LINEAR SYSTEMS WITH TIME VARYING PARAMETERS

by

James A. Bautz

A Thesis Submitted to the Faculty of the

DEPARTMENT OF ELECTRICAL ENGINEERING

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

1964

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: James A. Batty

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Donald G. Schultz — Oct 24, 1964

DONALD G. SCHULTZ — Date

Associate Professor of Electrical Engineering

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Dr. Donald G. Schultz for his patient guidance in the planning and completion of this thesis.

TABLE OF CONTENTS

ABSTRACT

An analysis is made of various methods that have been developed to solve linear differential equations with time varying parameters and how these techniques apply to the solution of Control System problems. The various techniques are evaluated to determine their range of application and relative worth when applied to a particular system. In this way criteria are established to determine which of the various techniques will be most advantageous in solving a particular problem.

CHAPTER 1

INTRODUCTION

The type of ordinary, linear differential equations with constant coefficients has been studied in great detail by many individuals, and a good number of techniques have been developed to solve them. The situation is altered considerably when the coefficients of the equation variables become time varying also. The once useful methods of time invarient differential equation solution become difficult with which to work. Consequently, attempts have been made to seek new approaches to obtain solutions.

In an effort to find some convenient means of attacking the problem, investigators have tried both analytical and numerical methods of solution. The purpose of the ensuing chapters will be to discuss each of these areas in sufficient detail to give a general picture of various numerical and analytical techniques that have been proposed for solving time varying differential equations, along with their good and bad points. The ultimate objective is to choose one or more methods that are useful in solving a wide range of problems in a relatively simple manner. Analytical methods are discussed first, followed by a chapter on numerical methods. Finally, the entire paper is summed up and conclusions are drawn, based on the presented facts. The following paragraphs give a general idea of the material to be covered in each chapter.

Chapter 2 deals with analytical methods, those that seek to apply the standard theory for the solution of differential equation with constant coefficients to the time varying case. Both time and frequency domain approaches are presented.

The time domain approach involves breaking the answer up into a complementary function and a particular integral and finding the solution for each. However, even though the solution to a problem may exist, it is sometimes extremely difficult to determine by means of this approach due to the complicated integrals that are encountered. Approximation techniques have been developed to get around this problem. Some of these are presented to demonstrate how the time domain approach may be used.

Various transform techniques are investigated as well. The Laplace Transform, which has been used so effectively in working with time invarient equations, is discussed, along with other transform attempts in an effort to point out their strengths and weaknesses.

Chapter 3 discusses various numerical methods that are available for the solution of time varying differential equations. Unlike analytical methods, which supply a mathematical expression as the solution to a problem, numerical techniques yield a series of points which give a graphical representation of the answer. When time varying equations are involved, the varying element is approximated in a stepwise manner with a change occurring each time a point is calculated. The exactness of this approximation depends on the length of time between points.

General background theory for numerical methods is represented first, along with several elementary techniques that are useful in hand calculating problems of low order. The selection of a particular method depends on the length of time one is willing to spend, and the accuracy desired.

The solutions to higher order equations can also be determined by hand calculation; however, the amount of labor involved is prohibitive. Therefore, three methods are presented which are readily adapted to use on a digital computer.

The mathematics involved is quite simple, and the repetitive nature of the calculation necessary to obtain the desired points makes them ideal for solution by digital machine.

Chapter 4 is devoted to the actual application of a sample problem to a digital computer. The steps involved in determining the solution of the problem will be outlined followed by an explanation of the program that translates the mathematics into language which can be understood by the computer.

Finally, the results are summed up and conclusions are drawn.

CHAPTER 2

ANALYTICAL TECHNIQUES

2.1 Introduction

Much work has been done on the solution of ordinary linear differential equations with constant coefficients by analytical techniques. Books have been written describing methods that are applicable for solving equations in the time domain. Solution in the frequency domain can be handled very well by the familiar Laplace Transform.

These well-known and developed ideas have been put to work in an attempt to solve ordinary linear differential equations with time varying coefficients. It has been found that the presence of variable elements changes the complexion of the problem. By presenting various methods that have been used to solve time varying equations, Chapter 2 shows some of the problems involved in applying these analytic techniques.

The time domain approach is presented first, followed by the frequency domain or transform approach. In each case, various methods are covered in some detail to reveal the actual complexity that is encountered.

2.2 Classical Solution

The term classical solution is used here to refer to those methods of solution of linear differential equations with time varying parameters that are a reapplication of the time domain techniques

used to solve linear differential equations with constant coefficients. Included in this set are some approximate solutions that have been developed to make computation easier.

When one is confronted with the problem of solving a linear differential equation with time varying coefficients, the natural impulse is to seek to apply the theory that has already been developed for the solution of linear differential equations with constant coefficients. Theoretically, exact solutions for equations can be found in this way; however, "in most cases details of finding the solution become exceedingly complicated." (1)

By way of example, consider the following standard form for a first order differential equation,

$$\frac{dx}{dt} + P(t)x = Q(t) \qquad (2.1)$$

where $Q(t)$ and $P(t)$ are functions of time.

The standard form for the solution of this equation is known to be

$$x = Ce^{-\int P(t)dt} + e^{-\int P(t)dt} \int Q(t)e^{\int P(t)dt}\, dt \qquad (2.2)$$

where C is a constant determined by initial conditions.

The above solution is not always difficult to apply, although the calculation may become involved depending on the problem in question. It can be used to solve problems involving current or charge distribution in time varying circuits. Practical examples would include the case of a carbon microphone or condenser transmitter circuit.

Solved examples of such circuits may be found in a book by Cunningham(1) and a paper by Pipes (2).

Since most problems of practical interest involve equations of higher order than first, the classical solution to second order problems is the next item of interest (1).

The general equation

$$\ddot{x} + 2P(t)\dot{x} + R^2(t)x = Q(t) \tag{2.3}$$

is a second order equation with a forcing function. Its solution consists of a complementary function and a particular integral.

When the equation is solved for the complementary function, it might appear as

$$x_c = Ax_1(t) + Bx_2(t) \tag{2.4}$$

where $x_1(t)$ and $x_2(t)$ are the independent parts required for a second order equation and A and B are arbitrary constants.

In this case, the method of variation of parameters will be used to determine the particular integral. This method is "based on the observation that the particular integral of a linear equation can be written as the product of one or more parts of the complementary function and one or more functions of the independent variable." (1)

First, reduce Eq. 2.3 to a set of first order equations by letting $z = \dot{x}$.

$$\begin{aligned} \dot{x} &= z \\ \dot{z} &= -2P(t)z - R^2(t)x + Q(t) \end{aligned} \tag{2.5}$$

The complementary function, Eq. 2.4, can be written

$$x = Ax_1 + Bx_2 \qquad (2.6)$$
$$z = A\dot{x}_1 + B\dot{x}_2$$

The particular integral has the form of Eq. 2.6 except that the terms A and B are functions of time. With this in mind, substitute Eq. 2.6 into Eq. 2.5. Manipulation of the equations yields

$$\dot{A}x_1 + \dot{B}x_2 = 0$$
$$\dot{A}\dot{x}_1 + \dot{B}\dot{x}_2 = Q(t) \qquad (2.7)$$

If the equations are solved simultaneously for $\dot{A}$ and $\dot{B}$, the result is

$$\dot{A} = \frac{-Qx_2}{x_1\dot{x}_2 - \dot{x}_1x_2}$$
$$\dot{B} = \frac{Qx_1}{x_1\dot{x}_2 - \dot{x}_1x_2} \qquad (2.8)$$

Integrating these values and substituting into Eq. 2.4 yields

$$x = \left(-\int \frac{Qx_2}{x_1\dot{x}_2 - \dot{x}_1x_2}\,dt + C_1\right)x_1 + \left(\int \frac{Qx_1}{x_1\dot{x}_2 - \dot{x}_1x_2}\,dt + C_2\right)x_2 \qquad (2.9)$$

where C_1 and C_2 are arbitrary constants.

The evaluation of the integrals can become quite complex. This can be demonstrated by sample problems (1).

In order to make computation easier, approximate solutions have been developed to cover problems of a special type. For instance, a

technique has been developed that applies to time varying equations in which the varying coefficient changes by a small amount about a large mean value. It is known as the WKBJ approximation (1) (2).

The equation to which the WKBJ approximation applies is of the form

$$\frac{d^2x}{dt^2} + G^2(t)x = 0 \tag{2.10}$$

where $G^2(t)$ includes the varying coefficient that has a large mean value with small variation. No forcing function exists in this equation. If there is a forcing function, the variation of parameters method can be used to find the particular integral. The solution to Eq. 2.10 would be a complementary integral.

The WKBJ approximation comes from a combination of the names Wentzel, Kramers, Brillouin, and Jeffreys. The approximation solution has the form

$$x = G(t)^{-\frac{1}{2}}\left[C_1 e^{j\phi(t)} + C_2 e^{-j\phi(t)}\right] \tag{2.11}$$

where $\phi(t) = \int G(t)dt$ and j is the imaginary unit. If Eq. 2.11 is differentiated twice, it can be shown to satisfy the following equation,

$$\frac{d^2x}{dt^2} + \left[G^2 + \frac{\ddot{G}}{2G} - \frac{3}{4}\left(\frac{\dot{G}}{G}\right)^2\right]x = 0 \tag{2.12}$$

where $\dot{G} = \frac{dG}{dt}$ and $\ddot{G} = \frac{d^2G}{dt^2}$

If $G^2(t)$ is such that

$$G^2 >> \left| \frac{\ddot{G}}{2G} - \frac{3}{4}\left(\frac{\dot{G}}{G}\right)^2 \right|$$

then Eqs. 2.10 and 2.12 have essentially the same form, and Eq. 2.11 is an approximate solution for Eq. 2.10.

Finally, if $G(t)$ and $\phi(t)$ are both real, Eq. 2.11 may be written in the following form.

$$x = G(t)^{-\frac{1}{2}} \left[A \cos \phi(t) + B \sin \phi(t)\right] \qquad (2.13)$$

where A and B are constants and $\phi(t) = \int G(t)dt$. Equation 2.13 then is the WKBJ approximation for Eq. 2.10. Examples of this technique are given by Cunningham and Pipes.

The use of this approximation, however, does not alleviate the situation to the point where problems can be worked with ease (1). Complicated, if not impossible, integrals are encountered even in apparently simple problems. For this reason, the classical approach in the time domain is not often used in practice today. Investigators have turned their attention away from this method toward something that gives more promise.

2.3 Transform Methods

Under this category are included both attempts to utilize existing transform methods as well as develop new ones. This section covers three approaches along this line, starting with the most familiar, the Laplace Transform.

Since Laplace Transforms have proved so useful in the solution of constant coefficient differential equations, it is logical to determine the efficacy of the Laplace Transform approach in dealing with time varying equations.

The first concern is the development of the transforms. Since no tables exist specifically for time varying cases, they have to be developed by the individual solving the problem.

The defining equation for the Laplace Transform assumes the form

$$\mathcal{L}\{g(t)\ f(t)\} = \int_0^\infty e^{-st}\ g(t)\ f(t)\ dt$$

This may not be difficult to evaluate in such familiar cases as $tf(t)$ and $e^{-\alpha t} \sin \alpha t$, but, conceivably, cases could arise in which the evaluation of the integral would prove difficult.

However, the actual determination of the transform does not put an end to the difficulty involved. It happens that the Laplace Transform of a time varying equation leads to a differential equation in s which may be nonlinear. Thus, the original problem is transformed into another problem of equal or greater magnitude.

As an example, consider the following equation (3).

$$t\frac{d^2x}{dt^2} + \frac{dx}{dt} + tx = 0 \qquad (2.14)$$

where $x = x(t)$

Taking the Laplace Transform of both sides yields

$$-\frac{d}{ds}\left[s^2x(s) - sx(0) - x'(0)\right] + sx(s) - x(0) - \frac{d}{ds}x(s) = 0 \quad (2.15)$$

If $x'(0) = 0$, this reduces to

$$(s^2 + 1)x'(s) + sx(s) = 0 \quad (2.16)$$

Equation 2.16 is a nonlinear differential equation. Thus, nothing has been gained by using Laplace Transforms.

The solution to Eq. 2.16 is

$$x(s) = \frac{k}{\sqrt{s^2 + 1}} \quad (2.17)$$

where k is a constant of integration.

Equation 2.17 may be written as

$$x(s) = ks^{-1}\left(1 + \frac{1}{s^2}\right)^{-\frac{1}{2}} \quad (2.18)$$

or

$$x(s) = k\sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} \frac{(-1)^n}{s^{2n+1}}$$

Taking the inverse transform term by term, the result is

$$x(t) = k\sum_{n=0}^{\infty} \frac{(2n)!\,(-1)^n}{2^{2n}(n!)^2} L^{-1}\left[\frac{1}{s^{2n+1}}\right] \quad (2.19)$$

or

$$x(t) = k\sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{2^{2n}(n!)^2}$$

The amount of work involved here may be considerable. Since by this method, one problem, perhaps more difficult than the original, is simply substituted for another, there does not seem to be any point in pursuing this line further.

J. A. Aseltine has developed a general transform approach to solve time varying equations in which a transform can be developed to handle each type of problem encountered (4). It is possible by this method to generate both the Laplace and Mellin transforms. Although there is a number of situations in which this method can be used, "the complications for equations higher than second order usually make other methods more attractive." (4)

The integral transformation is of the form

$$L\{q(t)\} = Q(\gamma) = \int_0^\infty q(t)\, h(\gamma,t)dt \qquad (2.20)$$

where $h(\gamma,t)$, the kernel, is a function of time and the transformation variable γ. $Q(\gamma)$ is the transform of $q(t)$.

A requirement is placed upon the transformation to insure that it is an algebraic equation in the transform of the dependent variable, $Q(\gamma)$. Thus, in working with an equation of the form

$$a(t)q'' + b(t)q' + d^2q = e(t) \qquad (2.21)$$

where $a(t)$, $b(t)$ and $e(t)$ are functions of time, and d^2 is a constant, the transformation should be of the form

$$f(\gamma)\, Q(\gamma) + d^2Q(\gamma) = E(\gamma) + (\text{initial conditions}) \qquad (2.22)$$

where $f(\rho)$ is an arbitrary function of the transform variable. The nature of the resultant transform depends on the function of ρ that is selected.

Once the kernel $h(\rho,t)$ is evaluated, the nature of the transform is determined. For instance, if $h(\rho,t) = e^{-\rho t}$, then Eq. 2.20 has the form

$$L\{q(t)\} = \int_0^\infty q(t)e^{-\rho t}\,dt$$

which is the familiar Laplace Transform.

The disadvantages of this technique are the fact that it is only good for problems of first or second order, and it requires that a new table of transforms be developed for each problem. The lack of interest by other investigators is obvious for the method has not been developed further in the literature since its proposal.

A later paper by F. R. Gerardi (5) suggests the use of Mellin Transforms as a means of attacking the problem. It can be used similarly to the Laplace Transform.

The defining equation is

$$F(s) = \int_0^\infty f(t)t^{s-1}\,dt = M\left[f(t)\right]$$

where

s = complex variable of Mellin Transform

M = Mellin Transform operation

t = time

The inversion integral is

$$f(t) = \frac{1}{2\pi i} \int_{c-j\infty}^{c+j\infty} F(s)t^{-s}\, ds = M^{-1}\left[F(s)\right]$$

where M^{-1} = inverse Mellin Transform.

Since the actual application of the Mellin Transform is similar to the Laplace Transform, the details of its use are not covered here. Transform tables can be built up from the defining equation and applied in the usual way.

The difficulty encountered in the application of this method is apparent upon reading the paper by Gerardi who restricts the problems to Euler-Cauchy differential equations which have the form,

$$\sum_{n=0}^{N} A_n t^n \frac{d^n x}{dt^n} = f(t)$$

Due to its limited scope, the Mellin Transform does not appear to be worth further consideration.

2.4 Conclusion

This chapter examines both time and frequency domain solutions to time varying equations. Practically all of them are found to be applicable only to first or second order equations. In most cases the amount of labor involved in the solution of even a second order equation is not only great but sometimes of a complicated nature.

None of the methods mentioned in this section are in general use by people working with time varying problems. The work that must

be done to get an answer and the lack of generality inherent in these analytical techniques have led investigators to find new approaches to the problem. Up to the present time, the search has centered around some sort of numerical rather than analytical technique.

CHAPTER 3

NUMERICAL METHODS

3.1 Introduction

The methods that were discussed previously were found lacking to such an extent that it is logical to abandon them and move on to a different area of study. This new area includes a group of techniques which are numerical rather than analytical in nature and are referred to by the general title, numerical methods.

Numerical methods do not give quick, easy solution; however, even though much calculation is involved, the mathematics is extremely simple, and solving a problem becomes a question of repeating a procedure over and over again. Also, some of these methods have a means of checking for arithmetic mistakes.

Unlike the analytical approach, a numerical solution is not a mathematical statement, but, rather, a series of points which, when connected, produces a curve that describes the way in which the variable in question behaves. The solution to the equation is found at each of these points, hereafter referred to as sample points. The time interval between each of the points will be known as the sampling period or sampling interval.

When a time varying coefficient is involved, the variation is approximated by assuming that it is constant during the sampling period and changes at each sample point. The closer the sample points are,

the better the approximation is. One way to check if the sampling interval is small enough (11) is to do the problem twice, using a shorter sampling interval the second time. If no discernible change occurs, then it can be assumed that the interval chosen is good enough for the problem. This is simply a brute-force method for determining the sampling interval; however, there is no convenient analytical way of determining this value.

This chapter deals first with numerical methods that can be computed by hand or with the help of a desk calculator. These methods are applicable to the solution of low order problems and are included to give a more complete picture. Higher order problems require a more sophisticated approach; therefore, a few methods are presented which can be solved by means of a digital computer. These methods are then compared in order to provide a basis for Chapter 4, which will cover the actual writing of a program.

3.2 Basic Hand Techniques

There are many numerical methods which have been developed to solve differential equations. There is not room to present them all, and since most are variations of the same idea, the following discussion consists of a summary of the basic theory involved and a presentation of a few of the more common methods.

Basically, these methods consist of utilization of information concerning one or more points on a curve, along with the slope between these points. By knowing something about these, it is possible to predict the location of the next point on the curve.

The Taylor series expansion (7) of a function is often used because it gives a relationship between the function at a point and the derivatives of a function at a previous point.

A Taylor series can be written as (10)

$$f(s) = \sum_{k=0}^{\infty} a_k (s - s_o)^k \qquad (3.1)$$

where

$$a_k = \frac{(f^{(k)}{}_{s_o})}{k!}$$

If this expression is expanded, the result is

$$f(s) = f(s_o) + \frac{(s-s_o)}{1!} f'(s_o) + \frac{(s-s_o)^2}{2!} f''(s_o) + \ldots + \frac{(s-s_o)^n}{n!} f^n(s_o) \qquad (3.2)$$

If Eq. 3.2 is written in terms of x and y coordinates, the result is

$$y_{n+1} = y_n + h \frac{dy_n}{dx} + \frac{h^2}{2!} \frac{d^2 y_n}{dx^2} + \frac{h^3}{3!} \frac{d^3 y_n}{dx^3} + \ldots \qquad (3.3)$$

where $h = (s-s_o)$ is the sampling interval, y_{n+1} is the point to be calculated, and y_n is a known point upon which the calculation is based. The main sources of errors in predicting points by series methods comes from:

(1) Inherited error - produced in the calculation of y_{n+1} due to some error in the calculation of y_n

(2) Truncation error - due to cutting off the higher order terms of the series.

The simplest numerical method, known as the Euler Method (10), uses only the first two terms of this series to approximate the solution of the differential equation.

$$y_{n+1} = y_n + h\frac{dy_n}{dx} \tag{3.4}$$

If the initial point is known, an interval, h, can be selected and the value of $y'_n = \frac{dy_n}{dx}$ computed from the differential equation in question. Then y_{n+1} can be calculated from Eq. 3.4. This point becomes the y_n for the next computation.

The above method has limited practical use since the error increases exponentially. However, it is sometimes used to start computation of a solution since it requires that only one point be known. Later on it is seen that the knowledge of several points is necessary for more sophisticated methods. The Euler Method can be used to get this information, and another method used to complete the computation.

A better approximation is obtained quite simply by the following procedure (6) (7). The Taylor series given in Eq. 3.3 can be written

$$y_{n+1} = y_n + hy'_n + \frac{h^2}{2!}y''_n + \frac{h^3}{3!}y'''_n + \cdots \tag{3.5}$$

where $y'_n = \frac{dy_n}{dx}$.

The series for the point y_{n-1} will differ from Eq. 3.5 in that the sign of all the odd derivative terms will be negative. This is due to the fact that $(s-s_o)$ is less than zero. Therefore,

$$y_{n-1} = y_n - hy'_n + \frac{h^2}{2!} y''_n - \frac{h^3}{3!} y'''_n + \cdots \tag{3.6}$$

Subtracting Eq. 3.6 from Eq. 3.5 yields

$$y_{n+1} - y_{n-1} = 2hy'_n - \frac{h^3}{3!} y''_n + (\text{higher order terms}) \tag{3.7}$$

This may be rewritten as

$$y_{n+1} = y_{n-1} + 2hy'_n + (\text{higher order terms}) \tag{3.8}$$

If $h < 1$, this approximation is better than the previous one since the first term neglected is a third order rather than second order term.

Accuracy may be improved when using Eq. 3.8 by means of the trapezoidal formula, Eq. 3.9, which considers the average of the slopes of the curve before and after the point in question.

$$y_{n+1} = y_n + \frac{h}{2} (y'_{n+1} + y'_n) \tag{3.9}$$

When solving a problem using Eq. 3.8 and Eq. 3.9, the following procedure should be used:

(1) predict y_{n+1} from Eq. 3.5

(2) find y'_{n+1} from the differential equation in question

(3) Correct y_{n+1} by Eq. 3.6

(4) Correct y'_{n+1} by use of the differential equation

(5) Repeat steps (3) and (4) until no further changes occur.

The value of h should be chosen such that the above calculation converges to some value within the first one or two attempts.

Example 3.1. To illustrate the above procedure, consider the following differential equation:

$$y' = -2xy^2$$

Use the initial values $x_o = 0$, $y_o = 1$, and let $h = .1$.

In order to start the problem, use the first three terms of Eq. 3.3. This is necessary since information is available about one point only.

$$y_{n+1} = y_n + hy'_n + \frac{h^2}{2!} y''_n$$

Differentiate the problem to get y".

$$y'' = -2y^2 - 4xy\, y'$$

Evaluate y' and y" .

$$y' = 2(0)\,(1)^2 = 0$$

$$y'' = -2(1)^2 - 4(0)\,(1)\,(0) = -2$$

Therefore

$$y_1 = 1 + 0 \cdot h - h^2 = .990 \text{ at } x = .1 .$$

The next computation can be made using Eq. 3.8 since two points are now known.

$$y_{n+1} = y_{n-1} + 2hy'_n$$

From the given equation,

$$y'_1 = -2(.1)\,(.990)^2 = -.196$$

So $$y_2 = 1.0 + 2(.1)\,(-.196) = .961$$

The computation of y_3 proceeds as follows,

$$y_3 = y_1 + 2hy'_2$$

$$y_3 = .990 + 2(.1)\,(-2(.2)(.961)^2)$$

$$y_3 = .990 - .074 = .916.$$

In order to use the correction formula, the steps outlined previously will be followed,

$$y'_3 = -2(.3)(.916)^2 = .504$$

From the trapezoidal formula,

$$y_3 = .961 + \tfrac{1}{2}(-.504 - .370) = .917$$

Therefore, the corrected value of y'_3 is

$$y'_3 = -2(.3)(.917)^2 = .504$$

The two values of y'_3 agree so closely that the difference cannot be

detected on a slide rule. If greater accuracy is desired, a desk calculator can be used.

The data calculated thus far is tabulated below.

x	y	y'
0	1	0
.1	.990	-.196
.2	.962	-.370
.3	.917	-.506

The procedure can be carried on to give as many points as desired.

The procedures just presented are quite simple and can be derived directly from a Taylor series. There are many methods which are much more accurate than these. The development of these alternate methods is not nearly as simple as the ones given above. An example of one of these will be presented below to give an idea of a more sophisticated numerical method that can be computed by hand with relative ease.

The method to be discussed is known as the Milne Method (6) (8). It is based on a rather lengthy derivation, the salient features of which are shown below.

The roots of the Milne Method are found in Newton's Formula for Forward Interpolation (8). If $y = f(x)$ is a function which takes values $y_0, y_1, y_2, \ldots, y_n$ for the equidistant values $x_0, x_1, x_2, \ldots, x_n$ of the independent variable x, let $\phi(x)$ denote a polynomial of the nth degree,

$$\phi(x) = a_o + a_1(x-x_o) + a_2(x-x_o)(x-x_1) + \ldots$$
$$+ a_n(x-x_o)(x-x_1)(x-x_2) \ldots (x-x_{n-1}) \qquad (3.10)$$

Determine the coefficients a_o, a_1, a_2,...,a_n to make

$$\phi(x_o) = y_o,\ \phi(x_1) = y_1,\ldots,\phi(x_n) = y_n$$

Therefore, substitute the successive values x_o, x_1,...,x_n into Eq. 3.10 and put $\phi(x_o) = y_o \ldots \phi(x_n) = y_n$ to determine the various values of a. Remember that $(x_1-x_o) = h$ and $(x_2-x_o) = 2h$, etc. This gives

(1) $y_o = a_o$ or $a_o = y_o$

(2) $y_1 = a_o + a_1 (x_1 - x_o) = y_o + a_1 h$

Therefore, $a_1 = \frac{y_1-y_o}{h} = \frac{\Delta y_o}{h}$

(3) $y_2 = a_o + a_1(x_2-x_o) + a_2(x_2-x_o)(x_2-x_1)$

$$y_2 = y_o + \frac{y_1-y_o}{h}(2h) + a_2(2h)(h)$$

Therefore, $a_2 = \frac{y_2-2y_1+y_o}{2h^2} = \frac{\Delta^2 y_o}{2h^2}$

where $\Delta y_n = y_n - y_{n-1}$

and $\Delta^2 y_n = \Delta y_n - \Delta y_{n-1} = y_n - 2y_{n-1} + y_{n-2}$

Continuing in this manner, all the values a_o,...,a_n can be computed, giving as the final result

$$\phi(x) = y_o + \frac{\Delta y_o}{h}(x-x_o) + \frac{\Delta^2 y_o}{2h^2}(x-x_o)(x-x_1) \tag{3.11}$$

$$+ \frac{\Delta^3 y_o}{3!h^3}(x-x_o)(x-x_1)(x-x_2) + \cdots + \frac{\Delta^n y_o}{n!h^n}(x-x_o)(x-x_1)\cdots(x-x_{n-1})$$

Using the substitution

$$\frac{x-x_o}{h} = u$$

$$\phi(x) = y_o + u\Delta y + \frac{u(u-1)}{2!}\Delta^2 y_o + \frac{u(u-1)(u-2)}{3!}\Delta^3 y_o + \cdots$$

$$+ \frac{u(u-1)\cdots(u-n+1)}{n!}\Delta^n y_o \tag{3.12}$$

Equation 3.12 is the statement of Newton's Formula for Forward Interpolation.

In order to obtain the actual Milne Formula, integrate Eq. 3.12 from u = 0 to u = 4. If Eq. 3.12 is written in terms of y', the result is

$$(y_4-y_1) = h(4y_o' + 8\Delta y_o' + \frac{20}{3}\Delta^2 y_o' + \frac{8}{3}\Delta^3 y_o' + \frac{28}{90}\Delta^4 y_o') \tag{3.13}$$

Now remove the Δ^n terms by replacing them with their equivalent values, i.e., $\Delta y_o = y_o - y_o$, and so forth. The result is

$$(y_4-y_1) = \frac{4h}{3}(2y_1' - y_2' + 2y_3') + \frac{28}{90}h\Delta^4 y_o' \tag{3.14}$$

Equation 3.14 is Milne's Formula. It can also be written in a different manner,

$$y_{n+1} = y_{n-3} + \frac{4h}{3}(2y'_{n-2} - y'_{n-1} + 2y'_n) + (\text{remainder}) \qquad (3.15)$$

A checking formula can be developed quite easily to check for arithmetic mistakes and to provide a means of calculating an error compensation formula.

To obtain the checking formula, integrate eq. 3.12 over the range u = 0 to u = 2. The result is

$$(y_2 - y_o) = h(2y'_o + 2\Delta y'_o + \frac{1}{3}\Delta^2 y'_o - \frac{1}{90}\Delta^4 y'_o) \qquad (3.16)$$

Replacing the Δ^n values as before yields

$$y_2 - y_o = \frac{h}{3}(y'_o + 4y'_1 + y'_2) - \frac{h}{90}\Delta^4 y'_o \qquad (3.17)$$

The formula is more commonly written as

$$y_{n+1} = y_{n-1} + \frac{h}{3}(y'_{n-1} + 4y'_n + y'_{n+1}) \qquad (3.18)$$

Equations 3.15 and 3.18 give the pair of equations used in the Milne Method. By examining the error involved in using these equations, it is possible to develop a formula to compensate for the error at each point.

Let $y(1)_{n+1}$ and $y(2)_{n+1}$ represent the results of Eqs. 3.15 and 3.18, respectively. Then, since the remainder terms have been neglected (see Eqs. 3.14 and 3.17), a more exact value of y at $x = x_{n+1}$ is either

$$y = y(1)_{n+1} + \frac{28}{90} h \Delta^4 y'$$

or

$$y = y(2)_{n+1} - \frac{h}{90}\Delta^4 y'$$

Equating these values yields

$$y(1)_{n+1} + \frac{28}{90} h \Delta^4 y' = y(2)_{n+1} - \frac{h}{90} \Delta^4 y' \qquad (3.19)$$

$$y(1)_{n+1} - y(2)_{n+1} = \frac{-29}{90} h \Delta^4 y' = 29E \qquad (3.20)$$

where E is the principal part of the error in Eq. 3.13. Therefore,

$$E = \frac{1}{29} (y(1)_{n+1} - y(2)_{n+1}) \qquad (3.21)$$

If the value of Eq. 3.21 is large with respect to the desired accuracy, a smaller sampling interval should be chosen.

Now the Milne Method is complete. It has a fairly accurate formula and a method of checking for arithmetic mistakes and correcting for inherent error. This method, however, is not the most accurate of the numerical methods nor the most difficult to derive. It is presented here as an example of a reasonably accurate (by mathematicians' standards) and popular technique which is not merely a point slope method. There are probably as many numerical methods as there are men working in the field. The final choice of one depends not only on accuracy requirements, but also on length of computation involved, and on pure chance, when several equally acceptable formulas are available.

The numerical methods of the type just presented will give an accurate solution to time varying differential equations, and can be conveniently used on systems of low order when the number of points to be calculated is not too great. The application of these techniques to equations of second or higher order require that they be broken into

a set of first order equations in order to be solved. This could lead to a large amount of calculation, and, since much of this calculation is repetitive, the use of the digital computer has become widespread to carry out the tedious labor. The next section will deal with methods that can be conveniently programmed on a digital computer, thus allowing use of more involved formulas and the computation of a greater number of sample points.

3.3 Computer Methods

The introduction of the digital computer affected the field of numerical methods significantly. Techniques which had fallen into disuse due to the amount of computation involved were resurrected and proved to be quite valuable. Also, new techniques that lend themselves to computer solution were developed.

This section deals with methods of solving differential equations with time varying coefficients that can be programmed for solution by a digital computer. Although these methods can be calculated by hand, the amount of labor involved is enormous even for low order problems. The greatest usefulness of these techniques is realized in the solution of higher order equations, which are difficult to solve using the methods presented previously.

The methods in question are the Runge-Kutta, Simpson's Rule, Z Form, and State Variable. The theory behind these methods is discussed, followed where practical, by a sample problem as a means of demonstration.

3.3.1 Runge-Kutta Method and Simpson's Rule

The Runge-Kutta Method (6) (7) (8) is one of the older techniques that was not widely used for hand calculation due to the labor involved. It has become popular in recent years, however, because of its applicability to computers.

This method differs from the ones presented previously in that the actual computation of derivatives is not carried out. Also, the calculation of a value for y_{n+1} is made without reference to some previous value of y; therefore, no separate starting equation is necessary. The method is "inherently stable, and many steps of computation may be carried out." (6) This means that many points may be computed and the result will still be accurate. The error stays small and does not increase exponentially as it does in some methods.

This approach also possesses advantages that make it particularly suitable for a computer. Since no special starting formula is required, the entire computation can be put in one program. Also, the required memory capacity of the computer is reduced, since data from previous steps is not required.

The formulas for this method are presented below. The derivation for these is extremely involved and is generally not given in standard texts. Most authors refer the reader to the original work done on the subject (9).

There are actually several sets of formulas which have been derived. One of the most commonly used of these is (8)

$$\Delta y = (k_1 + 2k_2 + 2k_3 + k_4)/6 \tag{3.22}$$

or $$y_{n+1} = y_n + (k_1 + 2k_2 + 2k_3 + k_4)/6$$

where $$k_1 = hf(x_n, y_n)$$

$$k_2 = hf(x_n + \frac{h}{2}, y_n + \frac{k_1}{2})$$

$$k_3 = hf(x_n + \frac{h}{2}, y_n + \frac{k_2}{2})$$

$$k_4 = hf(x_n + h, y_n + k_3)$$

The actual calculation of the values of y is laborious but straightforward.

To obtain greater accuracy, the following set of formulas can be used. These should give an accuracy to the fifth order in h. However, it should be noted that the calculation becomes more difficult.

$$y_{n+1} = y_n + \frac{1}{192}(23k_1 + 125k_3 - 81k_5 + 125k_6)$$

$$k_1 = hf(x_n, y_n) \tag{3.23}$$

$$k_2 = hf(x_n + \frac{h}{3}, y_n + \frac{k_1}{3})$$

$$k_3 = hf(x_n + \frac{2h}{5}, y_n + \frac{6k_2 + 4k_1}{25})$$

$$k_4 = hf(x_n + h, y_n + \frac{15k_3 - 12k_2 + k_1}{4})$$

$$k_5 = hf(x_n + \frac{2h}{3}, y_n + \frac{8k_4 - 50k_1 + 90k_2 + 6k_1}{81})$$

$$k_6 = hf(x_n + \frac{4h}{5}, y_n + \frac{8k_4 + 10k_3 + 36k_2 + 6k_1}{75})$$

For second order equation, a different set of equations is necessary.

Consider a second order equation of the type

$$y'' = f(x, y, y')$$

Then

$$\Delta y = h\left[y_n' + \frac{1}{6}(k_1 + k_2 + k_3)\right] \tag{3.24}$$

$$\Delta y' = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

$$k_1 = hf(x_n, y_n, y_n')$$

$$k_2 = hf(x_n + \frac{h}{2}, y_n + \frac{h}{2} y_n' + \frac{h}{8} k_1, y_n' + \frac{k_1}{2})$$

$$k_3 = hf(x_n + \frac{h}{2}, y_n + \frac{h}{2} y_n' + \frac{h}{8} k_1, y_n' + \frac{k_2}{2})$$

$$k_4 = hf(x_n + h, y_n + hy_n' + \frac{h}{2} k_3, y_n' + k_3)$$

For the special case of equations of the form

$$y'' = f(x, y)$$

the solution is given by (3.25)

$$\Delta y = h\left[y_n' + \frac{1}{6}(k_1 + 2k_2)\right]$$

$$\Delta y' = \frac{1}{6}(k_1 + 4k_2 + k_3)$$

$$k_1 = hf(x_n, y_n)$$

$$k_2 = hf(x_n + \frac{h}{2}, y_n + \frac{h}{2} y_n' + \frac{h}{8} k_1)$$

$$k_3 = hf(x_n + h, y_n + hy_n' + \frac{h}{2} k_2)$$

When the Runge-Kutta Method is applied to a time varying equation, the assumption that the varying parameter is constant during the sampling interval is utilized. At each sample point the value of the variable quantity changes.

If an equation of order higher than second is to be solved, then it is necessary to break it up into a set of first order equations and solve each one independently.

Simpson's Rule is a widely used numerical method that provides accurate solution to differential equations. It is much easier to use than the Runge-Kutta method, and if a short sampling internal is used, it may be just as accurate.

This technique is derived from Newton's Formula for Forward Interpolation (Eq. 3.12). If Eq. 3.12 is integrated from x_o to x_o+nh (where n is the number of sample points to be calculated and h is the sampling interval) in steps of two, and the results are added together, it can be shown (8) that

$$\int_{x_o}^{2h} y\,dx = \frac{h}{3}(y_o + 4y_1 + y_2)$$

$$\int_{x_2}^{x_2+2h} y\,dx = \frac{h}{3}(y_2 + 4y_3 + y_4) \qquad (3.26)$$

and

$$\int_{x_o}^{x+nh} y\,dx = \frac{h}{3}(y_o+4y_1+2y_2+4y_3+2y_4+\ldots+3y_{n-2}+4y_{n-1}+y_n)$$

The application is simple. If the initial conditions are known, then the points can be calculated one at a time to give as many as desired. If a higher order equation is involved, it can be written as a set of first order equations. In this case, the application of Simpson's Rule yields a set of simultaneous algebraic equations which must be solved at each point. This is a situation in which a digital computer is extremely helpful.

3.3.2 Z Form Method

The Z Form Method is another numerical technique that may be used to obtain the response of a time varying system. It possesses the advantage that, although the approach utilizes an operational form of the problem in question, inverse transforms do not have to be obtained. Thus, the good features of an operational approach are retained while the part that may become cumbersome is avoided. Also, the method may be programmed for solution by a digital computer.

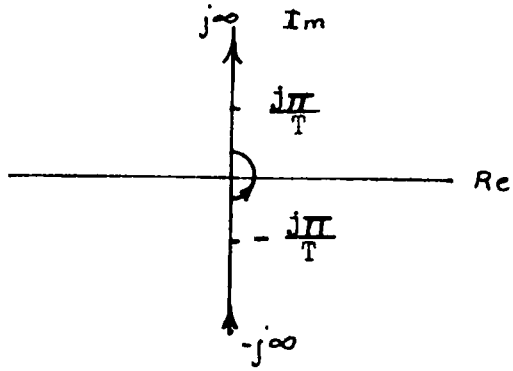
The theory involved with this method starts with the definition of the inverse Laplace Transform.

$$c(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} C(s)e^{st}\, ds \qquad (3.27)$$

This equation represents contour integration in the s plane.

If only stable systems are considered, the poles of C(s) are confined to the left half plane. Then the $j\omega$ axis can be utilized as a path of integration and $c + j\infty$ is equal to $j\infty$.

The path of integration is shown below



The integration around the infinite semicircle will be zero for functions which meet the following conditions (10).

If C(s) is an analytic function and

$$\lim_{R\to\infty} RC(Re^{j\theta}) = 0 \qquad \frac{-\pi}{2} \leq \theta \leq \frac{\pi}{2} \quad \text{or} \quad \frac{\pi}{2} \leq \theta \leq \frac{3\pi}{2}$$

uniformly with respect to θ, then

$$\lim_{R\to\infty} \int_{c_1} C(s)ds = 0$$

$$\lim_{R\to\infty} \int_{c_2} C(s)ds = 0$$

where c_1 and c_2 are semicircles in the right and left hand planes, respectively, centered at the origin and of radius R.

Assuming that these conditions are met, the integration along the imaginary axis can be broken up as follows.

$$c(t) = \frac{1}{2\pi j}\int_{-j\pi/T}^{j\pi/T} C(s)e^{st}\,ds + \frac{1}{2\pi j}\int_{j\pi/T}^{j\infty} C(s)e^{st}\,ds$$

$$+ \frac{1}{2\pi j}\int_{-j\infty}^{j\pi/T} C(s)e^{-st}\,ds \qquad (3.28)$$

If the period T is chosen such that the value $j\pi/T$ is much larger than the values of the poles on the $j\omega$ axis, then the second and third terms on the right side of Eq. 3.28 may be ignored.

Let $t = nT$

$$c(nT) = \frac{1}{2\pi j}\int_{-j\pi/T}^{j\pi/T} C(s)e^{snT}\,ds \qquad (3.29)$$

In the development of z transform theory (12), the following relationship is defined:

$$z = e^{Ts}$$

or

$$s = \frac{1}{T}\ln z$$

Change variables in Eq. 3.29 by substituting the z term defined above.

$$c(nT) = \frac{1}{2\pi j}\int_{e^{-j\pi}}^{e^{j\pi}} C(\tfrac{1}{T}\ln z)\, e^{nT(\frac{1}{T}\ln z)}\, d(\tfrac{1}{T}\ln z) \qquad (3.30)$$

Equation 3.30 may be written

$$c(nT) = \frac{1}{2\pi j}\frac{1}{T}\oint C(\frac{1}{T}\ln z)\, z^{n-1}\, dz \qquad (3.31)$$

where the contour is the unit circle in the z plane.

Equation 3.31 has the same form as the expression for the inverse z transform except for the term $\frac{1}{T}$. If a means can be found to evaluate the expression $C(\frac{1}{T}\ln z)$, then the solution is an approximation of the Laplace Transform.

The evaluation of $C(\frac{1}{T}\ln z)$ is difficult to do by ordinary means since $\frac{1}{T}\ln z$ cannot be expanded by synthetic division into powers of $\frac{1}{z}$. Therefore, ln z is approximated by

$$\ln z = 2(u + \frac{1}{3}u^3 + \frac{1}{5}u^5 + \ldots) \qquad (3.32)$$

where

$$u = \frac{1 - z^{-1}}{1 + z^{-1}}$$

From before,

$$s^{-1} = \frac{T}{\ln z}$$

Therefore,

$$s^{-1} = \frac{\frac{T}{2}}{(u + \frac{u^3}{3} + \frac{u^5}{5} + \ldots)} \qquad (3.33)$$

Use long division to obtain the following series:

$$s^{-1} = \frac{T}{2}(\frac{1}{u} - \frac{u}{3} - \frac{4u^3}{45} - \frac{44u^5}{945} - \ldots) \qquad (3.34)$$

The above series may be approximated by its principal part.

$$s^{-1} \approx \frac{T}{2u} \approx \frac{T}{2} \frac{1 + z^{-1}}{1 - z^{-1}} \tag{3.35}$$

If both sides of Eq. 3.34 are squared, the result is

$$s^{-2} = \frac{T^2}{4} \left[\frac{1}{u^2} - \frac{2}{u} \left(\frac{u}{3} + \frac{4u^3}{45} + \frac{45u^5}{945} + \ldots \right) + \ldots \right] \tag{3.36}$$

Approximate Eq. 3.36 by its principal part,

$$s^{-2} \approx \frac{T^2}{4} \left[\frac{1}{u^2} - \frac{2}{3} \right] \tag{3.37}$$

$$s^{-2} \approx \frac{T^2}{2} \left[\frac{1 + 10z^{-1} + z^{-2}}{(1 - z^{-1})^2} \right] \tag{3.38}$$

Similarly, series expansions for any s^{-k} can be obtained by raising both sides of Eq. 3.34 to the kth power. These expressions are known as Z forms. A table of these Z forms can be constructed and applied to various problems.

Now a relationship exists between Laplace Transforms and Z forms. An expression in terms of s can be changed into a function of z by using this relationship. The function of z still must be evaluated. This may be done by using a standard means of evaluating the inverse z transform and dividing by 1/T since the Z forms were derived from Eq. 3.31.

The Z forms may be evaluated by expanding the function of z, G(z), into a power series of z^{-n} terms by long division. The coefficient of the z^{-n} term corresponds to the time function g(t) at the nth sampling instant.

By definition, (12)

$$g(t) = \sum_{n=0}^{\infty} g(nT)\delta(t - nT) \tag{3.39}$$

$$g(t) = g(0)\delta(t) + g(T)\delta(t - T) + g(2T)\delta(t - 2T) + \ldots + g(nT)\delta(t - nT) \tag{3.40}$$

where g(t) is some pulsed output.

The Z Transform of Eq. 3.40 can be shown to be

$$G(z) = g(0) + g(T)z^{-1} + g(2T)z^{-2} + \ldots + g(nT)z^{-n} \tag{3.41}$$

Thus, the "coefficient of any z^{-n} is the value of g(t) at the nth sampling instant." (12)

The theory developed above may be expressed in a set of convenient rules. These may be applied to solve for the response of a system, whose transfer function is given in terms of s, by means of Z forms.

(1) Express C(s) as a rational fraction in powers of 1/s by dividing numerator and denominator by s^m, the highest power of s.

(2) Substitute for each s^{-k} a rational fraction z^{-1} and rearrange C(s) as a rational fraction in powers of z^{-1}.

(3) Divide the resulting expression by T, the time interval between the points at which the solution is desired.

(4) Expand the fraction by synthetic division into a series of the form,

$$D_o + D_1z^{-1} + D_2z^{-2} + \ldots D_kz^{-k} + \ldots$$

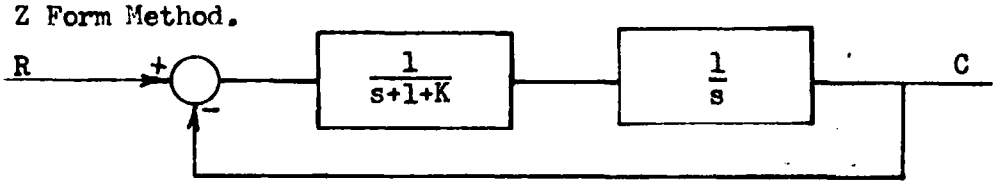
(5) The desired values are the coefficients D_o, D_1, ... D_k, ...

It would probably be wise to examine the error involved in the derivation of Z forms. Error can originate from the assumption that Eq. 3.29 represents the integral properly, the series approximation for ln z (Eq. 3.32), truncation of the series to obtain the Z form and round off error in the synthetic division to obtain the coefficients that give the answer.

Boxer and Thaler (11) suggest that the error can be minimized by shortening the sampling interval until no further change is observed. J. E. Gibson (6), however, disagrees, saying that although shortening the sampling interval will decrease the error involved in Eq. 3.29, it will increase the round off error because more points must be calculated. He concludes that "a more complete analysis of the errors involved in the Z form and an improved method of picking T (the sampling interval) would be called for." It would seem that the Z form method cannot be used with absolute confidence in the answers until a better analysis of the error is available.

The Z Form Method is now applied to a sample problem as a means of demonstration.

Example 3.2. Determine the step response of the following system by the Z Form Method.



where $K = e^{-.5t}$ and $R = u(t)$

Proceed by following the steps as outlined above.

(1) Find C(s) and divide the numerator and denominator by s^m.

$$C(s) = \frac{1}{s}\,\frac{1}{s(s+K)+1} = \frac{1}{s^3 + Ks^2 + s}$$

or

$$C(s) = \frac{s^{-3}}{1 + s^{-1}K + s^{-2}}$$

(2) Substitute the Z forms for each value of s^{-k}. These can be obtained from any convenient table.

$$C(z) = \frac{\frac{T^3}{2}\,\frac{z^{-1}+z^{-2}}{(1-z^{-1})^3}}{1 + \frac{KT}{2}\,\frac{1+z^{-1}}{1-z^{-1}}\,\frac{T^2}{12}\,\frac{1+10z^{-1}+z^{-2}}{(1-z^{-1})^2}}$$

(3) Divide the resulting expression by T

$$\frac{C(z)}{T} = \frac{T^2(z^{-1}+z^{-2})}{2(1-z^{-1})^3 + KT(1+z^{-1})(1-z^{-1})^2 + \frac{T^2}{6}(1+10z^{-1}+z^{-2})(1-z^{-1})}$$

$$\frac{C(z)}{T} = \frac{6T^2(z^{-1}+z^{-2})}{12(1-3z^{-1}+3z^{-2}-z^{-3}) + 6KT(1-z^{-1}-z^{-2}+z^{-3}) + T^2(1+9z^{-1}-9z^{-2}-z^{-3})}$$

$$\frac{C(z)}{T} = \frac{6T^2(z^{-1}+z^{-2})}{(12+6KT+t^2)+(-36-6KT+9T^2)z^{-1}+(36-6KT-9T^2)z^{-2}+(-12+6KT-T^2)Z^{-3}}$$

Choose T = 1

Then,

$$\frac{C(z)}{T} = \frac{6(z^{-1}+z^{-2})}{(13+6K) + (-27-6K)z^{-1} + (27-6K)z^{-2} + (-13+6K)z^{-3}}$$

(4) Expand the above fraction by synthetic division. The value of the time varying element must be adjusted for each sampling time.

The actual calculation of the values of the response is now just a question of tedious algebra. It takes several pages of hand calculation to solve a few points. Therefore, the actual results will not be shown here since this method is not ordinarily for hand calculation and there would be no advantage in presenting the results.

3.3.3 State Variable Method

The State Variable approach to the solution of differential equations is a more recent development. In this method, an nth order linear system is characterized by a set of n first order differential equations. The variables of this set of equations are called state variables since they "form a basis for an n-dimensional Euclidean Space. This space is called a state space, and a point in this space is called the state of the system." (17)

The steps used in this method are based on some physical meaning in the system; therefore, one might have more of a feeling for what is happening. There are no approximations in the development, and the mathematics is concise due to the use of vector-matrix notation which allows a set of first order equations to be represented by a single term. This method can easily be adapted to handle time varying problems.

To demonstrate one method of determining the state variables of a system, consider the following nth order differential equation.

$$\frac{d^n x_1}{dt^n} + a_n \frac{d^{n-1} x_1}{dt^{n-1}} + \cdots + a_2 \frac{dx_1}{dt} + a_1 x_1 = m(t) \qquad (3.42)$$

If $x_2 = \frac{dx_1}{dt}$, $x_3 = \frac{d^2 x_1}{dt} = \frac{dx_2}{dt}$, $x_4 = \frac{dx_3}{dt}$, ..., it is possible to write Eq. 3.42 as a set of first order differential equations. The quantities $x_1, x_2, \ldots, x_n$ are known as state variables.

The set of first order equations may be written in the following forms:

$$\dot{x}_1 = x_2 \qquad (3.43)$$

$$\dot{x}_2 = x_3$$

$$\vdots$$

$$\dot{x}_n = -a_n x_n - a_{n-1} x_{n-1} - \cdots - -a_2 x_2 - a_1 x_1 + m(t)$$

The above set of first order equations may be expressed in matrix form.

$$\dot{\underline{x}} = A\underline{x} + \underline{f(t)} \qquad (3.44)$$

where

$$\dot{\underline{x}} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \underline{f(t)} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ m(t) \end{bmatrix}$$

and $A =$

$$\begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ -a_1 & -a_2 & -a_3 & \cdot & \cdot & \cdot & -a_n \end{bmatrix}$$

If Eq. 3.44 is expanded according to the rules of matrix multiplication, Eqs. 3.43 are obtained.

The representation given for Eq. 3.42 is by no means unique. If the equation is expressed in block diagram form, a whole new set of state variables may be defined.

Example 3.3. Write the state variable representation for the following equation in the time and frequency domains.

$$\dddot{C} + (\alpha+\beta)\ddot{C} + \alpha\beta\dot{C} = ku(t)$$

where

$$\dot{C} = \frac{dC}{dt}$$

According to the development given for the general nth order equation, this can be written in the form

$$\dot{\underline{x}} = A\underline{x} + \underline{f(t)}$$

where

$$\dot{\underline{x}} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad f(t) = \begin{bmatrix} 0 \\ 0 \\ ku(t) \end{bmatrix}$$

and $$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -\alpha\beta & -(\alpha+\beta) \end{bmatrix}$$

If the equation is written in block diagram form, it appears as Fig. 3.1.

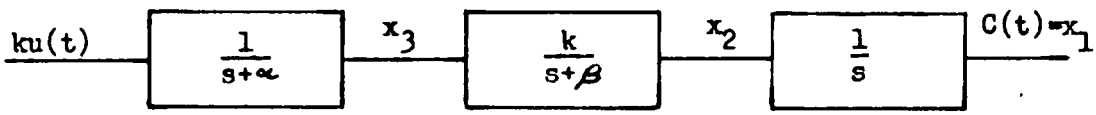


Fig. 3.1 Block diagram representation for Ex. 3.3.

The state variables can be expressed as

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = -\beta x_2 + kx_3$$

$$\dot{x}_3 = -\alpha x_3 + r(t)$$

These may be written in the form

$$\dot{\underline{x}} = A\underline{x} + \underline{f(t)}$$

In this case

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\beta & k \\ 0 & 0 & \alpha \end{bmatrix}$$

Two methods of representing the system by means of state variables have been presented. Other sets of equations can also be written. The representation can take many forms; however, all yield the same answer when the result is converted to some common basis.

It will be noted that both the time and frequency domain representations of the system may be written in the form of Eq. 3.44. The system is characterized by the matrix A. Solution of Eq. 3.44 may be determined in terms of A in both the time and frequency domains.

The solution of Eq. 3.44 in the frequency domain is started by taking the Laplace Transform of both sides.

$$s\underline{x}(s) - \underline{x}(0^+) = A\underline{x}(s) + \underline{F(s)} \tag{3.45}$$

Solve for $\underline{x}(s)$

$$\underline{x}(s)\ (s-A) = \underline{x}(0^+) + \underline{F(s)} \tag{3.46}$$

$$\underline{x}(s) = (sI-A)^{-1}\,\underline{x}(t_o^+) + (sI-A)^{-1}\,\underline{F(s)} \tag{3.47}$$

where $\underline{x}(t_o^+)$ denotes an initial condition of x(t) evaluated at $t = t_o$.

The matrix A includes only constants for constant coefficient equations; however, for time varying equations, it is necessary to alter the matrix at each sample point to account for the change in the time varying element. The matrix will be considered constant during the sampling interval, so that the theory discussed above is applicable.

The inverse Laplace Transform of $(sI-A)^{-1}$ is known as the state transition matrix (13)

$$\mathcal{L}^{-1}(sI-A)^{-1} = \phi(t-t_o) \tag{3.48}$$

Therefore,

$$\mathcal{L}^{-1}(\underline{x}(s)) = \underline{x}(t) = \phi(t-t_o)\underline{x}(t_o^+) + \int_t^{t_o} \phi(t-\tau)\underline{f}(\tau)d\tau \qquad (3.49)$$

Equation 3.44, therefore, gives a "description of the transition of state of the systems' state variables x_i from $x_i(t)$ to $x_i(t) + dx_i(t)$ when time is changed from t to t+dt. Since dt is an infinitesimal time change, the transition of a continuous data system takes place in a continuous fashion with respect to time." (13) Therefore, for a linear differential equation with constant coefficients, the answer at any time can be determined if the initial conditions and driving function are known. Enough information is known as is required to determine the entire future response of the system.

Example 3.4. - To demonstrate the theory presented above, an example involving a time varying element is shown in which the step response of a system is determined by the State Variable Method.

Consider the closed loop system for which the open loop transfer function will be

$$\frac{1}{s(s+1+k)}$$

where k varies as $e^{-.5t}$, and R = u(t).

The block diagram is given below

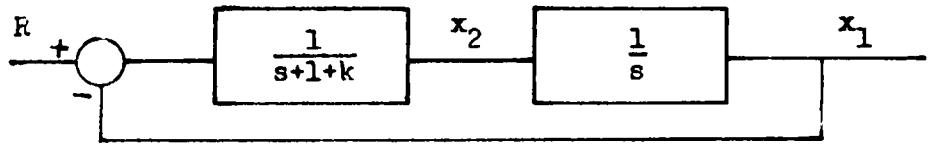


Fig. 3.2 Block diagram for Ex. 3.4.

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = -x_2(1 + k) + r(t)$$

Now the matrix A and the vector $\underline{f}(t)$ in Eq. 3.44 will be

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 1+k \end{bmatrix} \qquad \underline{f}(t) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The A matrix now contains a variable element. It is approximated by assuming the time varying term is constant for the sampling period and acquires a new value at each sampling instant.

To find the state transition equation, proceed in the usual way.

$$(sI - A) = \begin{bmatrix} s & -1 \\ 0 & s+1+k \end{bmatrix}$$

$$(sI - A)^{-1} = \begin{bmatrix} \frac{1}{s} & \frac{1}{s(s+1+k)} \\ 0 & \frac{1}{s+1+k} \end{bmatrix}$$

For a unit step input, R = 1 and $R(s) = \frac{1}{s}$. Therefore,

$$(sI - A)^{-1} R(s) = \begin{bmatrix} \dfrac{1}{s^2(s + 1 + k)} \\ \\ \dfrac{1}{s(s + 1 + k)} \end{bmatrix}$$

and the state transition equation is:

$$\underline{x}(s) = \begin{bmatrix} \dfrac{1}{s} & \dfrac{1}{s(s + 1 + k)} \\ \\ 0 & \dfrac{1}{s + 1 + k} \end{bmatrix} \begin{bmatrix} x_1(0^+) \\ \\ x_2(0^+) \end{bmatrix} + \begin{bmatrix} \dfrac{1}{s^2(s + 1 + k)} \\ \\ \dfrac{1}{s(s + 1 + k)} \end{bmatrix}$$

To find the values of the state variables at some time t, calculate the value of the time varying element at that time, find the inverse Laplace Transform of the preceding equation and expand the matrices according to the rules of matrix mathematics.

In this case, the output x_1 is computed at one second intervals starting at t = 0. It is assumed that the initial conditions are zero.

At t = 0

$$\underline{x}(s) = \begin{bmatrix} \dfrac{1}{s} & \dfrac{1}{s(s + 2)} \\ \\ 0 & \dfrac{1}{s + 2} \end{bmatrix} \begin{bmatrix} x_1(0^+) \\ \\ x_2(0^+) \end{bmatrix} + \begin{bmatrix} \dfrac{1}{s^2(s + 2)} \\ \\ \dfrac{1}{s(s + 2)} \end{bmatrix}$$

$$x_1(t_o) = \begin{bmatrix} 1 & .5 - .5e^{-2T} \\ \\ 0 & e^{-2T} \end{bmatrix} \begin{bmatrix} x_1(0^+) \\ \\ x_2(0^+) \end{bmatrix} + \begin{bmatrix} .5T - .25 + .25e^{-2T} \\ \\ .5 - .5e^{-2T} \end{bmatrix}$$

where $T = t - t_o$ = sampling period = 1 second.

Substitute T = 1 and $x_1(0^+) = x_2(0^+) = 0$. After calculation of the equation, it is determined that

$$x_1(t_o) = .284$$

$$x_2(t_o) = .432$$

These values of $x_1(t)$ and $x_2(t)$ become the initial conditions for the second calculation. Also $x_1(t_o)$ is the output of the system for t = 1.

It should also be noted that the matrix

$$\begin{bmatrix} 1 & .5-.5e^{-2t} \\ 0 & e^{-2t} \end{bmatrix}$$

is the state transition matrix of this equation for the first sampling period.

In order to carry out the rest of the computation, it is necessary to change the values of several terms in the state transition equation. The initial conditions change at each sampling point. Also, any term in the matrices that is affected by a change in the time varying element also changes. Finally, the value of the input is different since negative feedback is being used in the system.

Therefore, for t = 1 second, the state transition equation will change as shown below.

$$\underline{x}(s) = \begin{bmatrix} \frac{1}{s} & \frac{1 \quad 1}{s(s+1.606)} \\ 0 & \frac{1}{s+1.606} \end{bmatrix} \begin{bmatrix} x_1(0^+) \\ x_2(0^+) \end{bmatrix} + \begin{bmatrix} \frac{1}{s^2(s+1.606)} \\ \frac{1}{s(s+1.606)} \end{bmatrix} .716$$

and

$$x(t_1) = \begin{bmatrix} 1 & .625-.625e^{-1.606T} \\ 0 & e^{-1.606T} \end{bmatrix} \begin{bmatrix} .284 \\ .432 \end{bmatrix} + \begin{bmatrix} .625T-.388+.388e^{-1.606T} \\ .625-.625e^{-1.606T} \end{bmatrix} .716$$

$$x_1(t_1) = .724$$

$$x_2(t_1) = .437$$

In a similar manner, the points can be calculated for as many values of t as are desired. The solution to this problem is shown tabulated below.

TABLE I

SOLUTION TO EXAMPLE 3.4

t	$x_1(t)$	$x_2(t)$	$e(t) = r(t)-x_1(t)$
0^+	0	0	1
1	.284	.432	.716
2	.724	.437	.276
3	1.051	.260	-.051
4	1.182	.048	-.182
5	1.145	-.1	-.145
6	1.031	-.119	-.031
7	.947	-.062	.053
8	.926	.012	.074
9	.971	.057	.029

It can be seen that the value of $x_1(t)$ is approaching one as a limit.

An important point should be made concerning the physical significance of the state transition matrix. Examine this matrix in the preceding example. Note that each element a_{ij} of the matrix is the impulse response at the jth block due to a unit impulse input at the ith block. This not only provides an interesting piece of information, but also gives a quick means of determining the state transition matrix when the problem can be set up in a block diagram.

The state transition matrix helps in determining the response of a system since, by knowing the impulse response between all points in the system, the response to any input or set of initial conditions can be determined.

Linear equations with time varying coefficients can also be solved in the time domain using state variable techniques (14)(15).

Consider an equation of the form of Eq. 3.44.

$$\dot{\underline{x}} = A\underline{x} + \underline{f}(t) \tag{3.50}$$

A scalar equation of the same form has the solution,

$$y(t) = e^{a(t-t_o)} y(t_o) + \int_t^{t_o} e^{a(t-t')} f(t')\, dt' \tag{3.51}$$

where

$$e^{at} = \sum_{k=0}^{\infty} \frac{a^k t^k}{k!}$$

By analogy (15), the solution for equations of the form of Eq. 3.50 is given by

$$\underline{x}(t;x_o,t_o;f(t)\ ) = e^{A(t-t_o)}\ x_o + \int_t^{t_o} e^{A(t-t')}\ \underline{f}(t')\ dt' \qquad (3.52)$$

where e^{At} is defined as

$$e^{At} = \sum_{k=0}^{\infty} \frac{A^k t^k}{k!} \qquad (3.53)$$

The main problem in using this technique is the evaluation of e^{At}. The method used here is based on the Cayley-Hamilton theorm, which states that every matrix satisfies its own characteristic equation.

Assume $q(\lambda)$ is a polynomial and A is an arbitrary n-dimensional matrix whose characteristic equation is $g(A) = 0$. If $q(\lambda)$ is divided by $g(\lambda)$, a quotient $q_1(\lambda)$ is realized along with a remainder $r(\lambda)$.

$$q(\lambda) = g(\lambda)\ q_1(\lambda) + r(\lambda) \qquad (3.54)$$

Since Eq. 3.54 is an identity, it will also be valid if λ is replaced by the matrix A.

Therefore,

$$q(A) = g(A)\ q_1(A) + r(A) \qquad (3.55)$$

But $g(A)$, the characteristic equation, is zero. Thus, $q(A) = r(A)$.

Now suppose it is desired to find some $f(A)$ when $f(\lambda)$ is an analytic function of λ . Using a method similar to the one above,

look for a polynomial of degree n-1 such that

$$f(\lambda) = g(\lambda)\, h(\lambda) + r(\lambda) \tag{3.56}$$

where $h(\lambda)$ is an analytic function of λ . Substitute the eigenvalues of A $--\lambda_1, \lambda_2, \ldots, \lambda_n$ -- in Eq. 3.56. Since $g(\lambda_i) = 0 (1 \leq i \leq n)$, a set of n linear equations for the n coefficients of the polynomial $r(\lambda)$ is obtained.

$$\begin{aligned} f(\lambda_1) &= r(\lambda_1) \\ &\vdots \\ f(\lambda_n) &= r(\lambda_n) \end{aligned} \tag{3.57}$$

Next $r(\lambda)$ is determined to satisfy Eq. 3.57, and the resulting $h(\lambda)$

$$h(\lambda) = \frac{f(\lambda) - r(\lambda)}{g(\lambda)} \tag{3.58}$$

is an analytic function of λ , since it is the quotient of two analytic functions, and the zeros of the numerator are the same as the zeros of the denominator.

Since Eq. 3.56 is valid for all values of λ in a circle around the origin, A may be substituted for λ to obtain

$$f(A) = g(A)\, h(A) + r(A) \tag{3.59}$$

and since $g(A) = 0$

$$f(A) = r(A) \tag{3.60}$$

The result is formulated below (14).

If A is a matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$, arranged in order of increasing absolute value, and if $f(\lambda)$ is an analytic function of λ in a circle about the origin with radius greater than $|\lambda_n|$, then f(A) equals r(A), the polynomial of degree n-1 for which

$$f(\lambda_k) = r(\lambda_k) \quad k = 1, 2, \ldots, n$$

A complication arises when two or more of the eigenvalues of the matrix A have equal values. In the series of equations given by Eq. 3.57, not enough information is supplied to solve for the unknown, since two or more of the resulting equation are identical.

To remedy the situation, start with the following theorm (14).

"If a suitable basis is used, every operator on a finite dimensional space may be represented by a matrix in Jordan canonical form."

Jordan Canonical form of a matrix is

$$J = \begin{bmatrix} J_0 & 0 & \cdot & \cdot & 0 \\ 0 & J_1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & J_k \end{bmatrix} \tag{3.61}$$

where

$$J_o = \begin{bmatrix} \lambda_1 & 0 & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \circ & \cdot \\ \cdot & \cdot & \cdot & \cdot & \circ \\ \circ & \circ & \cdot & \cdot & 0 \\ 0 & \circ & \cdot & 0 & \lambda_n \end{bmatrix} \tag{3.62}$$

is a matrix containing the distinct eigenvalues of the system along its main diagonal, and

$$J_i = \begin{bmatrix} \lambda_{n+i} & 1 & 0 & \circ & 0 \\ 0 & \lambda_{n+i} & 1 & \circ & \circ \\ \circ & \circ & \cdot & \cdot & 0 \\ \cdot & \circ & \cdot & \cdot & 1 \\ 0 & \circ & \circ & 0 & \lambda_{n+i} \end{bmatrix} \tag{3.63}$$

where $i = 1, 2, \ldots, k$.

There is a J_i matrix for each set of repeating roots.

For a fourth order system, with two repeating roots,

$$J_o = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad \text{and} \quad J_1 = \begin{bmatrix} \lambda_3 & 1 \\ 0 & \lambda_3 \end{bmatrix}$$

If $f(J)$ is expanded in a Taylor series, the result is

$$f(J) = f(\lambda)I + f'(\lambda)(J - \lambda I) + \frac{f''(\lambda)}{2!}(J - \lambda I)^2 + \ldots \tag{3.64}$$

If J is of the form J_i

$$f(J_i) = f(\lambda_i)I + f'(\lambda_i)(J_i - \lambda_i I) + \ldots \tag{3.65}$$

$$(J - \lambda_i I) = \begin{bmatrix} 0 & 1 & \cdot & \cdot & 0 \\ \cdot & 0 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \\ 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix} \tag{3.66}$$

This is milpotent, i.e.

$$(J_i - \lambda_i I)^{\alpha} = \underline{0} \tag{3.67}$$

where α is the number of repeated eigenvalues.

From Eq. 3.56

$$f(x) = g(x)h(x) + r(x) \tag{3.68}$$

Expanding f(x) and r(x) in a Taylor series yields

$$f(\lambda) + f'(\lambda)\,(x-\lambda) + \frac{f'(\lambda)}{2!}(x-\lambda)^2 + \ldots = g(x)h(x) + r(\lambda)$$

$$+ r'(\lambda)\,(x-\lambda) + \frac{r''(\lambda)}{2!}(x-\lambda)^2 + \ldots \tag{3.69}$$

Write Eq. 3.68 in terms of J_i

$$f(\lambda_i)I + f'(\lambda_i)\,(J_i - \lambda_i I) + \ldots = g(J_i)\,h(J_i) + r(\lambda_i)I$$

$$+ r'(\lambda_i)\,(J_i - \lambda_i I) + \ldots \tag{3.70}$$

Since $(J - \lambda_i I)^{\alpha}$ is milpotent, there will only be as many terms in each series as there are repeated roots. Also the value

$g(J_i)$ is zero, since it is the characteristic equation for J_i.

Finally, equate like terms on either side of the equation to get

$$\begin{aligned} f(\lambda_i) &= r(\lambda_i) \\ f'(\lambda_i) &= r'(\lambda_i) \\ &\vdots \\ f^{(\alpha)}(\lambda_i) &= r^{(\alpha)}(\lambda_i) \end{aligned} \tag{3.71}$$

This set of equations can be used to determine $r(\lambda)$.

Example 3.5. To illustrate this procedure, Example 3.4 is solved in the time domain. The differential equation is

$$\ddot{x} + (1 + k)\dot{x} + \hat{x} = u(t)$$

where $k = e^{-.5t}$. This may be written as a set of first order equations.

$$\begin{aligned} x_1 &= x \\ \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -(1+k)x_2 - x_1 + u(t) \end{aligned}$$

which may also be put in the form

$$\dot{\underline{x}} = A\underline{x} + \underline{f}(t)$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -(1+k) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t)$$

Next find the eigenvalues of the matrix A.

$$\left| A - \lambda I \right| = \begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda-(1+k) \end{vmatrix}$$

$$\lambda^2 + \lambda(1+k) + 1 = 0$$

At t = 0, this equation will be

$$\lambda^2 + 2\lambda + 1 = 0$$

The eigenvalues are -1, -1. To evaluate e^{At}, let

$$f(A) = e^{At}$$

Select a polynomial of degree n-1 for r(A).

$$r(A) = \alpha_o I + \alpha_1 A$$

Write a set of relationships between $f(\lambda_i)$ and $r(\lambda_i)$.

$$f(\lambda_1) = r(\lambda_1)$$

$$f(\lambda_2) = r(\lambda_2)$$

Since the eigenvalues are equal, the second equation must follow the rule given in Eq. 3.64.

$$e^{-t} = \alpha_o - \alpha_1$$

$$te^{-t} = 0 - \alpha_1$$

Solve for α_1 and α_o to find that

$$\alpha_1 = te^{-t}$$

$$\alpha_o = e^{-t} + te^{-t} = e^{-t}(t + 1)$$

Use the equation

$$e^{At} = \alpha_o I + \alpha_1 A$$

to determine that

$$e^{At} = e^{-t}(t+1) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + te^{-t} \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}$$

Therefore,

$$e^{At} = \begin{bmatrix} e^{-t}(t+1) & te^{-t} \\ -te^{-t} & e^{-t}(-t+1) \end{bmatrix}$$

If this is substituted in Eq. 3.52, the free motion of the system in the the interval $(t-t_o)$ is described by

$$x_1(t) = e^{-(t-t_o)}((t-t_o)+1)x_1(0+) + (t-t_o)e^{-(t-t_o)}x_2(0+)$$

$$x_2(t) = -(t-t_o)e^{-(t-t_o)}x_1(0+) + e^{-(t-t_o)}(-(t-t_o)+1)x_2(0+)$$

The step input is accounted for by the second part of Eq. 3.52.

$$x_1(t) = \int_o^t (t-t')e^{-(t-t')}dt'$$

$$x_2(t) = \int_o^t e^{-(t-t')}(-(t-t')+1)dt'$$

The complete solution would be the sum of the free and forced motion.

As before, $x_1(t)$ and $x_2(t)$ become the initial conditions for the next calculation.

The solutions at various sample points are not presented here; therefore, a comparison with Ex. 3.4 is not possible at this time. In the following chapter, this problem is solved in the time domain by computer and a comparison can be made.

3.4 Conclusion

All of the numerical methods discussed in the previous section are in use today to solve differential equations. Their application to time varying equations is not discussed extensively in the literature; however, the same assumption concerning the time varying element is made for all of the methods; the value of the varying element is assumed constant during the sampling interval and changes at each sample point.

It is difficult, therefore, to say that one method is "better" than another, since all give accurate answers. However, since it is desired to select one method for computer solution, other things can be considered in making the choice.

For instance, the Runge-Kutta method was developed many years ago and has remained substantially unchanged. This does not mean it is not a useful approach; however, since so much has been done with it, it does not appear profitable to go into this method of solution more deeply at this time. Also, the nature of the method is such that feeling for the problem is lost, since the equations that are used are not connected with the problem.

Simpson's Rule is easier to use than the Runge-Kutta method and is also quite accurate. It is widely used and extremely helpful. However, it would be convenient to have a method which is more in line with the thinking of Electrical Engineers.

The Z Form or State Variable method appears at first to be desirable in this respect. They are both frequently used techniques which are known to Electrical Engineers. However, upon investigation, they are found lacking.

The Z Form method is not desirable due to the uncertainty of the error involved in the approximations that are made. The State Variable method does not have this drawback; however, the difficulties involved in writing a program for this method are great. The extension to solution of higher order equations is even harder. For these reasons, it seems that the difficulty involved in using these methods outweighs the advantages they possess.

The investigation of the State Variable and Z Form methods showed them to be undesirable. Therefore, Simpson's Rule will be chosen as a method to solve time varying problems. It is accurate, efficient, and widely known. In the next chapter, it is programmed and used to solve sample problems as a means of demonstration.

CHAPTER 4

APPLICATION OF DIGITAL COMPUTERS

4.1 Introduction

In previous chapters, the theoretical aspects of the solution of differential equations are discussed. Much of the theory may be readily applied to various problems; however, the work involved in hand calculation is usually too great for this to be considered practical. Therefore, digital computers are looked to as a means of overcoming this barrier. It is the purpose of this chapter to present a computer program that handles the calculation of time varying equations in an efficient manner, thus opening the door to the practical solution of many problems.

In the previous chapter, it was decided to program Simpson's Rule as a means of solving time varying differential equations. This method is a convenient one to use since it can be written without much difficulty, does not require a great amount of computer memory space, and can be extended to higher orders quite easily.

The next section discusses the program and is followed by the solution of sample problems.

4.2 Discussion of the Program

This section may best be followed by referring to the flow chart in Fig. 4.1 and the sample problems.

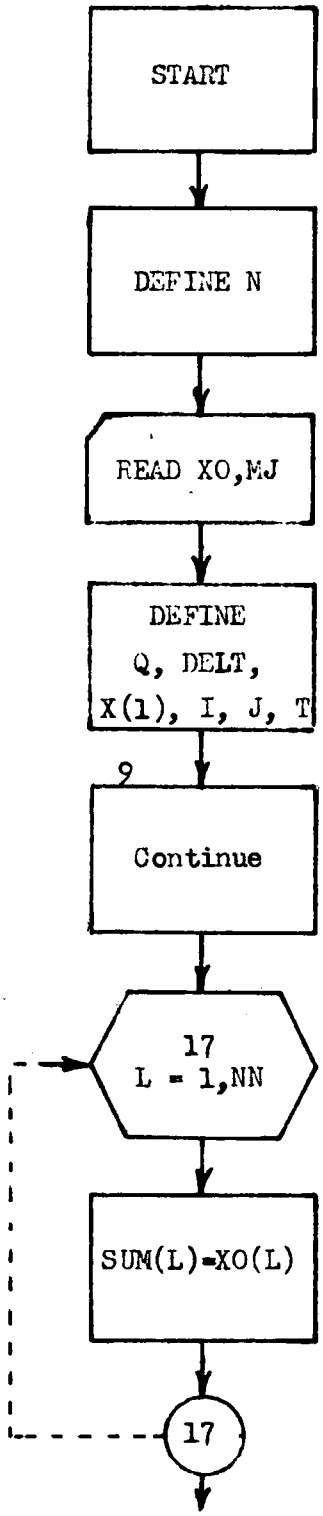


Figure 4.1 Flow Chart for Digital Computer Program

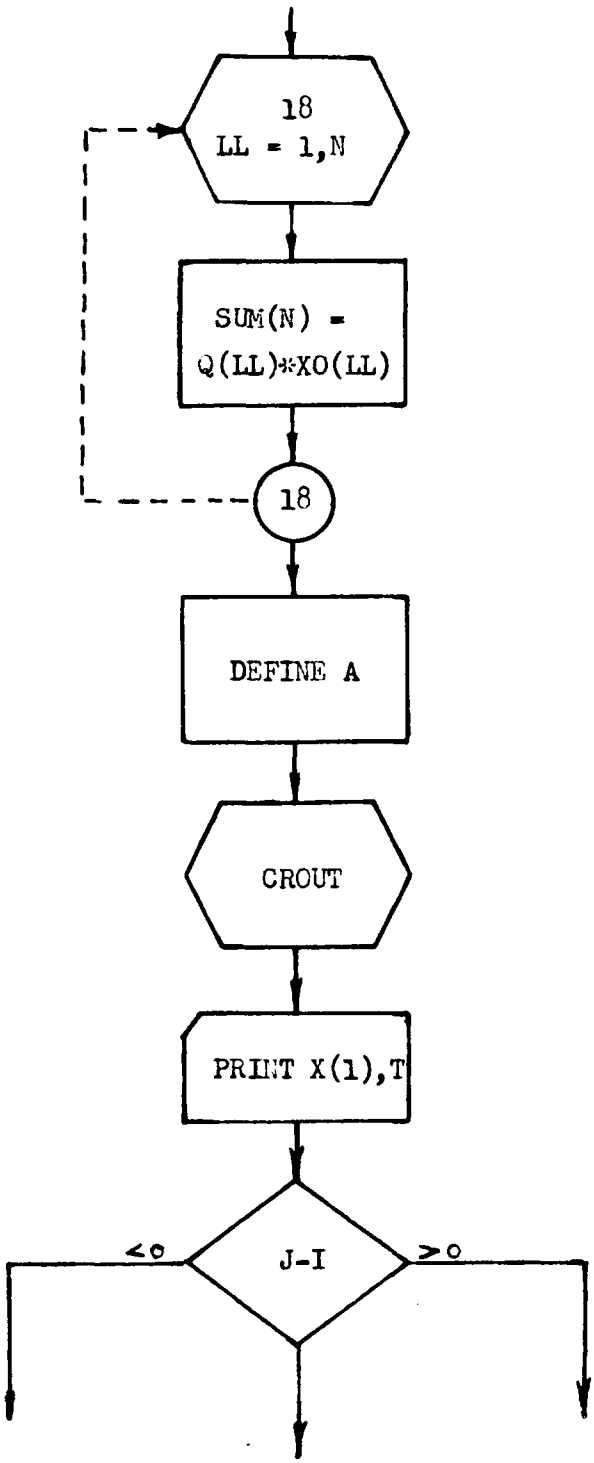


Figure 4.1--Continued

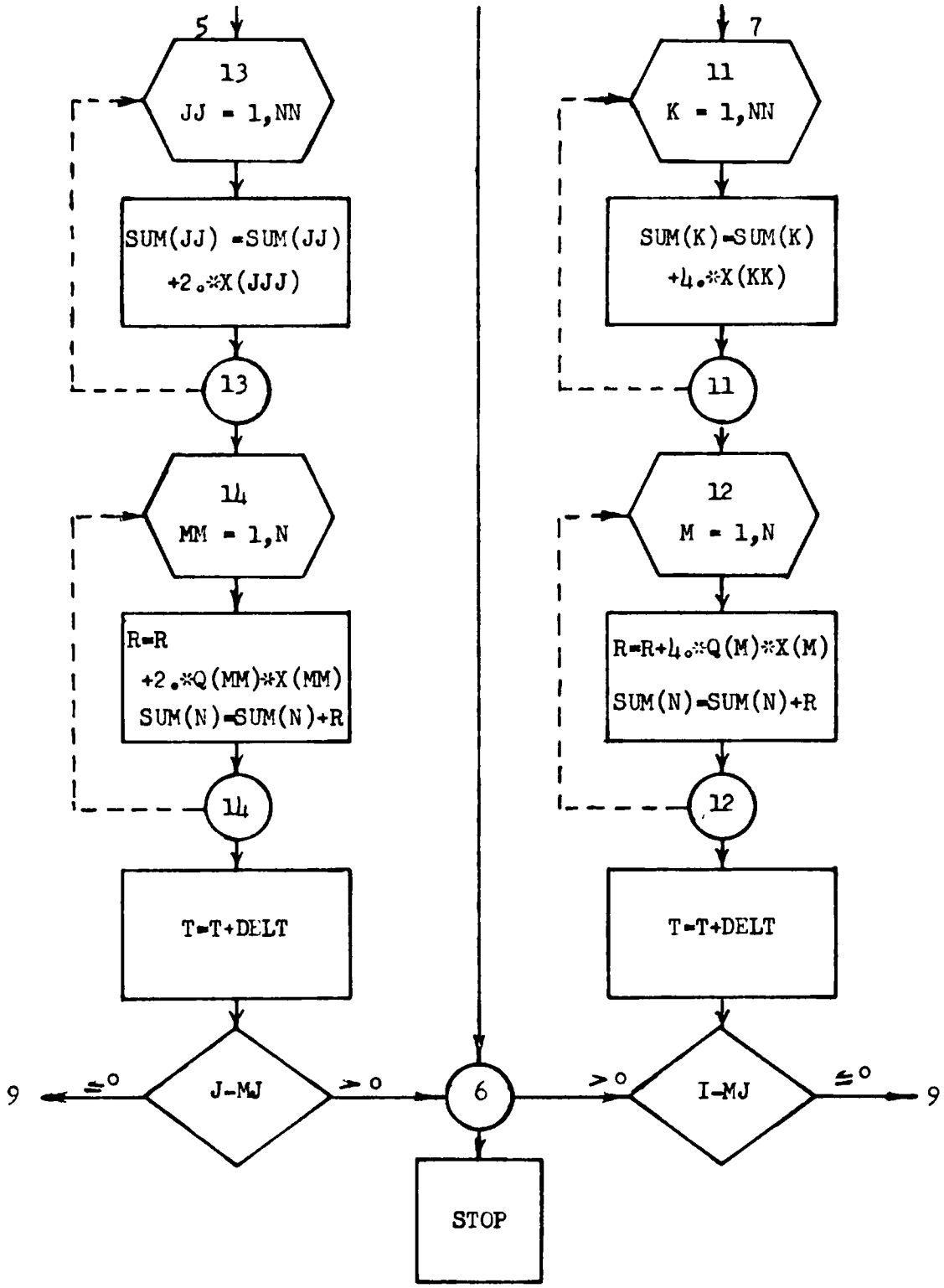


Figure 4.1--Continued

The following procedure is handled by the computer. When Simpson's Rule is applied to a set of first order equations, a set of linear algebraic equations is obtained (see examples). The coefficients on the left of the equations and the values on the right are set up in an array and fed to the subroutine CROUT which furnishes a solution to the set of equations and, consequently, determines the response at a sample point. The process is repeated any desired number of times.

The following terms are explained now to avoid confusion:

SUM = sum of all the sample points that have already been calculated. This must be changed each time by adding on each new sample point (see Ex. 4.1).

Q = coefficients of the variable in the last equation of the set of first order equations (see Ex. 4.1).

N = order of the problem

XO = initial conditions

DELT = sampling interval

MJ = number used to control the amount of sample points to be calculated

X(1) = value of sample point

To use the program, first dimension XO(N), SUM(N), and Q(N). Define N, read in XO and MJ, define DELT and set SUM equal to the initial conditions. Next define Q(1), ..., Q(N) and the elements of the array that will be fed into subroutine CROUT, A(1,1), A(1,2), ... Note that the value of Q and A must be punched on separate cards since some of them vary with time and must be recalculated each time. For

an equation with constant coefficients, these values may be read in.

Call CROUT. This subroutine solves the set of simultaneous algebraic equations and yields a sample point, X(1), with each calculation. Next, the result is printed along with the time.

Some of the sample points must be multiplied by four and some by two. The correct decision in this case is made by the IF statement following CROUT. After this is carried out, the values of SUM are changed and time is increased by DELT. The program is stopped when the desired number of points is computed by the final IF statement.

4.3 Sample Problems

Example 4.1. Solve Example 3.4 in the time domain by means of Simpson's Rule. The differential equation is

$$\ddot{x} + (1 + k)\,\dot{x} + x = u(t)$$

where $k = e^{-.5t}$

Let $x_1 = x$ and write the problem as a set of first order equations.

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = -x_1 - (1 + k)\,x_2 + u(t)$$

Applying Simpson's Rule to the set of equations yields

$$x_1^n - x_1^o = \frac{h}{3}\,(x_2^o + 4x_2^1 + 2x_2^2 + \ldots\, x_2^n)$$

$$x_2^n - x_2^o = \frac{h}{3}\,(x_1^o + 4x_1^1 + 2x_1^2 + \ldots x_1^n) - (1+k)\,\frac{h}{3}(x_2^o + \ldots x_2^n) + u(t)\cdot T$$

where x_2^o denotes the initial condition of x_2, x_2^1 is the first value of x_2, etc.

Bring the nth terms to the left side of the equation and let all other terms on the right inside the brackets be known as SUM.

$$x_1^n - x_2^n \left(\frac{h}{3}\right) = x_1^o + \frac{h}{3}\,(SUM(1))$$

$$x_1^n \left(\frac{h}{3}\right) + x_2^n \left(1 + \frac{h}{3}(1+k)\right) = x_2^o - \frac{h}{3}\,(SUM(2)) + u(t)\cdot T$$

where

$$SUM(1) = x_2^o + 4x_2^1 + \ldots + 4x_2^{n-1}$$

$$SUM(2) = -(x_1^o + 4x_1^1 + \ldots + 4x_1^{n-1}) - (1 + k)\,(x_2^o + \ldots + 4x_2^{n-1})$$

and T = sampling interval

The equations are now in the form of two simultaneous algebraic equations.

To use the program, first set N=2 and punch out the proper DIMENSION statement. Then, put in the values Q(1)=-1, Q(2)=-(1+k) and the terms $A(1,1)=1$, $A(1,2)=-\frac{h}{3}$, $A(1,3)=x_1^o+\frac{h}{3}(SUM(1))$, $A(2,1)=\frac{h}{3}$, $A(2,2)=1+\frac{h}{3}(1+k)$, $A(2,3)=x_2^o-\frac{h}{3}(SUM(2))+u(t)\cdot T$.

Call CROUT to get the solution X(1) and X(2). The point X(1) is the solution at the particular time in question. The values of SUM(1) and SUM(2) are changed, the time is moved up, and the computation proceeds for another point.

The results are tabulated below for various sampling intervals. The actual program is found in the Appendix.

TABLE II

SOLUTION TO EXAMPLE 4.1

T	X(1) DELT = .05	X(1) DELT = .1	X(1) DELT = 1
0	0	0	0
1	.2752	.2720	.2024
2	.6731	.6744	.6345
3	.9405	.9441	1.251
4	1.050	1.053	1.192
5	1.057	1.058	1.096
6	1.026	1.025	.9694
7	1.000	.9990	.9339
8	.9906	.9897	.9241
9	.9925	.9923	1.037
10	.9976	.9978	1.014

Note that the results for sampling intervals .1 and .05 agree extremely well while the results for DELT = 1 differ by a greater amount. This illustrates the rule of halving the interval until the data agrees within a required value. The long sampling interval produces a poor approximation of the curve.

A comparison with Example 3.4 would not be too meaningful since the data in both cases for a sampling interval of one second is not too accurate. However, both curves do have the same general shape.

Example 4.2. Solve the following third order equation by means of Simpson's Rule.

$$\dddot{C} + (3 + k)\ddot{C} + 2(1 + k)\dot{C} + C = u(t)$$

where $\dot{C} = \frac{dC}{dt}$ and $k = e^{-t}$

Let $x_1 = C$ and write

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = x_3$$

$$\dot{x}_3 = -x_1 - 2(1 + k)x_2 - (3 + k)x_3 + u(t)$$

Applying Simpson's Rule as before and manipulating the equations yields

$$x_1^n + x_2^n(-\tfrac{h}{3}) + x_3^n(0) = x_1^o + \tfrac{h}{3}(SUM(1))$$

$$x_1^n(0) + x_2^n + x_3^n(-\tfrac{h}{3}) = x_2^o + \tfrac{h}{3}(SUM(2))$$

$$x_1^n(\tfrac{h}{3}) + x_2^n(2)(1 + k)(\tfrac{h}{3}) + x_3^n(1 + \tfrac{h}{3}(3+k)) = x_3^o + \tfrac{h}{3}(SUM(3)) + u(t)\cdot T$$

The values of the array are read in and the answers at each point are multiplied by the proper coefficient as before.

Again, the results are tabulated below and the program is given in the Appendix. The agreement for the smaller sampling intervals is extremely good while the data for the interval of one second does not agree as well. These examples demonstrate the importance of using a small enough sampling interval to obtain a good approximation.

TABLE III

SOLUTION OF EXAMPLE 4.2

T	X(1) DELT = .05	X(1) DELT = .1	X(1) DELT = 1
0	0	0	0
1	.0691	.0662	.0451
2	.3008	.2981	.2064
3	.5925	.5921	.6932
4	.8500	.8521	.9141
5	1.022	1.026	1.197
6	1.102	1.106	1.224
7	1.112	1.114	1.180
8	1.083	1.083	1.120
9	1.042	1.042	.9924
10	1.009	1.007	.9634

4.4 Conclusion

The benefits attained by the use of a digital computer are obvious. Computation that might take hours, days, or perhaps be prohibitively long to do by hand can be completed in a short time. Greater accuracy can also be obtained. The error due to rounding off is smaller than that involved in hand calculation, and answers accurate to many places may be written out.

The programs as they stand can solve equations of second and third order. The extension to fourth or higher order problems can be done quite easily. The only difference will be the number of cards feeding information into the computer. The procedure is exactly the same for any order problems. The only theoretical limit is the subroutine CROUT which cannot handle a matrix greater than twenty by twenty.

The method then is quite useful in solving a large number of problems efficiently and accurately.

CHAPTER 5

CONCLUSION

The information presented in the first four chapters makes several things apparent.

1. Analytical methods do not present a convenient means of solving linear differential equations with time varying coefficients.
2. Numerical methods are the only means available today of solving these equations efficiently.
3. A digital computer is an extremely helpful, if not indispensible, aid in using numerical methods.

The inadequacy of available analytical techniques is demonstrated in Chapter 2. The computations are long, complicated, and sometimes impossible to work. Even approximation techniques, which apply to a restricted class of problems, do not alleviate the situation to the point where they are worth pursuing further. Whether the approach is in the time or frequency domain, the stumbling blocks are sufficiently large to cause one to search for new ways to attack the problem. This leads to an investigation of numerical methods.

When numerical methods are examined, it is found that they are very time consuming to do by hand; however, the mathematics is quite simple even in the more sophisticated cases. What makes the numerical

techniques especially attractive is their adaptability to solution by digital computer.

The State Variable method at first glance seems to be the most desirable one to apply to digital computer solution. However, the attempt to do it turned up difficulties which far outweigh the advantages of using the method. Chief among these are the length of the program and the problems involved in writing a program for higher order equations. For these reasons, it is decided to use Simpson's Rule as a means of solving time varying equations. This accurate, well-known technique can easily be programmed for high order equations with little trouble. The program developed in Chapter 4 may be used to solve problems up to twentieth order.

As far as application of these methods is concerned, an obvious one is the field of rocket research. Several time varying quantities are involved here. Among these are varying mass due to fuel consumption, change of air density on takeoff and especially re-entry from space, changing temperature, and variation of gravitational forces. In other areas such as circuits, time varying elements could appear. These range from such simple circuits as the carbon microphone to a super-regenerative receiver where the circuit resistance is caused to vary periodically by means of an external oscillator.

In the future, it is hoped that some convenient analytical method can be determined for selecting the sampling interval. The hit or miss operation used today is effective, but wastes valuable computer time. Also, an examination of the Z Form method to determine an

effective means of analyzing the error would make the technique more effective by eliminating uncertainty regarding its accuracy.

APPENDIX

PROGRAM TO SOLVE EXAMPLE 4.1

```
      SUBROUTINE CROUT
      DIMENSION A(50,51), X(50)
      COMMON A, X, N
      NP1 = N + 1
      NM1 = N - 1
      DO 1100 J=2,NP1
      A(1,J) = A(1,J)/A(1,1)
      A(2,J) = A(2,J) - A(2,1)*A(1,J)
1100  CONTINUE
      DO 1102 J=3,NP1
      A(2,J) = A(2,J)/A(2,2)
1102  CONTINUE
      DO 1140 I = 3,N
      IP1 = I + 1
      IM1 = I - 1
      DO 1104 J=2,IM1
      JM1 = J - 1
      DO 1103 K=1,JM1
      A(I,J) = A(I,J) - A(I,K)*A(K,J)
1103  CONTINUE
1104  CONTINUE
      DO 1120 J=I,NP1
      DO 1110 K=1,IM1
      A(I,J) = A(I,J) - A(I,K)*A(K,J)
1110  CONTINUE
1120  CONTINUE
      DO 1130 J = IP1,NP1
      A(I,J) = A(I,J)/A(I,I)
1130  CONTINUE
1140  CONTINUE
      X(N) = A(N,NP1)
      J1 = N
      DO 1160 I = 1,NM1
      J1 = J1 - 1
      J1P1 = J1+1
      X(J1) = A(J1,NP1)
      DO 1150 J = J1P1,N
      X(J1) = X(J1) - A(J1,J)*X(J)
1150  CONTINUE
1160  CONTINUE
      RETURN
      END
      FUNCTION DERIV(U,V,W,Y,Z,DELT)
      DERIV = (3.*U-16.*V+36.*W-48.*Y+25.*Z)/(12.*DELT)
      RETURN
      END
```

```
      DIMENSION A(50,51), X(50)
      DIMENSION XO(2),SUM(2),Q(2)
      N=2
      READ10,(XO(L),L=1,N),MJ
   10 FORMAT(2F10.0,I10)
      X(1)=0.
      J=1
      I=0
      T=0.
      DELT=1.
      NN=N-1
      DO17 L=1,NN
   17 SUM(L)=XO(L)
      DO18 LL=1,N
   18 SUM(N)=Q(LL)*XO(LL)
    9 CONTINUE
      Q(1)=-1.
      Q(2)=-(1.+EXPEF(-.5*T))
      A(1,1)=1.
      A(1,2)=-DELT/3.
      A(1,3)=XO(1)+DELT/3.*SUM(1)
      A(2,1)=DELT/3.
      A(2,2)=1.-DELT/3.*Q(2)
      A(2,3)=XO(2)+DELT/3.*SUM(2)+T
      COMMON(A,X,N)
      CALLCROUT
      PRINT15,X(1),X(2),X(3),T
   15 FORMAT(4E22.8)
      R=0.
      IF(J-I)5,6,7
    7 DO11 K=1,NN
      KK=K+1
   11 SUM(K)=SUM(K)+4.*X(KK)
      DO12 M=1,N
   12 R=R+4.*Q(M)*X(M)
      SUM(N)=SUM(N)+R
      T=T+DELT
      I=I+2
      IF(I-MJ)9,9,6
    5 DO13 JJ=1,NN
      JJJ=JJ+1
   13 SUM(JJ)=SUM(JJ)+2.*X(JJJ)
      DO14 MM=1,N
   14 R=R+2.*Q(MM)*X(MM)
      SUM(N)=SUM(N)+R
      T=T+DELT
      J=J+2
      IF(I-MJ)9,9,6
    6 STOP
      END
```

PROGRAM TO SOLVE EXAMPLE 4.2

```
      SUBROUTINE CROUT
            .
            .
            .
      DIMENSION A(50,51), X(50)
      DIMENSION XO(3),SUM(3),Q(3)
      N=3
      READ10,(XO(L),L=1,N),MJ
   10 FORMAT(3F10.0,I10)
      X(1)=0.
      J=1
      I=0
      T=0.
      DELT=.05
      NN=N-1
      DO17 L=1,NN
   17 SUM(L)=XO(L)
      DO18 LL=1,N
   18 SUM(N)=Q(LL)*XO(LL)
    9 CONTINUE
      Q(1)=-1.
      Q(2)=-2.*(1.+EXPEF(-T))
      Q(3)=-(3.+EXPEF(-T))
      A(1,1)=1.
      A(1,2)=-DELT/3.
      A(1,3)=0.
      A(2,1)=0.
      A(2,2)=1.
      A(2,3)=-DELT/3.
      A(3,1)=DELT/3.
      A(3,2)=-DELT/3.*Q(2)
      A(3,3)=1.-DELT/3.*Q(3)
      A(1,4)=XO(1)+DELT/3.*SUM(1)
      A(2,4)=XO(2)+DELT/3.*SUM(2)
      A(3,4)=XO(3)+DELT/3.*SUM(3)+T
      COMMON(A,X,N)
      CALLCROUT
      PRINT15,X(1),X(2),X(3),T
   15 FORMAT(4E22.8)
      R=0.
      IF(J-I)5,6,7
    7 DO11 K=1,NN
      KK=K+1
   11 SUM(K)=SUM(K)+4.*X(KK)
      DO12 M=1,N
   12 R=R+4.*Q(M)*X(M)
      SUM(N)=SUM(N)+R
      T=T+DELT
```

```
      I=I+2
      IF(I-MJ)9,9,6
    5 DO13 JJ=1,MM
      JJJ=JJ+1
   13 SUM(JJ)=SUM(JJ)+2.*X(JJJ)
      DO14 MM=1,N
   14 R=R+2.*COS(T*X(MM)
      SUM(I )=SUM(N )+R
      T=T+DELT
      J=J+2
      IF(T-MJ)9,9,6
    6 STOP
      END
```

REFERENCES

1. W. J. Cunningham, Nonlinear Analysis, McGraw-Hill Book Co., Inc., 1958.

2. L. A. Pipes, Four Methods for the Analysis of Time Varying Circuits, IRE Transactions on Circuit Theory, Mar., 1955, p. 4-12.

3. Hall, Maple and Vinograde, Introduction to the Laplace Transform, Appleton-Century-Crofts, Inc., 1959.

4. J. A. Aseltine, Transform Method for Time Varying Linear Systems, Journal of Applied Physics, June, 1954, p. 761-764.

5. F. R. Gerardi, Application of the Mellin and Hankel Transforms, IRE Transaction on Circuit Theory, June, 1959, p. 197-208.

6. J. E. Gibson, Nonlinear Automatic Control, McGraw-Hill Book Co., Inc., 1963.

7. W. E. Milne, Numerical Solution of Differential Equation, John Wiley & Sons, Inc., 1959.

8. J. B. Scarborough, Numerical Mathematical Analysis, Johns Hopkins Press, 1962.

9. W. Kutta, Zeitschrift für Math. und Phys., Vol. 46 (1901).

10. W. R. LePage, Complex Variables and the Laplace Transform for Engineers, McGraw-Hill Book Co., Inc., 1961.

11. R. Boxer and S. Thaler, A Simplified Method of Solving Linear and Nonlinear Systems, Proc. IRE, Vol. 44, 1956, p. 89-101.

12. J. T. Tou, Digital and Sampled-data Control Systems, McGraw-Hill Book Co., Inc., 1959.

13. B. C. Kuo, Analysis and Synthesis of Sampled-Data Control Systems, Prentice-Hall, Inc., 1963.

14. B. Friedman, Principles and Techniques of Applied Mathematics, John Wiley & Sons, Inc., 1956.

15. C. A. Desoer, An Introduction to State Space Techniques in Linear Systems, Joint Automatic Control Conference, 1962.

16. E. Coddington and N. Levinson, Theory of Ordinary Differential Equations, McGraw-Hill Book Co., Inc., 1955.

17. F. Kurzweil, Jr., The Analysis and Synthesis of Nonlinear Continuous and Sampled-data Systems Involving Saturation, Stanford Electronics Laboratory, Technical Report No. 2101-1, 1959.

18. J. T. Tou, Optimum Design of Digital Control System, Academic Press, 1963.