{1}$ is a <u>polynomial</u> in z if the RK method is explicit, i.e. if A is strictly lower triangular and is a rational function in the implicit case.

If A is strictly lower triangular, then $I - zA$ is unit lower triangular and so is its inverse $(I - zA)^{-1}$. Furthermore

$$(I - zA)^{-1} = I + zA + (-zA)^2 + \cdots + (-1)^{s-1} (zA)^{s-1}.$$

Indeed the sum is finite since A being $s \times s$ and strictly lower triangular, $A^s = A^{s+1} = \cdots = 0$. Hence $(I - zA)^{-1}$ is a matrix whose elements are polynomials of degree at most $s-1$ in z. It then follows that $\phi(z)$ is a polynomial of degree at most s in z.

Revisiting the modified Euler method with Tableau

0	0	0
1	0	1
$\frac{1}{2}$	$\frac{1}{2}$	

we see that

$$\phi(z) = 1 + z(I - zA)^{-1}\mathbb{1} = 1 + z + \frac{z^2}{2}.$$

we already know that the order of accuracy of this method is $m=2$. This is consistent with the fact that

$$\phi(z) = e^z + O(h^3).$$

In the implicit RK case, $I - zA$ will be invertible for small $|z|$. Moreover, letting $A = S^{-1}JS$ be the Jordan canonical decomposition of A, we have

$$I - zA = I - zS^{-1}JS = S^{-1}(I - zJ)S.$$

For z sufficiently small, $I - zJ$ is invertible. Hence $I - zA$ will be invertible for such z and

$$(I - zA)^{-1} = S^{-1}(I - zJ)^{-1}S.$$

At this point we appeal to the formula

$$(I - zJ)^{-1} = \frac{1}{\det(I - zJ)} \operatorname{adj}(I - zJ)$$

where $\operatorname{adj}(I - zJ)$ is the <u>adjugate</u> matrix formed by <u>cofactors</u>, i.e. $(s-1) \times (s-1)$ determinants of $I - zJ$. Now $I - zJ$ is upper triangular and bi-diagonal. Hence $\det(I - zJ)$ is a polynomial of degree at most s. Furthermore, $\operatorname{adj}(I - zJ)$ is an $s \times s$ matrix whose entries are $(s-1) \times (s-1)$ determinants of $I - zJ$, hence polynomials of degree at most $s-1$ in z. In conclusion, $(I - zJ)^{-1}$, and thus $(I - zA)^{-1}$ are

sxs matrices whose (i,j)-th entry is a rational function $P(z)/Q(z)$ with $\deg P \le s-1$ and $\deg Q \le s$.

Hence

$$\phi(z) = r(z) = 1 + z\, w^T (I - zA)^{-1} \mathbb{1} \equiv \frac{P(z)}{Q(z)}, \quad \deg P \le s, \ \deg Q \le s.$$

As an example, we consider the 2-stage IRK given by

$$\begin{array}{cc|c} \frac{1}{4} & \frac{1}{4} - \frac{1}{2\sqrt{3}} & \frac{1}{2} - \frac{1}{2\sqrt{3}} \\ \frac{1}{4} + \frac{1}{2\sqrt{3}} & \frac{1}{4} & \frac{1}{2} + \frac{1}{2\sqrt{3}} \\ \hline \frac{1}{2} & \frac{1}{2} & \end{array}.$$

we have $$I - zA = \begin{bmatrix} 1 - \frac{z}{4} & -\frac{z}{4}\left(1 - \frac{2}{\sqrt{3}}\right) \\ -\frac{z}{4}\left(1 + \frac{2}{\sqrt{3}}\right) & 1 - \frac{z}{4} \end{bmatrix}$$

$\Rightarrow$

$$(I - zA)^{-1} = \frac{1}{\left(1 - \frac{z}{4}\right)^2 - \frac{z^2}{16}\left(1 - \frac{4}{3}\right)} \begin{bmatrix} 1 - \frac{z}{4} & \frac{z}{4}\left(1 - \frac{2}{\sqrt{3}}\right) \\ \frac{z}{4}\left(1 + \frac{2}{\sqrt{3}}\right) & 1 - z/4 \end{bmatrix}.$$

Hence, after some algebra

$$\boxed{\phi(z) = r(z) = \frac{1 + \frac{z}{2} + \frac{z^2}{12}}{1 - \frac{z}{2} + \frac{z^2}{12}}}$$

This is the so-called $(2,2)$-Padé approximation of e^z. Indeed, it is easily seen that

$$\phi(z) = e^z + O(z^5)$$

Linear Multistep methods

For integer $k \geq 1$, a **k-step Linear Multistep (LM)** is specified by a set of $2k+2$ coefficients $\alpha_0, \alpha_1, \ldots, \alpha_k$ and $\beta_0, \beta_1, \ldots, \beta_k$. When applied to the IVP $y' = f(t,y)$, they yield approximations $\{y_n\}_{n=0}^{N}$ as follows

$$\begin{cases} \alpha_k y_{n+1} + \alpha_{k-1} y_n + \cdots + \alpha_0 y_{n+1-k} = h\left[\beta_k f(t_{n+1}, y_{n+1}) + \beta_{k-1} f(t_n, y_n) + \cdots + \beta_0 f(t_{n+1-k}, y_{n+1-k})\right] \\ \qquad\qquad n = k-1, \ldots, N-1. \\ y_0, y_1, \ldots, y_{k-1} \quad \text{given initial values} \end{cases}$$

Ex $\alpha_2 = 1, \ \alpha_1 = -4/3, \ \alpha_0 = 1/3$

$\beta_2 = \frac{2}{3}, \ \beta_1 = 0, \ \beta_0 = 0$

$$\begin{cases} y_{n+1} - \frac{4}{3} y_n + \frac{1}{3} y_{n-1} = \frac{2}{3} h f(t_{n+1}, y_{n+1}), \quad n = 1, \ldots, N-1 \\ y_0, y_1 \quad \text{given} \end{cases}$$

is a 2-step LM method.

Remarks (i) It is always possible to have $\alpha_k = 1$.

(ii) If $k = 1$, then method reduces to single-step.

(iii) If $\beta_k = 0$, the method is explicit. otherwise, implicit.

(iv)

Adams LM methods There are characterized by

$$\alpha_k = 1, \ \alpha_{k-1} = -1, \ \alpha_{k-2} = \cdots = \alpha_0 = 0$$

Adams methods can be explicit or implicit. ($\beta_k \neq 0$)
$\beta_k = 0$

The explicit ones are known as Adams-Bashforth methods.

The impl'r ones are known as Adams-Moulton methods

Adams-Bashforth methods:

$k=2$ $\quad \alpha_2=1,\ \alpha_1=-1,\ \alpha_0=0\ ;\ \beta_2=0,\ \beta_1=\frac{3}{2},\ \beta_0=-\frac{1}{2}$

$$\begin{cases} y_{n+1}-y_n = h\left[\frac{3}{2}f(t_n,y_n)-\frac{1}{2}f(t_{n-1},y_{n-1})\right], n=1,\cdots,N-1 \\ y_0, y_1 \text{ given} \end{cases}$$

$k=3$ $\quad \alpha_3=1,\ \alpha_2=-1,\ \alpha_1=\alpha_0=0$

$\beta_3=0,\ \beta_2=\frac{23}{12},\ \beta_1=\frac{-16}{12},\ \beta_0=\frac{5}{12}$

$$\begin{cases} y_{n+1}-y_n = h\left[\frac{23}{12}f(t_n,y_n)-\frac{16}{12}f(t_{n-1},y_{n-1})+\frac{5}{12}f(t_{n-2},y_{n-2})\right] \\ \qquad\qquad n=2,\cdots,N-1 \\ y_0, y_1, y_2 \text{ given} \end{cases}$$

$k=4$ $\quad \alpha_4=1,\ \alpha_3=-1,\ \alpha_2=\alpha_1=\alpha_0=0$

$\beta_4=0,\ \beta_3=\frac{55}{24},\ \beta_2=\frac{-59}{24},\ \beta_1=\frac{37}{24},\ \beta_0=-\frac{9}{24}$

$$\begin{cases} y_{n+1}-y_n = h\left[\frac{55}{24}f(t_n,y_n)-\frac{59}{24}f(t_{n-1},y_{n-1})+\frac{37}{24}f(t_{n-2},y_{n-2})\right. \\ \qquad\qquad \left.-\frac{9}{24}f(t_{n-3},y_{n-3})\right], n=3,\cdots,N-1 \\ y_0, y_1, y_2, y_3 \text{ given.} \end{cases}$$

Some Adams-Moulton methods

K=2 $\alpha_2=1,\ \alpha_1=-1,\ \alpha_0=0\ ;\ \beta_2=\frac{5}{12},\ \beta_1=\frac{8}{12},\ \beta_0=-\frac{1}{12}$

$$\begin{cases} y_{n+1}-y_n=h\left[\frac{5}{12}f(t_{n+1},y_{n+1})+\frac{8}{12}f(t_n,y_n)-\frac{1}{12}f(t_{n-1},y_{n-1})\right] \\ \qquad n=1,\dots,N-1 \\ y_0,\ y_1 \quad \text{given.} \end{cases}$$

K=3 $\alpha_3=1,\ \alpha_2=-1,\ \alpha_1=\alpha_0=0$

$$\beta_3=\frac{9}{24},\ \beta_2=\frac{19}{24},\ \beta_1=\frac{-5}{24},\ \beta_0=\frac{1}{24}.$$

$$y^{n+1}-y^n=h\left[\frac{9}{24}f(t_{n+1},y_{n+1})+\frac{19}{24}f(t_n,y_n)-\frac{5}{24}f(t_{n-1},y_{n-1})+\frac{1}{24}f(t_{n-2},y_{n-2})\right]$$

K=4 $\alpha_4=1,\ \alpha_3=-1,\ \alpha_2=\alpha_1=0=0$ $\qquad n=2,\dots,N-1$

$$\beta_4=\frac{251}{720},\ \beta_3=\frac{646}{720},\ \beta_2=-\frac{264}{720},\ \beta_1=\frac{106}{720},\ \beta_0=\frac{-19}{720}$$

$$y_{n+1}-y_n=\frac{h}{720}\left[251f_{n+1}+646f_n-264f_{n-1}+106f_{n-2}-19f_{n-3}\right]$$

$n=3,\dots,N-1$

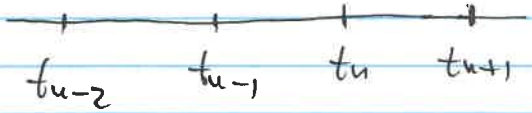
$y_0,\ y_1,\ y_2,\ y_3$ given.

LM methods are usually derived using <u>interpolation</u>.

<u>Ex.</u> 3-step Adams-Moulton method.

using the fundamental theorem of Calculus,

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



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

Let $P_3(t)$ be the cubic Lagrange interpolant of g on the nodes $\{t_{n-2}, t_{n-1}, t_n, t_{n+1}\}$. We have

$$P_3(t) = g(t_{n-2}) \frac{(t-t_{n-1})(t-t_n)(t-t_{n+1})}{(t_{n-2}-t_{n-1})(t_{n-2}-t_n)(t_{n-2}-t_{n+1})}$$

$$+ g(t_{n-1}) \frac{(t-t_{n-2})(t-t_n)(t-t_{n+1})}{(t_{n-1}-t_{n-2})(t_{n-1}-t_n)(t_{n-1}-t_{n+1})}$$

$$+ g(t_n) \frac{(t-t_{n-2})(t-t_{n-1})(t-t_{n+1})}{(t_n-t_{n-2})(t_n-t_{n-1})(t_n-t_{n+1})}$$

$$+ g(t_{n+1}) \frac{(t-t_{n-2})(t-t_{n-1})(t-t_n)}{(t_{n+1}-t_{n-2})(t_{n+1}-t_{n-1})(t_{n+1}-t_n)}.$$

We have

$$y(t_{n+1}) - y(t_n) = \int_{t_n}^{t_{n+1}} g(t)\,dt \approx \int_{t_n}^{t_{n+1}} P_3(t)\,dt$$

$$= h\left\{\frac{9}{24} g(t_{n+1}) + \frac{19}{24} g(t_n) - \frac{5}{24} g(t_{n-1}) + \frac{1}{24} g(t_{n-2})\right\}.$$

This serves as the motivation in the construction of the method

$$y_{n+1} - y_n = \frac{h}{24}\{9 f_{n+1} + 19 f_n - 5 f_{n-1} + 1 f_{n-2}\}.$$

Backward Differentiation methods

$$\begin{aligned}
\text{BDF1}: &\quad y_{n+1} - y_n &&= hf(t_{n+1}, y_{n+1}) \\
\text{BDF2}: &\quad y_{n+1} - \tfrac{4}{3}y_n + \tfrac{1}{3}y_{n-1} &&= \tfrac{2}{3}hf(t_{n+1}, y_{n+1}) \\
\text{BDF3}: &\quad y_{n+1} - \tfrac{18}{11}y_n + \tfrac{9}{11}y_{n-1} - \tfrac{2}{11}y_{n-2} &&= \tfrac{6}{11}hf(t_{n+1}, y_{n+1}) \\
\text{BDF4}: &\quad y_{n+1} - \tfrac{48}{25}y_n + \tfrac{36}{25}y_{n-1} - \tfrac{16}{25}y_{n-2} + \tfrac{3}{25}y_{n-3} &&= \tfrac{12}{25}hf(t_{n+1}, y_{n+1}) \\
\text{BDF5}: &\quad y_{n+1} - \tfrac{300}{137}y_n + \tfrac{300}{137}y_{n-1} - \tfrac{200}{137}y_{n-2} + \tfrac{75}{137}y_{n-3} - \tfrac{12}{137}y_{n-4} &&= \tfrac{60}{137}hf(t_{n+1}, y_{n+1}) \\
\text{BDF6}: &\quad y_{n+1} - \tfrac{360}{147}y_n + \tfrac{450}{147}y_{n-1} - \tfrac{400}{147}y_{n-2} + \tfrac{225}{147}y_{n-3} - \tfrac{72}{147}y_{n-4} + \tfrac{10}{147}y_{n-5} &&= \tfrac{60}{147}hf(t_{n+1}, y_{n+1})
\end{aligned}$$

BDF methods form a family of implicit linear multistep methods. Note that the first one, BDF1, is the implicit or Backward Euler method. We only list the first 6 members of the familty since these are the ones with satsifactory stability properties.

They are derived through Lagrange interpolation as follows. Consider the k-step method. We interpolate the data $\big(t_{n-k+1}, y(t_{n-k+1})\big), \ldots, \big(t_{n+1}, y(t_{n+1})\big)$. Let p_k denote the interpolant. We have

$$p_k(t) = \sum_{j=n-k+1}^{n+1} y(t_j) L_{k,j}(t)$$

where $L_{k,j}(t)$ are the canonical basis functions.

The Lagrange interpolation above implies $y'(t_{n+1}) \approx p_k'(t_{n+1})$ and consequently $f(t_{n+1}, y(t_{n+1})) \approx \sum_{j=n-k+1}^{n+1} y(t_j) L_{k,j}'(t_{n+1})$. Using this as motivation, the BDFk method is defined by

$$\sum_{j=n-k+1}^{n+1} \frac{L_{k,j}'(t_{n+1})}{L_{k,n+1}'(t_{n+1})} y_j = \frac{1}{L_{k,n+1}'(t_{n+1})} f(t_{n+1}, y_{n+1}),$$

where we have also used the scaling factor $\frac{1}{L_{k,n+1}'(t_{n+1})}$ in order for the term y_{n+1} on the left side to appear with coefficient 1. It is a simple exercise to show that $L_{k,n+1}'(t_{n+1}) = \frac{1}{h}\left[\frac{1}{k} + \frac{1}{k-1} + \cdots + 1\right]$.

k	1	2	3	4	5	6
$\frac{1}{L_{k,n+1}'(t_{n+1})}$	h	$\frac{2}{3}h$	$\frac{6}{11}h$	$\frac{12}{25}h$	$\frac{60}{137}h$	$\frac{60}{147}h$

It can be shown that the BDFk method has order of accuracy $O(h^k)$, $k \geq 1$. Indeed, the local truncation errors are given by

$$\begin{aligned}
\text{LTE(BDFk)} &= \frac{1}{hL_{k,n+1}'(t_{n+1})} \left| \sum_{j=n-k+1}^{n+1} L_{k,j}'(t_{n+1}) y(t_j) - y(t_{n+1}) \right| \\
&= \frac{1}{hL_{k,n+1}'(t_{n+1})} \left| p_k'(t_{n+1}) - y'(t_{n+1}) \right|.
\end{aligned}$$

Now $p_k(t) - y(t)$ vanishes ate the $k+1$ points $t_{n-k+1}, \ldots, t_{n+1}$. Hence $p_k'(t) - y'(t)$ vanishes at the k distinct points $\zeta_j \in (t_{n-k+j}, t_{n-k+j+1})$, $j = 1, \ldots, k$. Thus, $p_k'(t)$ is the Lagrange interpolant

of degree $k-1$ of $y'(t)$ at the nodes $\zeta_1, \ldots, \zeta_k$. Hence, it follows that for some $\xi \in (\zeta_1, \zeta_k) \subseteq (t_{n-k+j}, t_{n+1})$

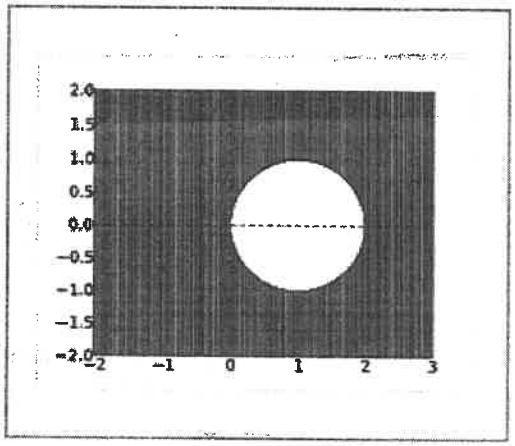
$$y'(t_{n+1}) - p_k'(t_{n+1}) = \frac{y^{(k+1)}(\xi)}{(k+1)!}(t_{n+1}-\zeta_1)\cdots(t_{n+1}-\zeta_k).$$

Since the term $\frac{1}{hL'_{k,n+1}(t_{n+1})}$ is $O(1)$, it follows that LTE(BDFk) $\leq c|y^{(k+1)}(\xi)|h^k$.

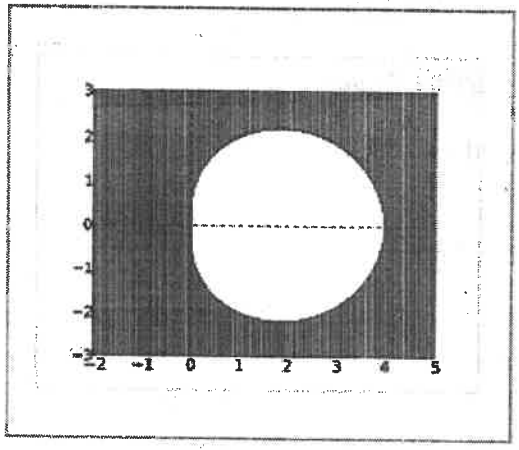
The BDF methods have good stability properties for $k = 1, \ldots, 6$. The plots below show the regions of absolute stability of these methods. They are all A_0 stable and the first two are A-stable. Indeed the region of absolute stability of BDFk contains the sector in the left-half complex plane $\{z \in \mathbb{C},\ \mathrm{Re}z < 0,\ \arg z \in (\pi - \theta_k, \pi + \theta_k)\}$ with the angles θ_k given by (in degrees)

k	1	2	3	4	5	6
θ_k	90	90	86.03	73.35	51.84	17.84

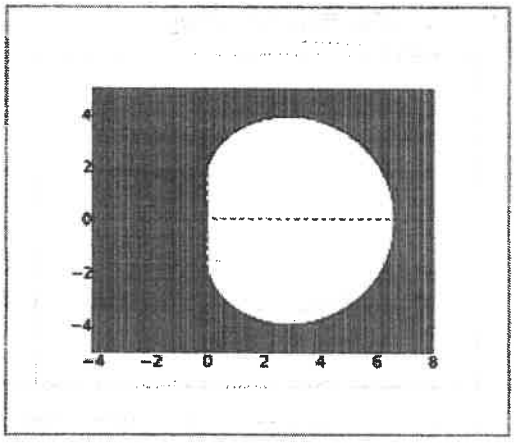
In fact methods with regions of absolute stability containing such sectors are called $A(\theta)$-stable.



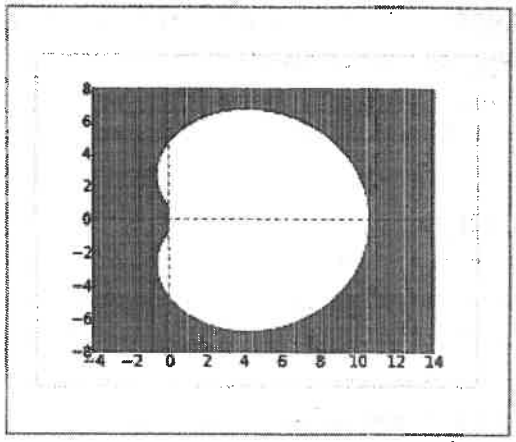
BDF1



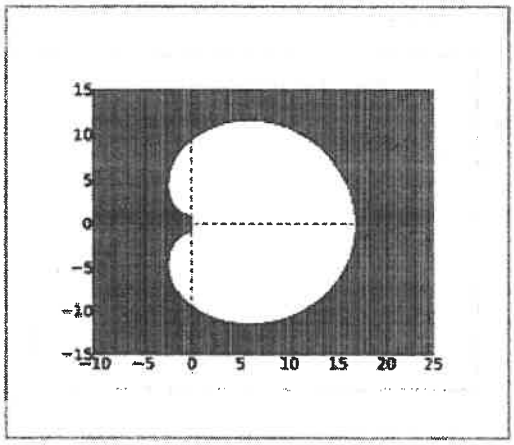
BDF2



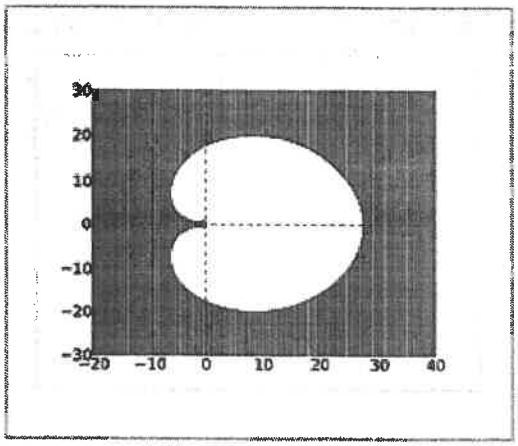
BDF3



BDF4


15
10
5
0
−5
−10
−15
−10 −5 0 5 10 15 20 25

BDF5


30
20
10
0
−10
−20
−30
−20 −10 0 10 20 30 40

BDF6