



A STUDY OF NUMERICAL METHODS FOR DIFFERENTIAL EQUATION

DIPAK LANJEWAR

RESEARCH SCHOLAR, CHHATRAPATI SHAHU JI MAHARAJ UNIVERSITY, KANPUR

GUIDE NAME: DR. R.K. VERMA

ABSTRACT

This paper studies a number of aspects of numerical methods for ordinary differential equations. The discussion includes the method of Euler and introduces Runge-Kutta methods and linear multistep methods as generalizations of Euler. Stability considerations arising from stiffness lead to a discussion of implicit methods and implementation issues. To the extent possible within this short survey, numerical methods are looked at in the context of problems arising in practical applications. Differential equations can describe nearly all systems undergoing change. They are ubiquitous in science and engineering as well as economics, social science, biology, business, health care, etc. Many mathematicians have studied the nature of these equations for hundreds of years and there are many well-developed solution techniques. Often, systems described by differential equations are so complex, or the systems that they describe are so large, that a purely analytical solution to the equations is not tractable. It is in these complex systems where computer simulations and numerical methods are useful. The techniques for solving differential equations based on numerical approximations were developed before programmable computers existed. During World War II, it was common to find rooms of people (usually women) working on mechanical calculators to numerically solve systems of differential equations for military calculations.

INTRODUCTION

Differential equations play a role in the modeling of almost every scientific discipline. However, it is relatively rare for a differential equation to have a solution that can be written in terms of elementary functions. Usually, the only information about the solution is that it is known to exist and to be unique, on theoretical

grounds, and that it can be approximated more or less accurately using computational techniques. In this review paper, we will consider some aspects of numerical methods for the solution of initial value problems in systems of ordinary differential equations. There are two standard forms for expressing such problems. The first of these is

$$y'(x) = f(x, y(x)), \quad y(x_0) = y_0. \quad (1)$$

Here the solution y is assumed to be a differentiable function on an interval $[x_0, \bar{x}]$ to a finite dimensional Euclidean space \mathbb{R}^N . The formulation (1) is very general and includes, for example, second and higher order differential equations; these are easily recast in

this way. By introducing an additional variable, if necessary, which always remains exactly equal to x , it is possible to reformulate the general problem as an 'autonomous' system of equations. This is the second standard form.

$$y'(x) = f(y(x)), \quad y(x_0) = y_0. \quad (2)$$

EULER METHOD AND RELATED METHODS

The most natural physical interpretation of a differential equation and its solution is in terms of distance and velocity. If x is interpreted as 'time' and $y(x)$ as the position of a moving particle at a particular time then the value of $f(x, y(x))$ is the velocity at this time. Hence, we can interpret the solution at $x_0 + h$,

$$y_1 + hf(x_0, y_0).$$

The numerical method originally proposed by Euler extends this idea by generating approximations at a sequence of points, $x_i =$

$$y_i = y_{i-1} + hf(x_{i-1}, y_{i-1}), \quad i = 1, 2, \dots \quad (4)$$

If it is required to find the solution at a known output point x then it is convenient to choose $h = (x - x_0)/n$, where n is an integer.

$$y(x_0 + h) = y(x_0) + hy'(x_0) + \frac{h^2}{2}y''(x_0) + \dots, \quad (5)$$

it should be expected that, at least for small values of h , the error in completing each step is approximately proportional to h^2 . Since the total number of steps is proportional to h^{-1} , this would mean that the total error committed by the time n steps have been completed to produce an approximation to $y(x)$, would be approximately proportional to h .

The fact that errors behave like the first power of h is regarded as a serious limitation of the Euler method, because reducing h , and therefore increasing the total computational effort, leads to only a modest improvement in accuracy. What are preferred are 'higher order' methods for which the error in a step

where h is a small time interval as being the value of y_0 to which has been added the product of the width of this interval and the average velocity over this interval. Of course, if we are given only the differential equation (1) we have no obvious means of calculating the average velocity; as a very first approximation, the average can be replaced by the velocity at the beginning of the interval; that is $f(x_0, y_0)$. Hence, if y_1 denotes an approximate solution at $x_1 = x_0 + h$, we have a possible numerical method based on the formula for its first step

(3)

$x_0 + hi$, $i = 1, 2, \dots$ where each point is related to the one next before it by the formula

Because (3) can be looked at as the first two terms of a Taylor expansion

behaves approximately like h^{p+1} and the total or global error behaves like hp , where the order p is 2 or more.

Using the velocity-distance interpretation of a differential equation and its solution, we can first consider how to overcome the limitation of having to use, instead of the average velocity in a time interval, the value at the beginning of the interval. Several approaches have been used for approximating the average velocity more accurately than in the Euler method. One idea is to somehow use an approximation at the mid-point of the interval, instead of at the left-hand end. A second idea is to use the mean of the values at the two ends



of the interval. The Runge-Kutta method, which we will explore in more detail in the next section, was originally based by Runge on each of these ideas. To obtain an approximation of the derivative at the centre or the end of the interval, it is possible to take a temporary step to the desired point and calculate the derivative using this approximate Mid-point rule method:

$$y_{i-1/2}^* = y_{i-1} + \frac{1}{2}hf(x_{i-1}, y_{i-1})$$

$$y_i = y_{i-1} + hf\left(x_{i-1} + \frac{1}{2}h, y_{i-1/2}^*\right)$$

(6 & 7)

Trapezoidal rule method:

$$y_i^* = y_{i-1} + hf(x_{i-1}, y_{i-1})$$

$$y_i = y_{i-1} + \frac{h}{2}(f(x_{i-1}, y_{i-1}) + f(x_{i-1} + h, y_i^*))$$

(8)

Each of the ‘mid-point rule Runge-Kutta method’, (5) and (6), and the ‘trapezoidal rule RungeKutta method’, (7) and (8), is of order $p = 2$ and is thus more accurate than the simple Euler method, as long as h is sufficiently small.

Another approach to approximating the average velocity within the interval $[x_{i-1}, x_i]$ is to note that between the beginning and end of the most recently completed step, from x_{i-2} to

x_{i-1} , the velocity had increased from approximately $f(x_{i-1}, y_{i-1})$ to $f(x_{i-2}, y_{i-2})$; that is, $f(x_{i-2}, y_{i-2}) - f(x_{i-1}, y_{i-1})$ per step. This suggests that within the next half step it will increase further by approximately $\frac{1}{2}(f(x_{i-2}, y_{i-2}) - f(x_{i-1}, y_{i-1}))$. Hence, it might be a reasonable approximation to the value of the average derivative within a step to add this quantity onto the derivative at the beginning: $f(x_{i-1}, y_{i-1})$. Hence the method constructed in this way can be written in the form

$$y_i = y_{i-1} + h\left(\frac{3}{2}f(x_{i-1}, y_{i-1}) - \frac{1}{2}f(x_{i-2}, y_{i-2})\right), \quad i = 2, 3, \dots \quad (9)$$

Of course, in the very first step, from x_0 to $x_1 = x_0 + h$, this method cannot be used. However, once the first step has been completed, it can be used in every later step.

The method given by (9), and known as an Adams-Bashforth method is also of order 2 but it differs from the two Runge-Kutta methods that have been given in two different ways. The first is that the value of the function f is evaluated only once in each step; thus it is less

computationally expensive. The second difference is that it is a multistep method; this means that the value computed in a step depends on two previous values; the method is therefore more complicated to use as we have already seen.

$$\begin{aligned}
 y(x_i) - y(x_{i-1}) - \frac{3}{2}hy'(x_{i-1}) + \frac{1}{2}hy'(x_{i-2}) \\
 &= y(x_{i-1} + h) - y(x_{i-1}) - \frac{3}{2}hy'(x_{i-1}) + \frac{1}{2}hy'(x_{i-1} - h) \\
 &= \left(y(x_{i-1} + hy'(x_{i-1}) + \frac{1}{2}y''(x_{i-1} + O(h^3))) - y(x_{i-1}) \right) \\
 &\quad - \frac{3}{2}hy'(x_{i-1}) + \frac{1}{2} \left(hy'(x_{i-1} + y''(x_{i-1} + O(h^3))) \right) \\
 &= O(h^3)
 \end{aligned}$$

Because the Adams-Bashforth methods give the value of y_i as a linear combination of y_{i-1} , $hf(x_{i-1}, y_{i-1})$ and the values of the same quantities but with other subscripts, they are known as 'linear multistep methods'. If the average derivative within step number i is computed as the mean of $hf(x_{i-1}, y_{i-1})$ and

$$y_i = y_{i-1} + \frac{h}{2} (f(x_i, y_i) + f(x_{i-1}, y_{i-1})), \quad i = 1, 2, \dots \quad (10)$$

Another special example of a linear multistep method, also of implicit type is written in the form

$$y_i = ay_{i-1} + by_{i-2} + chf(x_i, y_i).$$

By expanding by Taylor series and matching coefficients, it is found that the unique choice of a , b and c which gives order 2 is $a = \frac{4}{3}$, $b = -\frac{1}{3}$, $c = \frac{2}{3}$. The method that results from this choice:

$$y_i = \frac{4}{3}y_{i-1} - \frac{1}{3}y_{i-2} + \frac{2}{3}hf(x_i, y_i) \quad (11)$$

The order condition for (9) may be verified by expanding the difference of the two sides in Taylor series. That is,

$hf(x_i, y_i)$, we get the second order example of what is known as an 'Adams-Moulton method'. Like all methods in this family, this method is implicit (because y_i is given as the solution to an algebraic equation, rather than by an explicit formula).

is known as a BDF or backward difference method.

RUNGE-KUTTA METHODS

Runge-Kutta methods, as we have seen, are 'one step' in the sense that the result found at the end of a step is functionally dependent only on the result given at the end of the previous step. That is, if y_n denotes a computed approximation to $y(x_n)$, then y_n is given by a formula of the form



$$y_n = y_{n-1} + h \sum_{i=1}^s b_i F_i,$$

where the quantities F_1, F_2, \dots, F_s are derivatives computed from approximations Y_1, Y_2, \dots, Y_s to the solution at $x_{n-1} + hc_1, x_{n-1} + hc_2, \dots, x_{n-1} + hc_s$. That is, $F_i = f(x_{n-1} + hc_i, Y_i)$,

$i = 1, 2, \dots, s$ for the differential equation system (1) or $F_i = f(Y_i)$, $i = 1, 2, \dots, s$ for the autonomous system (2). The values of Y_i , $i = 1, 2, \dots, s$ are found from the equation

$$y_i = y_{n-1} + h \sum_{j=1}^s a_{ij} F_j, \quad i = 1, 2, \dots, s.$$

It turns out that the components of the c vector are related to the elements of the A matrix by

$$c_i = \sum_{j=1}^s a_{ij}, \quad i = 1, 2, \dots, s.$$

The number of stages s is the number of Y vectors needed to compute the solution in a method of this form and is a measure of the complexity of a particular method. The

characteristic coefficients of a specific Runge-Kutta method are conveniently displayed in a tableau as follows

c_1	a_{11}	a_{12}	\cdots	a_{1s}
c_2	a_{21}	a_{22}	\cdots	a_{2s}
\vdots	\vdots	\vdots		\vdots
c_s	a_{s1}	a_{s2}	\cdots	a_{ss}
	b_1	b_2	\cdots	b_s

Even though we have allowed for the possibility of implicit methods in this formulation, the traditional Runge-Kutta methods, for example those due to Runge, Kutta, Nyström and other early contributors,

have been explicit. This means that a_{ij} is precisely zero unless $i > j$ (and consequently $c_1 = 0$) because each quantity used in the computation should be functionally dependent only on other quantities already computed.

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



Note that here, as is usual for explicit methods, the zero elements on and above the diagonal of

$$\begin{array}{c|ccc}
 0 & & & \\
 \frac{1}{2} & \frac{1}{2} & & \\
 1 & -1 & 2 & \\
 \hline
 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6}
 \end{array}
 \qquad
 \begin{array}{c|ccc}
 0 & & & \\
 \frac{1}{2} & \frac{1}{2} & & \\
 \frac{1}{2} & 0 & \frac{1}{2} & \\
 1 & 0 & 0 & 1 \\
 \hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
 \end{array}$$

A have been omitted.

The fourth order method has become very popular since it was proposed by Kutta and is sometimes referred to as ‘the Runge-Kutta method’, as though it were the only one available. To verify the order of these and other Runge-Kutta methods it is necessary to

expand the exact and computed solutions in powers of h and to check that the terms up to and including those with an exponent p agree with each other. For example, it can be shown that the exact solution for (2) near an initial point (x0, y0) is given by the expansion

$$\begin{aligned}
 y(x_0 + h) = & y_0 + hf + \frac{h^2}{2}f'(f) + \frac{h^3}{6}(f''(f, f) + f'(f'(f))) \\
 & + \frac{h^4}{24}(f'''(f, f, f) + 3f''(f, f'(f)) + f'(f''(f, f)) + f'(f'(f'(f)))) + O(h^5), \quad (12)
 \end{aligned}$$

where $f = f(y_0)$, $f' = f'(y_0)$, $f'' = f''(y_0)$, $f''' = f'''(y_0)$. On the other hand, the

Taylor expansion for the numerical solution computed by an explicit Runge-Kutta with s = 4 is equal to

$$\begin{aligned}
 y_1 = & y_0 + h(b_1 + b_2 + b_3 + b_4)f + \frac{h^2}{2}(b_2c_2 + b_3c_3 + b_4c_4)f'(f) \\
 & + \frac{h^3}{6}(2(b_2c_2^2 + b_3c_3^2 + b_4c_4^2)f''(f, f) + (b_3a_{32}c_2 + b_4a_{42}c_2 + b_4a_{43}c_3)f'(f'(f))) \\
 & + \frac{h^4}{24}(6(b_2c_2^3 + b_3c_3^3 + b_4c_4^3)f'''(f, f, f) \\
 & + (b_3c_3a_{32}c_2 + b_4c_4a_{42}c_2 + b_4c_4a_{43}c_3)f''(f, f'(f)) \\
 & + (b_3a_{32}c_2^2 + b_4a_{42}c_2^2 + b_4a_{43}c_3^2)f'(f''(f, f)) \\
 & + b_4a_{43}a_{32}c_2f'(f'(f'(f)))) + O(h^5), \quad (13)
 \end{aligned}$$

A comparison of the terms in (12) and (13) shows that agreement up to h4 terms occurs if and only if



$$\begin{aligned}
 b_1 + b_2 + b_3 + b_4 &= 1, \\
 b_2c_2^2 + b_3c_3^2 + b_4c_4^2 &= \frac{1}{3}, \\
 b_2c_2^3 + b_3c_3^3 + b_4c_4^3 &= \frac{1}{4}, \\
 b_3a_{32}c_2^2 + b_4a_{42}c_2^2 + b_4a_{43}c_3^2 &= \frac{1}{12}, \\
 b_2c_2 + b_3c_3 + b_4c_4 &= \frac{1}{2}, \\
 b_3a_{32}c_2 + b_4a_{42}c_2 + b_4a_{43}c_3 &= \frac{1}{6}, \\
 b_3c_3a_{32}c_2 + b_4c_4a_{42}c_2 + b_4c_4a_{43}c_3 &= \frac{1}{8}, \\
 b_4a_{43}a_{32}c_2 &= \frac{1}{24}.
 \end{aligned}$$

These are easily seen to be satisfied by the values $a_{21} = a_{32} = c_2 = c_3 = \frac{1}{2}$, $a_{31} = a_{41} = a_{42} = 0$, $a_{43} = c_4 = 1$, $b_1 = b_4 = \frac{1}{6}$, $b_2 = b_3 = \frac{1}{3}$ to give the classical fourth order method. Explicit methods are known at least as high as order 10 but these require increasingly more stages. For example, for $p = 5$, $s = 6$ is necessary and for $p = 8$, $s = 11$ is necessary.

In contrast to the complicated relationship between the value of p and the minimal value of s to achieve this order for explicit methods, for implicit methods there is a simple relationship between s and p . This is that for any positive integer s there exists an implicit Runge-Kutta method with order $p = 2s$ (but no higher). In fact methods with these high orders are a generalization of Gaussian quadrature formulas and reduce to them in the special case of the trivial differential equation $y'(x) = f(x)$.

We give a single example of one of the Gauss family of methods, the method with $s = 2$ and $p = 4$.

$$\begin{array}{c|cc}
 \frac{1}{2} - \frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\
 \frac{1}{2} + \frac{\sqrt{3}}{6} & \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \\
 \hline
 & \frac{1}{2} & \frac{1}{2}
 \end{array} \tag{14}$$

LINEAR MULTISTEP METHODS: linear multistep methods make use of already computed solution values and derivatives over

several previous steps but only evaluate the function f once in each step. The general form of these methods is thus

$$y_n = \alpha_1 y_{n-1} + \alpha_2 y_{n-2} + \dots + \alpha_k y_{n-k} + \beta_0 f_n + \beta_1 f_{n-1} + \beta_2 f_{n-2} + \dots + \beta_k f_{n-k},$$

where f_i is defined as $f(x_i, y_i)$ for problem (1) and as $f(y_i)$ for problem (2). Note that if $\beta_0 \neq 0$, the method is implicit and the approximation y_n has to be determined as the solution of an algebraic equation. Assuming either $\alpha_k \neq 0$ or $\beta_k \neq 0$, the integer k is a measure of the complexity of these methods and the method is sometimes known as a k -step method.



The order of linear multistep methods is easy to determine by Taylor series analyses. In conditions for various orders, the α and β

$$\alpha_1 + \alpha_2 + \dots + \alpha_k = 1, \quad (15)$$

$$\alpha_1 + 2\alpha_2 + \dots + k\alpha_k = \beta_0 + \beta_1 + \beta_2 + \dots + \beta_k. \quad (16)$$

These conditions are of central importance to the study of linear multistep methods and are usually known as the 'consistency conditions'. It turns out that for a method to be capable of producing a sequence of approximations which converge to the exact solution as $h \rightarrow 0$, it is necessary and sufficient that the method be both consistent and 'stable' where a stable method is one for which the polynomial

$$z^k - \alpha_1 z^{k-1} - \dots - \alpha_k$$

has zeros only in the closed unit disc and repeated zeros only in the open unit disc.

Amongst explicit methods, the most important are the Adams-Bashforth methods. For these $\alpha_1 = 1$ and each of $\alpha_2, \dots, \alpha_k$ is zero. Furthermore, the values of $\beta_1, \beta_2, \dots, \beta_k$ are chosen in such a way as to give an order of $p = k$. Corresponding implicit Adams-Moulton methods have the same values of the α s and, because β_0 is an additional parameter, it is possible to obtain an order $p = k + 1$.

There are good reasons, which we will discuss later for combining an Adams-Bashforth with an Adams-Moulton method of the same order (or possibly with an order greater by 1) into a single algorithm. In these so-called 'predictor-corrector' pairs, the Adams-Bashforth predictor is used to obtain an approximate 'predicted' value of y_n from which an approximate value of f_n is computed. The approximation to t_n is then 'corrected' using the Adams-Moulton formula but with f_n

coefficients all occur linearly and therefore this sort of question is much simpler than for Runge-Kutta methods.

replaced by the value computed in the predictor stage of the algorithm. Many variants of this scheme are used in practical programs. It might be thought that the loss of generality in assuming the special form for the α coefficients is a disadvantage of the Adams methods. Even though by allowing a more general form for the method so that formally orders as high as $2k$ are actually possible, the stability restrictions generally makes these methods useless. The greatest order for a linear multistep method that is also stable and therefore convergent is $k + 2$ (or only $k + 1$ if k is an odd integer).

Instead of the choice of coefficients made in the Adams methods, it is also possible to use the values $\beta_1 = \beta_2 = \dots = \beta_k = 0$ with $\beta_0, \alpha_1, \alpha_2, \dots, \alpha_k$ selected to give order $p = k$. These backward difference methods, of which an example is (11), have a special role in the solution of stiff problems.

CONCLUSION: In this short survey of numerical methods for ordinary differential equations it has been possible to mention only selected aspects of this very active research area. We have not mentioned methods that make use of higher derivatives of the solution, for example Taylor series methods. We have also not mentioned methods that are both multistage (as in Runge-Kutta methods) and multivalued (as in linear multistep methods); these have many of the desirable properties of each of the main traditional classes which they generalize. It has also not been possible to



discuss recent exciting work on the solution of problems arising from a Hamiltonian formulation of mechanical problems. It is important to identify numerical methods with a 'symplectic' character because these are capable of accurately preserving theoretical invariants that, in general, cannot be held constant with other numerical methods.

REFERENCES

- [1]. J.C. Adams, Appendix in F. Bashforth, An attempt to test the theories of capillary action by comparing the theoretical and measured forms of drops of fluid. With an explanation of the method of integration employed in constructing the tables which give the theoretical form of such drops, by J.C.Adams. Cambridge Univ. Press. (1883).
- [2]. K. Burrage, J. C. Butcher and F. H. Chipman, An implementation of singly-implicit methods, BIT, 20 (1980), 326-340.
- [3]. J. C. Butcher, Coefficients for the study of Runge-Kutta integration processes, J. Austral. math. Soc., 3 (1963), 185-201.
- [4]. J. C. Butcher, Implicit Runge-Kutta Processes, Math. Comput., 18 (1964), 50-64.
- [5]. J. C. Butcher, The Numerical Analysis of Ordinary Differential Equations, John Wiley & Sons (1986).
- [6]. A.L. Cauchy, R esum e des Lecons donn ees  al'Ecole Royale Polytechnique. Suite du Calcul Infinitesimal; published: Equations diff erentielles ordinaires, ed. Chr. Gilain, Johnson 1981.
- [7]. G. Coriolis, M emoires sur le degr ed'approximation qu'on obtient pour les valeurs num eriques d'une variable qui satisfait  a une  equation diff erentielle, en employant pour calculer ces valeurs diverses  equations aux diff erences plus ou moins approch ees, J. de Math ematiques pures et appliqu ees (Liouville), 2 (1837), 229-244.
- [8]. C.F. Curtiss & J.O. Hirschfelder, Integration of stiff equations. Proc. Nat. Acad. Sci., 38 (1952), 235-243.
- [9]. G. Dahlquist, Convergence and stability in the numerical solution of ordinary differential equations, Math. Scand. 4 (1956), 33-53. [10] G. Dahlquist, A special stability problem for linear multistep methods, BIT 3 (1963), 27-43.
- [10]. B.L. Ehle, On Pad e approximations to the exponential function and A-stable methods for the numerical solution of initial value problems, Research Report CSRR 2010, Dept. AACS, Univ. of Waterloo, Ontario, Canada (1969).
- [11]. L. Euler, Institutionum Calculi Integralis. Volumen Primum, (1768), Opera Omnia, Vol.XI.