# NUMERICAL SOLUTIONS OF ORDINARY DIFFERENTIAL EQUATIONS

139609

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ÇANKAYA UNIVERSITY

BY

SİBEL ÇEVİK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE

JUNE 2003

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

139609

Approval of the Graduate School of Natural and Applied Sciences

Prof.Dr.Yurdahan GÜLER
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof.Dr.Yurdahan GÜLER
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof.Dr.Kenan TAŞ
Supervisor

Examining Committee Members

Prof. Dr. Kenan TAŞ

Prof. Dr. Şeref MİRASYEDİOĞLU

Asst. Prof. Dr. Emre SERMUTLU

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Prof. Dr. Kenan Taş for his patient guidance and support during the course of this study.

I would like to thank Prof. Dr. Şeref Mirasyedioğlu for being on my committee, for taking time to review this research, and for his advice.

I would also like to show my appreciation to Asst. Prof. Dr. Emre Sermutlu for also being on my committee and for his advice and direction with solving equations.

I am most grateful to my colleagues at the Department of Mathematics and Computer for the warm and supportive atmosphere they created for me during my work.

I would like to thank my parents for their continuous encouragements, their moral support, and their unending love.

Finally my deepest appreciation goes to Tolga Ersan, Nurdan Buz and all my friends. The completion of this work would not have been possible without their overwhelming love, support and sacrifices.

## ABSTRACT

### NUMERICAL SOLUTIONS OF ORDINARY DIFFERENTIAL EQUATIONS

Sibel Çevik
M.S., Department of Mathematics and Computer
Supervisor: Prof. Dr. Kenan Taş

June 2003, 65 pages

Since ordinary differential equations are useful in modelling the behavior of many physical processes, methods of solution for these equations are of great importance to engineers and scientists. Even though well-known analytical techniques can solve many important differential equations, a greater number of physically significant differential equations can not be solved using these techniques. Fortunately, the solutions of these equations can usually be generated numerically.

There are many methods for finding approximate solutions to differential equations. Throughout the thesis, numerical techniques for ordinary differential equations are considered. In the first chapter, basic concepts which are going to be used are given. Second chapter contains numerical methods, all of which do not generate exact solutions, only approximate ones. Finally, in the last chapter a new numerical integration technique inspired by the Runge-Kutta method to solve the initial value problem is given. The method presented adds higher order derivative terms to the Runge-Kutta stage equations resulting in a higher order method without increasing the number of stages.



Keywords: Ordinary Differential Equations, Numerical Analysis, Runge-Kutta Method.

# ÖZ

## ADİ DİFERENSİYEL DENKLEMLERİN NÜMERİK ÇÖZÜMLERİ

Sibel Çevik
Yüksek Lisans, Matematik-Bilgisayar Bölümü
Tez Yöneticisi: Prof. Dr. Kenan Taş

Haziran 2003, 65 sayfa

Diferensiyel denklemler, birçok fiziksel olayı modellemede kullanıldığından, bu denklemlerin çözüm metotları fizik ve mühendislik gibi alanlarda çalışanlar için büyük önem taşımaktadır. Bilinen analitik tekniklerle birçok denklem çözülebilmesine rağmen, önemli sayıda fiziksel uygulamalar için bu metotlar yetersiz kalmaktadır. Böyle denklemler ancak nümerik metotlarla çözülebilirler.

Diferensiyel denklemlerde yaklaşık sonuç bulan birçok metot bulunmaktadır. Bu tezde, tüm bu nümerik metotlar ele alınmıştır. Birinci bölümde, tezde kullanılacak temel kavramlar, ikinci bölümde ise; denklemlerin kesin sonuçlarını değil, ancak yaklaşık sonuçlarınıveren nümerik metotlar verilmiştir. Son bölümde ise Runge-Kutta metot baz alınarak geliştirilmiş başlangıç değer problemlerini çözen yeni bir metot incelenmiştir. Bu metot, Runge-Kutta metoduna, işlem sayısını arttırmadan, yüksek mertebeden türevler eklenerek elde edilmiştir.

Anahtar kelimeler: Adi diferensiyel denklemler, Nümerik analiz, Runge-Kutta metot.

## TABLE OF CONTENTS

## LIST OF TABLES

TABLE

## LIST OF FIGURES

FIGURE

# CHAPTER 1

# INTRODUCTION

## 1.1 Ordinary Differential Equations

An n-th order differential equation has the form

$$y^{(n)}(x) = F[x, y(x), y'(x), \ldots, y^{n-1}(x)] \tag{1.1}$$

where $y^{(k)} = d^k y/dx^k$. Here y is the dependent variable, and x is the independent variable. The order of the differential equation is defined as the order of the highest derivative appearing in it.

Equation (1.1) is a linear differential equation if F is a linear function of y and its derivatives. Formally, a linear differential equations may be written as

$$Ly(x) = f(x) \tag{1.2}$$

where L is a linear differential operator:

$$L = p_0(x) + p_1(x)\frac{d}{dx} + \cdots + p_{n-1}(x)\frac{d^{n-1}}{dx^{n-1}} + \frac{d^n}{dx^n}. \tag{1.3}$$

A nonlinear ordinary differential equation is an equation which is not linear. Nonlinear equations have a richer mathematical structure than linear equations and are generally more difficult to solve. Nevertheless, we'll see, the solution of many difficult-looking nonlinear equations is quite routine.

## 1.2 Initial-Value and Boundary-Value Problems

A solution $y(x)$ to a differential equation is not uniquely determined by the differential equation alone; the values of the n independent constants of integration must also be specified. These constants of integration may be specified in several disparate ways. In an initial-value problem one specifies y and its first

$n-1$ derivatives $y', \ldots, y^{(n-1)}$, at one point $x = x_0$:

$$\begin{aligned} y(x_0) &= a_0, \\ y'(x_0) &= a_1, \\ &\vdots \\ y^{(n-1)}(x_0) &= a_{n-1}. \end{aligned} \tag{1.4}$$

The numbers $a_0, a_1, \ldots, a_{n-1}$ are the n constants of integration of the initial-value problem.

In a boundary-value problem a total of n quantities are specified at two or more points. For example, for a fifth-order differential equation, one might impose the conditions: $y(x_1) = a_1, y'(x_2) = a_2, y'(x_3) = a_3, y''(x_3) = a_4, y'''(x_1) + [y(x_2)]^2 = a_5$. It may be that some or all of the points $x_1, x_2, \ldots$ are interior to and not on the boundary of the region in which $y(x)$ is to be found. In any case, the quantities $a_i$ are still the constants of integration for the problem.

Initial-value problems are much simpler than boundary-value problems. Initial-value problems may be subjected to local analysis to determine whether a unique solution $y(x)$ exists in a sufficiently small neighborhood of $x_0$. If it can be shown that the initial-value problem has a solution near $x_0$, the next step is to determine the global properties of this solution. In particular, one must determine the extent of the interval containing $x_0$ on which the local solution exists.

In example the solution to $y' = (\tan x)y + 1$ $[y(0) = 1]$ exists in the interval in which $\tan x$ is continuous, namely $(-\pi/2, \pi/2)$. However, the solution $y(x) = (1 + \sin x)/\cos x$ actually exists for $-3\pi/2 < x < \pi/2$

Boundary-value problems are inherently global. Existence and uniqueness theorems for solutions must be proved for an interval large enough to include all the points $x_1, x_2, \ldots$. Local analysis of the solution near any of these points is insufficient.

### 1.3 Methods Of Solution

When we say that we shall solve a differential equations we mean that we shall find one or more of its solutions. These equations can be solved by using analytical or numerical methods. Numerical methods give approximate values of the solution functions corresponding to selected values of the independent

variables. These methods are not so desirable as analytical methods because of the amount of work involved in them and because the results obtained from them are only approximate; But if analytical methods are not applicable, one has no choice but to turn to approximate methods.

Modern science and engineering problems continue to give rise to differential equations to which analytical methods do not apply, and approximate methods are becoming increasingly more important.

Some of the first order ordinary differential equations can be given

$$\begin{array}{lll} \text{Bernouilli Equation} & : & y' + P(x)y = Q(x)y^n \\ \text{Riccati Equation} & : & y'(x) = a(x)y^2(x) + b(x)y(x) + c(x) \\ \text{Clairaut's Equation} & : & f(xy' - y) = g(y') \end{array} \tag{1.5}$$

These equations are solved analytically. Other analytical method, transformations, may be used to reduce equations to more tractable types. To illustrate, equidimensional-in-x and equidimensional-in-y equations are left invariant under the transformation $x \to ax$ and $y \to ay$, where a is a constant. An equation of the former can be converted to an autonomous equation, in which there are no explicit occurrences of the independent variable, of the same order. And the latter can be converted to an equation of lower order.

# CHAPTER 2

# NUMERICAL METHODS

It is often difficult to find the analytic solution to many differential equations. There are many methods for finding approximate solutions to differential equations. These methods are referred to by a variety of different names including:numerical methods, numerical integration or approximate solutions.

None of the methods generate exact solutions, only approximate ones. Additionally, because these methods are based on computation, they provide solutions only at certain discrete time intervals. Specifically, the solutions are available at the time of the initial conditions and for every time interval, h, thereafter (i.e. at $t = t_0, t_0 + h, t_0 + 2h, \ldots, t_0 + kh$ where k is an integer.)

Firstly, quasilinearization technique which is going to be used to linearize the nonlinear ordinary differential equations is given and later numerical methods solving ordinary differential equations are considered.

## 2.1 Quasilinearization

Quasilinearization is the process of solving nonlinear problems by the following steps:

1. Linearizing the nonlinear ordinary differential equations around a solution which satisfies the specified boundary conditions.

2. Solving a sequence of linear problems in which the solution of the (k)th linear problem satisfies the specified boundary conditions and is taken for the (k+1)st linear problem.[1]

In the limit the solution of the linear problems converges to the solution of the original nonlinear problem.

Consider the second-order nonlinear ordinary differential equation

$$3yij + j^2 = 0 \tag{2.1}$$

with the boundary conditions

$$\begin{aligned} y(0) &= 0 \\ y(1) &= 1 \end{aligned} \tag{2.2}$$

To transform (2.1) into a system of first-order equations, let

$$y_1 = y \qquad\qquad y_2 = \dot{y_1}, \tag{2.3}$$

and replace (2.1) with the set

$$\dot{y_1} = y_2 = g_1(y_2), \qquad\qquad \dot{y_2} = -\frac{y_2^2}{3y_1} = g_2(y_1, y_2) \tag{2.4}$$

with the boundary conditions

$$y_1(0) = 0 \qquad\qquad y_1(1) = 1, \tag{2.5}$$

On expanding the right-hand side of (2.4) in a Taylor's series around solutions $y_1^k(t), y_2^k(t)$ we have

$$g_i(y_1, y_2) = g_i(y_1^k(t), y_2^k(t)) + (\frac{\partial g_i}{\partial y_1})\delta y_1^k + (\frac{\partial g_i}{\partial y_2})\delta y_2^k \qquad i = 1, 2 \tag{2.6}$$

where the partial derivatives are evaluated using $y_1^k(t), y_2^k(t)$. From (2.4) it follows that

$$\begin{aligned} \frac{\partial g_1}{\partial y_1} &= 0 & \frac{\partial g_1}{\partial y_2} &= 0 \\ \frac{\partial g_1}{\partial y_1} &= \frac{1}{3}(\frac{y_2^k(t)}{y_1^k(t)})^2 & \frac{\partial g_2}{\partial y_2} &= -\frac{2}{3}(\frac{y_2^k(t)}{y_1^k(t)}) \\ \delta y_1^k &= y_1^{k+1}(t) - y_1^k(t) & \delta y_2^k &= y_2^{k+1}(t) - y_2^k(t) \end{aligned} \tag{2.7}$$

The set of nonlinear differential equations (2.4) is then replaced by the system of linear equations, using (2.6) and (2.7):

$$\begin{aligned} \dot{y_1}^{k+1} &= y_2^k(t) + (y_2^{k+1}(t) - y_2^k(t)) \\ \dot{y_2}^{k+1} &= -\frac{(y_2^k(t))^2}{3y_1^k(t)} + \frac{1}{3}\left(\frac{y_2^k(t)}{y_1^k(t)}\right)^2 (y_1^{k+1}(t) - y_1^k(t)) - \frac{2}{3}\left(\frac{y_2^k(t)}{y_1^k(t)}\right)(y_2^{k+1}(t) - y_2^k(t)) \end{aligned} \tag{2.8}$$

This is a system linear in $y_1^{k+1}(t)$ and $y_2^{k+1}(t)$.

### 2.2 Euler's Method

For a first-order differential equation

$$\frac{d}{dt}y(t) = f(t, y(t)) \tag{2.9}$$

with the initial condition

$$y(t_0) = y_0 \tag{2.10}$$

Euler's method is the simplest of approximation techniques.

Consider an approximate solution of (2.9) over a small interval $h = \Delta t = t_{n+1} - t_n$ by writing the integral as [2]

$$\int_{t_n}^{t_{n+1}} f(t, y(t))dt \approx hf(t_n, y(t_n)) \tag{2.11}$$

to obtain

$$y(t_{n+1}) = y(t_n) + hf(t_n, y(t_n)) \tag{2.12}$$

or, in a more concise notation,

$$y_{n+1} = y_n + hf(t_n, y_n) = y_n + hf_n \tag{2.13}$$

We can integrate over any larger interval by subdividing the range into sections of width h and repeating (2.13) for each part. Equivalently we can consider that we have approximated the derivative with a forward difference

$$\frac{dy}{dt}|_n \approx \begin{cases} \frac{y_{n+1}-y_{n-1}}{2h} & centered \\ \\ \frac{y_n-y_{n-1}}{h} & backward \end{cases} \tag{2.14}$$

respectively.

How accurate is the Euler method? To quantify this we consider a Taylor expansion of y(t) around $t_n$

$$y_{n+1} = y_n + h\frac{dy}{dt} + \frac{h^2}{2}\frac{d^2y}{dt^2} + \dots \tag{2.15}$$

and substitute this into (2.13)

$$\begin{aligned} y_n + h\frac{dy}{dt}|_n + \frac{h^2}{2}\frac{d^2y}{dt^2}|_n + \dots &\approx y_n + hf(t_n, y_n) \\ &= y_n + h\frac{dy}{dt}|_n \end{aligned} \tag{2.16}$$

Hence, we see that the term in h in the expansion has been correctly reproduced by the approximation, but that the higher order terms are wrong. Therefore we describe the Euler method as 1-st order accurate. It follows that rather small increments h would be needed for high accuracy. The approximation with step size $\frac{h}{2}$ lies above the step size $h$, because the slope of the solution curves, near the true solution at $x_0 + \frac{h}{2}$, exceeds the slope at $x_0$. Thus step of size $\frac{h}{2}$ incur less error than step of size h. The same procedure is applied using $\frac{h}{3}, \frac{h}{4}, \cdots$ for higher accuracy.

We can also apply it to the approximation of differential equations of higher order. The trick is to break down the higher order differential equation into several first order differential equations.

Hence, to integrate (2.9) we iterate (2.13) and use the initial conditions from (2.10) for

$$\begin{aligned} y(t_0) &= y_0, \\ y(t_0 + h) &\simeq y(t_0) + hf(t_0, y_0(t_0)), \\ y(t_0 + 2h) &\simeq y(t_0 + h) + hf(t_0 + h, y_0(t_0 + h)), \\ y(t_0 + 3h) &\simeq y(t_0 + 2h) + hf(t_0 + 2h, y_0(t_0 + 2h)), \\ &\vdots \end{aligned} \tag{2.17}$$

**Example 1** Suppose we want to approximate the value of $y(1)$ when $y(t)$ is defined by

$$\frac{dy}{dt} = ty^{1/3}, \qquad y(0) = 1 \tag{2.18}$$

Since this equation is separable, the exact solution is known to be $y(t) = \sqrt[3]{(\frac{t^2}{3} + 1)^2}$. We can use this exact solution to compare the accuracy of the numerical approximation. It follows that rather small increments h would be needed for high accuracy.

With $t_0 = 0, y_0 = 1$, using $h = 0.1$, bring

$$\begin{aligned} y_1 &\simeq 1 + (0,1)(0,1) = 1,01 \\ y_2 &\simeq 1,01 + (0,1)(0,2)(1,01)^{1/3} = 1,03006 \\ &\vdots \end{aligned} \tag{2.19}$$

Table 2.1: Euler approximations with $h = 0.1$

| | | |
|---|---|---|
| T=0,1 | y=1,01000 | Exact Solution=1,00500 |
| T=0,2 | y=1,03006 | Exact Solution=1,02006 |
| T=0,3 | y=1,06035 | Exact Solution=1,04533 |
| T=0,4 | y=1,10113 | Exact Solution=1,08105 |
| T=0,5 | y=1,15276 | Exact Solution=1,12756 |
| T=0,6 | y=1,21567 | Exact Solution=1,18529 |
| T=0,7 | y=1,29037 | Exact Solution=1,25474 |
| T=0,8 | y=1,37746 | Exact Solution=1,33650 |
| T=0,9 | y=1,47759 | Exact Solution=1,43121 |
| T=1,0 | y=1,59148 | Exact Solution=1,53960 |

## 2.3 Analytic Continuation

A numerical approximation in the form of a Taylor series. Idea is; if the Taylor series of a function is known at a single point, then the Taylor series of that function may be found at another point. This process may be repeated until a particular value is reached.

Given a system of initial value ordinary differential equations, the method is to replace each dependent variable present by a Taylor series centered at a certain origin. The coefficients in each Taylor series are regarded as unknown quantities. The ordinary differential equations are used to obtain a set of recurrence relations from which the unknown coefficients may be calculated.

Thus, a formal power series solution may be determined to an initial value problem and the series will be convergent in some region about the origin. Then, the truncated power series are evaluated at some point within the region of convergence. At this new point, initial values for the system are obtained from the already obtained Taylor series. Using these initial values, the recurrence relations then yield a second series solution valid in a region about the new origin.

This procedure can be iterated and the solution at a given point may be determined via a sequence of Taylor series. This algorithm is a numerical version of the process of analytic continuation.

**Example 2** Suppose we have the system of ordinary differential equations

$$\begin{aligned} y' &= y^2 + z, \qquad y(0) = 1 \\ z' &= z^2, \qquad\quad z(0) = 1 \end{aligned} \tag{2.20}$$

This system can be rewritten as the differential/algebraic system

$$\begin{aligned} a &= y^2, \qquad b = a + z, \qquad c = z^2 \\ y' &= b, \qquad z' = c \end{aligned} \tag{2.21}$$

with $b = 2$, $a = c = y = z = 1$ when $t = 0$. If we define the Taylor series coefficients $\{a_k^{(j)}, b_k^{(j)}, c_k^{(j)}, y_k^{(j)}, z_k^{(j)}\}$ by the expansions

$$\begin{aligned} a(t) &= \sum_{k=0}^{\infty} a_k^{(j)} (t - t_j)^k, \qquad b(t) = \sum_{k=0}^{\infty} b_k^{(j)} (t - t_j)^k, \\ c(t) &= \sum_{k=0}^{\infty} c_k^{(j)} (t - t_j)^k, \qquad y(t) = \sum_{k=0}^{\infty} y_k^{(j)} (t - t_j)^k, \\ z(t) &= \sum_{k=0}^{\infty} z_k^{(j)} (t - t_j)^k, \end{aligned} \tag{2.22}$$

then, using (2.22) in (2.21), the following recurrence relations can be obtained

$$\begin{aligned} a_k^{(j)} &= \sum_{n=0}^{k} y_n^{(j)} y_{k-n}^{(j)}, \qquad b_k^{(j)} = a_k^{(j)} + z_k^{(j)}, \\ c_k^{(j)} &= \sum_{n=0}^{k} z_n^{(j)} z_{k-n}^{(j)}, \qquad y_k^{(j)} = b_k^{(j)}/(k+1), \\ z_k^{(j)} &= c_k^{(j)}/(k+1), \end{aligned} \tag{2.23}$$

The initial conditions give the starting values: $\{j = 0, t_0 = 0, a_0^{(0)} = c_0^{(0)} = y_0^{(0)} = z_0^{(0)} = 1, b_0^{(0)} = 2\}$. To determine the Taylor series about the point $t_0 = 0$, equation (2.23) is iterated for $k = 1, 2, \ldots, M$.

Then a new point $t_1$ is chosen. A Taylor series for each of a,b,c,y and z is then found about this new point by taking $j = 1$ and determining the initial conditions from.

$$a_0^{(1)} = \sum_{k=0}^{M} a_k^{(0)}(t_1 - t_0)^k, \qquad b_0^{(1)} = \sum_{k=0}^{M} b_k^{(0)}(t_1 - t_0)^k, \ldots \tag{2.24}$$

The recurrence relations in equation (2.23) are then iterated again. This process can be repeated indefinitely.

## 2.4 Runge-Kutta Methods

The differential equation

$$\frac{dy}{dx} = f(x, y) \tag{2.25}$$

determines a family of curves (the "characteristics") which do not intersect each other and of which one passes through every point in the plane. [3]

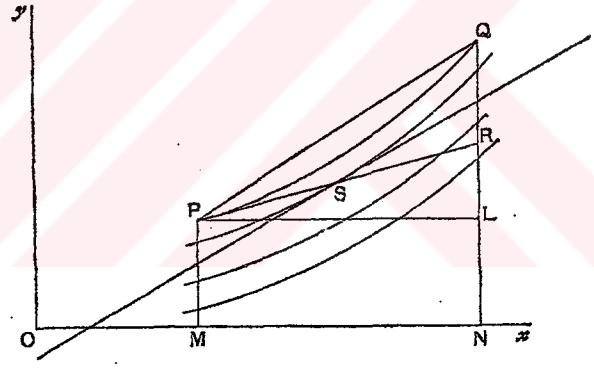


Figure 2.1: Family of curves

Given a point P(a,b), we know that the gradient of the characteristic through P is f(a,b), and we want to determine the $y = NQ$ of any other point on the same characteristic, given that $x = ON = a + h$. A first approximation is given by taking the tangent PR instead of the characteristic PQ, i.e. taking

$$y = NL + LR = NL + PL \tan \angle RPL = b + hf(a, b) = b + hf_0 \tag{2.26}$$

But unless h is very small indeed, the error RQ is far from negligible. A more reasonable approximation is to take the chord PQ as parallel to the tangent to the characteristic through S, the middle point of PR.

Since S is $(a+\frac{1}{2}h, b+\frac{1}{2}hf_0)$, this gives

$$y = NL + LQ = NL + PL\tan\angle QPL = b + hf(a+\frac{1}{2}h, b+\frac{1}{2}hf_0) \quad (2.27)$$

Suppose that the function of y defined by

$$\frac{dy}{dx} = f(x,y), \qquad y = b \qquad when \qquad x = a, \quad (2.28)$$

is denoted by y=F(x).

If this can be expanded by Taylor's theorem,

$$F(a+h) = F(a) + hF'(a) + \frac{h^2}{2!}F''(a) + \frac{h^3}{3!}F'''(a) + \ldots \quad (2.29)$$

Now

$$F'(x) = \frac{dy}{dx} = f(x,y) = f \quad (2.30)$$

We shall now take the total differential coefficient with respect to x (that is, taking the y in f to vary in consequence of the variation of x). Let us denote partial differential coefficients by

$$p = \frac{\partial f}{\partial x}, \qquad q = \frac{\partial f}{\partial y}, \qquad r = \frac{\partial^2 f}{\partial x^2}, \qquad s = \frac{\partial^2 f}{\partial x \partial y}, \qquad t = \frac{\partial^2 f}{\partial y^2}; \quad (2.31)$$

and their values when $x = a$ and $y = b$ by $p_0, q_0$,etc. Then,

$$F''(x) = \frac{df}{dx} = \left(\frac{\partial}{\partial x} + \frac{dy}{dx}\frac{\partial}{\partial y}\right) f = p + fq. \quad (2.32)$$

Similarly

$$\begin{aligned} F'''(x) &= \left(\frac{\partial}{\partial x} + \frac{dy}{dx}\frac{\partial}{\partial y}\right)(p+fq) \\ &= r + pq + fs + f(s + q^2 + ft) \end{aligned} \quad (2.33)$$

Thus

$$F(a+h)-F(a) = hf_0+\frac{1}{2}h^2(p_0+f_0q_0)+\frac{1}{6}h^3(r_0+2f_0s_0+f_0^2t_0+p_0q_0 + f_0q_0^2)+\ldots \tag{2.34}$$

The second approximation of (2.27), i.e.

$$y-b=hf(a+\frac{1}{2}h,b+\frac{1}{2}hf_0)=k_1 \tag{2.35}$$

may now be expanded and compared with (2.34).

Now, by Taylor's theorem for two independent variables,

$$f(a+\frac{1}{2}h,b+\frac{1}{2}hf_0) = f_0+\frac{1}{2}hp_0+\frac{1}{2}hf_0q_0+\frac{1}{2!}\left(\frac{1}{4}h^2r_0+\frac{1}{2}h^2f_0s_0+\frac{1}{4}h^2f_0^2t_0\right) + \ldots \tag{2.36}$$

giving

$$k_1=hf_0+\frac{1}{2}h^2(p_0+f_0q_0)+\frac{1}{8}h^3(r_0+2f_0s_0+f_0^2t_0)+\ldots \tag{2.37}$$

It is obvious that $k_1$ is at fault in the coefficient of $h^3$.

Our next step is suggested by the usual methods for the numerical integration of the simpler differential equation

$$\frac{dy}{dx}=f(x,y). \tag{2.38}$$

Now the next approximation discussed is generally Simpson's Rule, which may be written

$$y-b=\frac{1}{6}h\{f(a)+4f(a+\frac{1}{2}h)+f(a+h)\} \tag{2.39}$$

If we expand the corresponding formula in two variables, namely

$$\frac{1}{6}h\{f_0 + 4f(a + \frac{1}{2}h, b + \frac{1}{2}hf_0) + f(a + h, b + hf_0)\}, \tag{2.40}$$

we easily obtain

$$hf_0 + \frac{1}{2}h^2(p_0 + f_0q_0) + \frac{1}{6}h^3(r_0 + 2f_0s_0 + f_0^2t_0) + \ldots, \tag{2.41}$$

which is a better approximation than $k_1$, but even now has not the coefficient of $h^3$ quite in agreement with (2.34).

To obtain the extra terms in $h^3$, Runge replaces

$$hf(a + h, b + hf_0) \tag{2.42}$$

by $k''' = hf(a+h, b+k'')$, where $k'' = hf(a+h, b+hf_0)$. The modified formula may be briefly written $\frac{1}{6}\{k' + 4k_1 + k'''\}$, where $k' = hf_0$, or $\frac{2}{3}k_1 + \frac{1}{3}k_2 = k_1 + \frac{1}{3}(k_2 - k_1)$, where $k_2 = \frac{1}{2}(k' + k''')$.

Of course this method will give bad results if the series (2.34) converges slowly.

If $f_0 > 1$ numerically, we rewrite our equation

$$\frac{dx}{dy} = \frac{1}{f(x,y)} = F(x,y), \tag{2.43}$$

and now $F_0 < 1$ numerically, and we take y as the independent variable.

To avoid confusion, the calculations should be formed in some definite order, such as the following:

Calculate successively

$$\begin{aligned} k' &= hf_0, \\ k'' &= hf(a+h, b+k'), \\ k''' &= hf(a+h, b+k''), \\ k_1 &= hf(a+\frac{1}{2}h, b+\frac{1}{2}k'), \\ k_2 &= \frac{1}{2}(k'+k'''), \\ k &= k_1 + \frac{1}{3}(k_2 - k_1). \end{aligned} \tag{2.44}$$

Moreover, as $k_1$ is itself an approximation to the value required, it is clear that if the difference between k and $k_1$, namely $\frac{1}{3}(k_2 - k_1)$, is small compared with $k_1$ and k, the error in k is likely to be even smaller.

**Example 3** $\frac{dy}{dx} = \frac{y-x}{y+x}$ ; given that $y = 1$ when $x = 0$, find y when $x = 1$.

Divide the range into three parts, 0 to 0.2, 0.2 to 0.5, 0.5 to 1. We take a small increment for the first step because $f(x, y)$ is largest at the beginning.

*First step*

$$a = 0 \qquad b = 1 \qquad h = 0,2 \qquad f_0 = 1$$

$$\begin{aligned} k' &= hf_0 = 0,200, \\ k'' &= hf(a+h, b+k') = 0.143, \\ k''' &= hf(a+h, b+k'') = 0.140, \\ k_1 &= hf(a+\frac{1}{2}h, b+\frac{1}{2}k') = 0.167, \\ k_2 &= \frac{1}{2}(k'+k''') = 0.170, \\ k &= k_1 + \frac{1}{3}(k_2 - k_1) = 0.168. \end{aligned} \tag{2.45}$$

giving $y = 1,168$ when $x = 0.2$.

*Second step*

$$a = 0.2 \qquad b = 1.168 \qquad h = 0,3 \qquad f_0 = f(0.2, 1.168) = 0.708 \tag{2.46}$$

Proceeding as before we get $k_1 = 0.170$, $k_2 = 0.173$ and so $k = 0.171$,

giving $y = 1.168 + 0.171 = 1.339$ when $x = 0.5$

*Third step*

$$a = 0.5 \qquad b = 1.339 \qquad h = 0,5 \tag{2.47}$$

We find $k_1 = k_2 = k = 0.160$, giving $y = 1.499$ when $x = 1$.

Considering the k and $k_1$, the error in this result should be less than 0.001 on each of the first and second steps and negligible on the third, that is, less than 0.002 altogether.

As a matter of fact, the true value of y is between 1.498 and 1.499, so the error is less than 0.001. This true value of y is found by integrating the equation, leading to

$$\pi - 2\tan^{-1}\frac{y}{x} = \log_e(x^2 + y^2). \tag{2.48}$$

The method is easily extended to simultaneous equations.

**Example 4**

$$\begin{aligned} \frac{dy}{dx} &= 2z - \frac{y}{x} = f(x, y, z), \\ \frac{dz}{dx} &= \frac{y}{\sqrt{(1 - y^2)}} = g(x, y, z), \end{aligned} \tag{2.49}$$

given that $y = 0.2027$ and $z = 1.0202$ when $x = 0.2$, find y and z when $x = 0.4$. Here,

$$a = 0.2, b = 0.2027, c = 1.0202, f_0 = f(0.2, 0.2027, 1.0202) = 1.027,$$
$$g_0 = 0.2070, h = 0.2; \tag{2.50}$$

$$
\begin{aligned}
k' &= hf_0 = 0.2 * 1.027 = 0,2054; \\
l' &= hg_0 = 0.2 * 0.2070 = 0,0414; \\
k'' &= hf(a+h, b+k', c+l') = 0.2 * f(0.4, 0.4081, 1.0616) = 0.2206; \\
l'' &= hg(a+h, b+k', c+l') = 0.2 * g(0.4, 0.4081, 1.0616) = 0.0894; \\
k''' &= hf(a+h, b+k'', c+l'') = 0.2 * f(0.4, 0.4233, 1.1096) = 0.2322; \\
l''' &= hg(a+h, b+k'', c+l'') = 0.2 * g(0.4, 0.4233, 1.1096) = 0.0934; \\
k_1 &= hf(a+\frac{1}{2}h, b+\frac{1}{2}k', c+\frac{1}{2}l') = 0.2 * f(0.3, 0.3054, 1.0409) = 0.2128; \\
l_1 &= hg(a+\frac{1}{2}h, b+\frac{1}{2}k', c+\frac{1}{2}l') = 0.2 * g(0.3, 0.3054, 1.0409) = 0.0641; \\
k_2 &= \frac{1}{2}(k' + k''') = 0.2188; \\
l_2 &= \frac{1}{2}(l' + l''') = 0.0674; \\
k &= k_1 + \frac{1}{3}(k_2 - k_1) = 0.2128 + 0.0020 = 0.2148; \\
l &= l_1 + \frac{1}{3}(l_2 - l_1) = 0.0641 + 0.0011 = 0.0652.
\end{aligned}
\tag{2.51}
$$

giving

$$
\begin{aligned}
y &= 0.2027 + 0.2148 = 0.4175 \\
z &= 1.0202 + 0.0652 = 1.0854
\end{aligned}
\tag{2.52}
$$

probably correct to the third place of decimals.

It must be better to be given the best known Runge-Kutta formula. The result yielded these two formulas will be the same. So it is given by

$$
\begin{aligned}
k_1 &= hf(x_0, y_0), \\
k_2 &= hf(x_0 + \frac{1}{2}h, y_0 + \frac{1}{2}k_1), \\
k_3 &= hf(x_0 + \frac{1}{2}h, y_0 + \frac{1}{2}k_2), \\
k_4 &= hf(x_0 + h, y_0 + k_3).
\end{aligned}
\tag{2.53}
$$

where

$$
y(x_0 + h) = y(x_0) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{2.54}
$$

What is the advantage of Runge-Kutta formulas over the Taylor method?

Though approximately the same as the Taylor polynomial of degree four, these formulas do not require prior calculation of the higher derivatives of $y(x)$, as the Taylor method does. Since the differential equations arising in applications are often complicated, the calculation of derivatives can be onerous. The Runge-Kutta formulas involve computation of $f(x, y)$ at various positions instead and this function occurs in the given equation. The method is very extensively used.

### 2.5 Predictor-Corrector Methods

To integrate an ordinary differential equation from a point $x_n$ to a new point $x_{n+1} = x_n + h$, a single formula may be used to predict $y_{n+1}$. Alternately, the value of $y_{n+1}$ could be predicted by one formula, and then that value could be refined by an iterative formula (the "corrector").

For the first order ordinary differential equation $y' = f(x, y)$, suppose that the values of x and y are known at the sequence of $m + 1$ points $\{x_{n-m}, \ldots, x_{n-1}, x_n\}$. Then the values of $y'$ are known at those same points (since $y'$ is determined from x and y via $y' = f(x, y)$ ). A polynomial of degree m can be fitted to $m + 1$ values of x and $y'$. This polynomial can be used to predict the value of $y'$ in the interval $(x_n, x_{n+1})$. This, in turn, can be used to predict the value of $y_{n+1}$ by a numerical approximation of the relation

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} y'(x)dx \tag{2.55}$$

Such a formula is called an "predictor".

A modification of this step can be repeated. The values of x and $y'$ are now known at the $m + 1$ points $\{x_{n-m+1}, \ldots, x_n, x_{n+1}\}$. A polynomial can be fit through these points, and then the quantity in equation (2.55) can be re-computed. This formula, which furnishes a new estimate of $y_{n+1}$, is called a "corrector". The corrector may be used repeatedly.

A simple predictor-corrector pair is

$$\begin{aligned} y_{k+1} &\simeq y_k + hy'_k \\ y_{k+1} &\simeq y_k + \frac{1}{2}h(y'_k + y'_{k+1}) \end{aligned} \tag{2.56}$$

the predictor being Euler's formula and the corrector being known as the modified Euler formula.(The modified Euler and second-order Runge-Kutta methods are identical.[2]) Since $y'_k = f(x_k, y_k)$ and $y'_{k+1} = f(x_{k+1}, y_{k+1})$ the predictor first estimates $y_{k+1}$. This estimate then leads to a $y'_{k+1}$ value and then to a corrected $y_{k+1}$. Further corrections of $y'_{k+1}$ and $y_{k+1}$ successively can be made until a satisfactory result is achieved.

One set of predictor-corrector equations is the Adams-Bashforth predictor formula [4]

$$y_{n+1} = y_n + \frac{h}{24}(55y'_n - 59y'_{n-1} + 37y'_{n-2} - 9y'_{n-3}) \tag{2.57}$$

and the Adams-Moulton corrector formula

$$y_{n+1} = y_n + \frac{h}{24}(9y'_{n+1} + 19y'_n - 5y'_{n-1} + y'_{n-2}) \tag{2.58}$$

where h is the difference between adjacent x points (The x points are assumed to be equally spaced). These equations are fourth order accurate.

**Example 5** Apply modified Euler formula to the problem

$$y' = xy^{\frac{1}{3}}, \qquad y(1) = 1. \tag{2.59}$$

Assuming $y_k$ and $y'_k$ already in hand, the two equations

$$y_{k+1} \simeq y_k + \frac{1}{2}h(y'_k + y'_{k+1}) \qquad y'_{k+1} = f(x_{k+1}, y_{k+1}) \tag{2.60}$$

are used to determine $y_{k+1}$ and $y'_{k+1}$. Applied successively, beginning with $k = 0$, this algorithm generates sequences of values $y_k$ and $y'_k$. The simpler Euler formula will be used as a predictor. It provides a first estimate of $y_{k+1}$. Here, with $x_0 = 1$ and $h = 0.05$ it offers

$$y(1.05) \simeq 1 + (0.05)(1) = 1.05 \tag{2.61}$$

The differential equation then presents us with

$$y'(1.05) \simeq (1.05)(1.016) \simeq 1.0661 \tag{2.62}$$

Now the modified Euler formula serves as a corrector, yielding

$$y(1.05) \simeq 1 + (0.025)(1 + 1.0661) \simeq 1.05165 \tag{2.63}$$

With this new value the differential equation corrects $y'(1.05)$ to 1.0678, after which the corrector is reapplied and produces

$$y(1.05) \simeq 1 + (0.025)(1 + 1.0678) \simeq 1.0517 \tag{2.64}$$

**Example 6** Apply the Adams method to $y' = -xy^2$ with $y(0) = 2$ using $h = 0.2$.

Each step involving a prediction and then an iterative use of the corrector formula.

Table 2.2: Comparing the values of y and errors with h=0.2

| x | y (correct) | y (predicted) | Error | y (corrected) | Error |
|---|---|---|---|---|---|
| 0 | 2.000000 | | | | |
| 1 | 1.000000 | 1.000798 | -789 | 1.000133 | -133 |
| 2 | 0.400000 | 0.400203 | -203 | 0.400158 | -158 |
| 3 | 0.200000 | 0.200140 | -140 | 0.200028 | -28 |
| 4 | 0.117647 | 0.117679 | -32 | 0.117653 | -6 |
| 5 | 0.076923 | 0.076933 | -10 | 0.076925 | -2 |
| 6 | 0.054054 | 0.054058 | -4 | 0.054055 | -1 |
| 7 | 0.040000 | 0.040002 | -2 | 0.040000 | |
| 8 | 0.030769 | 0.030770 | -1 | 0.030769 | |
| 9 | 0.024390 | 0.024391 | -1 | 0.024390 | |
| 10 | 0.019802 | 0.019802 | | 0.019802 | |

## 2.6 Invariant Imbedding

Invariant imbedding is a type of continuation method. For the usual problems that are treated,the length of the interval of interest is considered to be the continuation parameter. Hence the end point in a two point boundary value problem is treated as a variable. By differentiating with respect to this variable, an initial value problem can be created.

We begin our study by treating simple linear systems and illustrating the techniques and results by an example. [5]

Suppose we have the system of ordinary differential equations

$$\begin{aligned} \dot{x}(t) &= a(t)x(t) + b(t)y(t) \\ \alpha_1 x(0) + \alpha_2 y(0) &= 0 \end{aligned} \tag{2.65}$$

$$\dot{y}(t) = c(t)x(t) + d(t)y(t) + f(t)$$
$$\alpha_3 x(T) + \alpha_4 y(T) = 1 \tag{2.66}$$

where a,b,c,d,f are continuous functions on $0 \leq t \leq T$, and $\{\alpha_i\}$ are constants, $i = 1, 2, 3, 4$.

Our goal is to derive an initial value, rather than two point boundary value problem which "represents" the solution to (2.65)-(2.66). By "represents", we mean that the solution of the initial value (or Cauchy) system uniquely determines the solution to the boundary value problem, and conversely.

The imbedding parameter we shall use to derive a Cauchy system is the interval length T. Consequently, we rewrite (2.65)-(2.66) to explicitly indicate the dependence of the solution upon T:

$$\dot{x}(t,T) = a(t)x(t,T) + b(t)y(t,T) \tag{2.67}$$
$$\alpha_1 x(0,T) + \alpha_2 y(0,T) = 0$$

$$\dot{y}(t,T) = c(t)x(t,T) + d(t)y(t,T) + f(t) \tag{2.68}$$
$$\alpha_3 x(T,T) + \alpha_4 y(T,T) = 1 \qquad 0 \leq t \leq T$$

We make use of linearity and consider the two systems:

SYSTEM I

$$\dot{u}(t,T) = a(t)u + b(t)v \tag{2.69}$$
$$\alpha_1 u(0,T) + \alpha_2 v(0,T) = 0$$
$$\dot{v}(t,T) = c(t)u + d(t)v + f(t) \tag{2.70}$$
$$\alpha_3 u(T,T) + \alpha_4 v(T,T) = 0, 0 \leq t \leq T$$

and

SYSTEM II

$$\dot{p}(t,T) = a(t)p + b(t)q$$
$$\alpha_1 p(0,T) + \alpha_2 q(0,T) = 0 \tag{2.71}$$
$$\dot{q}(t,T) = c(t)p + d(t)q$$
$$\alpha_3 p(T,T) + \alpha_4 q(T,T) = 1, 0 \leq t \leq T \tag{2.72}$$

The superposition principle for linear systems then allows us to write

$$x(t,T) = u(t,T) + p(t,T) \tag{2.73}$$
$$y(t,T) = v(t,T) + q(t,T), \qquad 0 \le t \le T \tag{2.74}$$

Let us first consider the functions u and v of system I. We wish to examine how the solution curves change at a fixed point t, $0 \le t \le T$ , as the interval length T is changed. Differentiating (2.69)-(2.70) with respect to T gives

$$\dot{u}_T(t,T) = a(t)u_T(t,T) + b(t)v_T(t,T) \tag{2.75}$$

$$\alpha_1 u_T(0,T) + \alpha_2 v_T(0,T) = 0$$

$$\dot{v}_T(t,T) = c(t)u_T(t,T) + d(t)v_T(t,T) \tag{2.76}$$

$$\alpha_3[\dot{u}(T,T) + u_T(T,T)] + \alpha_4[\dot{v}(T,T) + v_T(T,T)] = 0$$

Here a dot represents differentiation with respect to t, $()_T$ differentiation with respect to T. Comparing (2.75)-(2.76) with (2.71)-(2.72) we see that

$$u_T(t,T) = -[\alpha_3\dot{u}(T,T) + \alpha_4\dot{v}(T,T)]p(t,T) \tag{2.77}$$
$$v_T(t,T) = -[\alpha_3\dot{u}(T,T) + \alpha_4\dot{v}(T,T)]q(t,T), \qquad 0 \le t \le T \tag{2.78}$$

We now consider the bracketed term in equations (2.77)-(2.78). From equations (2.69)-(2.70) with t=T, we have

$$\dot{u}(T,T) = a(T)u(T,T) + b(T)v(T,T) \tag{2.79}$$
$$\dot{v}(T,T) = c(T)u(T,T) + d(T)v(T,T) + f(T) \tag{2.80}$$

Introduce the new variables m and n by

$$m(T) = u(T,T) \tag{2.81}$$
$$n(T) = v(T,T), \qquad T \ge 0$$

In view of (2.79)-(2.80), it suffices to determine the functions m and n. Differentiate equation (2.77) to obtain

$$\begin{aligned} m'(T) &= \dot{u}(T,T) + u_T(T,T) \\ &= a(T)m(T) + b(T)n(T) - \{\alpha_3[a(T)m(T) + b(T)n(T)] \\ &\quad +\alpha_4[c(T)m(T) + d(T)n(T) + f(T)]\}p(T,T) \end{aligned} \tag{2.82}$$

Similarly, for n we have

$$\begin{aligned} n'(T) &= \dot{v}(T,T) + v_T(T,T) \\ &= c(T)m(T) + d(T)n(T) - \{\alpha_3[a(T)m(T) + b(T)n(T)] \\ &\quad +\alpha_4[c(T)m(T) + d(T)n(T) + f(T)]\}q(T,T) \end{aligned} \tag{2.83}$$

Equations (2.82) and (2.83) show that we must consider the quantities p(T,T) and q(T,T).

Differentiate equations (2.71) and (2.72) with respect to T. This yields

$$\dot{p}_T(t,T) = a(t)p_T + b(t)q_T \tag{2.84}$$

$$\alpha_1 p_T(0,T) + \alpha_2 q_T(0,T) = 0$$

$$\dot{q}_T(t,T) = c(t)p_T + d(t)q_T \tag{2.85}$$

$$\alpha_3[\dot{p}(T,T) + p_T(T,T)] + \alpha_4[\dot{q}(T,T) + q_T(T,T)] = 0,$$

Comparing (2.71)-(2.72) with (2.84)-(2.85) shows that

$$p_T(t,T) = -[\alpha_3\dot{p}(T,T) + \alpha_4\dot{q}(T,T)]p(t,T) \tag{2.86}$$

and

$$q_T(t,T) = -[\alpha_3\dot{p}(T,T) + \alpha_4\dot{q}(T,T)]q(t,T) \tag{2.87}$$

To make use of these relations, we observe that from (2.71) and (2.72) with t=T

$$\begin{aligned} \dot{p}(T,T) &= a(T)p(T,T) + b(T)q(T,T) \\ \dot{q}(T,T) &= c(T)p(T,T) + d(T)q(T,T) \end{aligned} \tag{2.88}$$

Let the functions r and s be given by

$$r(T) = p(T,T) \tag{2.89}$$

$$s(T) = q(T,T) \tag{2.90}$$

We now derive a Cauchy system satisfied by r and s.

$$\begin{aligned} r'(T) &= \dot{p}(T,T) + p_T(T,T) \qquad (2.91) \\ &= a(T)r(T) + b(T)s(T) - \{\alpha_3[a(T)r + b(T)s] + \alpha_4[c(T)r + d(T)s]\}r \end{aligned}$$

$$s'(T) = c(T)r + d(T)s - \{\alpha_3[a(T)r + b(T)s] + \alpha_4[c(T)r + d(T)s]\}s \qquad (2.92)$$

Combining terms in (2.91)-(2.92) gives

$$\begin{aligned} r'(T) &= b(T)s + r[a(T) - \alpha_3 b(T)s - \alpha_4 d(T)s] - r^2[\alpha_3 a(T) + \alpha_4 c(T)] \qquad (2.93) \\ s'(T) &= c(T)r + s[d(T) - \alpha_3 a(T)r - \alpha_4 c(T)r] - s^2[\alpha_3 b(T) + \alpha_4 d(T)], \end{aligned}$$

The initial conditions at T=0 are obtained from equations (2.71) and (2.72) by solving the system

$$\begin{aligned} \alpha_1 r(0) + \alpha_2 s(0) &= 0 \\ \alpha_3 r(0) + \alpha_4 s(0) &= 1 \end{aligned} \qquad (2.94)$$

Obviously, this forces the compatibility condition $\alpha_1\alpha_4 - \alpha_2\alpha_3 \neq 0$ to insure a unique solution exists.
Knowledge of r and s allows us to obtain p and q from the initial value system (2.86) and (2.87). The equations are

$$\begin{aligned} p_T(t,T) &= -\{r[\alpha_3 a(T) + \alpha_4 c(T)] + s[\alpha_3 b(T) + \alpha_4 d(T)]\}p(t,T) \qquad (2.95) \\ q_T(t,T) &= -\{r[\alpha_3 a(T) + \alpha_4 c(T)] + s[\alpha_3 b(T) + \alpha_4 d(T)]\}q(t,T), \quad 0 \leq t \leq T \end{aligned}$$

The initial conditions at T=t are

$$\begin{aligned} p(t,t) &= r(t) \qquad (2.96) \\ q(t,t) &= s(t) \end{aligned}$$

Returning now to equations (2.82) and (2.83) for the functions m and n, we see that

$$\begin{aligned} m'(T) &= a(T)m + b(T)n - \{m[\alpha_3 a(T) + \alpha_4 c(T)] + n[\alpha_3 b(T) \\ &\quad + \alpha_4 d(T)] + f(T)\}r(T) \qquad (2.97) \end{aligned}$$

$$n'(T) = c(T)m + d(T)n + f(T) - \{m[\alpha_3 a(T) + \alpha_4 c(T)] + n[\alpha_3 b(T) + \alpha_4 d(T)] + f(T)\}s(T), \quad (2.98)$$

The initial conditions at T=0 are given by equations (2.69) and (2.70) as

$$\begin{aligned} m(0) &= 0 \\ n(0) &= 0 \end{aligned} \quad (2.99)$$

The equations for u and v are determined in terms of m and n by

$$u_T(t,T) = -\{m[\alpha_3 a(T) + \alpha_4 c(T)] + n[\alpha_3 b(T) + \alpha_4 d(T)] + f(T)\}p(t,T)$$
$$v_T(t,T) = -\{m[\alpha_3 a(T) + \alpha_4 c(T)] + n[\alpha_3 b(T) + \alpha_4 d(T)] + f(T)\}q(t,T) \quad (2.100)$$

At T=t, we have

$$u(t,t) = m(t) \quad (2.101)$$

$$v(t,t) = n(t) \quad (2.102)$$

This completes our derivation of the complete Cauchy system for determining the functions u,v,p and q necessary to obtain x and y, the solutions to (2.65) and (2.66).

Since the derivation was lengthy, we now summarize and collect in one place the relevant equations. The complete Cauchy system consists of the following equations for the functions r,s,m,n,p,q,u and v:

$$r'(T) = b(T)s + r[a(T) - \alpha_3 b(T)s - \alpha_4 d(T)s] - r^2[\alpha_3 a(T) + \alpha_4 c(T)]$$
$$s'(T) = c(T)r + s[d(T) - \alpha_3 a(T)r - \alpha_4 c(T)r] - s^2[\alpha_3 b(T) + \alpha_4 d(T)]$$

$$m'(T) = a(T)m + b(T)n - \{m[\alpha_3 a(T) + \alpha_4 c(T)] + n[\alpha_3 b(T) + \alpha_4 d(T)] + f(T)\}r(T)$$
$$n'(T) = c(T)m + d(T)n + f(T) - \{m[\alpha_3 a(T) + \alpha_4 c(T)] + n[\alpha_3 b(T) + \alpha_4 d(T)] + f(T)\}s(T) \quad (2.103)$$

$$p_T(t,T) = -\{r[\alpha_3 a(T) + \alpha_4 c(T)] + s[\alpha_3 b(T) + \alpha_4 d(T)]\}p(t,T)$$
$$q_T(t,T) = -\{r[\alpha_3 a(T) + \alpha_4 c(T)] + s[\alpha_3 b(T) + \alpha_4 d(T)]\}q(t,T)$$

$$u_T(t,T) = -\{m[\alpha_3 a(T) + \alpha_4 c(T)] + n[\alpha_3 b(T) + \alpha_4 d(T)] + f(T)\}p(t,T)$$
$$v_T(t,T) = -\{m[\alpha_3 a(T) + \alpha_4 c(T)] + n[\alpha_3 b(T) + \alpha_4 d(T)] + f(T)\}q(t,T)$$

The initial conditions are given by

$$\begin{aligned} \alpha_1 r(0) + \alpha_2 s(0) &= 0 \\ \alpha_3 r(0) + \alpha_4 s(0) &= 1 \\ m(0) &= 0 \\ n(0) &= 0 \\ p(t,t) &= r(t) \\ q(t,t) &= s(t) \\ u(t,t) &= m(t) \\ v(t,t) &= n(t) \end{aligned} \tag{2.104}$$

The solution curves for the original system are then given by

$$\begin{aligned} x(t,T) &= u(t,T) + p(t,T) \\ y(t,T) &= v(t,T) + q(t,T) \end{aligned} \tag{2.105}$$

Suppose the solution is desired at a set of abscissas $0 \leq t_1 < t_2 < t_3 < \cdots < t_N \leq T^*$, where $T^*$ is the interval length of interest. The solution procedure is to integrate the equations for r,s,m and n from $T = 0$ to $T = t_1$. At this point equation for the functions $p(t_1, T), q(t_1, T), u(t_1, T), v(t_1, T)$ are adjoined with the initial conditions given by (2.89),(2.90),(2.101) and (2.102). The entire system is then integrated from $T = t_1$ to $T = t_2$, at which point additional equation for the functions $p(t_2, T), q(t_2, T), u(t_2, T), v(t_2, T)$ are adjoined with the appropriate initial conditions. This procedure is carried out for each $t_i$ and the integration continues until $T = T^*$. At this point, the desired solution values are obtained from the functions $p(t_i, T^*), q(t_i, T^*), u(t_i, T^*), v(t_i, T^*)$ by means of the relations (2.73) and (2.74).

**Example 7** Suppose we want to turn the boundary value problem

$$\begin{aligned} \frac{dx}{dt} &= 10y, & x(0) &= 0 \\ \frac{dy}{dt} &= 10x, & y(2) &= 1 \end{aligned} \tag{2.106}$$

into an initial value problem. Using the above notation, we find that $\alpha_1 = \alpha_4 = 1, \alpha_2 = \alpha_3 = 0, a(t) = d(t) = f(t) = 0, b(t), c(t) = 10, T^* = 2$. Notice that since $f \equiv 0$ only the functions p and q are of interest. The Cauchy system

then becomes

$$\begin{aligned} r'(T) &= 10 - 10r^2, \qquad r(0) = 0 \\ p_T(t,T) &= -10r(T)p(t,T), \qquad p(t,t) = r(t), \\ q_T(t,T) &= -10r(T)q(t,T), \qquad q(t,t) = 1, \end{aligned} \tag{2.107}$$

The nature of the boundary conditions obviously implies $s(T) \equiv 1$. For this simple example, this can be seen from the closed form solutions

$$\begin{aligned} r(T) &= \tanh 10T, \qquad T \geq 0, \\ p(t,T) &= \frac{\sinh 10t}{cosh 10T} = x(t,T), \\ q(t,T) &= \frac{\cosh 10t}{cosh 10T} = y(t,T), \qquad 0 \leq t \leq T \end{aligned} \tag{2.108}$$

The numerical solution of **nonlinear problems** can be a challenging proposition. The invariant imbedding can be applied directly to the nonlinear problem. But the resulting system of equations are nonlinear partial differential equations. Frequently, these equations appear to be at least as difficult to handle as the original problem.(To circumvent this difficulty, we combine the technique of quasilinearization.)

Now consider the system of nonlinear ordinary differential equations

$$\dot{u} = F(u,v,t) \tag{2.109}$$

$$\dot{v} = G(u,v,t), \qquad 0 < t < T \tag{2.110}$$

subject to the boundary conditions

$$u(0) = 0 \tag{2.111}$$

$$v(T) = c \tag{2.112}$$

To indicate the dependence of the functions u and v upon c and T, as well as upon t, we shall write u(t,c,T) and v(t,c,T).

By differentiating equations (2.109)-(2.110) with respect to c, it is seen that

$$\begin{aligned} \dot{u}_c(t,c,T) &= F_u u_c + F_v v_c \\ \dot{v}_c(t,c,T) &= G_u u_c + G_v v_c \\ u_c(0,c,T) &= 0 \\ v_c(T,c,T) &= 1 \end{aligned} \tag{2.113}$$

Similarly, a differentiation in T yields

$$\begin{aligned} \dot{u}_T &= F_u u_T + F_v v_T \\ \dot{v}_T &= G_u u_T + G_v v_T \qquad 0 < t < T \\ u_T(0,c,T) &= 0 \\ \dot{v}(T,c,T) + v_3(T,c,T) &= 0 \end{aligned} \tag{2.114}$$

In the above equation, $\dot{v}$ is the derivative of v with respect to the first argument, while $v_3$ is the derivative with respect to the third argument.

To make use of these equations, note that from the differential equation (2.110), when $t = T$ we have

$$\begin{aligned} \dot{v}(T,c,T) &= G(u(T,c,T), v(T,c,T), T) \\ &= G(r(c,T), c, T) \end{aligned} \tag{2.115}$$

where the notation

$$r(c,T) = u(T,c,T) \tag{2.116}$$

has been introduced.

Comparing equations (2.113) with (2.114), it follows that

$$u_T(t,c,T) = -G(r(c,T),c,T)u_c(t,c,T) \tag{2.117}$$

$$v_T(t,c,T) = -G(r(c,T),c,T)v_c(t,c,T) \qquad 0 \le t \le T \tag{2.118}$$

Equations (2.117) and (2.118) are partial differential equations for u and v. The initial conditions at T=t are

$$\begin{aligned} u(t,c,t) &= r(c,t) \\ v(t,c,t) &= c \end{aligned} \tag{2.119}$$

It remains to consider the function r. Differentiate equation (2.116) with respect to T to obtain

$$r_T(c,T) = \dot{u}(T,c,T) + u_3(T,c,T) \tag{2.120}$$

From equations (2.109) and (2.117), we now see that

$$r_T(c,T) = F(r(c,T),c,T) - G(r(c,T),c,T)r_c(c,T) \tag{2.121}$$

This is the first order partial differential equation satisfied by r. From equation (2.111) we see that

$$r(c,0) = 0 \tag{2.122}$$

The equations for u,v and r, together with their initial conditions, constitute the initial value representation for the original nonlinear problem.

**Example 8** Consider $u'' = e^u$, $u(0) = u(b) = 0$

Applying the quasilinearization technique for b=1, we consider the sequence of approximations defined by

$$\begin{aligned} u''_{n+1} &= e^{u_n} + e^{u_n}(u_{n+1} - u_n) \\ u_{n+1}(0) &= u_{n+1}(1) = 0 \end{aligned} \tag{2.123}$$

Taking a most obvious initial approximation $u_0 = 0$, we compute the functions $u_1(x)$ and $u_2(x)$. Table 2.3 indicates the rapidity of convergence.

Table 2.3: The approach of $u_n(x)$ to $u(x)$

| $x$ | $u_0(x)$ | $u_1(x)$ | $u_2(x)$ | $u(x)$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0.1 | 0 | -0.0412 | -0.0414 | -0.0414 |
| 0.2 | 0 | -0,0729 | -0,0732 | -0,0732 |
| 0,3 | 0 | -0,0953 | -0,0958 | -0,0958 |
| 0,4 | 0 | -0,1087 | -0,1092 | -0,1092 |
| 0,5 | 0 | -0,1131 | -0,1137 | -0,1137 |
| 0,6 | 0 | -0,1087 | -0,1092 | -0,1092 |
| 0,7 | 0 | -0,0953 | -0,0958 | -0,0958 |
| 0,8 | 0 | -0,0729 | -0,0732 | -0,0732 |
| 0,9 | 0 | -0,0412 | -0,0414 | -0,0414 |
| 1 | 0 | 0 | 0 | 0 |

## 2.7 Continuation Method

We embed a given problem into a problem with a continuation parameter $\sigma$ in it. For one value of $\sigma$ (say $\sigma = 1$) we obtain the original equations, while for a different value of $\sigma$ (say $\sigma = 0$) we have an "easier" problem. We solve the simpler problem numerically and then slowly vary the continuation parameter from 0 to 1.[1]

After setting up the problem as described above, we define a metric that tells how well a function satisfies the problem when the continuation parameter is between 0 and 1. First, we numerically solve the easier problem (at $\sigma = 0$). Then the continuation parameter $\sigma$ is increased by a small amount, and a solution is found by using Newton's method or sometimes Newton-Raphson method (this is accomplished by making the metric as small as possible). We increase $\sigma$ some more, and repeat this step until we have arrived at $\sigma = 1$.

**Example 9** Suppose we wish to solve the following boundary value problem for y=y(x).

$$y_{xx} + e^y = 0, \qquad y(0) = 1, \qquad y(\frac{\pi}{2}) = 0 \tag{2.124}$$

We embed (2.124) into the problem for $v = v(x;\sigma)$,

$$v_{xx} + (1-\sigma)v + \sigma e^v = 0, \qquad v(0;\sigma) = 1, \qquad y(\frac{\pi}{2};\sigma) = 0 \tag{2.125}$$

Note that when $\sigma = 1$, the problem for $v(x;1)$ becomes identical to the original problem that we wanted to solve, (2.124). Note also that, when $\sigma = 0$, the problem for $v(x;0)$ becomes

$$v(x;0)_{xx} + v(x;0) = 0, \qquad v(0;0) = 1, \qquad v(\frac{\pi}{2};0) = 1 \tag{2.126}$$

with the solution $v(x;0) = \cos x$

The technique is to solve (2.125) numerically on a grid of values from 0 to $\frac{\pi}{2}$. We will start with $\sigma = 0$ and $v(x;0) = \cos x$ and then increase $\sigma$ by a small amount and allow $v(x;\sigma)$ to change accordingly.

We choose to solve (2.124) at the N+1 grid points: $\{x_n = hn/for\ n = 0,1,2,\ldots,N\}$ where $h = \frac{\pi}{2N}$, and we define $v_n^\sigma$ to be the numerical approximation to $v(x;\sigma)$ at the n-th grid point. We take $v_0^\sigma = 1$ and $v_N^\sigma = 0$ so that



T.C. YÜKSEKÖĞRETİM KURULU DOKÜMANTASYON MERKEZİ

the boundary conditions to (2.125) are always satisfied.

Now we must define the metric. We choose

$$\varepsilon_n^\sigma = \frac{v_{n+1}^\sigma - 2v_n^\sigma + v_{n-1}^\sigma}{h^2} + (1-\sigma)v_n^\sigma + \sigma e^{v_n^\sigma} \tag{2.127}$$

We choose this metric since, when $\varepsilon_n^\sigma$ is close to zero, (2.125) will be approximately satisfied. This metric was obtained by simply applying a centered second order difference formula to (2.125).

The procedure is now as follows (with $\sigma_0 = 0, k = 0$):

(A) Increase $\sigma$ by a small amount $\delta\sigma$ (i.e. $\sigma_{k+1} = \sigma_k + \delta\sigma$)

(B) Find $\{v_n^\sigma\}$ by making $\varepsilon_n^{\sigma_k} \simeq 0$. This is best accomplished by Newton's method. That is, we keep iterating

$$\begin{pmatrix} v_2^{\sigma_k} \\ v_3^{\sigma_k} \\ \vdots \\ v_{N-1}^{\sigma_k} \end{pmatrix}_{m+1} = \begin{pmatrix} v_2^{\sigma_k} \\ v_3^{\sigma_k} \\ \vdots \\ v_{N-1}^{\sigma_k} \end{pmatrix}_{m} - J^{-1} \begin{pmatrix} \varepsilon_2^{\sigma_k} \\ \varepsilon_3^{\sigma_k} \\ \vdots \\ \varepsilon_{N-1}^{\sigma_k} \end{pmatrix}_{m} \tag{2.128}$$

where J is the Jacobian matrix defined by $J = \frac{\partial(\varepsilon_2^{\sigma_k}, \varepsilon_3^{\sigma_k}, \ldots, \varepsilon_{N-1}^{\sigma_k})}{\partial(v_2^{\sigma_k}, v_3^{\sigma_k}, \ldots, v_{N-1}^{\sigma_k})}$, until the "difference" between

$$\begin{pmatrix} v_2^{\sigma_k} \\ v_3^{\sigma_k} \\ \vdots \\ v_{N-1}^{\sigma_k} \end{pmatrix}_{m+1} \quad and \quad \begin{pmatrix} v_2^{\sigma_k} \\ v_3^{\sigma_k} \\ \vdots \\ v_{N-1}^{\sigma_k} \end{pmatrix}_{m} \tag{2.129}$$

is smaller than some predefined constant.

Note that the Jacobian and the $\{\varepsilon_n^\sigma\}$ all depend on the values of $\{v_n^{\sigma_k}\}_m$. The initial values for $\{v_n^{\sigma_k}\}_0$ will be given by $\{v_n^{\sigma_{k-1}}\}$. If $\delta\sigma$ is small enough, then Newton's method should converge.

(C) If $\sigma_k \neq 1$, go back to step (A).

(D) If $\sigma_k = 1$, then we have found a numerical approximation to the solution of (2.124).

## 2.8 Shooting Method

The general procedure can be illustrated by studying a second order ordinary differential equation. Suppose we wish to numerically approximate the solution y(x) of the equation

$$\begin{aligned} & L(y'', y', y, x) = 0 \\ y(0) = 0, \qquad & y(1) = A \end{aligned} \tag{2.130}$$

where A is a given constant. If $z(x;\alpha)$ is defined to be the solution of

$$\begin{aligned} & L(z'', z', z, x) = 0 \\ z(0;\alpha) = 0, \qquad & z'(0;\alpha) = \alpha \end{aligned} \tag{2.131}$$

then $y(x)$ will be equal to $z(x;\alpha)$ for one or more values of $\alpha$. The parameter $\alpha$ in (2.131) must be determined so that

$$z(1;\alpha) = A \tag{2.132}$$

Since (2.131) is an initial value problem, it is straightforward to integrate it numerically from $x = 0$ to $x = 1$. To use the shooting method, we integrate (2.131) numerically for some arbitrary initial guess for $\alpha$, say $\alpha_0$. If $z(1;\alpha_0) = A$, then $y(x) = z(x;\alpha_0)$ and we are done.

If $z(1;\alpha_0) \neq A$, then a new value of $\alpha$ must be chosen, say $\alpha_1$. Equation (2.131) is then integrated for this new value of $\alpha$. The process of choosing new values for $\alpha$ is repeated until the value of $z(1;\alpha)$ is sufficiently close to A. If the new $\alpha$'s are chosen well, then $z(1;\alpha)$ will converge to A and a numerical approximation to (2.130) will have been obtained.

One way to choose the sequence of $\alpha$'s is by Newton's method:

$$\alpha_{n+1} = \alpha_n - \frac{z(1;\alpha_n) - A}{\frac{\partial}{\partial\alpha} z(1;\alpha)|_{\alpha=\alpha_n}} \tag{2.133}$$

A numerical way to implement (2.133) might be

$$\alpha_{n+1} = \alpha_n - \frac{z(1;\alpha_n) - A}{[z(1;\alpha_n + \varepsilon) - z(1;\alpha_n)]/\varepsilon}, \tag{2.134}$$

where $\varepsilon$ is a small number.

**Example 10** Suppose we have the nonlinear second order ordinary differential equation

$$\begin{aligned} y'' + 2(y')^2 &= 0 \\ y(0) = 1, \qquad y(1) &= \frac{1}{2} \end{aligned} \tag{2.135}$$

Analytically, the solution of (2.135) is found to be

$$y(x) = 1 + \frac{1}{2}\log(1 + \frac{1-e}{e}x) \tag{2.136}$$

Hence,

$$y'(0) = \frac{1-e}{2e} \simeq -0,31606 \tag{2.137}$$

The equation in (2.135) is turned into the two first order ordinary differential equations

$$\begin{aligned} \frac{dy}{dx} &= z, \\ \frac{dz}{dx} &= -2y^2, \end{aligned} \tag{2.138}$$

and then integrated by the use of Euler's method. An initial guess of $y'(0) = 0$ is used.

Table 2.4: The successive approximation of $y'(0)$.

| | | |
|---|---|---|
| Iteration number | 0 | Value of y'(0)=0 |
| Iteration number | 1 | Value of y'(0)=-0,50000 |
| Iteration number | 2 | Value of y'(0)=-0,49857 |
| Iteration number | 3 | Value of y'(0)=-0,49102 |
| Iteration number | 4 | Value of y'(0)=-0,46366 |
| Iteration number | 5 | Value of y'(0)=-0,40465 |
| Iteration number | 6 | Value of y'(0)=-0,34199 |
| Iteration number | 7 | Value of y'(0)=-0,31799 |
| Iteration number | 8 | Value of y'(0)=-0,31608 |
| Iteration number | 9 | Value of y'(0)=-0,31607 |

## 2.9 Box Method

We will illustrate the procedure on the general second order linear ordinary differential equation. The same technique can be used, with only slight modifications, to systems of higher order ordinary differential equations, with the boundary data given virtually anywhere in the interval of interest. For nonlinear equations or nonlinear boundary conditions, this method can be used iteratively by quasilinearizing the nonlinear terms at each step.

Given the second order linear ordinary differential equation

$$a(x)y'' + b(x)y' + c(x)y = d(x) \tag{2.139}$$

$$y(x_L) = y_L, \qquad y(x_U) = y_U \tag{2.140}$$

we introduce the variable $z(x) = y'(x)$ and write (2.139)-(2.140) as the system

$$\frac{d}{dx}\begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} z \\ \frac{d-cy-bz}{a} \end{pmatrix} \tag{2.141}$$

Now we choose a grid, not necessarily uniform, on the interval $(x_L, x_U)$, say $x_L = x_1 < x_2 < \cdots < x_N = x_U$. At each one of the grid points, some finite difference scheme is chosen to approximate the equations in (2.141). The scheme used can vary from point to point. For instance, if Euler's method is used for every point, then

$$\begin{pmatrix} y \\ z \end{pmatrix}_{k+1} = \begin{pmatrix} y \\ z \end{pmatrix}_k + (x_{k+1} - x_k)\begin{pmatrix} z \\ \frac{d-cy-bz}{a} \end{pmatrix}_k \tag{2.142}$$

to first order, where $y_k = y(x_k)$, $z_k = z(x_k)$, and similarly for $\{a_k, b_k, c_k, d_k\}$. From (2.141) the values $y_1 = y_L$ and $y_N = y_U$ are known.

To determine all of the $\{z_k\}$, and the remaining $\{y_k\}$, all of the relations in (2.142)(that is, for $k = 1, 2, \ldots, N$) should be combined into one large matrix equation. First, for ease of notation, define $h_k = x_{k+1} - x_k$, $e_k = d_k/a_k$, $f_k = c_k/a_k$ and $g_k = b_k/a_k$. In these new variables, equation (2.142) may be written as

$$\begin{aligned} y_{k+1} &= y_k + h_k z_k \\ z_{k+1} &= z_k + h_k(e_k - f_k y_k - g_k z_k). \end{aligned} \tag{2.143}$$

Combining all of the equations in (2.143) results in

$$\begin{pmatrix} 1 & h_1 & -1 & 0 & 0 & 0 & \dots \\ h_1 f_1 & -1+h_1 g_1 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & h_2 & -1 & 0 & \dots \\ 0 & 0 & h_2 f_2 & -1+h_2 g_2 & 0 & 1 & \\ \vdots & \vdots & & & & & \end{pmatrix} \begin{pmatrix} y_1 \\ z_1 \\ y_2 \\ z_2 \\ y_3 \\ \vdots \\ z_N \end{pmatrix} = \begin{pmatrix} 0 \\ h_1 e_1 \\ 0 \\ h_2 e_2 \\ 0 \\ \vdots \\ h_N e_N \end{pmatrix}$$

To this matrix equation should be added two more rows, one corresponding to $y_1 = y_L$ and one corresponding to $y_N = y_U$. With these two rows, there results an $2N \times 2N$ matrix equation. This equation can be solved to determine a numerical approximation to the solution at all of the grid points.

**Example 11** The second order linear ordinary differential equation

$$\begin{aligned} y'' + y &= 3, \\ y(0) = 3, \qquad y(\frac{\pi}{2}) &= 2 \end{aligned} \tag{2.144}$$

has the solution $y = 3 - \sin x$. We will use the box method to numerically approximate the solution. Writing (2.144) as a system results in

$$\frac{d}{dx}\begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} z \\ 3-y \end{pmatrix}. \tag{2.145}$$

We choose a uniform grid: $x_n = (n-1)h$ for $n = 1, 2, 3, 4$ with $h = \pi/6$. Defining $y_n = y(x_n)$ and $z_n = z(x_n)$, then, using Euler's method, (2.145) may be approximated as

$$\begin{aligned} y_{n+1} &= y_n + h z_n \\ z_{n+1} &= z_n + h(3 - y_n) \end{aligned} \tag{2.146}$$

combining all the equations in (2.146) for $n = 1, 2, 3, 4$ results in

$$\begin{pmatrix} 1 & h & -1 & 0 & 0 & 0 & 0 & 0 \\ h & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & h & -1 & 0 & 0 & 0 \\ 0 & 0 & h & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & h & -1 & 0 \\ 0 & 0 & 0 & 0 & h & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ z_1 \\ y_2 \\ z_2 \\ y_3 \\ z_3 \\ y_4 \\ z_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 3h \\ 0 \\ 3h \\ 0 \\ 3h \end{pmatrix} \tag{2.147}$$

Then the following two rows are added, to incorporate the known values of $y(0)$ and $y(\pi/2)$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ z_1 \\ y_2 \\ z_2 \\ y_3 \\ z_3 \\ y_4 \\ z_4 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \tag{2.148}$$

Here is the approximate solution:

3.000 -0.701 2.633 -0.701 2.266 -0.509 2.000 -0.124

Here is the exact solution:

3.000 -1.000 2.500 -0.866 2.134 -0.500 2.000 0.000

The values for $y_n$ are only accurate to one decimal place in this example. Putting more points in the interval would decrease the error, as would using a higher order method in place of Euler's method.

## 2.10 Differential-Algebraic Equations

### 2.10.1 Introduction

Most treatments of ordinary differential equations, both analytical and numerical, begin by defining the first order system

$$F(t, y(t), y'(t)) = 0, \tag{2.149}$$

where F and y are vector valued. (2.149) can be rewritten in the explicit form

$$y' = f(t, y(t)) \tag{2.150}$$

If (2.149) can, in principle, be rewritten as (2.150) with the same state variables y, then it will be referred to as a system of implicit ordinary differential equations. We are especially interested in those problems for which this rewriting is impossible or less desirable. In a system of differential-algebraic equations, there are algebraic constraints on the variables. The constraints may appear explicitly as in (2.152) of the system

$$F(x', x, y, t) = 0 \tag{2.151}$$
$$G(x, y, t) = 0 \tag{2.152}$$

where the Jacobian of F with respect to $x'$ (denoted by $\frac{\partial F}{\partial x'} = F_{x'}$) is nonsingular, or they may arise because $F_{y'}$ in (2.149) is singular.

Singular perturbation problems form a special class of problems containing a parameter $\varepsilon$. When this parameter $\varepsilon$ tends to zero, the differential equation becomes differential-algebraic. So in equation [6]

$$\varepsilon z'' + (z^2 - 1)z' + z = 0 \tag{2.153}$$

we insert the identity

$$\varepsilon z'' + (z^2 - 1)z' = \frac{d}{dx} \underbrace{(\varepsilon z' + (\frac{z^3}{3} - z))}_{=y} \tag{2.154}$$

so that (2.153) becomes

$$\begin{aligned} y' &= -z = f(y, z) \\ \varepsilon z' &= y - \left(\frac{z^3}{3} - z\right) = g(y, z) \end{aligned} \tag{2.155}$$

In order to approximate the solution for very small $\varepsilon$, we get $\varepsilon = 0$ in (2.155) and obtain

$$\begin{aligned} y' &= -z = f(y, z) \\ 0 &= y - \left(\frac{z^3}{3} - z\right) = g(y, z) \end{aligned} \tag{2.156}$$

While (2.155) has no analytic solution, (2.156) can easily be solved to give

$$y' = -z = (z^2 - 1)z' \qquad or \qquad \ln|z| - \frac{z^2}{2} = x + C \tag{2.157}$$

Equation (2.156) is a differential-algebraic equation, since it combines a differential equation (first line) with an algebraic equation(second line).

The general nonlinear differential-algebraic equation

$$F(t, y(t), y'(t)) = 0 \tag{2.158}$$

may be linear in the derivative

$$A(t, y(t))y'(t) + f(t, y(t)) = 0 \tag{2.159}$$

This system is sometimes referred to as linearly implicit. A special case of (2.159) is the semi-explicit nonlinear differential-algebraic equation.

$$\begin{aligned} x_1'(t) &= f_1(x_1(t), x_2(t), t) \\ 0 &= f_2(x_1(t), x_2(t), t) \end{aligned} \tag{2.160}$$

We shall sometimes refer to a system as semi-explicit if it is in the form

$$\begin{aligned} F(x'(t), x(t), y(t), t) = 0 \\ G(x(t), y(t), t) = 0 \end{aligned} \tag{2.161}$$

where $F_{x'}$ is nonsingular.

### 2.10.2 Index

For general differential-algebraic systems (2.149), the index along a solution $y(t)$ is the minimum number of differentiations of the system which would be required to solve for $y'$ uniquely in terms of y and t.

A property known as the index plays a key role in the classification and behavior of differential-algebraic equations.[7]

Consider the special case of a semi-explicit differential-algebraic equations

$$x' = f(x, y, t) \tag{2.162}$$

$$0 = g(x, y, t) \tag{2.163}$$

If we differentiate the constraint equation (2.163) with respect to t, we get

$$\begin{aligned} x' &= f(x,y,t) && (2.164)\\ g_x(x,y,t)x' + g_y(x,y,t)y' &= -g_t(x,y,t) && (2.165) \end{aligned}$$

If $g_y$ is nonsingular, the system (2.164)-(2.165) is an implicit ordinary differential equation and we say that (2.162)-(2.163) has index one. If the new system is not an implicit ordinary differential equation, we repeat the process. The number of differentiation steps required in this procedure is the index.

In example; the scalar equation $y = q(t)$ is a (trivial) index-1 differential-algebraic equation, because it takes one differentiation to obtain an ordinary differential equation for y.[8]

For the system

$$\begin{aligned} y_1 &= q(t), && (2.166)\\ y_2 &= y_1', \end{aligned}$$

we differentiate the first equation to get

$$y_2 = y_1' = q'(t) \qquad (2.167)$$

and then

$$y_2' = y_1'' = q''(t) \qquad (2.168)$$

The index is 2 because two differentiations of $q(t)$ were needed.

A similar treatment for the system

$$\begin{aligned} u &= q(t), && (2.169)\\ y_3 &= u'', \end{aligned}$$

necessitates three differentiations to obtain an ordinary differential equation for $y_3$, hence the index is 3.

### 2.10.3 Solutions of Differential-Algebraic Equations

Differential-algebraic equations are more difficult to solve than standard ordinary differential equations. There are a few analytic solution techniques for differential-algebraic equation, as the examples show.

**Example 12** The differential equation

$$y = f(y') = (y')^5 + (y')^3 + y' + 5 \tag{2.170}$$

for $y(x)$ is an example of a differential-algebraic equation. It is impossible for (2.170) to be analytically written in the form $y' = g(x, y)$.

However, it is possible to solve differential equations of the form $y = f(y')$ parametrically. The solution may be written as

$$y = f(t), \qquad x = \int t^{-1} f'(t) dt + C, \tag{2.171}$$

where C is an arbitrary constant. Hence, equation (2.170) has the solution

$$\begin{aligned} x &= \frac{5}{4}t^4 + \frac{3}{2}t^2 + \log t + C, \\ y &= t^5 + t^3 + t + 5. \end{aligned} \tag{2.172}$$

**Example 13** If a differential-algebraic equation is of the form $x = f(y')$, then the solution may be written parametrically as

$$x = f(t), \qquad y = \int t f'(t) dt + C, \tag{2.173}$$

where C is an arbitrary constant. Thus, the equation $x = (y')^3 - y' - 1$ has the parametric solution

$$\begin{aligned} x &= t^3 - t - 1, \\ y &= \frac{3}{4}t^4 - \frac{1}{2}t^2 + C. \end{aligned} \tag{2.174}$$

Differential-algebraic equations are solved exclusively by numerical means. One common numerical technique is to use the backwards Euler method.[7]

That is, (2.149) is approximated by

$$F\left(t_n, y_n, \frac{y_n - y_{n-1}}{h}\right) \tag{2.175}$$

and then the resulting system of nonlinear equations is solved for $y_1$, then $y_2$, etc.

For the simplest class of nonlinear differential-algebraic equations, namely, semi-explicit index-1,

$$\begin{aligned} x' &= f(t, x, z) \\ 0 &= g(t, x, z), \end{aligned} \tag{2.176}$$

where $g_z$ is nonsingular.

First we assume that, there exists a function $\tilde{g}$ such that $z = \tilde{g}(t, x)$. Thus the differential-algebraic equation (2.176) is equivalent to the ordinary differential equation $x' = f(t, x, \tilde{g}(t, x))$.

Now, consider the backward Euler method applied to (2.176),

$$\frac{x_n - x_{n-1}}{h} = f(t_n, x_n, z_n) \tag{2.177}$$

$$0 = g(t_n, x_n, z_n) \tag{2.178}$$

Solving for $z_n$ in (2.178) and substituting into (2.177) yields

$$\frac{x_n - x_{n-1}}{h} = f(t_n, x_n, \tilde{g}(t_n, x_n)) \tag{2.179}$$

So if we consider the differential-algebraic equation

$$\begin{aligned} \frac{dy_2}{dt} &= y_1 + \eta_1(t) \\ 0 &= y_2 + \eta_2(t) \end{aligned} \tag{2.180}$$

Backward Euler gives

$$\begin{aligned} y_1(t_n) &= \frac{y_2(t_n) - y_2(t_{n-1})}{h} - \eta_1(t_n) \\ y_2(t_n) &= -\eta_2(t_n). \end{aligned} \tag{2.181}$$

(Note that higher order techniques such as Runge-Kutta methods are generalizations of this simple idea.)[9]

## 2.11 Other Methods

### 2.11.1 Weighted Residual Methods

This method changes the numerical calculation of an ordinary differential equation to the numerical calculation of a set of algebraic equations. We approximate the solution by taking a linear combination of an arbitrarily chosen set of functions. The coefficients of the functions, which may be constants or functions themselves, are unknown. We may use any of a number of schemes to find the numerical values for the unknown coefficients. More details can be found in [10].

The method of weighted residuals can be described in its generality by considering the operator equation

$$A(u) = f \tag{2.182}$$

in domain $\Omega$ , where A is an operator, often a differential operator, acting on the dependent variable u, and f is a known function of the independent variables.

The solution u is approximated by the expression

$$u_N = \sum_{j=1}^{N} c_j \phi_j + \phi_0 \tag{2.183}$$

Substitution of the approximate solution $u_N$ into the left-hand side of (2.182) gives a function $f_N \equiv A(u_N)$ that, in general, is not equal to the specified function f. The difference $A(u_N) - f$, called the residual of the approximation, is nonzero:

$$R \equiv A(u_N) - f = A(\sum_{j=1}^{N} c_j \phi_j + \phi_0) - f \neq 0 \tag{2.184}$$

In the weighted-residual method, the parameters $c_j$ are determined by requiring the residual R to vanish in the weighted-integral sense:

$$\int_\Omega \psi_i(x,y) R(x,y,c_j) dxdy = 0 \qquad (i = 1, 2, \ldots, N) \tag{2.185}$$

where $\Omega$ is a two-dimensional domain and $\psi_i$ are weight functions, which, in general, are not the same as the approximation functions $\phi_i$. The set $\{\psi_i\}$ must be a linearly independent set; otherwise, the equations provided by (2.185) will not be linearly independent and hence will not be solvable.

Now we choose N weighting functions, $\{\psi_i(x)\}$. It is the choice of the weighting functions that defines the method. For example;

$$\begin{aligned} Petrov-Galerkin &: \psi_i \neq \phi_i, \\ Galerkin &: \psi_i = \phi_i, \\ Least-squares &: \psi_i = \frac{\partial R}{\partial c_i} \end{aligned} \tag{2.186}$$

**Example 14** Consider the differential equation

$$-\frac{d^2u}{dx^2} - u + x^2 = 0, \qquad u(0) = 0, \qquad u'(1) = 1 \tag{2.187}$$

For a weighted-residual method, $\phi_0$ and $\phi_i$ should satisfy the following conditions:

$\phi_0(0) = 0$, $\phi_0'(1) = 1$ (satisfy actual boundary conditions)
$\phi_i(0) = 0$, $\phi_i'(1) = 0$ (satisfy homogeneous form of the specified boundary conditions)

For a choice of algebraic polynomials, we assume $\phi_0(x) = a + bx$ and use the two conditions on $\phi_0$ to determine the constants a and b. We obtain $\phi_0(x) = x$

Since there are two homogeneous conditions, we must assume at least a three-parameter polynomial to obtain a nonzero function, $\phi_1 = a + bx + cx^2$. Using the conditions on $\phi_i$, we obtain $\phi_1 = -x(2 - x)$

For $\phi_2$, we can assume $\phi_2 = a + bx + dx^3$ with $d \neq 0$; $\phi_2$ does not contain all-order terms in either case, but the approximate solution is complete because $\{\phi_1, \phi_2\}$ contains all terms up to degree three. We obtain $\phi_2 = x^2(1 - \frac{2}{3}x)$.

The residual in the approximation of the equation is

$$\begin{aligned} R &= -\left(\sum_{i=1}^{N} c_i \frac{d^2\phi_i}{dx^2}\right) - \left(\phi_0 + \sum_{i=1}^{N} c_i\phi_i\right) + x^2 \\ &= c_1(2 - 2x + x^2) + c_2(-2 + 4x - x^2 + \frac{2}{3}x^2) - x + x^2 \end{aligned} \tag{2.188}$$

First consider Petrov-Galerkin method.

Let the weight functions be $\psi_1 = x$, $\psi_2 = x^2$. Then

$$\int_0^1 xRdx = 0, \qquad \int_0^1 x^2 Rdx = 0 \tag{2.189}$$

Solving for $c_i$, we obtain $c_1 = \frac{103}{682}$ and $c_2 = -\frac{15}{682}$; the solution becomes

$$u = 1.3020x - 0.1730x^2 - 0.0146x^3 \tag{2.190}$$

Galerkin method: Taking $\psi_i = \phi_i$, we have

$$\int_0^1 x(2-x)Rdx = 0, \qquad \int_0^1 x^2(1-\frac{2}{3}x)Rdx = 0 \tag{2.191}$$

Hence, the solution becomes $u = 1.2894x - 0.1398x^2 - 0.0032x^3$

Least-squares method: Taking $\psi_i = \partial R/\partial c_i$, we have

$$\int_0^1 (2-2x+x^2)Rdx = 0, \qquad -\int_0^1 (2-4x+x^2-\frac{2}{3}x^3)Rdx = 0 \tag{2.192}$$

So we have $u = 1.2601x - 0.0801x^2 - 0.0332x^3$

For this problem, the Petrov-Galerkin method gives the most accurate solution.

### 2.11.2 Finite Element Method

The finite element method is one version of the method of weighted residuals. The present method is characterized by having "local elements".

Given a differential equation that comes from a variational principle, and a domain in which the equation is to be solved, the steps are as follows:[11]

[1 ] Discretize the domain into simple shapes (these are the "finite elements"). Define a basis function $\phi_k(x)$ on each of the finite elements. These basis functions should have bounded support.

[2 ] Assemble the stiffness matrix and the load matrix. These only depend on the finite elements chosen and not on the differential equation to be approximated.

[3 ] Write the given differential equation as a variational principle. Approximate the unknown in the variational principle by a linear combination of the functions defined on the finite elements; i.e. $u(x) \simeq u_N(x) = \sum_{k=1}^{N} c_k \phi_k(x)$. In this last expression, the $\{c_k\}$ are unknown and must be determined.

[4 ] Construct element stiffness matrices and load vectors, element by element. Then assemble these together into the global stiffness matrix A and the global load vector f.

[5 ] Relate the minimization in the variational principle to the minimization of the quadratic functional

$$I[u_N] = c^T A c - 2c^T f \tag{2.193}$$

When A is symmetric, the minimization of (2.193) will occur when c is the solution of the system: $Ac = f$. If the original differential equation was nonlinear, then $A = A(c)$ or $f = f(c)$.

**Example 15** Suppose we have the constant coefficient second order linear ordinary differential equation

$$L[u] = -\frac{d}{dx}\left(p(x)\frac{du}{dx}\right) + q(x)u = f(x) \tag{2.194}$$

on the interval $0 \leq x \leq 1$. For simplicity, we take $p(x)$ and $q(x)$ to be constants. For this equation, we take the natural boundary conditions

$$u(0) = u(1) = 0 \tag{2.195}$$

Then

$$I[u] = \int_0^1 [p(u'(x))^2 + qu^2(x) - 2f(x)u(x)]dx \tag{2.196}$$

It is straightforward to show that the first variation of I[u] yields (2.194) and (2.195).[10]

Now we define the interval $(x_k, x_{k+1})$ to be "finite element number k." We choose as basis functions on the finite elements the functions $\phi_k(x)$ defined by

$$\phi_k(x) = \begin{cases} \frac{x - x_{k-1}}{h}, & for \quad x_{k-1} \leq x \leq x_k, \\ \frac{x_{k+1} - x}{h}, & for \quad x_k \leq x \leq x_{k+1}, \\ 0, & otherwise. \end{cases} \tag{2.197}$$

Note that

$$\phi_k'(x) = \begin{cases} \frac{1}{h}, & for \qquad x_{k-1} \le x \le x_k, \\ -\frac{1}{h}, & for \qquad x_k \le x \le x_{k+1}, \\ 0, & otherwise. \end{cases} \tag{2.198}$$

Now we approximate the function $u(x)$ by a linear combination of the $\phi_k(x)$. We take

$$u(x) \simeq u_N(x) = \sum_{k=1}^{N} c_k \phi_k(x) \tag{2.199}$$

where the unknowns $\{c_k\}$ must be determined. Once the $\{c_k\}$ are known, then the approximation to $u(x)$ at any point can be found from (2.199).

Hence, on finite element k (i.e., for $x_k < x < x_{k+1}$)

$$\begin{aligned} u_N(x) &= c_k\phi_k(x) + c_{k+1}\phi_{k+1}(x) \\ u_N'(x) &= \frac{-c_k + c_{k+1}}{h} \end{aligned} \tag{2.200}$$

Using $u_N(x)$ in (2.196) results in

$$\begin{aligned} I[u_N] &= \sum_{k=0}^{N} \int_{x_k}^{x_{k+1}} [p(u_N')^2 + qu_N^2 - 2fu_N]dx \\ &= \sum_{k=0}^{N} [I_k^s + I_k^m + I_k^l], \end{aligned} \tag{2.201}$$

where

$$\begin{aligned} I_k^s &= \int_{x_k}^{x_{k+1}} p(u_N')^2 dx = \begin{pmatrix} c_k & c_{k+1} \end{pmatrix} K_k^s \begin{pmatrix} c_k \\ c_{k+1} \end{pmatrix}, \\ I_k^m &= \int_{x_k}^{x_{k+1}} q(u_N)^2 dx = \begin{pmatrix} c_k & c_{k+1} \end{pmatrix} K_k^m \begin{pmatrix} c_k \\ c_{k+1} \end{pmatrix}, \\ I_k^l &= \int_{x_k}^{x_{k+1}} 2f(x)u_N(x)dx \end{aligned} \tag{2.202}$$

Here $K_k^s$ and $K_k^m$ are element stiffness matrices, they are defined by

$$K_k^s = \frac{p}{h}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \qquad K_k^m = \frac{qh}{6}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}. \tag{2.203}$$

A numerical integration is required to determine $I_k^l$. If, on finite element number k, $f(x)$ is approximated by $f(x) \simeq f_k\phi_k(x)+f_{k+1}\phi_{k+1}(x)$, then we find $I_k^l = \begin{pmatrix} f_k^l \end{pmatrix}^T \begin{pmatrix} c_k \\ c_{k+1} \end{pmatrix}$, where the load vector is defined by $f_k^l = \frac{h}{3}\begin{pmatrix} 2f_k + f_{k+1} \\ f_k + 2f_{k+1} \end{pmatrix}$.

The system can now be assembled, element by element. That is, we write a single matrix equation representing (2.201). For this example, we find that

$$I[u_N] = c^T(K + M)c - 2f^T c, \tag{2.204}$$

where $c = (c_1, c_2, \ldots, c_N)^T$, $f = \frac{h}{6}(f_0 + 4f_1 + f_2, f_1 + 4f_2 + f_3, \ldots, f_{N-2} + 4f_{N-1} + f_N)^T$, and

$$K = \frac{p}{h}\begin{pmatrix} 2 & -1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 2 & -1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 2 & -1 & & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & & -1 & 2 & -1 & 0 \\ 0 & 0 & \ldots & 0 & -1 & 2 & -1 \\ 0 & 0 & \ldots & 0 & 0 & -1 & 2 \end{pmatrix}, \tag{2.205}$$

$$M = \frac{qh}{6}\begin{pmatrix} 4 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 4 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 4 & 1 & & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & & 1 & 4 & 1 & 0 \\ 0 & 0 & \ldots & 0 & 1 & 4 & 1 \\ 0 & 0 & \ldots & 0 & 0 & 1 & 4 \end{pmatrix}. \tag{2.206}$$

To minimize the expression in (2.204), c should be chosen (since the global stiffness matrix K+M is symmetric in this example) to satisfy the matrix equation $(K + M)c = f$. It may be solved by standard linear algebra routines.

# CHAPTER 3

# GENERALIZATIONS OF RUNGE-KUTTA METHOD

## 3.1 Introduction

We introduce a form of Runge-Kutta in which it is assumed that the user will evaluate both f and $f'$ in solving $y' = f(x, y)$ numerically. This allows us to introduce new Runge-Kutta parameters that increase the order of accuracy of the solution with evaluations of $f'$ replacing evaluations of f.

We assume that $f(x, y)$ is a continuous function with domain D in $\mathbf{R}^{n+1}$ where $x \in \mathbf{R}$, $y \in \mathbf{R}^n$ and $(x, y) \in D$. We assume that

$$||f(x, y_1) - f(x, y_2)||_2 \leq L||y_1 - y_2||_2 \tag{3.1}$$

for all $(x, y_1)$, $(x, y_2) \in D$; thus the problem

$$\begin{aligned} y' &= f(x, y) \\ y(x_0) &= y_0 \qquad with \qquad (x_0, y_0) \in D \end{aligned} \tag{3.2}$$

has a unique solution.

The classical Runge-Kutta method is an algorithm designed to approximate the Taylor series solution by using a linear combination of values of $f(x, y)$ to approximate $y(x)$. This linear combination is matched up with the Taylor series for $y(x)$ to obtain agreement up to a $h^p$ term, giving methods of order p.

For completeness and later comparison, the following outlines the development of Runge-Kutta formulas by deriving the classical Runge-Kutta second-order method. The technique employed in this derivation extends easily to the development of all Runge-Kutta type formulas. This method requires two evaluations of the first derivatives to obtain agreement with the Taylor series solution through terms order $h^2$.

Given the first-order ordinary differential equation $y' = f(x, y)$ with the initial condition $y(x_0) = y_0$, this classical method seeks values $b_1$, $b_2$, $a_2$ and $c_2$ so as to match

$$\hat{y}_{n+1} = y_n + b_1 k_1 + b_2 k_2 \tag{3.3}$$

where

$$\begin{aligned} k_1 &= hf(x_n, y_n), \\ k_2 &= hf(x_n + hc_2, y_n + a_2 k_1) \end{aligned} \tag{3.4}$$

with the Taylor series expansion

$$\bar{y}_{n+1} = y_n + hy'(x_n, y_n) + \frac{1}{2!}h^2 y''(x_n, y_n) + \frac{1}{3!}h^3 y'''(x_n, y_n) + \ldots \tag{3.5}$$

up to and including terms containing $h^2$. In what follows, all arguments of f and its derivatives will be suppressed when evaluated at $(x_n, y_n)$.

First, one must expand the function f in the $k_2$ equation in a Taylor series for a function of two variables and drop all terms in which the exponent of h is greater than two

$$\begin{aligned} k_2 &= hf(x_n + hc_2, y_n + a_2 k_1) \\ &= hf(x_n + hc_2, y_n + a_2 hf) \end{aligned} \tag{3.6}$$

So the object is to match

$$\begin{aligned} \hat{y}_{n+1} &= y_n + b_1 k_1 + b_2 k_2 \\ &= y_n + b_1 hf + b_2 h[f + hc_2 f_x + a_2 hf f_y + O(h^2)] \\ &= y_n + (b_1 + b_2)hf + b_2 h^2 (c_2 f_x + a_2 f f_y) + O(h^3) \end{aligned} \tag{3.7}$$

with the Taylor series expansion

$$\begin{aligned} \bar{y}_{n+1} &= y_n + hy' + \frac{1}{2!}h^2 y'' + O(h^3) \\ &= y_n + hf + \frac{1}{2!}h^2 f' + O(h^3) \\ &= y_n + hf + \frac{1}{2!}h^2 (f_x + f f_y) + O(h^3) \end{aligned} \tag{3.8}$$

Equating coefficients of like terms in the above expression for $\hat{y}_{n+1}$ and $\bar{y}_{n+1}$, one obtains agreement in the following terms

$$\begin{aligned} h^1 f: &\quad b_1 + b_2 = 1 \\ h^2 f_x: &\quad b_2 c_2 = 1/2 \\ h^2 f f_y: &\quad b_2 a_2 = 1/2 \end{aligned} \tag{3.9}$$

These equations can be solved in terms of an arbitrary parameter, $b_2$, to give

$$\begin{aligned} b_1 &= 1 - b_2 \\ a_2 &= c_2 = 1/(2b_2) \end{aligned} \tag{3.10}$$

This gives a one-parameter family of explicit two-stage, second order, one-step Runge-Kutta method, provided $b_2 \neq 0$. When $b_2 = 0$, the initial equation defining this method, $\hat{y}_{n+1} = y_n + b_1 k_1 + b_2 k_2$, collapses to the first-order Euler's method. Well-known second-order methods are obtained with $b_2 = 1/2$ and 1. For $b_2 = 1/2$, the method is also called the improved Euler's method

$$\begin{aligned} y_{n+1} &= y_n + (k_1 + k_2)/2 \\ k_1 &= hf(x_n, y_n), \\ k_2 &= hf(x_n + h, y_n + k_1) \end{aligned} \tag{3.11}$$

For $b_2 = 1$, the method is also called the modified Euler's method

$$\begin{aligned} y_{n+1} &= y_n + k_2 \\ k_1 &= hf(x_n, y_n), \\ k_2 &= hf(x_n + h/2, y_n + k_1/2) \end{aligned} \tag{3.12}$$

### 3.2 Third-order method

Most efforts to increase the order of the Runge-Kutta methods have been accomplished by increasing the number of Taylor's series used. New method introduces new terms involving higher order derivatives of f in the Runge-Kutta $k_i$ terms ($i > 1$) to achieve a higher order of accuracy without a corresponding increase in evaluations of f, but with the addition of evaluations or approximations of $f'$.

The third-order method, for autonomous systems, assigns

$$\begin{aligned} y_{n+1} &= y_n + b_1 k_1 + b_2 k_2 \\ k_1 &= hf(y_n) \end{aligned} \tag{3.13}$$

However, additional terms are introduced by assigning

$$k_2 = hf(y_n + a_{21} k_1 + h a_{22} k_1) \tag{3.14}$$

The general procedure employed in all Runge-Kutta methods is to equate the Taylor series expansion of the method equations with the Taylor series expansion of $y_{n+1}$. Thus this method requires the Taylor series expansion of $k_2$ to $O(h^4)$. Its expansion is as follows

$$k_2 = hf + h(a_{21}h + a_{22}h^2)ff_y + \frac{1}{2}h(a_{21}h + a_{22}h^2)^2 f^2 f_{yy} + O(h^4) \quad (3.15)$$

Thus one must match

$$\begin{aligned} y_{n+1} &= y_n + b_1k_1 + b_2k_2 \\ &= y_n + b_1hf + b_2hf(y_n + a_{21}k_1 + ha_{22}k_1) \quad (3.16) \\ &= y_n + b_1hf + b_2(hf + h(a_{21}h + a_{22}h^2)ff_y + \frac{1}{2}h(a_{21}h + a_{22}h^2)^2 f^2 f_{yy}) \\ &\quad + O(h^4) \\ &= y_n + (b_1 + b_2)hf + b_2a_{21}h^2ff_y + b_2a_{22}h^3ff_y + \frac{1}{2}b_2a_{21}^2h^3f^2f_{yy} + O(h^4) \end{aligned}$$

with the Taylor series expansion

$$\begin{aligned} \bar{y}_{n+1} &= y_n + hy' + \frac{1}{2!}h^2y'' + \frac{1}{3!}h^3y''' + O(h^4) \\ &= y_n + hf + \frac{1}{2!}h^2f' + \frac{1}{3!}h^3f'' + O(h^4) \quad (3.17) \\ &= y_n + hf + \frac{1}{2!}h^2ff_y + \frac{1}{3!}h^3(ff_y^2 + f^2f_{yy}) + O(h^4) \end{aligned}$$

Equating coefficients of like terms for the above expressions for $y_{n+1}$ and $\bar{y}_{n+1}$, one obtains the following

$$\begin{array}{lll} h^1f & : & b_1 + b_2 = 1 \\ h^2ff_y & : & b_2a_{21} = 1/2 \\ h^3ff_y & : & b_2a_{22} = \frac{1}{6}f_y \\ h^3f^2f_{yy} & : & \frac{1}{2}b_2a_{21}^2 = 1/6 \end{array}$$

This reduces to the following system of equations

$$\begin{aligned} b_1 + b_2 &= 1 \\ b_2a_{21} &= 1/2 \quad (3.19) \\ b_2a_{22} &= \frac{1}{6}f_y \\ b_2a_{21}^2 &= 1/3 \end{aligned}$$

of which the only solution is $b_1 = 1/4$, $b_2 = 3/4$, $a_{21} = 2/3$ and $a_{22} = \frac{2}{9}f_y$.

In summary, the following specifies a third-order method, satisfying the Taylor series expansions to $O(h^3)$

$$\begin{aligned} k_1 &= hf(y_n) \\ k_2 &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}hf_y k_1) \\ y_{n+1} &= y_n + \frac{1}{4}k_1 + \frac{3}{4}k_2 \end{aligned} \tag{3.20}$$

### 3.2.1 Utilizing $f_y$

The previous section developed a two-stage, third-order method; however, it introduced a term with $f_y$. The result is the addition of a higher derivative term to the classical Runge-Kutta method. The following describes three methods to utilize $f_y$.

**Method 1:** If one knows or can generate $f_y$, and if the evaluation of $f_y$ is cheaper than the evaluation of f, then savings can be realized. For example, with a linear system of equations, $y' = Ay$, $f_y$ is known and constant.

**Method 2:** Since $y'' = f' = f_y f$ for autonomous equations and since $k_1 = hf$, $k_2$ can be replaced with

$$\begin{aligned} k_2 &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}hf_y k_1) \\ &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}hf_y hf) \\ &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}h^2 f_y f) \\ &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}h^2 f') \end{aligned} \tag{3.21}$$

or

$$k_2 = hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}h^2 y'')$$

Again, savings can be realized if one can formulate $y''$ (or $f'$) and if it is cheaper to evaluate than f.

**Method 3:** It suffices to approximate $f'$ to some given accuracy by using the current and previous evaluations of f. For the third-order method above, an approximation of $O(h)$ is enough to retain the third order in the computed

$y_{n+1}$. This can be seen by comparing the computations using the exact $f'$ with that using a first-order approximation. Let

$$\bar{y}_n = y_n + \frac{2}{3}k_1 + \frac{2}{9}hf_y k_1 \tag{3.22}$$

or

$$\bar{y}_n = y_n + \frac{2}{3}k_1 + \frac{2}{9}h^2 f'_n \tag{3.23}$$

and let

$$\hat{y}_n = y_n + \frac{2}{3}k_1 + \frac{2}{9}h^2 \hat{f}'_n \tag{3.24}$$

where

$$\hat{f}'_n = \frac{f(y_n) - f(y_{n-1})}{h} \tag{3.25}$$

However, since $f(y_{n-1}) = f(y_n) - hf'(y_n) + O(h^2)$ then

$$\hat{f}'_n = f'(y_n) + O(h) = f'_n + O(h) \tag{3.26}$$

thus

$$\hat{y}_n = y_n + \frac{2}{3}k_1 + \frac{2}{9}h^2 f'_n + O(h^3) \tag{3.27}$$

or

$$\hat{y}_n = \bar{y}_n + O(h^3) \tag{3.28}$$

Thus, if

$$y_{n+1} = y_n + \tfrac{1}{4}k_1 + \tfrac{3}{4}hf(\bar{y}_n) \quad \text{is replaced by} \quad \hat{y}_{n+1} = y_n + \tfrac{1}{4}k_1 + \tfrac{3}{4}hf(\hat{y}_n)$$

we have

$$\hat{y}_{n+1} - y_{n+1} = \frac{3}{4}h(f(\hat{y}_n) - f(\bar{y}_n)) \tag{3.29}$$

By the Lipschitz condition,

$$||f(\hat{y}_n) - f(\bar{y}_n)||_2 \leq L||\hat{y}_n - \bar{y}_n||_2 \tag{3.30}$$

so

$$||\hat{y}_{n+1} - y_{n+1}||_2 \leq \frac{3}{4}hL||\hat{y}_n - \bar{y}_n||_2 \tag{3.31}$$

Thus,

$$||\hat{y}_{n+1} - y_{n+1}||_2 \leq O(h^4) \tag{3.32}$$

Since an $O(h)$ approximation of $f'$ is given by $f' = (f_n - f_{n-1})/h$, one can approximate $k_2$ as follows:

$$\begin{aligned} k_2 &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}h^2 f') \\ &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}h^2(f_n - f_{n-1})/h) \\ &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}h(f_n - f_{n-1})) \end{aligned} \tag{3.33}$$

Since $f_n$ is calculated in the current step in the evaluation of $k_1$, one only has to store the previous value, $f_{n-1}$. In effect, by using previous values for the approximation, one has created a multistep Runge-Kutta method.

### 3.3 Fourth-order method

For the fourth-order formula let

$$y_{n+1} = y_n + b_1k_1 + b_2k_2 + b_3k_3 \tag{3.34}$$

and

$$\begin{aligned} k_1 &= hf(y_n) \\ k_2 &= hf(y_n + a_{21}k_1 + ha_{22}f_yk_1) \\ k_3 &= hf(y_n + a_{31}k_1 + a_{32}k_2 + ha_{33}f_yk_1 + ha_{34}f_yk_2) \end{aligned} \tag{3.35}$$

The following system of equations can be shown to solve the Taylor series

expansion of the above

$$
\begin{aligned}
&b_1 + b_2 + b_3 = 1 \\
&b_2 a_{21} + b_3(a_{31} + a_{32}) = 1/2 \\
&b_2 a_{21}^2 + b_3(a_{31} + a_{32})^2 = 1/3 \\
&b_2 a_{21}^3 + b_3(a_{31} + a_{32})^3 = 1/4 \\
&b_2 a_{22} + b_3(a_{21} a_{32} + a_{33} + a_{34}) = 1/6 \\
&b_3(a_{21} a_{34} + a_{22} a_{32}) = 1/24 \\
&b_2 a_{21} a_{22} + b_3[a_{21} a_{32}(\tfrac{1}{2} a_{21} + a_{31} + a_{32}) + (a_{31} + a_{32})(a_{33} + a_{34})] = 1/6
\end{aligned}
\tag{3.36}
$$

However, in order to utilize $f' = f f_y$ fully, one must restrict the solution with $a_{34} = 0$. The general solution to the above system of equations (with $a_{34} = 0$) has been found and example solutions are shown in Table 3.1.

Table 3.1: Example of fourth-order autonomous solutions

| $b_1$ | $b_2$ | $b_3$ | $a_{21}$ | $a_{22}$ | $a_{31}$ | $a_{32}$ | $a_{33}$ |
|---|---|---|---|---|---|---|---|
| $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{2}{3}$ | 1 | $\frac{1}{2}$ | $\frac{3}{8}$ | $\frac{1}{8}$ | 0 |
| $\frac{1}{6}$ | $\frac{2}{3}$ | $\frac{1}{6}$ | $\frac{1}{2}$ | $\frac{1}{8}$ | -1 | 2 | $-\frac{1}{2}$ |
| $\frac{1}{6}$ | $\frac{2}{3}$ | $\frac{1}{6}$ | $\frac{1}{2}$ | $-\frac{1}{8}$ | 3 | -2 | $\frac{5}{2}$ |
| $\frac{1}{10}$ | $\frac{1}{2}$ | $\frac{2}{5}$ | $\frac{1}{3}$ | $\frac{1}{18}$ | $-\frac{25}{24}$ | $\frac{15}{8}$ | $-\frac{5}{18}$ |
| $\frac{1}{10}$ | $\frac{1}{2}$ | $\frac{2}{5}$ | $\frac{1}{3}$ | $-\frac{1}{6}$ | $\frac{35}{24}$ | $-\frac{5}{8}$ | $\frac{5}{6}$ |

### 3.4 Fifth-order method

For the fifth-order formula let

$$y_{n+1} = y_n + b_1 k_1 + b_2 k_2 + b_3 k_3 + b_4 k_4 \tag{3.37}$$

and

$$
\begin{aligned}
k_1 &= h f(y_n) \\
k_2 &= h f(y_n + a_{21} k_1 + h a_{22} f_y k_1) \\
k_3 &= h f(y_n + a_{31} k_1 + a_{32} k_2 + h a_{33} f_y k_1) \\
k_4 &= h f(y_n + a_{41} k_1 + a_{42} k_2 + a_{43} k_3 + h a_{44} f_y k_1)
\end{aligned}
\tag{3.38}
$$

The following system of equations can be shown to solve the Taylor series expansion of the above

$$
\begin{aligned}
&b_1 + b_2 + b_3 + b_4 = 1 \\
&b_2 a_{21} + b_3(a_{31} + a_{32}) + b_4(a_{41} + a_{42} + a_{43}) = 1/2 \\
&b_2 a_{21}^2 + b_3(a_{31} + a_{32})^2 + b_4(a_{41} + a_{42} + a_{43})^2 = 1/3 \\
&b_2 a_{22} + b_3(a_{21}a_{32} + a_{33}) + b_4(a_{21}a_{42} + a_{43}(a_{31} + a_{32}) + a_{44}) = 1/6 \\
&b_2 a_{21}^3 + b_3(a_{31} + a_{32})^3 + b_4(a_{41} + a_{42} + a_{43})^3 = 1/4 \\
&b_2 a_{21} a_{22} + b_3(\tfrac{1}{2}a_{21}^2 a_{32} + (a_{31} + a_{32})(a_{21}a_{32} + a_{33})) + \tfrac{1}{2}b_4(a_{21}^2 a_{42} \\
&+a_{43}(a_{31} + a_{32})^2 + 2(a_{41} + a_{42} + a_{43})(a_{21}a_{42} + (a_{31} + a_{32})a_{43} + a_{44})) = 1/6 \\
&b_3 a_{22} a_{32} + b_4(a_{21}a_{32}a_{43} + a_{22}a_{42} + a_{33}a_{43}) = 1/24 \\
&b_2 a_{21}^4 + b_3(a_{31} + a_{32})^4 + b_4(a_{41} + a_{42} + a_{43})^4 = 1/5 \\
&3b_2 a_{21}^2 a_{22} + b_3(a_{21}^3 a_{32} + 3(a_{31} + a_{32})^2(a_{21}a_{32} + a_{33})) + b_4(a_{21}^3 a_{42} \\
&+(a_{31} + a_{32})^3 a_{43} + 3(a_{41} + a_{42} + a_{43})^2(a_{21}a_{42} + (a_{31} + a_{32})a_{43} + a_{44})) = 7/20 \\
&b_3 a_{21}^2 a_{32}(a_{31} + a_{32}) + b_4(a_{41} + a_{42} + a_{43})(a_{21}^2 a_{42} + (a_{31} + a_{32})^2 a_{43}) = 1/15 \\
&\tfrac{1}{2}b_2 a_{22}^2 + b_3(a_{21}a_{32}(\tfrac{1}{2}a_{21}a_{32} + a_{22} + a_{33}) + a_{22}a_{32}(a_{31} + a_{32}) + \tfrac{1}{2}a_{33}^2) \\
&+b_4(\tfrac{1}{2}a_{21}^2(a_{32}a_{43} + a_{42}^2) + (a_{31} + a_{32})(a_{21}(a_{32}a_{43} + a_{42}a_{43}) + a_{43}(a_{33} + a_{44}) \\
&+\tfrac{1}{2}(a_{31} + a_{32})a_{43}^2) + a_{21}a_{42}(a_{22} + a_{44}) + (a_{21}a_{32}a_{43} + a_{22}a_{42} \\
&+a_{33}a_{43})(a_{41} + a_{42} + a_{43}) + \tfrac{1}{2}a_{44}^2) = 11/120 \\
&b_4 a_{22} a_{32} a_{43} = 1/120
\end{aligned}
\tag{3.39}
$$

The solution verified using the above system of equations is

$$
\begin{aligned}
k_1 &= hf(y_n) \\
k_2 &= hf(y_n + \frac{1}{3}k_1 + \frac{1}{18}hf_y k_1) \\
k_3 &= hf(y_n - \frac{152}{125}k_1 + \frac{252}{125}k_2 - \frac{44}{125}hf_y k_1) \\
k_4 &= hf(y_n + \frac{19}{2}k_1 - \frac{72}{7}k_2 + \frac{25}{14}k_3 + \frac{5}{2}hf_y k_1) \\
y_{n+1} &= y_n + \frac{5}{48}k_1 + \frac{27}{56}k_2 + \frac{125}{336}k_3 + \frac{1}{24}k_4
\end{aligned}
\tag{3.40}
$$

Additional solutions can be found in Table 3.2.

Table 3.2: Example of fifth-order autonomous solutions

| | | | |
|---|---|---|---|
| $b_1$ | 1/24 | 5/54 | 1/14 |
| $b_2$ | 125/336 | 250/567 | 32/81 |
| $b_3$ | 27/56 | 32/81 | 250/567 |
| $b_4$ | 5/48 | 1/14 | 5/54 |
| $a_{21}$ | 1/5 | 3/10 | 1/4 |
| $a_{22}$ | 1/50 | 9/200 | 1/32 |
| $a_{31}$ | -52/27 | -9/8 | -329/250 |
| $a_{32}$ | 70/27 | 15/8 | 252/125 |
| $a_{33}$ | -8/27 | -9/32 | -259/1000 |
| $a_{41}$ | 43/5 | 17/3 | 209/35 |
| $a_{42}$ | -64/7 | -490/81 | -32/5 |
| $a_{43}$ | 54/35 | 112/81 | 10/7 |
| $a_{44}$ | 13/10 | 23/18 | 11/10 |

## 3.5 Non-autonomous derivations

If we proceed as above for $y' = f(x, y)$ we need to augment the terms involving $f_y k_i$ with $h f_x$.

### 3.5.1 Third-order method

A third-order method with two functional evaluations would have

$$y_{n+1} = y_n + b_1 k_1 + b_2 k_2 \tag{3.41}$$

where

$$\begin{aligned} k_1 &= h f(x_n, y_n), \\ k_2 &= h f(x_n + h c_{21}, y_n + a_{21} k_1 + h a_{22}(f_y(x_n, y_n) k_1 + h f_x(x_n, y_n))). \end{aligned} \tag{3.42}$$

Again, utilizing Taylor series expansion techniques, the above is satisfied as follows

$$\begin{aligned} k_1 &= h f(x_n, y_n) \\ k_2 &= h f(x_n + \frac{2}{3}h, y_n + \frac{2}{3}k_1 + \frac{2}{9}h(f_y(x_n, y_n)k_1 + h f_x(x_n, y_n))) \\ y_{n+1} &= y_n + \frac{1}{4}k_1 + \frac{3}{4}k_2 \end{aligned} \tag{3.43}$$

Here $hf' = f_y(x_n, y_n)k_1 + hf_x(x_n, y_n)$ and the method is the similar to the method derived for the autonomous case.

### 3.5.2 Fourth-order method

Similarly, for a single non-autonomous equation, the fourth-order method is derived by letting

$$y_{n+1} = y_n + b_1k_1 + b_2k_2 + b_3k_3 \tag{3.44}$$

where

$$\begin{aligned} k_1 &= hf(x_n, y_n) \\ k_2 &= hf(x_n + hc_{21} + h^2c_{22}f_y, y_n + a_{21}k_1 + ha_{22}f_yk_1) \\ k_3 &= hf(x_n + hc_{31} + h^2c_{32}f_y, y_n + a_{31}k_1 + a_{32}k_2 + ha_{33}f_yk_1 + ha_{34}f_yk_2) \end{aligned} \tag{3.45}$$

The following system of equations can be shown to solve the Taylor series expansion of the above:

$$\begin{aligned} &b_1 + b_2 + b_3 = 1 \\ &b_2c_{21} + b_3c_{31} = 1/2 \\ &b_2c_{21}^2 + b_3c_{31}^2 = 1/3 \\ &b_2c_{21}^3 + b_3c_{31}^3 = 1/4 \\ &b_2c_{22} + b_3[c_{21}a_{32} + c_{32}] = 1/6 \\ &b_3c_{21}c_{31}a_{32} = 1/8 \\ &2b_2c_{21}c_{22} + b_3[c_{21}^2a_{32} + 2c_{31}c_{32}] = 1/12 \\ &b_3[c_{22}a_{32} + c_{21}a_{34}] = 1/24 \\ &c_{31} = a_{31} + a_{32} \\ &c_{32} = a_{33} + a_{34} \\ &c_{21} = a_{21} \\ &c_{22} = a_{22} \end{aligned} \tag{3.46}$$

Solutions to this system of equations have been found and example solutions are shown in Table 3.3.

Table 3.3: Example fourth-order non-autonomous solutions

| $b_1$ | $b_2$ | $b_3$ | $c_{21}$&$a_{21}$ | $c_{22}$&$a_{22}$ | $c_{31}$ | $c_{32}$ | $a_{31}$ | $a_{32}$ | $a_{33}$ | $a_{34}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1/6 | 1/6 | 2/3 | 1 | -1/2 | 1/2 | 0 | 1/8 | 3/8 | -1/4 | 1/4 |
| 1/6 | 2/3 | 1/6 | 1/2 | 3/32 | 1 | -1/8 | -1/2 | 3/2 | -11/32 | 7/32 |
| 1/10 | 1/2 | 2/5 | 1/3 | 1/90 | 5/6 | 5/90 | -7/24 | 9/8 | -53/180 | 7/20 |

### 3.6 Stability analysis

If the step size used is too small, excessive computation time and errors occur. On the other hand, if the step size is too large, the method can become numerically unstable, and the numerical solution produced no longer corresponds qualitatively with the exact solution. Stability analysis is important for studying the growth of numerical errors in a computed solution to a differential equation. Stability analysis of Runge-Kutta and other numerical methods is carried out using a "test equation", $y' = \lambda y$, where $\lambda$ is a complex constant. Then the region of stability is defined to be the set of values of $h$ and $\lambda$ for which a perturbation in a single value $y_n$ will produce a change in subsequent values which does not increase from step to step.

If one attempts to solve the standard test problem using the autonomous third-order method presented here while using a functional evaluation of $f_y$, one obtains the following

$$\begin{aligned} k_1 &= hf(y_n) \\ k_2 &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}hf_y k_1) \\ y_{n+1} &= y_n + \frac{1}{4}k_1 + \frac{3}{4}k_2 \end{aligned} \tag{3.47}$$

since $y' = f = \lambda y$ and $f_y = \lambda$

$$\begin{aligned} k_1 &= h\lambda y_n \\ k_2 &= h\lambda(y_n + \frac{2}{3}h\lambda y_n + \frac{2}{9}h\lambda h\lambda y_n) \end{aligned} \tag{3.48}$$

or

$$k_2 = h\lambda y_n + \frac{2}{3}h^2\lambda^2 y_n + \frac{2}{9}h^3\lambda^3 y_n$$

thus

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{4}h\lambda y_n + \frac{3}{4}(h\lambda y_n + \frac{2}{3}h^2\lambda^2 y_n + \frac{2}{9}h^3\lambda^3 y_n) \\ &= y_n + h\lambda y_n + \frac{1}{2}h^2\lambda^2 y_n + \frac{1}{6}h^3\lambda^3 y_n \\ &= y_n(1 + h\lambda + \frac{1}{2}h^2\lambda^2 + \frac{1}{6}h^3\lambda^3) \end{aligned} \tag{3.49}$$

or

$$y_{n+1}/y_n = 1 + h\lambda + \frac{1}{2}h^2\lambda^2 + \frac{1}{6}h^3\lambda^3$$

Letting $z = h\lambda$ and $w = y_{n+1}/y_n = y_n/y_{n-1}$, then

$$w = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 \tag{3.50}$$

or

$$w - (1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3) = 0 \tag{3.51}$$

This is the same stability equation as the classical third-order Runge-Kutta method. In fact, when using a functional evaluation of $f_y$, all of the new methods presented here have the same stability as the classical Runge-Kutta method of equivalent order.

However, the region of stability shrinks if one uses an approximation to $f_y$. For example, using a linear approximation to $f_y$

$$\begin{aligned} k_1 &= hf(y_n) \\ k_2 &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}h(f_n - f_{n-1})) \\ y_{n+1} &= y_n + \frac{1}{4}k_1 + \frac{3}{4}k_2 \end{aligned} \tag{3.52}$$

Then, using $f = \lambda y$

$$\begin{aligned} k_2 &= hf(y_n + \frac{2}{3}k_1 + \frac{2}{9}h(f_n - f_{n-1})) \\ &= hf(y_n + \frac{2}{3}h\lambda y_n + \frac{2}{9}h(\lambda y_n - \lambda y_{n-1})) \\ &= h\lambda(y_n + \frac{2}{3}h\lambda y_n + \frac{2}{9}h(\lambda y_n - \lambda y_{n-1})) \end{aligned} \tag{3.53}$$

and

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{4}k_1 + \frac{3}{4}k_2 \\ &= y_n + \frac{1}{4}h\lambda y_n + \frac{3}{4}h\lambda(y_n + \frac{2}{3}h\lambda y_n + \frac{2}{9}h(\lambda y_n - \lambda y_{n-1})) \\ &= y_n + h\lambda y_n + \frac{1}{2}h^2\lambda^2 y_n + \frac{1}{6}h^2\lambda^2 y_n - \frac{1}{6}h^2\lambda^2 y_{n-1} \\ &= y_n + h\lambda y_n + \frac{2}{3}h^2\lambda^2 y_n - \frac{1}{6}h^2\lambda^2 y_{n-1} \end{aligned} \tag{3.54}$$

or

$$\begin{aligned} y_{n+1}/y_n &= 1 + h\lambda + \frac{2}{3}h^2\lambda^2 - \frac{1}{6}h^2\lambda^2 y_{n-1}/y_n \\ w &= 1 + h\lambda + \frac{2}{3}h^2\lambda^2 - \frac{1}{6}h^2\lambda^2/w \\ w &= 1 + z + \frac{2}{3}z^2 - \frac{1}{6}z^2/w \\ w^2 &= (1 + z + \frac{2}{3}z^2)w - \frac{1}{6}z^2 \\ w^2 &- (1 + z + \frac{2}{3}z^2)w + \frac{1}{6}z^2 = 0 \end{aligned} \tag{3.55}$$

Figure 3.1 compares the stability region of the new third-order method using a linear approximation to $f'$ with classical Runge-Kutta methods. One can see that the stability region is smaller than the region of both third and second-order Runge-Kutta methods.

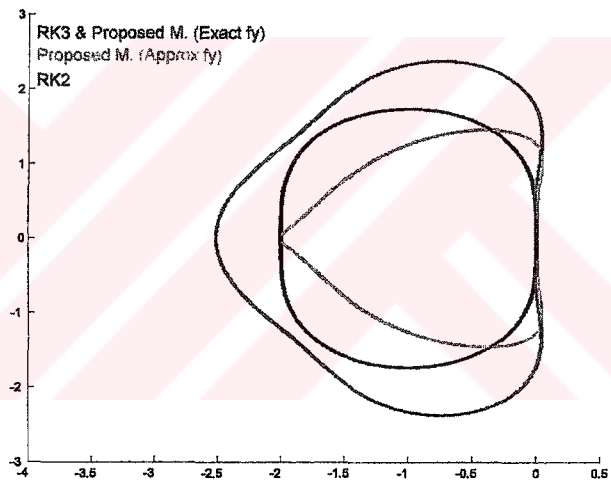


Figure 3.1: Stability vs. Runge-Kutta

Since using previous values for the approximation has created a multistep Runge-Kutta method, it is interesting to compare this method against other multistep methods. Figure 3.2 compares the new third-order method using a linear approximation to $y''$ with the multistep Adams-Bashforth methods. The figure shows that the region of stability for the new method is larger than that of the Adams-Bashforth methods.

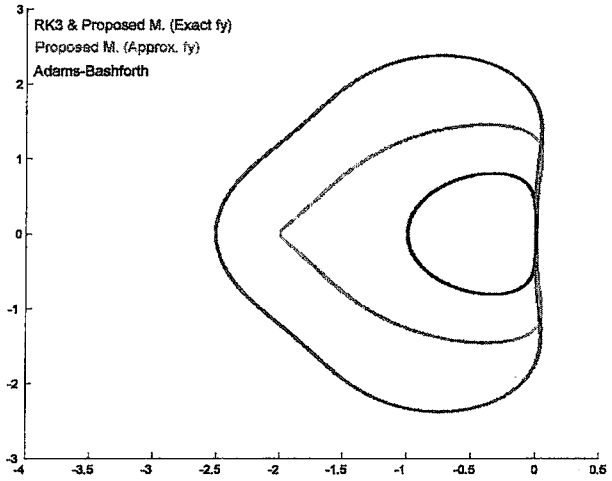


Figure 3.2: Stability vs. Adams-Bashforth

### 3.7 Numerical Results

To verify that the new methods are of the order claimed, the nonlinear autonomous equation

$$y' = y - \frac{y^2}{40} \qquad y(0) = 1 \tag{3.56}$$

whose exact solution is $y(x) = 40/1 + 39e^{(-x)}$, has been solved using the new fourth-order method and classical fourth-order Runge-Kutta methods. Both the exact and an approximated solution to $y''$ were computed. Results are comparable to a classical Runge-Kutta solution of equal order, thus demonstrating the claim.
It is applied three times these formulas with $h = 1$, writing the results with four decimal places. Results show the accuracy of the new method both the exact and approximated solution to $f_y$.

Table 3.4: Comparing methods and the exact solution with $h = 1$

| $x_n$ | Classical Runge-Kutta Method | Proposed Method Exact $f_y$ | Proposed Method Approximating $f'$ | $\frac{40}{1+39e^{-x}}$ |
|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 1 |
| 1 | 2.5987 | 2.5998 | 2.6050 | 2.6063 |
| 2 | 6.3414 | 6.3483 | 6.3059 | 6.3714 |
| 3 | 13.5333 | 13.5482 | 13.4500 | 13.5976 |

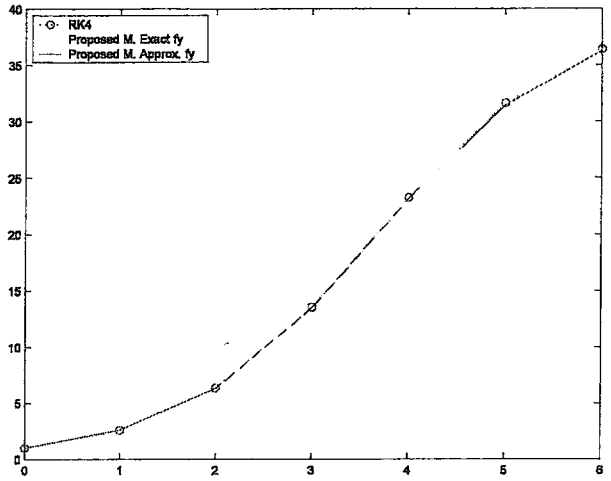


Figure 3.3: Equation $y' = y - \frac{y^2}{40}$

Table 3.5: Error in methods for $y' = y - \frac{y^2}{40}$ with $y(0) = 1$

| $x_n$ | Classical Runge-Kutta Method | Proposed Method Exact $f_y$ | Proposed Method Approximating $f'$ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 1 | 0.0076 | 0.0065 | 0.0013 |
| 2 | 0.0300 | 0.0231 | 0.0655 |
| 3 | 0.0642 | 0.0495 | 0.1476 |