FINITE ELEMENT ANALYSIS OF EIGENVALUE PROBLEMS

IN THE

STABILITY OF FLUID MOTIONS

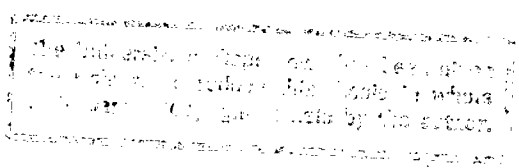
A thesis submitted to the
UNIVERSITY OF CAPE TOWN
in fulfilment of the requirements for the degree of
MASTER OF SCIENCE

by

Helena du Toit B.Sc. (Hons)

Department of Applied Mathematics
University of Cape Town
Rondebosch
South Africa

September 1986



To My Parents

DECLARATION OF CANDIDATE

I hereby declare that this thesis is my own work and that it has not been submitted for a degree at any other university.

HFdutoit

H. F. du Toit

ACKNOWLEDGEMENTS

I would like to express my most sincere thanks and appreciation to Professor B.D.Reddy, supervisor of this project. I would also like to thank my fiancé for the support he has given me in completing this thesis.

SUMMARY

Variational eigenvalue problems for linear and energy stability theory of buoyancy-driven flow are studied. Critical Rayleigh numbers are determined by the finite element method. The penalty method is used to approximate the incompressibility condition. We consider the stability of Boussinesq flows in a two-dimensional box in which internal heat sources are present. The influence of side walls are studied for various boundary conditions and width-to-height ratios. The temperature boundary conditions include fixed heat flux at the side walls, fixed temperature and fixed heat flux at the bottom surface, and a general convective exchange at the upper surface which includes fixed temperature and fixed heat flux as special cases. The velocity boundary conditions include rigid side walls and rigid and free upper and lower surfaces.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

Stability of fluid motions

Fluids in motion exhibit varied and sometimes complex flow patterns. In order to understand such phenomena, mathematical formulations of the laws that govern the behaviour of the fluid are sought. The equations describing the actual fluid motion allow some patterns of flow as solutions. These patterns of flow are only possible for certain ranges of the parameters that characterize them. Outside these ranges the patterns of flows are replaced by other flows, with transition from one type of flow to another occurring as a result of the instability of the former. Stability theory enables us to obtain critical values of parameters which separate the different types of flows.

For example, consider flows of viscous incompressible fluids, which are described by the Navier-Stokes equations. We start with a solution of these equations satisfying the boundary and initial conditions; this solution is called the basic motion. Suppose now that the initial condition is disturbed; this gives rise to a disturbed solution which also satisfies the same equations and boundary conditions. By subtracting the equations for the basic flow from those for the disturbed flow, we obtain the equations governing the evolution of the disturbance. When investigating the stability of the basic motion to disturbances we want to know whether the disturbance grows or decays with time.

In the case of the Navier-Stokes equations the parameter governing the stability of the basic solution is the Reynolds number, Re. When Re is less than a certain critical value, Re_E, all solutions tend monotonically to a single flow, the basic

flow, and the energy of any disturbance of this flow will decay from the initial instant. If Re > Re_E, disturbances of the basic flow will grow at first, and may later die away. This critical Reynolds number, which separates the monotonically decaying disturbances from those whose energy increase initially is called the energy stability limit (see Figure 1.1). There is a second critical Reynolds number Re_G, called the global stability limit, where $Re_G \geq Re_E$. When Re < Re_G the basic flow is stable since disturbances, whatever their size will die away eventually. When Re > Re_G the basic flow is unstable to some disturbances although it may be (conditionally) stable to small disturbances. The third critical Reynolds number, indicated in Figure 1.1, is called the linear stability limit, Re_L, where $Re_L \geq Re_G$. When Re > Re_L, all disturbances of the basic motion will grow, no matter how small they are. In this case the basic flow loses its stability to more complicated flows. As the Reynolds number is increased further, these flows may in turn lose their stability to other more complicated flows, and so on.

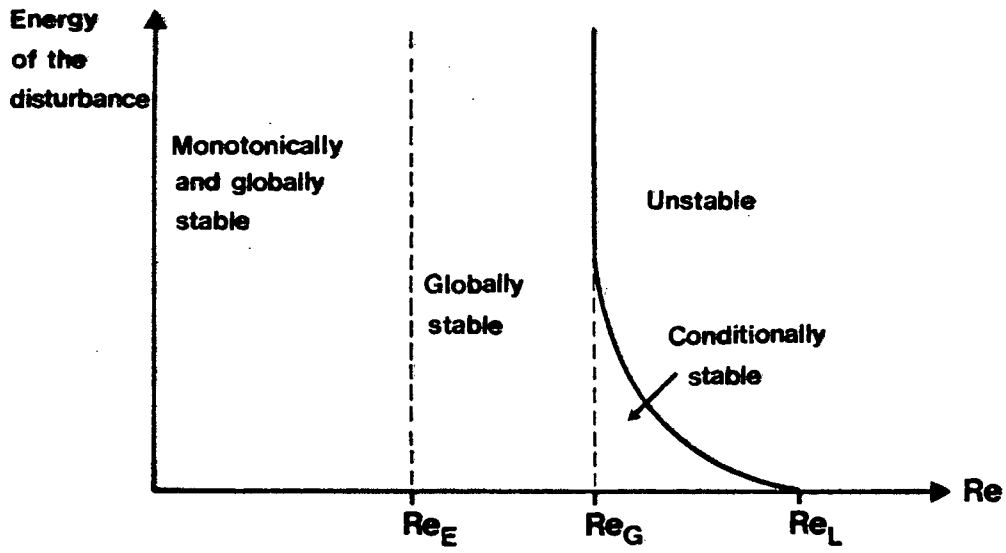


Figure 1.1

There are two well-known methods for obtaining critical values separating different types of flows. The linear theory of stability results when the disturbance is assumed small, so that a linearized version of the equations governing the disturbance may be used. In the energy theory of stability the size of the disturbance is unrestricted.

In the linear theory we consider only infinitesimal disturbances so that all terms which are nonlinear in the disturbance may be neglected in the equations governing the evolution of the disturbance. Dependence on time is eliminated in the resulting linearized equation by seeking a separable solution which is proportional to exp(σt). This leads to an eigenvalue problem for σ , where σ may be complex. The condition for stability is that the real part of σ be negative. We have neutral stability when the real part of σ is zero, and instability when it is positive (see Chandrasekhar (1961)). For any Reynolds number, Re, we can work out a smallest eigenvalue $\overline{\sigma}$. The critical Reynolds number of linear theory, Re_L, is the value for which the real part of $\overline{\sigma}$ is zero. Linear stability theory gives no prediction about instability due to sufficiently large disturbances, which may occur even when this theory indicates stability. Thus, the criteria of linearized theory can only give sufficient conditions for instability since a flow which is judged stable by the linear theory may become unstable to disturbances of finite size.

The energy theory leads to an eigenvalue problem for the critical Reynolds number of energy theory, and gives a sufficient condition for stability. In this theory one starts with the disturbance equations and, by applying the divergence theorem, obtains the equations governing the evolution of the energy of the disturbance. The energy equation consists of the average

dissipation and energy production which couples the basic flow to the disturbance. Osborne Reynolds was the first to derive such an energy equation in 1895; however, his energy equation did not apply to the basic motion and a disturbance, but rather to fluctuations from a mean motion (see Joseph (1976 I)). The modern theory of energy stability dates from studies by Serrin in 1959. He was the first to apply the variational method to the problem, which results from decomposing the motion into a basic motion and a disturbance. Serrin formulated a stability theorem proving the existence of a stability limit, Re_E, such that if $Re < Re_E$ the energy of the disturbance decreases at each and every instant. The proof that $Re < Re_E$ implies exponential decay is given in the paper by Joseph and Shir (1966).

When the Reynolds number is small, the basic flow is globally stable. This flow will lose its stability, however, if the Reynolds number is increased past the level of the critical Reynolds number of linear theory. Out of the instability of the basic flow new solutions develop. Bifurcation is the phenomenon which occurs when two solutions are possible for a given Reynolds number. According to the criterion of linear theory the energy of all sufficiently small disturbances decays to zero if $Re < Re_L$. However, if $Re_E < Re_L$, solutions exist whose energy do not decay even though the stability criterion of linear theory is satisfied. Such bifurcating solutions, which exist for values of Re judged stable by the linear theory, are called subcritical.

A comprehensive treatment and review of the stability problem for the Navier-Stokes equations has been given by Joseph (1976 I).

Fluid flows with heat conduction

Consider flows in which temperature variations are introduced through temperature differences at the fluid boundaries or by internal heat generation. The temperature variations give rise to variations in the properties of the fluid, such as density. The exact Navier-Stokes equations can be simplified by ignoring variations in the fluid density except for the variations of density in the buoyant force term. The equations which result as a consequence of these simplifications, together with an equation governing the conduction of heat, are called the Oberbeck-Boussinesq equations. In the Oberbeck-Boussinesq equations the parameters governing the stability of the basic solution are the Reynolds number, Re, and another non-dimensional number called the Rayleigh number, R, which is a measure of the "average" temperature gradient found in the fluid. The problem of finding a stability limit now consists of finding the largest region in the Rayleigh-Reynolds number plane for which the fluid will be globally stable. This problem is simplified by introducing a positive parameter, c, defined by Re = c R, and seeking stability limits for fixed c, thereby eliminating the explicit dependence on the Reynolds number.

A problem which is adequately modelled by the Oberbeck-Boussinesq approximations, and whose stability characteristics have been investigated extensively, is the Bénard problem; a horizontal layer of fluid in which an adverse temperature gradient is maintained by heating the lower surface. The fluid at the top of the layer will be heavier than the fluid at the bottom due to density differences caused by thermal gradients. Such a situation is potentially unstable. The instability is opposed by the frictional action of the viscosity of the fluid and also by its thermal conductivity which tends to remove temperature

differences in the fluid. Instability will only set in when the adverse temperature gradient is strong enough to overcome these.

Bénard convection has been extensively investigated, both experimentally and theoretically. The earliest experiments demonstrating the onset of thermal instability were those by Bénard in 1900 (see Chandrasekhar (1961)). His experiments were carried out on very thin layers of fluid, about a millimeter in depth, or less, standing on a levelled metallic plate maintained at a constant temperature. The upper surface, which was in contact with air, was at a lower temperature. Bénard was particularly interested in the viscosity of the fluids with which he experimented. He found that a critical adverse temperature gradient had to be exceeded before instability set in, and that the resulting motion had a cellular character.

Experimental work by Bénard led to theoretical investigation on buoyancy-driven instability by Lord Rayleigh in 1916 (see Joseph (1976 II)). His buoyancy theory was, however, not applicable to Bénard's observations as was assumed at the time. It has only been shown recently that Bénard cells were primarily induced by the surface tension gradients resulting from temperature variations across the free upper surface of the layer of fluid. Rayleigh showed that the stability, or otherwise, of a layer of fluid heated from below depends on the value of the Rayleigh number R and that instability sets in when R exceeds a certain critical value R_{crit}. Rayleigh obtained the equations expressing the conditions of neutral stability by neglecting second order terms and setting all time variations to zero.

Pellew and Southwell (1940) established that in seeking the conditions for maintained convective motion only real exponential time-factors need be considered. In fact, Chandrasekhar (1961)

shows that for the Bénard problem σ is real. If the imaginary part of σ is zero when the real part is zero the principle of exchange of stabilities is valid. Since σ is real it follows that the transition from stability to instability must occur via a stationary state. The equations governing the state of neutral stability are therefore obtained by setting $\sigma = 0$ in the linearized equations. The solution to the stability problem is then obtained by determining the lowest eigenvalue for R, called R_{crit} at which instability will occur. In the Bénard problem the linear and energy stability limits coincide and subcritical solutions are not possible. The smallest Rayleigh number, R_{crit}, for which $\sigma = 0$ is the global stability limit, i.e., we have global stability when $R < R_{crit}$ and instability when $R > R_{crit}$.

Sparrow, Goldstein and Jonsson (1964) determined analytically the conditions for the onset of convection in a horizontal layer of fluid bounded only in the vertical direction, for a broad range of thermal boundary conditions. They further investigated the effect of a non-linear temperature distribution on the stability of the fluid. The non-linear distribution arises when heat is generated uniformly throughout the fluid. In this situation instability sets in when a modified Rayleigh number exceeds a critical value. Joseph (1965) investigated a similar problem by the energy method. He compares his results to those of Sparrow et al. (1964) for the linear theory and shows that for the Bénard problem the linear and energy stability limits coincide. In this case, all stable disturbances, whatever their size, will decay exponentially from the initial instant. When the fluid layer contains heat sources, Joseph and Shir (1966) found that the linear and energy stability limits differ only by a small amount and that subcritical instabilities are confined to a narrow band of Rayleigh numbers between the energy and linear stability limits. Shir and Joseph (1968) obtained an improved

criterion for stability and uniqueness through the formulation of a variational maximum problem.

Pellew and Southwell (1940) investigated the effect of side walls on the stability of flows for a variety of boundary conditions. Mathematical difficulties arise when horizontally bounded domains are considered because conditions imposed on the side walls cannot be satisfied and separation of variables is made impossible in the analytic solution (see Chandrasekhar (1961)). Davis (1967) investigated the influence of side walls on the convective process in a rectangular box. He used a Galerkin procedure to obtain approximate critical Rayleigh numbers whose corresponding approximate eigenfunctions satisfy all boundary conditions and continuity exactly. Charlson and Sani (1970) gave a similar analysis when the fluid is contained in a right circular cylinder. Hall and Walton (1977) investigated linear and non-linear stability in a two-dimensional box for a certain class of boundary conditions.

Davis (1969) developed an energy theory for motion in a horizontal, heated layer subject to both buoyancy and surface tension effects and showed that the equations governing the energy theory are the symmetric part of the time-independent linear theory problem, and that surface tension terms behave like a bounded perturbation to the Bénard problem.

The problem of Bénard convection in an infinite horizontal layer can be solved exactly (see Chandrasekhar (1961)). However, exact solutions are not always possible for bounded domains, and approximate solutions have to be sought. The finite element method is a general technique for constructing approximate solutions to boundary-value problems. This method is a special case of the Galerkin method for solving problems approximately in

finite-dimensional subspaces; in the finite element method the subspaces are conventionally spanned by piecewise polynomials. The method is rapidly gaining popularity in computational fluid mechanics (see, for example, Chung (1978) and Taylor and Hughes (1981)).

Finite element solutions for problems in fluid stability have received a limited treatment. Jackson and Winters (1984) used a standard finite element program for buoyancy driven-flow to obtain critical Rayleigh numbers for the Bénard problem as a function of the width to height ratio of the container for various different boundary conditions. They also studied the flow at Rayleigh numbers greater than the critical value. Van Steeg and Wesseling (1978) studied the accuracy of several finite elements by solving the Bénard problem and determining the critical Rayleigh number. The convergence of the finite element approximation of the critical Rayleigh number and cell size for the onset of Bénard convection in an infinite horizontal layer was investigated by Winters and Cliffe (1985).

Outline of this work

The aim of this study is to show how the finite element method can be used to obtain critical Rayleigh numbers for buoyancy-driven stability problems corresponding to the linear and energy stability theory. We consider problems in bounded domains, and for which internal heat sources are present. The influence of side walls are studied for various different boundary conditions and geometries. We restrict our work to the study of flows which satisfy the Oberbeck-Boussinesq equations; that is, we consider only incompressible flow (pressure variations do not produce any significant density variations), and fluids of constant

viscosity. The penalty method is used to approximate the incompressibility condition by one of small compressibility. This has the advantage that the hydrostatic pressure is eliminated from the formulation, thereby reducing the number of unknowns.

In Chapter 2 we formulate the equations governing the evolution of the disturbance of the basic flow, which leads to variational eigenvalue problems for the linear and energy stability theory. Finite element approximations of these eigenvalue problems are discussed in Chapter 3. The computer program used to solve the finite element approximations to the eigenvalue problems is discussed in Chapter 4. Results obtained are discussed in Chapter 5 and compared to published work where appropriate. We conclude with a short discussion in Chapter 6.

CHAPTER 2. THEORY OF CONVECTIVE STABILITY

When investigating the stability of a fluid, we want to know how the fluid reacts to disturbances. Specifically, we ask: if the basic flow is disturbed, does the disturbance decay or grow with time? If the disturbance decays, we say the basic flow is stable with respect to the particular disturbance. The basic flow is said to be conditionally stable if it is stable only for disturbances smaller than a given magnitude, otherwise it is globally stable. If the energy of the disturbance decreases for all time from the initial instant, we say that the basic motion is monotonically stable.

The stability of motion of a fluid subject to thermal gradients depends on the Reynolds number Re and on the non-dimensional quantity R known as the Rayleigh number; R is defined by

$$R = \sqrt{\frac{\alpha g\ T'\ d^3}{\kappa \nu}}$$

where d is the characteristic length, T' is a characteristic temperature, g is the acceleration due to gravity, and α , ν and κ are respectively the coefficient of thermal expansion, kinematic viscosity, and thermal diffusivity. In order to find quantitative stability criteria we need to calculate critical values for the Rayleigh number. There are two criteria for obtaining critical values of R : one is based on the linear theory of stability and the other on energy theory.

The linear stability theory enables us to make quantitative predictions about when instability sets in. In this method only infinitesimal disturbances are considered. When R exceeds the critical stability limit, R_{crit}, obtained by the linear method

the disturbance will grow exponentially with time. When $R < R_{crit}$ *the basic flow is conditionally stable, since it may be unstable to finite disturbances. Thus, the linear theory leads to a criterion which is sufficient for instability. It does not guarantee stability.*

The energy stabilty theory allows for disturbances of finite size. The critical stability limit R_{crit}, *obtained by the energy method, separates the monotonically decaying disturbances from those whose energy increases initially, as illustrated in Figure 2.1. When* $R < R_{crit}$ *all disturbances of the basic flow will decay from the initial instant. When* $R > R_{crit}$ *the energy of the disturbance will increase for a time and may ultimately decay. The energy method leads to a sufficient condition for the global stability of the basic flow. It is silent about instability.*

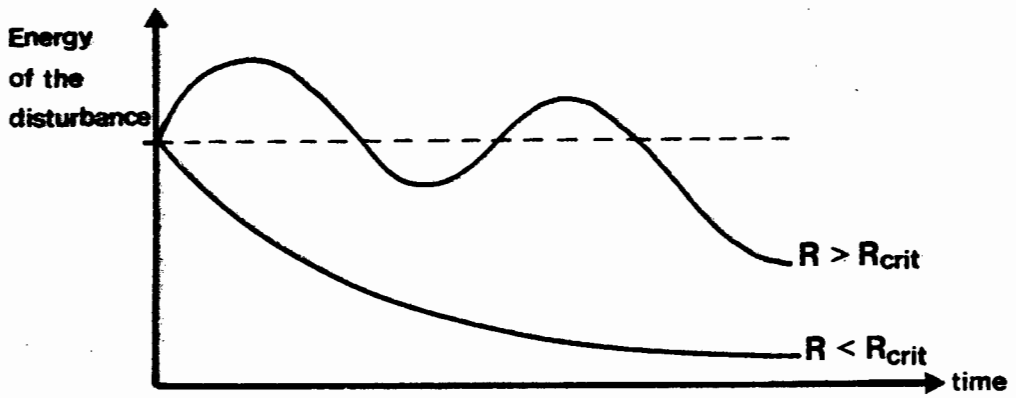


Figure 2.1

In Section 2.1 we formulate the equations describing the basic flow of the fluid. We disturb the initial conditions and obtain the equations governing the evolution of these disturbances. The linear stability theory is given in Section 2.2. We obtain the equations governing the disturbances of the basic flow by neglecting products and powers of the disturbances

and retain only the terms which are linear in them. We then arrive at an eigenvalue problem for the linear theory. In Section 2.3 we formulate the energy stability theory. We define the energy of the disturbance as a linear combination of the kinetic energy and the "thermal" energy. To study the stability problem we first fix the coupling constant in the linear combination and define a maximum problem to find a critical stability number for global stability. This leads to a variational eigenvalue problem for the energy theory. Since each choice of the coupling constant gives a different stability number , this leads to the problem of finding the optimal coupling constant, i.e. that which gives the largest stability number. Sufficient conditions are given for the existence of the maximum. The formulation of the equations follows that of Joseph (1976 II) and Shir and Joseph (1968), though our derivation of the weak formulation appears not to have been carried out previously.

2.1 Equations of flow

Temperature variations within a fluid give rise to variations in density as well as material parameters such as viscosity. Simplifications are possible when the variations of these parameters and the coefficient of volume expansion are sufficiently small. In the Boussinesq equations [Joseph (1976 II)] the fluid is assumed to have a uniform density, i.e. the motion is as if incompressible. Density variations are recognized only in the gravitational term in the equation of motion.

Accordingly, we can treat the density, ρ, as a constant in all terms in the equation of motion except in the gravitational term. The Boussinesq approximation further assumes a linear dependence of ρ on T, the temperature. As an extension to the above simplifications the fluid is assumed to have a constant heat capacity C, and thermal conductivity k.

Let Ω be a two-dimensional region with boundary Γ and $\underline{x} = \{x_i\}$, $i=1,2$ a general point in Ω. The equations for incompressible, heat conducting and convective flow of a viscous fluid in terms of the foregoing approximations are given by the Oberbeck-Boussinesq equations, in terms of the velocity $\underline{U}(\underline{x},t)$, temperature $T(\underline{x},t)$ and pressure $P(\underline{x},t)$. These equations are:

(a) the incompressibility condition:

$$div\ \underline{U} = 0 \qquad on\ \Omega \quad ; \qquad (2.1)$$

(b) the equation of state:

$$\rho = \rho_o\ (\ 1 - \alpha(\ T - T_o\ )) \qquad on\ \Omega \qquad (2.2)$$

where α is the coefficient of volume expansion and

T_o is the temperature at which $\rho = \rho_o$;

(c) the equation of motion: (Navier-Stokes equation)

$$\rho \underline{a} = \rho \underline{g} + div \underline{T} \qquad on\ \Omega$$

where $\underline{T} = -p\underline{I} + 2\mu \underline{D}[\underline{U}]$ is the stress ,

$\underline{D}[\underline{U}] = {}^1/_2\ (\nabla \underline{U} + (\nabla \underline{U})^T)$ is the deformation rate tensor,

and $\underline{g}$ *is the gravitational acceleration .*

Substituting for $\underline{a}$ *and* $\underline{T}$, *we have*

$$(\partial \underline{U}/\partial t + (\nabla \underline{U})\underline{U} = \rho \underline{g} - \nabla P + 2\mu \, div \underline{D}[\underline{U}].$$

The above equation in terms of the approximations and equation (2.2) becomes

$$\partial \underline{U}/\partial t + (\nabla \underline{U})\underline{U} = -1/\rho_o \nabla P + (1 - \alpha(T-T_o))\underline{g} + 2\nu \, div(\underline{D}[\underline{U}]) \qquad (2.3)$$

where $\nu = \mu/\rho_o$ *is the kinematic viscosity .*

(d) <u>*the heat conduction equation:*</u>

$$\partial T/\partial t + (\nabla T)\underline{U} = \kappa \nabla^2 T + Q \qquad \text{on } \Omega \qquad (2.4)$$

where $\kappa = k/\rho_o C$ *is the coefficient of thermal diffusivity and*
Q *is a prescribed heat source field;*

(e) <u>*the boundary conditions:*</u>

(i) temperature conditions:

$$T = T_o \quad \text{(prescribed temperature)} \quad \text{on } \Gamma_T ,$$

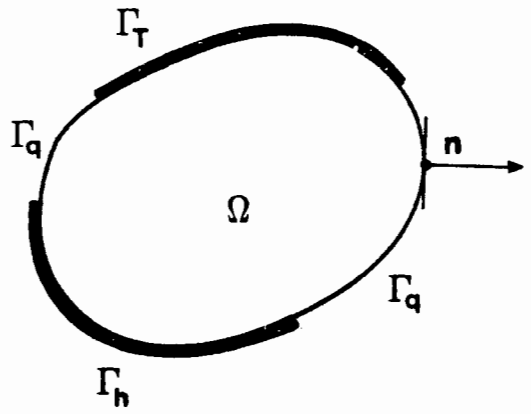
$$\partial T/\partial n = q_o \quad \text{(prescribed heat flux)} \quad \text{on } \Gamma_q , \qquad (2.5)$$

and $\partial T/\partial n + h_T T = r_o$ *(mixed boundary condition) on* Γ_h ,

where Γ_T, Γ_q *and* Γ_h *are disjoint subsets of* Γ *such that*

$\Gamma_T \cup \Gamma_q \cup \Gamma_h = \Gamma$, *and*

h_T *is the heat transfer coefficient called the Nusselt number;*



(ii) velocity boundary conditions:

$$\underline{U} = \underline{U}_o \qquad \text{on} \qquad \Gamma_U \,, \tag{2.6}$$

and

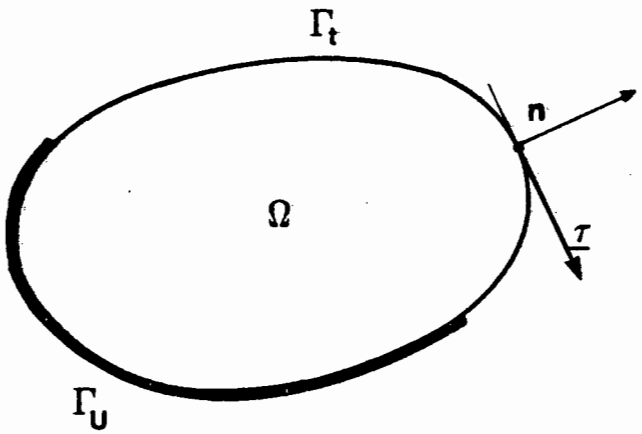
$$\left.\begin{aligned} \underline{U}.\underline{n} &= U_n \\ \underline{\tau}(\underline{U}) &= \underline{\tau}_o \end{aligned}\right\} \quad \text{on} \quad \Gamma_t = \Gamma - \Gamma_U \tag{2.7, 2.8}$$

where $\underline{\tau}(\underline{U})$ *is the tangential surface traction, given by*

$$\underline{\tau} = \underline{T}\underline{n} - (\underline{T}\underline{n}.\underline{n})\underline{n} \,,$$

i.e.,

$$\underline{\tau}(\underline{U}) = 2\mu[\underline{I} - (\underline{n}\otimes\underline{n})]\,\underline{D}[\underline{U}]\underline{n} \,;$$



(f) the initial conditions:

$$\left.\begin{aligned} \underline{U}(\underline{x},0) &= \underline{U}_0(\underline{x}) \\ T(\underline{x},0) &= T_0(\underline{x}) \\ P(\underline{x},0) &= P_0(\underline{x}) \end{aligned}\right\} \quad \text{in } \Omega . \qquad (2.9)$$

To investigate the stability of the basic motion, $[\underline{U}, T, P]$, we disturb the initial condition so that we have an altered motion

$$[\underline{U} + \underline{u},\ T + \theta\ ,\ P + p\ ],$$

satisfying the Oberbeck-Boussinesq equations, corresponding boundary conditions and initial conditions. The equations for the difference motion are formed by subtracting equations (2.1) - (2.9) for the basic flow from those for the disturbed flow, and are

$$\partial \underline{u} / \partial t + (\nabla \underline{u})\underline{U} + (\nabla \underline{U})\underline{u} + (\nabla \underline{u})\underline{u} = -\alpha\theta\underline{g} - {}^{1}/_{\rho_0}\nabla p + 2\nu\, div(\underline{D}[\underline{u}]) \qquad ..(2.10)$$

$$\partial\theta / \partial t + (\nabla\theta)\underline{U} + (\nabla T)\underline{u} + (\nabla\theta)\underline{u} = \kappa\nabla^2\theta \qquad (2.11)$$

$$div\ \underline{u} = 0 \ . \qquad (2.12)$$

The boundary conditions are:

$$\theta = 0 \qquad \text{on } \Gamma_T ,$$

$$\partial\theta / \partial n = 0 \qquad \text{on } \Gamma_q , \qquad (2.13)$$

and

$$\partial\theta / \partial n + h_T\theta = 0 \qquad \text{on } \Gamma_h .$$

On the surface Γ_U *we have*

$$\underline{\boldsymbol{u}} = \underline{\boldsymbol{0}} \tag{2.14}$$

and on the surface $\Gamma - \Gamma_U$

$$\underline{\boldsymbol{u}}.\underline{\boldsymbol{n}} = 0 \tag{2.15}$$

and

$$\underline{D}[\underline{\boldsymbol{u}}]\underline{\boldsymbol{n}} - \underline{\boldsymbol{n}}(\underline{\boldsymbol{n}}.\underline{D}[\underline{\boldsymbol{u}}]\underline{\boldsymbol{n}}) = 0 \quad . \tag{2.16}$$

The initial conditions are:

$$\left.\begin{aligned} \underline{\boldsymbol{u}}(\underline{\boldsymbol{x}},0) &= \underline{\boldsymbol{u}}_0(\underline{\boldsymbol{x}}) \\ \theta(\underline{\boldsymbol{x}},0) &= \theta_0(\underline{\boldsymbol{x}}) \end{aligned}\right\} \quad \text{in } \Omega \ . \tag{2.17}$$

The dimensionless description of the basic motion is obtained by dividing

$$[\ \underline{\boldsymbol{U}},\ \underline{D}[\underline{\boldsymbol{U}}],\ T,\ \nabla T,\ \underline{\boldsymbol{g}},\ \underline{\boldsymbol{x}},\ t\ ]$$

by

$$[U',\ {}^{U'}/_{d},\ T',\ {}^{T'}/_{d},\ g,\ d,\ {}^{d^2}/_{\nu}\ ]$$

where U', T' *and* d *are typical values for* $\underline{\boldsymbol{U}}$, T *and the length of the domain, and by dividing*

$$[\underline{\boldsymbol{u}},\ \underline{D}[\underline{\boldsymbol{u}}],\ \theta\ ,\ p]$$

by

$$\left[\ \nu/d,\quad \nu/d^2,\ \sqrt{\frac{\nu^3 T'}{g\, d^3 \kappa \alpha}}\ ,\ \frac{\rho_o \nu^2}{d^2}\ \right].$$

Since we will be concerned exclusively with the stability of motionless states later, we assume henceforth that the basic state is <u>*motionless*</u>*, so that* $\underline{U} = \underline{0}$ *in equations (2.10) - (2.11). The set of equations (2.10) - (2.17) governing the evolution of the disturbance of the motionless state, in dimensionless variables, are*

$$\partial \underline{u}/\partial t + (\nabla \underline{u})\underline{u} = - R\theta \underline{g} - \nabla p + 2div(\underline{D}[\underline{u}]) \qquad (2.18)$$

$$Pr\ \partial\theta/\partial t + Pr(\nabla\theta)\underline{u} + R\,(\nabla T)\underline{u} = \nabla^2\theta \qquad (2.19)$$

$$div\ \underline{u} = 0 \qquad (2.20)$$

on Ω

where $R = \sqrt{\dfrac{\alpha g\, T'\, d^3}{\nu \kappa}}$ *is the Rayleigh number and*

$Pr = \nu/\kappa$ *is the Prandtl number ,*

satisfying the boundary conditions (2.13) - (2.16) and initial conditions (2.17). The dimensionless variables $\underline{u}$, $\underline{g}$, p, *etc. are denoted by the same notation as their dimensional counterparts.*

2.2 Linear stability theory

In the linear theory only infinitesimal disturbances are considered, so that only linear terms in (2.18) - (2.19) need be retained.

Suppose that the disturbances $\underline{u}$, θ *and* p *have the form*

$$\underline{u}(\underline{x}, t) = \underline{u}(\underline{x}) e^{\sigma t}$$
$$\theta(\underline{x}, t) = \theta(\underline{x}) e^{\sigma t}$$
$$p(\underline{x}, t) = p(\underline{x}) e^{\sigma t}$$

where σ *may be complex.*

The linear equations that govern the infinitesimal disturbance of the motionless state, obtained from equations (2.18), (2.19) and (2.20), are

$$\sigma \underline{u} = - R \theta \underline{g} - \nabla p + 2 div(\underline{D}[\underline{u}]) \qquad (2.21)$$

$$Pr \sigma \theta = - R (\nabla T) \underline{u} + \nabla^2 \theta \qquad (2.22)$$

$$div \underline{u} = 0 . \qquad (2.23)$$

Solutions of these equations must satisfy the boundary conditions (2.13) - (2.16) and initial condition (2.17).

The growth or decay of the disturbance depends on the sign of the real part of σ *. If* $Re(\sigma)$ *is positive, the original flow breaks down and instability sets in. If* $Re(\sigma)$ *is negative the original flow is stable to all infinitesimal disturbances. The flow is marginally or neutrally stable if* $Re(\sigma) = 0$*. States of marginal stability can be one of two kinds: small disturbances*

can grow (or decay) aperiodically; or they can grow (or decay) with oscillations of increasing (or decreasing) amplitude. In the former case, $Im(\sigma) = 0$ and we say that the principle of exchange of stabilities is valid. In the latter case, $Im(\sigma) \neq 0$ and we say that we have a case of overstability. Assume that the principle of exchange of stability holds so that $Im(\sigma) = 0$ when $Re(\sigma) = 0$. Consequently the terms on the left-hand side of equations (2.21) - (2.22) may be deleted. The value of R corresponding to the marginal state is then found from (2.21) - (2.23) with $\sigma = 0$. The linear method determines when the motionless state is unstable to infinitesimal disturbances. This leads to sufficient conditions for instability since flow concidered "stable" by the linear method may be unstable to finite disturbances.

In order to construct finite element approximations (Chapter 3) of the marginal state we will need to construct a variational formulation of (2.21) - (2.23) with $\sigma = 0$. First we define a number of spaces of functions which will be required, starting with the spaces $L_2(\Omega)$ of Lebesque-square integrable functions and the Sobolev space $H^1(\Omega)$, the latter being defined by

$$H^1(\Omega) = \{ \phi \in L_2(\Omega) : \partial\phi/\partial x_i \in L_2(\Omega) \} .$$

Both $L_2(\Omega)$ and $H^1(\Omega)$ are Hilbert spaces with inner products $(.,.)$ and norms $\|.\|$ defined by

$$(\theta, \phi)_{L_2} = \int_\Omega \theta \phi \, dx , \quad \|\theta\|_{L_2} = \sqrt{(\theta, \theta)_{L_2}} ,$$

$$(\theta, \phi)_{H^1} = \int_\Omega (\theta \phi + \nabla\theta . \nabla\phi) \, dx , \quad \|\theta\|_{H^1} = \sqrt{(\theta, \theta)_{H^1}} .$$

We also define the spaces Q and $\underline{V}$ by

$$Q = \{ \phi \in H^1(\Omega) : \phi = 0 \text{ on } \Gamma_T \},$$

$$\underline{V} = \{ \underline{v} = (v_1, v_2) : v_i \in H^1(\Omega), \underline{v} = \underline{0} \text{ on } \Gamma_U ,$$

$$\underline{v}.\underline{n} = 0 \quad \text{on } \Gamma - \Gamma_U \} .$$

Q and $\underline{V}$ are closed subspaces of $H^1(\Omega)$ and $\underline{H}^1 = [H^1(\Omega)]^2$ respectively, and are thus Hilbert spaces with the inner products $(.,.)_L$ and $(.,.)_{\underline{H}}$, where

$$(\underline{u},\underline{v})_{\underline{H}^1} = \int_\Omega (\underline{u}.\underline{v} + \nabla \underline{u}.\nabla \underline{v}) \, dx$$

$$= \int_\Omega (u_i v_i + \partial u_i / \partial x_j \; \partial v_i / \partial x_j \;) \, dx .$$

It is convenient to define the product space $\overline{V}$ by

$$\overline{V} = \underline{V} \times Q ;$$

this is a Hilbert space with inner product

$$(\overline{u},\overline{v})_{\overline{V}} = (\underline{u},\underline{w})_{\underline{H}^1} + (\theta, \phi)_{L_2}$$

where $\overline{u} = (\underline{u}, \theta)$, $\overline{v} = (\underline{v}, \phi)$.

A weak form of the boundary value problem may be obtained by multiplying (2.21) and (2.22) by an arbitrary member $\overline{v} \in \overline{V}$, integrating over the domain Ω and applying Green's Theorem . This yields a pair of equations

$$R \int_\Omega \theta \underline{v}.\underline{g} = \int_\Omega p \; div \; \underline{v} - 2 \int_\Omega \underline{D}[\underline{u}].\underline{D}[\underline{v}] \qquad (2.24)$$

and

$$R \int_\Omega \phi (\nabla T) \underline{u} = - \{ \int_\Omega \nabla\theta . \nabla\phi + \int_{\Gamma_h} h_T \theta \phi \} \qquad (2.25)$$

where the scalar product $\underline{D}.\underline{D} = D_{ij}D_{ij}$.

We eliminate the incompressibility constraint by introducing a small amount of compressibility, i.e. by setting

$$p_\varepsilon = -1/\varepsilon \quad div\ \underline{u}_\varepsilon \tag{2.26}$$

where ε *is a penalty parameter.*

Thus instead of (2.24) and (2.25) we consider the following penalised variational eigenvalue problem *for the linear theory: find* $\bar{u}_\varepsilon \in \bar{V}$, $R_\varepsilon \in \mathbb{R}$ *such that*

$$2\int_\Omega \underline{D}[\underline{u}_\varepsilon].\underline{D}[\underline{v}] + 1/\varepsilon \int_\Omega div\ \underline{u}_\varepsilon . div\ \underline{v} = - R_\varepsilon \int_\Omega \theta_\varepsilon \underline{v}.\underline{g} \tag{2.27}$$

$$\int_\Omega \nabla\theta_\varepsilon . \nabla\phi + \int_{\Gamma_h} h_T \theta_\varepsilon \phi = - R_\varepsilon \int_\Omega \phi\, \underline{u}_\varepsilon . \nabla T \tag{2.28}$$

for all $\bar{v} \in \bar{V}$.

If we define the bilinear forms

$$a:\ \underline{V}\times\underline{V} \to \mathbb{R}\ ,\quad a(\underline{u},\underline{v}) = 2\int_\Omega \underline{D}[\underline{u}].\underline{D}[\underline{v}] + 1/\varepsilon \int_\Omega div\ \underline{u}.div\ \underline{v}\ ,$$

$$b:\ Q\times Q \to \mathbb{R}\ ,\quad b(\theta,\phi) = \int_\Omega \nabla\theta.\nabla\phi + \int_{\Gamma_h} h_T\theta\phi\ ,$$

$$c:\ Q\times\underline{V} \to \mathbb{R}\ ,\quad c(\theta,\underline{v}) = \int_\Omega -\theta\underline{v}.\underline{g}\ , \tag{2.29}$$

$$d:\ \underline{V}\times Q \to \mathbb{R}\ ,\quad d(\underline{u},\phi) = \int_\Omega -\phi\underline{u}.\nabla T$$

and

$$A:\ \bar{V}\times\bar{V} \to \mathbb{R}\ ,\quad A(\bar{u},\bar{v}) = a(\underline{u},\underline{v}) + b(\theta,\phi)\ ,$$

$B: \bar{V} \times \bar{V} \rightarrow \mathbb{R}$, $B(\bar{u}, \bar{v}) = c(\theta, \underline{w}) + d(\underline{u}, \phi)$,

then the resulting variational eigenvalue problem can be expressed in the form: find $\bar{u}_\varepsilon \in \bar{V}$ *and* $R_{L,\varepsilon} \in \mathbb{R}$ *such that*

$$A(\bar{u}_\varepsilon, \bar{v}) = R_{L,\varepsilon}\, B(\bar{u}_\varepsilon, \bar{v}) \quad \text{for all } \bar{v} \in \bar{V} \qquad (2.30)$$

where the lowest eigenvalue, $R_{L,\varepsilon}$, *obtained from the eigenvalue problem corresponds to the critical stability number obtained by the linear stability theory.*

The advantage of the above formulation is that the pressure is eliminated as an unknown; furthermore, it is expected that $p_\varepsilon \rightarrow p$, $\bar{u}_\varepsilon \rightarrow \bar{u}$ *and* $R_{L,\varepsilon} \rightarrow R_L$ *as* $\varepsilon \rightarrow 0$. *Proof of convergence, however, is beyond the scope of this work (see Geveci, Reddy and Pearce (1986) for a treatment of convergence in the case of the penalised eigenvalue problem for the Stokes operator).*

If A(.,.) and B(.,.) are symmetric, the eigenvalues associated with A and B are all real. In the linear stability theory, however, B(.,.) is in general not symmetric. In this case, necessary and sufficient conditions for eigenvalues of the variational eigenvalue problem (2.30) to be real are not known, though our numerical studies show that real eigenvalues exist.

2.3 Energy stability theory

An alternative method for assessing the stability of a fluid is the energy method; a fluid is called stable if the energy of any disturbance, no matter how large, of the given motion decays. In order to obtain critical energy stability parameters below which the hydrodynamical system is stable, we introduce the kinetic energy

$$K = \int_{\Omega} {}^{1}/_{2} \, |\underline{u}|^2 \tag{2.31}$$

and the "thermal energy"

$$\Theta = \int_{\Omega} {}^{1}/_{2} \, \theta^2 \tag{2.32}$$

of the disturbance.

If K and $\Theta \rightarrow 0$ as $t \rightarrow \infty$, we say the basic motion is stable. Let $\bar{u} = (\underline{u}, \theta)$ be a solution of the OB equations (2.18) -(2.19); then it is easily shown [Joseph (1976 I)] that the rate of change of K and Θ are given by

$$\begin{aligned} {}^{dK}/_{dt} &= \int_{\Omega} \underline{u} \cdot ({}^{\partial \underline{u}}/_{\partial t} + (\nabla \underline{u}) \underline{u}) \\ &= - \int_{\Omega} R \, \theta \, \underline{u} \cdot \underline{g} + 2 \underline{D}[\underline{u}] \cdot \underline{D}[\underline{u}] \end{aligned} \tag{2.33}$$

and

$$\begin{aligned} {}^{d\Theta}/_{dt} &= \int_{\Omega} \theta ({}^{\partial \theta}/_{\partial t} + (\nabla \theta) \underline{u}) \\ &= - {}^{1}/_{Pr} \left\{ \int_{\Omega} (\nabla \theta \cdot \nabla \theta + R \, \theta \, \underline{u} \cdot \nabla T) + \int_{\Gamma_h} h_T \, \theta^2 \right\} . \end{aligned} \tag{2.34}$$

We are now able to define an "energy" E_λ of the system which is a linear combination of the kinetic energy, K, and the "thermal energy", Θ :

$$E_\lambda(\bar{v}) = {}^1/_2 \int_\Omega \underline{v}.\underline{w} + \lambda Pr\ \phi^2 \quad . \tag{2.35}$$

Here, $\lambda > 0$ is a coupling parameter.

Using equations (2.33) and (2.34) we define the bilinear forms

$$J_1 : \underline{V} \times \underline{V} \to \mathbb{R} \ , \qquad J_1(\underline{u},\underline{v}) = 2 \int_\Omega \underline{D}[\underline{u}].\underline{D}[\underline{v}] \ ,$$

$$J_2 : Q \times Q \to \mathbb{R} \ , \qquad J_2(\theta,\phi) = \int_\Omega \nabla\theta.\nabla\phi + \int_{\Gamma_h} h_T \theta\phi \ ,$$

$$J : \bar{V} \times \bar{V} \to \mathbb{R} \ , \qquad J(\bar{u},\bar{v}) = J_1(\underline{u},\underline{v}) + \lambda J_2(\theta,\phi) \ ,$$

$$I_1 : \bar{V} \times \bar{V} \to \mathbb{R} \ , \qquad I_1(\bar{u},\bar{v}) = - {}^1/_2 \int_\Omega (\underline{u}\phi + \theta\underline{v}).\underline{g} \ ,$$

$$I_2 : \bar{V} \times \bar{V} \to \mathbb{R} \ , \qquad I_2(\bar{u},\bar{v}) = - {}^1/_2 \int_\Omega (\underline{u}\phi + \theta\underline{v}).\nabla T \ ,$$

$$I : \bar{V} \times \bar{V} \to \mathbb{R} \ , \qquad I(\bar{u},\bar{v}) = I_1(\bar{u},\bar{v}) + \lambda I_2(\bar{u},\bar{v}) \ ,$$

where λ is a given constant, $0 < \lambda < \infty$.

If $\bar{u}$ is a solution of equations (2.18) - (2.20) then it can be shown [Joseph (1976 II)] that the rate of change of energy can be written as

$$dE_\lambda(\bar{u})/dt = -J(\bar{u},\bar{u}) + R\, I(\bar{u},\bar{u}) \tag{2.36}$$

$$= J(\bar{u},\bar{u}) \left\{ -1 + R \left(I(\bar{u},\bar{u})/J(\bar{u},\bar{u}) \right) \right\}$$

$$\leq J(\bar{u},\bar{u}) \left\{ -1 + R/\rho_\lambda \right\} \tag{2.37}$$

where $\rho_\lambda^{-1} = \sup\left(I(\bar{v},\bar{v}) / J(\bar{v},\bar{v}) \right)$, *the supremum being taken over all functions in* $\bar{V}$ *satisfying div* $\underline{v} = 0$.

We are now in a position to state the basic energy stability theorem. First, we define the spaces $\underline{W}$ *and* $\bar{W}$ *by*

$$\underline{W} = \{ \underline{v} \in \underline{V} : \ div\ \underline{v} = 0 \},$$

$$\bar{W} = \underline{W} \times Q \quad .$$

Both of these are Hilbert spaces with inner products $(.,.)_{\underline{H}^1}$ *and* $(.,.)_{\bar{V}}$, *respectively.*

Theorem 2.1. Suppose there exists a constant $\alpha > 0$ *such that*

$$J(\bar{v},\bar{v}) \geq \alpha \| \bar{v} \|^2_{\bar{V}} \qquad \text{for all } \bar{v} \in \bar{V}.$$

Suppose further that there are constants $\lambda_1, \lambda_2 > 0$ *and* Pr_1, $Pr_2 > 0$ *such that* $0 < \lambda_1 \leq \lambda \leq \lambda_2$ *and* $0 < Pr_1 \leq Pr \leq Pr_2$. *Then there is a constant* $\gamma > 0$ *such that*

$$E_\lambda(\bar{u}(t)) \leq E_\lambda(\bar{u}(0)) \exp\left\{ -\int_0^t \left(1 - R/\rho_\lambda(s)\right) \gamma\, ds \right\},$$

provided $R < \rho_\lambda(t)$ *in the time interval* $[0,t]$, *where* $E_\lambda(\bar{u}(0))$ *is the initial energy of the difference motion. If* $R < \rho_\lambda(t)$ *for all* t *then* $E_\lambda(\bar{u}(t)) \to 0$ *as* $t \to \infty$ *and the flow is asymptotically stable.*

Proof. We have

$$E_\lambda(\bar{u}(t)) = {}^1/_2 \int_\Omega \underline{u}.\underline{u} + \lambda Pr\, \theta^2$$

$$\leq {}^1/_2\, C \int_\Omega \underline{u}.\underline{u} + \theta^2$$

$(C = max(1, \lambda_2 Pr_2))$

$$= {}^1/_2\, C \left(\sum_i \|u_i\|^2_{L_2} + \|\theta\|^2_{L_2} \right)$$

$$\leq {}^1/_2\, C \|\bar{u}\|^2_{\bar{V}} \leq {}^C/_{2\alpha}\, J(\bar{u},\bar{u}) \ .$$

Hence from (2.37)

$$dE_\lambda(\bar{u}(t))/_{dt} \leq -\left(1 - {}^R/_{\rho_\lambda}\right) J(\bar{u},\bar{u})$$

$$\leq -\left(1 - {}^R/_{\rho_\lambda}\right) {}^{2\alpha}/_C\, E_\lambda(\bar{u}(t)) \ .$$

By integrating over the time interval [0,t] we obtain

$$E_\lambda(\bar{u}(t)) \leq E_\lambda(\bar{u}(0))\ exp\left\{ -\int_0^t \left(1 - {}^R/_{\rho_\lambda(s)}\right) \gamma\, ds \right\} \qquad (2.38)$$

where $\gamma = {}^{2\alpha}/_C$.

The maximization problem

$${}^1/_{\rho_\lambda} = \sup_{\bar{v} \in \bar{W}} \left[I(\bar{v},\bar{v}) / J(\bar{v},\bar{v}) \right] \qquad (2.39)$$

is equivalent to the problem

$${}^1/_{\rho_\lambda} = \sup_{\bar{v} \in \bar{W}} \left\{ I(\bar{v},\bar{v}) \right\} \qquad (2.40)$$

subject to

$$J(\bar{v},\bar{v}) = 1. \qquad (2.41)$$

The energy studies clearly depend on the existence of the maximum of $I(\bar{v})/J(\bar{v})$. *This is equivalent to showing the existence of* $\bar{u} \in S$ *such that*

$$I(\bar{u}) \geq I(\bar{v}) \quad \text{for all } \bar{v} \in S$$

where $S = \{\bar{v} \in \bar{W} : J(\bar{v},\bar{v}) = 1\}$,

$I(\bar{u}) = I(\bar{u},\bar{u})$ *and*

$I(\bar{v}) = I(\bar{v},\bar{v})$.

Set

$$I(\bar{u},\bar{v}) = {}^1/_2 \int_\Omega \underline{a}.(\underline{u}\,\phi + \underline{v}\,\theta)$$

where $\underline{a} = -(\underline{g} + \lambda \nabla T)$ *and*

$\underline{a} \in [\, L_\infty(\Omega)\,]^2$.

We start the existence proof with the following lemmas.

<u>*Lemma 2.1*</u> *Suppose that there exists a constant* a_2 *such that*

$$\max_i \|a_i\|_{L_\infty} \leq a_2$$

and constants λ_1 *and* λ_2 *such that*

$$0 < \lambda_1 \leq \lambda \leq \lambda_2 < \infty .$$

then

$$I(\bar{u},\bar{v}) \leq C \|\bar{u}\|_{\bar{L}_2} \|\bar{v}\|_{\bar{L}_2} \leq C \|\bar{u}\|_{\bar{V}} \|\bar{v}\|_{\bar{V}} \qquad (2.42)$$

Proof.

$$I(\bar{u},\bar{v}) = \tfrac{1}{2}\int_\Omega a_i\, v_i\theta + \tfrac{1}{2}\int_\Omega a_i\, u_i\phi \ .$$

where

$$\int_\Omega a_i\, v_i\theta \le \Big|\int_\Omega a_i\, v_i\theta\Big|$$

$$\le a_2\,\big|(\theta,\textstyle\sum_i v_i)_{L_2}\big| \le a_2\textstyle\sum_i |(\theta,v_i)_{L_2}|$$

$$\le a_2\,\|\theta\|_{L_2}\textstyle\sum_i\|v_i\|_{L_2} \le 2a_2\,\|\theta\|_{L_2}\|\underline{v}\|_{\underline{L}_2}$$

$$\le 2a_2\,\|\theta\|_{L_2}\|\underline{v}\|_{\underline{L}_2} \le 2a_2\,\|\bar{u}\|_{\overline{L}_2}\|\bar{v}\|_{\overline{L}_2}$$

$$\le 2a_2\,\|\bar{u}\|_{\overline{V}}\,\|\bar{v}\|_{\overline{V}}$$

and, similarly

$$\int_\Omega a_i\, u_i\phi \le 2a_2\,\|\bar{u}\|_{\overline{V}}\,\|\bar{v}\|_{\overline{V}} \ .$$

Set $C = 2a_2$ *and add to get result.*

Lemma 2.2 *Assume that* $h_T \in L_\infty(\Gamma_h)$ *and that* $h_T(\underline{x}) \ge 0$, $\|h_T\|_{L_\infty(\Gamma_h)} \le h_2$. *Then there exists constants* α *and* K *such that*

$$J(\bar{v}) \ge \alpha\,\|\bar{v}\|^2_{\overline{V}} \qquad (2.43)$$

and

$$J(\bar{u},\bar{v}) \le K\,\|\bar{u}\|_{\overline{V}}\,\|\bar{v}\|_{\overline{V}} \qquad (2.44)$$

where $J(\bar{u},\bar{v}) = 2\int_\Omega \underline{D}[\underline{u}].\underline{D}[\underline{v}] + \lambda\int_\Omega \nabla\theta.\nabla\phi + \lambda\int_{\Gamma_h} h_T\,\theta\,\phi$.

Proof. *We have*

$$J(\bar{v}) = 2 \int_{\Omega} \underline{D}[\underline{v}] . \underline{D}[\underline{v}] + \lambda \int |\nabla \phi|^2 + \lambda \int_{\Gamma_h} h_T \phi^2$$

$$\geq k \| \underline{v} \|^2_{\underline{H}^1} + \lambda_1 \int |\nabla \phi|^2 + \lambda_1 \int_{\Gamma_h} h_T \phi^2$$

(using Korn's inequality for the first term [Duvaut and Lions (1976)])

$$\geq k \| \underline{v} \|^2_{\bar{V}} + \lambda_1 \| \phi \|^2_Q$$

$$\geq \alpha \| \underline{v} \|^2_{\bar{V}}$$

where $\alpha = min(k, \lambda_1)$.

The proof of (2.44) is similar to that of (2.42). We have

$$J(\bar{u}, \bar{v}) = \int_{\Omega} (u_{i,j} v_{i,j} + u_{j,i} v_{i,j}) + \lambda \int_{\Omega} \theta_{,i} \phi_{,i} + \lambda \int_{\Gamma_h} h_T \theta \phi$$

$$\leq (u_{i,j}, v_{i,j})_{L_2} + (u_{j,i}, v_{i,j})_{L_2} + \lambda_2 (\theta_{,i}, \phi_{,i})_{L_2} + \lambda_2 h_2 (\theta, \phi)_{L_2(\Gamma)}$$

$$\leq \| u_{i,j} \|_{L_2} \| v_{i,j} \|_{L_2} + \| u_{j,i} \|_{L_2} \| v_{i,j} \|_{L_2} + \lambda_2 \| \theta_{,i} \|_{L_2} \| \phi_{,i} \|_{L_2}$$

$$+ \lambda_2 h_2 \| \theta \|_{L_2(\Gamma)} \| \phi \|_{L_2(\Gamma)}$$

(we extend h by zero to all of Γ *)*

$$\leq 4 \| \underline{u} \|_{\underline{H}^1} \| \underline{v} \|_{\underline{H}^1} + 2 \lambda_2 \| \theta \|_{H^1} \| \phi \|_{H^1} + \lambda_2 h_2 C \| \theta \|_{H^1} \| \phi \|_{H^1}$$

(using the trace theorem in the last term and assuming that the problem is two-dimensional)

$$\leq K \| \bar{u} \|_{\bar{V}} \| \bar{v} \|_{\bar{V}} .$$

On the basis of Lemma 2.1 and 2.2 we have the following existence

result [Galdi and Straughan (1985)], [Rionero(1968)].

Theorem 2.2 *Assume that the conditions of Lemmas 2.1 and 2.2 and Theorem 2.1 hold. Then there exists* $\bar{u} \in S$ *such that*

$$\sup_{\bar{v} \in W} \frac{I(\bar{v},\bar{v})}{J(\bar{v},\bar{v})} = \frac{I(\bar{u},\bar{u})}{J(\bar{u},\bar{u})} = 1/\rho(\lambda) < \infty \ .$$

If $R < \rho(\lambda)$ *for all time* t *then* $E_\lambda(\bar{u}(t)) \to 0$ *as* $t \to \infty$ *and the flow is asymptotically stable.*

The constrained maximisation problem (2.39) is reduced to an unconstrained problem by introducing the Lagrange multipliers ρ_λ *and* p, *and defining the Lagrangian* L *by*

$$L: \bar{V} \times \mathbb{R} \times L_2 \to \mathbb{R} \ ,$$

$$L(\bar{v},m,q) = I(\bar{v},\bar{v}) - m(J(\bar{v},\bar{v}) - 1) + \int_\Omega q \ div \ \underline{v}$$

(see Oden and Kikuchi (1982)).

Then a necessary and sufficient condition for $I(\bar{v},\bar{v})$ *to be a maximum is that the first variation of* L *vanishes, i.e.*

$$\langle DL(\bar{u}, \rho_\lambda, p), (\bar{v}, m, q)\rangle = \lim_{h \to 0} {}^{d}/_{dh} \ L(\bar{u} + h\bar{v}, \rho_\lambda + hm, p + hq) = 0$$

This leads to the variational eigenvalue problem of finding $\bar{u} \in \bar{V}$, $p \in L_2(\Omega)$ *and* $\rho_\lambda \in \mathbb{R}$ *such that*

$$I(\bar{u},\bar{v}) - {}^{1}/_{\rho_\lambda} \ J(\bar{u},\bar{v}) + \int_\Omega p \ div \ \underline{v} = 0 \quad , \qquad (2.45)$$

$$\int_\Omega q \ div \ \underline{u} = \underline{0} \ ,$$

$$J(\bar{u},\bar{u}) - 1 = 0,$$

for all $\bar{v} \in \bar{V}$, $q \in L_2(\Omega)$.

As in Section 2.2, we approximate the pressure p in (2.45) by

$$p_\varepsilon = -1/\varepsilon \ div\ \underline{u}_\varepsilon$$

where $\varepsilon > 0$ *is the penalty parameter.*

Substituting the above into equation (2.45) and expanding, we obtain the variational eigenvalue problem for the energy stability theory: find $\bar{u}_\varepsilon \in \bar{V}$ *and* $\rho_{\lambda,\varepsilon} \in \mathbb{R}$ *such that*

$$2\int_\Omega \underline{D}[\underline{u}_\varepsilon].\underline{D}[\underline{v}] + 1/\varepsilon \int_\Omega div\underline{u}_\varepsilon . div\underline{v} = -\rho_{\lambda,\varepsilon}\int_\Omega 1/2\ \theta_\varepsilon \underline{v}(\underline{g} + \lambda\nabla T) \quad (2.46)$$

and

$$\lambda\int_\Omega \nabla\theta_\varepsilon.\nabla\phi + \lambda\int_{\Gamma_h} h_T\theta_\varepsilon\phi = -\rho_{\lambda,\varepsilon}\int_\Omega 1/2\ \phi\,\underline{u}_\varepsilon.(\underline{g} + \lambda\nabla T) \quad (2.47)$$

for all $\bar{v} \in \bar{V}$.

In addition to the bilinear forms (2.29) defined in Section 2.2, we define

$$e: \underline{V}\times Q \to \mathbb{R}\ ,\ e(\underline{u},\phi) = \int_\Omega -1/2\ \phi\,\underline{u}.(\underline{g} + \lambda\nabla T)\ ,$$

$$A': \bar{V}\times\bar{V} \to \mathbb{R}\ ,\ A'(\bar{u},\bar{v}) = a(\underline{u},\underline{v}) + \lambda b(\theta,\phi)\ ,$$

and

$$B': \bar{V}\times\bar{V} \to \mathbb{R}\ ,\ B'(\bar{u},\bar{v}) = e(\underline{u},\phi) + e(\theta,\underline{v}).$$

The eigenproblem can now be written as follows: for a given $\lambda > 0$, *find* $\bar{u}_\varepsilon \in \bar{V}$ *and* $\rho_{\lambda,\varepsilon} \in \mathbb{R}$ *such that*

$$A'(\bar{u}_\varepsilon, \bar{v}) = \rho_{\lambda\varepsilon} B'(\bar{u}_\varepsilon, \bar{v}) \quad \text{for all } \bar{v} \in \bar{V} \qquad (2.48)$$

where $\rho_{\lambda\varepsilon}$ corresponds to the lowest eivenvalue. As in Section 2.2, we expect that the solution to the penalised problem (2.48) converges to the solution of (2.45) as $\varepsilon \to 0$, i.e., that $\bar{u}_\varepsilon \to \bar{u}$, $p_\varepsilon \to p$ and $\rho_{\lambda\varepsilon} \to \rho_\lambda$ as $\varepsilon \to 0$ [Geveci, Reddy and Pearce (1986)].

Since both $A'(.,.)$ and $B'(.,.)$ are symmetric the eigenvalues of (2.48) are all real. The left-hand side of (2.48) is equivalent to the left-hand side of the eigenvalue problem (2.30) defined in Section 2.2 if we set the coupling parameter λ equal to one. Moreover, when $\underline{g} = \nabla T$ (the Bénard problem), $B(\bar{u}, \bar{v}) = B'(\bar{u}, \bar{v})$ and the two eigenvalue problems coincide.

The energy theory gives sufficient conditions for stability, while the linear theory gives sufficient conditions for instability. When the linear and energy stability limits coincide, all stable disturbances, whatever their size, will decay exponentially from the initial instant. In this situation subcritical instabilities cannot occur. Thus, when $R_L = \rho_\lambda$ we have a <u>neccesary and sufficient condition for stability</u>.

Generally we may select λ to give the best possible limit for stability. This is equivalent to the problem of finding R_E such that

$$R_E = \max_{\lambda>0} [\rho_\lambda] = \max_{\lambda>0} [\min_{\bar{v} \in \bar{V}} J(\bar{v}, \bar{v}) / I(\bar{v}, \bar{v})]. \qquad (2.49)$$

Each choice of coupling constant gives a different energy and leads to a different stability number ρ_λ. The largest value of ρ_λ over λ leads to the best possible limit for stability, R_E, as illustrated in Figure 2.2.

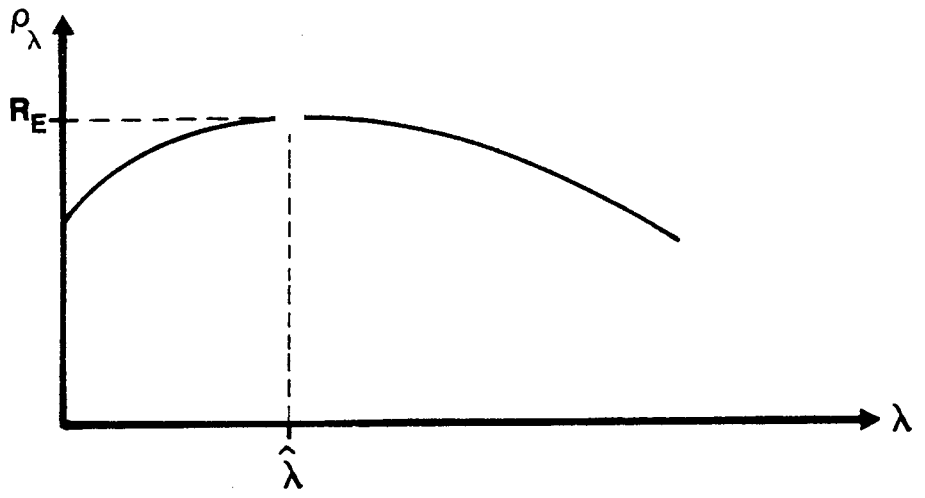


Figure 2.2: The optimum stability limit R_E.

We conclude this Chapter with an examination of the conditions under which a unique value of R_E exists. Our procedure follows that of Shir and Joseph (1968), who investigated this problem for the case when temperature and concentration gradients are present. First, we seek the λ at which ρ_λ is a maximum. From Figure 2.2 we see that the best λ must appear as a root of the equation $d\rho_\lambda / d\lambda = 0$. Second, we need to show that there exists one and only one value λ satisfying the above equation, i.e. that there is only one such a maximum.

Define

$$\begin{bmatrix} \underline{u} \\ \phi \end{bmatrix} = \tilde{u} \in \bar{W} \quad , \quad \begin{bmatrix} \underline{v} \\ \psi \end{bmatrix} = \tilde{v} \in \bar{W} \quad , \quad \phi = \sqrt{\lambda}\,\theta$$

and set $\underline{G} = {}^1/_{2\sqrt{\lambda}} \; (\underline{q} + \lambda \nabla T)$;

note that $\tilde{u}$ and $\underline{G}$ are functions of λ.

The Euler-Lagrange equations corresponding to (2.46) and (2.47)

using the above notation are

$$\rho_\lambda \underline{G}\phi = 2 \; div \; \underline{D}[\underline{u}] \qquad (2.50)$$

and

$$\rho_\lambda \underline{G}.\underline{u} = \nabla^2\phi \qquad . \qquad (2.51)$$

If we introduce the symmetric matrix

$$\hat{\underline{A}} = \begin{bmatrix} 0 & \underline{G} \\ \underline{G} & 0 \end{bmatrix}$$

and the operator $\underline{L}$ defined by

$$\underline{L}\tilde{u} = \begin{bmatrix} 2 \; div \; \underline{D}[\underline{u}] \\ \nabla^2\phi \end{bmatrix} \qquad ,$$

then equations (2.50) and (2.51) can be written as

$$\rho_\lambda \hat{\underline{A}}\tilde{u} = \underline{L}\tilde{u} \qquad on \; \Omega \qquad (2.52)$$

or

$$\underline{M}\tilde{u} = (\underline{L} - \underline{A})\tilde{u} = 0 \qquad on \; \Omega \qquad (2.53)$$

where $\underline{A} = \rho_\lambda \hat{\underline{A}}$.

Define an inner product $(.,.)$ on $\bar{W}$ by

$$(\tilde{u}, \tilde{v}) = \int_\Omega \underline{u}.\underline{v} + \phi\psi \quad .$$

Then

$$(\tilde{u}, \hat{\underline{A}}\tilde{v}) = \int_\Omega \underline{G}.(\underline{u}\psi + \underline{v}\phi)$$

$$= I(\tilde{u}, \tilde{v})$$

and

$$(\tilde{u}, \underline{L}\tilde{v}) = \int_{\Omega} \underline{u}.2\ div\ \underline{D}[\underline{v}] + \phi.\nabla^2\psi$$

$$= - J(\tilde{u}, \tilde{v})$$

for $\tilde{u}, \tilde{v} \in \overline{W}$, *assuming* $\tilde{u}$ *to be the solution of (2.52), (2.53).*

Using the above results, the maximum problem (2.39) may be written as

$$1/\rho_\lambda = \max_{\tilde{v}} - \frac{(\tilde{v}, \underline{\hat{A}}\tilde{v})}{J(\tilde{v}, \tilde{v})} \geq - \frac{(\tilde{v}, \underline{\hat{A}}\tilde{v})}{J(\tilde{v}, \tilde{v})} \quad . \qquad (2.54)$$

We assume that $\rho_\lambda \geq 0$ *and set*

$$H(\tilde{u}, \tilde{v}) = J(\tilde{u}, \tilde{v}) + \rho_\lambda (\tilde{u}, \underline{\hat{A}}\tilde{v}) \ ;$$

then $$H(\tilde{v}, \tilde{v}) \geq 0 \ .$$

Henceforth, $\tilde{u} = (\underline{u}, \phi)$ *denotes the actual solution and* $\tilde{v}$ *is arbitrary.*

Differentiating (2.53) n times with respect to λ, *we get*

$$\underline{M}\ d^n\tilde{u}/d\lambda^n + \sum_{k=1}^{n} \binom{n}{k}\ d^k\underline{M}/d\lambda^k\ \ d^{n-k}\tilde{u}/d\lambda^{n-k} = 0 \ . \qquad (2.55)$$

Next, multiplying (2.53) by $d^n\tilde{u}/d\lambda^n$ *and integrating over* Ω, *we obtain*

$$(d^n\tilde{u}/d\lambda^n\ , \underline{M}\tilde{u}) = 0 \ .$$

Because $\underline{M}$ *is self-adjoint we have*

$$(\tilde{u}, \underline{M}\, d^n\tilde{u}/d\lambda^n) = 0 \quad . \tag{2.56}$$

From (2.53) we have

$$d^n\underline{M}/d\lambda^n = -d^n\underline{A}/d\lambda^n \tag{2.57}$$

because $\underline{L}$ *does not depend on* λ .

Putting n=1 in equations (2.55) and (2.56) it follows that

$$\underline{M}\, d\tilde{u}/d\lambda + d\underline{M}/d\lambda\, \tilde{u} = 0$$

and

$$(\tilde{u}, \underline{M}\, d\tilde{u}/d\lambda) = 0 \quad .$$

Substitute for $\underline{M}\, d\tilde{u}/d\lambda$ *in the above equation to get*

$$(\tilde{u}, d\underline{M}/d\lambda\, \tilde{u}) = 0$$

and using (2.57) with n=1 we obtain

$$(\tilde{u}, d\underline{A}/d\lambda\, \tilde{u}) = 0 \tag{2.58}$$

where

$$d\underline{A}/d\lambda = \begin{bmatrix} 0 & \rho_\lambda\, d\underline{G}/d\lambda + \underline{G}\, d\rho_\lambda/d\lambda \\ \rho_\lambda\, d\underline{G}/d\lambda + \underline{G}\, d\rho_\lambda/d\lambda & 0 \end{bmatrix} . \tag{2.59}$$

The best λ *which solves (2.49) must appear at the root of the equation*

$$d\rho_\lambda/d\lambda = 0 \ . \qquad (2.60)$$

From (2.59) and (2.60) we get

$$d\underline{\underline{A}}/d\lambda = \rho_\lambda \begin{bmatrix} 0 & d\underline{G}/d\lambda \\ d\underline{G}/d\lambda & 0 \end{bmatrix} .$$

Substituting the above into (2.58) we get

$$0 = \int_\Omega 2\,\rho_\lambda \ d\underline{G}/d\lambda \ . \underline{u}\phi$$

$$= \int_\Omega -\,1/4 \ \lambda^{-3/2} \ \rho_\lambda \ (\underline{g} - \lambda \ \nabla T) . \underline{u}\phi .$$

Thus, the value of λ *leading to the optimum stability limit satisfies*

$$\hat{\lambda} = \frac{\int \underline{g} . \underline{u}\phi}{\int \nabla T . \underline{u}\phi} \qquad (2.61)$$

where $\underline{u}$ *and* ϕ *are the velocity and temperature corresponding to this optimal value.*

We need to show that there exists only one such value satisfying equation (2.60). To establish that the stationary point is a global maximum it is necessary to show that

$$d^2\rho_\lambda/d\lambda^2 \ < 0 \qquad (2.62)$$

at any point where $d\,\rho_\lambda/d\lambda = 0$, *so that a minimum cannot exist.*

Put n=2 in equation (2.55) to get

$$\underline{M}\, d^2\tilde{u}/d\lambda^2 + 2\, d\underline{M}/d\lambda\; d\tilde{u}/d\lambda + d^2\underline{M}/d\lambda^2\; \tilde{u} = 0 \ .$$

Multiply by $\tilde{u}$ *and integrate over* Ω *to obtain*

$$(\tilde{u},\ \underline{M}\, d^2\tilde{u}/d\lambda^2\ ) + 2(\tilde{u},\ d\underline{M}/d\lambda\ d\tilde{u}/d\lambda\ ) - (\tilde{u},\ d^2\underline{A}/d\lambda^2\ \tilde{u}) = 0 \quad (2.63)$$

using the result obtained in (2.57).

From (2.55) and setting n=1 we have

$$(\ d\tilde{u}/d\lambda\ ,\ \underline{M}\, d\tilde{u}/d\lambda\ ) + (\ d\tilde{u}/d\lambda\ ,\ d\underline{M}/d\lambda\ \tilde{u}) = 0 \ .$$

Since $d\underline{M}/d\lambda$ *is symmetric*

$$\begin{aligned}(\tilde{u},\ d\underline{M}/d\lambda\ d\tilde{u}/d\lambda\ ) &= (\ d\tilde{u}/d\lambda,\ d\underline{M}/d\lambda\ \tilde{u}) \\ &= (\ d\tilde{u}/d\lambda\ ,\ \underline{M}\, d\tilde{u}/d\lambda\ ) \\ &= H(\ d\tilde{u}/d\lambda\ ,\ d\tilde{u}/d\lambda\ ) \qquad (2.64)\end{aligned}$$

Substituting (2.64) into (2.63) and using (2.56) with n=2, we get

$$(\tilde{u},\ d^2\underline{A}/d\lambda^2\ \tilde{u}) = 2\, H(\ d\tilde{u}/d\lambda\ ,\ d\tilde{u}/d\lambda\ ) \qquad (2.65)$$

where $d^2\underline{A}/d\lambda^2 = d^2\rho_\lambda/d\lambda^2\ \hat{\underline{A}} + 2\, d\rho_\lambda/d\lambda\ d\hat{\underline{A}}/d\lambda + \rho_\lambda\, d^2\hat{\underline{A}}/d\lambda^2 \ .$

Then with $\rho_\lambda(\tilde{u},\ \hat{\underline{A}}\tilde{u}) = -1$, *one finds that*

$$\begin{aligned}d^2\rho_\lambda/d\lambda^2 = -\rho_\lambda \{\ &2H(d\tilde{u}/d\lambda\ ,\ d\tilde{u}/\ d\lambda\ ) - 2\, d\rho_\lambda/d\lambda\ (\tilde{u},\ d\hat{\underline{A}}/d\lambda\ \tilde{u}) \\ &- \rho_\lambda(\tilde{u},\ d^2\hat{\underline{A}}/d\lambda^2\ \tilde{u})\ \} \ . \qquad (2.66)\end{aligned}$$

At a stationary point $d\rho_\lambda / d\lambda = 0$ *and*

$$d^2\hat{\underline{A}}/d\lambda^2 = \begin{bmatrix} 0 & d^2\underline{G}/d\lambda^2 \\ d^2\underline{G}/d\lambda^2 & 0 \end{bmatrix}$$

where $d^2\underline{G}/d\lambda^2 = {}^3/_8 \lambda^{-5/2} \underline{g} - {}^1/_8 \lambda^{-3/2} \nabla T$.

Substitute in (2.66) to get

$$d^2\rho_\lambda/d\lambda^2 = -\rho_\lambda \{ 2H(d\tilde{u}/d\lambda , d\tilde{u}/d\lambda) + J_1(\phi, \phi)/2\lambda^2 \} .$$

Since we assume ρ_λ *is positive,* $H(d\tilde{u}/d\lambda , d\tilde{u}/d\lambda)$ *and* $J_1(\phi, \phi)$ *are positive definite, and (2.62) is proven, i.e.* ρ_λ *is a local maximum at every stationary point. It follows that* ρ_λ *is a concave function of* λ *so that the local maximum is unique, and is a global maximum for* $\lambda > 0$. *We summarise these results in the following theorem.*

Theorem 2.3. Assume that $\rho_\lambda > 0$ *and* $\hat{\lambda} > 0$, *where* $\hat{\lambda}$ *is the best value of the coupling parameter.*

Then (i) the best value satisfies the relationship

$$\hat{\lambda} = \int \underline{g}.\underline{u}\,\phi \Big/ \int \nabla T.\underline{u}\,\phi \qquad (2.67)$$

where $\tilde{u} = \begin{bmatrix} \underline{u} \\ \phi \end{bmatrix}$ *is the maximizing function for (2.54) ;*

(ii) there exists exactly one optimal value R_E *of the*

stability number, where

$$R_E = \rho_{\hat{\lambda}} \geq \rho_\lambda \qquad \text{for all } \lambda > 0.$$

(iii) Provided that $R < R_E$, the basic solution ($\underline{U}$, T, P) will be globally monotonically stable.

Proof that ρ_λ is positive and that ρ_λ attains a maximum for a positive value of λ is beyond the scope of this work and is therefore omitted

In Chapter 2 we formulated the penalised variational eigenvalue problems for the linear stability theory (2.30) and the energy stability theory (2.48) in the form: find a function $\bar{u} \in \bar{V}$, $R \in \mathbb{R}$ such that

$$A(\bar{u},\bar{u}') = R\, B(\bar{u},\bar{u}') \qquad \text{for all } \bar{u}' \in \bar{V} \qquad (3.1)$$

where $\bar{u} = (u, v, \theta)$ and

$\bar{u}' = (u', v', \theta')$.

Here and henceforth the subscripts ε are omitted for clarity.

It is difficult to solve (3.1) since the space of admissable functions $\bar{V}$ is infinite-dimensional: we have

$$\bar{V} = \text{span}\{\phi_i\}_{i=1}^{\infty}$$

where $\{\phi_i, \ i=1, \ \ldots, \infty\}$ is a basis for $\bar{V}$.

The Galerkin method is a method for constructing an approximate solution to the variational eigenvalue problem (3.1) in a finite-dimensional subspace of $\bar{V}$ of admissible functions, rather than in the whole space. Instead of posing the problem in $\bar{V}$ we define a space $\bar{V}^h$ to be a finite-dimensional subspace of $\bar{V}$ spanned by a finite number of linearly independent functions ϕ_i, $i=1, \ \ldots, \ N$, i.e.

$$\bar{V}^h \in \bar{V}, \quad \text{span}\{\phi_i\}_{i=1}^{N} = \bar{V}^h \ .$$

The index h is a parameter that lies between 0 and 1, and is a measure of how close $\bar{V}^h$ is to $\bar{V}$ (e.g. $h = 1/\dim \bar{V}^h$). Having defined the space $\bar{V}^h$ the variational eigenvalue problem now

becomes: find $\bar{u}^h \in \bar{V}^h$, $R^h \in \mathbb{R}$ *that satisfies*

$$A(\bar{u}^h, \bar{u}'^h) = R^h\, B(\bar{u}^h, \bar{u}'^h) \quad \text{for all } \bar{u}'^h \in \bar{V}^h \ . \tag{3.2}$$

Since

$$\bar{u}^h = \sum_{i=1}^{N} a_i \phi_i \quad , \quad \bar{u}'^h = \sum_{j=1}^{N} b_j \phi_j \tag{3.3}$$

we have

$$\sum_{i=1}^{N} \sum_{j=1}^{N} A(\phi_i, \phi_j) a_i b_j = R^h \sum_{i=1}^{N} \sum_{j=1}^{N} B(\phi_i, \phi_j) a_i b_j$$

or

$$\sum_{j=1}^{N} b_j \left(\sum_{i=1}^{N} K_{ij} a_i - R^h M_{ij} a_i \right) = 0 \tag{3.4}$$

where

$$K_{ij} = A(\phi_i, \phi_j)$$

and (3.5)

$$M_{ij} = B(\phi_i, \phi_j)$$

are $N \times N$ *matrices.*

Since $\bar{u}'^h$ *is arbitrary, so are the coefficients* b_j, *and the problem is reduced to one of solving the set of simultaneous linear equations*

$$\sum_{i=1}^{N} K_{ij} a_i = R^h \sum_{i=1}^{N} M_{ij} a_i \quad , \ j=1, \ \ldots, \ N \ . \tag{3.6}$$

Once these equations are solved, the approximate solution $\bar{u}^h$ *can be found from (3.3). We would like to choose families of basis functions* ϕ_i *in such a way that as N gets larger,* $\bar{V}^h$ *approaches*

the space $\bar{V}$ and $\bar{u}^h$ approaches the exact solution $\bar{u}$.

In the Galerkin method there is no systematic way of constructing reasonable basis functions. A poor choice may produce ill-conditioned element matrices resulting in inaccurate solutions. These difficulties can be overcome by using the finite element method, which is a special case of the Galerkin method. In the finite element method the basis functions are generated in a systematic manner in such a way that the family of spaces $\bar{V}^h (h \in (0,1))$ defined by the finite element procedure has the property that $\bar{V}^h$ approaches $\bar{V}$ as h approaches zero.

In Section 3.1 we construct a finite element mesh representing Ω and piecewise-polynomial basis functions defined on the mesh, which generate a finite-dimensional subspace of $\bar{V}$. To simplify calculations we construct a master element $\hat{\Omega}$ so that every element Ω_e in the domain can be generated by a map T_e. Having constructed the sequences of coordinate maps T_e from $\hat{\Omega}$ into elements Ω_e we can approximate the linear and energy eigenvalue problems obtained in Chapter 2. In Section 3.2 we calculate the element matrices for the eigenvalue problems. The numerical integration schemes used to calculate these matrices are discussed in Section 3.3. We conclude this chapter by giving an algorithm for solving the eigenvalue problems. We are interested only in the lowest eigenvalue and corresponding eigenvector, and an effective technique of finding these is the Inverse Iteration Method, which is described in Section 3.4. Our exposition in this Chapter of the finite element method follows that of Becker, Carey and Oden (1981) and Reddy (1986).

The finite element method is a technique for constructing an approximate solution to the variational eigenvalue problem (3.6) in a finite dimensional subspace $\vec{V}^h$. *The method involves dividing the domain of solution up into a finite number of subdomains, the finite elements, and constructing an approximation of the solution over the collection of finite elements. The basis functions are defined piecewise over the finite elements and are chosen to be very simple functions. To construct the piecewise basis functions, we first partition the domain* Ω *of our problem into a finite number* E *of subdomains* Ω_1, Ω_2, ..., Ω_E, *called finite elements which satisfy*

$$\Omega_e \cap \Omega_f = 0 \quad \text{for } e \neq f, \qquad \bigcup_{e=1}^{E} \bar{\Omega}_e = \bar{\Omega} \quad .$$

For simplicity, we assume that the domain Ω *is in* $\mathbb{R}^2$ *and that* Ω *is polygonal, so that the domain can be covered exactly by polygonal elements, as shown in Figure 3.1. Within each element, certain points are identified, called nodes or nodal points, which play an important role in the finite element method. Nodes are allocated at least at the vertices of elements.*

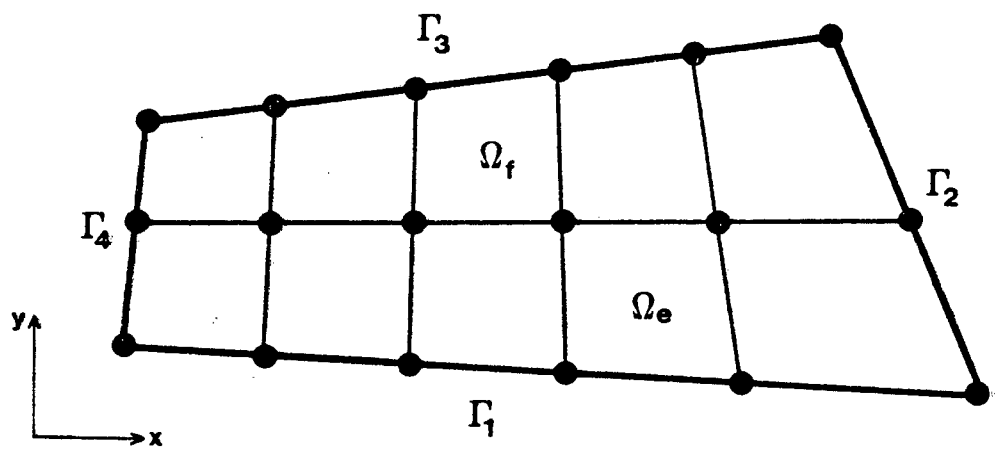


Figure 3.1

To improve the approximation additional nodes may be introduced at the midpoints of the sides of elements and at the centre of elements, as in Figure 3.2. There are a total of N nodes, numbered 1, 2, ..., N which have position vectors $\underline{x}_1$, $\underline{x}_2$, ..., $\underline{x}_N$. The set of elements and nodes that make up the domain is called the finite element mesh.

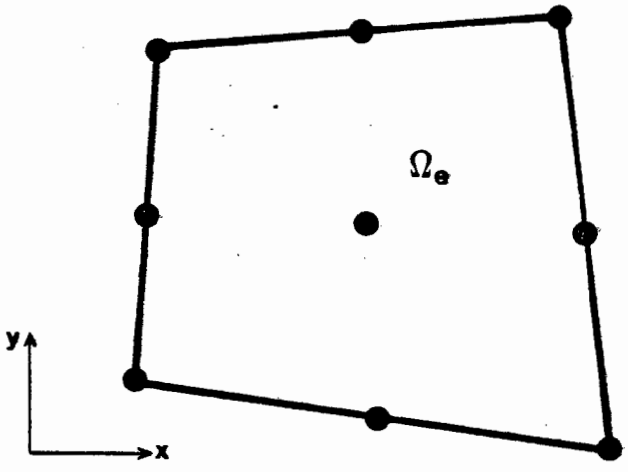


Figure 3.2

In the finite element method we set up basis functions which are piecewise polynomials, and which are non-zero only on a small part of the domain. There are a total of N basis functions $\phi_1(\underline{x}_j)$, $\phi_2(\underline{x}_j)$, ..., $\phi_N(\underline{x}_j)$ such that span $\{\phi_i\}_{i=1}^N = \bar{V}^h$, and for which

$$\phi_i(\underline{x}_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad (3.7)$$

where $(\underline{x}_j)$ are the coordinates of nodal points in the finite element mesh. When (3.7) holds we have, for $v^h \in \bar{V}^h$,

$$v^h(\underline{x}_j) = \sum_{i=1}^{N} b_i \phi_i(\underline{x}_j) = b_j ,$$

so that b_j is the value of v^h at node j.

Local basis functions $\psi_i^{(e)}$ *are constructed over each element* Ω_e *such that when patched together they produce the basis functions. The local basis functions satisfy the property*

$$\psi_i^{(e)}(\underline{x}_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad (3.8)$$

where i and j run over all nodes in Ω_e. *Thus we have* $\psi_i^{(e)} = \phi_i \big|_{\Omega_e}$.

We can now write (3.5) as

$$K_{ij} = A(\phi_i, \phi_j) = \sum_{e=1}^{E} A^{(e)}(\phi_i, \phi_j) = \sum_{e=1}^{E} \underbrace{A^{(e)}(\psi_i^{(e)}, \psi_j^{(e)})}_{K_{ij}^{(e)}}$$

and, similarly (3.9)

$$M_{ij} = B(\phi_i, \phi_j) = \sum_{e=1}^{E} B^{(e)}(\phi_i, \phi_j) = \sum_{e=1}^{E} \underbrace{B^{(e)}(\psi_i^{(e)}, \psi_j^{(e)})}_{M_{ij}^{(e)}}$$

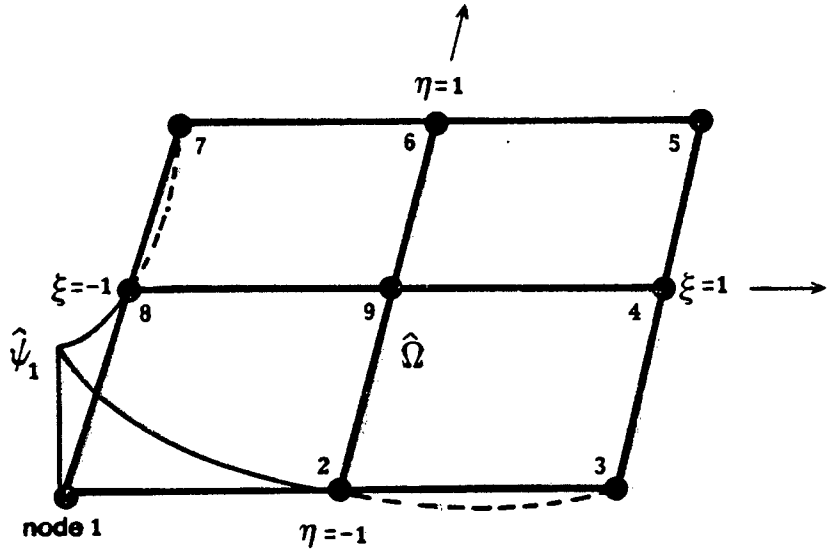


Figure 3.3: The master element $\hat{\Omega}$.

Instead of defining local basis functions for each element, we can simplify matters by setting up a master element $\hat{\Omega}$ *which is isolated from the actual finite element mesh and which has its own coordinate system. For simplicity, we choose* $\hat{\Omega}$ *to be the square* $(-1,1) \times (-1,1)$*, as shown in Figure 3.3. The master element has the same system of nodal points as the elements* Ω_e *in the actual mesh (nine nodes in this case). An invertible transformation is used to map points from the master element onto points in each element, as shown in Figure 3.4. We can now introduce a map* T_e *of* $\hat{\Omega}$ *onto* Ω_e*:*

$$T_e: \hat{\Omega} \rightarrow \Omega_e \qquad \begin{aligned} x &= x(\xi, \eta) \\ y &= y(\xi, \eta) \end{aligned} \qquad \text{or } T_e \underline{\xi} = \underline{x} \qquad (3.10)$$

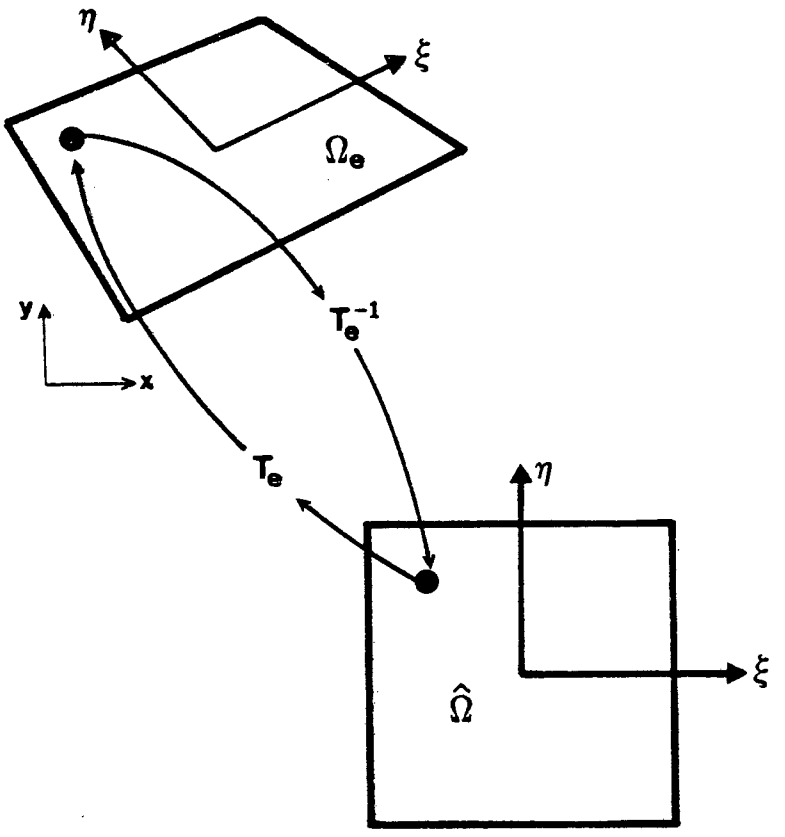


Figure 3.4: The element Ω_e *as the image of* $\hat{\Omega}$ *under the map* T_e

The advantage of setting up a master element in this way is that we can define the local basis functions $\hat{\psi}_i$ *on* $\hat{\Omega}$ *and simply use (3.10) to map* $\hat{\psi}_i$ *to* $\psi_i^{(e)}$ *by defining* $\psi_i^{(e)}$ *to be functions on* Ω_e *satisfying*

$$\psi_i^{(e)}(x,y) = \hat{\psi}_i(\xi,\eta). \tag{3.11}$$

For a master element with nine nodes the local basis functions that satisfy (3.8) are

$$\left.\begin{aligned}
\hat{\psi}_1 &= {}^1/_4\,(\xi^2-\xi)(\eta^2-\eta)\\
\hat{\psi}_2 &= {}^1/_2\,(1-\xi^2)(\eta^2-\eta)\\
\hat{\psi}_3 &= {}^1/_4\,(\xi^2+\xi)(\eta^2-\eta)\\
\hat{\psi}_4 &= {}^1/_2\,(\xi^2+\xi)(1-\eta^2)\\
\hat{\psi}_5 &= {}^1/_4\,(\xi^2+\xi)(\eta^2+\eta)\\
\hat{\psi}_6 &= {}^1/_2\,(1-\xi^2)(\eta^2+\eta)\\
\hat{\psi}_7 &= {}^1/_4\,(\xi^2-\xi)(\eta^2+\eta)\\
\hat{\psi}_8 &= {}^1/_2\,(\xi^2-\xi)(1-\eta^2)
\end{aligned}\right\} \tag{3.12}$$

$$\hat{\psi}_9 = (1-\xi^2)(1-\eta^2)$$

The piecewise biquadratic polynomial basis functions ϕ_i *are formed by patching together the local basis functions* $\psi_i^{(e)}$ *associated with node i.*

It is now possible to transform the operations on finite elements Ω_e so that they hold on $\hat{\Omega}$. The complete finite element mesh containing E elements is generated by a sequence of transformations $\{T_1, T_2, \ldots, T_E\}$ in which each element Ω_e is the image of the fixed master element under a coordinate map T_e, as illustrated in Figure 3.5.

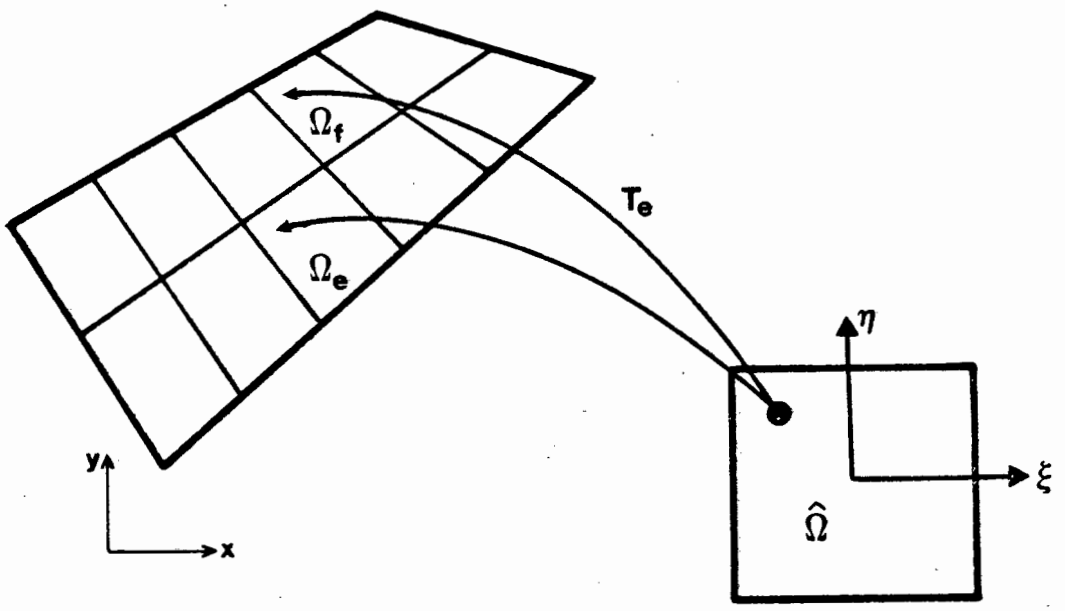


Figure 3.5: Generation of the finite element mesh.

The map T_e can be constructed using the finite element local basis functions:

$$T_e \underline{\xi} = \underline{x}: \qquad x(\xi, \eta) = \sum_{i=1}^{n} \hat{\psi}_i(\xi, \eta) x_i \qquad (3.13)$$

$$y(\xi, \eta) = \sum_{i=1}^{n} \hat{\psi}_i(\xi, \eta) y_i$$

where x and y are the coordinates at any point of the element,

x_i and y_i, $i = 1, \ldots, n$ are the coordinates of the n element nodes and

$\hat{\psi}_i$ are the local basis functions defined in the natural coordinate system of the element.

The variables u, v and θ at any point within the element are given by

$$u(\xi, \eta) = \sum_{i=1}^{n} \hat{\psi}_i(\xi, \eta) u_i$$

$$v(\xi, \eta) = \sum_{i=1}^{n} \hat{\psi}_i(\xi, \eta) v_i \qquad (3.14)$$

$$\theta(\xi, \eta) = \sum_{i=1}^{n} \hat{\psi}_i(\xi, \eta) \theta_i$$

where u_i, v_i and θ_i, $i = 1, \ldots, n$ are values of u, v and θ at node i.

The bilinear form $A(.,.)$ in (3.9) contains derivatives of u, v and θ which need to be evaluated with respect to the local (x,y) coordinates. Because u, v and θ are defined in the natural coordinate system as defined in (3.14), we need to relate the x-y derivatives to the ξ-η derivatives. Since the map T_e is invertible, we may define the inverse map T_e^{-1} of the x-y coordinates into the ξ-η coordinates:

$$T_e^{-1} : \begin{aligned} \xi &= \xi(x,y) \\ \eta &= \eta(x,y) \end{aligned} \qquad (3.15)$$

To obtain the derivatives $\partial/\partial x$ and $\partial/\partial y$ in terms of $\partial/\partial \xi$ and $\partial/\partial \eta$ we use the chain rule in the form

$$\left.\begin{aligned} \partial/\partial x &= \partial\xi/\partial x \;\; \partial/\partial \xi + \partial\eta/\partial x \;\; \partial/\partial \eta \\ \partial/\partial y &= \partial\xi/\partial y \;\; \partial/\partial \xi + \partial\eta/\partial y \;\; \partial/\partial \eta \end{aligned}\right\} \quad (3.16)$$

and the derivatives are evaluated as follows

$$\begin{bmatrix} \partial/\partial \xi \\ \partial/\partial \eta \end{bmatrix} = \begin{bmatrix} \partial x/\partial \xi & \partial y/\partial \xi \\ \partial x/\partial \eta & \partial y/\partial \eta \end{bmatrix} \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \end{bmatrix} \quad (3.17)$$

Here, the 2×2 matrix of partial derivatives is called the Jacobian matrix , J, of the transformation (3.15). We want to solve for $\partial/\partial x$ and $\partial/\partial y$ in terms of $\partial/\partial \xi$ and $\partial/\partial \eta$ and use

$$\begin{bmatrix} \partial/\partial x \\ \partial/\partial y \end{bmatrix} = J^{-1} \begin{bmatrix} \partial/\partial \xi \\ \partial/\partial \eta \end{bmatrix}$$

or (3.18)

$$\begin{bmatrix} \partial/\partial x \\ \partial/\partial y \end{bmatrix} = {}^{1}/_{detJ} \begin{bmatrix} \partial y/\partial \eta & -\partial y/\partial \xi \\ -\partial x/\partial \eta & \partial x/\partial \xi \end{bmatrix} \begin{bmatrix} \partial/\partial \xi \\ \partial/\partial \eta \end{bmatrix}$$

which requires that the inverse of J exist. This inverse exists if there is a one-to-one correspondence between the natural and

local coordinates of the element. We can now evaluate the partial derivatives in terms of the master element and can therefore construct the matrices defined in equation (3.6).

3.2 Finite element calculations

In the finite element method the variables $\underline{u}^h$ *and* θ^h *are completely described by values at nodal points. Consider the two-dimensional nine-noded element shown in Figure 3.6.*

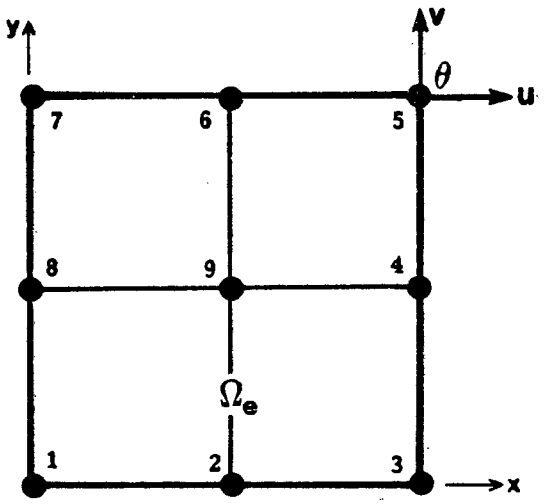


Figure 3.6

At each node i, i=1, ..., 9 we have 3 unknowns: the horizontal and vertical components of velocity, u and v, and the temperature θ*. The nodal variables for each element can then be written as*

$$\underline{\delta}^e = \begin{bmatrix} \underline{\delta}_1 \\ \underline{\delta}_2 \\ \vdots \\ \underline{\delta}_9 \end{bmatrix} \qquad (3.19)$$

where $\underline{\delta}_i$ *at node i is given by*

$$\underline{\delta}_i = \begin{bmatrix} u_i \\ v_i \\ \theta_i \end{bmatrix} \quad . \tag{3.20}$$

The variables u,v and θ *at any point inside the element are given in terms of their nodal point values using the local basis functions (3.12). Thus* $\underline{\delta}$ *at any point within the element can be expressed, in terms of the element basis functions, as*

$$\underline{\delta} = \underline{\psi}\,\underline{\delta}^e = \sum_{i=1}^{9} \underline{\psi}_i \; \underline{\delta}_i \tag{3.21}$$

using (3.11).

The derivatives of u,v and θ *with respect to the local coordinate system are collected together in a vector* $\underline{\varepsilon}$ *, defined by*

$$\underline{\varepsilon} = \begin{bmatrix} \partial u / \partial x \\ \partial v / \partial y \\ (\partial u / \partial y + \partial v / \partial x) \sqrt{2} / 2 \\ \partial \theta / \partial x \\ \partial \theta / \partial y \end{bmatrix} \quad . \tag{3.22}$$

Substitution for u,v and θ from (3.21) results in

$$\underline{\varepsilon} = \underline{B}\,\underline{\delta}^{e} = \sum_{i=1}^{9} \underline{B}_i\,\underline{\delta}_i \qquad (3.23)$$

where

$$\underline{B}_i = \begin{bmatrix} \partial\hat{\psi}_i/\partial x & 0 & 0 \\ 0 & \partial\hat{\psi}_i/\partial y & 0 \\ \sqrt{2}/2\,\partial\hat{\psi}_i/\partial y & \sqrt{2}/2\,\partial\hat{\psi}_i/\partial x & 0 \\ 0 & 0 & \partial\hat{\psi}_i/\partial x \\ 0 & 0 & \partial\hat{\psi}_i/\partial y \end{bmatrix} . \qquad (3.24)$$

If we set

$$\Delta = \partial u/\partial x + \partial v/\partial y \qquad (3.25)$$

then substituting for u and v from (3.21) we get

$$\Delta = \underline{C}\;\underline{\delta}^{e} = \sum_{i=1}^{9} \underline{C}_i\;\underline{\delta}_i \qquad (3.26)$$

where

$$\underline{C}_i = \begin{bmatrix} \partial\hat{\psi}_i/\partial x \\ \partial\hat{\psi}_i/\partial y \end{bmatrix} \qquad (3.27)$$

Next, define

$$\underline{E}_i = \begin{bmatrix} 0 \\ 0 \\ \hat{\psi}_i \end{bmatrix} \quad \text{and} \quad \underline{F}_i = \begin{bmatrix} 0 \\ \hat{\psi}_i \\ 0 \end{bmatrix} \quad ; \quad (3.28)$$

then we can write

$$\theta = \underline{E}\ \underline{\delta}^e \quad \text{and} \quad v = \underline{F}\ \underline{\delta}^e .$$

We define $\underline{\delta}'^e$, $\underline{\varepsilon}'^e$, $\underline{C}'^e$, $\underline{E}'^e$ *and* $\underline{F}'^e$ *in terms of* $\bar{u}'$ *similarly, i.e.,*

$$\underline{\delta}'_i = \begin{bmatrix} u'_i \\ v'_i \\ \theta'_i \end{bmatrix} \quad (3.29)$$

and

$$\begin{aligned} \underline{\varepsilon}' &= \underline{B}'\ \underline{\delta}'^e , \\ \Delta' &= \underline{C}'\ \underline{\delta}'^e , \\ \theta' &= \underline{E}'\ \underline{\delta}'^e , \\ v' &= \underline{F}'\ \underline{\delta}'^e . \end{aligned} \quad (3.30)$$

(i) The penalised eigenvalue problem for the linear stability theory obtained in Section 2.2 has the form

$$A(\bar{u}_\varepsilon, \bar{u}') = R_{L,\varepsilon}\ B(\bar{u}_\varepsilon, \bar{u}') \quad (3.31)$$

where $A(\bar{u}_\varepsilon, \bar{u}') = a(\underline{u}_\varepsilon, \underline{u}') + b(\theta_\varepsilon, \theta')$ and (3.32)

$B(\bar{u}_\varepsilon, \bar{u}') = c(\theta_\varepsilon, \underline{u}') + d(\underline{u}_\varepsilon, \theta')$. (3.33)

Equations (3.32) and (3.33) in terms of the approximations (3.19) - (3.30) are, for element e,

$$A^e(\bar{u}_\varepsilon, \bar{u}') = \int_{\Omega_e} (\underline{\delta}'^e)^T \{2\underline{B}'^T\underline{B} + {}^1/_\varepsilon \underline{C}'^T\underline{C}\} \underline{\delta}^e + \int_{\Gamma_h^e} (\underline{\delta}'^e)^T h_T \underline{E}'^T\underline{E}\ \underline{\delta}^e \quad \ldots(3.34)$$

and

$$B^e(\bar{u}_\varepsilon, \bar{u}') = -\int_{\Omega_e} (\underline{\delta}'^e)^T \{\underline{E}'^T\underline{F}(\underline{g}.\underline{i}) + \underline{F}'^T\underline{E}(\nabla T.\underline{i})\}\ \underline{\delta}^e \quad (3.35)$$

where $\underline{i}$ is the unit vector in the direction of increasing y.

The eigenvalue problem is obtained by summing (3.34) and (3.35) over all E elements in the mesh. Thus, (3.31) becomes

$$\sum_{e=1}^{E} (\underline{\delta}'^e)^T \{\int_{\Omega_e} (\underline{K} + {}^1/_\varepsilon \underline{H}) + \int_{\Gamma_h^e} h_T \underline{K}^b\} \underline{\delta}^e = R_{L,\varepsilon}^h \sum_{e=1}^{E} (\underline{\delta}'^e)^T \int_{\Omega_e} \underline{M}\ \underline{\delta}^e \quad \ldots(3.36)$$

where $\underline{K} = 2\underline{B}'^T\underline{B}$,

$\underline{H} = \underline{C}'^T\underline{C}$,

$\underline{K}^b = \underline{E}'^T\underline{E}$ and

$\underline{M} = -\underline{E}'^T\underline{F}(\underline{g}.\underline{i}) - \underline{F}'^T\underline{E}(\nabla T.\underline{i})$.

The integration in equation (3.36) is defined in the local (x,y) coordinate system. The differential dxdy in terms of the natural coordinates is

$$dxdy = \det J\ d\xi\ d\eta$$

where det J is the determinant of the Jacobian operator. The

integration extends over the area $-1 \leq \xi \leq +1$ *and* $-1 \leq \eta \leq +1$. *For the one-dimensional case* $J = {}^{L}/_{2}$ *where* L *is the length of the element boundary. The integration defined in the natural* (ξ, η) *coordinate system extends over one of the boundaries of the master element.*

Substituting into (3.36), we get

$$\sum_{e=1}^{E} (\underline{\delta}'^{e})^{T} \underline{K}^{e} \underline{\delta}^{e} = R_{L,\varepsilon}^{h} \sum_{e=1}^{E} (\underline{\delta}'^{e})^{T} \underline{M}^{e} \underline{\delta}^{e} \qquad (3.37)$$

where $\underline{K}^{e} = \int_{-1}^{+1}\int_{-1}^{+1} (\underline{K} + {}^{1}/_{\varepsilon}\, \underline{H}) \det J \, d\xi \, d\eta + \int h_T \underline{K}^{b} ({}^{L}/_{2}) \, d\hat{s}$,

$$\underline{M}^{e} = \int_{-1}^{+1}\int_{-1}^{+1} \underline{M} \det J \, d\xi \, d\eta \quad ,$$

$\underline{K} = \underline{K}(x(\xi, \eta), y(\xi, \eta))$,

$\underline{M} = \underline{M}(x(\xi, \eta), y(\xi, \eta))$, *etc. and*

$\hat{s}$ *measures the distance along one of the master boundaries.*

(ii) The penalised eigenvalue problem for the energy stability theory obtained in Section 2.3 has the form

$$A'(\bar{u}_{\varepsilon}, \bar{u}') = \rho_{\lambda\varepsilon} B'(\bar{u}_{\varepsilon}, \bar{u}') \qquad (3.38)$$

where $A'(\bar{u}_{\varepsilon}, \bar{u}') = a(\underline{u}_{\varepsilon}, \underline{u}') + b(\theta_{\varepsilon}, \theta')$ *and* (3.39)

$$B'(\bar{u}_{\varepsilon}, \bar{u}') = e(\underline{u}_{\varepsilon}, \theta') + e(\theta_{\varepsilon}, \underline{u}') \qquad (3.40)$$

The bilinear form $A'(\bar{u}_\varepsilon, \bar{u}')$ differs from $A(\bar{u}_\varepsilon, \bar{u}')$ in the linear theory and the vector $\underline{\varepsilon}$ defined in (3.22) must be modified slightly.

Define

$$\underline{\varepsilon}_\lambda = \begin{bmatrix} \partial u / \partial x \\ \partial v / \partial y \\ (\partial u / \partial y + \partial v / \partial x) \sqrt{2} / 2 \\ \sqrt{\lambda}\, \partial\theta / \partial x \\ \sqrt{\lambda}\, \partial\theta / \partial y \end{bmatrix}$$

and $\underline{B}_\lambda$ similarly.

Equations (3.39) and (3.40) in terms of the finite element approximations are, for element e

$$A'^e(\bar{u}_\varepsilon, \bar{u}') = \int_{\Omega_e} (\underline{\delta}'^e)^T \{ 2\underline{B}_\lambda'^T \underline{B}_\lambda + 1/\varepsilon \, \underline{C}'^T \underline{C} \} \underline{\delta}^e + \int_{\Gamma_h^e} (\underline{\delta}'^e)^T \lambda h_T \underline{E}^T \underline{E} \, \underline{\delta}^e \qquad (3.41)$$

and

$$B'^e(\bar{u}_\varepsilon, \bar{u}') = - \int_{\Omega_e} (\underline{\delta}'^e)^T \, 1/2 (\underline{F}'^T \underline{E} + \underline{E}'^T \underline{F}) (\underline{g} + \lambda \nabla T) \cdot \underline{i} \, \underline{\delta}^e \qquad (3.42)$$

where $\underline{i}$ is a unit vector in the direction of increasing y.

The eigenvalue problem approximating (3.38) is given by

$$\sum_{e=1}^{E} (\ \underline{\delta}'^{e})^T \left\{ \int_{\Omega_e} (\underline{K}_\lambda + {}^1/_\varepsilon\ \underline{H} + \int_{\Gamma_h^e} \lambda h_T \underline{K}^b \right\} \underline{\delta}^e = \rho_{\lambda\varepsilon}^h \sum_{e=1}^{E} (\ \underline{\delta}'^{e})^T \int_{\Omega_e} \underline{M}_\lambda\ \underline{\delta}^e \qquad (3.43)$$

where $\underline{K}_\lambda = 2\ \underline{B}'^T_\lambda\ \underline{B}_\lambda$ *and*

$$\underline{M} = -\ (\underline{F}'^T \underline{E} + \underline{E}'^T \underline{F})\ (\underline{q} + \lambda \nabla T).\underline{i}\ .$$

Set $dx\ dy = det\ J\ d\xi\ d\eta$

then equation (3.43) may be written in the form

$$\sum_{e=1}^{E} (\ \underline{\delta}'^{e})^T \quad \underline{K}_\lambda^e \quad \underline{\delta}^e = \rho_{\lambda\varepsilon}^h \sum_{e=1}^{E} (\ \underline{\delta}'^{e})^T \quad \underline{M}_\lambda^e \quad \underline{\delta}^e \qquad (3.44)$$

where $\underline{K}_\lambda^e = \int_{-1}^{+1}\int_{-1}^{+1} (\underline{K} + {}^1/_\varepsilon\ \underline{H})\ det\ J\ d\xi\ d\eta + \int \lambda h_T \underline{K}^b\ ({}^L/_2)\ d\hat{s}$ *and*

$$\underline{M}^e = \int_{-1}^{+1}\int_{-1}^{+1} \underline{M}\ \ det\ J\ d\xi\ d\eta\ .$$

3.3 Numerical integration

In the finite element method integrals of the form (3.37) and (3.44) are evaluated using numerical techniques because of their advantage over analytical procedures [Bathe (1982)]. In particular, we choose Gauss quadrature for the ease with which it can be implemented and for its high accuracy.

The element matrices,

$$\underline{\mathbf{K}}^e = \int_{-1}^{+1}\int_{-1}^{+1} f(\xi,\eta)\, d\xi\, d\eta + \int_{-1}^{+1} g(\xi,\eta)\, d\xi \Big|_{\eta=1}$$

and

$$\underline{\mathbf{M}}^e = \int_{-1}^{+1}\int_{-1}^{+1} h(\xi,\eta)\, d\xi\, d\eta \qquad (3.45)$$

can now be evaluated as

$$K_{ij}^e = \sum_{ij}^{p} a_i\, a_j\, f(\xi_i, \eta_j) + \sum_{i}^{p} a_i\, g(\xi_i)$$

and

$$M_{ij}^e = \sum_{ij}^{p} a_i\, a_j\, h(\xi_i, \eta_j) \qquad (3.46)$$

where f, g and h are the element matrices evaluated at the integration points,

ξ_i *and* η_j *are the coordinates of the integration points,*

a_i *and* a_j *are the corresponding weighting factors and*

p is the total number of sampling points.

The integration (or sampling) points and weighting factors are chosen to obtain maximum accuracy in the integration. In Gauss quadrature formulas, a polynomial of order (2p -1) is integrated exactly for p sampling points. The sampling points and

weights depend on the interval of integration. The sampling points and weights for the master element with interval -1 to +1 are given in Table 3.1 for values of p = 1, 2 and 3.

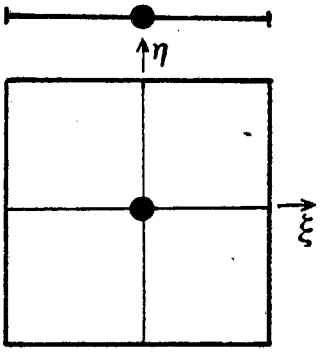
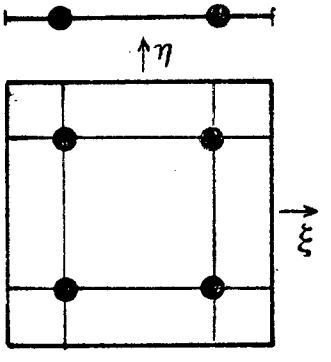
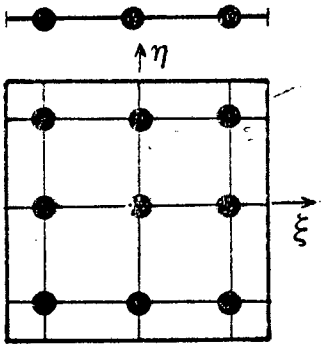
Order of integration	*Sampling points*	*weighting factors*	*Location of integration points*
p = 1	*0.0*	*2.0*	
p = 2	*± 0.5773503*	*1.0*	
p = 3	*± 0.7745967* *0.0*	*0.5555556* *0.8888889*	

Table 3.1

The choice of the order of numerical integration is important because the results are sensitive to different integration orders. Since all integrands are of polynomial type, all matrices are evaluated exactly if the order of the numerical integration is high enough. It has been shown that the use of a lower integration order when evaluating the penalty matrix, $\int \underline{C}'^{T}\underline{C}$ *, is necesary for reasons of stability [Oden, Kikuchi and Song (1982)]. Therefore, reduced integration may lead to improved results. The appropriate integration orders in the evaluation of a 9-noded element are given in Table 3.2.*

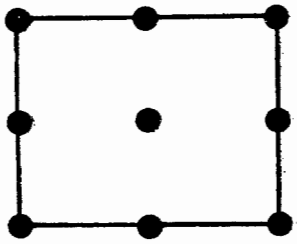
Element	Exact integration order	Reduced integration order
	3 × 3	2 × 2 or 1 × 1

Table 3.2

3.4 Solution method of the eigenvalue problem

We need to find the solution to the eigenvalue problem

$$\underline{\boldsymbol{K}}\ \underline{a} = R^{h}\ \underline{\boldsymbol{M}}\ \underline{a} \qquad (4.47)$$

where $\underline{K}$ *and* $\underline{\boldsymbol{M}}$ *are, respectively, the stiffness matrix and mass matrix of the finite element assemblage.*

In particular, we want to find the smallest eigenvalue R^h_1 *and corresponding eigenvector* $\underline{a}_1$. *The vector iteration method uses the property that*

$$\underline{\boldsymbol{K}}\ \underline{a}_i = R^h_i\ \underline{\boldsymbol{M}}\ \underline{a}_i \tag{4.48}$$

where $\underline{\boldsymbol{K}}$ *is positive definite.*

A vector iteration technique that calculates an eigenvector and corresponding eigenvalue very effectively is the inverse iteration method. The basic steps of the inverse iteration method are as follows [Bathe (1982)] :

Assume a starting vector $\underline{x}_1$ *so that* $\underline{y}_1 = \underline{\boldsymbol{M}}\underline{x}_1$, *and evaluate in each iteration step k=1, 2, ...*

$$\underline{\boldsymbol{K}}\ \underline{x}_{k+1} = \underline{y}_k \tag{3.49}$$

$$\underline{y}_{k+1} = \underline{\boldsymbol{M}}\ \underline{x}_{k+1} \tag{3.50}$$

$$\rho(\underline{x}_{k+1}) = \frac{\underline{x}^T_{k+1}\ \underline{y}_k}{\underline{x}^T_{k+1}\ \underline{y}_{k+1}} \tag{3.51}$$

$$\underline{y}_{k+1} = \frac{\underline{y}_{k+1}}{(\underline{x}^T_{k+1}\ \underline{y}_{k+1})^{1/2}} \tag{3.52}$$

where, provided that $\underline{y}^T_1\ \underline{a}_1 \neq 0$,

$$\underline{y}_{k+1} \rightarrow \underline{\boldsymbol{M}}\ \underline{a}_1 \quad \text{and} \quad \rho(\underline{x}_{k+1}) \rightarrow R^h_1 \quad \text{as } k \rightarrow \infty\ .$$

In (3.51) we obtain an approximation to the eigenvalue R^h_1 *given by the Rayleigh quotient* $\rho(\underline{x}_{k+1})$. *This approximation to* R^h_1 *is used to determine convergence in the iteration. We say we have convergence when*

$$\frac{\left|\rho(\underline{x}_{k+1}) - \rho(\underline{x}_k)\right|}{\rho(\underline{x}_{k+1})} \leq tol \qquad (3.53)$$

where tol is a specified tolerance and

$\rho(\underline{x}_{k+1})$ *denotes the current approximation to* R^h_1.

If m is the last iteration, then we have

$$R_1^h \simeq \rho(\underline{x}_{m+1}) \qquad (3.54)$$

and

$$\underline{a}_1 \simeq \frac{\underline{x}_{m+1}}{(\underline{x}_{m+1}^T \underline{y}_{m+1})^{1/2}}. \qquad (3.55)$$

In the previous Chapter we obtained the equations arising from the finite element approximation of the eigenvalue problems for the linear and energy stability theory. These can be written in the form: find $\underline{a} \in \vec{V}_h$ *and* $R^h \in \mathbb{R}$ *such that*

$$\underline{\boldsymbol{K}}\,\underline{a} = R^h \quad \underline{\boldsymbol{M}}\,\underline{a} \tag{4.1}$$

where R^h *is the lowest eigenvalue,*

$\underline{a}$ *is the eigenvector corresponding to* R^h,

$\underline{\boldsymbol{K}}$ *is the stiffness matrix (which includes the contribution due to the penalty approximation) and*

$\underline{\boldsymbol{M}}$ *is the mass matrix.*

In this Chapter we set up the finite element program to calculate the eigenvalue, R^h, *and corresponding eigenvector,* $\underline{a}$, *in (4.1) numerically. The programming details are based on Hinton and Owen (1977) and Bathe (1982).*

In the finite element method individual element stiffness and mass matrices are calculated separately and then assembled into the global stiffness and mass matrices, respectively. The global matrices are obtained by summing the element matrices, so that we have

$$\underline{\boldsymbol{K}} = \sum_{e=1}^{E} \underline{\boldsymbol{K}}^e \qquad \text{and} \qquad \underline{\boldsymbol{M}} = \sum_{e=1}^{E} \underline{\boldsymbol{M}}^e \tag{4.2}$$

where $\underline{\boldsymbol{K}}^e$ *is the stiffness matrix of element e and*

$\underline{\boldsymbol{M}}^e$ *is the mass matrix of element e.*

The element stiffness and mass matrices, $\underline{\boldsymbol{K}}^e$ *and* $\underline{\boldsymbol{M}}^e$, *are of the same order as the global stiffness and mass matrices,* $\underline{\boldsymbol{K}}$ *and* $\underline{\boldsymbol{M}}$,

and have non-zero entries only in those rows and columns that correspond to nodes in element e. Therefore, we only need to store the compacted element matrices, which are of the order of the element degrees of freedom, together with an array that relates to each element degree of freedom the corresponding assemblage degree of freedom. This vector is called the connectivity array, LEQNS, and is of the order of the number of degrees of freedom per element. For example, consider the mesh shown in Figure 4.1.

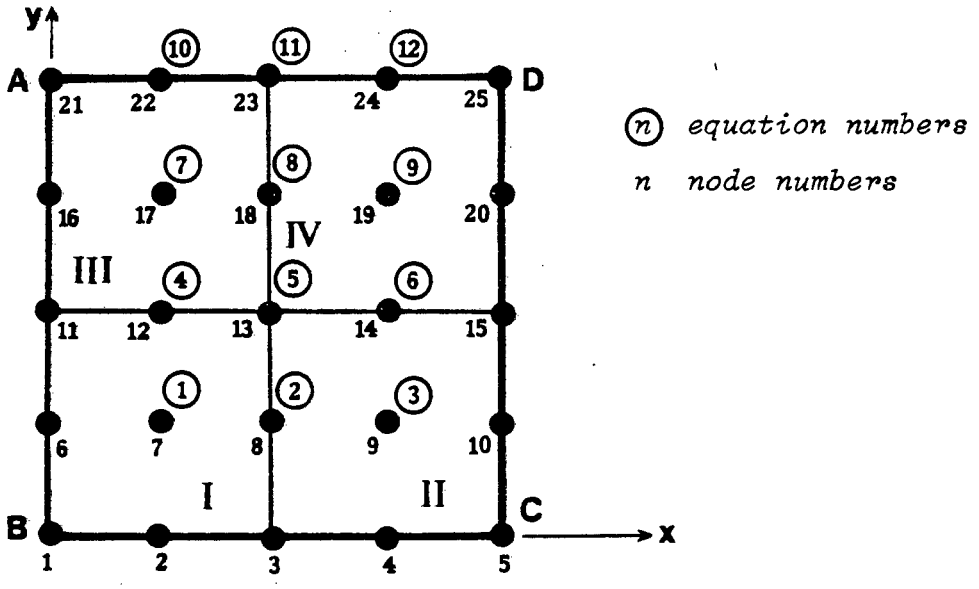


Figure 4.1

Assume for ease of explanation that only one degree of freedom exists at each node, and that a_i=0 along the edges AB, BC and CD. There are thus a total number of 12 equations, a_i (i=1, ..., 12), in this example. The compacted element stiffness and mass matrices are of the order of the element degrees of freedom, i.e., they are 9 × 9 matrices. For element I in Figure 4.1. the connectivity array has the following entries

row, column	1	2	3	4	5	6	7	8	9
LEQNS =	0	0	0	2	5	4	0	0	1

where a zero means that the corresponding row and column of the compacted element matrix are ignored and do not enter the global matrix. Similarly, the connectivity array for element II is

$$LEQNS = \boxed{0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 6 \quad 5 \quad 2 \quad 3} \quad ,$$

for element III

$$LEQNS = \boxed{0 \quad 4 \quad 5 \quad 8 \quad 11 \quad 10 \quad 0 \quad 0 \quad 7} \quad ,$$

and for element IV

$$LEQNS = \boxed{5 \quad 6 \quad 0 \quad 0 \quad 0 \quad 12 \quad 11 \quad 8 \quad 9} \quad .$$

The connectivity array of an element is determined from the nodal points and the equation numbers that have been assigned to it. Once the connectivity array has been defined, the compact element matrices may be added to the global matrices.

In practice, we never actually assemble the complete global stiffness and mass matrices since this results in wasted computer storage. The global matrices possess special features which allow us to economise on computer storage and computational time. These special features are: (i) the symmetry and

(ii) the bandedness of the global matrices.

In the energy problem, discussed in Section 2.3, both the stiffness and mass matrices are symmetric. In the linear problem, discussed in Section 2.2, the stiffness matrix, $\underline{\mathbf{K}}$, is symmetric, but the mass matrix, $\underline{\mathbf{M}}$, is generally unsymmetric. The symmetry allows us to process only that part of the matrix on and above the leading diagonal. In addition, the matrices are banded due to the systematic numbering of the nodal points. The non-zero

elements in the matrix form a band which contains the main diagonal, as illustrated in Figure 4.2. Let m(i), (i = 1, ..., n) be the row number of the first non-zero element in column i; we refer to m(i), (i = 1, ..., n) as the skyline of the matrix. The bandedness allows us to store and manipulate only that part of the matrix which lies below the skyline. All zero elements outside the skyline are ignored in the solution procedure. Zero elements within the skyline of the matrix are stored and operated on, since they may become non-zero during matrix reduction.

An effective storage scheme for the stiffness and mass matrices is to store only the elements below the skyline of the matrices $\underline{\mathbf{K}}$ *and* $\underline{\mathbf{M}}$ *in one-dimensional arrays* $\underline{\mathbf{A}}$ *and* $\underline{\mathbf{B}}$*, as illustrated in Figure 4.3. In addition, we also define an array MAXAI which stores the addresses of the diagonal elements of* $\underline{\mathbf{K}}$ *in* $\underline{\mathbf{A}}$ *and, similarly, of* $\underline{\mathbf{M}}$ *in* $\underline{\mathbf{B}}$*. Thus, the address of the* i^{th} *diagonal element of* $\underline{\mathbf{K}}$*,* K_{ii}*, in* $\underline{\mathbf{A}}$ *is given by MAXAI(i).*

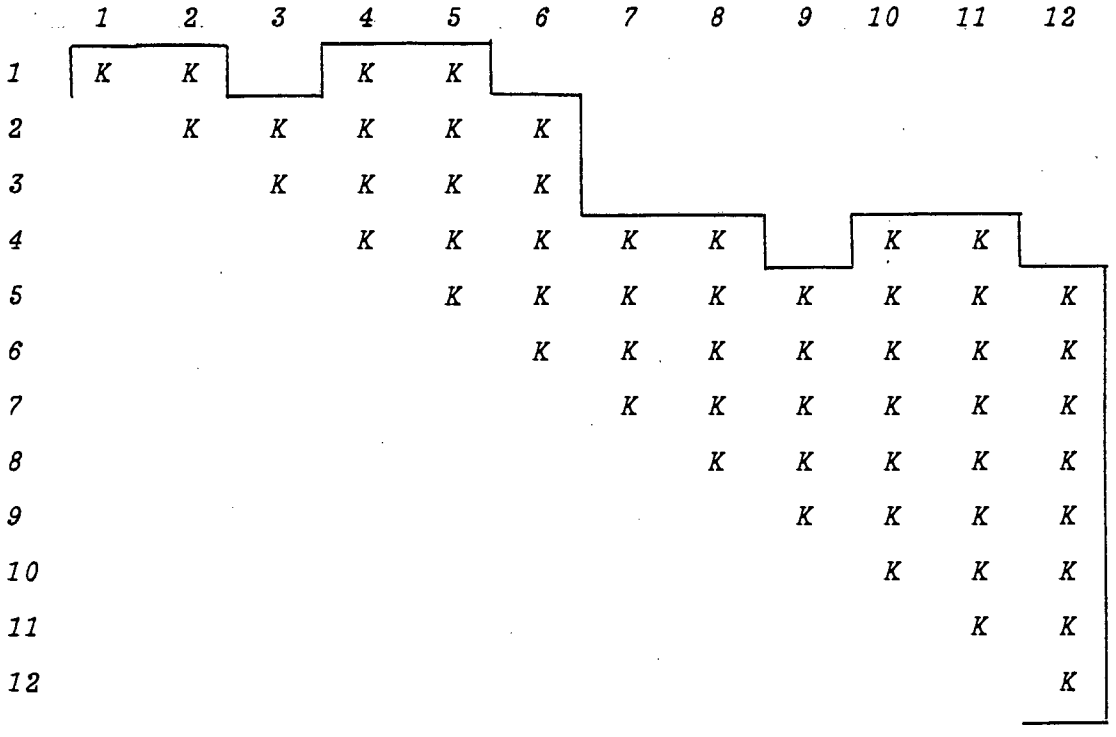


Figure 4.2: The upper triangle of $\underline{\mathbf{K}}$*.*

	1	2	3	4	5	6	7	8	9	10	11	12
1	A_1	A_3		A_9	A_{14}							
2		A_2	A_5	A_8	A_{13}	A_{19}						
3			A_4	A_7	A_{12}	A_{18}						
4				A_6	A_{11}	A_{17}	A_{23}	A_{28}		A_{40}	A_{48}	
5					A_{10}	A_{16}	A_{22}	A_{27}	A_{33}	A_{39}	A_{47}	A_{56}
6						A_{15}	A_{21}	A_{26}	A_{32}	A_{38}	A_{46}	A_{55}
7							A_{20}	A_{25}	A_{31}	A_{37}	A_{45}	A_{54}
8								A_{24}	A_{30}	A_{36}	A_{44}	A_{53}
9									A_{29}	A_{35}	A_{43}	A_{52}
10										A_{34}	A_{42}	A_{51}
11											A_{41}	A_{50}
12												A_{49}

$MAXAI^T =$ | 1 | 2 | 4 | 6 | 10 | 15 | 20 | 24 | 29 | 34 | 41 | 49 |

Figure 4.3: Array $\underline{A}$ *storing elements of the upper triangle of* $\underline{K}$.

In the linear stability problem the mass matrix is generally unsymmetric. When dealing with an unsymmetric mass matrix we need to store the lower triangle of $\underline{M}$ *in the same way as the upper triangle. In this case the elements below the skyline of* $\underline{M}$ *are stored in a one-dimensional array,* $\underline{B}_u$, *and the elements below the skyline of* $\underline{M}^T$ *are stored in a one-dimensional array,* $\underline{B}_l$. *The diagonal elements of the upper and lower triangles have the same*

addresses, i.e., only one array MAXAI is needed to address both $\underline{B}_u$ and $\underline{B}_l$, as illustrated in Figure 4.4.

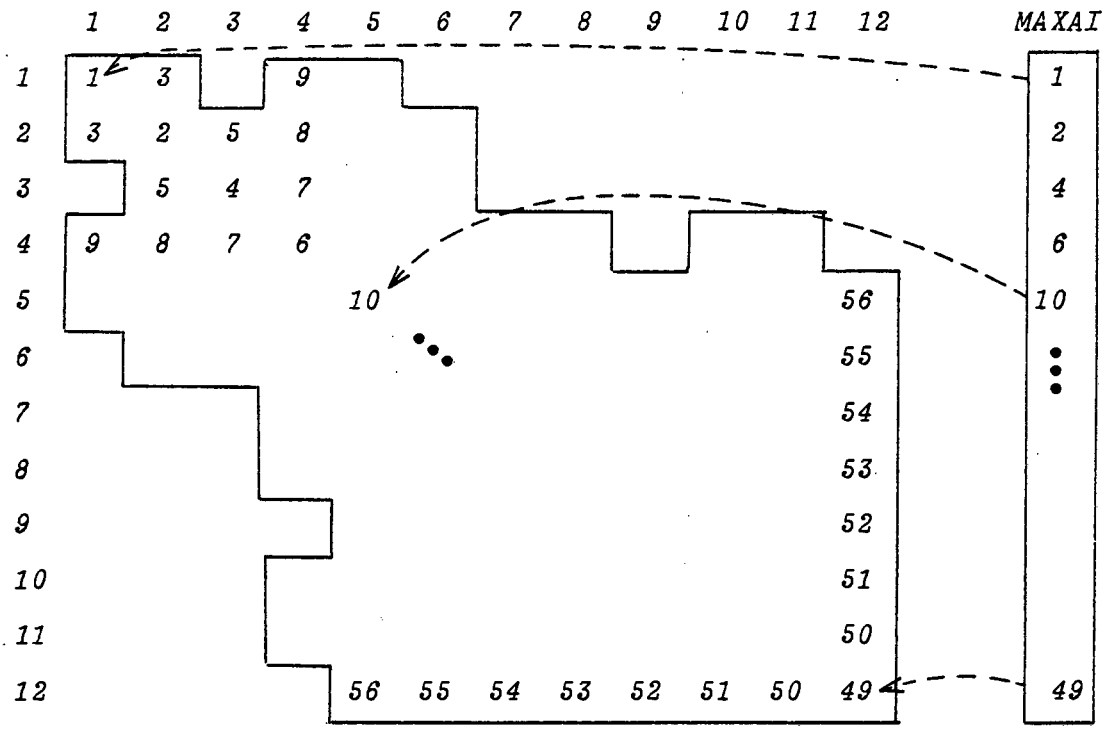


Figure 4.4: The unsymmetric mass matrix $\underline{M}$.

The basic structure of the finite element program is given in Section 4.1. The input required by the finite element analysis is discussed in Section 4.2. The addresses of the elements in the global matrices are calculated in Section 4.3. In Section 4.4 the compacted element matrices, $\underline{K}^e$ *and* $\underline{M}^e$, *are calculated and added to the global vectors* $\underline{A}$ *and* $\underline{B}$. *The eigenvalue solution routine is discussed in Section 4.5.*

4.1 The program

A modular approach is adopted where separate subroutines perform the various main finite element operations. The basic finite element steps are performed by primary subroutines which rely on auxiliary subroutines to carry out secondary operations. The master or main routine organises the calling of the primary routines as outlined in the flow diagram Figure 4.5. Subroutine **INPUT** reads the mesh data. A separate subroutine is not used to output the results. The results are output as soon as they are obtained. Subroutine **LINKIN** [Hinton and Owen (1977)] links the rest of the program with the solution routine, i.e., it generates all the information needed by the eigenvalue solving routine for matrix manipulation. **GSTIFF** calculates the global stiffness and mass matrices in compacted form. **EIGSOL** [Bathe (1982)] solves the eigenvalue problem using inverse iteration.

A brief description of the auxiliary routines are given below:

Subroutine **NODEXY**

Calculates the coordinates of the midside nodes which lie on a straight line connecting two adjacent midside nodes.

Subroutine **GAUSSQ**

Sets up the sampling point positions and weighting factors for numerical integration. The Gauss quadrature routines utilised in our problem are one- two- and three-point integration rules. The numerical integration scheme used is discussed in Chapter 3.

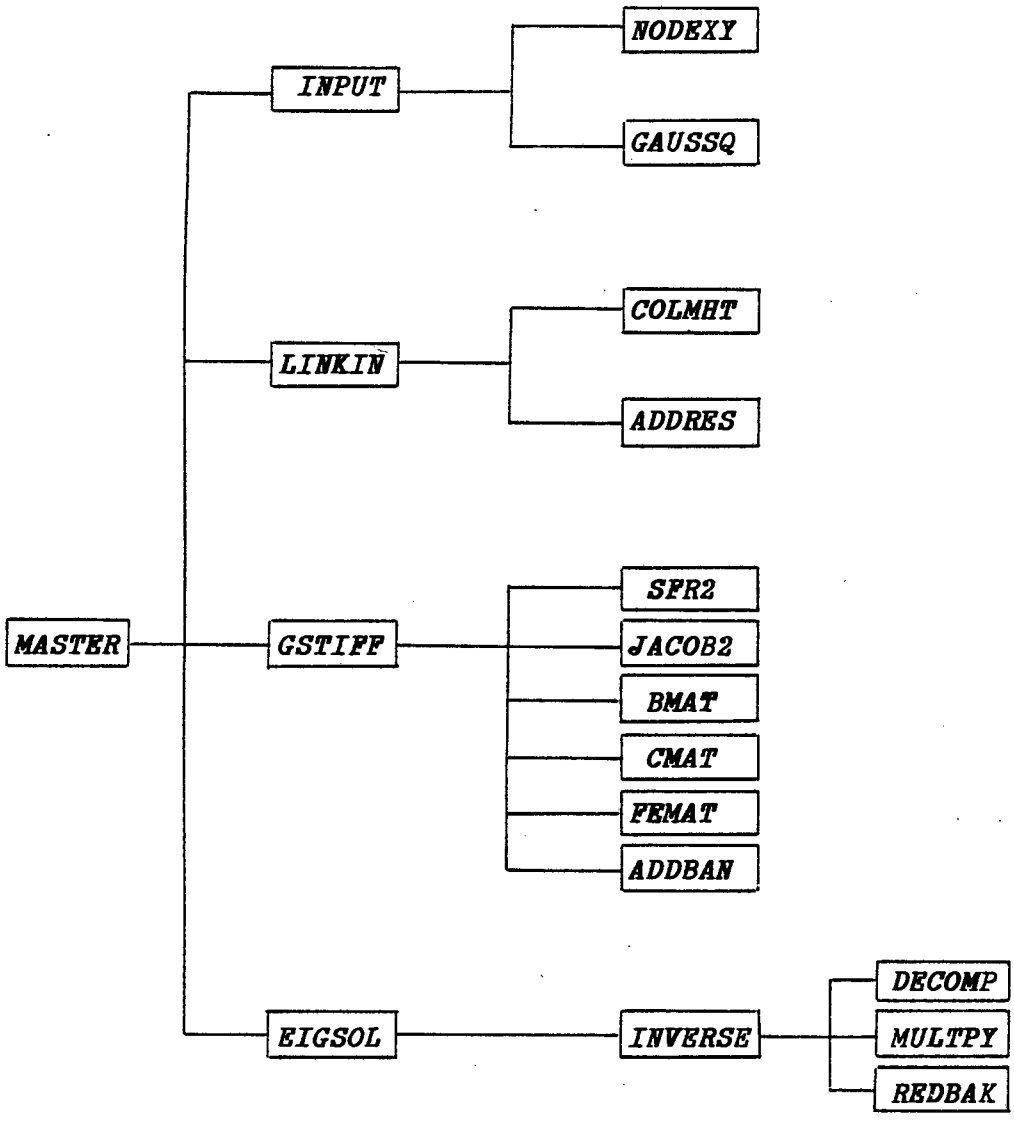

MASTER
INPUT
NODEXY
GAUSSQ
LINKIN
COLMHT
ADDRES
GSTIFF
SFR2
JACOB2
BMAT
CMAT
FEMAT
ADDBAN
EIGSOL
INVERSE
DECOMP
MULTPY
REDBAK

Figure 4.5.

Subroutine ***COLMHT***

Calculates the vertical column heights above the diagonal of the global stiffness and mass matrices.

Subroutine ***ADDRES***

This routine addresses the diagonal entries of the global matrices using the column heights.

Subroutine ***SFR2***

The shape functions and their local derivatives are computed for the two-dimensional problem.

Subroutine ***JACOB2***

The Jacobian matrix and its inverse and the derivatives of the shape functions are computed for the two-dimensional problem.

Subroutine ***BMAT, CMAT, FEMAT***

These subroutines compute the matrices $\underline{B}$, $\underline{C}$, $\underline{F}$ *and* $\underline{E}$ *obtained in Section 3.2*

Subroutine ***ADDBAN***

This routine assembles the element stiffness and mass matrices into the global stiffness and mass matrices in compacted form.

Subroutine ***INVERSE***

This routine solves for the lowest eigenvalue and eigenvector

using inverse iteration.

Subroutine ***DECOMP***

This routine factorises a symmetric positive-definite matrix into lower, diagonal and upper matrices.

Subroutine ***MULTPY***

This routine evaluates the product of a square matrix and an array.

Subroutine ***REDBAK***

This routine solves the equations after the stiffness matrix is decomposed in the form $\boldsymbol{LDL}^T$ *using forward and backward substitution.*

4.2 Subroutine ***INPUT***

The input data required for a finite element analysis consists of the control data and data required to define the geometry of the structure. Subroutine ***INPUT*** *reads and writes the control parameters, nodal point coordinates, element connectivities, boundary conditions and parameter values used in the matrix calculations. In addition it calls the subroutines* ***NODEXY*** *and* ***GAUSSQ****.*

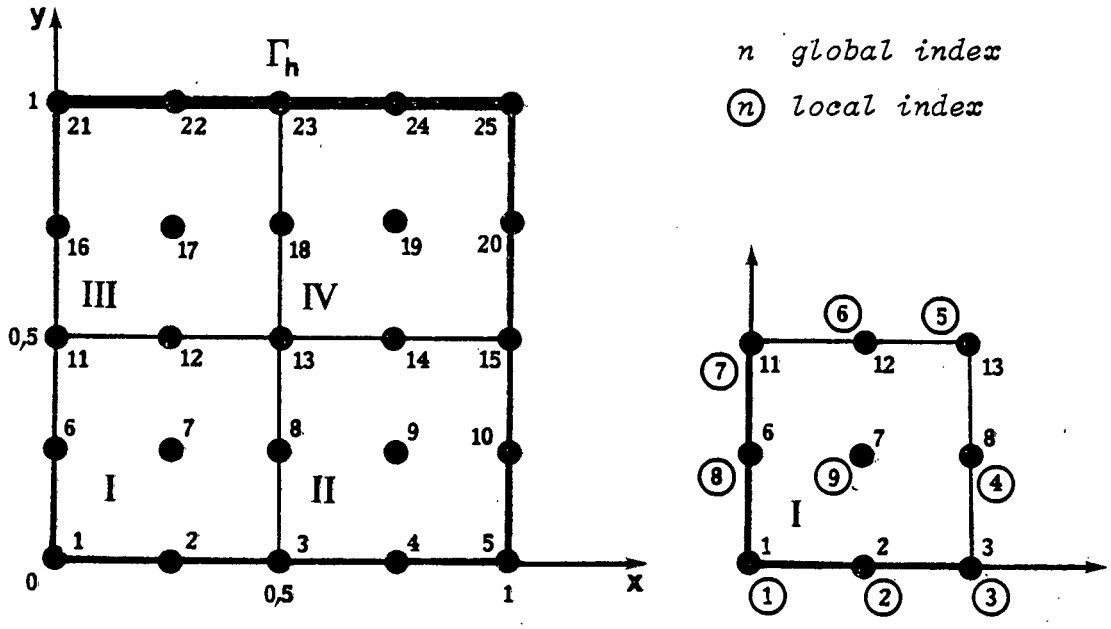


Figure 4.6

(a) Control data:

The following list of control parameters are required as input:

NPOIN	*Total number of nodal points.*
NELEM	*Total number of elements in the domain.*
NVFIX	*Total number of boundary points, i.e., nodal points at which one or more degrees of freedom are restrained.*
NBELEM	*Total number of elements with boundary terms that enter the matrix calculations explicitly.*
ELEM	*Typical length of an element in the mesh.*

For example, *NPOIN = 25 ,*
NELEM = 4 ,
NVFIX = 16 ,
NBELEM = 2 and
ELEM = 0.5

in Figure 4.6.

Once the domain Ω has been discretised into a number of finite elements, the geometry must be defined numerically. Each node is identified by prescribing a number to each nodal point as in Figure 4.6. The nodal points are numbered in an anticlockwise sequence beginning at a corner node. Each element in the mesh is numbered in the order in which it is intended to be processed.

(b) Specification of element connections:

The geometry of each element must be specified by listing in a systematic way the numbers of the nodal points which define the element. The element numbering is read into the array

LNODS(NUMEL,INODE)

where NUMEL, (NUMEL = 1,NELEM) is the number of the element under consideration and
INODE (INODE = 1, 9) runs over all nodes in the element (9 nodes in our case).

LNODS will have the following entries for the example given in Figure 4.6.

element I:	1	2	3	8	13	12	11	6	7 ,
element II:	3	4	5	10	15	14	13	8	9 ,
element III:	11	12	13	18	23	22	21	16	17
and									
element IV	13	14	15	20	25	24	23	18	19 .

(c) Specification of the spatial coordinates of each node:

The coordinates of each nodal point defined in the local (x-y) coordinate system are read into the array

COORD(IPOIN,IDIME)

where IPOIN, (IPOIN = 1, NPOIN) corresponds to the number of the nodal point and

IDIME, (IDIME = 1, 2) refers to the x-y components.

(d) Boundary conditions

The nodes at which one or more degrees of freedom are restrained are read into the array

IFPRE(IDOFN,IPOIN)

where IDOFN, (IDOFN = 1, 3) ranges over the number of degrees of freedom per node (in our case u,v and θ).

A unit value in the relevant column indicates a fixed degree of freedom whereas a zero entry indicates no restraint of that particular component. Thus, IFPRE may have the following values.

1 1 1	*u, v and θ restrained,*
1 1 0	*u and v restrained,*
0 1 1	*v and θ restrained and*
0 1 0	*normal component of velocity restrained.*

Nodal points on which natural boundary conditions hold that enter the matrix calculations explicitly are input, along with any prescribed parameter associated with that boundary, in the array

IBNODS(NUMEL,JNODE)

where NUMEL, (NUMEL = 1, NBELEM) is the number of the element

which side coincides with the boundary under consideration and

JNODE, (JNODE = 1, 3) runs over the nodes that lie on that part of the boundary

Elements which have sides that coincide with the particular boundary are recognized by entering the parameter value associated with the natural boundary condition into the array

BELEM(NUMEL)

For example, in the eigenvalue problems obtained in Chapter 2 we have a boundary term of the form $\int h_T \theta \phi$ *. Let* $\int h_T \theta \phi$ *be non-zero on the boundary* Γ_h *in Figure 4.6, then the input data have the form*

3	*21*	*22*	*23*	h_T
4	*23*	*24*	*25*	h_T .

(d) <u>*Parameters*</u>

The various parameters that entered the formulation of the problem in Chapter 2 must be initialized. These are:

ε	*the penalty parameter,*
λ	*the coupling constant,*
H_s	*the heat-source parameter (see Chapter 5) and*
g	*the gravitational acceleration.*

(e) <u>*Integration points and weighting factors*</u>

The integration (or Gauss) points, defined in Section 3.3, for

one-, two-, and three-point integration are read into the arrays

POSG1(IGAUS) , POSG2(IGAUS) and POSG3(IGAUS)

respectively, where IGAUS = 1 for one-point integration,
IGAUS = 1, 2 for two-point integration and
IGAUS = 1, 2, 3 for three-point integration.

The weighting factors for one-, two-, and three-point integration are read into the arrays

WEIG1(IGAUS) , WEIG2(IGAUS) and WEIG3(IGAUS)

respectively. The integration points and weighting factors are given in Table 3.1.

*4.3 Subroutine **LINKIN***

*This routine calculates the equation numbers from the array IFPRE which stores the information about the restrained degrees of freedom. It calls **COLMHT** which calculates the vertical column heights above the diagonal of the global matrices using equation numbers and the total number of degrees of freedom of an element. It also calls **ADDRES** which addresses the diagonal elements of the global matrices using the column heights.*

(a) calculate equation numbers

Equation numbers are assigned to the IFPRE vector which have zero entries. If IFPRE is non-zero then IFPRE is reassigned as zero. The total number of equations, NEQNS, is equal to the number of unrestrained degrees of freedom.

(b) Evaluate the connectivity array

Equation numbers corresponding to each degree of freedom at each node in the element are assigned to the vector

LEQNS(IEVAB, IELEM)

where IEVAB, (IEVAB = 1, 27) runs over all the degrees of freedom in each element in the mesh. (In our case, 3 degrees of freedom at each of the 9 nodes).

(c) Calculate the column heights

Subroutine ***COLMHT*** *calculates the column heights above the diagonal of the global mass and stiffness matrices using the connectivity array, LEQNS, and stores it in the array*

MHIGH(IEQNS)

where IEQNS is the number of the equation under consideration.

(d) Address diagonal elements

Subroutine ***ADDRES*** *assigns the locations of the diagonal entries in the global stiffness and mass matrices, using the column heights, to the array*

MAXAI(IEQNS)

The total number of entries under the skyline of the global matrices are stored in NWKTL.

4.4 Subroutine GSTIFF

This routine generates the compacted stiffness and mass matrices from the element stiffness and mass matrices defined in Section 3.2. The element stiffness matrix is symmetric for both the linear and energy problems. The element mass matrix is symmetric for the energy problem but this is not always the case for the linear problem. The stiffness and mass matrices are both square and of size NEVAB × NEVAB where NEVAB = 27 for a 9-noded element with 3 degrees of freedom per node.

*Subroutine **GSTIFF** calls **BMAT**, **CMAT** and **FEMAT** which calculate the arrays **B**, **C**, **F** and **E**. It also calls **SFR2** and **JACOB2** which calculate the shape functions, their derivatives and the Jacobian matrix. After the element matrices are calculated, **ADDBAN** is called which assembles the element matrices into the global matrices in compacted form. The structure of **GSTIFF** is illustrated in the flow diagram Figure 4.7.*

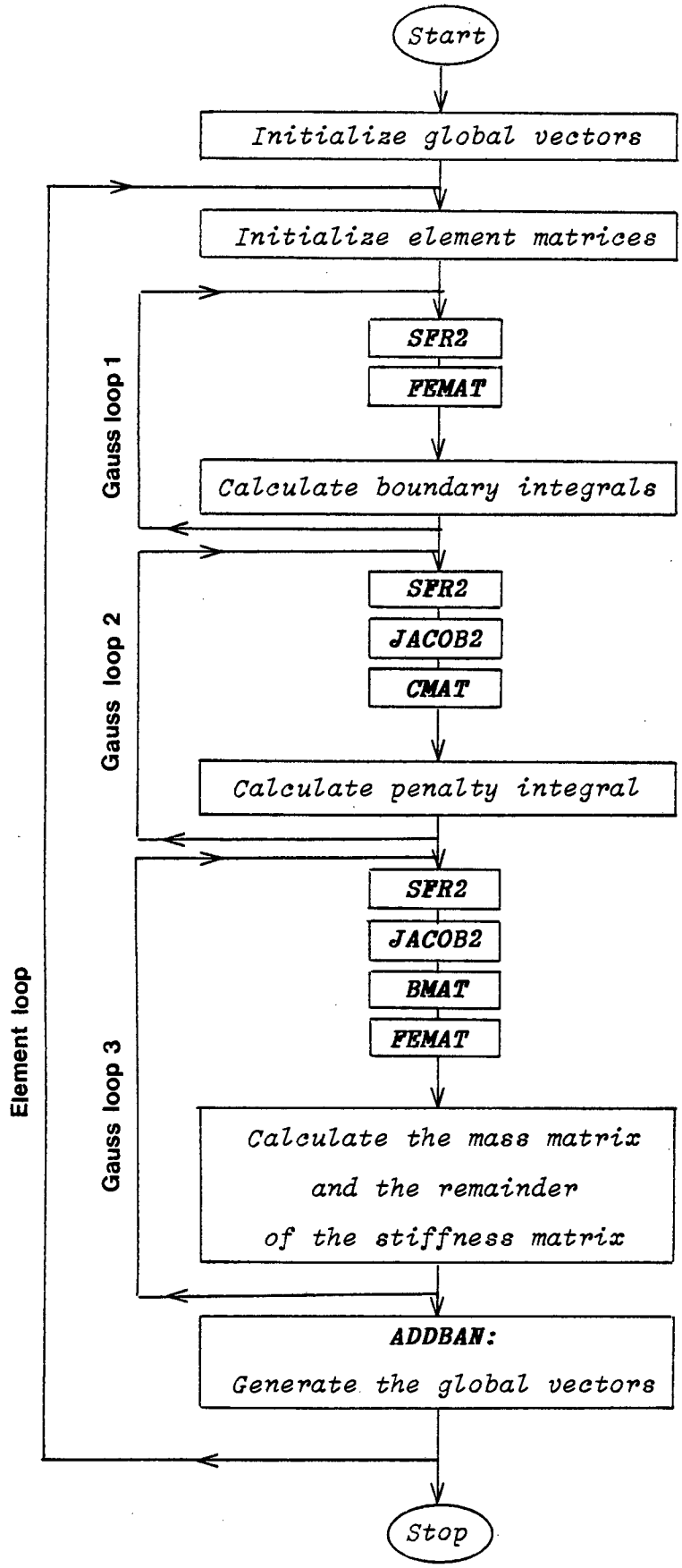

Start
Initialize global vectors
Initialize element matrices
SFR2
FEMAT
Calculate boundary integrals
SFR2
JACOB2
CMAT
Calculate penalty integral
SFR2
JACOB2
BMAT
FEMAT
Calculate the mass matrix
and the remainder
of the stiffness matrix
ADDBAN:
Generate the global vectors
Stop
Gauss loop 1
Gauss loop 2
Gauss loop 3
Element loop

Figure 4.7

(a) Shape functions and their derivatives

Subroutine ***SFR2*** *calculates the shape functions, as defined in (3.12), and their derivatives for a two-dimensional element. The shape functions associated with the nodes of the element under consideration sampled at any Gauss point within the element are stored in the vector*

SHAPE(INODE)

where INODE, (INODE = 1, 9) an element with nine nodes.

The derivatives of the shape functions with respect to the natural coordinate system are stored in

DERIV(INODE,IDIME)

where IDIME, (IDIME = 1, 2) for a two-dimensional element.

(b) Jacobian and cartesian shape function derivatives

Subroutine ***JACOB2*** *calculates the Jacobian matrix, J, defined in (3.17) and its inverse, J^{-1}. The cartesian shape function derivatives associated with the nodes of the current element sampled at any Gauss point within the element are calculated using (3.18) and stored in the array*

CARTD(IDIME,INODE)

(c) Calculation of $\underline{B}$ *,* $\underline{C}$ *,* $\underline{F}$ *and* $\underline{E}$

Subroutine ***BMAT*** *calculates* $\underline{B}$ *and* $\underline{B}^T$ *defined in (3.24), using the cartesian shape function derivatives, and stores them in*

BMATX(ISTRE,IEVAB) *and* *BMATT(IEVAB,ISTRE)*

respectively,
where ISTRE, (ISTRE = 1, 5) and
IEVAB, (IEVAB = 1, 27) runs over all the degrees of freedom in the element.

Subroutine ***CMAT*** *calculates the penalty matrix* $\underline{C}$ *defined in (3.27), using the cartesian shape function derivatives, and stores it in*

CMATX(IEVAB)

Subroutine ***FEMAT*** *calculates the vectors* $\underline{F}$ *and* $\underline{E}$ *defined in (3.28), using the shape functions, and stores them in*

FMATX(IEVAB) *and* *EMATX(IEVAB)* .

(d) *Evaluation of element stiffness and mass matrices*

The compacted element matrices are evaluated numerically using Gauss quadrature, as discussed in Section 3.3. The sampling positions and weighting factors are set up in subroutine ***GAUSSQ*** *for one-, two- and three-point integration.*

The finite element stiffness matrix for the element under consideration is stored in the array

ESTIF(IEVAB,JEVAB)

where IEVAB, (IEVAB = 1, 27) and
JEVAB, (JEVAB = 1, 27) for a nine-noded element with three

degrees of freedom per node .

The finite element mass matrix for the current element is stored in

EMASS(IEVAB,JEVAB) .

(e) Generation of global matrices

Subroutine ***ADDBAN*** *assembles the upper triangle of an element matrix into the global matrix in compacted form. This routine uses the information generated by subroutine* ***LINKIN*** *to assemble the global matrix.*

The upper triangle of the element stiffness matrix is assembled into the one-dimensional array

STIFF(IWKTL)

where IWKTL, (IWKTL = 1, NWKTL) is equal to the current equation number.

For the case when the element mass matrix is symmetric (energy problem), the upper triangle is assembled into the one-dimensional array

MASS(IWKTL) .

When the element mass matrix is not symmetric (linear problem), the upper triangle is assembled in

MASSU(IWKTL)

and the lower triangle is assembled into the array

MASSL(IWKTL) .

4.5 Subroutine ***EIGSOL***

This routine : (i) reads the following parameters

ISTAR — *an output flag and*

μ — *the shift (see (b) below),*

and initializes various parameters,

(ii) calls subroutine ***INVERSE*** *which solves for the lowest eigenvalue and corresponding eigenvector of the problem*

$$\underline{K}\,\underline{a} = R\ \underline{M}\,\underline{a} \qquad (4.3)$$

where $\underline{K}$ *is symmetric and positive definite, and*

(iii) prints the lowest eigenvalue and corresponding eigenvector if they exist.

(a) Subroutine ***INVRSE***

The inverse iteration method, described in Section 3.4, is an efficient and simple method to calculate the lowest eigenvalue and corresponding eigenvector. A shift is used in inverse iteration when the stiffness matrix, $\underline{K}$*, is positive definite and the mass matrix,* $\underline{M}$ *has zero diagonal elements. Subroutine* ***INVRSE*** *calls* ***DECOMP*** *which factorises the symmetric, positive definite stiffness matrix into the form* $\underline{L}\underline{D}\underline{L}^T$*.* ***REDBAK*** *is called in every*

iteration to solve the equations using forward and backward substitution. ***MULTPY*** *is also called in every iteration to evaluate the product of a vector and the mass matrix. The structure of* ***INVRSE*** *is illustrated in the flow diagram Figure 4.8.*

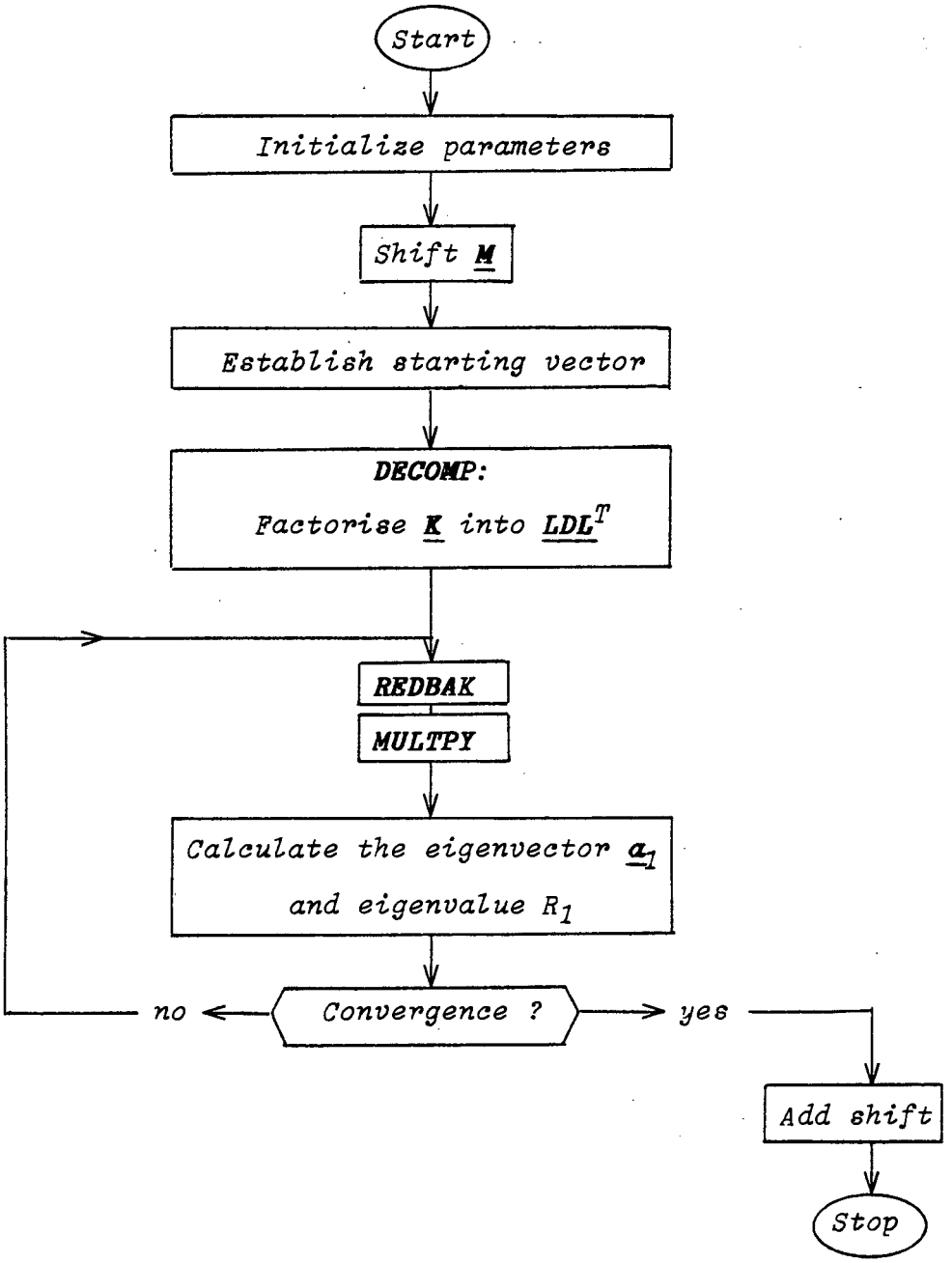


Figure 4.8

(b) Shifting

In the solution of (4.3) where $\underline{\underline{M}}$ *has zero diagonal elements, we perform a shift on* $\underline{\underline{M}}$ *to remedy the situation. The shifted eigenvalue problem has the form*

$$\underline{\underline{K}}\ \underline{a}' = R'\ (\underline{\underline{M}}\ \underline{a}' + \mu\ \underline{\underline{K}}\ \underline{a}') \qquad (4.4)$$

where R' *is the eigenvalue corresponding to the shifted problem, and* $\underline{a}'$ *is the corresponding eigenvector.*

To identify how the eigenvalue and eigenvector of (4.3) are related to that of problem (4.4), we rewrite (4.4) in the form

$$\underline{\underline{K}}\ \underline{a}' = [\ R'/(1 - \mu\ R')]\ \underline{\underline{M}}\ \underline{a}' \qquad (4.5)$$

which is equivalent to the eigenvalue problem (4.3) if we put

$$R = [\ R'/(1 - \mu\ R')]\ . \qquad (4.6)$$

The shift does not affect the eigenvector $\underline{a}$*, so that* $\underline{a}' = \underline{a}$*, because the shift causes a change in size, not direction.*

(c) Factorisation of stiffness matrix

In subroutine ***DECOMP*** *the stiffness matrix* $\underline{\underline{K}}$*, stored in the one-dimensional array STIFF, is factorised into lower , diagonal and upper matrices using Gauss reduction. The* $\underline{\underline{LDL}}^T$ *matrices are stored in the array*

STIFF(IWKTL) .

An error message is output if the stiffness matrix is not

positive definite.

(d) Multiplication of mass matrix and a vector: ($\underline{y}_{k+1} = \underline{\underline{M}}\underline{x}_{k+1}$)

Subroutine ***MULTPY*** *evaluates the product of the square symmetric matrix, MASS, and an iteration vector and stores the resulting vector in*

FINAL(IEQNS)

where IEQNS, (IEQNS = 1, NEQNS) runs over all the equation numbers.

If the mass matrix is not symmetric (linear stability problem), then subroutine ***MULTPY*** *evaluates the product of (MASSU + MASSL) and the iteration vector and stores the result in FINAL.*

(e) Solution of equations: ($\underline{\underline{K}}\underline{x}_{k+1} = \underline{y}_k$)

Subroutine ***REDBAK*** *solves the equations, after the stiffness matrix* $\underline{\underline{K}}$ *has been decomposed. The matrix STIFF and the iteration vector FORCE is input and the solution vector is stored in*

FORCE(IEQNS) .

4.6 Sample input

The input for a simple problem is given in this Section. The domain Ω *in* $\mathbb{R}^2$ *is divided into four nine-noded elements, as illustrated in Figure 4.9. At each nodal point there are three*

degrees of freedom, u, v and θ . Assume the following boundary conditions :

$$\underline{u} = \underline{0} \quad \text{on} \quad \Gamma,$$

$$\theta = 0 \quad \text{on} \quad \Gamma_1 \cup \Gamma_3 \quad , \text{ and}$$

$$\partial\theta/\partial n = 0 \quad \text{on} \quad \Gamma_2 \cup \Gamma_4 \ .$$

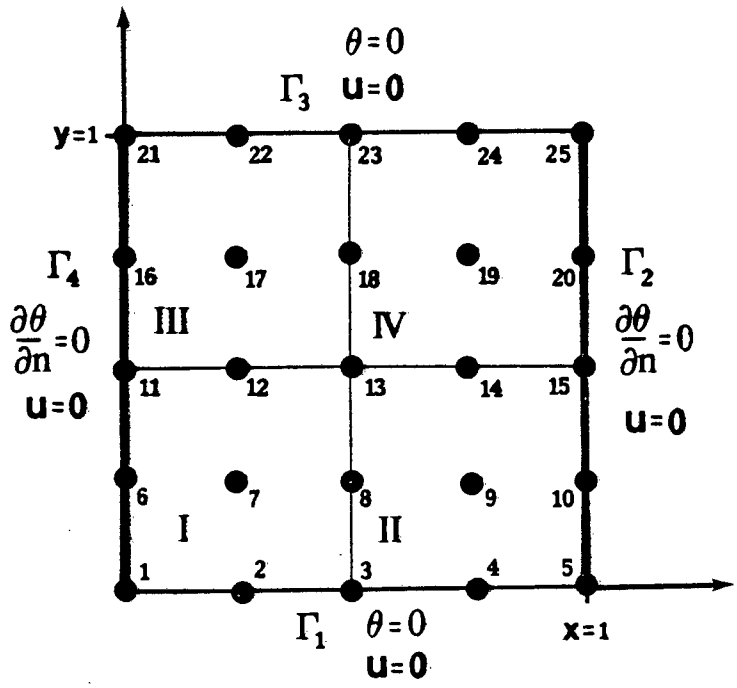


Figure 4.9

The data is read in free format and values may be separated by a space(s) or a comma. The input for the above situation would be the following:

DATA LINE 1. Control data - One line containing

NPOIN NELEM NVFIX NBELEM ELEM

25	4	16	4	0.5

The first four values are integers, the last real.

DATA SET 2. Element Connections - One line for each element, total of NELEM lines. The node numbers are read in an anticlockwise sequence.

NUMEL	*LNODS(ELEM,1)*	. . . . .	*LNODS(IELEM,9)*

1	*1*	*2*	*3*	*8*	*13*	*12*	*11*	*6*	*7*
2	*3*	*4*	*5*	*10*	*15*	*14*	*13*	*8*	*9*
3	*11*	*12*	*13*	*18*	*23*	*22*	*21*	*16*	*17*
4	*13*	*14*	*15*	*20*	*25*	*24*	*23*	*18*	*19*

The element number and nodal point numbers are all integer.

DATA SET 3. Nodal point coordinates - One line for each node. The last node (IPOIN = NPOIN) is read at the end.

IPOIN	*COORD(IPOIN,1)*	*COORD(IPOIN,2)*

1	*0.0*	*0.0*
2	*0.25*	*0.0*
3	*0.5*	*0.0*
.	.	.
:	:	:
25	*1.0*	*1.0*

The x and y coordinates are real.

DATA SET 4. Restrained nodes. One line for each restrained node, total of NVFIX lines.

IPOIN	IFPRE(IVFIX,1)	IFPRE(IVFIX,2)	IFPRE(IVFIX,3)
1	1	1	0
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	0
6	1	1	0
10	1	1	0
.	.	.	.
:	:	:	:
25	1	1	0

All values must be integers.

DATA SET 5. Unrestrained boundary nodes - One line for each element containing unrestrained boundary nodes, total of NBELEM lines.

NUMEL	IBNODS(NUMEL,1)	IBNODS(NUMEL,2)	IBNODS(NUMEL,3)	HNUS
1	1	6	11	0.0
2	5	10	15	0.0
3	11	16	21	0.0
4	15	20	25	0.0

All values are integers, except for the Nusselt number HNUS. If BELEM = 0 then data set 5 is omitted.

DATA LINE 6. Parameter values - One line containing

ε	λ	H_s	g

where all the parameter values are real.

DATA LINE 7. Parameter values - One line containing

ISTAR	μ

ISTAR is an integer which contains 0 on entry and is set to 1 if additional output is required. The shift μ *is a real value.*

We consider a rectangular container of stationary fluid heated from below and internally. It is assumed that the extension in the z-direction is sufficiently large so that the three-dimensional problem can be reduced to one of two dimensions, in the x-y plane, as shown in Figure 5.1. The width and depth of the layer are denoted by l and d respectively.

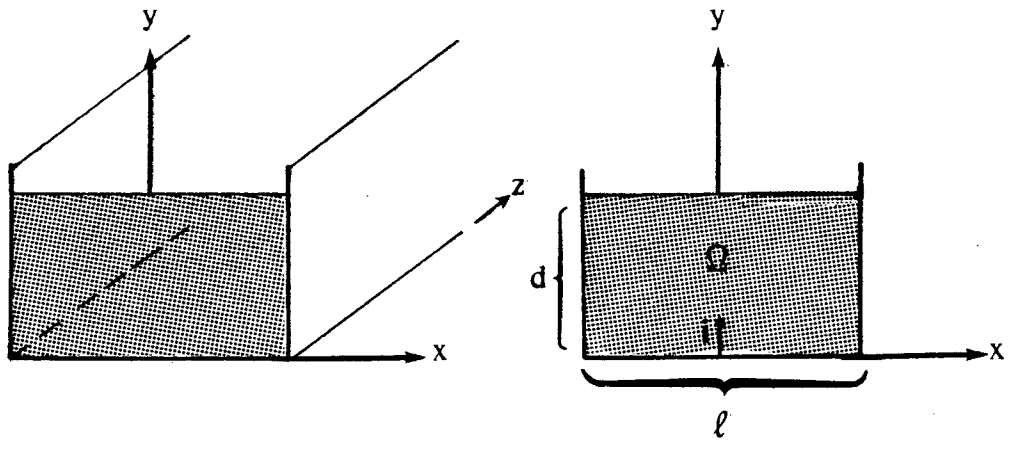


Figure 5.1

Let the unit vector $\underline{i}$ point in the direction of y increasing. We consider situations in which the temperature gradient, ∇T, and the dimensionless gravity field, $\underline{g}$ are parallel vectors. For our problem we take $\underline{g}$ as a constant gravity field pointing "down", i.e. in the direction of $-\underline{i}$. Thus, we have

$$\nabla T = \underline{i}\, {}^{dT}/_{dy} \quad \text{and} \quad \underline{g} = -\underline{i}$$

(recall that $\underline{g}$ is dimensionless). As in Sparrow, Goldstein and Jonsson (1964) we consider a quadratic temperature distribution which depends only on y, and which may be written as

$$T = -{}^{1}/_{2}\,({}^{s}/_{\kappa})\,y^2 + Ay + B \qquad (5.1)$$

where s is the internal heat-source intensity and

κ is the thermal conductivity.

Equation (5.1) in dimensionless variables is given by

$$T'T = -{}^{1}/_{2}\,({}^{s}/_{\kappa})\,d^2y^2 + Ady + B \qquad (5.2)$$

where (dimensionless T) $= {}^{T}/_{T'}$ and

(dimensionless y) $= {}^{y}/_{d}$;

T' is the characteristic temperature, which will be chosen shortly.

Let T_1 and T_2 be the temperature at the bottom and top of the layer of fluid respectively, with $T_1 > T_2$. Then

$$B = T_1$$

and

$$A = [T_2 - T_1 + {}^{1}/_{2}({}^{s}/_{\kappa})d^2]\,{}^{1}/_{d}$$

Substitution of the above into equation (5.2) results in

$$T'T = -{}^{1}/_{2}({}^{s}/_{\kappa})d^2y^2 + [T_2 - T_1 + {}^{1}/_{2}({}^{s}/_{\kappa})d^2]\,y + T_1 \qquad (5.3)$$

If we set

$$H_s = {}^{1}/_{2}({}^{s}/_{\kappa})d^2/(T_1 - T_2)$$

where $H_s > 0$ is the heat-source parameter, then the derivative of (5.3) with respect to y is given by

$$T'\,{}^{dT}/_{dy} = (T_1 - T_2)\,[H_s\,(1 - 2y) - 1] \qquad (5.4)$$

We choose T' such that

$$T' = (T_1 - T_2)[H_s + 1]$$

to obtain

$$\frac{dT}{dy} = \frac{[H_s(1 - 2y) - 1]}{[H_s + 1]} \qquad (5.5)$$

Thus

$$(\underline{g}.\underline{i}) = -1 \qquad (5.6)$$

and

$$(\nabla T.\underline{i}) = \frac{[H_s(1 - 2y) - 1]}{[H_s + 1]} \qquad (5.7)$$

in the expressions for the eigenvalue problems for the linear and energy stability theory obtained in Chapter 3.

In Section 5.1 we obtain results for the Bénard problem (H_s=0) and we determine how accurate the finite element method approximates the solutions to the penalised eigenvalue problems obtained in Chapter 2. We further use the Bénard problem to show that the solutions to the penalised problem converge to a solution of the unpenalised problem as the penalty paramater $\varepsilon \rightarrow 0$. Once we have found a suitable working mesh size and penalty parameter we study the Bénard problem for a variety of width-to-height ratios ($^{l}/_{d}$) and boundary conditions.

The stability problems for the linear and energy theory are studied in Section 5.2 for a fluid heated from below and

internally. We first determine how the linear stability limit R_L varies with the heat-source parameter H_s for different width-to-height ratios. Stability limits for the energy problem, ρ_λ, are obtained for different H_s and coupling parameters λ. The optimal stability limit, R_E, for any heat-source parameter H_s is given by the maximum of ρ_λ over λ. Critical stability limits for the linear and energy problems, R_L and R_E, are compared to determine the region for possible subcritical instabilities. Most results are illustrated graphically and compared to published results where appropriate.

5.1 The Bénard problem

In the Bénard problem the temperature distribution of the motionless state is taken as linearly decreasing with height (no internal heat sources present) so that

$$\nabla T = \underline{g} = -\underline{i}$$

The critical stability limits, R_L and R_E, for the linear and energy stability theory coincide. If the Rayleigh number, R, is less that the critical stability limit there is no flow and we have global stability. If R is greater than the critical value we have instability.

In Chapter 2 we formulated the penalised eigenvalue problems for the linear and energy stability theory in the form: find $\bar{u}_\varepsilon \in \bar{V}$ and $R_\varepsilon \in \mathbb{R}$ satisfying

$$A(\bar{u}_\varepsilon, \bar{v}) = R_\varepsilon B(\bar{u}_\varepsilon, \bar{v}) \qquad \text{for all } \bar{v} \in \bar{V} . \tag{5.8}$$

We expect that the solutions $(\bar{u}_\varepsilon, p_\varepsilon, R_\varepsilon)$ to the penalised problem will converge to a solution $(\bar{u}, p, R)$ of the unpenalised problem as ε tends to zero, although proof of convergence has not been established. To construct a finite element approximation of (5.8), we constructed a family $\{\bar{V}^h\}$ of finite-dimensional subspaces of $\bar{V}$ using piecewise polynomial basis functions. In the finite element method these basis functions are chosen in such a way that $\bar{V}^h$ approaches $\bar{V}$ as h approaches zero. As the mesh is refined the dimension of the finite element solution space is increased to contain ultimately the solution. Thus, the finite element solution should converge to the solution of the penalised eigenvalue problem as the number of elements is increased.

The finite element approximation to the eigenvalue problem (5.8) has the form

$$(\underline{\boldsymbol{K}} + {}^{1}/_{\varepsilon}\ \underline{\boldsymbol{H}})\ \underline{\boldsymbol{a}} = R_\varepsilon^h\ \underline{\boldsymbol{M}}\ \underline{\boldsymbol{a}} \qquad (5.9)$$

where $\underline{\boldsymbol{H}}$ is the matrix arising from the penalty term,

R_ε^h is the lowest eigenvalue, and

$\underline{\boldsymbol{a}}$ is the corresponding eigenvector.

We need to know how accurate our approximate solution is and how the error in the approximation is affected by the parameters ε and h. In order to obtain information about the error we solve (5.9) for the smallest eigenvalue R_ε^h for various values of ε and h. Let the domain Ω in which (5.9) is solved be the unit square $(0,1)\times(0,1)$, as shown in Figure 5.2. All four walls are rigid and adhesion is assumed so both velocity components, u and v, are zero there. The side walls x=0,1 are taken to be perfect insulators and the upper and lower surfaces are isothermal. The domain is subdivided into a uniform mesh of square elements, so

the mesh parameter is given by

$$h = \{ l/(d*\text{number of elements}) \}^{1/2}$$

where $l = 1 = d$ *for a unit square.*

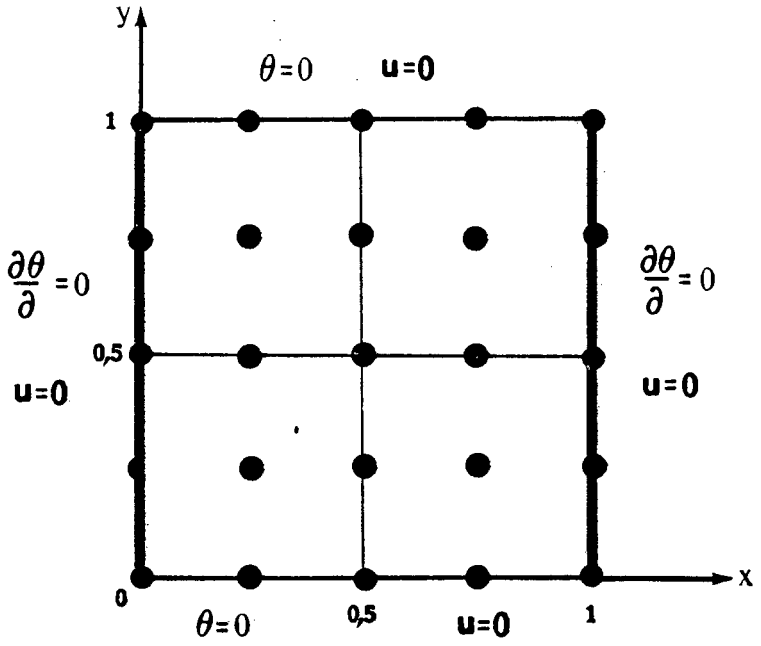


Figure 5.2

The matrices $\underline{K}$ *and* $\underline{M}$ *in (5.9) are evaluated using 3×3 Gauss numerical integration. The penalty matrix,* $\underline{H}$*, should be evaluated with a lower integration than that which is necessary to integrate it exactly [Oden, Kikuchi and Song (1982)], for reasons of stability. Using a nine-noded element the penalty matrix is evaluated using 2×2 Gauss numerical integration which corresponds to a linear interpolation of the pressure.*

Geveci, Reddy and Pearce (1986) obtained estimates of the error for the smallest eigenvalue of the form

$$\left| R_\varepsilon^h - R^h \right| \leq C_1 R^2 \varepsilon \qquad (5.10)$$

and

$$\left| R_\varepsilon^h - R \right| \leq C_2 (R^2 \varepsilon + R^3 h^4) \qquad (5.11)$$

in the penalised eigenvalue problem for the Stokes operator,

using the nine-noded element. Error estimates are not available for our problem but we assume that estimates of the form (5.10) and (5.11) hold in our case, i.e. we assume that estimates of the form

$$\left| R_{\varepsilon}^{h} - R^{h} \right| \leq C(R, \Omega)\varepsilon^{\mu} \qquad (5.12)$$

and

$$\left| R_{\varepsilon}^{h} - R \right| \leq c(R, \Omega)\ (\varepsilon^{\mu} + h^{\beta}) \qquad (5.13)$$

hold, for $\mu > 0$ *and* $\beta > 0$.

Since the actual error cannot be calculated unless the exact solution is known, relative errors are calculated instead. We shall first examine the dependence of the relative error on the penalty parameter ε *as follows: we fix some values of* ε *and* h, $\hat{\varepsilon}$ *and* $\hat{h}$ *say, and obtain the relative error*

$$\left| R_{\hat{\varepsilon}}^{\hat{h}} - R_{\varepsilon}^{\hat{h}} \right| \leq \left| R_{\hat{\varepsilon}}^{\hat{h}} - R^{\hat{h}} \right| + \left| R_{\varepsilon}^{\hat{h}} - R^{\hat{h}} \right|$$

$$\leq C(R, \Omega)\ (\varepsilon^{\mu} + \hat{\varepsilon}^{\mu})$$

$$\simeq C\ \varepsilon^{\mu} \qquad (5.14)$$

using (5.12), where $R^{\hat{h}}$ *is the lowest eigenvalue of (5.9), and assuming that* $\hat{\varepsilon} << \varepsilon$.

Results for $h=0.5$ *and* $\hat{\varepsilon} = 0.00001$ *are given in Table 5.1, and a plot showing the results is given in Figure 5.3. The slope* m *of the line through the data points is a measure of the order of convergence* μ. *If* m *is positive, the relative error approaches zero as* ε *tends to zero. The points in Figure 5.3 appear to indicate that* $\mu = 1$, *i.e. that the linear convergence obtained by Geveci et al. is reproduced in this problem. The*

point corresponding to ε =0.0001 lies far off the straight line, but this is to be expected since ε =0.0001 is very close to $\hat{\varepsilon}$, and the assumption $\hat{\varepsilon} << \varepsilon$ is not valid for this case.

ε	$R_{\varepsilon}^{\hat{h}}$; $\hat{h}$=0.5
0.1	35.10
0.01	49.56
0.001	50.42
0.0001	50.52
0.00001	50.64

Table 5.1

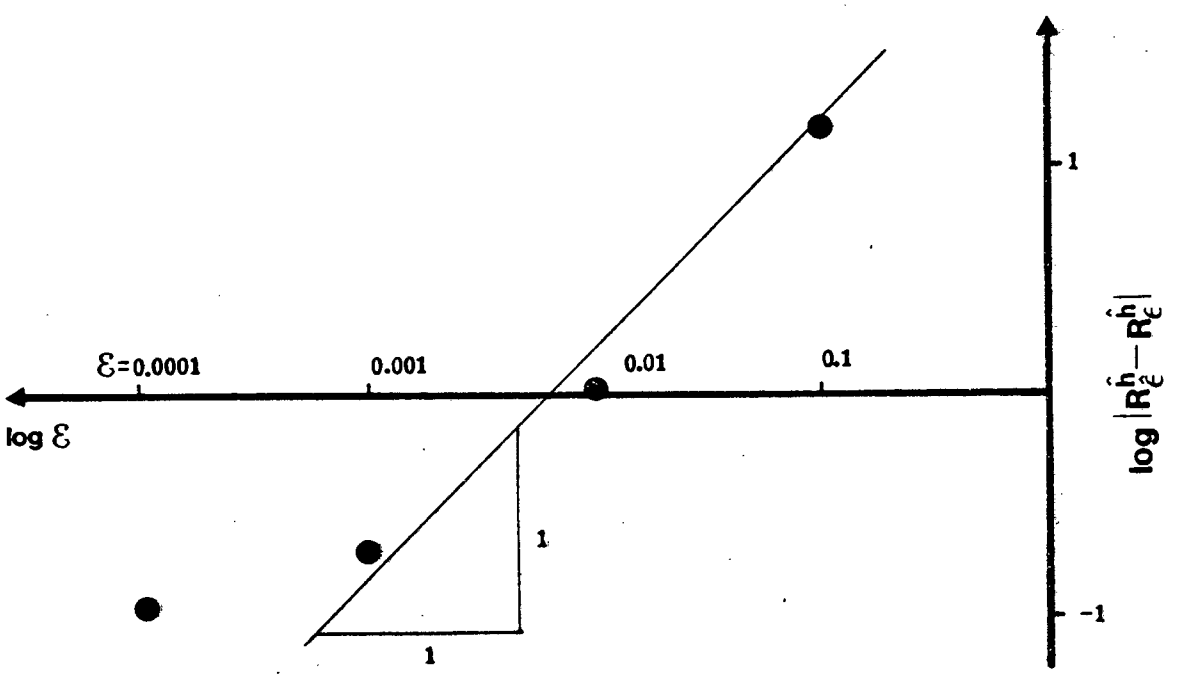


Figure 5.3

Next, we examine convergence as mesh size goes to zero, for fixed $\hat{\varepsilon}$ *. In order to determine the rate at which* R_ε^h *converges to* R_ε *as* h *tends to zero it is necessary to run the analysis with progressively finer meshes until changes in the smallest eigenvalue are sufficiently small. We determine how fast the relative error decreases with* h *by fixing the penalty parameter at a value* $\hat{\varepsilon}$ *and obtaining* $R_{\hat{\varepsilon}}^h$ *for a sequence of values of* h*. Using (5.13) we obtain the relative error*

$$\begin{aligned} \left| R_{\hat{\varepsilon}}^{\hat{h}} - R_{\hat{\varepsilon}}^{h} \right| &\leq \left| R_{\hat{\varepsilon}}^{\hat{h}} - R \right| + \left| R_{\hat{\varepsilon}}^{h} - R \right| \\ &\leq c(R,\Omega)\{(2\hat{\varepsilon}^{\mu} + \hat{h}^{\beta}) + h^{\beta}\} \\ &\simeq c\, h^{\beta} \end{aligned} \tag{5.15}$$

for $\hat{\varepsilon}$ *and* $\hat{h}$ *sufficiently small.*

The results for $\hat{h}=1/8$ *and* $\hat{\varepsilon}=0.0001$ *are given in Table 5.2 and plotted in Figure 5.4. The point in Figure 5.4 corresponding to* $h=1/6$ *may be ignored since* $(1/6)^4$ *is close to* $\hat{h}^4$*. We appear to have convergence of the order* $2/3$*, which does not correspond to the convergence obtained by Geveci et al. for the Stokes problem. Also, since the error estimate is valid only as* $h \to 0$*, points corresponding to* $h=1/2,\ 1/3$ *may not be reliable. This leaves us only two points from which to estimate* β*, which is unsatisfactory. We see nevertheless from both Figures 5.3 and 5.4 that the relative error decreases monotonically with decreasing* ε *and* h*, respectively, and we conclude that* R_ε^h *will ultimately converge to* R *as* ε *and* h *tend to zero.*

h	$R^h_{\hat{\epsilon}}$; $\hat{\epsilon}=0.0001$
1/2	50.52
1/3	50.65
1/4	50.69
1/5	50.78
1/6	50.87
1/8	51.06

Table 5.2

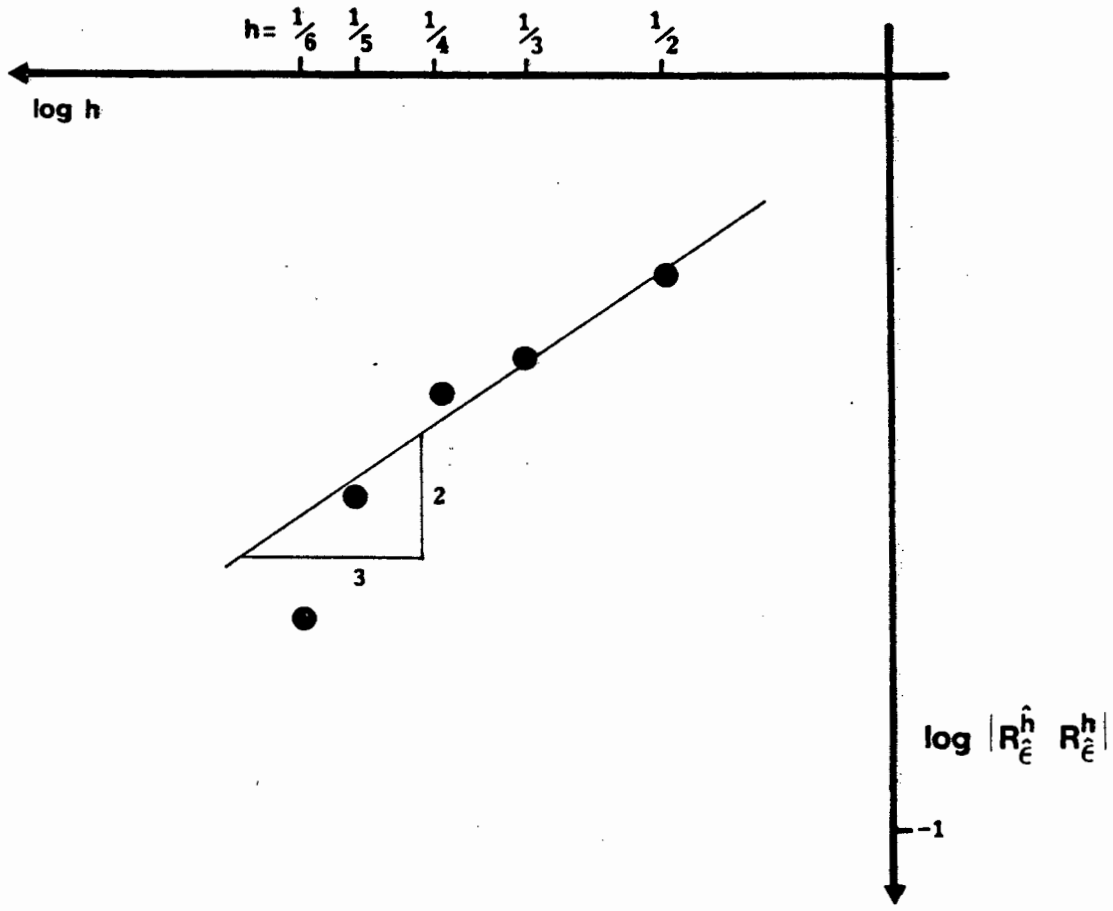


Figure 5.4

The results given in Tables 5.1 and 5.2 can be used to make an acceptable choice of mesh size and penalty parameter for use in the rest of this work. We must choose parameter values ε and h that give accurate results and which economise on computation time. From Table 5.1 we see that a large value of ε leads to inaccurate results. From Table 5.2 we see that a rather coarse mesh (h=0.5) produces relatively accurate results, whereas a fine mesh gives only slightly better results although computation time is increased considerably. We choose ε =0.0001 and h=0.5, as shown in Figure 5.5, which should lead to accurate and efficient solutions.

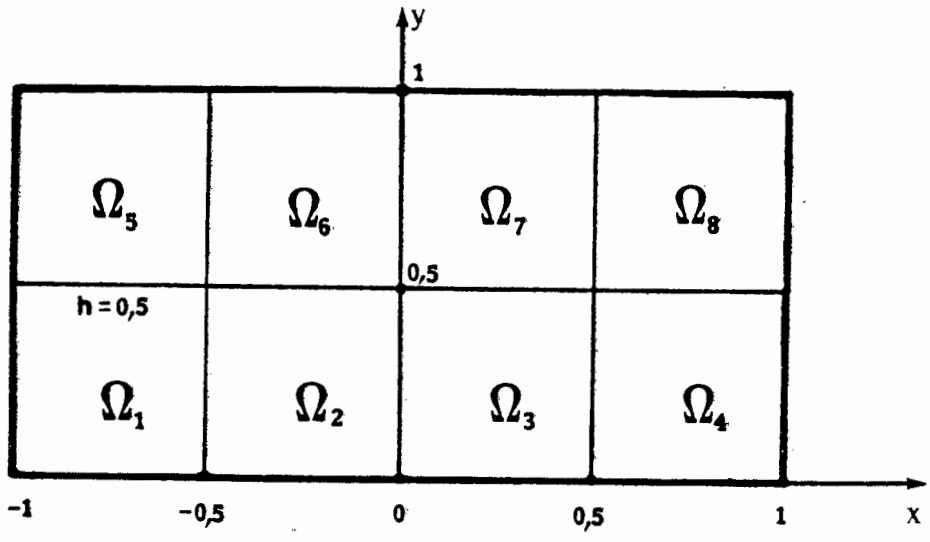


Figure 5.5

Having found suitable working values for ε and h we proceed to study the Bénard problem for various geometries and investigate the dependence of the stability of the motionless state on the temperature and velocity boundary conditions given in Chapter 2. The temperature boundary conditions (2.13) include a fixed temperature, fixed heat flux, and a general convective exchange with the environment at the bounding surface.

Consider first a fluid layer with a fixed temperature at both upper and lower bounding surfaces. We will investigate the following velocity boundary conditions: (a) the upper and lower bounding surfaces are both rigid, (b) the lower surface is rigid while the upper surface is free, and (c) the upper and lower surfaces are free (this condition does not correspond to a real physical situation but may be of theoretical interest).

We require that the side walls be rigid ($\underline{u}=\underline{0}$) and perfect insulators ($\partial\theta/\partial x = 0$) throughout this study.

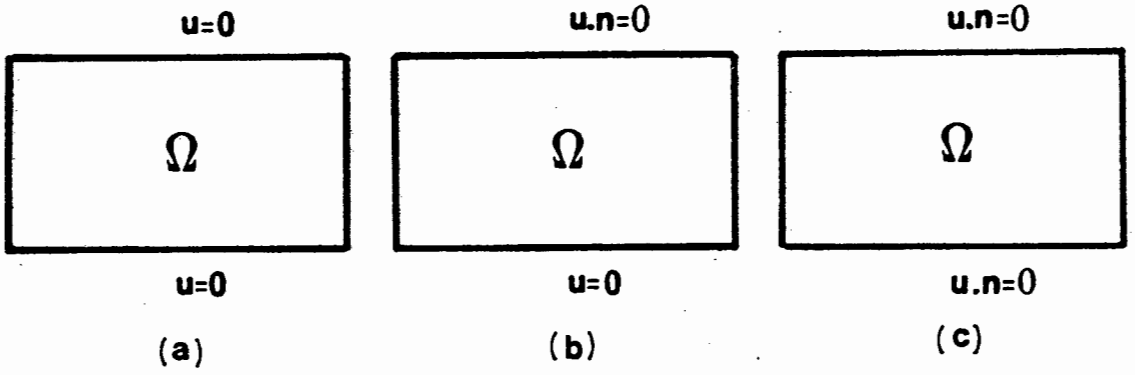


Figure 5.6

The critical Rayleigh numbers for the motionless solution to the Bénard problem with rigid walls on all sides are presented in Table 5.3 (a) for l/d in the range 1 to 10. The results for the rigid-free case are presented in Table 5.3 (b) and for the free-free case in Table 5.3 (c). Results obtained by Hall and Walton (1976), Jackson and Winters (1984), and Sparrow, Goldstein and Jonsson (1964) are presented for comparison. Jackson and Winters computed critical Rayleigh numbers using the finite element

method with six-node quadratic triangles to model velocities and temperature, and three-node linear triangles to model pressure. A mesh size of h=0.25 was used compared to h=0.5 in our calculations. Critical Rayleigh numbers were obtained analytically by Hall and Walton for the free-free case and various geometries. Results obtained by Sparrow et al. include rigid-rigid and rigid-free horizontal bounding surfaces but are for the case of an infinitely long container.

Ratio	R_{crit}	Published R_{crit}
1	50.52	51.11 *
2	46.55	45.24 *
4	44.97	42.94 *
10	43.81	
∞		41.33 #

Table 5.3 (a): Rigid-rigid case.

Ratio	R_{crit}	Published R_{crit}
1	43.87	45.62 *
2	38.45	38.54 *
4	35.24	34.58 *
10	34.44	
∞		33.18 #

Table 5.3 (b): Rigid-free case.

Ratio	R_{crit}	Published R_{crit}
1	39.30	41.29 *
2	29.44	29.90 *
		29.75 "
4	27.19	27.17 *
		27.09 "
10	26.24	25.91 "

Table 5.3 (c): Free-free case.

* *Jackson and Winters (1984)*

\# *Sparrow, Goldstein and Jonsson (1964)*

" *Hall and Walton (1976)*

The critical Rayleigh numbers for onset of convection decrease monotonically as the width increases, and we expect that R_{crit} *will ultimately converge to the critical value obtained by Sparrow et al. (1964) for the case of an infinitely long channel. Plots of critical Rayleigh number versus width-to-height ratio for the rigid-rigid, rigid-free and free-free cases are given in Figure 5.7 (a), (b) and (c). Comparison of our results with those of others gets progressively better as the velocity boundary conditions are relaxed. This is probably as a result of the coarse mesh used (h=0.5).*

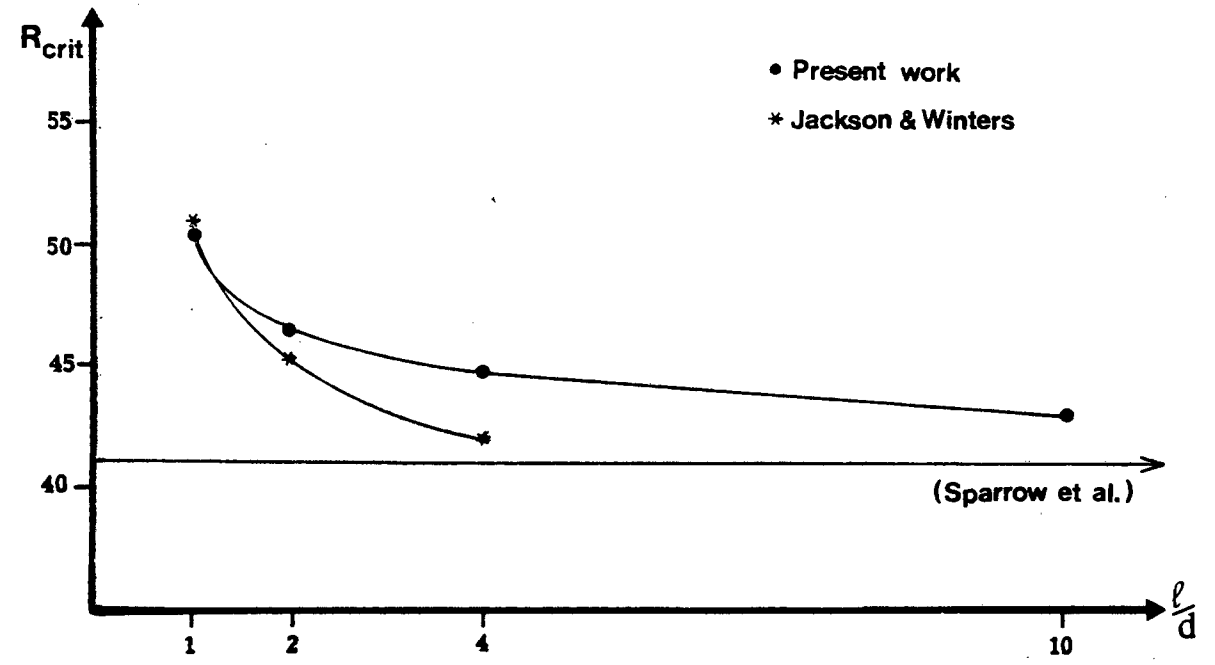


Figure 5.7 (a): Rigid-rigid case.

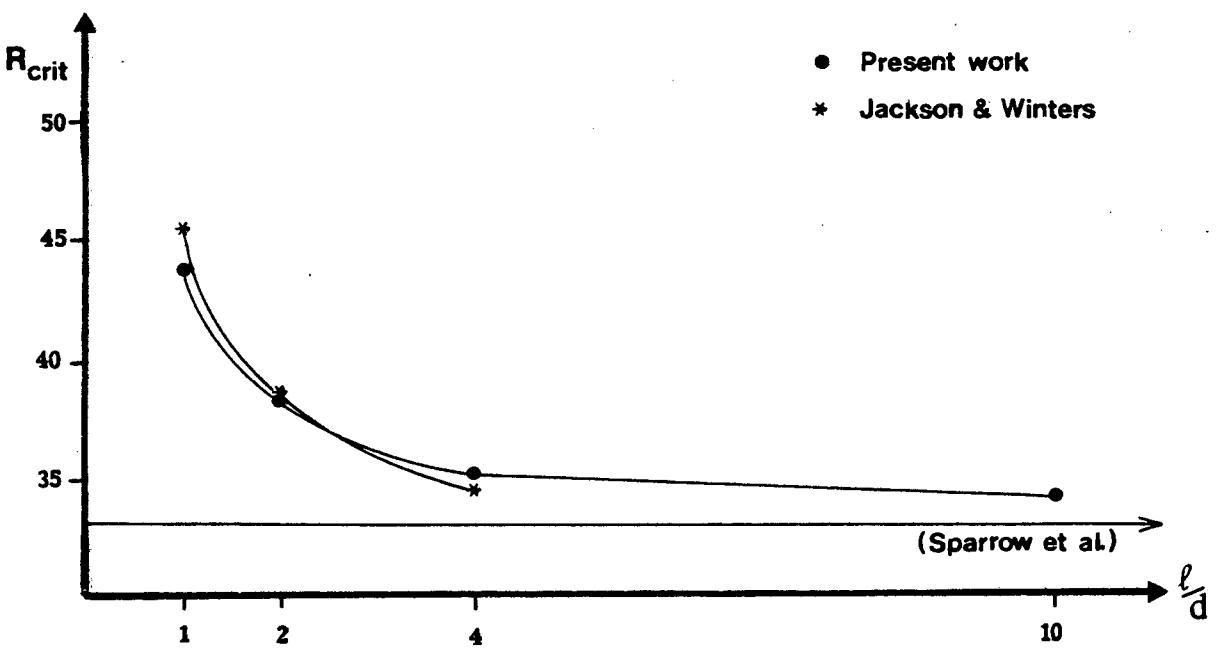


Figure 5.7 (b): Rigid-free case.

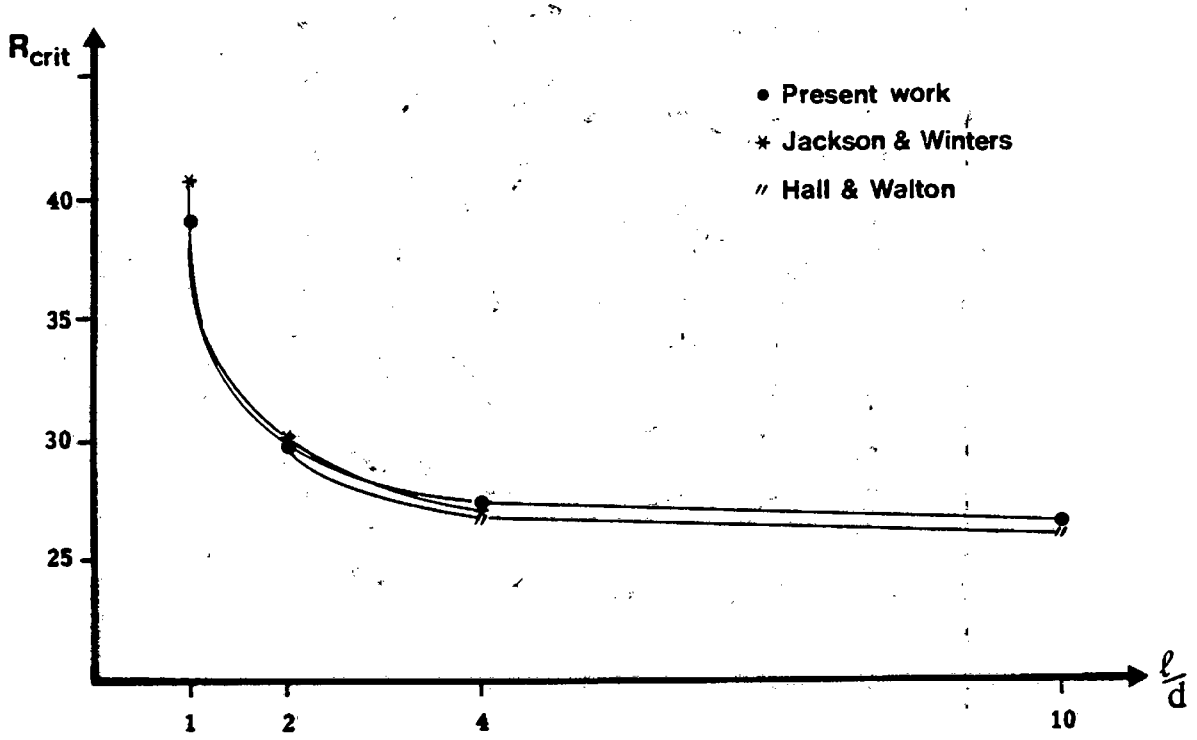


Figure 5.7 (c): Free-free case.

At the critical Rayleigh number convection occurs and the motion takes the form of cells. Figure 5.8 shows the velocity vector field for $^{l}/_{d}$ = 1 and 2. Figure 5.9 shows the isotherms corresponding to $^{l}/_{d}$ = 1.

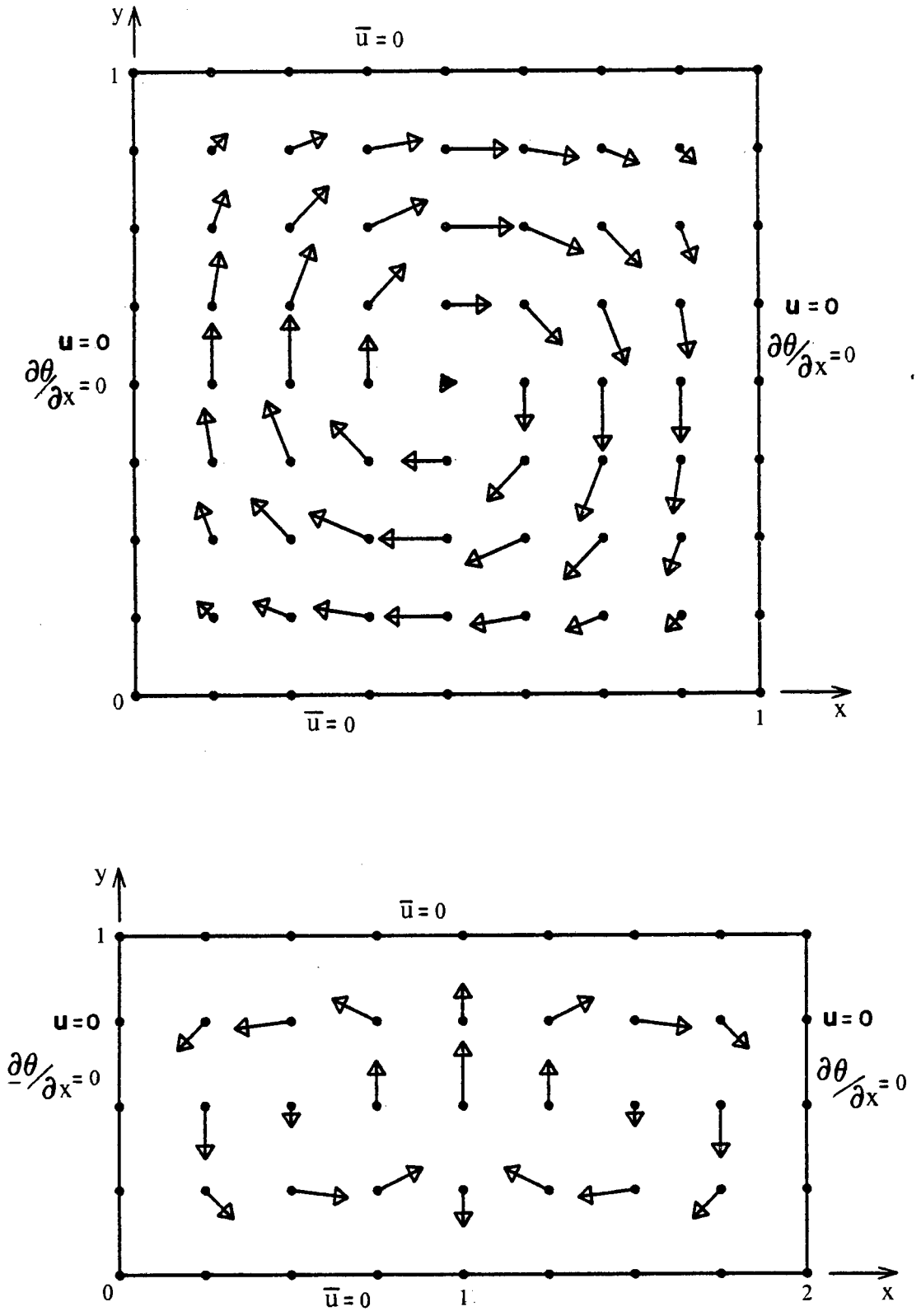


Figure 5.8: Velocity vector fields for l/d = 1, 2.

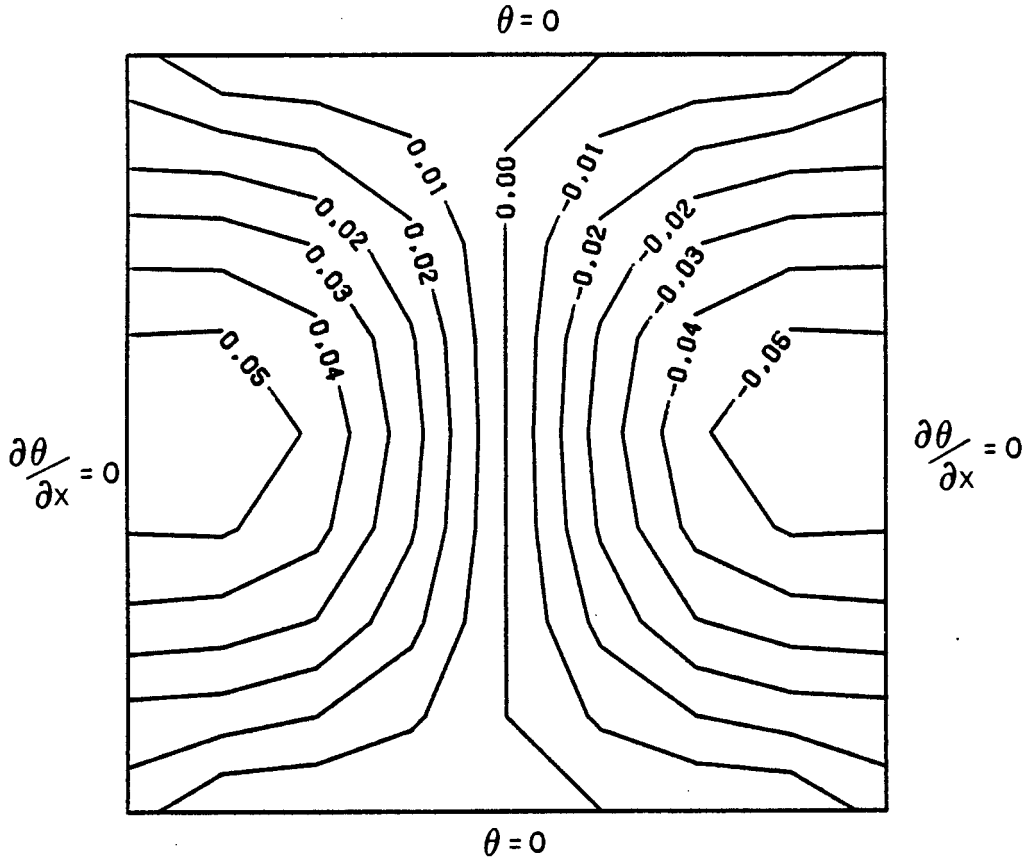


Figure 5.9: Isotherms for $^{l}/_{d}$ = 1.

We now investigate the stability of a fluid for a broad range of temperature boundary conditions. These include a fixed temperature (θ =0) and fixed heat flux ($^{d\theta}/_{dy}$=0) at the lower bounding surface and a general convective exchange at the upper surface ($^{d\theta}/_{dy}+h_T\theta$ =0). The last condition includes fixed temperature and fixed heat flux as special cases. We require that the lower bounding surface of the fluid layer be rigid. The upper surface may either be a free surface or a rigid surface.

For each Nusselt number, h_T, there is a critical Rayleigh number below which the motionless state is stable. The critical

Rayleigh numbers marking the onset of instability for a fluid layer with $^{l}/_{d}$ = 10 are presented in Figure 5.10 and Figure 5.11. In Figure 5.10 the lower surface is at a fixed temperature and in Figure 5.11 the lower surface is at a fixed heat flux. The transparencies show the results obtained by Sparrow et al (1964) for an infinitely long fluid layer. On each of the graphs there are two curves; the curve corresponding to the free upper surface is referred to the left-hand ordinate scale; the curve corresponding to a rigid upper surface is referred to the right-hand ordinate scale.

From Figures 5.10 and 5.11 it is seen that, for a given velocity boundary condition at the upper surface, the critical Rayleigh number increases monotonically with increasing Nusselt number h_T. Thus, the most stable situation corresponds to a fixed temperature ($h_T \rightarrow \infty$). From Figure 5.10 we note that the critical Rayleigh numbers for the rigid surface exceed those for the free surface by a nearly uniform amount. Furthermore, the curves corresponding to $^{l}/_{d}$=10 exhibit the same qualitative shape as those of Sparrow et al.

To show convergence to Sparrow's results for the rigid-rigid case with fixed temperature at the lower surface, a plot of critical Rayleigh number versus $^{l}/_{d}$ is shown in Figure 5.12. There are two curves; the one corresponds to an upper surface with h_T=1 and the other corresponds to an upper surface with h_T=10. Both curves approach the limits obtained by Sparrow et al. as $^{l}/_{d}$ gets larger.

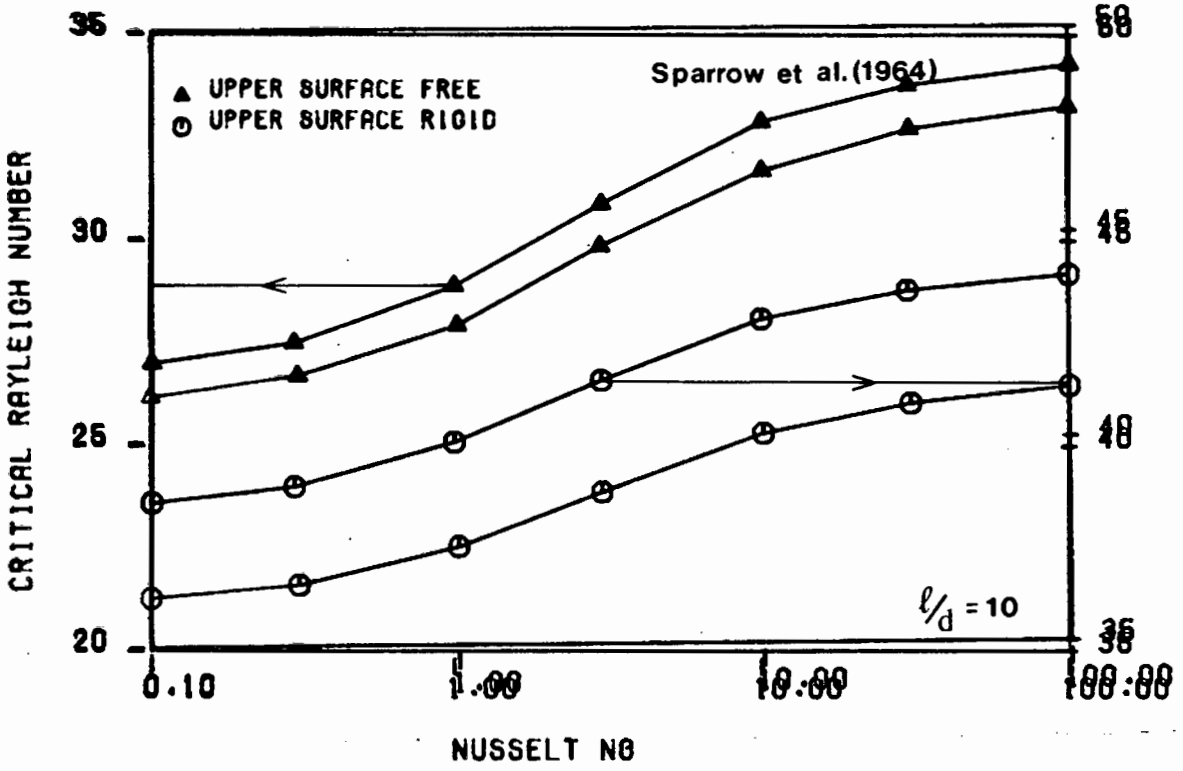


Figure 5.10: Fixed temperature at the lower bounding surface.

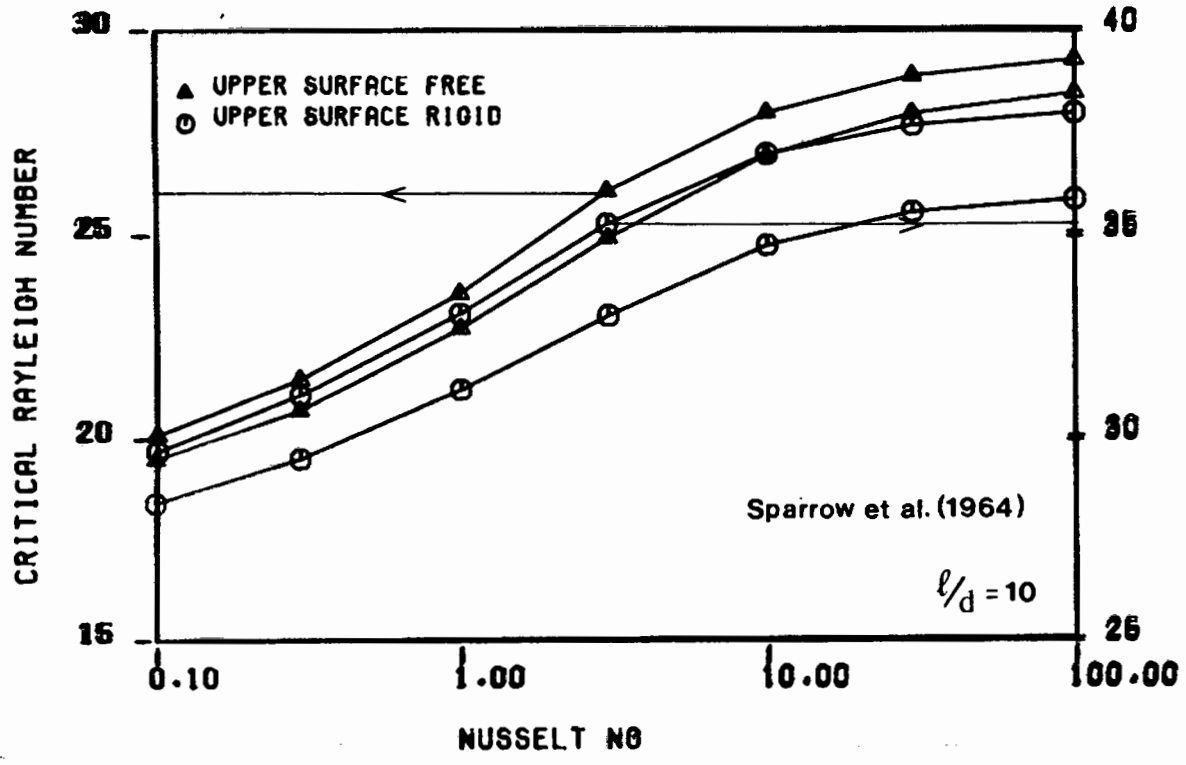


Figure 5.11: Fixed heat flux at the lower bounding surface.

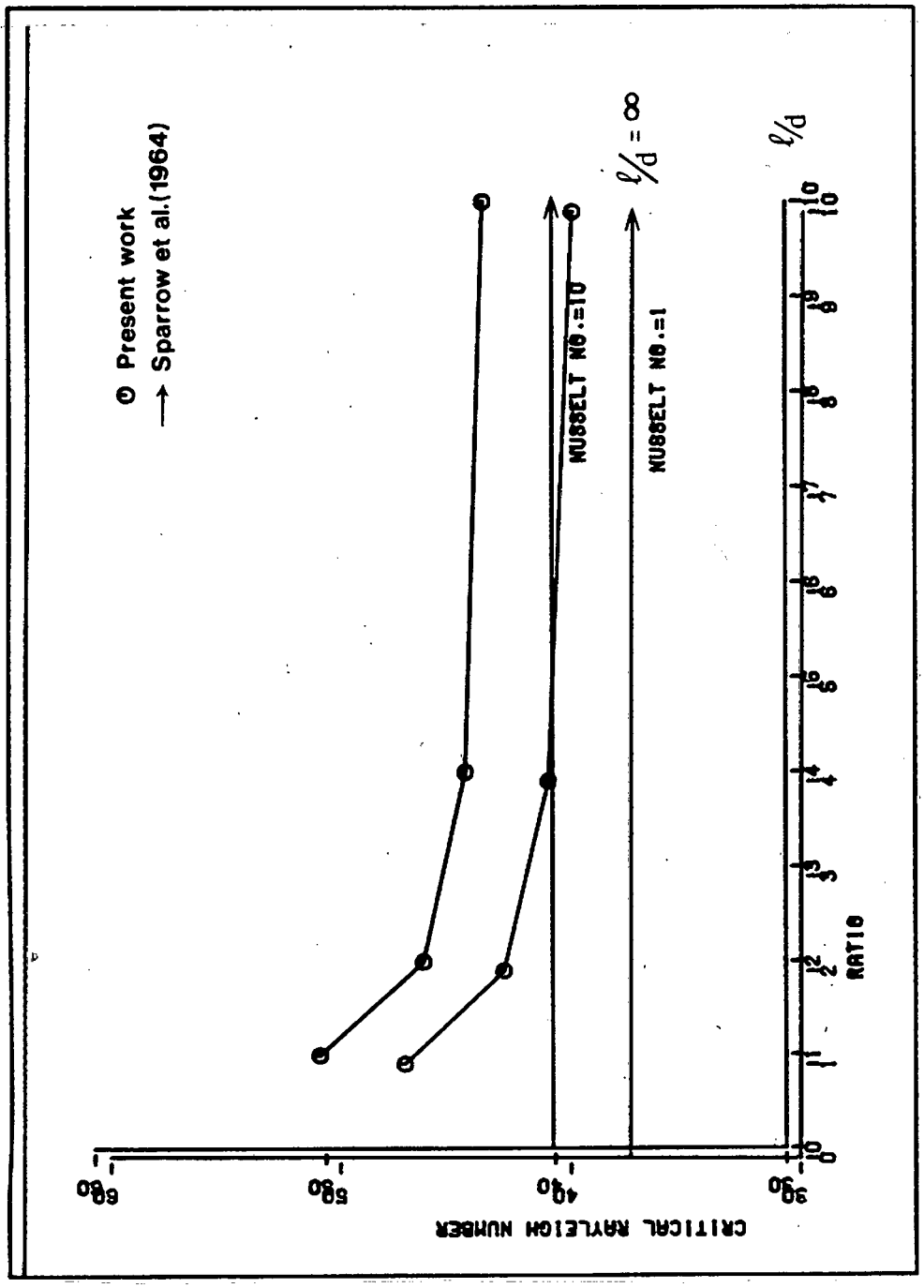

Present work
Sparrow et al.(1964)
NUSSELT NO.=10
NUSSELT NO.=1
$\ell/d = \infty$
ℓ/d
RATIO
CRITICAL RAYLEIGH NUMBER
39_{30}
40_{0}
58_{30}
68_{30}

Figure 5.12

5.2 Non-linear temperature distrubution

We now investigate how the stability of a fluid in the motionless state is affected by the temperature distribution. In Section 5.1 we examined the stability of a fluid in which the temperature distribution decreased linearly with height. In this Section the temperature distribution is non-linear due to an internal heat source in the fluid layer. The temperature gradient ∇T, obtained in (5.5), is given by

$$\nabla T = \frac{H_s(1 - 2y) - 1}{(H_s + 1)} \quad .$$

With $H_s=0$ the above equation reduces to a linear temperature distribution, discussed in Section 5.1. The magnitude of H_s is thus a rough measure of the degree of nonlinearity. When $H_s\neq 0$ the critical stability limits for the linear and energy theory, R_L and R_E, do not coincide. By the criterion of the linear theory the motionless state is stable if $R<R_L$. The motionless state is unstable by the criterion of the energy theory if $R>R_E$. When $H_s>0$, $R_E<R_L$ and solutions exist whose energy does not decay even though the stability criterion of the linear theory is satisfied. Such solutions are called subcritical.

In order to compare results obtained in this Section with the previous Section we divide the critical Rayleigh number corresponding to a non-linear temperature distribution by a factor $(H_s+1)^{1/2}$. This is necessary since the critical Rayleigh number corresponding to a linear temperature distribution is given by

$$R = \sqrt{\frac{\alpha g d^3 T'}{\kappa \nu}}$$

where $T' = (T_1 - T_2)$ is the temperature difference accross the fluid layer. However, for the case when we have a non-linear temperature distribution, T' is defined by

$$T' = (T_1 - T_2)\ (H_s + 1)\ .$$

Thus, for $H_s > 0$ we have

$$R = \sqrt{\frac{\alpha g\ d^3\ T'}{\kappa \nu}} \times \frac{1}{H_s + 1}$$

We first calculate the critical values of the Rayleigh number for the linear theory. Consider a fluid layer with $^l/_d = 1$ and 10, and the boundary conditions: horizontal bounding surfaces both rigid and isothermal, and vertical bounding surfaces rigid and perfect insulators. The critical Rayleigh numbers, R_L, are plotted in Figure 5.13 for values of heat-source parameter H_s in the range 1 to 100. The results obtained by Sparrow et al. (1964) for an infinitely long layer are shown on the transparency for comparison. The critical Rayleigh number for $H_s = 1$ is very close to that corresponding to a linear temperature distribution and, as H_s increases the critical Rayleigh number decreases monotonically. Thus, the non-linear temperature has a destabilizing effect on the fluid. Comparing the three sets of results, we see that the curves differ by an almost uniform amount and we expect that this value will approach zero as the ratio $^l/_d$ approaches infinity.

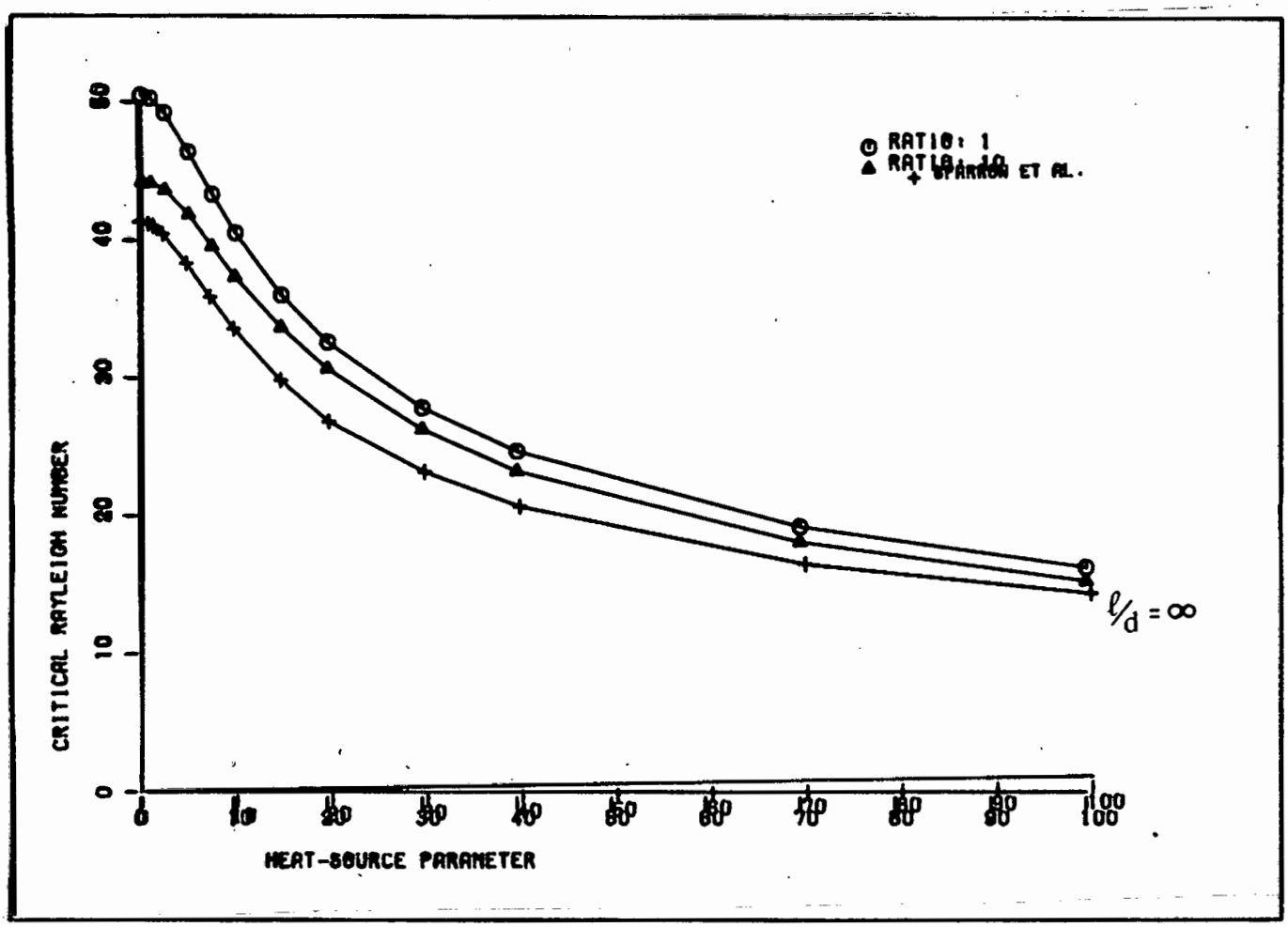


Figure 5.13: Critical Rayleigh numbers corresponding to a non-linear temperature distribution for the rigid-rigid case.

We next obtain critical values of the Rayleigh number for the energy theory. In the energy theory stability is guaranteed if $R < \rho(\lambda)$ *for a fixed value of* H_s *and fixed* $\lambda > 0$. *A set of minimum eigenvalues* ρ_λ *are found for different* λ *keeping* H_s *fixed. The* λ *which produces the maximum value of* ρ_λ *determines the critical stability limit* R_E. *The variation of* ρ_λ *with* λ *is given in Figure 5.14 for values of* H_s *in the range 0 to 100. The dashed line gives the optimal values of* ρ_λ *corresponding to the critical Rayleigh number* R_E, *over the range of values of* H_s. *In Figure 5.15 the critical Rayleigh numbers,* R_L *and* R_E, *of the linear and energy theory are compared. When* $H_s=0$, $R_L=R_E$, *and no subcritical instabilities exist. For* $H_s>0$ *the critical Rayleigh*

numbers for the energy theory are slightly less than those given by the linear theory, the difference increasing from zero with the magnitude of the heat-source intensity. The curve corresponding to the linear theory defines a boundary above which the flow is certainly unstable. The curve corresponding to the energy theory defines a boundary below which the flow is certainly stable. The region between these two curves is open to subcritical instabilities. Results obtained by Joseph and Shir (1966) and Sparrow et al. (1964) are reproduced on the transparency and in Table 5.6 for comparison.

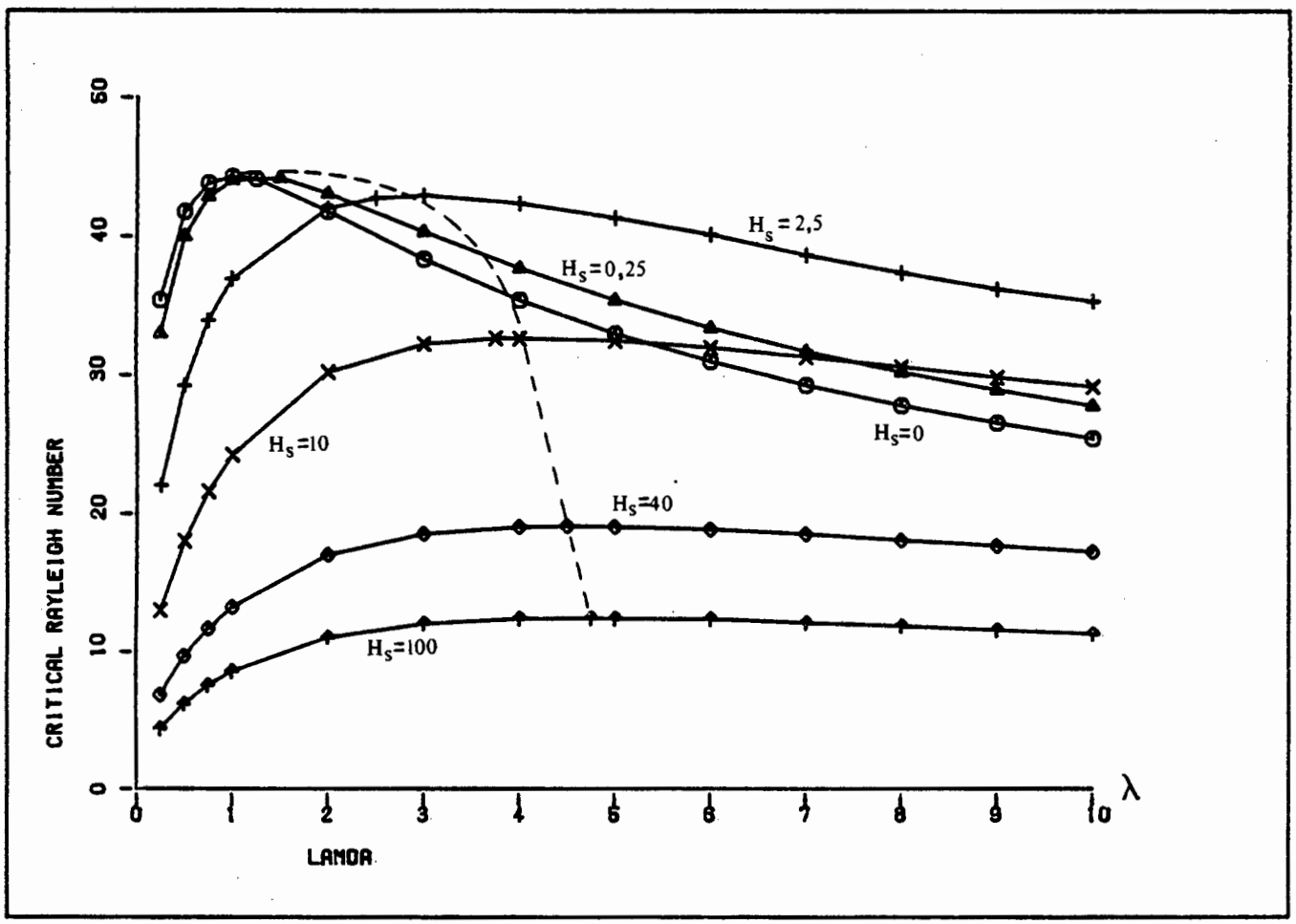


Figure 5.14: Critical Rayleigh numbers ρ_λ *as a function of* λ *. The optimal values,* R_E*, are given by the maxima of the curves.*

H_s	R_E	Joseph et al.	R_L	Sparrow et al.
0	44.292	41.326	44.292	41.326
0.25	44.180	41.304	44.276	41.315
2.5	42.952	39.306	43.712	40.409
10.0	32.662	29.933	37.380	33.443
40.0	19.117	17.464	23.280	20.213
100.0	12.483	11.402	15.345	13.302

Table 5.6

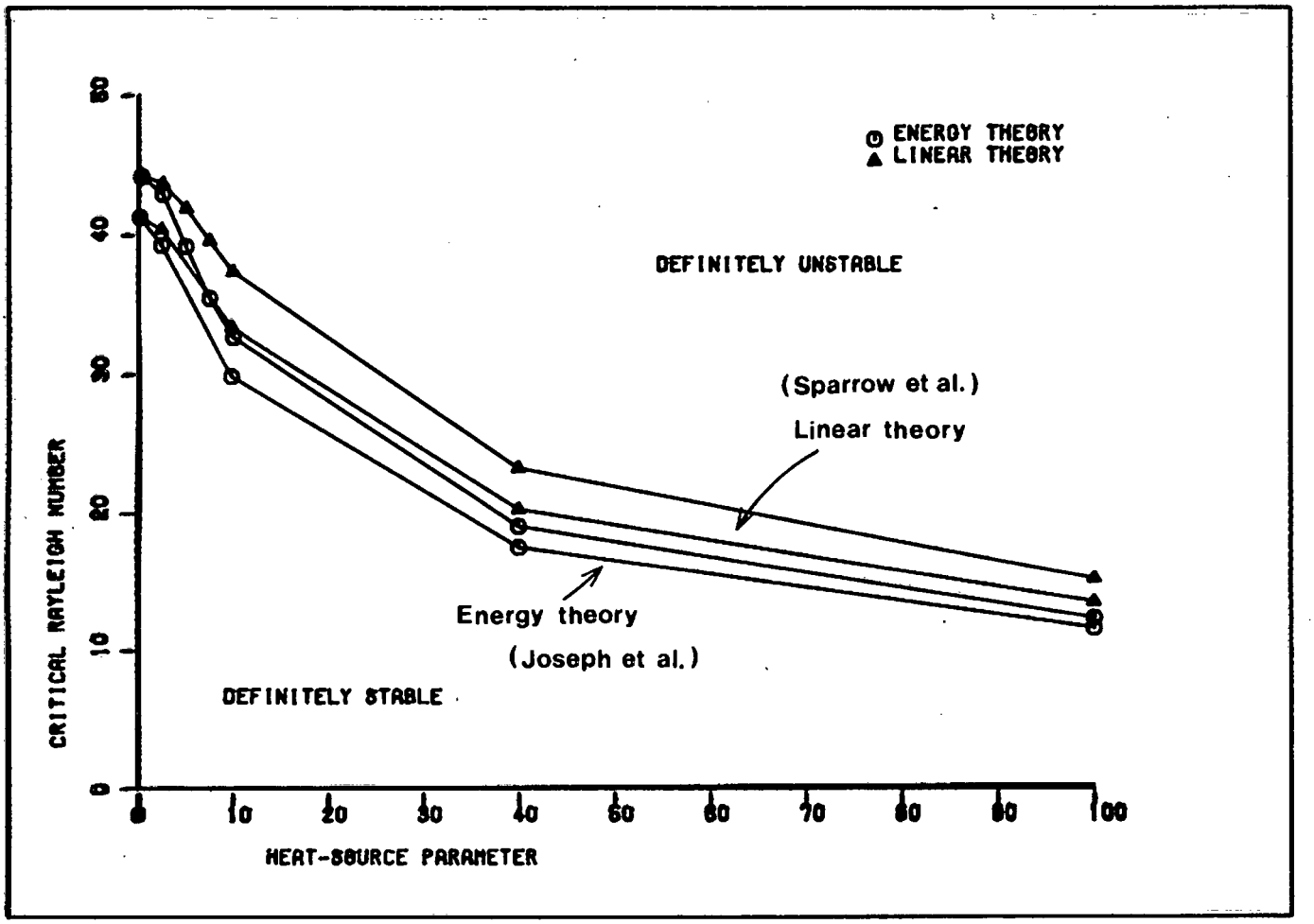


Figure 5.15: Stability boundaries for the linear and energy theory.

The penalised variational eigenvalue problems for the linear and energy theory have been derived from the Oberbeck-Boussinesq equations for heat conducting and convective flow. The existence of eigenvalues have been considered and, for the energy theory, proof of existence and uniqueness of an optimal critical Rayleigh number given. Proof of existence for real eigenvalues of the unsymmetric eigenvalue problem corresponding to the linear stability theory is beyond the scope of this work and has therefore been omitted. Finite element approximations have been constructed of the penalised eigenvalue problems and solved using Inverse Iteration. A nine-noded quadrilateral element with biquadratic interpolation of the velocity and temperature used.

The stability of Boussinesq flows in a two-dimensional box in which a negative temperature gradient occurs have been studied. Results are presented for width-to-height ratios in the range 1 to 10 and for a number of boundary conditions. The temperature conditions include fixed heat flux at the vertical bounding surfaces, fixed temperature and fixed heat flux at the bottom surface, and a general convective exchange at the upper surface which includes fixed temperature and fixed heat flux as special cases. The velocity boundary conditions include rigid vertical surfaces and rigid and free horizontal surfaces.

The results are summarized below:

- The critical Rayleigh number marking the onset of convection decreases monotonically with increasing width-to height ratio.
- The critical Rayleigh number is greatest for a boundary condition of fixed temperature and decreases as the condition of fixed heat flux is approached. For the velocity boundary

conditions, the most stable situation corresponds to a rigid bounding surface.

- *As the heat-source intensity within the layer is increased the critical Rayleigh number decreases, i.e., the nonlinear temperature distribution has a destabilizing influence.*
- *The critical Rayleigh numbers given by the energy theory are slightly less than those given by the linear theory, the difference increasing from zero (the Bénard problem) with the heat-source intensity.*
- *Comparison of the linear and energy stability limits yields a range of Rayleigh numbers in which subcritical instabilities are possible.*
- *Results obtained in this work compare well with existing results and critical Rayleigh numbers for increasing width-to-height ratios approach stability limits for the ∞ case smoothly.*

In the energy stability theory, discussed in Chapter 2, we introduced a positive coupling constant. Each choice of coupling constant gives a different critical Rayleigh number, and a maximum problem was formulated to obtain the optimal coupling constant leading to the largest critical Rayleigh number. In obtaining sufficient conditions for the existence and uniqueness of a maximum critical Rayleigh number corresponding to a optimal coupling constant we assumed that, first, all critical Rayleigh numbers are positive, and second, that the optimal coupling constant at which the critical Rayleigh number is a maximum, is positive. Proof that the global maximum is obtained at a positive value of coupling constant is lacking, however, and is a problem which is worthy of future investigation.

The convergence of the finite element approximation of the penalized problem was studied only computationally. Error estimates are not available for our problem. A theoretical

investigation necessary to obtain estimates was not attempted and error estimates of the form obtained by Geveci, Reddy and Pearce (1986) for the Stokes operator were assumed. The results show a similar trend to those of Geveci et al. (1986) for the error dependence on the penalty parameter. However, the results obtained in investigating the dependence of the error on mesh size were not adequate. Relative errors corresponding to finer meshes are needed, which results in very large matrices, causing storage problems. For such large problems an alternative storage scheme will have to be developed. A theoretical study of convergence of finite element approximations would also be worth investigating, though this is not a trivial problem due to the non-coersiveness of the form I(.,.) in Chapter 2 and due to its non-symmetry for the linear theory.

We considered only fluids which are initially motionless. This study can be extended without difficulty to include fluids in which a basic flow is present. The eigenvalue problems for the linear and energy theory will then have an extra term, each involving the basic flow, and a stability parameter, called the Reynolds number. The explicit dependence of the Reynolds number, Re, is eliminated by introducing a positive constant c such that Re = c R. This results in eigenvalue problems which are in terms of Rayleigh number only. The critical Rayleigh number is found by fixing c and obtaining R_{crit}, in the usual way, the best value of c leading to the optimal value R_{crit} corresponding to the altered problem.

We conclude that the finite element method lends itself well to the solution of the eigenvalue problems for the linear and energy stability theory. No difficulties occurred due to the presence of unsymmetric matrices in the linear theory. The boundary conditions, initial conditions and geometry of the

problem can be changed with ease. However, care must be taken when working with problems which result in large matrices since storage problems may arise.

REFERENCES

Bathe, K.J. (1982), Finite Element Procedures in Engineering Analysis. Prentice-Hall, (New Jersey)

Becker, E.B., Carey, G.F. & Oden, J.T. (1981), Finite Elements: An Introduction I. Prentice-Hall, (New-Jersey)

Chandrasekhar, S. (1961), Hydrodynamic and Hydromagnetic Stability. Oxford University Press, (London)

Charlson, G.S. & Sani, R.L. (1970), Thermoconvective Instability in Bounded Cylindrical Fluid Layer. Int. J. Heat Mass Transfer, 13, pp.1479-1496.

Chung, T.J. (1978), Finite Element analysis in Fluid dynamics. McGraw-Hill, (New York)

Davis, S.H. (1967), Convection in a box: linear theory. J. Fluid Mech., 30, pp. 465-478.

Davis, S.H. (1969), Buoyancy-surface tension instability by the method of energy. J. Fluid Mech., 39, pp.347-359.

Duvaut, G. and Lions, J.L. (1976), Inequalities in Mechanics and Physics. Springer, (Berlin)

Galdi, G.P., & Straughan, B. (1985) Exchange of stabilities, symmetry and nonlinear stability. Arch. Ration. Mech. Anal., 89, pp.211-228

Geveci, T., Reddy, B.D. & Pearce, H.T. (1986) The Approximation of the Spectrum of the Stokes Operator. *RAIRO Analyse Numérique* *(to appear).*

Hall, P. & Walton, I.C. (1977), The Smooth Transition to a Convective Régime in a Two-dimensional Box. *Proc. R. Soc. Lond. A.*, **358**, *pp. 199-221.*

Hinton, E. & Owen, D.R.J. (1977), *Finite Element Programming.* *Academic Press, (London)*

Jackson, C.P. & Winters, K.H. (1984), A Finite -Element Study of the Bénard Problem using Parameter-Stepping and Bifurcation Search. *International Journal for Numerical Methods in Fluids*, **4**, *pp. 127-145.*

Joseph, D.D. (1965), On the Stability of the Boussinesq Equations. *Arch. Ration. Mech. Anal.*, **20**, *pp.59-71.*

Joseph, D.D. (1976 I), *Stability of Fluid Motions*, *Volume I. Springer, (Berlin)*

Joseph, D.D. (1976 II), *Stability of Fluid Motions*, *Volume II. Springer, (Berlin)*

Joseph, D.D., & Shir, C.C. (1966), Subcritical convective instability. *J. Fluid Mech.*, **26**, *pp. 753-768.*

Oden, J.T., & Kikuchi, N. (1982), Finite Element Methods for Constrained Problems in Elasticity. *International Journal for Numerical Methods in Engineering*, **18**, *pp. 701-725.*

Oden, J.T., Kikuchi, N. & Song, Y.N. (1982), Penalty-Finite Element Methods for Analysis of Stokesian Flows. Computer Methods in Applied Mechanics and Engineering, 31, pp. 297-329.

Pellew, A. & Southwell, R.V. (1940), On maintained convective motion in a fluid heated from below. Proc. R. Soc. Lond. A., 176, pp. 312-343.

Reddy, B.D. (1986), Functional Analysis and Boundary-Value Problems: An Introductory Treatment. Longman, (London)

Rionero, S. (1968), Metodi variazionali per la stabilità asintotica in media in magneto-idrodinamica. Ann. Mat. Pura. Appl., 78, pp. 339

Shir, C.C. & Joseph, D.D. (1968), Convective Instability in a Temperature and Concentration Field. Arch. Ration. Mech. Anal., 30, pp. 38-80

Sparrow, E.M., Goldstein, R.J. & Jonsson, V.K. (1964), Thermal instability in a horizontal fluid layer: effect of boundary conditions and non-linear temperature profile. J. Fluid Mech., 18, pp. 513-528.

Taylor, C. & Hughes, T.G. (1981), Finite Element Programming of the Navier-Stokes Equations. Pineridge, (Swansea)

Van Steeg, J.G. & Wesseling, P. (1978), Solutions of the Boussinesq Equations by means of the Finite Element Method. Computers and Fluids, 6, pp. 93-101.

Winters, K.H. & Cliffe, K.A. (1985), Convergence Properties of the Finite-Element Method for Bénard Convection in an Infinite Layer. Journal of Computational Physics, **60**, *pp.346-351.*