

FOR REFERENCE

NOT TO BE TAKEN FROM THIS ROOM

**EVALUATION OF TIME INTEGRATION ALGORITHMS
FOR THE SEMIDISCRETIZED HEAT CONDUCTION EQUATION**

by

SABRI ÜNCO

B.S. in M.E., Boğaziçi University, 1984

B.S. in Mathematics, Boğaziçi University, 1986

Bogazici University Library



39001100314197

14

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of

Master of Science

in

Mechanical Engineering

Boğaziçi University

1986

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis supervisor, Y.Doç.Dr. Vahan Kalenderođlu for his invaluable help, guidance and kind cooperation throughout this study, to Prof. Dr. Akın Tezel for his informative advices, to Mine Kalenderođlu for her kindness and patience in the painstaking typing of the manuscript, and to Engin Atik for his utmost help in the computer work.

Sabri Őncü

ABSTRACT

Finite element space discretization of heat conduction equation is briefly outlined, stability and convergence characteristics of relevant time integration methods are discussed. For the numerical integration of the semidiscretized heat conduction equation, several computer programs implementing the adopted one and two step methods are developed. A series of experiments are made and overall nine different integration algorithms are compared on the basis of accuracy and computational efficiency. Based on the results of the experiments, it has been concluded that the Crank-Nicholson method in conjunction with the smoothing process proposed by Zienkiewicz and the two-step Galerkin and Liniger methods with the starting procedure which involves starting the method with the known steady conditions prevailing for time less than zero are superior to the other methods investigated.

Ö Z E T

Isı iletimi denkleminin sonlu elemanlar yöntemi ile uzayda ayrıklanması kısaca özetlendi, uzayda ayrıklanmış ısı iletimi denkleminin zamanda tümlevlenmesi amacıyla kullanılan yöntemlerin kararlılık ve yakınsaklık özellikleri tartışıldı. Yarı-ayrıklanmış ısı iletimi denkleminin seçilen yöntemlerle zamanda tümlevlenmesi için bilgisayar programları geliştirildi. Yapılan deneylerle, toplam dokuz tümlev alma algoritması, doğruluk ve işlemsel verimlilik temelinde karşılaştırıldı. Deneyler sonucunda, sıfır anının yarım-zaman adımı gerisinden başlatılan Crank-Nicholson yöntemi ile sıfır anının bir zaman adımı gerisinden başlatılan iki adımlı Galerkin ve Liniger yöntemlerinin, diğer yöntemlere oranla daha iyi sonuç verdikleri görüldü.

TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
" ÖZET	v
LIST OF FIGURES	vi
LIST OF TABLES	vii
LIST OF SYMBOLS	xii
I. INTRODUCTION	1
II. THE SEMIDISCRETIZATION OF LINEAR TRANSIENT HEAT CONDUCTION PROBLEM	3
2.1 Classical Formulation of the Heat Conduction Problem	3
2.2 Weak Formulation of the Heat Conduction Problem and the Galerkin Approach	4
III. LINEAR MULTISTEP METHODS	9
3.1 Preliminary Definitions	10
3.2 Convergence, Consistency, Zero-Stability	16
3.3 Absolute Stability	24
3.4 Problem of Stiffness	31
IV. MULTISTEP DISCRETIZATION OF THE SEMIDISCRETIZED HEAT CONDUCTION EQUATION	39
4.1 The General Linear k-Step Method Applied to Heat Conduction Equation	39

	<u>Page</u>
4.2 One-Step Methods	42
4.3 Two-Step Methods	48
4.4 The Problem Associated with the Interpretation of the Forcing Functions	56
4.5 Smoothing Methods for Discontinuous Forcing Functions	59
V. RESULTS	64
VI. CONCLUSION	94
APPENDICES	
APPENDIX 1 - LINEAR DIFFERENCE EQUATIONS WITH CONSTANT COEFFICIENTS	96
APPENDIX 2 - DERIVATION DETAILS FOR ONE- AND TWO-STEP METHODS	100
REFERENCES	104

LIST OF FIGURES

	<u>Page</u>	
FIGURE 3.1	Successive steplength halving.	17
FIGURE 3.2	An absolute stability region.	33
FIGURE 3.3	Absolute stability region of A-stable methods.	36
FIGURE 3.4	Absolute stability region of $A(\alpha)$ -stable methods.	38
FIGURE 4.1	Basis functions for one-step methods.	42
FIGURE 4.2	Basis functions for two-step methods.	49
FIGURE 4.3	Step change in \tilde{h} or a_ρ .	60
FIGURE 4.4	Ramp change in \tilde{h} or a_ρ .	60
FIGURE 4.5	Smoothing of \tilde{h} or a_ρ as proposed by Zienkiewicz.	60
FIGURE 4.6	Smoothing of \tilde{h} or a_ρ by an exponential function.	61
FIGURE 4.7	Interpolation of the forcing function.	62
FIGURE 4.8	Interpolation of the forcing function.	62
FIGURE 5.1	Test problem.	65
FIGURE 5.2	10x10 Finite element mesh with triangular elements.	66

LIST OF TABLES

		<u>Page</u>
TABLE 5.1	One-step Methods, solution obtained for discontinuous boundary temperature, $x = (0.0,0.0)$, $\Delta t = 0.1$	68
TABLE 5.2	One-step Methods, solution obtained for discontinuous boundary temperature. $x = (0.5,0.5)$, $\Delta t = 0.1$	69
TABLE 5.3	One-step Methods, solution obtained for discontinuous boundary temperature, $x = (0.0,0.0)$, $\Delta t = 0.001$	70
TABLE 5.4	One-step Methods, solution obtained for discontinuous boundary temperature, $x = (0.5,0.5)$, $\Delta t = 0.001$	71
TABLE 5.5	One-step Methods, solution obtained for ramp smoothed boundary temperature, $x = (0.0,0.0)$, $\Delta t = 0.01$	72
TABLE 5.6	One-step Methods, solution obtained for ramp smoothed boundary temperature, $x = (0.5,0.5)$, $\Delta t = 0.01$	73
TABLE 5.7	One-step Methods, solution obtained for ramp smoothed boundary temperature, $x = (0.0,0.0)$, $\Delta t = 0.001$	74
TABLE 5.8	One-step Methods, solution obtained for ramp smoothed boundary temperature, $x = (0.5,0.5)$, $\Delta t = 0.001$	75
TABLE 5.9	One-step Methods, solution obtained for smoothed boundary temperature as proposed by Zienkiewicz, $x = (0.0,0.0)$, $\Delta t = 0.01$	76
TABLE 5.10	One-step Methods, solution obtained for smoothed boundary temperature as proposed by Zienkiewicz, $x = (0.5,0.5)$, $\Delta t = 0.01$	77
TABLE 5.11	Crank-Nicholson Method, solution obtained for averaging first two successive temperatures, $x = (0.0,0.0)$, $\Delta t = 0.01$	78

TABLE 5.12	Crank-Nicholson Method, solution obtained for averaging first two successive temperatures, $\tilde{x} = (0.5, 0.5)$, $\Delta t = 0.01$	79
TABLE 5.13	Crank-Nicholson Method, solution obtained for averaging first two successive temperatures, $\tilde{x} = (0.0, 0.0)$, $\Delta t = 0.001$	80
TABLE 5.14	Crank-Nicholson Method, solution obtained for averaging first two successive temperatures, $\tilde{x} = (0.5, 0.5)$, $\Delta t = 0.001$	81
TABLE 5.15	Crank-Nicholson method, solution obtained for exponential smoothed boundary temperature, $\tilde{x} = (0.0, 0.0)$, $\Delta t = 0.01$	82
TABLE 5.16	Crank-Nicholson Method, solution obtained for exponential smoothed boundary temperature, $\tilde{x} = (0.5, 0.5)$, $\Delta t = 0.01$	83
TABLE 5.17	Crank-Nicholson Method, solution obtained for exponential smoothed boundary temperature, $\tilde{x} = (0.0, 0.0)$, $\Delta t = 0.001$	84
TABLE 5.18	Crank-Nicholson Method, solution obtained for exponential smoothed boundary temperature, $\tilde{x} = (0.5, 0.5)$, $\Delta t = 0.001$	85
TABLE 5.19	Two-step Methods, solution obtained by starting with the Crank-Nicholson Method, $\tilde{x} = (0.0, 0.0)$, $\Delta t = 0.01$	86
TABLE 5.20	Two-step Methods, solution obtained by starting with the Crank-Nicholson Method, $\tilde{x} = (0.5, 0.5)$, $\Delta t = 0.01$	87
TABLE 5.21	Two-step Methods, solution obtained by starting with the Crank-Nicholson Method, $\tilde{x} = (0.0, 0.0)$, $\Delta t = 0.001$	88
TABLE 5.22	Two-step Methods, solution obtained by starting with the Crank-Nicholson Method, $\tilde{x} = (0.5, 0.5)$, $\Delta t = 0.001$	89
TABLE 5.23	Two-step Methods, solution obtained by starting with known steady conditions, $\tilde{x} = (0.0, 0.0)$, $\Delta t = 0.01$	90

	<u>Page</u>
TABLE 5.24 Two-step Methods, solution obtained by starting with known steady conditions, $\tilde{x} = (0.5, 0.5)$, $\Delta t = 0.01$	91
TABLE 5.25 Two-step Methods, solution obtained by starting with known steady conditions, $\tilde{x} = (0.0, 0.0)$, $\Delta t = 0.001$	92
TABLE 5.26 Two-step Methods, solution obtained by starting with known steady conditions, $\tilde{x} = (0.5, 0.5)$, $\Delta t = 0.001$	93

LIST OF SYMBOLS

\tilde{A}	Coefficient matrix of linear first order systems.
A	Coefficient matrix in Eq. 4.2.
\tilde{a}	Vector of unknown nodal parameters for temperature.
$\tilde{\tilde{a}}$	Vector of all nodal parameter for temperature.
C	Complex numbers.
\tilde{C}	Modified heat capacity matrix.
$\tilde{\tilde{C}}$	Heat capacity matrix.
C_q	Coefficients in Eq. 3.7.
\tilde{c}_l	Coefficient matrix of the derivative of specified nodal parameters.
e_n	Global error at time t_n .
\tilde{e}_n	Truncation error at time t_n .
\tilde{f}	Modified forcing vector.
f	Right-hand vector of first order systems.
$\tilde{\tilde{f}}$	Forcing vector.
g	Rate of heat generation per unit mass.
\underline{g}	Forcing vector in Eq. 4.2.
h	Heat transfer coefficient.
\tilde{h}	Condensated forcing vector.
I	Time interval.
$I_{\Delta t}$	Discrete time interval.
\tilde{J}	Jacobian matrix.

K	Modified heat conductivity matrix.
\tilde{K}	Heat conductivity.
K_N	N-dimensional Euclidean space.
k	Step number of linear multistep methods.
k_{ij}	Conductivity tensor.
k_ℓ	Coefficient vector of specified nodal parameters.
L	Lipschitz constant.
\tilde{L}	Difference operator associated with a linear multistep method.
N_i	Basis functions for finite element time discretization.
n	Time node.
\tilde{n}	Unit normal vector.
q_i	Heat flux vector.
q_n	Vector $\tilde{S}^{-1} e_n$
R	Absolute stability region.
R_{n+k}	Local round-off error at time t_{n+k} .
\mathbb{R}^+	Positive real numbers.
r	Parameter of stability polynomial.
r_n	Round-off error.
S	Boundary of the domain Ω .
\tilde{S}	Similarity transformation matrix.
S_1, S_2, S_3	Subboundaries of the boundary S .
t_n	Time at n^{th} time node.
u_i	Eigenvectors of the matrix A .
x	Point in three-dimensional Euclidean space.
x, z	Points in N-dimensional Euclidean space.
\tilde{x}	Unknown vector of first order systems.

$\alpha_\gamma, \beta_\gamma$	Coefficients of linear multistep methods.
β, γ	Parameters of Zienkiewicz three-level schemes.
Δt	Steplength.
$\tilde{\eta}$	Initial vector of first-order systems.
$\tilde{\eta}_i$	Starting vectors for linear multistep methods.
θ	Temperature distribution.
θ	Parameter of θ -methods.
θ_0	Initial temperature distribution.
θ_s	Specified temperature on S_1 .
θ_∞	Ambient temperature.
$\tilde{\Lambda}$	Diagonal form of of Jacobian matrix.
λ	Eigenvalue of the test equation.
λ_i	Eigenvalues of first-order systems.
ξ	Local parameters for finite element time discretization.
π	Stability polynomial.
ρ	Mass density.
ρ	First characteristic polynomial.
σ	Second characteristic polynomial.
ϕ	Trial function.
ϕ_ℓ	Basis functions for finite element space discretization.
ϕ_{n+k}	Local error at time t_{n+k} .
Ω	Bounded domain in three-dimensional Euclidean space.
ψ_{n+k}	Local truncation error.
ζ	Parameter of first and second characteristic polynomials.

I. INTRODUCTION

One of the exciting areas in heat transfer applications today is the digital simulation of the transient heat conduction problem. By means of digital simulation, it is possible to solve both linear and nonlinear problems with irregular geometries, which cannot be handled by the classical methods of analysis. In the modelling of the heat conduction problem for digital simulation, the governing equations are first discretized in space. This procedure is referred to as the semidiscretization and yields a system of first order differential equations in time. Because of its flexibility with regard to geometry and specification of initial and boundary conditions, the finite element method has been the most widely used method for the semidiscretization of the heat conduction equation.

The study of computational methods for the numerical integration of the semidiscrete systems has deepened considerably in recent years. In the mean time, several new approaches with greater performance and computational efficiency have been developed. Being of considerable importance, transient heat conduction problem has been one of the focal points of this study for the first order systems.

Although much development has been achieved during the last two decades, effective engineering analysis of many important problems is still hindered by the large cost of time integration of computer models. Thus, further study for understanding and improving numerical integration algorithms is still a necessary and important task to be taken over. Present state of the art of the numerical integration methods for the semidiscretized heat conduction equation is given in the works of Zienkiewicz [1,2], Zlamal [3], Wood [4], and Hogge [5].

The application of the finite element method to heat conduction problems was first made by Wilson and Nickell [6] for the linear transient case. The main development in the numerical integration of the semidiscretized heat conduction equation, on the other hand, is due to Zienkiewicz [1].

The purpose of this study is to discuss and compare several methods that have been proposed for the numerical integration of semidiscretized heat conduction equation with regard to computational efficiency and accuracy.

The present work consists of mainly three parts. The first part briefly deals with the statement of the heat conduction problem and its finite element discretization, the subject of Chapter two. The second part is devoted to the study of the multistep methods and their application to heat conduction problem, the subject of the Chapters three and four. The third part deals with the results of the experiments made with the one step θ -methods and the Zienkiewicz three-level methods, a brief summary of which can be found in Chapter four.

II. THE SEMIDISCRETIZATION OF LINEAR TRANSIENT HEAT CONDUCTION PROBLEM

2.1 CLASSICAL FORMULATION OF THE HEAT CONDUCTION PROBLEM

The initial-boundary-value problem associated with linear transient heat conduction is to find a temperature distribution $\theta = \theta(\underline{x}, t)$ in $\Omega \times [0, T)$, satisfying the heat conduction equation

$$(k_{ij}\theta_{,j})_{,i} + \rho g = c \frac{\partial \theta}{\partial t} \quad \text{in} \quad \Omega \times [0, T) \quad (2.1)$$

subject to the initial condition

$$\theta(\underline{x}, 0) = \theta_0 \quad \text{in} \quad \Omega \quad (2.2)$$

and the boundary conditions

$$\theta(\underline{x}, t) = \theta_s \quad \text{on} \quad S_1 \times (0, T) \quad (2.3)$$

$$(k_{ij}\theta_{,j} - q_i)n_i = 0 \quad \text{on} \quad S_2 \times (0, T) \quad (2.4)$$

$$k_{ij}\theta_{,j} - h(\theta_\infty - \theta) = 0 \quad \text{on} \quad S_3 \times (0, T) \quad (2.5)$$

where Ω is a bounded domain in three-dimensional Euclidean space;

$S = S_1 \cup S_2 \cup S_3$ is the boundary of the domain Ω with subboundaries S_1, S_2, S_3 ; $\underline{x} = (x_1, x_2, x_3)$ is a point of the domain Ω ; $[0, T)$ is a time interval where T is any positive scalar; $k_{ij} = k_{ij}(\underline{x})$, $i, j = 1, 2, 3$ is the distribution of the conductivity tensor in Ω ; $\rho = \rho(\underline{x})$ is the distribution of the mass density in Ω ; $C = C(\underline{x})$ is the distribution of the specific heat capacity in Ω ; $\theta_0 = \theta_0(\underline{x})$ is the initial temperature distribution in Ω ; $g = g(\underline{x}, t)$ is the distribution of the rate of heat generation per unit mass in Ω ; $\theta_s = \theta_s(\underline{x}, t)$ is the specified temperature distribution on $S_1 \times (0, T)$; $q_i = q_i(\underline{x}, t)$, $i = 1, 2, 3$ is the heat flux distribution on $S_2 \times (0, T)$; $h = h(\underline{x}, t)$ is the heat transfer coefficient distribution and $\theta_\infty = \theta_\infty(\underline{x}, t)$ is the ambient temperature distribution on $S_3 \times (0, T)$; $\underline{n} = (n_1, n_2, n_3)$ is the unit normal vector on boundary S .

2.2 WEAK FORMULATION OF THE HEAT CONDUCTION PROBLEM AND THE GALERKIN APPROACH

Many of the finite element methods that have been developed for linear transient heat conduction problem are based on the method of weighted residuals. Among the various forms of this method, Galerkin approach (or the Galerkin weighting procedure) is the most popular one used for the finite element space discretization of the heat conduction problem. The discretization of transient problems in space is called a semidiscretization and yields a set of ordinary differential equations in time.

In order to formulate the heat conduction problem in a weak form and to apply the Galerkin method, we introduce the space H^1 .

H^1 is the space of real functions which together with their first order derivatives are square integrable over Ω . Now, we consider the space V of trial functions ϕ , consisting of elements $\phi \in H^1$ such that $\phi|_{S_1} = 0$. Then the integral form of the heat conduction equation can be written as

$$\int_{\Omega} \phi \{ (k_{ij} \theta_{,j})_{,i} + \rho g - \rho c \frac{\partial \theta}{\partial t} \} d\Omega = \int_{S_2} \phi \{ (k_{ij} \theta_{,j} - q_i) n_i \} dS + \int_{S_3} \phi \{ (n_i k_{ij} \theta_{,j} - h(\theta_{\infty} - \theta)) \} dS \quad (2.6)$$

If we apply the Green-Gauss theorem to the first term of the first integral in Eq. (2.6), we get

$$\int_{\Omega} \phi (k_{ij} \theta_{,j})_{,i} d\Omega = - \int_{\Omega} \phi_{,i} k_{ij} \theta_{,j} d\Omega + \int_{S_2 \cup S_3} \phi n_i k_{ij} \theta_{,j} dS \quad (2.7)$$

Substituting Eq. (2.7) in Eq. (2.6) and, then, rearranging the terms in the resulting equation, we can write the weak form of the heat conduction problem as

$$\int_{\Omega} \phi_{,i} k_{ij} \theta_{,j} d\Omega + \int_{\Omega} \phi \rho c \frac{\partial \theta}{\partial t} d\Omega + \int_{S_3} \phi' h \theta dS = \int_{\Omega} \phi \rho g d\Omega + \int_{S_2} \phi n_i q_i dS + \int_{S_3} \phi h \theta_{\infty} dS \quad (2.8)$$

Now for any finite element partition of the domain of Ω , let h be a positive number which characterizes this partition and is defined as the largest diameter of all elements. Let $\{\phi_{\ell}(\underline{x})\}$ be a finite element C^0 basis such that each $\phi_{\ell}(\underline{x})$ vanishes over all but a fixed number of elements and is continuous across the boundaries of any two adjacent elements, and one of the nodal parameters of each

$\phi_\ell(\tilde{x})$ equals one at a node whereas the others are zero. Then we can approximate θ as

$$\theta \approx \hat{\theta} = \sum_{\ell=1}^n \phi_\ell(\tilde{x}) a_\ell(t) = \tilde{\phi}^T \tilde{a} \quad (2.9)$$

where all or some of the parameters $a_\ell(t)$ are unknowns depending on the existence of the specified temperature boundary condition.

It is important to note that the choice of basis functions are restricted to be from the space V of trial functions ϕ consisting of elements $\phi \in H^1$ such that $\phi|_{S_1} = 0$. On the other hand, the approximation

$$\theta \approx \hat{\theta} = \sum_{\ell} \phi_\ell(\tilde{x}) a_\ell(t)$$

requires that none of the basis functions vanish on the boundary S_1 . Therefore, approximation for θ must be in number equal only to the unknown nodal values. However, we will agree to form all the equations for all the nodal values, and subsequently to constrain the resulting equations such that the specified temperature boundary condition is satisfied. The reason for adopting this approach is that it is much easier to insert specified boundary temperatures at the end of discretization.

Substituting Eq. (2.9) in Eq. (2.8) and using the basis functions $\phi_1, \phi_2, \dots, \phi_n$ for ϕ , we obtain

$$\int_{\Omega} \tilde{\phi}^T \rho c \tilde{\phi} \frac{\partial \tilde{a}}{\partial t} d\Omega + \int_{\Omega} \tilde{\phi}_{,i} k_{ij} \tilde{\phi}_{,j} d\Omega + \int_{S_3} \tilde{\phi} h \tilde{\phi} dS = \int_{\Omega} \tilde{\phi} \rho g d\Omega + \int_{S_2} \tilde{\phi} n_i q_i dS + \int_{S_3} \tilde{\phi} h \theta_{\infty} dS \quad (2.10)$$

or, in matrix form, we get

$$\tilde{C}\tilde{a} + \tilde{K}\tilde{a} = \tilde{f} \quad (2.11)$$

where \tilde{C} is the heat capacity matrix defined as

$$\tilde{C} = \int_{\Omega} \rho c \phi \phi^T d\Omega$$

\tilde{K} is the heat conductivity matrix defined as

$$\tilde{K} = \int_{\Omega} \phi_{,i} k_{ij} \phi_{,j}^T d\Omega + \int_{S_2} h \phi \phi^T dS$$

and \tilde{f} is the forcing function defined as

$$\tilde{f} = \int_{\Omega} \rho g \phi d\Omega + \int_{S_2} \phi n_i q_i dS + \int_{S_3} \phi h \theta_{\infty} dS$$

The initial condition $\theta(x, 0) = \theta_0(x)$ on Ω gives the initial condition for $\tilde{a}(t)$

$$\tilde{a}(0) = \tilde{a}_0$$

where \tilde{a}_0 is the vector determined as $\theta_0(x) = \phi^T \tilde{a}_0$.

So far we have obtained the semidiscretized equation (2.11) of the heat conduction problem in space for all the nodal values. If the number of specified nodal parameters is m and the number of all nodal parameters is n , then the effect of modification for specified nodal parameters yields the system of first order differential equations for $(n-m)$ free nodal parameters defined as

$$\tilde{C}\tilde{a} + \tilde{K}\tilde{a} = \tilde{f} \quad (2.12)$$

where \tilde{a} is the vector of $(n-m)$ unknown nodal parameters, \tilde{C} and \tilde{K} are

condensated form of the matrices of \tilde{C} and \tilde{K} by simply deleting the rows and columns of \tilde{C} and \tilde{K} that corresponds to the specified nodal parameters. In this case the vector \tilde{f} takes of the form

$$\tilde{f} = \tilde{h} - a_{\ell\tilde{\ell}} c_{\ell} - a_{\ell\tilde{\ell}} k_{\ell} \quad , \quad \ell = 1, 2, \dots, m \quad (2.13)$$

where \tilde{h} is the condensated form of the forcing function \tilde{f} by simply deleting the rows of \tilde{f} that corresponds to specified nodal parameters, c_{ℓ} and k_{ℓ} are, respectively, the relevant partitions of the ℓ 'th columns of the matrices \tilde{C} and \tilde{K} .

III. LINEAR MULTISTEP METHODS

A large class of algorithms commonly used for the numerical integration of initial value problems is the class of linear multistep methods. In this work, we have employed a subclass of such methods for the solution of finite element semidiscretized heat conduction equation. It therefore seems appropriate to give a brief review of the properties of this class of methods, rather than dealing with specific properties of any member of it.

This chapter is intended to provide a summary of the fundamental concepts in the theory of linear multistep methods. The discussion is mainly based on the celebrated works of Henrici [7,8], Lambert [9] and Gear [10]. We begin, in Section 3.1, by introducing the preliminary definitions necessary for later considerations. In Section 3.2, we discuss the concepts of convergence, consistency and zero-stability. This enables the presentation of the fundamental theorem of Dahlquist [11], which establishes the necessary and sufficient conditions for convergence. The main subject of Section 3.3 is absolute stability. The study of absolute stability provides an answer for the question of accuracy, when the method is applied with a fixed steplength. Finally Section 3.4 deals with stiffness problem that arises in forcing the absolute stability criterion to be satisfied.

3.1 PRELIMINARY DEFINITIONS

Let K_N be the complex N -dimensional Euclidean space, \underline{z} be a point in K_N where $\underline{z} = [z_1, z_2, \dots, z_N]^T$. We set

$$||\underline{z}|| = |z_1| + |z_2| + \dots + |z_N|$$

Also let $I = [a, b]$ be a closed finite time interval where $-\infty < a < b < \infty$ and t be a typical point in I .

Consider the initial value problem for a system of first order differential equations

$$\begin{aligned} \dot{\underline{x}} &= \underline{f}(t, \underline{x}) & (\dot{\cdot} &\equiv \frac{d}{dt}) \\ \underline{x}(a) &= \underline{\eta} \end{aligned} \tag{3.1}$$

where \underline{x} and \underline{f} are functions from $I \times K_N$ to K_N .

THEOREM 1 (Henrici [7]): Let $\underline{f}(t, \underline{x})$ be defined and continuous for all points (t, \underline{x}) in $I \times K_N$ and let there exist a constant L , called the Lipschitz constant of \underline{f} . For every $t, \underline{x}, \underline{z}$ such that (t, \underline{x}) and (t, \underline{z}) are both in $I \times K_N$, if

$$||\underline{f}(t, \underline{x}) - \underline{f}(t, \underline{z})|| \leq L ||\underline{x} - \underline{z}|| \tag{3.2}$$

then, for all $\underline{\eta} \in K_N$ there exists a unique solution $\underline{x}(t)$ of the initial value problem (3.1) where $\underline{x}(t)$ is continuous and differentiable for all (t, \underline{x}) in $I \times K_N$.

Consider now the sequence of points

$$I_{\Delta t} = \{t_n = a + n\Delta t \mid n = 0, 1, \dots, (b-a)/\Delta t, \Delta t \in \mathbf{R}^+\}$$

on which we seek an approximate solution to $\underline{x}(t)$. The parameter Δt is called steplength and assumed to be constant throughout. Let us define \underline{x}_n as an approximation to $\underline{x}(t_n)$ and let $\underline{f}_n = f(t_n, \underline{x}_n)$.

Definition 1: A computational method of the form

$$\sum_{j=0}^k \alpha_j \underline{x}_{n+j} = \Delta t \sum_{j=0}^k \beta_j \underline{f}_{n+j} \quad (3.3)$$

to which the sequence $\{\underline{x}_n\}$ is a solution is called a linear k -step method or linear multistep method of step number k , where α_j and β_j are constants. We shall assume that $\alpha_k \neq 0$ and α_0 and β_0 are not both zero.

The polynomials $\rho(\zeta)$ and $\sigma(\zeta)$ defined as

$$\begin{aligned} \rho(\zeta) &= \sum_{j=0}^k \alpha_j \zeta^j \\ \sigma(\zeta) &= \sum_{j=0}^k \beta_j \zeta^j \end{aligned} \quad (3.4)$$

are called respectively, the first and second characteristic polynomials of the linear multistep method. A linear multistep method is called explicit if $\beta_k = 0$ and implicit if $\beta_k \neq 0$.

In order to compute the sequence $\{\underline{x}_n\}$ through Eq.(3.3), we need a set of starting vectors $\underline{x}_0, \underline{x}_1, \dots, \underline{x}_{k-1}$. Given any set of arbitrary starting vectors $\underline{x}_0, \underline{x}_1, \dots, \underline{x}_{k-1}$ for Eq. (3.3) Henrici [7] proves that there exists a unique solution $\{\underline{x}_n\}$ if Δt satisfies the condition

$$0 \leq \Delta t \leq \frac{1}{|\alpha_k \beta_k^{-1} L^{-1}|} = \Delta t_0$$

where L denotes the Lipschitz constant of \underline{f} .

Methods for obtaining starting vectors for the linear multi-step method (3.3) are called starting algorithms. Formally, starting procedures are defined as done in Henrici [7] as follows:

Definition 2: A starting procedure Σ is a set of vector-valued functions $\eta_0(\Delta t), \eta_1(\Delta t), \dots, \eta_{k-1}(\Delta t)$ defined for sufficiently small $\Delta t > 0$. The starting algorithm then consists of setting

$$\underline{x}_m = \underline{\eta}_m(\Delta t) \quad (m = 0, 1, \dots, k-1).$$

It should be noted that the above definition does not necessarily require to choose \underline{x}_0 as $\underline{\eta}$, namely the initial vector of the problem (3.1). However, in practice, the starting algorithm is frequently realized by setting $\underline{x}_0 = \underline{\eta}$ and calculating $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_{k-1}$ by some one-step method.

Let us define a vector-valued difference operator associated with a given multistep method as

$$\underline{L}\{\underline{y}(t); \Delta t\} = \sum_{j=0}^k [\alpha_j \underline{y}(t+j\Delta t) - \Delta t \beta_j \dot{\underline{y}}(t+j\Delta t)] \quad (3.5)$$

where $\underline{y}(t)$ is an arbitrary vector-valued function that is continuously differentiable on I . Expanding $\underline{y}(t+j\Delta t)$ and $\dot{\underline{y}}(t+j\Delta t)$ into Taylor's Series about t , substituting the results in Eq. (3.5) and collecting the like terms, we get

$$\underline{L}\{\underline{y}(t); \Delta t\} = C_0 \underline{y}(t) + C_1 \Delta t \underline{y}^{(1)}(t) + \dots + C_q \Delta t^q \underline{y}^{(q)}(t) + \dots \quad (3.6)$$

where

$$\begin{aligned}
 C_0 &= \sum_{j=0}^k \alpha_j \\
 C_1 &= \sum_{j=0}^k j\alpha_j - \sum_{j=0}^k \beta_j \\
 C_q &= \frac{1}{q!} \sum_{j=0}^k j^q \alpha_j - \frac{1}{(q-1)!} \sum_{j=0}^k j^{(q-1)} \beta_j
 \end{aligned} \tag{3.7}$$

Definition 3: The order of the difference operator $\underline{L}\{y(t); \Delta t\}$ and its associated linear multistep method is the largest integer p such that for all t

$$\|\underline{L}\{y(t); \Delta t\}\| = O(\Delta t^{p+1}).$$

It can be shown that the difference operator (3.5) and its associated linear multistep method is of order p if and only if $C_0 = C_1 = \dots = C_p = 0$, and $C_{p+1} \neq 0$. Thus if the method (3.3) is of order p , then

$$\underline{L}\{y(t); \Delta t\} = C_{p+1} \Delta t^{p+1} y^{(p+1)}(t) + O(\Delta t^{p+2}) \tag{3.8}$$

Let us reconsider the difference equation

$$\sum_{j=0}^k \alpha_j x_{n+j} = \Delta t \sum_{j=0}^k \beta_j f_{n+j} \tag{3.3}$$

Since Eq. (3.3) can be multiplied on both sides by the same constant without altering the relationship, the coefficients α_j and β_j are arbitrary to the extent of a constant multiplier. Thus by multiplying across Eq. (3.3) by an arbitrary constant, which will not alter the performance the linear multistep method, it is possible arbitrarily

to change the value of C_{p+1} in Eq. (3.8) and thus rob it of its significance. Fortunately, this arbitrariness can be removed by fixing α_k as 1 which is equivalent to divide across Eq. (3.3) by α_k which is normally different than one. The C_{p+1} term to be obtained from the resulting equation does not suffer from the arbitrariness discussed above, since the coefficients are not arbitrary anymore. On the other hand, it is also possible to leave the coefficients α_j and β_j arbitrary to the extent of a constant multiplier and consider the ratio

$$\frac{C_{p+1}}{k \sum_{j=0}^k \beta_j},$$

where C_{p+1} is the first nonvanishing term obtained from Eq. (3.3) without any restriction on the coefficients, rather than the coefficient C_{p+1} itself. It should be noted that although C_{p+1} is arbitrary to the extent of a constant multiplier, the ratio $C_{p+1} / \sum_{j=0}^k \beta_j$ is invariant since $\sum_{j=0}^k \beta_j$ is also arbitrary to the extent of the same constant multiplier.

Still another way of removing the arbitrariness of the coefficients in Eq. (3.3), is to agree to divide the Eq. (3.3) by $\sum_{j=0}^k \beta_j$. Then the resulting coefficients β_j , $j = 0, 1, \dots, k$ will satisfy the condition

$$\sum \beta_j = 1$$

and

$$C_{p+1} \left| \begin{array}{l} k \\ \sum_{j=0}^k \beta_j = 1 \end{array} \right. = \frac{C_{p+1} \left| \begin{array}{l} k \\ \text{coefficients arbitrary} \end{array} \right.}{\sum_{j=0}^k \beta_j}$$

Depending on the way in which the arbitrariness of the C_{p+1} term is removed, either the coefficient involved in this term or the ratio $C_{p+1} / \sum_{j=0}^k \beta_j$ is called the error constant. The error constant in the case of $\alpha_k = 1$ or $\sum_{j=0}^k \beta_j = 1$ will simply be the coefficient C_{p+1} obtained from the modified Eq. (3.3) such that either $\alpha_k = 1$ or $\sum_{j=0}^k \beta_j = 1$. The error constant in the case in which the coefficients are left arbitrary is

$$\frac{C_{p+1}}{\sum_{j=0}^k \beta_j}$$

where C_{p+1} is the coefficient associated with Eq. (3.3) in which the coefficients are arbitrary.

Let

$$e_n = \tilde{x}_n - x(t_n)$$

be the global error at $t_n \in I$ where \tilde{x}_n is the computed solution and $x(t_n)$ is the exact solution at t_n . For a given multistep method, it can be shown that

$$e_n = O(\Delta t^p)$$

i.e. the error approaches to zero with order Δt^p , where p is the order of the method. Thus, p may also be referred to as the order of accuracy or rate of convergence. Thus, the order of a multistep method give the functional dependence of error on Δt . The error constant, on the other hand, as discussed in Lambert [9], provides a measure for comparing the error in two different methods of the same order.

3.2 CONVERGENCE, CONSISTENCY, ZERO-STABILITY

Definition 4: The linear multistep method (3.3) is called convergent if, for all initial value problems (3.1) subjected to the hypothesis of Theorem 1, we have that

$$\lim_{\substack{\Delta t \rightarrow 0 \\ n\Delta t = t-a}} ||\tilde{x}_n - \tilde{x}(t_n)|| = 0$$

holds for all $t \in I$ and for all solutions $\{\tilde{x}_n\}$ of the Eq. (3.3) satisfying the starting conditions

$$\tilde{x}_m = \eta_m(\Delta t) \quad \text{for which} \quad \lim_{\Delta t \rightarrow 0} \eta_m(\Delta t) = \tilde{\eta}, \quad m = 0, 1, \dots, k-1.$$

Lambert [9] explains the above definition by the following argument:

"It is inappropriate to consider n as remaining fixed while $\Delta t \rightarrow 0$. For example, consider a fixed point $t = \tau$, and let the initial choice of steplength Δt_1 be such that $\tau = a + 2\Delta t_1$. In the special case when the steplength is successively halved, the situation illustrated by Fig. 3.1 holds.

If we use the notation $\tilde{x}_n(\Delta t)$ to denote the value \tilde{x}_n given by the linear multistep method (3.3) when the steplength is Δt , then we are interested not in the convergence of the sequence $\tilde{x}_2(\Delta t_1), \tilde{x}_2(\Delta t_1/2), \tilde{x}_2(\Delta t_1/4), \dots$, but in the convergence of the sequence $\tilde{x}_2(\Delta t_1), \tilde{x}_4(\Delta t_1/2), \tilde{x}_8(\Delta t_1/4), \dots$, to $\tilde{x}(\tau)$. Thus the limit

$$\lim_{\substack{\Delta t \rightarrow 0 \\ n\Delta t = t-a}}$$

means, $n \rightarrow \infty$ as $\Delta t \rightarrow 0$ and $n\Delta t = t-a$ remains fixed."

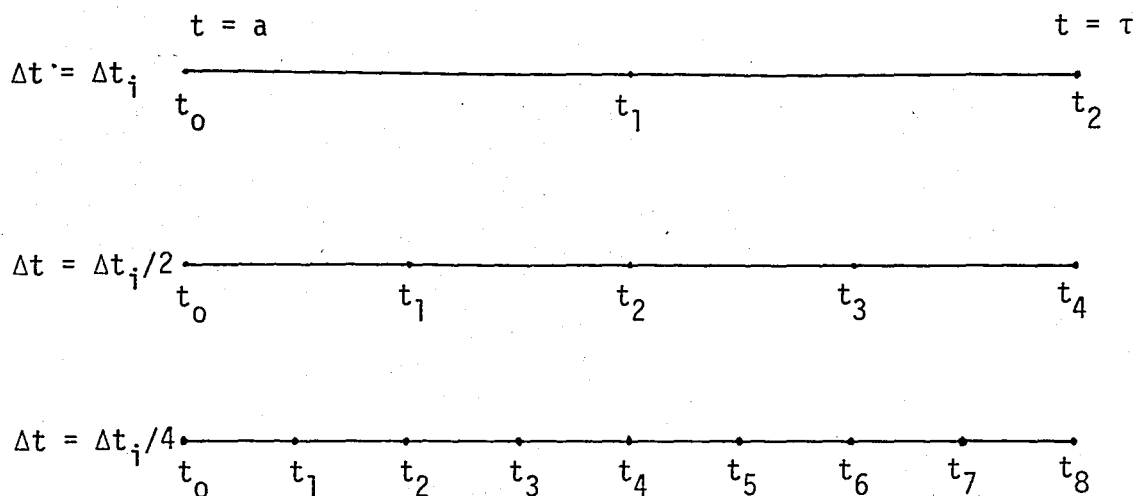


FIGURE 3.1 - Successive steplength halving.

Definition 5: A linear k -step method is called consistent if for any solution $\tilde{x}(t)$ of the initial value problem (3.1) subjected to the hypothesis of Theorem 1 we have that

$$\max_{n=0,1,\dots,n_0} \|\tilde{L}\{\tilde{x}(t_n); t\}\| = o(\Delta t)$$

where $n_0 = [(b-a)/\Delta t - k]$.

The following theorem given in Henrici [7] proves that the consistency of a linear multistep method is equivalent to purely algebraic conditions imposed on the first and second characteristic polynomials of the method.

THEOREM 2: A linear multistep method is consistent if and only if the polynomials $\rho(\zeta)$ and $\sigma(\zeta)$ satisfy the following conditions

$$\rho(1) = 0 \quad , \quad \rho'(1) = \sigma(1) \quad \left(' \equiv \frac{d}{d\zeta} \right) ,$$

or more explicitly,

$$\sum_{j=0}^k \alpha_j = 0 \quad , \quad \sum_{j=0}^k j\alpha_j = \sum_{j=0}^k \beta_j \quad (3.9)$$

Therefore, the condition (3.9) is equivalent to say that $C_0 = C_1 = 0$ where C_0 and C_1 are the constants given by the identities (3.7). Hence, any consistent linear multistep method is of order at least one.

In order to clarify the concept of consistency, let us continue with the following argument;

Assume that the linear multistep method (3.3) is convergent.

i) Consider the scalar initial value problem $\dot{x}(t) = 0$, $x(0) = 1$, whose exact solution is $x(t) = 1$. If we apply the method (3.3) to this problem we get the following difference equation;

$$\alpha_k x_{n+k} + \alpha_{k-1} x_{n+k-1} + \dots + \alpha_0 x_n = 0 \quad (3.10)$$

Let us choose the starting values $x_m = 1$, $m = 0, 1, \dots, k-1$, which automatically satisfy the condition $\lim_{\Delta t \rightarrow 0} \eta_m(\Delta t) = 1$, $m = 0, 1, \dots, k-1$.

As the method is convergent, we must have

$$\lim_{\substack{\Delta t \rightarrow 0 \\ n\Delta t = t}} x_n = 1$$

holds for all $t > 0$. Since the solution, x_n , of the difference

Eq. (3.10) does not depend on Δt , for any Δt , $x_n = 1$ for all t . Thus for any Δt , at any stage of computation $x_{n+k} = x_{n+k-1} = \dots = x_n = 1$ which in turn requires that

$$\alpha_k + \alpha_{k-1} + \dots + \alpha_0 = 0,$$

the first condition of consistency.

ii) Consider the scalar initial value problem $\dot{x}(t) = 1$, $x(0) = 0$, whose exact solution is $x(t) = t$. Then the Eq. (3.3) takes the form

$$\alpha_k x_{n+k} + \alpha_{k-1} x_{n+k-1} + \dots + \alpha_0 x_n = \Delta t (\beta_k + \beta_{k-1} + \dots + \beta_0) \quad (3.11)$$

Since the method (3.3) is assumed to be convergent, every solution started with $x_m = \eta_m(\Delta t)$ such that

$$\lim_{\Delta t \rightarrow 0} \eta_m(\Delta t) = 0, \quad m = 0, 1, \dots, k-1 \quad (3.12)$$

must also satisfy

$$\lim_{\substack{\Delta t \rightarrow 0 \\ n\Delta t = t}} x_n = t \quad (3.13)$$

Let the sequence $\{x_n\}$ be a solution of Eq. (3.11) defined as $x_n = Kn\Delta t$, where $K > 0$ is a constant. The starting condition (3.12) is immediately satisfied by $x_m = \eta_m(\Delta t) = Km\Delta t$, $m = 0, 1, \dots, k-1$, since $\lim_{\Delta t \rightarrow 0} Km\Delta t = 0$ for all $m = 0, 1, \dots, k-1$.

If we substitute $x_{n+j} = K(n+j)\Delta t$, $j = 0, 1, \dots, k$ in Eq. (3.11) we get

$$K\Delta t \{(n+k)\alpha_k + (n+k-1)\alpha_{k-1} + \dots + n\alpha_0\} = \Delta t (\beta_k + \beta_{k-1} + \dots + \beta_0)$$

or equivalently

$$\begin{aligned} K\{k\alpha_k + (k-1)\alpha_{k-1} + \dots + \alpha_1\} + K_n\{\alpha_k + \alpha_{k-1} + \dots + \alpha_0\} \\ = \beta_k + \beta_{k-1} + \dots + \beta_0 \end{aligned} \quad (3.14)$$

As it was shown in part (i), for any convergent linear multistep method, $\alpha_k + \alpha_{k-1} + \dots + \alpha_0 = 0$. Therefore, Eq. (3.14) reduces to

$$K\{k\alpha_k + (k-1)\alpha_{k-1} + \dots + \alpha_1\} = \beta_k + \beta_{k-1} + \dots + \beta_0 \quad (3.15)$$

Assume that $k\alpha_k + (k-1)\alpha_{k-1} + \dots + \alpha_1 = 0$. Then, the equality (3.15) holds if and only if $\beta_k + \beta_{k-1} + \dots + \beta_0 = 0$. In this case, the sequence $\{x_n\}$ defined as $x_n = Kn\Delta t$ will be the solution of (3.11) for every $K > 0$. This is a contradiction since,

$$\lim_{\substack{\Delta t \rightarrow 0 \\ n\Delta t = t}} Kn\Delta t = Kt \neq t$$

unless $K = 1$. Therefore we must have that $k\alpha_k + (k-1)\alpha_{k-1} + \dots + \alpha_1 \neq 0$.

If we solve K from Eq. (3.15) we obtain

$$K = \frac{\beta_k + \beta_{k-1} + \dots + \beta_0}{k\alpha_k + (k-1)\alpha_{k-1} + \dots + \alpha_1}$$

and the condition (3.13) is satisfied if and only if $K = 1$. Hence, we reach the identity

$$\sum_{j=0}^k j\alpha_j = \sum_{j=0}^k \beta_j$$

which is the second condition of consistency.

We can conclude from the above argument that any convergent linear multistep method is necessarily consistent. However, as it

will become clear later, consistency alone is not a sufficient condition for convergence.

Definition 6: A linear multistep method is called zero-stable if, for all functions, f subjected to the condition of Theorem 1, the following statement holds:

Let $\eta_0(\Delta t), \eta_1(\Delta t), \dots, \eta_{k-1}(\Delta t)$ denote k functions from the interval $(0, \Delta t_0)$ (where Δt_0 is previously defined as $\Delta t_0 = |\alpha_k \beta_k^{-1} L^{-1}|$) to K_N such that

$$\|\eta_m(\Delta t)\| \leq M \quad 0 < \Delta t < \Delta t_0, \quad m = 0, 1, \dots, k-1$$

for some constant M , and denote, for each $\Delta t \in (0, \Delta t_0)$, the solution of Eq. (3.3) having the starting values

$$\tilde{x}_m = \eta_m(\Delta t), \quad m = 0, 1, \dots, k-1$$

by $\{x_n(\Delta t)\}$. Then the family of these solutions is uniformly bounded as $\Delta t \rightarrow 0$, $n\Delta t = t$ there exists a constant M' such that

$$\max_{\substack{0 < \Delta t < \Delta t_0 \\ n=0, 1, \dots, (b-a)/\Delta t}} \|x_n(\Delta t)\| \leq M'$$

A zero-stable linear multistep method is frequently called simply "stable". The word "zero" was suggested by Lambert [9], since zero-stability is related with the limiting process as $\Delta t \rightarrow 0$. Some authors also call zero-stability as Dahlquist- (or shorty D-)stability since it was originally defined by Dahlquist.

By the following theorem, Henrici [7] proves that the zero-stability of a linear multistep method is equivalent to the purely algebraic property of the first characteristic polynomial of the method.

THEOREM 3: A linear multistep method is zero-stable if and only if the roots of the polynomial

$$\rho(\zeta) = \alpha_k \zeta^k + \alpha_{k-1} \zeta^{k-1} + \dots + \alpha_0$$

satisfy $|\zeta_i| \leq 1$, $i = 1, 2, \dots, k$, and the roots satisfying $|\zeta_i| = 1$ have multiplicity one. The above condition is frequently called as the root condition.

In order to show the significance of zero-stability, we will proceed with the following discussion:

Again assume that the linear multistep method (3.3) is convergent. Consider the scalar initial value problem $\dot{x}(t) = 0$, $x(0) = 0$, whose exact solution is $x(t) = 0$. If we apply the method (3.3) to that problem, we get the difference equation with constant coefficients

$$\alpha_k x_{n+k} + \alpha_{k-1} x_{n+k-1} + \dots + \alpha_0 x_n = 0 \quad (3.16)$$

As the method (3.3) is convergent, for all solutions of Eq. (3.10) started with $x_m = \eta_m(\Delta t)$ such that

$$\lim_{\Delta t \rightarrow 0} \eta_m(\Delta t) = 0, \quad m = 0, 1, \dots, k-1 \quad (3.17)$$

must also satisfy

$$\lim_{\substack{\Delta t \rightarrow 0 \\ n\Delta t = t}} x_m = 0 \quad (3.18)$$

for all $t > 0$.

A detailed discussion on the solution of linear difference equations with constant coefficients is given in Appendix 1. In view of the analysis given in Appendix 1, let us examine the following two cases;

i) Let ζ_r be a real root of $\rho(\zeta)$ with multiplicity one. Then the sequence $\{x_n\}$ defined as $x_n = \Delta t \zeta_r^n$ is a solution of Eq. (3.16). The condition (3.17) is also satisfied by $x_m = \eta_m(\Delta t) = \Delta t \zeta_r^m$, $m = 0, 1, \dots, k-1$, since $\lim_{\Delta t \rightarrow 0} \Delta t \zeta_r^m = 0$ for all $m = 0, 1, \dots, k-1$. Note that

$$\lim_{\substack{\Delta t \rightarrow 0 \\ n\Delta t = t}} \Delta t \zeta_r^n = t \lim_{n \rightarrow \infty} (\zeta_r^n / n) = 0 \quad \text{iff} \quad |\zeta_r| \leq 1.$$

Hence the condition (3.18) holds if and only if $|\zeta_r| \leq 1$.

ii) Let ζ_r be a real root of $\rho(\zeta)$ with multiplicity greater than one. Then the sequence $\{x_n\}$ defined as $x_n = \Delta t n \zeta_r^n$ is a solution of Eq. (3.16). Again the condition (3.17) is satisfied by $x_m = \eta_m(\Delta t) = \Delta t m \zeta_r^m$, $m = 0, 1, \dots, k-1$, since $\lim_{\Delta t \rightarrow 0} \Delta t m \zeta_r^m = 0$ for all $m = 0, 1, \dots, k-1$. Note that, in this case

$$\lim_{\substack{\Delta t \rightarrow 0 \\ n\Delta t = t}} \Delta t n \zeta_r^n = t \lim_{n \rightarrow \infty} \zeta_r^n = 0 \quad \text{iff} \quad |\zeta_r| < 1.$$

Hence, the condition (3.18) holds if and only if $|\zeta_r| < 1$.

It can be found in Henrici [8] that, for a convergent linear multistep method, the root condition is also satisfied when the roots of $\rho(\zeta)$ are complex. Hence, zero-stability is a necessary condition for convergence.

We will complete Section (3.2) by stating the fundamental theorem of Dahlquist, whose proof is also given in Henrici [7].

THEOREM 4: A linear multistep method is convergent if and only if it is consistent and zero-stable.

3.3 ABSOLUTE STABILITY

Upto this point, we have only dealt with the behaviour of the approximate solutions obtained by linear multistep methods as $\Delta t \rightarrow 0$. However, in practice, we compute approximate solutions with a fixed positive Δt . In order to decide on the accuracy of computed solutions when the steplength is fixed, we have to get some information about the error.

There are mainly two types of errors in any method of approximate solution. The first is the truncation error, and the second is the round-off error.

The truncation error of the linear multistep method is defined by

$$\tilde{\epsilon}_n = x_n - \tilde{x}(t_n), \quad t_n \in I$$

This error is unique for the initial value problem, the linear multistep method and the starting procedure.

The round-off error is that kind of error originating from discretizing the number system, hence depends also on the computer and the details of the programs used. It is defined by

$$r_n = \tilde{x}_n - x_n$$

where \tilde{x}_n is the computed solution at t_n . Therefore, we can define the global error as

$$e_n = \tilde{e}_n + r_n = \tilde{x}_n - x(t_n)$$

Let us assume that we have no round-off error and no previous truncation errors have been made. Then the expression $L\{\tilde{x}_n; \Delta t\}$ given by Eq. (3.5) will be the local truncation error of the associated linear multistep method at t_{n+k} , when $x(t)$ is the theoretical solution. We put

$$\psi_{n+k} = L\{x(t_n); \Delta t\} \quad (3.19)$$

for the local truncation error.

Let us assume that we have no truncation error and no previous round-off errors have been made. Then the expression $L\{\tilde{x}_n; \Delta t\}$ given by Eq. (3.5) will be the local round-off error of the associated linear multistep method at t_{n+k} , when $\{\tilde{x}_n\}$ is the computed solution. We put

$$R_{n+k} = L\{\tilde{x}_n; \Delta t\} = \sum_{j=0}^k \{\alpha_j \tilde{x}_{n+j} - \Delta t \beta_j f(\tilde{x}_{n+j}, t_{n+j})\} \quad (3.20)$$

Rewriting the Eqs. (3.19) and (3.20) as

$$\sum_{j=0}^k \alpha_j x(t_{n+j}) = \Delta t \sum_{j=0}^k \beta_j f(t_{n+j}, x(t_{n+j})) + \psi_{n+k} \quad (3.21)$$

$$\sum_{j=0}^k \alpha_j \tilde{x}_{n+j} = \Delta t \sum_{j=0}^k \beta_j f(t_{n+j}, \tilde{x}_{n+j}) + R_{n+k} \quad (3.22)$$

and subtracting Eq. (3.21) from Eq. (3.22) we get

$$\sum_{j=0}^k \alpha_j e_{n+j} = \Delta t \sum_{j=0}^k \beta_j \{f(t_{n+j}, \tilde{x}_{n+j}) - f(t_{n+j}, x(t_{n+j}))\} + \phi_{n+k} \quad (3.23)$$

where $e_{n+j} = \tilde{x}_{n+j} - x(t_{n+j})$ and $\phi_{n+k} = R_{n+k} - \psi_{n+k}$. If we apply the mean value theorem for a function of several variables to f in Eq.

(3.23) (See Lambert [9], page 220), we obtain

$$\sum_{j=0}^k \alpha_j e_{n+j} = \Delta t \sum_{j=0}^k \beta_j (\partial f / \partial x) e_{n+j} + \phi_{n+k} \quad (3.24)$$

where $\partial f / \partial x$ is the $N \times N$ Jacobian matrix of f .

Let us assume that $\partial f / \partial x = J$ is a constant $N \times N$ matrix. Then we can rewrite Eq. (3.24) as

$$\sum_{j=0}^k (\alpha_j I - \Delta t \beta_j J) e_{n+j} = \phi_{n+k} \quad (3.25)$$

where I is the $N \times N$ unit matrix. If we further assume that the eigenvalues of the matrix J are distinct, then there exists a similarity transformation which reduces J to diagonal form such that

$$\tilde{S}^{-1} J \tilde{S} = \tilde{\Lambda} = \begin{vmatrix} \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 \\ \vdots & \vdots & & & \\ \vdots & \vdots & & & \\ 0 & 0 & \dots & 0 & \lambda_N \end{vmatrix}$$

where λ_ℓ , $\ell = 1, 2, \dots, N$ are the eigenvalues of J .

Defining q_n by $q_n = \tilde{S}^{-1} e_n$ and H_{n+k} by $H_{n+k} = \tilde{S}^{-1} \phi_{n+k}$, we get

$$\sum_{j=0}^k (\alpha_j \tilde{I} - \Delta t \beta_j \tilde{\Lambda}) \tilde{q}_{n+j} = \tilde{H}_{n+k} \quad (3.26)$$

Since \tilde{I} and $\tilde{\Lambda}$ are diagonal matrices, the components of Eq. (3.26) are uncoupled, hence we may write Eq. (3.26) as

$$\sum_{j=0}^k (\alpha_j - \Delta t \beta_j \lambda_\ell) q_{n+j, \ell} = h_{n+k, \ell}, \quad \ell = 1, 2, \dots, N \quad (3.27)$$

where $q_{n, \ell}$ and $h_{n+k, \ell}$ are the components of \tilde{q}_n and \tilde{H}_{n+k} respectively.

At this point, it is convenient to define the polynomial

$\pi(r, \Delta t \lambda)$ as

$$\pi(r, \Delta t \lambda) = \sum_{j=0}^k (\alpha_j - \Delta t \beta_j \lambda) r^j \quad (3.28)$$

or, in terms of the first and second characteristic polynomials of the linear multistep method, as

$$\pi(r, \Delta t \lambda) = \rho(r) - \Delta t \lambda \sigma(r) \quad (3.29)$$

The polynomial $\pi(r, \Delta t \lambda)$ is frequently called as the stability polynomial.

The solution of Eq. (2.27) is, then

$$q_{n, \ell} = \sum_{j=1}^s \sum_{i=0}^{p_j-1} C_{j, i+1} [n! / (n-i)!] r_{\ell, j}^n + d_{n, \ell}, \quad \ell = 1, 2, \dots, N \quad (3.30)$$

where $r_{\ell, j}$, $j = 1, 2, \dots, s$ are the roots of polynomial $\pi(r_\ell, \Delta t \lambda_\ell)$, each root $r_{\ell, j}$ being a root of multiplicity p_j where $\sum_{j=1}^s p_j = k$, and $d_{n, \ell}$ is some particular solution of Eq. (3.27). Here $C_{j, i}$, $i = 1, 2, \dots, p_j$, $j = 1, 2, \dots, s$ are arbitrary constants (See Appendix 1).

If the system (3.1) is linear i.e.,

$$\dot{\tilde{x}} = \tilde{A} \tilde{x} + \tilde{\phi}(t)$$

then, we can replace Eqs. (3.21) and (3.22) by

$$\sum_{j=0}^k \alpha_j x_{n+j} = \Delta t \sum_{j=0}^k \beta_j [Ax_{n+j} + \tilde{\phi}(t_{n+j})] + \psi_{n+k} \quad (3.31)$$

$$\sum_{j=0}^k \alpha_j \tilde{x}_{n+j} = \Delta t \sum_{j=0}^k \beta_j [A\tilde{x}_{n+j} + \tilde{\phi}(t_{n+j})] + R_{n+k} \quad (3.32)$$

Subtracting Eq. (3.31) from Eq. (3.32) we get

$$\sum_{j=0}^k \alpha_j e_{n+j} = \Delta t \sum_{j=0}^k \beta_j A e_{n+j} + \phi_{n+k} \quad (3.33)$$

or

$$\sum_{j=0}^k (\alpha_j - \Delta t \beta_j A) e_{n+j} = \phi_{n+k} \quad (3.34)$$

where $\phi_{n+k} = R_{n+k} - \psi_{n+k}$

Let us assume that the eigenvalues of A are distinct. Then we can diagonalize A by a similarity transformation such as

$$\tilde{S}^{-1} A \tilde{S} = \tilde{\Lambda} = \begin{vmatrix} \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \lambda_N \end{vmatrix}$$

where λ_ℓ , $\ell = 1, 2, \dots, N$ are the eigenvalues of A . Defining q_n by

$q_n = \tilde{S}^{-1} e_n$ and H_{n+k} by $H_{n+k} = \tilde{S}^{-1} \phi_{n+k}$, we get

$$\sum_{j=0}^k (\alpha_j I - \Delta t \beta_j \tilde{\Lambda}) q_{n+j} = H_{n+k} \quad (3.35)$$

or in the component form

$$\sum_{j=0}^k (\alpha_j I - \Delta t \beta_j \lambda_\ell) q_{n+j, \ell} = h_{n+k, \ell}, \quad \ell = 1, 2, \dots, N \quad (3.35)$$

where $q_{n,\ell}$ and $h_{n+k,\ell}$ are the components of q_n and H_{n+k} respectively.

The solution of Eq. (3.36) is then

$$q_{n,\ell} = \sum_{j=1}^s \sum_{\ell=0}^{p_j-1} C_{j,\ell+1} [n!/(n-1)!] r_{\ell,j}^n + d_{n,\ell}, \quad \ell = 1, 2, \dots, N \quad (3.37)$$

where $r_{\ell,j}$, $j = 1, 2, \dots, s$ are the roots of polynomial $\pi(r_{\ell}, \Delta t \lambda_{\ell})$, each root $r_{\ell,j}$ being a root of polynomial $\pi(r_{\ell}, \Delta t \lambda_{\ell})$ where $\sum_{j=1}^s p_j = k$, and $d_{n,\ell}$ is some particular solution. Here $C_{j,i}$, $i = 1, 2, \dots, p_j$, $j = 1, 2, \dots, s$ are arbitrary constants.

Note that, in both cases, if q_n grows or decays with n , then e_n does and vice versa. We therefore define a linear multistep absolutely stable if, for a given initial value problem and a given positive Δt , the global error remains bounded as the computation proceeds.

It is evident from Eqs. (3.30) and (3.37) that, in both cases, if any $|r_{\ell,j}| \geq 1$, then the global error will grow as increases. Hence, the method will be absolutely unstable for that $\Delta t \lambda_{\ell}$. Thus absolute stability can formally be defined as follows.

Definition 7: A linear multistep method is said to be absolutely stable for $\lambda \Delta t$, such that λ is any eigenvalue of the Jacobian of the function f or in the linear case is any eigenvalue of the coefficient matrix A , if all roots of the stability polynomial $\pi(r, \lambda \Delta t)$ associated with the method satisfy

$$|r_{\ell,j}| < 1, \quad j = 1, 2, \dots, k, \quad \ell = 1, 2, \dots, N$$

Obviously, the above definition is strictly dependent on the initial value problem to which we apply the linear multistep method.

Let us now consider the differential equation

$$\dot{\tilde{x}} = \lambda \tilde{x} \quad (3.38)$$

where λ is a complex constant. For this Eqs.(3.19) and (3.20) reduces to

$$\sum_{j=0}^k \alpha_j \tilde{x}(t_{n+j}) = \Delta t \sum_{j=0}^k \beta_j \lambda \tilde{x}(t_{n+j}) + \psi_{n+k} \quad (3.39)$$

$$\sum_{j=0}^k \alpha_j \tilde{x}_{n+j} = \Delta t \sum_{j=0}^k \beta_j \lambda \tilde{x}_{n+j} + R_{n+k} \quad (3.40)$$

respectively. Subtracting Eq. (3.39) from Eq. (3.40), we get

$$\sum_{j=0}^k \alpha_j e_{n+j} = \Delta t \sum_{j=0}^k \beta_j \lambda e_{n+j} + \phi_{n+k} \quad (3.41)$$

or, in component form,

$$\sum_{j=0}^k \alpha_j e_{n+j,i} = \Delta t \sum_{j=0}^k \beta_j \lambda e_{n+j,i} + \phi_{n+k,i}, \quad i = 1, 2, \dots, N \quad (3.42)$$

Thus

$$\sum_{j=0}^k (\alpha_j - \Delta t \lambda \beta_j) e_{n+j,i} = \phi_{n+k,i}, \quad i = 1, 2, \dots, N \quad (3.43)$$

which gives us the stability polynomial

$$\sum_{j=0}^k (\alpha_j - \Delta t \lambda \beta_j) r^j = 0$$

The region $R \subset \mathbf{C}$ defined as

$$R = \{ \lambda \Delta t \in \mathbf{C} \mid |r_j| < 1, r_j \text{ being a root of the stability polynomial, } j = 1, 2, \dots, k \}$$

is called the absolute stability region of the employed multistep method.

Clearly, if the method is absolutely stable in the region R , then the region

$$R' = \{ \lambda_\ell \Delta t \in \mathbf{C}, \ell = 1, 2, \dots, N \mid |r_{\ell, j}| < 1, r_{\ell, j} \text{ being a root} \\ \text{of the stability polynomial, } \ell = 1, 2, \dots, N, \\ j = 1, 2, \dots, k \}$$

is a subset of R . It should be noted that the absolute stability region of the multistep is defined in term of the differential equation (3.38). This may give the false impression that the definition of the region is in terms of a specific problem. That this is not the case, it follows from the fact that both λ and Δt arbitrary. It should be kept in mind that though λ and Δt are arbitrary $\lambda \Delta t$ is not.

We could have defined the absolute stability region of the method in terms of the region R' . This however will restrict the region to the specific initial value problem involved and therefore this approach is deliberately avoided. The equation

$$\dot{\tilde{x}} = \lambda \tilde{x}$$

is referred to as the "test equation".

3.4 THE PROBLEM OF STIFFNESS

Let us consider the $N \times N$ linear system of equations

$$\dot{\tilde{x}} = \underline{\underline{A}}\tilde{x} + \underline{\underline{\phi}}(t) \quad (3.44)$$

and assume that the matrix $\underline{\underline{A}}$ has distinct eigenvalues λ_i , $i = 1, 2, \dots, N$. Let $\underline{\underline{C}}_i$ be the corresponding eigenvector of the eigenvalue λ_i . Then, it can be shown that Eq. (3.44) has a general solution of the form

$$\tilde{x}(t) = \sum_{i=1}^N k_i e^{\lambda_i t} \tilde{c}_i + \underline{\psi}(t)$$

where k_i , $i = 1, 2, \dots, N$ are constants and $\underline{\psi}(t)$ is the particular solution. If

$$\operatorname{Re} \lambda_i < 0 \quad i = 1, 2, \dots, N$$

then the term

$$\sum_{i=1}^N k_i e^{\lambda_i t} \tilde{c}_i \rightarrow 0 \quad \text{as } t \rightarrow \infty$$

It is for this reason that the term $\sum_{i=1}^N k_i e^{\lambda_i t} \tilde{c}_i$ is called the transient term and the particular is called the steady-state term. It is important that the steady-state term should not be confused with the steady-state solution. The latter is the limiting value of the former as $t \rightarrow \infty$.

Let λ_m and λ_n be two eigenvalues of \underline{A} such that

$$|\operatorname{Re} \lambda_m| \leq |\operatorname{Re} \lambda_i| \leq |\operatorname{Re} \lambda_n| \quad , \quad i = 1, 2, \dots, N.$$

If our aim is to find numerically the steady-state solution, then we must pursue the numerical solution of Eq. (3.44) at least until the slowest decaying exponential in the transient term, namely $e^{\lambda_m t}$, is negligible. Thus the smaller $|\operatorname{Re} \lambda_m|$, the longer will be the range of integration. On the other hand, in the case of a method with absolute stability region as shown in Fig. 3.2, the presence of eigenvalues of \underline{A} far out to the left in the complex plane, i.e. far away from the imaginary axis, will force us to use excessively small steplength in order that $\lambda_i \Delta t$, $i = 1, 2, \dots, N$, will lie within the region of absolute stability region of the method of our choice (unless, of course, the

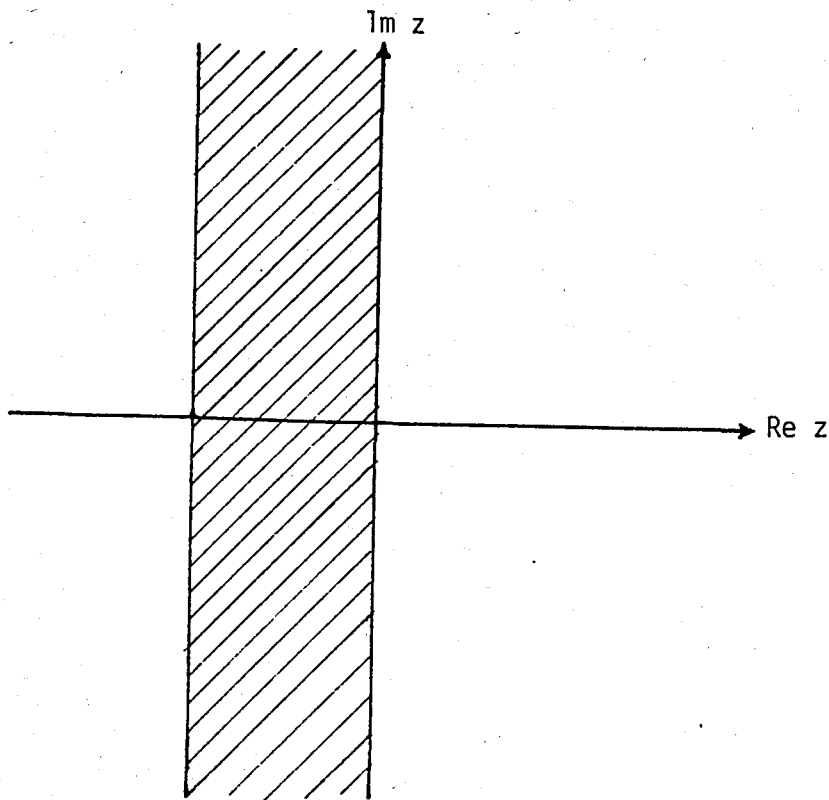


FIGURE 3.2 - An absolute stability region.

region is infinite and includes the left-hand half plane). In the case of the specified absolute stability region, a precise measure of this difficulty is the magnitude of $|\text{Re}\lambda_n|$. If $|\text{Re}\lambda_n| \gg |\text{Re}\lambda_m|$, we are forced into the highly undesirable computational situation of having to integrate numerically over a long range using a steplength which is everywhere excessively small relative to the interval involved. The difficulty of being forced to integrate numerically over a long range using a steplength which is excessively small relative to the interval involved is the problem of stiffness.

It should be emphasized that the magnitude of $|\text{Re}\lambda_n|$ is not in general a precise measure of the difficulty involved in the stiffness problem, but only a rough measure. That this is the case, can best be

seen if a multistep method with an arbitrary absolute stability region is considered. In such cases, the identification of a Δt such that all $\lambda_i \Delta t$, $i = 1, 2, \dots, N$ fall into the stability region is governed by the real parts as well as the imaginary parts of the eigenvalues and the geometry of the stability region. In view of this fact, the magnitude of the $|\operatorname{Re}\lambda_i|$, in general, can only be regarded as a rough rather than a precise measure of the difficulty involved in the stiffness problem.

In view of the above discussion, we focus our attention to a class of problems called stiff problems that can be rather defined as follows.

Definition 8: The linear system $\dot{\underline{x}} = \underline{A}\underline{x} + \underline{\phi}(\underline{x})$ is said to be stiff if

- i) $\operatorname{Re}\lambda_i < 0$, $i = 1, 2, \dots, N$
- ii) $\max_{i=1,2,\dots,N} |\operatorname{Re}\lambda_i| \gg \min_{i=1,2,\dots,N} |\operatorname{Re}\lambda_i|$,

where λ_i , $i = 1, 2, \dots, N$, are the eigenvalues of \underline{A} . The ratio

$$\left[\max_{i=1,2,\dots,N} |\operatorname{Re}\lambda_i| \right] : \left[\min_{i=1,2,\dots,N} |\operatorname{Re}\lambda_i| \right]$$

is called the stiffness ratio.

Non-linear systems $\dot{\underline{x}} = \underline{f}(\underline{x}, t)$ exhibit stiffness if the eigenvalues of the Jacobian $\partial \underline{f} / \partial \underline{x}$ behave in a similar fashion. The eigenvalues are no longer constant but depend on the solution, and therefore vary with t . Accordingly we say that the system $\dot{\underline{x}} = \underline{f}(\underline{x}, t)$ is stiff in an interval I of t if, for $t \in I$, the eigenvalues $\lambda_i(t)$ of $\partial \underline{f} / \partial \underline{x}$ satisfy (i) and (ii) above.

Note that if the partial derivatives appearing in the Jacobian $\partial \underline{f} / \partial \underline{x}$ are continuous and bounded in an appropriate region, then the Lipschitz constant L of the system $\dot{\underline{x}} = \underline{f}(t, \underline{x})$ may be taken to be $L = \|\partial \underline{f} / \partial \underline{x}\|$ for any matrix \underline{A} , $\|\underline{A}\| \geq \rho(\underline{A})$, where $\rho(\underline{A})$, the spectral radius, is defined to be

$$\max_{i=1,2,\dots,N} |\lambda_i|$$

λ_i , $i = 1, 2, \dots, N$, being the eigenvalues of \underline{A} . If $\max_{i=1,2,\dots,N} |\operatorname{Re} \lambda_i| \gg 0$ it follows that $L \gg 0$. Thus stiff systems are occasionally referred to as "systems with large Lipschitz constants".

The basic difficulty, but not the only one, in the numerical solution of stiff systems is the satisfaction of the requirement of absolute stability with an integrationwise economically feasible time step. To overcome this difficult several definitions, which require for the method to possess some "adequate" region of absolute stability, have been proposed. The following is a brief account of such definitions given by Lambert [9].

Definition 9: (Dahlquist [12]). A linear multistep method is called A-stable if all solutions of (3.3) tend to zero as $n \rightarrow \infty$, when the method is applied with a fixed steplength Δt to any differential equation of the form

$$\dot{\underline{x}} = \lambda \underline{x}$$

where λ is a complex constant with a negative real part. In other words, a linear multistep method is A-stable if its region of absolute stability contains the whole of the left-hand complex plane.

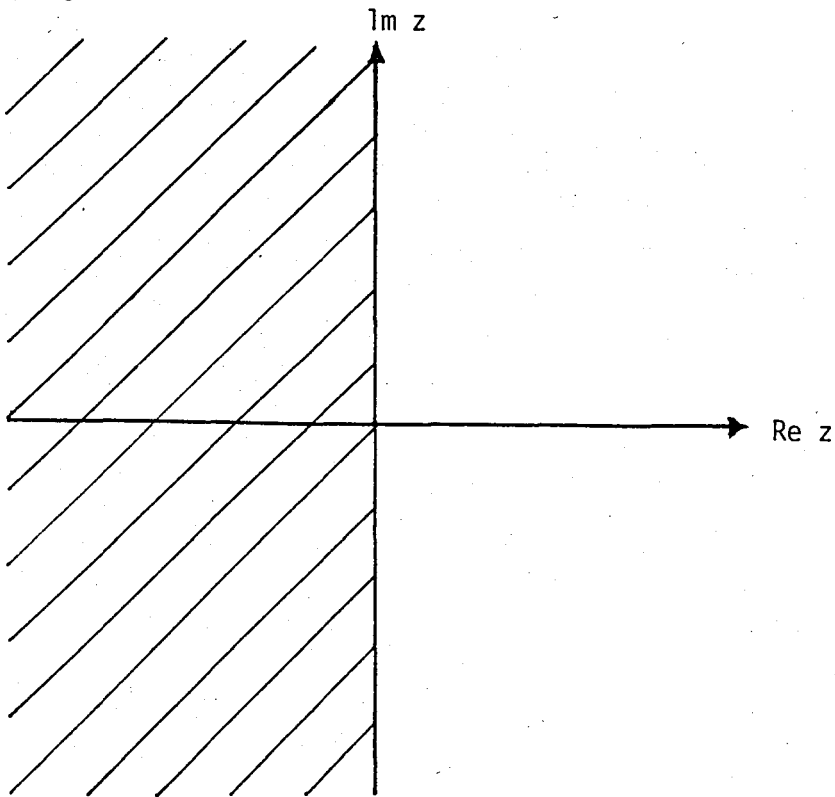


FIGURE 3.3 - Absolute stability region of A-stable methods.

If an A-stable method is applied to a stiff system, then the difficulties associated with the stiffness problem disappear, since, no matter how large $\max_{i=1,2,\dots,N} |\operatorname{Re} \lambda_i|$, no stability restriction on Δt can result. However, A-stability is a severe requirement to ask of a numerical method, as the following somewhat depressing theorem of Dahlquist [12] shows.

THEOREM 5: i) An explicit linear multistep method cannot be A-stable.

- ii) The order of an A-stable implicit linear multistep method cannot exceed two.
- iii) The second-order A-stable implicit linear multistep method with smallest error constant is the Crank-Nicholson method.

The restriction on order implied by (ii) is a severe one (Note that the Crank-Nicholson method is the one-step method with coefficients

$$\alpha_1 = 1, \quad \alpha_0 = -1, \quad \beta_1 = 1/2, \quad \beta_0 = 1/2$$

so that $C_0 = C_1 = C_2 = 0$ and

$$C_3 = (1/3!) \alpha_1 - (1/2!) \beta_1$$

$$C_3 = (1/6) - (1/4) = -1/12.$$

Then

$$\mathcal{L}\{\tilde{x}(t); \Delta t\} = -\frac{1}{12} \Delta t^3 \tilde{x}^{(3)}(t) + O(\Delta t^4);$$

for a stiff system, we can expect the components $\tilde{x}^{(3)}(t)$ to be very large, at least in an interval on which the transient solution is not negligible). In view of this, several less demanding stability definitions have been proposed; we present two here.

Definition 10: (Windlund [13]). A linear multistep method is $A(\alpha)$ -stable, $\alpha \in (0, \pi/2)$, if all solutions of (3.3) tend to zero as $n \rightarrow \infty$ when the method is applied with a fixed positive Δt to any differential equation of the form

$$\dot{\tilde{x}} = \lambda \tilde{x}$$

where λ is a complex constant which lies in the set.

$$S_\alpha = \{z \mid -\alpha < \pi - \arg(z) < \alpha, z \neq 0\}$$

A method is $A(\pi/2)$ -stable if it is $A(\alpha)$ -stable for all $\alpha \in (0, \pi/2)$ and $A(0)$ -stable if it is $A(\alpha)$ -stable for some sufficiently small $\alpha \in (0, \pi/2)$

The above definition is equivalent to say that a linear multi-step method is $A(\alpha)$ -stable, $\alpha \in (0, \pi/2)$, if its region of absolute stability contains the infinite wedge.

$$S_\alpha = \{z \mid -\alpha < \pi - \arg(z) < \alpha, z \neq 0\}$$

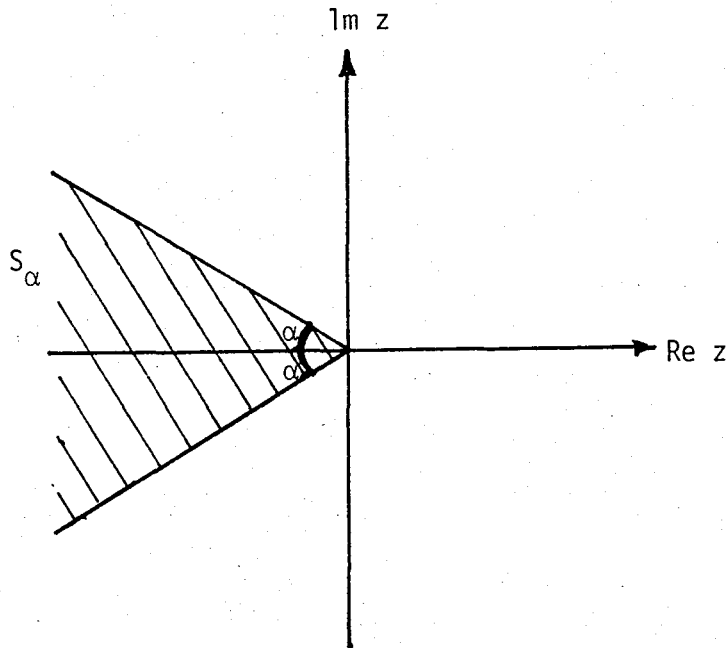


FIGURE 3.4 - Absolute stability region of $A(\alpha)$ -stable methods.

An alternative weakening of A -stability defined by Cryer [14] is A_0 -stability.

Definition 11: (Cryer [14]). A linear multistep method is called A_0 -stable if all solutions of (3.3) tend to zero as $n \rightarrow \infty$, when the method is applied with a fixed positive Δt to any differential equation of the form

$$\dot{\underline{x}} = \lambda \underline{x}$$

where $\lambda \in (0, +\infty)$. Hence, the absolute stability region of an A_0 -stable method contains the open negative real axis.

VI. MULTISTEP DISCRETIZATION OF THE SEMIDISCRETIZED HEAT CONDUCTION EQUATION

4.1 THE GENERAL LINEAR k-STEP METHOD APPLIED TO HEAT CONDUCTION EQUATION

In the second chapter we have shown that the finite element semidiscretized heat conduction equation is of the form

$$\underline{\underline{C}}\dot{\underline{a}} + \underline{\underline{K}}\underline{a} = \underline{f} \quad (4.1)$$

where

$$\underline{a}(0) = \underline{a}_0$$

The semidiscretized heat conduction equation (4.1) can be expressed in the form of (3.1) as

$$\dot{\underline{a}} = -\underline{\underline{A}}\underline{a} + \underline{g} \quad (4.2)$$

where $\underline{\underline{A}} = \underline{\underline{C}}^{-1}\underline{\underline{K}}$ and $\underline{g} = \underline{\underline{C}}^{-1}\underline{f}$. Since $\underline{\underline{C}}$ and $\underline{\underline{K}}$ are symmetric positive definite matrices, $\underline{\underline{A}}$ has real positive eigenvalues. The eigenvalue spectrum of $\underline{\underline{A}}$ is usually very large (Zlamal [3,15], Wood [16]) and hence, system (4.2) is a stiff system.

The general form of the linear k-step method applied to semi-discretized heat conduction is therefore

$$\sum_{j=0}^k (\alpha_j I + \beta_j \Delta t A) \tilde{a}_{n+j} = \Delta t \sum_{j=0}^k \beta_j \tilde{g}_{n+j} \quad (4.3)$$

If we multiply Eq. (4.3) by \tilde{C} from left, we obtain

$$\sum_{j=0}^k (\alpha_j \tilde{C} + \beta_j \Delta t \tilde{K}) \tilde{a}_{n+j} = \Delta t \sum_{j=0}^k \beta_j \tilde{f}_{n+j} \quad (4.4)$$

Since Eqs. (4.3) and (4.4) are identical, they can be used interchangeably. On the other hand, the form of Eq. (4.4) is much more preferable for computational purposes as far as core requirement and computation time is concerned. This is because of the fact that the matrices \tilde{C} and \tilde{K} are symmetric and banded whereas \tilde{A} is not.

In this chapter we will introduce some of the methods proposed for the numerical integration of problem (4.1). The most well-known and commonly used algorithms for this purpose are linear multistep methods upto step number two.

If a linear multistep method has a bounded region of absolute stability, then the absolute stability requirement will impose a step-length restriction. Such methods are called conditionally stable in the sense of absolute stability. Due to the excessively small steplength with respect to the range of numerical integration, these method, in general, yields a highly undesirable computational situation in the case of stiff systems. Therefore, we should restrict the choice of linear multistep methods to those for which there is no steplength

restriction imposed by absolute stability i.e. to unconditionally stable ones.

In the case of semidiscretized heat conduction equation the eigenvalues of A are real and positive, and, therefore, linear multistep methods with the region of absolute stability $(-\infty, 0)$ be general enough and impose no restriction on the size of the steplength to be used. Hence we will restrict the choice of a linear multistep method to be from the set of A_0 -stable ones. It should be clear that once we decide for such a method, in order to assure convergence we should further restrict it to meet the requirements of consistency and zero-stability.

4.2 ONE-STEP METHODS

It is possible to derive any specific multistep method in a number of different ways. One possible way, as proposed by Zienkiewicz, is to use the finite element time discretization technique. This technique yields the most popular methods used for the solution of the semidiscretized heat conduction equation. The following is a brief account of the derivation of such methods.

Let us approximate the vector \underline{a} as

$$\underline{a} \approx \hat{\underline{a}} = \sum_i N_i \underline{a}_i$$

where \underline{a}_i is the vector of unknown nodal parameters at time t_i and $\{N_i(t)\}$ is a finite element C^0 basis.

For one-step discretization, we will use a time element of length Δt with nodal points identified as n and $n+1$. A typical

element is shown in Fig. 4.1. The basis functions in terms of local variables are then given as

$$N_n = 1 - \xi \quad , \quad N_{n+1} = \xi \quad (4.5)$$

where

$$0 \leq \xi \leq 1 \quad , \quad \xi = t/\Delta t$$

Then

$$\dot{N}_n = -1/\Delta t \quad , \quad \dot{N}_{n+1} = 1/\Delta t$$

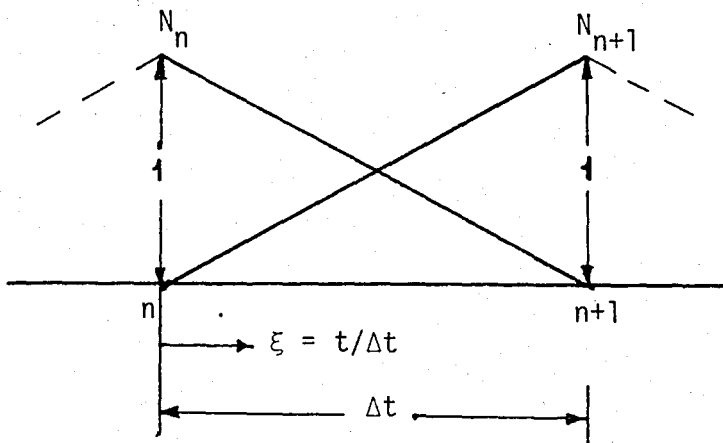


FIGURE 4.1 - Basis functions for one-step methods.

With the assumption that f can be expanded in terms of the same basis functions N_i , we get

$$\tilde{f} \approx \hat{f} = \sum_i N_i f_{i\sim i}$$

where $f_{i\sim i}$ is the value of the function f at time $t_{i\sim i}$. For the time element shown in Fig. 4.1, the weighted residual equation associated with the system of equations 4.1 can be written as

$$\int_0^1 w_j \{ C(\tilde{a}_{n+1} \dot{N}_{n+1} + \tilde{a}_n \dot{N}_n) + K(\tilde{a}_{n+1} N_{n+1} + \tilde{a}_n N_n) - (f_{n+1} N_{n+1} + f_n N_n) \} d\xi = 0 \quad (4.6)$$

(j = 1)

If we substitute the expressions for the basis functions given by Eqs. (4.5) in Eq. (4.6) and multiply the result by Δt , we obtain

$$\begin{aligned} & \{ C \int_0^1 w_j d\xi + \Delta t K \int_0^1 w_j \xi d\xi \} \tilde{a}_{n+1} + \{ -C \int_0^1 w_j d\xi + \Delta t K \int_0^1 w_j (1-\theta) d\xi \} \tilde{a}_n \\ & = \Delta t \{ f_{n+1} \int_0^1 w_j \xi d\xi + f_n \int_0^1 w_j (1-\theta) d\xi \} \quad (j=1) \end{aligned} \quad (4.7)$$

Dividing Eq. (4.7) by $\int_0^1 w_j d\xi$ and defining θ as

$$\theta = \frac{\int_0^1 w_j \xi d\xi}{\int_0^1 w_j d\xi} \quad (4.8)$$

we get

$$\{ C + \theta \Delta t K \} \tilde{a}_{n+1} + \{ -C + (1-\theta) \Delta t K \} \tilde{a}_n = \Delta t \{ \theta f_{n+1} + (1-\theta) f_n \} \quad (4.9)$$

If we premultiply Eq. (4.8) by C^{-1} , we can put it in the form of Eq. (4.3) as

$$\{ I + \theta \Delta t A \} \tilde{a}_{n+1} + \{ -I + (1-\theta) \Delta t A \} \tilde{a}_n = \Delta t \{ \theta g_{n+1} + (1-\theta) g_n \} \quad (4.10)$$

Eq. (4.10) defines a class of one-step methods with coefficients

$$\begin{aligned} \alpha_1 &= 1 & , & & \beta_1 &= \theta \\ \alpha_0 &= -1 & , & & \beta_0 &= 1 - \theta. \end{aligned}$$

A member of this class is frequently referred to as the θ -method.

The first characteristic polynomial of a θ -method is

$$\rho(\zeta) = \zeta - 1$$

so that its root is $\zeta_1 = 1$. Thus, any θ -method is necessarily zero-stable. Since

$$\sum_{j=0}^1 \alpha_j = 0, \quad \sum_{j=0}^1 j\alpha_j = 1, \quad \sum_{j=0}^1 \beta_j = 1,$$

any θ -method is also necessarily consistent. Therefore, any θ -method is consistent and zero-stable, thus convergent.

For the study of A_0 -stability, it will be convenient to reduce system (4.2) to a typical single degree of freedom model equation which can be achieved with the application of the modal analysis technique.

Let us assume that \underline{A} has a complete set of normalized orthogonal eigenvectors. Then, the vector \underline{a} can be written as

$$\underline{a} = \sum_{i=1}^n \underline{u}_i y_i \quad (4.11)$$

where \underline{u}_i are normalized eigenvectors of \underline{A} and y_i are modal participation variables. If expression (4.11) is substituted into Eq. (4.2) and the result is premultiplied by \underline{u}_i^T ($i = 1, 2, \dots, n$) we get

$$\dot{y}_i = -\lambda_i y_i + g_i \quad (4.12)$$

where $g_i = \underline{u}_i^T \underline{g}$ and λ_i are the eigenvalues of \underline{A} . For notational simplicity, we will drop the i 'th modal subscript in Eq. (4.12) and continue our analysis with the following single degree of freedom model equation

$$\dot{y} = -\lambda y + g \quad (4.13)$$

where λ represents an eigenvalue of the matrix A . Application of the θ -method to Eq. (4.13) yields

$$(1 + \lambda\Delta t\theta)y_{n+1} + (-1 + \lambda\Delta t(1-\theta))y_n = \Delta t(\theta g_{n+1} + (1-\theta)g_n) \quad (4.14)$$

The stability polynomial of the θ -method is then

$$\pi(r, \lambda\Delta t) = (1 + \lambda\Delta t\theta)r - 1 + \lambda\Delta t(1 - \theta) \quad (4.15)$$

If we set polynomial (4.15) equal to zero and solve r , we obtain

$$r = \frac{1 - (1-\theta)\lambda\Delta t}{1 + \theta\lambda\Delta t} \quad (4.16)$$

Hence, a θ -method is A_0 -stable if and only if r has modulus less than one for all $\lambda\Delta t > 0$. The requirement $|r| < 1$ implies

$$-1 - \theta\lambda\Delta t < 1 - (1-\theta)\lambda\Delta t < 1 + \theta\lambda\Delta t$$

or

$$\begin{aligned} \lambda\Delta t(2\theta-1) &> -2 \\ \lambda\Delta t &> 0 \end{aligned} \quad (4.17)$$

Condition (4.17) is always satisfied when $\theta \geq 1/2$. Therefore, a θ -method is A_0 -stable if and only if $\theta \geq 1/2$.

With the proper choice of weight functions in Eq. (4.8) we can obtain the following A_0 -stable θ -methods:

$$w_j = \delta\left(\xi - \frac{1}{2}\right) \quad \text{or} \quad w_j = 1, \quad \theta = 1/2 \quad (\text{Crank-Nicholson})$$

$$w_j = \xi, \quad \theta = 2/3 \quad (\text{Galerkin})$$

$$w_j = \delta(\xi - 1), \quad \theta = 1 \quad (\text{Backward-difference})$$

Here $\delta(\xi - \xi_0)$ is the Dirac function. The derivation of the above θ -values are given in Appendix 2.

With the intention of stating another θ -method, let us assume that the term g in Eq. (4.13) is zero. Then Eq. (4.14) yields the recurrence relation

$$y_{n+1} = ry_n, \quad r = \frac{1 - (1 - \theta)\lambda\Delta t}{1 + \theta\lambda\Delta t} \quad (4.18)$$

The homogeneous solution of Eq. (4.13), on the other hand is

$$y(t) = y(0)e^{-\lambda t} \quad (4.19)$$

and satisfies the recurrence relation

$$y[(n+1)\Delta t] = e^{-\lambda\Delta t}y(n\Delta t) \quad (4.20)$$

Comparing the Eqs. (4.18) and (4.20) we conclude that r approximates $e^{-\lambda\Delta t}$. It is proposed by Liniger [17] to choose θ in such a way that

$$\max_{0 < \lambda\Delta t < \infty} \left| e^{-\lambda\Delta t} - \frac{1 - (1 - \theta)\lambda\Delta t}{1 + \theta\lambda\Delta t} \right|$$

is minimum, and this happens when $\theta = 0.878$. Here on, the θ -method corresponding to the value $\theta = 0.878$ will be considered as the fourth θ -method and referred to as the Liniger method.

Let us rewrite the recurrence relation (4.18) as

$$y_n = r^n y_0, \quad r = \frac{1 - (1 - \theta)\lambda\Delta t}{1 + \theta\lambda\Delta t} \quad (4.21)$$

We have shown that when $\theta \geq 1/2$, the method (4.10) is A_0 -stable so that $|r| < 1$ for all $\lambda\Delta t > 0$. Note that if $\theta \geq 1$, then r will be positive for all values of $\lambda\Delta t > 0$ and produce no oscillation. On the other hand, for $1/2 \leq \theta < 1$ positiveness of r is conditional and if $-1 < r < 0$, the recurrence relation (4.21) produces an oscillatory solution. One possible approach to prevent this oscillation is to choose Δt in such a way that for $1/2 \leq \theta < 1$, r stays positive. If we write this requirement as

$$\frac{1 - (1 - \theta)\lambda\Delta t}{1 + \theta\lambda\Delta t} > 0, \quad 1/2 \leq \theta < 1 \quad (4.22)$$

we get the condition

$$\Delta t < \frac{1}{(1 - \theta)\lambda}, \quad 1/2 \leq \theta < 1 \quad (4.23)$$

Since λ stands for any eigenvalue of the matrix A , the condition (4.23) is equivalent to

$$\Delta t < \Delta t_{\text{crit}} = \frac{1}{(1 - \theta)\lambda_{\text{max}}} \quad (4.24)$$

where λ_{max} is the maximum eigenvalue of the matrix A .

For the semidiscretized heat conduction equation, which we have seen to be a stiff system, the critical value of the steplength will most probably be excessively small with respect to the range of integration. Therefore, it is impracticable to prevent oscillations by reducing the magnitude of the steplength and some artificial smoothing techniques should be preferred.

It should further be noted that

$$\lim_{\lambda\Delta t \rightarrow \infty} r = (\theta - 1)/\theta$$

is less than zero for all $1/2 < \theta < 1$ and approaches to the value -1 for $\theta = 1/2$. Therefore, the θ -method with the value $\theta = 1/2$, which is known as the Crank-Nicholson scheme, is inaccurate for larger values of $\lambda\Delta t$ and only marginally stable (Liniger [17]). On the other hand, it is proved by Cryer [14] that the Crank-Nicholson scheme with $p = 2$, $k = 1$ is the only A_0 -stable linear k -step method of order $p \geq k+1$. For this reason, it is of great importance and special precautions should be taken in using it for the solution of stiff systems.

4.3 TWO-STEP METHODS

We will adopt the finite element time discretization technique to Eq. (4.1) to obtain a class of two-step methods proposed by Zienkiewicz [2] for the numerical integration of first order systems.

Let us approximate the vector \underline{a} as

$$\underline{a} \approx \hat{\underline{a}} = \sum_j N_j \underline{a}_j$$

where \underline{a}_j is the vector of unknown nodal parameters at time t_j and $\{N_j(t)\}$ is a finite element C^1 basis.

For two-step discretization we will use a time element of length $\lambda\Delta t$ with nodal points identified as $n, n+1, n+2$. A typical element is shown in Fig. 4.2. The standard parabolic basis functions in terms of local variables can then be written as

$$N_n = -\xi(1-\xi)/2, \quad N_{n+1} = (1-\xi)(1+\xi), \quad N_{n+2} = \xi(1+\xi)/2 \quad (4.25)$$

where

$$-1 \leq \xi \leq 1, \quad \xi = t/\Delta t$$

Therefore

$$\dot{N}_n = (-(1/2) + \xi)/\Delta t, \quad \dot{N}_{n+1} = -2\xi/\Delta t, \quad \dot{N}_{n+2} = ((1/2) + \xi)/\Delta t$$

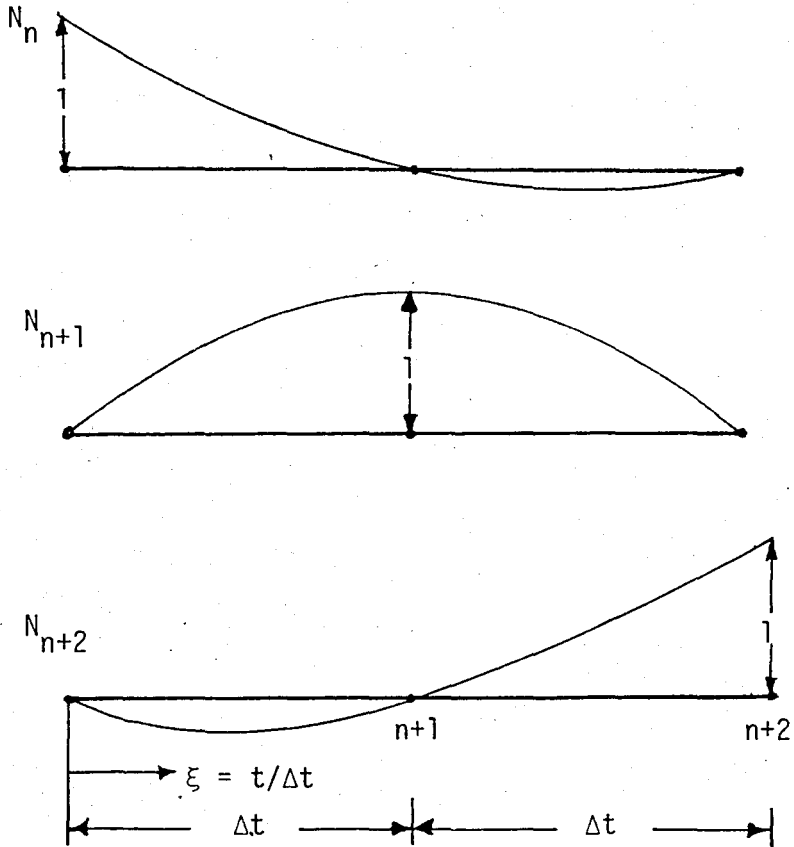


FIGURE 4.2 - Basis functions for two-step methods.

Assuming further that the same approximation as that of \underline{a} can be applied to the function f , we get

$$\tilde{f} \approx \hat{\tilde{f}} = \sum_i N_i f_i$$

for the element shown in Fig. 4.2, the weighted residual equation associated with the system of equations (4.1) can be written as

$$\int_{-1}^1 w_j \{ C(\tilde{a}_{n+2} \tilde{N}_{n+2} + \tilde{a}_{n+1} \tilde{N}_{n+1} + \tilde{a}_n \tilde{N}_n) + K(\tilde{a}_{n+2} \tilde{N}_{n+2} + \tilde{a}_{n+1} \tilde{N}_{n+1} + \tilde{a}_n \tilde{N}_n) - (\tilde{f}_{n+2} \tilde{N}_{n+2} + \tilde{f}_{n+1} \tilde{N}_{n+1} + \tilde{f}_n \tilde{N}_n) \} d\xi = 0 \quad (j = 1) \quad (4.26)$$

Substituting the expressions for the basis functions given by Eqs. (4.25), we get

$$\int_{-1}^1 w_j \{ C[\tilde{a}_{n+2}((1/2)+\xi)/\Delta t + \tilde{a}_{n+1}(-2\xi)/\Delta t + \tilde{a}_n((-1/2)+\xi)/\Delta t] + K[\tilde{a}_{n+2}\xi(1+\xi)/2 + \tilde{a}_{n+1}(1-\xi)(1+\xi) + \tilde{a}_n\xi(\xi-1)/2] - [\tilde{f}_{n+2}\xi(1+\xi)/2 + \tilde{f}_{n+1}(1-\xi)(1+\xi) + \tilde{f}_n\xi(\xi-1)/2] \} d\xi = 0 \quad (j = 1) \quad (4.27)$$

Let us define γ and β as follows:

$$\gamma = \frac{\int_{-1}^1 w_j((1/2) + \xi)d\xi}{\int_{-1}^1 w_j d\xi} \quad (4.28)$$

$$\beta = \frac{\int_{-1}^1 w_j \xi(1 + \xi)d\xi}{\int_{-1}^1 w_j d\xi}$$

If we divide Eq. (4.27) by $(1/\Delta t) \int_{-1}^1 w_j d\xi$ and rearrange the terms, we get

$$\begin{aligned} & \{\gamma C + \beta \Delta t K\} \tilde{a}_{n+2} + \{(1 - 2\gamma)C + ((1/2) - 2\beta + \gamma)\Delta t K\} \tilde{a}_{n+1} \\ & \quad + \{(\gamma - 1)C + ((1/2) + \beta - \gamma)\Delta t K\} \tilde{a}_n \\ & = \Delta t \{ \beta \tilde{f}_{n+2} + ((1/2) - 2\beta + \gamma) \tilde{f}_{n+1} + ((1/2) + \beta - \gamma) \tilde{f}_n \} \quad (4.29) \end{aligned}$$

Premultiplying Eq. (4.29) by \tilde{C}^{-1} , we can put it in the form of Eq. (4.3) as

$$\begin{aligned} & \{\gamma \tilde{I} + \beta \Delta t A\} \tilde{a}_{n+2} + \{(1-2\gamma) \tilde{I} + ((1/2)-2\beta+\gamma) \Delta t A\} \tilde{a}_{n+1} + \{(\gamma-1) \tilde{I} \\ & + ((1/2)+\beta-\gamma) \Delta t A\} \tilde{a}_n = \Delta t \{ \beta \tilde{g}_{n+2} + ((1/2)-2\beta+\gamma) \tilde{g}_{n+1} \\ & + ((1/2)+\beta-\gamma) \tilde{g}_n \} \end{aligned} \quad (4.30)$$

Eq. (4.30) defines a class of two step methods with coefficients

$$\begin{aligned} \alpha_2 &= \gamma & , & \beta_2 = \beta \\ \alpha_1 &= 1 - 2\gamma & , & \beta_1 = ((1/2) - 2\beta + \gamma) \\ \alpha_0 &= \gamma - 1 & , & \beta_0 = ((1/2) + \beta - \gamma) \end{aligned}$$

This class of two-step methods are commonly referred to as Zienkiewicz three-level schemes.

Application of the two-step method (4.30) to Eq. (4.13) gives

$$\begin{aligned} & [\gamma + \lambda \Delta t \beta] y_{n+2} + [(1-2\gamma) + \lambda \Delta t ((1/2)-2\beta+\gamma)] y_{n+1} + [(\gamma-1) \\ & + \lambda \Delta t ((1/2)+\beta-\gamma)] y_n = \Delta t \{ \beta g_{n+2} + ((1/2)-2\beta+\gamma) g_{n+1} \\ & + ((1/2)+\beta-\gamma) g_n \} \end{aligned} \quad (4.31)$$

The first and second characteristic polynomials of the method (4.30) are

$$\rho(\zeta) = \gamma \zeta^2 + (1 - 2\gamma) \zeta + \gamma - 1 \quad (4.32)$$

$$\sigma(\zeta) = \beta \zeta^2 + ((1/2) - 2\beta + \gamma) \zeta + ((1/2) + \beta - \gamma)$$

Since

$$\sum_{j=0}^2 \alpha_j = 0, \quad \sum_{j=0}^2 j\alpha_j = 1, \quad \sum_{j=0}^2 \beta_j = 1$$

or equivalently $\rho(1) = 0$, and $\rho'(1) = \sigma(1)$, then the method is consistent. The requirement of zero-stability is fulfilled if the roots of first characteristic polynomial $\rho(\zeta)$ have modulus less than or equal to one. These roots are

$$\zeta_1 = 1, \quad \zeta_2 = (1-\gamma)/\gamma$$

and the condition $|\zeta_i| \leq 1$, $i = 1, 2$, is always satisfied if $\gamma \geq 1/2$.

To get the conditions for A_0 -stability, we require that modulus of the roots of the stability polynomial

$$\begin{aligned} \pi(r, \lambda\Delta t) = & [\gamma + \lambda\Delta t\beta]r^2 + [(1 - 2\gamma) + \lambda\Delta t((1/2) - 2\beta + \gamma)]r \\ & + [(\gamma - 1) + \lambda\Delta t((1/2) + \beta - \gamma)] \end{aligned} \quad (4.33)$$

is less than one for all $\lambda\Delta t$. Application of the transformation $r = (1+z)/(1-z)$, which maps the interior of the unit circle into left-half complex plane, makes it possible to use the well-known Routh-Hurwitz criterion that gives the necessary and sufficient conditions for the roots of a polynomial to have negative real parts. Under this transformation, we obtain the following polynomial equation for the stability polynomial (4.33);

$$\{4\gamma - 2 + \lambda\Delta t(4\beta - 2\gamma)\} z^2 + \{2 - \lambda\Delta t(1 - 2\gamma)\}z + \lambda\Delta t = 0 \quad (4.34)$$

The necessary and sufficient conditions for the roots of Eq. (4.34) to have negative real parts as given by Routh-Hurwitz criterion are as follows.

$$4\gamma - 2 + \lambda\Delta t(4\beta - 2\gamma) > 0$$

$$2 - \lambda\Delta t(1 - 2\gamma) > 0$$

$$\lambda\Delta t > 0$$

These conditions are satisfied for all $\lambda\Delta t > 0$ if and only if $\gamma \geq 1/2$ and $\beta > \gamma/2$.

We conclude that the method (4.30) is consistent, zero-stable if $\gamma \geq 1/2$ and, A_0 -stable if $\gamma \geq 1/2$ and $\beta > \gamma/2$. Two popular A_0 -stable methods of type (4.30) are obtained from Eqs. (4.28) with the following specific choices of weight functions:

$$w_j = \delta(\xi - 1) \quad , \quad \gamma = 3/2 \quad , \quad \beta = 1 \quad (\text{Fully implicit})$$

$$w_j = \xi(\xi + 1)/2 \quad , \quad \gamma = 3/2 \quad , \quad \beta = 4/5 \quad (\text{Galerkin})$$

Derivation details are given in Appendix 2.

Note that the homogeneous solution of the difference equation (4.31) is

$$y_n = C_1 r_1^n + C_2 r_2^n \quad (4.35)$$

where r_1, r_2 are the roots of the stability polynomial (4.33) and C_1, C_2 are constants. It should be clear that C_1 and C_2 are to be determined from the starting values y_0 and y_1 . It is evident that a negative root or complex roots can produce oscillation. In this case, some artificial smoothing techniques may reduce oscillation and improve

accuracy. Lambert [9] shows that for one of the roots of stability polynomial, say r_1 ,

$$r_1 = e^{-\lambda\Delta t} + O(\lambda\Delta t^{p+1}) \quad \text{as } \lambda\Delta t \rightarrow 0$$

where p is the order of the linear multistep method. Therefore the value $e^{-\lambda\Delta t}$ is being approximated here by the root r_1 , which is called the principal root, whereas the other root is spurious. Liniger [17] suggests an exponential fitting by demanding that $r_1 = e^{-\lambda\Delta t}$, $r_2 = 0$ for some particular $\lambda\Delta t = C > 0$ and shows that fitting is compatible with A_0 -stability if $\lambda\Delta t \geq 2$. If we choose $\lambda\Delta t = 3$, then $r_1 = e^{-3}$ and $r_2 = 0$ will satisfy the polynomial equation $\pi(r_1, 3) = 0$, i.e.,

$$\begin{aligned} \{\gamma + 3\beta\}r^2 + \{(1 - 2\gamma) + 3((1/2) - 2\beta + \gamma)\}r \\ + \{(\gamma - 1) + 3((1/2) + \beta - \gamma)\} = 0 \quad (4.36) \end{aligned}$$

For one of the roots to be zero, we should have

$$\gamma - 1 + 3((1/2) + \beta - \gamma) = 0$$

and therefore

$$e^{-6}\{\gamma + 3\beta\} + e^{-3}\{(1 - 2\gamma) + 3((1/2) - 2\beta + \gamma)\} = 0$$

From above equations it follows that $\gamma = 1.2184$, $\beta = 0.646$. The method with these γ and β values will be referred to as the Liniger's two-step method. Two other useful A_0 -stable two-step methods are

$$\gamma = 1/2 \quad , \quad \beta = 1/3$$

proposed by Lees [18] and called the Lees' algorithm and

$$\gamma = 1 \quad , \quad \beta = 3/4$$

due to Dupont [19] and referred to as the Dupont scheme.

If one demands to make the roots of the stability polynomial (4.33) real, then any probable oscillation because of the complex roots can be prevented. The roots will be real, if the discriminant of the polynomial (4.33) satisfies the condition

$$(\lambda\Delta t)^2\{(1 + 2\gamma)^2 - 16\beta\} + 4\lambda\Delta t(1 - 2\gamma) + 4 \geq 0 \quad (4.37)$$

Condition (4.37) is satisfied for all $\lambda\Delta t$ if $\beta \leq \gamma/2$. Therefore, any A_0 -stable Zienkiewicz three-level scheme has complex roots for some interval of $\lambda\Delta t$ and oscillation can be expected in this interval. In order to have real roots for all $\lambda\Delta t$, we may accept marginal stability (Wood [4]) with $\beta = \gamma/2$. For example the method

$$w_j = |\xi| \quad , \quad \gamma = 1/2 \quad , \quad \beta = 1/4$$

which is known as the method of average acceleration may be used for this purpose. However, any method with $\beta = \gamma/2$ is equivalent to the well-known Crank-Nicholson method so that the benefit from using a two-step method diminishes.

As we have discussed in Chapter three, two-step methods require the determination of starting values a_0, a_1 and it is desirable that these values be calculated to an accuracy as high as the order of accuracy of the two-step method. Defining the coefficients C_0, C_1, \dots as in Chapter three, we have for the Zienkiewicz three level schemes

$$C_0 = C_1 = C_2 = 0 \quad \text{and} \quad C_3 = \frac{6\gamma - 12\beta - 1}{12}$$

It is clear that none of the methods we have given in this section make $C_3 = 0$ so that they are all second order accurate. Therefore, it is possible to start these methods by setting $\underline{a}_0 = \underline{a}(0)$ (the initial vector of the problem (4.1)) and calculating \underline{a}_1 by the second order accurate Crank-Nicholson method. An alternative way of starting multistep methods is to translate the problem to that of an initially steady system to which a perturbation is applied.

4.4 THE PROBLEM ASSOCIATED WITH THE INTERPOLATION OF THE FORCING FUNCTIONS

As we have stated in Chapter two, the forcing function \underline{f} is of the form

$$\underline{f} = \underline{h} - \underline{C}_\ell \dot{\underline{a}}_\ell(t) - \underline{k}_\ell \underline{a}_\ell(t) \quad , \quad \ell = 1, 2, \dots, m$$

where m is the number of specified nodal parameters, \underline{h} is the condensed form of the function $\tilde{\underline{f}}$ defined as

$$\tilde{\underline{f}} = \int_{\Omega} \rho c \phi \phi^T d\Omega + \int_{s_2} \phi q_i n_i ds + \int_{s_3} \phi h \theta_\infty ds$$

by simply deleting the correspondant rows of $\tilde{\underline{f}}$ to specified nodal parameters \underline{C}_ℓ and \underline{k}_ℓ are, respectively, the partition of the ℓ 'th columns of the $\tilde{\underline{C}}$ and $\tilde{\underline{K}}$ matrices assembled with the assumption that the specified nodal parameters are free and the partitioning consists of the condensation of the relevant column vectors by simply omitting the m rows corresponding to the specified nodal parameters \underline{a}_ℓ and $\dot{\underline{a}}_\ell$ respectively.

With the intention of deriving multistep methods, in sections 4.2 and 4.3 we have interpolated the forcing function f as

$$\tilde{f} \approx \hat{\tilde{f}} = \sum_i N_i f_i$$

where N_i is the i 'th basis function and f_i is the value of the function f at time t_i . Clearly, the above interpolation requires the knowledge of f at each t_i . If the functions $a_\ell(t)$ are discontinuous at some t_i , then the function \tilde{f} will be undefined for this t_i . In order to overcome this difficulty, it should be preferred to interpolate the functions a_ℓ and h separately.

Defining the function \bar{f} as in Zienkiewicz, we have

$$\bar{f} = \frac{\int_{t_n/\Delta t}^{t_{n+k}} w_j \hat{f} d\xi}{\int_{t_n/\Delta t}^{t_{n+k}} w_j d\xi} \quad (j = 1)$$

where k is the stepnumber of the method applied to the heat conduction equation. We can therefore rewrite Eqs. (4.9) and (4.29) as

$$\{C + \theta \Delta t K\} a_{n+1} + \{-C + (1 - \theta) \Delta t K\} a_n = \Delta t \bar{f} \quad (4.38)$$

where

$$\bar{f} = \frac{\int_0^1 w_j \hat{f} d\xi}{\int_0^1 w_j d\xi}$$

and

$$\begin{aligned} & \{\gamma \underline{c} + \beta \Delta t \underline{k}\} \underline{a}_{n+2} + \{(1 - 2\gamma) \underline{c} + ((1/2) - 2\beta + \gamma) \Delta t \underline{k}\} \underline{a}_{n+1} \\ & + \{(\gamma - 1) \underline{c} + ((1/2) - \beta + \gamma) \Delta t \underline{k}\} \underline{a}_n = \Delta t \bar{f} \end{aligned} \quad (4.39)$$

where

$$\bar{f} = \frac{\int_{-1}^1 w_j \hat{f} d\xi}{\int_{-1}^1 w_j d\xi}$$

Let us interpolate a_ℓ as

$$a_\ell \approx \hat{a}_\ell = \sum_i N_i a_{\ell,i}$$

and h as

$$h \approx \hat{h} = \sum_i N_i h_i$$

where $a_{\ell,i}$ and h_i are the values of the functions a_ℓ and h at time t_i .

In this case the approximation of f will be

$$\bar{f} \approx \hat{f} = \sum_i N_i h_i - \sum_i N_i \underline{c}_\ell a_{\ell,i} - \sum_i N_i \underline{k}_\ell a_{\ell,i}$$

With the above interpolation we can obtain \bar{f} for Eq. (4.38) as

$$\bar{f} = \theta [h_{n+1} + \underline{k}_\ell a_{\ell,n+1}] + (1-\theta) [h_n + \underline{k}_\ell a_{\ell,n}] - \frac{1}{\Delta t} \underline{c}_\ell [a_{\ell,n+1} - a_{\ell,n}]$$

and for Eq. (4.39) as

$$\begin{aligned} \bar{f} &= \beta [h_{n+2} + \underline{k}_\ell a_{\ell,n+2}] + ((1/2) - 2\beta + \gamma) [h_{n+1} + \underline{k}_\ell a_{\ell,n+1}] \\ &+ ((1/2) + \beta - \gamma) [h_n + \underline{k}_\ell a_{\ell,n}] \\ &- \frac{1}{\Delta t} \underline{c}_\ell [\gamma a_{\ell,n+2} + (1-2\gamma) a_{\ell,n+1} + (\gamma-1) a_{\ell,n}] \end{aligned}$$

It can be seen that the above interpolation is completely equivalent to, first, to treat the specified nodal variables as free and assemble the matrix equation associated with the multistep methods, and, then, modify the resulting equation such that specified nodal variable conditions hold.

4.5 SMOOTHING METHODS FOR DISCONTINUOUS FORCING FUNCTIONS

We have shown that the forcing function \tilde{f} is of the form

$$\tilde{f} = \tilde{h} - \tilde{c}_l \dot{a}_l - \tilde{k}_l a_l \quad (4.40)$$

It should immediately be noted that \tilde{f} will be of a singular form if discontinuous jumps of boundary values a_l occur or if impulsive heat fluxes or heat generation are imposed.

In finite difference approximations, or in finite element formulation in which the matrix \tilde{C} is lumped (diagonal), the second term of Eq. (4.40) does not occur. \tilde{f} then becomes discontinuous when merely step changes of boundary values are specified.

Both types of variation of the forcing term are deleterious to the performance of time integration schemes derived for the solution of semidiscretized heat conduction equation, and result in violent oscillation of the solution for which artificial smoothing methods have to be devised. The following is a brief summary of the smoothing methods that have been incorporated for one-step methods.

The first possible method to smooth discontinuity is to replace the step function shown in Fig. 4.3 with a ramp function in the first time interval as shown in Fig. 4.4.

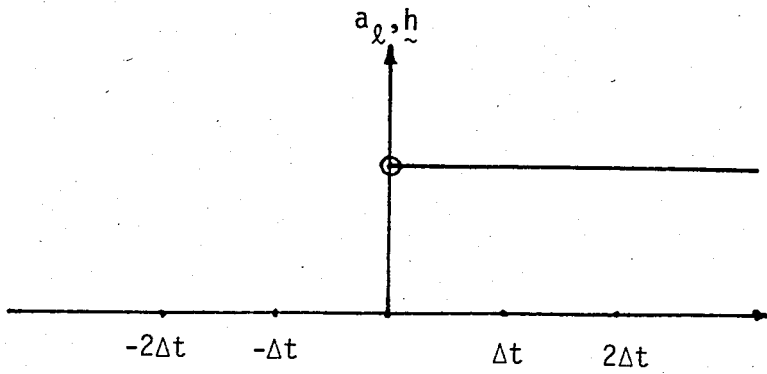


FIGURE 4.3 - Step change in h or a_l .

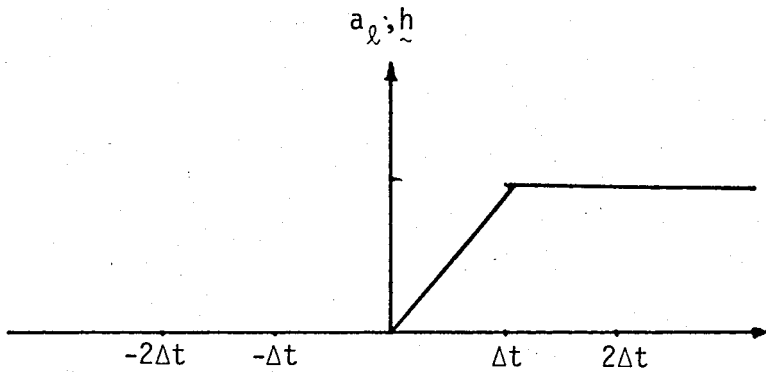


FIGURE 4.4 - Ramp change in h or a_l .

An alternative method of smoothing as proposed by Zienkiewicz [2] is to start numerical integration from $-\Delta t/2$ and to assume a ramp function variation in the first interval. The method is illustrated in Fig. 4.5.

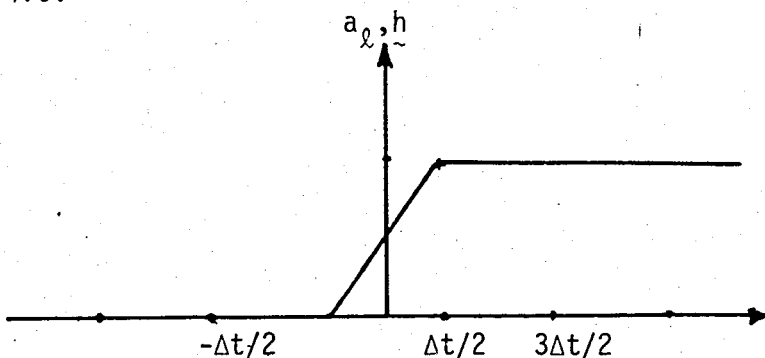


FIGURE 4.5 - Smoothing of h or a_l as proposed by Zienkiewicz.

Two other smoothing methods, particularly proposed for the Crank-Nicholson method, are

- a) the smoothing of the discontinuity with an exponential function, $(1 - e^{-\alpha t})$, as shown in Fig. 4.6 [16], and
- b) the averaging the values of the variable at the beginning and the end of the first interval and continuing numerical integration from there [20].

Based on numerical experiments, it is found that for exponential smoothing the values $\alpha = 2/\Delta t$ and $\alpha = 4/\Delta t$, improves the performance of the Crank-Nicholson method considerably.

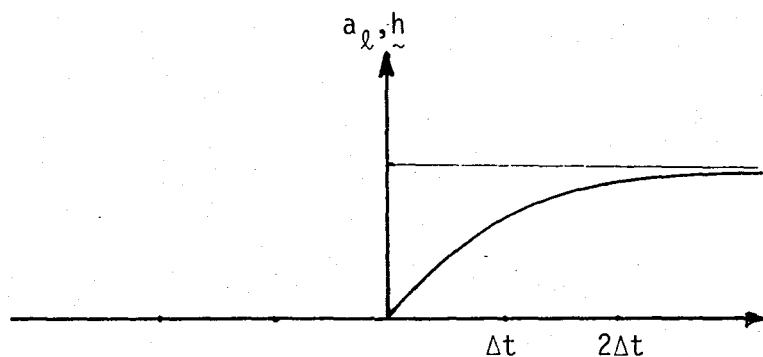


FIGURE 4.6 - Smoothing of h or a_l by an exponential function.

In the case of two-step methods there exist two different methods of smoothing the discontinuity, and are illustrated by Fig. 4.7 and Fig. 4.8.

It should be noted that in the case of the smoothing method illustrated in Fig. 4.7, the forcing function is interpolated in terms of the functional values at $-\Delta t$, 0 , Δt values with the same interpolating function that are used for the interpolation of the nodal

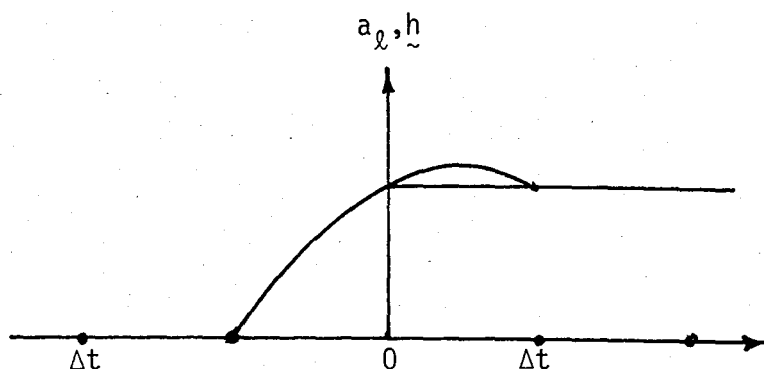


FIGURE 4.7 - Interpolation of the forcing function.

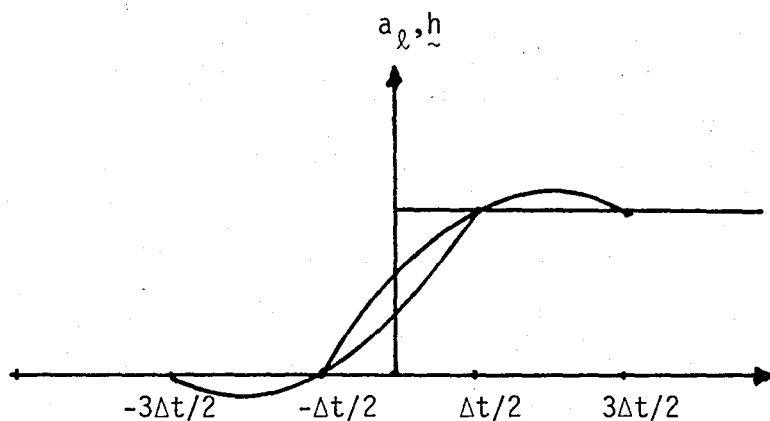


FIGURE 4.8 - Interpolation of the forcing function.

variables. Since the functional value of the function is zero at $-\Delta t$ the interpolation functions used are $N|_{t=0}$ and $N|_{t=\Delta t}$.

In the case of the smoothing illustrated in Fig. 4.8, the forcing function is interpolated in terms of the functional values at $-3\Delta t/2, -\Delta t/2$ with the same interpolating functions that are used for the interpolation of the nodal variables. Since the functional

values of the forcing function is zero at $-3\Delta t/2$ and $-\Delta t/2$, the only interpolation function used is $N|_{t=\Delta t/2}$.

In the case of the smoothing methods, Fig. 4.7 and Fig. 4.8, it is natural to start the method with the values of the nodal variables at $-\Delta t$ and 0 and with the values at $-3\Delta t/2$ and $-\Delta t/2$, respectively.

V. RESULTS

The test problem we have considered for the comparison of numerical integration methods discussed is given by the equation

$$\frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2} = \frac{\partial \theta}{\partial t} \quad 0 \leq x < 1, \quad 0 \leq y < 1, \quad t > 0 \quad (5.1)$$

subject to the initial condition

$$\theta(x, y, 0) = 0 \quad (5.2)$$

and the boundary conditions

$$(\partial/\partial y)\theta(x, y, t) = 0 \quad \text{at } x = 0, \quad \theta(x, y, t) = 100 \quad \text{at } x = 1 \quad (5.3)$$

$$(\partial/\partial x)\theta(x, y, t) = 0 \quad \text{at } y = 0, \quad \theta(x, y, t) = 100 \quad \text{at } y = 1 \quad (5.4)$$

The analytic solution of this problem is given in Arpacı [21] as

$$\theta(x, y, t) = 100 - 400 \left(\sum_{n=0}^{\infty} \frac{(-1)^n}{\lambda_n} e^{-\lambda_n^2 t} \cos \lambda_n x \right) \left(\sum_{n=0}^{\infty} \frac{(-1)^n}{\lambda_n} e^{-\lambda_n^2 t} \cos \lambda_n y \right) \quad (5.5)$$

where $\lambda_n = (2n+1)\pi/2$, $n = 0, 1, 2, \dots$

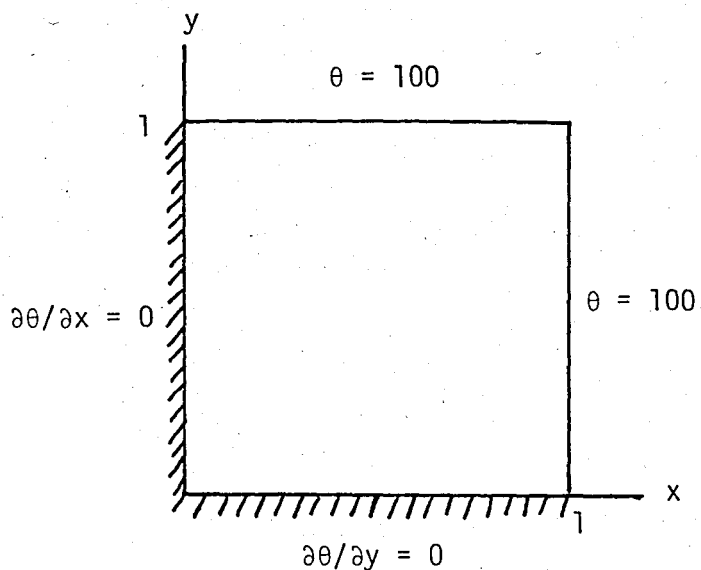


FIGURE 5.1 - Test problem.

The finite element method used for the solution of the problem (5.1)-(5.4) is shown in Fig. 5.2. The elements used are linear plane triangular elements.

The test problem is first solved by the following four θ -methods:

- i) $\theta = 1/2$ Crank-Nicholson
- ii) $\theta = 2/3$ Galerkin
- iii) $\theta = 0.878$ Liniger
- iv) $\theta = 1$ Backward-Difference.

For smoothing the discontinuity resulting from the suddenly applied boundary temperature, the following techniques are employed for all the θ -methods listed above:

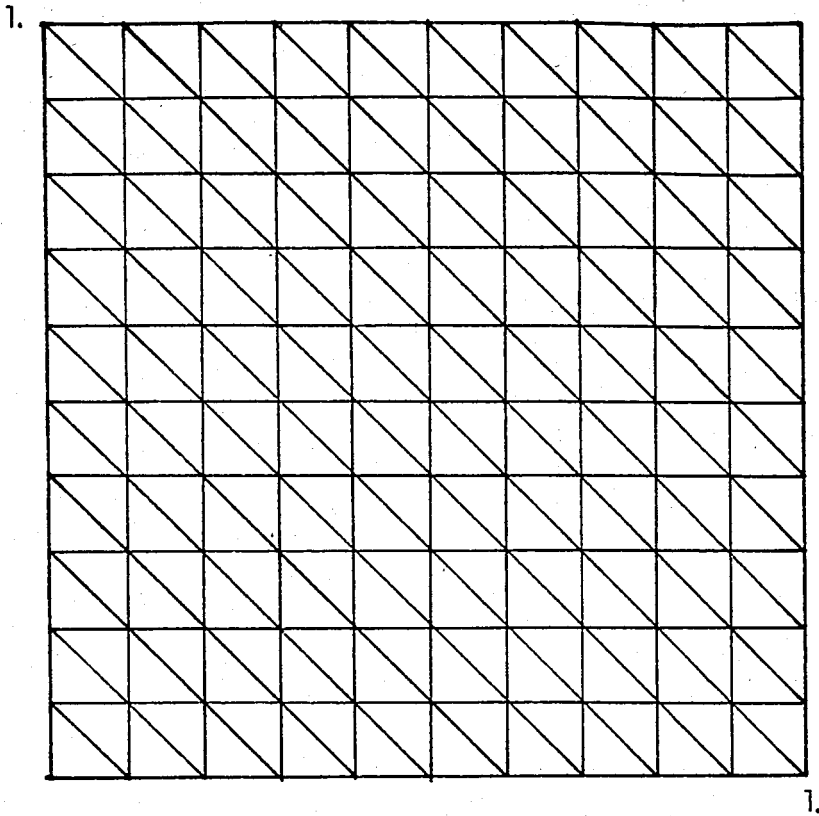


FIGURE 5.2 - 10x10 Finite Element Mesh with Triangular Elements.

- a) Smoothing the suddenly applied boundary temperature with a ramp function in the first time interval.
- b) Smoothing the suddenly applied boundary temperature with the procedure proposed by Zienkiewicz.

The following procedures are particularly adopted for the Crank-Nicholson method in order to eliminate the difficulties associated with discontinuous surface temperature.

- a) Averaging the value of the variable at the beginning and end of the first time step and continuing with Crank-Nicholson from there.

- b) Smoothing the suddenly applied boundary temperature with an exponential function.

The test problem is then solved by the following five Zienkiewicz three-level methods:

- | | | | |
|------|--------------------|-----------------|----------------|
| i) | $\gamma = 3/2$, | $\beta = 4/5$ | Galerkin |
| ii) | $\gamma = 3/2$, | $\beta = 1$ | Fully implicit |
| iii) | $\gamma = 1.218$, | $\beta = 0.646$ | Liniger |
| iv) | $\gamma = 1$, | $\beta = 3/4$ | Dupont |
| v) | $\gamma = 1/2$, | $\beta = 1/3$ | Lees |

The starting procedures employed are

- calculating a_{ℓ} by Crank-Nicholson method.
- starting the method with the known steady conditions for time less than zero, as shown in Fig. 4.7.

Results of the numerical experiments are given in the following tables for points (0,0) and (0.5,0.5). The error given in these tables is defined as

$$[\text{error}] = [\text{numerical solution}] - [\text{analytical solution}].$$

TABLE 5.3 ONE-STEP METHODS, SOLUTION OBTAINED FOR DISCONTINUOUS BOUNDARY TEMPERATURE,

$\tilde{x} = (0.0, 0.0), \Delta t = 0.001$

time	Analytic Solution	$\theta = 1/2$		$\theta = 2/3$		$\theta = 0.0878$		$\theta = 1$	
		Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.62 E-9	0.39 E-5	0.39 E-5	-0.30 E-6	-0.30 E-6	-0.75 E-7	-0.76 E-7	-0.14 E-5	-0.14 E-5
0.02	0.23 E-3	-0.27 E-3	-0.50 E-3	-0.44 E-3	-0.67 E-3	-0.29 E-3	-0.52 E-3	-0.17 E-4	-0.25 E-3
0.03	0.18 E-1	0.50 E-2	-0.13 E-1	0.12 E-1	0.61 E-2	-0.21 E-1	-0.33 E-2	0.27 E-1	0.92 E-2
0.04	0.16	0.15	-0.15 E-1	0.17	0.98 E-2	0.20	0.41 E-1	0.22	0.06
0.05	0.63	0.66	0.03	0.70	0.75 E-1	0.75	0.13	0.79	0.16
0.06	1.55	1.68	0.13	1.74	0.18	1.80	0.25	1.84	0.29
0.07	2.99	3.25	0.26	3.30	0.32	3.38	0.39	3.42	0.43
0.08	4.91	5.30	0.40	5.36	0.45	5.42	0.52	5.46	0.56
0.09	7.23	7.76	0.53	7.81	0.58	7.87	0.64	7.90	0.67
0.10	9.88	10.53	0.65	10.57	0.69	10.62	0.73	10.64	0.76
0.20	40.35	41.38	1.03	41.34	0.99	41.30	0.95	41.28	0.92
0.30	63.18	64.00	0.82	63.96	0.78	63.90	0.72	63.87	0.69
0.40	77.49	78.08	0.59	78.04	0.56	78.00	0.51	77.97	0.49
0.50	86.25	86.67	0.41	86.64	0.39	86.60	0.35	86.58	0.33

TABLE 5.1 ONE-STEP METHODS, SOLUTION OBTAINED FOR DISCONTINUOUS BOUNDARY TEMPERATURE,

$$\tilde{x} = (0.0, 0.0), \quad \Delta t = 0.01$$

		$\theta = 1/2$		$\theta = 2/3$		$\theta = 0.878$		$\theta = 1$	
time	Analytic Solution	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.62 E-9	0.39 E-3	0.39 E-3	0.25 E-2	0.25 E-2	0.10 E-1	0.10 E-1	0.18 E-1	0.18 E-1
0.02	0.23 E-3	0.64 E-2	0.62 E-1	0.26 E-1	0.26 E-1	0.74 E-1	0.73 E-1	0.11	0.11
0.03	0.18 E-1	0.48 E-1	0.31 E-1	0.13	0.11	0.28	0.26	0.37	0.36
0.04	0.16	0.22	0.59 E-1	0.44	0.28	0.73	0.57	0.90	0.79
0.05	0.63	0.70	0.77 E-1	1.09	0.47	1.54	0.91	1.77	1.14
0.06	1.55	1.67	0.12	2.18	0.63	2.74	1.19	3.03	1.48
0.07	2.99	3.20	0.21	3.75	0.77	4.36	1.37	4.67	1.68
0.08	4.91	5.24	0.33	5.77	0.86	6.36	1.45	6.66	1.76
0.09	7.23	7.69	0.46	8.15	0.92	8.68	1.45	8.96	1.72
0.10	9.88	10.46	0.58	10.83	0.95	11.26	1.38	11.50	1.62
0.20	40.35	41.37	1.02	41.83	0.68	40.63	0.28	40.42	0.06
0.30	63.18	64.01	0.83	63.58	0.40	63.05	-0.13	62.75	-0.43
0.40	77.49	78.08	0.60	77.73	0.24	77.27	-0.21	77.01	-0.47
0.50	86.25	86.67	0.42	86.40	0.15	86.05	-0.20	85.85	-0.41

TABLE 5.2 ONE-STEP METHODS, SOLUTION OBTAINED FOR DISCONTINUOUS BOUNDARY TEMPERATURE,

$x = (0.5, 0.5), \Delta t = 0.01$

time	Analytic Solution	$\theta = 1/2$		$\theta = 2/3$		$\theta = 0.878$		$\theta = 1$	
		Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.81 E-1	0.23	0.15	0.54	0.46	0.99	0.91	1.24	1.16
0.02	2.47	2.06	-0.41	3.12	0.65	4.09	1.62	4.53	2.06
0.03	8.08	7.50	-0.57	8.42	0.34	9.14	1.07	9.46	1.39
0.04	14.83	15.41	-0.58	15.12	0.29	15.15	0.32	15.21	0.38
0.05	21.47	23.31	-0.84	21.73	0.25	21.26	-0.21	21.09	-0.39
0.06	27.57	28.36	-0.79	27.77	0.20	27.05	-0.52	26.71	-0.86
0.07	33.01	31.03	1.02	33.17	0.17	32.33	-0.68	31.90	-0.11
0.08	37.82	38.61	0.79	37.96	0.14	37.09	-0.79	36.62	-1.20
0.09	42.09	43.00	0.91	42.21	0.13	41.35	-0.79	40.88	-0.12
0.10	45.88	46.60	0.72	46.00	0.12	45.18	-0.71	44.72	-1.17
0.20	69.40	69.82	0.42	69.50	0.10	69.08	-0.32	68.83	-0.57
0.30	81.52	81.83	0.30	81.59	0.07	81.29	-0.23	81.12	-0.40
0.40	88.74	88.96	0.23	88.78	0.04	88.55	-0.19	88.41	-0.33
0.50	93.13	93.29	0.16	93.15	0.03	92.97	-0.15	92.87	-0.25

TABLE 5.4 ONE-STEP METHODS, SOLUTION OBTAINED FOR DJSCONTINUOUS BOUNDARY TEMPERATURE,

$$\tilde{x} = (0.5, 0.5), \quad \Delta t = 0.001$$

time	Analytic Solution	$\theta = 1/2$		$\theta = 2/3$		$\theta = 0.878$		$\theta = 1$	
		Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.08	-0.19	-0.28	-0.13	-0.21	-0.51 E-1	-0.13	-0.85 E-2	-0.09
0.02	2.47	2.25	-0.22	2.38	-0.09	2.53	0.06	2.61	0.14
0.03	8.08	8.41	0.33	8.45	0.37	8.50	0.42	8.53	0.45
0.04	14.83	15.50	0.68	15.48	0.65	15.46	0.63	15.44	0.61
0.05	21.47	22.30	0.83	22.25	0.78	22.18	0.71	22.14	0.67
0.06	27.57	28.44	0.87	28.37	0.80	28.28	0.71	28.23	0.66
0.07	33.01	33.87	0.86	33.79	0.78	33.70	0.69	33.64	0.64
0.08	37.82	38.64	0.82	38.57	0.75	38.47	0.65	38.42	0.60
0.09	42.09	42.86	0.77	42.79	0.70	42.70	0.61	42.64	0.56
0.10	45.88	46.60	0.72	46.53	0.65	46.45	0.57	46.40	0.52
0.20	69.40	69.80	0.40	69.77	0.37	69.73	0.33	69.71	0.31
0.30	81.52	81.82	0.30	81.80	0.27	81.77	0.25	81.75	0.23
0.40	88.74	88.96	0.22	88.94	0.20	88.92	0.18	88.90	0.16
0.50	93.13	93.29	0.16	93.27	0.14	93.25	0.12	93.24	0.12

TABLE 5.5 ONE-STEP METHODS, SOLUTION OBTAINED FOR RAMP SMOOTHED BOUNDARY TEMPERATURE,

$\tilde{x} = (0.0, 0.0), \Delta t = 0.01$

time	Analytic Solution	$\theta = 1/2$		$\theta = 2/3$		$\theta = 0.878$		$\theta = 1$	
		Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.62 E-9	0.13 E-3	0.13 E-3	0.13 E-2	0.13 E-2	0.73 E-2	0.73 E-2	0.15 E-1	0.15 E-1
0.02	0.23 E-3	0.25 E-2	0.22 E-2	0.15 E-1	0.14 E-1	0.56 E-1	0.56 E-1	0.97 E-1	0.97 E-1
0.03	0.18 E-1	0.21 E-1	0.30 E-2	0.81 E-1	0.63 E-1	0.22	0.20	0.33	0.31
0.04	0.16	0.11	-0.05	0.29	0.13	0.61	0.44	0.82	0.66
0.05	0.63	0.39	-0.24	0.77	0.15	1.31	0.69	1.64	1.01
0.06	1.55	1.04	-0.52	1.65	0.10	2.41	0.85	2.83	1.28
0.07	2.99	2.20	-0.79	2.98	-0.01	3.91	0.92	4.41	1.42
0.08	4.91	3.90	-1.01	4.78	-0.13	5.80	0.89	6.35	1.44
0.09	7.23	6.08	-1.16	6.98	-0.26	8.03	0.79	8.59	1.36
0.10	9.88	8.64	-1.24	9.51	-0.37	10.54	0.66	11.10	1.21
0.20	40.35	35.50	-0.85	39.65	-0.70	39.85	-0.50	39.98	-0.35
0.30	63.18	62.81	-0.37	62.68	-0.50	62.54	-0.64	62.46	-0.72
0.40	77.49	77.35	-0.14	77.17	-0.31	76.95	-0.53	76.83	-0.66
0.50	86.25	86.23	-0.03	86.06	-0.19	85.85	-0.40	85.73	-0.60

TABLE 5.6 ONE-STEP METHODS, SOLUTION OBTAINED FOR RAMP SMOOTHED BOUNDARY TEMPERATURE,

$$\tilde{x} = (0.5, 0.5), \quad \Delta t = 0.01$$

time	Analytic Solution	$\theta = 1/2$		$\theta = 2/3$		$\theta = 0.878$		$\theta = 1$	
		Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.08	0.78 E-1	0.31 E-2	0.27	0.19	0.70	0.62	1.04	0.95
0.02	2.47	0.84	-1.62	1.83	-0.69	3.20	0.73	3.98	1.51
0.03	8.08	3.88	-4.20	5.77	-2.31	7.69	-0.38	8.65	0.57
0.04	14.83	10.15	-4.68	11.78	-3.05	13.43	-1.40	14.26	-0.56
0.05	21.47	17.73	-3.74	18.44	-3.03	19.52	-1.96	20.12	-1.35
0.06	27.57	24.35	-3.21	24.77	-2.80	25.40	-2.17	25.79	-1.78
0.07	33.01	30.27	-2.73	30.49	-2.52	30.83	-2.18	31.06	-1.95
0.08	37.82	35.58	-2.24	35.59	-2.23	35.73	-2.09	35.86	-1.97
0.09	42.09	40.09	-1.99	40.11	-1.98	40.14	-1.95	40.19	-1.90
0.10	45.88	44.22	-1.66	44.13	-1.76	44.09	-1.79	44.10	-1.79
0.20	69.40	68.75	-0.65	68.69	-0.71	68.61	-0.79	68.56	-0.84
0.30	81.52	81.21	-0.31	81.13	-0.39	81.02	-0.50	80.97	-0.56
0.40	88.74	88.59	-0.15	88.50	-0.24	88.38	-0.35	88.32	-0.42
0.50	93.13	93.06	-0.06	92.98	-0.15	92.87	-0.25	92.81	-0.31

TABLE 5.7 ONE-STEP METHODS, SOLUTION OBTAINED FOR RAMP SMOOTHED BOUNDARY TEMPERATURE,

$\tilde{x} = (0.0, 0.0), \Delta t = 0.001$

time	Analytic Solution	$\theta = 1/2$		$\theta = 2/3$		$\theta = 0.878$		$\theta = 1$	
		Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.62 E-9	0.21 E-7	0.20 E-7	0.33 E-6	0.33 E-6	0.44 E-7	0.44 E-7	-0.16 E-6	-0.16 E-6
0.02	0.23 E-3	0.77 E-4	-0.15 E-3	-0.18 E-3	-0.41 E-3	-0.32 E-3	-0.55 E-3	-0.25 E-3	-0.48 E-3
0.03	0.18 E-1	-0.13 E-2	-0.19 E-1	0.34 E-2	-0.14 E-1	0.11 E-1	-0.68 E-2	0.16 E-1	-0.17 E-2
0.04	0.16	0.92 E-1	-0.71 E-1	0.12	-0.45 E-1	0.15	-0.12 E-1	0.17	0.83 E-2
0.05	0.63	0.51	-0.12	0.56	-0.65 E-1	0.63	0.98 E-3	0.66	0.39 E-1
0.06	1.55	1.42	-0.13	1.49	-0.59 E-1	1.59	0.34 E-1	1.64	0.88 E-1
0.07	2.99	2.87	-0.11	2.96	-0.27 E-1	3.07	0.08	3.13	0.14
0.08	4.91	4.83	-0.07	4.93	0.21 E-1	5.04	0.14	5.11	0.20
0.09	7.23	7.22	-0.13 E-1	7.31	0.78 E-1	7.43	0.19	7.49	0.26
0.10	9.88	9.93	0.05	10.02	0.14	10.13	0.25	10.19	0.31
0.20	40.35	40.79	0.44	40.81	0.45	40.82	0.47	40.83	0.48
0.30	63.18	63.63	0.45	63.61	0.43	63.60	0.42	63.59	0.41
0.40	77.49	77.85	0.36	77.83	0.35	77.81	0.32	77.80	0.31
0.50	86.25	86.53	0.27	86.51	0.26	86.49	0.24	86.48	0.23

TABLE 5.8 ONE-STEP METHODS, SOLUTION OBTAINED FOR RAMP SMOOTHED BOUNDARY TEMPERATURE,

$x = (0.5, 0.5) \quad \Delta t = 0.001$

time	Analytic Solution	$\theta = 1/2$		$\theta = 2/3$		$\theta = 0.878$		$\theta = 1$	
		Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.08	-0.11	-0.19	-0.10	-0.18	-0.77 E-1	-0.16	-0.56 E-1	-0.14
0.02	2.47	1.23	-1.23	1.45	-1.02	1.71	-0.76	1.85	-0.61
0.03	8.08	6.92	-1.15	7.09	-0.98	7.31	-0.77	7.43	-0.64
0.04	14.83	13.99	-0.84	14.09	-0.73	14.23	-0.60	14.30	-0.52
0.05	21.47	20.91	-0.56	20.97	-0.50	21.04	-0.43	21.09	-0.39
0.06	27.57	27.21	-0.36	27.23	-0.33	27.27	-0.30	27.29	-0.28
0.07	33.01	32.78	-0.22	32.79	-0.21	32.80	-0.20	32.81	-0.19
0.08	37.82	37.69	-0.13	37.69	-0.13	37.69	-0.13	37.69	-0.13
0.09	42.09	42.01	-0.74 E-1	42.01	-0.08	42.00	-0.09	42.00	-0.09
0.10	45.88	45.85	-0.34 E-1	45.84	-0.04	45.83	-0.05	45.83	-0.56 E-1
0.20	69.40	69.47	-0.74 E-1	69.47	0.07	69.46	-0.06	69.46	0.56 E-1
0.30	81.52	81.63	0.10	81.62	0.10	81.61	-0.09	81.60	0.79 E-1
0.40	88.74	88.84	0.10	88.83	0.09	88.82	-0.08	88.81	0.77 E-1
0.50	93.13	93.22	0.09	93.21	0.08	93.20	-0.07	93.19	0.65 E-1

TABLE 5.9 ONE-STEP METHODS, SOLUTION OBTAINED FOR SMOOTHED BOUNDARY TEMPERATURE AS PROPOSED BY ZIENKIEWICZ,

$$\underline{x} = (0.0, 0.0), \quad \Delta t = 0.01$$

time	Analytic Solution	$\theta = 1/2$		$\theta = 2/3$		$\theta = 0.878$		$\theta = 1$	
		Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.005	0.64 E-11	0.13 E-3	0.13 E-3	0.13 E-2	0.13 E-2	0.73 E-2	0.73 E-2	0.15 E-1	0.15 E-1
0.015	0.31 E-5	0.25 E-2	0.25 E-2	0.15 E-1	0.15 E-1	0.56 E-1	0.56 E-1	0.97 E-1	0.97 E-1
0.025	0.31 E-2	0.21 E-1	0.18 E-1	0.81 E-1	0.78 E-1	0.22	0.22	0.33	0.33
0.035	0.63 E-1	0.11	0.47 E-1	0.29	0.23	0.61	0.55	0.82	0.76
0.045	0.34	0.39	0.05	0.77	0.43	1.31	0.97	1.64	1.30
0.055	1.02	1.04	0.02	1.65	0.63	2.41	1.39	2.83	1.81
0.065	2.21	2.20	-0.01	2.98	0.77	3.91	1.70	4.41	2.20
0.075	3.89	3.90	0.01	4.78	0.89	5.80	1.91	6.35	2.46
0.085	6.02	6.08	0.06	6.98	0.96	8.03	2.01	8.59	2.57
0.095	8.52	8.64	0.12	9.51	0.99	10.54	2.02	11.10	2.58
0.195	38.95	39.50	0.55	39.65	0.70	39.85	-0.10	39.98	1.03
0.295	62.27	62.80	0.53	62.68	0.41	62.54	0.27	62.46	0.19
0.395	76.92	77.35	0.43	77.17	0.25	76.95	0.03	76.83	-0.09
0.495	85.91	86.23	0.32	86.06	0.15	85.85	-0.06	85.73	-0.18

TABLE 5.10 ONE-STEP METHODS, SOLUTION OBTAINED FOR SMOOTHED BOUNDARY TEMPERATURE AS PROPOSED BY ZIENKIEWICZ,

$\tilde{x} = (0.5, 0.5), \Delta t = 0.01$

time	Analytic Solution	$\theta = 1/2$		$\theta = 2/3$		$\theta = 0.878$		$\theta = 1$	
		Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.005	0.11 E-1	0.78 E-1	0.67 E-1	0.27	0.26	0.70	0.69	1.04	1.03
0.015	0.78	0.84	0.06	1.83	1.05	3.20	2.42	3.98	3.20
0.025	5.01	3.88	-1.13	5.77	0.76	7.69	2.68	8.65	3.64
0.035	11.41	10.15	-1.36	11.78	0.37	13.43	2.02	14.26	2.75
0.045	18.20	17.73	-0.47	18.44	0.24	19.52	1.32	20.12	1.92
0.055	24.60	24.35	-0.25	24.77	0.17	25.46	0.86	25.79	1.19
0.065	30.37	30.27	-0.10	30.49	0.12	30.83	0.46	31.06	0.69
0.075	35.49	35.58	0.09	35.59	0.10	35.73	0.24	35.86	0.37
0.085	40.02	40.09	0.07	40.11	0.09	40.14	0.12	40.19	0.17
0.095	44.04	44.22	0.18	44.13	0.09	44.09	0.05	44.10	0.06
0.195	68.59	68.75	0.16	68.69	0.10	68.01	0.42	68.56	-0.03
0.295	81.06	81.21	0.15	81.13	0.07	81.02	-0.04	80.97	-0.09
0.395	88.46	8.59	0.14	88.50	0.04	88.38	-0.08	88.32	-0.14
0.495	92.95	93.06	0.11	92.98	0.03	92.87	-0.08	92.81	-0.14

TABLE 5.11 CRANK-NICHOLSON METHOD, SOLUTION OBTAINED FOR
 AVERAGING FIRST TWO SUCCESSIVE TEMPERATURES,
 $\tilde{x} = (0.0, 0.0)$, $\Delta t = 0.01$

time	Analytic Solution	AVERAGING	
		Numeric Solution	Error
0.005	0.64 E-11	0.66 E-4	0.66 E-4
0.015	0.31 E-5	0.13 E-2	0.13 E-2
0.025	0.31 E-2	0.11 E-1	0.08 E-2
0.035	0.62 E-1	0.64 E-1	0.02 E-1
0.045	0.34	0.25	-0.09
0.055	1.02	0.71	-0.31
0.065	2.21	1.62	-0.59
0.075	3.89	3.05	-0.84
0.085	6.02	4.99	-2.03
0.095	8.52	7.36	-1.16
0.195	38.95	38.05	-0.90
0.295	62.27	61.87	-0.40
0.395	76.92	76.77	-0.15
0.495	85.91	85.87	-0.04

TABLE 5.12 CRANK-NICHOLSON METHOD, SOLUTION OBTAINED FOR
 AVERAGING FIRST TWO SUCCESSIVE TEMPERATURES,
 $\tilde{x} = (0.5, 0.5)$, $\Delta t = 0.01$

time	Analytic Solution	AVERAGING	
		Numeric Solution	Error
0.005	0.11 E-3	0.39 E-1	0.39 E-1
0.015	0.78	0.46	-0.32
0.025	5.01	2.36	-2.65
0.035	11.41	7.01	-4.40
0.045	18.20	13.94	-4.26
0.055	24.60	21.04	-4.56
0.065	30.37	27.31	-3.06
0.075	36.49	32.93	-3.56
0.085	40.02	37.84	-2.18
0.095	44.04	42.15	-1.89
0.195	68.59	67.91	-0.68
0.295	81.06	80.73	-0.33
0.395	88.46	88.30	-0.16
0.495	92.95	92.88	-0.07

TABLE 5.13 CRANK-NICHOLSON METHOD, SOLUTION OBTAINED FOR
 AVERAGING FIRST TWO SUCCESSIVE TEMPERATURES,
 $\tilde{x} = (0.0, 0.0)$, $\Delta t = 0.001$

time	Analytic Solution	AVERAGING	
		Numeric Solution	Error
0.0095	0.17 E-9	-0.19 E-5	-0.19 E-5
0.0195	0.16 E-3	0.93 E-4	-0.7 E-4
0.0295	0.15 E-1	-0.17 E-2	-0.17 E-1
0.0395	0.15	0.08	-0.07
0.0495	0.59	0.48	-0.11
0.0595	1.49	1.36	-0.13
0.0695	2.90	2.79	-0.11
0.0795	4.80	4.73	-0.07
0.0895	7.11	7.09	-0.02
0.0995	9.74	9.79	0.05
0.1995	40.21	40.65	0.44
0.2995	63.09	63.54	0.45
0.3995	77.43	77.79	0.36
0.4995	86.21	86.49	0.20

TABLE 5.14 CRANK-NICHOLSON METHOD, SOLUTION OBTAINED FOR
 AVERAGING FIRST TWO SUCCESSIVE TEMPERATURES,
 $\tilde{x} = (0.5, 0.5)$, $\Delta t = 0.001$

time	Analytic Solution	AVERAGING	
		Numeric Solution	Error
0.0095	0.57 E-1	-0.60 E-1	-0.11
0.0195	2.26	1.04	-1.22
0.0295	7.75	6.59	-1.16
0.0395	14.48	13.63	-0.85
0.0495	21.15	20.58	-0.57
0.0595	27.27	26.91	-0.36
0.0695	32.75	32.52	-0.23
0.0795	37.60	37.48	-0.12
0.0895	41.89	41.81	-0.08
0.0995	45.70	45.66	-0.04
0.1995	69.32	69.37	0.05
0.2995	81.48	81.58	0.10
0.3995	88.71	88.81	0.10
0.4995	93.11	93.20	0.09

TABLE 5.15 CRANK-NICHOLSON METHOD, SOLUTION OBTAINED FOR
 EXPONENTIAL SMOOTHED BOUNDARY TEMPERATURE,

$$\underline{x} = (0.0, 0.0), \quad \Delta t = 0.01$$

time	Analytic Solution	$1-e^{-4t/\Delta t}$		$1-e^{-2t/\Delta t}$	
		Numeric Solution	Error	Numeric Solution	Error
0.01	.62 E-9	.13 E-3	.13 E-3	0.11 E-3	0.11 E-3
0.02	.23 E-3	.24 E-2	.21 E-2	0.21 E-2	0.19 E-2
0.03	.18 E-1	.21 E-1	.27 E-2	0.18 E-1	0.50 E-3
0.04	.16	.11	-.06	0.96 E-1	0.67 E-1
0.05	.63	.38	-.24	0.35	-0.28
0.06	1.55	1.02	-.53	0.94	-0.61
0.07	2.99	2.17	-.81	2.03	-0.96
0.08	4.91	3.87	-1.04	3.65	-1.26
0.09	7.23	6.04	-1.20	5.75	-1.48
0.10	9.88	8.59	-1.29	8.25	-1.63
0.20	40.35	39.45	-0.91	39.05	-1.31
0.30	63.18	62.77	-0.41	62.51	-0.67
0.40	77.49	77.33	-0.16	77.17	-0.32
0.50	86.25	86.21	-0.04	86.11	-0.14

TABLE 5.16 CRANK-NICHOLSON METHOD, SOLUTION OBTAINED FOR
 EXPONENTIAL SMOOTHED BOUNDARY TEMPERATURE,
 $\underline{x} = (0.5, 0.5), \Delta t = 0.01$

time	Analysis Solution	$1-e^{-4t/\Delta t}$		$1-e^{-2t/\Delta t}$	
		Numeric Solution	Error	Numeric Solution	Error
0.01	0.08	0.77 E-1	-0.46 E-2	0.07	-0.01
0.02	2.47	0.83	-1.64	0.74	-1.73
0.03	8.08	3.82	-4.25	3.45	-4.62
0.04	14.83	10.03	-4.79	9.29	-5.58
0.05	21.47	17.59	-3.88	16.58	-4.89
0.06	27.57	24.23	-3.34	23.30	-4.27
0.07	33.01	30.16	-2.84	29.33	-3.68
0.08	37.82	35.48	-2.34	34.79	-3.09
0.09	42.09	40.01	-2.08	39.37	-2.72
0.10	45.88	44.14	-1.74	43.56	-2.32
0.20	69.40	68.72	-0.68	68.48	-0.91
0.30	81.52	81.19	-0.33	81.06	-0.47
0.40	88.71	88.58	-0.16	88.50	-0.29
0.50	93.13	93.06	-0.07	93.01	-0.12

TABLE 5.17 CRACK-NICHOLSON METHOD, SOLUTION OBTAINED FOR
 EXPONENTIAL SMOOTHED BOUNDARY TEMPERATURE,

$$\underline{x} = (0.0, 0.0), \quad \Delta t = 0.001$$

time	Analytic Solution	$1-e^{-4t/\Delta t}$		$1-e^{-2t/\Delta t}$	
		Numeric Solution	Error	Numeric Solution	Error
0.01	0.62 E-9	-0.49 E-7	-50 E-7	-0.42 E-6	-0.42 E-6
0.02	0.23 E-3	-0.77 E-4	-0.15 E-3	0.81 E-4	-0.15 E-3
0.03	0.18 E-1	-0.13 E-2	-0.19 E-1	-0.14 E-2	-0.02
0.04	0.16	0.91 E-1	-0.07	0.88 E-1	-0.07
0.05	0.63	0.51	-0.12	0.50	-0.13
0.06	1.55	1.42	-0.14	1.40	-0.15
0.07	2.99	2.87	-0.12	2.85	-0.14
0.08	4.91	4.83	-0.08	4.80	-0.10
0.09	7.23	7.21	-0.02	7.18	-0.05
0.10	9.88	9.93	0.05	9.89	0.01
0.20	40.35	40.79	0.44	40.75	0.40
0.30	63.18	63.62	0.44	63.60	0.42
0.40	77.49	77.85	0.36	77.83	0.35
0.50	86.25	86.53	0.27	86.52	0.26

TABLE 5.18 CRANK-NICHOLSON METHOD, SOLUTION OBTAINED FOR
 EXPONENTIAL SMOOTHED BOUNDARY TEMPERATURE,
 $\underline{x} = (0.5, 0.5), \quad \Delta t = 0.001$

time	Analytic Solution	$1 - e^{-4t/\Delta t}$		$1 - e^{-2t/\Delta t}$	
		Numeric Solution	Error	Numeric Solution	Error
0.01	0.81 E-1	-0.11	-0.19	-0.95.E-1	-0.18
0.02	2.47	1.23	-1.24	1.18	-1.29
0.03	8.08	6.91	-1.17	6.82	-1.26
0.04	14.83	13.98	-0.85	13.88	-0.95
0.05	21.47	20.90	-0.57	20.81	-0.66
0.06	27.57	27.20	-0.37	27.11	-0.46
0.07	33.01	32.77	-0.23	32.70	-0.31
0.08	37.82	37.68	-0.14	37.62	-0.20
0.09	42.09	42.01	-0.08	41.95	-0.14
0.10	45.88	45.84	-0.04	45.79	-0.09
0.20	69.40	69.47	0.07	69.45	0.05
0.30	81.52	81.62	0.10	81.61	0.09
0.40	88.74	88.84	0.10	88.83	0.09
0.50	93.13	93.22	0.09	93.21	0.08

TABLE 5.19 TWO-STEP METHODS, SOLUTION OBTAINED BY STARTING WITH THE CRANK-NICHOLSON METHOD,

$$\underline{x} = (0.0, 0.0), \quad \Delta t = 0.01$$

		$\gamma = 3/2 \quad \beta = 4/5$		$\gamma = 3/2 \quad \beta = 1$		$\gamma = 1.2184 \quad \beta = 0.646$		$\gamma = 1 \quad \beta = 3/4$		$\gamma = 1/2 \quad \beta = 1/3$	
time	Analytic Solution	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.62 E-9	0.39 E-3	0.30 E-3	0.39 E-3	0.39 E-3	0.39 E-3	0.39 E-3	0.39 E-3	0.39 E-3	0.39 E-3	0.39 E-3
0.02	0.23 E-3	0.75 E-2	0.73 E-2	0.13 E-1	0.13 E-1	0.74 E-2	0.72 E-2	0.18 E-1	0.18 E-1	0.13 E-1	0.13 E-1
0.03	0.18 E-1	0.55 E-1	0.37 E-1	0.85 E-1	0.67 E-1	0.55 E-1	0.37 E-1	0.99 E-1	0.81 E-1	0.73 E-1	0.55 E-1
0.04	0.16	0.24	0.08	0.31	0.15	0.24	0.08	0.33	0.17	0.28	0.12
0.05	0.63	0.73	0.10	0.83	0.20	0.73	0.10	0.85	0.22	0.76	0.13
0.06	1.55	1.69	0.14	1.77	0.22	1.69	0.14	1.77	0.22	1.69	0.14
0.07	2.99	3.20	0.21	3.21	0.22	3.20	0.21	3.18	0.19	3.15	0.16
0.08	4.91	5.22	0.31	5.15	0.24	5.22	0.31	5.11	0.20	5.17	0.26
0.09	7.23	7.66	0.43	7.53	0.30	7.66	0.43	7.49	0.26	7.61	0.38
0.10	9.88	10.43	0.56	10.27	0.49	10.43	0.56	10.24	0.46	10.40	0.53
0.20	40.35	41.37	1.02	41.34	0.99	41.37	1.02	41.35	1.00	41.37	1.02
0.30	63.18	64.02	0.84	64.04	0.86	64.02	0.84	64.04	0.86	64.02	0.84
0.40	77.49	78.09	0.60	78.11	0.62	78.09	0.60	78.11	0.62	78.09	0.60
0.50	86.25	86.68	0.43	86.69	0.44	86.88	0.43	86.69	0.41	86.68	0.43

TABLE 5.20 TWO-STEP METHODS, SOLUTION OBTAINED BY STARTING WITH THE CRANK-NICHOLSON METHOD,

$x = (0.5, 0.5), \Delta t = 0.01$

time	Analytic Solution	$\gamma = 3/2 \quad \beta = 4/5$		$\gamma = 3/2 \quad \beta = 1$		$\gamma = 1.2184 \quad \beta = 0.646$		$\gamma = 1 \quad \beta = 3/4$		$\gamma = 1/2 \quad \beta = 1/3$	
		Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.81 E-1	0.23	0.15	0.23	0.15	0.23	0.15	0.23	0.15	0.23	0.15
0.02	2.47	2.13	-0.34	2.32	-0.15	2.12	-0.33	2.39	-0.08	2.32	-0.15
0.03	8.08	7.45	-0.63	7.21	-0.87	7.44	-0.62	7.03	-1.05	6.97	-1.11
0.04	14.83	15.07	0.24	14.13	0.70	15.11	0.74	13.82	-1.01	14.61	-0.22
0.05	21.47	22.19	0.72	21.46	-0.01	22.24	0.77	21.38	-0.09	22.37	0.90
0.06	25.57	28.38	0.81	28.16	0.59	28.41	0.83	28.37	0.80	28.77	1.20
0.07	33.01	33.95	0.94	33.93	0.92	33.96	0.95	34.20	1.19	33.90	0.89
0.08	37.82	38.71	0.89	38.85	1.03	38.71	0.89	39.00	1.18	38.50	0.68
0.09	42.08	42.97	0.89	43.11	1.03	42.96	0.88	43.13	1.05	43.26	1.18
0.10	45.88	46.69	0.81	46.85	0.97	46.68	0.80	46.82	0.94	46.61	0.73
0.20	69.40	69.83	0.43	69.87	0.47	69.83	0.43	69.87	0.47	69.86	0.46
0.30	81.52	81.83	0.31	81.85	0.33	81.83	0.31	81.85	0.33	81.79	0.27
0.40	88.74	88.97	0.23	88.98	0.24	88.96	0.22	88.98	0.24	88.99	0.25
0.50	93.13	93.29	0.16	93.30	0.17	93.29	0.16	93.30	0.17	93.29	0.16

TABLE 5.21 TWO-STEP METHODS, SOLUTION OBTAINED BY STARTING WITH THE CRANK-NICHOLSON METHOD,

$$\underline{x} = (0.0, 0.0), \quad \Delta t = 0.001$$

		$\gamma = 3/2$	$\beta = 4/5$	$\gamma = 3/2$	$\beta = 1$	$\gamma = 1.2184$	$\beta = 0.646$	$\gamma = 1$	$\beta = 3/4$	$\gamma = 1/2$	$\beta = 1/3$
time	Analytic Solution	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.62 E-9	0.24 E-5	0.24 E-5	0.14 E-6	0.14 E-6	0.26 E-5	0.26 E-5	0.15 E-7	0.14 E-7	-0.97 E-8	-0.10 E-7
0.02	0.23 E-3	-0.27 E-3	-0.50 E-3	-0.29 E-3	-0.52 E-3	-0.27 E-3	0.50 E-3	-0.27 E-3	-0.50 E-3	-0.25 E-3	-0.48 E-3
0.03	0.18 E-1	0.53 E-2	-0.12 E-1	0.66 E-2	-0.01	0.53 E-2	-0.13 E-1	0.67 E-2	-0.01	0.56 E-2	-0.12 E-1
0.04	0.16	0.15	-0.01	0.15	-0.01	0.15	-0.01	0.15	-0.01	0.15	-0.15 E-1
0.05	0.63	0.66	0.03	0.66	0.03	0.66	0.03	0.66	0.03	0.66	0.03
0.06	1.55	1.68	0.13	1.68	0.13	1.68	0.13	1.68	0.13	1.68	0.13
0.07	2.99	3.25	0.26	3.25	0.26	3.25	0.26	3.25	0.26	3.25	0.26
0.08	4.91	5.30	0.40	5.30	0.39	5.30	0.39	5.30	0.40	5.30	0.40
0.09	7.23	7.76	0.53	7.76	0.53	7.76	0.53	7.76	0.53	7.76	0.53
0.10	9.88	10.53	0.65	10.53	0.65	10.53	0.65	10.53	0.65	10.53	0.65
0.20	40.35	41.38	1.03	41.38	1.03	41.38	1.03	41.38	1.03	41.38	1.03
0.30	63.18	64.00	0.82	64.00	0.82	64.00	0.82	64.00	0.82	64.00	0.82
0.40	77.49	78.08	0.59	78.08	0.59	78.08	0.59	78.08	0.59	78.08	0.59
0.50	86.25	86.67	0.41	86.67	0.41	86.67	0.41	86.67	0.41	86.67	0.41

TABLE 5.22 TWO-STEP METHODS, SOLUTION OBTAINED BY STARTING WITH THE CRANK-NICHOLSON METHOD,

$\underline{x} = (0.5, 0.5), \Delta t = 0.001$

		$\gamma = 3/2$	$\beta = 4/5$	$\gamma = 3/2$	$\beta = 1$	$\gamma = 1.2184$	$\beta = 0.646$	$\gamma = 1$	$\beta = 3/4$	$\gamma = 1/2$	$\beta = 1/3$
time	Analytic Solution	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.81 E-1	-0.19	-0.27	-0.16	-0.24	-0.19	-0.27	-0.16	-0.24	-0.19	-0.27
0.02	2.47	2.25	-0.22	2.23	-0.24	2.25	-0.22	2.23	-0.24	2.25	-0.22
0.03	8.08	8.41	0.33	8.39	0.32	8.41	0.33	8.39	0.32	8.41	0.33
0.04	14.83	15.50	0.67	15.50	0.67	15.50	0.67	15.50	0.67	15.50	0.68
0.05	21.47	22.31	0.84	22.30	0.83	22.31	0.84	22.30	0.83	22.31	0.83
0.06	27.57	28.44	0.88	28.44	0.88	28.44	0.88	28.44	0.88	28.44	0.88
0.07	33.01	33.87	0.86	33.87	0.86	33.87	0.86	33.87	0.86	33.87	0.86
0.08	37.82	38.64	0.82	38.65	0.81	38.64	0.82	38.65	0.81	38.64	0.82
0.09	42.09	42.86	0.77	42.86	0.77	42.86	0.77	42.86	0.77	42.86	0.77
0.10	45.88	46.60	0.72	46.60	0.72	46.60	0.72	46.60	0.72	46.60	0.72
0.20	69.40	69.81	0.41	69.81	0.41	69.81	0.41	69.81	0.41	69.81	0.41
0.30	81.52	81.82	0.30	81.82	0.30	81.82	0.30	81.82	0.30	81.82	0.30
0.40	88.74	88.96	0.22	88.96	0.22	88.96	0.22	88.96	0.22	88.96	0.22
0.50	93.13	93.29	0.16	93.29	0.16	93.29	0.16	93.29	0.16	93.29	0.16

TABLE 5.23 TWO-STEP METHODS, SOLUTION OBTAINED BY STARTING WITH KNOWN STEADY CONDITIONS,

$$\underline{x} = (0.0, 0.0), \quad \Delta t = 0.01$$

		$\gamma = 3/2$	$\beta = 4/5$	$\gamma = 3/2$	$\beta = 1$	$\gamma = 1.2184$	$\beta = 0.646$	$\gamma = 1$	$\beta = 3/4$	$\gamma = 1/2$	$\beta = -1/3$
time	Analytic Solution	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.62 E-9	0.52 E-3	0.52 E-3	0.17 E-2	0.17 E-2	0.53 E-3	0.53 E-3	0.35 E-2	0.35 E-2	0.29 E-2	0.29 E-2
0.02	0.23 E-3	0.77 E-2	0.75 E-2	0.18 E-1	0.17 E-1	0.79 E-2	0.77 E-2	0.27 E-1	0.26 E-1	0.22 E-1	0.21 E-1
0.03	0.18 E-1	0.52 E-1	0.36 E-1	0.87 E-1	0.69 E-1	0.54 E-1	0.36 E-1	0.11	0.94 E-1	0.10	0.84
0.04	0.16	0.22	0.59 E-1	0.29	0.13	0.23	0.07	0.34	0.18	0.33	0.17
0.05	0.63	0.66	0.03	0.74	0.12	0.69	0.06	0.81	0.18	0.87	0.24
0.06	1.55	1.55	-0.38 E-2	1.56	0.01	1.60	0.05	1.66	0.11	1.83	0.28
0.07	2.99	2.95	-0.04	2.82	0.17	3.06	0.07	2.95	-0.04	3.36	0.34
0.08	4.91	4.85	-0.06	4.55	-0.35	5.02	0.11	4.74	-0.16	5.38	0.47
0.09	7.23	7.18	-0.04	6.73	-0.50	7.41	0.18	6.99	-0.24	7.89	0.66
0.10	9.88	9.87	-0.01	9.28	-0.60	10.14	0.26	9.62	-0.26	10.13	0.75
0.20	40.35	40.73	0.38	40.10	-0.25	41.05	0.70	40.61	0.26	41.41	1.06
0.30	63.18	63.59	0.41	63.22	0.04	63.80	0.62	63.55	0.38	63.9	0.72
0.40	77.49	77.84	0.36	77.62	0.13	77.96	0.48	77.82	0.33	77.90	0.42
0.50	86.25	86.52	0.27	86.39	0.14	86.60	0.35	86.51	0.26	86.47	0.22

TABLE 5.24 TWO-STEP METHODS, SOLUTION OBTAINED BY STARTING WITH KNOWN STEADY CONDITIONS,

$\underline{x} = (0.5, 0.5), \Delta t = 0.01$

		$\gamma = 3/2$	$\beta = 4/5$	$\gamma = 3/2$	$\beta = 1$	$\gamma = 1.2184$	$\beta = 0.646$	$\gamma = 1$	$\beta = 3/4$	$\gamma = 1/2$	$\beta = 1/3$
time	Analytic Solution	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.81 E-1	0.25	0.16	0.39	0.31	0.26	0.18	0.54	0.45	0.63	0.55
0.02	2.46	1.96	-0.50	2.19	-0.27	2.05	-0.45	2.44	-0.02	2.66	0.20
0.03	8.08	6.70	-1.38	6.16	-1.92	7.02	-0.15	6.43	-0.17	7.95	-0.13
0.04	14.82	13.71	-1.11	12.00	-2.82	14.34	-0.49	12.40	-2.42	14.82	0.00
0.05	21.47	20.67	-0.80	18.68	-2.79	21.43	-0.04	19.43	-2.04	23.33	1.86
0.06	27.57	26.96	-0.61	25.25	-0.23	27.69	0.12	26.37	1.20	28.61	1.05
0.07	33.01	32.64	-0.37	31.21	-1.79	33.31	0.30	32.48	-0.52	34.95	1.95
0.08	37.82	37.56	-0.26	36.46	-1.37	38.15	0.33	37.61	-0.21	38.19	0.37
0.09	42.09	41.94	-0.15	41.03	-1.06	42.45	0.36	41.97	-0.12	44.21	2.12
0.10	45.88	45.79	-0.09	45.03	-0.85	46.24	0.36	45.80	-0.08	46.29	0.41
0.20	69.40	69.46	0.06	69.16	-0.25	69.64	0.24	69.44	0.04	69.60	0.20
0.30	81.52	81.61	0.09	81.43	-0.09	81.72	0.20	81.40	0.08	81.60	0.08
0.40	88.74	88.84	0.10	88.73	-0.01	88.90	0.16	88.93	0.09	88.82	0.08
0.50	93.13	93.21	0.08	93.15	0.02	93.25	0.12	93.21	0.08	93.16	0.03

TABLE 5.25 TWO-STEP METHODS, SOLUTION OBTAINED BY STARTING WITH KNOWN STEADY CONDITIONS.

$$\underline{x} = (0.0, 0.0), \quad \Delta t = 0.001$$

	$\gamma = 3/2$	$\beta = 4/5$	$\gamma = 3/2$	$\beta = 1$	$\gamma = 1.2184$	$\beta = 0.646$	$\gamma = 1$	$\beta = 3/4$	$\gamma = 1/2$	$\beta = 1/3$	
time	Analytic Solution	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.62 E-9	0.27 E-5	0.27 E-5	0.64 E-7	0.64 E-7	0.27 E-5	0.27 E-5	0.37 E-9	-0.25 E-7	-0.76 E-7	-0.76 E-7
0.02	0.23 E-3	0.33 E-3	-0.56 E-3	-0.32 E-3	-0.54 E-3	-0.29 E-3	-0.52 E-3	-0.25 E-3	-0.48 E-3	-0.17 E-3	-0.40 E-3
0.03	0.18 E-1	0.67 E-2	-0.11 E-1	0.77 E-2	-0.10 E-1	0.59 E-2	-0.01	0.62 E-2	-0.12 E-1	0.34 E-2	-0.14 E-1
0.04	0.16	0.16	-0.24 E-2	0.16	-0.50 E-2	0.15	-0.01	0.14	-0.02	0.13	-0.03
0.05	0.63	0.69	0.66 E-1	0.68	0.05	0.67	0.04	0.64	0.01	0.61	-0.14 E-1
0.06	1.55	1.74	0.19	1.72	0.17	1.71	0.16	1.65	0.10	1.60	0.05
0.07	2.99	3.33	-0.35	3.30	0.31	3.28	0.29	3.20	0.21	3.13	0.14
0.08	4.91	5.41	0.51	5.37	0.46	5.35	0.44	5.25	0.34	5.16	0.25
0.09	7.23	7.89	0.66	7.84	0.61	7.82	0.58	7.70	0.47	7.60	0.37
0.10	9.88	10.68	0.79	10.62	0.73	10.59	0.71	10.46	0.58	10.36	0.48
0.20	40.35	41.52	1.17	41.46	1.11	41.44	1.08	41.31	0.96	41.31	0.95
0.30	63.18	64.09	0.91	64.05	0.88	64.04	0.86	63.96	0.78	69.07	0.89
0.40	77.49	78.13	0.65	78.11	0.63	78.10	0.61	78.05	0.57	78.21	0.73
0.50	86.25	86.70	0.45	86.89	0.43	86.68	0.43	86.61	0.40	86.82	0.57

TABLE 5.26 TWO-STEP METHODS, SOLUTION OBTAINED BY STARTING WITH KNOWN STEADY CONDITIONS,

$x = (0.5, 0.5), \Delta t = 0.001$

		$\gamma = 3/2$	$\beta = 4/5$	$\gamma = 3/2$	$\beta = 1$	$\gamma = 1.2184$	$\beta = 1$	$\gamma = 3/4$	$\beta = 1/2$	$\gamma = 1/2$	$\beta = 1/3$
time	Analytic Solution	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error	Numeric Solution	Error
0.01	0.81 E-1	-0.21	-0.29	-0.17	-0.25	-0.20	-0.28	-0.14	-0.22	-0.12	-0.20
0.02	2.47	2.49	0.02	2.39	-0.07	2.35	-0.12	2.13	-0.34	1.88	-0.59
0.03	8.08	8.77	0.69	8.62	0.55	8.56	0.48	8.22	0.15	7.93	-0.15
0.04	14.83	15.87	1.05	15.72	0.90	15.65	0.83	15.32	0.49	15.14	0.32
0.05	21.47	22.64	1.17	22.51	1.04	22.44	0.97	22.14	0.67	22.11	0.64
0.06	27.57	28.74	1.17	28.62	1.06	28.56	0.99	28.30	0.73	28.39	0.83
0.07	33.01	34.43	1.12	34.03	1.02	33.98	0.97	33.74	0.73	33.93	0.92
0.08	37.82	38.87	1.05	38.78	0.96	38.74	0.91	38.53	0.71	38.79	0.97
0.09	42.09	43.06	0.97	42.98	0.89	42.94	0.85	42.76	0.67	43.06	0.97
0.10	45.88	46.78	0.90	46.71	0.83	46.67	0.79	46.51	0.63	46.84	0.96
0.20	69.40	69.88	0.48	69.85	0.45	69.84	0.44	69.77	0.37	70.05	0.65
0.30	81.52	81.86	0.34	81.85	0.32	81.84	0.32	81.80	0.27	81.99	0.47
0.40	88.74	88.99	0.25	88.97	0.24	88.97	0.23	88.94	0.21	89.09	0.35
0.50	93.13	93.30	0.17	93.30	0.17	93.29	0.17	93.28	0.15	93.39	0.26

VI. CONCLUSIONS

A. One-Step Methods

1. Whether discontinuous or ramp smoothed, Liniger and Backward-difference methods give comparatively inaccurate results. Crank-Nicholson and Galerkin methods are of acceptable accuracy in the discontinuous case.
2. In the case of larger steplengths, averaging and exponential smoothing techniques reduce the accuracy of the Crank-Nicholson method. Accuracy is improved by reducing the steplength. On the other hand, computation time increases considerably.
3. Galerkin method when applied with smoothing the discontinuity by a ramp function gives better results than do all the above methods.
4. Crank-Nicholson method when applied with smoothing as proposed by Zienkiewicz gives excellent results.

B. Two-Step Methods

1. Irrespective of the method of starting, Lees, Dupont and fully implicit methods give less accurate results than Liniger and Galerkin methods. Galerkin and Liniger methods show acceptable accuracy in the case of Crank-Nicholson starting.
2. Liniger method shows better performance in the case of starting from an initially steady state. Galerkin method with this starting technique gives excellent results.

APPENDIX 1
LINEAR K^{TH} ORDER DIFFERENCE
EQUATIONS WITH CONSTANT COEFFICIENTS

A linear k^{th} order difference equation with constant coefficients is of the form

$$a_k y_{n+k} + a_{k-1} y_{n+k-1} + \dots + a_0 y_n = b_{n+k} \quad (\text{A.1.1})$$

Solving the above equation consists of finding a sequence of numbers $\{y_n\}$, where a_k, a_{k-1}, \dots, a_0 are constants and b_{n+k} are given functions of n defined for $n = 0, 1, 2, \dots$. It will be assumed that $a_k \neq 0$ and $a_0 \neq 0$.

The general solution $\{y_n\}$ can be written of the form $\{\hat{y}_n + \psi_n\}$, where $\{\hat{y}_n\}$ is the solution of the homogeneous equation

$$a_k y_{n+k} + a_{k-1} y_{n+k-1} + \dots + a_0 y_n = 0$$

and $\{\psi_n\}$ is same particular solution of the nonhomogeneous equation.

A possible method of finding a particular solution $\{\psi_n\}$, is to solve the nonhomogeneous equation with the starting values $\psi_0 = \psi_1 = \dots = \psi_{k-1} = 0$. For $m = 0, 1, 2, \dots$ let the sequence $\{y_{n,m}\}$ satisfy the conditions

$$\begin{aligned} y_{n,m} &= 0 & n < m \\ y_{m,m} &= 1/\alpha_n \end{aligned}$$

and let it be a solution of homogeneous equation for $n > m$. It is proved in Henrici [8] that, the particular solution of the difference equation (A.1.1) is

$$\psi_n = \sum_{m=k}^n b_n y_{n,m}, \quad n \geq k$$

The next step is to find the homogeneous solution of Eq. (A.1.1). One can easily verify that, $y_n = \xi^n$ is a homogeneous solution of the difference equation if and only if ξ is a root of the polynomial

$$P(\xi) = a_k \xi^k + a_{k-1} \xi^{k-1} + \dots + a_0.$$

It is clear that $\xi = 0$ is not a root of $P(\xi)$, since $a_0 \neq 0$ [8].

If the polynomial $P(\xi)$ has k distinct roots, $\xi_1, \xi_2, \dots, \xi_k$, then it can be proved that the solutions are linearly independent, thus form a fundamental system [8]. Hence the homogeneous solution

$$\hat{y}_n = \sum_{j=1}^k c_j \xi_j^n$$

and, therefore, the general solution is

$$y_n = \sum_{j=1}^k c_j \xi_j^n + \psi_n$$

where c_j , $j = 1, 2, \dots, k$ are arbitrary constants.

If the polynomial $P(\xi)$ has a root, say ξ_1 , of multiplicity p , then it can be shown that $(n!/(n-i)!) \xi_1^n$, $i = 1, 2, \dots, p-1$ are also solutions of the homogeneous equation and the set of solutions $\{(n!/(n-i)!) \xi_1^n\}$, $i=0, 1, \dots, p-1$ and $\{\xi_j\}$, $j = p+1, p+2, \dots, k$ form a fundamental system [8]. The homogeneous solution in this case will be

$$\hat{y}_n = \sum_{i=0}^{p-1} c_{i+1} \frac{n!}{(n-i)!} \xi_1^n + \sum_{j=p+1}^k c_j \xi_j^n$$

and, therefore, the general solution will be

$$y_n = \sum_{i=0}^{p-1} c_{i+1} \frac{n!}{(n-i)!} \xi_1^n + \sum_{j=p+1}^k c_j \xi_j^n + \psi_n$$

where c_j , $j = 1, 2, \dots, k$ are arbitrary constants.

For the most general case, where the polynomial $P(\xi)$ has roots ξ_j , $j = 1, 2, \dots, r$ and each root ξ_j is a root of multiplicity P_j , $\sum_{j=1}^r P_j = k$, then the general solution will be

$$y_n = \sum_{j=1}^r \sum_{i=0}^{P_j-1} c_{j,i+1} \frac{n!}{(n-i)!} \xi_j^n + \psi_n$$

where $c_{j,i}$, $i = 1, 2, \dots, P_j$, $j = 1, 2, \dots, r$ are arbitrary constants [8].

APPENDIX 2

**DERIVATION DETAILS FOR θ -METHODS AND
ZIENKIEWICZ THREE-LEVEL SCHEMES**

θ Values for the θ -Methods

We have shown that

$$\theta = \frac{\int_0^1 w_j \xi d\xi}{\int_0^1 w_j d\xi}$$

For the weight function $w_j = \delta(\xi - (1/2))$, we get

$$\int_0^1 w_j d\xi = \int_0^1 \delta(\xi - (1/2)) d\xi = 1$$

Thus

$$\theta = \int_0^1 \delta(\xi - (1/2)) \xi d\xi = 1/2$$

which corresponds to the Crank-Nicholson method. For the weight function $w_j = 1$, we get

$$\int_0^1 w_j d\xi = \int_0^1 \xi d\xi = 1$$

Thus

$$\theta = \int_0^1 w_j \xi d\xi = \int_0^1 \xi d\xi = (1/2) \xi^2 \Big|_0^1 = 1/2$$

which also corresponds to the Crank-Nicholson method. For the weight function $w_j = \xi$, we get

$$\int_0^1 w_j d\xi = \int_0^1 \xi d\xi = (1/2) \xi^2 \Big|_0^1 = 1/2$$

Thus

$$\theta = 2 \int_0^1 w_j \xi d\xi = 2 \int_0^1 \xi^2 d\xi = 2(1/3)\xi^3 \Big|_0^1 = 2/3$$

which corresponds to the Galerkin method.

For the weight function $w_j = \delta(\xi-1)$, we get

$$\int_0^1 w_j d\xi = \int_0^1 \delta(\xi-1) d\xi = 1$$

Thus

$$\theta = \int_0^1 \delta(\xi-1) \xi d\xi = 1$$

which corresponds to the Backward-difference method.

γ and β Values for the Zienkiewicz Three-Level Schemes

We have shown that

$$\gamma = \frac{\int_{-1}^1 w_j ((1/2) + \xi) d\xi}{\int_{-1}^1 w_j d\xi}$$

$$\beta = \frac{(1/2) \int_{-1}^1 w_j \xi (1 + \xi) d\xi}{\int_{-1}^1 w_j d\xi}$$

For the weight function $w_j = \delta(\xi-1)$, we get

$$\int_{-1}^1 w_j d\xi = \int_{-1}^1 \delta(\xi-1) d\xi = 1$$

Thus

$$\gamma = \int_{-1}^1 \delta(\xi-1) ((1/2) + \xi) d\xi = (1/2) + 1 = 3/2$$

$$\beta = (1/2) \int_{-1}^1 \delta(\xi-1)\xi(1+\xi)d\xi = (1/2)(1+1) = 1$$

which corresponds to the fully-implicit two-step method.

For the weight function $w_j = \delta(\xi+1)/2$, we get

$$\int_{-1}^1 w_j d\xi = (1/2) \int_{-1}^1 \xi(\xi+1)d\xi = (1/2) \left((\xi^3/3) + (\xi^2/2) \right) \Big|_{-1}^1 = 1/3$$

Thus

$$\gamma = 3 \int_{-1}^1 (1/2)\xi(\xi+1)((1/2)+\xi)d\xi = (3/2) \int_{-1}^1 (\xi^3 + (3/2)\xi^2 + \xi)d\xi$$

$$\gamma = (3/2) \left((\xi^4/4) + (\xi^3/2) + (\xi^2/4) \right) \Big|_{-1}^1 = 3/2$$

$$\beta = (3/2) \int_{-1}^1 (1/2)\xi(\xi+1)\xi(\xi+1)d\xi = (3/4) \int_{-1}^1 (\xi^4 + 2\xi^3 + \xi^2)d\xi$$

$$\beta = (3/4) \left((\xi^5/5) + (\xi^4/2) + (\xi^3/3) \right) \Big|_{-1}^1 = 4/5$$

which corresponds to Galerkin two-step method.

REFERENCES

1. O.C. Zienkiewicz, "The Finite Element Method", 3rd ed., McGraw-Hill, London, 1977.
2. O.C. Zienkiewicz, J.M. Bettencourt and G. Cantin, "Consistent Use of Finite Elements in Time and the Performance of Various Recurrence Schemes for the Heat Diffusion Equation", Int. J. Num. Meth. Eng., Vol. 17, 931-938 (1981).
3. M. Zlamal, "Finite Element Methods in Heat Conduction Problems", Proc. Conf. Finite Element Meth. Brunel (1975).
4. W.L. Wood, "On the Zienkiewicz Three- and Four-Time Level Schemes Applied to the Numerical Integration of Parabolic Equations", Int. J. Num. Meth., Vol. 12, 1717-1726 (1978).
5. M.A. Hogge, "A Comparison of Two- and Three-Level Integration Schemes for Non-Linear Heat Conduction", Num. Meths. in Heat Trans. (R.W. Lewis, K. Morgan and O.C. Zienkiewicz, Eds.), Wiley, New York, 1981, pp. 75-90.
6. E.L. Wilson and R.E. Nickell, "Application of Finite Element Method to Heat Conduction Analysis", Nuclear Engineering Design, Vol. 4, 276-286 (1966).
7. R.P. Henrici, "Error Propagation for Difference Methods", Wiley, New York-London (1963).
8. R.P. Henrici, "Discrete Variable Methods in Ordinary Differential Equations", Wiley, New York-London (1962).
9. J.D. Lambert, "Computational Methods in Ordinary Differential Equations", Wiley, London (1973).
10. C.W. Gear, "Numerical Initial Value Problems in Ordinary Differential Equations", Prentice-Hall, Englewood Cliffs, N.J. (1971).
11. G. Dahlquist, "Convergence and Stability in the Numerical Integration of Ordinary Differential Equations", Math. Scand., Vol. 4, 33-53 (1956).
12. G. Dahlquist, "A Special Stability Problem for Linear Multistep Methods", BIT, Vol. 3, 27-43 (1963).

13. O.B. Widlund, "A Note on Unconditionally Stable Linear Multistep Methods", BIT, Vol. 7, 65-70 (1967).
14. C.W. Cryer, "A New Class of Highly-Stable Methods. A_0 -Stable Methods", BIT, Vol. 13, 153-159 (1973).
15. M. Zlamal, "Finite Element Multistep Discretizations of Parabolic Boundary Value Problems", Math. Comp., Vol. 29, 350-359 (1975).
16. W.L. Wood, "Control of Crank-Nicholson Noise in the Numerical Solution of the Heat Conduction Equation", Int. J. Num. Meth., Vol. 11, 1059-1065 (1977).
17. W. Liniger, "Global Accuracy and A-Stability of One- and Two-step Integration Formulae for Stiff Ordinary Differential Equations", Proc. Conf. Num. Solution Differential Equations, Dundee (1969).
18. M. Lees, "A Linear Three-Level Difference Scheme for Quasilinear Parabolic Equations", Math. Comp., Vol. 20, 516-622 (1966).
19. T. Dupont, et.al., "Three-Level Galerkin Methods for Parabolic Equations", SIAM J. Num. Analysis, Vol. 11, 392-410 (1974).
20. W.L. Wood and R.W. Lewis, "A Comparison of Time Marching Schemes for the Transient Heat Conduction Equation", Int. J. Num. Meth., Vol. 9, 679-689 (1975).
21. V.S. Arpaci, "Conduction Heat Transfer", Addison-Wesley, Massachusetts (1966).