AF

THE BEHAVIOUR OF OPTIMAL LYAPUNOV FUNCTIONS

by

Derek Nicholas Shields BSc., D.I.S. (Loughborough)

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of the Loughborough University of Technology.

October 1973

Supervisor: Prof. C. Storey,

Department of Mathematics.

Summary

The use of Lyapunov's direct method in obtaining regions of asymtotic stability of non-linear autonomous systems is well known. This thesis is an investigation into the optimization of some function of these systems over different classes of Lyapunov functions.

In Chapter 2 bounds on the transient response of two systems are optimized over a subset of quadratic Lyapunov functions and numerical work is carried out to compare several bounds.

Zubov's equation is the subject of Chapter 3. The non-uniformity of the series-construction procedure is studied analytically and a new approach is made to the solution of the equation by finite difference methods.

Chapters 4, 5 and 6 have a common theme of optimizing the RAS over a class of Lyapunov functions. Chapter 4 is restricted to optimal quadratics which are investigated analytically and numerically, two algorithms being developed. An optimal quadratic algorithm and a RAS algorithm are proposed in Chapter 5 for high order systems. Extensions are made in Chapter 6 to optimal Lyapunov functions of general degree and relay control systems and systems of Lure' form are considered.

Acknowledgements

The author would like to thank Professor C. Storey for his guidance and encouragement in carrying out this work. His thanks are also due to the efficient work of the LUT computing centre and his wife, Greta, for typing this thesis.

Contents

CHAPTER 1

INTRODUCTION

Chapter 1

Introduction

1.1. Discussion

The classical idea of stability originated in the motion of rigid bodies in mechanics. An equilibrium was said to be stable if a body returned to its original position after a small displacement. In the last twenty years this idea of stability has been extended considerably both in depth and scope, and powerful tools now exist to treat the stability of a large number of dynamic motions or systems. The most striking development has been the direct method of Lyapunov with its many theoretical and applicative aspects. This thesis is mainly concerned with 'optimum problems' in the use of so called Lyapunov functions to find estimates of transient response and of the domain of attraction of nonlinear autonomous differential equations.

1.2 Preliminaries

In what follows the usual notations for vectors and matrices in n-dimensional Euclidean space will apply throughout (See (14)).

Elements of E^n will be denoted by $\underline{x}$, $\underline{y}$ etc. and will be treated as column vectors, although written as rows in long hand i.e.

$$\underline{x} = (x_1, x_2, \ldots\ldots, x_n)$$

$\|\cdot\|$ is the Euclidean norm defined as

$$\|\underline{x}\|^2 = \underline{x}^T\underline{x} = \sum_{i=1}^{n} |x_i|^2$$

The elements of matrices A, B etc. will be denoted by $a_{i,j}$ and $b_{i,j}$ respectively.

E^n may also be called the <u>statespace</u> or <u>phase space</u> depending upon the nature of $\underline{x}$. For n = 2 we may write $\underline{x} = (x,y)$.

We will be concerned with the vector differential equation

$$\dot{\underline{x}} = \frac{d\underline{x}}{dt} = \underline{f}(\underline{x}, t) \qquad 1.1.1$$

where $\underline{x} \in E^n$; t is an independent parameter, usually the time; and $\underline{f} \in E^n$, whose components $f_i(\underline{x}, t)$ are functions of x_i and t (specifically $\underline{f}$ is the map $\underline{f} : E^n \times R \rightarrow E^n$, R the real line).

The form of 1.1.1 is quite general and it can be considered as a system of n first order equations. The important nth order scalar equation,

$$y^{(n)} = f\,(y^{(n-1)}, y^{(n-2)}, \ldots, y^{(1)}, y, t)$$

where

$$y^{(n)} = \frac{d^n y}{dt^n} \qquad 1.2.1.$$

is reducible to this form by defining

$$\begin{aligned} x_1 &= y \\ \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ &\vdots \\ \dot{x}_{n-1} &= x_n \\ \dot{x}_n &= f\,(x_n, x_{n-1}, \ldots, x_1, t) \end{aligned} \qquad 1.2.2.$$

If $\underline{f}$ in 1.1.1 is independent of t we have

$$\underline{\dot{x}} = \underline{f}\,(\underline{x}) \qquad 1.2.3$$

an <u>autonomous system</u> with special properties (See later or Zubov (1)).

If the right hand side of 1.1.1 is <u>continuous</u> and <u>the existence</u> and <u>uniqueness</u> of solutions is assured together with their <u>continuous dependence</u> on initial values, $\underline{f}$ will be said to be of class E, $\underline{f} \in E$.

Let $(\underline{x}_o, t_o)$ be the initial values and let $\underline{f} \in E$, then define $\underline{z}\,(t, \underline{x}_o, t_o)$ as the solution of 1.1.1, i.e.

$$\frac{d\underline{z}}{dt}\,(t, \underline{x}_o, t_o) = \underline{f}\,(\underline{z}\,(t, \underline{x}_o, t_o), t) \qquad 1.2.4.$$

with $\underline{z}\,(t_o, \underline{x}_o, t_o) = \underline{x}_o$

The <u>singular</u> or equilibrium points of 1.1.1 are the constant solutions $\underline{z}\,(t, \underline{x}_o, t_o) = \underline{x}_o$ or, equivalently, the solutions $\underline{x}$ satisfying $\underline{f}\,(\underline{x}, t) = \underline{0}$.

By simple transition of co-ordinates any singular point may be brought to the origin, $\underline{x} = 0$. Henceforth, we assume $\underline{f}(\underline{o},t) = o$ and that the <u>origin</u> is an <u>isolated</u> singular point (i.e. no other such point exists in a neighbourhood of $\underline{x} = o$).

Define the (n + 1) - dimensional space of quantities $(\underline{x}, t)$ as the <u>motion space</u>, then a <u>motion</u> of 1.1.1 is the continuous path formed by the set $(\underline{z}(t,\underline{x}_o,t_o), t)$. A <u>trajectory</u> is the projection of this path onto the phase space and a <u>half trajectory</u> is a trajectory defined for some $t \geqq t_o$ (or $t \leqq t_o$).

1.3 Basic Definitions of Stability in the Sense of Lyapunov

Denote by R(h) the region $R(h): \{\underline{x} / \|\underline{x}\| \leq h\}$, or more generally, let the set $\{(\underline{x}, t) / \|\underline{x}\| \leq h, t \geq t_o\}$ be denoted by $R(h, t_o)$. Suppose in $R(h,t_o)$

$$\dot{\underline{x}} = \underline{f}(\underline{x}, t) \qquad (\underline{f}(\underline{o},t) = o , \underline{f} \in E) \qquad 1.3.2.$$

Then two definitions are basic to Lyapunov's direct method:

<u>Def. 1.3.1</u>

The origin of the differential equation (d.e.) 1.3.2. is said to be <u>stable</u> if there exists for any $\epsilon > o$ a number $\delta > o$ such that.

$$\|\underline{x}_o\| < \delta$$

implies

$$\|\underline{z}(t, \underline{x}_o, t_o)\| < \epsilon , \forall t \geq t_o$$

Def. 1.3.2

The origin of the d.e. 1.3.2 is said to be asymptotically stable (a.s.) if it is stable and there exists a δ_o such that for

$$\|\underline{x}_o\| < \delta_o \quad , \quad \delta_o > 0$$

follows

$$\lim_{t \to \infty} \underline{z}(t, \underline{x}_o, t_o) = 0 \qquad 1.3.3$$

If 1.3.3 holds for all $\underline{x}_o \in E^n$ in Def. 1.3.2 the origin is said to be a.s. in the whole. Further definitions are given in Lefshetz (14), Zubov (1) etc., including instability definitions. Good critical treatments of these and other definitions are given in Hahn (15) and Lehnigk (17).

Finally, the domain of attraction (DOA) of an a.s. system 1.3.2 is the set U defined by

$$U(t_o) : \left\{ \underline{x}_o \,/\, \lim_{t \to \infty} \underline{z}(t,\underline{x}_o,t_o) = 0 \right\} \qquad 1.3.4.$$

For the autonomous case U is independent of t_o.

1.4. The Autonomous Case

The main system in the following chapters is system 1.2.3 namely,

$$\dot{\underline{x}} = \underline{f}\,(\underline{x}) \qquad (\underline{f}(\underline{o}) = o,\ \underline{f} \in E) \qquad 1.4.1.$$

Two forms of 1.4.1, one particular and one general, are the following:

a) the linear system

$$\dot{\underline{x}} = A\underline{x} \qquad 1.4.2.$$

where A is an n x n matrix which is said to be stable if its eigenvalues, λ_i, i=1,n, have negative real parts. The system is called significant if $Re(\lambda_i) \neq o$

b) The system

$$\dot{\underline{x}} = A\,\underline{x} + \underline{g}\,(\underline{x}) = \underline{f}\,(\underline{x}) \qquad 1.4.3.$$

where $\underline{g}\,(\underline{x})$ possesses a convergent power series expansion about the origin whose terms are of degree two and greater. Here, A $\underline{x}$ is called the first approximation or linear part of 1.4.3. A can be regarded as the Jacobian of $\underline{f}$ at $\underline{x}$ = o,

$$A = \left.\frac{\partial \underline{f}(\underline{x})}{\partial \underline{x}}\right|_{\underline{x} = o} = \left\{ \left.\frac{\partial f_i}{\partial x_j}\right|_{\underline{x} = o} \right\} \qquad 1.4.4.$$

1.5. The Theorems of Lyapunov

The second method of Lyapunov attempts to determine the stability of the equilibrium without prior knowledge of the solutions of differential equations. It introduces the idea of a certain function called a Lyapunov function which possesses properties analogous to those of the total energy of a dissipative dynamic system. The energy in the latter is positive and non-increasing near a stable equilibrium.

Formally, let $V(\underline{x})$ be a continuous scalar function defined in some region $R(h)$ and possessing continuous first partial derivatives. Then the following definitions and theorems are pertinent to the autonomous system 1.4.1 (See Hahn (15) for general case of system 1.3.2)

Def. 1.5.1.

The function $V(\underline{x})$ is positive (negative) definite if in some region $R(h_1)$,

$$V(\underline{x}) > o \ (< o) \quad \text{and} \quad V(\underline{o}) = o. \quad \text{If}$$

$$V(\underline{x}) \geq o \ (\leq o) \quad \text{and} \quad V(\underline{o}) = o \quad \text{it is}$$

positive (negative) semi-definite.

$V(\underline{x})$ is called strictly positive definite if $V(\underline{x}) > o$ for $\underline{x} \in E^n$, $\underline{x} \neq o$; and radially unbounded if $\|\underline{x}\| \longrightarrow \infty$ implies $V(\underline{x}) \longrightarrow \infty$.

If in some $R(h)$, $V(\underline{x})$ is positive definite and its total derivative $\dot{V}$, where $\dot{V} = \nabla V^T \underline{f}(\underline{x}) = \sum_{i=1}^{n} \frac{\partial V}{\partial x_i} f_i(\underline{x})$, is negative semi-definite, it is called a Lyapunov function (LF) for system 1.4.1.

Theorem 1.5.1.

The equilibrium of system 1.4.1. is

(a) stable if there exists a Lyapunov function and

(b) a.s. if $V(\underline{x})$ is positive definite and $\dot{V}$ is negative definite with respect to 1.4.1.

Theorem 1.5.2. (Barbashin(18))

The equilibrium of 1.4.1 is a.s. if

(a) $V(\underline{x})$ is positive definite and

(b) $\dot{V}$ is negative semi-definite and does not vanish identically on any non-trivial trajectory of 1.4.1.

These two theorems are purely local in character and give little information as to the size of the actual stability regions. In this respect the following theorem is of great practical importance.

Theorem 1.5.3. (Lefshetz (14))

Trajectories of 1.4.1 which start from a region D containing the origin will be a.s.* if there exists a function $V(\underline{x})$ with the properties:

(a) $V(\underline{x})$ is positive def. in D,

(b) $\dot{V}$ with respect to system 1.4.1 is at least neg. semi-definite,

(c) $\dot{V}(\underline{x}) \not\equiv 0$ on any trajectory of 1.4.1 in D except $\underline{x} = 0$,

*(An a.s. trajectory is one originating from some initial $\underline{x}_0 \in D$, an RAS)

(d) $\nabla V(\underline{x}) \neq o$ in D except for $\underline{x} = o$,

(e) one of the level surfaces V = constant bounds D.

Let $V(\underline{x}) = c$ be a level surface bounding a region D,

D: $\{ \underline{x} / V(\underline{x}) < c, c > o \}$

Henceforth, the region D_{max} bounded by the surface $V(\underline{x}) = C_{max}$, where C_{max} denotes the largest c for which properties (a) to (e) hold, will be called the <u>region of asymptotically stability</u> (RAS) of the Lyapunov function $V(\underline{x})$ for the system 1.4.1.

If this RAS is unbounded then 1.4.1 is <u>a.s. in the whole</u>.

From Theorem A1.1 we see that the level $V(\underline{x})$ surfaces will be <u>closed</u> and hence bounded in some neighbourhood of the origin.

<u>Theorem 1.5.4.</u> (Linear Case)

The linear system 1.4.2 is a.s. iff A is a stability matrix.

<u>Theorem 1.5.5.</u>

The origin of 1.4.2 is a.s. iff there exists a positive definite symmetric matrix P which is the unique solution of the Lyapunov matrix equation.

$$A^T P + PA = - Q \qquad 1.5.1.$$

for any positive definite symmetric Q.

(The theorem also holds with P and Q p.d. hermitian matrices).

Theorem 1.5.6

If the stability of the first approximation of 1.4.3 is significant then the stability behaviour of the linear part (1.4.2) and the complete system (1.4.3) are the same.

For less restrictive conditions on $\underline{g}(\underline{x})$ in 1.4.3. see Hahn (15) or Lehnigk (17). A comprehensive treatment of the matrix equation 1.5.1 is given by Barnett (20).

1.6. A Practical RAS For the Autonomous Case

Let A be a stability matrix in system 1.4.3. Then a general class of Lyapunov functions for this sytem can be generated as follows. Let $V(\underline{x})$ be of the form

$$V(\underline{x}) = \underline{x}^T P \underline{x} + V_o(\underline{x}) \qquad 1.6.1.$$

where P is a unique solution of 1. 5.1. for some positive definite Q and where $V_o(\underline{x})$ can be expanded as a Taylor series with terms of degree three and greater. Since

$$\dot{V} = -\underline{x}^T Q \underline{x} + 2\underline{x}^T P \underline{g}(\underline{x}) + \nabla V_o(\underline{x})^T (A \underline{x} + \underline{g}(\underline{x})) \qquad 1.6.2$$

by the assumptions on $\underline{g}(\underline{x})$ and V_o and the properties of quadratic forms, $\dot{V} < o$ in R(h) for $\underline{x} \neq o$ and h small. We have

Theorem 1.6.1

Let E_V be the set

$$E_V : \{ \underline{x} / \dot{V}(\underline{x}) = o , \underline{x} \neq o \} \qquad 1.6.3$$

Then the RAS indicated by the Lyapunov function (LF) $V(\underline{x})$ of 1.6.1 is given by D where

$$D : \{ \underline{x} / V(\underline{x}) < C_{min} \} \qquad 1.6.3.$$

and where

$$C_{min} = \min V(\underline{x}), \quad \underline{x} \in E_V \qquad 1.6.4.$$

Appendix 1 (A1) gives some useful definitions on the closed contours of Lyapunov functions.

1.7. Motivation

Two main problems are inherent in using Lyapunov methods to find regions of a.s. of autonomous differential equations,

(a) the construction of a suitable LF,

(b) the determination of the RAS indicated by that LF.

Many methods exist to solve (a) (see Tait (37) and Brockett (38) for bibliographies). Some were developed for a specific form of differential equation, others were more general. Hewit (2) has developed computer algorithms for their construction and compared their RAS's.

The main motivation of this thesis has been that whereas a great deal of research has centred on constructing LF's little attention has been paid to finding the 'best' LF of a given class. Here, 'best' need not be interpreted solely in terms of the RAS but also in terms of the transient response or some other function of the system.

Emphasis has therefore been placed on finding 'optimum results' where possible and in showing what properties if any, these 'optimal Lyapunov functions' possess.

Some attention has been centred on finding analytic results for simple specific cases, which have given a lead to the development of numerical algorithms needed to study more complex cases.

1.8 Contents of Chapters and Background Material

Many authors, including Kalman and Bertram (16), Vogt (19), Zubov (1) and Wiberg (23), have used the fact that if $V(\underline{x})$ is a Lyapunov function for the autonomous system 1.4.3 (or the more general system 1.3.2) giving asymptotic stability, then minimizing the expression $(-\dot{V}/V)$,

$$\alpha = \min_{\underline{x} \in R(h)} (-\dot{V}/V)$$

for some sufficiently small region $R(h)$, implies the inequality

$$V(\underline{x}(t)) \leq e^{-\alpha t} V(\underline{x}(0))$$

The quantity α^{-1} may be interpreted as the largest time-constant over the region $R(h)$ of the phase space and is therefore a figure of merit of the system.

In Chapter 2 we extend some work of Wiberg (23) and maximize α over a sub-class of quadratic Lyapunov functions (with given $R(h)$) for the system 1.4.3 and the more general system

$$\dot{\underline{x}} = A\underline{x} + G(\underline{x},t)\underline{x} + \underline{u}(t) \qquad \text{(A stable)}$$

An optimizing condition is found when A is in companion form (CF) with real eigenvalues and some useful bounds are proposed, the latter being tested by numerical work. Some numerical work is also conducted in determining whether some bounds of Vogt (19) are useful in locating the real parts of the eigenvalues of A.

In Chapter 3 we consider Zubov's (1) partial differential equation (PDE) for the autonomous system 1.4.3,

$$\sum_{i=1}^{n} \frac{\partial V}{\partial x_i} f_i = -\phi(1 - V)$$

A wealth of numerical experience on the application of its solution by the series procedure (mainly by Hewit (2), Margolis and Vogt (4), Rao and De Sarker (5), and Yu and Vongsuriya (6)) has shown that the RAS's of the high degree LF's are often inferior to those of lower degree. No analytic study of this non-uniformity has yet been attempted.

In the first part of this chapter a system showing this non-uniformity is investigated analytically and an important question emerges concerning the region of convergence of the series-Lyapunov function. In the second part we look at an alternative way of solving the PDE by finite-difference methods. By using polar co-ordinates an initial value problem results and solution by a Crank-Nicholson-type difference scheme is possible. Numerical examples show that the solution breaks down near the DOA boundary. Some other methods are also considered.

Chapters 4, 5 and 6 have a common theme in that for the system of the form 1.4.3, and also for relay systems in chapter 6, the problem of maximizing the RAS over a class of Lyapunov functions is considered. The stated problem is

$$\max_{\underline{a}} \rho(\underline{a}) \ , \quad \underline{a} \in C \qquad 1.8.1$$

where C is a parameter class determining the LF and $\rho(\underline{a})$ is a measure of the size of the RAS

$$V(\underline{x},\underline{a}) < V_m(\underline{a}) \qquad 1.8.2$$

where

$$V_m(\underline{a}) = \min_{\underline{x}} V(\underline{x},\underline{a}) \qquad 1.8.3$$

subject to

$$\dot{V}(\underline{x},\underline{a}) = 0 \ , \quad \underline{x} \neq 0 \qquad 1.8.4$$

The main difficulty is the RAS determination of finding V_m in 1.8.2. Research on the subject is divided into two main camps of either treating 1.8.3 analytically as an equality-constrained optimization problem, or geometrically, as a tangency between two hypersurfaces, $V(\underline{x},\underline{a}) = V_m(\underline{a})$ and $\dot{V}(\underline{x},\underline{a}) = 0$. Rodden (3) gave a method for the latter which Hewit (2) has improved and applied to some second order systems. Hewit (2) has used the method to optimize the average radius of the RAS for a number of LF's determined by his construction procedures for the methods of Zubov, Szegö, Ingwerson and Krasovski.

Advocates of the equality-optimization approach have been in the main Szegö (52), Geiss (50, 51), Julich (57), Muddle (58) and Lapidus and Berger (49). They all use the penalty function method. For instance, Geiss, Julich and Szegö use either

$$F_1 = V + K\dot{V}^2 / \|\underline{x}\|^2$$

or

$$F_2 = V + K\dot{V}^2$$

and for increasing K minimize F_1 or F_2 via some powerful minimization technique such as Davidon -Fletcher-Powell (60), thus reducing 1.8.3 to a sequence of unconstrained minimizations which hopefully converge to the constrained minima (Fiacco and McCormick (59) discuss convergence). Once V has been chosen therefore, the problem is taken out of the realm of Lyapunov theory and into that of non-linear optimization where new powerful algorithms can be applied. Variations are the choice of penalty function and the minimization method.

In Chapter 4 the class of Lyapunov functions is that of quadratics, $V = \underline{x}^T P \underline{x}$, with P determined through the matrix equation 1.5.1 for Q p.d. Due to the complexity of the RAS determination, analytic results in applying problem 1.8.1 to various systems have been scarce. Geiss (50) found an optimal quadratic for a second order Duffing equation but the example gives little insight into the nature of the problem. It is surprising that the numerical algorithms developed so far for maximizing $\rho(\underline{a})$ for a given LF are based on little understanding of the relationship between the optimal RAS boundary , $V = V_m$ say, and its constraint contour, $\dot{V} = 0$. (Wilson (61) gives a topological account of the V-contours but the constraint contour is not investigated). In Chapter 4 five systems are studied analytically as far as possible through the Lagrange equations. Although only two optimal quadratics are obtained, sufficient insight is gained on which to base a numerical investigation. An efficient and accurate algorithm is developed for RAS determination for a restricted class of second order systems, and optimal quadratics are found via Powell's (33) conjugate gradient algorithm for a number of systems. Extension to higher order systems is made. The analytic and confirming numerical results exhibit an 'equal tangency property', namely, that for many systems a subset of quadratics exist such that their RAS boundaries have at least two points of contact (not radially symmetric) with their constraint contours, $\dot{V}(\underline{x}) = 0$.

In Chapter 5 two algorithms are proposed. The first is an optimal quadratic algorithm based on an idea of

Davidson and Kurak (47) who reduce the optimum quadratic problem to one of a constrained optimization which they solve via use of Rosenbrock's (32) method. The proposed algorithm takes into account the work of Chapters 2 and 4 and replaces Rosenbrock's hill-climber by a variation of the Complex method of Box (55). A comparison is made between the two methods for 2 second order and 4 third order systems, showing the proposed method is superior.

The second algorithm incorporates some features of the previous algorithm and determines the precise RAS for a quadratic LF via a special penalty function of Miele (62) (reference is made to one of the four authors) which is minimized by the Fletcher Powell (60) method of conjugate directions. The drawback of many penalty function methods - noteably those of Julich (57) and Lapidus (49) - is that no automatic method is proposed to find the 'global' minimum. An exception is that of Geiss (50) who encloses a possible RAS boundary, $\underline{x}^T P \underline{x} = V_m$, with an n-dimensional box inside which $\dot{V}$ is determined at random points. However the method is time consumming. The proposed method finds the 'global' minimum of the penalty function automatically and has good convergence.

Chapter 6 is an extension of Chapter 4 in that the optimal properties of general Lyapunov functions are investigated. The work is divided into three sections which correspond to the three different systems considered; the general autonomous system 1.4.3, a relay control system and a system of Lure' form.

The problem in 1.8.1 of maximizing $\rho(\underline{a})$ for a LF of degree mv

$$V = \sum_{i=2}^{mv} V_i \qquad (V_i \text{ poly. of deg. i}) \qquad 1.8.5$$

has been considered by Szegö (52) and Hewit (2) and others. However, due to the number of independent parameters involved for a LF of degree mv $(2 + (mv + 5)(mv - 2)/2)$ little numerical experience has resulted, even on comparing their RAS's. Szegö (52) proposed solving the RAS problem of 1.8.3 and 1.8.4 via a penalty function approach using the Fletcher-Powell (60) minimization routine. He then maximizes r, the distance of the nearest point of the RAS boundary, $V = V_m$, to the origin, over the co-efficients of the V_i terms by Powell's (33) method. Disappointingly, only an optimal quadratic is obtained and that for a simple example where a global search for the minimum in 1.8.3 is not required. In Chapter 6 the problem is investigated fully via Rodden's (3) method and Nelder and Mead (34) Simplex optimization on $\rho(\underline{a})$ (average radius). The 2mth degree LF of the form

$$V_{2m} = \prod_{k=1}^{m} \underline{x}^T P_k \underline{x}$$

is also considered. The optimal RAS's are compared and a multiple tangency phenomenon is exhibited for a number of second order systems.

In the following section we extend some work of Weissenberger (48, 66) for the relay system

$$\dot{\underline{x}} = A\underline{x} + \underline{b}\,\text{sgn}\,\sigma \quad , \quad \sigma = \underline{d}^T\underline{x}$$

who showed that under certain conditions LF's of the forms:

a) $V = \underline{x}^T P \underline{x} + |\underline{d}^T \underline{x}|$

and

$$b)\ V = \sum_{i=1}^{m} |\underline{c}_i^T \underline{x}|$$

could be used. A number of second order systems are studied numerically for LF (a) showing 'equal tangency properties'. Using this property an optimum RAS is found analytically for LF (b).

In the final section a connection between the work of Walker and McClamroch (73) and that of Weissenberger (72) is found concerning an optimal quadratic for the Lure' system

$$\dot{\underline{x}} = A\underline{x} + \underline{b}\,\mathrm{sgn}\,\sigma, \quad \sigma = \underline{c}^T\underline{x}$$

where the sector condition

$$0 < f(\sigma)/\sigma < K$$

is satisfied only for some region $\sigma_2 \leq \sigma \leq \sigma_1$. Some extensions are considered.

The computing times given in this thesis are all in terms of mill/secs. (ICL 1905) and serve only as a comparison, all other conditions being equal. All programs were written in FORTRAN 4 and only a listing of the optimal quadratic algorithm of C5* is included. Several programs were written using graph plotter routines to trace the required Lyapunov contours, points on which were joined by straight line segments. Diagrams and tables are included in the text for continuity, whereas the figures appear at the back of each chapter and, as far as is convenient, in numbered order.

* Chapter 5

CHAPTER 2

OPTIMAL BOUNDS ON THE RESPONSE OF NON-LINEAR STABLE SYSTEMS.

Chapter 2

Optimal Bounds On The Response Of Non-linear Stable Systems.

2.1 Introduction

In the design of a control system it is useful to predict a conservative bound on the response of the system, which takes into account noise and perturbation effects, or to predict a crude approximation to the domain of attraction.

Consider the two systems

$$\dot{\underline{x}} = A\underline{x} + G(\underline{x},t)\underline{x} + \underline{u}(t) \qquad 2.1.1$$

$$= \underline{f}(\underline{x},t) \qquad (\underline{f} \in E)$$

and

$$\dot{\underline{x}} = A\underline{x} + \underline{g}(\underline{x}) \qquad 2.1.2$$

with A stable.

In the former $G\underline{x}$ is regarded as the non-linear or perturbation term and $\underline{u}$ the input to the system; both are assumed to be bounded,

$$\|\underline{u}(t)\| \leq c_1, \quad \forall \quad t \qquad 2.1.3$$

$$\|G(\underline{x},t)\| \leq c_o, \quad \forall \quad t \qquad 2.1.4$$

We assume that no real $\underline{y}$ exists such that $A + G(\underline{y},t) = 0$ inside some $R(h)$, then the origin will be an isolated singularity. System 2.1.2 is that of 1.4.3.

For both systems choose the LF

$$V = \underline{x}^T P\underline{x} \qquad 2.1.5$$

where P solves

$$A^T P + PA = -Q \qquad 2.1.6$$

and where Q and hence P may be general positive definite symmetric matrices. Then for 2.1.1

$$\dot{V} = -\underline{x}^T Q\underline{x} + 2(\underline{x}^T G(\underline{x},t) + \underline{u}^T(t))P\underline{x} \qquad 2.1.7$$

and for 2.1.2

$$\dot{V} = -\underline{x}^T Q\underline{x} + 2\underline{x}^T P\underline{g}(\underline{x})\underline{x} \qquad 2.1.8$$

Following Kalman and Bertram (16) we use several matrix inequalities and the Schwartz inequality to give, respectively

$$\dot{V} \leq (-\eta + 2c_o\sqrt{\mu})V + 2c_1\sqrt{M(P)V} \qquad 2.1.9$$

and

$$\dot{V} \leq (-\eta + 2\mu\|\underline{g}(\underline{x})\| / \|\underline{x}\|)V \qquad 2.1.10$$

Here $\lambda(A)$ denotes an eigenvalue of A, $M(A) = \max \mathrm{Re}\,\lambda(A)$ and $m(A) = \min \mathrm{Re}\,\lambda(A)$. Also

$$\eta = \min_{\underline{x} \neq 0} \frac{\underline{x}^T Q\underline{x}}{\underline{x}^T P\underline{x}} \qquad 2.1.11$$

and

$\mu = \mu(P) = M(P)/m(P)$. It can be shown that (16) $\eta = m(P^{-1}Q)$.

Changing the variable in 2.1.9 to $\sqrt{V}$, by dividing by $\sqrt{V}$, and assuming $\|\underline{g}\| \leq c_2\|\underline{x}\|$, ($c_2$ constant) the two equations 2.1.9/2.1.10 can be integrated to give the bounds

$$\sqrt{V}(t) \leq \sqrt{V}(0)\exp(-\alpha t) + \frac{c_1}{\alpha} M(P)(1-\exp(-\alpha t)) \qquad 2.1.13$$

and

$$V(t) \leq V(0)\exp(-\beta t) \qquad 2.1.14$$

where

$$\alpha = \eta/2 - \sqrt{\mu}c_o = \sqrt{\mu}\left(\frac{\eta}{2\sqrt{\mu}} - c_o\right) \qquad 2.1.15$$

and

$$\beta = \eta/2 - \mu c_2 = \mu\left(\frac{\eta}{2\mu} - c_2\right) \qquad 2.1.16$$

Finally, by use of the inequality

$$\|\underline{x}\|^2 m(P) \leq \underline{x}^T P \underline{x} \leq M(P)\|\underline{x}\|^2 \qquad 2.1.17$$

the two bounds in V(t) above give respectively

$$\|\underline{x}(t)\| \leq \sqrt{\mu}\left(\|\underline{x}(0)\| \exp(-\alpha t) + \frac{c_1}{\alpha}(1-\exp(-\alpha t))\right) \qquad 2.1.18$$

and

$$\|\underline{x}(t)\| \leq \sqrt{\mu}\|\underline{x}(0)\| \exp(-\beta t) \qquad 2.1.19$$

which imply, if $\alpha > 0$, $\beta > 0$, that

$$\|\underline{x}(t)\| \leq \sqrt{\mu}\max(\|\underline{x}(0)\|, c_1/\alpha) \qquad 2.1.20$$

and

$$\|\underline{x}(t)\| \leq \sqrt{\mu}\|\underline{x}(0)\| \qquad 2.1.21$$

(Note for brevity we have written $\underline{x}(t) \doteq \underline{z}(t, \underline{x}_o, t_o)$, $t_o = 0$ from 1.2.4). The above work follows that of Wiberg (23) with some corrections, namely the bound 2.1.18. A crude RAS from the bounds is given in A3. We add that if c_o is given á priori too large, α may be negative which destroys the bound; but if $\underline{G}(\underline{x}, t) \longrightarrow 0$ as $\|\underline{x}\| \longrightarrow 0$ then the bound 2.1.18 will always hold in some region R(h) for sufficiently small h.

Generally, α^{-1} and β^{-1} behave as time constants for the respective systems in some region R(h), where c_o and c_2 are considered fixed. As μ and η are complex functions of Q, obtaining the analytic maximum of α or β is difficult and we resort to the following <u>sub-optimum</u> in each case:

a) max η over $\overline{Q}$ then

b) minimize μ over a subspace of $\overline{Q}$,

where $\overline{Q}$ is the space of p.d.s. matrices.

Problem (a) was first solved by Lewis and Taussky (24). Problem (b) arises because Q and P giving the maximum η are not unique, and is the main content of this chapter.

Section 2.2 generalizes some previous work of Wiberg (23). The results of the remaining sections are believed to be new.

2.2 The Optimum

Assume A has linear elementary divisors (21), i.e there exists a transformation matrix S such that

$$S^{-1}AS = C \qquad 2.2.1$$

where C is a diagonal matrix of eigenvalues of A and the columns of S, $\underline{s}_i$, i = 1, n, are their eigenvectors chosen so that $\|\underline{s}_i\| = 1$, all i.

Select P as

$$P = ((SD)(SD)^*)^{-1} = (SD)^{-*}(SD)^{-1} \qquad 2.2.2$$

where * denotes conjugate transpose and $D = \text{diag}(d_1, d_2, \ldots \ldots, d_n)$ is an arbitrary diagonal matrix with $d_i \neq 0$. Then P is a p.d. hermitian matrix and substitution into 2.1.6 yields

$$Q = -(SD)^{-*}(C + C^*)(SD)^{-1} \qquad 2.2.3$$

which is also p.d. hermitian (p.d.h.) and not necessarily real. Then

$$\begin{aligned} \lambda(P^{-1}Q) &= -\lambda[(SD)(C + C^*)(SD)^{-1}] \\ &= -\lambda(C + C^*) \qquad 2.2.4 \end{aligned}$$

and

$$m(P^{-1}Q) = \eta = -2M(A) \qquad 2.2.5$$

$$M(P^{-1}Q) = -2m(A)$$

Vogt (19) has shown that for P and Q satisfying 2.1.6 we have the inequalities

$$2M(A) \leq -m(P^{-1}Q) \qquad 2.2.6$$

$$2m(A) \geq -M(P^{-1}Q) \qquad 2.2.7$$

the former showing that η in 2.2.5 is the maximum possible.

Statement 2.2.4 shows with suitable ordering of eigenvalues that $\lambda(P^{-1}Q) = -2\mathrm{Re}\,\lambda(A)$.

For a more complete statement consider a result of Barnett (20) who showed that if A is stable, then given a p.d. Q,

$$A = P^{-1}(R - \tfrac{1}{2}Q) \qquad 2.2.8$$

where P solves 2.1.6 and R is a skew-symmetric matrix found from 2.2.8, but which also solves

$$A^TR + RA = \tfrac{1}{2}(A^TQ - QA) \qquad 2.2.9$$

He gave the bound

$$|\beta_j| \leq M(iP^{-1}S) \quad (i=\sqrt{-1}) \qquad 2.2.10$$

where $\lambda_j(A) = \alpha_j + i\beta_j$, $j = 1,\ldots,n$.

It easily follows from 2.2.8 and the P in 2.2.2 that

$$2\lambda(P^{-1}R) = \lambda(C - C^*) = 2\mathrm{Im}\,\lambda(A)$$

showing that equality also holds in 2.2.10. Infact with suitable ordering of eigenvalues

$$\lambda(A) = \lambda(P^{-1}S) - \tfrac{1}{2}\lambda(P^{-1}Q)$$

Finally, if A is real, P in 2.2.2 may be taken real. For the $\underline{s}_i$ appear, if complex, in conjugate pairs. Suppose $\underline{s}_i$ and $\underline{s}_j$ are such a pair, then choose $d_i = d_j$. Then $\exists$ a symmetric permutation matrix T (30) (i.e. $T^T = T^{-1} = T$) such that

$$\overline{SD} = SDT$$

then
$$\overline{SD}^T = (SD)^* = T(SD)^T \quad \text{and}$$

$$P^{-1} = (SD)(SD)^*$$

$$= (SD)T(SD)^T$$

Hence P^{-1}, and thus P, are symmetric and thus real.

2.3 The Minimum Condition Number, (P)

The matrix P in 2.2.2 contains essentially n-1 arbitrary parameters to a multiplicative constant. Regarding μ as a function of P and thus D, problem (b) is restated as

$$\min_{D} \mu(P) \qquad 2.3.1$$

This quantity μ is called the P-condition number of P and is related to the condition number K(X) of a general matrix X defined as (Bauer (22))

$$K(X) = \|X\| \, \|X^{-1}\| \qquad 2.3.6$$

where $\|X\| = \sup_{x \neq 0} \|Xx\| / \|x\|$ 2.3.7

$$= (M(X^*X))^{\frac{1}{2}} \qquad 2.3.8$$

For Hermitian X, $\mu(X) = K(X)$ and for P in 2.2.2

$$K(P) = K(P^{-1}) = K((SD)(SD)^*)$$

$$= K((SD)^*(SD))$$

$$= K^2(SD) \qquad 2.3.9$$

Thus $\min_{D} \mu(P) = \min_{D} K^2(SD)$ 2.3.10

Now <u>define</u> $|X|$ as the matrix whose i,j th element is $|x_{i,j}|$. Such a matrix is called <u>non-negative</u>. We say X has <u>checkerboard sign distributions</u> (CSD) if matrices E_1 and E_2 exist such that $S = E_1 |S| E_2$ with $|E_1| = |E_2| = I$. We shall need the following theorem due to Bauer (22).

<u>Theorem 2.3</u>

For the matrix norm $\|\cdot\|$ in 2.3.8

$$\min_{D} K(SD) \leq \| \, |S| \, |S^{-1}| \, \| \qquad 2.3.11$$

with equality holding if both S and S^{-1} have CSD.

This theorem gives a useful bound on the minimum in 2.3.10

2.4 A Conjecture of Wiberg (23)

Wiberg conjectured that $\mu(P)$ is minimized if $D = I$ (i.e $\|s_i\| = 1$). Although this is true for $n = 2$ it is false for $n > 2$.

Choose $A = S^{-1}CS$ where

$$S = \begin{bmatrix} 1, & 1/\sqrt{2}, & 1/\sqrt{6} \\ 0, & 1/\sqrt{2}, & 2/\sqrt{6} \\ 0, & 0\ , & 1/\sqrt{6} \end{bmatrix} \qquad C = \text{diag}\,(-1, -2, -3)$$

With $D = I$, $\mu(P) = K^2(SD) = \underline{39.52}$

Now

$$S^{-1} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & \sqrt{2} & -2\sqrt{2} \\ 0 & 0 & \sqrt{6} \end{bmatrix}$$

Since both S and S^{-1} have CSD, by Theorem 2.3

$$\min_D K(SD) = \| |S| \; |S^{-1}| \| = (3 + \sqrt{10})$$

Then $\mu(P) = K^2(SD) = (3 + \sqrt{10})^2 = \underline{37.974}$ which disproves the conjecture.

2.5 An Optimum Class of Matrices

The diagonal matrix D giving the upper bound in 2.3.11 is given in A3 and would appear a better choice than $D = I$. A natural question is what form must A have in 2.1.1 and 2.1.2 so that S has CSD. We have:

Theorem 2.5

If A is a stable matrix in companion form with distinct real eigenvalues then S and S^{-1} may be chosen to have CSD.

Proof

For A in companion form (CF)

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & \cdot & & & & & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 1 \\ -a_n & -a_{n-1} & & \cdot & \cdot & \cdot & \cdot & -a_1 \end{bmatrix} \qquad 2.5.1$$

with $\lambda_i(A) = -\alpha_i < 0, i = 1, \ldots n$, we choose $S = V_n$ (and disregard unit columns for convenience) where V_n is the Van der Monde matrix (30)

$$V_n = \begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \lambda_1 & \lambda_2 & \cdot & \cdot & \cdot & \cdot & \cdot & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdot & \cdot & \cdot & \cdot & \cdot & \lambda_n^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & & & & & \cdot \\ & & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \lambda_1^{n-1} & \lambda_2^{n-1} & & & & & & \lambda_n^{n-1} \end{bmatrix} \qquad 2.5.2$$

V_n is clearly CSD i.e.

$$E_1 V_n = |V_n| \qquad 2.5.3$$

with $E_1 = \mathrm{diag}(1, -1, 1, \ldots, (-1)^{n-1})$ 2.5.4

To show that V_n^{-1} is CSD let σ_m , where

$$\sigma_m = \sigma_m(y_1, y_2, \ldots, y_n) \qquad 1 \le m \le n$$

$$= \sum y_{v_1} y_{v_2} \ldots y_{v_n}$$

$$\sigma_o = 1$$

be the m^{th} elementary symmetric function of any n variables y_i (30), e.g. $\sigma_1 = \sum^n y_i$ and $\sigma_n = \prod^n y_i$. Also define

$$\sigma_m^p = \sigma_m^p(y_1, y_2, \ldots y_{p-1}, y_{p+1}, \ldots y_n)$$

as the m^{th} elementary symmetric function of the y_i with y_p missing. Then the i,j^{th} element of V_n^{-1}, v_{ij}^{-1}, is given by (28)

$$v_{ij}^{-1} = (-1)^{j-1}\frac{\sigma_{n-j}^{i}(\lambda_1, \lambda_2, \ldots, \lambda_n)}{\prod_{k\neq i}^{n}(\lambda_k - \lambda_i)} \qquad 2.5.5$$

Numerically stable formulae exist to invert V_n (39)

Assume that $0 \leq \alpha_1 < \alpha_2 \ldots < \alpha_n$ and let

$$r_i = \prod_{j=1}^{i-1}(\alpha_i - \alpha_j)\prod_{j=i+1}^{n}(\alpha_j - \alpha_i) > 0 \qquad 2.5.6$$

Then in terms of α_i, V_n^{-1} may be written

$$V_n^{-1} = RQ \qquad 2.5.7$$

with $R = \mathrm{diag}(\frac{1}{r_1}, -\frac{1}{r_2}, \ldots, \frac{(-1)^{n-1}}{r_n})$

and

$$Q = \begin{bmatrix} \sigma_{n-1}^{1} & \sigma_{n-2}^{1} & \cdot & \cdot & \cdot & \sigma_1^{1} & 1 \\ \sigma_{n-1}^{2} & \sigma_{n-2}^{2} & & & & \sigma_1^{2} & 1 \\ \cdot & \cdot & & & & \cdot & 1 \\ \cdot & \cdot & & & & \cdot & 1 \\ \sigma_{n-1}^{n} & \sigma_{n-2}^{n} & \cdot & \cdot & \cdot & \sigma_1^{n} & 1 \end{bmatrix} \qquad 2.5.8$$

with $q_{i,j} = \sigma_{n-j}^{i}(\alpha_1, \ldots, \alpha_n) > 0$. Clearly, all elements of Q are non-negative and

$$V_n^{-1} = E_1 |V_n^{-1}| = |E_1| RQ$$

with E_1 from 2.5.4. Hence the theorem is proved.

In general the CSD property does not hold. It is not satisfied, for example, when A is in CF with complex $\lambda(A)$. In such cases one resorts to a non-linear programming technique to minimize μ (section 2.8) or chooses some upper bound such as that of Bauer, 2.3.11. In this respect some simple bounds can be found.

2.6 Bounds on $\mu(P)$

The best bound found by the author was from a result of Marcus and Haynsworth(27). They showed

$$K(P) = \mu(P) \leq \frac{1 + \sqrt{1 - D_1}}{1 - \sqrt{1 - D_1}} \qquad 2.6.1$$

where

$$D_1 = d(P)/(\frac{t(P)}{n})^n \qquad 2.6.2$$

and $d(P)$, $t(P)$ are the determinant and trace of P respectively. Let

$$F(P) = \frac{1 + \sqrt{1 - D_1}}{1 - \sqrt{1 - D_1}} \qquad 2.6.3$$

then in general $F(P) \neq F(P^{-1})$ and two bounds are possible in 2.6.1. Using the arithmetic-geometric mean inequality (2.6.12), each bound is minimized, for the choice of P in 2.2.2, when $D = I$ and $d_i = \|r_i\|$ respectively and we have

$$\mu(P^{-1}) \leq F(P^{-1}) \qquad 2.6.4$$

and

$$\mu(P) \leq F(P) \qquad 2.6.5$$

where

$$D_1(P^{-1}) = |d(S)|^2/\prod_{i=1}^{n}\|\underline{s}_i\|^2$$
$$= |d(S)|^2 \qquad 2.6.6$$

and

$$D_1(P) = (|d(S)|^2\prod_{i=1}^{n}\|\underline{r}_i\|^2)^{-1} \qquad 2.6.7$$

Here $\underline{r}_i$ = row i of S^{-1}.

The only difficult calculation in 2.6.6 or 2.6.7 is $d(S)$. Simplification arises with A in CF for then

$$d(V_n) = \prod_{i<j}(\lambda_j - \lambda_i) \text{ and}$$

$$D_1(P^{-1}) = \prod_{i<j}(|\lambda_j - \lambda_i|^2)/\prod_{i=1}^{n}(\sum_{j=1}^{n}|\lambda_i|^{2(j-1)}) \qquad 2.6.8$$

$$D_1(P) = \prod_{i<j} |\lambda_j - \lambda_i|^2 / \prod_{j=1}^{n} \left(\sum_{j=1}^{n} |\sigma^i_{n-j}|^2 \right) \qquad 2.6.9$$

(The latter follows because in 2.5.7 $|d(R)| = |d(V_n)|^{-2}$)

Note that equality occurs in 2.6.1 for n = 2.

A <u>looser</u> bound than 2.6.1 is proved as follows. Consider P in 2.2.2 with unit vectors $\underline{s}_i$, $\|\underline{s}_i\| = 1$, and D = I. Then let u_i, i = 1, ... , n, be the eigenvectors of T (T = S*S) ordered as $u_1 \geq u_2 \geq \ldots \geq u_n$.

Since $\mu(P) = K(T)$,

$$K(T) = \frac{u_1}{u_n} \qquad 2.6.10$$

If u is any eigenvalue of T, by Gershgorin's Theorem (30)

$$|u - |\underline{s}_i^* \underline{s}_i|| = |u - 1| \leq \rho_i, \quad i = 1,n$$

where

$$\rho_i = \sum_{j \neq i}^{n} |t_{i,j}| = \sum_{j \neq i}^{n} |\underline{s}_i^* \underline{s}_j|$$

Then $$u_1 \leq 1 + \max_i \rho_i = 1 + \bar{\rho} \qquad 2.6.11$$

Now $$d(T) = \prod_{j=1}^{n} u_i = u_n \prod_{j=1}^{n-1} u_i$$

Using the Arithmetic-Geometric mean inequality

$$\left(\prod^{n} u_i\right)^{\frac{1}{n}} = G_n \leq A_n = \left(\sum_{1}^{n} u_i\right)/n \qquad 2.6.12$$

we have, since $\sum u_i = n$,

$$d(T) \leq u_n \left(\frac{n - u_n}{n - 1}\right)^{n - 1}$$

$$\leq u_n \left(\frac{n}{n - 1}\right)^{n - 1} \qquad 2.6.13$$

Substitution of 2.6.11 and 2.6.13 into 2.6.10 yields the bound

$$\mu(P) \leq \frac{1}{|d(S)|^2} \left[\frac{n}{n - 1}\right]^{n - 1} (1 + \bar{\rho}) \qquad 2.6.14$$

An <u>invalid bound</u> was obtained by Wiberg (23) who used Gershgorin's inequality to obtain

$$\mu(P) \leq \frac{1 + \bar{\rho}}{1 - \bar{\rho}} \qquad 2.6.15$$

To show the bound is sometimes invalid consider A and S where

$$A = -\begin{bmatrix} 1 & 12/5 & 24/6 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad S = \begin{bmatrix} 1 & 12/13 & 12/13 \\ 0 & 5/13 & 0 \\ 0 & 0 & 5/13 \end{bmatrix}$$

for which $S^{-1}AS = \text{diag}(-1, -2, -3)$. Clearly, $\bar{\rho} = 24/13 > 1$ and the bound becomes negative.

2.7 Non-uniqueness of P in the Form 2.2.2

The form of P in 2.2.2 which satisfies the condition

$$\eta = m(P^{-1}Q) = -2M(A) \qquad 2.7.1$$

is not generally unique. We only need consider A stable and in the form

$$A = \left[\begin{array}{c|c} -a & \underline{0} \\ \hline \underline{0} & A_1 \end{array}\right] \qquad 2.7.2$$

where A_1 is $(n - 1) \times (n - 1)$ such that $S_1^{-1}A_1S_1^{-1} = C_1$. Let $M(A) = -a$, $M(A_1) = -a_1$ and $m(A_1) = -a_2$ $(a < a_1)$. P and Q are of the form

$$P = \left[\begin{array}{c|c} 1 & \underline{0} \\ \hline \underline{0} & P_1 \end{array}\right] \qquad Q = \left[\begin{array}{c|c} 2a & \underline{0} \\ \hline \underline{0} & Q_1 \end{array}\right] \qquad 2.7.3$$

where $A_1^TP_1 + P_1A_1 = -Q_1$. Then condition 2.7.1 is met with $P_1^{-1} = (S_1D_1)(S_1D_1)^*$ for which $\eta(Q_1) = m(P_1^{-1}Q_1) = 2a_1$ and

$M(P_1^{-1}Q_1) = 2a_2$. By varying P_1 and Q_1 and noting that ν depends continuously on the elements of P_1, a P_1 not generally of the form 2.2.2 can be found such that

$$2a < m(P_1^{-1}Q_1) < 2a_1 \qquad 2.7.4$$

Then we still have $m(P^{-1}Q) = 2a$, if $a < a_1$, but P not of the form 2.2.2.

Example

Choose

$$A = \left[\begin{array}{c|cc} -a & 0 & 0 \\ \hline 0 & 0 & 1 \\ 0 & -1 & -1 \end{array}\right] \qquad \lambda = -a,\ -\tfrac{1}{2}(1 \mp \sqrt{3}i)$$

Here, the only real minimizing P_1, to a constant factor, is

$$P_1 = \sqrt{\frac{4}{5}}\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} \quad , \quad \mu(P_1) = 3$$

but with the choice

$$P_1 = \frac{1}{2}\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$

$M(P_1^{-1}Q_1) = 1 + 1/\sqrt{5}$ and $m(P_1^{-1}Q_1) = 1 - 1/\sqrt{5}$ and 2.7.4 is satisfied if $2a < 1 - 1/\sqrt{5}$.

Also

$$\mu(P) = \frac{\sqrt{5} + 1}{\sqrt{5} - 1} < 3$$

and this P_1 gives smaller values of α or β in 2.1.15/16. Thus problem (b) does not necessarily give 'best' sub-optima of α or β.

The arbitrariness of the form 2.2.2 is due to the somewhat non-unique transformation S in 2.2.1. It can generate a subspace of p.d matrices (P) giving the same value of η . To see this, consider a solution P of 2.1.6, for a given p.d.s Q, in terms of the $\underline{s}_i$ and λ_i of A (31)

$$P = (S^*)^{-1}P_1S^{-1} \qquad 2.7.5$$

where

$$(P_1)_{ij} = -\underline{s}_i^*Q\underline{s}_j/(\bar{\lambda}_i + \lambda_j) \qquad 2.7.6$$

Let $\eta_p = m(P^{-1}Q)$, then P_o and Q_o also solve 2.1.6, with D arbitrary and diagonal, where $P_o = (SD)^{-*}P_1(SD)^{-1}$ and $Q_o = (SD)^*(S^*QS)(SD)^{-1}$. Further

$$P_o^{-1}Q_o = (SD)P_1^{-1}S^*QS(SD)^{-1}$$

and due to similarity properties

$$\lambda(P_o^{-1}Q_o) = \lambda(P_1^{-1}S^*QS) = \lambda(SP_1^{-1}S^*Q)$$
$$= \lambda(P^{-1}Q)$$

Then $\eta_p = \eta_{p_o}$

2.8 Numerical Optimum of the Condition Number

In view of the fact that for n = 2, the P of the form 2.2.2 which minimizes μ is real, it is reasonable to conjecture that this holds generally. To test the conjecture, Powell's conjugate gradient algorithm (A4) was used to minimize $\mu(P)$ over the N-dimensional space of elements of D, where for

(a) real P, N = n-k-1;

(b) hermitian P, N = n-1;

where $\lambda_j = r_j \mp is_j$, j = 1, 2, ... , k

$\lambda_j = -r_j < 0$, j = 2k+1, ... , n

are the eigenvalues of A. The bounds 2.3.11, 2.6.4 and 2.6.5 were also calculated.

Table 2.1 shows some results for a number of third and fourth order matrices (A1 to A10). The average number of function evaluations of $\mu(P)$ for an accuracy of 10^{-3}

TABLE 2.1

MATRIX A (LAST ROW ONLY GIVEN IF A IS IN C.F.)	NAME OF MATRIX	F(P)	$F(P^{-1})$	$\lVert \lvert S \rvert \lvert S^{-1} \rvert \rVert$	$P=(SS^*)^{-1}$ $\lVert s_i \rVert = 1$ (WIBERG)	MIN $\mu(P)$ P REAL	N	MIN $\mu(P)$	N
-8 1 5 4 -4 2 -18 5 7 $\lambda = -2 \mp 4i, -1$	A1	.1208 x 10^4	.14799 x 10^3	.19261 x 10^3	.14435 x 10^3	.14345 x 10^3	1	.14345 x 10^3	2
-33 -16 -72 24 10 57 8 4 17 $\lambda = -1, -2, -3$	A2	.58282 x 10^5	.17661 x 10^8	.2789 x 10^5	.3238 x 10^5	.26178 x 10^5	2	SAME	2
* -4 -1 -1 -2 -4 -1 0 -1 -4 $\lambda = -3, -3, -6$	A3	.29242 x 10^4	.21561 x 10^3	.22456 x 10^3	.21304 x 10^3	.200781 x 10^3	2	SAME	2
-101, -103, -3 $\lambda = -1, -1 \mp 10i$	A4	.242 x 10^4	.3165 x 10^3	.23774 x 10^3	.26098 x 10^3	.231718 x 10^3	1	.231756 x 10^3	2
-200, -202, -102 $\lambda = -100, -1 \mp i$	A5	93.085	48.984	50.886	42.269	41.7843	1	41.7854	2

* Derogatory

TABLE 2.1 (contd.)

-4,-10,-10,-5 λ= -2,-1,-1∓i	A6	.54903 x 10^{7}	.19123 x 10^{5}	.37229 x 10^{4}	.34319 x 10^{4}	.315129 x 10^{4}	2	.315984 x 10^{4}	3
-10,-18,-15,-8 λ=-5.94,-1,-.532∓1.184i	A7	.485 x 10^{3}	.2255 x 10^{3}	.20411 x 10^{3}	.15029 x 10^{3}	.14704 x 10^{3}	2	.14704 x 10^{3}	3
-202,-402,-304,-103 λ= -100,-1,-1∓i	A8	.19598 x 10^{5}	.96673 x 10^{3}	.44055 x 10^{3}	.3453 x 10^{3}	.34353 x 10^{3}	2	.34354 x 10^{3}	3
-24,-50,-35,-10 λ= -1,-2,-3,-4	A9	.13794 x 10^{12}	.33835 x 10^{8}	.27130 x 10^{6}	.35344 x 10^{6}	.27130 x 10^{6}	3	.27130 x 10^{6}	3
-202,-206,-107,-4 λ= -1∓10i,-1∓i	A10	.7066 x 10^{6}	.63389 x 10^{4}	.13486 x 10^{4}	.16222 x 10^{4}	.12751 x 10^{4}	1	.12747 x 10^{4}	3

in the minimum was 54 and 76 for $N = 2$ and $N = 3$ respectively.

Comparison of the minimum μ for real and complex P indicates that the conjecture may be true; certainly no advantage is gained by minimizing over the higher dimensional space of hermitian P as Wiberg suggests (23).

Of the two bounds $F(P)$ and $F(P^{-1})$, the latter gave better results and, except for matrices A_1, A_3, and A_5, both were inferior to Bauer's bound. Choosing μ with $D = I$ (Wiberg) gave some better bounds than that of Bauer (6 against 4), but as shown in theory, the latter gave the exact minimum for matrix A_9.

The examples were chosen with differing eigenvalue spreads and, for most, the minima for μ seem quite large. One might therefore expect quite crude estimates of the system response, independant of the non-linear terms in 2.1.1 and 2.1.2.

2.9 C in Jordan Form

When the elementary divisors of A are non-linear the analysis in 2.2 remains the same except that C is now in Jordan form and the relation 2.2.5 is replaced by (Vogt (19))

$$-2M(A) - \epsilon \leq \eta \leq -2M(A) + \epsilon$$

with $\epsilon > 0$ as small as desired. In theory one has to find the Jordan transformation S. In practice it is simpler to perturb the elements of A slightly, and one can replace A by $\bar{A}$ where $\bar{A} = A - D$ and D a diagonal matrix with sufficiently small positive elements, d_i, such that $M(\bar{A}) < 0$ and the eigenvalues $\lambda(\bar{A})$ are distinct. From continuity arguments, the optima of η and μ for $\bar{A}$ will differ only slightly from those of A. Sometimes, non-linear divisors cause little bother since round off errors in computation render the divisors linear (see example A3 Table 2.1).

2.10 Some Numerical Experiments

Prior to the work of this chapter the optimization techniques of Rosenbrock and Powell were applied to determine the minima of the quantities: (a) $s(P^{-1}Q)$, (b) $\mu(P)$, (c) $-\eta$ and (d) $-\eta^2/\mu$; the latter being important quantities in 2.1.15 and 2.1.16. The quantity $s(P^{-1}Q)$ is the spread of $P^{-1}Q$,

$$s(P^{-1}Q) = M(P^{-1}Q) - m(P^{-1}Q)$$

which is a bound on $M(A) - m(A)$, the spread of $\mathrm{Re}\,\lambda(A)$. An algorithm for minimizing (a) to (d) over p.d.s Q is given in A3. The main motivation was to test the usefulness of the bounds 2.2.6 and 2.2.7 in locating $\mathrm{Re}\,\lambda(A)$.

Table 2.2 shows some results applied to third order matrices, u_i, $i = 1, 3$ being the eigenvalues of $\frac{1}{2}P^{-1}Q$ ($u_1 \leq u_2 \leq u_3$). Here $s(P^{-1}Q)$ is minimized and average computation time was 110 mill/sec (I.C.L 1905) for 300 function evaluations (FE). In each case Q = I initially. The bounds for matrices A_3, B_3 and B_4 are good whereas for A_1, A_2 and B_5 there was a tendency for two u_i to become equal thus destroying the bound. This was particularly so for conjugate $\lambda(A)$ and suggests some theoretical reason for occuring. In all cases the minimizing Q was non-unique.

The results of minimizing quantities (b) to (d) are shown graphically in Fig 2.1 to 2.3 by plotting $\eta/2$ vs μ at stages in the minimization. Three matrices (A_1 to A_3)

TABLE 2.2

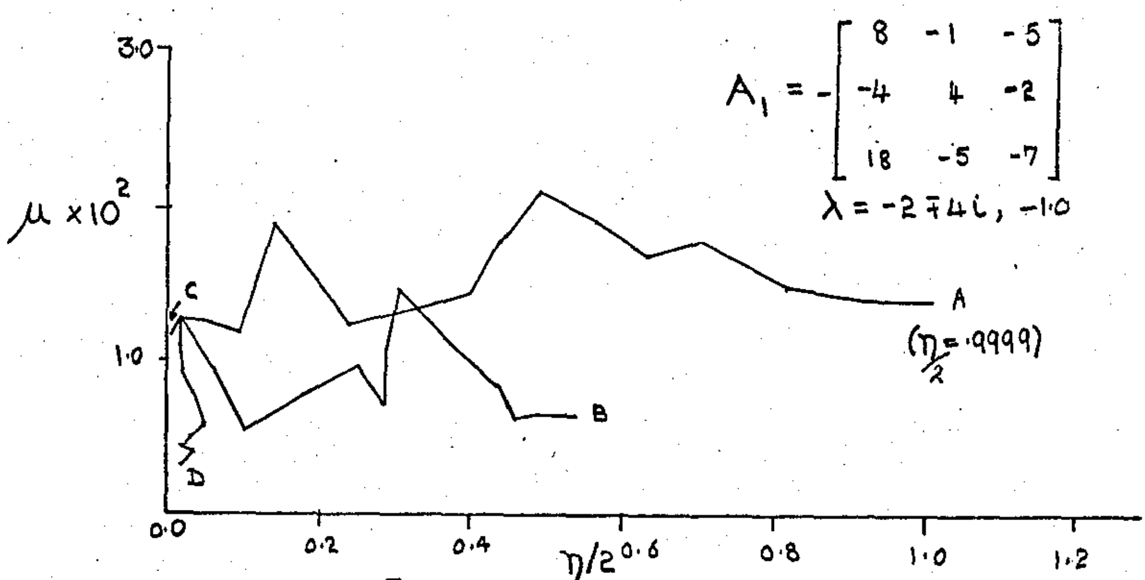
Minimization of $s(P^{-1}Q)$

A (LAST ROW GIVEN IF A IN CF (2.5.1))	NAME OF MATRIX	$\lambda(A)$	$\frac{1}{2}\lambda(P^{-1}Q)$ u_1, u_2, u_3	$\frac{1}{2}s(P^{-1}Q)$ 300 FUNCTION EVALUATIONS
10^{-6},0.01002,-.2001	B1	-10^{-4}, -.1, -.1	.1002,.0998,6×10^{-6}	.09988
-10^{4},1110.0,-111.0	B2	-100.0,-10.0,-1	100.35,10.64,.006	100.35
A3 TABLE 2.1	A3	-6, -3, -3	6.004,3.0042,2.9959	3.0034
A1 "	A1	-2 $\mp$ 4i, -1	2.9906,2.4906,.019	2.4716
A2 "	A2	-1, -2, -3	3.003,2.997,10^{-4}	3.002
-10 0 0 3 -3 0 2 1 -1	B3	-10, -3, -1	9.9899,3.024,.9957	8.0942
-1 0 -.01 -0.1 -1 0 0 -1 -1	B4	-1.1,-.95$\mp\frac{.05\ 3i}{2}$	1.102,.949,.949	.09153
-200,220,-21	B5	-10 $\mp$ 10i, -1	10.46,10.46,.089	10.371

are taken from table 2.1.

Consider maximizing η for A_3 (Fig 2.3). Initially $\eta = 4.8$ and $\mu = 1.54$, but after 300 FE's $\eta = 6.00$ and $\mu = 4.0 \times 10^3$ which give an α inferior to that of Table 2.1. However, assuming c_o constant in 2.1.15, initially, $\alpha = \eta/2 - \sqrt{\mu}\, c_o = 2.4 - \sqrt{1.54}\, c_o$. Taking values from Table 2.1, $\alpha = 3.0 - 200.78 c_o$ which is grossly inferior. This casts some doubt on the usefulness of problem (b). The matrix A_2 causes some lack in convergence as in Table 2.2. The quantity η^2/μ determines the size of the crude RAS and optimum values were superior to those calculated from Table 2.1.

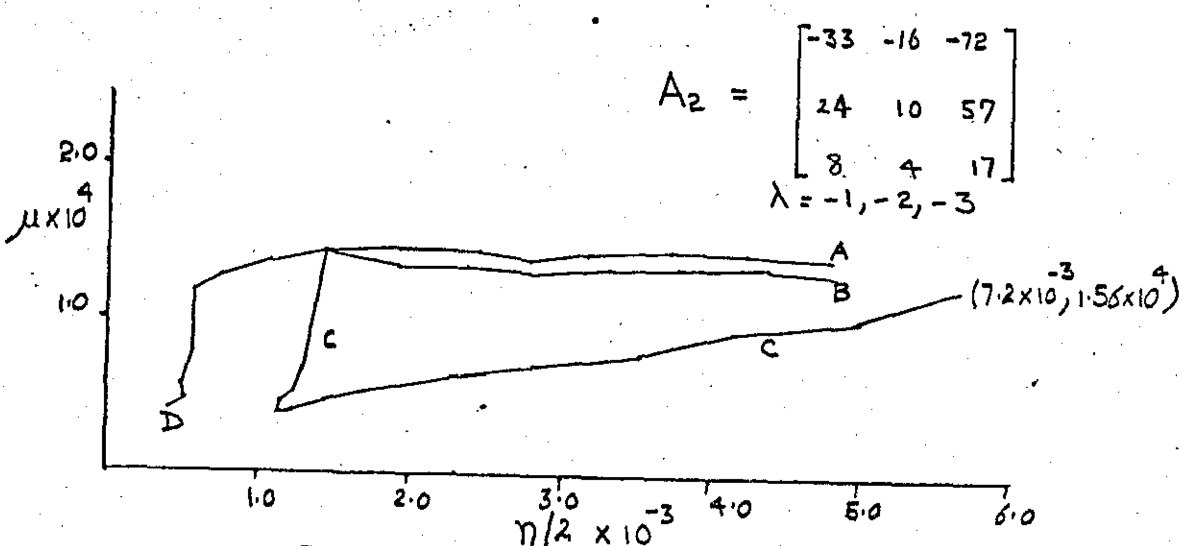
Finally, the use of the algorithm seems infeasible in locating the spread $M(A) - m(A)$ by minimizing $s(P^{-1}Q)$, due to high computing time (100 times greater for $n = 3$ than straight calculation of $\lambda(A)$) and the lack of convergence to the minima.



$$A_1 = -\begin{bmatrix} 8 & -1 & -5 \\ -4 & 4 & -2 \\ 18 & -5 & -7 \end{bmatrix}$$

$\lambda = -2 \mp 4i, -1.0$

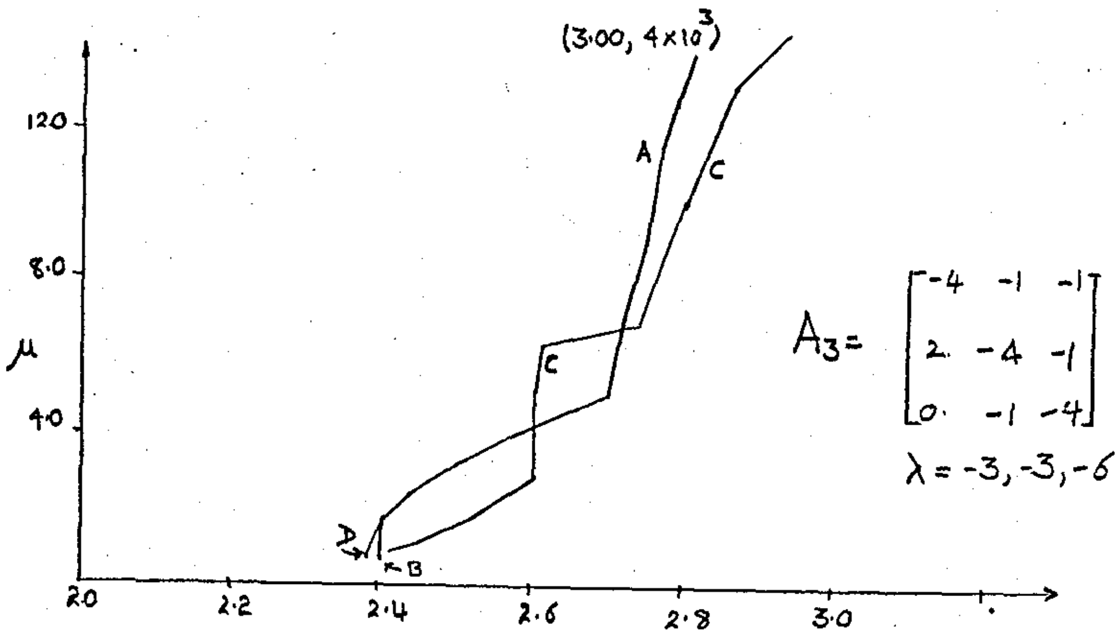
FIG. 2.1



$$A_2 = \begin{bmatrix} -33 & -16 & -72 \\ 24 & 10 & 57 \\ 8 & 4 & 17 \end{bmatrix}$$

$\lambda = -1, -2, -3$

FIG. 2.2



$$A_3 = \begin{bmatrix} -4 & -1 & -1 \\ 2 & -4 & -1 \\ 0 & -1 & -4 \end{bmatrix}$$

$\lambda = -3, -3, -6$

FIG. 2.3 $\eta/2$

A: $\min(-\eta)$ B: $\min(-\eta^2/\mu)$

C: $\min[M(P^{-1}Q) - m(P^{-1}Q)]$ D: $\min \mu(P)$

CHAPTER 3.

THE DETERMINATION OF STABILITY REGIONS BY ZUBOV'S APPROACH

Chapter 3

The Determination of Stability Regions by Zubov's approach

3.1 Introduction

Although applicable to the theory of dynamic systems in general, the work of Zubov (1) has shown greatest use in the determination of regions of asymptotic stability (RAS) of autonomous non-linear differential equations. It reduces the choice of a Lyapunov function to the solution of a partial differential equation (PDE), the exact solution of which determines the precise D.O.A.

3.2 The Main Theorem

Consider the autonomous differential equation (d.e.)

$$\dot{\underline{x}} = \underline{f}(\underline{x}) \qquad 3.2.1.$$

where $\underline{f}(\underline{o}) = o$ and $\underline{f} \in E$, $\quad \underline{x} \in E^n$

and where $\underline{x} = o$ is an asymptotically stable equilibrium point. Then the core of Zubov's treatment lies in the following theorem (1).

Theorem 3.1.

Let U be an open region containing the origin and $\bar{U}$ its closure.

Then a necessary and sufficient condition for U to be the domain of attraction (DOA) of system 3.2.1. is the existence of two functions $W(\underline{x})$ and $\Psi(\underline{x})$ with the properties:

a) $W(\underline{x})$ is defined and continuous in U

b) $\Psi(\underline{x})$ is positive definite and continuous in E^n

c) $0 < W(\underline{x}) < 1$ for $\underline{x} \in U$, $\underline{x} \neq 0$

d) if $\underline{y} \in B$ $(B \equiv \bar{U} - U)$ then $\lim_{\underline{x} \to \underline{y}} W(\underline{x}) = 1$

and if $\|\underline{x}\| \to \infty$ for $\underline{x} \in U$, $\lim_{\|\underline{x}\| \to \infty} W(\underline{x}) = 1$

e) $\frac{dW}{dt} = \dot{W} = \sum_{i=1}^{n} \frac{\partial W_i}{\partial x_i} f_i = -\psi(\underline{x})(1 - W(\underline{x}))(1 + \|\underline{f}\|^2)^{\frac{1}{2}}$

By assuming the f_i terms are bounded the factor $(1 + \|\underline{f}\|^2)^{\frac{1}{2}}$ in (e) may be removed giving the main PDE

$$\dot{W}(\underline{x}) = \sum_{i=1}^{n} \frac{\partial W}{\partial x_i} \cdot f_i = -\phi(\underline{x}) \ (1 - W(\underline{x})) \qquad 3.2.2$$

where W satisfies (c), which we call the regular equation.

By defining another p.d. function

$$V(\underline{x}) = -\ln\ (1 - W(\underline{x}))$$

3.2.2 may be transformed into

$$\dot{V} = \sum_{i=1}^{n} \frac{\partial V_i}{\partial x_i} f_i = -\phi(\underline{x}) \qquad 3.2.3$$

which we call the modified equation. The solutions W and V of 3.2.2 and 3.2.3 for arbitrary $\phi(\underline{x})$ then give for the boundary of the D.O.A., B, either of the sets

$$W(\underline{x}) = 1 \qquad 3.2.4$$

or

$$V(\underline{x}) = \infty \qquad 3.2.5$$

In what follows we concentrate on 3.2.3 for convenience (P.D.E. 3.2.2 can be treated similarly).

3.3. The Construction Procedure

We make the following assumptions:

(a) the components of $\underline{f}$ may be expanded as convergent power series about $\underline{x} = o$.

i.e.

$$f_i(x) = \sum a\,(i_1, i_2, .., i_n) x_1^{i_1} x_2^{i_2} .. x_n^{i_n}$$

(b) the linear part is asymptotically stable.

In all but the simplest cases the analytic solution for V of 3.2.3 is impossible. Consequently, Zubov proposed the following procedure.

Express V and ϕ as power series of arbitrary degree m and l respectively, and write

$$V_1^m = V_2 + V_3 + \quad . \quad . \quad + V_m \qquad 3.3.5$$

$$\phi^l = \phi_2 + \phi_3 + \quad . \quad . \quad + \phi_l \qquad 3.3.6$$

where V_i, ϕ_i are homogeneous polynomials of degree i and ϕ_2 is positive definite.

Substitute these series for V, ϕ and $\underline{f}$ into 3.2.3 and equate coefficients of like-powers. There then results a set of simultaneous equations for the coefficients of the powers of the V_i terms, which may be solved successively for those of V_2, V_3, onwards. Since ϕ_2 is p.d. V_2 is by (b) above and termination of the series 3.3.5 for any **finite** m will result in an m - th degree Lyapunov function (2). An RAS, with boundary $V^m = C_m$, can then be obtained with recourse to Theorem 1.6.1.

A wealth of experience on the determination of stability regions, (2) - (6), has shown that the series procedure may be non-uniformly convergent in that higher degree Lyapunov functions can result in inferior R A S 's. It is difficult to show why or how this happens in general, but the following example throws some light on the issue.

3.4 An Example (Zubov)

The system is

$$\begin{aligned} \dot{x} &= -x + 2x^2y \\ \dot{y} &= -y \end{aligned} \qquad 3.4.1.$$

whose D.O.A. is $xy < 1$

Write 3.3.5 as

$$V^m(x,y) = \sum_{i=2}^{m} \sum_{j=1}^{i+1} a_{ij}\, x^{i-j+1}\, y^{j-1} = \sum_{i=2}^{m} V_i \qquad 3.4.2.$$

and choose $\phi = 2\,(x^2 + y^2)$

Then substitution into 3.2.3. gives

$$\frac{\partial V^m}{\partial x}(-x + 2x^2y) + \frac{\partial V^m}{\partial y}(-y) = -2(x^2 + y^2)$$

or

$$-\sum_{i=2}^{m} i\, V_i + 2x^2y \sum_{i=2}^{m} \frac{\partial V_i}{\partial x} = -2(x^2+y^2) \qquad 3.4.3.$$

Equating coeffts.of like powers and putting $u=xy$ gives the 2m-th deg. Lyapunov Function (LF),

$$V^{2m} = y^2 + x^2(1 + u + u^2 \quad . \; . + u^{m-1}) \qquad 3.4.4$$

Assuming $|u| < 1$, and taking the limit, actually gives the analytic solution

$$V = y^2 + \frac{x^2}{1-xy} \qquad 3.4.5$$

However numerical calculation of the RAS's for various m by the method of Rodden (3) gives the RAS boundaries in Fig. 3.1. ($m = 51$ was near to the overflow value of m in computation).

The average radius (1.14) for $m = 50$ is slightly smaller than that for the quadratic (1.24), and the procedure is highly non-uniform. Considering the stability boundaries for m odd and m even seperately, we appear to have uniform convergence in each case, those for m odd being superior. Analysing the example further indicates why this should happen.

For $m = 2$ and $m = 4$ analytic calculations give the following RAS boundaries and tangency points,

$$V^4 = y^2 + x^2\,(1 + u) = \frac{4\sqrt{3}}{9} \qquad 3.4.6$$

with Tan. pts. $\mp\left(-\sqrt{\frac{2}{\sqrt{3}}} \,,\, \sqrt{\frac{2}{3\sqrt{3}}} \right)$

and

$$V^8 = y^2 + x^2\,(1 + u + u^2 + u^3) = \cdot 93913 \qquad 3.4.7$$

with Tan. pts. $\mp$ (-1.0642, -·68486)

These are also the largest <u>closed</u> contours and ∇V, the gradient of V, vanishes at the respective tangency points.

We now show that for any even m there are points where $\nabla V^{2m}(\underline{x}) = 0$ for $u > -1$, and that the RAS boundaries lie inside $|u| < 1$.

From 3.4.4 the gradient vanishes when

$$\frac{\partial V^{2m}}{\partial x} = V_x^{2m} = x(2P(u) + Q(u)) = 0 \qquad 3.4.8$$

and

$$\frac{\partial V^{2m}}{\partial y} = V_y^{2m} = 2u + x^4 Q(u) \quad = 0 \qquad 3.4.9$$

where

$$P(u) = (1 - u^m)/(1 - u) > 0 \quad , u > -1 \qquad 3.4.10$$

and

$$Q(u) = 1 + 2u + 3u^2 \ldots + (m - 1)u^{m-2}.$$

Clearly $\exists$ a u, $-1 < u < 0$, such that

$$F(u) = 2P(u) + Q(u) = 0 \qquad 3.4.11$$
$$= 2 + 3u + 4u^2 + \ldots + (m + 1)u^{m-1}$$

since $F(0) = 2$ and $F(-1) < 0$.

If $F(\bar{u}) = 0$ then 3.4.11 implies by 3.4.10 that

$$Q(\bar{u}) = -2P(\bar{u}) < 0$$

Then with $u = \bar{u}$, equation 3.4.9 will vanish for some $x = \bar{x} \neq 0$ because this implies

$$\bar{x}^4 = \frac{-2\bar{u}}{Q(\bar{u})} = \frac{\bar{u}^2}{P(\bar{u})} > 0$$

Thus, $\exists$ at least two radially symmetric points where

$\nabla V = 0$, namely $\mp (\bar{x}, \bar{u}/\bar{x})$.

Let $V^{2m}(x,y) = C_m$ be the largest closed RAS boundary. Then it lies in $u > -1$. For otherwise $\exists$ two distinct pts., (x_1,y_1), (x_2,y_2) say, in common with $xy = -1$ and this boundary. By 3.4.4, at these points.

$$V^{2m}(x_1,y_1) = y_1^{\,2} = V^{2m}(x_2,y_2) = y_2^{\,2}$$

then $y_1 = y_2 \Rightarrow x_1 = x_2$, a contradiction. Finally the boundary lies inside $u < 1$ since this is the D.O.A.

Consider now m odd. Summing the geometric series in the u terms in 3.4.4 gives

$$V^{2m} = y^2 + x^2 (1 - u^m) / (1-u) \qquad 3.4.12$$

the contours of which will always be closed since it is p. d. and radially unbounded. Hence there is no restriction to the convergence of the RAS. to the DOA for increasing m odd.

Comment

Let R be the region of convergence of the series 3.5.5 with partial sum V^m. We have shown an example where R ($|u| < 1$) is a subset of the DOA, where some truncations (m even) always give inferior RAS's, and where others (m odd) give better ones but converge slowly to the DOA (m = 51, still poor?). We ask ' is convergence of the RAS to the DOA always non-uniform if $R \subset U$?' No answer has yet been forthcoming.

3.5. Direct Numerical Solution of the P.D.E.

Due to the non-uniformity of the Zubov procedure and the infeasibility of solving the series solution for large n(2), even though formulation is possible (7), it seems reasonable to look for other methods of solving 3.2.3. In this section we consider solving the P.D.E. by a finite difference method. We restrict the problem to two dimensions.

Write system 3.2.1. as

$$\begin{aligned} \dot{x} &= f(x,y) \\ \dot{y} &= g(x,y) \end{aligned} \qquad 3.5.1.$$

Then we write the P.D.E.'s 3.2.2 and 3.2.3 as one equation.

$$\frac{\partial V}{\partial x} f + \frac{\partial V}{\partial y} g = -\phi \quad + \quad [\phi V] \qquad 3.5.2.$$

where [ ] is included for 3.2.2 only, the regular equation.

From a classification point of view, 3.5.2 is a linear PDE solvable by the method of characteristics (8). Unfortunately this leads us back to the solution of the system trajectories.

Consequently, we express 3.5.2 in polar co-ordinates

$$\frac{\partial V}{\partial r} \cdot F(r, \theta) + \frac{\partial V}{\partial \theta} \cdot G(r,\theta) = -\phi + [\phi V] \qquad 3.5.2$$

where $F = (f \cos\theta + g \sin\theta)$

$$G = \frac{1}{r}(g \cos\theta - f \sin\theta)$$

Consider the mesh in Diag. 1.

where

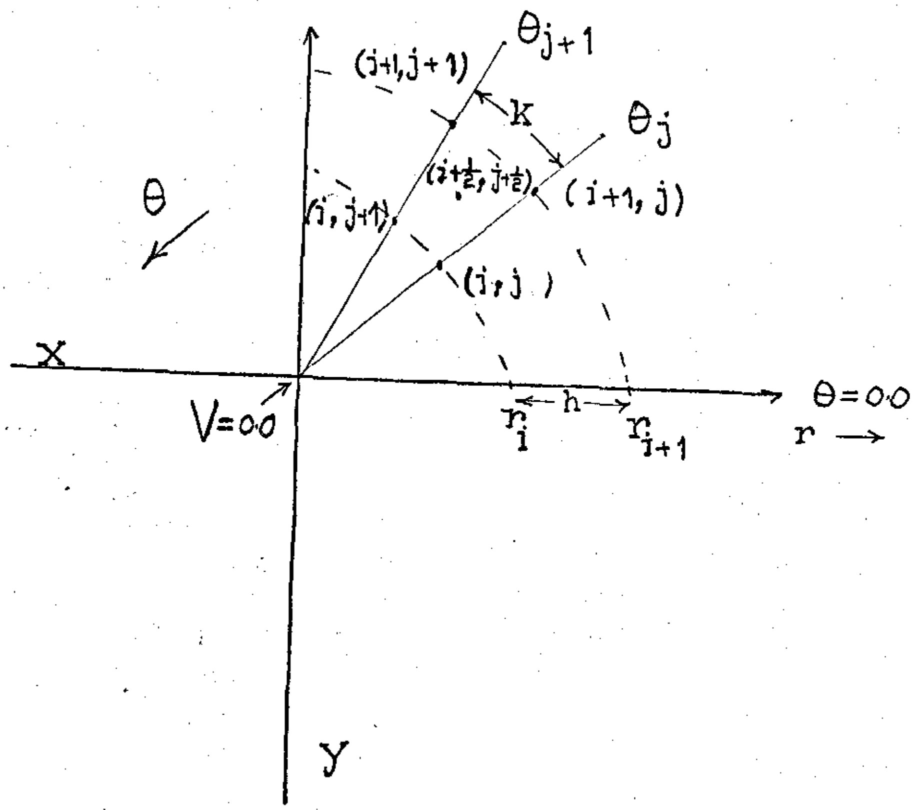
$r_i = ih, \; i = 1,2,..$

$\theta_j = (j-1)k, \; j=1, ., N+1$

$k = 2\pi / N, \; \theta_1 = \theta_{N+1}$

and $V_{i,j} = V(ih, (j-1)k)$

Diag. 1



Choosing $\phi = \lambda(ax^2 + 2bxy + cy^2)$ and making the approximations about the mesh point $(i + \frac{1}{2},\ j + \frac{1}{2})$

$$\left(\frac{\partial V}{\partial r}\right) \simeq \frac{1}{2h}((V_{i+1,j} + V_{i+1,j+1}) - (V_{i,j} + V_{i,j+1})) + 0(h,k)$$

$$\left(\frac{\partial V}{\partial \theta}\right) \simeq \frac{1}{2k}((V_{i,j+1} + V_{i+1,j+1}) - (V_{i,j} + V_{i+1,j})) + 0(k,h)$$

substitution into 3.5.3 gives the following Crank-Nicholson type difference equations (9) on neglecting second order terms in h and k,

$$V_{i+1,j}a_j + V_{i+1,j+1}b_j = c_j \qquad 3.5.6$$

$$j = 1, \ldots, N$$

where $V_{N+1} = V_1$

and

$$a_j = F - pG - z \qquad 3.5.7$$

$$b_j = F + pG - z \qquad 3.5.8$$

$$c_j = V_{i,j}(b_j + [2z]) + V_{i,j+1}(a_j + [2z]) - 4z \qquad 3.5.9$$

$$p = h/k, \quad z = h\phi/2$$

The terms in [ ] are zero for the modified eqn. (ϕ, F and G are evaluated at $(i+\frac{1}{2},\ j+\frac{1}{2})$ in 3.5.7 and 3.5.8). Since $V(0,0) = 0$, we can assume

$$V(r_1,\theta_j) = V(h,jk) = \epsilon, \quad j = 1, \ldots, N \qquad 3.5.10$$

for initial conditions ($\epsilon > 0$, small), or, make

$$V(h,jk) = V_2(h,jk) \quad j = 1, \ldots, N$$

V_2 being obtained from the series 3.3.5.

We have an initial value problem (9) and by writing 3.5.6 as a linear system of equations

$$A\underline{V}_{i+1} = \underline{c} \qquad 3.5.11$$

where $\underline{c} = (c_1, c_2, \quad . \quad , c_N)$,

$$\underline{V}_{i+1} = (V_{i+1,1}, \ldots\ldots\ldots V_{i+1,N})$$

and A an N x N coefficient matric, an efficient computer algorithm results for their solution (A2).

3.6 Numerical Examples

The following examples show features of the method.

Example 3.6.1

The system is

$$\dot{x} = -x + y + x\,(x^2 + y^2)$$
$$\dot{y} = -y - x + y\,(x^2 + y^2)$$

With $\phi = 2r^2$, $N = 100$, $h = .0125$, $r_1 = .25$

Solution of the regular equation gave V = 1.0000 for all mesh points on r = 1.0. The analytic solution is $V = r^2$ giving $r^2 = 1.0$. for B. With ϕ changed to $\phi = 10\,x^2 + y^2$, V = 1.0000 again on $(1.0, \theta_j)$. Constructing the V contours shows they are no longer concentric circles as above but ellipses, approaching $r^2 = 1$ as $r \to 1$. Fig. 3.2 shows variation of V with θ for fixed r. Note how the numerical solution of V becomes negative for $r > 1$, indicating that the analytic solution is not defined for such r.

Example 3.62.

A Van der Pol equation.

$$\dot{x} = y$$
$$\dot{y} = -x - y + x^2 y$$

Fig. 3.5 shows attempted constructions of some V contours from mesh values, for the regular equation, with h =.0125,

$r_1 = \cdot 25$, $\phi = r^2$, and $N = 100$. Singularities occured at the DOA boundary, B, as in 3.6.1, with V taking on large negative values ($V \to -\infty$)
Choice of mesh was important since
$f = y = o$ for $\theta = o$. If $\theta_{j+\frac{1}{2}} = o$ for some j, the method breaks down.

Example 3.6.3

The system is 3.4.1. With $\phi = 2(x^2 + 2xy + 2y^2)$ and initially $V(r_1,\theta) = V_2 = \phi/2$, $h = \cdot 0125$ $r_1 = \cdot 25$ and $N = 100$, errors between the numerical and analytic results were less than 10^{-4} for the regular equation, for $r < 1\cdot 375$. As $r \to 1\cdot 4$, the $V_{i,j} \to -\infty$ rendering computed values useless. The reason is found in the analytic solution.

$$V = 1 - e^{-(y^2 + x^2/(1-xy))}$$

since $V \longrightarrow -\infty$ as $xy \longrightarrow 1+$
Fig. 3.3 shows a construction of the V contours, while Fig. 3.4, the variation of V along a ray to B for various λ

Comment

If numerical values are correct the mesh points at which $V_{i,j} < 1$ lie in U, the DOA. Unfortunately the method breaks down near the DOA boundary and the values at such mesh points, $V_{i,j}$, are such that $V_{i,j} \to -\infty$. The accuracy of other points where $V_{i,j} < 1$ is then decreased.

Suitable choice of λ is also a problem. Large values of λ means $V \to 1$ quickly (Analytically, if V solves $\dot{V} = -\phi$ then $\hat{V} = \lambda V$ solves $\dot{\hat{V}} = -\lambda\phi$) and conversely for λ small (Fig.3.4).

Numerically, a criterion such as

$$0 < 1 - V_{i,j} < 10^{-t}$$

gives a point 'near' D. In general t will depend on λ.

Finally, we add that other difference schemes for solving 3.5.3 have given similar results.

3.7 Recent Methods

It is pertinent in view of section 3.6 to mention other methods of solving Zubov's PDE 3.2.2/3.2.3. Burnand and Sarlos (1968) (10), treat 3.2.3 and system 3.2.1 as two differential equations. Using Lie series, the two equations in question,

$$\dot{\underline{x}} = \underline{f}(\underline{x}) \qquad 3.7.1.$$

$$\dot{V} = -\phi(\underline{x}) \qquad 3.7.2.$$

with ϕ p.def., are integrated with reverse time from an initial point $\underline{x}$ (o) near the origin (typically $V(\underline{x}(0)) = 0.0$) until $V > K$, where K is a large positive number (e.g. K = 20). Hopefully the final point $\underline{x}(t)$ is near B. The procedure is repeated for a series of initial points until B is traced out.

Kormanik and Li (1972) (13) have extended the method and fit an algebraic curve to the final set of points by a pattern classification algorithm.

Troch (1972) (12), with computation efficiency in mind, integrates system 3.7.1. and 3.7.2 by analogue computer. The termination criterion $V>K$ determines a point on B and a more accurate digital computation integrates a trajectory through this point. The trajectory (or a series of such) will trace out B.

<u>Comment</u>

From limited computer experience we found Burnand/ Sarlos's method suffered from two drawbacks:

a) although accurate, computer time was high for the Lie series computation,

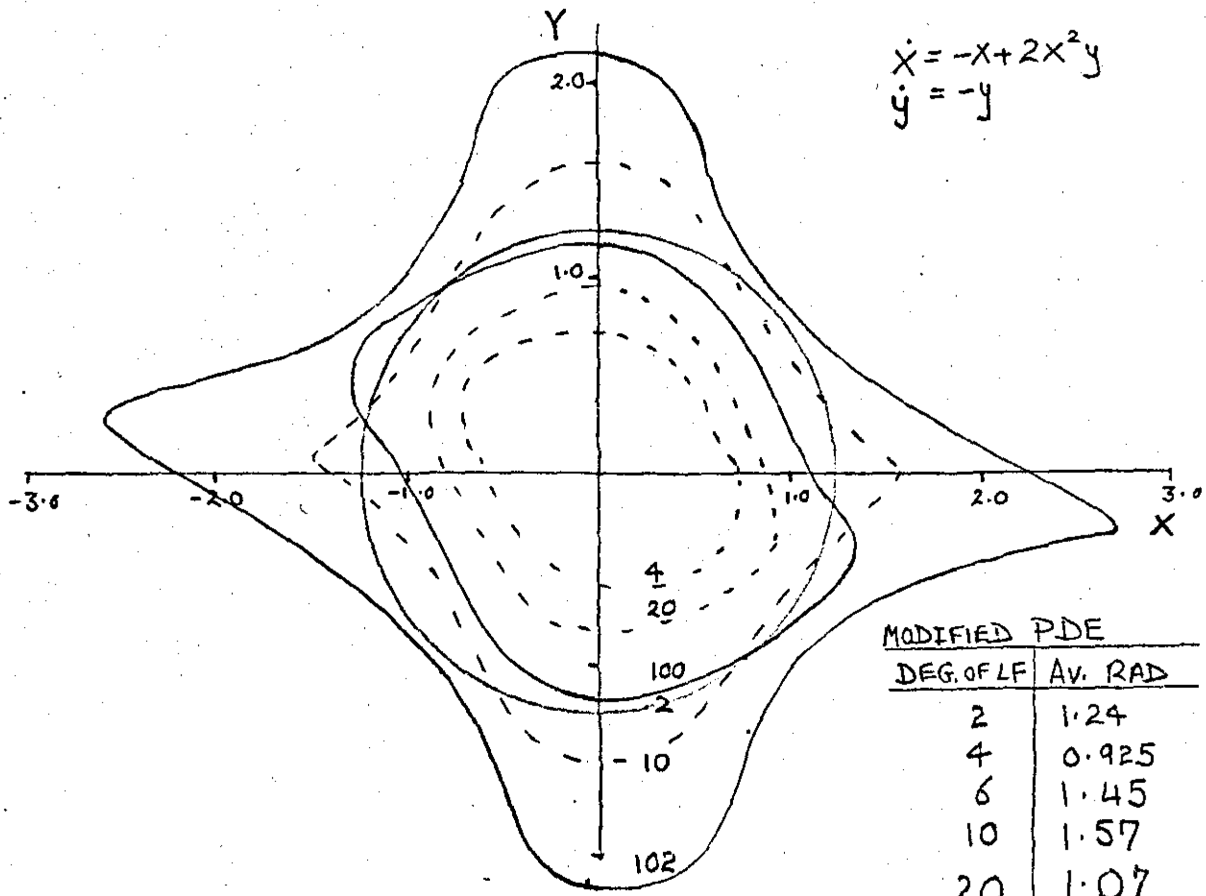
b) calculation of the <u>recursive terms</u> (10) in the Lie series was infeasible for complex r.h.s. of 3.7.1.

As a criticism of both methods we ask, is the additional solution of 3.7.2 really necessary since the criterion $V>k$ (or '$1 - W < \epsilon$, where $W = 1-\bar{e}^{V}$) is somewhat arbitrary as a criterion for 'nearness' to B? The non Lyapunov method of Davidson and Cowan (1969) (11), restricted to n = 2, gives a partial answer. For their criterion they use the function

$$V\ (\underline{x}\ (to)) = \|\underline{x}\ (t_o)\| - \|\underline{x}\ (t_o + T)\|$$

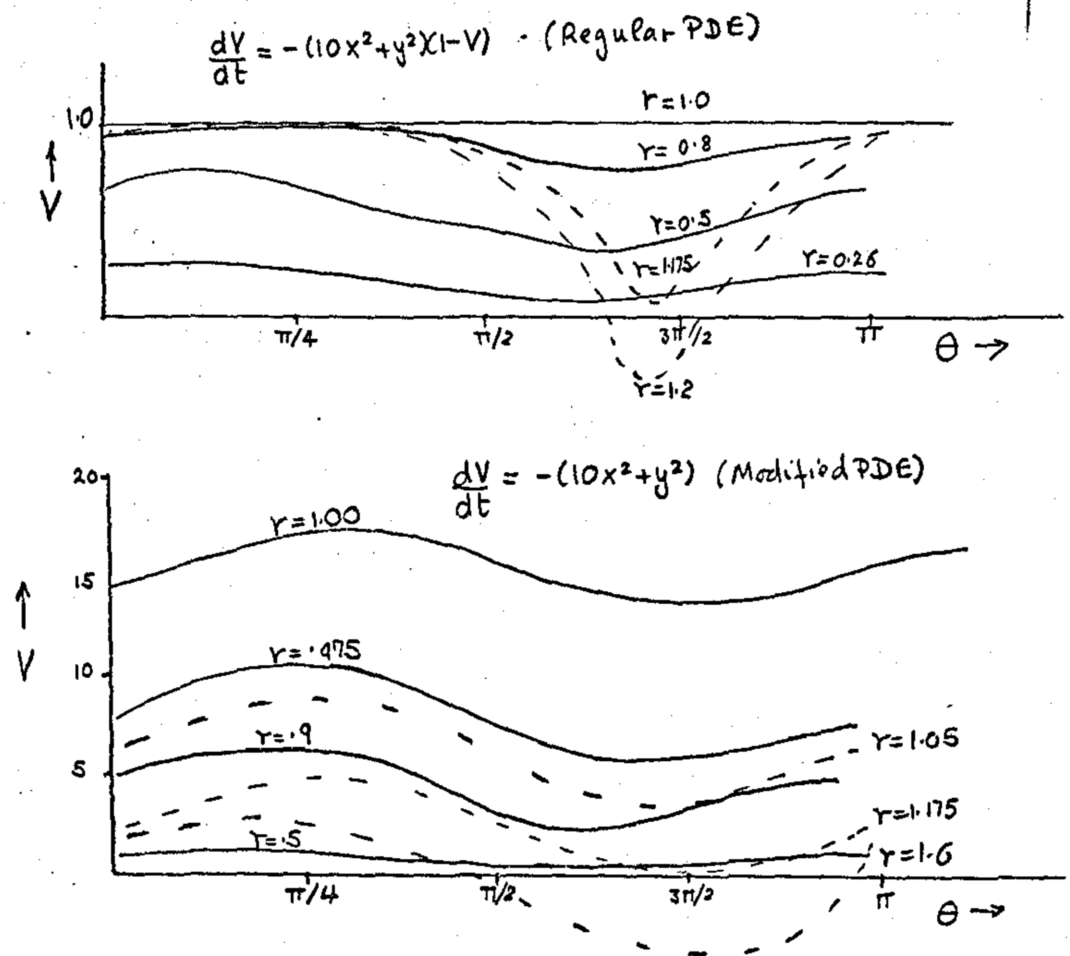
where T is the periodic time (for limit cycle) or an upper bound (for nodal type systems). Equation 3.7.1 is integrated by a fourth order Runge Kutta formula and a point satisfying $V(\underline{x}(t)) = 0$ determines a point on B(13).

Finally, it is questionable whether any of the above methods are useful for $n > 2$ in view of the computation time involved.



MODIFIED PDE	
DEG. OF LF	AV. RAD
2	1.24
4	0.925
6	1.45
10	1.57
20	1.07
(m=50) 100	1.14
(m=51) 102	1.81

FIG. 3.1



System: $\begin{cases} \dot{x} = -x + y + x(x^2+y^2) \\ \dot{y} = -x - y + y(x^2+y^2) \end{cases}$

FIG. 3.2

VARIATION OF V ALONG RAY $\theta = 41^\circ 40'$ FOR VARIOUS λ.

REGULAR PDE $\frac{dV}{dt} = -\phi(1-V)$

SYSTEM:

$\dot{x} = -x + 2x^2 y$

$\dot{y} = -y$

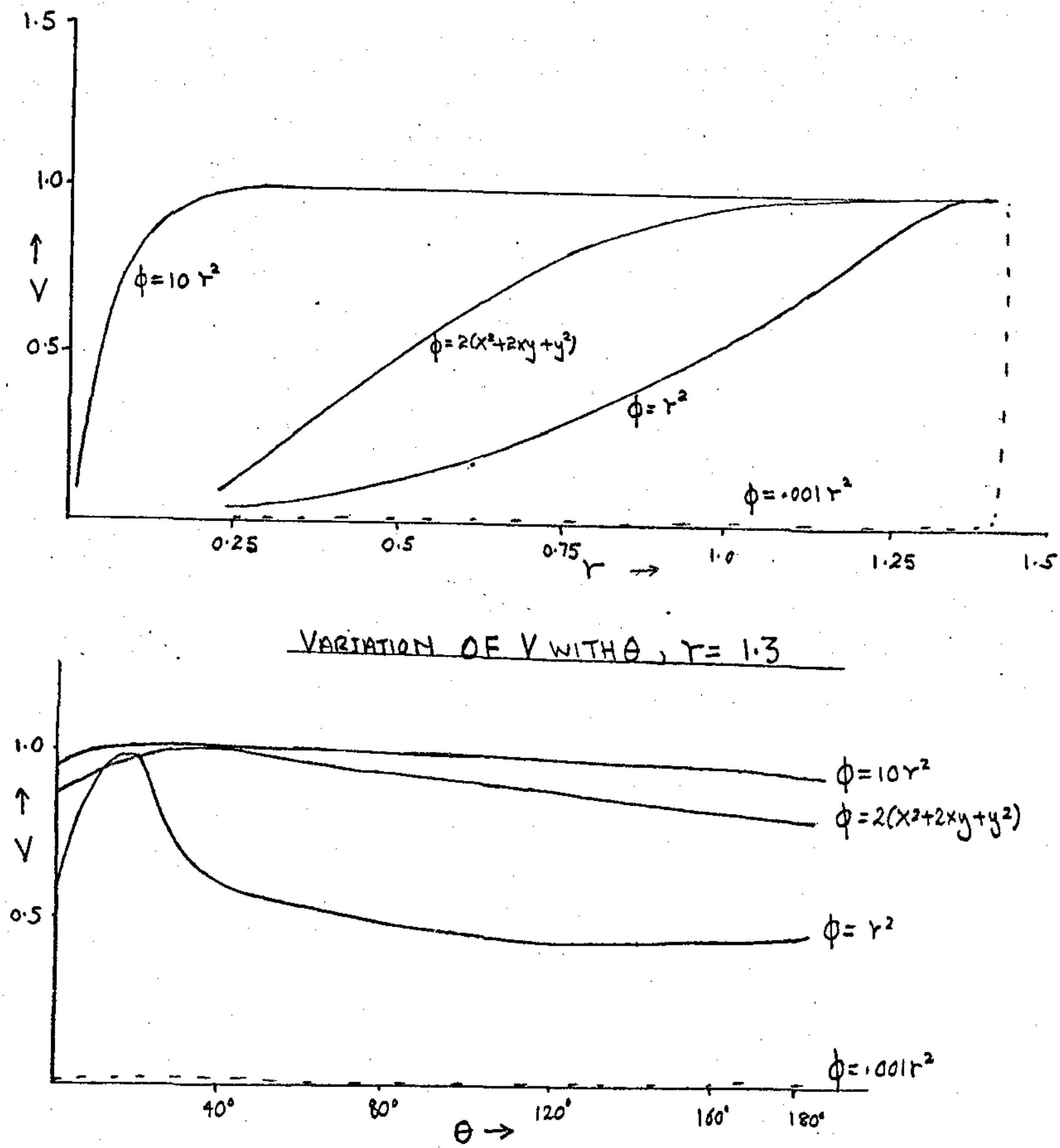


FIG. 3.4

VAN DER POL SYSTEM

ATTEMTED CONSTRUCTION OF V-CONTOURS FROM MESH POINTS.

$$\frac{dV}{dt} = -(x^2+y^2)(1-V)$$

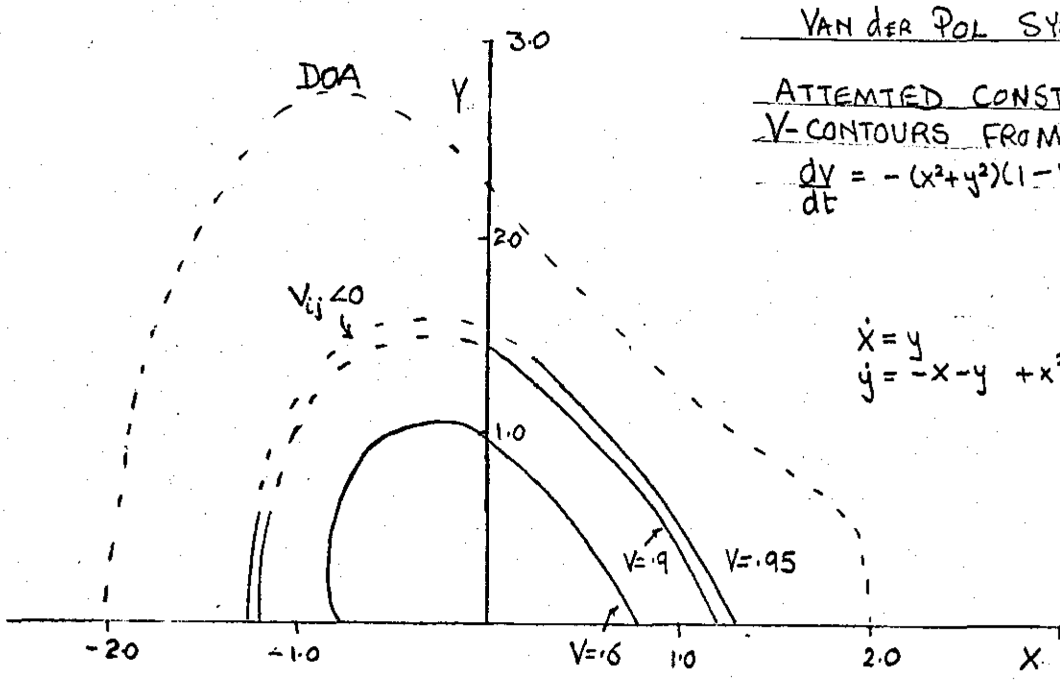


FIG 3.5

SYSTEM:

$\dot{x} = -x + 2x^2 y$

$\dot{y} = -y$

ZUBOV EXAMPLE

ATTEMTED CONSTRUCTION OF V-CONTOURS FOR REG. EQN. , $\frac{dV}{dt} = -2(x^2+2xy+y^2)(1-V)$

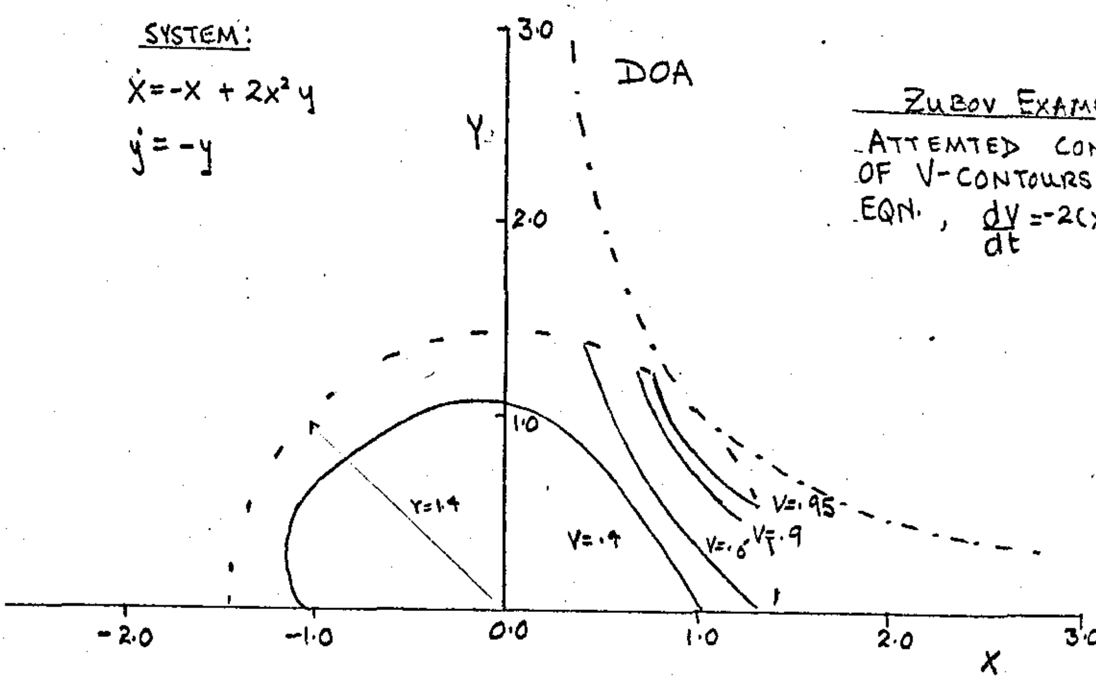


FIG. 3.3

CHAPTER 4

OPTIMAL QUADRATIC LYAPUNOV FUNCTIONS

CHAPTER 4

Optimal Quadratic Lyapunov Functions

4.1 Introduction

We will be concerned with the autonomous system 1.4.3 namely,

$$\dot{\underline{x}} = A\underline{x} + \underline{g}(\underline{x}) \qquad 4.1.1$$

where A is stable and $\underline{g}$ is expandable as a power series with terms of at least degree two in x_i, the coefficients of $\underline{x}$.

A quadratic LF for 4.1.1, $V(\underline{x}) = \underline{x}^T P\underline{x}$, is determined via Theorem 1.6.1 whose RAS is given by

$$\underline{x}^T P\underline{x} < V_m \qquad 4.1.2$$

where

$$V_m = \min_{\underline{x}} V(\underline{x}) \text{ with } \underline{x}\in E_V \qquad 4.1.3$$

Here

$$E_V : (\underline{x}/\dot{V} = -\underline{x}^T Q\underline{x} + 2\underline{x}^T P\underline{g}(\underline{x}) = o,\ \underline{x} \neq o) \qquad 4.1.4$$

and Q is any p.d.s. matrix such that

$$A^T P + PA = -Q \qquad 4.1.5$$

Let $\rho(V)$ be a measure of the size of the RAS, 4.1.2. Then a problem inherent in this chapter is to maximize $\rho(V)$ over the class of quadratic LF's,

$$\max_{Q} \rho(V) \qquad 4.1.6$$

subject to

$$Q > o \ (Q \text{ p.d.s.}) \qquad 4.1.7$$

(ρ will be some function of P and thus Q; $\rho = \rho(P) = \rho(Q)$)
The two usual choices of ρ are the generalized volume, which for an ellipsoidal region is

$$\rho(P) = \pi V_m^{n/2}/d(P)^{\frac{1}{2}}, \qquad 4.1.8$$

and the numerical average radius

$$\rho(P) = \left(\sum_{i=1}^{N} \underline{y}_i\right)/N \qquad 4.1.9$$

where the $\underline{y}_i$ are points (usually equally spaced) lying on the RAS boundary. i.e. $V(\underline{y}_i)=V_m$. (For a general LF we have $\rho(V) = \int_S w(\underline{x})dv$, a general volume measure with S the RAS , $w(\underline{x})$ a weighting factor and dv a volume element).

The optimum problem of 4.1.6. is highly nonlinear due to the associated RAS determination and researchers have therefore concentrated on the numerical side of its solution. The works of Weissenberger (48), Geiss (51), Szego (52) and Lapidus (49) have this emphasis and depend upon the formulation of 4.1.3 as variants of the constrained minimization problem

$$\min_{\underline{x} \neq 0} V(\underline{x}) = V_m \qquad 4.1.10$$

subject to

$$\dot{V}(\underline{x}) = 0 \qquad 4.1.11$$

They chose various numerical optimization techniques to solve the problem.

The following sections of this chapter fill a need in that for various systems either an optimal quadratic is found analytically or some useful RAS is determined. Some new properties are found which are confirmed to hold for general and higher order systems by efficient numerical algorithms.

4.2 Examples on the Determination of Optimal Quadratics and Simple RAS's

It is well known that necessary conditions for a solution of the constrained problem 4.1.10 are the (n+1) Lagrange equations in the coefficients x_i and the Lagrange multiplier λ

$$\nabla V(\underline{x}) + \lambda \nabla \dot{V}(\underline{x}) = 0 \qquad 4.2.1$$

$$\dot{V}(\underline{x}) = 0 \qquad 4.2.2$$

Geometrically, if $\underline{x}^*$ and λ^* satisfy these equations then $\nabla V(\underline{x}^*) = -\lambda^* \nabla \dot{V}(\underline{x}^*)$, which implies that the contours $V(\underline{x})=V(\underline{x}^*)$ and $\dot{V}(\underline{x})=0$ are 'tangential' at $\underline{x}=\underline{x}^*$. Consequently, 4.2.1 will be called the tangency equation and $\underline{x}^*$ a tangency point. Of all such $\underline{x}^*$, $x^* \neq o$, we require the one minimizing V on 4.2.2., i.e. the global minimum. This will be called the valid tangency point (In some cases no valid tangency exists e.g. $\dot{V}<o$, $\underline{x} \neq o$; then ρ is unbounded and E^n the DOA) The equations 4.2.1 and 4.2.2 also hold for a general LF but valid tangency points may exist for which $\nabla V(\underline{x}^*)=o$.

Analytically, 4.2.1 implies the n-1 independent equations

$$\frac{V_{x_1}}{\dot{V}_{x_1}} = \frac{V_{x_2}}{\dot{V}_{x_2}} = \cdot \quad \cdot \quad \frac{V_{x_n}}{\dot{V}_{x_n}} \qquad 4.2.3$$

with $$V_{x_i} = \frac{\partial V}{\partial x_i}$$

For descriptive convenience we call the contour $\dot{V}(\underline{x})=0$, $\underline{x} \neq o$, (i.e. E_V) the constraint contour and as its components we use the intuitive definition as the subsets of E_V which form continuous connected curves (n=2) or surfaces (n=3). If two components of E_V are radially symmetric (RS) we will call this one RS component and likewise for tangency points. In the examples which follow we choose ρ as the volume (or area) measure.

Example 4.2.1

Consider the system from Zubov (1) in 3.4.1, namely,

$$\dot{x}_1 = -x_1 + 2x_1^2 x_2 \qquad S1$$

$$\dot{x}_2 = -x_2$$

with DOA $x_1 x_2 < 1$. Any p.d. quadratic is a LF for this system so let $V = ax_1^2 + 2bx_1x_2 + cx_2^2$ with $a > o$, $ac - b^2 > o$. Then 4.2.2 gives

$$\dot{V} = 2(-V + 2x_1^2 x_2 (ax_1 + bx_2)) = o \qquad 4.2.4$$

and after some manipulation 4.2.3 gives

$$x_1^2 - 3\frac{c}{a}x_1x_2^2 - 2\frac{bc}{a^2}x_2^3 = 0 \qquad 4.2.5$$

Let $u = x_1/x_2$, then dividing 4.2.5 by x_2^3 gives the tangency equation

$$u^3 - 3\frac{cu}{a} - 2\frac{bc}{a^2} = 0 \qquad 4.2.6$$

It is well known that the standard cubic $u^3 + pu + q = 0$ has real roots if $p < o$ and $4p^3 + 27q^2 < 0$. In this case they are given by

$$u = x_1/x_2 = 2\sqrt{\frac{c}{a}}\cos(l) \qquad 4.2.8$$

where $l = \frac{1}{3}\cos^{-1}(z)$, $l = \frac{1}{3}(2\pi + \cos^{-1}(z))$

with $z = b/\sqrt{ac}$. Substituting 4.2.8 into 4.2.4 gives

$$\bar{x}_2^{\,2} = (a/c)^{\frac{1}{2}} \frac{z\ (4\cos^2(1) + 4z\cos(1) + 1)}{8\cos^2(1)\ (2\cos(1) + z)} \qquad 4.2.9$$

and thus

$$V_m = \frac{(ac)^{\frac{1}{2}}\ (4\cos^2(1) + 4z\cos(1) + 1)^2}{8\ \cos^2(1)\ (2\cos(1) + z)} \qquad 4.2.10$$

The area $\rho(V) = \rho(a,b,c)$ of the RAS is

$$\rho = \frac{\pi}{8}\frac{(4\cos^2(1) + 4z\cos(1) + 1)^2}{\sqrt{1-z^2}\ \cos^2(1)\ (2\cos(1) + z)} \qquad 4.2.11$$

which, interestingly, is a function of one parameter, z.

For p.d. V we require $|z|<1$ and $a>0$. Given such a z, by varying a, b and c to satisfy $z=b/(ac)^{\frac{1}{2}}$, an infinite number of quadratics exist all giving the same area for ρ. Since this is true for the optimizing z, z^* say, the optimal quadratic is in this sense <u>non unique</u>. It can be shown that $z^* = 1/\sqrt{2}$ giving $\rho = 4\pi$.

By inspection of 4.2.4 the lines $x_1=0$, $x_2=0$ and $ax_1 + bx_2=0$ separate the phase space into regions where RS components for E_V lie. Further, we have for

a) $z\leq 0$, 1 RS component and one tangency point.

b) $z>0$, 2 RS components and 2 tangency points.

(See Fig. 4.1). Also ρ is small for $z<0$ and increases to 4π with increasing z until $z=z^*$, thereafter decreasing. The valid tangency point for $0<z<1/\sqrt{2}$ lies on a RS component in $x_1x_2>0$, whereas for $\frac{1}{\sqrt{2}}<z<1$, on a RS component in $x_1x_2<0$. When $z=z^*=1/\sqrt{2}$ we have the <u>important property</u> that <u>two valid tangency points</u> exist and thus an optimal quadratic boundary $V=V_m$ touches both RS components.

There are an infinite number of such boundaries satisfying $\sqrt{2}\, b = \sqrt{ac}$ which, in the limit, sweep out an open region. Fig. 4.2 shows some of these together with two loci obtained by eliminating the variable quantity $(a/c)^{\frac{1}{2}}$ between 4.2.8 and 4.2.9 giving a locus of tangency points as

$$x_1 x_2 = \frac{(4\cos^2(l) + 4z\cos(l) + 1)}{\cos(l)\quad(2\cos(l)+z)} \qquad 4.2.12$$

For $z = z^*$, the two valid values of l are $l_1 = \pi/12$ and $l_2 = 7\pi/12$ which give the loci $x_1x_2 = \sqrt{3} - 1$ and $x_1x_2 = -(\sqrt{3}+1)$ respectively. Since a quadratic RAS is a convex region an estimate of the DOA is given by points satisfying $x_1x_2 > -\sqrt{3}-1$ and $x_1x_2 < \sqrt{3} - 1$. A better estimate are the loci of extreme points on the major and minor axes given by $(R_1\cos\theta_1, R_1\sin\theta_1)$ and $(R_2\cos\theta_2, R_2\sin\theta_2)$ where for $b > 0$

$$R_1 = \sqrt{\frac{2}{b^3}\left(2b^2+1+\sqrt{1+4b^2}\right)}, \qquad R_2 = \sqrt{\frac{2}{b^3}\left(2b^2+1-\sqrt{1+4b^2}\right)}$$

and $\theta_1 = (\pi + a)/2$, $\theta_2 = a/2$, $a = \tan^{-1}(2b/(1-2b^2))$

A System of Special Form

A simple quadratic RAS may be determined for a second order system of the form 4.1.1. with $\underline{g}(\underline{x})$ an homogeneous polynomial and A of the form

$$A = \begin{bmatrix} 0 & 1 \\ -\alpha & -\beta \end{bmatrix} \quad \text{with } \beta^2 - 4\alpha < 0 \quad (\alpha, \beta > 0)$$

For then a p.d.s. P exists such that $A^T P + PA = -\lambda P$, namely, to a constant factor

$$P = \begin{bmatrix} \alpha & \beta/2 \\ \beta/2 & 1 \end{bmatrix}, \quad \lambda = \beta \qquad 4.2.13$$

Equations 4.2.2 and the tangency equation then give

$$\dot{V} = -\lambda V + 2\,\underline{x}^T P\,\underline{g}(\underline{x}) \qquad 4.2.14$$

and

$$\frac{V_{x_1}}{V_{x_2}} = \frac{(\underline{x}^T P \underline{g})x_1}{(\underline{x}^T P \underline{g})x_2} \qquad 4.2.15$$

The significance is that 4.2.15 reduces to an homogeneous polynomial in x_1 and x_2 as in 4.2.5 with reduction to a polynomial in $\left(\frac{x_1}{x_2}\right)$. Its roots determine straight lines as in 4.2.8, and tangency points exist where such lines intersect the constraint contour, $\dot{V} = o$. An example is the Van der Pol equation:

Example 4.2.2.

The system may be written

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\epsilon(1-x_1^2)x_2 - x_1 \end{aligned} \qquad S2$$

Its DOA is a limit cycle region depending upon $\epsilon > o$. Choosing P in 4.2.13, 4.2.14 and 4.2.15 reduce to

$$x_1^2 x_2\,(x_1 + 2x_2) - (x_1^2 + \epsilon x_1 x_2 + x_2^2) = o \qquad 4.2.16$$

and

$$\epsilon x_1^3 + x_1^2 x_2\,(4-\epsilon^2) - 3\epsilon x_1 x_2^2 - 4x_2^3 = o \qquad 4.2.17$$

where $V = x_1^2 + \epsilon x_1 x_2 + x_2^2$. Substituting $z = x_1/x_2 + \frac{R}{3}$ $(R = 4/\epsilon - \epsilon)$ in 4.2.17 gives the standard cubic

$$z^3 + pz + q = o$$

For real roots $4p^3 + 27q^2 < o$ which holds when $\epsilon < 2$ (here $\beta^2 - 4\alpha = \epsilon^2 - 4$). The roots are given as

$$x_1 = x_2\,(2\cos(l)\sqrt{9+R^2} - R)/3 \qquad 4.2.18$$

where $l = \frac{1}{3}\cos^{-1}(T)$ or $l = \frac{1}{3}(2\pi \mp \cos^{-1}(T))$

and $T = -\dfrac{(2R^3 - 27\epsilon)}{2(9+R^2)^{3/2}}$

It appears there are 2 RS components of E_V for a given ϵ and only two possible RS tangency points (or values of l). Substitution of 4.2.18 into 4.2.16 gives at tangency

$$V_m = \frac{(d^2 + \epsilon d + 1)^2}{d(\epsilon d + 2)}$$

with $d = \frac{1}{3}(2\cos(l)(9+R^2)^{\frac{1}{2}} - R)$. Valid values of l must give $d(\epsilon d + 2) > o$. When $\epsilon = .1$, for example, two values of l are 57°44' and 62°16' giving $V_m = 1.74$, the valid boundary value, and $V_m = 2.14$ respectively.

The optimum quadratic is impossible to obtain analytically and is found numerically. (See fig. 4.5(a) and 4.5(b)).

The RAS is a reasonable estimate of the DOA.

Example 4.2.3.

Another practical example is the Duffing equation

$$\dot{x}_1 = x_2 \qquad S3$$

$$\dot{x}_2 = -\alpha x_1 - \beta x_2 + \epsilon x_1^3$$

with $\alpha, \beta, \epsilon > o$. The singular points are P_1 (0,0) and P_2 $(\sqrt{\frac{\alpha}{\epsilon}}, 0)$, P_3 $(-\sqrt{\frac{\alpha}{\epsilon}}, 0)$.

Consider the particular case $\alpha = \beta = \epsilon = 1.0$. By choosing

$$Q = 2\begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

in 4.1.5. with Q p.d a general quadratic LF is found to be

$$V = Rx_1^2 + 2ax_1x_2 + (a+c)x_2^2 \qquad 4.2.19$$

with $R = 2(a-b) + c$. The constraint and tangency equations give

$$\dot{V} = 2\left[-(ax_1{}^2+2bx_1x_2+cx_2{}^2)+x_1{}^3(ax_1+x_2(a+c))\right] = o \qquad 4.2.20$$

and

$$-2\left[x_1{}^2(Rb-a^2) + x_1x_2(Rc-a(a+c))+x_2{}^2(ac-b(a+c))\right]$$
$$+x_1{}^2\left[x_1{}^2(R(a+c)-4a^2)-x_1x_2 6a(a+c)-3x_2{}^2(a+c)^2\right] = 0$$
$$4.2.21$$

As a first estimate, 4.2.14 and 4.2.15 give us, with $V = x_1{}^2+x_1x_2+x_2{}^2$,

$$\dot{V} = -V + x_1{}^3(x_1+2x_2) \qquad 4.2.22$$

and

$$x_2(x_1+x_2) = o \qquad 4.2.23$$

The <u>only</u> RS tangency point corresponds to $x_2=o$ in the latter i.e. $\mp(1.o,o)$. It gives as RAS $V<1.o$ with $\rho = 2\pi/\sqrt{3}$.

Trying various values of a, b, and c and solving 4.2.20 and 4.2.21 for the RS tangency points - which involves the real roots of a 4th degree polynomial - it is found that (a) only one RS component of E_V and one RS tangency point exists and that (b) $\rho(P)$ increases when a = const. and b, c $\to$ o. In the limit 4.2.20 and 4.2.21 give with $V = a(2x_1{}^2+2x_1x_2+x_2{}^2)$,

$$\dot{V} = ax_1{}^2(x_1{}^2+x_1x_2-1) = o \qquad 4.2.24$$

and

$$a^2\left[2x_1(x_1+x_2) - x_1{}^2(2x_1{}^2+6x_1x_2+3x_2{}^2)\right] = o$$

Clearly, $x_1 = 0$ is not a trajectory of the system except for $\underline{x} = 0$.

Thus theorem 1.5.3. may be used to give an RAS.

From 4.2.24 $x_2 = \frac{1-x_1^2}{x_1}$ and use in the last equation gives

$\mp$ (1,o) as the valid RS tangency point with $\rho(P) = 2\pi$ and with RAS boundary $2x_1^2 + 2x_1x_2 + x_2^2 = 2.0$. (Numerical work confirms that this is the unique optimal boundary). Fig. 4.3 shows the boundary in relation to its constraint contour.

The example gives as an optimizing Q

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

which is p.s.d. Using the same Q a simple RAS boundary is easily obtained for the general case of system S3 as

$$V = x_1^2 (\alpha + \beta^2) + 2 x_1x_2 + x_2^2 = \frac{2\beta\alpha^{3/2}}{\epsilon}$$

with $\rho = 2\alpha\pi\beta/\epsilon$. Thus we see that increasing β, the damping coefficient, will increase the area of stability, as will decreasing ϵ the forcing term coefficient.

<u>Example 4.2.4</u> (The n-Dimensional Case)

System S3 is a particular case of the n-dimensional system,

$$\dot{\underline{x}} = \begin{bmatrix} 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & 1 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ & & & & & \\ -a_1 & -a_2 & -a_3 & . & . & -a_n \end{bmatrix} \underline{x} + \epsilon \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 0 \\ x_1^3 \end{bmatrix} \qquad S4$$

with A stable. A RAS is obtained by solving 4.1.5. with $Q = \underline{e}_1 \underline{e}_1^T$ where $\underline{e}_1 = (1,0,0,\ldots,0)$. It can be shown (see Lefshetz (43)) that P is p.d. if the matrix $C = (\underline{e}_1, A^T\underline{e}_1, \ldots, A^{T^{n-1}}\underline{e}_1)$ has rank n.

For this A, C = I and hence P is p.d. (Smith (53) gives a method for the determination of the p_{ij})

Consider the case for n = 3 where for A stable the following Routh-Hurwitz conditions must hold (17)

$$a_3 > 0 \ , \ a_3 a_2 - a_1 > 0 \ , \ a_1 > 0$$

For the moment put $\epsilon = 1.0$ then the required P is

$$P = \begin{bmatrix} a_1 d + a_2 & a_2 d & 1 \\ \cdot & \frac{d(a_3^3 + a_1)}{a_3^2} & d \\ \cdot & \cdot & \frac{d}{a_3} \end{bmatrix} \frac{1}{2a_1} \ , \quad d = \frac{a_3^2}{(a_3 a_2 - a_1)} > 0 \qquad 4.2.26$$

With $V = \underline{x}^T P \underline{x}$ the constraint equation is

$$\dot{V} = -x_1^2 \left(1 - \frac{1}{a_1} \left(x_1^2 + d x_1 x_2 + \frac{d}{a_3} x_1 x_3\right)\right) = 0 \qquad 4.2.27$$

Of the tangency equations, the equation

$$\frac{V_{x_2}}{V_{x_3}} = \frac{\dot{V}_{x_2}}{\dot{V}_{x_3}} \qquad 4.2.28$$

gives $x_2 = -a_3 x_1$. The other equation

$$\frac{V_{x_1}}{V_{x_2}} = \frac{\dot{V}_{x_1}}{\dot{V}_{x_2}} \qquad 4.2.29$$

is more complex. However since $\dot{V}$ in 4.2.27 is non-vanishing on a non-trivial trajectory of the system, the only component of interest is

$$x_3 = \frac{a_3}{d}\left(\frac{a_1}{x_1} - x_1 - dx_2\right)$$

Substituting the latter and the result of 4.2.28 into 4.2.29 we easily obtain an equation in $\bar{x}_1$ only, reducing to $x_1^4 = a_1(a_3\, a_2 - a_1)$. Thus the valid RS tangency point is $\mp \bar{\underline{x}}$, where

$$\bar{x}_1 = (a_1\,(a_3\, a_2 - a_1))^{\frac{1}{4}}$$

$$\bar{x}_2 = -\, a_3\, \bar{x}_1$$

$$\bar{x}_3 = \frac{\bar{x}_1^3}{a_1 a_3}(a_1 + \bar{x}_1^2\,(da_3 - 1))$$

Also $V_m = (a_1/d)^{\frac{1}{2}}$ and since $d\,(P) = \left(\frac{da_1}{a_3^2}\right)^2 \frac{1}{8a_1^3}$, the determinant of P,

$$\rho(P) = \pi 2\sqrt{2}\;\; a_3^2\,\left(\frac{a_1^5}{d^7}\right)^{\frac{1}{4}}$$

For any ϵ an RAS is obtained with recourse to the following theorem which holds when $\underline{g}(\underline{x}) = \epsilon \underline{g}_k(\underline{x})$, with g_k homogeneous of degree k.

Theorem 4.1.

If $V = \underline{x}^T P \underline{x} = V_m$ is a quadratic boundary for 4.1.1 which has a tan. pt. at $\underline{x} = \underline{x}^*$ for $\epsilon = 1.0$, then for any $\epsilon > 0$, 4.1.1. has a tangency point $\underline{x} = \underline{x}^* / \epsilon^{\frac{1}{k-1}}$ which lies on the boundary $V = V_m / \epsilon^{\frac{2}{k-1}}$.

Proof:

For system 4.1.1 with $\underline{g} = \varepsilon \underline{g}_k$ we have

$$\dot{V}(\underline{x}/\varepsilon^{(\frac{1}{k-1})}) = \frac{1}{\varepsilon^{(\frac{2}{k-1})}}(-\underline{x}^T Q\underline{x} + 2\varepsilon^{(\frac{k}{k-1})}\underline{x}^T P\underline{g}_k(\underline{x}/\varepsilon^{\frac{1}{k-1}}))$$

and since $\underline{g}_k(\underline{x})$ is of degree k

$$\dot{V}(\underline{x}/\varepsilon^{(\frac{1}{k-1})}) = \frac{1}{\varepsilon^{(\frac{2}{k-1})}}\dot{V}(\underline{x}) \qquad 4.2.29a$$

Let $\underline{W}(\underline{x}) = \underline{\nabla}V(\underline{x}) + \lambda\underline{\nabla}\dot{V}(\underline{x})$, then

$$\underline{W}(\underline{x}/\varepsilon^{(\frac{1}{k-1})}) = \frac{1}{\varepsilon^{(\frac{1}{k-1})}}\Big[P\underline{x} + \lambda(-Q\underline{x} + \varepsilon^{(\frac{k}{k-1})}P\underline{g}_k(\underline{x}/\varepsilon^{(\frac{1}{k-1})}) + \varepsilon J^T(\underline{g}_k(\underline{x}/\varepsilon^{(\frac{1}{k-1})}))P\underline{x})\Big]$$

Since $J(\underline{g}_k(\underline{x}))$, the Jacobian of $\underline{g}_k$, is of degree k-1 we have

$$\underline{W}(\underline{x}/\varepsilon^{(\frac{1}{k-1})}) = \frac{1}{\varepsilon^{(\frac{1}{k-1})}}\underline{W}(\underline{x}) \qquad 4.2.29b$$

Hence if $\underline{W}(\underline{x}) = 0$ and $\dot{V}(\underline{x}) = 0$ are satisfied for $\varepsilon = 1$, $\underline{x} = \underline{x}^*$ and some $\lambda = \lambda^*$, they are satisfied for all $\varepsilon > 0$.

. . . .

For the example above only one tangency point exists for $\varepsilon > 0$ and this occurs at $\underline{x} = \mp\underline{x}/\sqrt{\varepsilon}$.

Example 4.2.5

As a final example consider a generalization of system S1

$$\dot{x}_1 = -x_1 + \epsilon\left(x_1^{i_1}\right)\left(x_2^{i_2}\right). \quad .\left(x_n^{i_n}\right) = -x_1 + g_1$$

$$\dot{x}_2 = -x_2$$

$$\vdots \qquad\qquad\qquad\qquad S5$$

$$\dot{x}_n = -x_n$$

with $i_1 > 1$. Consider a simple quadratic LF

$$V = \underline{x}^T D\, \underline{x} = \sum_{j=1}^{n} x_j^2\, d_j, \; d_j > o$$

The constraint and tangency equations reduce to

$$\dot{V} = 2\,(-V + d_1\, \epsilon x_1\, g_1) = o \qquad 4.2.30$$

and the (n-1) equations

$$x_j^2 = \frac{d_1\, x_1^2}{(1 + i_1)} \left\{ \frac{i_j}{d_j} \right\} \qquad j=2, \; . \; , n$$

4.2.31

Substitution of 4.2.31 into 4.2.30 gives an eqn. in x_1 from which the tangency points follow. Then V_m is given as

$$V_m = d_1 \left[\sum_{j=1}^{n} \left(\frac{d_j}{d_1}\right)^{i_j} \right]^{\frac{1}{\left(\sum_{j=1} i_j - 1\right)}} \cdot \; K$$

where

$$K = \frac{(1+\sum_{j=1}^{n} i_j)}{(1+i_1)} \left[\frac{(1+\sum_{j=1}^{n} i_j)(1+i_1)^{(\frac{1}{2}\sum_{j=2}^{n} i_j - 1)}}{\epsilon \prod_{j=2}^{n} (i_j)^{i_j/2}} \right]^m$$

$$m = 2/(\sum_{j=1}^{n} i_j - 1)$$

and for the volume $\rho(D) = \dfrac{\pi V_m^{\,n/2}}{\sqrt{\prod d_j}}$

The example shows that for some choices of the i_j the volume becomes unbounded. For, select

$$n = 2, \quad i_1 = 2, \quad i_2 = 2, \quad \epsilon = 2.$$

The system is

$$\dot{x}_1 = -x_1 + 2x_1^2 x_2^2$$

$$\dot{x}_2 = -x_2$$

and $V = K\, d_1 \left(\frac{d_2}{d_1}\right)^{2/3}$ with $\rho = \pi K \left(\frac{d_2}{d_1}\right)^{1/6}$

Put $d_1 = 1.0$ and let $d_2 \to \infty$; then $\rho \to \infty$ but the quadratic boundary is of poor shape approaching the straight line $x_2 = 0$. The actual DOA is $x_1^2 x_2 < 3/2$.

The system S5 is useful for test purposes since its DOA is always an open region given by

$$x_1^{i_1-1} x_2^{i_2} \cdot \cdot x_n^{i_n} (i_1 - 1) + (1 - \sum_1^n i_j) < 0.$$

4.3 Numerical Determination of A RAS and an Optimal Quadratic

Clearly, finding the optimum RAS for a given class of LF's even as simple as quadratics is difficult analytically. In view of the work in the previous section a more comprehensive study of second order systems is needed to investigate the relationship between the optimum quadratic boundary, the number of valid tangency points and the components of the constraint contour. Consequently, a special algorithm, both efficient and accurate, was developed to study the restrictive class of system

$$\dot{x}_1 = f\,(x_1, x_2) = f_1 + f_2 + \cdots + f_{nf}$$
$$\dot{x}_2 = g\,(x_1, x_2) = g_1 + g_2 + \cdots + g_{nf} \qquad 4.3.1$$

where f_i, g_i are homogeneous of degree i in x_1, x_2 and the linear part is assumed stable. The algorithm considers a general LF of the usual form

$$V\,(x_1,x_2) = V_2 + V_3 + \cdots + V_{mv} \qquad 4.3.2$$

where $V_i = \sum_{j=1}^{i+1} a_{ij} x_1^{i-j+1} x_2^{j-1}$, $V_2 = \underline{x}^T P \underline{x}$ and P solves 4.1.5.

Consider V and $\dot{V}$ in polar coordinates (r, θ) then

$$V(r,\theta) = \sum_{i=2}^{mv} \psi_i(\theta) r^i \qquad 4.3.4$$

where

$$\psi_i(\theta) = \sum_{j=1}^{i+1} a_{ij} \cos(\theta)^{i+j-1} \sin(\theta)^{j-1}$$

and

$$\dot{V} = r^2 \sum_{i=2}^{m} \phi_i(\theta) r^{i-2}, \quad m = m_V - nf - 1 \qquad 4.3.5$$

where for

$$i \leq mv, \; \phi_i = \sum_{j=1}^{\min(i-1,nf)} \left(\frac{\partial V_{i+1-j}}{\partial x_1} \cdot f_j + \frac{\partial V_{i+1-j}}{\partial x_2} \cdot g_j \right) \qquad 4.3.6$$

$$i > mv, \; \phi_i = \sum_{j=i+1-mv}^{\min(i-1,nf)} \left(\frac{\partial V_{i+1-j}}{\partial x_1} \cdot f_j + \frac{\partial V_{i+1-j}}{\partial x_2} \cdot g_j \right)$$

where for example $f_j = f_j\,(\cos\theta, \sin\theta)$ and

$$\frac{\partial V_j}{\partial x_1} = \sum_{j=1}^{i} a_{ij}\,(i + 1 - j)\,\cos(\theta)^{i-j} \sin(\theta)^{j-1}$$

$$\frac{\partial V_j}{\partial x_2} = \sum_{j=2}^{i+1} a_{ij}\,(j-1)\,\cos(\theta)^{i+1-j} \sin(\theta)^{j-2}$$

These equations hold by comparing like powers of r. Given θ, a point on the contour $\dot{V} = 0$ is then found by solving, via 4.3.5, an algebraic equation of degree $m-2$ in r for the root of smallest magnitude, $\bar{r}$ say. At the point $(\bar{r}, \theta)$, $V = V(\bar{r}, \theta)$ and $\dot{V}(\bar{r}, \theta) = 0$.
Thus the problem 4.1.10 reduces to

$$\min_{\theta} V(\bar{r}, \theta) \qquad 4.3.9$$

where

$$\sum_{i=2}^{m} \phi_i(\theta)\, \bar{r}^{i-2} = 0 \qquad 4.3.10$$

Henceforth we restrict 4.3.1 and 4.3.2 such that $m = mv + nf - 1 \leq 6$ so that 4.3.10 is at most a polynomial of degree 4 in $\bar{r}$ (Note $\phi_2, \psi_2 \neq 0, \forall\theta$ for Q positive definite) We now describe the main sub-algorithms:

Root Finding Algorithm

Solution of 4.3.10 for its roots r_i $(i = 1, m - 2)$ is possible by iterative techniques (Froberg (8)) but for $m \leq 6$ all roots are obtainable analytically; those of third and fourth degree being solved by Cardan's and Descarte's method respectively (see Froberg (8)).

When searching along the constraint contour ($\dot{V} = o$) for tangency it is necessary to keep successive solutions of 4.3.10 on the same component. This is done by finding the root of smallest magnitude, $\bar{r}$, of the same sign as the previous root, r_o say. It is also convenient to replace a large or complex root by some upper bound, Rmax say. Thus $|r_j| >$ Rmax for some θ indicates that no point on $\dot{V} = o$ lies along the ray (r, θ) except $r = 0$.

Similar considerations apply when solving

$$\sum_{2}^{mv} \psi_i r^i = c$$

for a point on V = constant say.

A general flow diagram is indicated in Fig. 4.12 where IV is an indicator set to o or 1 and ϵ a small number testing the leading coefficient of 4.3.10 for zero value. (typically $\epsilon = 10^{-6}$).

Direct Search Algorithm

Problem 4.3.9 can now be regarded as finding the minimum of a scalar function $V(\theta) = V(\bar{r}(\theta), \theta)$. Starting from some initial point (r_o, θ_o) on $\dot{V} = 0$, a sequence of values θ_j is determined. Let $V_j = V(\theta_j) = V(\bar{r}_j, \theta_j)$

and s_1 be a suitable step length. Then Fig. 4.13 gives the flow diagram which derives 3 values of θ, $\theta_1 < \theta_2 < \theta_3$, which bracket θ^*, the value giving a relative minimum of V (θ), V^* say. A predicted minimum at θ_m is found via a well known quadratic fit (see I.C.I monograph (41)) procedure, which can be repeated on the best three values of θ_m, θ_1, θ_2, θ_3 that bracket the minimum (not now equally spaced). Convergence is obtained when either a) $|\theta_m - \bar{\theta}_m| < 10^{-t}$, where θ_m and $\bar{\theta}_m$ are sucessive predicted minima of θ^* or b) the number of repetitions exceeds some integer, I say; whichever comes first (typically I = 5 and t = 3). Finally, an upperbound is placed on the step length s_1 of Smax which if exceeded restarts the initial search for the bracketing values. (Smax = .15 rad. is suitable).

Algorithm for RAS Determination

The main steps in finding a RAS for the LF 4.3.2 are as follows:

(a) with $s = \pi/N$ (N = 10 say) let $\theta_j = js$, $j = 0, 1, . . , N$. Determine $\min_j V(\theta_j) = \min_j V(\bar{r}_j, \theta_j)$ and let θ^* be the minimizing θ_j.

If all $|r_j| \geq$ Rmax increase N and repeat (a) otherwise (b).

(b) with $\theta_0 = \theta^*$ ($\dot{V}(\bar{r}^*, \theta^*) = 0.0$) find the nearest θ_m via the direct search algorithm. Then $(\bar{r}_m, \theta_m)$ is a tangency point and $V = V(\bar{r}_m, \theta_m) = V_m$ a possible RAS boundary.

Choose a conservative boundary (since tan. pt. not exact) of $V = V(r, \theta_m) = \bar{V}_m$ with $r = \bar{r}_m (1-\gamma)$ ($\gamma = .01$ in practice). Test tan. pt. via step (c).

(c) Select $s_1 = 2\pi/N1$. For $\theta_j = \theta_m + js_1$, $j = 1, \ldots, N1$, find $\bar{r}_j$ <u>the valid root</u> of $\sum_{i=2}^{mv} \psi_i(\theta_j) r^i = \bar{V}_m$. Calculate $\dot{V}_j = \dot{V}(\bar{r}_j, \theta_j)$.

If for some j, $\dot{V}_{j-1} < o$ but $\dot{V}_j > o$, tangency is invalid; put $\theta^* = \theta_j$ and repeat (b). Otherwise (d).

(d) Find $\rho(V)$, plot contours of $V = V_m$ or $\dot{V} = o$ if required.

Step (a) insures a valid initial point on $\dot{V} = o$ near the origin and (c) insures a valid tangency (typically, N1 = 50). Also the search in (c) is reduced if the smallest root of opposite sign to $\bar{r}_j$ is found, then $s_1 = \pi/N1$ with N1 = 25 say. The algorithm, called REGION, was programmed in FORTRAN IV and determines RAS's of quite general LF's.

<u>Optimal Quadratic Algorithm</u>

Algorithm REGION, in conjuction with Powell's conjugate gradient algorithm (33), was used to obtain optimal quadratics for system 4.3.1. The constraint on Q (Q p.d.) was avoided by maximizing ρ over an upper triangular matrix L s.t. $Q = L^T L + \epsilon I$, $\epsilon > o$.

Thus

$$L = \begin{bmatrix} 1 & t_1 \\ o & t_2 \end{bmatrix}, \qquad Q = \begin{bmatrix} 1 & t_1 \\ t_1 & t_1^2 + t_2^2 \end{bmatrix} \geqslant 0$$

Considered as a function of t_1 and t_2, $\rho(t_1, t_2)$ was found by REGION for any (t_1, t_2) which was a function value input to Powell's routine. Fig. 4.14 shows the interaction of the various routines.

4.4. Numerical Results

In view of Example 4.2.1. and the results that follow we make the following definition w.r.t. measure $\rho(V)$.

Def 4.4

An asymptotically stable system has property A w.r.t. an optimal quadratic (OQ) and its (RS) constraint contour if a) at least two valid (RS) tangency points exist or b) only one (RS) valid tangency point exists, there being only one tangency point for any general quadratic. (Note, an OQ having two RS tangency points but only one RS valid tangency point would not satisfy property A).

System S1, satisfying (a), and S3, satisfying (b) with $\alpha = \beta = \epsilon = 1.0$, were shown to have this property. In what follows the property will hold to a certain accuracy in that if $\underline{x}_1$ is valid tangency point and $\underline{x}_2$ any other then the property holds if

$$\frac{|V(\underline{x}_1) - V(\underline{x}_2)|}{V(\underline{x}_1)} < \epsilon \qquad (\epsilon = 10^{-3} \text{ say})$$

In order to generate systems with say k components of E_V, and thus k possible tangency points, consider the system

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

$$\dot{x}_2 = -\alpha x_1 - \beta x_2 + f_k(x_1, x_2) \qquad 4.4.1.$$

(f_k homog. deg. k). Then E_V consists of points (r, θ) where

$$r^{(k-1)} = \frac{\underline{y}^T Q\, \underline{y}}{2(p_{21}y_1 + p_{22}y_2)\, f_k} = \frac{G(\theta)}{F(y_1, y_2)}$$

(p_{ij}, elements of P)

with $y_1 = \cos\theta$, $y_2 = \sin\theta$. Clearly, F may be factored as $F = b \prod_{j=1}^{k+1} (y_2 - a_j\, y_1)$; and by suitable choice of a_j (real or conj. complex) the lines $y_2 = a_j\, y_1$ partition E^2 into regions in which either 1, 2, . . ,k or k + 1 (RS) components of E_V lie for k odd (even).

System S6 (Davies (44))

$$\dot{x}_1 = 6x_2 - 2x_2^{\,2}$$

$$\dot{x}_2 = 10x_1 - x_2 + 4x_1^{\,2} + 2x_1x_2 + 4x_2^{\,2} \qquad S6$$

The singular points are $P_1(0,0)$ and $P_2(0,2.5)$.
The DOA is a limit cycle region, $(x_1 - .5)^2 + x_2^{\,2} < 1.0$

For any quadratic the system has one component of E_V but Fig. 4.4. shows an initial quadratic with one valid tangency and the unique OQ with two. The OQ gives a poor RAS.

System S2 (Van der Pol)

Consider Example 4.2.2. with $\epsilon = 1.0$. Fig. 4.5(a) shows the OQ boundary in relation to the two RS components of E_V. The OQ is unique with two valid tangency points.

Fig. 4.5(b) shows the variation of the OQ with ϵ. Its RAS increases both in area and elongation and reflects the change in the DOA. Property A held in all cases.

System S7

$$\dot{x}_1 = x_2 \qquad \text{S6}$$

$$\dot{x}_2 = -x_1 - 4x_2 + \tfrac{1}{4}(x_2 - .5x_1)(x_2 - 2x_1)(x_2 + 2x_1)(x_2 + x_1)$$

The singular points are $P_1(o,o)$, $P_2(\sqrt[3]{2},o)$ the latter beingasaddle point.

For a general quadratic E_v may have 5 components and the valid tangency alternates between two of these near the optimum. The OQ's obtained via Powell for different initial (t_1, t_2), varied considerably. Three supposedly OQ's are shown in Fig. 4.6. Each has two valid tangency points so we might suspect a non-unique optimum, as for system S1. However, consider Fig. 4.7 which shows the contours of ρ (ρ = av radius) in the $t_1 - t_2$ plane. There exists an equal tangency curve - shown AB - such that any point lying on it produces a quadratic with two valid tangency points. Infact the three quadratics V_1, V_2 and V_3 in Fig. 4.6 correspond to the points P_1, P_2, P_3 on this curve - tangency occurring on component S for points near and above AB and on T for points below AB.

The reason for the bad convergence of Powell is the 'sharp corners' of the ρ - contours on AB. For then the assumption that ρ may be approximated by a quadratic, on which Powell is based, breaks down with a consequent loss in search direction.

In part, this explains some results of Bream (45) who showed the inferiority of Powell compared to the simplex method of Nelder and Mead (46) when optimizing ρ for Zubov functions obtained from series 3.2.3. For higher degree LF's this lack of convergence would be even more prominent.

The Simplex method, not relying on gradients or quadratic fits, still retains a flexible search direction even when AB is reached.

System S7 also exhibits a problem when using any optimization routine. Consider Fig. 4.7. with a poor initial guess, C say. Searching along the gradient or coordinate directions in turn, we arrive at E or C_1 on AB. Further improvement is only made by moving along AB in an oscillatory fashion. As the valid tangency alternates between component S and T, repeated use of step (c) is needed in algorithm REGION. This increases computer time.

In contrast to S7, Fig. 4.8 shows the ρ-contours for the Van der Pol system. Again, the equal tangency curve is present - AB - but Powell's method has little difficulty in reaching the optimum since the contours are quite smooth and AB is almost parallel to the t_2 axis.

System S8

A system giving 3 components for E_V is

$$\dot{x}_1 = -2x_1 + x_1x_2$$
$$\dot{x}_2 = -x_2 + x_1x_2 \qquad S8$$

with singular points $P_1(o,o)$ and $P_2(1,2)$. Fig. 4.9 shows three quadratic boundaries with their associated constraint contours, while Fig. 4.10, the ρ-contours (ρ = area). In the latter V_1 corresponds to the initial point I (1,1); V_2 to the point J(·206,1·0), the point reached after searching along t_2 = 1·0; and V_3 to the point O(·233, ·449), the optimal point. The fact that I and J lie on different sides of AB is illustrated in Fig. 4.10 where V_1 and V_2 have valid tangency on different components. The point O lies on AB and 'property A' holds. (Note, there are quadratics which have valid tangency with the third component, but these have small area).

System S9

Consider a more general system

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = -2x_1 - 3x_2 + \frac{x_2^2}{4} + \frac{x_1^2x_2}{4} \qquad S9$$
$$+ x_2^2(x_1^2 + x_2^2)/4$$
$$+ (x_2 - x_1/2)^2 (x_2 - 4x_1)(x_2 + \cdot 5x_1)$$

For a general quadratic LF there exists 5 components of the constraint contour of which only two are of interest near the optimum. Fig. 4.11 shows an initial quadratic, V_1, with one valid tangency point and the optimum, V_o, with two.

Comment

Many other second order examples have shown similar results. Certainly, the examples shown here have indicated that property A is quite general and that it holds for ρ chosen as the average radius or the area.

4.5. Optimal Quadratics for a Restricted Class of High Order Systems

The optimal quadratic algorithm previously outlined can be generalized to the class of systems

$$\dot{\underline{x}} = A\underline{x} + \underline{g}_2 + \underline{g}_3 + \cdot\cdot + \underline{g}_{nf} \qquad 4.5.1$$

with $nf \leq 5$, by using the general polar coordinate system in $(r, \theta_1, \theta_2, \quad , \theta_{n-1})$ defined as

$$\begin{aligned} x_1 &= r\, c_1\, c_2 \quad \cdot\ \cdot \quad c_{n-1} \\ x_2 &= r\, c_1\, c_2 \quad \cdot\ \cdot \quad c_{n-2}\, s_{n-1} \\ &\cdot \\ &\cdot \\ x_i &= r\, c_1\, c_2 \quad \cdot\ \cdot \quad c_{n-i}\, s_{n-i-1} \\ &\cdot \\ &\cdot \\ x_n &= r\, s_1 \end{aligned} \qquad 4.5.2$$

where $c_i = \cos(\theta_i)$, $s_i = \sin(\theta_i)$ and $r = \|\underline{x}\|$.

Let $y_i = x_i/r$, $i = 1, \, . \, ,n$. Then given θ_i $(i=1, \cdot ,n-1)$, $\underline{y}$ is determined, and a point $\underline{x}$ on the RAS boundary, $\underline{x}^T P\underline{x} = V_m$, or the constraint surfa ce, $\dot{V} = o$, is found by calculating

$$r = (V_m / \underline{y}^T P \underline{y})^{\frac{1}{2}} \qquad 4.5.3$$

<u>or</u> by solving the polynominal in r

$$\sum_{i=1}^{n_f} \phi_i(\underline{y}) \; r^{i-1} = o \qquad 4.5.4$$

where $\phi_1 = -\underline{y}^T Q \underline{y}$ and $\phi_i = 2\underline{y}^T P \; \underline{g_i} \; (\underline{y}), \; i > 1$.
The root of smallest magnitude with a given sign of 4.5.4, $\bar{r}$ say, is found by using the ROOTV algorithm and forms the basis of a new algorithm for RAS determination, called DNRE^G.

For RAS and tangency point determination the problem of minimizing V on the constraint contour is replaced by

$$\min_{\underline{y}} \quad V(\bar{r}, \underline{y}) \qquad 4.5.5$$

with $\bar{r}$ the required root of 4.5.4. The minimization is over the n-1 variables of θ_i and for this Powell's[33] method was used, replacing the direct search method mentioned previously. The OQ was determined by maximizing ρ by the Simplex Method of Nelder and Mead (46) over the m - dimensional space of the elements of the upper triangular matrix L, with $Q = L^T L + \epsilon I$ and $m = n(n+1)/2-1$.

The algorithm was especially developed to study <u>third order</u> systems but is easily extended to n>3. In this case the steps for DNREG are those of REGION except that (a) and (c) are now 2 - dimensional searches:

a) Let $s_1 = \pi/2N_1$ (N1 = 5 say)

$s_2 = 2\pi/N_2$ (N2 = 15 say)

With $\underline{y} = \underline{y}(\theta_1, \theta_2)$, let

$$\underline{y}_{i,j} = \underline{y}(i\, s_1,\ j\, s_2)$$

$$i = 0,1, \cdot, N1 \qquad\qquad 4.5.6$$

$$j = 0,1, \cdot, N2$$

and determine $\min_{i,j} V(\underline{y}_{i,j}) = V(\overline{r}_{i,j}, \underline{y}_{i,j})$ with $\underline{y}^*$ the minimizing $\underline{y}_{i,j}$. If all $|\overline{r}_{i,j}| \geq$ Rmax increase N1 and N2 and repeat (a) until some $|\overline{r}_{i,j}| <$ Rmax. Otherwise step (b).

(c) Tangency is valid if $\max_{\underline{x}} \dot{V} < 0$ on $V(\underline{x}) = V_m$. Let $Y: (\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n)$ be a set of unit vectors, determined as in 4.5.6 but with a finer mesh (N1 = 15, N2 = 50 say with N = 16×51). Determine

$$\dot{V}_M = \max_j \left[\dot{V}\left[\mp \left(\frac{V_M}{\underline{y}_j^T P \underline{y}_j} \right)^{\frac{1}{2}} \underline{y}_j \right] \right] \qquad 4.5.7$$

If $\dot{V}_M < 0$ go to (d) otherwise let $\underline{y}^*$ be the maximizing $\underline{y}_j$ and repeat step (b).

The unit vectors were stored through the Simplex maximization, and also, step (a) was only used for the initial quadratic, the initial $\underline{y}$ for the Powell routine being that obtained from the tangency point of the previous quadratic.

The flow diagram of Fig. 4.15 shows how the complete OQ routine was divided into subroutines for FORTRAN IV Programming.

An initial L determines an initial point of the simplex (a set of m+1 points in m-space) written $\underline{p}_1 = (t_1, t_2, \ldots, t_m)$ where

$$L = \begin{bmatrix} t_1 & t_2 & t_4 \\ 0 & t_3 & t_5 \\ 0 & 0 & 1.0 \end{bmatrix} \qquad (t_6 = 1.0,\ m = 5)$$

4.5.8

The remaining m points of the simplex were

$$p_{j+1} = (t_1, t_2, \ldots, t_{j-1}, t_j + h, t_{j+1}, \ldots t_m)$$

$$j = 1, \ldots, m$$

with h a parameter.

In the following examples only 3 iterations (30 function evaluations of $\overline{F}$) of Powell's method were needed for step (b).

4.6. Numerical Results

System S10

The third order system is

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = x_3 \qquad \text{S10}$$
$$\dot{x}_3 = - x_1(1+x_1^2) - 2x_2 - x_3(1-x_3^2)$$

For a general quadratic, $\underline{x}^T P\underline{x} = V$, there appears to be two RS tangency points and two RS components of Ev, which lie in E^3 for which

$$(p_{31}x_1 + p_{32}x_2 + p_{33}x_3)\ (x_3 - x_1) > 0.$$

With h = .3 and N = 616 the initial quadratic boundary, with Q = I, was

$$2.5x_1^2 + 5x_1x_2 + x_1x_3 + 3x_1x_3 + 5x_2^2 + 2x_3^2 = .4675$$

with ρ = .32799 (Vol.) and valid tangency

$\mp$ (·005, - ·004, ·523)

After 61 evaluations of ρ the best boundary was $\underline{x}^T\bar{P}\underline{x}$ = ·847 with ρ = ·842 and tangency $\mp$ (·7145, - ·345, - ·367), with

$$\bar{P} = \begin{bmatrix} 2.144 & 2.232 & .47 \\ \cdot & 4.56 & 1.14 \\ \cdot & \cdot & 2.06 \end{bmatrix} \qquad 4.6.1.$$

The other tan. pt. $\mp$ (- ·062, - ·091, ·682) gave a boundary close to this of V = ·856.

In general it was found difficult to obtain the OQ accurately due to the high dimension of the problem and the approximations involved. However, the example shows property A is present. Firstly, Fig. 4.16 shows sections of the initial and best quadratics in relation to their constraint contours and DOA's. The section through x_3 = ·3677 shows the extra point of contact of $V = V_m$ with E_V for the OQ. Secondly, Fig. 4.17 shows how valid tangency alternates between the two possible tangency points as the maximization of ρ progresses, and indicates an equal tangency surface. The RAS is a good estimate of the DOA.

System S11 (Davidson (47))

A system possessing a limit cycle is the following:

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

$$\dot{x}_2 = x_3 \qquad S11$$

$$\dot{x}_3 = - \cdot 915\, x_1 + x_2\,(1 - \cdot 915\, x_1^2) - x_3$$

Again there are two possible RS tangency points for a general quadratic, and two RS components of E_V.
With $h = .1$, $N = 616$ and an initial P of $P = (SS^*)^{-1}$ (Eqn. 2.2.2) the RAS boundary was

$$24\cdot 2x_1^2 - 26\cdot 4x_1x_2 + 26\cdot 4x_1x_3 + 25\cdot 8x_2^2 - 27\cdot 6x_2x_3 + 27x_3^2 = 2\cdot 07$$

with $\rho = \cdot 105$ and tan. pt. $\mp(\cdot 249, \cdot 171, -\cdot 223)$.
After 90 evaluations of ρ the best OQ boundary was $\underline{x}^T P \underline{x} = 4\cdot 4856$ with $\rho = \cdot 767$ and

$$P = \begin{bmatrix} 18.1 & -15.86 & 1.88 \\ \cdot & 31.94 & -16.59 \\ \cdot & \cdot & 17.19 \end{bmatrix} \quad \text{tan.pt.} \ \mp(.376, .271, -.241)$$

The other tan. pt. gives the boundary $\underline{x}^T P \underline{x} = 4.534$ which is indistinguishable, graphically, from the OQ. Fig. 4.18 shows sections through the initial and best boundaries and constraint contours. Those through $x_3 = \cdot 158$ and $x_3 = \cdot 241$, parallel to the $x_1 - x_2$ plane, show the closeness of the OQ boundary to its two components of E_V.
In Fig. 4.19 the oscillatory effect of the valid tangency is shown again.

System S12

A system giving one RS tan. pt. is

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \qquad\qquad \text{S12} \\ \dot{x}_3 &= -x_1 - 2x_2 - x_3 + x_1^3 \end{aligned}$$

which is of the form S4. Using P in 4.2.26 gives the RAS boundary

$$3x_1^2 + 4x_1x_2 + 2x_1x_3 + 2x_2^2 + 2x_2x_3 + 2x_3^2 = 2.0$$

with $\rho = 8{\cdot}86$ & tan. pt. $\mp$ (1, -1, 1). With h = ·1, N = 616 and Q = I, initially, the best quadratic gave ρ = 17·5 after 100 ρ-evaluations. Due to the single tangency, convergence to the optimum was rapid. The best boundary was

$$89{\cdot}4x_1^2 + 96{\cdot}9x_1x_2 + 63x_1x_3 + 89x_2^2 + 48{\cdot}9x_1x_3 + 52x_3^2 = 85{\cdot}45$$

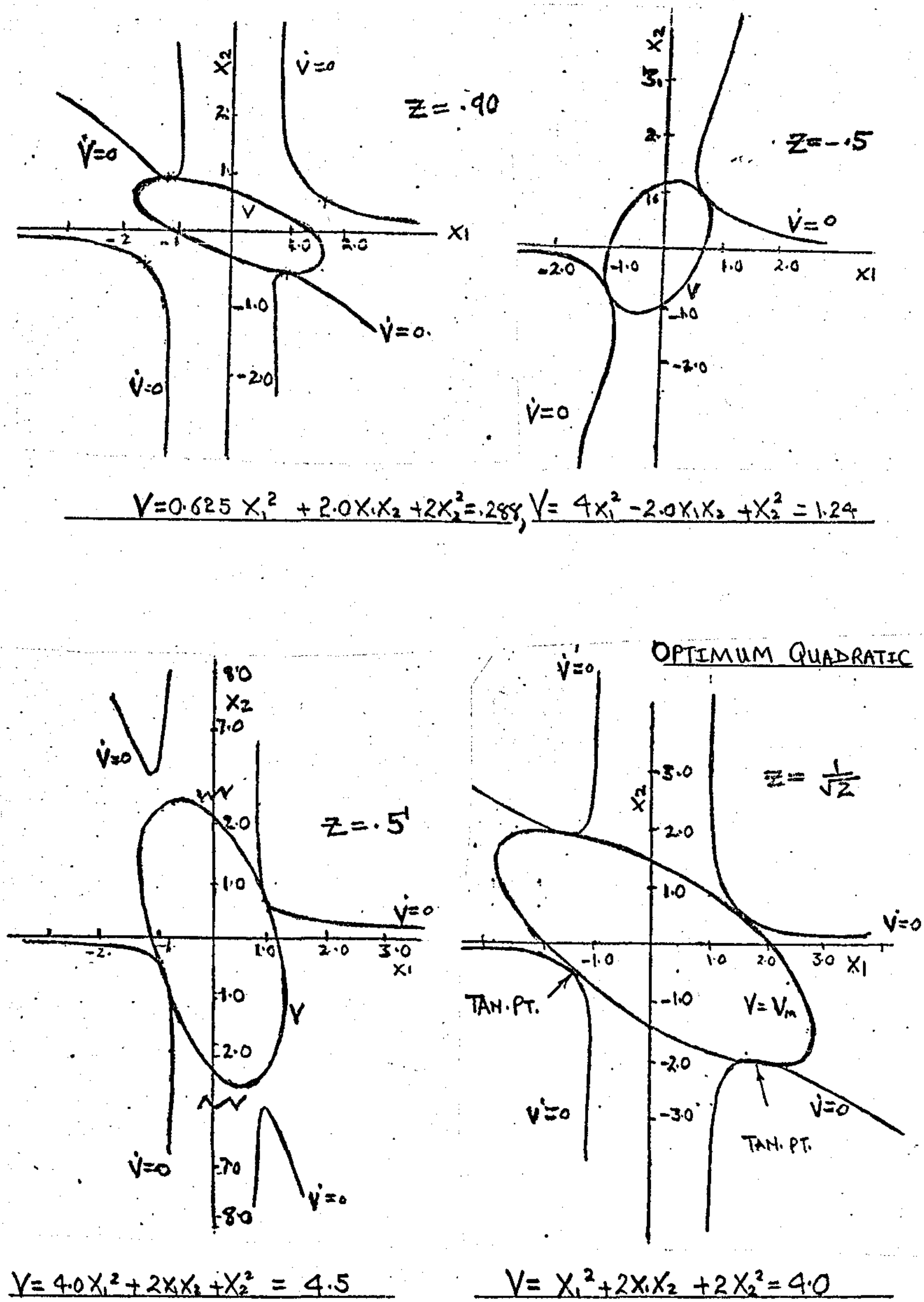
with tan. pt. $\mp$ (1·162, - ·437, ·0953)

Comments

Computation times for the 3 examples were 1000, 4000 and 1800 secs. respectively (ICL1905), an average function (ρ) evaluation being 16,67 and 18 secs respectively.

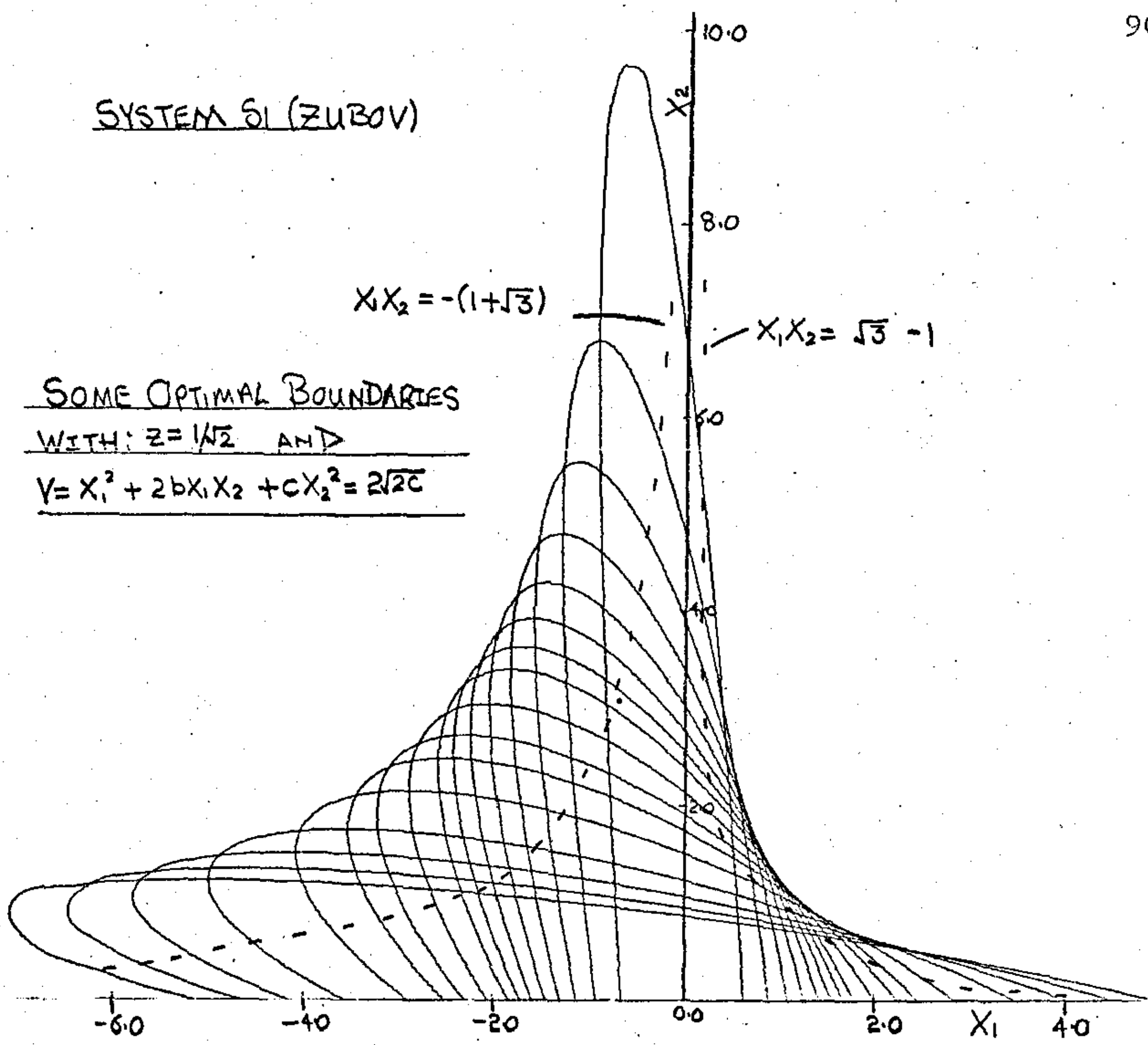
Comparison of computation times for second and third order system depends upon the degree and complexity of the system considered and on the values of N1 and N in step (c) of the algorithms. A two fold increase in N1 (with increase in accuracy) may often double computing time. For the second order case with N1 = 30 an average evaluation took 1 sec, whereas 5 sec. for N1 = 50.

Generally a compromise must be made between computing time and accuracy of the valid tangency point. For high order systems the validification of the latter via step (c) seems a big drawback, especially so near the optimum due to the oscillatory effect of the valid tangency and repetition of this step.



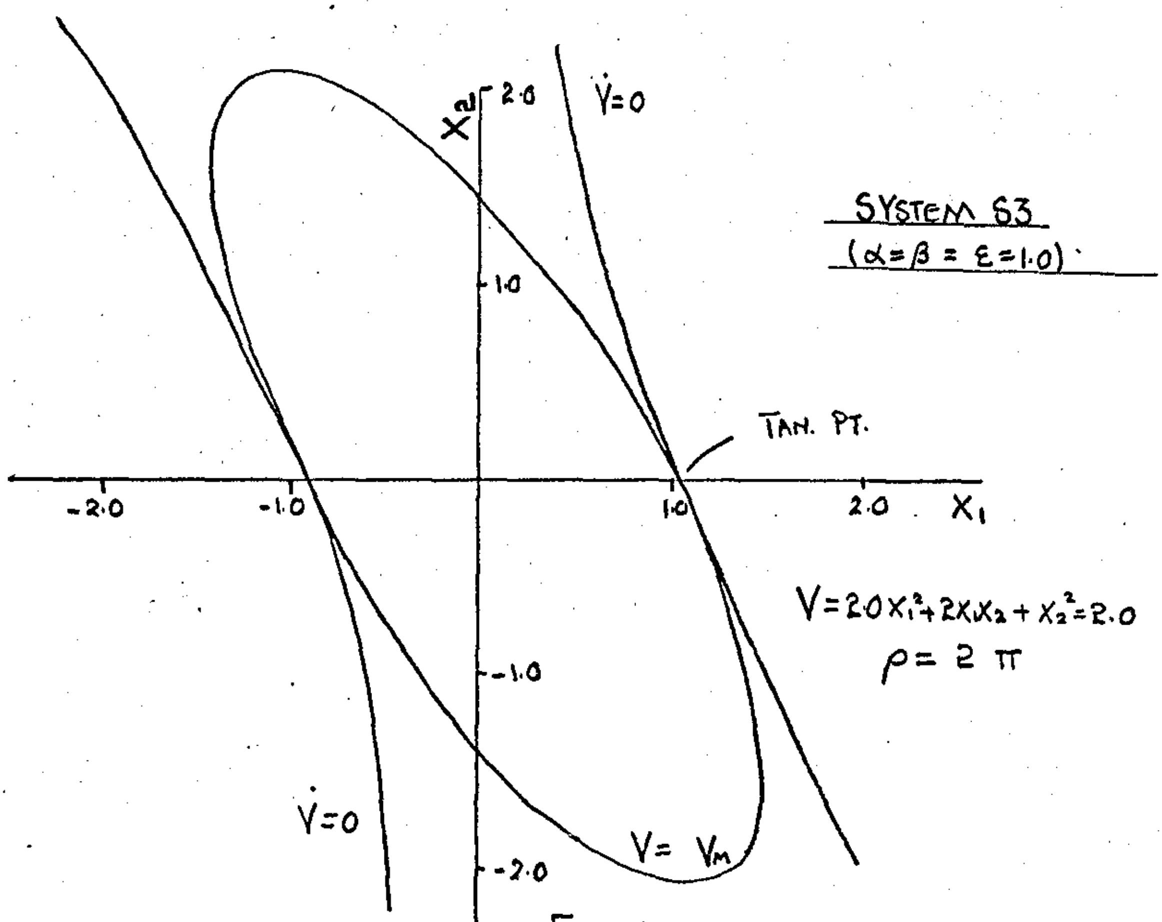
RAS BOUNDARIES AND CONSTRAINT CONTOURS (ZUBOV EXPLE.)

FIG. 4·1



SYSTEM S1 (ZUBOV)
$X_1X_2 = -(1+\sqrt{3})$
$X_1X_2 = \sqrt{3} - 1$
SOME OPTIMAL BOUNDARIES
WITH: $z = 1/\sqrt{2}$ AND
$V = X_1^2 + 2bX_1X_2 + cX_2^2 = 2\sqrt{2c}$
X_2
10.0
8.0
6.0
4.0
2.0
-6.0
-4.0
-2.0
0.0
2.0
X_1
4.0

FIG 4.2



X_2
2.0
1.0
-1.0
-2.0
$\dot{V}=0$
$\dot{V}=0$
SYSTEM S3
$(\alpha = \beta = \varepsilon = 1.0)$
TAN. PT.
-2.0
-1.0
1.0
2.0
X_1
$V = 2.0X_1^2 + 2X_1X_2 + X_2^2 = 2.0$
$\rho = 2\pi$
$V = V_M$

FIG 4.3

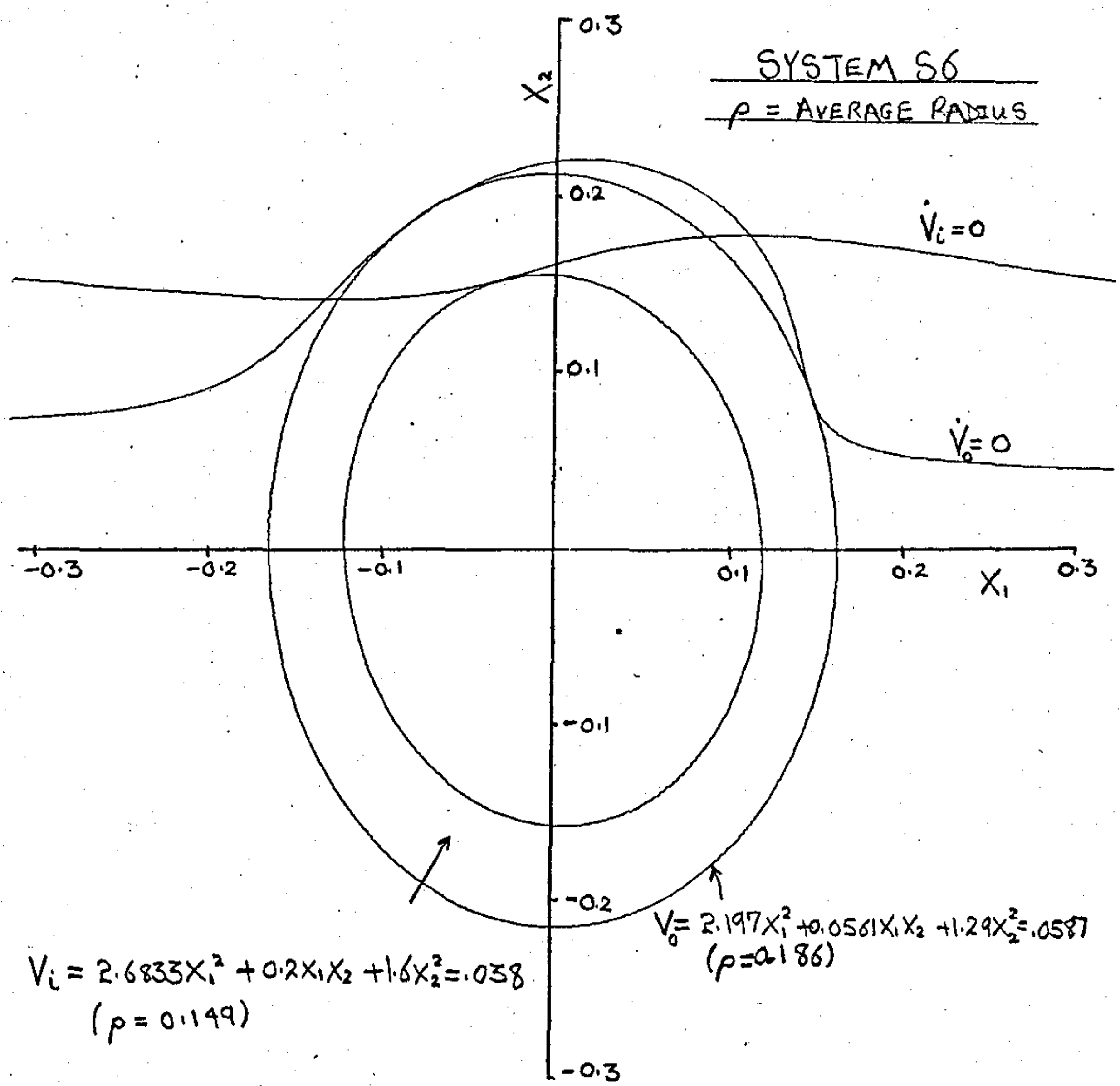


FIG. 4.4

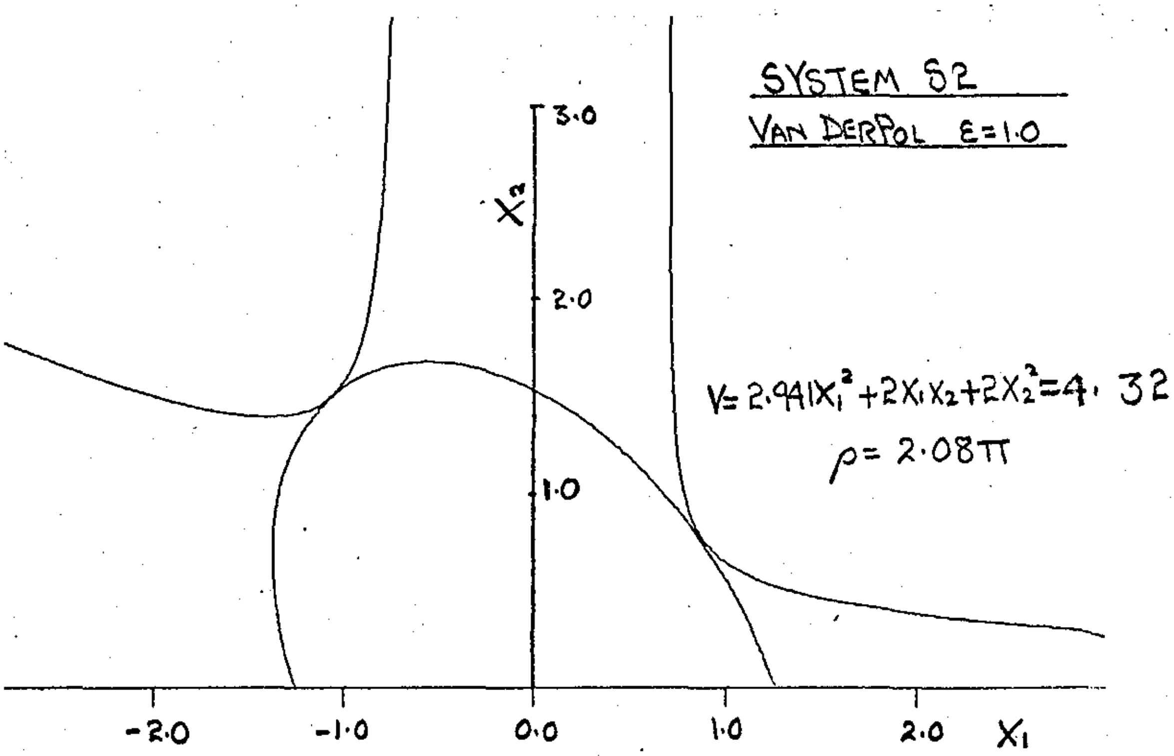


FIG. 4.5(a)

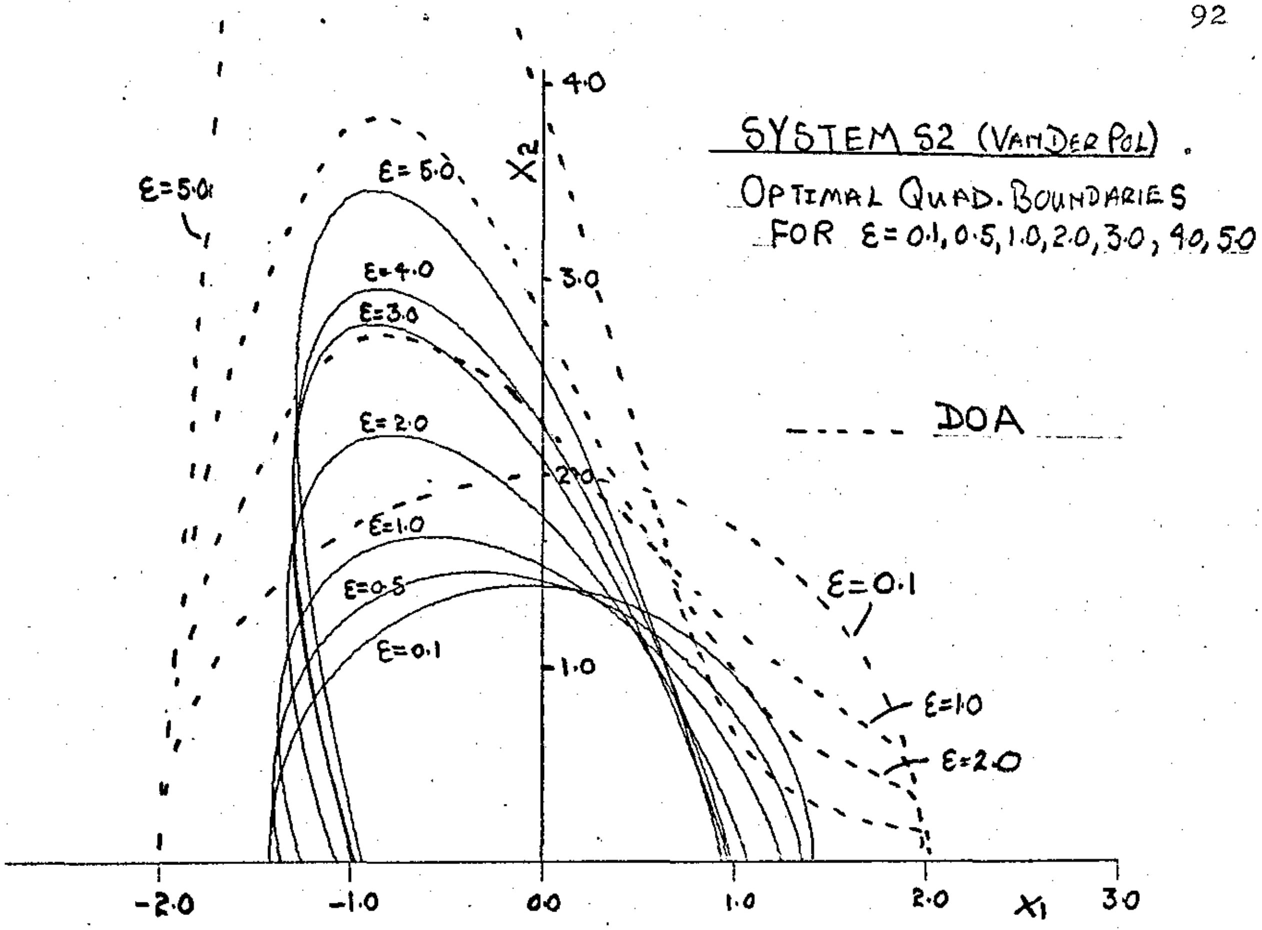


FIG. 4.5(b)

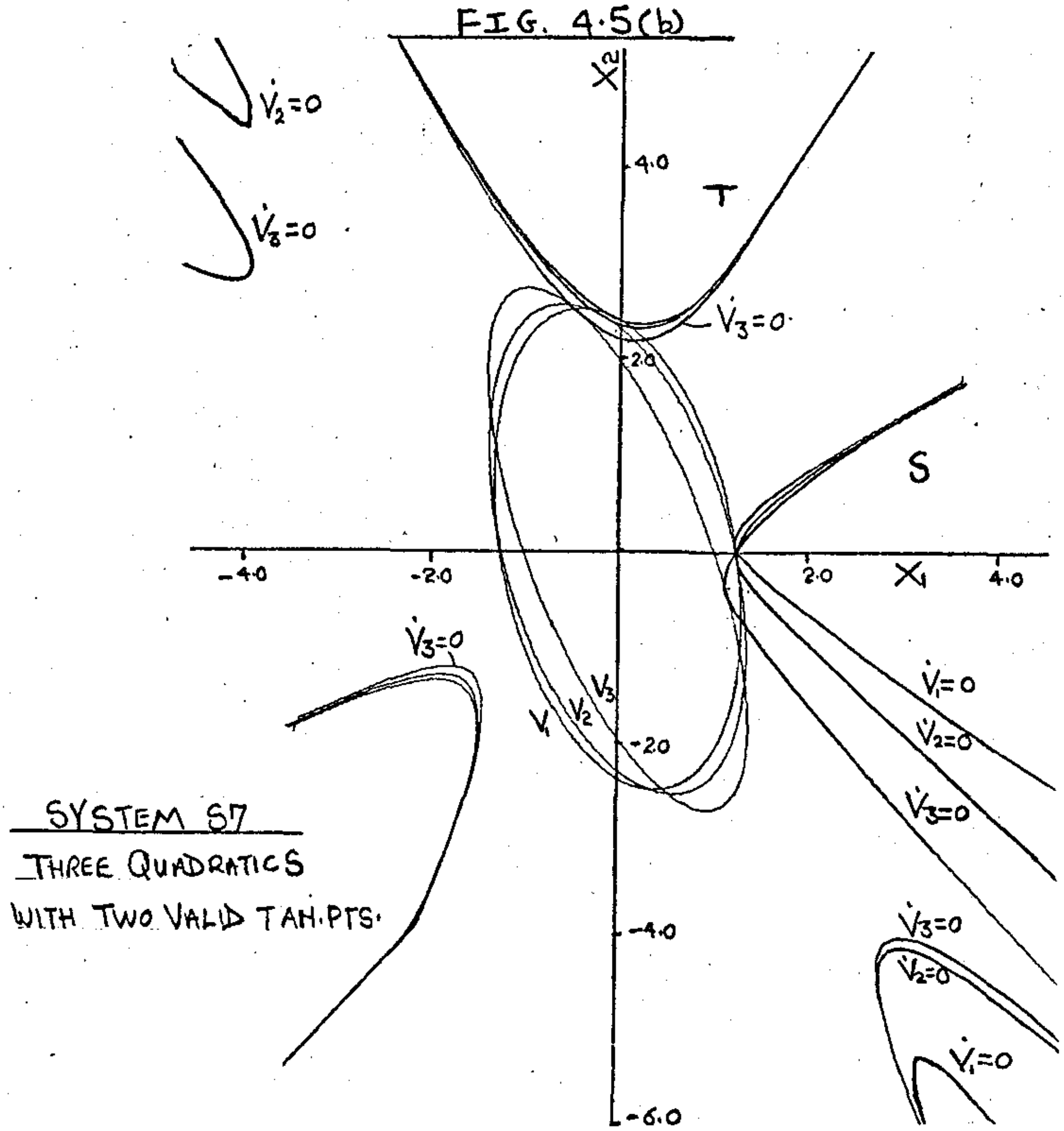


FIG. 4.6

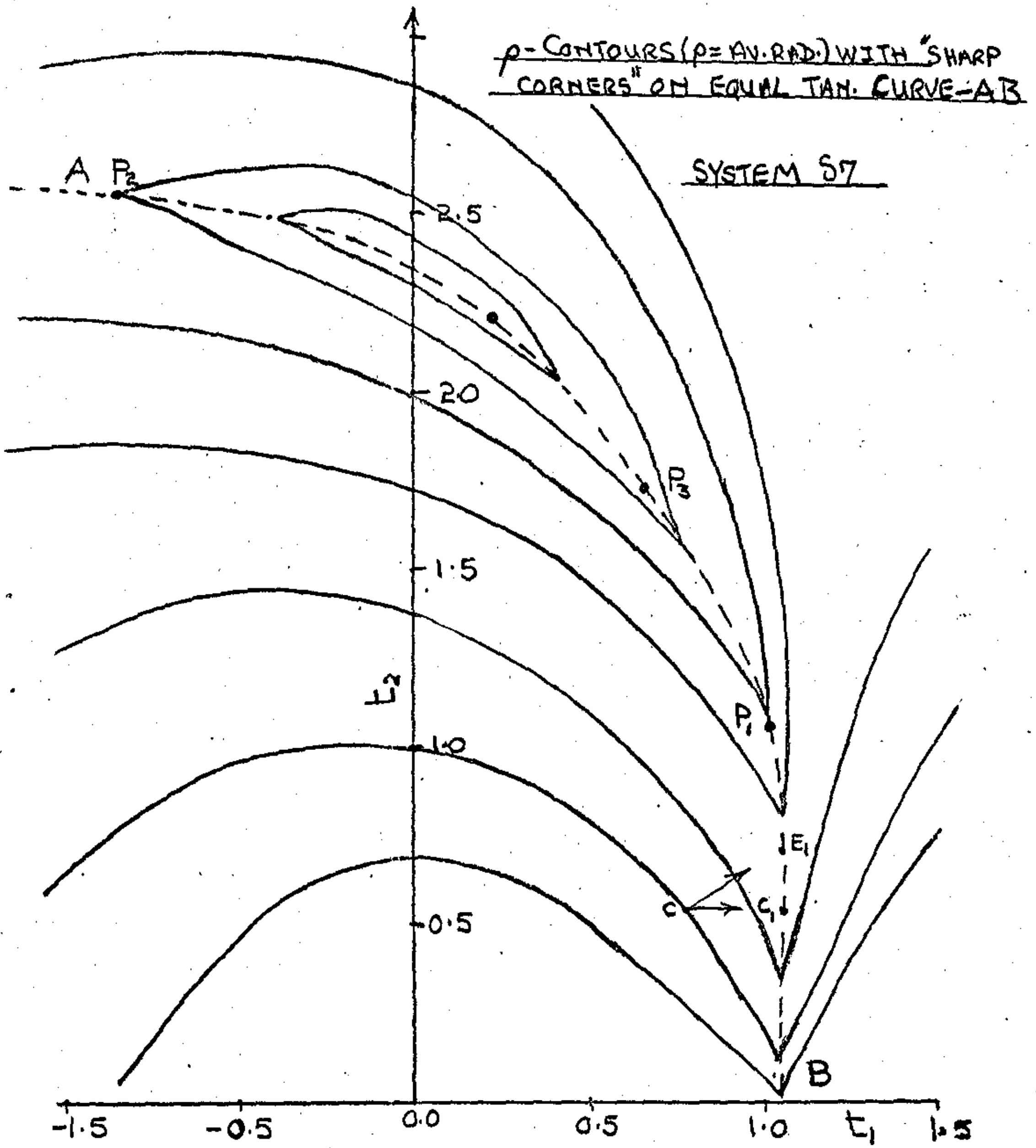

ρ-CONTOURS (ρ=AV.RAD.) WITH "SHARP CORNERS" ON EQUAL TAN. CURVE-AB
SYSTEM S7
A
P2
P3
P1
E1
C
C1
B
2.5
2.0
1.5
1.0
0.5
-1.5
-0.5
0.0
0.5
1.0
1.5
t1

FIG. 4.7

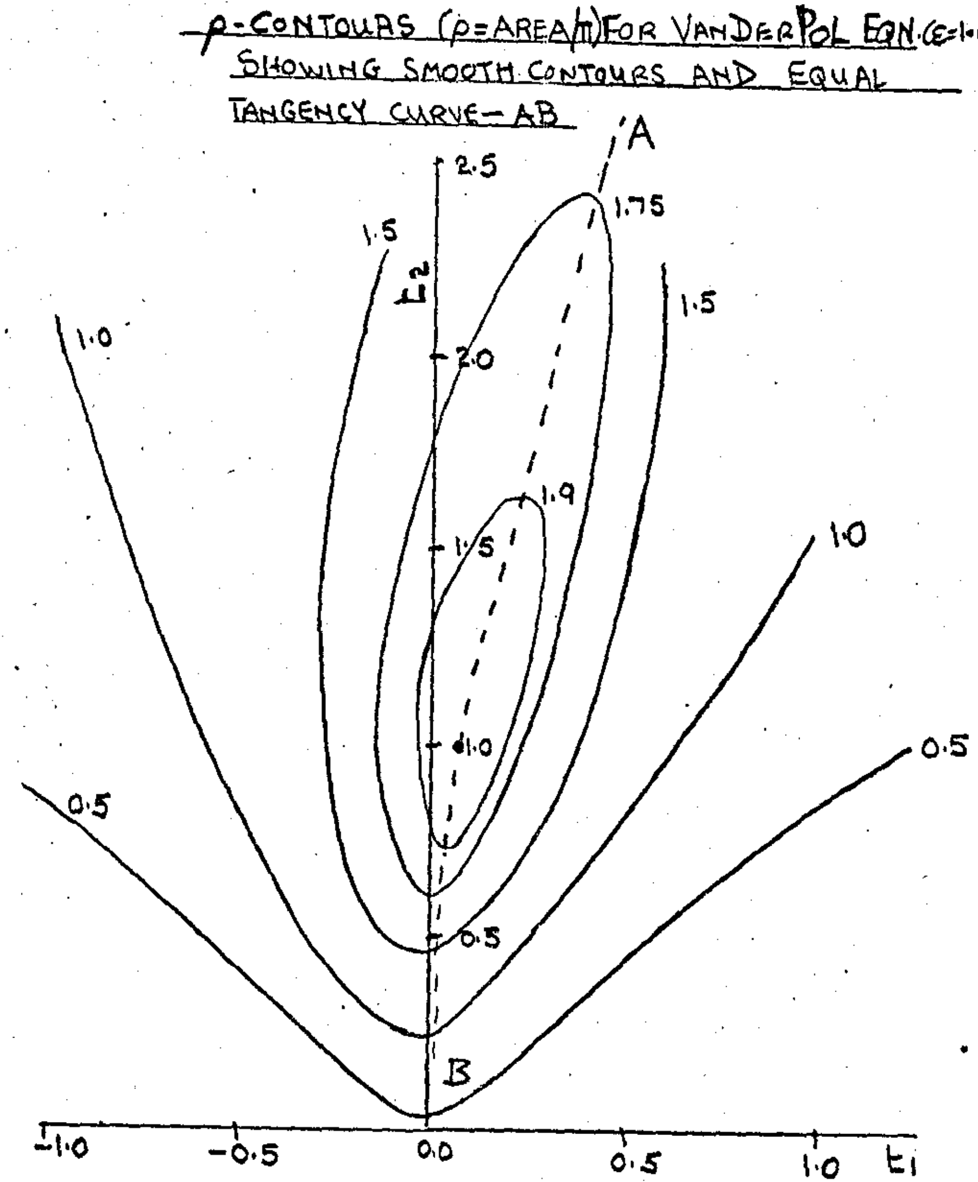

ρ-CONTOURS (ρ=AREA/π) FOR VAN DER POL EQN. (ε=1.0) SHOWING SMOOTH CONTOURS AND EQUAL TANGENCY CURVE-AB
A
B
1.75
1.5
1.5
1.9
1.0
1.0
0.5
0.5
2.5
2.0
1.5
1.0
0.5
-1.0
-0.5
0.0
0.5
1.0
t1

FIG. 4.8

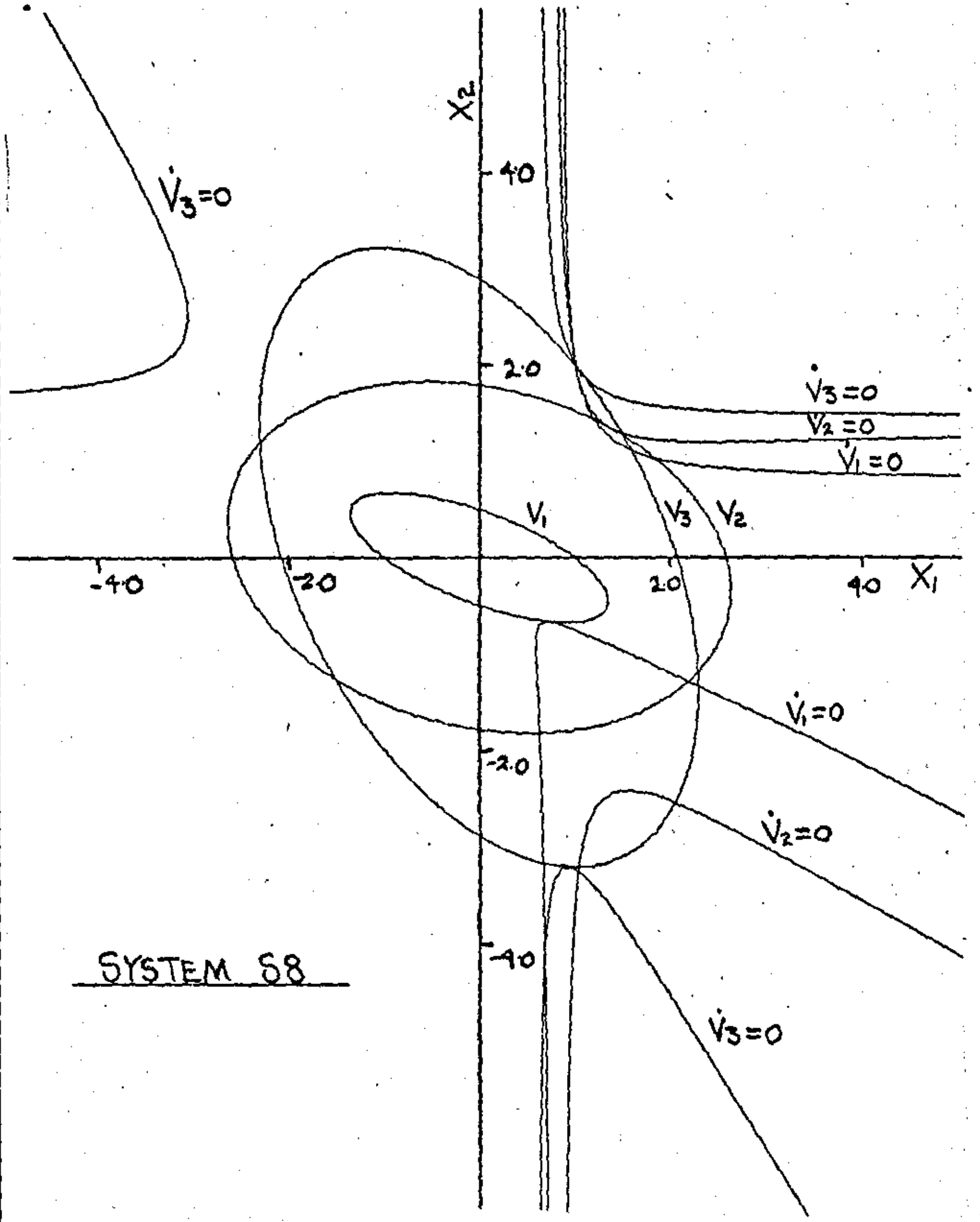


FIG 4.9

ρ CONTOURS FOR SYSTEM S8 (ρ=AREA/π)

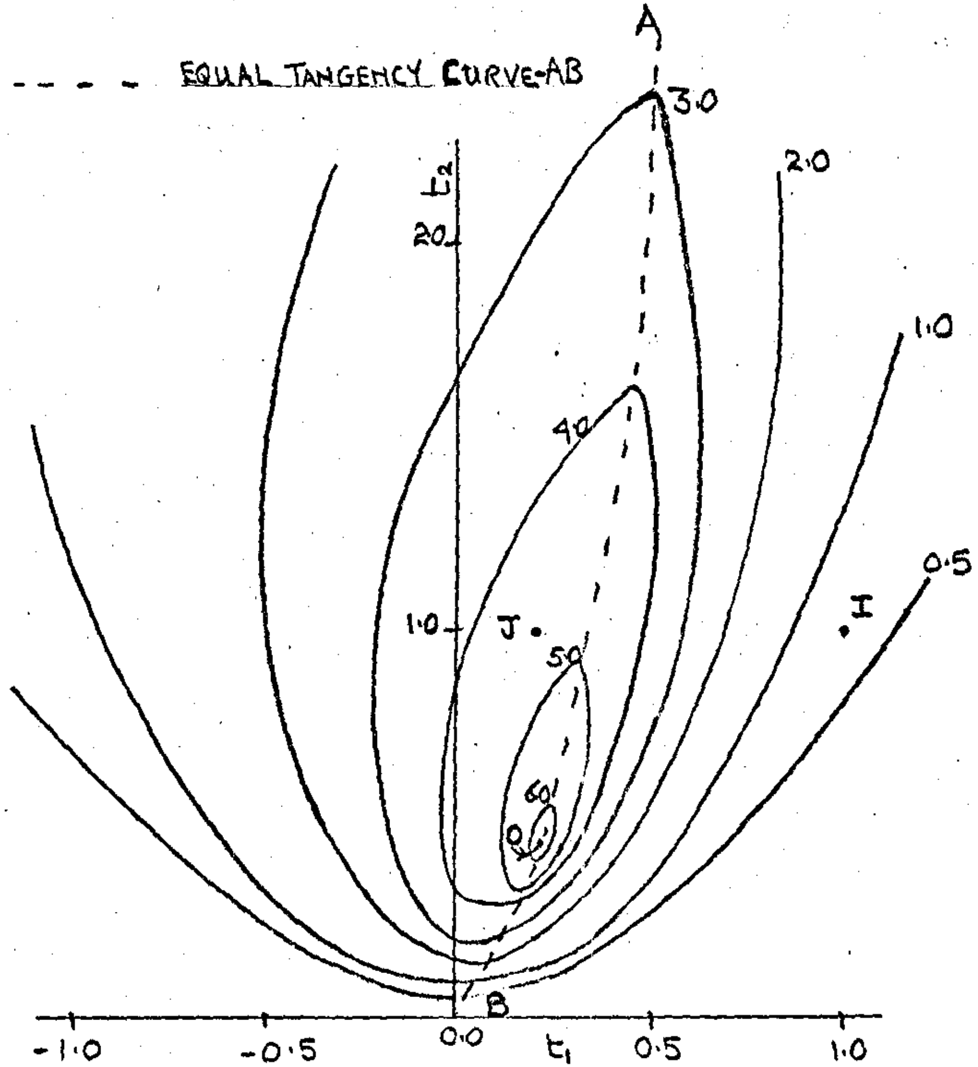


FIG 4.10

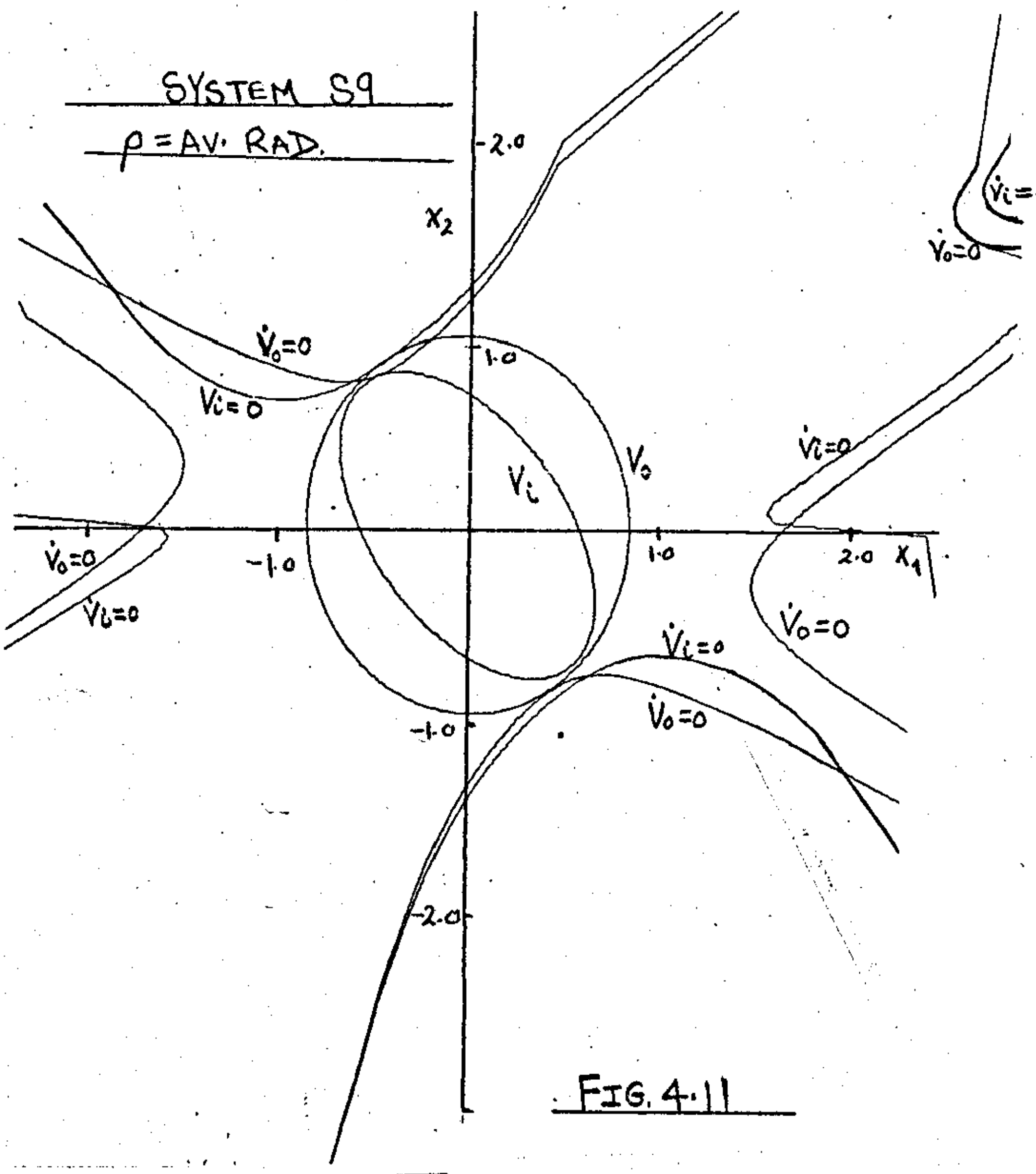


FIG. 4.11

SUBROUTINES FOR SECOND ORDER OPTIMAL QUAD. ALGORITHM.

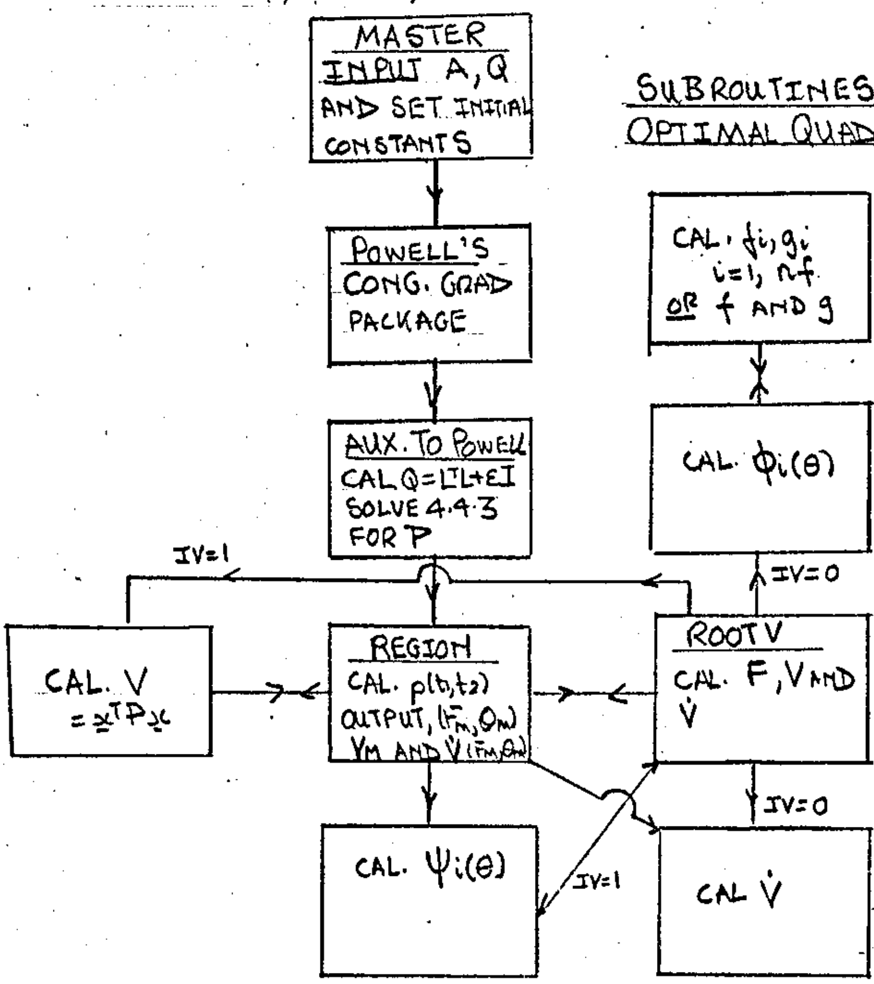


FIG. 4.14

ALGORITHM ROOTV

GIVEN θ, AN INDICATOR IV AND PREVIOUS ROOT r_0, IT CALTS. A POINT (r, θ) ON

a) $V(x) = V_m$ IF IV=1 OR

b) $\dot{V}(x) = 0.0$ IF IV=0

SUCH THAT SIGN(r)=SIGN(r_0) WITH r OF SMALLEST MAG. $V(r,\theta)$ OR $\dot{V}(r,\theta)$ ARE CALTD.

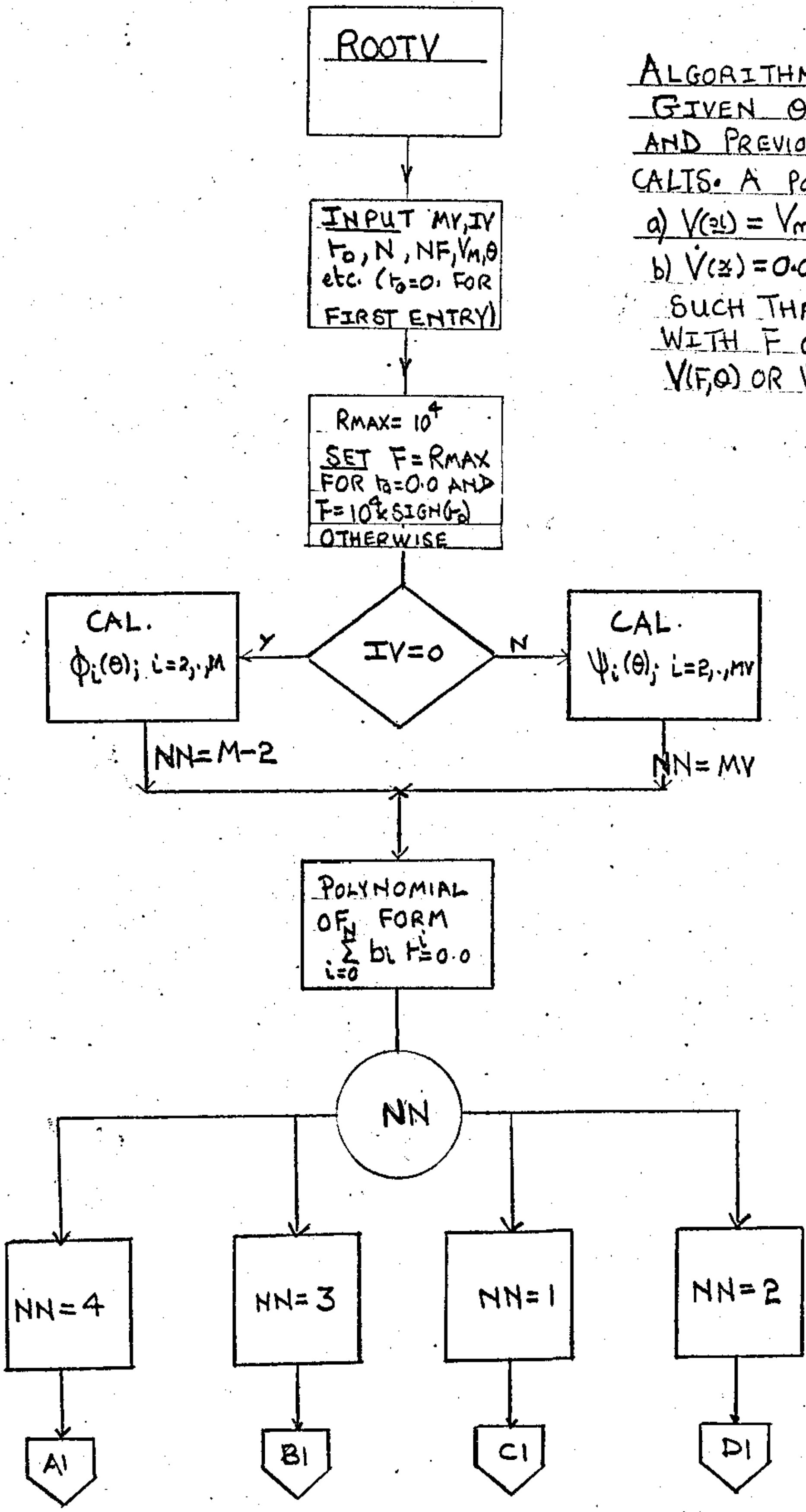


FIG. 4.12 (P.T.O)

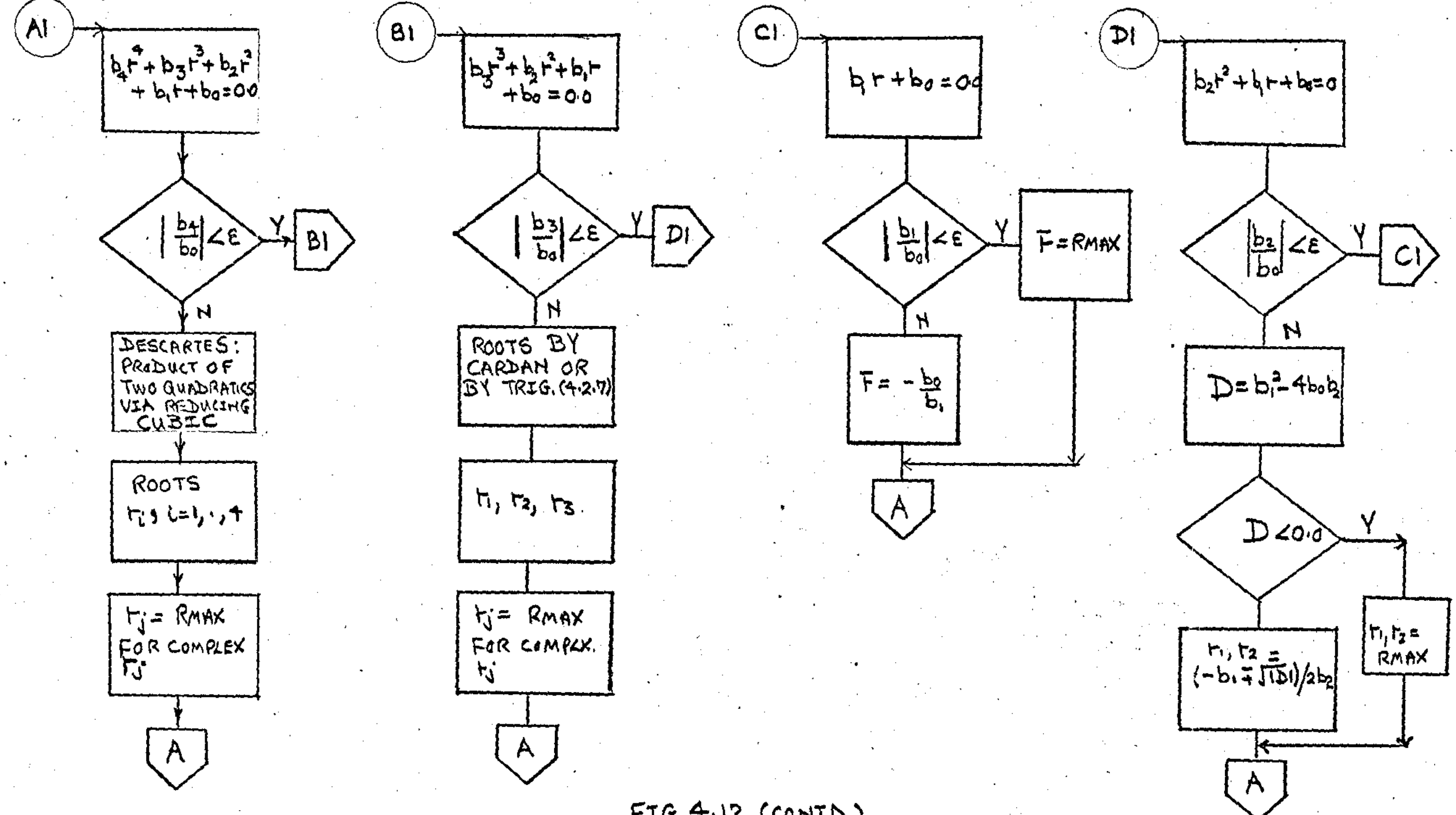


FIG. 4.12 (CONTD.)

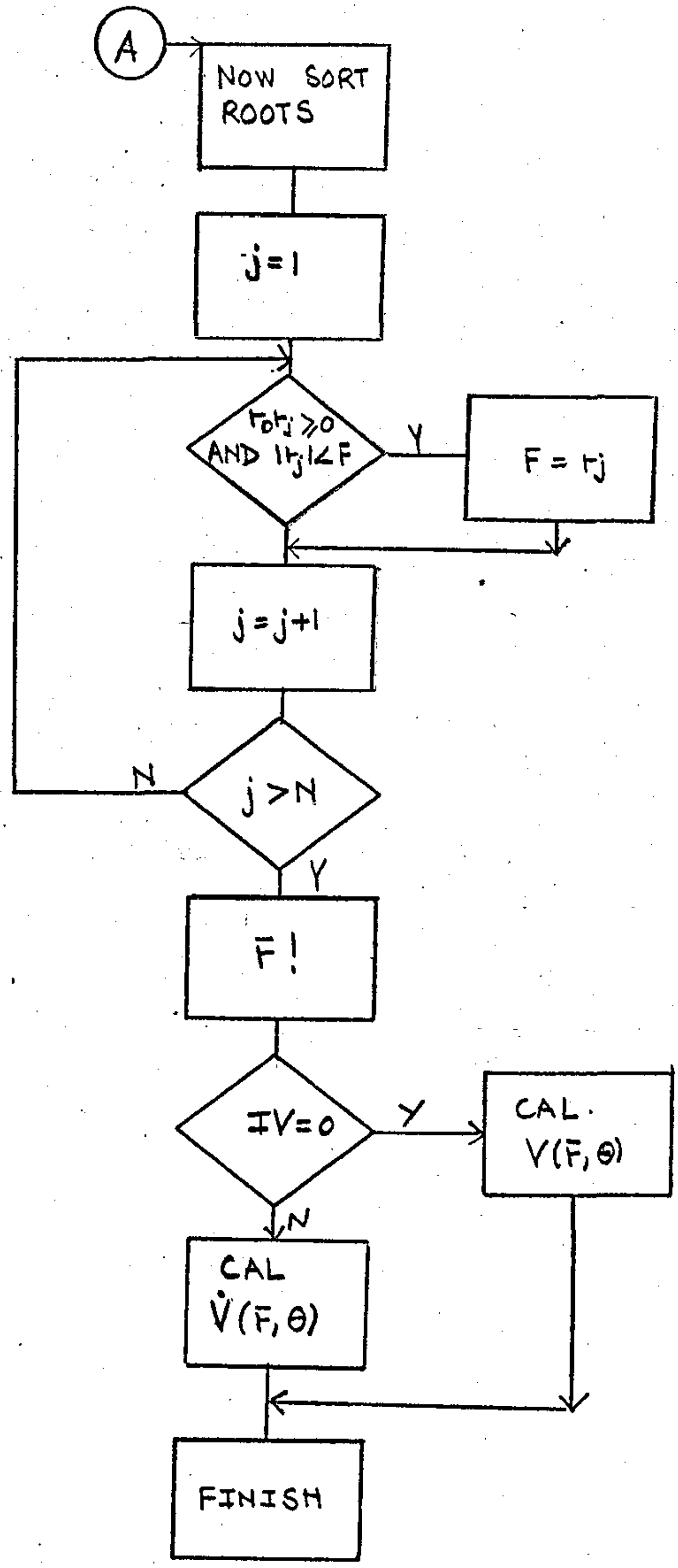


FIG. 4.12 (CONTD.)

ONE-DIMENSIONAL SEARCH ROUTINE

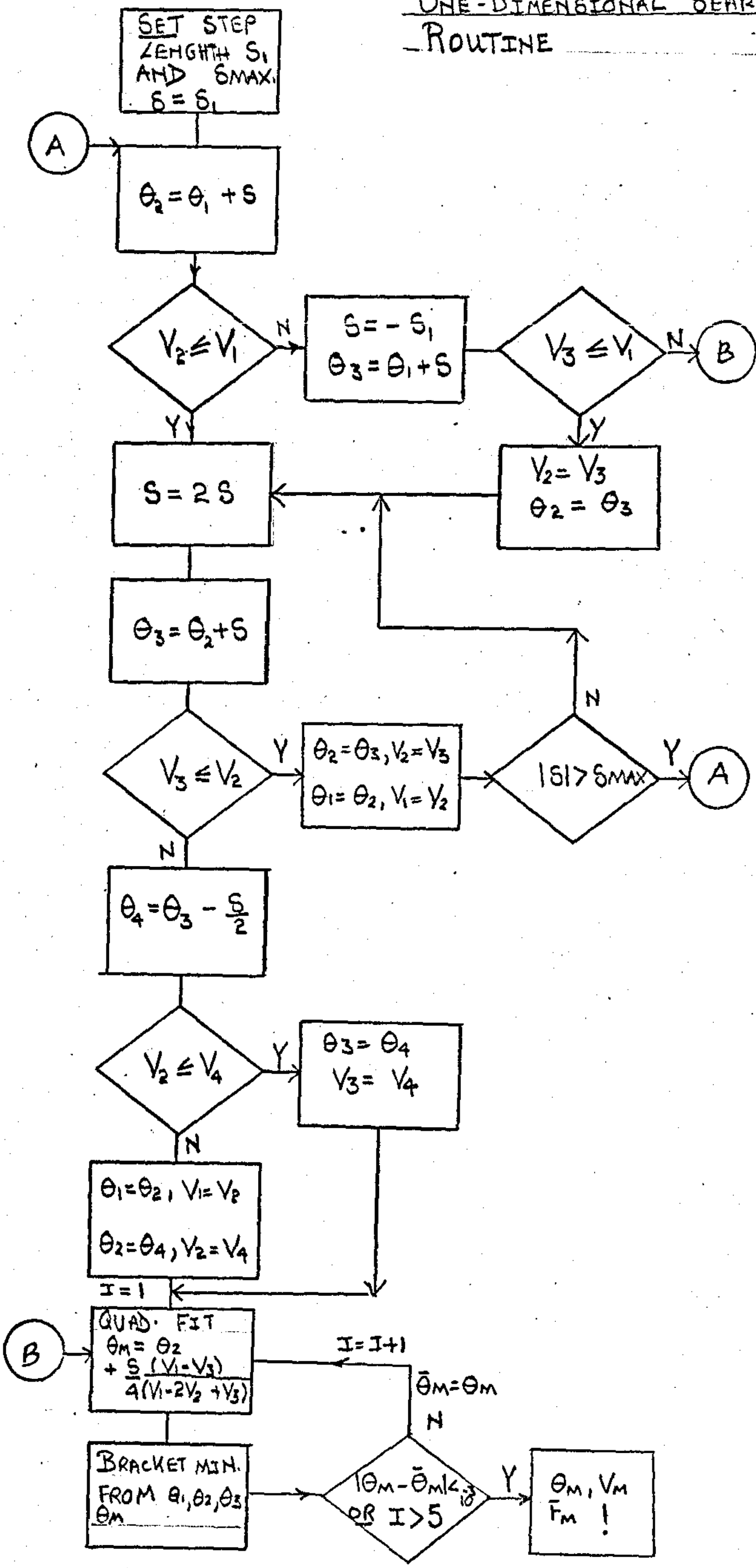
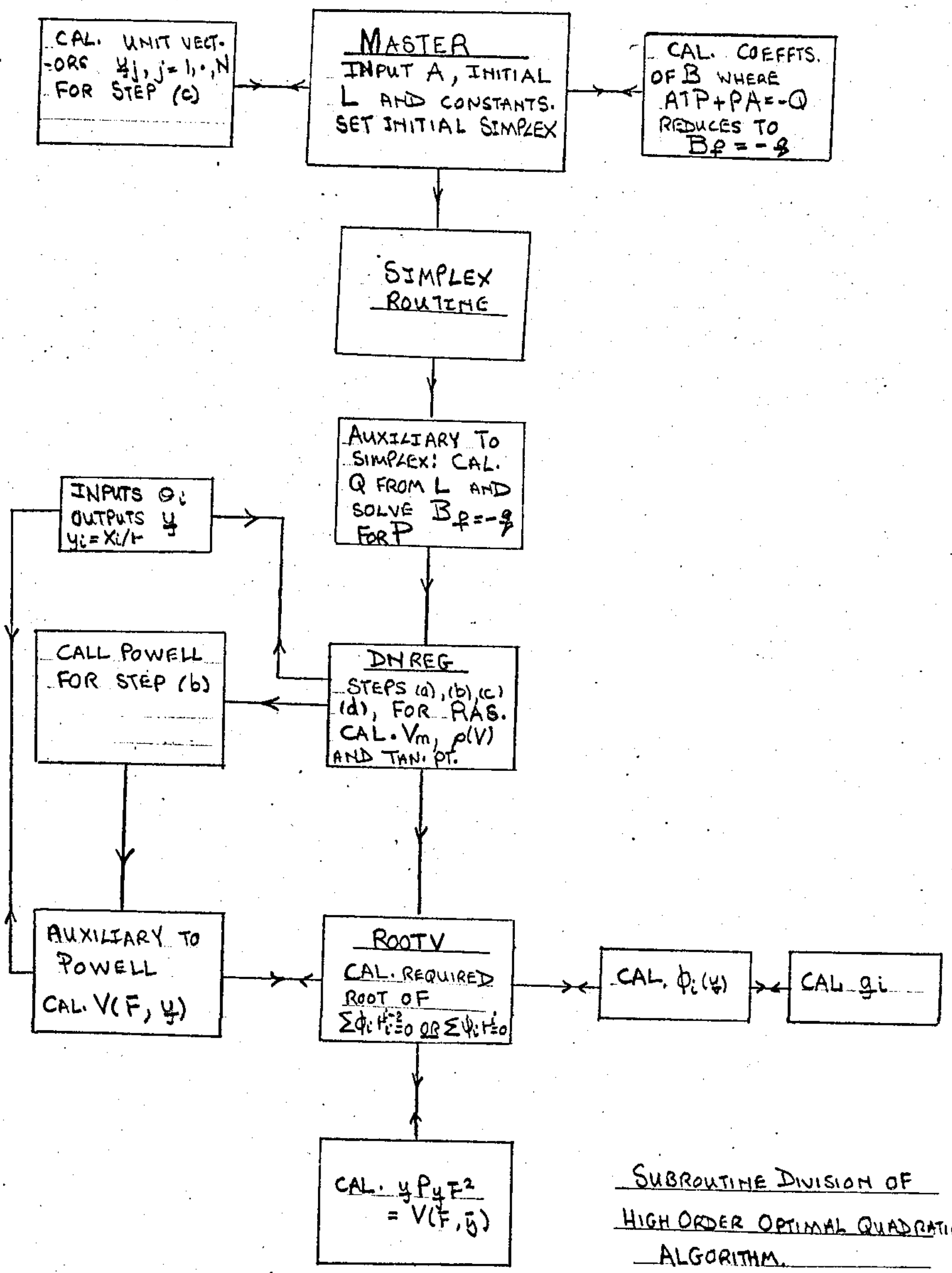
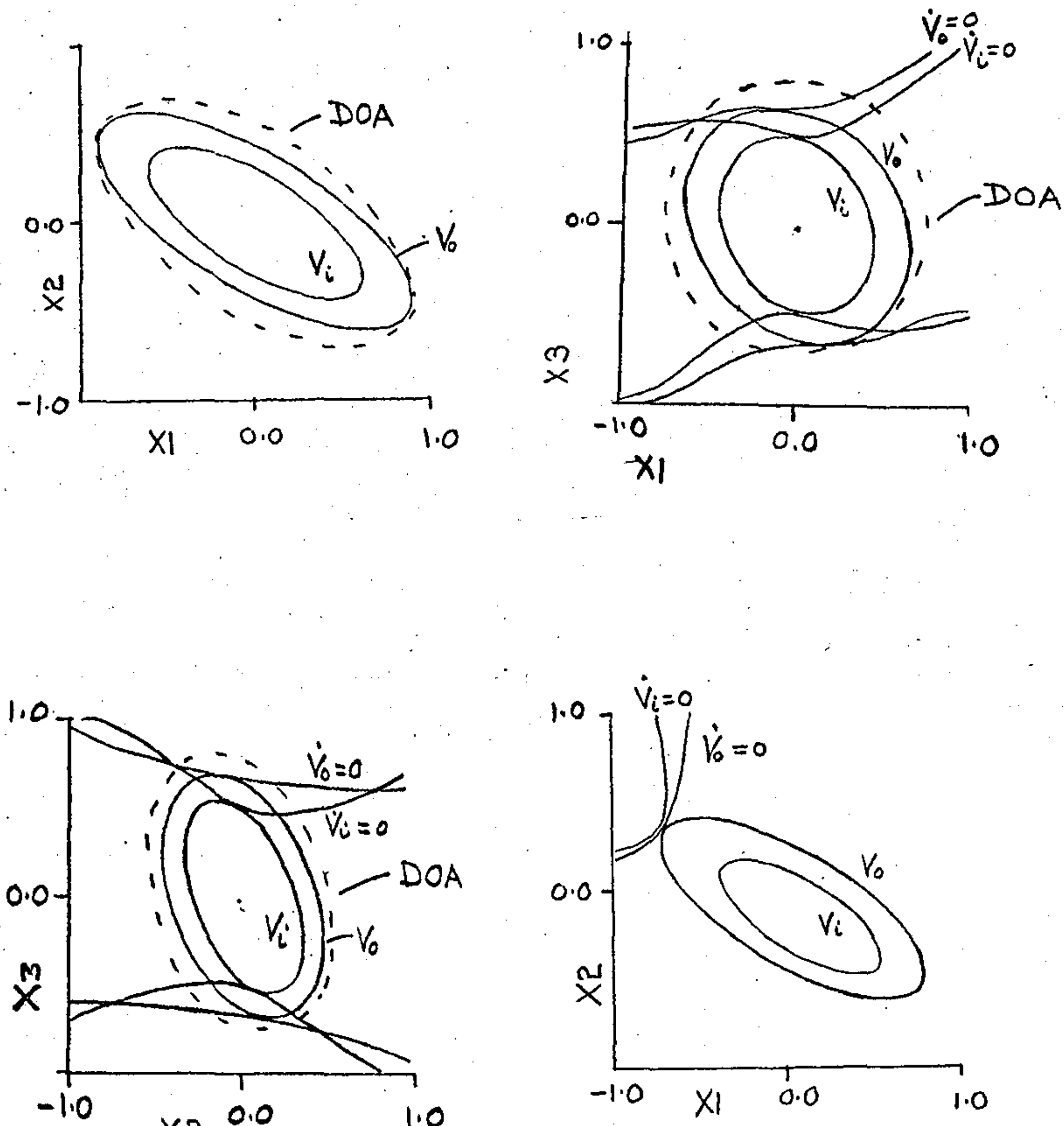


FIG. 4.13



SUBROUTINE DIVISION OF HIGH ORDER OPTIMAL QUADRATIC ALGORITHM.

FIG 4.15



SYSTEM S10

V_i, V_o INITIAL AND FINAL RAS BOUNDARIES

FIG 4.16

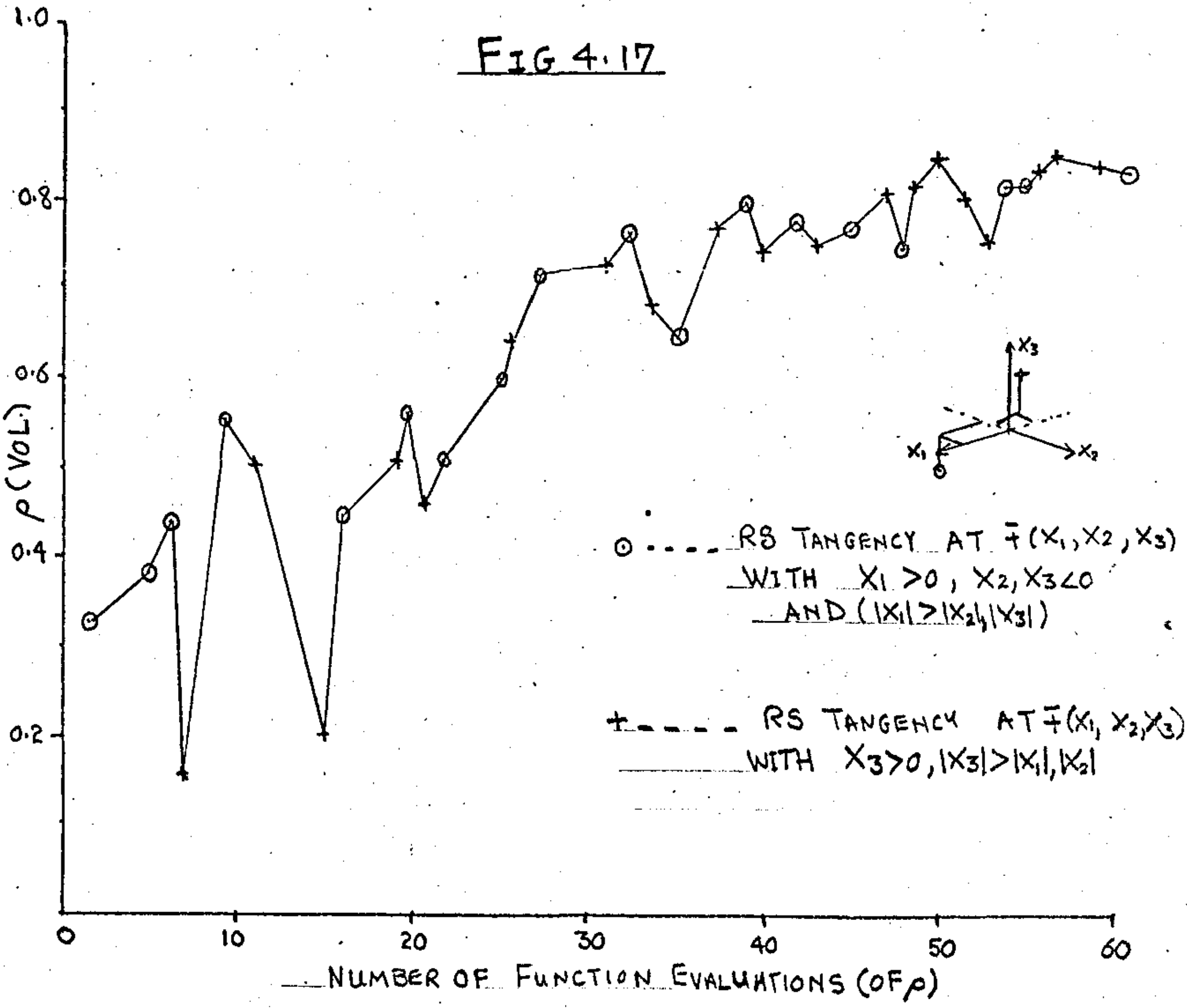


FIG 4.17

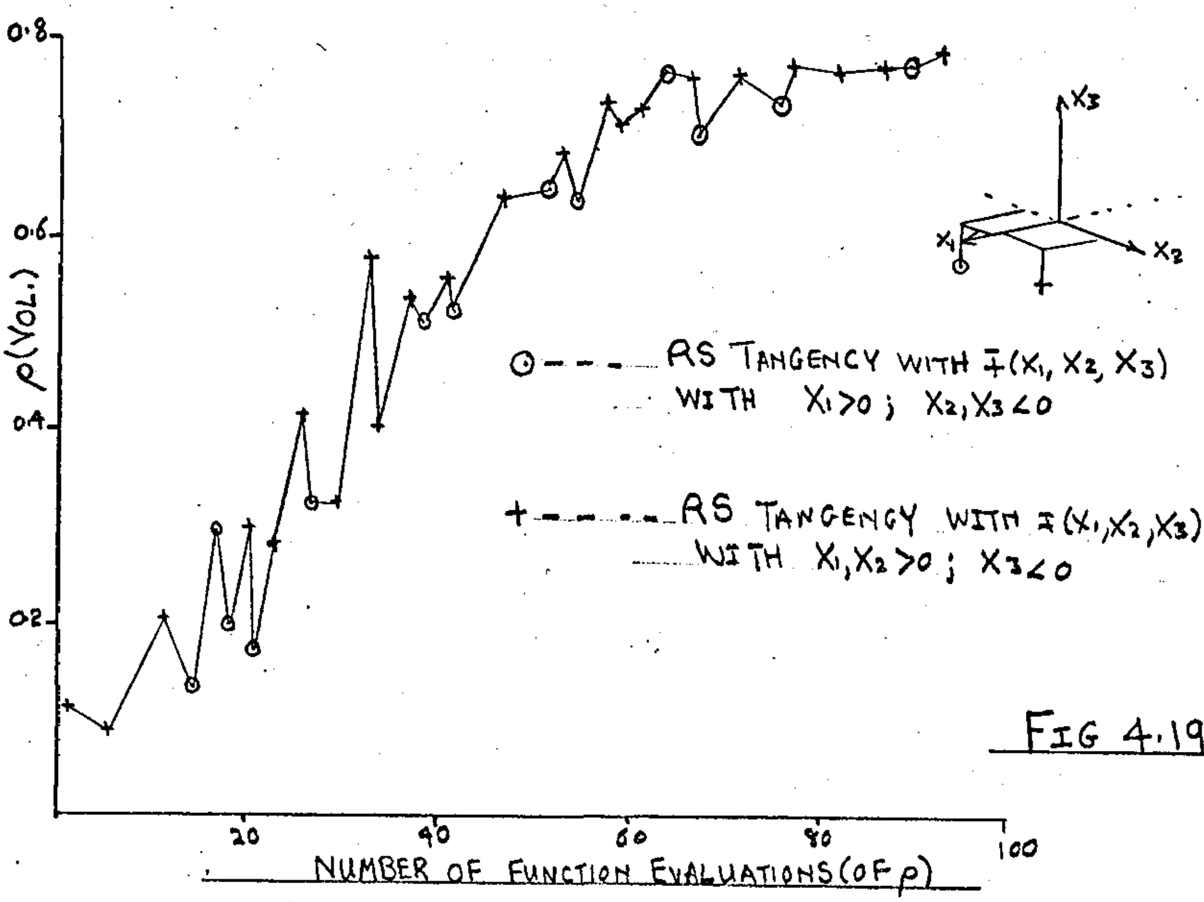
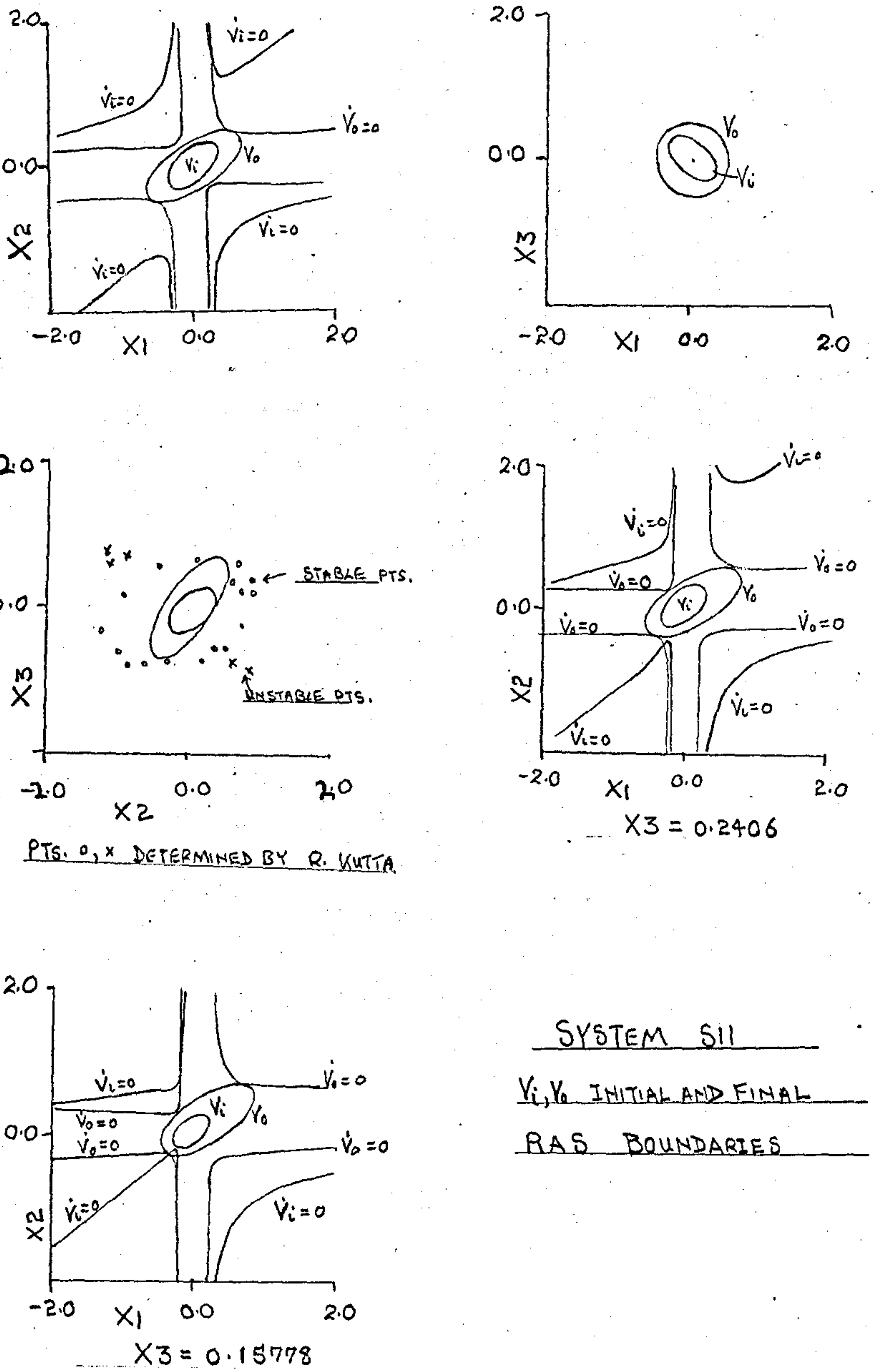


FIG 4.19



SYSTEM S11

V_i, V_o INITIAL AND FINAL

RAS BOUNDARIES

FIG 4.18

CHAPTER 5

COMPUTATIONAL METHODS FOR OPTIMAL QUADRATIC AND RAS DETERMINATION FOR GENERAL NON-LINEAR SYSTEMS.

Chapter 5

Computational Methods For Optimal Quadratic and RAS Determination For General Non-linear Systems.

5.1 Introduction

In this chapter two algorithms are proposed. The first is a method for optimal quadratic determination which does not rely on penalty functions as in Geiss (51) or tangency between hypersurfaces as in Rodden (3). It is somewhat heuristic and is based on an idea of Davidson and Kurak (47) who have developed a method which uses the special properties of a quadratic. The second is a method which determines an RAS for a general quadratic Lyapanov function via a penalty function approach. It automatically finds the valid tangency point to a desired accuracy.

5.2 An Optimal Quadratic Algorithm

Development

The method deals with the system

$$\underline{\dot{x}} = \underline{f}(\underline{x}) \qquad (\underline{f} \in E) \qquad 5.2.1$$

Let A be the Jacobian of $\underline{f}(\underline{x})$ at $\underline{x} = 0$, assumed stable. Then as in chapter 4 the class of quadratic LF's is determined as $V = \underline{x}^T P \underline{x}$, where P solves the matrix equation $A^T P + PA = -Q$ for p.d.s. Q. By Theorem 1.5.3 the region W given by $W : (\underline{x}/\underline{x}^T P \underline{x} < c)$, is a stability region for c sufficiently small, the RAS corresponding to c_{max}. Replacing P by P/c the boundary of W may be written more conveniently as

$$\underline{x}^T P \underline{x} = 1 \qquad 5.2.2$$

which bounds the volume

$$\rho = \pi/\sqrt{d(P)} \qquad 5.2.3$$

The OQ problem with respect to ρ is then formulated as

$$\min d(P) = \prod_{i=1}^{n} \lambda_i(P) \qquad 5.2.4$$

subject to the constraints

(1) P positive def. symmetric $\qquad$ 5.2.5

(2) $\underline{f}^T(\underline{x})P\underline{x} < 0, \; \forall \; \underline{x} \in W,$ $\qquad$ 5.2.6

with W now, $W : (\underline{x}/\underline{x}P\underline{x} < 1, \; \underline{x} \neq 0)$. (Note we have assumed $\|\underline{f}\| \rightarrow 0$ as $\|\underline{x}\| \rightarrow 0$).

Essentially, ρ is optimized over the p.d.s matrices(P) while constraint (2) ensures that the boundary $\underline{x}^T P\underline{x} = 1$ is a stability boundary. The main problems are then, the choice of optimization technique and the evaluation of the constraints in (1) and (2).

Constraint Evaluation

Since any p.d. P is expressible as $P = L^T L$ where L, $d(L) \neq 0$, is an upper triangular matrix, constraint (1) is avoided by optimizing ρ over L. Constraint (2) can be replaced by

$$\dot{V}_M(P) < 0 \qquad 5.2.7$$

where

$$\dot{V}_M(P) = \max_{\underline{x}} \underline{f}(\underline{x})^T P\underline{x}, \qquad \underline{x} \in W \qquad 5.2.8$$

In order to evaluate the latter the following approximation is made : store pre-determined unit vectors,

$\underline{y}_j$, j = 1, 2, ... , N_n, which ideally cut a closed surface

($\underline{x}^T P \underline{x} = 1$ or $\underline{x}^T \underline{x} = 1$ say) at equally spaced points. A point on the surface $\underline{x}^T P \underline{x} = 1$ is then given by

$$\frac{\underline{y}_j}{\|L\underline{y}_j\|} = \frac{\underline{y}_j}{\sqrt{\underline{y}_j^T P \underline{y}_j}} \quad ; \quad j = 1, \ldots , N_n$$

Hence

$$\underline{f}^T P \underline{x} \Big|_{\underline{x} \in W} = \frac{K}{K_f} \left[L\underline{f}\left(\frac{K\underline{y}_j}{K_f \; \|L\underline{y}_j\|}\right) \right]^T \frac{L\underline{y}_j}{\|L\underline{y}_j\|}$$

$$j = 1, \ldots , N_n$$
$$K = 1, 2, \ldots K_f,$$

where K_f is the number of grid points occuring along the straight line between $\underline{x} = 0$ and $\underline{y}_j / \|L\underline{y}_j\|$. Although $K_f = 5$ was chosen by Davidson (see later), it was found sufficient only to evaluate $\dot{V}$ on the surface $\underline{x}^T P \underline{x} = 1$ $(K_f = 1)$, mainly because of the starting procedure that follows. In short, the problem is now

$$\min_L d(L) \qquad (\text{i.e.} \quad \max \rho) \qquad 5.2.9$$

subject to

$$\dot{V}_M(L) = \max_{j=1,\ldots,N_n} \left[L\underline{f}(\underline{y}_j / \|L\underline{y}_j\|) \right]^T \frac{L\underline{y}_j}{\|L\underline{y}_j\|} < 0$$

5.2.10

provided the initial boundary, $\|L\underline{x}\|^2 = 1$, is a stability boundary. (The situation $d(L) = 0$ is rarely encountered, but $P = L^T L + \epsilon I$ can always replace constraint (1)).

The main points of the algorithm are as follows :

(a) Choice of Minimization Routine

Box's (55) Complex method is used to minimize d(L) subject to $\dot{V}_{M(L)} < 0$, and the variation used is given in A4. The method is simple to program and often works well for non-convex constraints.

(b) Initial L

In the absence of any other quadratic a natural choice is to choose $Q = I$ and solve

$$A^T P_1 + P_1 A = -Q \qquad 5.2.11$$

for P_1. For many practical systems this often leads to an initial boundary, $\underline{x}^T P_1 \underline{x} = \epsilon > 0$, being extremely eccentric which implies that the components of $\dot{V} = 0$ are usually long thin surfaces. The approximation in 5.2.10 is then only accurate with an extremely fine mesh and computation time is excessive. In this case the choice $P_1 = (SS^*)^{-1}$ is made from 2.2.2 where S has unit column vectors.

(c) Initial Boundary

If P_1 is factored as $P_1 = L_1^T L_1$ then

$$\underline{x}^T P_1 \underline{x} = \|L_1 \underline{x}\|^2 = l_o \quad (l_o = 10^{-5} \text{ say}) \qquad 5.2.12$$

gives a stability boundary ($\dot{V} < 0$ for $\|\underline{x}\| < \epsilon$, ϵ small). The vectors $\dfrac{\underline{y}_j}{\|L\underline{y}_j\|}$ are calculated and stored in $\underline{y}_j$ $(j = 1, 2, \ldots, N_n)$. A one-dimensional search is now made on l, starting with $l = l_o$, so as to increase the initial boundary in 5.2.12 :

$$\max\ (l) \qquad 5.2.13$$

subject to

$$\dot{V}_M(l) = \max_{j=1,\ldots,N_n} \left[L_1 \underline{f}(\sqrt{l}\,\underline{y}_j) \right]^T L_1 \underline{y}_j < 0$$

In practice a simple bisection method was adequate for maximizing l such that if finally l_2 and l_3 are bracketing values – $\dot{V}_M(l_2) < 0$ and $\dot{V}_M(l_3) > 0$ – and

$$\left| \frac{l_3 - l_2}{l_2} \right| < er \qquad 5.2.14$$

the required l was obtained.

In detail the steps are :

(1) put $l_1 = l_o$ (if $\dot{V}_M(l_o) > 0$ reduce l_o further) and with suitable s put $l_2 = l_1 + s$

(2) If $\dot{V}_M(l_2) < 0$, $s = 2s$, $l_1 = l_2$ and $l_2 = l_1 + s$ are made and (2) is repeated until $\dot{V}_M(l_2) > 0$

(If $s > s_{max}$, an upper bound, repeat (1) with larger s)

(3) then select $s = s/2$, $l_3 = l_1 + s$ and if $\dot{V}_M(l_3) < 0$ put $l_1 = l_3$, otherwise $l_2 = l_3$.

Repeat (3) until 5.2.14 is satisfied. (Typically, $s = .1$, $s_{max} = 10.0$, $er = .3$; the latter being sufficiently large to give a good initial complex).

The final boundary is $\underline{x}^T P \underline{x} = l_2$ or $\underline{x}^T P \underline{x} = 1.0$ with $P = P_1 / l_2$ (l_2 corresponds to l_1 in step (3))

(d) Choice of Unit Vectors

(1) A simple choice is the set of vectors from the origin to equally spaced points on the unit n-sphere via polar co-ordinates in 4.5.1.

(2) For highly eccentric surfaces it is advisable to transform $\underline{x}^T P_1 \underline{x} = 1$ to $\sum \lambda_i z_i^2 = 1$ via an orthogonal transformation $\underline{x} = T\underline{z}$ ($T^T = T^{-1}$) with $T = (\underline{z}_1, \underline{z}_2, .., \underline{z}_n)$

and $\underline{z}_i^T \underline{z}_i = 1$. Form an n-dimensional box with axes along $\underline{z}_i$ and sides of lengths $2/\sqrt{\lambda_i(P_1)}$. Divide its surface into an equally spaced mesh, then the points give the required vectors when normalized. (Note that the initial vectors need not necessarily be unit vectors).

(e) Initial Complex

The initial boundary, $\underline{x}^T P \underline{x} = 1$, (in (c)) gives an initial feasible point

$$\underline{p}_o = (t_1, t_2, \dots , t_m)$$

where m = n(n + 1)/2 - the number of elements of L, $P = L^T L$ and L is written as

$$L = \begin{bmatrix} t_1 & t_2 & t_4 & \cdot & \cdot & t_{m-n+1} \\ 0 & t_3 & t_5 & \cdot & \cdot & \cdot \\ 0 & 0 & t_6 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & t_m \end{bmatrix}$$

A complex of 2m + 1 points is used, the remaining 2m feasible points being obtained in the following way :

Introduce an m x 2m matrix D, partitioned as $D : (I_m, D_1) = (\underline{d}_1, \underline{d}_2, \dots , \underline{d}_{2m})$ with I the m x m unit matrix and D_1 a matrix whose columns, $\underline{d}_i$ (i = m+1, ... , 2m), have random elements of zeros or ones. Let $s_o = h \max_i |t_i|$, then a tentative j^{th} feasible point, $\underline{p}_j$ (j = 1, ... , 2m), is determined as

$$\underline{p}_j = \underline{p}_o + \underline{d}_j s \ , \quad s = s_o$$

If $\dot{V}_M(\underline{p}_j) < 0$ from 5.2.10, the point $\underline{p}_j$ is 'feasible' and is accepted, but if $\dot{V}_M(\underline{p}_j) \geq 0$ we find $\dot{V}_M(\underline{p}_j^*) = \dot{V}_M(\underline{p}_o - \underline{d}_j s)$. If $\dot{V}_M(\underline{p}_j^*) < 0$, $\underline{p}_j^*$ is accepted, otherwise s is halved and the process repeated until a feasible point is found or until $s < .01h$, say, when the complex is deemed too small ($h = .2$ was found practical).

It was found that when er was made too small in obtaining the initial boundary, either the resulting initial complex was also very small or that function values at points $\underline{p}_j$ were much less than that at $\underline{p}_o$. The idea of Box (55) and others of generating random and possibly large $\underline{p}_j$, was impractical due to the presence of the equal tangency surface (c.f. Chapter 4). A point lying on a line drawn between two feasible points on opposite sides of this surface was usually infeasible (no matter how close the points).

(f) Other Features

The unit vectors need only span a half space of E^n, $x_3 > 0$ say, for then 5.2.10 is evaluated as

$$\dot{V}_M(L) = \max_{j=1,\ldots,N_n} \mp\left[L\underline{f}\left(\mp\frac{\underline{y}_j}{\|L\underline{y}_j\|}\right)\right]^T \frac{L\underline{y}_j}{\|L\underline{y}_j\|} \qquad 5.2.15$$

A quick routine is used to calculate $L\underline{x}$, the resulting vector being stored in $\underline{x}$, thus

$$x_i = \sum_{j=i}^{n} l_{ij} x_j$$

Also if $\underline{f}(-\underline{x}) = -\underline{f}(\underline{x})$ computation of $\dot{V}_M(L)$ is halved due to radial symmetry of the contour $\dot{V}(\underline{x}) = 0$.

On exit from the Complex minimization the final boundary, $\underline{x}^T P \underline{x} = 1$, is an approximation to the RAS boundary, $\underline{x}^T P \underline{x} = V_M$. The former boundary is then either verified by using a finer mesh to calculate $\dot{V}_M(L)$ or the exact RAS is obtained by an algorithm such as that in section 5.4.

The number of calculations required to evaluate 5.2.15 is effectively $N_n(\frac{3}{2}n^2 + n(2c_1 + 11/2))$ with c_1 the average number of calculations to evaluate each f_i. Assuming $N_n = c^n$ the approximate constraint and optimal quadratic computation time for large n will vary as $c^n(\frac{n^2 3}{2} + 2c_1 n)$ and $\frac{K_1 n^2 c^n}{2}(\frac{3n^2}{2} + 2c_1 n)$ respectively, the optimization being over m ($\simeq n^2/2$) parameters (in practice $5 < c > 10$).

The algorithm has been programmed in Fortran IV and Fig's 5.1 and 5.2 show flow diagrams for the master and connecting segments. (A listing is given in A 5).

<u>Summary of Algorithm</u>

For completeness we summarize the main steps of the new algorithm along side those of Davidson and Kurak's in which

$$\dot{V}_M(P) = \max_{\substack{j=1,\ldots,N_n \\ K=1,\ldots,K_f}} \underline{f}^T\left(\frac{K\underline{y}_j}{K_f\sqrt{\underline{y}_j^T P \underline{y}_j}}\right) P \underline{y}_j \qquad 5.2.16$$

Step	New	Davidson & Kurak
(1)	Evaluate $A = \left.\frac{\partial \underline{f}}{\partial \underline{x}}\right\vert_{\underline{x}=0}$ and solve	
	$A^TP_1 + P_1A = -Q$ for P_1 with	
	a) $Q = I$ or b) Choose $P_1 = (SS^*)^{-1}$	a) $Q = I$
(2)	Starting with $l = l_o$ (= 10^{-5} say) max (l) subject to	
	$\dot{V}_M(l) < 0$ (5.2.13) by bisection method (in(c) This gives $\underline{x}^TP\underline{x} = 1$ $P = P_1/l_2 = L_1^TL_1/l_2$	$\dot{V}_M(P_1/l) < 0$ (5.2.16) by Rosenbrock (in A4) This gives $\underline{x}^T\bar{P}\underline{x} = 1$ with $\bar{P} = P_1/l_{max}$
(3)		Starting with this $\bar{P}$ min $M(\bar{P})$ subject to $m(\bar{P}) > 0$ and $\dot{V}_M(\bar{P}) < 0$ (5.2.16)
(4)	Starting with $L = L_1/\sqrt{l_2}$ min $d(L)$ (max ρ) subject to $l_{ii} \neq 0$ and $\dot{V}_M(L) < 0$ by the Complex optimization routine (A4)	Starting with $P = \bar{P}$ min $\prod \lambda_i(P)$ subject to $m(P) > 0$ and $\dot{V}_M(P) < 0$ by Rosenbrock ($K_f = 5$)
(5)	Verify $\dot{V} < 0$ in W : ($\underline{x}$ $\underline{x}^TP\underline{x} < 1$, $\underline{x} \neq 0$) by evaluating	
	a) $\dot{V}_M(L)$ with fine mesh or b) exact RAS (Section 5.4)	a) $\dot{V}_M(P)$ with fine mesh

Steps (3) and (4) of Davidson's algorithm require the evaluation of the eigenvalues of P (Jacobi's method used).

5.3 Numerical Results

Table 5.1 shows details of a comparison of the new method with that of Davidson for the following systems :

System S13 (Yu(6))

Consider the following equations for a synchronous generator

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

$$\dot{x}_2 = -R - Dx_2 \qquad \text{S13}$$

where $D = 0.0372 + 0.152\cos(2x_1 + 1.7744)$

and $R = \sin(x_1 + .8872) - \sin(.8872)$

$- .1291(\sin(2x_1 + 1.7744) - \sin(1.7744))$

With $Q = I$, $1_o = 10^{-5}$ and $N_2 = 30$ the initial boundary was expanded to

$$24.71x_1^2 + 1.4624x_1x_2 + 36.1x_2^2 = 1.1$$

Fig 5.3 shows property A is present with the final boundary

$$11.477x_1^2 + .928x_1x_2 + 17.02x_2^2 = 1.0$$

having two points of contact with its constraint contour. The RAS is not a good estimate of the DOA.

System S14 (Hewit (2))

The system is a variation of a surge-tank system found in Hewit (2)

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

$$\dot{x}_2 = -x_1(1 - a(2 + x_1))/(1 + x_1)^2 \qquad \text{S14}$$

$$- \frac{a^2}{b}x_2^2 - \frac{b}{a}x_2\left(\frac{2a^2}{b}(1 + x_1) - 1\right)/(1 + x_1)^2$$

$$a = .2, \quad b = .075$$

Fig 5.4 shows the OQ boundary has not two but three points of contact with its constraint contour, the initial quadratic having only one. (i.e the OQ has 3 valid tan.pts).

TABLE 5.1

D = DAVIDSON & KURAK				C = NEW COMPLEX METHOD			
FOR D : $h=.2, l_o=10^{-5}, er=0.3$				FOR C : $l_o = 10^{-5}$, $z = -1.0$ (A4)			
SYSTEM		INITIAL P OR Q	N_n	INITIAL VOL.	FINAL VOL.	NO. OF CONSTRAINT EVALUATIONS ($\dot{V}_M(L)$)	TIME IN MILL/SEC.
S13	D	Q = I	30	0.116	0.22104	688	350
	C	"	"	"	0.2249	360	120
S14	D	Q = I	30	0.69×10^{-3}	1.169×10^{-3}	732	280
	C	"	"	"	1.179×10^{-3}	420	160
S15	D	Q = I	180	0.516	0.960	1300	2000
	C		"	"	2.85	1400	1220
	C	RESTART	"	2.85	3.075	460	780
S16	D	Q = I	180	3.5	25.0	1370	2000
	C	"	370	3.5	50.1	800	2000
S11	D	$P=(SS^*)^{-1}$	180	.104	0.26313	1230	3000
	C	"	"	"	0.383	1300	3000
	D	Q = I	180	0.10	0.146	1200	2900
	C	"	"	"	0.285	800	1500
	C	$P=(SS^*)^{-1}$	370	0.104	0.376	900	1900
S17	C	$P=(SS^*)^{-1}$	1240	$.12\times10^{-3}$	$.165\times10^{-3}$	1040	5000

System S15

A system having a fixed constraint contour, $\underline{x}^T B \underline{x} = 1$, is

$$\dot{\underline{x}} = -\underline{x} + (\underline{x}^T B \underline{x})\underline{x} \qquad \textbf{S15}$$

with DOA, $\underline{x}^T B \underline{x} < 1$. We have chosen

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

Step (3) of Davidson's algorithm is redundant since P = I with Q = 2I and Table 5.1 exhibits the poor convergence of the algorithm. In fact the new algorithm was found superior to the latter for all third order systems tried.

A reason for the inferiority of Davidson was the lack of convergence of Rosenbrock's routine on the constrained problem. As the volume increased it was found that values of P were reached, exhibiting an equal tangency surface, where a large amount of time was spent by the search directions oscillating between the boundary zones, determined by Rosenbrock's penalty function (Appendix 4), and the feasible region $\dot{V}_M(P) < 0$. (Box (55) has noted this phenomenon).

The new Complex method gave better convergence although it seemed to stick near the optimum. Closer inspection showed that the vector $\underline{y}_j$, giving $\dot{V}_M(L)$, alternated between two or more places on the constraint contour. After restarting the procedure a final volume of $\rho = 3.07$ was obtained with

$$P = \begin{bmatrix} 1.035 & 1.028 & -0.009 \\ 1.028 & 2.025 & 0.992 \\ -0.009 & 0.992 & 2.003 \end{bmatrix}$$

(For the OQ boundary, $\underline{x}^T B \underline{x} = 1$, $\rho = \pi$).

System S16

The system is that of S4 (Chapter 4) with $a_1 = 1$, $a_2 = 2$ and $a_3 = 3$. Again the new method is superior. The best volume achieved by the third order algorithm of chapter 4 was $\rho = 68.01$ which is superior to the results in Table 5.1.

System 11 (C4)

For this system a lack of convergence was found for both methods which was attributed to the presence of the equal tangency surface, with two RS components of the constraint contour. For the new algoritm, whatever the initial L_1 (or P_1), a point near this surface was obtained which tended to shrink the complex of points so that further progress was poor.

Step (3) of Davidson's method of minimizing $M(\bar{P})$, originally intended to make the initial boundary less eccentric and give a good starting P for the final minimization of step (4), proved unsatisfactory. In many cases the procedure gave a smaller starting volume. Infact, Davidson (4) gives the OQ boundary as

$$\underline{x}^T P \underline{x} = \underline{x}^T \begin{bmatrix} 12.5 & -8.1 & 3.0 \\ \cdot & 20.8 & -8.5 \\ \cdot & \cdot & 13.4 \end{bmatrix} = 1.0$$

which is incorrect (the RAS for this P satisfies $\underline{x}^T P \underline{x} < 10^{-4}$) and results from the poor initial P from step (3). We note that the boundary $\underline{x}^T P \underline{x} = 1$ can be made less eccentric, more effectively, by minimizing $\mu(P) = M(P)/m(P)$, but in so doing the corresponding Q may approach a positive semi-definite matrix with an invalid LF. However, a system where the procedure is useful is the following approximation to a relay system

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

$$\dot{x}_2 = -\tanh(100(x_1 + x_2)) + x_2$$

with

$$A = \begin{bmatrix} 0 & 1 \\ -101 & -99 \end{bmatrix}$$

The initial (expanded) boundary for Q = I is

$$\underline{x}^T \begin{bmatrix} 1.005 & 0.0049 \\ \cdot & 0.00504 \end{bmatrix} \underline{x} = 2.55 \text{ X } 10^{-5}$$

which is very eccentric. After 970 constraint evaluations of $\dot{V}_M(\bar{P})$ in minimizing $M(\bar{P})$ we have

$$\bar{P} = \begin{bmatrix} 4.96 & 1.30 \\ \cdot & 2.42 \end{bmatrix}$$

However the choice $P = (SS^*)^{-1}$ gives

$$P = \begin{bmatrix} 1.516 & .5413 \\ .5413 & .4844 \end{bmatrix}$$

which is a good initial estimate of the optimal P with boundary

$$\underline{x}^T \begin{bmatrix} 4.81 & 1.33 \\ \cdot & 2.41 \end{bmatrix} \underline{x} = 1.0$$

System S17 (Rao (5))

As a final example consider the system of a synchronous machine swinging against an infinite busbar

$$\dot{x}_1 = x_2 \qquad \text{S17}$$

$$\dot{x}_2 = 28.61 - 84.99(b + x_3)\sin(x_1 + a) + 21.53\sin 2(x_1 + a)$$

$$\dot{x}_3 = 0.36 - 0.621(b + x_3) + .421\cos(x_1 + a)$$

The quantities a and b result from a shift of origin (solutions of $\dot{x}_1 = \dot{x}_2 = \dot{x}_3 = 0$) and Rao gives a = 0.478

b = 1.18 for the stable singularity. As seen in Table 5.1 convergence is very slow and no reasonable OQ has been obtained. Two important points however, are raised :

(a) that an accuracy of 10^{-3} in determining a and b is insufficient, giving an invalid boundary for $l_o = 10^{-5}$ when infact the boundary is valid and

(b) that spurious quadratics were obtained with $N_3 = 370$ due to inadequate mesh size, the $\dot{V} = 0$ contours being long thin surfaces.

The defects were remedied by more exact calculation of a and b (a = .4779930, b = 1.1816655) and the choice $N_n = 1240$, with large increase in computing time. Sections of the best quadratic boundary through the co-ordinate planes and the tangency point (0.0327, 0.0229, -.0483) are shown in Fig 5.5 The RAS is poor.

Comment

The application of the algorithms described certainly shows that property A and its associated tangency surface are present for varied practical systems and can cause problems in convergence to the OQ. For high order systems the latter also suffers from the high dimensional problem of calculating $\dot{V}_M(L)$ and the storage of the unit vectors. (Geiss (51) has taken 24 min. for one RAS determination for a satellite problem with n = 9 and m = 45).

5.4 A Method For Quadratic RAS Determination For High Order Systems

Given the system 5.2.1 the problem of determining a RAS for a quadratic, $V = \underline{x}^T P \underline{x}$, has been formulated as the constrained problem (C4)

$$\min V(\underline{x}) \ , \ \underline{x} \in E_v \qquad 5.4.1$$

$$E_v : (\dot{V} = 0, \ \underline{x} \neq 0) \qquad 5.4.2$$

When using a penalty function (PF) approach two main points must be considered :

1) ensuring that the PF, or sequence of PF's, approach a local minimum of 5.4.1 excluding the trivial solution $\underline{x} = 0$;
2) ensuring that, if the minimum point is not a valid tangency point, a sequence of decreasing minima is obtained which eventually give the global minimum, and thus the valid tangency point.

To satisfy these criteria an algorithm has been developed, the main points of which are the following :

(a) Choice of Initial LF

Unless a particular RAS is required P is chosen as

$$P = (SS^*)^{-1}$$

(b) Initial Starting Point of Minimization

A good initial starting point, $\underline{x}_o$, for the PF minimization is found by storing unit vectors as in the Complex algorithm and then expanding the boundary in 5.2.12 until 5.2.14 is satisfied to a given accuracy (er = .1 say). The vector $\underline{y}_j$ giving the maximum value of $\dot{V}$ on $\underline{x}^T P \underline{x} = 1_2$ is then taken as $\underline{x}_o$.

(c) Choice of Penalty Function and Minimization Routine

To avoid the trivial solution $\underline{x} = 0$ the constraint in 5.4.2 is posed as

$$e(\underline{x}) = \dot{V}(\underline{x})/V^z = 0 \qquad (z = 1 \text{ or } 2) \qquad 5.4.3$$

Then a PF from Miele (62) is chosen as

$$W(\underline{x}) = V(\underline{x}) + \lambda e(\underline{x}) + Ke^2(\underline{x}) \qquad 5.4.4$$
$$= R(\underline{x}, \lambda) + Ke^2(\underline{x})$$

where λ is essentially an approximation to the Lagrange multiplier and $K > 0$ a penalty constant.

Starting with $\underline{x} = \underline{x}_o$, W is minimized for several cycles by the algorithm of Fletcher and Powell (60, 64) of conjugate directions, where a 'cycle' consists of n iterations of the latter with K and λ kept constant. After each cycle λ and K are updated and convergence is obtained when the conditions

$$|e| < \epsilon$$

and 5.4.5

$$\|\underline{\nabla} R(\underline{x}, \lambda)\| < \epsilon$$

are satisfied (typically $\epsilon = 10^{-4}$). If however NC, the number of cycles, exceeds some upper bound (NC > NIT say) the minimization is termed a failure and another initial $\underline{x}_o$ is sought (see A4 for updating of λ and K and Fletcher-Powell routine).

Since the Fletcher-Powell routine requires gradients of W we assume $J(\underline{f})$, the Jacobian of $\underline{f}$, can be suitably calculated. Then for $z = 1$

$$\underline{\nabla} W = \underline{\nabla} V + \underline{\nabla} e(\lambda + 2Ke) \qquad 5.4.6$$

where $\nabla e = \frac{1}{V}2(V\nabla\dot{V} - \dot{V}\nabla V)$

and $\nabla\dot{V} = 2(P\underline{f}(\underline{x}) + J(\underline{f})^T P\underline{x})$

Prior to the minimization, scale factors for $\underline{x}$, V and e are derived (xsc_i, vsc and esc) such that at the initial point $\underline{x} = \underline{x}_o$ the scaled values (x_i/xsc_i, V/vsc, e/esc) give W and its first and second derivatives approximately unit magnitude (Haarhoff and Buys (65)). In this case K may be chosen relatively small and often kept constant throughout the minimization (K = 20 is suitable).

(d) Validification of Tangency Point

For convergence let $\underline{x} = \bar{\underline{x}}$ be the minimizing value of the PF for the last cycle. Since $\dot{V}(\bar{\underline{x}})<0$ may not be satisfied $\bar{\underline{x}}$ is repeatedly multiplied by a scalar γ, close to one, until $\dot{V}(\gamma^j\bar{\underline{x}}) < 0$ for some j (e.g $\gamma = 1 - 10^{-4}$). This gives $\underline{x}_m = \gamma^j\bar{\underline{x}}$ as a modified tangency point and a possible stability boundary of

$$\underline{x}^T P\underline{x} = \underline{x}_m^T P\underline{x}_m = V_m \qquad 5.4.7$$

The validity of the latter is tested by evaluating $\dot{V}_M(L)$ in 5.2.15 with a fine mesh. If $\dot{V}_M(L) > 0$ the maximizing vector giving $\dot{V}_M(L)$ is found, $\underline{z}$ say, and the interval (0, 1.0) is successively halved until bracketing numbers E_1 and E_2 are found satisfying $0 < E_1 < E_2 < 1.0$ such that

$$\left|\frac{V(E_2\underline{z}) - V(E_1\underline{z})}{V(E_2\underline{z})}\right| < .1$$

where $\dot{V}(E_1\underline{z}) < 0$ and $\dot{V}(E_2\underline{z}) > 0$. The vector $\underline{x}_o = E_1\underline{z}$ determines a new starting point for the penalty function and the minimization is repeated.

Since Vm in 5.4.7 decreases for successive tangency points obtained, the method converges to a valid RAS boundary. However, if it is suspected that two tangency points exist giving almost the same boundary in 5.4.7, then either both of these must be found or an extremely fine mesh is required to evaluate $\dot{V}_M(L)$.

<u>Comment</u>

A slightly quicker, but less useful, initial point $\underline{x}_o$ in (b) may be obtained by expanding the sphere $\underline{x}^T\underline{x} = 1$ instead of the boundary $\underline{x}^T P\underline{x} = 1$.

The choice of PF in 5.4.4 is not arbitrary. Infact that of Miele was compared to one of Fletcher and Lill (63). They use a PF which has a stationary point at a solution of the constrained problem thus avoiding a sequential approach. With suitable q they minimize

$$W = V - e\left[\frac{\underline{a}^T\underline{b}}{\underline{a}^T\underline{a}}\right] + \frac{q}{2}\,\frac{e^2}{(\underline{a}^T\underline{a})} \qquad 5.4.8$$

with $\underline{a} = \nabla e$ and $\underline{b} = \nabla V$. The second term is here a <u>continuous</u> approximation to the Lagrange multiplier. However, with limited experience the method was not as good as Miele's.

Fig 5.6 gives a general flow diagram for the algorithm which was programmed in Fortran IV. Table 5.2 shows some typical features when applied to the system S1, for a simple quadratic, and S11 for the best quadratic of Chapter 4, the latter showing the effect of two tangency points. Computation time seems large but this is due to the chosen values of N_n, $N_2 = 31$ and $N_3 = 160$, for the initial boundary. ($N_2 = 10$ and $N_3 = 50$ would suffice, although giving a worse starting point $\underline{x}_o$).

TABLE 5.2

SYSTEM	S1	S11
P	I	18.101 -15.853 1.878 • 31.936 -16.586 • • 17.194
EVALUATIONS OF $\dot{V}_M(1)$	8	8
FINAL 1	1.2	0.375
N_n	31	160
CONVERGENCE VALUE : (5.4.5)	$\epsilon = 10^{-4}$	$\epsilon = 10^{-4}$
SCALE FACTORS (xsc_i , vsc , esc)	NONE	$.05, .05, .053, 4.996, 1.4 \times 10^{-3}$ *0.11, .05, .05, 4.33, .029
INITIAL $\underline{x}_0$	1.05, 0.575	-.229, .287, .0745 *-.420, -.384, .116
INITIAL v, λ, K	1.44,0.0,1.0	4.996,0.0,20.0 *4.33,0.0,20.0
NO. OF CYCLES	3	5 *10
FINAL $\underline{x}$ (TAN.PT.)	1.0746,0.6204	-.22101,.2937,.1447(INVALID) *-.36841,-.2614,.2496(VALID)
FINAL v, λ, K	1.5365,-.769, 1.0	4.630,-.36961,20.0 *4.467,-1.156,20.0
FINAL VOL	4.8271 (AREA)	0.763
TIME (MILL/SECS.)	15	207

* SECOND MINIMIZATION

MASTER SEGMENT FOR HIGH ORDER OPTIMAL QUADRATIC ALGORITHM

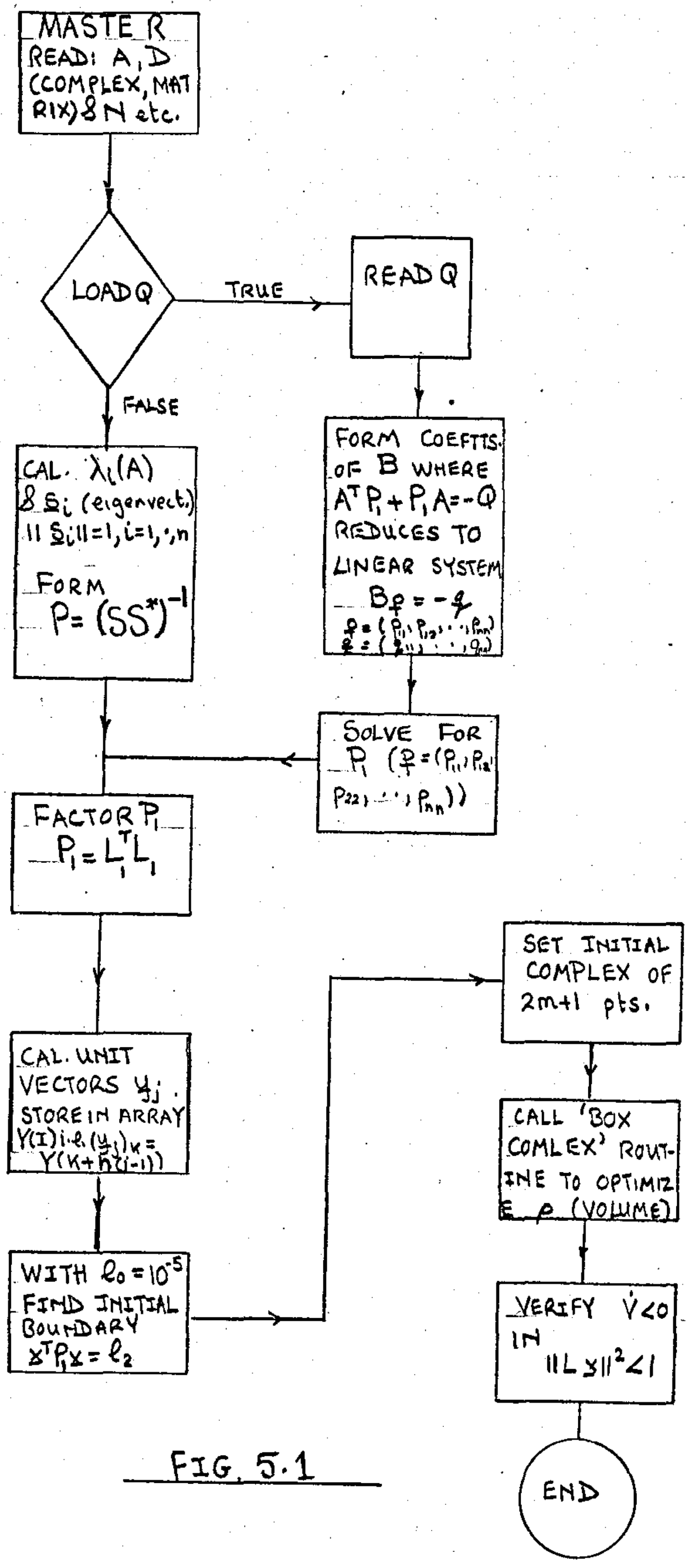


FIG. 5.1

Subroutine Division of High Order Optimal Quadratic Algorithm

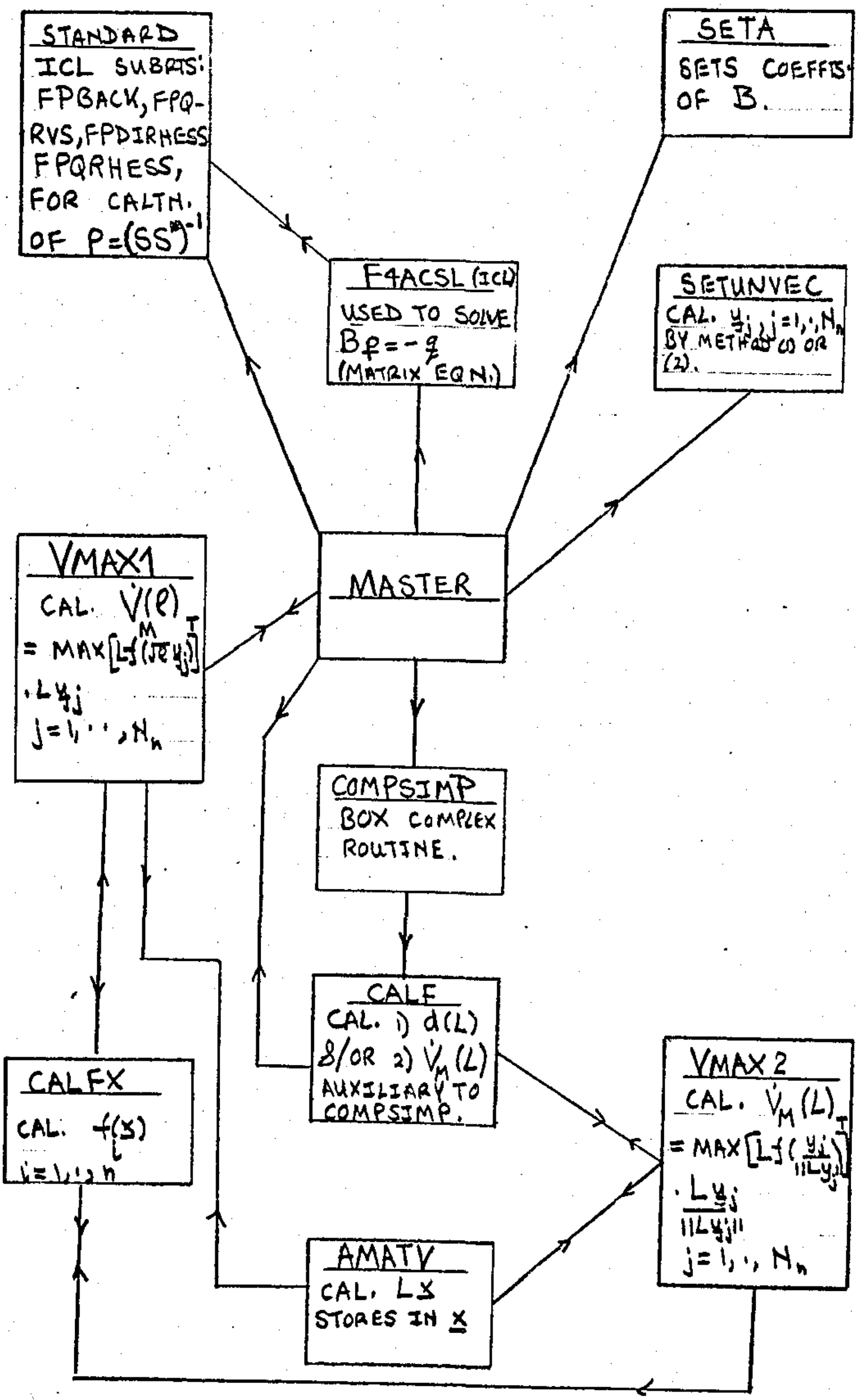


FIG. 5.2

SYSTEM S13

INITIAL AND FINAL QUADRATICS, (V_i AND V_o)

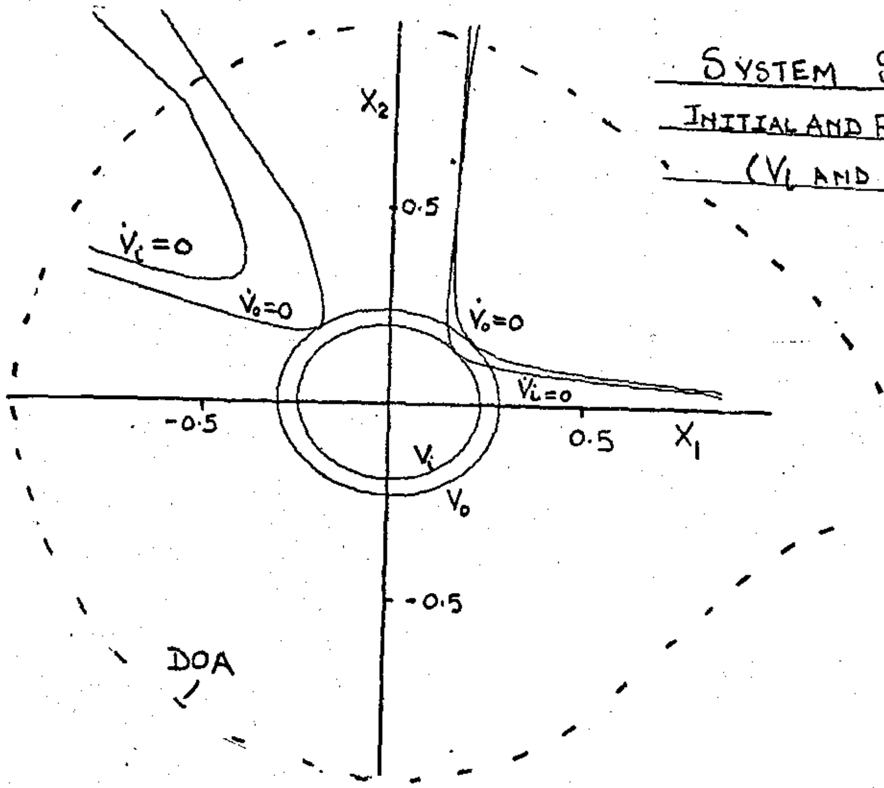
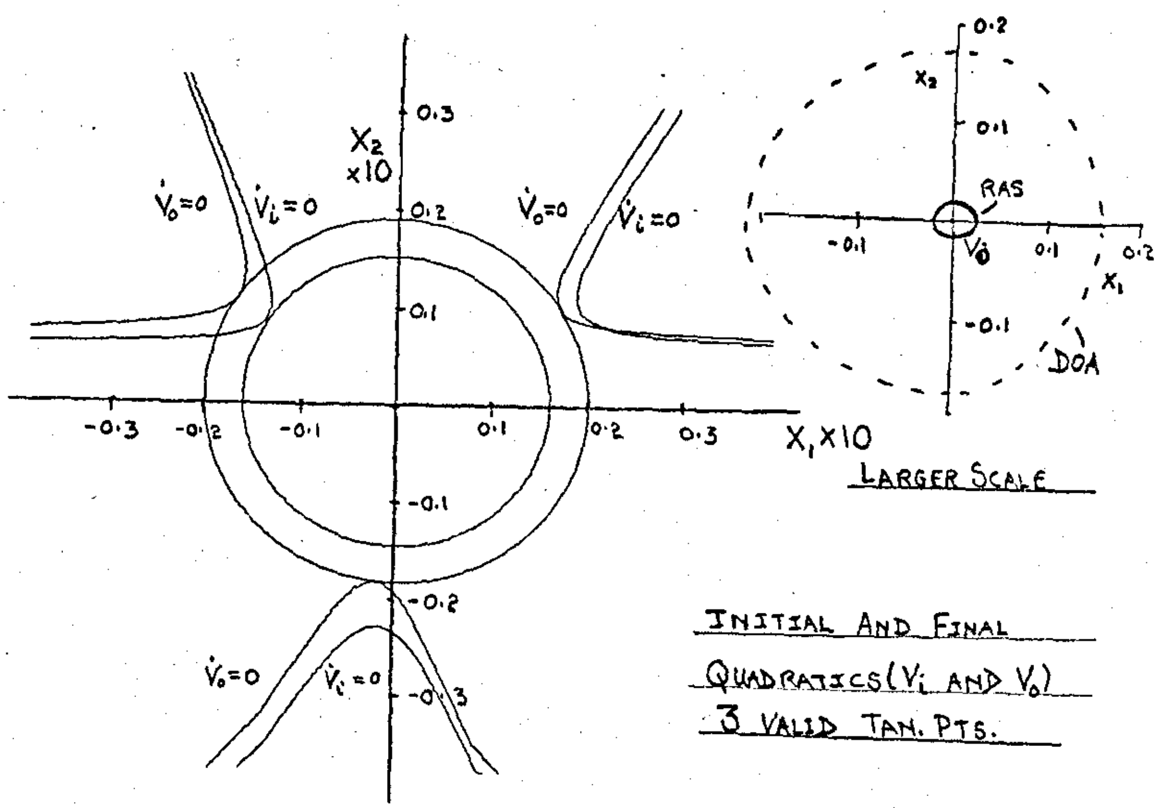


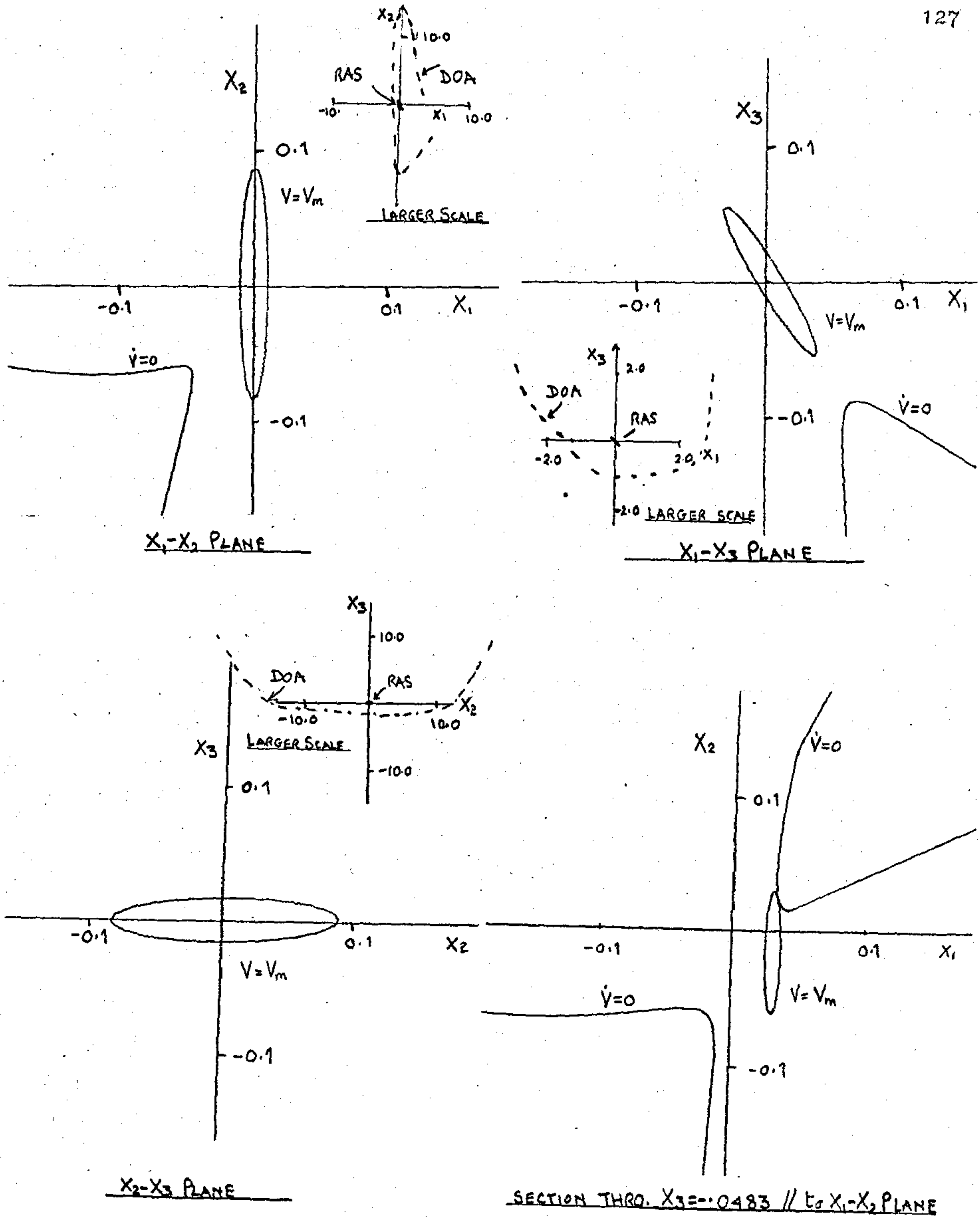
FIG. 5.3

SYSTEM S14



INITIAL AND FINAL QUADRATICS (V_i AND V_o)
3 VALID TAN. PTS.

FIG. 5.4



SYSTEM S17 (RAO (5))

FIG. 5.5

FLOW DIAG. FOR QUADRATIC RAS DETERMINATION

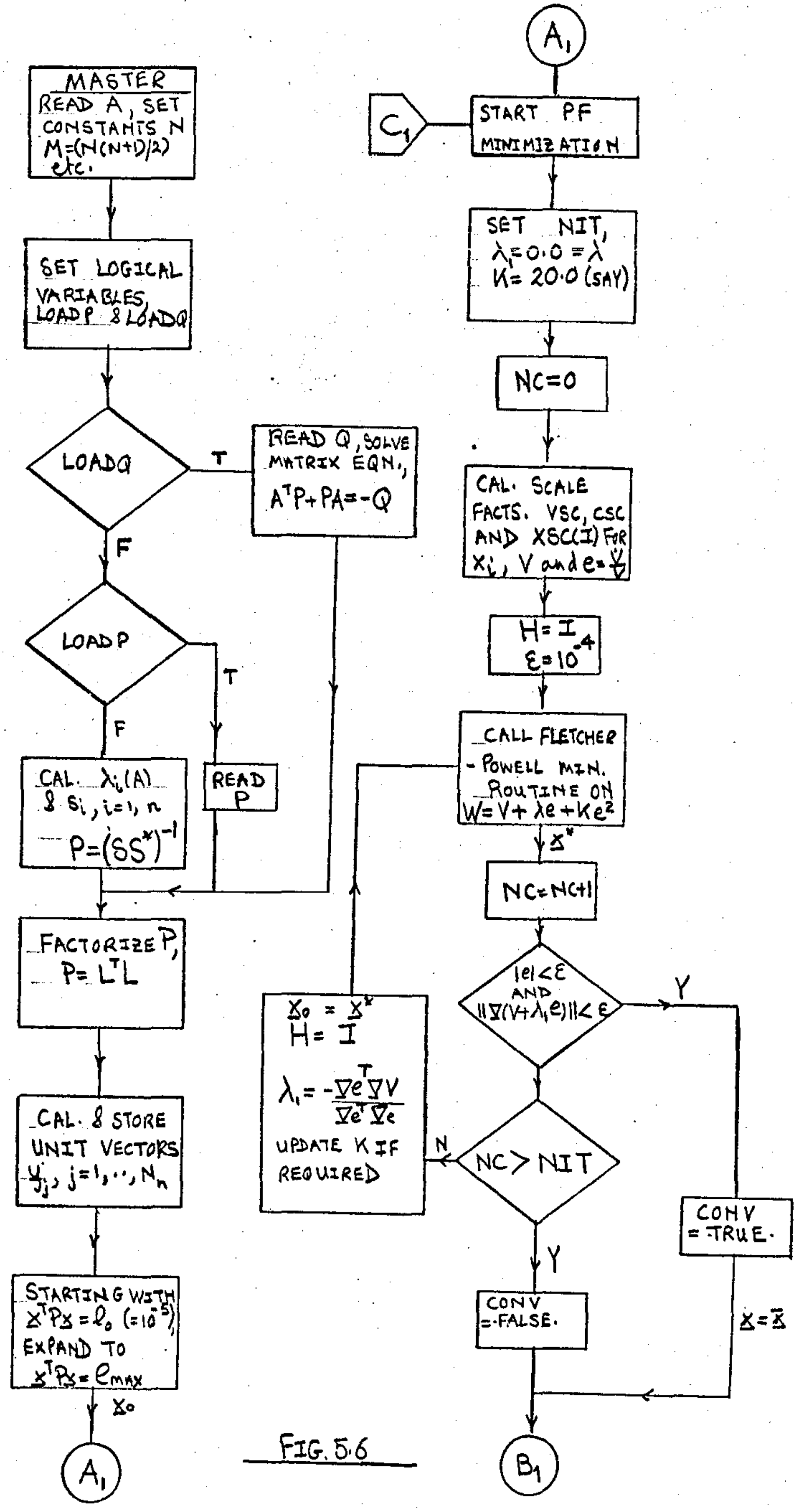


FIG. 5.6

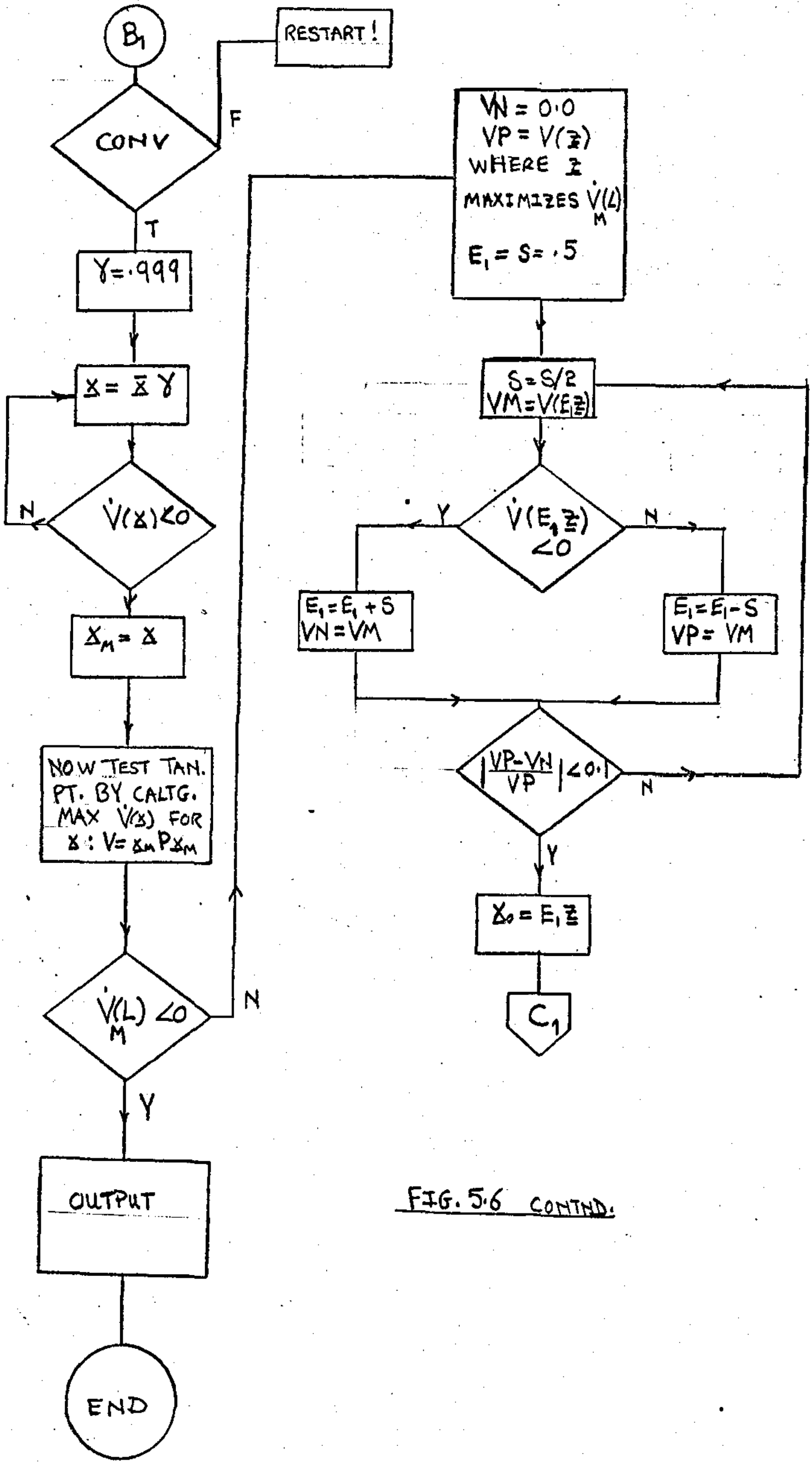


FIG. 5.6 CONTND.

CHAPTER 6

GENERAL OPTIMAL LYAPUNOV FUNCTIONS FOR NON-LINEAR SYSTEMS INCLUDING THOSE OF LURE' FORM AND RELAY CONTROL SYSTEMS.

Chapter 6

6.1 High Degree Lyapunov Functions for Autonomus Non-Linear Systems.

Introduction

A natural extension of a quadratic LF for the stable system of form 4.1.1 in Chapter 4, namely,

$$\dot{\underline{x}} = A\underline{x} + \underline{g}(\underline{x}) \qquad 6.1.1$$

is the LF of degree mv,

$$V = V_2 + V_3 + \dots + V_{mv} \qquad 6.1.2$$

where V_i is a homogeneous polynomial of degree i and V_2, $V_2 = \underline{x}^T P \underline{x}$, is itself a LF for the system. For second order systems we write

$$V = \sum_{i=2}^{mv} \sum_{j=1}^{i+1} v_{ij} x_1^{i-j+1} x_2^{j-1} \qquad 6.1.3$$

which involves to a multiplicative constant $m = 2 + (mv + 5)(mv - 2)/2$ parameters, v_{ij}, which is large for $mv \geq 5$ $(m \geq 17)$.

A less obvious class is the product of quadratics

$$V_{2mv} = \prod_{k=1}^{mv} \underline{x}^T P_k \underline{x} \qquad 6.1.4$$

where the P_k are p.d.s. matrices which solve

$$A^T P_k + P_k A = -Q_k \qquad (k = 1, \dots, mv)$$

for p.d.s. Q_k. This LF has the simplicity of the quadratic and only m = 2mv parameters need determining.

Two questions are apparent:

a) How do the optimal RAS'S for high degree LF's compare with those of the quadratic?

b) Are any 'equal tangency phenomena' present such as property A for optimal quadratics?

In order to give answers one must resort to numerical means as the RAS problem is intractable analytically. Henceforth we discuss some of the numerical features in determining the optimal LF's.

RAS Determination and Optimization Methods

The RAS problem in 4.1.10 and 4.1.11 of minimizing V on the constraint $\dot{V} = 0$ is more involved for a general LF of the form 6.1.3 and can cause difficulty with numerical methods. The latter can be attributed to the fact that (a) ∇V may vanish at a relative minimum and contact between the boundary $V = V_m$ and the constraint contour is not smooth, (b) the contours of V and $\dot{V} = 0$ may have several components and (c) V may not be p.d. in E^n. (Note (a), (b) and (c) are not unrelated and also that points satisfying $\nabla V(\underline{x}) = 0$ give tangency points)

Given the LF 6.1.3 a RAS was obtained for a second order system via Hewit's (2) version of Rodden's method (3). In essence the main features are:

(1) a search for the constraint contour by spiralling out from the origin until a point $\underline{x}_1$ is reached s.t. $\dot{V}(\underline{x}_1) > 0$, followed by an accurate location of the constraint at $\underline{x}_2$ say.

(2) A step is taken along a vector tangential to $\dot{V} = 0$ at $\underline{x}_2$ in a direction of decreasing V followed by a search along $\nabla\dot{V}$ to relocate the constraint. This is repeated until tangency, and a possible stability boundary is found.

(3) Steps along this tentative stability boundary, V = c, are made and $\dot{V}(\underline{x})$ is evaluated at each step. If $\dot{V}(\underline{x}) > 0$ the

previous tangency point is invalid, and having relocated the constraint, step (2) is repeated. Unless the method breaks down a conservative RAS is obtained of $V = c < V_m$ and its average radius is found.

To determine the optimal LF of 6.1.3 with respect to the numerical average radius ρ, the Nelder and Mead (46) simplex method was used. The optimization was over the m parameters

$$t_1, \ t_2 \text{ and } v_{ij}, \ i = 3, \ldots, mv$$
$$j = 1, \ldots, i+1$$

where t_1 and t_2 determine Q as $Q = L^T L + \epsilon I$, and P solves the matrix equation 4.1.5, the latter ensuring 6.1.3 is a LF for 6.1.1. In the numerical results which follow Rodden's method continually broke down for higher degree LF's due to difficulties (a) to (c) mentioned above and necessary features were incorporated:

a) an upper bound on points, $\underline{x}_i$ say, which search the constraint contour in step (2), $\|\underline{x}_i\| > R_{max}$,

b) a test for $V(\underline{x}_i) > V_{max}$ at tangency,

c) a test for a pt. $\underline{x}_2$ lying on $V = c$ (stability boundary), of $\|\underline{x}_2\| > R_{max}$ (i.e a test for an open boundary),

d) a test to discover whether step (2) of Rodden's method was repeatedly converging to the same tangency point.

If any of the bounds in (a) to (c) were satisfied the RAS function value of ρ for Nelder & Mead was penalized (i.e ρ = average radius = -10 say). In case (d) the best tangency point out of three was found.

The LF in 6.1.4 was investigated for the restricted class of systems of C4 such that, with minor modifications, the 2nd order optimal quadratic algorithm could be used. The optimization of ρ, the average radius, was over the 2mv values

$$t_{k1}, t_{k2} \qquad ; \qquad k = 1, mv$$

where $Q_k = L_k^T L_k + \epsilon I > 0$ and

$$L_k = \begin{bmatrix} 1 & t_{k1} \\ 0 & t_{k2} \end{bmatrix}$$

As for the quadratic, a boundary V = c is easily traced out for if $\underline{y}$ is any vector then $V(a\underline{y}) = c$ where

$$a = \left(c/\prod_{k=1}^{mv} \underline{y}^T P_k \underline{y}\right)^{\frac{1}{2mv}}$$

This fact also means that the complex algorithm of C5 can be extended to this LF to study high order systems, the average radius being easily obtained from the evaluation of $\dot{V}_M(L_k)$ in 5.2.15.

<u>Numerical Examples of Optimal LF's for $V_{mv} = \prod_{k=1}^{mv} \underline{x}^T P_k \underline{x}$</u>

In the examples that follow the average number of function evaluations of ρ by Powell's conjugate direction algorithm for the LF's of degree 2, 4 and 6 were respectively 70, 140 and 220, giving an accuracy of 10^{-2} or more for the optimum ρ (average radius). The latter was calculated from 80 points on the boundary $V = V_m$.

Table 6.1 shows a comparison of ρ for 4 systems taken from C4. Some marked improvement exists for systems S1, S2 and S7 where the DOA's are radially symmetric, but for S6 the difference between the optimal ρ for V_2 and V_6 is poor.

It was important in the optimization of ρ for V_4 and V_6 not to choose all initial $L_k(k = 1, 3)$ equal ($L_k = I$ say or $L_k = L_o$ where L_o gives the optimal quadratic.) This latter choice usually corresponded to a local minimum where the Powell optimization tended to stick due to the two valid tan. pts. of the OQ (Note the RAS's of $V^m(\underline{x})$ and $V(\underline{x})$ are the same).

Fig. 6.1 shows the relationship between the optimal boundaries of V_2, V_4 and V_6 and their constraint contours for system S1. For the 4^{th} degree LF there are 2 valid RS tangency points and 3 RS constraint components, whereas for the 6^{th} degree the numbers are 2 and 4 respectively. As for the quadratic case, these optimal LF's were non-unique. From numerical evidence it seems highly probable that at the exact optima for V_4 and V_6 there are 3 and 4 RS valid tangency points respectively. Interestingly, the non-convex shapes of the higher degree boundaries resemble those of the DOA more clearly.

Similar considerations apply to the corresponding boundaries for system S7, shown in Fig. 6.2, with an equal tangency property holding for V_4 and V_6 with two RS valid tangency points and 3 and 5 RS constraint components respectively. (There is almost a third point of contact of $V_6 = V_m$ with $\dot{V}_6 = 0$). All the optimal boundaries pass through the singular point (1,2) of system S7.

TABLE 6.1

SYSTEM	ρ = AVERAGE RADIUS		
	m = 2	m = 4	m = 6
	V_2	V_4	V_6
S1 (ZUBOV)	1.883	2.219	2.292
S2 (VAN.POL)	1.44	1.51	1.57
S6 (DAVIES)	0.1857	0.1873	0.1874
S7	2.52	2.75	3.21

Numerical Examples of Optimal LF's For V in Series Form (6.1.3).

An obvious initial choice for the coefficients v_{ij} in 6.1.3 for $mv > 2$ is

a) for V_{mv} select as initial values for

$$v_{ij}, \quad i = 2, \ldots, mv-1$$
$$j = 1, \ldots, i+1$$

those of the optimal LF of degree mv-1 and

b) put $v_{mvj} = 0.0$, $j = 1, \ldots, mv+1$

or c) put $v_{mvj} \neq 0.0$, " " "

The choice of a) and b) was unsatisfactory since the optimization started on an equal tangency surface and convergence was poor. Although the other choice is better - a) and c), arbitrarily selecting all coefficients of the same magnitude proved as good.

For convenience we write the LF in 6.1.3 as $V_{2,3,..,mv}$, then Figures 6.3 to 6.6 show the behaviour of the optimal LF's with respect to the following systems:

System S2 (Van der Pol)

Since the DOA for the system is radially symmetric it was reasonable to choose the LF accordingly and so the forms V_2, $V_{2,4}$ and $V_{2,4,6}$ were investigated for which m = 2, 7 and 14 respectively. The Simplex optimization was terminated when either the difference between function values of the simplex were less than 10^{-3} in magnitude or when I, the number of function evaluations exceeded an upper bound I_{MAX}, the latter being 50, 190 and 290 for the respective cases.

For the LF $V_{2,4}$ convergence was obtained in 184 function evaluations (FE's) but not without many of the breakdowns of Rodden ((a) to (d)), this being the case for the following systems as well. Valid tangency alternated between 3 distinct RS points near the optimum and is reflected in Fig. 6.3 where the optimal boundary has 3 RS valid tangency points in contact with its constraint contour, the latter having 2 RS components which are unlike those of the quadratic.

For the LF $V_{2,4,6}$ convergence was very poor in that ρ tended to stick to a value near the optimum, which was in part due to the high dimension of the space of parameters a_{ij} (m = 14) for which the Simplex method becomes inefficient. However, after several restarts an accurate optimum was obtained and Fig. 6.3 shows that the optimal boundary, on which 3 valid RS tangency points lie, gives a good RAS both in shape and size. The corresponding table shows the large computing time for $V_{2,4,6}$.

System S18

Consider the system

$$\dot{x}_1 = x_2 \qquad \text{S18}$$

$$\dot{x}_2 = -x_1(1 - x_1^2) - x_2(1 - x_2^2)$$

with singular points (0,0) and $\mp(1,0)$.

Investigating the same LF's shows the situation is somewhat different, since for all LF's, only one RS constraint component exists near the origin and the respective optimal boundaries all have two RS points of contact with the latter (Fig. 6.4). As for the previous system the average radius of the 6^{th} degree LF is only fractionally better than that of the 4^{th} and the RAS's are poor compared to the DOA.

System S19

A system with a non-symmetric DOA is

$$\dot{x}_1 = x_2 \qquad \text{S19}$$

$$\dot{x}_2 = -x_1(1 + x_2) - x_2(1 - x_2^2)$$

In this case the LF's V_2, $V_{2,3}$ and $V_{2,3,4}$ were investigated. Fig. 6.5 shows the optimal boundaries of the 3^{rd} and 4^{th} degree curves having 3 points of contact with their constraint contours (the non-optimal ones having one or two). Their respective RAS's have the usual failing for systems with open DOA's, being very poor .

System S13 (Yu (6))

The system is that in 5.3 and caused some difficulty in convergence of the average radius to the optimum for $V_{2,3,4}$. This is shown in Fig. 6.6 where although the optimal boundary for $V_{2,3}$ has almost 3 points of contact with its constraint contour, that of $V_{2,3,4}$ has one. Also, due to continued breakdown of the Rodden method (mainly tangency points giving open boundaries), the latter required 6000 mill/sec for 59 FE's after restarting near the optimum. However, some marked improvement is shown in the RAS.

Comment

It is evident that high degree optimal LF's do in some cases give much improved RAS's (systems S1, S7, S2, S19), but the equal tangency phenomenon which is also present and contributes to lack of convergence, thereby making computing time excessive, offsets this advantage.

All systems studied had two or more valid tangency points for the optimal LF's of degree greater than two. The direct relationship between the number of tangency points and valid tangency points is complicated by the multiple components of both the V_{mv} = const. and $\dot{V}_{mv} = 0$ contours. Depending on the system, the number of tangency points, valid tangency points and constraint components tend to increase with mv.

6.2 Optimum Lyapunov Functions For Relay Control Systems

Introduction

The application of Lyapunov's direct method to differential equations with discontinuous right hand sides has been studied in particular by Alimov (69), Ansonov (68) and Weissenberger (48,66). In this section we shall extend some work of the latter and consider the relay system

$$\dot{\underline{x}} = A\underline{x} + \underline{b}\,\mathrm{sgn}\,\sigma, \qquad \sigma = \underline{d}^T\underline{x} \qquad 6.2.1$$

where

$$\mathrm{sgn}\,\sigma = \begin{cases} +1, & \sigma > 0 \\ \epsilon, & \sigma = 0, \quad -1 \leq \epsilon \leq 1 \\ -1, & \sigma < 0 \end{cases} \qquad 6.2.2$$

We assume, as for most practical systems, that

$$\underline{d}^T\underline{b} < 0 \qquad 6.2.3$$

which implies two types of motion: regular switching in which trajectories in $\sigma > 0$ connect on $\sigma = 0$ with trajectories in $\sigma < 0$, and sliding, in which segments of the trajectories lie in the space $\sigma(\underline{x}) = 0$. A complete discussion of these motions is given in reference (48) and we mention that the main definitions and theorems - including Theorem 1.5.3 - of Lyapunov's direct method are applicable to relay systems provided the stability of the two motions are treated separately.

For regular switching the motion is described by 6.2.1, while for sliding, which occurs for $|\underline{d}^T A\underline{x}| < |\underline{d}^T\underline{b}|$ on $\underline{d}^T\underline{x} = 0$, by the linear system

$$\dot{\underline{x}} = \bar{A}\underline{x} \qquad 6.2.4$$

where

$$\bar{A} = (I - \frac{\underline{b}\underline{d}^T}{\underline{d}^T\underline{b}})A \qquad 6.2.5$$

By considering an orthogonal transformation from $\underline{x}$ to $\underline{y}$ via

$$\underline{y} = G\underline{x} \qquad (G^T = G^{-1}) \qquad 6.2.6$$

such that y_n is normal to the sliding plane, $\sigma = 0$ i.e

$$(0, 0, \ldots, 0, 1) = G\underline{d}/\ \|\underline{d}\|$$

the system 6.2.4 may be written

$$\dot{\underline{y}} = G\bar{A}G^T\underline{y} \qquad 6.2.7$$

Further, since $\dot{y}_n = 0$, by defining a new vector $\underline{z}$ as $\underline{z} = (y_1, y_2, \ldots, y_{n-1})$, we obtain

$$\dot{\underline{z}} = A'\underline{z} \qquad 6.2.8$$

where A' is the $(n-1) \times (n-1)$ matrix obtained by deleting the last row and column of $G\bar{A}G^T$.

Two candidates for a LF for 6.2.1 are

a) the <u>piecewise quadratic</u> LF

$$V = \underline{x}^T P\underline{x} + \int_0^{\sigma} \mathrm{sgn}\ s\ ds \qquad 6.2.9$$

$$= \underline{x}^T P\underline{x} + |\underline{d}^T\underline{x}|$$

for which $\dot{V} = \underline{x}^T Q\underline{x} + \underline{K}^T\underline{x}\ \mathrm{sgn}\sigma + \underline{d}^T\underline{b}$ $\qquad$ 6.2.10

where $\quad Q = A^TP + PA$

and $\quad \underline{K} = A^T\underline{d} + 2P\underline{b}$

(By 6.2.3 $V > 0$ and $\dot{V} < 0$ for some $R(h)$, h small, $\sigma \neq 0$). Here Weissenberger has shown that V is a LF for 6.2.1 iff the symmetric P has the property

$$A'^T P' + P'A' = -Q' < 0 \qquad 6.2.11$$

where A' is from 6.2.8 and P' is formed by deleting the last row and column of GPG^T. Here, V may still be a LF even though P is indefinite and/or A is unstable.

b) the piecewise linear LF

$$V = \sum_{i=1}^{m} |\underline{c}_i^T \underline{x}| \qquad 6.2.12$$

originally used by Rosenbrock (70), where for 6.2.1 $\underline{c}_i = \underline{d}$ for some i and, to ensure V is p.d. the vectors $\underline{c}_i$ span E^n. The contours of V = constant are composed of at most 2^m hyperplanes whose normals satisfy

$$\underline{n}_j = \sum_{i=1}^{m} \mp \underline{c}_i \,, \quad j = 2, \ldots , 2^m$$

Then since 6.2.1 is linear for $\sigma > 0$, by Rosenbrock's analysis, if

$$\dot{\underline{x}}^T \underline{n}_i \leq 0 \qquad 6.2.13$$

at each vertex of the face whose normal is $\underline{n}_i$, for all faces of V = c, then V < c is an estimate of the RAS provided 6.2.13 also holds for sliding.

The Piecewise Quadratic LF

An RAS determination for the LF 6.2.9 is essentially the same as in 6.1 of minimizing V on the constraint contour $\dot{V} = 0$ from 6.2.10, where in the plane $\sigma = 0$ the latter is discontinuous and given by

$$\dot{V} = \underline{x}^T Q \underline{x} \mp \underline{Kx} + \underline{d}^T \underline{b} = 0$$

Weissenberger (48) has given a second order example where he optimizes the area of the RAS for LF 6.2.9. He produces an optimal boundary having 2 RS points of contact with its constraint, but goes no further. Here we explore the situation more fully.

Since points on $\dot{V} = 0$ are easily determined by solving a quadratic, the second order algorithm of Chapter 4 is again applicable. The major modification in RAS determination is the replacement of the one-dimensional search routine using, continuity for quadratic fits, by the method of search by Golden Section (56). This is due to the two types of tangency which occur:

a) a smooth tangency at which V and $\dot{V} = 0$ are continuous and

b) a corner tangency lying on $\sigma = 0$ where the contours of $\dot{V} = 0$ may 'jump' discontinuously.

The optimization of the area of the RAS was made over the space of the elements of P (p_{11}, p_{12}, p_{22}) subject to 6.2.11.

Numerical Examples

System S20

A relay system having a limit cycle is

$$\dot{x}_1 = x_2 \qquad \text{S20}$$

$$\dot{x}_2 = - x_1 + x_2 - \operatorname{sgn}(x_1 + x_2)$$

Optimization of the RAS area by Powell's method showed a marked oscillatory effect of the valid tangency point between a corner and a smooth tangency. Fig. 6.7 shows an initial LF boundary

$$2x_1^2 - x_1x_2 + x_2^2 + |x_1 + x_2| = 1.705$$

(with $\rho = 2.094$) with one smooth RS valid tangency and the best LF boundary

$$1.589x_1^2 - x_1x_2 + x_2^2 + |x_1 + x_2| = 1.953$$

with a corner and a smooth RS valid tangency point, ($\rho = 2.73$). In this case the constraint contours are closed.

That one equal tangency surface exists, similar to the curve in the OQ case (C4), is seen in Fig. 6.8a and 6.8b where the actual area contours have been plotted through sections of the 3 dim. space of elements of P parallel to the $p_{11} - p_{22}$ plane for $p_{12} = 0.0$ and $p_{12} = -.5$ respectively (Shown AB). Values of P giving smooth tangency lie above surface AB, the optimal P lying on the section of AB in Fig. 6.8b.

System S21

An example showing a marked discontinuity of $\dot{V}$ on $\sigma = 0$ is

$$\dot{x}_1 = x_2 \qquad \text{S21}$$

$$\dot{x}_2 = -2x_1 + x_2 - 2\,\mathrm{sgn}(2x_1 + x_2)$$

Fig. 6.9 shows the initial LF used in the previous example with one valid RS corner tangency point, and again the optimal LF with the two types of valid tangency. The RAS boundary is

$$3.544x_1^2 - 0.8707x_1x_2 + 1.0x_2^2 + |2x_1 + x_2| = 3.9941$$

$$(\rho = 4.17)$$

Convergence was degraded once Powell's method hit the equal - tangency surface and various initial P's were tried to get the optimum.

System S22

A case where the optimum piecewise quadratic LF approaches a piecewise linear LF is the following

$$\dot{x}_1 = x_2 \qquad \text{S22}$$

$$\dot{x}_2 = x_1 - \operatorname{sgn}(x_1 + 2x_2)$$

Fig. 6.10 shows an initial LF (P = I) compared to the optimum in which the $\dot{V} = 0$ constraint forms part of the RAS boundary, $V = V_m$. The latter caused difficulty since the precise tangency at the optimum was not clear. (A check was made to see if $\underline{x}_1$ was the valid tangency where $\nabla V(\underline{x}_1)=0$).

Comment

Weissenberger uses a gradient method to optimize the RAS volume for 3rd order systems. The complex method of C5 could also be applied, with better convergence properties.

The examples given here certainly show that the equal tangency property of optimal LF's is not restricted to continuous systems.

The Piecewise Linear LF

Weissenberger (66) gives an analytic expression for the optimal piecewise quadratic LF of 6.2.9 for the system

$$\dot{\underline{x}} = \begin{bmatrix} 0 & 1 \\ p_1 & p_2 \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} \operatorname{sgn} \sigma \qquad 6.2.14$$

$$\sigma = d_1 x_1 + d_2 x_2, \; p_1 > 0$$

We show here that a good RAS boundary for the piecewise linear LF of 6.2.12 can be obtained by assuming an 'equal tangency property' for the optimum LF.

Consider the LF

$$V = |\underline{c}^T \underline{x}| + |\underline{d}^T \underline{x}| \qquad 6.2.15$$

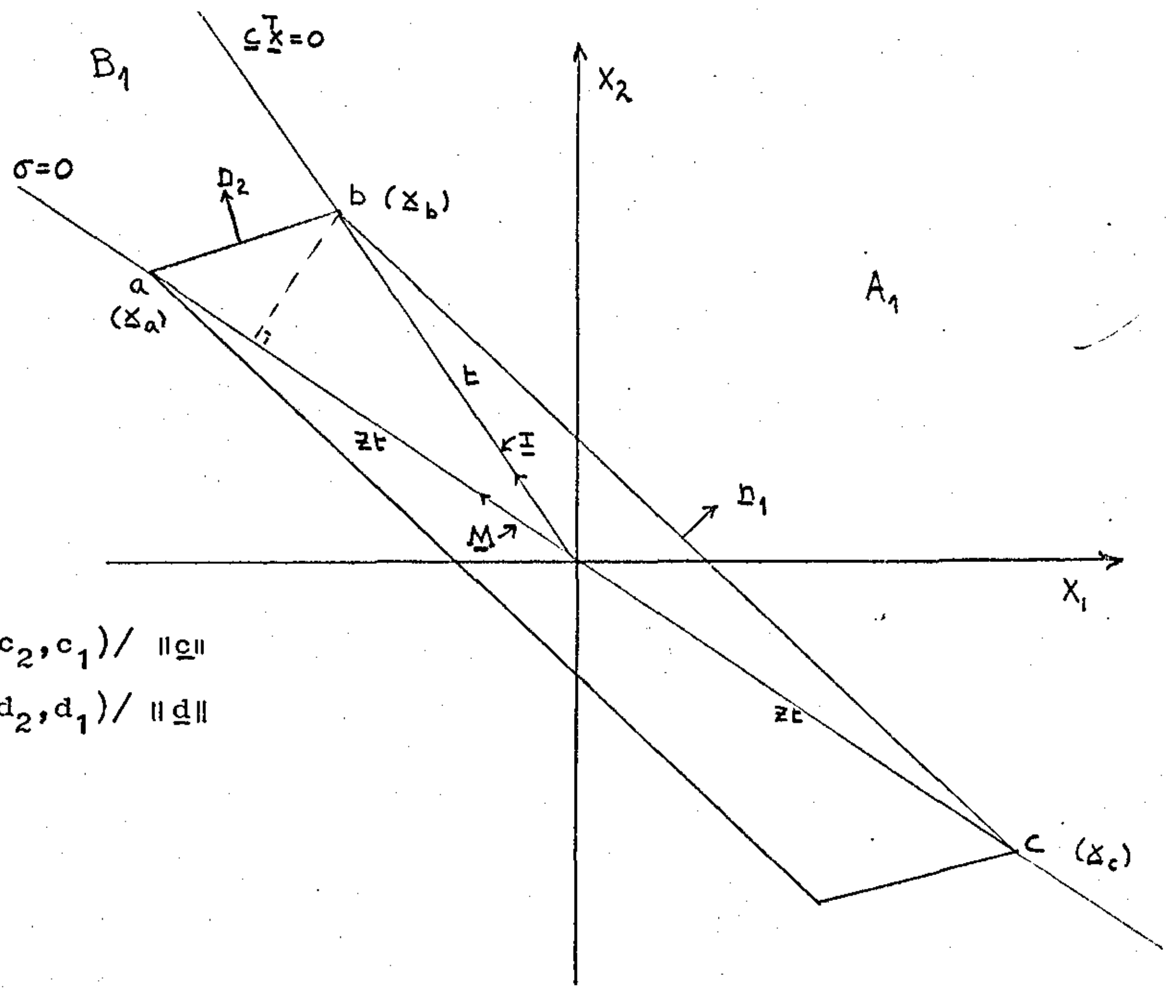
and assume sliding for 6.2.14 is a.s., i.e.

$$d_1,\ d_2 > 0$$

Then without loss of generality the situation is shown in Diag. 6.1 with a contour, $V = \alpha$ = const., having vertices a, b and c of interest in $\sigma > 0$ (by symmetry the analysis is considered for $\sigma > 0$ only) and normals

$$\underline{n}_1 = \underline{d} + \underline{c}$$
$$\underline{n}_2 = \underline{d} - \underline{c}$$

DIAG 6.1



$\underline{I} = (-c_2, c_1)/\ \|\underline{c}\|$

$\underline{M} = (-d_2, d_1)/\ \|\underline{d}\|$

For $V = \alpha$ to be a stability boundary we require

$$\underline{n}_i^T \dot{\underline{x}} \leq 0 \qquad\qquad i = 1, 2\ .$$

at the vertices a, b and c, or,

$$\left.\begin{array}{l} \underline{n}_2^{T}\dot{x}_a \leq 0 \\ \underline{n}_2^{T}\dot{x}_b \leq 0 \\ \underline{n}_1^{T}\dot{x}_b \leq 0 \\ \underline{n}_1^{T}\dot{x}_c \leq 0 \end{array}\right\} \qquad 6.2.16$$

and also for sliding, which on $\sigma = 0$ gives

$$\dot{V} = \underline{n}_2^{T}\dot{x} = -\underline{c}^{T}\dot{x} = -\frac{x_2}{d_2}(d_2c_1 - d_1c_2) \qquad \text{for } x_2 > 0$$

and $\dot{V} = \underline{n}_1^{T}\dot{x} = -\underline{n}_2^{T}\dot{x}$ for $x_2 < 0$.

Hence we require

$$d_2c_1 - d_1c_2 > 0 \qquad 6.2.17$$

which also implies V p.d. For local stability we also require $\underline{n}_i^T\underline{b} < 0$ which gives

$$c_2 + d_2 > 0 \quad \text{and} \quad d_2 - c_2 > 0 \qquad 6.2.18$$

Let $\underline{M}$ and $\underline{I}$ be unit vectors to vertices a and b and zt and t the respective distances
($\underline{x}_a = zt\underline{M}$, $\underline{x}_b = t\underline{I}$, $\underline{x}_c = -\underline{x}_a$) then

$$t = \frac{\|\underline{c}\| \; \alpha}{(d_2c_1 - c_2d_1)} \qquad 6.2.19$$

and the area of $V < \alpha$ is $\rho = t\alpha/\|\underline{c}\|$ or

$$\rho = \left(\frac{t}{\|\underline{c}\|}\right)^2 (d_2c_1 - d_1c_2) \qquad 6.2.19b$$

The RAS is determined by the maximum α satisfying 6.2.16, or equivalently, by the minimum of the t (>0) giving equality in all 4 relations of 6.2.16. The 4 values of t are respectively

$$\frac{-\underline{n}_2^T\underline{b}}{z\underline{n}_2^T A\underline{M}}, \quad \frac{-\underline{n}_2^T\underline{b}}{\underline{n}_2^T A\underline{I}}, \quad \frac{-\underline{n}_1^T\underline{b}}{\underline{n}_1^T A\underline{I}}, \quad \& \quad \frac{\underline{n}_1^T\underline{b}}{z\underline{n}_1^T A\underline{M}} \qquad 6.2.20$$

where $z = \|\underline{d}\| / \|\underline{c}\|$.

As the boundary $V = \alpha$ is expanded its vertices will have a first contact with the $\dot{V} = 0$ constraint - in this case two RS straight lines found in regions A_1 and B_1. (Diag. 6.1). In view of the equal tangency properties which usually hold for optimum LF's we assume that the optimum boundary in this case will touch $\dot{V} = 0$ at all vertices (and thus $\dot{V} = 0$ on all sides of the boundary). For $\dot{V} = 0$ along bc we require the last two terms in 6.2.20 to be equal i.e

$$\frac{-\underline{n}_1^T\underline{b}}{\underline{n}_1^T A\underline{I}} = \frac{\underline{n}_1^T\underline{b}}{z\underline{n}_1^T A\underline{M}} \qquad 6.2.21$$

which gives

$$(d_1 + c_1)^2 - p_1(d_2 + c_2)^2 + p_2(d_1 + c_1)(d_2 + c_2) = 0$$

or

$$u^2 + p_2 u - p_1 = 0 \qquad 6.2.22$$

with $u = (c_1 + d_1)/(c_2 + d_2)$.

Let $r = p_2/2 - \sqrt{p_2^2/4 + p_1} < 0$ be the negative ~~of one~~ root of 6.2.22, then c_1 and c_2 must satisfy

$$c_1 + rc_2 = -(rd_2 + d_1) \qquad 6.2.23$$

in order that 6.2.21 holds (the other root gives $c_1, c_2 < 0$). The RAS boundary for such c_1, c_2 is determined as

$$V = \left(\frac{t}{\|\underline{c}\|}\right)(d_2 c_1 - c_2 d_1)$$

from 6.2.19, where, from the substitution of 6.2.23 into 6.2.21, we have

$$\frac{t}{\|\underline{c}\|} = 1/(p_1 d_2 - p_2 d_1 + d_1 r) \qquad 6.2.24$$

Using the fact that

$$c_1 d_2 - c_2 d_1 = -(d_2 + c_2)(rd_2 + d_1)$$

and

$$\frac{p_1}{r}(d_1 + rd_2) = p_1 d_2 - p_2 d_1 + d_1 r$$

from 6.2.23, substitution of 6.2.24 into 6.2.19b gives the area in terms of c_2 as

$$\rho = -\frac{r}{p_1}(c_2 + d_2) \qquad 6.2.25$$

The only restriction on c_2 is from 6.2.18 and ρ is maximized when $c_2 = d_2$ giving the boundary

$$|d_2 x_2 - (d_1 + 2rd_2)x_1| + |d_1 x_1 + d_2 x_2| = -\frac{2rd_2}{p_1} = \rho$$

(Note the first two terms in 6.2.20 are both zero, and $\dot{V} = 0$ on ab) Fig. 6.10 shows this boundary for system S22, and is the optimum (Weissenberger (66)).

6.3 Finite Regions of Attraction For The Problem of Lure'

Introduction

We shall be concerned with the system of Lure' form

$$\dot{\underline{x}} = A\underline{x} + \underline{b}f(\sigma) \qquad 6.3.1$$

$$\sigma = \underline{c}^T\underline{x}, \text{ A stable,}$$

where $f(\sigma)$ is a continuous non-linear function which may leave the sector

$$\left.\begin{array}{c} 0 < f(\sigma)/\sigma < K, \ \sigma \neq 0 \\ f(0) = 0 \end{array}\right\} \qquad 6.3.2$$

at some $\sigma \leq \sigma_1 < 0$ and/or $\sigma \geq \sigma_2 > 0$, but satisfies 6.3.2 for $\sigma_2 \leq \sigma \leq \sigma_1$. The latter restrictions often occur in practical systems where $f(\sigma)$ is unknown for large σ. Even with these restrictions a region of attraction exists if the Popov condition

$$\mathrm{Re}(1 + i\omega q)G(i\omega) + \frac{1}{K} > 0, \qquad \forall\,\omega \geq 0 \tag{6.3.3}$$

holds for some real q where

$$G(s) = \underline{c}^T(A - sI)^{-1}\underline{b} \tag{6.3.4}$$

and A is stable (Aizerman and Gantmacher (71)). The standard LF associated with this problem is

$$V = \underline{x}^T P \underline{x} + q\int_0^{\sigma} f(s)ds \tag{6.3.5}$$

with P p.d. If such a LF can be found which proves g.a.s. for 6.3.2 without restriction, then with restriction, $V < V_m$ will be a region of attraction if it lies in $\sigma_2 \leq \sigma \leq \sigma_1$. The largest boundary $V = V_m$ will be tangential to $\sigma = \sigma_1$ or $\sigma = \sigma_2$ and thus

$$V = V_m = \min_i M_i \tag{6.3.6}$$

where

$$M_i = \sigma_i^2/\underline{c}^T P^{-1}\underline{c} + q\int_0^{\sigma_i} f(s)ds \tag{6.3.7}$$

An Optimal Quadratic For The Infinite Sector

For q = 0 and $K = \infty$ a quadratic $V = \underline{x}^T P \underline{x}$ (if it exists) giving $\dot{V} \leq 0$ without restriction on f in 6.3.2 is given by

$$P\underline{b} = -\underline{c}/2 \tag{6.3.8}$$

and

$$A^T P + PA = -Q \leq 0 \tag{6.3.9}$$

(Then $\dot{V} = -\underline{x}^T Q \underline{x} - \sigma f(\sigma) \lessapprox 0$.) Weissenberger (72) has proposed the optimization of the volume of region 6.3.6,

$$\rho = \max_{P \in \bar{P}} \pi \left[\frac{V_m^n}{d(P)} \right]^{\frac{1}{2}} \qquad 6.3.10$$

with $\bar{P}$: (P/ 6.3.8 and 6.3.9 satisfied), to obtain the optimal quadratic when the Popov condition holds, ensuring $\bar{P}$ is non-empty. Walker and McClamroch (73), however, suggest a single LF obtained via the Kalman construction procedure (67). For the general case ($q \neq 0$) the assumptions made are (1) the pairs $(A,\underline{b})$ and $(A^T,\underline{c})$ are completely controllable (see (43)) and (2) the Popov condition holds for some real q with $d(qA + I) \neq 0$. Then a LF of form 6.3.5 exists where P solves

$$A^T P + PA = -\underline{u}\,\underline{u}^T \qquad 6.3.11$$

and $\underline{u}$ is a real vector determined by writing

$$\mathrm{Re}\,(1 + i\omega q)G(i\omega) + \frac{1}{K} = \left| \frac{\theta(i\omega)}{d(i\omega I - A)} \right|^2 \qquad 6.3.12$$

and setting

$$\underline{u}^T (sI - A)^{-1} \underline{b} = \sqrt{z} - \theta(s)/d(sI - A) \qquad 6.3.13$$

(θ is a real polynomial of degree n and $z = 1/K - q\underline{c}^T\underline{b}$. Kalman chooses roots of $\theta(\omega)$ to have neg. real parts).

The following relation now holds between the two approaches above:

<u>Theorem 6.1</u>

For n = 2 the maximizing P in 6.3.10 is determined via the Kalman construction procedure i.e. 6.3.11

With a general A, $\underline{b}$ and $\underline{c}$ the proof is long and tedious. However, since it is assumed the pair (A, $\underline{b}$) is completely controllable a transformation exists (see Lefshetz (43)) transforming 6.3.1 to cononical form without changing the problem. Hence without loss of generality we assume the system

$$\dot{\underline{x}} = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} f(\sigma) \qquad 6.3.14$$

Here

$$G(s) = (c_2 s + c_1)/(s^2 + as + b)$$

and the relation for θ in 6.3.12 reduces to

$$r + \omega^2 t = \theta(i\omega)\theta(-i\omega)$$

with $r = c_1 b > 0$ and $t = ac_2 - c_1 > 0$. The two choices of θ are $\theta(i\omega) = \sqrt{r} + i\omega\sqrt{t}$ and $\theta(i\omega) = \sqrt{r} - i\omega\sqrt{t}$. With the former, $\underline{u} = -(\sqrt{r}, \sqrt{t})$ giving

$$\underline{u}\underline{u}^T = \begin{bmatrix} r & \sqrt{rt} \\ \cdot & t \end{bmatrix} \qquad 6.3.15$$

It is easy to show that the Q corresponding to the optimal P in 6.3.10 is just this. For 6.3.8 implies P of the form

$$P = \begin{bmatrix} \alpha & c_1/2 \\ \cdot & c_2/2 \end{bmatrix} \qquad \text{with } d(P) = (2\alpha c_2 - c_1^2)/4$$

and

$$Q = \begin{bmatrix} r & -(\alpha - \tfrac{1}{2}(ac_1 + bc_2)) \\ \cdot & t \end{bmatrix}$$

Hence we require to minimize α subject to Q p.d. or

$$-\sqrt{rt} < \alpha - \tfrac{1}{2}(ac_1 + bc_2) < \sqrt{rt}$$

Hence the optimal Q is that in 6.3.15.

Interestingly, the P minimizing the area in 6.3.10 corresponds to the other choice of θ.

The result was also shown valid for specific 3rd order examples and is conjectured to hold for general n.

Variation of ρ with q ($K = \infty$)

The system 6.3.1 will be globally asymptotically stable if the modified frequency response (X vs. Y) with

$$G(i\omega) = X(\omega) + iY(\omega)/\omega$$

lies to the right of the Popov line

$$X - qY + \frac{1}{K} = 0$$

for some q in the X - Y plane. A range of such q, $q_2 < q < q_1$ will usually exist and the question arises of whether a search for the q giving the best stability region in 6.3.6 is profitable, with V obtained via the Kalman construction procedure. We consider 4 second order examples for the infinite sector which give different types of frequency response and range of q.

For system 6.3.14 and LF 6.3.5 the modified frequency response is a convex curve and the Popov lines tangential to it give the range of q. Let $\alpha = ac_2 - c_1$ and $\beta = c_2b - c_1a$ then the tangency values give

$$q_1 = \frac{1}{\beta^2}(\alpha\beta + 2bc_1c_2) + \sqrt{4c_1c_2b + (\alpha\beta + c_1c_2b)}$$

and

$$q_2 = \quad " \quad - \quad "$$

The following ranges then hold

Case 1: $\alpha \leq 0, \ \beta > 0 \ ; \quad q_2 < q < q_1$

Case 2: $\alpha \geq 0, \ \beta > 0 \ ; \quad 0 \leq q < q_1$

Case 3: $\alpha \leq 0, \ \beta < 0 \ ; \quad q_2 < q < \infty$

Case 4: $\alpha > 0, \ \beta < 0 \ ; \quad 0 \leq q < \infty$

Fig. 6.11 to 6.14 show examples of the 4 cases for $f = \sigma - \sigma^3$ $(-1 < \sigma < 1)$. Evidently, marked increases in the size of stability regions are found for some q and in only one case (Case 3) is the optimum q near one of the extreme points $q = q_1$ or $q = q_2$ (i.e $q = 10^4$ in Fig. 6.13).

SYSTEM S1 (ZUBOV)
OPTIMAL LF'S FOR

$$V_{2mv} = \prod_{k=1}^{mv} \underline{x}^T P_{2k}$$

LF	NO. TAN. PTS.
V_2	2
V_4	2 (ALMOST 3)
V_6	2 (" 4)

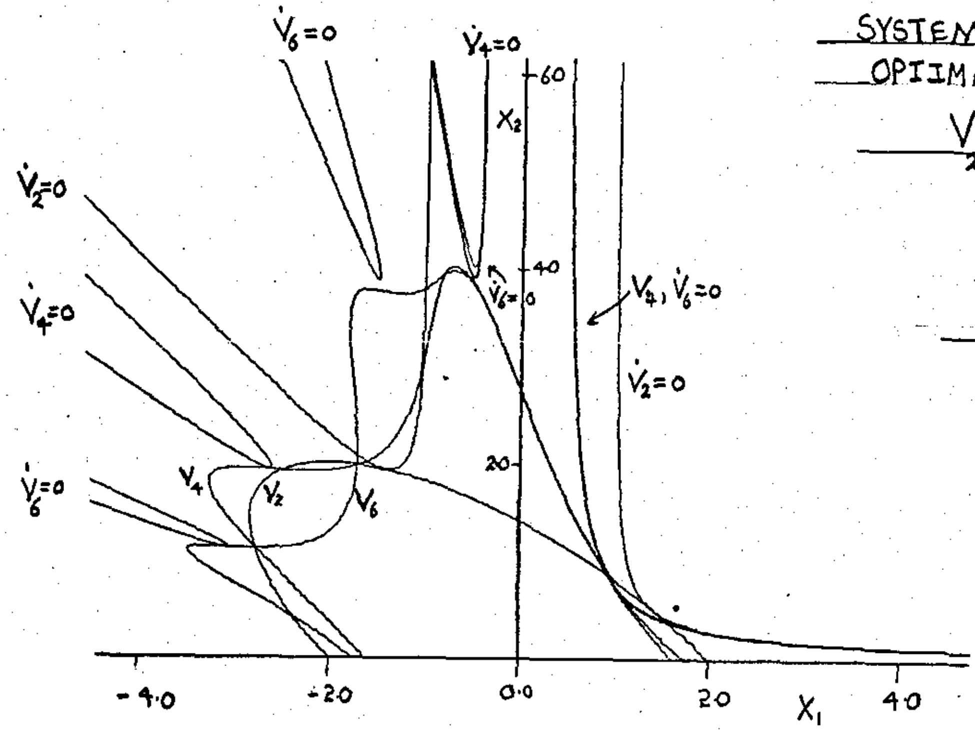


FIG. 6.1

SYSTEM S7

LF	NO. TAN. PTS.
V_2	2
V_4	2 (ALMOST 3)
V_6	2 (ALMOST 3)

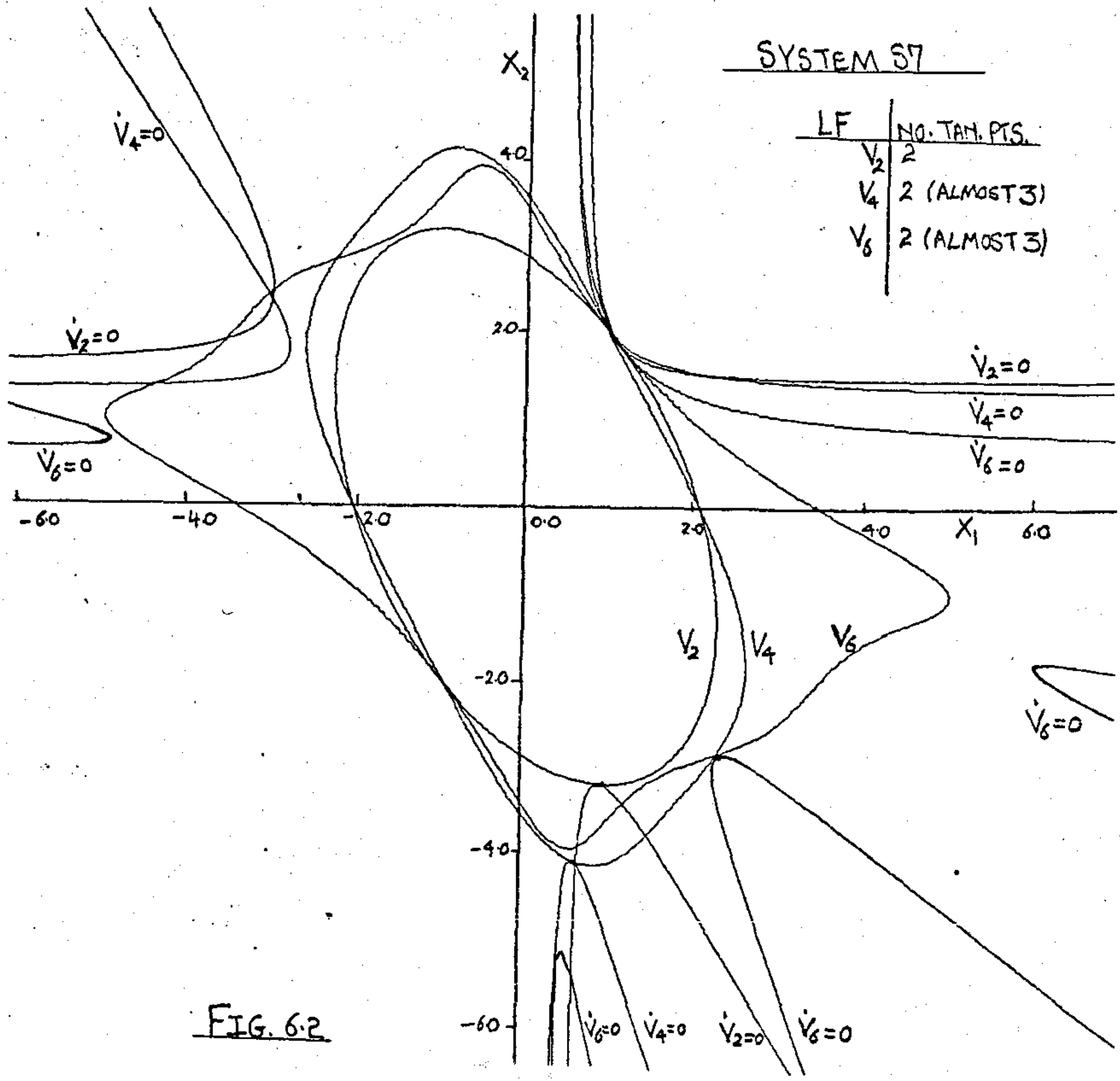
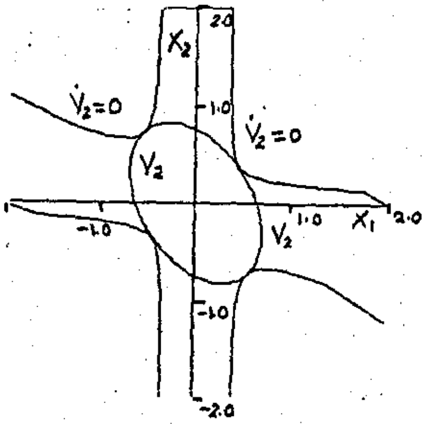
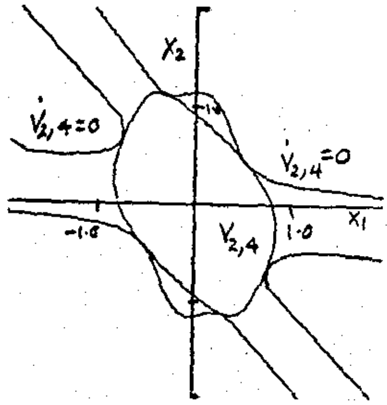


FIG. 6.2

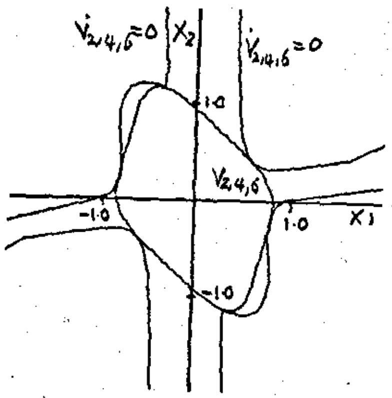


V = V2
2 VALID RS TAN. PTS.



V= V2 + V4
3 VALID RS TAN. PTS.

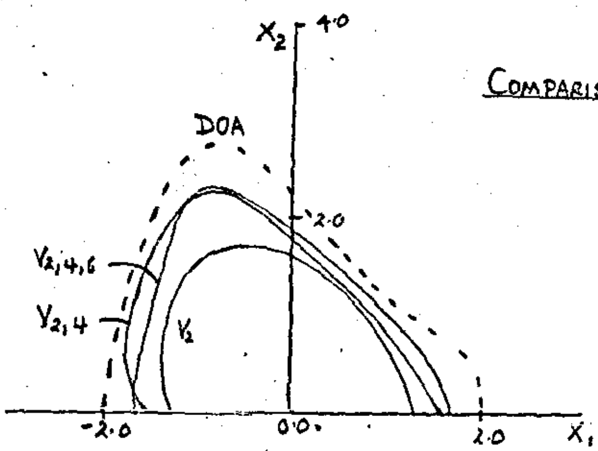
OPTIMAL CONTOURS FOR
SERIES LF (6.1.3)
SYSTEM S2 (VAN DER POL, ε=1.0)



V= V2 + V4 + V6
3 VALID RS TAN. PTS.

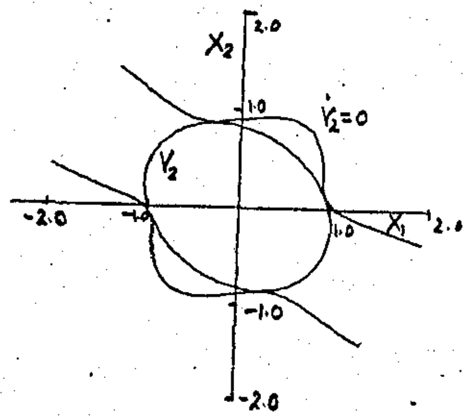
V	ρ AV. RAD	NO. OF FE'S	TIME (MILL/SEC)
V_2	1.44	70	760
$V_{2,4}$	1.798	184	883
$V_{2,4,6}$	1.805	290*	3050

* RESTARTED



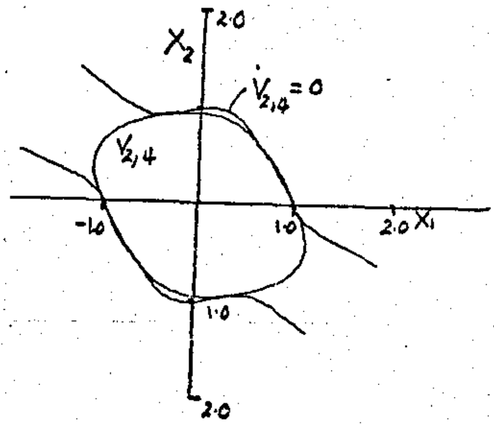
COMPARISON OF RAG'S WITH DOA

FIG. 6.3



$V = V_2$

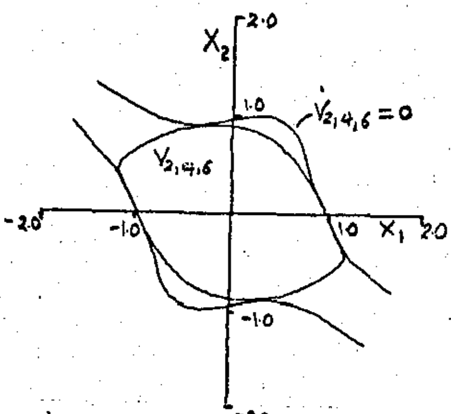
2 VALID TAN. PTS



$V = V_2 + V_4$

2 VALID TAN. PTS.

· OPTIMAL CONTOURS FOR SERIES LF (6.1.3) SYSTEM S18



$V = V_2 + V_4 + V_6$

2 VALID TAN. PTS.

V	ρ (AV. RAD.)	NO. OF FE'S	TIME (MILLISEC)
V_2	0·903	70	630
$V_{2,4}$	0·997	136	1024
$V_{2,4,6}$	1·00	288	2592

COMPARISON OF RAS'S WITH DOA

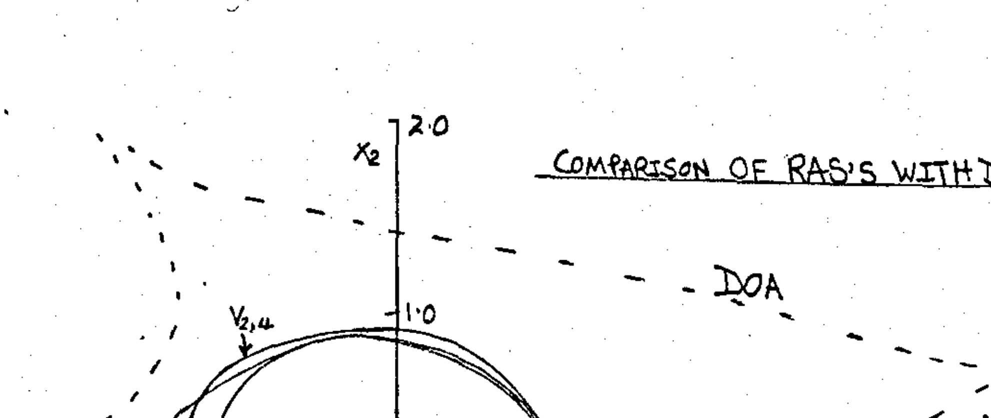
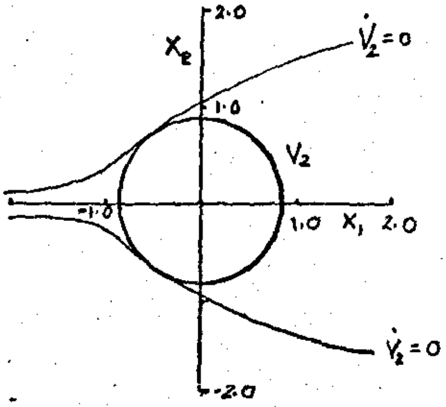
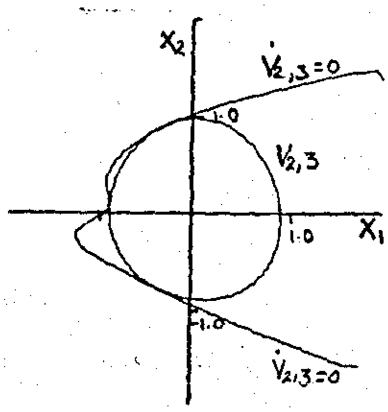


FIG. 6.4



$V = V_2$

2 VALID TAN. PTS.

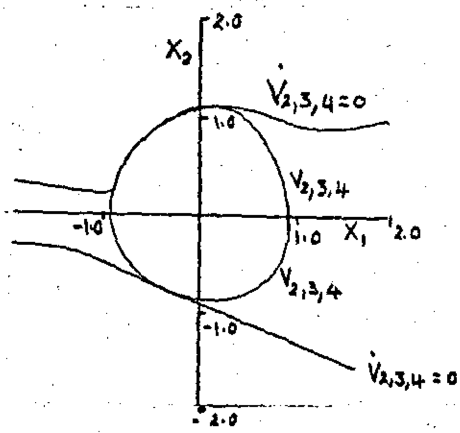


$V = V_2 + V_3$

3 VALID TAN. PTS.

OPTIMAL CONTOURS FOR SERIES LF (6.1.3)

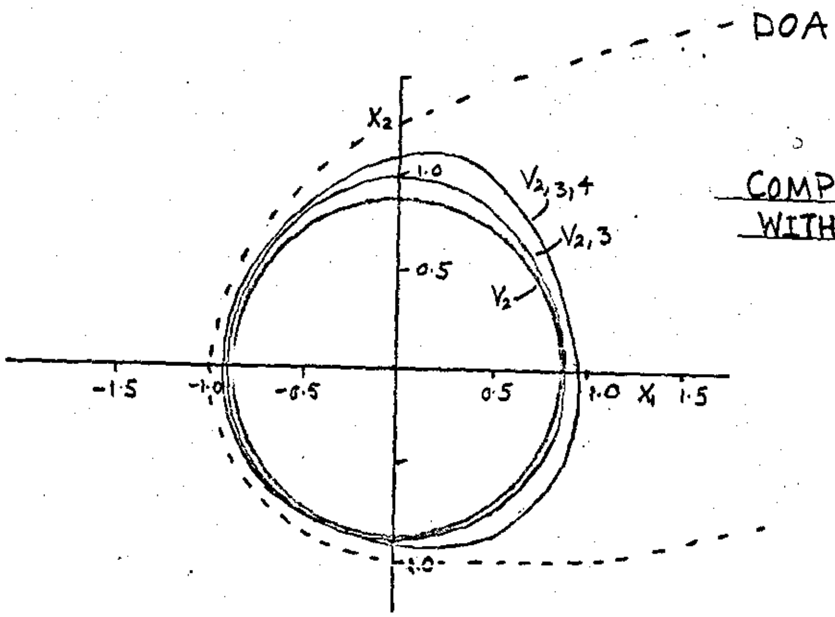
SYSTEM S19



$V = V_2 + V_3 + V_4$

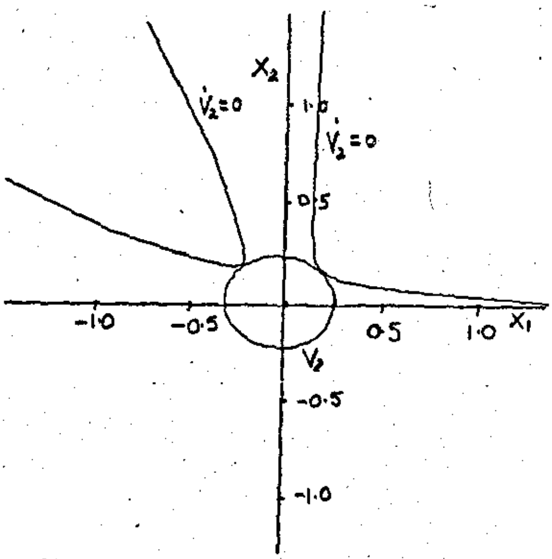
3 VALID TAN. PTS.

V	ρ AV. RAD.	NO. OF FE'S	TIME (MILLISEC)
V_2	0.86	40	174
$V_{2,3}$	0.903	101	590
$V_{2,3,4}$	0.951	318	1900



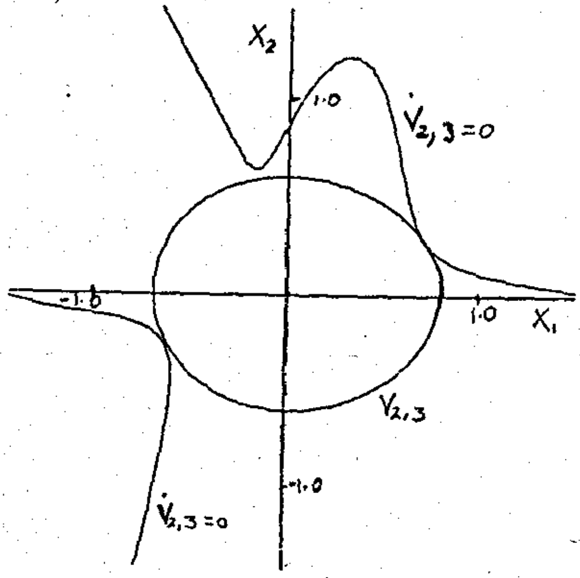
COMPARISON OF RAS'S WITH DOA

FIG. 6.5



$V = V_2$

2 VALID TAN. PTS.

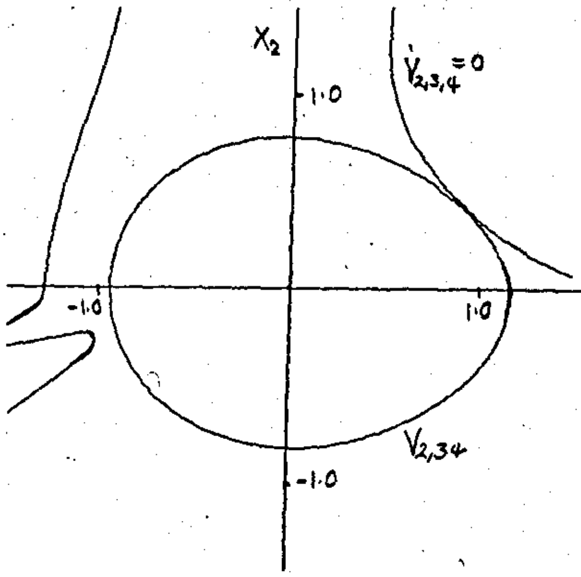


$V = V_2 + V_3$

2 VALID TAN. PTS.
(NOT CONVERGED)

OPTIMAL CONTOURS OF SERIES LF (6.1.3)

SYSTEM (Yu) 513

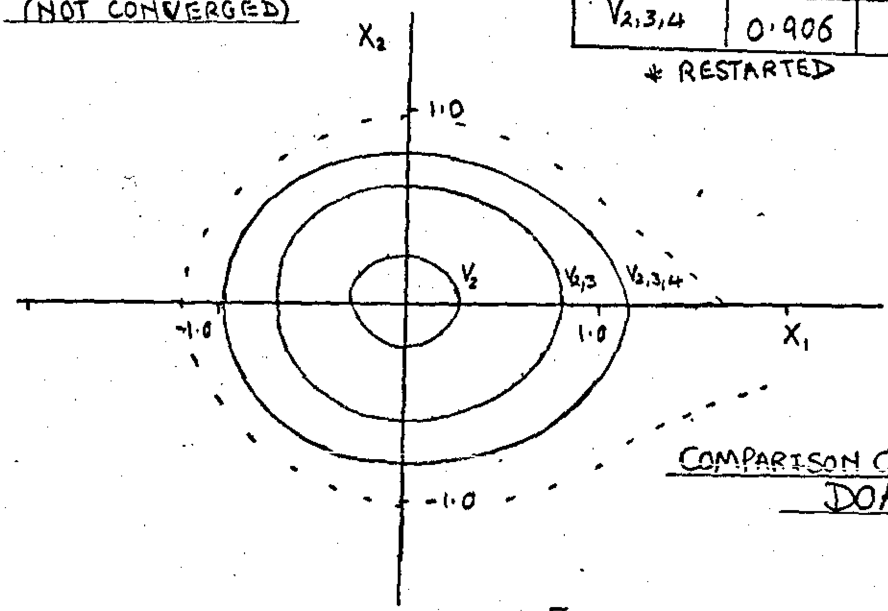


$V = V_2 + V_3 + V_4$

1 VALID TAN. PT.
(NOT CONVERGED)

V	ρ (AV. RAD.)	NO. OF FE'S.	TIME (MILL/SEC)
V_2	0.265	70	600
$V_{2,3}$	0.665	200	2400
$V_{2,3,4}$	0.795	226	2700
	0.906	59	6000*

* RESTARTED



COMPARISON OF RAS'S WITH DOA

FIG. 6.6

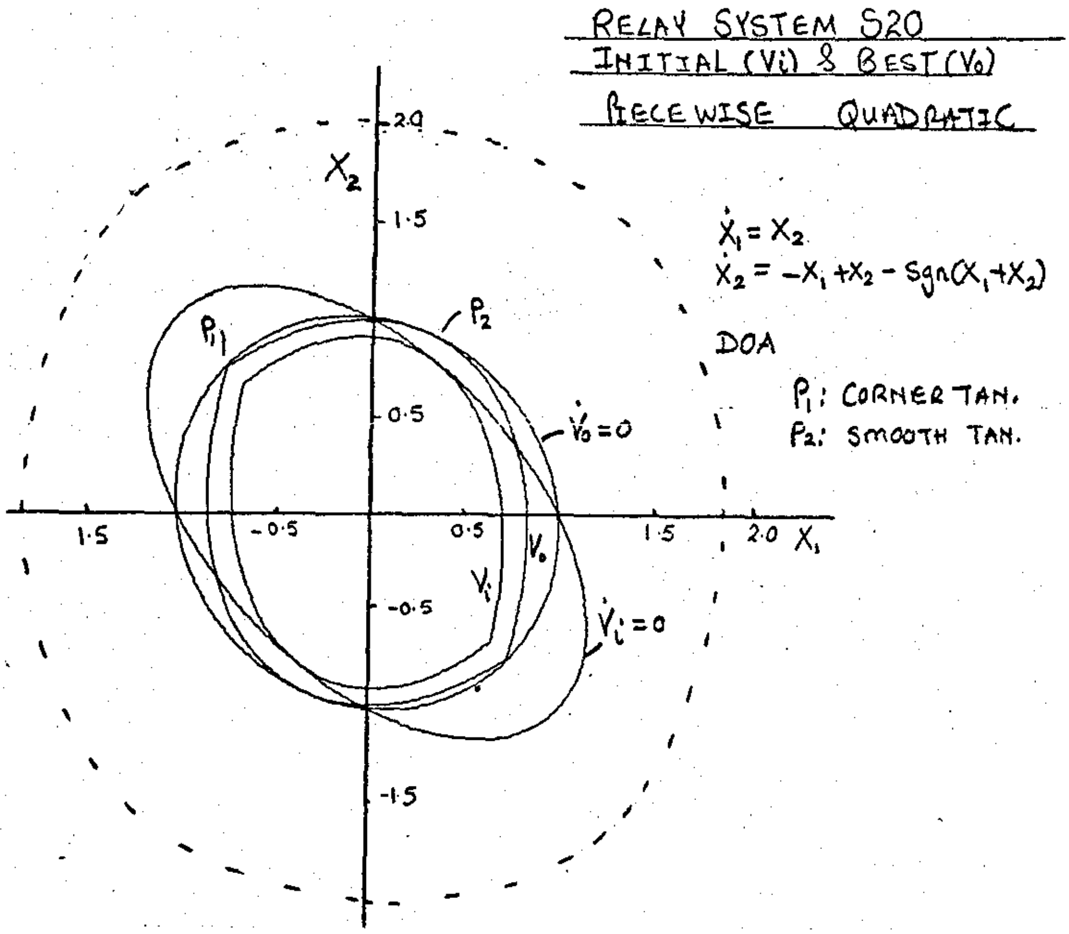

RELAY SYSTEM S20
INITIAL (Vi) & BEST (Vo)
PIECEWISE QUADRATIC
$\dot{X}_1 = X_2$
$\dot{X}_2 = -X_1 + X_2 - \mathrm{Sgn}(X_1 + X_2)$
DOA
P_1: CORNER TAN.
P_2: SMOOTH TAN.
X_2
2.0
1.5
0.5
-0.5
-1.5
1.5
-0.5
0.5
1.5
2.0
X_1
P_1
P_2
$\dot{V}_o = 0$
V_o
V_i
$\dot{V}_i = 0$

FIG. 6.7

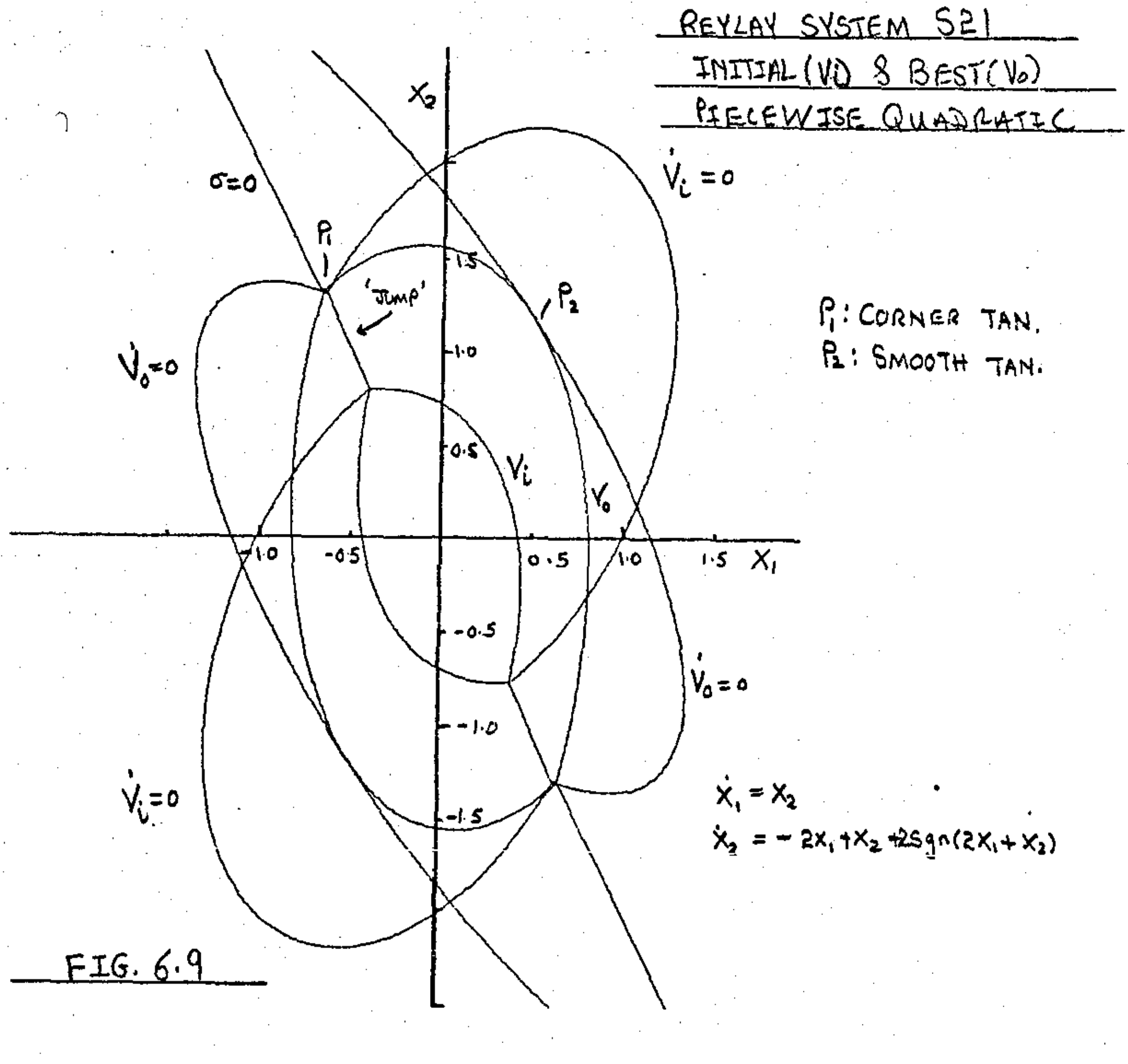
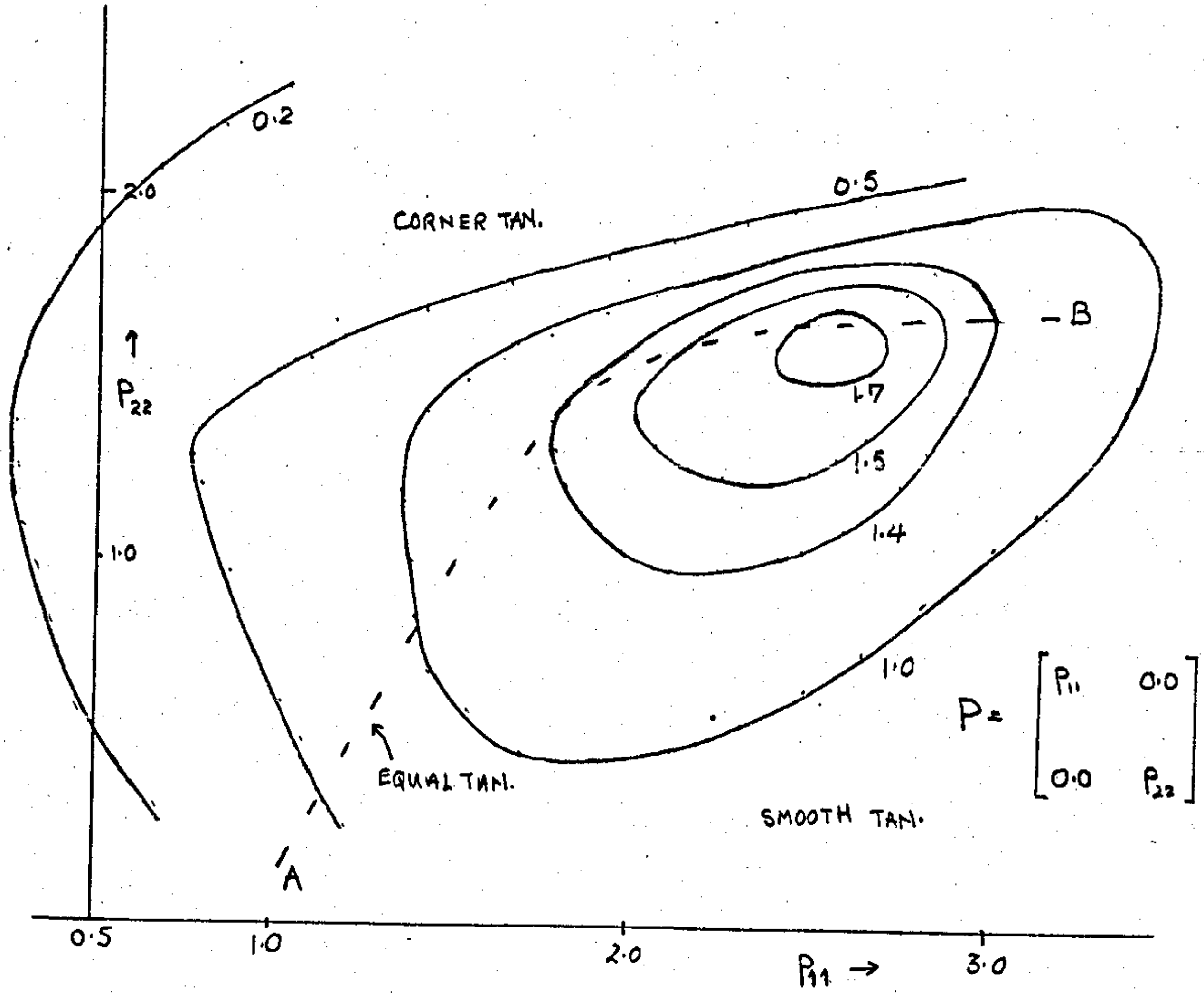
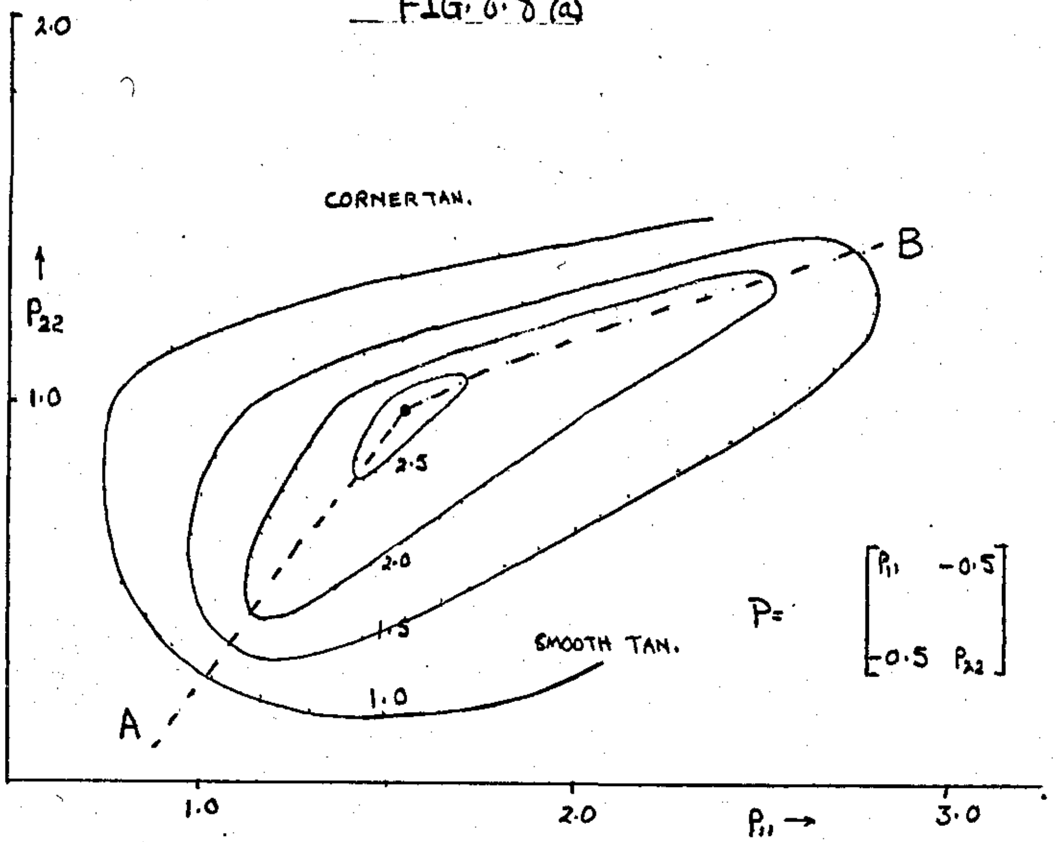

REYLAY SYSTEM S21
INITIAL (Vi) & BEST (Vo)
PIECEWISE QUADRATIC
X_2
$\sigma = 0$
P_1
'JUMP'
P_2
$\dot{V}_i = 0$
$\dot{V}_o = 0$
P_1: CORNER TAN.
P_2: SMOOTH TAN.
1.5
1.0
0.5
V_i
V_o
-1.0
-0.5
0.5
1.0
1.5
X_1
-0.5
-1.0
-1.5
$\dot{V}_o = 0$
$\dot{V}_i = 0$
$\dot{X}_1 = X_2$
$\dot{X}_2 = -2X_1 + X_2 + 2\mathrm{Sgn}(2X_1 + X_2)$

FIG. 6.9



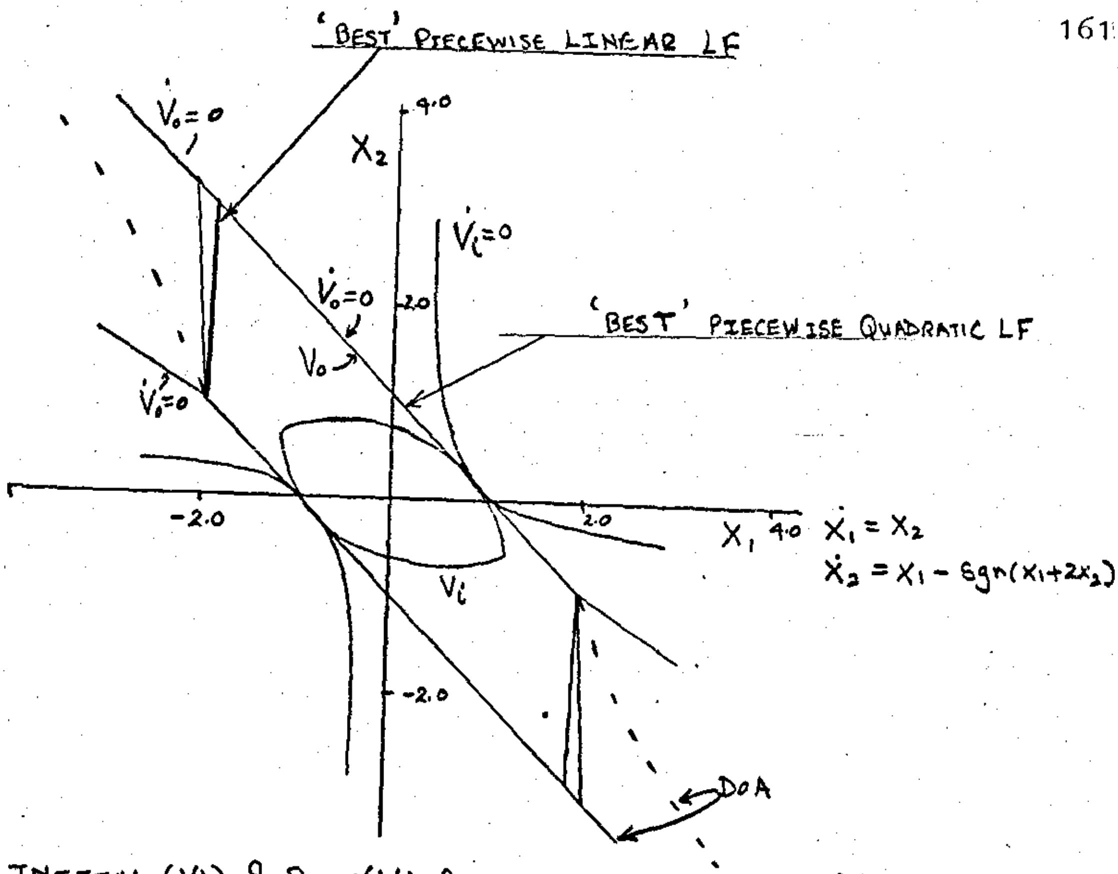
AREA CONTOURS FOR PIECEWISE QUADRATIC, SYSTEM S20 (REYLAY)

FIG. 6.8 (a)



SECTION THROUGH 'BEST P' (P_{12}=−0.5)

FIG. 6.8(b)



INITIAL (V_i) & BEST(V_o) PIECEWISE QUADRATIC, SYSTEM S22

FIG. 6.10

$$V = \underline{x}^T P \underline{x} + \int_0^{\sigma} f(s)\, ds = V_m$$

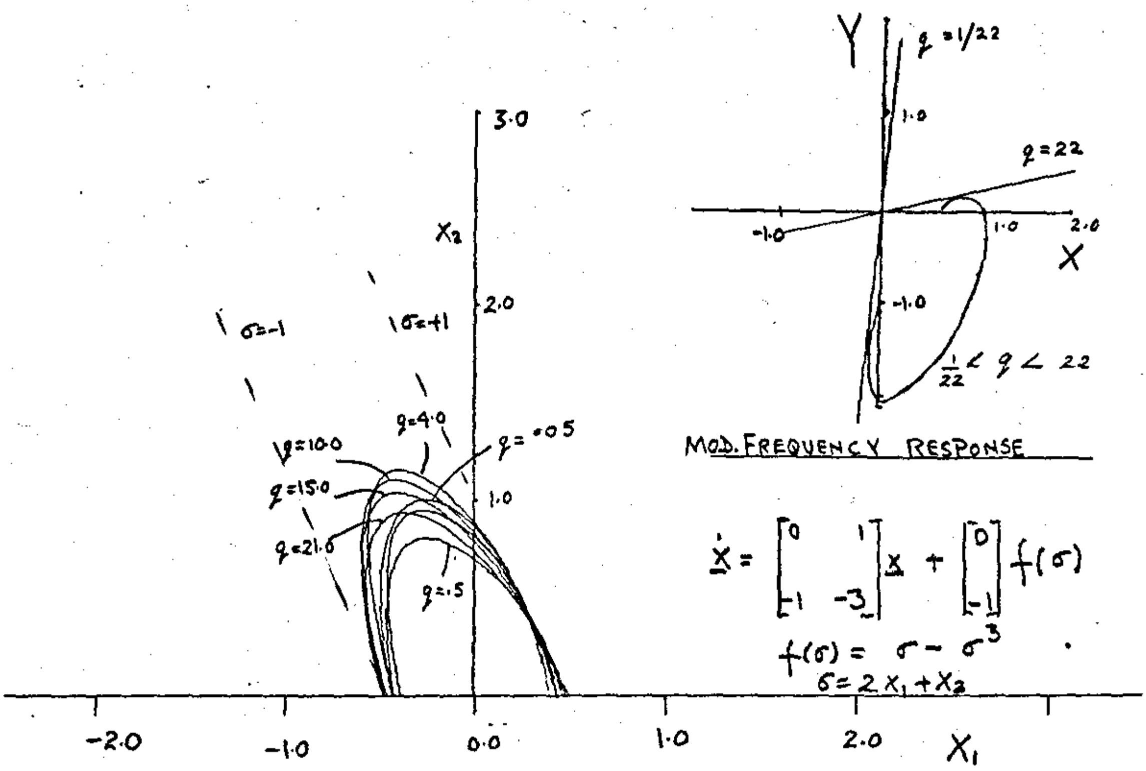


FIG. 6.11

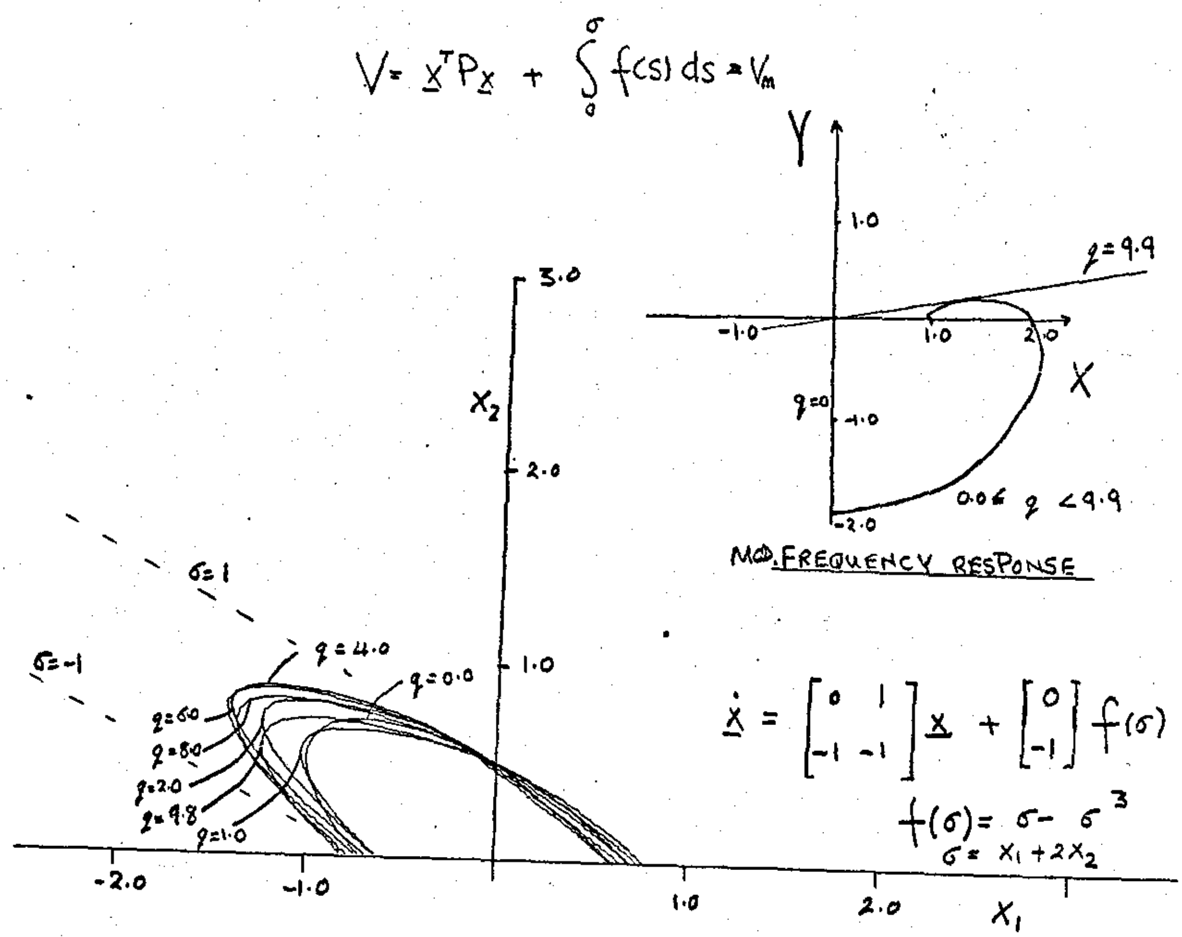


FIG. 6.12

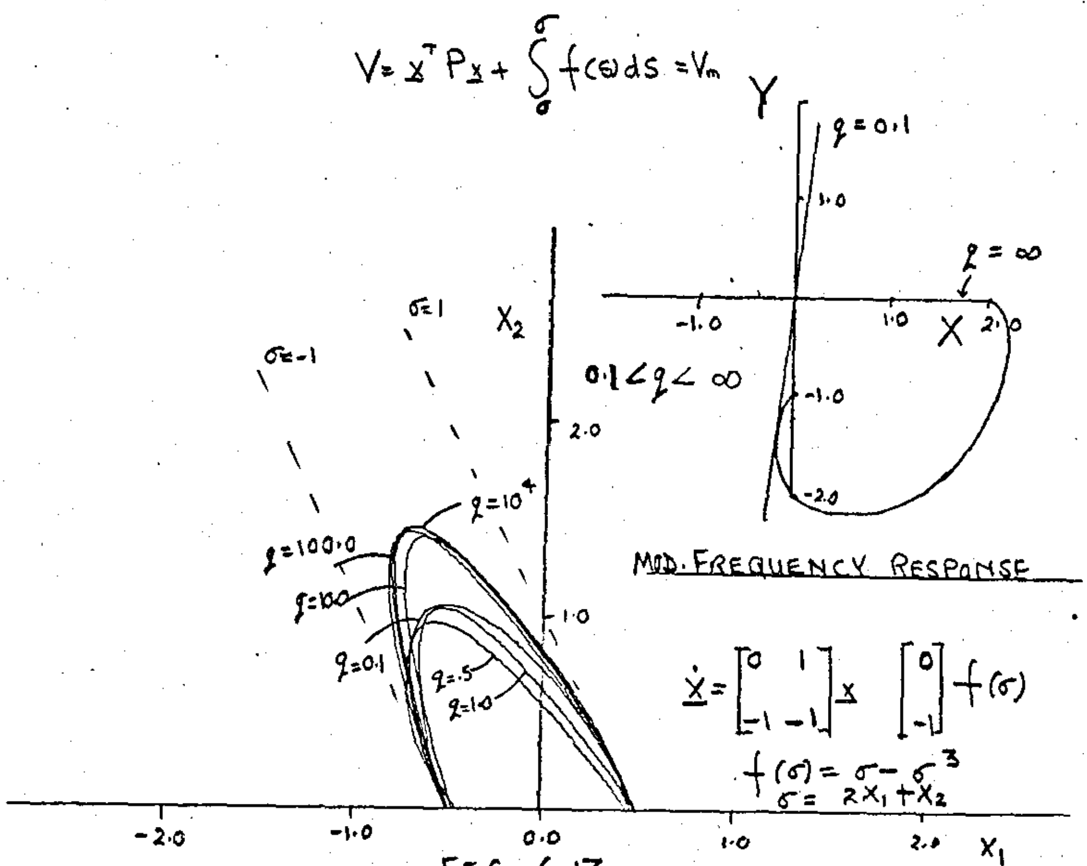
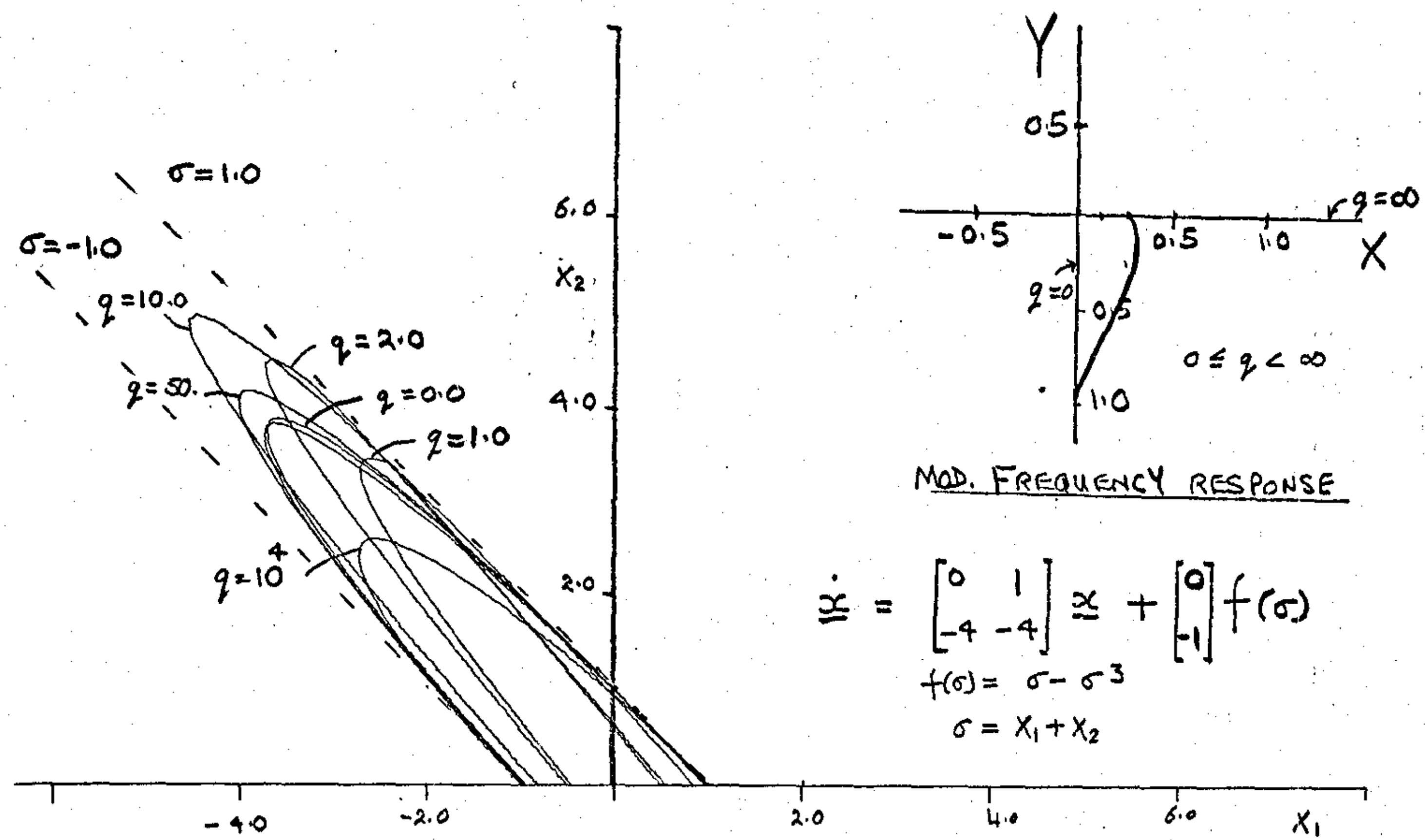


FIG. 6.13

$$V = x^T P x + \int_0^\sigma f(s)\,ds$$



$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -4 & -4 \end{bmatrix} x + \begin{bmatrix} 0 \\ -1 \end{bmatrix} f(\sigma)$$

$$f(\sigma) = \sigma - \sigma^3$$

$$\sigma = X_1 + X_2$$

FIG. 6.14

CONCLUSIONS

Conclusions

The continuous theme throughout this thesis has been the search for improved results when applying Lyapunov functions to autonomous non-linear systems. We intend here to bring together the main points arising from this search and add some suggestions for future work.

The problem of improving a crude RAS or an estimate of the transient response of a non-linear system was considered in Chapter 2. For the two forms of system considered the rate at which the trajectories approach the origin can be estimated through quantities α or β where $\alpha = \eta/2 - \sqrt{\mu}c_o$ and $\beta = \eta/2 - \mu c_2$; where α^{-1} and β^{-1} behave as 'time-constants' of the systems. The sub-optimum problem of maximizing η $(m(P^{-1}Q))$ and then minimizing $\mu(P)$ $(M(P)/m(P))$ over the space of p.d.s. matrices $\bar{Q}$ which results, was solved for a special case where A was in companion form with real eigenvalues. A conjecture and a bound of Wiberg were also proved invalid.

Although the theory of condition numbers is well established there are a surprisingly small number of useful bounds that are obtainable for $\mu(P)$. The main drawback of the ones proposed, namely, $F(P)$, $F(P^{-1})$ and that of 2.6.14, is the calculation of $d(s)$, although simplification arises with A in companion form.

The numerical work shows that the procedure of minimizing $\mu(P)$ $(P^{-1} = (SD)(SD)^*)$ over the extended space of hermitian matrices gives no improvement over that of the real matrices. The results in Table 2.1 show the extra effort in computing the estimates $\| |S| |S^{-1}| \|$ and $\mu(SS^*)$ $(\|s_i\| = 1)$ may be worthwhile. The evidence in Table 2.2 causes some doubt on the sub-optimum problem as simple choices of P or Q often give larger values of α or β.

Although the choice of a quadratic Lyapunov function gives rise to many simplifications in obtaining bounds for the transient response, the optimizations of quantities such as α^{-1} or β^{-1} are essentially independent of the non-linear parts of the system thus giving crude bounds. Unless one considers a very special type of system, LF's incorporating the non-linearity are hard to find.

An analytic solution of the Zubov eqn. in Chapter 3 would give the optimum result in that the exact DOA would be prescribed, $V < 1$ (regular equation). The impossibility of analytic solutions in general has led to the Zubov construction procedure with its non-uniform convergence of the RAS to the DOA. The analytic example given has thrown some light on the reason for this. However, the problem of determining whether the convergence of the RAS to the DOA is always non-uniform when the region of convergence of the power series LF is a subset of the DOA, would appear intractable analytically, and may be so numerically in view of the difficulty in finding the region of convergence.

The breakdown of the difference method near the DOA boundary when solving Zubov's PDE, is disappointing. It is due to the possibility of the analytic solution for V not being defined outside the DOA (Example 3.6.1 shows that with $\phi = r^2$, V is defined everywhere, whereas for $\phi = 10x^2 + y^2$, only in the DOA). The validity of the difference approach is also questionable and breakdowns occur. A more critical comparison of the methods of Troch, Burnand & Sarlos and Davidson and Cowan may prove interesting.

In the author's view the work in Chapters 4, 5 & 6 contains the essential result of this thesis. It has been shown, both from an analytical and numerical standpoint, that given a class of Lyapunov functions the optimal LF, V say, can have more than one valid tangency point with its constraint contour, $\dot{V} = 0$. Although the number of such points depends upon the system considered, for a quadratic LF, the property can be suitably defined as 'property A'. Analytic evidence has shown the intractability of proving that the property A always holds for autonomous systems not asymtotically stable in the whole. However, the numerical evidence gives some overwhelming evidence for such a property. A contrary result has not yet been encountered.

Defining a similar property for higher degree LF's is difficult due to the number of special cases - for example, a system giving only one tangency point for any LF - but the view that multiple valid tangency is a general phenomenon encountered with any class of suitable LF's is certainly supported by the evidence in this thesis.

In the following paragraphs we give the main points of each chapter.

The problem of RAS determination and its associated search for the global minima of a constrained function is a big obstacle in using Lyapunov's direct method. With this in mind Chapter 4 investigates the simplest class of Lyapunov functions, the quadratics. Even with very simple non-linear second order systems RAS determination is impossible. The systems studied show this in detail but in two cases optimal quadratics are obtained, one of which gives a lead to the equal tangency property, property A.

It is hoped that the second and higher order optimal quadratic algorithms will be of use to future researchers. They have certainly shown the existance of property A for a number of second and third order examples where accuracy was needed. Future work would be to compare some of the best optimization methods (e.g Fletcher and Lill, Nelder and Mead, Powell and DCS) on a varied selection of optimal quadratic problems. The degrading of the search directions near an equal tangency curve as for Powell's method, should affect these other methods to varying degrees. Future optimization algorithms should be chosen or developed on the fact that, although a good direction of search is along the equal tangency curve or surface, the RAS determination for a quadratic corresponding to a point on this curve is also degraded because of the extra effort in verifying the valid tangency point.

An extension of Chapter 4 is the problem of optimizing over the class of quadratic LF's in order to obtain a large region of the origin in which trajectories have a certain exponential stability. For example we maximize the volume inside a boundary $V = V_m$ in which also $\dot{V} + 2\lambda V < 0$ $(2\lambda < m(P^{-1}Q))$. The problem is similar to the one of the optimal quadratic and it is easily shown that for system S1 equal tangency exists.

In Chapter 5 the superiority of the 'complex' optimal quadratic algorithm over that of Davidson and Kurak was mainly due to some unsatisfactory steps in the latter algorithm, namely, solving the constrained problem by Rosenbrock's hill-climber and minimizing the eccentricity of the initial stability boundary. The main reason for the latter step was to avoid local minima of the objective function ρ (volume). In our experience local minima were rarely encountered. The main point was that of a bad initial quadratic which in the space of elements of L ($P = L^T L$) meant that the search for the optimum engaged an equal tangency surface someway from the optimum with little hope of reaching it.

It would be advantageous to find a large class of systems which for a quadratic LF have only one RS tangency point where the above maladies do not occur. A member is the system S4.

The penalty function approach to the quadratic RAS problem in Chapter 5 has given good results. Further experience with the penalty function method of Fletcher and Lill (63) for high order systems is required in view of the claims of the method by the authors.

The extensions made in Chapter 6 of obtaining quite general optimal LF's for non-linear systems, including relay systems, give conclusive numerical evidence for the equal tangency phenomenon being a general one, and not solely attributed to quadratics.

The problem of determining optimal LF's in the series form (6.1.3) has numerous pitfalls. Besides those of the quadratic, the greater number of valid tangency points for mv increasing, the possible open nature of the V_{mv} = constant boundaries and the vanishing of ∇V_{mv} at tangency, all cause serious problems in RAS determination for $mv > 2$. From a purely mathematical standpoint the question of why multiple valid tangencies occur is striking. As the degree of the LF increases so do the degrees of freedom of its contours, as shown in Figures 6.1 to 6.6, but the direct relationship between these contours and those of $\dot{V}_{mv} = 0$ at tangency, is obscured by the RAS determination. Finding the number of valid tangency points a' priori for a given LF of degree mv and a given system is an unsolved problem. It is hoped that the numerical evidence presented in this thesis has made some contribution to its solution.

By assuming an equal tangency property a useful sub-optimum piecewise linear LF has been obtained in Chapter 6 for a second order relay system. Perhaps this idea could be extended to other systems. The proof for general n that the Kalman construction procedure gives the optimal quadratic would be worthwhile in view of the ease in obtaining the region of attraction.

The numerous practical systems have shown that the optimal RAS of a quadratic often gives a good estimate of the DOA when the latter is radially symmetric, but that in general it is poor, especially for an open DOA (Figures 4.4, 5.3, 5.4, 5.5, 6.4 and 6.5 show this) Hewit (2), after comparing the best RAS's for a number of constructed Lyapunov functions (those of Krasovski , Ingwerson, Zubov and Szegö) concluded that Lyapunov's direct method was unsatisfactory when applying numerical construction procedures to it. The presence of the equal tangency phenomenon for second and high order systems and its associated convergence problems, gives even more evidence against Lyapunov's direct method as a tool for estimating the DOA.

We can conclude from the work in this thesis that the search for optimal results has been profitable and that applications range from simple autonomous systems to systems of discontinuous right-hand sides. It is hoped that the optimal phenomena shown have contributed to a better understanding of

Lyapunov's direct method.

REFERENCES

1. Zubov V.I, Methods of A.M.Lyapunov and Their Applications , Noordhoff Ltd, Gronigen, Netherlands, 1964.
2. Hewit J.R, A Comparison of Numerical Methods in Non-Linear Stability Analysis , Ph.D Thesis, 1968.
3. Rodden J.J, Numerical Applications of Lyapunov Stability Theory , J.A.C.C, Session IX, Paper 2, Stanford University, California, June 1964.
4. Margolis S.G, Vogt W.G., Control Engineering Applications of V.I.Zubov's Construction Procedures for Lyapunov Functions, I.E.E.E, AC-8, 1963, pp104-113.
5. De Sarker A.K, RAO N.D; Zubov's Method and Transient-stability Problems of Power Systems, I.E.E Proc. Vol 118, pp1035-1040, Aug. 1971.
6. Yu Y.N and Vongsuriya, Non-Linear Power System Stability Study by Lyapunov Function and Zubov's Method , I.E.E.E, Trans. Power App. Syst., Vol PA.S86, pp1480-1485, Dec. 1967.
7. Ferguson D.R, Generation of Zubov's Construction Procedure for Lyapunov Functions , Elec. Letters, Vol. 6, no.3, Feb. 1970.
8. Froberg C.E, Introduction to Numerical Analysis , Addison-Wesley, 1965.
9. Richtmyer R.D and Morton K.W., Difference Methods for Initial-Value Problems , Interscience Pub., 2nd Ed. 1969.
10. Burnand G, Sarlos G., Determination of the Domain of Stability , J.Math. Anal. Appl., Vol 23, pp714-722, 1968.
11. Davidson E.J, Cowan K.C., A Computational Method for Determining the Stability Region of a Second-order Non-linear Autonomous System , Int. J. Control, 1969, Vol.9, No.3, pp349-357.

12. Troch I., The Evaluation of the Domain of Attraction of Non-linear Control Systems with Hybrid Computing Systems , Procd. 5^{th} IFAC Cong., Part 4 No.32.1.

13. Kormanik J., Li C.C., Decision Surface Estimate of Non-linear System Stability Domain by Lie Series Method , IEEE, AC-17, Oct. 1972, pp666-669.

14. La Salle J.P. & Lefshetz S., Stability by Lyapunov's Direct Method with Applications , Acad. Press New York (1961).

15. Hahn W., Theory and Applications of Lyapunov's Direct Method, Prentice-Hall, 1963.

16. Kalman R.E & Bertram J.E., Control Synthesis and Design via the Second Method of Lyapunov, Part 1, A.S.M.E, D82, p371, 1960.

17. Lehnigk S.H., Stability Theorems for Linear Motions, Prentice-Hall 1966.

18. Barbashin E.A., Introduction to the Theory of Stability , Wolten-Noordhoff, 1970.

19. Vogt W.G., Transient Response from the Lyapunov Stability Eqn. , Joint Automatic Control Conference, 1965.

20. Barnett S., Solution of some Algebraic Problems Arising in the Theory of Stability and Sensitivity of Systems, with Particular Reference to the Lyapunov Matrix Equation , Ph.d Thesis (1967) loughborough University.

21. Barnett S. & Storey C., Matrix Methods in Stability Theory , Nelson, 1970.

22. Bauer F.L, Optimally Scaled Matrices, Numerische Mathematik S.73-87, 1963.

23. Wiberg D.M., A Simple Bound for Stable Non-linear Systems with Input , Int. J. Control, Vol.6, No.4, 343-350, 1967.

24. Lewis D.C., Taussky O, On the Eigenvalues of The Lyapunov Matrix Operator, J. Math. Phs., 1, 234, 1960.

25. Bass R.W, Control Synthesis & Optimization, NASA Langley Research Centre, Aug. 1961.

26. Bellman R.E., Stability Theory of Diff. Eqns., New York : McCraw-Hill, 1953, p35.

27. Marcus M & Haynsworth E, Bound for the P Condition No. of Matrices with Positive Roots, Nat. B. Stds., 65B, No 1, 1961, 13-14.

28. Gautshi W., On Inverses of Vandermonde & Confluent Vandermonde Matrices, Num. Math. 4, 113-123, 1962.

29. Reiss G., Inversion of the Vandermonde Matrix, I.E.E.E, AC-12, p793, 1967.

30. Marcus M., Minc H., A Survey of Matrix Theory and Matrix Inequalities, Allyn and Bacon, Boston, 1964.

31. I.S. Pace & S. Barnett, Comparison of Numerical Methods for Solving Lyapunov Matrix Eqns., I.J.C, Vol 15, pp907-915, 1972.

32. Rosenbrock H.H., An Automatic Method For Finding The Greatest or Least Value of a Function, Comp. Jnl., 3, pp175-84, 1960.

33. Powell M.J.D., An Efficient Method of Finding the Minimum of a Function of Several Variables Without Calculating Derivatives, Comp. Jnl., 7, pp155-62, 1964.

34. Nelder J.A., Mead R., A Simplex Method for Function Minimization, Comp. Jnl., 7, pp308-13, 1965.

35. Wilkinson J. H., Solution of Real and Complex Systems of Linear Equations, Num. Math., 8, pp217-234, 1966.

36. Wilkinson J. H., Three Contributions to Handbook Series Linear Algebra, Num. Math., 4, pp354-376, 1962.

37. Tait K. E., Stability Properties of Discrete-Continuous Feedback Control Systems With Signal Dependent Sampling, Int. J. Control, Vol 8, No 6, pp591-620, 1968.

38. Brockett R. E., The Status of Stability Theory For Deterministic Systems, I.E.E.E., Trans. Aut. Control, AC-11, 596, 1966.

39. Reiss G. G., Inversion of The Vandermonde Matrix, I.E.E.E., AC-12, pp793, 1967.

40. Benny L.B., Mathematics for Students of Engineering and Applied Science, Oxford Univ. Press, 1966.

41. Box M. J., Davies D., Swann W.H., Non-Linear Optimization Techniques, I.C.I Monograph No 5, Oliver and Boyd, 1969.

42. Powell M.J.D., A Method for Non-linear Constraints in Minimization Problems , Optimation Symposium, Academic Press, 1968.

43. Lefshetz S., Stability of Non-linear Control Systems, Academic Press, New York, 1965.

44. Davies H.L, Introduction to Non-linear Differential Eqns. & Integral Eqns., (1962), Dover Pub. Inc.

45. Bream R.C., M Sc. Thesis, Optimal Lyapunov Functions , Loughborough University, 1970.

46. Nelder J.A & Mead N., A Simplex Method for Function Minimization , Computer Journal, Vol 7, No.4, Jan 1965.

47. Davidson E.J.& Kurak E.M., A Computational Method for Determining Quadratic Lyapunov Functions for Non-linear Systems , Automatica, Vol 7, pp627-636, 1971.

48. Weissenberger S., Stability Analysis of Relay Control Systems via the Direct Method of Lyapunov , NASA, CR 320, 1965.

49. Lapidus L & Berger A, Stability of High Dimensional Non-linear Systems using Krasovski's Theorem , A.I.CH.E. Journal 15, 1969, pp171-173.

50. Geiss G.R, Estimation of the DOA, Grunmman Res. Dept., Report RM 3167, Betgpage, NY, 1966.

51. Geiss, Cohen and Rothschild : Development of an Algorithm for NL Stability Analysis of an Orbiting Astronomical Abs. Control System , Grunmman R.D, Dep RE-307 1967.

52. Szego G.P, On the Numerical Construction of Lyapunov Functions , IFAC, No.20, pp3-7, June 1969.

53. Smith R.A, Matrix Calculations For Lyapunov Quadratic forms, Journal Diff. Eqns., Vol 3, 2, pp208-217, 1966.

54. Fletcher R., Function Minimization Without Calculating Derivatives - A Review, Comp. Jnl., 8, pp33-41, 1965.

55. Box M. J., A New Method of Constrained Optimization and a Comparison With Other Methods, Comp. Jnl, 8, pp42-52, 1965.

56. Box M. J., A Comparison of Several Current Optimization Methods, and the use of Transformations in Constrained Problems, Comp. Jnl., 9, pp67-77, 1966.

57. Julich P. M., On Estimating the Region of Asymptotic Stability Using Lyapunov's Method, I.E.E.E., AC-12, pp787-788, 1967.

58. Muddle R. P., Determination of Stability Boundaries Using Lyapunov Functions and Numerical Optimization, R.A.E. Report No. 67148, 1967.

59. Fiacco A. V. & McCormick G. P., The Sequential Unconstrained Minimization Technique for Non-Linear Programming, a Primal-Dual Method, Mange. Sci., 10, pp360, 1964.

60. Fletcher R. & Powell M. J., A Rapidly Convergent Descent Method For Minimization, Compt. Jnl., 6, pp163-68, 1963.

61. Wilson F., The Structure of the Level Surfaces of a Lyapunov Function, Journal Diff. Eqns., 3, pp323-329, 1967.

62. Miele A. - Coggins G. M., On the Method Of Multipliers for Math. Programming Problems, JOTA, 10, No.1, 1972, pp1-32.

63. Fletcher R. & Lill S., Proceed. of Non-Linear Programming Symp., Madison, 1970.

64. Fletcher R., Certification of Algorithm 251 - Function Minimization, Comm. ACM, 9, pp686, 1966.

65. Haarhoff P. C. & Buys J. D., A New Method for the Optimization of a Non-Linear Function Subject to Non-Linear Constraints, Comp. Jnl., Vol 13, 2, pp178, 1970.

66. Weissenberger S., Piecewise-Quadratic and Piecewise-Linear Lyapunov Functions For Discontinuous Systems, Int. Jnl. Control, 10, No.2, pp171-180, 1969.

67. Kalman R. E., Lyapunov Functions for the Problem of Lure' in Automatic Control, Vol. 49, 2, pp201-205, 1963.

68. Anosov D. V., Stability of the Equilibrium Positions in Relay Systems, Autmn. Remote Control, Vol. 20, 2, pp130-143, 1959.

69. Alimov Y. I., On The Applications of Lyapunov's Direct Method To Differential Equations With Ambiguous Right Sides, Aut. Rem. Control, Vol. 22, No. 7, pp713-725.

70. Rosenbrock H. H., Some Conditions for the Stability of Nonlinear Time-dependent Differential Equations, Jnl. SIAM Control, 2, pp171, 1964.

71. Weissenberger S., Design for Stability Region Maximization, I.E.E.E., AC-14, pp314-315, 1969.

72. Walker J.A. & McClamroch N. H., Finite Regions of Attraction For The Problem of Lure', Int. J. Control, Vol. 6, No.4, pp331-336, 1967.

Appendix 1.

The following definition is important in defining the closed surfaces or contours of a continuous function $V(x)$ in some region $R(h)$.

Def. A1.1

The level surface $V(\underline{x}) = c$ (the set of points satisfying $V = c = \text{constant}$) is closed in $R(h)$ if any continuous line from the origin meets this surface in at least one point.

Theorem A1.1

If $V(\underline{x})$ is a p.d. (n.d.) function in $R(h)$ then $\exists$ a c, such that for $0 < c < c_1$ $(c_1 < c < 0)$ the level surfaces $V = c$ are closed.

Theorem A1.2 (Barbashin (18))

If $V(\underline{x})$ is p.d. (n.d.) and radially unbounded i.e.

$$\lim_{\|x\| \to \infty} V(\underline{x}) = \infty \ (-\infty)$$

its level surfaces are closed in E^n.

Appendix 2

Solving a Set of Linear Equations

Since we are interested only in the i th stage of the calculation of the V_{ij} we omit the i subscript. The set of linear equations 4.1. 6 gives ($n = N$ in C3)

$$\begin{aligned}
V_1 a_1 + V_2 b_1 \quad &= c_1 \\
V_2 a_2 + V_3 b_2 \quad &= c_2 \\
\cdot \quad \cdot \quad &\quad \cdot \\
\cdot \quad \cdot \quad &\quad \cdot \\
\cdot \quad \cdot \quad &\quad \cdot \\
V_{n-1} a_{n-1} + V_n b_{n-1} \quad &= c_{n-1} \\
V_1 b_n \quad \cdot \quad \cdot \quad \cdot + V_n a_n \quad &= c_n
\end{aligned} \qquad \text{A2.1}$$

or $\qquad A\underline{V} = \underline{c}$

Consider the following two algorithms:

a) starting with 1st equation express V_j in terms of V_1 up to the n-1 th equation; substition into the last equation then gives

$$V_1 \left(1 - (-1)^n \prod_{i=1}^{n} \frac{a_i}{b_i}\right) = \left(\frac{c_n}{b_n} - \frac{a_n \, c_{n-1}}{b_n \, b_{n-1}} + \frac{a_n \, a_{n-1} \, c_{n-2}}{b_n \, b_{n-1} \, b_{n-2}} \cdot \cdot \cdot \cdot \; (-1)^{n+1} \frac{a_n \, a_{n-1} \cdot \cdot \cdot a_2 \, c_1}{b_n \cdot \cdot \cdot b_2 \, b_1} \right)$$

If $b_i \neq 0$ & $1 - (-1)^n z \neq 0$ back substition gives

$$V_j = \frac{1}{b_{j-1}} (c_{j-1} - a_{j-1} V_{j-1}), j = 2, N$$

where
$$z = \prod^n \frac{a_i}{b_i}$$

b) starting with the last equation in A2.1 express V_j in terms of V_1 down to the 2nd equation substition into the 1st then gives

$$V_1 \left(1 - \frac{(-1)^n}{z}\right) = \left(\frac{c_1}{a_1} - \frac{b_1 c_2}{a_1 a_2} + \frac{b_1 b_2 c_3}{a_1 a_2 a_3} \cdots \right.$$
$$\left. \cdots (-1)^{n+1} \frac{b_1 \cdot \cdot b_{n-1} c_n}{a_1 a_2 \cdot \cdot a_{n-1} a_n} \right)$$

Provided $a_i \neq 0$ $1 - \frac{(-1)^n}{z} \neq 0$ back substition gives,

$$V_{n-j+1} = \frac{1}{a_{n-j}} \left(c_{n-j} - b_{n-j} V_{n-j+2} \right)$$

$$V_{N+1} = V_1$$

Thus provided $d(A) = \left(\prod^n b_i\right)(z - 1) \neq 0$ and with round off errors in view, choose algorithm

(a) if $|z| \leq 1$

and (b) if $|z| > 1$

Appendix 3

A Crude R.A.S.

With the assumptions on $\underline{g}(\underline{x})$ in 2.1, the bound 2.1.19 holds in some region R(h) for which $\beta > 0$ i.e.

$$\|\underline{g}(\underline{x})\| \leq \frac{h}{2\mu} \|\underline{x}\| \qquad \text{A 3.5.}$$

It is easily seen that A 3.5 is satisfied for all solutions of 2.1.2 with

$$\|\underline{x}(0)\| < h/\sqrt{\mu} \qquad \text{A 3.6}$$

where $h = \min_{\underline{x}} \|\underline{x}\|$ for $\underline{x}$ satisfying A 3.5

For then $\|\underline{x}(t)\| < h$ by 2.1.21 which implies A 3.5 and hence $\beta > 0$.

When $\underline{u} = 0$ and $G(\underline{x},t) = G(\underline{x})$ in 2.1.1, a possibly sharper estimate can be obtained if $\lim_{\|\underline{x}\| \to 0} G(\underline{x}) = 0$.

The crude R.A.S. is given by A. 3.6 with

$$h = \min_{\underline{x}} \|\underline{x}\|$$

for $\underline{x}$ satisfying $\|G(\underline{x})\| = h/2\sqrt{\mu}$

The Minimizing D (Bauer (22))

The diagonal matrix D giving the bound in 2.3.11 is determined as follows.

Let

$$M = |S^*S| \; |(S^*S)^{-1}|$$

and for this non-negative matrix let $\underline{e}^T Y_1$ and $X_1 \underline{e}$ be left and right eigenvectors (Perron) of M corresponding to the largest positive eigenvalue of M, where $\underline{e} = (1,1 \ldots, 1)$ and Y_1 and X_1 are diagonal matrices. Then

$$D = Y_1^{\frac{1}{2}} X_1^{-\frac{1}{2}} \qquad \text{A 3.9}$$

<u>A Minimization Algorithm</u>

Let p_{ij} and $q_{i,j}$ be the i,j th elements of P and Q where Q is any p.d.s. matrix with $m(= n(n+1)/2)$ elements. The optimization problems are to minimize

(a) $s(P^{-1}Q) = M(P^{-1}Q) - m(P^{-1}Q)$

(b) $\mu(P)$

(c) $-\eta$

(d) $-\eta^2/\mu$

over the (m-1) dimensional space of elements q_{ij} subject to

$$Q > 0 \text{ (positive definite)} \qquad \text{A 3.15}$$

where $A^T P + PA = -Q$ A 3.16

(Note one element of Q can be arbitrarily chosen)

The non-linear optimization algorithms of Rosenbrock (32) or Powell (33) are well suited to these problems and standard subroutines are available. (See A4). These routines require an auxiliary routine to evaluate the objective function and in this case the main steps were:

(1) from L, a lower triangular matrix input to the routine, calculate Q via

$$Q = L^T L + \epsilon I \qquad \text{A 3.17}$$

which is p.d. for $\epsilon > 0$;

(2) Solve A 3.16 for P by the direct method (20) i.e. solve

$$B\underline{p} = -\underline{q} \qquad \text{A 3.18}$$

where $\underline{p} = (p_{11}, p_{12}, p_{22}, p_{13}, p_{23}, \cdots \; p_{nn})$

$\underline{q} = (q_{11}, q_{12}, q_{22}, \cdots\cdots\cdots\cdots \; q_{nn})$

and B the m x m matrix of coefficients (See Barnett (20));

(3) compute R and R^{-1} where $P = R^{T}R$ and R a lower triangular matrix;

(4) compute $M(P^{-1}Q) = M(T)$ and $\eta = m(T)$ where $T = R^{-T}Q\,R^{-1}$ $(T \sim P^{-1}Q)$;

(5) For (b) and (d) compute m (P) and M (P).

The introduction of L avoids the n implied constraints of A 3.15 and the minimization is thus over the m-1 elements (l_{ij}, $i = 2; n, j = 1, 2 \,.., i$) of L($l_{11} = 1.0$ say)

(In practice A 3.18 was solved by a Crout iterative method (35) and the eigenvalues of T by the method of Householder (36). The choice of $\epsilon = 10^{-10}$ was sufficient in A 3.17.

Appendix 4

The optimization methods used in this thesis concern the programming problem

$$\min F(\underline{x}) \qquad \text{A4.1}$$

subject to (a) an equality constraint $e(\underline{x}) = 0$

or (b) an inequality constraint $c(\underline{x}) \leq 0$

The class of methods used to solve A4.1 depends on the form of F and the constraints. Fletcher (54) and Box (56) give good accounts. The methods of Rosenbrock (32), Powell(33) and Nelder and Mead (34) are well known and standard packages were used written in Fortran IV. We only mention that Rosenbrock deals with the constraint , $c \leq 0$, by introducing an implicit variable $x_{n+1} = c$ which is deemed feasible if

$$z(1 - 10^{-4}) \leq x_{n+1} \leq z10^{-4}$$

where z is a lower bound for c. Should this not be satisfied but x_{n+1} lies in one of the boundary zones $z10^{-4} < x_{n+1} \leq 0$ or $z \leq x_{n+1} < z(1 - 10^{-4})$ a special penalty function is used otherwise the value is termed a failure.

Box's Complex Method For Constraint (b)

A method modifying the Simplex method, it consists of forming a complex - a set of K ($\geq n+1$) independent points $\underline{p}_j (j = 1, \ldots , K)$, at which F is evaluated. Initially, assume all points are feasible i.e. $c(\underline{p}_j) \leq 0$. Let $\underline{p}_h$ correspond to F_h, the highest value of F in the complex, and $\bar{\underline{p}}_h$ the centroid of points $\underline{p}_j$ excluding $\underline{p}_h$. Then a stage consists of the following : with $\alpha = \alpha_o$ and

$$\underline{p} = \bar{\underline{p}}_h + \alpha(\bar{\underline{p}}_h - \underline{p}_h)$$

(an over-reflection of $\underline{p}_n$ through the centroid) evaluate $F(\underline{p})$ and (1) if $c(\underline{p}) > 0$, α is halved and the calculation repeated until $c(\underline{p}) \leq 0$, then (2) if $F(\underline{p}) < F_h$ is not satisfied α is again halved until it is. The final values of $\underline{p}$ and $F(\underline{p})$ then replace $\underline{p}_h$ and F_h and new values of $\underline{p}_h$ and F_h are found with new centroid

$$\bar{\underline{p}}_h(\text{new}) = \bar{\underline{p}}_h(\text{old}) + (\underline{p} - \underline{p}_h)/(K - 1)$$

This completes a stage. The calculation is repeated until after I stages, $I \geq I_{max}$ or $|F_h - F_{lowest}| < \epsilon$. (In C5, $\alpha_o = 1.3$, $K = 2n + 1$, $\epsilon = 10^{-4}$).

Penalty Function Updating

With respect to the penalty function of Miele (62),

$$W(\underline{x}) = R(\underline{x}, \lambda) + Ke^2(\underline{x})$$

the following rules are used for the conjugate gradient (DFP) algorithm :

(1) Initially select $\lambda = \lambda_1 = 0$, $K = K_1$ ($K_1 = 20$ say)

(2) With some initial $\underline{x}_o$ and $H = I$ perform n iterations of DFP on $W(\underline{x})$. Let $\underline{x}^*$ be the final point.

(3) If $\|\nabla R(\underline{x}^*, \lambda_1)\| < \epsilon$ and $|e(\underline{x}^*)| < \epsilon$ convergence is assumed, otherwise

(4) An updated λ is chosen to minimize $\|\nabla R(\underline{x}^*, \lambda_2)\|^2$, of the form $\lambda_2 = \lambda_1 + 2\gamma e$, which gives

$$\lambda_2 = \lambda_1 - \frac{\nabla e^T \nabla R(\underline{x}^*, \lambda_1)}{\|\nabla e\|^2} = - \frac{\nabla e^T \nabla V}{\|\nabla e\|^2}$$

The value of K is either kept constant or a criterion of reducing $\|\nabla R\|$ and $|e|$ at the same rate is employed which gives for the updated K, K_2;

$$K_2 = \min(K_o, K_1) \text{ if } |e| \leq \|\nabla R(\underline{x}^*, \lambda_2)\|$$

$$K_2 = \max(K_o, \beta K_1) \text{ if } |e| > \|\nabla R(\underline{x}^*, \lambda_2)\|$$

where $\beta \geq 1$ and $K_o = |\lambda_2/e|$ where the latter makes the order of the term Ke^2 the same as that of R. Step (2) is now repeated.

Fletcher Powell Method (DFP)

When derivatives of A5.1 are available the algorithm determines a sequence of iterates $\underline{x}_i$ and directions $\underline{s}_i$ as follows : with some initial p.d.s. H, H = I say, and initial $\underline{x}_o$,

$$\underline{x}_{i+1} = \underline{x}_i + \lambda \underline{s}_i = \underline{x}_i + \underline{z}_i$$

where $\underline{s}_i = -H_i \underline{g}_i$, $\underline{g}_i = \nabla F(\underline{x}_i)$

and λ determines the minimum of F along a line

$$\left.\frac{\partial F}{\partial \alpha}(\underline{x}_i + \alpha \underline{s}_i)\right|_{\alpha = \lambda} = 0.$$

Here H_i is updated as

$$H_{i+1} = H_i + \frac{\underline{z}_i \underline{z}_i^T}{\underline{z}_i^T \underline{u}_i} - \frac{H_i \underline{u}_i \underline{u}_i^T H_i}{\underline{u}_i^T H_i \underline{u}_i}$$

where $\underline{u}_i = \underline{g}_{i+1} - \underline{g}_i$.

A converted Algol program of Fletcher (64) was used where cubic interpolation is advocated to determine the min. along a line. The initial step, $\lambda = s_1$, along the latter was restricted to

$$s_1 = \frac{0.1 \ \|\underline{x}_o\| \ \underline{s}_i}{\|\underline{s}_i\|}$$

for the problem in Chapter 5.

Appendix 5

The Complex Optimal Quadratic Algorithm

The program, designated M459, is written in Fortran 4 and flow diagrams are given in Figures 5.1 and 5.2. Access to the ICL library of scientific subroutines is needed to run the program (FPQRVS, FPDIRHESSF, FPQRHESSE, FPBACK, F4ACSL and F4ROOT1 are from it). The instructions are intended for guidance in running the program.

Input

The sequence of input variables are as follows:

1) N and LOADQ are set in the Master segment; n being stored in N and LOADQ being of logical value .TRUE. or .FALSE. .

2) The initial complex matrix D is input and stored in array DEL (i,j) (j = 1, 2n, i = 1, n)

3) The elements of A, a_{ij}, are input in array A(I) such that A(k) contains a_{ij}, k = n(j-1) + i. (In the Rao system in the listing, A(k) have been calculated in the program for accuracy).

The r.h.s. of the system $\underline{f}(\underline{x})$ are set through subroutine CALFX (N, X, F) where X and F are one-dim. arrays containing x_i and f_i. Unit vectors $\underline{y}_j$ are stored in array Y(i) such that the i^{th} component of $\underline{y}_j$ is contained in Y(n(j-1)+i). They are calculated through a call to SETUNVEC (S, NV, N, Y, NX, P) where S is the step in Θ_1 and S/cosΘ_2 the step in Θ_2 for unit vectors calculated via polar co-ordinates 4.5.1; NV is the number of unit vectors calculated; NX = N(N+1)/2, and the array P(i) contains the initial P such that p_{ij} (j=1,n, i=1,j) = P(i+j(j-1)/2). If the 'box' method is used S is immaterial and N1, the number of mesh points along a half-edge of the box, is set in SETUNVEC (N1 is set after statement 61. S is set after statement 100 in the master segment).

Output

After some checking output the following sequence is observed:

1) The elements of initial P and Q are output in the order that the p_{ij} are read.

2) After each evaluation of $\dot{V}_M(1)$ the values of $\underline{y}_{max}$, $\dot{V}_M(1)$ and 1 are output in order, where $\underline{y}_{max}$ gives $\dot{V}_M(1)$.

3) The final 1, 1_2, is output (called E) followed by the volume of $\underline{x}^T P \underline{x} < 1_2$.

4) Output from the setting up of the initial feasible complex - rows containing ρ (vol), $\dot{V}(\underline{p}_j)$ and S (initially 2max $|t_i|$). If no feasible complex is found the words

'IFEASIBLE SIMPLEX'

are output.

5) The complex points $\underline{p}_j$ (j=1, 2n+1) are output.

6) After every 10 evaluations of $\dot{V}_M(L)$, $\underline{y}_{max}$, p_{ij} (j=1,n, i=1,j), $\dot{V}_M(L)$ and $-\rho$ are output. Then t_i (i=1,NX) are output on a new line (elements of L). The co-efficients of $\underline{y}_{max}$ follow the word 'MAXVDOT'.

```
      MASTER M459
      LOGICAL LOADQ
      COMPLEX SS(4,4),SUN
      DIMENSION DEL(6,12)
      DIMENSION T(110),A(16),XA(16),REINT(5),X(16),P(10)
      DIMENSION Y(4000)
      DIMENSION INT(4),        AA(16),EX(4),EY(4),ITS(4),AAA(16)
      DIMENSION G(14),H(14)
      N=2
      N=3
      NX=(N*(N+1))/2
      NK=2*NX
      NSQ=N*N
      READ(1,1)((DEL(I,J),I=1,NX),J=1,NK)
      LOADQ=.TRUE.
      LOADQ=.FALSE.
      READ(1,1)(A(I),I=1,NSQ)
    1 FORMAT (6E12.5)
      Z2=COS(.478)
      Z1=SIN(.478)
       AA(1),AA(6),AA(7)=0.
      AA(4)=1
      AA(2)=43.06*COS(.956)-84.99+1.18*Z2
      AA(3)=0.421*Z1
      AA(3)=-AA(3)
      AA(5)=0.803*Z2*Z2+1.04*Z1*Z1
      AA(5)=-AA(5)
      AA(8)=-84.99*Z1
      AA(9)=-.621
      DO 83 I=1,NSQ
   83 A(I)=AA(I)
      IF(LOADQ) GO TO 61
C                                     READ A BY COLS
      DO 10 I=1,NSQ
   10 AA(I)=A(I)
      IF(N-2)0,0,49
      Z1=(A(1)+A(4))/2.
      Z2=A(1)*A(4)-A(2)*A(3)
      D=Z1*Z1-Z2
      D1=SQRT(ABS(D))
      IF(D)0,52,52
      EX(1),EX(2)=Z1
      EY(1)=-D1
      EY(2)=D1
      GO TO 53
   52 EY(1),EY(2)=0.
      EX(1)=Z1-D1
      EX(2)=Z1+D1
   53 CALL FPORVS(N,AA(1),AAA(1),EX(1),EY(1),T(1))
      GO TO 48
   49 CONTINUE
      CALL FPOIRHESSE(N,AA(1),INT(1))
      IVS=0
      CALL FPORHESSE (N,AA(1),ITS(1),EX(1),EY(1),AAA(1),IVS)
C                                     EIGENVALUES ARE EX+EY
      CALL FPORVS(N,AA(1),AAA(1),EX(1),EY(1),  T(1))
      CALL  FPBACK(N,AA(1),AAA(1),EY(1),INT(1))
   48 CONTINUE
      WRITE(2,44)((AAA(I+N*(J-1)),J=1,N),I=1,N)
   44 FORMAT(3E12.5)
C                                     FIND S
```

```
      J=1
    6 IF(ABS(EY(J))-1.0E-10)0,0,2
      DO 3 I=1,N
      K=I+N*(J-1)
    3 SS(I,J)=CMPLX(AAA(K),0.0)
      J=J+1
      GO TO 5
    2 DO 4 I=1,N
      K=I+N*(J-1)
      SS(I,J)=CMPLX(AAA(K),AAA(K+N))
    4 SS(I,J+1)=CMPLX(AAA(K),-AAA(K+N))
      J=J+2
    5 IF (J-N) 6,6,0
C                                      FIND UNIT  VECTORS
      DO 41 J=1,N
      SM=0.0
      DO 42 I=1,N
      AMD  =CABS(SS(I,J))
   42 SM=SM + AMD*AMD
      SUN =CMPLX(SQRT(SM),0.0)
      DO 41 I=1,N
   41 SS(I,J)=SS(I,J)/SUN
      WRITE(2,43)((SS(I,J),J=1,N),I=1,N)
   43 FORMAT(3(2E12.5))
C                                      FIND P=S*S, REUSE AAA FOR STORAGE
C                                      RE-USE AA TO SOLVE AAA S*S=I
      DO 7 J=1,N
      DO 7 I=1,N
      SUN=(0.0,0.0)
      DO 8 K=1,N
    8 SUN=SUN+CONJG(SS(J,K))*SS(I,K)
      KK=I+N*(J-1)
      WRITE(2,51)SUN
   51 FORMAT(2E12.5)
    7 AAA(KK)=REAL(SUN)
C                                      INVERT; STORE UNIT MATRIX IN AA
      IN=1
      DO 11 K=1,NSQ
   11 AA(K)=0.0
      DO 12 I=1,N
   12 AA(I+N*(I-1))=1.0
      CALL F4ACSL(AAA,AA,N,NSQ,NSQ,IN,X,D,ID,IT,XA,T,REINT)
      DO 13 J=1,N
      DO 13 I=1,J
      K=I+(J*(J-1))/2
   13 P(K)=X(I+N*(J-1))
       GO TO 62
   61 NSQ2=NSQ*NSQ
      CALL SETA(N,NSQ,NSQ2,T,A)
C     READ Q  MATRIX BY ROWS
C        DIM   OF  Q     =    N*N
      READ(1,1)(XA(I),I=1,NSQ)
      IN=1
      CALL F4ACSL(T,XA,NSQ,NSQ2,NSQ,IN,AA,D,ID,IT,Y,AAA,A)
      K=0
      DO 63 I=1,N
      DO 63 J=1,I
      K=K+1
   63 P(K)=-AA(N*(I-1)+J)
      WRITE(2,65)
   65 FORMAT(10X,16HINITIAL Q MATRIX)
      WRITE(2,1)((XA(N*(I-1)+J),J=1,I),I=1,N)
```

```
   62 CONTINUE
      WRITE(2,45)
   45 FORMAT(1X,9HINITIAL P)
      WRITE(2,44)(P(K),K=1,NX)
      X(1)=SQRT(P(1))
      DO 14 J=2,N
      K=(J*(J-1))/2+1
   14 X(K)=P(K)/X(1)
      DO 16 I=2,N
      K1=(I*(I+1))/2
      SUM=P(K1)
      K2=(I*(I-1))/2
      DO 15 L=1,I-1
   15 SUM=SUM-X(K2+L)*X(K2+L)
      X(K1)=SQRT(SUM)
      DO 16 J=I+1,N
      K=(J*(J-1))/2
      SUM=P(I+K)
      DO 17 L=1,I-1
   17 SUM=SUM-X(L+K2)*X(L+K)
   16 X(I+K)=SUM/X(K1)
      WRITE(2,47)(X(I),I=1,NX)
   47 FORMAT(1X,9HINITIAL X,3E12.5)
C                                   START INITIAL VDOT CONTACT
  100 CONTINUE
      FR1=.2
      FR1=.3
      S=.1
      S=.2
      CALL SETUNVEC(S,NV,N,Y,NX,P)
      II=0
      E1=1.0E-5
      E1=1.0E-10
      E1=.01
      NNV=N*NV
      SMAX=10.0
      S=.1
  101 CONTINUE
      CALL VMAX1(N,NV,Y,NNV,X,NX,II,E1,SM1)
      IF (SM1) 20,0,0
      E1=.1*E1
      IF(E1-1.0E-10)0,101,101
      WRITE (2,21)
   21 FORMAT(1X,14HSMALL VOL STOP)
      Z=1./0.
      STOP
   20 E2=E1+S
      CALL VMAX1(N,NV,Y,NNV,X,NX,II,E2,SM2)
      IF (SM2) 0,23,23
      E1=E2
      SM1=SM2
      IF (S-SMAX) 25,25,0
      WRITE (2,24) E1
   24 FORMAT (1X  ,16HVDOT NOT REACHED,E12.5)
      STOP
   25 S=2.*S
      GO TO 20
   23 CONTINUE
      S=S/2.
   27 E3=E1+S
      CALL VMAX1(N,NV,Y,NNV,X,NX,II,E3,SM3)
      S=S/2.
```

```
      IF (SM3) 0,28,28
      SM1=SM3
      E1=E3
      GO TO 29
   28 SM2=SM3
      E2=E3
   29 E=SM1/E1
      IF(2.*S/E2-ER1)0,27,27
      WRITE(2.46)E1
   46 FORMAT(1X,10HVALUE OF E,E12.5)
      NW=1000
      NW=2000
      E=SQRT(E1)
      Y(NW)=NV
      Y(NW-1)=E
      Y(NW-2)=N
      IV=0
      CALL CALF(NX,X,F,M,IV,XM,Y,NW)
      XA(1)=F
      WRITE(2.1)F
      EE=1./E
      DO 81 I=1,NX
   81 T(I),X(I)=X(I)*EE
      STP=.1
      Y(NW-1)=1.
      STP=0.
      DO 82I=1,NX
   82 STP=AMAX1(STP,ABS(X(I)))
       STP=.1*STP
C     CAL. PTS    OF   NON REGULAR  SIMPLEX
      K=NX+1
      IV=1
      DO 84 J=1,NK
      HS=STP
   95 NEG=0
   93 NEG=NEG+1
      DO 86 I=1,NX
   86 G(I)=X(I)+DEL(I,J)*HS
   87 CALL CALF(NX,G,F,M,IV,XM,Y,NW)
      XA(J+1)=F
      WRITE(2.1)F,XM,HS
      IF(XM)85,85,0
      IF(NEG-1)0,0,94
      HS=-HS
      GO TO 93
   94 HS=-HS/2.
      IF(STP*.01-HS)95,92,92
   85 DO 89 I=1,NX
      T(K)=G(I)
   89 K=K+1
   84 CONTINUE
      GO TO 88
   92 CONTINUE
      WRITE(2.90)
   90 FORMAT(1X,18HINFEASIBLE SIMPLEX)
      STOP
   88 K2=K2+NX
      NK=NK+1
      WRITE(2.1)(H(I),I=1,NX)
      WRITE(2.1)(T(I),I=1,NX*NK)
      Y(NW-1)=1.0
      MF=500
```

```
      MF=300
      MF=1000
      EPS=ABS(XA(1))*1.0E-3
      EPS=1.0E-4
      EPS=ABS(XA(1))*1.0E-4
      NK=2*NX+1
      NP=(NK+1)*NX
      CALL COMPSIM(X,F,EPS,T,XA,NX,NP,MF,Y,NW,NK)
      GO TO 100
      STOP
      END
```

```
      SUBROUTINE AMATV(X,NX,B,N)
      DIMENSION X(NX),B(N)
C                                   ROUTINE MULTIPLY XB-X UPPER TRIANGULA
C                                   BY COLS. STORE RESULT IN B - A VECTOR
      DO 2 I=1,N
      SUM=0.0
      DO 1 J=I,N
      K=I+(J*(J-1)/2)
    1 SUM=SUM+X(K)*B(J)
    2 B(I)=SUM
      RETURN
      END
```

```
      SUBROUTINE  CALFX(N,X,F)
      DIMENSION X(N),F(N)
      E=X(3)+1.181665534
      D=X(1)+.4779929599
      A=SIN(D)
      B=COS(D)
      F(1)=X(2)
      F(2)=28.61-84.99*E*A+21.53*SIN(2.*D)-X(2)*(
     11.04*A*A+0.802*B*B)
      F(3)=.36-.621*E+.421*B
      RETURN
      END
```

```
      SUBROUTINE COMPSIM(X,F,EPS,P,FP,N,NP,MF,Y,NW,NK)
      DIMENSION X(N),P(NP),FP(NK)
      DIMENSION Y(NW)
      DIMENSION XS(10)
      IC=10
        NNN=Y(NW-2)
        APO=1.5
      APO=1.3
      NN=N*NK
      N1=NK
      IV=0
      ISH=0
      M=1
C     SORT IH(IGHEST) AND IL(OWEST)
   35 IF(FP(1)-FP(2))0,0,5
      IH=2
      IL=1
       GO TO 6
    5 IH=1
      IL=2
    6 DO 7 I=3,N1
      IF(FP(I)-FP(IH))0,0,8
      IF(FP(I)-FP(IL))0,7,7
      IL=I
      GO TO 7
    8 IH=I
    7 CONTINUE
C     ISH=0 REFLECTION,FIND CENTROID OF SUBSIMPLEX
      IF(ISH)0,0,30
      XN=NK-1
   50 K1=NN
       DO 9 I=1,N
      K=I
      S=0
      DO 10 J=1,N1
      IF(J-IH)0,10,0
      S=S+P(K)
   10 K=K+N
      K1=K1+1
    9 P(K1)=S/XN
      WRITE(2,71)(P(NN+I),I=1,N)
   71 FORMAT(1X,6E12.5)
  155 K=(IH-1)*N+1
      Y(NW-3)=IC
   67 IV=1
      AP=APO
   62 AP1=AP+1.
      K0=NN+1
      DO 16 I=1,N
      X(I)=AP1*P(K0)-AP*P(K)
      K=K+1
   16 K0=K0+1
      K=K-N
      CALL CALF(N,X,F,M,IV,XM,Y,NW)
      M=M+1
      IF(IV)0,61,0
      IF(XM)61,61,0
      AP=AP/2
      GO TO 62
   61 IF(F-FP(IH))64,0,0
      AP=AP/2
```

```
      IV=0
      GO TO 62
   64 CONTINUE
      DO 65 I=1,N
      P(K)=X(I)
   65 K=K+1
      K=K-N
      FP(IH)=F
      ISH=1
      GO TO 35
   30 XN=NK-1
      IF((FP(IH)-FP(IL))-EPS)100,100,0
      IF(M-MF-1)0,100,100
C          NEW    CENTROID
      K1=(IH-1)*N+1
      K2=NN+1
      DO  66 I=1,N
      P(K2)=P(K2)+(X(I)-P(K1))/XN
      K1=K1+1
   66 K2=K2+1
      IC=IC+1
      IF(IC-10)155,0,0
       IC=0
      GO TO 102
      DO 81 I=1,N
   81 XS(I)=P(NN+I)
      IV=1
      CALL  CALF(N,XS,FC,M,IV,XMC,Y,NW)
      WRITE(2,72)(P(NN+I),I=1,N),FC,XMC
   72 FORMAT(1X,1HC,3E12.5)
      WRITE(2,73)IH,IL
   73 FORMAT(1X,2I5)
      K1=(IL-1)*N
      WRITE(2,2)(P(K1+I),I=1,N),FP(IL)
    2 FORMAT(1X,3HXL=,6E12.5)
  102 CONTINUE
      WRITE(2,3)(X(I),I=1,N),XM,F
    3 FORMAT(1X,3HXR=,6E12.5)
      K1,K2=0
      DO 104 I=1,NNN
      K2=K1
      DO 105 J=I,NNN
      K=I+K2
      XS(K)=0
      DO 106 L=1,I
  106 XS(K)=XS(K)+X(K1+L)*X(K2+L)
  105 K2=K2+J
  104 K1=K1+I
      WRITE(2,3)(XS(I),I=1,N)
      GO TO 155
      K1=0
      DO 91 I=1,NK
      WRITE(2,71)(P(K1+J),J=1,N),FP(I)
   91 K1=K1+N
  100 CONTINUE
      RETURN
      END

SEGMENT, LENGTH   622, NAME  COMPSIM
```

```
      SUBROUTINE  SETUNVEC(S,NV,N,Y,NX,P)
      LOGICAL SYM
      DIMENSION T(3,3),Z(9),W(3),XX(12)
      DIMENSION P(NX)
      DIMENSION Y(4000)
C                                      SETS UNIT VECTORS FOR 3 DIMENSIONS
      PI2=3.14159267*2.0
      P2=PI2/4.
      PI=2.*P2
      II=N-1
      GO TO (20,30,40,50),II
   20 I=1
      T1=0.0
      K=-N
    4 K=K+N
      Y(K+1)=COS(T1)
      Y(K+2)=SIN(T1)
      T1=T1+S
      I=I+1
      IF(T1-PI)4,0,0
      NV=  I-1
      GO TO 50
      RETURN
   30 M=1
      GO TO 61
      K=0
      TF=PI
      T1=0.0
      IF=0
      C1=1.0
      S1=0.0
      Y(1)=1.0
      Y(2),Y(3)=0.0
    1 T2=0.0
      SS=S/COS(T1)
    2 T2=T2+SS
      K=K+N
      C2=COS(T2)
      S2=SIN(T2)
      Y(K+1)=C2*C1
      Y(K+2)=S2*C1
      Y(K+3)=S1
      M=M+1
      IF(T2-TF)2,0,0
      TF=PI2
      T1=T1+S
      C1=COS(T1)
      S1=SIN(T1)
      IF(T1-P2)1,0,0
      NV=M
      GO TO 40
      IF (IF.EQ.0) IF=M
      IFF=IF+N
      NV=M*N
      DO 3 I=1,M-IF
      K=N*(I-1)
      Y(NV+K+1)=-Y(IFF+K+1)
      Y(NV+K+2)=-Y(IFF+K+2)
    3 Y(NV+K+3)=-Y(IFF+K+3)
      NV=2*M-IF
   40 CONTINUE
```

```
   61 CONTINUE
      NP=2000
      NP=1000
      NP=1500
      NP=3000
      NP=4000
      SYM=.TRUE.
      SYM=.FALSE.
      N1=15
      N1=5
       N1=8
      N1=6
      N1=5
      N1=10
      M4=4*N
      M2=N*N
      IVS=0
      NE=3
      CALL F4ROOT1(P,N,1.0E-6,IVS,Z,W,NE,XX,1,N,1,R1,R2,M2,NX,M4)
      DO 62 J=1,N
      DO 62 I=1,N
      K=N*(N-J)+I
   62 T(I,J)=Z(K)
       B=SQRT(W(N)/W(1))
      A=SQRT(W(N)/W(2))
      XN=N1
       N2,N3=N1
      S,S1=1./XN
      S2=A/XN
      S3=B/XN
      WRITE(2,51)(W(I),I=1,N),A,B,S
      WRITE(2,52)N1,N2,N3
   52 FORMAT(/1X,4I5/)
C     BASE    OF    CYLNDR.
       K=0
      X=-1.
      DO 73 I=1,2
      DO 74 J=1,N1+1
      XJ=J-1
      Y(K+1)=X
       Y(K+2)=XJ*S2
      Y(K+3)=0.
   74 K=K+N
   73 X=1.
      DO 75 I=1,2*N1+1
      XI=I-1
      Y(K+1)=-1.+S1*XI
      Y(K+3)=0.
      Y(K+2)=A
   75 K=K+N
C     SIDES    OF    CLDR.
      DO 76 KK=2,N1
      XK=KK-1
      X=-1.
      DO 77 I=1,2
      DO 78 J=1,2*N1+1
      XJ=J-1
      Y(K+1)=X
      Y(K+2)=-A+S2*XJ
      Y(K+3)=S3*XK
   78 K=K+N
   77 X=1.
```

```
      X=-A
      DO 79 J=1,2
      DO 80 I=1,2*N1+1
      XI=I-1
       Y(K+1)=-1.+XI*S1
      Y(K+2)=X
      Y(K+3)=S3*XK
   80 K=K+N
   79 X=A
   76 CONTINUE
C     TOP    OF      CYLNDR,
      DO 81 I=1,2*N1+1
      XI=I-1
      DO 81 J=1,2*N1+1
      XJ=J-1
      Y(K+1)=-1.+S1*XI
       Y(K+2)=-A+S2*XJ
      Y(K+3)=B
   81 K=K+N
      NV=K/N
      WRITE(2,82)NV,K
   82 FORMAT(//1X,29HNO VECTORS ABOVE  X    YPLANE,I5,8HSTORAGE=,
      WRITE(2,51)(Y(J),J=1,NV*N)
      IF(NV*N LE.NP)GO TO 83
      STOP
C     FIND  UNIT   VECTORS    AND   TRANSFORM
   83 DO 90 I=1,NV
      K=N*(I-1)
      SUM=0.
      DO 91 J=1,N
      Z(J)=Y(J+K)
   91 SUM=SUM+Z(J)*Z(J)
      SUM=1./SQRT(SUM)
      DO 92 J=1,N
   92 Z(J)=Z(J)*SUM
      DO 93 L=1,N
      SUM=0.
      DO 94 J=1,N
   94 SUM=SUM+T(L,J)*Z(J)
   93 Y(K+L)=SUM
   90 CONTINUE
      RETURN
   50 CONTINUE
   51 FORMAT(1X,9F8.4)
      RETURN
      END

F SEGMENT, LENGTH   919, NAME  SETUNVEC
```

```
      SUBROUTINE VMAX1(N,NV,Y,NNV,X,NX,II,FE,SM)
      DIMENSION B(4),C(4),F(4),F2(4)
      DIMENSION Y(NNV),X(NX)
      IF (II.EQ.1) GO TO 9
C                                          SET Y
      K=-N
      DO 4 I=1,NV
      K=K+N
      DO 2 J=1,N
    2 B(J)=Y(J+K)
C                                          CALCULATE LB, STORE IN B
      CALL AMATV(X,NX,B,N)
      SUM=0.0
      DO 3 L=1,N
    3 SUM=SUM+B(L)*B(L)
      SUM=SQRT(SUM)
      DO 4 J=1,N
    4 Y(J+K)=Y(J+K)/SUM
C                                          THIS INITIALISES UNIT VECTORS
C                                          NOW FIND MAX. VDOT
    9 E=SQRT(FE)
      SM=-1.0E60
      K=-N
      DO 5 I=1,NV
      K=K+N
      DO 6 J=1,N
      B(J)=Y(J+K)*E
    6 C(J)=-B(J)
      CALL CALFX(N,B,F)
      CALL CALFX(N,C,F2)
      CALL AMATV(X,NX,B,N)
      CALL AMATV(X,NX,F2,N)
      CALL AMATV(X,NX,F,N)
C                                          LF STORED IN F
C                                          LB STORED IN B
      S1=0.0
      S2=0.
      DO 7 L=1,N
      S2=S2-F2(L)*B(L)
    7 S1=S1+F(L)*B(L)
      S1=2.*S1
      S2=2.*S2
C                                          CHECK MAX
      IF(S2-SM)13,13,0
      SM=S2
      NM=K
      AL=-1.
   13 IF(S1-SM)5,5,0
      SM=S1
      NM=K
      AL=1.
    5 CONTINUE
      AL=AL*E
      DO 12 I=1,N
   12 B(I)=Y(I+NM)*AL
      WRITE(2,11)(B(I),I=1,N),SM,FE
   11 FORMAT(6E12.5)
      II=1
      RETURN
      END
```

```
      SUBROUTINE VMAX2(N,NV,Y,NNV,X,NX,SM,E,IC)
      DIMENSION Y(NNV),X(NX)
      DIMENSION B(4),F(4),F1(4),C(4),C1(4)
      SM=-1.0E60
      K=-N
      DO 4 I=1,NV
      K=K+N
      DO 2 J=1,N
    2 B(J)=Y(J+K)
C                                         CALCULATE LB. STORE IN B
      CALL AMATV(X,NX,B,N)
      SUM=0.0
      DO 3 L=1,N
    3 SUM=SUM+B(L)*B(L)
      SUM=SQRT(SUM)
      AL=E/SUM
      DO 6 J=1,N
      C(J)=Y(J+K)*AL
    6 C1(J)=-C(J)
C                                         CALCULATE LF & STORE IN F
      CALL CALFX(N,C,F)
      CALL CALFX(N,C1,F1)
      CALL AMATV(X,NX,F,N)
      CALL AMATV(X,NX,F1,N)
      S1=0.0
      S2=0.
      DO 5 L=1,N
      S2=S2-B(L)*F1(L)
    5 S1=S1+F(L)*B(L)
      IF(AMAX1(S1,S2))12,12,0
      SM=1.
      RETURN
   12 IF(S2-SM)13,13,0
      SM=S2
      NM=K
      ALM=-AL
   13 IF(S1-SM)4,4,0
      NM=K
      ALM=AL
      SM=S1
    4 CONTINUE
      IF(IC-9)11,0,0
      DO 8 I=1,N
    8 C(I)=Y(NM+I)*ALM
      WRITE(2,7)(C(I),I=1,N)
    7 FORMAT(1X,8HMAXDOTV=,5E12.5)
   11 CONTINUE
      RETURN
      END

F SEGMENT, LENGTH   276, NAME  VMAX2
```

```
      SUBROUTINE SETA(N,NSQ,NSQ2,T,A)
      DIMENSION T(NSQ2),A(NSQ)
      K3=0
      DO 2 I=1,NSQ2
    2 T(I)=0.0
      DO 4 K=1,N
      DO 4 L=1,N
      K1=NSQ*N*(L-1)
      K3=K3+1
      DO 1 I=1,N
      K2=NSQ*(I-1)+I+(K-1)*N
    1 T(K1+K2)=A(K3)
      IF(L-K)4,0,4
      KK=0
      DO 3I=1,N
      DO 3 J=1,N
      KK=KK+1
      K2=NSQ*(J-1)+I+(K-1)*N
    3 T(K1+K2)=T(K1+K2)+A(KK)
    4 CONTINUE
      WRITE(2,5)(T(I),I=1,NSQ2)
    5 FORMAT(1X,4E12.5)
      RETURN
      END
```

```
      SUBROUTINE CALF(NX,X,F,M,IV,XM,Y,NW)
      DIMENSION X(NX),Y(NW)
      N=Y(NW-2)
      NV=Y(NW)
      E=Y(NW-1)
      NNV=N*NV
      IF(IV-1)2,0,2
      IC=Y(NW-3)
      CALL VMAX2(N,NV,Y,NNV,X,NX,XM,E,IC)
    2 K=1
      F=X(1)
      DO 1I=2,N
      K=K+I
    1 F=F*X(K)
      F=3.14159267/F
      XN=N
      F=F*F**XN
      F=-F
      RETURN
      END
```