FINAL REPORT

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
CONTRACT NO. NAS8-11264

RESEARCH IN THE GENERAL AREA OF NON-LINEAR DYNAMICAL SYSTEMS

JUNE, 1967

CENTER FOR DYNAMICAL SYSTEMS

DIVISION OF APPLIED MATHEMATICS

BROWN UNIVERSITY

CENTER FOR DYNAMICAL SYSTEMS
Division of Applied Mathematics
BROWN UNIVERSITY
Providence, Rhode Island 02912

FINAL REPORT

NATIONAL AERONAUTICS AND SPACE
ADMINISTRATION - CONTRACT NAS8-11264

PERIOD COVERED:
JUNE 8, 1965 - JUNE 8, 1967

RESEARCH IN THE GENERAL AREA OF NON-LINEAR DYNAMICAL SYSTEMS

RESEARCH INVESTIGATORS:

Henry Hermes (now at University of Colorado)
J. J. Hurt (NASA Fellow)
E. F. Infante
J. P. LaSalle, Principal Investigator
K. R. Meyer
R. K. Miller
Carlos Perello (now at Centro de Investigacion del I.P.N., Mexico)
San Wan

Control No. DCN 1-6-75-0088(1F)

June 20, 1967

INTRODUCTION

This is a final report on Contract No. NAS8-11264, Research in the General Area of Non-Linear Dynamical Systems, covering the period June 8, 1965 through June 8, 1967.

In Section I a brief account is given of the research completed which was supported in part by this contract. In almost all instances a more detailed account of this research can be found in the papers and reports contained in the Appendix. The numbers in square brackets ([]) appearing in the text refer to the papers and reprints in the Appendix.

In Section II some areas for future research which we feel are of importance and worth further investigation are indicated.

The Appendix contains all of the papers and reports that have been prepared to date on this research. Some of the papers have already been published and references to their publication are given. In other cases the papers have been accepted or submitted for publication and the journals in which they are to appear are noted.

TABLE OF CONTENTS

I

RESEARCH COMPLETED

1. Control theory

One of the main theoretical tools used to consider feedback controls is the Hamilton-Jacobi equation. The major difficulties in applying this technique are discontinuities in controls and large dimensionality. An approximation technique to overcome the first difficulty has been studied by Hermes in several papers; in particular "The equivalence and approximation of optimal control problems", J. Diff. Eqs., Vol. 1, No. 4, 1965 and "Attainable sets and generalized geodesic spheres", J. Diff. Eqs., Vol. 3, No. 2, 1967. The latter paper (see [1], Appendix), incorporating results of the former, studies properties of the set of attainability (or reachable set) for a class of nonlinear control problems formulated as contingent equations.

The attainable set $\mathfrak{A}(t)$ is defined as the set of all states attainable at time t, from a fixed initial state x_o, by trajectories of the control system using all possible controls. Particular emphasis is placed on the study of the boundary of $\mathfrak{A}(t)$, denoted $\partial\mathfrak{A}(t)$. A trajectory is an extremal if it satisfies the conditions of the maximum principle. The set $S(t)$ of points attainable by extremals at time t (i.e. the geodesic sphere of radius t) is studied in relation to $\partial\mathfrak{A}(t)$. It is shown that $\partial\mathfrak{A}(t) \subset S(t)$ for all $t \geqq 0$. (The advantage is that $S(t)$ is easier to compute). Under appropriate assumptions and for t sufficiently small, these sets are equal and $S(t)$ is an imbedded sphere. As t increases $S(t)$ may become an immersed sphere, in which case the degree of its Gauss map remains one. Also, as t increases one may encounter conjugate points, in which case $S(t)$ ceases to be immersed. Other properties and several examples are given in the paper.

A further effort in the application of functional analytic methods in control theory has been carried out by Hermes. This has resulted in the paper entitled "On the closure and convexity of attainable sets in finite and infinite dimensional spaces", scheduled to appear in Vol. 5, No. 3, SIAM J. on Control. (See[2], Appendix.) Here, the role of the weak* topologies in the Filippov existence conditions is shown, together with some preliminary results which consider the admissible controls as a given set of functions, rather than the usual case of specifying only the values a control may assume at a given time. In many ways, the former may be the more practical problem when one has a limited set of function generators.

2. Stability of systems defined over a finite time.

In cooperation with L. Weiss a comprehensive study of the stability of systems defined over a finite time interval was completed. The theory follows essentially the viewpoint of the Direct Method of Liapunov and was motivated by two purposes: the desire to bring the physical concept of stability within a mathematical framework similar to the definitions of Liapunov, and the hope of easing the difficult task of constructing Liapunov functions. The large majority of physical systems operate or are observed for only finite periods of time; yet the definitions of stability of Liapunov are based on the assumption that it is possible to observe a system for an infinite time. Secondly, the stability definitions of Liapunov do not correspond precisely to the intuitive concept of stability for dynamical systems. The principal purpose of the work of Infante and Weiss was to define a stability concept which would circumvent these two objections and to establish a set of theorems that parallels those of the classical theory. The results of this investigation can be found in [3] of the Appendix.

3. Stability of a singular point of a control system.

In the theory of control systems for which the mathematical model is a system of differential equations

$$\dot{x} = f(x,u(x)) = F(x),$$

x is an n-vector (the error in control), u is the control law and is a function on R^n to R^r, and f is a function on R^{n+r} to R^n. It is often possible, using Liapunov theory (or, equivalently, dynamic programming) to determine control laws that stabilize the system. If the origin is an equilibrium point, then one has stability in the sense of Liapunov and the error $x(t) \to 0$ as $t \to \infty$.

In many cases, however, one obtains control laws that reduce the error to zero in finite time. The control then must be discontinuous at the origin, and the origin is a singular point and is not an equilibrium state. Liapunov's definitions and his classical theorems have no meaning here, and it seems worth while to extend the definitions of stability and the basic Liapunov theorems to cover the stability of singular points. LaSalle has done this and from it has obtained sufficient conditions for optimal control. A brief account of this has appeared in notes prepared by LaSalle on lectures he gave last summer on control theory.

4. An extension of Liapunov's Direct Method. An invariance principle for ordinary differential equations.

LaSalle has completed a unification and extension of Liapunov's

Direct Method for the study of the stability systems defined by ordinary differential equations. The theory shows first of all why the study of the stability of nonautonomous (time-varying) systems is inherently more difficult than that of autonomous (stationary) systems. What is exploited in this new theory is the invariance property of the limit sets of solutions of autonomous systems. This makes possible the introduction of a more general concept of a Liapunov function and in terms of such Liapunov functions there is one fundamental theorem that includes all of the classical Liapunov theorems on stability and instability as well as more recent extensions of those theorems.

Besides giving a complete unity to the theory and identifying the essential nature of a Liapunov function, this new theory has also some practical consequences. Here Liapunov functions are not, for instance, required to be positive definite and this means that the class of suitable Liapunov functions is greatly enlarged and this eases the problem of constructing Liapunov functions. Examples show that rather precise information about asymptotic stability can be obtained by functions which are not positive definite. This also makes application easier for outside of quadratic forms we have no computable criterion for determining when a function is positive definite.

Of more significance may also be the fact that this new theory for ordinary differential equations has already guided Hale in providing an adequate stability theory for functional differential equations (delay differential equations ; for example, see Hale, J., Sufficient conditions for stability and instability of autonomous functional differential equations, J. Diff. Eqs. 1, 452-482 (1966)) and Hale and Infante in their search for a stability theory for partial differential equations. (See paragraph 6 below and [6] in Appendix). For a more complete account of this research see [4] and [5], Appendix.

A complete paper on this subject is being prepared for publication.

5. An invariance principle for difference equations with applications to numerical analysis.

Hurt developed a theory of stability similar to paragraph 4 above for difference equations. To illustrate the application of this theory he derived a region of convergence for the Newton-Raphson and secant iteration methods. A modification of one of these theorems is also given and is applied to study the effect of round-off errors in the Newton-Raphson and Gauss-Seidel iteration methods. The theory is also used to derive a well-known necessary and sufficient condition that spectral radius of a matrix be less than one. This theory should also have applications to the study of the stability of sampled data systems. See [16], Appendix.

6. An invariance principle for partial differential equations and generalized dynamical systems.

For dynamical systems described by functional differential equations and partial differential equations the state space is no longer, as is the case for ordinary differential equations, a finite dimensional Euclidean space but is an infinite dimensional function space. This raises the difficulty of compactness (bounded sets are not necessarily compact nor is the space locally compact) and there is also the additional difficulty that solutions while unique in the future may not be unique in the past. This means the flow in state space defined by solutions will not in general define a group of motions but only a semi-group. It is for these reasons that the stability theories for "generalized" dynamical systems (Zubov, and Auslander and Seibert, for example) have failed to be satisfactory for either functional differential or partial differential equations. Zubov came closest but seems neither to have been aware of an "invariance principle" nor did he solve the difficulties concerning the notion of limit sets and their invariance.

Hale and Infante have overcome these difficulties and achieve an invariance principle for generalized dynamical systems that yields satisfactory stability theory with broad applications. Furthermore, this general theory makes it possible to bring to bear on these problems the well developed concepts and theorems about Sobolev spaces. A publication will appear shortly on this subject. A preliminary version of this paper is included in the Appendix, [6].

7. Stability of linear time-varying systems.

A considerable amount of work has been devoted during the past ten years to the study of the stability of nonlinear systems. Much less effort has been expended on the study of the stability of systems whose parameters vary with time in an imprecisely known manner. Infante has considered two different types of problems along this line. In one, stability conditions are obtained for a linear system with time-varying coefficients under the assumption that the range of the magnitude of these coefficients is known. The results obtained generalize the well-known circle criterion. In a second problem it is assumed that the expectation of the value of the coefficients is known and stability conditions are obtained by imposing restrictions on the expectations of the variations of the coefficients. The techniques used in this second problem are equally applicable to stochastic and to deterministic problems. The results of these investigations have partly appeared in one publication (see [7], Appendix) and further results have been submitted for a second paper (see [8], Appendix).

8. The theory of matroids with applications to electrical network theory.

A complete exposition of the application of matroids to certain problems in graph theory, and in particular to the theory of electrical networks and network flows, has been completed. No basically new results were obtained, but the theory was shown not only to be an elegant way of approaching network theory but was also shown to be a promising tool in the solution of certain problems in automata theory which are of a combinatorial nature. Some of this work will appear in a monograph on graph theory principally authored by Professor S. Lefschetz.

9. Global Liapunov functions for Morse-Smale systems.

Meyer has considered the problem of global Liapunov functions (or energy functions) for Morse-Smale systems. He has been able to construct global Liapunov functions for any Morse-Smale system and also show that in a sense these functions are unique. In the two-dimensional case one gets a one-to-one correspondence between topological equivalence classes of structurally stable fields and energy functions. See [9], Appendix.

10. Level curves of Liapunov functions.

Miller has studied the problem of characterizing the level curves of Liapunov functions for a nonlinear autonomous ordinary differential equation in the neighborhood of an equilibrium point. If the dimension of the system of equations is greater than seven it is shown that a level curve plus its interior is diffeomorphic to a disk.

11. Asymptotic equivalence.

Miller has studied the behavior of solutions of a certain nonlinear second order perturbed ordinary differential equation. Under certain assumptions he obtains results on the asymptotic equivalence of the perturbed and the unperturbed equations.

12. Volterra integral equations.

Miller and G. R. Sell of the University of Minnesota are jointly studying some problems in the theory of dynamical systems. They are studying the qualitative behavior of a class of nonlinear Volterra integral equations by formulating these problems in the language of dynamical systems. In order to do this they have the problem of obtaining the necessary theorems on existence, uniqueness and continuity with respect to parameters. (See [17])

Miller has been studying the asymptotic behavior of solutions of a nonlinear convolution type Volterra integral equation. He has obtained theorems which justify linearization of these equations near critical points. Under certain stability assumptions on the linear equations one obtains local stability results for the nonlinear system. This theory has been applied in some specific examples. See [10], Appendix.

13. Periodic solutions.

Recently Smale has introduced a class of vector fields on a manifold that hopefully will play the role in n-dimensions that structurally stable fields play in two dimensions. Meyer has been able to give an estimate for the number of periodic solutions these systems can have in the case when the fields admit a global cross section. He is now working on the general case. See [11], Appendix.

14. Functional integral equations.

Meyer (with J. K. Hale) has studied a very general class of linear functional integral equations which arise as a natural generalization of neutral differential difference equations and pure difference equations. The aim of this research was to carry forth the study of the linear equation to a sufficient degree that the usual theorems for weakly nonlinear ordinary differential equations could be proved for the corresponding weakly nonlinear functional integral equation in an analogous way. Indeed the theorems on stability by the first approximation and integral manifolds were established using the developed linear theory.

The basic elements of the linear theory that were discussed were (1) the variation of constants formula, (2) the decomposition of the space into invarient subspaces (eigen spaces) and (3) sharp exponential bounds on the growth of solutions on these invarient subspaces. See [12], Appendix.

15. Contact transformations.

Meyer has written a short note on contact transformations and generating functions. It is often quoted in the literature that not all contact transformations can be derived from a generating function. Meyer shows that if one first makes a linear orthogonal change of variables then any contact transformation can be written as a transformation arising from a generating function. See [13], Appendix.

16. Geometric theory of ordinary differential equations.

Poincare gave a simple geometric procedure for computing the index of a critical point in the plane. Meyer has given a simple geometric procedure which generalizes this formula which will allow one to compute the index in dimensions 2, 3 and 4. See [14], Appendix.

17. Periodic solutions of difference differential equations.

Perello has shown that the method of Cesari and Hale for the study of periodic solutions can be extended to difference differential equations. An application of the result is made to the study of a control system with a delay in the feedback. See [15], Appendix.

II

SOME IMPORTANT AREAS FOR FUTURE RESEARCH

1. Stability of time-varying systems.

Our knowledge of the stability of linear and nonlinear time-varying systems is at the present time at about the same stage of development as the Lur'e Problem was five years ago. The well-known circle criterion needs to be further generalized, but especially it would seem appropriate to develop stability criteria that depend on the whole history of the time-varying coefficients and not just on their instantaneous behavior. Very little effort has been made to go in this direction, and yet it would appear that this is the manner in which the time-varying problem should be formulated.

2. Stability of dynamical systems defined by partial differential equations.

Up to the present time there has not been a completely satisfactory Liapunov type stability theory for partial differential equations, and it is our hope that the research reported above in I - 6 will be a "break-through". A great deal remains to be done in the way of theory but even more attention needs to be paid to applications to such problems as the stability of structures, of fluid flows, of the oscillations of plasmas and transmission lines, etc. This seems to us a most fruitful area of research.

3. Dynamic programming via Liapunov's Direct Method.

That a relationship exists between dynamic programming, Pontryagin's maximum principle and Liapunov's direct method is fairly well known. The result reported in I-3 above was motivated by this relationship and provides a sufficient condition for optimal control over a finite interval of time.

The invariance principle has now provided a general stability theory for ordinary differential equations, difference equations, functional differential equations, and now partial differential equations.

It therefore seems natural to exploit both this new knowledge and the known relationship between optimal control and Liapunov's method. For instance, it is not clear that optimal control problems have been properly formulated for functional differential equations and for partial differential equations. This is the first thing that we would expect to learn something about.

4. Choice of performance criterion.

Given a control system and a class of admissible controls u the set of controllability Γ is the set of initial states from which the system can be brought to the origin (zero control error) in finite time. Define Σ, called the set of stabilization, to be the set of all initial states for which there is a control such that the system reaches the origin in finite time or approaches the origin as $t \to \infty$. The system is said to be controllable if the origin is an interior point of Γ and stabilizable if the origin is an interior point of Σ. It is obvious that if a system is controllable then $\Gamma = \Sigma$. A system can, however, be stabilizable without being controllable.

It is known that if a performance criterion is properly selected then optimality (assuming there is an optimal "feedback" control $u^o(x)$) implies stability. The choice of the performance criterion will, in general, affect the size and shape of the region R^o of asymptotic stability. What is the relation between R^o, Σ, and A? Does this give us a way of judging "good" performance criteria? For example, we might say it is "good" only if $R^o = \Sigma$ or perhaps require only that R^o contain some preassigned set Ω.

Letov states in his paper presented at the 1966 IFAC Congress in London that "the main difficulties encountered by engineers (in the analytic design of optimal controllers) consist in the choice of a desirable optimizing functional." Letov then attempts to formulate the proper choice of a "payoff" function as a problem in the theory of control.

In any case this is an interesting problem and the above approach (less ambitious than Letov's) might shed some light on the choice of performance criteria and the relative merits of different criteria. The first step will be to construct some simple examples to show that the questions being asked are meaningful. Some work was done by San Wan under the direction of LaSalle on this problem but we were unable to devote sufficient effort to it. This is an important problem and its investigation should be renewed.

APPENDIX - PUBLICATIONS

TABLE OF CONTENTS

PAPER [1]

Journal of Differential Equations, 3(1967), 256-270

ATTAINABLE SETS AND GENERALIZED GEODESIC SPHERES

by H. Hermes*

INTRODUCTION

For each point $(t,x) \in E^1 \times E^n$ let $R(t,x)$ be a nonempty compact subset of E^n. (E^n denotes Euclidean real n-dimensional space.) A contingent equation has the form

$$(1) \qquad \dot{x}(t) \in R(t,x(t)) \quad , \quad x(0) = x^0 \quad ; \quad (\dot{x}(t) = \frac{dx(t)}{dt})$$

a solution is any absolutely continuous function φ such that $\varphi(0) = x^0$, $\dot{\varphi}(t) \in R(t,\varphi(t))$ for almost all t. The attainable set at time $t \geq 0$ for (1) is defined as

$$\mathcal{Q}(t) = \{\varphi(t): \varphi \text{ is a solution of (1)}\} .$$

We shall be concerned with the study of $\mathcal{Q}(t)$, its topological boundary which will be denoted $\partial\mathcal{Q}(t)$, and a set $S(t)$ which is related to $\partial\mathcal{Q}(t)$ and can be thought of as a geodesic sphere of

*Research supported in part by the National Aeronautics and Space Administration under Contract No. NAS8-11264 and in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant Nr. 693-65.

radius t.

To clarify the last statement, let $G(x)$ be a symmetric positive definite matrix valued function and $F(x,r) = [r'G(x)r]^{\frac{1}{2}}$; the prime denoting transpose. Then, for each x, $F(x,\cdot)$ determines a Riemannian metric structure on the tangent space to E^n at x, which gives rise to the standard geodesic problem on the manifold E^n. This is the usual case of Riemannian geometry and $S(t)$ would be the geodesic sphere of radius t. If, instead, $F(x,\cdot)$ is the support or Minkowski functional of a strictly convex set, it determines a Minkowski metric geometry in the tangent space at x. This gives rise to the geodesic problem in a Finsler space [1,p.82] and again $S(t)$ would be the geodesic sphere of radius t. In either of these cases the function F would determine an autonomous equation of the form (1) by defining $R(x) = \{r \in E^n: F(x,r) \leq 1\}$. Conversely if we were given a set valued function $R(x)$ which, for each x, was an ellipsoid centered about the origin, it could be used to construct $F(x,\cdot)$ and hence induce a Riemannian metric structure in the tangent space at x.

From time optimal control theory, there is a natural notion of extremal for the equation (1) even when R does not have so special a form as to induce a metric in the tangent spaces. In this case $S(t)$ will again denote the set of all points which are attained at time t by an extremal initiating from x^o at time zero. Thus we introduce the nomenclature generalized geodesic

sphere.

To see the relation between equation (1) and time optimal control theory, consider the controlled system of differential equations.

$$(2) \qquad \dot{x}(t) = f(t,x(t),u(t)) \qquad , \qquad x(0) = x^0$$

where $x(t) \in E^n$, $f \in C^1$ (once continuously differentiable) with u, the control function, belonging to $\Omega = \{u : u \text{ measurable}, u(t) \in U\}$ where U is a compact subset of E^n. Given a target, i.e. a continuous function $z: [0,\infty) \to E^n$, a time optimal control problem would be to find a $u \in \Omega$ which "steers" the solution of (2) to the target z in minimum time. Here, in a formulation similar to (1), $R(t,x) = \{f(t,x,u) : u \in U\}$ while a solution is possible only if $\mathcal{Q}(t) \cap \{z(t)\} \neq \emptyset$, the empty set, for some $t \geqq 0$. If this is satisfied, the existence of an optimal control depends on the compactness of $\mathcal{Q}(t)$; in the formulation of our problem in §1, we shall impose conditions on $R(t,x)$ which insure this. It is natural, then, to take as extremals arcs which satisfy the maximum principle [2].

Another problem which fits within the scope of our formulation is that of finite time stability for systems of differential equations experiencing persistent perturbations. Again, consider the equations (2), but now it is more natural to take $f(t,x,u) = g(t,x) + u$ with U a compact subset of E^n of the form

$U = \{y : |y| \leq \epsilon\}$ where ϵ measures the maximum amplitude of the persistent perturbation u. Given a $T > 0$, the problem is to obtain bounds at time T for the possible differences in the perturbed and unperturbed $(u \equiv 0)$ solutions. Obviously, a precise knowledge of $\mathcal{Q}(T)$ would give complete information for this problem.

In §1 conditions will be imposed on the set valued function R which make a reasonable amount of analysis of $\mathcal{Q}(t)$, $\partial\mathcal{Q}(t)$ and $S(t)$ possible. In §2, properties of these sets, the meaning of conjugate points, and examples are given and a start is made on the problem of classifying generalized geodesic spheres.

The author would like to acknowledge many helpful discussions with Drs. J. McAlpin and F. W. Wilson and the help in the numerical computations received from Mr. J. Hurt.

§1. Formulation of the Problem; Properties of R.

We wish to formulate (1) in such a way that for each $t \geq 0$, $\mathcal{Q}(t)$ is nonempty, compact, and $\mathcal{Q}(\cdot)$ is continuous when considered as a set valued function in the Hausdorff metric topology. From results obtained by Filippov [3], these properties of $\mathcal{Q}$ will follow if R is continuous as a set valued function in the Hausdorff topology; for each t,x, $R(t,x)$ is convex, and there exists a $c > 0$ such that for any function $r(t,x)$ with values in $R(t,x)$,

the inner product $(r(t,x),x) \leq c[1 + |x|^2]$.

In order to deal analytically with $\partial\mathcal{Q}(t)$ further restrictions are needed. These are motivated by the maximum principle which can be formulated as follows for a time optimal control problem associated with (1). Define $H(p,r) = (p,r)$ for $r \in R(t,x)$, $p \in E^n-\{0\}$ and

$$(3) \qquad H^*(t,x,p) = \max\{(p,r) : r \in R(t,x)\} .$$

Then a necessary condition that a solution φ be optimal (for some problem) is that there exist an absolutely continuous function ψ so that φ, ψ satisfy, respectively, the equations

$$(4) \qquad \begin{aligned} \dot{x}(t) &= \frac{\partial}{\partial p} H^*(t,x,p) \qquad && x(0) = x^0 \\ \dot{p}(t) &= -\frac{\partial}{\partial x} H^*(t,x,p) \qquad && p(0) \in S^{n-1} \end{aligned}$$

where S^{n-1} denotes the unit $n-1$ sphere. Intuitively this is a necessary condition that $\varphi(t)$ belong to the boundary of $\mathcal{Q}(t)$.

With the proper conditions on $R(t,x)$, the algebraic maximization which produces H^* will define a "smooth" function $r^*(t,x,p)$ such that $H^*(t,x,p) = (p,r^*(t,x,p))$. We will want r^* to be continuous in t, once continuously differentiable in p and twice continuously differentiable in x. The continuity in t

and x will follow from continuity conditions on the set valued function R; the continuity with respect to p, however, depends completely on the "shape" of the set $R(t,x)$. It is shown in [4] that if $R(t,x)$ is strictly convex and contains more than a single point, $r^*(t,x,p)$ is continuous in p. Furthermore, assume $R(t,x)$ is strictly convex and the Gauss map $\eta : \partial R(t,x) \to S^{n-1}$, which associates with each point on $\partial R(t,x)$ the unit outward normal at that point, is well defined and continuously differentiable in terms of local coordinates. Then since $r^*(t,x,p)$ maximizes (p,r) for $r \in R(t,x)$ it is clear that r^* is defined implicitly by the requirement $\eta(r^*(t,x,p)) = p/|p|$. Hence defining $G(r,p) = \eta(r)-p/|p|$ for $r \in \partial R(t,x)$, $p \in E^n-\{0\}$, the implicit function theorem applied to $G(r,p) = 0$ will yield a function $r^*(t,x,p)$ which is C^1 in p and satisfies $G(r^*(t,x,p),p) \equiv 0$ if the Jacobian matrix $G_r(r,p)$ has the required rank. The requirement on the rank of this Jacobian, see [4], can be shown to be equivalent to the condition that the Gauss map, in terms of local coordinates on $\partial R(t,x)$, have non-vanishing Jacobian.

With the previous conditions and their implications in mind, we shall now give a precise representation and formulation of properties which will be assumed for $R(t,x)$.

Let $\Omega(t,x,r)$ be a real valued C^2 function defined on $E^1 \times E^n \times E^n$ which satisfies

(5) $\Omega_{rr}(t,x,r)$ is a positive definite matrix

(6) $\Omega(t,x,0) = 0$.

Define

$$Q(t,x) = \{r \in E^n : \Omega(t,x,r) \leq 1\} .$$

Then $Q(t,x)$ is nonempty, compact, and Q is continuous as a set valued function in the Hausdorff topology. Property (5) implies the second fundamental form, in terms of local coordinates on $\partial Q(t,x)$, is definite. But the second fundamental form is a representation of the differential of the Gauss map. The strict convexity and nonvanishing of the Gauss map are thereby implied, yielding the desired continuity properties of the function r^* which maximizes (p,r) for $r \in Q(t,x)$.

From (6), we see zero always belongs to $Q(t,x)$, a condition which need not be imposed on $R(t,x)$. Let g: $E^1 \times E^n \to E^n$ be a C^2 function and define

$$R(t,x) = \{g(t,x) + r : r \in Q(t,x)\} . \tag{7}$$

In what follows, it will always be assumed that R admits a representation as in (7) and that there exists a $c > 0$ such that $(g(t,x)+r,x) \leq c[1+|x|^2]$ for any $r \in R(t,x)$; i.e. the corresponding trajectories of (1) will not escape in finite time.

Actually, this formulation is quite general. It contains Finsler geometry (and therefore Riemannian geometry) as the case $g \equiv 0$ and Ω independent of t. (Compare [1,p.84]). Also, it is shown in [4] that any time optimal control problem which satisfies the Filippov existence conditions can be approximated arbitrarily closely (in the sense that solutions of the approximating problem are uniformly close to those of the original problem) by a problem with R of the form considered in (7).

From the maximization it follows that $r^*(t,x,p)$ is that unique point on $\partial Q(t,x)$ where the outward normal has the direction p, i.e. $\Omega_r(t,x,r^*(t,x,p)) = kp$ for some $k > 0$. Then, since $\Omega(t,x,r^*(t,x,p)) \equiv 1$, $\Omega_r r^*_p = 0$ or

$$(8) \qquad p r^*_p(t,x,p) = 0.$$

(We will not use primes to designate transpose of a vector when this is obvious from its placement.) The equations (4) now become

$$(9) \qquad \dot{x}(t) = g(t,x) + r^*(t,x,p) \quad , \qquad x(0) = x^0$$

$$(10) \qquad \dot{p}(t) = -p[g_x(t,x) + r^*_x(t,x,p)] \; , \qquad p(0) \in S^{n-1}.$$

The formulation is such that the right sides of these equations are C^1 hence they can be used constructively rather than to just state

necessary conditions. Also, it suffices to consider $p(0) \in S^{n-1}$ since from its definition $r^*(t,x,\alpha p) = r^*(t,x,p)$ for $\alpha > 0$ hence if ψ is a solution of (10) so is $\alpha\psi$. Now let $\xi = (\xi_1, \ldots \xi_{n-1})$ be local coordinates on S^{n-1}; for any $p(0) = \xi \in S^{n-1}$ the equations (9) and (10) have unique solutions, denoted $\varphi(\cdot,\xi)$, $\psi(\cdot,\xi)$ respectively. Define

$$S(t) = \{\varphi(t,\xi) : \xi \in S^{n-1}\} . \tag{11}$$

For each $\xi \in S^{n-1}$, $\varphi(\cdot,\xi)$ is an extremal in the sense that it satisfies the maximum principle. We may also consider $\varphi(\cdot,\xi)$ as playing the equivalent role of the exponential map in the classical geodesic problem. Since the right sides of (9),(10) are C^1, solutions are differentiable with respect to initial data. Geometrically $S(t)$ may be viewed as the projection, onto the first n coordinates, of the diffeomorphic image of S^{n-1} under the flow of (9),(10) in E^{2n}.

For later use it will be convenient to have an equivalent representation of $R(t,x)$ of the form

$$R(t,x) = \{g(t,x) + f(t,x,u) : |u| \leq 1\} . \tag{12}$$

This is easily obtained as follows. Let $\rho(t,x,\cdot)$ be the support function of $Q(t,x)$. [Note: Ω was not required to satisfy

$\Omega(t,x,\alpha r) = \alpha\,\Omega(t,x,r)$ for $\alpha > 0$ and therefore need not be a support function.] Define $f(t,x,u) = u/\rho(t,x,u/|u|)$ for $u \neq 0$ and $f(t,x,0) = 0$. This f yields the representation (12) for $R(t,x)$ and satisfies the continuity conditions needed for the maximum principle.

§2. Properties of $\mathcal{A}(t)$, $\partial\mathcal{A}(t)$ and $S(t)$.

Let $\mathcal{F}(t) = \cup_{0\leq\tau\leq t}\mathcal{A}(\tau)$; this is sometimes referred to as the attainable funnel. The first three properties are immediate consequences of the problem formulation and results of Filippov [3],[5].

Property 1 For each $t \geq 0$, $\mathcal{A}(t)$ is a nonempty compact set.

Property 2 The set valued function $\mathcal{A}(\cdot)$ is continuous in the Hausdorff metric topology.

Property 3 For each $t \geq 0$, $\mathcal{F}(t)$ is a compact set in $E^1 \times E^n$.

Property 4 For each $t_1 \geq 0$, $\mathcal{A}(t_1)$ is arcwise connected.

Proof: Let φ^0 and φ^1 be any two solutions of $\dot{x} \in R(t,x)$, $x(0) = x^0$. Using the representation (12) we have for almost all $t \in [0,t_1]$ and $i = 0,1$, $\dot{\varphi}^i(t) = g(t,\varphi^i(t))+f(t,\varphi^i(t),u^i(t))$ for some $|u^i(t)| \leq 1$. By a lemma of Filippov [3], the functions u^i may be assumed measurable. Now for each $\alpha \in [0,1]$ define u^α

by $u^\alpha(t) = \alpha u^1(t)+(1-\alpha)u^o(t)$. Then u^α is measurable, $|u^\alpha(t)| \leq 1$ and the equation $\dot{x} = g(t,x)+f(t,x,u^\alpha)$, $x(0) = x^o$ has a unique solution; denote it by φ^α. From the continuity properties which the solution possesses with respect to parameters, as α varies continuously from 0 to 1, $\varphi^\alpha(t_1)$ traces out a continuous arc joining $\varphi^o(t_1)$ to $\varphi^1(t_1)$ in $\mathcal{Q}(t_1)$.

<u>EXAMPLE 1.</u> Consider the following two dimensional problem. $\dot{x}(t) \in R(x(t))$, $x(0) = (-1,0)$ where $R(x) = \{r \in E^2 : |r| \leq |x|\}$. Here $\Omega(x,r) = |r|/|x|$; we may either consider $x = 0$ not in the domain of definition, or define $R(0) = \{0\}$. Here we deal with Riemannian geometry since $R(x)$ determines the metric $[\dot{x}G(x)\dot{x}]^{\frac{1}{2}}$ where $G(x) = \begin{pmatrix} 1/|x|^2 & 0 \\ 0 & 1/|x|^2 \end{pmatrix}$.

On the unit circle $|x| = 1$, $R(x)$ is a unit disc implying that it is possible to traverse the unit circle, with unit velocity in either direction. Also, the point $x = 0$ is not attainable from $(-1,0)$ in finite time. Thus for t_1 slightly larger that π one expects $\mathcal{Q}(t_1)$ to look as follows.

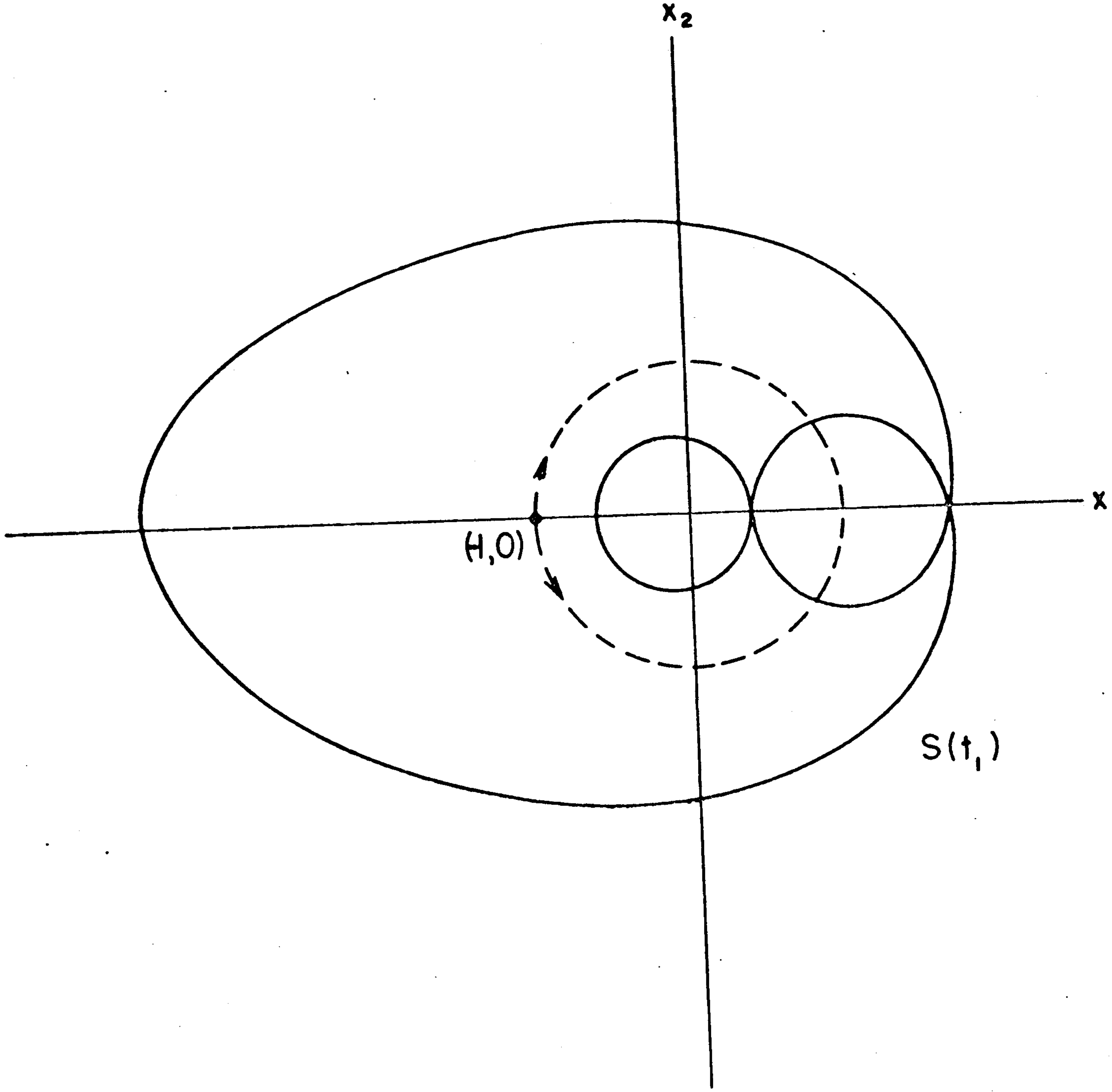


FIGURE 1

Here $S(t_1)$ is an immersed sphere, $\mathcal{Q}(t_1)$ is not simply connected and the topological boundary of $\mathcal{Q}(t_1)$, i.e. $\partial\mathcal{Q}(t_1)$, is a proper subset of $S(t_1)$.

The time optimal point to point transfer problem for $\dot{x} \in R(x)$ is equivalent to the geodesic problem on the manifold determined by $G(x)$. Indeed, since a geodesic will satisfy $|\dot{x}(t)| = |x(t)|$, its length will be $\int_0^t [\dot{x}G(x)\dot{x}]^{\frac{1}{2}} d\tau = t$.

The previous example shows that $\mathcal{Q}(t)$ need not be simply connected; insight into how this can occur with increasing time may be gained from the following.

<u>Property 5</u> <u>For any</u> $t_1 > 0$, $\partial\mathcal{Q}(t_1) \subset S(t_1)$ <u>and for every point</u> $y \in \partial\mathcal{Q}(t_1)$ <u>there exists a continuous function</u> $z : [0,t_1] \to E^n$ <u>such that</u> $z(t_1) = y$ <u>and</u> $\mathcal{Q}(t) \cap \{z(t)\} = \emptyset$ <u>for</u> $0 \leq t < t_1$.

Proof: Assume $y \in \partial\mathcal{Q}(t_1)$ and $x(\cdot)$ is a solution of $\dot{x} \in R(t,x)$, $x(0) = x^o$, such that $x(t_1) = y$. Suppose there exists no arc $z : [0,t_1] \to E^n$ with $z(t_1) = y$, $\mathcal{Q}(t) \cap \{z(t)\} = \emptyset$ for $0 \leq t < t_1$. Then for some $t' \in [0,t_1)$ we must have $\mathcal{Q}(t')$ has a non empty n dimensional interior to which $x(t')$ belongs. Indeed if $x(t) \in \partial\mathcal{Q}(t)$ for all $t \in [0,t_1]$, using the compactness of the funnel $\mathcal{F}(t_1)$, it is easy to construct a continuous function z, with values $z(t)$ in a neighborhood of $x(t)$, such that $\{(t,z(t)) : 0 \leq t \leq t_1\}$ and $\mathcal{F}(t_1)$ have only the point (t_1,y)

in common.

Now, using the representation (12) and the lemma of Filippov [3], we may assume the existence of a measurable function u^*, $|u^*(t)| \leq 1$, such that $\dot{x}(t) = g(t,x(t))+f(t,x(t),u^*(t))$ almost everywhere. Since $x(t')$ is in the interior of $\mathcal{Q}(t')$, there is some neighborhood $N(x(t'))$ contained in $\mathcal{Q}(t')$. Keeping u^* fixed and using points in $N(x(t'))$ as data at time t' for the differential equation $\dot{x} = f(t,x,u^*)$, the solutions evaluated at time t_1 provide a neighborhood of y which must belong to $\mathcal{Q}(t_1)$. This contradicts the assumption $y \in \partial\mathcal{Q}(t_1)$.

We next show that $\partial\mathcal{Q}(t_1) \subset S(t_1)$. Let $y \in \partial\mathcal{Q}(t_1)$; by the result obtained in the first part of this proof there exists a continuous function $z : [0,t_1] \to E^n$ such that the problem of hitting z in minimum time by a solution of $\dot{x} \in R(t,x)$, $x(0) = x^o$, has a solution with the optimal time being t_1 and the intercept occuring at the point y. The maximum principle, which is a necessary condition, then shows that if φ is the solution, there exists a function ψ such that the pair φ,ψ satisfy equations (9) and (10) respectively for some $\psi(0) \in S^{n-1}$. Thus $y \in S(t_1)$.

Property 5 shows that for every point on $\partial\mathcal{Q}(t)$, there is a time optimal control problem for which the optimal interception occurs at that point.

<u>Property 6</u> <u>In</u> E^n, <u>for</u> $t_1 > 0$ <u>and sufficiently small</u>, $S(t_1)$ <u>is the diffeomorphic image of</u> S^{n-1}(i.e. $\varphi(t_1,\cdot)$ <u>is</u>

an imbedding of $S^{n-1} \to E^n$) ; $\psi(t_1,\xi)$ *is an outward normal to* $S(t_1)$ *at* $\varphi(t_1,\xi)$; $\mathcal{Q}(t_1)$ *is an* n *disc and* $\partial\mathcal{Q}(t_1) = S(t_1)$.

Proof: Remembering that $\xi_1,\ldots,\xi_{n-1}$ are local coordinates on S^{n-1}, let $\overline{p}(\xi)$ be the inclusion map of $S^{n-1} \to E^n$. We shall adopt the notation that for any vector function (row or column) $v(\xi)$, v_ξ denotes the matrix of partial derivatives $(v^i_{\xi_j}(\xi))$.

For $\varphi(t,\xi) \in S(t)$,

$$\varphi(t,\xi) = x^o + \int_o^t [g(\tau,\varphi(\tau,\xi)) + r^*(\tau,\varphi(\tau,\xi),\psi(\tau,\xi))]d\tau \ ;$$

(13) $$\varphi_\xi(t,\xi) = \int_o^t [(g_x + r^*_x)\varphi_\xi(\tau,\xi) + r^*_p\psi_\xi]d\tau \ ;$$

$\varphi_\xi(0,\xi) = 0$. If φ_ξ has rank $n-1$ its rows span the tangent space of $S(t)$ at the point $\varphi(t,\xi)$. Since $\varphi_\xi(0,\xi) = 0$ and the Gauss map having nonvanishing Jacobian determinant on $\partial R(0,x^o)$ implies $r^*_p(0,\varphi(0,\xi),\psi(0,\xi))\psi_\xi(0,\xi)$ has rank $n-1$, it follows from observation of the integrand in (13) that for t sufficiently small, but positive, $\varphi_\xi(t,\xi)$ has rank $n-1$.

Also, the continuity properties allow equation (13) to be differentiated with respect to t, showing that $\varphi_\xi(\cdot,\xi)$ satisfies the matrix differential equation

$$(14)\qquad \dot{\varphi}_{\xi}(t,\xi) = [g_x(t,\varphi(t,\xi))+r_x^*(t,\varphi(t,\xi),\psi(t,\xi))]\varphi_{\xi}(t,\xi)+r_p^*\psi_{\xi}$$

with data $\varphi_{\xi}(0,\xi) = 0$.

Let $\Psi(\cdot,\xi)$ denote a fundamental solution matrix to $\dot{p} = -p[g_x(t,\varphi(t,\xi))+r_x^*(t,\varphi(t,\xi),\psi(t,\xi))]$; then $\psi(t,\xi)$ can be written as $\psi(t,\xi) = \overline{p}(\xi)\Psi(t,\xi)$ and we have the representation

$$(15)\qquad \varphi_{\xi}(t,\xi) = \int_0^t \Psi^{-1}(t,\xi)\Psi(\tau,\xi)r_p^*(\tau,\varphi(\tau,\xi),\psi(\tau,\xi))\psi_{\xi}(\tau,\xi)d\tau.$$

Multiplying both sides of this by $\overline{p}(\xi)\Psi(t,\xi)$ on the left and using (8) which shows that $\psi(t,\xi)r_p^*(t,\varphi(t,\xi),\psi(t,\xi)) \equiv 0$, we get

$$(16)\qquad \psi(t,\xi)\varphi_{\xi}(t,\xi) = 0 \ .$$

Now this holds for all t, even if rank $\varphi_{\xi}(t,\xi) < n-1$. In particular, if t_1 is sufficiently small so that rank $\varphi_{\xi}(t_1,\xi) = n-1$ the rows of $\varphi_{\xi}(t_1,\xi)$ span the tangent space to $S(t_1)$ at $\varphi(t_1,\xi)$ and (1) shows $\psi(t_1,\xi)$ is a normal to $S(t_1)$ at $\varphi(t_1,\xi)$.

We have shown that for each $t_1 > 0$ and sufficiently small, $\varphi(t_1,\cdot)$, as a map of $S^{n-1} \to E^n$, is regular (i.e. a C^1 map with Jacobian of rank $n-1$) and therefore $S(t_1)$ is an immersed sphere. We must show the mapping is globally one-one

(or that $S(t_1)$ is the homeomorphic image of a sphere) in order to conclude that $\varphi(t_1,\cdot)$ is actually an imbedding and $S(t_1)$ an imbedded sphere.

From our formulation, $\psi(0,\cdot)$ is the inclusion map of S^{n-1} into E^n. From (10) one sees that for initial data $p(0) = 0$, $p \equiv 0$ is a solution; from uniqueness it follows that for all $t \geq 0$ and $\xi \in S^{n-1}$, $|\psi(t,\xi)| \neq 0$. For each $t \geq 0$ define

(17) $$n(t,\cdot)\colon S^{n-1} \to S^{n-1} \text{ by } n(t,\xi) = \psi(t,\xi)/|\psi(t,\xi)|.$$

Before proceeding with the remainder of the proof of property 6, we shall need

LEMMA 1. *For each* $t \geq 0$, *the degree of the map* $n(t,\cdot)$ *is one.*

Proof: $n(0,\cdot)$ is the identity on S^{n-1} hence has degree one. Also, $n : [0,t] \times S^{n-1} \to S^{n-1}$ is a smooth homotopy; the degree is a homotopy invariant hence the degree of $n(t,\cdot)$ is one.

Since $n(0,\cdot)$ is the identity map on S^{n-1}, rank $n_\xi(0,\xi) =$ $= n-1$; by continuity for $t_1 > 0$ and sufficiently small, rank $n_\xi(t_1,\xi) = n-1$. This shows $n(t_1,\cdot)$ is an immersion of $S^{n-1} \to S^{n-1}$ of degree one, it must therefore be a diffeomorphism. Indeed, if not, there must be points ξ^0, $\xi^1 \in S^{n-1}$ with $n(t_1,\xi^0) = n(t_1,\xi^1)$ and sign $[\det n_\xi(t_1,\xi^0)] = -$ sign $[\det n_\xi(t_1,\xi^1)]$. Now join ξ^0 to ξ^1 by an arc on S^{n-1}; at some point of this arc $\det[n_\xi(t_1,\xi)] = 0$, a contradiction.

Now if $t_1 > 0$ is sufficiently small so that $\varphi(t_1,\cdot)$ is an immersion and $n(t_1,\cdot)$ is a diffeomorphism, we will show that $\varphi(t_1,\cdot)$ is actually an imbedding. Suppose not, i.e. $S(t_1)$ has a self intersection, in particular there exist $\xi^o, \xi^1 \in S^{n-1}$, $\xi^o \neq \xi^1$ and $\varphi(t_1,\xi^o) = \varphi(t_1,\xi^1)$. Let P be a hyperplane orthogonal to $n(t_1,\xi^o)$ at $\varphi(t_1,\xi^o)$; without loss of generality we assume the origin of E^n to be at $\varphi(t_1,\xi^o)$. Let h be the height function $h: S(t_1) \to R^1$ defined as the length of the projection of a point of $S(t_1)$ on $n(t_1,\xi^o)$. We note that a critical point of h is a point where the normal to $S(t_1)$ has direction $\pm n(t_1,\xi^o)$. There are three possibilities. a) $S(t_1)$ has points on either side of P; b) $S(t_1)$ lies in P ; c) $S(t_1)$ lies on one side of P. In case a) there must be at least one critical point of h in each of the open half spaces formed by P, i.e. there is a $\xi^2 \in S^{n-1}$ such that the normal to $S(t_1)$ at $\varphi(t_1,\xi^2)$ has direction $n(t_1,\xi^o)$ which contradicts the fact that $n(t_1,\cdot)$ is a diffeomorphism. In case b) we must lose the property that $\varphi_\xi(t_1,\xi)$ has rank $n-1$ at several points. In case c) we must have $n(t_1, \xi^o) = n(t_1,\xi^1)$,(i.e. a point of

second order contact with P) which is again a contradiction to $n(t_1,\cdot)$ being a diffeomorphism. This shows for $t_1 > 0$ and sufficiently small, $\varphi(t_1,\cdot)$ is an imbedding.

From property 5, $\partial\,\mathcal{Q}(t_1) \subset S(t_1)$; certainly $S(t_1) \subset \mathcal{Q}(t_1)$. For $t_1 > 0$ but small enough so that $S(t_1)$ is an imbedded sphere, it follows that $\partial\,\mathcal{Q}(t_1) = S(t_1)$ and $\mathcal{Q}(t_1)$ is the unique disc bounded by $S(t_1)$.

In keeping with the classical geodesic problem, $\partial\,\mathcal{Q}(t_1) = S(t_1)$ for $t_1 > 0$ and sufficiently small and property 5 imply that locally $(0 \leq t \leq t_1)$ every extremal is minimizing (optimizing). This is not true if $R(x)$ is merely convex!

In general it is not true that an immersion of $S^{n-1} \to E^n$ have a unique extension to an immersion of the disc D^n. Therefore, even if $\varphi(t_1,\cdot)$ is an immersion which extends to a disc immersion, one cannot conclude that $\mathcal{Q}(t_1)$ is necessarily the image of the disc under this immersion. However, in our case, even when $\varphi(t_1,\cdot)$ is not an immersion, we have:

<u>Property 7</u> <u>The mapping $\varphi(t_1,\cdot) : S^{n-1} \to E^n$ extends naturally to a continuous map of the disc $D^n \to E^n$ such that the image of D^n is $\mathcal{Q}(t_1)$.</u>

Proof: Modify the equations (9),(10) as follows:

$$\dot{x} = g(t,x) + \alpha r^*(t,x,p) \quad , \qquad x(0) = x^o \; , \; 0 \leq \alpha \leq 1$$
$$\dot{p} = -p[\, g_x(t,x) + \alpha r_x^*(t,x,p)] \quad , \qquad p(0) = \xi \in S^{n-1} \, .$$

Denote a solution pair of these modified equations by $\varphi(\cdot,\xi,\alpha)$, $\psi(\cdot,\xi,\alpha)$; certainly $\varphi(\cdot,\xi,1) = \varphi(\cdot,\xi)$. One may note that for each α, the modified equations are associated with the contingent equation $\dot{x} \in R(t,x,\alpha)$ where $R(t,x,\alpha) = \{g(t,x)+\alpha r : r \in Q(t,x)\}$. We will show the map $\varphi(t_1,\xi,\alpha)$ for (ξ,α) considered as polar coordinates in the disc D^n, is the required extension of $\varphi(t_1,\cdot)$.

For each $\alpha \in [0,1]$ let $\mathcal{Q}(t_1,\alpha)$ denote the attainable set at time t_1 for $\dot{x} \in R(t,x,\alpha)$, $x(0) = x^o$, and let $S(t_1,\alpha) = \{\varphi(t_1,\xi,\alpha) : \xi \in S^{n-1}\}$.

Now $R(t,x,\alpha) \subset R(t,x,1) = R(t,x)$ hence $\mathcal{Q}(t_1,\alpha) \subset \mathcal{Q}(t_1)$ for each $\alpha \in [0,1]$, or $\{\varphi(t_1,\xi,\alpha) ; \xi \in S^{n-1}, 0 \leq \alpha \leq 1\} \subset \mathcal{Q}(t_1)$. To complete the proof, the reverse inclusion must be shown.

Using property 5, for each $\alpha \in [0,1]$, $\partial\mathcal{Q}(t_1,\alpha) \subset S(t_1,\alpha)$ hence $\cup_{\alpha \in [0,1]}\partial\mathcal{Q}(t_1,\alpha) \subset \cup_{\alpha \in [0,1]}S(t_1,\alpha) = \{\varphi(t_1,\xi,\alpha): \xi \in S^{n-1}, 0 \leq \alpha \leq 1\}$. The proof will be complete if we show $\mathcal{Q}(t_1) = \cup_{\alpha \in [0,1]}\partial\mathcal{Q}(t_1,\alpha)$.

Certainly $\cup_{\alpha \in [0,1]}\partial\mathcal{Q}(t_1,\alpha) \subset \mathcal{Q}(t_1)$; to obtain the reverse inclusion, suppose $y \in \mathcal{Q}(t_1)$; we will show it belongs to $\partial\mathcal{Q}(t_1,\alpha)$ for some $\alpha \in [0,1]$. It is easy to check that $\mathcal{Q}(t_1,\alpha)$ is a continuous function of α in the Hausdorff metric topology, with $\mathcal{Q}(t_1,1) = \mathcal{Q}(t_1)$ and $\mathcal{Q}(t_1,0)$ consisting of a single point which is the unique solution of $\dot{x} = g(t,x)$, $x(0) = x^o$, evaluated at time t_1. Thus $\{\alpha \in [0,1] : y \in \mathcal{Q}(t_1,\alpha)\}$ is a closed interval; it has a least member, say α^*, and $y \in \partial\mathcal{Q}(t_1,\alpha^*)$.

In property 6 it is shown that for $t_1 > 0$ and sufficiently small, $\varphi(t_1,\cdot)$ is an imbedding of $S^{n-1} \to E^n$. Example 1 shows that for t_1 large enough, $\varphi(t_1,\cdot)$ may cease to be an imbedding and become an immersion. It is also possible, see equation (13), that as t_1 increases, the rank of $\varphi_\xi(t_1,\xi)$ becomes less than n-1. Since $\varphi_\xi(t_1,\xi)$ plays the equivalent role of the differential of the exponential map in the classical geodesic problem, it is natural to define a conjugate point as follows.

Definition. A point $\varphi(t_1,\xi^1)$ is conjugate to the point x^0 along the extremal $\varphi(\cdot,\xi^1)$ if rank $\varphi_\xi(t_1,\xi^1) < n-1$.

By this definition, conjugate points occur when the mapping $\varphi(t_1,\cdot)$ ceases to be an immersion. Thus, in example 1, there will be no conjugate points, since the equivalent geodesic problem is in a manifold of negative curvature. (See [6,pp. 100-102].) While this notion of conjugate point agrees with the classical notion, it is not equivalent to either of the definitions of conjugate points given in [7] or [8].

It would be interesting to classify those immersed spheres which could occur as the image of S^{n-1} under $\varphi(t_1,\cdot)$ for some contingent equation with $R(t,x)$ as in (7). Of course it would be of even more interest to allow pseudo-immersions [9], so that the case rank $\varphi_\xi(t,\xi) < n-1$ can also be considered.

From equation (16) in the proof of property 6 we see that

$\psi(t,\xi)\varphi_\xi(t,\xi) = 0$ even if rank $\varphi_\xi(t,\xi) < n-1$, hence $\psi(t,\xi)$ yields a generalization of the usual notion of an outward normal to an immersed $n-1$ manifold in E^n. In particular, if $S(t_1)$ is an immersed sphere then $\psi(t_1,\xi^1)$ is an outward normal to $S(t_1)$ at $\varphi(t_1,\xi^1)$. From this observation and lemma 1 we obtain

Property 8 *If $S(t_1)$ is an immersed sphere, (i.e. $\varphi(t_1,\cdot)\colon S^{n-1} \to E^n$ is an immersion) the degree of its normal (or Gauss) map is one.*

Proof: Let $\eta\colon S(t_1) \to S^{n-1}$ be the normal map. The conclusion is a consequence of the commutivity of the following diagram.

$$\begin{tikzcd}
 & S(t_1) \arrow[dr, "\eta(\cdot)"] & \\
S^{n-1} \arrow[ur, "\varphi(t_1\text{,}\cdot)"] \arrow[rr, "n(t_1\text{,}\cdot)"'] & & S^{n-1}
\end{tikzcd}$$

Actually, we can think of $n(t,\xi)$ as a generalization of a unit normal to $S(t)$ at $\varphi(t,\xi)$ even if a normal (in the usual sense) does not exist. Lemma 1 shows that the degree of this generalized Gauss map is always one even if $\varphi(t,\cdot)$ is not an immersion.

The property of having normal degree one is not alone enough to classify the immersed spheres which can be generalized geodesic spheres (i.e. $S(t_1)$ for some contingent problem with $R(t,x)$ of the form (7).) In [10], Smale classifies immersed spheres up to regular homotopy. Following property 9, we will show that this is also not a fine enough property to distinguish generalized geodesic spheres.

<u>Property 9</u> <u>If</u> $y \in \partial \mathcal{Q}(t_1)$ <u>then there exists a closed neighborhood</u> N <u>contained in</u> $\mathcal{Q}(t_1)$ <u>with</u> $y \in \partial N$. $(t_1 > 0.)$

Proof: Let $y = \varphi(t_1, \xi^1)$, and $y^o = \varphi(t_1 - \epsilon, \xi^1)$ for $\epsilon > 0$ and sufficiently small so that (using property 6) the attainable set at time t_1 from "initial" data $x(t_1 - \epsilon) = y^o$ rather than $x(0) = x^o$ for equation (9), is a disc. Then y belongs to the boundary of this disc and the disc belongs to $\mathcal{Q}(t_1)$.

The following figure shows an immersion of $S^1 \to E^2$ of normal degree one which cannot be a generalized geodesic sphere (see the point y) yet is regularly homotopic to the immersion of S^1 obtained in example 1. The regular homotopy is obtained by "pushing" along the arrows.

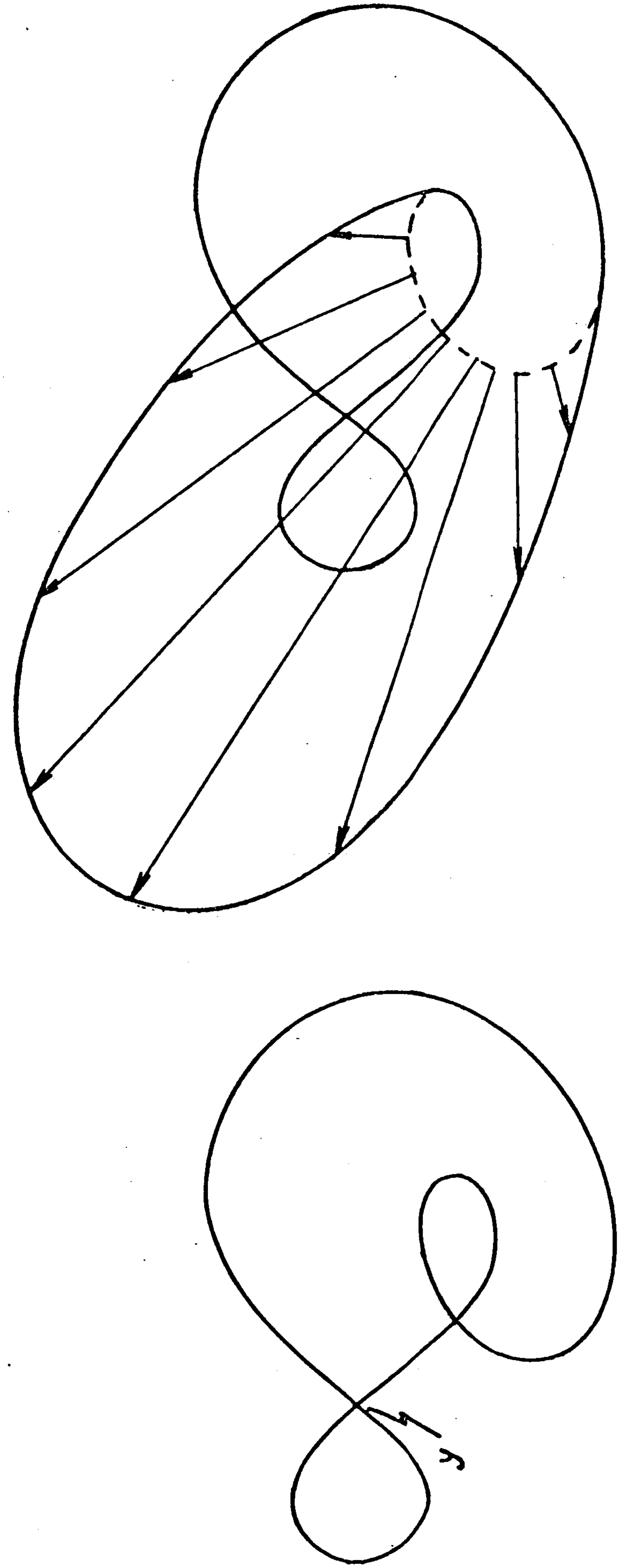


FIGURE 2

The following gives an application to a time optimal control problem, which in its original formulation does not fit the theory. It also shows that $S(t_1)$ can be computed numerically even if "singular" arcs occur.

EXAMPLE 2. Consider the controlled, two dimensional system

$$\dot{x}_1 = u_1 \qquad , \quad x_1(0) = 0$$

$$\dot{x}_2 = 1 + x_1^2 x_2 u_1 \qquad , \quad x_2(0) = 0$$

with control constraints $|u_1(t)| \leq 1$. As it stands, the corresponding set $R(t,x)$ does not admit the representation we require, however by adding a "small" second component of control one may consider

$$\dot{x}_1 = u_1 \qquad , \quad x_1(0) = 0$$

$$\dot{x}_2 = 1 + x_1^2 x_2 u_1 + u_2 \qquad , \quad x_2(0) = 0$$

where now we require $\epsilon^2 u_1^2 + u_2^2 \leq \epsilon^2$, i.e. the controls come from an ellipse with semi-major axis one, semi-minor axis ϵ. The problem is now within our formulation, equations (9) and (10) become

$$\dot{x}_1 = u_1^* \qquad , \quad \dot{p}_1 = -2p_2 x_2 x_1 u_1^*$$

$$\dot{x}_2 = 1 + x_1^2 x_2 u_1^* + u_2^* \qquad , \quad \dot{p}_2 = -p_2 x_1^2 u_1^*$$

where

$$u_1^*(x,p) = [p_1+p_2x_2x_1^2][\epsilon^2 p_2+(p_1+p_2x_2x_1^2)^2]^{-\frac{1}{2}}$$

$$u_2^*(x,p) = \epsilon^2 p_2[\epsilon^2 p_2^2+(p_1+p_2x_2x_1^2)^2]^{-\frac{1}{2}}$$

One can compute $S(t)$ by numerical integration of an initial value problem; a reasonable spacing of the initial data on S^1 can be obtained by noting that for ϵ small and p_1 near zero, u^* may change rapidly with the remaining variables. The following figures were easily obtained numerically.

In each figure $p(0) = \xi \epsilon S^1$ was given in angular measure with $p(0) = (1,0)$ corresponding to $0°$; $p(0) = (0,1)$ corresponding to $90°$ etc. and the computation carried out at each degree for five degrees on either side of $90°$ and $270°$, while increments of 5 to 15 degrees were used elsewhere.

In figure 3, at $t = 1$, one would still expect that $\varphi(1,\cdot)\colon S^1 \to E^2$ is an imbedding. In figure 4, at $t = 4$, it is not even an immersion, i.e. conjugate points have occured. The sharp corners which seem apparent in figure 4 may well exist since $S(t_1)$ is merely the projection to E^2 of a diffeomorphic image of S^1 in E^4.

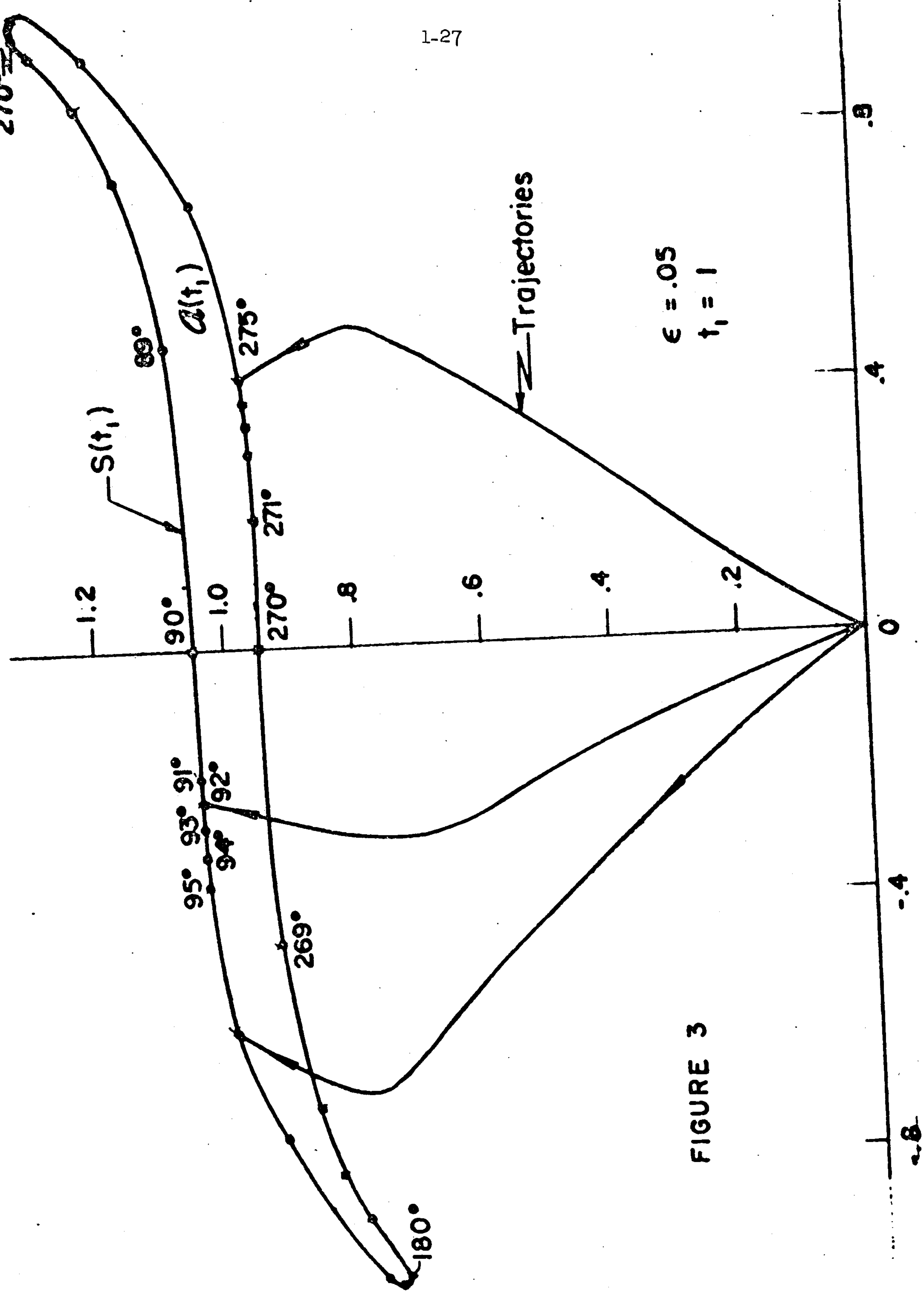
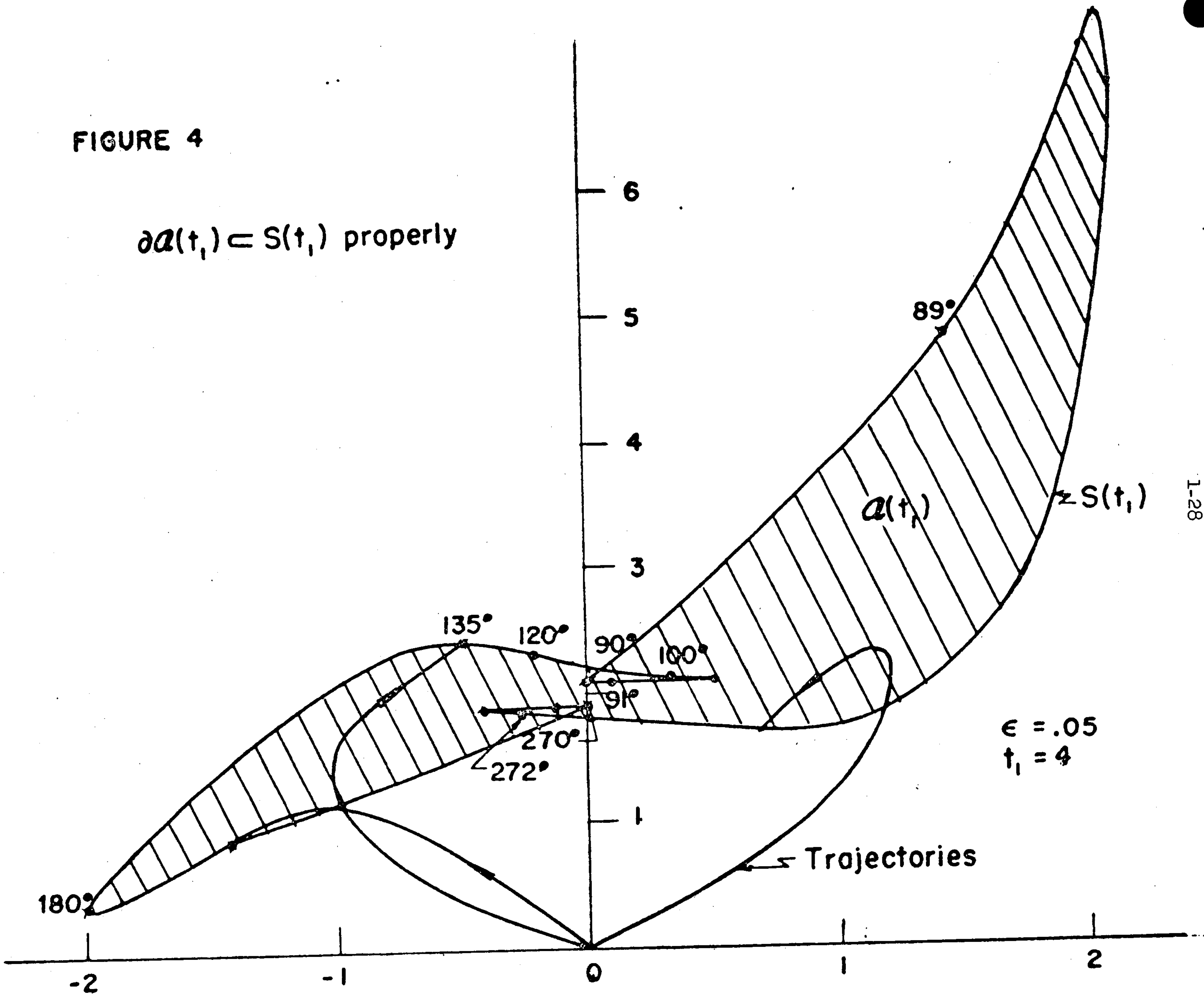

S(t₁)
𝒜(t₁)
Trajectories
ε = .05
t₁ = 1
270°
89°
275°
271°
90°
270°
91°
92°
93°
94°
95°
269°
180°
1.2
1.0
.8
.6
.4
.2
0
.8
.4
-.4
-.8

FIGURE 3

FIGURE 4

$\partial \mathcal{A}(t_1) \subset S(t_1)$ properly



REFERENCES

[1] Busemann, H., The Geometry of Geodesics, Academic Press, Inc., New York, New York (1955).

[2] Pontriagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V. and Mischenko, E.F., Mathematical Theory of Optimal Processes, Interscience, New York (1962).

[3] Filippov, A.F., On Certain Questions in the Theory of Optimal Control (English trans.) S.I.A.M. J. on Control, Ser. A. 1 (1962) 76-84.

[4] Hermes, H., The Equivalence and Approximation of Optimal Control Problems, J. of Diff. Eqs., Vol. 1, No. 4(1965) 409-426.

[5] Filippov, A.F., Differential Equations with Many-Valued Discontinuous Right-Hand Sides, Dokl. Trans., Jan. 1964, 941-945.

[6] Milnor, J., Morse Theory, Annals of Math. Studies, No. 51, Princeton Univ. Press, Princeton, N.J.(1963).

[7] Bushaw, D., Dynamical Polysystems and Optimization, Contr. to Diff. Eqs., Vol. 2, (1963), 351-365.

[8] Halkin, H., Topological Aspects of Optimal Control of Dynamical Polysystems, Contr. to Diff. Eqs., Vol. 3, No. 4, (1964), 377-385.

[9] Poenaru, V., On Regular Homotopy in Codimension 1, Ann. of Math., Vol. 83, No. 2, (1966), 257-265.

[10] Smale, S., The Classification of Immersions of Spheres in Euclidean Spaces, Ann. of Math., Vol. 69, No. 2, (1959), 327-344.

PAPER [2]

To appear in SIAM Journal on Control

ON THE CLOSURE AND CONVEXITY OF ATTAINABLE SETS

H. Hermes*

* This research was supported by the National Aeronautics and Space Administration under Contract No. NAS8-11264.

INTRODUCTION

We shall consider the control system

$$\dot{x}(t) = f(t,x(t),u(t)) , \qquad x(0) = x^0 \tag{1}$$

where $x(t)$ is an n vector, $u(t)$ an r vector. Our interest will be focused on solutions which exist within some fixed finite time interval $[0,T]$.

The control functions u will be assumed to belong to a control set Ω which may be given in either of the two following ways.

(i) For each $t \in [0,T]$ let $U(t)$ be any subset of Euclidean n space E^n and $\Omega = \{u \in \mathcal{L}_\infty[0,T]: u(t) \in U(t)\}$. We assume the sets $U(t)$ are contained in some fixed bounded sphere in E^r for $t \in [0,T]$.

(ii) Ω is a bounded subset of r vector valued functions with components in $\mathcal{L}_\infty[0,T]$.

The first case is that which is usually considered in control theory. It is not necessarily the practical case, since the admissible control functions may be a-priori restricted by the electronic and mechanical ability of function generators.

In the latter case, rather than introduce the awkward notation that Ω is contained in the direct product of $\mathcal{L}_\infty[0,T]$ taken with itself r times, we will merely write $\Omega \subset \mathcal{L}_\infty[0,T]$. The number of components a control $u \in \Omega$ has will be unimportant or clear from context.

There will be need to consider $\mathcal{L}_\infty$ with its norm topology, its weak topology, and its weak* topology (or $\mathcal{L}_1$ topology of $\mathcal{L}_\infty$). For writing ease both notations for the weak* topology will be used.

We will assume:

f is continuous on $[0,T] \times E^n \times E^r$ and once continuously differentiable in the x argument, unless explicitly stated elsewise. (2)

There exists a constant $c > 0$ such that $x \cdot f(t,x,u) \leq c[1 + |x|^2]$ for all t,x,u in the domain of definition of f. (This prevents finite escape time.) (3)

REMARK: Actually (3) implies any solution of (1) will remain in the sphere $|x| \leq [1 + |x^o|^2]\exp(cT)$, therefore the condition (2) can be relaxed by replacing E^n in the domain of continuity by this sphere, and E^r by the bounded sphere within which the controls take their values.

With these assumptions, for each $u \in \Omega$ equation (1) has a unique solution defined on $[0,T]$ which will be denoted $x(\cdot,u)$.

Define:

$$\mathcal{F} = \{f(\cdot,x(\cdot,u),u(\cdot)) \in \mathcal{L}_\infty[0,T] : u \in \Omega\}$$

$$\mathcal{Q}(t) = \{x(t,u) \in E^n: u \in \Omega\} = \{x^o + \int_0^t z(\tau)d\tau : z \in \mathcal{F}\}.$$

If Ω has the representation (i), it is possible to define

$$F(t) = \{f(t,x(t,u),u(t)) \in E^n : u \in \Omega\} .$$

From the representation (i) it follows that an equivalent definition is

$$F(t) = \{f(t,a,\sigma): a \in \mathcal{Q}(t), \ \sigma \in U(t)\} .$$

One could define a set analogous to $F(t)$ in its first representation for the case Ω given as in (ii) by introducing the notion of approximate continuity [1,pp. 261-2] to circumvent the difficulty that the value $u(t)$ of an element $u \in \Omega$ is ambiguous. However in this case the second representation would not be an equivalent representation. In what follows this would be of no use, therefore whenever reference is made to $F(t)$ it will be understood that Ω has the representation (i).

$$\mathcal{B}(t) = \{x^0 + \int_0^t z(\tau)d\tau : z \text{ measurable, } z(\tau) \in F(\tau) \text{ for } 0 \leqq \tau \leqq t\}.$$

REMARKS: a) From the assumptions on f and Ω it follows that $\mathcal{F}$ is a bounded subset of $\mathcal{L}_\infty[0,T]$. It is the set of derivatives of admissible trajectories.

b) $\mathcal{Q}(t)$ is commonly referred to as the attainable set at time t. We always have $\mathcal{Q}(t) \subset \mathcal{B}(t)$, one of the things we shall be interested in is when are these sets equal.

We shall next summarize results. In doing so several theorems from other references will be stated, at times the statements of these may be somewhat different from the form in which they originally appeared. In these cases the verification of the equivalence will be included in §1 where proofs of the results are given.

I. [2, Theorem 1] $\mathcal{B}(t)$ is convex for each $t \in [0,T]$.

II. [2, Theorem 4] If $F(\tau)$ is closed for each $t \in [0,t]$ (our assumptions imply it is bounded) then $\mathcal{B}(t)$ is convex and compact for each $t \in [0,T]$.

III. [3, Theorem 1 and Lemma]. Suppose Ω has the representation (i) with $U(t)$ a nonempty compact subset of E^r for each $t \in [0,T]$ which is continuous in the Hausdorff topology as a function of t. Suppose further that for each t, x, $\{f(t,x,\sigma): \sigma \in U(t)\}$ is convex. Then $F(t)$ is closed for each $t \in [0,T]$.

IV. (Restatement of Theorem 1, [3]). Assume the hypotheses of III. Then $\mathcal{F}$ is a weak* compact subset of $\mathcal{L}_\infty[0,T]$.

REMARK: From this it immediately follows that $\mathcal{Q}(t)$ is compact. Indeed the mapping $L : \mathcal{L}_\infty \to E^n$ defined by $Lz = \int_0^t z(\tau)d\tau$ is weak* continuous hence the image of $\mathcal{F}$ is compact.

V. Assume the hypotheses of III and that for each $\tau \in [0,T]$ and $a, a' \in \mathcal{Q}(\tau)$

$$\{f(\tau,a,\sigma): \sigma \in U(\tau)\} = \{f(\tau,a',\sigma): \sigma \in U(\tau)\}. \tag{4}$$

Then $\mathcal{Q}(t) = \mathcal{B}(t)$ for each $t \in [0,T]$.

VI. (Combining II, III, and V) If the hypotheses of V are satisfied $\mathcal{Q}(t)$ is compact and convex for each $t \in [0,T]$.

REMARK: Compactness of $\mathcal{Q}(t)$ is essential in existence theorems, the convexity of $\mathcal{Q}(t)$ rules out "conjugate points" and thus simplifies sufficiency conditions.

The next few results pertain to the case where Ω has the representation (ii).

Let X^* be the dual of a Banach space X, then every closed and bounded (in norm) convex set in X^* is closed in the X^{**} (or weak) topology of X^*. Also, a subset of X^* is compact in the X topology of X^* if and only if it is bounded in the norm topology and closed in the X topology. (See [6, pp. 422-424].)

In [4,p 881] Klee shows: Every nonreflexive separable Banach space contains two disjoint closed bounded convex sets which cannot be separated. As remarked in [4], the separability is not essential since every nonreflexive space has a nonreflexive closed separable subspace within which one could apply the result. Using Klee's result one easily obtains

VII. If X^* is a nonreflexive Banach space which is the dual of a Banach space X, it continas a closed, bounded, convex subset Ω_1 which is not closed in the X topology of X^*.

For any $y = (y_1,\ldots,y_n)$ with components in $\mathcal{L}_1$ let $L(y)$ denote the linear operator from $\mathcal{L}_\infty$ to E^n defined by

$$y(u) \equiv L(y)u = \int_0^T y(\tau)u(\tau)d\tau \tag{5}$$

(We assume u is scalar valued.)

VIII. There exists a $y \in \mathcal{L}_1[0,T]$ such that the image of the closed, bounded, convex set $\Omega_1 \subset \mathcal{L}_\infty$ under the continuous linear map $L(y)$ is not closed in E^n.

Equivalently, it readily follows from this that there exists a linear control system of the form

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) , \quad x(0) = x^0 \tag{6}$$

where A, B have components in $\mathcal{L}_1[0,T]$ and $u \in \Omega_1$, a closed, bounded, convex subset of $\mathcal{L}_\infty[0,T]$, for which $\mathcal{Q}(T)$ is not compact.

§1. VERIFICATION OF STATED RESULTS.

III. $F(t)$ is closed.

For Ω as given in III, $F(t) = \{f(t,x,\sigma): x \in \mathcal{Q}(t), \sigma \in U(t)\}$. Filippov's theorem [3] shows $\mathcal{Q}(t)$ is compact, $U(t)$ is given compact and f is continuous therefore $F(t)$ is compact.

IV. $\mathcal{F}$ is weak* closed.

Let $z^n(\cdot) = f(\cdot,x(\cdot,u^n),u^n(\cdot)) \in \mathcal{F}$ and z^n converge to z in the weak* topology. We will show $z \in \mathcal{F}$.

Since $\mathcal{F}$ is bounded in norm it is easily shown that z^n converges to z in the weak* topology if and only if

$\int_0^t z^n(\tau)d\tau \to \int_0^t z(\tau)d\tau$ for each $t \in [0,T]$ (Exercise 27,[6,p.342]). Letting $x(t) = x^o + \int_0^t z(\tau)d\tau$, $t \in [0,T]$, the hypotheses imply $x(\cdot,u^n)$ converges to x uniformly. But then by the Filippov argument [3, proof of Theorem 1] x is an admissible trajectory, i.e., there exists an admissible control u such that $z(t) = \dot{x}(t) = f(t,x(t),u(t))$ for almost all $t \in [0,T]$ showing $z \in \mathcal{F}$.

V. (Proof). We already know $\mathcal{A}(t) \subset \mathcal{B}(t)$, now let $x^o + \int_0^t z(\tau)d\tau$ be any element in $\mathcal{B}(t)$,i.e., for each $s \in [0,t]$, $z(s) \in F(s)$. We must show $x^o + \int_0^\tau z(\tau)d\tau$ is an admissible trajectory for $\tau \in [0,t]$.

By the representation of Ω and property (4) of the hypotheses of V, $F(s) = \{f(s,a,\sigma) : \sigma \in U(s)\}$ for any $a \in \mathcal{A}(s)$. Pick any admissible control u^o, let $x(\cdot,u^o)$ be its corresponding trajectory. Then for each $s \in [0,t]$, $z(s) \in \{f(s,x(s,u^o),\sigma) : \sigma \in U(s)\}$ hence by the Filippov lemma, [3], there exists an admissible control u^1 such that $z(s) = f(s,x(s,u^o),u^1(s))$ almost everywhere. Using u^1 in the place of u^o we may proceed inductively to generate a sequence of trajectories $\{x(\cdot,u^n)\}$ and corresponding sequence of controls $\{u^{n+1}\}$ such that $z(s) = f(s,x(s,u^n),u^{n+1}(s))$ almost everywhere in $[0,t]$. From the sequence $\{x(\cdot,u^n)\}$ choose a uniformly convergent subsequence (the original sequence is easily seen to be an equicontinuous family), for notational ease assume it is the original sequence. Define $z^n(s) = f(s,x(s,u^n),u^n(s))$, then

$$|z^n(s)-z(s)| = |f(s,x(s,u^n),u^n(s))-f(s,x(s,u^{n-1}),u^n(s))|$$

$$\leq K \sup_{0 \leq s \leq t} |x(s,u^n)-x(s,u^{n-1})|$$

for almost all $s \in [0,t]$. But $x(\tau,u^n) = x^o + \int_0^\tau z^n(s)ds$ is an admissible trajectory hence from the preceding estimate $x(\tau,u^n)$ converges uniformly to $x^o + \int_0^\tau z(s)ds$. Therefore $x^o + \int_0^\tau t(s)ds$ is the uniform limit of admissible trajectories, by Filippovs Theorem [3] it is an admissible trajectory which completes the argument.

VII. In [4], Klee shows that every nonreflexive separable Banach space contains two disjoint, closed, bounded, convex sets which cannot be separated. The separability of the space is inconsequential since, as commented in [4], every nonreflexive Banach space X^* has a separable nonreflexive closed subspace. Let E^* denote this subspace; consider A,B closed bounded and convex in E^* and such that they cannot be separated. Then as subsets of X^* they are also closed, bounded, convex and cannot be separated by a hyperplane since $E^* \subset X^*$ implies $X^{**} \subset E^{**}$; i.e., any continuous linear functional on X^* is a continuous linear functional on E^*.

Now suppose either A or B is closed in the X topology of X^*. Then by corollary 3, [6, p. 424] it is compact in the X topology of X^*. This implies we have two closed, disjoint, convex sets in a locally convex linear topological space (X^* with its X topology) one of which is compact. By corollary 11 [6, p. 418] there exists a nonzero continuous linear functional f which separates them. But if f is continuous in the weak* topology it is continuous in the norm topology of X^*, i.e., $f \in X^{**}$. This implies f separates A and B in X^*, a contradiction. Thus neither A or B can be closed in the X topology of X^*.

VIII. (Proof). It suffices to consider y, as in (5), to be scalar valued, i.e., $L(y): \mathcal{L}_\infty \to E^1$. We will consider only real linear spaces.

Let X be a Banach space and K^* a bounded, X closed, convex subset of X^* (i.e., an X compact subset of X^*.) Then K^* has continuous (in the X topology of X^*) nonzero, tangent functionals. In fact it is known that these exist at each point of a dense subset of its boundary (see [6], exercise 13, p. 459). Explicitly, let D^* be the (nonempty) subset of the boundary of K^* at which continuous tangent functionals exist, i.e., for each $x_1^* \in D^*$ there exists a nonzero $g \in X$ and real constant c_g such that $g(K^*) \leq c_g$, $g(x_1^*) = c_g$. Such a g determines a support hyperplane h_g to K^* at x_1^* where $h_g = \{x^* \in X^* : g(x^*) = c_g\}$ and a corresponding closed half space $H_g = \{x^* \in X^* : g(x^*) \leq c_g\}$ which contains K^*. Let G be the family of continuous tangent functionals so determined by elements of D^*.

<u>LEMMA</u>. K^* is uniquely determined as the intersection of the half spaces H_g, i.e., $K^* = \cap_{g \in G} H_g$.

<u>Proof of lemma</u>: If $x^* \in K^*$ then $x^* \in H_g$ for every g hence $K^* \subset \cap_{g \in G} H_g$.

To obtain the reverse inclusion, suppose $x_1^* \in \cap_{g \in G} H_g$ but $x_1^* \notin K^*$. Since K^* is closed and convex there exists a continuous linear functional $x \in X$ which separates x_1^* and K^*, suppose $x(x_1^*) = c$, $x(x^*) < c$ for $x^* \in K^*$. Let $c_x = \sup\{x(x^*): x^* \in K^*\}$, since K^* is compact in the X topology of X^*, $c_x < c$ and there exists an $x_2^* \in K^*$ such that $x(x_2^*) = c_x$. But then $x_2^* \in D^*$ and $x \in F$ and since $x(x_2^*) = c > c_x$

we have a contradiction to $x_1^* \in \bigcap_{g \in G} H_g$.

REMARK: The existence of even a single support plane for a bounded, closed, convex subset of a Banach space is still an open question. See [7, p. 98].

We now continue the proof of VIII. Let Ω_1 be the bounded, closed (in $\mathcal{L}_\infty$) convex set which is not weak* closed, as shown to exist in VII. Let $\overline{\Omega}_1$ denote the weak* closure of Ω_1, then $\overline{\Omega}_1 - \Omega_1$ is not empty. Applying the preceding lemma to $\overline{\Omega}_1$ we see it is uniquely determined by its support planes; since $\Omega_1 \neq \overline{\Omega}_1$ there must be a support plane P to $\overline{\Omega}_1$ which is not a support plane of Ω_1. Let $y \in \mathcal{L}_1$ be the continuous, linear, (tangent) functional which determines P, i.e., $y(x^*) \leq c$ for $x^* \in \overline{\Omega}_1$ and $y(x_1^*) = c$ for some $x_1^* \in \overline{\Omega}_1$. Since P is not a support plane for Ω_1, $y(x^*) < c$ for all $x^* \in \Omega_1$ but since x_1^* is in the weak* closure of Ω_1 there exists a sequence $\{z_\nu^*\} \subset \Omega_1$ with $\lim_{\nu \to \infty} y(z_\nu^*) = c$. This shows c is in the closure of $L(y)\Omega_1$ but not in $L(y)\Omega_1$.

REMARK: Using the theorem of Lyapunov on the range of a rector measure, one can show there do exist closed subsets of $\mathcal{L}_\infty$, e.g., $\{u \in \mathcal{L}_\infty[0,T]: |u(t)| = 1\}$ which have the property that their image under any map of the form $L(y)$ is compact. (See, for example [2, Theorem 3] or [5, Theorem 1].)

§2. EXAMPLES.

a) Any linear system of the form (6) with Ω as given in (i) and $U(t)$ convex and compact for each $t \in [0,T]$ can be transformed into an equivalent system which satisfies the hypotheses of V.

Indeed, let $X(t)$, $X(0) = I$, be a fundamental solution of the homogeneous system and make the change of variable $y(t) = X^{-1}(t)x(t)$. Then x satisfies (6) if and only if y satisfies $\dot{y}(t) = X^{-1}(t)B(t)u(t)$, $y(0) = x^{o}$. This transformed system obviously satisfies the hypotheses of V, therefore, as is well known, the associated set $\mathcal{Q}(t)$ is compact and convex.

b) Consider

$$\dot{x}_1 = 1 + \sin x_2 u \qquad , \; x_1(0) = \pi$$

$$\dot{x}_2 = 1 - \sin x_2 u \qquad , \; x_2(0) = \pi \quad , \qquad 0 \leqq u(t) \leqq 2$$

Since $\dot{x}_1 \geqq 0$, $\dot{x}_2 \geqq 0$, $x \in \mathcal{Q}(t)$ implies $x_1 \geqq \pi$, $x_2 \geqq \pi$ therefore $\{f(x,u) : u \in U\}$ is independent of $x \in \mathcal{Q}(t)$. (It is the segment of the line $y_1 + y_2 = 2$ with $y_1 \geqq 0$, $y_2 \geqq 0$.) The hypotheses of VI are satisfied and the attainable set will be compact and convex.

REFERENCES

[1] Natanson, I. P.; Theory of Functions of a Real Variable I, Fredrick Ungar Pub. Co., New York (1955).

[2] Aumann, R. J.; Integrals of Set Valued Functions; J. Math. Anal. and Appl., Vol. 12, No. 1, August, 1965.

[3] Filippov, A. F.; On Certain Questions in the Theory of Optimal Control; J. SIAM on Control, Ser. A, 1, (1962), pp. 76-84.

[4] Klee, V. L. Jr.; Convex Sets in Linear Spaces II, Duke Math. J., 18, (1951), pp. 875-883.

[5] Hermes, H.; A Note on the Range of a Vector Measure; Application to the Theory of Optimal Control; J. Math. Anal and Appl., Vol. 8, No. 1, 1964, pp. 78-83.

[6] Dunfard, N; Schwartz, J. T.; Linear Operators I, Interscience Inc., New York, New York, (1958).

[7] Klee, V.L. Jr.; Extremal Structure of Convex Sets II, Math. Z., 69, (1958), pp. 90-104.

PAPER [3]

FINITE TIME STABILITY UNDER PERTURBING FORCES AND ON PRODUCT SPACES

BY

L. WEISS AND E. F. INFANTE

Reprinted from IEEE TRANSACTIONS
ON AUTOMATIC CONTROL
Volume AC-12, Number 1, February, 1967
pp. 54-59

Finite Time Stability Under Perturbing Forces and on Product Spaces

LEONARD WEISS, MEMBER, IEEE, AND E. F. INFANTE, MEMBER, IEEE

Abstract—**This paper continues the development of a qualitative theory of stability, recently initiated by the authors, for systems operating over finite time intervals. The theory is motivated by 1) the need for a more practical concept of stability than is provided by the classical theory; and 2) the search for methods for investigating stability of a system trajectory (either analytically or numerically given) without the necessity of performing complicated transformations of the differential equations involved.**

The systems studied in this paper are nonautonomous, i.e., they are under the influence of external forces, and the concept of finite time stability (precisely defined in the paper) in this case involves the bounding of trajectories within specified regions of the state space during a given finite time interval. (The input is assumed to be bounded by a known quantity during this time interval.)

Sufficient conditions are given for various types of finite time stability of a system under the influence of perturbing forces which enter the system equations linearly. These conditions take the form of existence of "Liapunov-like" functions whose properties differ significantly from those of classical Liapunov functions. In particular, there is no requirement of definiteness on such functions or their derivative.

The remainder of the paper deals with the problem of determining finite time stability properties of a system from knowledge of the finite time stability properties of lower-order subsystems which, when appropriately coupled, form the original system.

An example is given which illustrates some of the concepts discussed in the paper.

I. Introduction

IN MANY CASES of practical interest, there is concern with the behavior of systems over a fixed interval of time; e.g., will a given system exhibit a response to given stimuli which is contained within certain specified bounds during the fixed time interval? Among the multitudinous problems which fall into this category are: the problem of assuring that a space vehicle will remain in a specified orbit for a given length of time in order to complete a set of experiments; the problem of sending a rocket from a neighborhood of a point A to a neighborhood of a point B over some nominal trajectory; the problem, in a chemical process, of keeping the temperature or pressure or some other parameter within specified bounds. It appears reasonable to consider such questions within a stability framework, i.e., a system is "stable" if it operates within the prespecified bounds and is "unstable" if it does not. However, it is evident that the classical theory of stability requires strong modification in order to be relevant toward the resolution of such stability questions.

Taking inspiration from the discussion of "practical stability" in the monograph by LaSalle and Lefschetz [1], the authors began, in a previous paper [2], the development of a qualitative theory of this type of stability, which is called *finite time stability*. The theory developed thus far parallels, to a certain extent, the classical Liapunov theory of stability, but differs from it in a number of significant respects which are evident from the definitions and theorems.

Finite time stability and instability of systems of the form

$$\dot{x} = f(x, t), \tag{1}$$

where x is a real n vector (the state vector), was discussed in [2]. This paper is concerned with finite time stability of systems under the influence of perturbing forces, i.e., the systems considered are of the form

$$\dot{x} = f(x, u, t), \tag{2}$$

where u is a vector representing a forcing function and, in general, $u = u(x, t)$.

It is assumed that the usual smoothness conditions are present so that there is no difficulty with questions of existence, uniqueness, and continuity of solutions with respect to initial data.

Finally, it is *not* required that $f(0, 0, t) = 0$, so that stability with respect to a set rather than a point can be discussed without resorting to complicated transformations.

II. Notational Preliminaries

Let X be the state space for (2). Then define

$$B(a) = \{x \in X; \|x\| < a\}$$

$$\overline{B}(a) = \{x \in X; \|x\| \leqq a\}$$

$$\mathfrak{J} = [t_0, t_0 + T) \qquad \text{where } t_0, T \in R^1$$

$V : X \times \mathfrak{J} \to R^1$ and $V(x, t)$ is C^1 in x and C^0 in t.

$$V_M^{\alpha}(t) = \max_{\|x\| = \alpha} V(x, t)$$

$$V_m^{\beta}(t) = \min_{\|x\| = \beta} V(x, t)$$

$$V(x, t) = \operatorname{grad} V(x, t) \cdot \frac{dx}{dt} + \frac{\partial V}{\partial t} \cdot$$

Manuscript received April 4, 1966; revised October 19, 1966. This research was supported partially by the National Aeronautics and Space Administration under Contract NAS8-11264 and partially by the Alfred P. Sloan Foundation in the form of a research fellowship to L. Weiss. An early version of this work was presented at the International Symposium on Differential Equations, Mayaguez, Puerto Rico, December 1965.

The authors are with the Center for Dynamical Systems, Brown University, Providence, R. I.

III. Definition of Stability

Let (2) be defined on a time interval $[t_0, t_0+T)$. Then consider

Definition 1

A system (2) is *stable under perturbing forces with respect to the set* $(\alpha, \beta, \epsilon, t_0, T, \|\cdot\|)$, $\alpha \leqq \beta$, if, for any trajectory $x(t)$, the conditions $\|x(t_0)\| < \alpha$ and $\|u(x, t)\| \leqq \epsilon$ for all $t \in \mathfrak{I}$, all $x \in (\overline{B}(\beta) - B(\alpha))$ imply $\|x(t)\| < \beta$ for $t \in \mathfrak{T}$.[1]

Remarks: 1) Obviously, a somewhat more general definition can be made in which stability is defined with respect to regions in the state space which are not necessarily balls in the norm topology. Later on, it becomes useful to make this type of modification of the above definition. We emphasize that the symbol $\|\cdot\|$ need not indicate a true norm.

2) For $\epsilon = 0$, Definition 1 reduces to that of finite time stability of (1) as given in reference [2].

3) It is strongly emphasized that the numbers α, β, t_0, T are all specified a priori in a given problem. Hence, although there is some analogy to the usual classical definition of stability under perturbations, it is clear that with respect to the aforementioned set, a system which is stable in the classical sense may be unstable in the sense of Definition 1 and vice versa. The next definitions are finite time analogs of asymptotic stability under perturbations. Since the word "asymptotic" has little meaning in the finite time context, the word "contractive" is used instead.

Definition 2

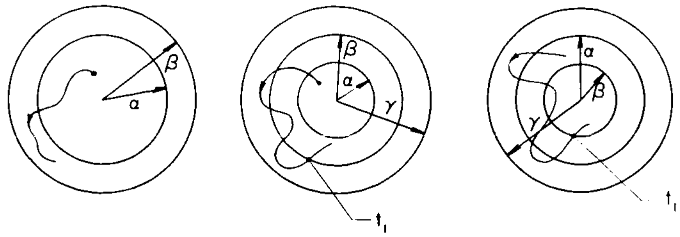
A system (2) is *quasi-contractively stable under perturbing forces with respect to* $(\alpha, \beta, \gamma, \epsilon, t_0, T, \|\cdot\|)$, $\alpha \leqq \beta < \gamma$, if, for any trajectory $x(t)$, the conditions $\|x(t_0)\| < \alpha$ and $\|u(x, t)\| \leqq \epsilon$ for all $x \in (\overline{B(\gamma)} - B(\alpha))$, all $t \in \mathfrak{I}$ imply 1) stability under perturbing forces with respect to $(\alpha, \gamma, \epsilon, t_0, T, \|\cdot\|)$; 2) there exists $t_1 \in (t_0, t_0+T)$ such that $\|x(t)\| < \beta$ for all $t \in (t_1, t_0+T)$.

Definition 3

A system (2) is *contractively stable under perturbing forces with respect to* $(\alpha, \beta, \gamma, \epsilon, t_0, T, \|\cdot\|)$, $\beta < \alpha < \gamma$, if, for any trajectory $x(t)$, the conditions $\|x(t_0)\| < \alpha$ and $\|u(x, t)\| \leqq \epsilon$ for all $x \in (\overline{B(\gamma)} - B(\beta))$, all $t \in \mathfrak{I}$ imply 1) stability under perturbing forces with respect to $(\alpha, \gamma, \epsilon, t_0, T, \|\cdot\|)$ and 2) there exists $t_1 \in (t_0, t_0+T)$ such that $\|x(t)\| < \beta$ for all $t \in (t_1, t_0+T)$.

Note: Definition (2) with $\epsilon = 0$ does not correspond to the definition of quasi-contractive stability given in [2] for system (1). It is believed that the concept of quasi-contractive stability defined above is the more natural one.

[1] For convenience, we have used the same symbol to denote the "length" of $x(t_0)$ as well as $u(x, t)$. The same measure of length need not, however, be applied to both of them.



$\|u(x, t)\| \leqq \epsilon$, $x \epsilon (\overline{B(\beta)} - B(\alpha))$, $t \in \mathfrak{I}$ — $\|u(x, t)\| \leqq \epsilon$, $x \epsilon (\overline{B(\gamma)} - B(\alpha))$, $t \in \mathfrak{I}$ — $\|u(x, t)\| \leqq \epsilon$, $x \epsilon (\overline{B(\gamma)} - B(\beta))$, $t \in \mathfrak{I}$

Fig. 1. Stability—Def. 1 Stability—Def. 2 Stability—Def. 3

Illustrations of the three types of stability defined above are given in Fig. 1 in terms of trajectory behavior.

IV. Theorems on Stability under Perturbing Forces

The definitions of stability given in Section III apply to every general type of system (2). The theorems given in this section yield sufficient conditions for stability of a special case of (2), namely, systems of the form

$$\dot{x} = f(x, t) + u(x, t). \tag{3}$$

Here u is an n vector which, as indicated above enters the system equation linearly.

It turns out to be convenient to separate the case $\alpha < \beta$ from the case $\alpha = \beta$ when discussing sufficient conditions for a system to be stable in the sense of definition.

A theorem for the former case will be stated and proved. The case $\alpha = \beta$ is left as an exercise.

Theorem 1

A system (3) is stable under perturbing forces with respect to $(\alpha, \beta, \epsilon, t_0, T)$, $\|\cdot\|)$, $\alpha < \beta$, if there exists a real-valued function $V(x, t)$ and real-valued functions $\phi(t)$, $\rho(t)$ integrable on $\mathfrak{I}$ such that

1) $\|\text{grad } V(x, t)\| \leqq \rho(t) \quad \text{for } x \in (\overline{B(\beta)} - B(\alpha)), t \in \mathfrak{I}$

2) $V_f(x, t) < \phi(t) \qquad \text{for } x \in (\overline{B(\beta)} - B(\alpha)), \quad t \in \mathfrak{I}$

$$\text{where } V_f = V\big|_{u(x,t)=0}$$

3) $$\int_t^{t_2} [\phi(t) + \epsilon\rho(t)]dt \leqq V_m^{\beta}(t_2) - V_M^{\alpha}(t_1),$$

$$t_1, t_2 \in \mathfrak{I}, \; t_2 > t_1.$$

Proof: Let $x(t)$ be an arbitrary trajectory of (3) such that $\|x(t_0)\| < \alpha$. Assume there exists $t_2 \in (t_0, t_0+T)$, the first such point in $\mathfrak{I}$ such that $\|x(t_2)\| = \beta$. Then there exists $t_1 < t_2$, $t_1 \in \mathfrak{I}$, such that $\|x(t_1)\| = \alpha$. Then

$$V(x(t), t) = V(x(t_1), t_1) + \int_{t_1}^{t} \dot{V}(x(\tau), \tau)d\tau, \quad t_1 \leqq t$$

$$\leqq V_M^{\alpha}(t_1) + \int_{t_1}^{t} \dot{V}(x(\tau), \tau)d\tau, \qquad t_1 \leqq t \leqq t_2.$$

Hence

$$V(x(t_2), t_2) \leqq V_M{}^\alpha(t_1) + \int_{t_1}^{t_2} V_f(x(\tau), \tau) d\tau + \int_{t_1}^{t_2} \text{grad } V \cdot u \, d\tau,$$

where $\dot{V}_f$ is as defined in the hypotheses. Making use of hypotheses 1) and 2),

$$V(x(t_2), t_2) < V_M{}^\alpha(t_1) + \int_{t_1}^{t_2} \phi(t) dt + \epsilon \int_{t_1}^{t_2} \|\text{grad } V\| dt$$

$$< V_M{}^\alpha(t_1) + \int_{t_1}^{t_2} [\phi(t) + \epsilon\rho(t)] dt,$$

and finally, by hypothesis 3),

$$V(x(t_2), t_2) < V_m{}^\beta(t_2)$$

which implies that $\|x(t_2)\| \neq \beta$, a contradiction to the original assumption. Hence $t_2 \in \mathfrak{I}$ and therefore $\|x(t)\| < \beta$ for all $t \in \mathfrak{I}$. Since this argument is independent of the exact value of $x(t_0)$ or the particular trajectory chosen, it holds for all trajectories emanating from $B(\alpha)$, and the theorem is proved.

Remark: One of the more interesting aspects of this theorem is the not too surprising fact that, unlike the classical case, there is no requirement of definiteness or semidefiniteness on either V or $\dot{V}$.

The same is true for the following theorems which deal with quasi-contractive and contractive stability of (3).

Theorem 2

A system (3) is quasi-contractively stable under perturbing forces with respect to $(\alpha, \beta, \gamma, \epsilon, t_0, T, \|\cdot\|)$, $\alpha < \beta < \gamma$, if there exists a real-valued function $V(x, t)$, which is C^1 in x and C^0 in t, and four real-valued functions of time $\phi_1, \rho_1, \phi_2, \rho_2$ which are integrable over $\mathfrak{I}$, such that

1) $\|\text{grad } V(x, t)\| \leqq \rho_1(t), \quad t \in \mathfrak{I}, \quad x \in (\overline{B(\gamma)} - B(\alpha))$

2) $\|\text{grad } V(x, t)\| \leqq \rho_2(t), \quad t \in \mathfrak{I}, \quad x \in (\overline{B(\gamma)} - B(\beta))$

3) $V_f(x, t) < \phi_1(t), \quad t \in \mathfrak{I}, \quad x \in (\overline{B(\gamma)} - B(\alpha))$

4) $V_f(x, t) < \phi_2(t), \quad t \in \mathfrak{I}, \quad x \in (\overline{B(\gamma)} - B(\beta))$

5) $\displaystyle\int_{t_1}^{t_2} [\phi_1(t) + \epsilon\rho_1(t)] dt \leqq V_m{}^\gamma(t_2) - V_M{}^\alpha(t_1),$
$t_1, t_2 \in \mathfrak{I}, \; t_2 > t_1$

6) $\displaystyle\int_{t_1}^{t_0+T} [\phi_2(t) + \epsilon\rho_2(t)] dt < V_m{}^\beta(t_0 + T) - V_M{}^\beta(t_1),$
$t_1 \in \mathfrak{I}$

7) $V(x, t_0 + T) \geqq V_m{}^\beta(t_0 + T), \quad x \in (B(\gamma) - B(\beta)).$

Proof: The system is stable under perturbing forces with respect to $(\alpha, \gamma, t_0, T, \|\cdot\|)$ by hypotheses 1), 3), 5). Now, if there is no trajectory $x(t)$ where $\|x(t_0)\| < \alpha$ which passes the boundary of $B(\beta)$, there is nothing to prove.

Hence, consider an arbitrary trajectory $x(t)$ where $\|x(t_0)\| < \alpha$, and suppose there exists $t_1 \in \mathfrak{I}$ such that $\|x(t_1)\| = \beta$, and $\|x(t)\| \geqq \beta$ for all $t \in (t_1, t_0 + T)$. Then,

$$V(x(t_0 + T), t_0 + T) = V(x(t_1), t_1) + \int_{t_1}^{t_0+T} V(x(\tau), \tau) d\tau$$

$$= V(x(t_1), t_1) + \int_{t_1}^{t_0+T} V_f(x(\tau), \tau) d\tau + \int_{t_1}^{t_0+T} \text{grad } V \cdot u \, d\tau.$$

From hypotheses 2), 4), and 6) plus the definition of $V_M{}^\beta$,

$$V(x(t_0 + T), t_0 + T)$$

$$\leqq V_M{}^\beta(t_1) + \int_{t_1}^{t_0+T} \phi_2(\tau) d\tau + \epsilon \int_{t_1}^{t_0+T} \|\text{grad } V\| d\tau$$

$$\leqq V_M{}^\beta(t_1) + \int_{t_1}^{t_0+T} [\phi_2(\tau) + \epsilon \rho_2(\tau)] d\tau$$

$$< V_M{}^\beta(t_1) + V_m{}^\beta(t_0 + T) - V_M{}^\beta(t_1) = V_m{}^\beta(t_0 + T).$$

This is a contradiction by hypothesis 7).

Hence there exists $t_2 \in \mathfrak{I}$ such that $\|x(t)\| < \beta$, $t \in (t_2, t_0 + T)$.

Q.E.D.

Remarks: 1) Hypothesis 6) implies that

$$V_M{}^\beta(t_0 + T) = V_m{}^\beta(t_0 + T).$$

2) If 1) above is difficult to satisfy, one can restate the theorem with δ replacing β in the hypotheses, where $\delta < \beta$.

Theorem 3

A system (3) is contractively stable under perturbing forces with respect to $(\alpha, \beta, \gamma, \epsilon, t_0, T, \|\cdot\|)$, $\beta < \alpha < \gamma$, if there exists a real-valued function $V(x, t)$ and four real-valued functions of time $\phi_1, \rho_1, \phi_2, \rho_2$ which are integrable over $\mathfrak{I}$ such that

1) $\|\text{grad } V(x, t)\| \leqq \rho_1(t), \quad t \in \mathfrak{I}, \quad x \in (\overline{B(\gamma)} - B(\alpha))$

2) $\|\text{grad } V(x, t)\| \leqq \rho_2(t), \quad t \in \mathfrak{I}, \quad x \in (B(\gamma) - B(\beta))$

3) $V_f(x, t) < \phi_1(t), \quad t \in \mathfrak{I}, \quad x \in (\overline{B(\gamma)} - B(\alpha))$

4) $V_f(x, t) < \phi_2(t), \quad t \in \mathfrak{I}, \quad x \in (B(\gamma) - B(\beta))$

5) $\displaystyle\int_{t_1}^{t_2} [\phi_1(t) + \epsilon \rho_1(t)] dt \leqq V_m{}^\gamma(t_2) - V_M{}^\alpha(t_1),$
$t_1, t_2 \in \mathfrak{I}, \; t_2 > t_1$

6) $\int_{t_0}^{t_0+T} [\phi_2(t) + \epsilon \rho_2(t)]dt < V_m^\beta(t_0 + T) - V_{M_0}$

where $V_{M_0} = \max_{x \in (\overline{B(\alpha)} - B(\beta))} V(x, t_0)$

7) $\int_{\tau}^{t_0+T} [\phi_2(t) + \epsilon \rho_2(t)]dt \leqq V_m^\beta(t_0 + T) - V_M^\beta(\tau), \quad \tau \in \mathfrak{J}$

8) $V(x, t_0 + T) \geqq V_m^\beta(t_0 + T), \quad x \in (B(\gamma) - B(\beta))$.

Proof: By hypotheses 1), 3), 5), the system is stable under perturbing forces with respect to $(\alpha, \gamma, t_0, T, \|\cdot\|)$.

Consider an arbitrary trajectory $x(t)$, where $\|x(t_0)\| < \alpha$, and suppose $\|x(t)\| > \beta$ for all $t \in \mathfrak{J}$. Then,

$$\begin{aligned} V(x(t), t) &= V(x(t_0), t_0) + \int_{t_0}^{t} \dot{V}(x(\tau), \tau)d\tau \\ &= V(x(t_0), t_0) + \int_{t_0}^{t} \dot{V}_f(x(\tau), \tau)d\tau \\ &\quad + \int_{t_0}^{t} [\text{grad } V(x(\tau), \tau)] \cdot [u(x(\tau), \tau)]d\tau \\ &\leqq V(x(t_0), t_0) + \int_{t_0}^{t} \dot{V}_f(x(\tau), \tau)d\tau \\ &\quad + \int_{t_0}^{t} \|\text{grad } V(x(\tau), \tau)\| \cdot \|u(x(\tau), \tau)\|d\tau \\ &< V_{M_0} + \int_{t_0}^{t} [\phi_2(\tau) + \epsilon \rho_2(\tau)]d\tau. \end{aligned}$$

Hence, from hypothesis 6),

$$V(x(t_0 + T), t_0 + T) < V_{M_0} + V_m^\beta(t_0 + T) - V_{M_0} = V_m^\beta(t_0 + T).$$

But, by hypothesis 8), this is a contradiction; hence, there exists $t_1 \in \mathfrak{J}$ for which $\|x(t_1)\| < \beta$. The remainder of the proof now follows that of Theorem 2 using hypotheses 2), 4), 8), 7).

V. Finite Time Stability on Product Spaces

One of the desirable goals in the development of any theory of stability is to be able to determine the stability properties of a complicated system by knowing stability properties of lower-order (simpler) subsystems which, when coupled together in appropriate fashion, form the original system.

In general this is rather difficult to achieve, but certain results along this line are immediately available in the case of finite time stability.

Consider a system (1) and suppose the state vector x is partitioned as

$$x = \begin{pmatrix} w \\ z \end{pmatrix}$$

so that (1) can be written as

$$\dot{w} = g(w, z, t) \tag{4a}$$

$$\dot{z} = h(z, w, t). \tag{4b}$$

The question to be answered is the following. If it is known that the systems (4a), (4b) are finite time stable over some given time interval in the sense of Definition 1 (with respect to certain fixed parameters), what does that imply about the finite time stability characteristics of (1) over the same given time interval? Following Lefschetz's [3] terminology for classical stability, this problem is referred to as one in finite time stability on product spaces.

To answer the preceding question precisely, it is convenient to define a concept of finite time stability for system (1) which takes account of the decomposition (4a), (4b).

Let $\|\cdot\|^*$ be a functional on R^n such that

$$\|x(t)\|^* = \|w(t)\|^* + \|z(t)\|^*.$$

For example, $\|\cdot\|^*$ might represent the square of the Euclidian norm or the sum of the absolute values of the components of $x(t)$.

Let $D_{a\times b} \subset R^n$ be a set such that

$$x(t) \in D_{a\times b} \Rightarrow \|x(t)\|^* < a + b, \ \|w(t)\|^* < a, \ \|z(t)\|^* < b.$$

Then the following definition is given:

Definition 4

A system (1) *with decomposition* (4a), (4b) *is stable with respect to* $(D_{\alpha_1\times\alpha_2}, D_{\beta\times\gamma}, t_0, T, \|\cdot\|^*)$, $\alpha_1 \leqq \beta$, $\alpha_2 \leqq \gamma$, if for any trajectory $x(t)$, $x(t_0) \in D_{\alpha_1\times\alpha_2}$ implies $x(t) \in D_{\beta\times\gamma}$ for all $t \in \mathfrak{J}$.

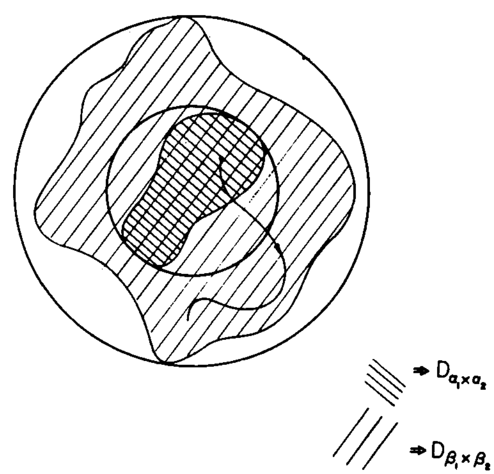


Fig. 2. Stability—Def. 4.

Stability in the sense of Definition 4 is illustrated in Fig. 2.

Contractive stability can also be defined in this context as follows:

Definition 5

A system (1) *with decomposition* (4a), (4b) *is contractively stable with respect to* $(D_{\alpha_1 \times \alpha_2}, D_{\beta_1 \times \beta_2}, D_{\gamma_1 \times \gamma_2}, t_0, T, \|\cdot\|^*)$, $\beta_1 < \alpha_1 \leq \gamma_1$, $\beta_1 < \alpha_2 \leq \gamma_2$, if

1) it is stable with respect to $(D_{\alpha_1 \times \alpha_2}, D_{\gamma_1 \times \gamma_2}, t_0, T, \|\cdot\|^*)$

2) for every trajectory $x(t)$ where $x(t_0) \in D_{\alpha_1 \times \alpha_2}$, there exists $t_1 \in (t_0, t_0+T)$ such that $x(t) \in D_{\beta_1 \times \beta_2}$ for all $t \in (t_1, t_0+T)$.

The following results are easy consequences of the above definitions.

Theorem 4

If (4a) is stable under perturbing forces with respect to $(\alpha_1, \beta, \gamma, t_0, T, \|\cdot\|^*)$, $\alpha_1 \leq \beta$, and (4b) is stable under perturbing forces with respect to $(\alpha_2, \gamma, \beta, t_0, T, \|\cdot\|^*)$, $\alpha_2 \leq \gamma$, then the system (1) with decomposition (4a), (4b) is stable with respect to $(D_{\alpha_1 \times \alpha_2}, D_{\beta \times \gamma}, t_0, T, \|\cdot\|^*)$.

Proof: Let $x(t)$ be an arbitrary trajectory of (1) such that $x(t_0) \in D_{\alpha_1 \times \alpha_2}$. We are considering $z(t)$ in (4a) and $w(t)$ in (4b) as forcing functions. Assume there exists $t_1 \in \mathfrak{I}$ the first such time at which $\|z(t_1)\|^* = \gamma$. Then $\|z(t)\|^* < \gamma$ for $t_0 \leq t < t_1$. But by the first stated hypothesis of the theorem, this implies that $\|w(t)\|^* < \beta$ for all $t \in [t_0, t_1]$. Then if $t_2 \in \mathfrak{I}$ is the first value of t for which $\|w(t)\|^* = \beta$, obviously $t_2 > t_1$.

Now consider (4b). By the second hypothesis of the theorem plus the above fact, $\|z(t_1)\|^* < \gamma$. This contradicts the earlier assumption about t_1, i.e., that $t_1 \in \mathfrak{I}$. Therefore $\|w(t)\|^* < \beta$, $\|z(t)\|^* < \gamma$ for all $t \in \mathfrak{I}$ and so $x(t) \in D_{\beta \times \gamma}$ for all $t \in \mathfrak{I}$.

Theorem 5

If (4a) is contractively stable under perturbing forces with respect to $(\alpha_1, \beta_1, \gamma, \delta, t_0, T, \|\cdot\|^*)$ and (4b) is contractively stable under perturbing forces with respect to $(\alpha_2, \beta_2, \delta, \gamma, t_0, T, \|\cdot\|^*)$, $\beta_2 < \alpha_2 \leq \delta$, then the system (1) with decomposition (4a), (4b) is contractively stable with respect to $(D_{\alpha_1 \times \alpha_2}, D_{\gamma \times \delta}, D_{\beta_1 \times \beta_2}, t_0, T, \|\cdot\|^*)$.

Proof: Stability of (1) with respect to $(D_{\alpha_1 \times \alpha_2}, D_{\gamma \times \delta}, t_0, T, \|\cdot\|^*)$ follows immediately from Theorem 4.

Now let $x(t)$ be any trajectory of (1) such that $x(t_0) \in D_{\alpha_1 \times \alpha_2}$. From the hypothesis on (4a), there exists $t_1 \in (t_0, t_0+T)$ such that $\|w(t)\|^* < \beta_1$ for all $t \in (t_1, t_0+T)$; and from the hypothesis on (4b), there exists $t_2 \in (t_0, t_0+T)$ such that $\|z(t)\|^* < \beta_2$ for $t \in (t_2, t_0+T)$. Hence $x(t) \in D_{\beta_1 \times \beta_2}$ for all $t \in (\max(t_1, t_2), t_0+T)$.

Remarks: 1) A definition and theorem can easily be stated which is analogous to Definition 5 and Theorem 5, for quasi-contractive stability. This is left as an exercise for the reader.

2) The usefulness of Theorems 4 and 5 depends, of course, on having means available for testing (4a) and (4b) for finite time stability under perturbing forces. In the specific case where these forces enter the system equation linearly as in (3), i.e., in the case where (4a) and (4b) are *linearly* coupled to form (1), Theorem 1 (or 2 or 3 depending on what is being sought) can, in principle, be utilized to obtain information about stability under perturbing forces for (4a) and (4b). A straightforward application of Theorem 4 or 5 then yields the desired information about the finite time stability characteristics of (1).

VI. An Example

The following simple example, a modification of one given by Cesari [4], is presented to illustrate some of the salient features of finite time stability. Consider the system

$$\dot{x}_1 = f_1(x_1, x_2, t) + u_1(x_1, x_2, t)$$
$$\dot{x}_2 = f_2(x_1, x_2, t) + u_2(x_1, x_2, t).$$

Let the system, with the perturbing terms u_i set to zero, be such that it can be written in the form

$$\frac{\dot{r}}{r} = \frac{1}{h(t, \phi)} \frac{\partial h(t, \phi)}{\partial t}, \qquad \dot{\phi} = 0$$

where $h = 1 + t^3 \sin^2 \phi / 1 + t + t^4 \sin^4 \phi$. It can be easily seen that this unperturbed system is unstable in the classical sense. Yet, this set of equations displays characteristics which, over finite periods of time, make it resemble a stable system. In fact, the above perturbed system is contractively stable with respect to (α, $\alpha/2$, 2α, $\sqrt{2}\alpha/10$, 0, $\sqrt{2}/3$, Euclidian norm). This can be seen by making use of Theorem 3 and letting $V = x_1^2 + x_2^2$, $\rho_1(t) = \rho_2(t) = 4\alpha$, and $\phi_1(t) = \phi_2(t) = 8\alpha^2(-1 + t + 2t^2)$. (The reader can easily check that conditions 5), 6), 7), and 8) of the theorem are satisfied.)

VII. Conclusions

The qualitative theory of finite time stability has been extended, in this paper, to systems under the influence of external forces. In so doing, the following important facts should be noted:

1) The definitions of stability upon which the theory is based are of a much more practical nature than those of classical stability. Moreover, as indicated earlier, and particularly in the example in Section VI, systems which are stable in the classical sense may be unstable in the finite time sense and vice versa. (It might well be desirable to put a rocket into an unstable orbit if that orbit is particularly well suited for performance of certain experiments, providing the orbit is finite time stable over the interval of time needed to complete the experiments.)

2) The sufficient conditions given for determining finite time stability under perturbing forces involve the existence of "Liapunov-like" functions, whose required properties are significantly less stringent (e.g. from the point of view of computerization) than those for class-

ical Liapunov functions. Furthermore, the problem of finite time stability with respect to an analytically or numerically defined trajectory can easily be handled within the format of the theorems presented in this paper without any recourse to the complicated transformations which are needed in the classical theory.

3) It should be apparent that *any* example in classical stability in which a Liapunov function is exhibited can be converted into a finite time stability example in which the Liapunov function plays the role of the function V in our theorems.

4) All the theorems in this paper yield sufficient conditions for finite time stability. To date, no converse theorems have been developed for this type of stability so that the determination of the necessity of the stated hypotheses is an open research problem.

References

[1] J. P. LaSalle and S. Lefschetz, *Stability by Liapunov's Direct Method with Applications*. New York: Academic, 1961.

[2] Leonard Weiss and E. F. Infante, "On the stability of systems defined over a finite time interval," *Proc. Nat'l Acad. Sci.*, vol. 54, pp. 44–48, July 1965.

[3] S. Lefschetz, *Differential Equations: Geometric Theory*. New York: Interscience, 1957.

[4] L. Cesari, *Asymptotic Behavior and Stability Problems in Ordinary Differential Equations*. Berlin: Springer-Verlag, 1953.

PAPER [4]

STABILITY PROBLEMS OF SOLUTIONS OF DIFFERENTIAL EQUATIONS,
Proceedings of the NATO Advanced Study Institute, Padua,
Italy, Edizioni "Oderisi", Gubbio, 1966, 95-106.

LIAPUNOV'S SECOND METHOD*

by

J. P. LaSalle

Center for Dynamical Systems
Brown University

1. Introduction.

In the preceding paper [1] an introduction to Liapunov's second or direct method was given based on ideas originally introduced by the author of this paper, and we wish to continue in that direction. We shall give first of all a statement of what can be called the fundamental theorem of stability (Theorem 1) which extends somewhat and includes the results of [1]. This fundamental theorem is based on a broader definition of a Liapunov function and makes use of the invariance property of limit sets of solutions of autonomous differential equations. It also has an important bearing on the extension of stability theory to more general dynamical systems and to applications of the theory.

By means of a simple example we will illustrate that this theorem takes us beyond the classical theory of Liapunov and shows how one may study the qualitative behavior of systems in the large. The techniques are not unknown but Theorem 1 now brings them within

*This research was supported in part by the National Aeronautics and Space Administration under Grant No. NGR-40-002-015 and under Contract No. NAS8-11264, in part by the United States Air Force through the Air Force Office of Scientific Research under Grant No. AF-AFOSR-693-65 and in part by the United States Army Research Office, Durham, under Contract No. DA-31-124-ARO-D-270.

the domain of Liapunov's method while at the same time unifying the whole theory. The previous paper [1] demonstrated this unification for theorems on stability. We will indicate -- and this is shown also by the example -- how one obtains from this same fundamental theorem criteria for instability. Up to this point we shall have confined ourselves to autonomous systems and basic to Theorem 1 is the fact that the limit sets of solutions of autonomous systems are invariant sets. We certainly expect therefore that Theorem 1 will have to be modified for nonautonomous systems and following Yoshizawa in [2] we give in Theorem 2 the analogous theorem for nonautonomous systems. As to be expected the information given by Liapunov functions is now less precise but by means of an example it is shown that the conclusion of the theorem given here is, however, the "best possible". There are types of nonautonomous systems where the limit sets of solutions have an invariance property that enables one to improve Theorem 2. This is discussed in Section 3.

More recently Hale in [3] has shown that properly interpreted the solutions of autonomous functional differential equations have limit sets which are invariant. With modifications this gives him a stability theory quite similar to that for autonomous differential equations. Functional differential equations which include delay-differential equations are mathematical models for systems whose future behavior depends upon a portion or all of its past history.

They can be expected to be of increasing importance in economics, biology, and control. Hale's work carries us so far beyond our geometric intuition that it is here that we can appreciate the necessity of a theory to guide us and his work suggests how the theory can be developed for more general dynamical systems. Since his paper [3] is complete, and is well illustrated by examples, his results are not summarized here.

2. Autonomous systems.

For the sake of simplicity we shall assume with some exceptions that all functions introduced are C^1 and as much as possible adopt the notations and definitions of [1]. With f an arbitrary C^1 function on R^n to R^n we consider first the ordinary differential equation $(\dot{x} = \frac{dx}{dt})$

$$\dot{x} = f(x) \; . \tag{1}$$

In order not to have to confine ourselves to bounded solutions we compactify R^n by adding the point at infinity where the distance $d(\infty, x)$ of x to infinity is $|x|^{-1}$. Thus, if Γ is a set in R^n and we define $\Gamma^* = \Gamma \cup \{\infty\}$, then a function $\varphi(t)$ is said to <u>approach</u> Γ^* if $d(\Gamma^*, \varphi(t)) \to 0$ as $t \to \infty$. This also gives a meaning to the statement that ∞ is a limit point of $\varphi(t)$, which is not necessarily the same as saying $\varphi(t) \to \infty$ as $t \to \infty$. When

$\varphi(t)$ is a solution of (1) it may happen that its maximum positive interval of definition is $[0, \tau)$. This causes no difficulty. We need only replace ∞ by τ. Understanding this we will usually ignore this point and speak as though all solutions are defined on $(-\infty, \infty)$.

Let G be an arbitrary set in R^n and let V be a C^1 function on R^n to R. We shall say that V is a Liapunov function on G for the system (1) if $\dot{V} = (\text{grad } V)\cdot f$ does not change sign on G. We define ($\overline{G}$ is the closure of G)

$$E = \{x \; ; \; \dot{V}(x) = 0, \; x \in \overline{G}\} \; ;$$

M will denote the largest invariant set in E and $M^* = M \cup \{\infty\}$. It then follows easily from the invariance property of limit sets of solutions of (1) that

Theorem 1. If V is a Liapunov function on G for the system (1) and if a solution $x(t)$ of (1) remains in G for all $t > 0$ $(t < 0)$, then $x(t)$ approaches M^* as $t \to \infty$ $(t \to -\infty)$. If M is bounded, then either $x(t) \to M$ or $x(t) \to \infty$ as $t \to \infty$ $(t \to -\infty)$.

This theorem states that a Liapunov function V on G locates all possible positive and negative limit sets of solutions which remain in G for $t > 0$ or $t < 0$. The problem in applying the theorem is to find a "good" Liapunov function. A constant function

V is always a Liapunov function for the whole state space R^n but, of course, gives no information. Here $E = M = R^n$. The theorem does, however, make it possible to obtain more information about the asymptotic behavior of systems with Liapunov functions not as severly restricted as those of the classical theory. It is also true that every C^1 function V is a Liapunov function on the region $\dot{V} \leqq 0$ (or $\dot{V} \geqq 0$) but this may or may not be helpful. The following simple example illustrates some features of this result and how it may be applied and how one obtains additional information by using more than one Liapunov function. It is not always this easy, and this example was manufactured for this purpose. In actuality it is often easier using Liapunov functions to synthesize a system to have a particular behavior than it is to analyze a given system, and this is proving to be true in the design of control systems.

The second order system

$$\text{(2)} \qquad \begin{aligned} \dot{x} &= -2xy \\ \dot{y} &= -x + y + xy - y^3 \end{aligned}$$

has three equilibrium points: $P_1 = (0, 1)$, $P_2 = (0, -1)$ and $P_3 = (0, 0)$. The eigenvalues of the linear approximation about P_1 are $-2, -2$; about P_2 they are $-2, 2$ and about P_3 are $0, 1$. Thus P_1 is asymptotically stable, P_2 is a saddle point and is unstable, and P_3 is unstable. The linear approximation does not given any information about the region of attraction (of asymptotic stability) about P_1 or about the character of the equilibrium point P_3. (See Figure 1.)

We have first of all for each of the four quadrants the obvious Liapunov function $V_1 = x$ since $\dot{V} = -2xy$. For each of these quadrants E_1 is the union of the x- and y-axes, and since $\dot{x} = 0$ when $x = 0$ and $\dot{y} = -x$ when $y = 0$, we see that M_1 is the y-axis. It is clear, for example, that no solution starting in the 4th quadrant can leave for $t > 0$ and cannot have a limit point on M_1. Hence all solutions starting in the 4th quadrant approach ∞ as $t \to \infty$.

Another Liapunov function is $V_2 = x - y^2$; $\dot{V}_2 = 2y^2(y^2-x-1) =$ $= -2y^2(V_2 + 1)$ and V_2 is a Liapunov function for the regions $G_1 : V_2 < -1$ and $G_2 : V_2 > -1$. Here E_2 is the x-axis and the parabola $V_2 = -1$, which is an integral, and M_2 is the curve $V_2 = -1$ and the origin P_3. The region G_1 and G_2 are invariant sets. In G_1, $\dot{V}_2 > 0$ and no solution can approach M_2 as $t \to -\infty$. Therefore every solution starting in G_1 approaches ∞ as $t \to -\infty$. Note next that each solution starting in $x < 0$ remains in this region and is bounded for $t > 0$. Therefore the only possible positive limit points are the intersection of M_1 and M_2 which consists of the three equilibrium points P_1, P_2 and P_3. To the left of P_2, $\dot{x} < 0$ and to the left of P_3, $\dot{V}_2 < 0$ so that every solution starting in the left-half plane $x < 0$ must approach P_1 as $t \to \infty$. Similarly, one can see that every solution starting in this half-plane inside G_2 approaches P_3 as $t \to -\infty$. Also it is easy to see that every solution in the 1st quadrant above $V_2 = 0$ approaches P_1 as $t \to \infty$. Hence the 2nd and 3rd quadrants and this portion of the 1st quadrant are in the region of attraction of P_1. Below $V_2 = 0$ in the 1st quadrant there must be a solution which approaches P_3 as $t \to \infty$ and this solution is the boundary of the region of attraction of P_1. We know this must happen since the boundary of the region of attraction is an invariant set and the region of attraction does not include the 4th quadrant.

The following corollary is a direct consequence of Theorem 1 and illustrates how instability results can be obtained:

Corollary 1. Assume inside a set G that $V \dot{V} > 0$ and on the boundary of G that $V = 0$. Then every solution of (1) starting in G approaches ∞ as $t \to \infty$ (or possibly in finite time).

Proof: The assumptions imply that every solution starting in G remains inside G for $t > 0$ and in fact cannot even have a positive limit point on the boundary of G. Since $G \cap M$ is the empty set, it must be that every solution approaches ∞ as $t \to \infty$ (it could have finite escape time).

In a manner similar to the above proof one can obtain

Četaev's instability theorem as a corollary of Theorem 1.

Corollary 2. Let G_o be an open set, let p be an equilibrium point on the boundary of G_o, and let N be a neighborhood of p. If $V(x)\dot{V}(x) > 0$ for x in $G = G_o \cap N$ and $V(x) = 0$ for x on the boundary of G_o inside N, then p is unstable.

From the point of view of applications the following is one of the most useful results.

Corollary 3. Assume that a component G of the set defined by $V(x) < \ell$ is bounded, $\dot{V}(x) \leq 0$ for $x \in G$, and $M^o \subset G$ where $M^o = \overline{M \cap G}$. Then M^o is an attractor as $t \to \infty$ and G is in the region of attraction to M^o. If V is constant on the boundary of M^o, then M^o is a stable attractor (is asymptotically stable).

Thus in the above corollary when M^o is a single point p, V is constant on M^o and the point p will be asymptotically stable with G providing an estimate of its stability. This is without any assumption that V be positive definite. However, in applying this theorem where the Liapunov function is itself to provide a positively invariant set one will usually look for a Liapunov function that is positive definite relative to p. Unless the set E where $\dot{V}$ vanishes contains a positively invariant set other than p, the point p will be a minimum of V so for this purpose one might expect "good" Liapunov functions to be positive definite. On the other hand the simple example above demonstrated this may not always be the best procedure and one can often do better using more than one Liapunov function none of which need be positive definite.

3. Nonautonomous systems.

In this section we follow fairly closely the ideas of Yoshizawa in [2] although we will not present them with as great a generality as he acheived. We concern ourselves with the system

$$(3) \qquad \dot{x} = f(t, x)$$

where f is continuous for (t, x) in $\mathscr{D} = [0, \infty) \times R^n$ and is C^1 on $\mathscr{D}$ with respect to x (or any other of the known conditions that imply existence and uniqueness of solutions). Here limit sets of solutions are still defined but they will not in general be invariant sets. Hence we cannot expect a result as strong as Theorem 1. Theorem 2 below is a modified version of Theorem 1 and is closely related to Yoshizawa's Theorem 6 in [2].

Let $V(t, x)$ be a C^1 function on $[0, \infty) \times R^n$ to R. We shall say that V is a <u>Liapunov function on</u> a set G of R^n if $V(t, x) \geq 0$ and $\dot{V}(t, x) \leq -W(x) \leq 0$ for all $t > 0$ and all x in G where W is continuous on R^n to R. We define

$$(4) \qquad E = \{x \; ; \; W(x) = 0, \; x \in \overline{G}\} .$$

Here $\dot{V} = \frac{\partial V}{\partial t} + \sum_{i=1}^{n} \frac{\partial V}{\partial x_i} f_i(t, x)$.

We then have

Theorem 2. If V is a Liapunov function on G for equation (3), then each solution $x(t)$ of (3) that remains in G for all $t > t_0$ approaches $E^* = E \cup \{\infty\}$ as $t \to \infty$, provided one of the following conditions is satisfied:

(i) For each $p \in \overline{G}$ there is a neighborhood N of p such that $|f(t, x)|$ is bounded for all $t > 0$ and all x in N

(ii) W is C^1 and $\dot{W}(t, x)$ is bounded from above or below along each solution which remains in G for all $t > 0$.

If E is bounded, then each solution of (3) remaining in G for $t > 0$ either approaches E or ∞ as $t \to \infty$.

Thus, this theorem is quite similar to Theorem 1 except that M is replaced by the set E. E is in general larger than M and the information given is not as precise. Condition (i) is essentially the same as that used by Yoshizawa. The following example illustrates a case where (ii) is satisfied and (i) is not and also shows that in general even for linear nonautonomous systems this is the best result one can hope to have.

Consider $\ddot{x} + p(t)\dot{x} + x = 0$ where $p(t) \geq \delta > 0$. An equivalent system is

$$\dot{x} = y$$
$$\dot{y} = -x - p(t)y .$$

Since we do not assume that $p(t)$ is bounded from above, condition (i) is not satisfied. With $2V = x^2 + y^2$, $\dot{V} = -p(t)y^2 \leq -\delta y^2$. Thus V is a Liapunov function on the entire state space R^2 and $W = \delta y^2$. It is then clear that each solution is bounded for $t > 0$. Now $\dot{W} = 2\delta y\dot{y} = -2\delta(xy + y^2p(t)) < -2\delta xy$. Hence condition (ii) is satisfied. E corresponds to $y = 0$ and we can conclude that for each solution $y(t) = \dot{x}(t) \to 0$ as $t \to \infty$. Since the equation $\ddot{x} + (2 + e^t)\dot{x} + x = 0$ has a solution $x(t) = 1 + e^{-t}$, this we see is the best possible result without further restrictions on $p(t)$. It also shows that Theorem 1 is not true for nonautonomous systems. Here M is the origin and if Theorem 1 held for nonautonomous systems this would imply that the origin is asymptotically stable which in the example it certainly is not.

In using Theorem 2 it is necessary to be able to identify solutions which remain in G for all positive t. We now look at this problem. If the Liapunov function $V(x)$ does not depend on t, define

$$Q_\ell = \{x \; ; \; V(x) \leq \ell\} .$$

It is then clear that the following is true:

Lemma 1. If $\dot{V}(x) \leqq 0$ for all $t \geqq 0$ and all x in G and Ω_ℓ is a component of Q_ℓ which is contained in G, then each solution $x(t)$ of (3) starting in Ω_ℓ at some time $t_0 \geqq 0$ remains in Ω_ℓ for all $t > t_0$.

If the Liapunov function $V(t, x)$ depends on t, define

$$Q_\ell = \{x \; ; \; V(t, x) \leqq \ell \quad \text{for all} \quad t \geqq 0\}$$
$$Q_\ell^0 = \{x \; ; \; V(0, x) \leqq \ell\}$$
$$Q_\ell^+ = \{x \; ; \; V(t, x) \leqq \ell \quad \text{for some} \quad t \geqq 0\}.$$

It is clear that $Q_\ell \subset Q_\ell^0 \subset Q_\ell^+$. Let Ω_ℓ denote a component of Q_ℓ; then Ω_ℓ^0 will be the component of Q_ℓ^0 and Ω_ℓ^+ will be the component of Q_ℓ^+ which contain Ω_ℓ. We then have

Lemma 2. If $\dot{V}(t, x) \leqq 0$ for all $t \geqq 0$ and all $x \in G$ and Ω_ℓ^+ is contained in G then

a. Each solution starting in Ω_ℓ^0 at time $t = 0$ remains in Ω_ℓ^+ for all $t \geqq 0$.

b. Each solution starting in Ω_ℓ at any time $t_0 \geqq 0$ remains in Ω_ℓ^+ for all $t > t_0$.

These two lemmas combined with Theorem 2 give methods for estimating the region of attraction of equilibrium points of non-autonomous systems and for studying their asymptotic behavior in general. One can also derive from these results sufficient conditions for instability but it still remains true that nonautonomous systems are more difficult to study and relatively few significant problems have been solved.

3. Special classes of nonautonomous systems.

Although we cannot in general expect to go beyond Theorem 2 for nonautonomous systems there are some types of such systems where the invariance properties of the limit sets of their solutions enable us to obtain precise information on their asymptotic behavior using Liapunov functions. The simpliest of these are periodic systems (see [4])

$$\dot{x} = f(t, x) \tag{5}$$

where $f(t + T, x) = f(t, x)$ for all t and x. Here the limit sets of solutions have an invariance property somewhat different from autonomous system. Suppose that $\Gamma \subset R^n$ is a limit set of a solution $x(t)$ of (5). Then Γ is invariant in the following sense: if p is contained in Γ, then there is a solution of (5) which remains in Γ for all t in $(-\infty, \infty)$. This means that if one starts a solution at p at the proper time it will remain in Γ for all t. However, this is sufficient to obtain a theorem quite similar to Theorem 1.

If $V(t, x)$ is C^1 on $R \times R^n$ and is periodic of period T and G is an arbitrary set in R^n, we say that V is a <u>Liapunov function on</u> G for the periodic system (5) if $\dot{V}$ does not change sign for x in G and all t. Define $E = \{(t, x); \dot{V}(t, x) = 0, x \in \overline{G}\}$ and let M be the union of all solutions $x(t)$ of (5) with the property that $(t, x(t))$ is in E for all t. M is called the <u>largest invariant set relative to</u> E. One then obtains the following theorem for periodic systems:

Theorem 3. If V is a Liapunov function on G for the periodic system (5), then each solution of (5) which remains in G for all $t > 0$ $(t < 0)$ approaches $M^* = M \cup \{\infty\}$ as $t \to \infty$ $(t \to -\infty)$. If M is bounded, then either $x(t) \to M$ or $x(t) \to \infty$ as $t \to \infty$ $(t \to -\infty)$.

Recently in [5] Miller has shown that the limit sets of almost periodic systems have an invariance property and one then obtains a similar theorem for almost periodic systems. These results provide improved methods for studying these classes of nonautonomous systems. This periodic version and Miller's almost periodic version of Theorem 1 are not as well known as they should be in spite of the fact it would seem that the difficulty in applying them is not much greater than for autonomous system.

A simple example is the following:

$$\dot{x} = y$$

$$y = -(a + \cos t)x - by .$$

With

$$V = x^2 + (a + \cos t)^{-1}y^2 ,$$

$$\dot{V} = -\frac{1}{2}(a + \cos t)^{-1}\left(2b - \frac{\sin t}{a + \cos t}\right)y^2 ,$$

If $a > 1$ and $2b\sqrt{a^2-1} > 1$, then $\dot{V} \leq 0$ and V is a Liapunov function on the plane R^2. The form of V implies that the origin is stable and that all solutions are bounded for $t > 0$. Here $E = \{(t, x, 0); -\infty < t < \infty , -\infty < x < \infty\}$ but M is simply the origin. Therefore for $a > 1$ and $2b\sqrt{a^2-1} > 1$ the origin is asymptotically stable in the large.

As has been shown by Opial in [6] and Markus in [7] the solution of what may be called "asymptotically autonomous" systems have limit sets with an invariance property which we will explain in a minute. In [2] Yoshizawa used this invariance property and obtained a result similar to Theorem 4 below.

A system of the form

(6) $$\dot{x} = f(t, x) = F(x) + f_1(t, x) + f_2(t, x)$$

will be said to be <u>asymptotically</u> <u>autonomous</u> if (i)(Markus) $f_1(t, x)$ approaches zero as $t \to \infty$ uniformly for x in an arbitrary compact set of R^n, (ii)(Opial) $\int_0^\infty |f_2(t, \varphi(t)|dt < \infty$ for all $\varphi(t)$ continuous and bounded on $[0, \infty)$ to R^n. The combined results of Markus and Opial then state that the positive limit sets of solutions of (6) are invariant sets of $\dot{x} = F(x)$. This then leads immediately, as a consequence of Theorem 2 to the following:

Theorem 4. If V is a Liapunov function on G for the asymptotically autonomous system (6), then each solution of (6) which remains in G for all $t > 0$ approaches $M^* = M \cup \{\infty\}$, where M is the largest invariant set of $\dot{x} = F(x)$ in E, provided f_2 satisfies condition (i) of Theorem 2 or W satisfies condition (ii) of Theorem 2.

It turns out to be useful in order to apply this result to nonautonomous systems (3) which are not asymptotically autonomous to give also the following version of this theorem.

Theorem 4. If in addition to the conditions of Theorem 2 it is known that the positive limit set of $x(t)$ is an invariant set of $\dot{x} = G(x)$, then $x(t) \to M^* = M \cup \{\infty\}$ where M is the largest invariant set of $\dot{x} = G(x)$ in E.

The example

$$\text{(7)} \qquad \begin{aligned} \dot{x} &= y \\ \dot{y} &= - x - p(t)y \qquad , \; 0 < \delta \leq p(t) \end{aligned}$$

considered before can again be used to illustrate the above theorem and to show how it can be applied even when the original system is not asymptotically autonomous. Let $(\overline{x}(t), \overline{y}(t))$ be any solution of (7). As shown previously we know it is bounded for $t > 0$ and that $\overline{y}(t) \to 0$ as $t \to \infty$. Assume now in addition that $p(t)$ is bounded from above: $0 < \delta \leq p(t) \leq m$ for all $t > 0$. Then consider for this particular solution the system

$$\dot{x} = y$$
$$\dot{y} = - x - p(t)\overline{y}(t) .$$

Certainly $\overline{x}(t), \overline{y}(t)$ is a solution, and this system is asymptotically autonomous to (*) $\dot{x} = y$, $\dot{y} = - x$. Therefore the positive limit set of $(\overline{x}(t), \overline{y}(t))$ is an invariant set of (*) and must also lie on the x-axis. Hence its positive limit set is the origin. This means that when $0 < \delta \leq p(t) \leq m$ for all $t > 0$ the system (7) is asymptotically stable in the large.

References

[1] Strauss, Aaron, "A geometric introduction to Liapunov's second method", this volume, (Proc. of NATO SUMMER Instit., Padua, Italy, 1965).

[2] Yoshizawa, T., "Asymptotic behavior of solutions of a system of differential equations", Contrib. to Diff. Eq., 1(1963), 371-387.

[3] Hale, J., "Sufficient conditions for stability and instability of autonomous functional differential equations", J. of Differential Equations, 1(1965), 452-482.

[4] LaSalle, J. P., "Asymptotic stability criteria",Proc. of Symposia in Applied Math., Vol. 13, Hydrodynamic Instability, Amer. Math. Soc., 1962, 299-307.

[5] Miller, R. "On almost periodic differential equations", Bul. Amer. Math. Soc., 70(1964), 792-794.

[6] Opial, Z., "Sur la dépendence des solutions d'un système d'équations differentielles de leur second membres. Application aux systèmes presque autonomes", Ann. Polonici Math., 8(1960), 75-89.

[7] Markus, L., "Asymptotically autonomous differential equations", Contrib. to the Theory of Nonlinear Oscillations, Vol. III, Annals of Math. Studies No. 36, Princeton University, 1956, 75-89.

PAPER [5]

DIFFERENTIAL EQUATIONS AND DYNAMICAL SYSTEMS, Proceedings of the International Symposium, Puerto Rico, Academic Press, New York, 1967, 277-286 and CDS-TR 66-1.

AN INVARIANCE PRINCIPLE IN THE THEORY OF STABILITY

by

J. P. LaSalle

Center for Dynamical Systems
Brown University

1. Introduction.

The purpose of this paper is to give a unified presentation of Liapunov's theory of stability that includes the classical Liapunov theorems on stability and instability as well as their more recent extensions. The idea being exploited here had its beginnings some time ago. It was, however, the use made of this idea by Yoshizawa in [1] in his study of nonautonomous differential equations and by Hale in [2] in his study of autonomous functional differential equations that caused the author to return to this subject and to adopt the general approach and point of view of this paper. This produces some new results for dynamical systems defined by ordinary differential equations which demonstrate the essential nature of a Liapunov function and which may be useful in applications. Of greater importance, however, is the possibility, as already indicated by Hale's results for functional differential equations,

*This research was supported in part by the National Aeronautics and Space Administration under Grant No. NGR-40-002-015 and under Contract No. NAS8-11264, in part by the United States Air Force through the Air Force Office of Scientific Research under Grant No. AF-AFOSR-693-65 and in part by the United States Army Research Office, Durham, under Contract No. DA-31-124-ARO-D-270.

that these ideas can be extended to more general classes of dynamical systems. It is hoped, for instance, that it may be possible to do this for some special types of dynamical systems defined by partial differential equations.

In section 2 we present some basic results for ordinary differential equations. Theorem 1 is a fundamental stability theorem for nonautonomous systems and is a modified version of Yoshizawa's Theorem 6 in [1]. A simple example shows that the conclusion of this theorem is the best possible. However, whenever the limit sets of solutions are known to have an invariance property then sharper results can be obtained. This "invariance principle" explains the title of this paper. It had its origin for autonomous and periodic systems in [3] - [5], although we present here improved versions of those results. Miller in [6] has established an invariance property for almost periodic systems and obtains thereby a similar stability theorem for almost periodic systems. Since little attention has been paid to theorems which make possible estimates of regions of attraction (regions of asymptotic stability) for nonautonomous systems results of this type are included. Section 3 is devoted to a brief discussion of some of Hale's recent results [2] for autonomous functional differential equations.

2. Ordinary differential equations.

Consider the system

$$\dot{x} = f(t,x) \qquad (1)$$

where x is an n-vector, f is a continuous function on R^{n+1} to R^n and satisfies any one of the conditions guaranteeing uniqueness of solutions. For each x in R^n we define $|x| = (x_1^2 + \dots + x_n^2)^{\frac{1}{2}}$; and for E a closed set in R^n we define $d(x,E) = \text{Min}\,\{|x-y| : y \text{ in } E\}$. Since we do not wish to confine ourselves to bounded solutions, we introduce the point at ∞ and define $d(x,\infty) = |x|^{-1}$. Thus when we write $E^* = E \cup \{\infty\}$, we shall mean $d(x,E^*) = \text{Min}\{d(x,E), d(x,\infty)\}$. If $x(t)$ is a solution of (1), we say that $x(t)$ approaches E as $t \to \infty$ if $d(x(t),E) \to 0$ as $t \to \infty$. If we can find such a set E, we have obtained information about the asymptotic behavior of $x(t)$ as $t \to \infty$. The best that we could hope to do is to find the smallest closed set Ω that $x(t)$ approaches as $t \to \infty$. This set Ω is called the <u>positive limit set</u> of $x(t)$ and the points p in Ω are called the <u>positive limit points</u> of $x(t)$. In exactly the same way one defines $x(t) \to E$ as $t \to -\infty$, negative limit sets, and negative limit points. This is exactly G. D. Birkhoff's concept of limit sets. A point p is a positive limit point of $x(t)$ if and only if there is a sequence of times t_n approaching ∞ as $n \to \infty$ and such that $x(t_n) \to p$ as $n \to \infty$. In the above it may be that the maximal interval of definition of $x(t)$ is $[0,\tau)$. This causes no difficulty since in the results to be presented here we need only with respect to time t replace ∞ by τ. We usually ignore this possibility and speak as though our solutions are defined on $[0,\infty)$ or $(-\infty,\infty)$.

Let $V(t,x)$ be a C^1 function on $[0,\infty) \times R^n$ to R, and let G be any set in R^n . We shall say that V is a Liapunov function on G for equation (1) if $\dot{V}(t,x) \geqq 0$ and $V(t,x) \leqq -W(x) \leqq 0$ for all $t > 0$ and all x in G where W is continuous on R^n to R and

$$\dot{V} = \frac{\partial V}{\partial t} + \sum_{i=1}^{n} \frac{\partial V}{\partial x_i} f_i \ .$$

We define ($\overline{G}$ is the closure of G)

$$E = \{x, W(x) = 0, \ x \text{ in } \overline{G}\}.$$

The following result is then a modified but closely related version of Yoshizawa's Theorem 6 in [1].

THEOREM 1. If V is a Liapunov function on G for equation (1), then each solution $x(t)$ of (1) that remains in G for all $t > t_0 \geqq 0$ approaches $E^* = E \cup \{\infty\}$ as $t \to \infty$, provided one of the following conditions is satisfied:

(i) For each p in $\overline{G}$ there is a neighborhood N of p such that $|f(t,x)|$ is bounded for all $t > 0$ and all x in N.

(ii) W is C^1 and $\dot{W}$ is bounded from above or below along each solution which remains in G for all $t > t_0 \geqq 0$.

If E is bounded, then each solution of (1) that remains in G for $t > t_o \geqq 0$ either approaches E or ∞ as $t \to \infty$.

Thus this theorem explains precisely the nature of the information given by a Liapunov function. A Liapunov function relative to a set G defines a set E which under the conditions of the theorem contains (locates) all the positive limit sets of solutions which for positive time remain in G. The problem in applying the result is to find "good" Liapunov functions. For instance, the zero function $V = 0$ is a Liapunov function for the whole space R^n and condition (ii) is satisfied but gives no in-information since $E = R^n$. It is trivial but useful for applications to note that if V_1 and V_2 are Liapunov functions on G, then $V = V_1 + V_2$ is also a Liapunov function and $E = E_1 \cap E_2$. If E is smaller than either E_1 or E_2, then V is a "better" Liapunov function than either E_1 or E_2 and is always at least as "good" as either of the two.

Condition (i) of Theorem 1 is essentially the one used by Yoshizawa. We now look at a simple example where condition (ii) is satisfied and condition (i) is not. The example also shows that the conclusion of the theorem is the best possible. Consider $\ddot{x} + p(t)\dot{x} + x = 0$ where $p(t) \geqq \delta > 0$. Define $2V = x^2 + y^2$, where $y = \dot{x}$. Then $\dot{V} = -p(t)y^2 \leqq -\delta y^2$ and V is a Liapunov function on R^2. Now $W = \delta y^2$ and $\dot{W} = 2\delta y\dot{y} = -2\delta(xy + p(t)y^2) \leqq -2\delta xy$. Since all solutions are evidently bounded for all $t > 0$, condition (ii) is satisfied. Here E is the x-axis $(y = 0)$ and for each solution $x(t)$, $y(t) = \dot{x}(t) \to 0$ as $t \to \infty$. Noting that the equation $\ddot{x} + (2 + e^t)\dot{x} + x = 0$ has a solution $x(t) = 1 + e^{-t}$, we see that this is the best possible result without further restrictions on p.

In order to use Theorem 1 there must be some means of determining which solutions remain in G. The following corollary, which is an obvious consequence of Theorem 1, gives one way of doing this and also provides for nonautonomous systems a method for estimating regions of attraction.

Corollary 1. Assume that there exist continuous functions $u(x)$ and $v(x)$ on R^n to R such that $u(x) \leqq V(t,x) \leqq v(x)$ for all $t \geqq 0$. Define $Q_\eta^+ = \{x \,;\, u(x) < \eta\}$ and let G^+ be a component of Q_η^+. Let G denote the component of $Q_\eta = \{x \,;\, v(x) < \eta\}$ containing G^+. If V is a Liapunov function on G for (1) and the conditions of Theorem 1 are satisfied, then each solution of (1) starting in G^+ at any time $t_0 \geqq 0$ remains in G for all $t > t_0$ and approaches E^* as $t \to \infty$. If G is bounded and $E^0 = \overline{E \cap G} \subset G^+$, then E^0 is an attractor and G^+ is in its region of attraction.

In general we know that if $x(t)$ is a solution of (1)--in fact, if $x(t)$ is any continuous function on R to R^n--then its positive limit set is closed and connected. If $x(t)$ is bounded, then its positive limit set is compact. There are, however, special classes of differential equations where the limit sets of solutions have an additional invariance property which makes possible a refinement of Theorem 1. The first of these are the autonomous systems

$$\dot{x} = f(x) \tag{3}$$

The limit sets of solutions of (3) are invariant sets. If $x(t)$ is defined on $[0,\infty)$ and if p is a positive limit point of $x(t)$, then the points on the solution through p on its maximal interval of definition are positive limit points of $x(t)$. If $x(t)$ is bounded for $t > 0$, then it is defined on $[0,\infty)$, its positive limit set Ω is compact, nonempty and solutions through points p of Ω are defined on $(-\infty,\infty)$ (i.e., Ω is invariant). If the maximal domain of definition of $x(t)$ for $t > 0$ is finite, then $x(t)$ has no finite positive limit points: that is, if the maximal interval of definition of $x(t)$ for $t > 0$ is $[0,\beta)$, then $x(t) \to \infty$ as $t \to \beta$. As we have said before, we will always speak as though our solutions are defined on $(-\infty,\infty)$ and it should be remembered that finite escape time is always a possibility unless there is, as for example in Corollary 2 below, some condition that rules it out. In Corollary 3 below, the solutions might well go to infinity in finite time.

The invariance property of the limit sets of solutions of autonomous systems (3) now enables us to refine Theorem 1. Let V be a C^1 function on R^n to R. If G is any arbitrary set in R^n, we say that V is a <u>Liapunov function on</u> G for equation (3) if $\dot{V} = (\text{grad } V) \cdot f$ does not change sign on G. Define $E = \{ x ; \dot{V}(x) = 0 , x \text{ in } \bar{G} \}$, where $\bar{G}$ is the closure of G. Let M be the largest invariant set in E. M will be a closed set. The fundamental stability theorem for autonomous systems is then the following:

THEOREM 2. If V is a Liapunov function on G for (3), then each solution $x(t)$ of (3) that remains in G for all $t > 0$ $(t < 0)$ approaches $M^* = M \cup \{\infty\}$ as $t \to \infty$ $(t \to -\infty)$. If M is bounded, then either $x(t) \to M$ or $x(t) \to \infty$ as $t \to \infty$ $(t \to -\infty)$.

This one theorem contains all of the usual Liapunov like theorems on stability and instability of autonomous systems. Here however, there are no conditions of definiteness for V or $\dot{V}$, and it is often possible to obtain stability information about a system with these more general types of Liapunov functions. The first corollary below is a stability result which for applications has been quite useful and the second illustrates how one obtains information on instability. Četaev's instability theorem is similarly an immediate consequence of Theorem 2 (see section 3).

COROLLARY 2. Let G be a component of $Q_\eta = \{ x ; V(x) < \eta \}$. Assume that G is bounded, $\dot{V} \leq 0$ on G, and $M^0 = \overline{M \cap G} \subset G$. Then M^0 is an attractor and G is in its region of attraction. If, in addition, V is constant on the boundary of M^0, then

M^o is a stable attractor.

Note that if M^o consists of a single point p, then p is asymptotically stable and G provides an estimate of its region of asymptotic stability.

COROLLARY 3. Assume that relative to (3) that $V\dot{V} > 0$ on G and on the boundary of G that $V = 0$. Then each solution of (3) starting in G approaches ∞ as $t \to \infty$ (or possibly in finite time).

There are also some special classes of nonautonomous systems where the limit sets of solutions have an invariance property. The simplest of these are periodic systems (see [3]).

$$\dot{x} = f(t,x) \quad , \; f(t + T,x) = f(t) \quad \text{for all } t \text{ and } x \,. \qquad (4)$$

Here in order to avoid introducing the concept of a periodic approach of a solution of (4) to a set and the concept of a periodic limit point let us confine ourselves to solutions $x(t)$ of (4) which are bounded for $t > 0$. Let Ω be the positive limit set of such a solution $x(t)$, and let p be a point in Ω. Then there is a solution of (4) starting at p which remains in Ω for all t in $(-\infty,\infty)$; that is, if one starts at p at the proper time the solution remains in Ω for all time. This is the sense now in which Ω is an invariant set. Let $V(t,x)$ be C^1 on $R \times R^n$ and periodic in t of period T. For an arbitrary set G of R^n we say that V is a <u>Liapunov function on</u> G for

for the periodic system (4) if $\dot{V}$ does not change sign for all t and all x in G. Define $E = \{ (t,x); \dot{V}(t,x) = 0, x \text{ in } \bar{G} \}$ and let M be the union of all solutions $x(t)$ of (4) with the property that $(t,x(t))$ is in E for all t. M could be called "the largest invariant set relative to E". One then obtains the following version of Theorem 2 for periodic systems:

THEOREM 3. If V is a Liapunov function on G for the periodic system (4), then each solution of (4) that is bounded and remains in G for all $t > 0$ $(t < 0)$ approaches M as $t \to \infty$ $(t \to -\infty)$.

In [6] Miller showed that the limit sets of solutions of almost periodic systems have a similar invariance property and from this he obtains a result quite like Theorem 3 for almost periodic systems. This then yields for periodic and almost periodic systems a whole chain of theorems on stability and instability quite similar to that for autonomous systems. For example, one has

COROLLARY 4. Let $Q_\eta^+ = \{ x; V(t,x) < \eta, \text{ all } t \text{ in } [0,T] \}$, and let G^+ be a component of Q_η^+. Let G be the component of $Q_\eta = \{ x; V(t,x) < \eta \text{ for some } t \text{ in } [0,T] \}$ containing G^+. If G is bounded, $\dot{V} \leq 0$ for all t and all x in G, and if $M^o = \overline{M \cap G} \subset G^+$, then M^o is an attractor and G^+ is in its region of attraction. If $V(t,x) = \varphi(t)$ for all t and all x on the boundary of M^o, then M^o is a stable attractor.

Our last example of an invariance principle for ordinary differential equations is that due to Yoshizawa in [1] for "asymptotically autonomous" systems. It is a consequence of Theorem 1 and results by Markus and Opial (see [1] for references) on the limit sets of such systems. A system of the form

$$\dot{x} = F(x) + g(t,x) + h(t,x) \tag{5}$$

is said to be <u>asymptotically autonomous</u> if (i) $g(t,x) \to 0$ as $t \to \infty$ uniformly for x in an arbitrary compact set of R^n, (ii) $\int_0^\infty |h(t,\varphi(t))|\, dt < \infty$ for all φ bounded and continuous on $[0,\infty)$ to R^n. The combined results of Markus and Opial then state that the positive limit sets of solutions of (5) are invariant sets of $\dot{x} = F(x)$. Using this, Yoshizawa then improved Theorem 1 for asymptotically autonomous systems.

It turns out to be useful, as we shall illustrate in a moment on the simplest possible example, in studying systems (1) which are not necessarily asymptotically autonomous to state the theorem in the following manner:

THEOREM 4. If, in addition to the conditions of Theorem 1, it is known that a solution $x(t)$ of (1) remains in G for $t > 0$ and is also a solution of an asymptotically autonomous system (5), then $x(t)$ approaches $M^* = M \cup \{\infty\}$ as $t \to \infty$, where M is the largest invariant set of $\dot{x} = F(x)$ in E.

It can happen that the system (1) is itself asymptotically autonomous in which case the above theorem can be applied. However, as the following example illustrates, the original system may not itself be asymptotically autonomous but it still may be possible to construct for each solution of (1) an asymptotically autonomous system (5) which it also satisfies.

Consider again the example

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= -x - p(t)y \qquad , \quad 0 < \delta \leq p(t) \leq m \\ & \qquad\qquad\qquad\qquad \text{for all} \quad t > 0 \end{aligned} \tag{6}$$

Now we have the additional assumption that $p(t)$ is bounded from above. Let $(\bar{x}(t), \bar{y}(t))$ be any solution of (6). As was argued previously below Theorem 1, all solutions are bounded and $\bar{y}(t) \to 0$ as $t \to \infty$. Now $(\bar{x}(t), \bar{y}(t))$ satisfies $\dot{x} = y$, $y = -x - p(t)\bar{y}(t)$, and this system is asymptotically autonomous to (*) $\dot{x} = y$, $\dot{y} = -x$. With the same Liapunov function as before, E is the x-axis and the largest invariant set of (*) in E is the origin. Thus for (6) the origin is asymptotically stable in the large.

3. Autonomous functional differential equation.

Difference differential equations of the form

$$\dot{x}(t) = f(t,x(t),x(t-r)) \qquad , \qquad r > 0 \tag{7}$$

have been studied almost as long as ordinary differential equations and these as well as other types of systems are of the general form

$$\dot{x}(t) = f(t,x_t) \tag{8}$$

where x is in R^n and x_t is the function defined on $[-r,0]$ by $x_t(\tau) = x(t+\tau)$, $-r \leq \tau \leq 0$. Thus x_t is the function that describes the past history of the system on the interval $[t-r,t]$ and in order to consider it as an element in the space C of continuous functions all defined on the same interval $[-r,0]$, x_t is taken to be the function whose graph is the translation of the graph of x on the interval $[t-r,t]$ to the interval $[-r,0]$. Since such equations have had a long history it seems surprising that it is only within the last 10 years or so that the geometric theory of ordinary differential equations has been successfully carried over to functional differential equations. Krasovskii [8] has demonstrated the effectiveness of a geometric approach in extending the classical Liapunov theory, including the converse theorems, to functional differential equations. An account of other aspects of their theory which have yielded to this geometric approach can be found in the paper [9] by Hale. What we wish to do here is to present Hale's extension in [2] of the results of Section 2 of this paper to autonomous functional differential equations

$$\dot{x} = f(x_t) \ . \tag{9}$$

It is this extension that has had so far the greatest success in studying stability properties of the solutions of systems (9), and it is possible that this may lead to a similar theory for special classes of systems defined by partial differential equations.

With $r \geq 0$ the space C is the space of continuous functions φ on $[-r,0]$ to R^n with $\|\varphi\| = \max\{|\varphi(\tau)|;\ -r \leq \tau \leq 0\}$. Convergence in C is uniform convergence on $[-r,0]$. A function x defined on $[-r,\infty)$ to R^n is said to be a _solution_ of (9) _satisfying the initial condition_ φ at time $t = 0$ if there is an $a > 0$ such that $\dot{x}(t) = f(x_t)$ for all t in $[0,a)$ and $x_0 = \varphi$. Remember $x_0 = \varphi$ means $x(\tau) = \varphi(\tau)$, $-r \leq \tau \leq 0$. At $t = 0$, $\dot{x}$ is the right hand derivative. The existence uniqueness theorems are quite similar to those for ordinary differential equations. If f is locally Lipschitzian on C, then for each φ in C there is one and only one solution of (9) and the solution depends continuously on φ. The solution can also be extended in C for $t > 0$ as long as it remains bounded. As in Section 2, we will always speak as though solutions are defined on $[-r,\infty)$. The space C is now the state space of (9) and through each point φ of C there is the motion or flow x_t starting at φ defined by the solution $x(t)$ of (9) satisfying at time $t = 0$ the initial condition φ; x_t, $0 \leq t < \infty$, is a curve in C which starts at time $t = 0$ at φ. In analogy to Section 2 with C replacing R^n, x_t replacing $x(t)$, and $\|x_t\|$ replacing $|x(t)|$, we define the _distance_ $d(x_t,E)$ of x_t from a closed set E of C to be $d(x_t,E) = \min\{\|x_t-\psi\|;\ \psi \in E\}$. The positive limit set of x_t is then defined in a manner completely analogous to Section 2. Because there are some important differences we shall be satisfied here with restricting ourselves to motions

x_t bounded for $t > 0$. One of the differences here is that in C closed and bounded sets are not always compact. Another is that although we have uniqueness of solutions in the future two motions starting from different initial conditions can come together in finite time $t_0 > 0$; after this they coincide for $t \geq t_0$. (The motions define semi-groups and not necessarily groups.)

Hale in [2] has, however, shown that the positive limit sets Ω of bounded motions x_t are nonempty, compact, connected, invariant sets in C. Invariance here is in the sense that, if x_t is a motion starting at a point of Ω, then there is an extension onto $(-\infty,-r]$ such that $x(t)$ is a solution of (9) for all t in $(-\infty,\infty)$ and x_t remains in Ω for all t. With this result he is then able to obtain a result which is similar to Corollary 1 of Section 2.

For $\varphi \in C$ let $x_t(\varphi)$ denote the motion defined by (9) starting at φ. For V a continuous function on C to R define $\dot{V}$ and Q_ℓ by

$$\dot{V}(\varphi) = \overline{\lim_{\tau \to 0+}} \frac{1}{\tau}[V(x_\tau(\varphi))-V(\varphi)]. \tag{10}$$

and

$$Q_\ell = \{\varphi \; ; \; V(\varphi) < \ell\} .$$

THEOREM 5. If $\dot{V}$ does not change sign on G for (9) and x_t is a trajectory of (9) which remains in G and is bounded for $t > 0$, then $x_t \to M$ as $t \to \infty$.

Hale has also given the following more useful version of this result.

COROLLARY 5. Define $Q_\eta = \{\varphi;\ V(\varphi) < \eta\}$ and let G be Q_η or a component of Q_η. Assume that V is nonpositive on G for (9) and that either (i) G is bounded or (iii) $|\varphi(0)|$ is bounded for φ in G. Then each trajectory starting in G approaches M as $t \to \infty$.

The following is an extension of Četaev's instability theorem. This is a somewhat simplified version of Hale's Theorem 4 in [2], which should have stated "$V(\varphi) > 0$ on U when $\varphi \neq 0$ and $V(0) = 0$" and at the end "... intersect the boundary of C_γ ...". This is clear from his proof and is necessary since he wanted to generalize the usual statment of Četaev's theorem to include the possibility that the equilibrium point be inside U as well as on its boundary.

COROLLARY 6. Let $p \in C$ be an equilibrium point of (9) contained in the closure of an open set U and let N be a neighborhood of p. Assume that (i) V is nonnegative on $G = U \cap N$, (ii) $M \cap G$ is either the empty set or p, (iii) $V(\varphi) > \eta$ on G when $\varphi \neq p$, and (iv) $V(p) = \eta$ and $V(\varphi) = \eta$ on that part of the boundary of G inside N. Then p is unstable. In fact, if N_0 is a bounded neighborhood of p properly contained in N then each trajectory starting at a point of $G_0 = G \cap N_0$ other than p leaves N_0 in finite time.

Proof. By the conditions of the corollary and Theorem 5 each trajectory starting inside G_0 at a point other than p must either leave G_0, approach its boundary or approach p. Conditions (i) and (iv) imply that it cannot reach or approach that part of the boundary of G_0 inside N_0 nor can it approach p as $t \to \infty$. Now (ii) states that there are no points of M on that part of the boundary of N_0 inside G. Hence each such trajectory must leave N_0 in finite time. Since p is either in the interior or on the boundary of G, each neighborhood of p contains such trajectories, and p is therefore unstable.

In [2] it was shown that the equilibrium point $\varphi = 0$ of

$$\dot{x}(t) = ax^3(t) + bx^3(t-r)$$

was unstable if $a > 0$ and $|b| < |a|$. Using the same Liapunov function and Theorem 6 we can show a bit more. With

$$V(\varphi) = -\frac{\varphi^4(0)}{4a} + \frac{1}{2}\int_{-r}^{0} \varphi^6(\theta)d\theta ,$$

$$V(x_t) = -\frac{x^4(t)}{4a} + \frac{1}{2}\int_{t-r}^{t} x^6(\theta)d\theta$$

and

$$\dot{V}(\varphi) = -\frac{1}{2}(\varphi^6(0) + 2\,\frac{b}{a}\,\varphi^3(0)\varphi^3(-r) + \varphi^6(-r))$$

which is nonpositive when $|b| < |a|$ (negative definite with respect to $\varphi(0)$ and $\varphi(-r)$); that is, V is a Liapunov function on C and $E = \{\varphi; \varphi(0) = \varphi(-r) = 0\}$. Therefore M is simply the null function $\varphi = 0$. If $a > 0$, the region $G = \{\varphi; V(\varphi) < 0\}$

is nonempty, and no trajectory starting in G can have $\varphi = 0$ as a positive limit point nor can it leave G. Hence by Theorem 5 each trajectory starting in G must be unbounded. Since $\varphi = 0$ is a boundary point of G, it is unstable. It is also easily seen [2] that if $a < 0$ and $|b| < |a|$, then $\varphi = 0$ is asymptotically stable in the large.

In [2] Hale has also extended this theory for systems with infinite lag $(r = \infty)$, and in that same paper gives a number of significant examples of the applications of this theory.

References

[1] Yoshizawa, T., Asymptotic behavior of solutions of a system of differential equations, Contrib. to Diff. Eq., 1(1963), 371-387.

[2] Hale, J., Sufficient conditions for stability and instability of autonomous functional differential equations, J. of Diff. Eq., 1(1965), 452-482.

[3] LaSalle, J., Some extensions of Liapunov's second method, IRE Trans. on Circuit Theory, CT-7(1960), 520-527.

[4] LaSalle, J., Asymptotic stability criteria, Proc. of Symposia in Applied Mathematics, Vol. 13, Hydrodynamic Instability, Amer. Math. Soc., Providence, R.I., 1962, 299-307.

[5] LaSalle, J., and Lefschetz, S., Stability by Liapunov's Direct Method with Applications, Academic Press, New York, 1961.

[6] Miller, R., On almost periodic differential equations, Bul. Amer. Math. Soc., 70(1964), 792-795.

[7] Miller, R., Asymptotic behavior of nonlinear delay-differential equations, J. of Diff. Eq., 3(1965), 293-305.

[8] Krasovskii, N.N., Stability of Motion, Stanford University Press, 1963(translation of 1959 Russian Edition).

[9] Hale, J., "Geometric theory of functional differential equations", Proc. of An International Symposium on Differential Equations and Dynamical Systems, University of Puerto Rico, Mayaguez, P. R., Dec. 1965, Academic Press, New York (to appear).

PRELIMINARY

PAPER [6]

To appear in the Proceedings of the National Academy of Sciences

EXTENDED DYNAMICAL SYSTEMS AND STABILITY THEORY

by

J. K. Hale[1] and E. F. Infante[2]

Center for Dynamical Systems
Brown University
Providence, Rhode Island

[1] This Research was supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force under AFOSR Grant No. 693-67 and in part by the National Aeronautics and Space Administration under Grant No. NGR 40-002-015.

[2] This Research was supported by the National Aeronautics and Space Administration under Grant No. NGR 40-002-015 and under Contract No. NAS8-11264.

Extended Dynamical Systems and Stability Theory

The term dynamical system, as used in this note, is used to describe a one-parameter family of operators with certain properties defined in an appropriate space and is a natural generalization of differential equations, functional differential equations and certain partial differential equations. Zubov[1] has shown that the stability theorems of Liapunov as well as their converses are applicable to dynamical systems. These results play an important role in theoretical studies of stability but, unfortunately, are not easy to apply to particular problems.

For ordinary differential equations and functional differential equations LaSalle[2] and Hale[3] have shown that the limiting sets of trajectories which lie in a compact subset of the space are contained in the largest invariant set where the derivative of the Liapunov function V vanishes. The purpose of the present paper is to extend this result and other related stability results to dynamical systems. In this manner the invariance principle and the stability theorems obtained are also applicable to a large class of partial differential equations. The natural setting for the study of dynamical systems is a Banach space, which can be considered as the space of continuous functions over a finite interval in the case of functional differential equations, as the Euclidean space in the case of differential equations, and as a Sobolev space for certain hyperbolic partial differential equations.

Let R^+ denote the interval $[0,\infty)$ and $\mathcal{A}$ a Banach space with $\|\varphi\|_{\mathcal{A}}$ the norm of an element φ of this space.

Definition 1. We say u is a dynamical system on a Banach space $\mathcal{B}$ if u is a continuous mapping of $R^+ \times \mathcal{B}$ into $\mathcal{B}$, $u(t,\varphi)$ is uniformly continuous in t for t,φ in bounded sets, $u(0,\varphi) = \varphi$ and $u(t+\tau,\varphi) = u(t,u(\tau,\varphi))$ for all $t,\tau \geq 0$, φ in $\mathcal{B}$. The positive orbit $0^+(\varphi)$ through φ in $\mathcal{B}$ is defined as $0^+(\varphi) = \bigcup_{t\geq 0} u(t,\varphi)$. We say φ is an equilibrium point if $0^+(\varphi) = \varphi$.

Zubov[1] has discussed systems of this type, without the uniform continuity condition on bounded sets, and referred to them as generalized dynamical systems. In the theory of dynamical systems on n-dimensional vector spaces the concept of invariant sets is basic since the limits of orbits are invariant sets. Zubov defines an invariant set of his generalized dynamical system as a set M such that, for any φ in M, $0^+(\varphi)$ belongs to M. Since u is defined only on R^+ this appears at first sight to be a reasonable definition; however, this definition does not impart any special significance to the limit set of an orbit and appears unreasonable since it generally occurs that trajectories having limits can be used to define functions on $(-\infty,\infty)$. We shall therefore modify the definition of invariant set.

If u is a dynamical system on $\mathcal{B}$, then one can be assured that $0^+(\varphi)$ has a nonempty limit set if $0^+(\varphi)$ belongs to a compact subset of $\mathcal{B}$. In ordinary differential equations and functional differential equations it is possible to show that $0^+(\varphi)$ belonging to a bounded set implies $0^+(\varphi)$ belongs to a compact set (see, for example ref. 3) and thus the limit set is nonempty. However, for many partial differential equations, this is not the case. On the other hand, for certain partial differential equations bounded orbits in $\mathcal{B}$ will belong to a compact set of a larger Banach space $\mathcal{L}$.

It is this latter property which we wish to exploit in detail. More specifically, if we know that every bounded orbit in $\mathcal{B}$ belongs to a compact set in $\mathcal{L}$, then we can discuss the limit of the orbit in $\mathcal{L}$ (thus extending the dynamical system) and as a consequence hope to obtain more specific information about trajectories than would be possible by remaining only in $\mathcal{B}$. These remarks provide the motivation for the following discussion. The reader should contrast this approach with the one of Auslander and Seibert[4] in which it is assumed that the space $\mathcal{B}$ is locally compact.

Let $\mathcal{B}, \mathcal{L}$ be Banach spaces, $\mathcal{B} \subset \mathcal{L}$ and let there exist a constant $K > 0$ such that $\|\varphi\|_{\mathcal{L}} \leq K\|\varphi\|_{\mathcal{B}}$.

<u>Definition 2</u>. Let u be a dynamical system on $\mathcal{B}$. Let $\mathcal{B}^*$ be the set of φ in $\mathcal{L}$ such that there is a sequence φ_n in $\mathcal{B}$ and a function $u^*(t,\varphi)$ in $\mathcal{L}$ for t in R^+, such that $\|\varphi_n-\varphi\|_{\mathcal{L}} \to 0$, $\|u(t,\varphi_n)-u^*(t,\varphi)\|_{\mathcal{L}} \to 0$ as $n \to \infty$ uniformly on compact subsets of R^+. We refer to the function $u^*: R^+ \times \mathcal{B}^* \longrightarrow \mathcal{B}^*$ as the <u>extension of the dynamical system u to $\mathcal{B}^*$</u> or simply as the <u>extended dynamical system</u>.

The function u^* is clearly an extension of u. In fact, if φ is in $\mathcal{B}$, then there exists a sequence φ_n in $\mathcal{B}$ such that $\|\varphi_n-\varphi\|_{\mathcal{B}} \to 0$ (and therefore $\|\varphi_n-\varphi\|_{\mathcal{C}} \to 0$) as $n \to \infty$. This fact and the continuity of u implies $\|u(t,\varphi_n)-u(t,\varphi)\|_{\mathcal{B}} \to 0$ as $n \to \infty$ and therefore $\|u(t,\varphi_n)-u(t,\varphi)\|_{\mathcal{C}} \to 0$ as $n \to \infty$. Thus $u^*(t,\varphi) = u(t,\varphi)$ for φ in $\mathcal{B}$. Furthermore it is easy to prove

Lemma 1. The function $u^*(t,\varphi)$ is continuous in t and $u^*(0,\varphi) = \varphi$, $u^*(t+\tau,\varphi) = u^*(t,u^*(\tau,\varphi))$ for t,τ in R^+ and φ in $\mathcal{B}^*$.

We now give a definition of invariance of a different nature from the one given by Zubov:

Definition 3: A set M in $\mathcal{B}^*$ is an invariant set of the dynamical system if for each φ in M there is a function $U(t,\varphi)$ defined and in M for t in $(-\infty,\infty)$ such that, for any σ in $(-\infty,\infty)$, $u^*(t,U(\sigma,\varphi)) = U(t+\sigma,\varphi)$ for all t in R^+.

Definition 4: For any φ in $\mathcal{B}$, the ω-limit set $\Omega(\varphi)$ of the orbit through φ is the set of ψ in $\mathcal{C}$ such that there is a nondecreasing sequence $\{t_n\}$, $t_n > 0$, $t_n \to \infty$ as $n \to \infty$ such that $\|u(t_n,\varphi)-\psi\|_{\mathcal{C}} \to 0$ as $n \to \infty$.

It should be noted that sets are invariant according to the above definition relative to the interval $(-\infty,\infty)$ and that the ω-limit set of an orbit is obtained relative to convergence in $\mathcal{C}$ and not in $\mathcal{B}$. With these definitions it is then possible to prove the fundamental

<u>Lemma 2</u>: Let φ in $\mathcal{B}$ be such that $0^+(\varphi)$ belongs to a bounded set of $\mathcal{B}$ and a compact subset of $\mathcal{E}$. Then the ω-limit set $\Omega(\varphi)$ of the orbit through φ is a nonempty, compact, connected set in $\mathcal{B}^*$, invariant with respect to the extended dynamical system and $\mathrm{dist}_{\mathcal{E}}(u(t,\varphi),\Omega(\varphi)) \to 0$ as $t \to \infty$.

<u>Proof</u>: Since $0^+(\varphi)$ belongs to a compact subset of $\mathcal{E}$, it is clear that $\Omega(\varphi)$ is nonempty and belongs to a compact subset of $\mathcal{E}$. We shall show below that it belongs to $\mathcal{B}^*$. Suppose that ψ in $\Omega(\varphi)$ is given and that $\{t_n\}$, nondecreasing, $t_n \geqq 0$, $t_n \to \infty$ as $n \to \infty$ satisfies $\|u(t_n,\varphi)-\psi\|_{\mathcal{E}} \to 0$ as $n \to \infty$. For a given τ in $[0,\infty)$ there exists an $n_0(\tau)$ such that $t_n - \tau \geqq 0$ for $n \geqq n_0(\tau)$ and it is therefore meaningful to consider the sequence $u(t+t_n,\varphi)$; $n \geqq n_0(\tau)$, t in $[-\tau,\tau]$. By hypothesis there exists an M such that $\|u(t,\varphi)\|_{\mathcal{B}} \leqq M$ for all t in $[0,\infty)$. Thus $\|u(t,\varphi)\|_{\mathcal{E}} \leqq KM$ for $n \geqq n_0(\tau)$, t in $[-\tau,\tau]$. Also, since $u(t,\varphi)$ is uniformly continuous in t for t,φ in bounded sets, for any $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\|u(t+s+t_n,\varphi)-u(t+t_n,\varphi)\|_{\mathcal{E}} \leqq K\|u(s,u(t+t_n,\varphi))-u(0,u(t+t_n,\varphi))\|_{\mathcal{B}} \leqq \epsilon$$

for $n \geqq n_0(\tau)$, $0 \leqq s \leqq \delta$. This proves that the sequence $\{u(t+t_n,\varphi)\}$, t in $[-\tau,\tau]$ is uniformly bounded and equicontinuous in $\mathcal{E}$. Since this sequence belongs by hypothesis to a compact subset of $\mathcal{E}$, Ascoli's theorem implies the existence of a subsequence which we again label by t_n such that it converges uniformly on $[-\tau,\tau]$; that is, there exists a function $U(t,\varphi)$ continuous in t such that $\lim_{n\to\infty} \|U(t,\varphi)-u(t+t_n,\varphi)\|_{\mathcal{B}} = 0$ uniformly on $[-\tau,\tau]$. Obviously $U(0,\varphi) = \psi$. Letting now $\tau = 1,2,\ldots$ successively and using the familiar triangularization procedure we determine a subsequence which is relabled by t_n and a continuous function $U(t,\varphi)$ defined for t in $(-\infty,\infty)$ such that $\lim_{n\to\infty} \|U(t,\varphi)-u(t+t_n,\varphi)\|_{\mathcal{B}} = 0$ uniformly on compact subsets of $(-\infty,\infty)$. Applying this in particular to $[0,\infty)$ we obtain that ψ belongs to $\mathcal{B}^*$. Furthermore, it is clear that $U(t,\varphi)$ is in $\Omega(\varphi)$.

Let now σ be an arbitrary real number in $(-\infty,\infty)$. We claim that $U(t+\sigma,\varphi) = u^*(t,U(\sigma,\varphi))$, $t \geqq 0$. For this particular σ We have $\lim_{n\to\infty} \|u(\sigma+t_n,\varphi)-U(\sigma,\varphi)\|_{\mathcal{B}} = 0$ and $\lim_{n\to\infty} \|u(t,u(\sigma+t_n,\varphi)) - U(t+\sigma,\varphi)\|_{\mathcal{B}} = 0$ uniformly on compact subsets of $[0,\infty)$. But this is precisely the manner in which $u^*(t,U(\sigma,\varphi))$ was defined. This shows that $\Omega(\varphi)$ is invariant with respect to the extended dynamical system. It is clear that $\Omega(\varphi)$ is connected.

We now show that $\Omega(\varphi)$ is closed. Let ψ_n in $\Omega(\varphi)$ be such that $\psi_n \to \psi$ as $n \to \infty$. Then for any ϵ-neighborhood of ψ in $\mathcal{B}$ there exists a $t_\epsilon, t_\epsilon \to \infty$ as $\epsilon \to 0$ such that $\|u(t_\epsilon,\varphi) - \psi\|_{\mathcal{B}} < \epsilon$. Hence closure.

Finally, assume there exists a sequence $\{t_n\}$, nondecreasing, $t_n \to \infty$ as $n \to \infty$ and an $\alpha > 0$ such that $\|u(t_n,\varphi)-\psi\|_{\mathcal{B}} \geqq \alpha$

for all ψ in $\Omega(\varphi)$. By assumption $\{u(t_n,\varphi)\}$ belongs to a compact set of $\mathcal{B}$ and therefore there exists a subsequence which converges to $\bar{\psi}$ in $\mathcal{B}$. But then $\bar{\psi}$ belongs to $\Omega(\varphi)$ by definition, contradicting the assumption and the proof is complete.

We now define the concepts of stability with respect to these spaces:

Definition 5: If zero is an equilibrium point, then we say that zero is stable if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\|\varphi\|_{\mathcal{B}} < \delta$ implies $\|u(t,\varphi)\|_{\mathcal{B}} < \epsilon$ for all $t \geqq 0$. If, in addition, there exists a $b > 0$ such that $\|\varphi\|_{\mathcal{B}} < b$ implies $\|u(t,\varphi)\|_{\mathcal{B}} \to 0$ as $t \to \infty$ then the origin is said to be asymptotically stable $(\mathcal{B}, \mathcal{B})$. The origin is called unstable if it is not stable.

It is remarked that asymptotic stability is defined by taking limits in $\mathcal{B}$, as is to be expected from the definition of ω-limit sets.

If V is a continuous scalar functional defined on $\mathcal{B}$, we define

$$\dot{V}(\varphi) = \overline{\lim_{t \to 0^+}} \frac{1}{t}[V(u(t,\varphi))-V(\varphi)].$$

Following LaSalle[5] we give

Definition 6: We say a scalar functional V is a Liapunov functional on a set G in $\mathcal{B}$ if V is continuous and bounded below on bounded sets of G and $\dot{V}(\varphi) \leq 0$ for φ in G. We define sets R, M as follows:

$$R = \{\varphi \text{ in } \mathcal{C} : \text{there exists } \{\varphi_n\} \text{ in } G \text{ with } \lim_{n\to\infty} \|\varphi_n - \varphi\|_{\mathcal{C}} = 0$$

$$\text{and } \lim_{n\to\infty} \dot{V}(\varphi_n) = 0\},$$

and M is the largest set in R which is invariant with respect to the extended dynamical system.

With the above definitions and with the fundamental Lemma 2 it is now possible to prove stability theorems which are direct generalizations of those given for functional differential equations and differential equations[3,4].

Theorem 1: Suppose every orbit $O^+(\varphi)$ which is in a bounded set in $\mathcal{B}$ also belongs to a compact set in $\mathcal{C}$. If V is a Liapunov functional on G and an orbit $O^+(\varphi)$ belongs to G and is in a bounded set in $\mathcal{B}$, then $u(t,\varphi) \to M$ in $\mathcal{C}$ as $t \to \infty$.

Corollary 1: Suppose that every orbit which belongs to a bounded set in $\mathcal{B}$ also belongs to a compact set in $\mathcal{C}$. Assume V is a continuous scalar functional defined on $\mathcal{B}$, $S_\rho = \{\varphi \text{ in } \mathcal{B}: V(\varphi) < \rho\}$ and let G be S_ρ or a component of S_ρ. If V is a Liapunov functional on G and any orbit remaining in G belongs to a bounded set in $\mathcal{B}$, then φ in G implies $u(t,\varphi) \to M$ in $\mathcal{C}$ as $t \to \infty$.

Note, in this corollary, that if zero is in G and M consists of only the point zero, then the origin is an "attractor" but we have not shown it to be stable. The following result gives conditions that insure stability. The part of the corollary which does not follow directly from Theorem 1 is proved as in the usual Liapunov theory.

<u>Corollary 2</u>: If the conditions of Corollary 1 are satisfied and V is a continuous positive definite functional on G, then zero is stable. If, in addition, $M = \{0\}$, then zero is asymptotically stable $(\mathcal{B}, \mathcal{C})$. If, in addition, $\dot{V}$ is negative definite, then zero is asymptotically stable $(\mathcal{B}, \mathcal{B})$.

The stronger form of asymptotic stability given in the last part of this corollary should be noted. Unfortunately, for any given system it is very difficult to construct a Liapunov functional with these characteristics.

<u>Theorem 2</u>: Suppose that every orbit which is in a bounded set in $\mathcal{B}$ also belongs to a compact set in $\mathcal{C}$. Let zero be an equilibrium point contained in the closure of an open set U and let N be a neighborhood of zero. Assume that

(i) V is a Liapunov functional on $G = U \cap N$,

(ii) $M \cap G$ is either the empty set or is zero,

(iii) $V(\varphi) < \eta$ on G when $\varphi \neq 0$

(iv) $V(0) = \eta$ and $V(\varphi) = \eta$ when φ is in that part of the boundary of G inside N.

Then zero is unstable. More precisely, if N_o is a bounded neighborhood of zero properly contained in N, then $\varphi \neq 0$ in $G_o = G \cap N_o$ implies that there exists a $\tau > 0$ such that $u(\tau, \varphi)$ belongs to the boundary of N_o.

The proofs of these theorems and corollaries follow closely those previously given for ordinary differential equations[5].

The lemmas and theorems displayed above are in terms of two spaces, $\mathcal{B}$ and $\mathcal{E}$. If the space $\mathcal{B}$ is a Hilbert space then a considerable simplification occurs.

Lemma 3: If $\mathcal{B}$ is a Hilbert space and $\mathcal{E}$ is a Banach space, $\mathcal{B} \subset \mathcal{E}$, $\|\varphi\|_{\mathcal{E}} \leq K\|\varphi\|_{\mathcal{B}}$ for some constant $K > 0$, then the unit ball in $\mathcal{B}$ is closed in $\mathcal{E}$.

This lemma is a direct consequence of the Banach-Saks Theorem.

It follows that if $\mathcal{B}$ and $\mathcal{E}$ are Hilbert spaces, then the set $\mathcal{B}^*$ in Definition 2 is the same as $\mathcal{B}$ and therefore the extended dynamical system is the same as the original dynamical system. Therefore, the ω-limit sets will belong to $\mathcal{B}$ but the convergence of $u(t, \varphi)$ to its ω-limit set is in the sense of the topology of $\mathcal{E}$ and not, in general in $\mathcal{B}$. These remarks play an important role in the applications to certain partial differential equations.

REFERENCES

[1] Zubov, V.I., Methods of A.M. Liapunov and their Application, (translation - P. Noordhoff Ltd., 1964).

[2] LaSalle, J.P., "The Extent of Asymptotic Stability", Proc. Nat. Acad. Sci., 46(1960), 365.

[3] Hale, J.K., "Sufficient Conditions for Stability and Instability of Autonomous Functional-Differential Equations", J. Diff. Eqns., 1(1965), 452.

[4] Auslander, J., and P. Seibert, "Prolongations and Stability in Dynamical Systems", Ann. Inst. Fourier Grenoble, 14(1964), 237.

[5] LaSalle, J.P., "An Invariance Principle on the Theory of Stability", to appear in the Proc. of International Symposium, Differential Equations and Dynamical Systems, Academic Press, 1967, 277.

PAPER [7]

DIFFERENTIAL EQUATIONS AND DYNAMICAL SYSTEMS, Proceedings of the International Symposium, Puerto Rico, Academic Press, New York, 1967, 309-321.

Stability Criteria for n-th Order, Homogeneous Linear Differential Equations[†]

E. F. Infante[††]

Center For Dynamical Systems, Brown University

1. Introduction

This note is concerned with the homogeneous differential equation

$$x^{(n)} + \rho_1(t)x^{(n-1)} + \ldots + \rho_{n-1}(t)\dot{x} + \rho_n(t)x = 0, \qquad (1.1)$$

where the $\rho_i(t)$ are real continuous functions. It is desired to determine appropriate criteria for the stability of the origin, criteria dependent on the behavior of the functions $\rho_i(t)$ but not of their derivatives.

This problem has been previously studied by Starzinski [1,2,3] for particular forms of this equation up to the fourth order, and by Razumichin [4] for the general matrix equation $\dot{x} = A(t)x$. The approach of these authors has been to use the direct method of Liapunov, using a constant quadratic Liapunov function $V(x) = x'Bx$ which is generated by determining the $n(n+1)/2$ constant elements of the symmetric matrix B. The determination of all these elements requires very heavy algebraic computations, computations which are completely unreasonable for $n > 2$. Recently, Ghizzetti [5,6] has obtained simple stability criteria for (1.1) by using some appropriate majoration formulae for all the integrals of this equation. The

† This research was supported by the National Aeronautics and Space Administration under Contract NAS8-11264.

†† On leave of absence from Department of Mechanical Engineering, University of Texas.

particularly attractive aspect of these criteria is that they depend on only n constant parameters which locate a family of hyperellipsoids in the n-dimensional space of the $\rho_i(t)$. If the curve parametrically represented by the $\rho_i(t)$ is entirely contained within one of these hyperellipsoids, then (1.1) is asymptotically stable.

In §2 of this note the second method of Liapunov is used to obtain stability criteria for (1.1) that depend on only n parameters which determine a family of elliptic paraboloids in the n-dimensional space $\rho_i(t)$. It can be shown that these elliptic paraboloids completely contain the hyperellipsoids of Ghizzetti. In §3 a practical technique for the application of the stability criteria obtained is discussed and is applied in the last section to two examples. The stability conditions presented in this note are not necessary. Indeed, they are probably not the best possible conditions obtainable from a quadratic Liapunov function. The technique presented in this note was devised with particular emphasis on ease of computability of some simple criteria.

2. Stability Criteria

Consider Eq. (1.1) rewritten in state-space coordinates as

$$
\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= x_3 \\
&\vdots \\
\dot{x}_{n-1} &= x_n \\
\dot{x}_n &= -\rho_n(t)x_1 - \dots -\rho_1(t)x_n.
\end{aligned}
\tag{2.1}
$$

For the determination of the stability of the origin of (2.3), consider the Liapunov function $V(x) = x'Bx$, $B' = B = (\beta_{ij})$, β_{ij} = constant. Let b_n denote the n-th column of the matrix B, and

$$b_n = \begin{pmatrix} b_n^u \\ \beta_{ni} \\ b_n^\ell \end{pmatrix} . \tag{2.5}$$

The derivative $\dot{V}$ of the Liapunov function V in terms of (2.3) is given by

$$\dot{V} = x'(A'B + BA)x - x'(U'(t)B + BU(t))x, \tag{2.6}$$

or

$$-\dot{V} = x'Cx + x'(ub_n' + b_nu')x, \tag{2.7}$$

where $A'B + BA = -C$. If it were possible to determine a matrix B, positive definite, such that $-\dot{V}$ is positive definite for all $t \geqq 0$, then asymptotic stability of the origin of (2.1) will have been determined by the well known theorem of Liapunov [8]. For this purpose, consider the following simple lemma:

Lemma 2.1: Given the constant matrix A, defined by (2.4), for any constant positive semidefinite diagonal matrix $C \neq 0$ the equation $A'B + BA = -C$ has a unique solution B, and B is positive definite.

It is assumed that the $\rho_i(t)$, real continuous functions of time, satisfy the Routh-Hurwitz inequalities [7]. Let the n real numbers α_i, assumed to satisfy the Routh-Hurwitz inequalities, be associated to (2.1), which is rewritten as

$$\begin{aligned} \dot{x}_1 &= x_2 \\ &\vdots \\ \dot{x}_n &= -(\rho_n(t) - \alpha_n)x_1 - \dots -(\rho_1(t) - \alpha_1)x_n - \alpha_n x_1 - \dots - \alpha_1 x_n. \end{aligned} \tag{2.2}$$

For economy of notation, (2.2) is rewritten as

$$\dot{x} = Ax - U(t)x, \tag{2.3}$$

where

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -\alpha_n & -\alpha_{n-1} & -\alpha_{n-2} & \cdots & -\alpha_2 & -\alpha_1 \end{pmatrix},$$

$$U(t) = \begin{pmatrix} 0 \\ \\ \\ u' \end{pmatrix}, \quad u = \begin{pmatrix} \eta_n \\ \eta_{n-1} \\ \vdots \\ \eta_1 \end{pmatrix} = \begin{pmatrix} u^u \\ \\ \eta_{n+1-i} \\ u^\ell \end{pmatrix}, \tag{2.4}$$

and where $\eta_i \equiv \rho_i(t) - \alpha_i$.

Proof. The matrix B, obviously symmetric, is unique since all the eigenvalues of A have negative real parts. Now, let $V(x_o) = x_0'Bx_0 < 0$ for some $x_o \neq 0$, and define δ_o as the trajectory of $\dot{x} = Ax$ issuing from x_o at $t = 0$. Along δ we then have $V(x) \leq V(x_o) < 0$. But δ approaches the origin and $V(0) = 0$. Hence $V(x) \geq 0$. Similarly, let $V(x_1) = 0$, $x_1 \neq 0$, and δ_1 the trajectory emanating from x_1 at t_o. Since this trajectory approaches the origin, it must lie on the manifold $x'Cx = 0$. But this is clearly impossible with C diagonal and A in the form (2.4). Hence B is positive definite.

Hence, let the matrix B be generated by the diagonal matrix

$$C = \begin{pmatrix} C^u & & \\ & 0 & \\ & & C^\ell \end{pmatrix}, \tag{2.8}$$

where C^u and C^ℓ are constant nonsingular positive definite diagonal square matrices, and where the zero element in the diagonal is located in the i,i position. On the basis of the above lemma $V(x) = x'Bx$ will be positive definite. In this case, Eq. (2.7) then becomes

$$-\dot{V} = x'\begin{pmatrix} C^u & & \\ & 0 & \\ & & C^\ell \end{pmatrix}x + x'\begin{pmatrix} u^u b_n^{u'} + b_n^u u^{u'} & u^u \beta_{ni} + \eta_{n+1-i} b_n^u & u^u b_n^{\ell'} + b_n^{u'} u^{\ell'} \\ \eta_{n+1-i} b_n^{u'} + \beta_{ni} u^{u'} & 2\beta_{ni}\eta_{n+1-i} & \eta_{n+1-i} b_n^{\ell'} + \beta_{ni} u^{\ell'} \\ u^\ell b_n^{u'} + b_n^\ell u^{u'} & u^\ell \beta_{ni} + \eta_{n+1-i} b_n^\ell & u^\ell b_n^{\ell'} + b_n^\ell u^{\ell'} \end{pmatrix} x \tag{2.9}$$

Assume $\beta_{ni} > 0$ (it is always possible to find a $\beta_{ni} > 0$, namely β_{nn}) and consider the regular transformation $x = Sy$,

$$S = \begin{pmatrix} I & 0 & 0 \\ -\dfrac{b_n^{u'}}{\beta_{ni}} & 1 & -\dfrac{b_n^{\ell'}}{\beta_{ni}} \\ 0' & 0 & I \end{pmatrix}, \qquad (2.10)$$

where the unit element is in the i,i position and the I are unit matrices of appropriate dimensions. If this transformation is applied to Eq. (2.9), one obtains

$$-\dot{V} = y' \begin{pmatrix} C^u & & \\ & 0 & \\ & & C^\ell \end{pmatrix} y + y' \begin{pmatrix} 0 & \beta_{ni} u^u - \eta_{n+1-i} b_n^u & 0 \\ \beta_{ni} u^{u'} - \eta_{n+1-i} b_n^{u'} & 2\beta_{ni}\eta_{n+1-i} & \beta_{ni} u^{\ell'} - \eta_{n+1-i} b_n^{\ell'} \\ 0' & \beta_{ni} u^\ell - \eta_{n+1-i} b_n^\ell & 0 \end{pmatrix} y \qquad (2.11)$$

or

$$-\dot{V} = y' \begin{pmatrix} C^u & \beta_{ni} u^u - \eta_{n+1-i} b_n^u & 0 \\ \beta_{ni} u^{u'} - \eta_{n+1-i} b_n^{u'} & 2\beta_{ni}\eta_{n+1-i} & \beta_{ni} u^{\ell'} - \eta_{n+1-i} b_n^{\ell'} \\ 0' & \beta_{ni} u^\ell - \eta_{n+1-i} b_n^\ell & C^\ell \end{pmatrix} y \qquad (2.12)$$

It now becomes necessary to determine under what conditions (2.12) is positive definite. For this purpose, consider the second transformation. $y = Tz$,

$$T = \begin{pmatrix} I & -C^{u^{-1}} v^u & 0 \\ 0' & 1 & 0' \\ 0' & -C^{\ell^{-1}} v^{\ell} & I \end{pmatrix} , \qquad (2.13)$$

where the unit element is in the i,i position, the I are unit matrices of appropriate dimensions and $v^u = \beta_{ni}u^u - \eta_{n+1-i}b_n^u$, $v^{\ell} = \beta_{ni}u^{\ell} - \eta_{n+1-i}b_n^{\ell}$. This transformation is obviously regular and when applied to Eq. (2.12) yields

$$-\dot{V} = z' \begin{vmatrix} C^u & 0 & 0 \\ 0' & \omega & 0' \\ 0' & 0 & C^{\ell} \end{vmatrix} z \qquad (2.14)$$

where

$$\omega = 2\beta_{ni}\eta_{n+1-i} - (\beta_{ni}u^u - \eta_{n+1-i}b_n^u)'C^{u^{-1}}(\beta_{ni}u^u - \eta_{n+1-i}b_n^u) + - (\beta_{ni}u^{\ell} - \eta_{n+1-i}b_n^{\ell})'C^{\ell^{-1}}(\beta_{ni}u^{\ell} - \eta_{n+1-i}b_n^{\ell}) \qquad (2.15)$$

Since (2.14) is diagonal, it can then be concluded that $\dot{V}$ will be negative definite if $\omega \geq \delta > 0$.

On the basis of what has been said above, it is then possible to state:

Theorem 2.1: Given the homogeneous differential equation

$$x^{(n)} + \rho_1(t)x^{(n-1)} + \dots + \rho_{n-1}(t)\dot{x} + \rho_n(t)x = 0 \,, \tag{2.16}$$

with $\rho_i(t)$ real continuous functions for $t \geq 0$, associate with this equation the n real constants $\alpha_1,\dots,\alpha_n$ satisfying the Routh-Hurwitz inequalities, and define $\eta_i = \rho_i(t) - \alpha_i$. Let the matrix $B = (\beta_{ij})$ be the solution of the matrix equation

$$A'B + BA = -\begin{pmatrix} C^u & & \\ & 0 & \\ & & C^\ell \end{pmatrix} , \tag{2.17}$$

where C^u, C^ℓ are constant, positive definite diagonal matrices, and the zero element in the diagonal appears in the i,i position; and where

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdot\cdot & 0 & 0 \\ 0 & 0 & 1 & \cdot\cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot\cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot\cdot & 0 & 1 \\ -\alpha_n & -\alpha_{n-1} & -\alpha_{n-2} & \cdot\cdot & -\alpha_2 & -\alpha_1 \end{pmatrix} . \tag{2.18}$$

Let b_n denote the n-th column of B and define

$$b_n = \begin{pmatrix} b_n^u \\ \beta_{ni} \\ b_n^\ell \end{pmatrix} , \quad u = \begin{pmatrix} \eta_{n+1-1} \\ \vdots \\ \eta_{n+1-i} \\ \vdots \\ \eta_1 \end{pmatrix} = \begin{pmatrix} u_u \\ \eta_{n+1-i} \\ u^\ell \end{pmatrix} . \tag{2.19}$$

<u>Then</u>, <u>if for any</u> $\delta > 0$ <u>and any</u> $i = 1,\dots,n$

$$2\beta_{ni}\eta_{n+1-i} - (\beta_{ni}u^u - \eta_{n+1-i}b_n^u)'C^{u^{-1}}(\beta_{ni}u^u - \eta_{n+1-i}b_n^u) + \\ - (\beta_{ni}u^\ell - \eta_{n+1-i}b_n^\ell)'C^{\ell^{-1}}(\beta_{ni}u^\ell - \eta_{n+1-i}b_n^\ell) \geq \delta \tag{2.20}$$

<u>for all</u> $t \geq 0$, <u>the null solution of</u> (2.16) <u>is asymptotically stable</u>.

This theorem is not as general as it would have been possible to state, yet it is still too general for practical applications because of the generality of the matrices C^u and C^ℓ. Before restricting the theorem, it is desirable to make some remarks concerning the results so far obtained.

First of all we wish to point out that Eq. (2.20) represents, in the parameter space of the η's, an elliptic paraboloid. This can be easily seen by introducing the transformation of coordinates for the parameter space given by

$$v = \begin{pmatrix} \gamma_{n+1-1} \\ \vdots \\ \gamma_{n+1-i} \\ \vdots \\ \gamma_1 \end{pmatrix} = \begin{pmatrix} v^u \\ \\ \gamma_{n+1-i} \\ \\ v^\ell \end{pmatrix} = \begin{pmatrix} \beta_{ni}I & -b_n^u & 0 \\ & & \\ 0' & 1 & 0' \\ & & \\ 0' & -b_n^\ell & \beta_{ni}I \end{pmatrix} \begin{pmatrix} u^u \\ \\ \eta_{n+1-i} \\ \\ u^\ell \end{pmatrix} \tag{2.21}$$

This transformation is obviously regular if $\beta_{ni} > 0$, which as was previously pointed out, is no restriction. In the new coordinates, Eq. (2.20) becomes

$$2\beta_{ni}\gamma_{n+1-i} - v' \begin{pmatrix} C^{u^{-1}} & & \\ & 0 & \\ & & C^{\ell^{-1}} \end{pmatrix} v \geq \varepsilon . \qquad (2.22)$$

This is evidently the equation of an elliptic paraboloid. If $\beta_{ni} > 0$, as assumed, the domain defined in the parameter space by (2.22), hance by (2.20). is nonempty.

Secondly, it is evident that, for any C^u and C^ℓ satisfying the conditions of Theorem 2.1, the domain of the η parameter space defined by any of the (2.20) is strictly contained within the domain where the $\rho_i(t)$ satisfy the Routh-Hurwitz inequalities. On the other hand, it is easily shown that every point of the domain of the parameter space where the $\rho_i(t)$ satisfy the Routh-Hurwitz inequalities is contained in at least one of the domains defined by (2.20). To prove this, let $\rho_i(t) = \overline{\rho_i}$ = constants. Since the $\overline{\rho_i}$ satisfy the Routh-Hurwitz inequalities, it is possible to select the n numbers α_i, themselves satisfying these inequalities, and such that $\eta_{n+1-j} = \overline{\rho}_{n+1-j} - \alpha_{n+1-j} = \varepsilon > 0$ for some j and $\overline{\rho}_{n+1-i} - \alpha_{n+1-i} = 0$ for all $i \neq j$. Under these conditions Eq. (2.20) reduces to

$$2\beta_{nj}\eta_{n+1-j} - \eta_{n+1-j} b_n^{u'} C^{u^{-1}} b_n^u \eta_{n+1-j} - \eta_{n+1-j} b_n^{\ell'} C^{\ell^{-1}} b_n^\ell \eta_{n+1-j} \geq \delta. \quad (2.23)$$

But for any $\varepsilon > 0$ sufficiently small, a $\delta > 0$ can be found such that (2.23) is satisfied. Hence the remark.

Finally, it is noted that the continuity condition imposed by Theorem 2.1 on the $\rho_i(t)$ imply that Eq. (2.16) does not have a finite escape time. It is therefore possible on the basis of this remark and the two previous ones to state:

Corollary 2.1: Given the differential equation (2.16) with $\rho_i(t)$ real continuous functions for $t \geq 0$, if there exist a $\tau > 0$ such that for all $t \geq \tau$ (2.20) is satisfied for some $\delta > 0$ and some $i = 1,\ldots,n$, then the null solution of (2.16) is asymptotically stable.

Corollary 2.2: If, in Eq. (2.16), the $\rho_i(t)$ are real continuous functions for $t \geq 0$ and $\lim_{t \to \infty} \rho_i(t) = \overline{\rho_i}$, where the $\overline{\rho_i}$ satisfy the Routh-Hurwitz inequalities, then the null solution of (2.16) is asymptotically stable.

This last corollary is very well known [7], and can be traced directly to Liapunov.

3. Application of Stability Criteria

The positive definite diagonal matrices C^u and C^ℓ have not been so far specified. The first step in the application of the stability criteria obtained to a specific example is the selection of these two matrices, from which the matrix B is obtained as the solution of the equation $A'B + BA = -C$. Algorithms for the solution of this matrix equation are available. A particularly simple one has been recently given by Smith [9] in the case matrix A has the form (2.18).

It is particularly convenient, to obtain algebraically simple forms for B, to select the matrices C^u and C^ℓ to be composed of linear combinations of matrices of the form

$$C_1 = 2 \operatorname{diag} (\mu, 0,\ldots,0) \tag{3.1}$$

and

$$C_k = 2 \operatorname{diag}\left(0,\ldots,\frac{\mu}{\alpha_n},0,\ldots,0\right), \quad k \neq 1, \qquad (3.2)$$

where μ is the Hurwitz determinant [7] of the α:

$$\mu = \begin{vmatrix} \alpha_1 & \alpha_3 & \alpha_5 & \cdot & \cdot & \cdot & \alpha_{2n-1} \\ 1 & \alpha_2 & \alpha_4 & \cdot & \cdot & \cdot & \alpha_{2n-2} \\ 0 & \alpha_1 & \alpha_3 & \cdot & \cdot & \cdot & \alpha_{2n-3} \\ 0 & 1 & \alpha_2 & \cdot & \cdot & \cdot & \alpha_{2n-4} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \alpha_n \end{vmatrix} \qquad (3.3)$$

The matrix equation $A'B_k + B_kA = -C_k$, where A is given by (2.18) can be rapidly solved for B_k when C_k is of the suggested form. The matrices obtained in this manner for $n = 2,3$ are shown in Table 1. Ingwerson [10] previously published these matrices for $n = 2,3,4$. If C^u and C^ℓ are obtained, as suggested, by linear combinations of the C_k, then the matrix B will be the corresponding linear combination of the B_k.

Table I

n = 2

$$B_1 = \begin{pmatrix} \alpha_1^2 + \alpha_2 & \alpha_1 \\ \alpha_1 & 1 \end{pmatrix}, \qquad C_1 = \begin{pmatrix} 2\alpha_1\alpha_2 & 0 \\ 0 & 0 \end{pmatrix}$$

$$B_2 = \begin{pmatrix} \alpha_2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad C_2 = \begin{pmatrix} 0 & 0 \\ 0 & 2\alpha_1 \end{pmatrix}$$

n = 3

$$B_1 = \begin{pmatrix} \alpha_2(\alpha_1\alpha_2-\alpha_3)+\alpha_1^2\alpha_3 & \alpha_1^2\alpha_2 & \alpha_1\alpha_2-\alpha_3 \\ \alpha_1^2\alpha_2 & \alpha_1^3+\alpha_3 & \alpha_1^2 \\ \alpha_1\alpha_2-\alpha_3 & \alpha_1^2 & \alpha_1 \end{pmatrix}, \quad C_1 = \begin{pmatrix} 2\alpha_3(\alpha_1\alpha_2-\alpha_3) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$B_2 = \begin{pmatrix} \alpha_1\alpha_3 & \alpha_3 & 0 \\ \alpha_3 & \alpha_1^2+\alpha & \alpha_1 \\ 0 & \alpha_1 & 1 \end{pmatrix}, \qquad C_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2(\alpha_1\alpha_2-\alpha_3) & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$B_3 = \begin{pmatrix} \alpha_3^2 & \alpha_2\alpha_3 & 0 \\ \alpha_2\alpha_3 & \alpha_1\alpha_3+\alpha_2^2 & \alpha_3 \\ 0 & \alpha_3 & \alpha_2 \end{pmatrix}, \qquad C_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2(\alpha_1\alpha_2-\alpha_3) \end{pmatrix}$$

4. Two Examples

In this section, the stability criteria obtained is applied to two simple but illustrative example problems.

As a first example, consider the second order equation

$$\ddot{x} + p\dot{x} + q(t)x = 0 \tag{4.1}$$

or

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -q(t)x_1 - px_2 \end{aligned} \tag{4.2}$$

where, $p > 0$ is a constant and $0 < q_1 + \xi \leq q(t) \leq q_2 - \xi$, for $\xi > 0$. It is desired to determine conditions on q_1, q_2 and p that guarantee the asymptotic stability of the null solution of (4.2). This same problem has been treated by Ghizzetti [5], with whom we wish to compare our results.

In the case of a second order equation, inspection of the matrices B_1 and B_2 of table one indicates that, for $\beta_{ni} > 0$ one must select $i = 2$. With this choice one immediately obtains

$$C = \begin{pmatrix} 2\alpha_1\alpha_2 & 0 \\ 0 & 0 \end{pmatrix}, \qquad C_u^{-1} = \frac{1}{2\alpha_1\alpha_2},$$

$$b_n = \begin{pmatrix} \beta_{21} \\ \beta_{22} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ 1 \end{pmatrix}; \quad u = \begin{pmatrix} \eta_2 \\ \eta_1 \end{pmatrix} = \begin{pmatrix} q - \alpha_2 \\ p - \alpha_1 \end{pmatrix}, \tag{4.3}$$

upon which the stability equation given by (2.20) becomes

$$2(p - \alpha_1) - [(q - \alpha_2) - \alpha_1(p - \alpha_1)] \frac{1}{2\alpha_1\alpha_2} [(q - \alpha_2) - \alpha_1(p - \alpha_1)] \geq \delta > 0 \tag{4.4}$$

or, letting $\nu_1 = \frac{\alpha_1}{p}$, $\nu_2 = \frac{\alpha_2}{p^2}$ and $z(t) = \frac{q(t)}{p^2}$,

$$4\nu_1\nu_2(1 - \nu_1) - [z(t) - \nu_2 - \nu_1(1 - \nu_1)]^2 \geq \varepsilon > 0 \tag{4.5}$$

To determine the appropriate values of ν_1 and ν_2 for this expression, let

$$\begin{aligned} z_1 &= \frac{q_1}{p^2} = \nu_2 + \nu_1(1 - \nu_1) - 2\sqrt{\nu_1\nu_2(1 - \nu_1)} , \\ z_2 &= \frac{q_2}{p^2} = \nu_2 + \nu_1(1 - \nu_1) + 2\sqrt{\nu_1\nu_2(1 - \nu_1)} , \end{aligned} \tag{4.6}$$

and to maximize the difference between z_2 and z_1 let $\nu_1 = 1/2$. Then

$$\begin{aligned} z_1 &= \frac{1}{4} + \nu_2 - \sqrt{\nu_2} , \\ z_2 &= \frac{1}{4} + \nu_2 + \sqrt{\nu_2} . \end{aligned} \tag{4.7}$$

Solving now for ν_2 from the first of these equations

$$\nu_2 = \frac{1}{4} + z_1 + \sqrt{z_1} \tag{4.8}$$

is obtained. With these two particular values of ν_1 and ν_2 (4.7) yields

$$z_2 = z_1 + 2\sqrt{\frac{1}{4} + z_1 + \sqrt{z_1}} \quad .$$

Hence, if $0 < z_1 + \xi < z(t) < \xi$ for some $\xi > 0$, an $\varepsilon > 0$ can be found such that Eq. (4.5) is satisfied. Therefore, Eq. (4.1) is asymptotically stable if, for some $\xi > 0$,

$$0 < q_1 + \xi \leq q(t) \leq q_2 - \xi \tag{4.10}$$

and

$$\frac{q_2}{p^2} = \frac{q_1}{p^2} + 2\sqrt{\frac{1}{4} + \frac{q_1}{p^2} + \sqrt{\frac{q_1}{p^2}}} \quad . \tag{4.11}$$

This result is represented in graphical form in Figure 1: if $\frac{q(t)}{p^2}$ is strictly internal to the domain Δ of the parameter space q_1/p^2 vs. q_2/p^2, then Eq. (4.1) is asymptotically stable. The domain Δ obtained by Ghizzetti [5] is shown also.

As a second example, consider the differential equation

$$\dddot{x} + p\ddot{x} + \dot{x} + r(t)x = 0 \;, \tag{4.11}$$

where $p > 0$ is a constant and $0 < \xi \leq r(t) \leq r_2 - \xi$ for some $\xi > 0$. It is desired to determine conditions on r_2 to guarantee the asymptotic stability of the null solution of this equation. This equation has been studied by Starzinski [3], who generated a constant Liapunov function by determining, through a very laborious process, appropriate values for all six elements of the 3×3 B matrix.

Inspection of the third order matrices of Table 1 indicates that, for $\beta_{ni} > 0$ one must select either $i = 2$ or $i = 3$. Let $i = 3$ upon which the stability equation (2.20) becomes

$$2\beta_{33}(p - \alpha_1) - [\beta_{33}(r(t) - \alpha_3) - \beta_{31}(p - \alpha_1)]^2 C^{u^{-1}} + \\ - [\beta_{33}(1 - \alpha_2) - \beta_{32}(p - \alpha_1)]^2 C^{\ell^{-1}} \geq \delta > 0. \tag{4.12}$$

Since $i = 3$, let $C = C_1 + \lambda C_2$ where C_1 and C_2 are the two matrices shown in Table 1, and $\lambda > 0$. From Table 1, then

$$\beta_{31} = \alpha_1\alpha_2 - \alpha_3, \quad \beta_{32} = \alpha_1^2 + \lambda\alpha_1, \quad \beta_{33} = \lambda + \alpha_1 \\ C^{u^{-1}} = \frac{1}{2\alpha_3(\alpha_1\alpha_2-\alpha_3)}, \quad C^{\ell^{-1}} = \frac{1}{2\lambda(\alpha_1\alpha_2-\alpha_3)} \tag{4.13}$$

are immediately obtained. Equation (4.12) can be therefore rewritten as

$$4\alpha_3(\lambda + \alpha_1)(\alpha_1\alpha_2-\alpha_3)(p - \alpha_1) - [(\lambda + \alpha_1)(r(t) - \alpha_3) - (\alpha_1\alpha_2-\alpha_3)(p - \alpha_1)]^2 + \\ - \frac{\alpha_3}{\lambda}[(\lambda + \alpha_1)(1 - \alpha_2) - (\alpha_1^2 + \lambda\alpha_1)(p - \alpha_1)]^2 \geq \varepsilon > 0. \tag{4.14}$$

The second quadratic term vanishes if

$$(1 - \alpha_2) = \alpha_1(p - \alpha_1). \tag{4.15}$$

Furthermore, (4.14) can be satisfied as $r(t)$ becomes very small only if

$$p - \alpha_1 = \alpha_3 \frac{\lambda + \alpha_1}{\alpha_1\alpha_2 - \alpha_3} . \tag{4.16}$$

Assuming these two conditions, Eq. (4.14) yields

$$0 < \xi \le r(t) \le 4\alpha_3 - \xi , \tag{4.17}$$

where $\xi \to 0$ as $\varepsilon \to 0$. Equations (4.15) and (4.16) yield

$$\alpha_3 = \frac{\alpha_2 - \alpha_2^2}{\lambda + p} , \quad \alpha_1 = \frac{p + \sqrt{p^2 - 4 + 4\alpha_2}}{2} ; \tag{4.18}$$

therefore, let

$$\begin{aligned} \alpha_2 &= 1 - \left(\frac{p}{2}\right)^2 & &\text{if} \quad 0 < p \le \sqrt{2} \\ \alpha_2 &= \frac{1}{2} & &\text{if} \quad \sqrt{2} \le p \end{aligned} \tag{4.19}$$

upon which one obtains that Eq. (4.11) is asymptotically stable if

$$\begin{aligned} 0 < \xi \le r(t) &\le \frac{1}{\lambda + p}\left(p^2 + \frac{p^4}{4}\right) - \xi \quad &\text{if} \quad 0 < p \le \sqrt{2} \\ 0 < \xi \le r(t) &\le \frac{1}{\lambda + p} - \xi \quad &\text{if} \quad p \ge \sqrt{2} \end{aligned} \tag{4.20}$$

for some $\xi > 0$ and $\lambda > 0$, since the α's obtained from Eq. (4.18) and

(4.19) satisfy the Routh-Hurwitz inequalities.

This same result would have been obtained if the stability Eq. (2.20) for $i = 2$ had been used. The stability conditions (4.20) are identical to those obtained by Starzinski [3].

References

[1] Starzinski, V. M., "Sufficient Conditions for Stability of a Mechanical System with One Degree of Freedom", PMM, v. 16, 1952, p. 369, (Russian).

[2] Starzinski, V. M., "On the Stability of Unsteady Motion in One Case", PMM, v. 16, 1952, p. 500, (Russian).

[3] Starzinski, V. M., "On Stability of Unsteady Motions in a Special Case", PMM, v. 19, 1955, p. 471, (Russian).

[4] Razumichin, B. S., "On the Application of Liapunov's Method to Stability Problems", PMM, v. 22, 1958, p. 338.

[5] Ghizzetti, A., "Some Sufficient Stability Criteria for the Integrals of a Linear Homogeneous Differential Equation of Second Order", Rendiconti, Acc. Naz. dei Lincei, Series VIII, v. 33, 1962, p. 219, (Italian).

[6] Ghizzetti, A., "Majoration Formulae and Sufficient Criteria for the Stability of the Integrals of a Linear Homogeneous Differential Equation of n-th Order", Memorie, Acc. Naz. dei Lincei, Series VIII, v. 7, 1963, p. 17, (Italian).

[7] Cesari, L., Asymptotic Behavior and Stability Problems in Ordinary Differential Equations, Springer-Verlag, 1953, p. 21.

[8] LaSalle, J. P. and S. Lefschetz, Stability by Liapunov's Direct Method, Academic Press, 1961, p. 40.

[9] Smith, R. A., "Matrix Calculations for Liapunov Quadratic Forms", to appear, Journal of Differential Equations.

[10] Ingwerson, D. R., "A Modified Lyapunov Method for Nonlinear Stability Analysis", IRE Trans., v. AC-6, 1961, p. 199.

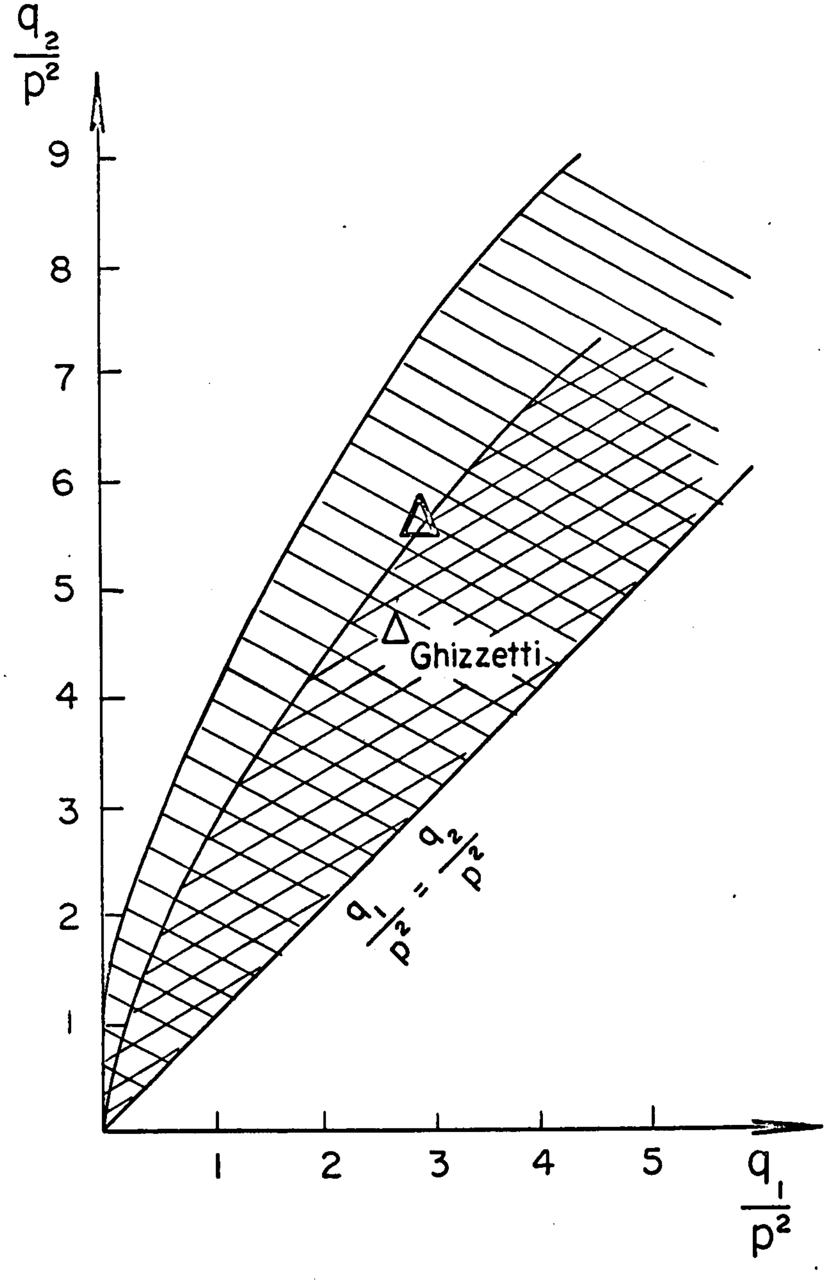


FIGURE 1

PAPER [8]

Submitted to the Journal of Applied Mechanics

ON THE STABILITY OF SOME LINEAR NONAUTONOMOUS SYSTEMS*

by

E. F. Infante

Center for Dynamical Systems
Division of Applied Mathematics
Brown University, Providence, Rhode Island 02912

* This research was supported in part by the National Aeronautics and Space Administration under Contract No. NAS8-11264 and in part by Grant No. NGR-40-002-015.

ON THE STABILITY OF SOME LINEAR NONAUTONOMOUS SYSTEMS

E. F. Infante

The stability of systems described by differential equations with time varying coefficients has been the subject of numerous mathematical studies, see for example [1]; however very limited sucess has been achieved from the practical viewpoint with the exception of the case in which the coefficients are periodic. Recently Kozin [2], Caughey and Gray [3] and Ariaratnam [4] among others have studied the stability of linear systems with stochastic coefficients; in these studies the principal tools used have been Gronwall's inequality and a norm used to reduce the vector differential equation to a scalar equation. Kozin [2] used the so-called taxicab norm, Caughey and Gray [3] used a very special quadratic norm and obtained results superior to those of Kozin. A natural problem within this context is to determine the optimum norm, among a certain class, for a specific problem.

The stability theorems given in [2] and [3] depend on the specific norms used in their proofs. The object of this paper is to extend these theorems so that they are applicable for any quadratic norm. This can be easily done through the use of well known results on pencils of quadratic forms [5], an application which seems to have been overlooked. The theorem obtained in this manner, and two corollaries, are then applied to the determination of conditions for the stability of second order equations, for which it is possible to obtain the optimum quadratic norm. The stability results obtained in this manner, which as expected represent sufficient but not necessary conditions, constitute a considerable improvement over those presented in [2] and [3], and are believed to be new. The examples are limited to second order systems since problems of this type are often reduced to them.

The notation used here follows that of [2] and [3], and emphasizes the application to stochastic processes. Naturally, the results are equally applicable to deterministic systems which satisfy the condition of Equation (2).

A STABILITY THEOREM

Consider the differential equation

$$\dot{x} = [A + F(t)]x , \tag{1}$$

where x is an n vector, A is a constant matrix and $F(t)$ is a matrix whose nonzero elements $f_{ij}(t)$ are stochastic processes, measurable, strictly stationary, and that they satisfy an ergodic property ensuring the equality of time averages and ensemble averages. If G is a measurable, integrable, function defined on $f_{ij}(t)$ then

$$E\{G(f_{ij}(t))\} = E\{G(f_{ij}(0))\} = \lim_{t\to\infty} \frac{1}{t-t_o} \int_{t_o}^{t} G(f_{ij}(\tau))d\tau \qquad (2)$$

exists with probability one. For simplicity let, in (1), $E\{F(t)\} = 0$ and denote by $\lambda_{max}[Q]$ the largest eigenvalue of the matrix Q, Q' the transpose of Q.

<u>THEOREM</u>: If, for some positive definite matrix B and some $\epsilon > 0$

$$E\{\lambda_{max}[A' + F'(t) + B(A+F(t))B^{-1}] \leq -\epsilon , \qquad (3)$$

then (1) is almost surely asymptotically stable in the large.

<u>Proof</u>: Consider the quadratic (Liapunov) function $V(x) = x'Bx$. Then, along the trajectories of (1), define

$$\lambda(t) = \frac{\dot{V}(x)}{V(x)} = \frac{x'[(A+F)'B + B(A+F)]x}{x'Bx} . \qquad (4)$$

From the extremal properties of pencils of quadratic forms [5] the inequality

$$\lambda_{min}[(A+F)'+B(A+F)B^{-1}] \leq \lambda(t) \leq \lambda_{max}[(A+F)'+B(A+F)B^{-1}] \qquad (5)$$

is obtained, where λ_{max} and λ_{min}, being the maximum and minimum eigenvalues of a pencil, are real. It follows from (4) and (5) that

$$V(x(t)) = V(x(t_o))e^{\int_{t_o}^{t}\lambda(\tau)d\tau} = V(x(t_o))e^{(t-t_o)[\frac{1}{t-t_o}\int_{t_o}^{t}\lambda(\tau)d\tau]}, \quad (6)$$

from which it follows that, if $E\{\lambda(t)\} \leqq -\epsilon$ for some $\epsilon > 0$, $V(x(t))$ is bounded and that $V(x(t)) \to 0$ as $t \to \infty$. This is the condition imposed by (3), which proves the result.

It is remarked that a necessary condition for (3) to hold is that the eigenvalues of matrix A have negative real parts. The eigenvalue computation specified by (3) is far simple. It is possible to obtain a result which is easier to compute, but not as sharp.

COROLLARY 1: If, for some positive definite matrix B and some $\epsilon > 0$

$$E\{\lambda_{max}[F'(t)+BF(t)B^{-1}]\} \leqq -\lambda_{max}[A'+BAB^{-1}]-\epsilon \quad , \qquad (7)$$

then (1) is almost surely asymptotically stable in the large.

Proof: The proof follows immediately from the theorem by noting that

$$\lambda(t) \leqq \lambda_{max}[(A+F)'+B(A+F)B^{-1}] \leqq \lambda_{max}[A'+BAB^{-1}]+\lambda_{max}[F'+BFB^{-1}], (8)$$

from which it follows upon the application of (7), that

$$E\{\lambda(t)\} \leq \lambda_{max}[A'+BAB^{-1}]+E\{\lambda_{max}[F'(t)+BF(t)B^{-1}]\} \leq -\epsilon \quad , \tag{9}$$

the desired result.

It is obvious that unless the second inequality in (8) is an equality the stability results obtained will not be as good as those given by the theorem. For computational purposes, it is desirable to further simplify the theorem. For this purpose let (1) be written as

$$\dot{x} = Ax + \sum_{i=1}^{R} f_i(t)C_i \quad , \tag{10}$$

where $R \leq n^2$, and recall that $E\{f_i(t)\} = 0$.

<u>COROLLARY 2</u>: If, for some positive definite matrix B and some $\epsilon > 0$

$$\begin{aligned} &\sum_{i=1}^{R} \tfrac{1}{2}E\{|f_i(t)|\}(\lambda_{max}[C_i'+BC_iB^{-1}]-\lambda_{min}[C_i'+BC_iB^{-1}]) \\ &\qquad \leq -\lambda_{max}[A'+BAB^{-1}]-\epsilon \quad , \end{aligned} \tag{11}$$

then (10) is almost surely asymptotically stable in the large.

<u>Proof</u>: In this case equation (4) of the theorem becomes

$$\lambda(t) = \frac{x'(A'B+BA)x}{x'Bx} + \sum_{i=1}^{R} f_i(t) \frac{x'(C_i'B+BC_i)x}{x'Bx} \quad . \tag{12}$$

Since $E\{f_i(t)\} = 0$ by assumption, define the two functions

$$f_i^+(t) = \begin{cases} f_i(t) & \text{if} \quad f_i(t) \geqq 0 \\ 0 & \text{if} \quad f_i(t) \leqq 0 \end{cases} ,$$

$$f_i^-(t) = \begin{cases} f_i(t) & \text{if} \quad f_i(t) \leqq 0 \\ 0 & \text{if} \quad f_i(t) \geqq 0 \end{cases} . \tag{13}$$

It then follows that

$$E\{f_i^+(t)\} = -E\{f_i^-(t)\} = \tfrac{1}{2}E\{|f_i(t)|\}, \tag{14}$$

and Equation (12) yields

$$E\{\lambda(t)\} \leqq \lambda_{max}[A + BAB^{-1}] + \sum_{i=1}^{R} \tfrac{1}{2}E\{|f_i(t)|\}(\lambda_{max}[C_i + BC_iB^{-1}]$$
$$-\lambda_{max}[C_i + BC_iB^{-1}]) \quad , \tag{15}$$

from which, through application of condition (11),

$$E\{\lambda(t)\} \leqq -\epsilon \tag{16}$$

is obtained, proving the corollary.

It is again to be expected that the results obtained from this corollary will seldom be as good as those given by either the Theorem or Corollary 1, since the majorizations used are rougher than the previous ones.

The above theorem and corollaries say nothing regarding how the matrix B should be chosen. If this matrix is chosen, as in [3], as the solution of the matrix equation $A'B+BA = -I$ then the stability condition of the Theorem, Equation (3), becomes

$$E\{\lambda_{max}[-B^{-1}+F'(t)+BF(t)B^{-1}]\} \leq -\epsilon \,, \tag{3'}$$

Corollary 1 yields the stability condition

$$E\{\lambda_{max}[F'(t)+BF(t)B^{-1}]\} \leq \frac{1}{\lambda_{max}[B]} -\epsilon \,, \tag{7'}$$

and the condition of Corollary 2 becomes

$$\sum_{i=1}^{R} \tfrac{1}{2}E\{|f_i(t)|\}(\lambda_{max}[C_i'+BC_iB^{-1}]-\lambda_{min}[C_i'+BC_iB^{-1}]) \leq \frac{1}{\lambda_{max}[B]} -\epsilon \,. \tag{11'}$$

The conditions implied by (7') and (11') are clearly satisfied if we majorize further in these equations by noting that, if $Q(t) = F'(t)+BF(t)B^{-1}$,

$$\lambda_{max}[Q(t)] \leq \sum_{i,j} |Q_{ij}| \,,$$

and further that

$$\tfrac{1}{2}(\lambda_{max}[C_i'+BC_iB^{-1}]-\lambda_{min}[C_i'+BC_iB^{-1}]) \leq |\mu^i|_{max} ,$$

where $|\mu^i|_{max}$ is the largest eigenvalue, in absolute value, of $C_i' + BC_iB^{-1}$. With these majorizations equations (7') and (11') become

$$E\{\sum_{i,j}|Q_{ij}|\} \leq \frac{1}{\lambda_{max}[B]} - \epsilon \tag{7''}$$

and

$$\sum_{i=1}^{R} E\{|f_i(t)|\}|\mu^i|_{max} \leq \frac{1}{\lambda_{max}[B]} - \epsilon , \tag{11''}$$

the stability conditions given by Caughey and Gray [3].

It is then seen that the use of well known results on pencils of quadratic forms yields stability theorems of time varying systems that include those of [3]. The natural question at this juncture is to demand a theorem which yields the optimal matrix B to be used. Unfortunately, this problem does not appear amenable to analysis, as the third example of the next section indicates. The purpose of the following section is to obtain the optimal matrix B of the Theorem and Corollaries 1 and 2 for the two most common second order equations of type (1). A third second order equation is analyzed to show that an optimal matrix B does not exist; finally an application of the theorem of this section to the study of the stability of a nuclear reactor is shown. The stability results thus obtained are compared with those given in [2] and [3], and indicate that the matrix of Caughey and Gray is, in general, not optimal.

SOME EXAMPLES

EXAMPLE 1: Consider the differential equation

$$\ddot{x} + 2\xi\dot{x} + (1+f(t))x = 0 \,, \tag{17}$$

studied by Kozin [2], Caughey and Gray [3] and Ariaratnam [4]. It is assumed that $E\{f(t)\} = 0$, and the equation is rewritten as

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -1 & -2\xi \end{bmatrix} x + f(t) \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} x \tag{18}$$

or, $\dot{x} = Ax + F(t)x$. Consider, for the matrix B, the most general quadratic positive definite form

$$B = \begin{bmatrix} \alpha_1^2 + \alpha_2 & \alpha_1 \\ \alpha_1 & 1 \end{bmatrix} \,, \quad \alpha_2 > 0 \tag{19}$$

where α_1 and α_2 are numbers to be determined.

Simple computations immediately yield that

$$B^{-1} = \frac{1}{\alpha_2} \begin{bmatrix} 1 & -\alpha_1 \\ \alpha_2 & \alpha_1^2 + \alpha_2 \end{bmatrix} \tag{20}$$

and that

$$A' + F'(t) + B(A+F(t))B^{-1} =$$

$$= \frac{1}{\alpha_2}\begin{bmatrix} -\alpha_1(1+f)-\alpha_1^2(\alpha_1-2\xi)-\alpha_1\alpha_2 & \alpha_1^2(1+f)+(\alpha_1^2+\alpha_2)[\alpha_1(\alpha_1-2\xi)+\alpha_2] \\ -(1+f)-\alpha_1(\alpha_1-2\xi) & \alpha_1(1+f)+(\alpha_1-2\xi)(\alpha_1^2+\alpha_2) \end{bmatrix}. \tag{21}$$

The maximum eigenvalue of this expression is computed as

$$\lambda_{max}[A'+F'+B(A+F)B^{-1}] = -2\xi + \sqrt{4(\xi-\alpha_1)^2 + \frac{1}{\alpha_2}[\alpha_2+\alpha_1^2-1-f(t)+2\alpha_1(\xi-\alpha_1)]^2} \tag{22}$$

and setting $f = 0$ in this equation

$$\lambda_{max}[A'+BAB^{-1}] = -2\xi + \sqrt{4(\xi-\alpha_1)^2 + \frac{1}{\alpha_2}\{\alpha_2+\alpha_1^2-1+2\alpha_1(\xi-\alpha_1)\}^2}. \tag{23}$$

is obtained. Finally,

$$F'(t) + BF(t)B^{-1} = \frac{f(t)}{\alpha_2}\begin{bmatrix} -\alpha_1 & \alpha_1^2 - \alpha_2 \\ -1 & \alpha_1 \end{bmatrix}, \tag{24}$$

from which the eigenvalue expression

$$\lambda_{max}[F'(t) + BF(t)B^{-1}] = \frac{1}{\sqrt{\alpha_2}}\,|f(t)| \tag{25}$$

is immediately computed.

In this particular example, then, the conditions for almost sure asymptotic stability given by the previous section become, for the theorem

$$E\{-2\xi + \sqrt{4(\xi-\alpha_1)^2 + \frac{1}{\alpha_2}[\alpha_2+\alpha_1^2-1-f(t)+2\alpha_1(\xi-\alpha_1)]^2}\} \leqq -\epsilon \qquad (26)$$

and from either of the two corollaries

$$E\{|f(t)|\}\frac{1}{\sqrt{\alpha_2}} \leqq 2\xi - \sqrt{4(\xi-\alpha_1)^2 + \frac{1}{\alpha_2}[\alpha_2+\alpha_1^2-1+2\alpha_1(\xi-\alpha_1)]^2} - \epsilon \qquad (27)$$

for some α_1, and some $\alpha_2 > 0$ and $\epsilon > 0$. If the stability conditions are desired in terms of $E\{|f(t)|\}$, the optimum values of α_1 and α_2 for equations (26) and (27) coincide and are easily computed as

$$\begin{aligned} &\alpha_1 = \xi\,, \qquad \alpha_2 = 1-\xi^2, \qquad \text{if} \quad \xi \leqq \frac{\sqrt{2}}{2}\,, \\ &\alpha_1 = \xi\,, \qquad \alpha_2 = \xi^2\,, \qquad \text{if} \quad \xi \geqq \frac{\sqrt{2}}{2}\,, \end{aligned} \qquad (28)$$

upon which the stability conditions (26) become

$$\begin{aligned} &E\{|f(t)|\} \leqq 2\xi\sqrt{1-\xi^2} - \epsilon\,, \qquad \xi \leqq \frac{\sqrt{2}}{2}\,, \\ &E\{|f(t)+1-2\xi^2|\} \leqq 2\xi^2 - \epsilon\,, \qquad \xi \geqq \frac{\sqrt{2}}{2}\,, \end{aligned} \qquad (29)$$

while conditions (27) yield

$$\begin{aligned} E\{|f(t)|\} &\leqq 2\xi\sqrt{1-\xi^2} - \epsilon , && \xi \leqq \frac{\sqrt{2}}{2} , \\ E\{|f(t)|\} &\leqq 1 - \epsilon , && \xi \geqq \frac{\sqrt{2}}{2} . \end{aligned} \tag{30}$$

As expected, conditions (29) are weaker than conditions (30); this is strongly emphasized by obtaining stability conditions from (29) and (30) in terms of $E\{f^2(t)\}$ through the use of Schwarz's inequality, remembering that $E\{f(t)\} = 0$. This process yields the stability conditions

$$\begin{aligned} E\{f^2(t)\} &\leqq 4\xi^2(1-\xi^2) - \epsilon , && \xi \leqq \frac{\sqrt{2}}{2} , \\ E\{f^2(t)\} &\leqq 4\xi^2-1 - \epsilon , && \xi \geqq \frac{\sqrt{2}}{2} , \end{aligned} \tag{29'}$$

from (29) and, from (30)

$$\begin{aligned} E\{f^2(t)\} &\leqq 4\xi^2(1-\xi^2) - \epsilon , && \xi \leqq \frac{\sqrt{2}}{2} , \\ E\{f^2(t)\} &\leqq 1 - \epsilon , && \xi \geqq \frac{\sqrt{2}}{2} , \end{aligned} \tag{30'}$$

a much more meager result.

If, at the outset, it is desired to obtain stability conditions as a function of $E\{f^2(t)\}$, then the values

$$\alpha_1 = \xi , \qquad \alpha_2 = \xi^2 + 1 \tag{31}$$

are optimal for equation (26) which yields

$$E\{f^2(t)\} \leq 4\xi^2 . \tag{32}$$

These results are a considerable improvement over those of [2] and [3]. Figure 1 displays these results and those of these two references in a pictorial form. It is of interest to note that either (29') or (32) show that, for almost sure asymptotic stability, it is possible to let $E\{f^2(t)\} \to \infty$ as the damping ξ increases; this result therefore answers a question raised by Mehr and Wang [6] in their discussion of [2].

<u>EXAMPLE 2</u>: As a second example consider the equation

$$\ddot{x} + (2\xi + g(t))\dot{x} + x = 0, \quad E\{g(t)\} = 0 , \tag{33}$$

which is rewritten in the usual companion form yielding, in the notation of (1),

$$A = \begin{bmatrix} 0 & 1 \\ -1 & -2\xi \end{bmatrix} , \qquad F(t) = g(t)\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \tag{34}$$

Using again the matrix B given by (19) a simple computation yields

$$\lambda_{max}[A'+F'+B(A+F)B^{-1}] = -2\xi - g(t) +$$

$$+\sqrt{(g(t)+2\xi-2\alpha_1)^2 + \frac{1}{\alpha_2}[\alpha_2+\alpha_1^2-1+\alpha_1 g(t)+2\alpha_1(\xi-\alpha_1)]^2} \tag{35}$$

and

$$\{\lambda_{max}[F'(t)+BF(t)B^{-1}] - \lambda_{min}[F'(t)+BF(t)B^{-1}]\} =$$

$$= |g(t)| \sqrt{1+\frac{\alpha_1^2}{\alpha_2}} \, . \tag{36}$$

Hence, in this case, the theorem of the previous section yields, for stabilty

$$E\{-2\xi+\sqrt{(g(t)+2\xi-2\alpha_1)^2+\frac{1}{\alpha_2}[\alpha_2+\alpha_1^2-1+\alpha_1 g(t)+2\alpha_1(\xi-\alpha)]^2}\} \leqq -\epsilon \; ; \tag{37}$$

either of the two corollaries give instead the condition

$$E\{|g(t)|\}\sqrt{1+\frac{\alpha_1^2}{\alpha_2}} \leqq 2\xi - \sqrt{4(\xi-\alpha_1)^2+\frac{1}{\alpha_2}[\alpha_2+\alpha_1^2-1+2\alpha_1(\xi-\alpha_1)]^2} - \epsilon. \tag{38}$$

A straightforward computation yields, in the case that stability conditions are desired as functions of $E\{|g(t)|\}$, that the optimum values for α_1 and α_2 for equations (37) and (38) coincide and are

$$\alpha_1 = \xi\,, \qquad \alpha_2 = 1-\xi^2\,, \qquad \text{if } \xi^2 \leq \frac{\sqrt{5}-1}{2}\,;$$
$$\alpha_1 = \frac{1}{\sqrt{\xi^2+1}}\,, \qquad \alpha_2 = \frac{\xi^2}{\xi^2+1}\,, \qquad \text{if } \xi^2 \geq \frac{\sqrt{5}-1}{2}\,; \tag{39}$$

and the conditions of stability become, for the theorem, equation (37),

$$E\{|g(t)|\} \leq 2\xi\sqrt{1-\xi^2} - \epsilon\,, \qquad \xi^2 \leq \frac{\sqrt{5}-1}{2}\,;$$
$$E\{|g(t)+2\xi - \frac{2}{\sqrt{\xi^2+1}}|\} \leq 2\xi\,\frac{\xi^2}{\sqrt{1+\xi^2}} - \epsilon, \quad \xi^2 \geq \frac{\sqrt{5}-1}{2}\,; \tag{40}$$

and for either of the corollaries, equation (38),

$$E\{|g(t)|\} \leq 2\xi\sqrt{1-\xi^2} - \epsilon\,, \qquad \xi^2 \leq \frac{\sqrt{2}-1}{2}\,;$$
$$E\{|g(t)|\} \leq 2\xi\left[\sqrt{1+\frac{1}{\xi^2}} - 1\right] - \epsilon, \qquad \xi^2 \geq \frac{\sqrt{5}-1}{2}\,; \tag{41}$$

It is noted that equation (40) gives weaker conditions for stability, since application of Schwarz's inequality to this equation gives the stability conditions

$$E\{g^2(t)\} \leq 4\xi^2(1-\xi^2) - \epsilon\,, \qquad \xi^2 \leq \frac{\sqrt{5}-1}{2}\,;$$
$$E\{g^2(t)\} \leq 4\,\frac{2\xi - \sqrt{\xi^2+1}}{\sqrt{\xi^2+1}} - \epsilon\,, \qquad \xi^2 \geq \frac{\sqrt{5}-1}{2}\,. \tag{40'}$$

If stability conditions are desired as a function of $E\{f^2(t)\}$, the optimum values

$$\alpha_1 = \frac{\xi}{1+\xi^2}\ , \qquad \alpha_2 = 1 - \frac{\xi^2}{(1+\xi^2)^2} \tag{42}$$

applied to equation (37) yield, after application of Schwarz's inequality, the stability condition

$$E\{g^2(t)\} \leq \frac{4\xi^2}{1+\xi^2} - \epsilon \tag{43}$$

These stability results are shown in a pictorial representation in Figure 2.

EXAMPLE 3: Consider, in this case, the differential equation

$$\ddot{x} + [2\xi + g(t)]\dot{x} + [1+f(t)]x = 0\ , \tag{44}$$

a generalization of the two previous differential equations. Using the same matrix B of equation (19) and repeating the computations indicated in the previous examples the following conditions for almost sure stability in the large are obtained: from the theorem

$$E\{-2\xi + \sqrt{(g(t)+2\xi-2\alpha_1)^2 + \frac{1}{\alpha_2}[\alpha_2+\alpha_1^2-1+\alpha_1 g(t)-f(t)+2\alpha_1(\xi-\alpha_1)]^2}\}$$
$$\leq -\epsilon\ , \tag{45}$$

from Corollary 1

$$E\{\sqrt{g^2(t)+\frac{1}{\alpha_2}[f(t)-\alpha_1 g(t)]^2} \leq$$
$$E\{2\xi-\sqrt{4(\xi-\alpha_1)^2+\frac{1}{\alpha_2}[\alpha_2+\alpha_1^2-1+2\alpha_1(\xi-\alpha_1)^2} - \epsilon\ , \tag{46}$$

and from Corollary 2

$$E\{|f(t)|\frac{1}{\sqrt{\alpha_2}}+|g(t)|\sqrt{1+\frac{\alpha_1^2}{\alpha_2}}\}$$
$$\leq\ +2\xi-\sqrt{4(\xi-\alpha_1)^2+\frac{1}{\alpha_2}[\alpha_2+\alpha_1^2-1+2\alpha_1(\xi-\alpha_1)]^2}\ . \tag{47}$$

Inspection of these last three equations shows that, in general, unless $f(t)$ and $g(t)$ are related no optimum matrix B exists. Indeed, if stability conditions as a function of $E\{f^2(t)\}$ and $E\{g^2(t)\}$ are desired, equation (45) yields, upon application of the Schwarz's inequality, the condition

$$\alpha_2 E\{g^2(t)\}+[\alpha_1 E\{g^2(t)\}^{\frac{1}{2}}+E\{f^2(t)\}^{\frac{1}{2}}]^2 \leq 4\alpha_2-[\alpha_2-\xi^2+1+(\xi-\alpha_1)^2]^2-\epsilon, \tag{48}$$

and it is immediately seen that, for fixed values of α_1 and $\alpha_2 > 0$, it is not possible to obtain simultaneously results which coincide with those given by equation (32), in the event that $g(t) \equiv 0$, and with equation (43), if $f(t) \equiv 0$. Hence, the choice of α_1 and α_2 depends on the relative magnitudes of $E\{f^2(t)\}$ and $E\{g^2(t)\}$.

The two extreme choices for α_1 and α_2 are given by equations (31) and (42), in which cases we obtain the stability conditions

$$
\begin{gathered}
(\xi^2+1)E\{g^2(t)\}+[\xi E\{g^2(t)\}^{\frac{1}{2}}+E\{f^2(t)\}^{\frac{1}{2}}]^2 \leqq 4\xi^2-\epsilon, \\
(1-\frac{\xi^2}{(1+\xi^2)^2})E\{g^2(t)\}+[\frac{\xi}{1+\xi^2}E\{g^2(t)\}^{\frac{1}{2}}+E\{f^2(t)\}^{\frac{1}{2}}]^2 \leqq \frac{4\;{}^2}{1+\xi^2}-\epsilon.
\end{gathered}
\tag{49}
$$

The first of these equations yields equation (32) if $g(t) = 0$, while the second becomes equation (43) for $f(t) = 0$. Appropriate choices of α_1 and $\alpha_2 > 0$ will give results bounded by these two extremes.

If results are desired as functions of $E\{|f(t)|\}$ and $E\{|g(t)|\}$, equation (47) can be optimized by the values

$$\alpha_1 = \xi \quad , \quad \alpha_2 = 1-\xi^2 \qquad \text{if} \quad \xi^2 \leqq \frac{\sqrt{5}-1}{2} \quad ,$$

in which case the stability condition becomes

$$E\{|f(t)|+|g(t)|\} \leqq 2\xi\sqrt{1-\xi^2}-\epsilon, \text{ if } \xi^2 \leqq \frac{\sqrt{5}-1}{2}. \tag{50}$$

For $\xi^2 \geqq \frac{\sqrt{5}-1}{2}$ it is not possible to optimize simultaneously, and one is again forced to consider the relative magnitudes of $E\{|f(t)|\}$ and $E\{|g(t)|\}$. To obtain extreme values the values for α_1 and α_2 of equations (28) and (39) are used yielding

$$E\{|f| + \sqrt{2}\,\xi|g|\} \leq 1-\epsilon \quad , \qquad \text{if} \quad \xi \geq \frac{\sqrt{2}}{2} \; ;$$

$$E\{|f| + |g|\} \leq 2\xi\sqrt{1+\frac{1}{\xi^2}} - 1 \; , \qquad \text{if} \quad \xi \geq \frac{\sqrt{5}-1}{2} \; . \tag{51}$$

Again, appropriate choices of α_1 and $\alpha_2 > 0$ will yield results between these extremes.

As indicated previously, the results of this example are rather disappointing since they indicate that an optimum quadratic norm does not exist. On the other hand, it appears that if a differential equation has only one time varying coefficient then the determination of such a norm does not appear amenable to simple analysis.

<u>EXAMPLE 4</u>: <u>An Application</u>. Consider the application of the theorem of the previous section to the study of the stability of the solutions of the differential equations of the kinetics of a simple nuclear reactor problem. A set of differential equations modeling such a system is

$$\dot{n} = \frac{p(t)-\beta}{\ell}\, n + \lambda c \quad ,$$
$$\dot{c} = \frac{\beta}{\ell}\, n - \lambda c \quad , \tag{52}$$

where

c = concentration of total delayed neutron precursors ($c \geq 0$)

ℓ = neutron effective lifetime ($\ell > 0$)

n = neutron density ($n \geqq 0$)

$p(t)$ = reactivity, a function of time

β = total delayed neutron fraction ($\beta > 0$)

λ = mean decay constant of delayed neutron precursors. ($\lambda > 0$)

This set of equations and its variants have been the subject of numerous studies [7]. In [8], for example, it was proved that if $p(t)$ is sinusoidal, for every frequency of the sinusoid and values of the parameters, the solutions of (52) are unstable.

For notational simplicity, let

$$x_1 = n\ ,\ x_2 = c\ ,\ a = \frac{1}{\ell}\ ,\ b = \frac{\beta}{\ell} \tag{53}$$

and define

$$E\{\frac{p(t)}{\ell}\} = -m\ ,\quad f(t) = \frac{p(t)}{\ell} + m\ . \tag{54}$$

Equations (52) then becomes

$$\dot{x} = \begin{bmatrix} -m-b & \lambda \\ b & -\lambda \end{bmatrix} x + f(t)\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x\ , \tag{55}$$

in the same form as given by (1). Application of the matrix B given by (19) yields, after some computations and application of Schwarz's inequality, that the theorem of the previous section will predict stability for some $\alpha_1, \alpha_2 > 0$ and $\epsilon > 0$ if

$$E\{f^2(t)\}(\alpha_2+\alpha_1^2) \leq 4\alpha_2\lambda[m+b+\lambda\alpha_1]-\{b+\lambda(\alpha_1+\alpha_2)-\alpha_1(m+b+\lambda\alpha_1)\}^2-\epsilon . \quad (56)$$

The optimum values of α_1 and α_2 are immediately found to be,

$$\alpha_1 = 0, \qquad \alpha_2 = \frac{b}{\lambda} , \quad (57)$$

upon which (56) becomes

$$E\{f^2(t)\} \leq 4m\lambda ; \quad (58)$$

or, in the notation of equation (52), the condition for almost sure asymptotic stability in the large becomes

$$E\{p(t)^2\} \leq E\{p(t)\}^2 - 4\lambda\ell E\{p(t)\}-\epsilon . \quad (59)$$

It is evident from this expression that $E\{p(t)\}$ must be negative for stability. In the specific case that the reactivity varies sinusoidally as

$$p(t) = -m + h \sin \omega t \quad (60)$$

stability condition (59) becomes

$$h^2 \leqq 8\, m\ell\lambda - \epsilon\,, \tag{61}$$

for some $\epsilon > 0$.

CONCLUSIONS

A simple theorem that gives sufficient conditions for the almost sure stability of linear time varying systems has been presented. As the applications of this theorem and its corollaries to examples show, the stability results obtained are quite good and simple to use. The question of determination of the optimum quadratic norm for a system of differential equation with only one time varying coefficient has not been resolved, and remains an open problem.

REFERENCES

[1] W. Hahn, Theory and Application of Liapunov's Direct Method, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

[2] F. Kozin, "On Almost Sure Stability of Linear Systems with Random Coefficients", Journal of Mathematics and Physics, 42(1963), 59-67.

[3] T.K. Caughey and A.H. Gray, Jr., "On the Almost Sure Stability of Linear Dynamic Systems with Stochastic Coefficients", Journal of Applied Mechanics, 32(1965), 365-372.

[4] S.T. Ariaratnam, "Dynamic Stability of a Column Under Random Loading", Dynamic Stability of Structures, Proceedings of International Conference, Pergamon Press, New York, 1967, 267-284.

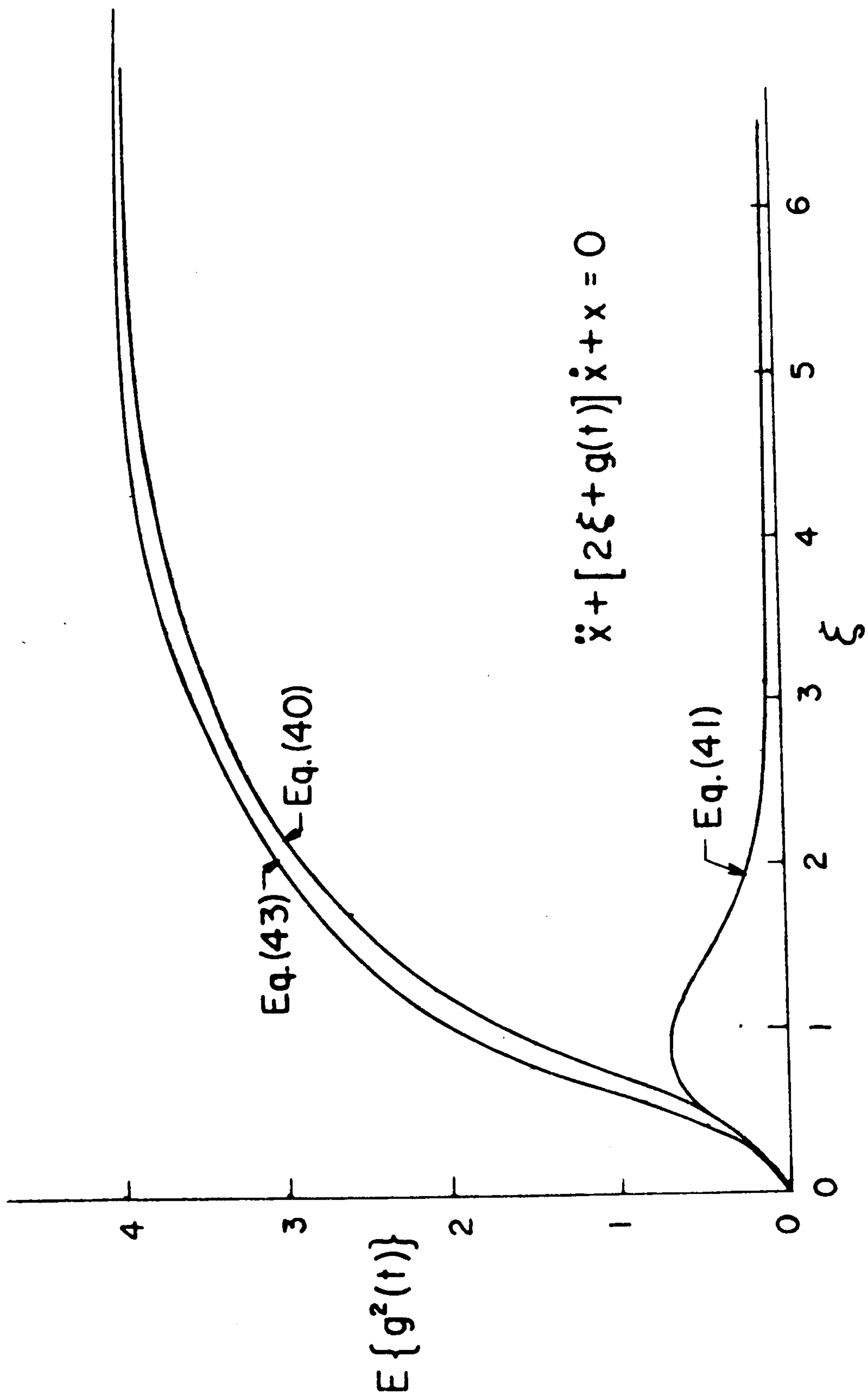
[5] F.R. Gantmacher, The Theory of Matrices, Chelsea Publishing Co., New York, 1959.

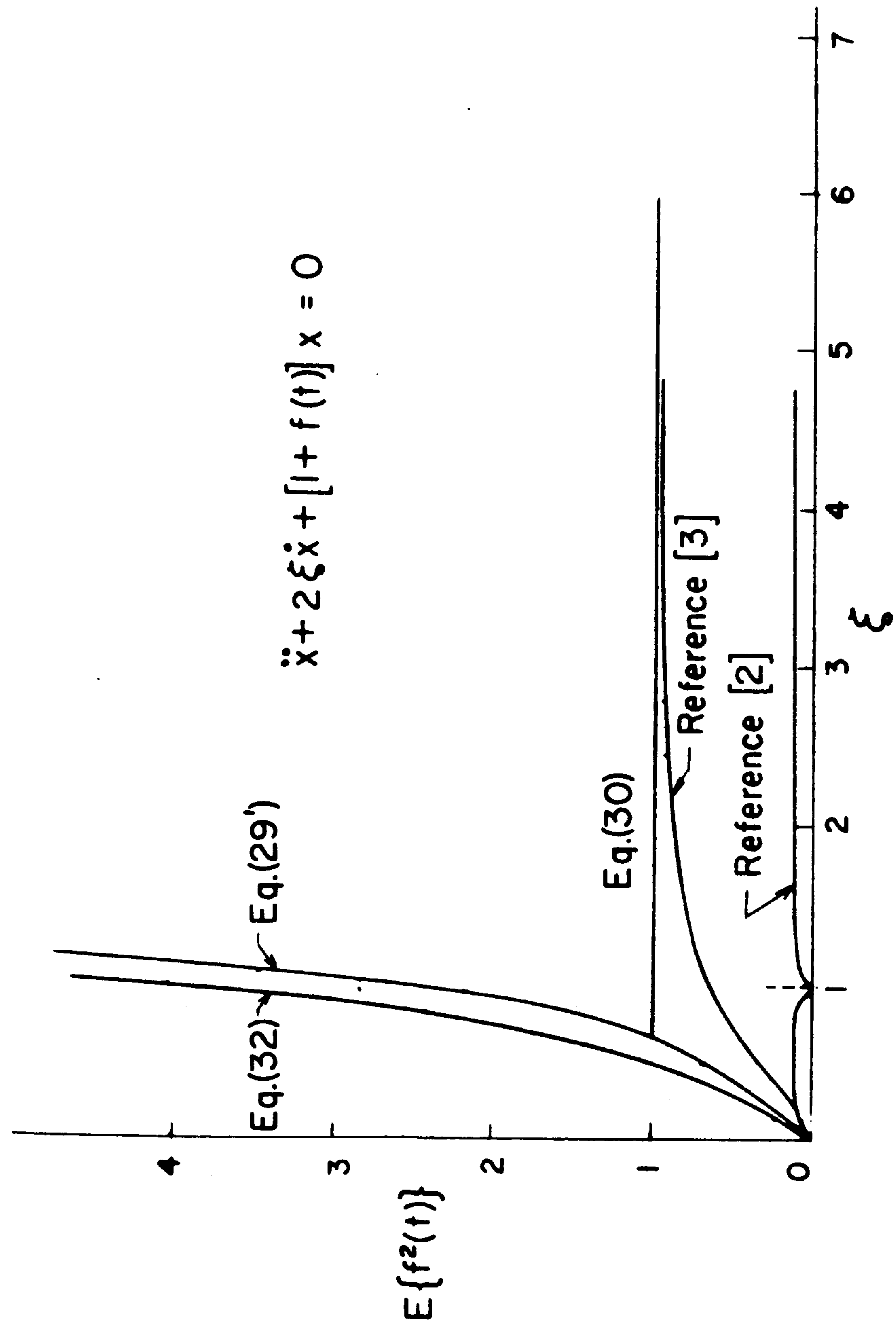
[6] C.B. Mehr and P.K.C. Wang, Discussion of [3], Journal of Applied Mechanics, 33(1966), 234-236.

[7] H.B. Smets, Problems in Nuclear Power Reactor Stability, Press Universitaires de Bruxelles, Belgium, 1962.

[8] A.L. Ruiz, "On the Instability of the Reactor Kinetics Equations with Sinusoidally Varying Reactivity", Argonne National Laboratory, Report, 1960.

LIST OF FIGURES


$\ddot{x} + [2\xi + g(t)]\dot{x} + x = 0$
$E\{g^2(t)\}$
0
1
2
3
4
ξ
0
1
2
3
4
5
6
Eq.(43)
Eq.(40)
Eq.(41)



$\ddot{x} + 2\xi\dot{x} + [1 + f(t)]\,x = 0$
$E\{f^2(t)\}$
ξ
0
1
2
3
4
5
6
7
Eq.(32)
Eq.(29')
Eq.(30)
Reference [3]
Reference [2]

PAPER [9]

Submitted to the American Journal of Mathematics

ENERGY FUNCTIONS FOR MORSE SMALE SYSTEMS

by

K. R. Meyer*

*This Research was supported by the National Aeronautics and Space Administration, Huntsville, under Contract Number NAS 8-11264.

Energy Functions for Morse Smale Systems

by

K. R. Meyer

I. Introduction.

In [1] Smale introduced a class of vector fields on a manifold that are similar to gradient fields generated by Morse functions and have since been called Morse-Smale systems. Morse-Smale systems are allowed to have a finite number of closed orbits and singular points but they share with gradient fields the property that the α and ω limit sets of every trajectory can only be a singular point or a closed orbit. Hence there is no complicated recurrent motion. A Morse-Smale system without closed orbits is called gradient like. In [2] it is shown that for every gradient like system there exists a Morse function that is decreasing along trajectories. In this paper a larger class of functions is considered, called $\mathcal{E}$-functions, and it is shown in Theorem 1 that for every Morse-Smale system there exists an $\mathcal{E}$-function that is decreasing along the trajectories of the system. This reminds one of the energy function associated to a dissipative system in mechanics and hence the name $\mathcal{E}$-function.

The construction of the $\mathcal{E}$-function requires little more effort but the added generality has suggested new questions that are discussed here. It is natural to ask if the association of an $\mathcal{E}$-function to a Morse-Smale field is unique in some sense. Theorem 2 establishes that the functions corresponding to a particular field are topologically equivalent.

Several interesting special results are also obtained when the manifold is compact and two dimensional. In this case one has a necessary and sufficient condition for structural stability in terms of $\mathcal{E}$-functions and moreover there is a one-to-one correspondence between topological equivalence classes of structurally stable fields and $\mathcal{E}$-functions.

II. Definitions and Preliminaries.

In this paper smooth will always mean C^{∞}. Let M be a closed smooth manifold of dimension m with a distance function d inherited from some Riemannian metric. R^n will be Euclidean n-space, S^n the unit sphere in R^{n+1} and B^n the open unit ball in R^n. If X is a smooth vector field on M then φ_t will denote the 1-parameter group of diffeomorphisms generated by X. If $p \in M$ then $\gamma(p)$ will denote the trajectory of X through p, i.e. $\gamma(p) = \bigcup_t \varphi_t(p)$. If $p \in M$ then the α and ω limit sets of $\gamma(p)$ are defined in the usual manner by $\alpha(p) = \bigcap_{\tau \leq 0} \bigcup_{t \leq \tau} \varphi_t(p)$ and $\omega(p) = \bigcap_{\tau \geq 0} \bigcup_{t \leq \tau} \varphi_t(y)$.

If A is a subset of M then A^{o} will denote the topological interior of A and A^{-} the topological closure of A.

Definition: A smooth vector field X will be called a Morse-Smale system (or field) provided

1) X has a finite number of singular points, say $\beta_1, \ldots, \beta_k$, each of generic type. A generic singular point is a singular point such that in local coordinates the matrix of partial derivatives of X has eigenvalues with non-zero real parts.

2) X has a finite number of closed orbits (i.e. periodic solutions), say $\beta_{k+1}, \ldots, \beta_n$, each of generic type. A generic orbit is a closed orbit such that all the characteristic multipliers, except the one corresponding to the orbit itself, have modulus different from one.

3) For any $p \in M$, $\alpha(p) = \beta_i$ and $\omega(p) = \beta_j$ for some i and j.

4) If β_i is a closed orbit then there is no $p \in M-\beta_i$ such that $\alpha(p) = \beta_i$ and $\omega(p) = \beta_i$.

5) The stable and unstable manifolds associated with the β_i have transversal intersection.

The sets $\beta_1, \ldots, \beta_n$ will be called the singular elements of the field X.

Let W_i and W_1^* denote the unstable and stable manifold associated to β_i. See [1] and [2] for a discussion of condition 5) and definition of W_i and W_i^*. Note that in [1] transversal intersection is called a normal intersection. A large number of the lemmas in [1] can be summarized by the following:

<u>Lemma</u>: Let X be a Morse-Smale system on M. Let $\beta_i \succ \beta_j$ mean that there is a trajectory not equal to β_i or β_j whose α-limit set is β_i and whose ω-limit set is β_j. Then $\succ$ satisfies:

1) it is never true that $\beta_i \succ \beta_i$

2) if $\beta_i \succ \beta_j$ and $\beta_j \succ \beta_\ell$ then $\beta_i \succ \beta_\ell$ (thus $\succ$ is a partial ordering of $\bigcup \beta_i$)

3) if $\beta_i \succ \beta_j$ then $\dim W_i \geq \dim W_j$ and equality can only occur if β_j is a closed orbit.

Let E be a smooth function from M into R and let Δ denote the set of critical points of E. Let Δ_i denote the set of points in Δ where the Hessian of E has nullity i. It is well known (see [3]) that Δ_o is a finite union of points, say

$\delta_1, \ldots, \delta_k$, and there exists a coordinate system (N_i, x_i) such that

$$E \circ x_i^{-1} = E(\delta_i) + Q(x)$$

where Q is a nonsingular quadratic form in x whose index is the same as the index of the Hessian of E at δ_i. For discussion and definitions relevent for these functions see [3].

Definition: A smooth function E from M into R will be called on $\mathscr{E}$-function for M provided

1) $\Delta = \Delta_o \cup \Delta_1$
2) Δ_1 is the disjoint union of a finite number of circles, i.e. closed connected one dimensional submanifolds of M, such that the Hessian of E is constant on each circle. Denote these circles by $\delta_{k+1}, \ldots, \delta_n$.
3) For $i = k+1, \ldots, n$ there exists a neighborhood N_i of δ_i and a diffemorphism x_i such that x_i maps N_i into the product of B^{m-1} and S^1 if N_i is orientable or into the twisted product of B^{m-1} and S^1 if N_i is nonorientable with the property that $E \circ x_i^{-1} = E(\delta_i) + Q(x)$ where Q is a nonsingular quadratic form in $x_1, \ldots, x_{m-1}$, the coordinates in B^{m-1}, and is periodic of period 1 in x_m, the coordinate in S^1. Moreover, for each point in S^1 the quadratic form has index equal to the index of

E on δ_i.

In this paper the connection between Morse-Smale systems and $\mathcal{E}$-functions is investigated. In this respect the $\mathcal{E}$-function is closely related to the field when $\mathcal{E}$ is decreasing along trajectories. To formalize this we need:

Definition: Let X be a smooth vector field on M. Then an $\mathcal{E}$-function, E, for M will be called an $\mathcal{E}$-function for X provided

1) $XE(p) < 0$ for all $p \in M-\Delta$, i.e. E is decreasing along the trajectories of X or the trajectories of X are transversal to the level lines of E
2) if p is a singular point of X then $p \notin \Delta_1$
3) there exists a constant $\kappa > 0$ such that on each N_i

$$-XE(p) \geq \kappa d(p,\delta_i)^2 \quad \text{for} \quad p \in N_i$$

III. Existence of $\mathcal{E}$-functions.

The first result is that Morse-Smale systems admit $\mathcal{E}$-functions, that is

Theorem 1: If X is a Morse-Smale system then there exists an $\mathcal{E}$-function for X.

Proof: The first step is to define the $\mathcal{E}$-function on the β_i and since E must be decreasing along trajectories this must be done in a consistent way. The lemma shows that this can be done, that is, one can find n real numbers α_i such that if $\beta_i \succ \beta_j$

then $\alpha_i > \alpha_j$. Thus we define E on the β_i by $E(\beta_i) = \alpha_i$ and then construct E globally so that $\delta_i = \beta_i$ and E is decreasing along trajectories. Next E must be extended to a neighborhood of the β_i in such a way that the nondegenerating conditions are satisfied. If β_i is a singular point then in local coordinates X has the form $\dot{x} = Ax + f(x)$ where $f(0) = df(0) = 0$ and the eigenvalues of A have nonzero real parts. By Liapunov theory there exists symmetric matrices Q and C, C positive definite and Q nonsingular such that $A'Q + QA = -C$. Moreover, the index of Q is equal to the number of eigenvalues of A with positive real part. If we define $E(x) = \alpha_i + x'Qx$ then by standard Liapunov arguments there exists a neighborhood sufficiently small and a constant $\kappa_i > 0$ such that $-XE(p) \geq \kappa_i d(x,p)^2$ in this neighborhood. Take the N_i sufficiently small that the above holds and so that they do not overlap.

Now around a closed orbit β_i one can choose a neighborhood N_i' and a diffeomorphism x_i' mapping N_i' into $B^{m-1} \times S^1$ or B^{m-1} twisted product with S' (if N_i' is non orientable) such that if y is the coordinates in B^{m-1} and θ is the coordinate in S^1 then X takes the form

$$\dot{\theta} = \omega + \Theta(\theta,y)$$
$$\dot{y} = A(\theta)y + Y(\theta,y)$$

where A is an $(m-1) \times (m-1)$ periodic matrix of period 1 i.e. A is a function on S^1. Θ and Y are periodic of period 1 in θ

$\Theta(\theta,0) = 0$ and $Y = o(\|y\|)$. By Floquet theory the fundamental matrix solution of $u' = -uA$ can be written in the form $e^{S\theta}P(\theta)$ where S is constant and P is either periodic or skew periodic of period 1 i.e. either $P(\theta) = P(\theta+1)$ or $P(\theta) = -P(\theta+1)$. By assumption S has no eigenvalue with zero real part and so by Liapunov theory there exists symmetric matrices Q and C, Q nonsingular and C positive definite such that $S^TQ + QS = C$. Define $E \circ x_i^{-1} = \alpha_i + y^TP(\theta)^TQP(\theta)y$ by direct computation then $XE = -y^TP^T(\theta)CP^T(\theta)y + \varepsilon(\theta,y)$ where $\varepsilon = o(\|y\|^2)$. We again restrict the neighborhood N_i so that they do not overlap and so that $-XE(p) \geq \kappa d(\delta_i, p)^2$ for $p \in N_i$.

Thus the $\mathcal{E}$-function is now defined in neighborhoods of the singular points and closed orbits of X. The extension of this function can now be accomplished by the same procedure as in [2]

As a partial converse of the above theorem we have

Proposition: Let X be a smooth vector field on M. If there exists an $\mathcal{E}$-function for X then X satisfies the conditions 1) 2) 3) and 4) in the definition of a Morse-Smale system. Moreover, the field X can be approximated arbitrarily closely in the C^r-topology for fields on M by a Morse-Smale system.

The first part follows by standard Liapunov arguments and the second part is established essentially the same way as proposition 2 in [4].

If M is compact and 2-dimensional the above result can be sharpened. In this case Morse-Smale systems are the same as structurally stable systems by a theorem of Peixoto [5]. If E is an $\mathcal{E}$-function for X such that all the sources of X lie in $E^{-1}(1)$; all saddle points of X lie in $E^{-1}(0)$ and all sinks of X lie in $E^{-1}(-1)$ then E will be called a special <u>$\mathcal{E}$-function for X</u>. It is clear from the above that if M is compact and two-dimensional then the construction of Theorem 2 could be made to yield a special $\mathcal{E}$-function for X.

If E is a special $\mathcal{E}$-function for X then there can be no trajectory joining saddle points of X since E is decreasing along trajectories. Thus the stable and unstable manifolds have transversal intersection. Hence

<u>Corollary</u>: If M is compact and two dimensional then a necessary and sufficient condition for X to be structurally stable is the existence of a special $\mathcal{E}$-function for X.

IV. <u>Uniqueness of $\mathcal{E}$-functions</u>.

Clearly the $\mathcal{E}$-function constructed in Theorem 1 is not unique but if one introduces the concept of topological equivalence a form of uniqueness can be established.

Recall (see [6]) that two functions E and E' from M to R are said to be topologically equivalent if there exists homeomorphisms f and g, $f:M \to M$ and $g:R \to R$ such that the following diagram commutes

$$\begin{tikzcd}
M \arrow[r, "E"] \arrow[d, shift right, "f"'] & R \arrow[d, shift left, "g"] \\
M \arrow[r, "E'"] \arrow[u, shift right, "f^{-1}"'] & R \arrow[u, shift left, "g^{-1}"]
\end{tikzcd}$$

Recall that two vector fields X and X' on M are said to be topologically equivalent if there exists a homeomorphism $h: M \to M$ which sends the trajectories of X into the trajectories of X'.

In general two $\mathcal{E}$-functions for two topologically equivalent fields are not topologically equivalent since the $\mathcal{E}$-functions are defined quite arbitrarily on the singular points and closed orbits. To obtain uniqueness some regularity on the way the $\mathcal{E}$-functions are defined on the β_i's is necessary. This could be done by uniquely specifying the way the functions are defined on the β_i's as was done in the definition of the special $\mathcal{E}$-functions for two-dimensional fields. Instead of this we assume that the $\mathcal{E}$-functions are defined correctly on the critical elements.

Since hence forth we shall be considering two fields and two functions we shall use the same symbols as before and all unprimed symbols will refer to one system and all primed to the other.

<u>Theorem 2</u>. Let X and X' be two Morse-Smale systems on M that are topologically equivalent under the homeomorphism h. Let E and E' be $\mathcal{E}$-functions for X and X' respectively. Then if the two $\mathcal{E}$-functions are equivalent on the singular elements, i.e. on the singular points and closed orbits, then they are topologically equivalent. That is to say if there exists a homeomorphism $g: R \to R$ such that the diagram

$$\begin{tikzcd}
\bigcup_1^n \beta_i \arrow[r, "E"] \arrow[d, shift right, "h"'] & R \arrow[d, shift left, "g"] \\
\bigcup_1^n \beta_i' \arrow[r, "E'"] \arrow[u, shift right, "h^{-1}"'] & R \arrow[u, shift left, "g^{-1}"]
\end{tikzcd}$$

commutes then E and E' are topologically equivalent.

Proof: Let β_i and β'_i be so numbered that $h(\beta_i) = \beta'_i$. Observe that g is by assumption a homeomorphism of R into R that must satisfy a finite number of other requirements, namely $g\circ E(\beta_i) = E'(\beta'_i)$. If such a g exists then a smooth $\tilde{g}$ exists satisfying the same conditions. Hence we can assume that E and E' agree the singular elements since otherwise we would consider E and $\tilde{g}\circ E'$.

We first define a special neighborhood of one singular element. Let β represent any one of the β_i or β'_i and N,x and E the corresponding N_i, N'_i, x_i, x'_i, E or E'. Then a $\mathcal{P}$-neighborhood, P, of β is defined as a closed neighborhood of β contained in N such that the boundary of P is the union of three sets: I a $m-1$ closed submanifold of M that lies in the level line $E^{-1}(E(\beta)+\varepsilon)$ for some $\varepsilon > 0$ or $I = \varphi$, O a $m-1$ closed submanifold of M that lies in the level line $E^{-1}(E(\beta)-\varepsilon)$ for some $\varepsilon > 0$ or $O = \varphi$ and U the union of trajectories that join the boundary of I to the boundary of O.

Such a neighborhood always exists as can be seen by the following. If β is a source take P to be the set of points in N where E is greater than $E(\beta)-\varepsilon$ with ε small and positive. If β is a sink P is defined similarly. Let β be a saddle point. Then $E(x) = E(\beta) + Q(x)$ in N where Q is a nonsingular quadratic form. Let T be the quadratic form that is equal to Q on the subspace of R^n where Q is negative definite and zero on the complement. For ε and δ sufficiently small the set I of points p where

$p \in E^{-1}(E(\beta) + \varepsilon)$ and $-T(p) \leq \delta$ is contained in the interior of N.

Moreover, if ε and δ are sufficiently small one can also fulfill the requirement that the set of all points p that lie on a trajectory through I and satisfy $E(\beta) - \varepsilon \leq E(p) \leq E(\beta) + \varepsilon$ is contained in N, let P be the closure of this set. It is easy to see that P is a closed neighborhood of β contained in N and that the boundary of P is composed of I as defined above, O and U where O and U satisfy the requirements of the definition of a $\mathcal{P}$-neighborhood. $\mathcal{P}$-neighborhoods for closed orbits are constructed in a similar way.

Let P_i be a $\mathcal{P}$-neighborhood for β_i, $i = k+1, k+2, \ldots, n$ and P_i^o its interior. We first construct f on $M - \bigcup_1^k \beta_i - \bigcup_{k+1}^n P_i^o$. Let $p \in M - \bigcup_1^k \beta_i - \bigcup_{k+1}^n P_i^o$ and define $f: p \to q$ where q is defined as the unique point on the X'-trajectory through $h(p)$ that satisfies $E'(q) = E(p)$. To make sure that this map is well defined observe that $E(\varphi_t(p))$ and $E'(\varphi_t'(h(p)))$ tend to the same limit as $t \to +\infty$ and the same limit as $t \to -\infty$ and moreover both are decreasing functions of t. Thus f is a homeomorphism taking level line into level line where it is defined.

Now f can be extended to the singular points by $f(\beta_i) = \beta_i'$ for $i = 1, \ldots, k$. To see that f is still a homeomorphism note that f maps $\mathcal{P}$-neighborhood of β_i onto $\mathcal{P}$-neighborhoods of β_i' and conversely. For closed orbits the extension is more difficult since the β_i no longer consist of single points.

The homeomorphism f is defined on the boundary of P_i, $i = k+1,\dots,n$ and maps the boundary of P_i into the boundary of a $\mathscr{P}$-neighborhood, P'_i, of β'_i. To see this observe that the image I'_i of I_i under f is contained in a level line of E' and similarly for the image of O. Moreover the image of U is the union of X' trajectories joining the boundary of I' to the boundary of O'. P'_i is defined once I' or O' are defined as can be seen by our construction of $\mathscr{P}$-neighborhoods.

We now show how to extend the definition of f to the interiors of P_i, $i = k+1,\dots,n$. Since we shall be working locally we shall drop the subscripts. For definiteness let us consider the case when the neighborhood of β and the stable and unstable manifolds of β are orientiable. The nonorientable cases are similarly treated.

First let β be a source or a sink. Let N be a neighborhood of β containing P and x a diffeomorphism $x:N \to B^m \times S^1$, $x = (y,\theta)$, $y:N \to B^{m-1}$, $\theta:N \to S^1$ such that in N, $E(x) = E(\beta) + y^Ty$. Let P', N', x', y', θ' be similarly defined. For simplicity let E be zero on β and 1 on the boundary of P. f is defined on the boundary of P and let $f = h$ on β. Let $p = (y,\theta) \in P^o - \beta$. p is on the curve $\tau(0,\theta) + (1-\tau)(\|y\|^{-1}y,\theta)$. Let $f(0,\theta) = (0,\theta'_o)$ and $f((\|y\|^{-1}y,\theta)) = (y'_1,\theta'_1)$ and let q be the unique point on the curve $\tau(y'_1,\theta'_1) + (1-\tau)(0,\theta^1_o)$ that satisfies $E(p) = E'(q)$. By defining $f(p) = q$ we see that f has been extended to the criterior of P as a homeomorphism taking level line into level line.

Now let β be a saddle type closed orbit. Let N be a neighborhood of β containing P and $x = (y,\theta)$ a diffeomorphism

$y:N \to B^{n-1}$ and $\theta:N \to S^1$ such that in these local coordinates $E(x) = E(\beta) + y^T Q y$ where $Q = \mathrm{diag}\,(1,\ldots,1,-1,\ldots,-1)$. Moreover, let N', x', y', θ' be similarly defined. Let Π be a $\mathscr{P}$-neighborhood of β completely interior to P. Define f on Π by $f:p \to q$ where $p \in \Pi$ and $q \in \Pi'$ and p and q have the same numerical coordinates in the unprimed and primed coordinates respectively.

Thus f must be extended to $P^o - \Pi$. This extension can be accomplished by dividing $P^o - \Pi$ into several parts each of which has a simple geometric type. Let a and b be the real numbers such that the region of the boundary of Π that is a region of ingress resp. egress is in the level line $E^{-1}(a)$ resp. $E^{-1}(b)$. Consider $K_1 = E^{-1}(\alpha) \cap (P-\Pi^o)$ and $K_2 = E^{-1}(b) \cap (P-\Pi^o)$. f is defined on the boundary of K_1 and K_2 and topologically K_1 and K_2 are just products of unit intervals and spheres. Let L_1 $(P-\Pi^o) \cap \{p \in M: E(I) \geq E(p) \geq a\}$, $L_2 = (P-\Pi^o) \cap \{p \in M: E(a) \geq E(p) \geq E(b)\}$ and $L_3 = (P-\Pi^o) \cap \{p \in M: E(b) \geq E(p) \geq E(0)\}$.

Topologically L_1, L_2 and L_3 are just the product of the unit interval and spheres. f is defined on the boundary of K_1 and K_2 and so we first extend f to K_1 and K_2. Now f is defined on the boundary of L_1, L_2 and L_3 and so f is then extended to their interiors.

Each extension is carried out in the same way as the extension was carried out for the source because in each case there is a set that acts as the center. That is if one of the sets is $I \times I \times S^1$ then $(0,0) \times S^1$ is the center.

The center is mapped homeomorphically on the center by fiat and then the extension is carried out by joining the center to the boundary by lines and carrying points proportionally.

Of course for special $\mathcal{E}$-functions the homeomorphism g may always be taken as the identity. In the case where M is compact and two dimensional the converse of Theorem 4 holds also. Namely

<u>Proposition:</u> Let M be a compact and two dimensional smooth manifold. Let X and X' be smooth vector field on M and let E and E' be special $\mathcal{E}$-functions for X and X' respectively. If E and E' are topologically equivalent then X and X' are topologically equivalent.

<u>Proof:</u> Let f be the homeomorphism of M that takes level lines of E into level lines of E' i.e. $E = E' \circ f$. f sets up a correspondence between the critical elements of E and E' let them be so numbered that $f(\delta_i) = \delta_i'$ and let the β_i and β_i' be numbered so that $\beta_i = \delta_i$ and $\beta_i' = \delta_i'$ as sets. Let $\Gamma = E^{-1}(0)$ and $\Gamma' = E'^{-1}(0)$. Then f is a homeomorphism of Γ onto Γ'. Define h to be equal to f on Γ.

The first thing to be established is that if $p \in \Gamma$ and $\alpha(p) = \beta_i$ and $\omega(p) = \beta_j$ then $\alpha'(f(p)) = \beta_i'$ and $\omega'(f(p)) = \beta_j'$. Let $p \in \Gamma$ and p not a saddle point and let $p^* = f(p)$. Consider the X'-trajectory through p^* and let it be reparameterized so that it is a map u from $(-1,1)$ into M where the new parameter is the value of E', this can be done since $E'(\varphi_t'(p^*))$ is a decreasing function of t. To be precise $u:(-1,1) \to M$ such that $u(\alpha)$, $\alpha \in (-1,1)$, is the unique point on the X' trajectory through p^* such that $E'(u(\alpha)) = \alpha$. In a similar manner let $v:(-1,1) \to M$ be the reparameterization of $f(\varphi_t(p))$ by values of E'. To be precise

$v(\alpha)$, $\alpha \in (-1,1)$, is the unique point on $f(\varphi_t(p))$ such that $E'(v(\alpha)) = \alpha$. We want to show that u and v are isotopic with an isotopy that moves points in a level line. That is we want to show that there exists a map $V:(-1,1) \quad [0,1] \to M$ such that $V(\cdot,0) = u$ and $V(\cdot,1) = v$ and moreover $E'(V(\alpha,t)) = \alpha$ for all $t \in [0,1]$. Clearly this will establish the fact that α and ω limit sets of trajectories correspond as described above.

Let A be a small disk about p' such that A contains no singular points of X'. For α different from zero the level lines $E'^{-1}(\alpha)$ is a smooth one manifold and so there is a unique arc a in A joining $u(\alpha)$ to $v(\alpha)$ of arc length $s(\alpha)$. Let $V(\alpha,t)$ be the unique point on the arc a such that the arc length from $u(\alpha)$ to $V(\alpha,t)$ is $ts(\alpha)$. Thus the isotopy V is defined so long as α is small but the extension is now obvious and so our claim is established.

The sets $\{p \in M:E'(p) \geq \frac{1}{2}\}$ and $\{p \in M:E' \leq -\frac{1}{2}\}$ are the disjoint union of $\mathscr{D}$-neighborhoods of all the sources and sinks respectively.

The homeomorphism h is now extended in the following way. Let p be a point of M not on a separatrix of X and such that $-\frac{1}{2} \leq E(p) \leq \frac{1}{2}$. The X trajectory through p meets at p^* let q be the unique point on the X'-trajectory through $h(p^*) = f(p^*)$ that satisfies $E(p) = E'(q)$. Now extend this map to all of $\{p \in E:-\frac{1}{2} \leq E(p) \leq \frac{1}{2}\}$ so that separatrix goes to separatrix and level line of E to level line of E'.

Thus the map f is defined on all but the interiors of $\mathscr{D}$-neighborhoods of the sources and sinks. The map f is defined on the boundaries of these $\mathscr{D}$-neighborhoods and takes the boundary of one particular $\mathscr{P}$-neighborhood of an X critical element into the boundary of a $\mathscr{P}$-neighborhood of an X' critical element of the same type.

But it is shown in [7] that if one is given two critical elements of the same type and an arbitrary homeomorphism of the boundaries of $\mathscr{P}$-neighborhoods for these two critical elements then the homeomorphism can be extended to the interior of the neighborhoods taking trajectories into trajectories. Thus f can be defined globally.

References

[1] Smale, S., Morse inequalities for dynamical systems, Bull. Ann. Math. Soc., 66 (1960), 43- 9.

[2] Smale, S., On gradient dynamical systems, Ann. Math., 74 (1961), 199-206.

[3] Milnor, J., Morse Theory, Princeton University Press, (1963).

[4] Peixoto, M.M., On an approximation theorem of Kupka and Smale, (to appear).

[5] Peixoto, M. M., Structural stability on two-dimensional manifolds, Topology, 1, (1962), 101-120.

[6] Thom, R., La stabilite topologique des applications polynomiales, L'Enseignement Mathematique, VIII, (1962), 24-33.

[7] Peixoto, M. C., and Peixoto, M.M., Structural stability in the plane with enlarged boundary conditions, Ann. Bras. Acad. Sci., 31 (1959), 135-160.

PAPER [10]

To be submitted

On the Linearization of Volterra Integral Equations

R. K. Miller*

Center for Dynamical Systems

Brown University
Providence, Rhode Island

May, 1967

*This Research was supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, U. S. Air Force under AFOSR Grant No., 693-67, in part by the National Aeronautics and Space Administration, Huntsville, under Contract No. NAS8-11264 and in part by the National Aeronautics and Space Administration under Grant No. NGR 40-002-015.

On the Linearization of Volterra Integral Equations

I. Introduction.

Given a nonlinear differential equation

$$(1) \qquad x' = Cx + o(|x|), \quad (' = d/dt)$$

it is well known that the asymptotic stability of the linear system $y' = Cy$ implies the local asymptotic stability of the trivial solution of (1). All known proofs of this fact depend on the fact that solutions of the linear system decay exponentially or the equivalent fact that there exists a quadratic Lyapunov function for the linear system.

Consider a system of n equations of the form

$$(2) \qquad x(t) = f(t) + \int_0^t a(t-s)g(x(s))ds, \quad t \geq 0$$

where x,f and g are n-vectors, $a(t)$ is an $n \times n$ matrix and $g(0) = 0$. If f is "small" this system is often replaced by the linear system

$$(3) \qquad y(t) = f(t) + \int_0^t a(t-s)Jy(s)ds,$$

where J is the Jacobian matrix $g'(0) = (\partial g_i(0)/\partial x_j)$.

Levin and Nohel have proved by example that solutions of equations of the form (3) need not decay exponentially, cf. [1, p. 350, line (2.11)]. Indeed there seems to be no known mathematical justification for linearization of Volterra integrodifferential equations. The purpose of this paper is to provide in Section II below mathematical justification for the linearization of equation (3) under certain conditions on the matrix $a(t)J$.

If $f(t) \equiv X_0$ is constant and $a(t) \equiv 1$, then equation (2) is equivalent to an ordinary differential equation. In this case our criterion reduces to the known condition that the eigenvalues of J have negative real parts.

The advantage of our method is that one can replace the local, nonlinear problem (2) by the linear equation (3) and the linear equation for its resolvent. These linear equations may be studied using known methods such as transform techniques. In Sections III, IV and V below we give some examples which illustrate this.

In the sequel we shall need the following notations and conventions. Let R^n denote real n-space with a norm $|x|$. Let $|D|$ denote the corresponding matrix norm. Let $BC[0,\infty)$ be the space of bounded continuous functions on $0 \leq t < \infty$ with norm

$$\|h\|_0 = \sup \{|h(t)|;\ 0 \leq t < \infty\}.$$

Similarly $BC(R)$ will be the space of bounded continuous functions on $-\infty < t < \infty$ with norm

$$\|h\|_1 = \sup \{|h(t)|;\ -\infty < t < \infty\}.$$

II. General Stability Conditions.

Concerning equation (2) we assume:

(A1) $a \in L^1(0,t)$ for each $t > 0$,

(A2) $f(t) \in BC[0,\infty)$,

(A3) $g(x) \in C^1(R^n), g(0) = 0$ and

(A4) the Jacobian matrix J is nonsingular.

Since we assume J is nonsingular, it is no loss of generality to assume J is the $n \times n$ identity matrix I. We need only replace $a(t)$ by $a(t)J$ and $g(x)$ by $J^{-1}g(x)$. Thus equation (3) may be rewritten in the form

$$(3') \qquad y(t) = f(t) + \int_0^t a(t-s)y(s)ds.$$

It is well known that the unique solution of equation (3') has the form

$$(4) \qquad y(t) = f(t) - \int_0^t b(t-s)f(s)ds, \quad (t \geq 0)$$

where the matrix b is the resolvent kernel determined by the matrix equation

$$(5) \qquad b(t) = -a(t) + \int_0^t b(t-s)a(s)ds.$$

We assume that

(A5) the matrix b determined by (5) exists for all $t > 0$ and $|b(t)| \in L^1(0,\infty)$.

<u>Theorem 1.</u> <u>If assumptions</u> (A1-5) <u>are satisfied then there exists</u> $\epsilon_0 > 0$ <u>and</u> $\epsilon_1 > 0$ <u>such that when the solution</u> $y(t)$ of (3')

satisfies $\|y\|_0 \leq \epsilon_0$ *the solution* $x(t)$ *of* (2) *exists for all* $t \geq 0$ *and* $\|x\|_0 \leq \epsilon_1$.

Proof. Since $b \in L^1(0,\infty)$ it follows that equation (2) is equivalent to the system

$$x(t) = y(t) - \int_0^t b(t-s)G(x(s))ds, \tag{6}$$

where y is defined by line (4) and

$$G(x) = g(x) - x = o(|x|). \qquad (|x| \to 0)$$

Pick $\epsilon_1 > 0$ such that if $|x| \leq \epsilon_1$, then

$$2|G(x)| \int_0^\infty |b(s)|ds \leq |x|,$$

and $\int_0^\infty |b(s)|ds|g'(x) - I| < 1$. Pick $\epsilon_0 = \epsilon_1/2$. Let $Tx(t)$ be the function defined by the right hand side of equation (6). Let

$$S(0,\epsilon_1) = \{h \in BC[0,\infty);\ \|h\|_0 \leq \epsilon_1\}.$$

Our estimates on ϵ_0 and ϵ_1 imply that T is a contradiction map on $S(0,\epsilon_1)$. This proves Theorem 1.

Corollary 1. *If* (A1-5) *are satisfied, then there exist* $\epsilon_1 > 0$ *and* $\epsilon_2 > 0$ *such that when* $\|f\|_0 \leq \epsilon_2$ *the solution* $x(t)$ *of* (2) *exists for all* $t \geq 0$ *and satisfies* $\|x\|_0 \leq \epsilon_1$.

Proof. Pick ϵ_2 such that

$$\epsilon_2(1 + \int_0^\infty |b(s)|ds) \leq \epsilon_1/2,$$

where ϵ_1 is the constant given in Theorem 1. Then equation (4) above implies $\|y\|_o \leq \epsilon_o$. Thus Corollary 1 follows from Theorem 1 above.

Theorem 2. _Let_ (A1-5) _hold and let_ ϵ_o _and_ ϵ_1 _be given by_ _Theorem_ 1 _above_. _If_ $\|y\|_o \leq \epsilon_o$ _and_ $y(t) \to 0$ _as_ $t \to \infty$, _then_ $x(t) \to 0$ _as_ $t \to \infty$.

Proof. Let Γ be the positive limit set of the solution $x(t)$, that is Γ is the smallest set such that $x(t)$ tends to Γ as $t \to \infty$. Since $x(t)$ is bounded it is easily shown that Γ is nonempty, compact and connected.

Since $x(t)$ solves equation (6), $y(t) \to 0$ and $b \in L^1(0,\infty)$ it follows from Theorem 1 of [2] that Γ is the union of solutions of

$$z(t) = -\int_{-\infty}^{t} b(t-s)G(z(s))ds, \tag{7.1}$$

$$|z(t)| \leq \epsilon_1. \qquad (-\infty < t < \infty) \tag{7.2}$$

Let $Tz(t)$ be the function defined by the right hand side of line (7.1) when $z \in BC(-\infty,\infty)$ and $\|z\|_1 \leq \epsilon_1$. The estimates in the proof of Theorem 1 above imply that T is a contraction map. Thus $z(t) \equiv 0$ is the unique solution of (7.1-2). This means that $\Gamma = \{0\}$. Thus $x(t) \to 0$ and the proof of Theorem 2 is complete.

Using Corollary 1 and Theorem 2 we obtain the following result.

Corollary 2. *Let* (A1-5) *hold and let* ϵ_1 *and* ϵ_2 *be given by Corollary* 1 *above. If* $\|f\|_0 \leq \epsilon_2$ *and* $y(t) \to 0$ *as* $t \to \infty$, *then* $x(t) \to 0$.

III. Applications: Integrable Kernels.

The purpose of this section is to apply the theory in Section II with the additional assumption that $a \in L^1(0,\infty)$. We shall need the following result.

Theorem 3 (Paley and Wiener). *Let* $a \in L^1(0,\infty)$. *Then the solution* b *of equation* (5) *is* $L^1(0,\infty)$ *if and only if the determinant*

$$\det\left(I-\int_0^\infty \exp(-st)a(t)dt\right) \neq 0, \tag{8}$$

in the right half plane $\mathrm{Re}\, s \geq 0$.

This theorem is proved by a trivial modification of the proof of Paley and Wiener of Theorem XVIII in [3, p. 60]. Paley and Wiener use Theorem 3 to study the asymptotic behavior of solutions of equation (3') in case $f(t) \to 0$ as $t \to \infty$. Their result has the following nonlinear generalization.

Theorem 4. *Suppose* (A1-4) *hold,* (8) *is satisfied for* $\mathrm{Re}\, s \geq 0$ *and* ϵ_2 *is given by Corollary* 1 *above. If* $\|f\|_0 \leq \epsilon_2$ *and* $f(t) \to 0$ *as* $t \to \infty$, *then* $x(t) \to 0$.

Proof. The solution of the linearized equation (3') is given by (4). Since $f(t) \to 0$ as $t \to \infty$ and $b \in L^1(0,\infty)$, the Lebesgue Dominated Convergence Theorem implies that $y(t) \to 0$. An application of Corollary 2 completes the proof of Theorem 4.

Levin [4] has obtained another nonlinear generalization of the Paley-Wiener result. His result is neither stronger nor weaker than Theorem 4 above. Levin studies a scalar equation $(n=1)$ while we allow $n > 1$. Our hypothesis on $a(t)$ is weaker than Levin's and our hypothesis on $g(x)$ stronger. Theorem 3 is a local result while Levin's result is global.

The condition $f(t) \to 0$ is essential to the proof of Theorem 4 above. If f has a different type of asymptotic behavior, it may still be possible to analyze the local behavior of solutions of equation (2). For example in Theorem 5 below, $f(t)$ is constant but not necessarily zero.

IV. Applications: Integrodifferential Equations.

The purpose of this section is to apply the theory of Section II to the study of the local behavior of integrodifferential equations of the form

$$(9) \qquad x'(t) = mg(x(t)) + \int_0^t k(t-s)g(x(s))ds, \; x(0) = x_0, \; (t \geqq 0)$$

where k is locally integrable and m is a constant. We allow $m = 0$. This system can be written in the form of equation (2) if one sets $f(t) \equiv x_0$ and

$$a(t) = m + \int_0^t k(s)ds.$$

We wish to investigate the asymptotic behavior of solutions of equation (9) when x_0 is small. We remark that the definitions of stability and asymptotic stability of the trivial solution $x = 0$ of (9) are the same as for ordinary differential equations.

Theorem 5. *Let f and a be as defined above. If (A3-4) hold, $a \in L^1(0,\infty)$ and (8) is true for $\operatorname{Re} s \geq 0$, then for x_0 sufficiently small*

(i) *the trivial solution of (9) is stable and*

(ii) *each solution of (9) tends to a constant as* $t \to \infty$.

Proof. It follows from the proof of Corollary 1 above that for each ϵ, $0 < \epsilon < \epsilon_1$, there exists $\delta > 0$ such that $\|x\|_0 \leq \epsilon$ when $|x_0| \leq \delta$.

To prove part (ii) note that if $|x_0| \leq \epsilon_2$ then $|x(t)| \leq \epsilon_1$ for all $t \geq 0$. Moreover

$$x(t) = (I-\int_0^t b(s)ds)x_0 - \int_0^t b(t-s)G(x(s))ds.$$

Since $b \in L^1(0,\infty)$,

$$\lim_{t\to\infty} (I-\int_0^t b(s)ds)x_0 = I - \int_0^\infty b(s)ds,$$

exists. By Theorem 1 of [2] the positive limit set of $x(t)$ is the union of solutions of

$$z(t) = (I-\int_0^\infty b(s)ds)x_0 - \int_{-\infty}^t b(t-s)G(z(s))ds, \tag{10.1}$$

$$|z(t)| \leq \epsilon_1. \qquad (-\infty < t < \infty) \tag{10.2}$$

Let $S(0,\epsilon_1)$ be the closed sphere in $BC(R)$ with center at the origin and radius ϵ_1. Let S_0 be the subset of $S(0,\epsilon_1)$

consisting of constant functions. The estimates on ϵ_1 in the proof of Theorem 1 imply that the right side of (10.1) defines a contraction map on $S(0,\epsilon_1)$ and on S_0. Therefore the unique solution of (10.1-2) is a constant function $z(t) \equiv z_0$. Thus the positive limit set of $x(t)$ is the single point z_0, $x(t) \to z_0$ as $t \to \infty$, and Theorem 5 is proved.

For x_0 small, the limit z_0 is obtained by solving the equation

$$z_0 = (I-\int_0^\infty b(s)ds)x_0 - (\int_0^\infty b(s)ds)G(z_0).$$

Let the solution be $z_0 = F(x_0)$. Then $F(0) = 0$ and F maps a neighborhood of $x_0 = 0$ diffeomorphically onto a neighborhood of $z_0 = 0$. This means that the trivial solution cannot be asymptotically stable.

V. Applications: A Reactor Problem.

The dynamic behavior of a continuous medium nuclear reactor can be described, under certain simplifying assumptions, by the following integrodifferential equations for the unknown u and T:

(11.1) $$u'(t) = -\int_{-\infty}^{\infty} \alpha(x)T(x,t)dx,$$

(11.2) $$T_t = T_{xx} + \eta(x)g(u(t)), \quad (-\infty < x < \infty,\ 0 < t < \infty)$$

with the initial conditions

(12) $$u(0) = u_0, \quad T(x,0) = f(x). \quad (-\infty < x < \infty)$$

These equations have been extensively studied by Levin and Nohel, in the linear case $g(u) = u$ c.f. [1,5] and in the nonlinear case cf. [6]. In the reactor problem $g(u) = \exp(u) - 1$.

We wish to study the asymptotic behavior of solutions of (11) using the theory of Section II. Our analysis depends heavily on the papers [1,5, and 6] both for motivation and techniques. Since Levin and Nohel have treated the uniqueness problem for (11-12) we do not consider it further.

Let $*$ denote the L^2 Fourier transform. If f, α, and η are $L^2(R)$, then an elementary application of transform theory shows that $u(t)$ satisfies the equation

$$(13) \qquad u'(t) = -\int_0^t m_1(t-s)g(u(s))ds - m_2(t), u(0) = u_0$$

where for $j = 1, 2$.

$$m_j(t) = (1/\pi) \int_0^\infty \exp(-x^2 t)h_j(x)dx,$$

and

$$h_1(x) = \mathrm{Re}\ \eta^*(x)\alpha^*(-x),\ h_2(x) = \mathrm{Re}\ f^*(x)\alpha^*(-x).$$

Using a Taubian theorem Levin and Nohel [1] study the linear equation

$$(14) \qquad v'(t) = -\int_0^t m_1(t-s)v(s)ds - m_2(t), v(0) = v_0.$$

They prove

<u>Theorem 6</u> (Levin and Nohel). <u>Suppose</u> f, α <u>and</u> η <u>satisfy</u>:

(A6) $\quad f(x), \eta(x), \alpha(x) \in O(\exp(-\lambda|x|)), \lambda > 0, |x| \to \infty$

(A7) $\sup_{-\infty<x<\infty} \{|\alpha(x)|, \eta(x)|, |f(x)|\} < \infty$.

(A8) $h_1(x) \geq 0$ and $h_1(0) \neq 0$.

Then the solution $v(t)$ *of* (14) *exists for all* $t \geq 0$ *and* $v(t) = O(t^{-3/2})$ *as* $t \to \infty$.

Corollary 3. *If the hypotheses of Theorem* 6 *are satisfied then there exists a positive constant* K_1 *(independent of* v_o *and* f) *such that for all* $t \geq 0$

$$|v(t)| \leq K_1(|v_o|+\|f\|), \|f\| = \int_{-\infty}^{\infty} |f(x)|dx.$$

Proof. Let $v_1(t)$ be the solution of (14) when $v_o = 1$ and $m_2(t) \equiv 0$ and let $v_2(t)$ be the solution when $v_o = 0$. Then the general solution is $v_1(t)v_o + v_2(t)$. By Theorem 6 $v_1(t)$ is bounded.

Let V be the Laplace transform of v_2. Using lines 5.28 and 5.32 of [1] we see that for $-\infty < y < \infty$

$$V(iy) = H(y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-(iy)^{1/2}|x-s|)\alpha(x)f(s)dxds$$

where H(y) is in $L^1(-\infty,\infty)$ and H depends only on α and η. Lemmas 5.1-5.6 of [1] show that V satisfies the hypotheses of Theorem 2 of [8, p. 266]. Therefore

$$|v_2(t)| \leq (2\pi)^{-1}\int_{-\infty}^{\infty}|H(y)|dy \int_{-\infty}^{\infty}|\alpha(x)|dx \int_{-\infty}^{\infty}|f(x)|dx.$$

This proves Corollary 3.

Using Theorem 2 and 6 we prove

Theorem 7. Let f, α and η satisfy (A6-8). Let g satisfy (A3) with $g'(0) = 1$. Then there exist $\delta > 0$ (depending only on η, α and g) such that when $|u_o| \leq \delta$ and $\|f\| \leq \delta$, then the solution $u(t)$ of (13) exists for all $t \geq 0$ and $u(t) \to 0$ as $t \to \infty$.

Proof. Let $b(t)$ be the resolvent kernel for equation (14), that is $b(t)$ solves (14) in the special case $v_o = 0$ and $f = \eta$. By Theorem 6 we see that $v(t)$ and $b(t) = \mathcal{O}(t^{-3/2})$ as $t \to \infty$. Thus b is in $L^1(0,\infty)$. We know from Corollary 3 that $|v(t)|$ is small when $|u_o|$ and $\|f\|$ are small. An application of Theorem 2 completes the proof of Theorem 7.

Corollary 4. Let the hypotheses of Theorem 7 hold. If δ is given by Theorem 7 and $\|f\|, |u_o| \leq \delta$, then $u(t) \in L^1(0,\infty)$.

Proof. Fix u_o and f with $|u_o|$ and $\|f\| \leq \delta$. Let $v(t)$ be the solution of (14). There exists $K > 0$ such that for all $t \geq 0$

$$|b(t)| \leq K(t+1)^{-3/2}, v(t)| \leq K(t+1)^{-3/2}.$$

Since $u(t) \to 0$, there exists $T > 0$ such that if $t \geq T$ then

$$|G(u(t))| = |g(u(t))-u(t)| \leq |u(t)|/(4K).$$

Let K_1 be a bound on $|G(u(t))|$ for $0 \leq t < \infty$. For all $t \geq 0$,

$$u(t+T) = v(t+T) - \int_0^T b(t+T-s)g(u(s))ds$$

$$- \int_0^t b(t-s)G(u(T+s))ds,$$

$$|u(t+T)| \leq K(t+T+1)^{-3/2} + KK_1 \int_0^T (t+T+1-s)^{-3/2}ds)$$

$$+ \int_0^t K(t+1-s)^{-3/2}|u(T+s)|/(4K)ds,$$

$$\leq K(t+T+1)^{-3/2} + 2KK_1((t+1)^{-1/2} - (t+T+1)^{-1/2})$$

$$+ \int_0^t (t+1-s)^{-3/2}|u(T+s)|/4ds$$

$$\leq H_1(t) + H_2(t) + \int_0^t H_3(t-s)|u(s+T)|ds.$$

The comparison theorem of Nohel [7, Theorem 2.1] implies that for $t \geq 0$, $|u(t+T)| \leq U(t)$, where U solves

$$U(t) = H_1(t) + H_2(t) + \int_0^t H_3(t-s)U(s)ds. \tag{15}$$

Since for any $t > 0$,

$$\int_0^t H_2(s)ds = 4KK_1(\sqrt{t+1} - \sqrt{t+T+1} - 1 + \sqrt{T+1}) \leq 4KK_1(\sqrt{T+1} - 1),$$

it follows that $H_2 \in L^1(0,\infty)$. Clearly H_1 and $H_3 \in L^1(0,\infty)$ and $\int_0^\infty H_3(s)ds \leq 1/2$. Thus the right hand side of equation (15) determines a contraction map on $L^1(0,\infty)$. Since $U(t)$ dominates $|u(t+T)|$, $u(t) \in L^1(0,\infty)$. This completes the proof of Corollary 3.

In order to study the asymptotic behavior of $T(x,t)$ we need the following additional assumption:

(A9) $f,\eta \in C(R)$ and η is locally Holder continuous.

<u>Theorem 8. Suppose</u> g <u>satisfies</u> (A3) <u>and</u> $g'(0) = 1$. <u>Let</u> f,α <u>and</u> η <u>satisfy</u> (A6-9). <u>Then for</u> u_0 <u>and</u> $\|f\|$ <u>sufficiently small problem</u> (11-12) <u>has a unique solution</u> $u(t), T(x,t)$. <u>Moreover,</u>

$$\sup_{-\infty<x<\infty} |T(x,t)| \to 0, \qquad (t \to \infty)$$

<u>and</u> $u(t) \to 0$ <u>as</u> $t \to \infty$ <u>with</u> $u \in L^1(0,\infty)$.

<u>Proof</u>. For u_0 and $\|f\|$ sufficiently small Theorem 7 and Corollary 3 imply the existence of a solution $u(t)$ of equation (13) such that $u \in L^1(0,\infty)$ and $u(t) \to 0$. Given this $u(t)$ define $T(x,t)$ on $-\infty < x < \infty$, $0 < t < \infty$ by

$$(16) \qquad T(x,t) = \int_{-\infty}^{\infty} G(x-y,t)f(y)dy + \int_0^t \int_{-\infty}^{\infty} G(x-y,t-s)\eta(y)g(u(s))dyds,$$

where $G(x,t) = (4\pi t)^{-1/2}\exp(-x^2/(4t))$. Using the same proof as in [7, p.264] we verify that the pair $u(t), T(x,t)$ is a solution of (11) and (12). Moreover, for any $t > 0$

$$|T(x,t)| \le (4\pi t)^{-1/2}\int_{-\infty}^{\infty} |f(y)|dy + (4\pi)^{-1/2}\int_{-\infty}^{\infty} |\eta(y)|dy \int_0^t (t-s)^{-1/2}|g(u(s))|ds.$$

Since $g(u(t))$ is $L^1(0,\infty)$ it follows by dominated convergence that

$$\int_0^t s^{-1/2}|g(u(t-s))|\,ds = \int_0^t (t-s)^{-1/2}|g(u(s))|\,ds \to 0.$$

Therefore $T(x,t) \to 0$ as $t \to \infty$ uniformly for $-\infty < x < \infty$. This proves Theorem 8.

Theorem 8 is neither stronger nor weaker than the results in [6]. The advantage of Theorem 8 is that we avoid a hypothesis on the interconnection of f, α and η, c.f. [6, line 1.16]. The main disadvantage of Theorem 8 is that the result is local while the results of [6] are global.

References

[1] J. J. Levin and J. A. Nohel, On a system of integrodifferential equations occurring in reactor dynamics, Journal of Math. Mech, 9 (1960), pp. 347-368.

[2] R. K. Miller, Asymptotic behavior of solutions of nonlinear Volterra equations, Bull. Amer. Math. Soc., 72 (1966), pp. 153-156.

[3] R. E. A. C. Paley and N. Wiener, Fourier Transforms in the Complex Domain, Amer. Math. Soc. 1934.

[4] J. J. Levin, The qualitative behavior of a nonlinear Volterra equation, Proc. Amer. Math. Soc., 16 (1965), pp. 711-718.

[5] J. J. Levin and J. A. Nohel, On a system of integrodifferential equations occurring in reactor dynamics II, Archive Rat. Mech. Anal. 11 (1962), pp. 210-243.

[6] ______________________, A system of nonlinear integro-differential equations, Mich. Math. J. 13 (1966), pp. 257-270.

[7] J. A. Nohel, Some problems in nonlinear Volterra integral equations, Bull. Amer. Math. Soc. 68 (1962), pp. 323-329.

[8] G. Doetsch, Theorie und Anwendung der Laplace - Transformation, Berlin (1937).

PAPER [11]

To appear in the Bulletin of American Mathematical Society

Periodic Points of Diffeomorphisms

by

K. R. Meyer†

Center for Dynamical Systems
Brown University
Providence, Rhode Island

†This Research was supported in part by the National Aeronautics and and Space Administration under Grant No. NGR 40-002-015 and in part by the National Aeronautics and Space Administration, Huntsville, under Contract No. NAS8-11264.

Periodic Points of Diffeomorphisms

I. Introduction:

In [1] Artin and Mazur prove that there is a dense set in the space of C^k mappings of a compact manifold into itself such that for each member of this set the number of fixed points under iteration grows at most exponentially. This estimate allows one to define an analytic ζ-function associated to the diffeomorphism that measures the number of fixed points of the diffeomorphism under iteration.

The theorem of Artin and Mazur gives no indication as to whether or not a specific diffeomorphism satisfies such an estimate.

In this note we announce (Theorem 1) that the number of fixed points of the general class of diffeomorphisms recently introduced by Smale [2,3] grows at most exponentially under iteration. It should be noted that this new theorem is neither contained in nor contains the theorem of Artin and Mazur.

The method of proof is quite simple. One need only show that the size of the domain where there is a unique fixed point of the diffeomorphism decreases at most exponentially by using an estimate on the domain of validity of the implicit function theorem. The complexity arises only from the necessity of checking uniformity at each step.

II. Notation and Theorem.

Let M be a compact C^2-Riemannian manifold and suppose that $f: M \to M$ is a diffeomorphism of M. A closed invariant set

$\Lambda \subset M$ is said to have a <u>hyperbolic structure</u> if the tangent bundle $T_\Lambda M$ of M restricted to Λ has a continuous invariant splitting $T_\Lambda M = E^u + E^s$ under df such that

$$df\colon E^u \to E^u \quad ; \quad df\colon E^s \to E^s$$

$$\|df^n(x)(u)\| < C\lambda^n\|u\|$$

$$\|df^n(x)(v)\| > C^{-1}\lambda^{-n}\|v\|$$

for some fixed constants $C > 0$, $0 < \lambda < 1$, where $x \in \Lambda$, $v \in E^s_x$, $u \in E^u_x$ and $n \in Z^+$.

If f is a diffeomorphism of M and $x \in M$ then x is called a <u>wandering point</u> if there exists a neighborhood U of x such that $\bigcap_{n \in Z} f^n(U) = \varphi$. A point x of M is called a <u>non-wandering point</u> if it is not a wandering point. Clearly the set of nonwandering points of f forms a compact invariant subset of M.

The class of diffeomorphisms introduced by Smale in [2,3] is the class of diffeomorphisms of M with a hyperbolic structure on the set of nonwandering points of f. This class of diffeomorphisms is sufficiently general to include all known examples of diffeomorphisms with globial stability properties (see [3] for a detailed discussion).

Let $N_n(f)$ be the number of fixed points of f^n. Then our main result is

Theorem 1. <u>If</u> f <u>is a</u> C^2<u>-diffeomorphism of</u> M <u>into itself with a hyperbolic structure on the set of nonwandering points of</u> f <u>then there exists a constant</u> $k > 0$ <u>such that</u>

$$N_n(f) \leq k^n \quad \underline{\text{for}} \quad n \in Z^+.$$

III. Outline of the Proof:

In what follows $|\cdot|$ will denote the usual Euclidean norm in E^m with respect to a fixed basis. The following lemma follows easily from the implicit function theorem given in [4], page 12.

Lemma 1: Let φ_n, $n \in Z^+$, be a C^2 map from the closed ball B_a of radius a about the origin in E^m into itself with $\varphi_n(0) = 0$. Let the supremum of the modulus of the second partials of φ be less than b^n on B_a. Let $|(d\varphi_n(0)-I)| \leq c^n$ and $|(d\varphi_n(0)-I)^{-1}| \leq c^n$. Then there exists a constant $d = d(a,b,c)$ such that φ has a unique point in the sphere of radius d^n about the origin.

Let (V_i,y_i) and (U_i,x_i), $i = 1,\ldots r$ be a finite number of coordinate systems for M such that $V_i \supset \overline{U}_i$, $\bigcup_1^r U_i \supset M$ and $x_i = y_i|U_i$. Consider the sets $y_i(V_i)$ and $x_i(U_i)$ in E^m. There exists a constant $a > 0$ such that each point of $x_i(\overline{U}_i)$ is contained in a sphere of radius a completely contained in $y_i(V_i)$. We shall count the number of fixed points of f^n in each $x_i(\overline{U}_i)$.

Let $\|\cdot\|$ denote the norm induced in $y_i(V_i)$ by the metric on M.

Lemma 2: Let x_o be a fixed point of $f^n, x_i o f^n o x_i^{-1}$, and let A be the Jacobian matrix of f^n evaluated at x, then there exist constants N and $c > 0$ such that

$$|(A-I)| \leq c^n \quad \text{and} \quad |(A-I)^{-1}| \leq c^n$$

for $n \geqq N$.

Comments on the Proof of Lemma 2.

At this point the strong uniformity of the hyperbolic structure on the set of nonwandering points is used. Because the set of nonwandering points is compact, and the splitting is continuous there exists a constant $e \geqq 1$ such that the norm $\|\cdot\|$ and the norm $|||\cdot|||$ defined by the coordinates such that A has the form $A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$ where A_j; $j = 1,2$, is the representation of the mapping df on $E^u_{x_0}$ and $E^s_{x_0}$ respectively satisfies the condition

$$e^{-1}|||\cdot||| \leqq \|\cdot\| \leqq e|||\cdot|||.$$

With this uniformity at hand the rest of the lemma follows by standard matrix methods.

Since the total volume of $x_i(\overline{U}_i)$ is finite and fixed points of f^n in $x_i(\overline{U}_i)$ can be covered by disjoint balls of radius $\frac{d^n}{3}$ the required estimate follows from the above two lemmas.

It seems likely that the general outline given above can be used to give a similar estimate for the number of periodic orbits for a flow on M having a hyperbolic structure on the set of nonwandering points. The author is presently working on this problem.

References

[1] Artin, M., and B. Mazur, On Periodic Points, Ann. of Math. 81 (1965), pp. 82-99.

[2] Smale, S., Dynamical Systems on n-Dimensional Manifolds, Proceedings of the Internation Symposium on Differential Equations and Dynamical Systems, Academic Press, to appear.

[3] Smale, S., Differential Dynamical Systems, to appear.

[4] Hartman, P., Ordinary Differential Equations, John Wiley, New York (1964).

PAPER [12]

To appear in the Memoirs of the American
Mathematical Society and CDS-TR 66-5.

A CLASS OF FUNCTIONAL EQUATIONS OF NEUTRAL TYPE

by

Jack K. Hale* and Kenneth R. Meyer**

*This research was supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFSOR Grant No. AF-AFOSR-693-66, by the United States Army Research Office, Durham, under Contract No. DA-31-124-ARO-D-270 and in part by National Aeronautics and Space Administration under Grant No. NGR-40-002-015.

**This research was supported by National Aeronautics and Space Administration under Contract No. NAS8-11264.

I. INTRODUCTION

In the study of weakly nonlinear systems, the most useful elements from the theory of linear non-homogeneous ordinary differential equations with autonomous homogeneous part are 1) the variation of constants formula, 2) the decomposition of Euclidean space into the direct sum of subspaces which are invariant with respect to the solutions of the homogeneous system (the Jordan canonical form) and 3) sharp exponential bounds on the growth of solutions on these invariant subspaces. Once these facts are well understood, many problems in the theory of stability, asymptotic behavior and nonlinear oscillations can be discussed. For delay differential equations of retarded type these three concepts have been developed and applied to problems of the above type (see,for example, [1],[2],[3,4],[5,6],[7]).

For delay differential equations of neutral type, the theory is not so well developed even though some results are contained in the book of Bellman and Cooke [1]. In equations of neutral type, the first difficulty arises because the derivative of a solution occurs with a retardation. This leaves much freedom in the choice of the topology on the solution space as well as on the space of initial conditions. The topology must be chosen in such a way as to obtain solutions which are at least continuous with respect to the initial data. That such a choice is not obvious may be easily seen by consulting the papers of Driver [8,9] where a general existence and uniqueness theorm is given for a rather broad class of neutral equations.

Our approach in this paper is to investigate a class of functional integral equations in the space of continuous functions. This class includes certain types of equations of neutral type and does include some equations which arise in the applications. For this class of equations, we obtain precise analogues of the above stated properties of ordinary differential equations. Furthermore, the decomposition of our space into invariant subspaces is given in a way that is amenable to computations. As specific applications of the theory, we give a stability theorem and extend the method of averaging to these systems.

The symbol [ ] indicates references in the bibliography, Roman numerals refer to sections and Arabic to formulas.

Let R^n be a real or complex n-dimensional linear space of column vectors with norm $|\cdot|$ and let $C([a,b],R^n)$ denote the Banach space of continuous functions from $[a,b]$ into R^n with norm $\|\cdot\|_{[a,b]}$ given by $\|\varphi\|_{[a,b]} = \sup\{|\varphi(\theta)| : \theta \in [a,b]\}$. Let r be a fixed non-negative number and let $C = C([-r,0],R^n)$ and $\|\cdot\| = \|\cdot\|_{[-r,0]}$.

Let $\mathcal{L}_p([a,b],R^n)$, $1 \leq p < \infty$, be the set of Lebesque integral functions from $[a,b]$ into R^n with the norm of any φ in $\mathcal{L}_p([a,b],R^n)$ defined by $[\int_a^b|\varphi(s)|^p ds]^{1/p}$. Also let $\mathcal{L}_\infty([a,b],R^n)$ denote the set of essentially bounded measurable functions from $[a,b]$ into R^n, with the norm of any φ in $\mathcal{L}_\infty([a,b],R^n)$ given by ess.sup $|\varphi(\theta)|$. We shall also use the space $\mathcal{L}_\infty([a,b],R^{n^2})$ of essentially bounded measurable functions into the space of $n \times n$ matrices with the norm defined in the obvious way.

Suppose τ is a given real number. We allow $\tau = -\infty$ and in this case the interval $[\tau,\infty)$ denotes the interval $(-\infty,\infty)$. Let g and f be continuous functions from $[\tau,\infty) \times C$ into R^n such that for each $t \in [\tau,\infty)$ the functions $f(t,\cdot)$ and $g(t,\cdot)$ are linear operators and there exist positive continuous functions K and L defined for all $t \geq \tau$ such that

(1) $$|g(t,\varphi)| \leq K(t)\|\varphi\| \quad \text{and} \quad |f(t,\varphi)| \leq L(t)\|\varphi\|$$

for all $\varphi \in C$ and $t \in [\tau,\infty)$.

By the Riesz representation theorem there exists $n \times n$ matrix valued functions μ and η defined on $[\tau,\infty) \times [-r,0]$ such that

$$(2)\qquad \begin{aligned} g(t,\varphi) &= \int_{-r}^{0} [d_\theta \mu(t,\theta)]\varphi(\theta) \\ f(t,\varphi) &= \int_{-r}^{0} [d_\theta \eta(t,\theta)]\varphi(\theta) \end{aligned}$$

for all $\varphi \in C$. Moreover for each fixed t the functions $\mu(t,\cdot)$ and $\eta(t,\cdot)$ are of bounded variation in $[-r,0]$.

For any $x \in C([-r,A),R^n)$, $A > 0$, define x_t, $0 \leq t < A$, as the element of C given by $x_t(\theta) = x(t + \theta)$; that is, x_t is the restriction of x to the interval $[t-r,t]$ shifted to $[-r,0]$.

For any $\varphi \in C$ and any σ in $[\tau,\infty)$ define $\gamma(\sigma,\varphi) = \varphi(0)-g(\sigma,\varphi)$. For any h, $h \in \mathcal{L}_1([\sigma,\nu),R^n)$ for every ν in $[\sigma,\infty)$, consider the following functional integral equation

$$(3)\qquad \begin{aligned} &\text{a)}\quad x_\sigma = \varphi \\ &\text{b)}\quad x(t) = \gamma(\sigma,\varphi)+g(t,x_t)+\int_\sigma^t f(s,x_s)ds+\int_\sigma^t h(s)ds\ ,\ t \in [\sigma,\infty). \end{aligned}$$

By a solution of (3) we shall mean an element of $C([\sigma-r,A),R^n)$, $\sigma < A \leq \infty$, that satisfies the relations in (3). We shall refer to φ as the initial function and to σ as the initial time.

If f and g are independent of t then (3) will be called autonomous and otherwise non-autonomous. If $h = 0$ the equation (3) will be called homogeneous and otherwise non-homogeneous.

If $g \equiv 0$ then (3) is equivalent to the functional differential equation of retarded type

$$\dot{x}(t) = f(t,x_t) + h(t)$$

with initial function at $t = \sigma$ given by φ.

If $f \equiv 0$ and $h \equiv 0$ then equation (3) is a functional difference equation of retarded type, and in particular, includes difference equations. For both f and g not identically zero, equation (3) corresponds to a functional differential equation of neutral type. Indeed, formal differentiation of the equation yields

$$\dot{x}(t) = g(t,\dot{x}_t) + \hat{f}(t,x_t) + h(t) \,, \tag{4}$$

where $\hat{f} = \partial g/\partial t + f$ and $\dot{x}_t$ is defined by $\dot{x}_t(\theta) = \dot{x}(t+\theta)$, $-r \leqq \theta \leqq 0$. Also, if one begins with (4) and defines a solution with initial function φ at σ to be a continuous function satisfying (4) almost everywhere, then an integration yields (3) with $\gamma(\sigma,\varphi) = \varphi(0)-g(\sigma,\varphi)$.

Notice that all differential difference equations of neutral type with variable coefficients and constant retardations can be written in the form (3) provided the coefficients of the terms involving the derivatives have an integrable first derivative.

Also, equation (3) contains as a special case some differential difference equations of neutral type with variable lags provided that the lags are bounded and satisfy some other reasonable conditions. For example, the equation $\dot{x}(t) = \dot{x}(\beta(t)) + \dot{x}(\gamma(t))$ can be written in the form (3) if β, $\dot{\beta} > 0$, γ are continuous, $\ddot{\beta}$ is integrable and there is a constant $r \geqq 0$ such that $t-r \leqq \beta(t) \leqq t$, $t-r \leqq \gamma(t) \leqq t$.

These last remarks are precisely the reason for considering equation (3). It one attempts to discuss the equation (4) directly, then the first problem encountered are precise definitions of a solution and the topology to be used on the space in which the solutions lie. To discuss (4) the topology must include information about the derivatives of functions whereas (3) can be discussed in the simpler space C.

Equation (3) would also include equations of advanced type unless some further restriction is made on the function g. This is due to the fact that the measure $\mu(t,\theta)$ in (2) may have a jump at $\theta = 0$ equal to the identity for some values of t. To avoid this difficulty, we shall assume that the measure μ is uniformly nonatomic at zero. More precisely, we assume that there exists a nonnegative, continuous, nondecreasing function δ defined on $[0,\varepsilon_0]$ for some $0 < \varepsilon_0 \leqq r$ such that

$$(5) \qquad \delta(0) = 0 \quad \text{and} \quad \left| \int_{-s}^{0} [d\mu(t,\theta)]\varphi(\theta) \right| \leq \delta(s)\|\varphi\|_{[-s,0]}$$

for all $\varphi \in C$, $t \in [\tau,\infty)$ and all $s \in [0,\varepsilon_0]$. In some cases it will be necessary to further restrict μ.

Observe that the solution $x(t,\sigma,\varphi)$ of (3) with initial function φ at σ satisfies

$$x(t,\sigma,\varphi) = x(t,s,x_s(\cdot,\sigma,\varphi)) \qquad t \geqq s \geqq \sigma \tag{6}$$

provided all the above solutions exist and are uniquely defined by initial values.

Also, at times it will be necessary to consider solutions of (3) that are matrix valued. In this case we define the action of f and g by (2) when φ is a continuous $n \times n$ matrix valued function of the scalar θ, $\theta \in [-r,0]$.

II. THE GENERAL LINEAR EQUATION.

This section deals with the general non-autonomous equation I(3). Existence and uniqueness of solutions and variation of constants formula are discussed.

THEOREM 1. For any given $\varphi \in C$, $\sigma \in [\tau,\infty)$ and h, where $h \in \mathcal{L}_1([\sigma,\nu),R^n)$ for every ν in $[\sigma,\infty)$, there exists a unique function $x(\sigma,\varphi)$ defined and continuous on $[\sigma-r,\infty)$ that satisfies I(3).

PROOF. Suppose $K(t)$, $L(t)$ are defined by I(1) and $\delta(s)$, s in $[0,\varepsilon_o]$ is defined by I(5). Let $\beta > \sigma$ be any fixed positive number and let K_β and L_β be the supremum on $[\sigma,\beta]$ of $K(t)$ and $L(t)$, respectively. Choose $A > 0$ so that $\delta(A)+L_\beta A < 1$ and $\sigma+A < \beta$, $A < \varepsilon_o$. Let $\Gamma = \{y \in C([\sigma-r,\sigma+A],R^n) : y_\sigma = \varphi\}$, and for any y in Γ, define

$$(Iy)(t) = \begin{cases} \varphi(t-\sigma) & \text{for } \sigma-r \leqq t \leqq \sigma \\ \gamma(\sigma,\varphi)+g(t,y_t)+\int_\sigma^t f(s,y_s)ds+\int_\sigma^t h(s)ds, & \sigma < t \leqq \sigma+A \end{cases}$$

Clearly $I\Gamma \subset \Gamma$. For any y and z in Γ

$$|Iy(t)-Iz(t)| \leqq |g(t,y_t-z_t)| + \int_\sigma^t |f(s,y_s-z_s)|ds$$

$$\leqq [\delta(A) + L_\beta A]\|y-z\|_{[\sigma-r,\sigma+A]}$$

and so I is contracting in Γ. Thus, I has a unique fixed point in Γ, which implies I(3) has a unique continuous solution defined on $[\sigma-r,\sigma+A]$. But A is a constant independent of the norm of φ and the solution can be extended to $[\sigma-r,\beta]$ by use of the above and relation I(6). Since β was arbitrary the theorem is proved.

If the operators f and g do not increase too fast with t we would expect that the solutions of I(3) are exponentially bounded. Indeed one has

LEMMA 1. *Suppose* $|g(t,\varphi)| \leq K\|\varphi\|$ *and* $|f(t,\varphi)| \leq L\|\varphi\|$ *for all* $\varphi \in C$ *and all* $t \in [\tau,\infty)$ *where* K *and* L *are constants. Then there exist constants* a, b *and* c *such that for any* σ *in* $[\tau,\infty)$

$$\|x_t(\sigma,\varphi)\| \leq \{a\|\varphi\| + b\int_\sigma^t |h(s)|ds\}e^{c(t-\sigma)}, \quad t \geq \sigma.$$

PROOF. In this proof, we let x_t designate $x_t(\sigma,\varphi)$. Let M be such that $K+M > 1$, $|\gamma(t,\varphi)| \leq M\|\varphi\|$ for all $t \in [\tau,\infty)$, $\varphi \in C$, and let A be a positive constant such that $1-\delta(A) > 0$. Define $b = (1-\delta(A))^{-1}$ and $a = (K+M)(1-\delta(A))^{-1}$. For any $t \in [\sigma,\sigma+A]$ one has $|g(t,x_t)| \leq K\|\varphi\| + \delta(A)\|x_t\|$ and so

$$|x(t)| \leq (M+K)\|\varphi\| + \delta(A)\|x_t\| + L\int_\sigma^t \|x_s\|ds + \int_\sigma^t |h(s)ds, \quad t \geq \sigma.$$

Since $K+M > 1$ and $x_\sigma = \varphi$, the right-hand side is an upper bound for

$\|x_t\|$. Solving the resulting inequality for $\|x_t\|$ and applying Gronwall's inequality, we obtain

$$\|x_t\| \leq \{a\|\varphi\| + b \int_\sigma^t |h(s)|\, ds\} e^{bL(t-\sigma)} \quad \text{for} \quad t \in [\sigma, \sigma+A].$$

We shall now show by an induction argument that the above inequality is valid for all $t \geq \sigma$ provided bL is replaced by a larger constant. Let c be so large that $ae^{(bL-c)A} \leq 1$ and $c > bL$. Assume that

$$\|x_t\| \leq \{a\|\varphi\| + b \int_\sigma^t |h(s)|\, ds\} e^{c(t-\sigma)} \quad \text{for} \quad t \in [\sigma, \sigma+kA].$$

From the above, this assumption is true if $k = 1$. If $t \in [\sigma+kA, \sigma+(k+1)A]$, then the above estimate yields

$$\|x_t\| \leq \{a\|x_{t-A}\| + b \int_{t-A}^t |h(s)|\, ds\} e^{bLA}$$

and by the induction hypothesis

$$\|x_t\| \leq \{a[a\|\varphi\| + b \int_\sigma^{t-A} |h(s)|\, ds] e^{c(t-\sigma-A)} + \int_{t-A}^t |h(s)|\, ds\} e^{bLA}$$

$$\leq \{a\|\varphi\| + b \int_\sigma^t |h(s)|\, ds\} e^{c(t-\sigma)} .$$

This completes the proof of the lemma.

COROLLARY 1. *Let* $x(\cdot,\sigma,\varphi,h)$ *be the unique solution of* I(3) *with initial function* φ *at* σ *and forcing function* $h \in \mathcal{L}_1([\sigma,t_1),R^n)$. *For fixed* t_1 *and* σ, $x(t_1,\sigma,\cdot,\cdot)$ *is a continuous function from* $C \times \mathcal{L}_1([\sigma,t_1),R^n)$ *into* R^n.

PROOF. The corollary is obvious from lemma 1 if f and g admit a constant bound as required by the lemma. Since changing f and g for $t \geq t_1$ does not effect the value of the solution in $[\sigma,t_1]$ one can define new f' and g' to be identical to f and g for $\sigma \leq t \leq t_1$ and to equal $f(t_1,\cdot)$ and $g(t_1,\cdot)$ for $t \geq t_1$. Applying the above theorem to equation I(3) with f and g replaced by f' and g' yields the result.

The next problem is to obtain a variation of constants formula for the solutions of I(3). This is accomplished by observing that the solutions of I(3) are linear operators on the forcing function h. In particular we have:

THEOREM 2. (*Variation of Constants Formula*). *If* $x(\sigma,\varphi,h)$ *is the solution of* I(3) *with forcing function* h, *where* $h \in \mathcal{L}_1([\sigma,\nu),R^n)$, *for all* $\nu \geq \sigma$, *and initial value* φ *in* C *at* σ, *then*

$$(1) \qquad x(\sigma,\varphi,h)(t) = x(\sigma,\varphi,0)(t) + \int_\sigma^t U(t,s)h(s)ds, \qquad t \geq \sigma,$$

where $U(t,s)$ *is defined for* $\tau \leq s \leq t+r$, $U(t,\cdot) \in \mathcal{L}_\infty([\sigma,t],R^{n^2})$ *for each* t, $U(t,s) = + \partial W(t,s)/\partial s$ *a.e., where* $W(t,s)$ *is the unique*

<u>solution of</u>

(2a) $$W_s(\cdot,s) = 0$$

(2b) $$W(t,s) = \int_{-r}^{0} \{d_\theta \mu(t,\theta)\} W(t+\theta,s) + \int_s^t \int_{-r}^{0} \{d_\theta \mu(\zeta,\theta)\} W(\theta+\zeta,s)\,d\zeta - (t-s)I \quad \text{for} \quad \tau \leq s \leq t.$$

PROOF. Let $h \in \mathcal{L}_1([\sigma,t],R^n)$ and let $u(\cdot,\sigma,h)$ be the solution of I(3) that satisfies $u_\sigma = 0$. For fixed t and σ it follows from Corollary 1 that $u(t,\sigma,\cdot)$ is a continuous linear operator from $\mathcal{L}_1([\sigma,t],R^n)$ into R^n. So there exists (see [10]) an $n \times n$ matrix valued function $U^*(t,\sigma,\cdot) \in \mathcal{L}_\infty([\sigma,t],R^{n^2})$, $t \geq \sigma$, such that

$$u(t,\sigma,h) = \int_\sigma^t U^*(t,\sigma,\theta)h(\theta)d\theta .$$

Let α be in $[\sigma,t]$ and let k be any element of $\mathcal{L}_1([\sigma,t],R^n)$ that satisfies $k(\theta) = 0$ for $\theta \in [\sigma,\alpha]$. Then $u(t,\sigma,k) = u(t,\alpha,k)$, $t \geq \alpha$, and $U^*(t,\sigma,\theta) = U^*(t,\alpha,\theta)$ a.e. Since α is an arbitrary element of $[\sigma,t]$, it follows that U^* is independent of σ. Define $U(t,\theta) = U^*(t,\sigma,\theta)$, $t \in [\tau,\infty)$, $\theta \in [\tau,t]$, $U(t,\theta) = 0$ for $t \leq \theta \leq t+r$. For any s in $[\tau,\infty)$, let $W(t,s) = -\int_s^t U(t,\theta)d\theta$ for $t \geq s$ and $W(t,s) = 0$ for $t \in [s-r,s]$. Clearly W satisfies (2a), (2b) and U is given as stated in the theorem.

CORollary 2. *If* f *and* g *are independent of* t *then*

$$x(\sigma,\varphi,h)(t) = x(\sigma,\varphi,0)(t) + \int_{\sigma}^{t} U(t-s)h(s)ds \tag{3}$$

where U *is defined on* $[-r,\infty)$, $U \in \mathcal{L}_{\infty}([-r,t),R^{n^2})$, *for each* t *in* $[-r,\infty)$, $U(t) = -dW(t)/dt$ *a.e.* *and* W *satisfies*

$$W_{o} = 0 \tag{4a}$$

$$W(t) = g(W_t) + \int_{o}^{t} f(W_s)ds + tI,\ t \in [0,\infty). \tag{4b}$$

III. THE AUTONOMOUS, HOMOGENEOUS EQUATION.

In this section we study equation I(3) when f and g are independent of t and $h \equiv 0$. Since, for the autonomous case it is no restriction to choose the initial time $\sigma = 0$, we consider

$$(1) \qquad \begin{aligned} &a) \quad x_0 = \varphi \\ &b) \quad x(t) = \gamma(\varphi) + g(x_t) + \int_0^t f(x_s)ds \quad \text{for} \quad t \geq 0 \end{aligned}$$

with $\varphi \in C$, $\gamma(\varphi) = \varphi(0) - g(\varphi)$ and

$$(2) \qquad g(\varphi) = \int_{-r}^{0} \{d\mu(\theta)\}\varphi(\theta), \quad f(\varphi) = \int_{-r}^{0} \{d\eta(\theta)\}\varphi(\theta),$$

where μ and η are functions of bounded variation in $[-r,0]$.

The aim of this section is to study the behavior of the solutions in C. By some general results from functional analysis we are able to introduce coordinates in C in such a way that the behavior of the solution of 1) on certain finite dimensional subspaces are determined by ordinary differential equations. An explicit characterization of these subspaces is given that is amenable to computations.

If φ is any given function in C and $x(\varphi)$ is the unique solution of (1) with initial function φ at zero then we define a mapping $T(t): C \to C$, for each fixed t, by the relation

(3) $$T(t)\varphi = x_t(\varphi) .$$

The following lemma is an immediate consequence of the discussion in section II.

LEMMA 1. The family $\{T(t)\}_{t \in [0,\infty]}$ forms a strongly continuous, semi-group of bounded linear operators from C into itself for all $t \geq 0$.

Since $T(t)$ is strongly continuous we may define the infinitesimal operator A of $T(t)$ (see Hille and Phillips [11], p.306) as

(4) $$A\varphi = \lim_{t \to 0} \frac{1}{t} [T(t)\varphi - \varphi]$$

whenever this limit exists in the norm topology of C. The infinitesimal generator of $T(t)$ is the smallest closed extension of A. By the strong continuity of $T(t)$ on $[0,\infty)$ it follows that the infinitesimal generator and infinitesimal operator are the same (see corollary, p. 344 and Theorem 10.61, p. 322 of Hille and Phillips [11]). From the above remarks and Theorem 10.3.1 of Hille and Phillips, page 307, the domain $\mathcal{D}(A)$ of A, is dense in C and the range $\mathcal{R}(A)$ of A is C. These remarks allow us to compute A directly from (4). In fact, we have

LEMMA 2. The infinitesimal generator A of the semi-group $\{T(t)\}_{t\in[0,\infty)}$

and its domain $\mathcal{D}(A)$ *are given by*

$$(5)\qquad \text{a)}\quad A\varphi(\theta) = \frac{d}{d\theta}\varphi(\theta) = \dot{\varphi}(\theta)$$

$$\text{b)}\quad \mathcal{D}(A) = \{\varphi \in C : \dot{\varphi} \in C,\ \dot{\varphi}(0) = g(\dot{\varphi}) + f(\varphi)\}\ .$$

Moreover, $\mathcal{D}(A)$ *is dense in* C *and, for* $\varphi \in \mathcal{D}(A)$,

$$(6)\qquad \frac{d}{dt}T(t)\varphi = T(t)A\varphi = AT(t)\varphi\ .$$

PROOF. Suppose φ is in $\mathcal{D}(A)$. Since $T(t)\varphi(\theta) = \varphi(t+\theta)$ when $-r \leq t+\theta \leq 0$, it follows directly from the definition (4) that $(A\varphi)(\theta) = \dot{\varphi}(\theta^+)$ for $\theta \in [-r,0)$, where $\dot{\varphi}(\theta^+)$ is the right-hand derivative of φ at θ.

Since $\lim_{t\to 0^+}[T(t)\varphi-\varphi]/t$ exists for φ in $\mathcal{D}(A)$, there are constants α and β such that $\|T(t)\varphi-\varphi\| \leq \beta t$ for $t \in [0,\alpha)$. Thus $|x(t+\theta)-\varphi(\theta)| \leq \beta t$ for $t \in [0,\alpha)$ and $\theta \in [-r,0]$. This implies

$$\frac{1}{t}\int_{-r}^{0} d\mu(\theta)\{x(t+\theta)-\varphi(\theta)\} = \int_{-r}^{-t} d\mu(\theta)\left\{\frac{\varphi(t+\theta)-\varphi(\theta)}{t}\right\} + \int_{-t}^{0} d\mu(\theta)\left\{\frac{x(t+\theta)-\varphi(\theta)}{t}\right\}$$

tends to $\int_{-r}^{0} d\mu(\theta)\dot{\varphi}(\theta^+)$ as $t \to 0^+$ since

$$\left|\int_{-t}^{0} d\mu(\theta)\left\{\frac{x(t+\theta)-\varphi(\theta)}{t}\right\}\right| \leq \delta(t)\beta \to 0, \text{ as } t \to 0^+\ .$$

From 1(b), it follows immediately that

$$\mathcal{A}\varphi(0) = g(\dot{\varphi}^{+}) + f(\varphi) .$$

Since $\mathcal{A}\varphi$ must be in C it follows that $d\varphi(\theta)/d\theta$ exists and is continuous. The rest of the lemma follows by Theorem 10.3.3 of Hille and Phillips [11], page 308.

We shall now proceed to analyze the spectrum of $\mathcal{A}$. Let B be any linear operator of a Banach space $\mathcal{B}$ into itself. The <u>resolvent set</u> $\rho(B)$ is defined as the set of λ in the complex plane for which $(\lambda I-B)$ has a bounded inverse in all of $\mathcal{B}$. The complement of $\rho(B)$ in the λ-plane is called the <u>spectrum of B</u> and is denoted by $\sigma(B)$. The <u>point spectrum</u>, $P\sigma(B)$, consists of those λ in $\sigma(B)$ for which $(\lambda I-B)$ does not have an inverse. The points of $P\sigma(B)$ are called <u>eigenvalues</u> of B and the nonzero $\varphi \in \mathcal{B}$ such that $(\lambda I-B)\varphi = 0$ are called <u>eigenvectors</u> of $\mathcal{A}$. The <u>null space</u> $\mathfrak{N}(B)$ of B is the set of all $\varphi \in \mathcal{B}$ for which $B\varphi = 0$. For any given $\lambda \in \sigma(B)$ the <u>generalized eigenspace of λ</u> is defined to be the smallest closed subspace of $\mathcal{B}$ containing the subspaces $\mathfrak{N}(\lambda I-B)^k$, $k = 1,2,\dots,$ and will be denoted by $\mathfrak{M}_\lambda(B)$.

One of our objects is to determine the nature of $\sigma(A)$ and $\sigma(T(t))$. We would hope to discuss most of the properties of $T(t)$ by using only properties of the known operator A.

<u>THEOREM 1. Let A be defined as in Lemma 2, then $\sigma(A) = P\sigma(A)$ and $\lambda \in \sigma(A)$ if and only if λ satisfies the characteristic equation</u>

$$(7) \qquad \det \Delta(\lambda) = 0, \quad \Delta(\lambda) = \lambda I - \int_{-r}^{0} \lambda e^{\lambda\theta} d\mu(\theta) - \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta) .$$

The roots of (7) have real parts bounded above and for any $\lambda \in \sigma(A)$, the generalized eigenspace $\mathfrak{M}_\lambda(A)$ is finite dimensional. Finally if λ is a root of (7) of multiplicity k, then $\mathfrak{M}_\lambda(A) = \mathfrak{N}(\lambda I - A)^k$ and $C = \mathfrak{N}(A - \lambda I)^k \oplus \mathscr{R}(A - \lambda I)^k$, where $\oplus$ is the direct sum. Moreover $T(t)$ is completely reduced by the two linear manifolds $\mathfrak{M}_\lambda(A)$ and $\mathscr{R}_\lambda(A)$; that is, $T(t)\mathfrak{M}_\lambda(A) \subset \mathfrak{M}_\lambda(A)$, $T(t)\mathscr{R}_\lambda(A) \subset \mathscr{R}_\lambda(A)$ for all $t \geq 0$.

PROOF. To prove that $\sigma(A) = P\sigma(A)$, we show that the resolvent set $\rho(A)$ consists of all λ except those that satisfy (7) and then show that any λ satisfying (7) is in $P\sigma(A)$. The constant λ will be in $\rho(A)$ if and only if the equation

$$(8) \qquad (A - \lambda I)\varphi = \psi$$

has a solution φ in $\mathscr{D}(A)$ for all ψ in C and the solution depends continuously on ψ. Thus, we must have $\dot{\varphi}(\theta) - \lambda\varphi(\theta) = \psi(\theta)$, $\theta \in [-r,0]$; that is,

$$(9) \qquad \varphi(\theta) = e^{\lambda\theta} b + \int_{0}^{\theta} e^{\lambda(\theta-\xi)} \psi(\xi) d\xi \quad , \quad \theta \in [-r,0].$$

But, φ will be in $\mathscr{D}(A)$ if and only if $\dot{\varphi}(0) = g(\dot{\varphi}) + f(\varphi)$ and this yields

$$(10) \qquad \Delta(\lambda)b = \{-\psi(0) + \int_{-r}^{0} d\mu(\theta)[\frac{d}{d\theta}\int_{0}^{\theta} e^{\lambda(\theta-\xi)}\psi(\xi)d\xi] + $$
$$+ \int_{-r}^{0} d\eta(\theta) \int_{0}^{\theta} e^{\lambda(\theta-\xi)}\psi(\xi)d\xi\} .$$

Thus, if $\det \Delta(\lambda) \neq 0$, (9) and (10) show that (8) has a solution for any ψ in C and the solution is a continuous linear operator on C. This operator, called the resolvent operator, will be denoted by $(A-\lambda I)^{-1}$ and is given by

$$(11) \qquad [(A-\lambda I)^{-1}\psi](\theta) = e^{\lambda\theta}b + \int_{0}^{\theta} e^{\lambda(\theta-\xi)}\psi(\xi)d\xi \qquad \theta \in [-r,0]$$

where b is given by (10) and $\det \Delta(\lambda) \neq 0$. Hence $\rho(A) \supset \{\lambda : \det \Delta(\lambda) \neq 0\}$.

If $\det \Delta(\lambda) = 0$, then (9) and (10) imply there exists a nonzero solution of (8) for $\psi = 0$; that is, λ is in $P\sigma(A)$. This proves the first part of the theorem.

As we have seen, if λ is such that $\det \Delta(\lambda) = 0$ and b is such that $\Delta(\lambda)b = 0$, then $be^{\lambda\theta}$ is an eigenvector of A and every eigenvector is of this form. But then $x(t) = e^{\lambda t}b$ is a solution of (1) and hence by Lemma II(1) the real parts of the roots of (7) are bounded above.

For fixed k, any element of $\mathfrak{N}(A-\lambda I)^k$ is of the form $\sum_{i=0}^{k-1} \theta^i e^{\lambda\theta}\alpha_i$ and since there are only a finite number of linearly independent vectors α_i the space $\mathfrak{N}(A-\lambda I)^k$ is finite dimensional.

Since $\det \Delta(\lambda)$ is an entire function of λ is follows that

$(A-\lambda I)^{-1}$ is a meromorphic function with poles only at the zeros of det $\Delta(\lambda)$. Thus we can apply Theorem 5.8-A of Taylor [12] to conclude that if λ is a zero of order $k > 0$ of det $\Delta(\lambda)$ then $C = \mathfrak{N}(A-\lambda I)^k \oplus \mathcal{R}(A-\lambda I)^k$. Furthermore, since A and $T(t)$ commute for all $t \geq 0$ it follows that $T(t)$ is completely reduced by the two linear manifolds $\mathfrak{N}(A-\lambda I)^k$ and $\mathcal{R}(A-\lambda I)^k$. Thus the theorem is proved.

Now let us consider these spaces in more detail. Let $\varphi_1^\lambda,\ldots,\varphi_d^\lambda$ be a basis for $\mathfrak{M}_\lambda(A) = \mathfrak{N}(A-\lambda I)^k$ and let $\Phi_\lambda = (\varphi_1^\lambda,\ldots,\varphi_d^\lambda)$. Since $A\mathfrak{M}_\lambda(A) \subseteq \mathfrak{M}_\lambda(A)$, there exists a $d \times d$ matrix B_λ such that $A\Phi_\lambda = \Phi_\lambda B_\lambda$ and the only eigenvalue of B_λ is λ. From the definition of A and the relation $A\Phi_\lambda = \Phi_\lambda B_\lambda$ it follows that $\Phi_\lambda(\theta) = \Phi_\lambda(0)e^{B_\lambda\theta}$. From this fact and (6), one obtains

$$T(t)\Phi_\lambda = \Phi_\lambda e^{B_\lambda t}, \quad t \in [0,\infty),$$

(12)

$$[T(t)\Phi_\lambda](\theta) = \Phi_\lambda(0)e^{B_\lambda(t+\theta)}, \quad \theta \in [-r,0),\ t \in [0,\infty).$$

This relation permits one to define $T(t)$ on $\mathfrak{M}_\lambda(A)$ for all values of $t \in (-\infty,\infty)$, and so on a generalized eigenspace the equation (1) has the same structure as an ordinary differential equation. By repeated application of the same process one obtains

COROLLARY 1. *Suppose* Λ *is a finite set* $\{\lambda_1,\ldots,\lambda_p\}$ *of eigenvalues* *of* (1) *and* let $\Phi_\Lambda = (\Phi_{\lambda_1},\ldots,\Phi_{\lambda_p})$, $B_\Lambda = \mathrm{diag}(B_{\lambda_1},\ldots,B_{\lambda_p})$, *where* Φ *and is a basis for* $\mathfrak{M}_{\lambda_i}(A)$ *and* B_{λ_i} *is the matrix defined by* $A\Phi_{\lambda_i} = \Phi_{\lambda_i} B_{\lambda_i}$, $i = 1,2,\ldots,p$. *Then the only eigenvalue of* B_{λ_i} *is* λ_i *and for any vector* a *of the same dimension as* Φ_Λ, *the solution* $T(t)\Phi a$ *with initial value* $\Phi_\Lambda a$ *at* $t = 0$ *may be defined on* $(-\infty,\infty)$ *by the relation*

$$(13) \qquad T(t)\Phi_\Lambda a = \Phi_\Lambda e^{B_\Lambda t} a\ , \quad \Phi_\Lambda(\theta) = \Phi_\Lambda(0)e^{B_\Lambda \theta}, \quad \theta \in [-r,0]\ .$$

Furthermore there exists a subspace Q_Λ *of* C *such that* $T(t)Q_\Lambda \subseteq Q_\Lambda$ *for all* $t \geq 0$ *and*

$$(14) \qquad C = P_\Lambda \oplus Q_\Lambda\ , \quad P_\Lambda = \{\varphi \in C\colon \varphi = \Phi_\Lambda a\ , \ \textit{for some fixed vector}\ \ a\}.$$

This corollary gives a very clear picture of the behavior of the solutions of (1). In fact on the generalized eigenspaces the system behaves much like an ordinary differential equation. The above decomposition of C allows one to introduce a coordinate system in C which plays the same role as the Jordan canonical form in ordinary differential equations.

Before obtaining estimates for $T(t)$ on the complementary subspace Q_Λ, we give an explicit characterization for Q_Λ. This could be obtained from the general theory of linear operators, by means of a contour integral, but we prefer to give this representation in terms of an operator "adjoint" to A relative to a certain bilinear form. This method leads to ease in computations and also provides a language more familiar to differential equationists. Let $C^* = C([0,r],R^{n*})$ where R^{n*} is the n-dimensional linear vector space of row vectors. For any φ in C, define

(15) $$(\alpha,\varphi) = \alpha(0)\varphi(0) - \int_{-r}^{0} \left[\frac{d}{d\zeta} \int_{0}^{\zeta} \alpha(s-\zeta)d\mu(\theta)\varphi(s)ds\right]_{\zeta=\theta} - \int_{-r}^{0}\int_{0}^{\theta} \alpha(s-\theta)d\eta(\theta)\varphi(s)ds$$

for all those α in C^* for which this expression is meaningful. In particular, (α,φ) will have meaning if α is continuously differentiable. The motivation for this bilinear form is not easy to understand, but it was first encountered in the proof of Theorem 1. In fact, equations (8), (9), (10) show that $(A-\lambda I)\varphi = \psi$ has a solution if and only if $(ae^{-\lambda\cdot}I,\psi) = 0$ for all row vectors a for which $a\Delta(\lambda) = 0$.

Without further ado, we use this bilinear form to try to determine an operator A^* with domain dense in C^* such that

(16) $$(\alpha,A\varphi) = (A^*\alpha,\varphi), \quad \text{for } \varphi \text{ in } \mathcal{D}(A), \ \alpha \text{ in } \mathcal{D}(A^*).$$

If we suppose α has a continuous first derivative and perform the standard type of calculations using an integration by parts, one shows that (16) is satisfied if A^* and the domain $\mathcal{D}(A^*)$ of A^* are defined by

(17a) $$(A^*\alpha)(s) = -d\alpha(s)/ds\,, \quad 0 \le s \le r$$

$$(17b)\qquad \mathcal{D}(A^*) = \{\alpha \in C^*;\ \dot{\alpha} \in C^*,$$

$$-\dot{\alpha}(0) = -\int_{-r}^{0} \dot{\alpha}(-\theta)d\mu(\theta) + \int_{-r}^{0} \alpha(-\theta)d\eta(\theta)\}\ .$$

Hereafter, we will take (17) as the defining relation for A^* and refer to A^* as the adjoint of A relative to the bilinear form (15).

For any α in C^*, consider the equation

$$(18a)\qquad y(s) = \alpha(s)\ ,\quad 0 \leqq s \leqq r,$$

$$(18b)\qquad y(s) = \alpha(0) - \int_{-r}^{0} \alpha(-\theta)d\mu(\theta) + \int_{-r}^{0} y(s-\theta)d\mu(\theta) - \int_{0}^{s}\Big[\int_{-r}^{0} y(u-\theta)d\eta(\theta)\Big]du,\quad s \leqq 0.$$

If we let y^s be the element of C^* defined by $y^s(\nu) = y(s+\nu)$, $0 \leqq \nu \leqq r$ and designate the solution of (18) by $y(\alpha)$, then the family of operators $T^*(s)$, $s \leqq 0$, defined by $y^s(\alpha) = T^*(s)\alpha$, $s \leqq 0$, is a strongly continuous semigroup for which $(-A^*)$ is the infinitesimal generator. We shall refer to (18) as the equation adjoint to (1).

Observe that α in $\mathcal{D}(A^*)$ implies that the solution $y(\alpha)$ of (18) on $(-\infty,r]$ is continuously differentiable and

$$(19)\qquad \dot{y}(s) = \int_{-r}^{0} \dot{y}(s-\theta)d\mu(\theta) - \int_{-r}^{0} y(s-\theta)d\eta(\theta)$$

for $s \leqq 0$.

LEMMA 3. <u>Suppose</u> $y(\alpha)$, $\alpha \in \mathcal{D}(A^*)$, <u>is the solution of</u> (18) <u>on</u> $(-\infty,r]$

and $x(\varphi)$ *is the solution of the nonhomogeneous equation*

(20a) $$x_\sigma = \varphi$$

(20b) $$x(t) = \gamma(\varphi) + g(x_t) + \int_\sigma^t f(x_s)ds + \int_\sigma^t h(s)ds, \quad t \geq \sigma.$$

Then for any $\nu \geq \sigma$,

(21) $$(y^{t-\nu}(\alpha), x_t(\varphi)) = (y^{\sigma-\nu}(\alpha), \varphi) + \int_\sigma^t y(s-\nu)h(s)ds, \quad \sigma \leq t \leq \nu.$$

PROOF: For simplicity in notation, let $z^t = y^{t-\nu}(\alpha)$, $t \leq \nu$, $x_t = x_t(\varphi)$, $t \geq 0$. Since α is in $\mathcal{D}(A^*)$, $z(t)$ is continuously differentiable and satisfies (19) for $t \leq \nu$. From the definition (15) and the fact that $x(\varphi)$ satisfies (20), one shows very easily that, for $0 \leq t \leq \nu$,

$$(z^t, x_t) = z(t)[\gamma(\varphi) + \int_\sigma^t f(x_s)ds + \int_\sigma^t h(s)ds] +$$

$$+ \int_{-r}^{o}\int_{t}^{t+\theta} \dot{z}(u-\theta)d\mu(\theta)x(u)du - \int_{-r}^{o}\int_{t}^{t+\theta} z(u-\theta)d\eta(\theta)x(u)du.$$

Consequently, (z^t, x_t) is differentiable in t and a simple calculation yields $d(z^t, x_t)/dt = z(t)h(t)$, $0 \leq t \leq \nu$. Integrating this expression from o to ν yields the formula (21) which proves Lemma 3.

LEMMA 4. *λ is in $\sigma(A)$ if and only if λ is in $\sigma(A^*)$. The operator A^* has only point spectrum and for any λ in $\sigma(A^*)$,*

<u>the generalized eigenspace of</u> λ <u>is finite dimensional</u>.

PROOF: The last part of the lemma is proved exactly as in Lemma 2 and the first part follows from the observation that λ is in $\sigma(A^*)$ if and only if $\alpha(\theta) = e^{-\lambda\theta}b$ where b is a nonzero row vector satisfying $b\Delta(\lambda) = 0$.

<u>LEMMA 5.</u> <u>A necessary and sufficient condition for the equation</u>

$$(A-\lambda I)^k\varphi = \psi \tag{22}$$

<u>to have a solution</u> φ <u>in</u> C, <u>or, equivalently, that</u> ψ <u>is in</u> $\mathcal{R}(A-\lambda I)^k$ <u>is that</u> $(\alpha,\psi) = 0$ <u>for all</u> α <u>in</u> $\mathfrak{N}(A^*-\lambda I)^k$. <u>Also</u>, $\dim \mathfrak{N}(A-\lambda I)^k = \dim \mathfrak{N}(A^*-\lambda I)^k$ <u>for every</u> k.

PROOF: First, we introduce some notation. With the matrix $\Delta(\lambda)$ given in (7), we define the matrices P_j as

$$P_{j+1} = P_{j+1}(\lambda) = \frac{\Delta^{(j)}(\lambda)}{j!}, \quad \Delta^{(j)}(\lambda) = \frac{d^j\Delta(\lambda)}{d\lambda^j}, \quad j = 0,1,2,\ldots,k \tag{23}$$

and the matrices A_k of dimension $kn \times kn$ as

$$A_k = \begin{bmatrix} P_1 & P_2 & \cdots & P_k \\ 0 & P_1 & \cdots & P_{k-1} \\ \vdots & & & \\ 0 & 0 & \cdots & P_1 \end{bmatrix} \tag{24}$$

Let us also define functions β_j by

(25) $$\beta_j(s) = \frac{(-s)^{k-j}}{(k-j)!} e^{-\lambda s}, \quad 0 \leqq s \leqq r, \; j = 1,2,\ldots,k.$$

If (22) is to have a solution, then necessarily $(\frac{d}{d\theta} - \lambda)^k \varphi(\theta) = \psi(\theta)$, $-r \leqq \theta \leqq 0$, or

$$\varphi(\theta) = \sum_{j=0}^{k-1} \gamma_{j+1}\beta_{k-j}(-\theta) + \int_0^\theta \beta_1(\xi-\theta)\psi(\xi)d\xi,$$

where the γ_{j+1} are arbitrary n-dimensional column vectors which must be determined so that φ belongs to $\mathcal{D}(A-\lambda I)^k$. We now derive these conditions on the γ_j.

A simple induction argument on m shows that

$$\varphi^{(m)}(\theta) \overset{\text{def}}{=} (\frac{d}{d\theta} - \lambda)^m \varphi(\theta) = \sum_{j=0}^{k-m-1} \gamma_{m+j+1}\beta_{k-j}(-\theta) + \int_0^\theta \beta_{m+1}(\xi-\theta)\psi(\xi)d\xi$$

for $0 \leqq m \leqq k-1$.

Next, observe that φ belongs to $\mathcal{D}(A-\lambda I)^k$ if and only if $\varphi^{(m)}$ belongs to $\mathcal{D}(A-\lambda I)$, $m = 0,1,\ldots,k-1$. Since a continuously differentiable φ belongs to $\mathcal{D}(A)$ if and only if $\dot{\varphi}(0) = g(\dot{\varphi})+f(\varphi)$, it follows from the definition of the function $\varphi^{(m)}$ and the matrices P_j that $\varphi^{(m)}$, $m < k-1$, belongs to $\mathcal{D}(A)$ if and only if

$$P_1\gamma_{m+1} + P_2\gamma_{m+2} + \cdots + P_{k-m}\gamma_k = -(\beta_{m+1}I_n, \psi)$$

where I_n is the $n \times n$ identity matrix and $(\, , \,)$ is the bilinear form defined in (15). Since $\dot{\varphi}^{(k-1)}(0) = \lambda\gamma_k+\psi(0)$, it follows that $\varphi^{(k-1)}$ belongs to $\mathcal{D}(A)$ if and only if

$$P_1\gamma_k = -(\beta_k I_n, \psi) \, .$$

If we introduce the additional notation $\gamma = \mathrm{col}(\gamma_1,\dots,\gamma_k)$, $B = \mathrm{diag}(\beta_1 I_n,\dots,\beta_n I_n)$, then equation (22) has a solution if and only if γ satisfies the equation $A_k\gamma = -(B,\psi)$. But this equation has a solution if and only if $b(B,\psi) = (bB,\psi) = 0$ for all row vectors b satisfying $bA_k = 0$. On the other hand, calculations very similar to the ones above show that a function α in C^* belongs to $\mathfrak{N}(A^*-\lambda I)^k$ if and only if $\alpha = bB$ for some b satisfying $bA_k = 0$. It is clear from the above that $\dim \mathfrak{N}(A-\lambda I)^k = \dim \mathfrak{N}(A^*-\lambda I)^k$ for every k and this completes the proof of the lemma.

In the proof of the above lemma, we have actually characterized $\mathfrak{N}(A-\lambda I)^k$, $\mathfrak{N}(A^*-\lambda I)^k$ in a manner which is convenient for computations. In fact,

(26a) $$\mathfrak{N}(A-\lambda I)^k = \{\varphi \in C\colon \varphi(\theta) = \textstyle\sum_{j=0}^{k-1}\gamma_{j+1}\beta_{k-j}(-\theta),\ -r \leqq \theta \leqq 0,$$
$$A_k\gamma = 0,\ \gamma = \mathrm{col}(\gamma_1,\dots,\gamma_k)\},$$

(26b) $$\mathfrak{N}(A^*-\lambda I)^k = \{\psi \in C^*\colon \psi(s) = \textstyle\sum_{j=1}^{k}\delta_j\beta_j(s),\ 0 \leqq s \leqq r$$
$$\delta A_k = 0,\ \delta = \mathrm{row}(\delta_1,\dots,\delta_k)\},$$

where A_k, β_j, $j = 1,2,\dots,k$, are defined by (23),(24),(25).

An important implication of the preceding lemma is

THEOREM 2. *For λ in $\sigma(A)$, let $\Psi_\lambda = \mathrm{col}(\psi_1,\dots,\psi_p)$, $\Phi_\lambda = (\varphi_1,\dots,\varphi_p)$ be bases for $\mathfrak{M}_\lambda(A)$, $\mathfrak{M}_\lambda(A^*)$, respectively, and let $(\Psi_\lambda,\Phi_\lambda) = (\psi_i,\varphi_j)$, $i,j = 1,2,\dots,p$. Then $(\Psi_\lambda,\Phi_\lambda)$ is nonsingular and may be taken to be the identity. The decomposition of C given by Lemma 2 may be written explicitly as*

$$C = P_\lambda \oplus Q_\lambda$$

$$Q_\lambda = \{\varphi \in C\colon (\Psi_\lambda,\varphi) = 0\}$$

$$P_\lambda = \{\varphi \in C\colon \varphi = \Phi_\lambda(\Psi_\lambda,\varphi)\}.$$

PROOF: If k is the smallest integer for which $\mathfrak{M}_\lambda(A) = \mathfrak{N}(A-\lambda I)^k$ then Lemma 5 implies that $\mathcal{R}(A-\lambda I)^k = Q_\lambda$. If there is a p-vector a such that $0 = (\Psi_\lambda,\Phi_\lambda)a = (\Psi_\lambda,\Phi_\lambda a)$, then $\Phi_\lambda a$ belongs to both $\mathfrak{N}(A-\lambda I)^k$ and $\mathcal{R}(A-\lambda I)^k$ which implies by Lemma 3 that $\Phi_\lambda a = 0$ and, thus, $a = 0$. Consequently, $(\Psi_\lambda,\Phi_\lambda)$ is nonsingular and a change of the basis Ψ_λ will result in the identity matrix for $(\Psi_\lambda,\Phi_\lambda)$. The remaining statements in the lemma are obvious.

It is interesting to note that $(\Psi_\lambda,\Phi_\lambda) = I$ and $A^*\Psi_\lambda = B^*_\lambda\Psi_\lambda$, $A\Phi_\lambda = \Phi_\lambda B_\lambda$ implies $B^*_\lambda = B_\lambda$. In fact,

$$(\Psi_\lambda, A\Phi_\lambda) = (\Psi_\lambda, \Phi_\lambda B_\lambda) = (\Psi_\lambda, \Phi_\lambda)B_\lambda = B_\lambda$$

$$= (A^*\Psi_\lambda, \Phi_\lambda) = (B_\lambda^*\Psi_\lambda, \Phi_\lambda) = B_\lambda^*(\Psi_\lambda, \Phi_\lambda) = B_\lambda^* .$$

The following lemma is also convenient.

LEMMA 6. *If* $\lambda \neq \mu$, λ, $\mu \in \sigma(A)$, *then* $(\psi,\varphi) = 0$ *for all* ψ *in* $\mathfrak{M}_\mu(A^*)$, $\varphi \in \mathfrak{M}_\lambda(A)$.

The proof of this is not difficult but tedious and may be supplied as in [5].

If $\Lambda = \{\lambda_1, \ldots, \lambda_p\}$ is a finite set of characteristic values of (1); that is, $\lambda_j \in \sigma(A)$, we let P_Λ be the linear extension of the $\mathfrak{M}_{\lambda_j}(A)$, $\lambda_j \in \Lambda$ and refer to this set as the generalized eigenspace of (1) associated with Λ. In a similar manner we define $P_\Lambda^* = \mathfrak{M}_{\lambda_1}(A^*) \oplus \ldots \oplus \mathfrak{M}_{\lambda_p}(A^*)$ as the generalized eigenspace of the adjoint equation (18) associated with Λ. If $\Phi_\Lambda, \Psi_\Lambda$ are bases for P_Λ, P_Λ^*, respectively, $(\Psi_\Lambda, \Phi_\Lambda) = I$, then

(27)
$$\begin{aligned} C &= P_\Lambda \oplus Q_\Lambda \\ P_\Lambda &= \{\varphi \in C\colon \varphi = \Phi_\Lambda b \text{ for some vector } b\} \\ Q_\Lambda &= \{\varphi \in C\colon (\Psi_\Lambda, \varphi) = 0\} \end{aligned}$$

and, therefore, for any φ in C

$$\varphi = \varphi^{P_\Lambda} \oplus \varphi^{Q_\Lambda}$$

(28)

$$\varphi^{P_\Lambda} = \Phi_\Lambda(\Psi_\Lambda, \varphi).$$

When this particular decomposition of C is used, we shall briefly express this by saying that C is <u>decomposed by Λ.</u>

Our next objective is to perform the above decomposition on the variation of constants formula for the solution of (20). From Corollary II.2, we know that the solution of (20) can be written as

$$\begin{aligned} x(t+\theta,\sigma,\varphi,h) &= x(t+\theta,\sigma,\varphi,0) + \int_\sigma^{t+\theta} U(t+\theta-s)h(s)ds \\ &= x(t+\theta,\sigma,\varphi,0) + \int_\sigma^{t+\theta} [d_s W(t+\theta-s)]h(s), \ t+\theta \geq \sigma. \end{aligned}$$

If we use our notation $x(t+\theta,\sigma,\varphi,0) = x(t+\theta-\sigma,0,\varphi,0) = [T(t-\sigma)\varphi](\theta)$ and the fact that $W_0 = 0$, then

$$x_t(\sigma,\varphi,h)(\theta) = [T(t-\sigma)\varphi](\theta) + \int_\sigma^t [d_s W_{t-s}(\theta)]h(s), \ -r \leq \theta \leq 0.$$

For simplicity we suppress the explicit dependence on θ and write this as

$$\begin{aligned} x_t(\sigma,\varphi,h) &= T(t-\sigma)\varphi + \int_\sigma^t [d_s W_{t-s}]h(s) \\ &= T(t-\sigma)\varphi + \int_\sigma^t U_{t-s}h(s)ds \end{aligned} \tag{29}$$

where U_t is defined in the obvious way.

Now, suppose that Λ is a finite set of characteristic values of (1) and C is decomposed by Λ as in formulas (27),(28). For simplitity in notation, let $\Phi = \Phi_\Lambda$, $\Psi = \Psi_\Lambda$ and let B be the matrix defined by $A\Phi = \Phi B$. We have remarked before that $(\Psi,\Phi) = I$ implies that $A^*\Psi = B\Psi$. Consequently, the matrix $e^{-Bt}\Psi(0)$ is a solution of the adjoint equation (18) on $(-\infty,\infty)$. If we let $x_t = x_t(\sigma,\varphi,h) = x_t^{P_\Lambda} + x_t^{Q_\Lambda}$ and apply Lemma 3, it therefore follows that

$$
\begin{aligned}
x_t^{P_\Lambda} &\overset{\text{def}}{=} \Phi(\Psi,x_t) = \Phi e^{Bt}(e^{-Bt}\Psi,x_t) \\
&= \Phi e^{Bt}[(e^{-B\sigma}\Psi,\varphi) + \int_\sigma^t e^{-Bs}\Psi(0)h(s)ds] \\
&= T(t-\sigma)\Phi(\Psi,\varphi) + \int_\sigma^t \Phi e^{B(t-s)}\Psi(0)h(s)ds \\
&= T(t-\sigma)\varphi^P + \int_\sigma^t [d_s(-\int_0^{t-s} \Phi e^{Bu}\Psi(0))]h(s)
\end{aligned} \tag{30}
$$

If $W_t = W_t^P + W_t^Q$, $W_t^P = \Phi(\Psi,W_t)$, $t \geq 0$, then by the same type of argument as above making use of Lemma 3 and the fact that W satisfies II(4), we obtain

$$
W_t^P \overset{\text{def}}{=} \Phi(\Psi,W_t) = -\int_0^t \Phi e^{B(t-s)}\Psi(0)ds = -\int_0^t \Phi e^{Bu}\Psi(0)du .
$$

Using this fact, equation (29),(30) and the formulas $x_t^Q = x_t - x_t^P$,

$\varphi^Q = \varphi - \varphi^P$, we have

(31a) $$x_t^P(\sigma,\varphi,h) = T(t-\sigma)\varphi^P + \int_\sigma^t [d_s W_{t-s}^P] h(s) ,$$

(31b) $$x_t^Q(\sigma,\varphi,h) = T(t-\sigma)\varphi^Q + \int_\sigma^t [d_s W_{t-s}^Q] h(s) , \quad t \geq 0.$$

From formula (29), it is obvious that if $x_t^P(\sigma,\varphi,h) = \Phi y(t)$, then $y(t)$ satisfies the ordinary differential equation

(32) $$\dot{y}(t) = B_\Lambda y(t) + \Psi(0)h(t) , \qquad t \geq 0.$$

<u>THEOREM 3</u>. <u>If</u> Λ <u>is a finite set of characteristic values of</u> (1) <u>and</u> C <u>is decomposed by</u> Λ <u>as in</u> (27),(28), <u>then the solution</u> $x(\sigma,\varphi,h)$ <u>of</u> (20) <u>satisfies</u> (31). <u>Furthermore, if</u> $x_t^P(\sigma,\varphi,h) = \Phi_\Lambda y(t)$, <u>then</u> $y(t)$ <u>satisfies</u> (32).

We now give an example to clarify the concepts discussed in this section. An easier illustration could be given by considering only a retarded equation, but the example to be given will be used later for other applications of the theory. Consider the homogeneous scalar equation

(33) $$\dot{x}(t) = \alpha_0 \dot{x}(t-r) - \beta x(t) - \alpha_0 \gamma x(t-r)$$

where $r > 0$, α_0, β, γ are constants and the associated nonhomogeneous equation

$$(34)\qquad \dot{x}(t) = \alpha_o\dot{x}(t-r) - \beta x(t) - \alpha_o\gamma x(t-r) + h$$

where h is some given function. For simplicity in notation, we are writing these equations in differential form, but it is always understood that solutions are defined by specifying a continuous initial function on an interval $[\sigma-r,\sigma]$ and solving the integrated form of the equation for x on $t \geq \sigma$.

The characteristic equation for (33) is

$$(35)\qquad \lambda - \alpha_o\lambda e^{-\lambda r} + \beta + \alpha_o\gamma e^{-\lambda r} = 0$$

and the associated bilinear form is

$$(36)\qquad (\psi,\varphi) = \psi(0)\varphi(0)-\alpha_o\psi(0)\varphi(-r)-\alpha_o\int_{-r}^{0}\dot{\psi}(\theta+r)\varphi(\theta)d\theta-\alpha_o\gamma\int_{-r}^{0}\psi(\theta+r)\varphi(\theta)d\theta.$$

Equation (34) was encountered by Brayton [13] in the study of transmission lines and he showed that for $\gamma > \beta > 0$ there are an infinite set of real pairs (α_o,ω_o), $\omega_o > 0$, $\alpha_o^2 < 1$, such that $\pm\, i\omega_o$ are simple roots of (35) and ω_o,α_o are related by the formulas

$$(37)\qquad \sin\omega_o r = \frac{\omega_o}{\alpha_o}\cdot\frac{\gamma+\beta}{\omega_o^2+\gamma^2}\quad,\quad \cos\omega_o r = \frac{1}{\alpha_o}\cdot\frac{\omega_o^2-\gamma\beta}{\omega_o^2+\gamma^2}\quad .$$

Let us assume that α_o is such a real number and compute the decomposition of C assording to the set $\Lambda = \{+ i\omega_o, - i\omega_o\}$.

If $\Phi = (\varphi_1, \varphi_2)$, $\varphi_1(\theta) = \sin \omega_o \theta$, $\varphi_2(\theta) = \cos \omega_o \theta$, $-r \leq \theta \leq 0$, then Φ is a basis for the generalized eigenspace of (33) associated with Λ since we are assuming these eigenvalues are simple. Furthermore, $A\Phi = \Phi B$ implies

(38)
$$B = (b_{ij}),\ b_{11} = b_{22} = 0,\ b_{12} = -\omega_o = -b_{21}.$$

The equation adjoint to (33) is

(39)
$$\dot{y}(t) = \alpha_o \dot{y}(t+r) + \beta y(t) + \alpha_o \gamma y(t+r)$$

and $\Psi^* = \mathrm{col}(\psi_1^*, \psi_2^*)$, $\psi_1^*(\theta) = \sin \omega_o \theta$, $\psi_2^*(\theta) = \cos \omega_o \theta$, $0 \leq \theta \leq r$ is a basis for the generalized eigenspace of (39) associated with Λ.

After some straightforward but tedious calculations using (37) one obtains

$$(\psi_1^*, \varphi_1) = (\psi_2^*, \varphi_2) = \frac{1}{2(\omega_o^2 + \gamma^2)} [\gamma(\gamma+\beta) + r\beta(\gamma^2 + \omega_o^2)]$$

$$(\psi_2^*, \varphi_1) = -(\psi_1^*, \varphi_2) = \frac{\omega_o}{2(\omega_o^2 + \gamma^2)} [\gamma + \beta + r(\gamma^2 + \omega_o^2)] .$$

If we now define $\Psi = (\Psi^*, \Phi)^{-1} \Psi^*$, then $(\Psi, \Phi) = I$, the identity and we are in a position to make our decomposition of C by Λ. Our main interest lies in formulas (31),(32) and in particular (31b) and (32). Consequently, we only need $\Psi(0)$ which is easily calculated from the above formulas and found to be

$$\Psi(0) = \mathrm{col}\left[\frac{D}{C^2+D^2}, \frac{C}{C^2+D^2}\right]$$

$$(40) \qquad C = \frac{1}{2(\omega_o^2+\gamma^2)}[\gamma(\gamma+\beta)+r\beta(\gamma^2+\omega_o^2)]$$

$$D = \frac{\omega_o}{2(\omega_o^2+\gamma^2)}[\gamma+\beta+r(\gamma^2+\omega_o^2)] .$$

Finally, equation (34) is equivalent to the following system

$$x_t = \Phi y(t) + x_t^Q$$

$$(41) \qquad \dot{y}(t) = By(t) + \Psi(0)h$$

$$x_t^Q = T(t-\sigma)\varphi^Q + \int_\sigma^t [d_s W_{t-s}]h , \qquad t \geq \sigma$$

where $\Psi(0)$ is given in (40) and B is defined in (38).

IV. THE CHARACTERISTIC EQUATION AND EXPONENTIAL BOUNDS.

In this section the zeros of the characteristic equation are discussed and estimates are obtained for the growth of the solutions on the compliment of the generalized eigenspaces.

In order to analyze the characteristic equation it is necessary to further restrict the functional g or equivalently the measure μ. It is known [10] that every function of bounded variation can be decomposed into three summands 1) a saltus function (essentially a step function with a countable number of discontinuities) 2) an absolutely continuous function and 3) a "singular function" that is a continuous function of bounded variation whose derivative is zero almost everywhere. We shall assume that the measure μ is without singular part.

Specifically, assume that

$$(1) \qquad g(\varphi) = \sum_{k=1}^{\infty} A_k \varphi(-\omega_k) + \int_{-r}^{0} A(\theta)\varphi(\theta)d\theta \quad , \quad \varphi \in C([-r,0],R^n)$$

where the A_k are $n \times n$ constant matrices with $\sum_{1}^{\infty} A_k$ absolutely convergent, the ω_k are a countable sequence of real numbers with $0 < \omega_k \leqq r$ for all k and $A(\theta) \in \mathcal{L}_1([-r,0],R^{n^2})$.

Under the above assumption $\Delta(\lambda)$ has the form

$$(2) \qquad \Delta(\lambda) = \lambda\{H_1(\lambda) + H_2(\lambda)\} + H_3(\lambda)$$

where

$$\text{a)} \quad H_1(\lambda) = I - \sum_{k=1}^{\infty} A_k e^{-\omega_k \lambda}$$

$$(3) \qquad \text{b)} \quad H_2(\lambda) = -\int_{-r}^{0} A(\theta)e^{\lambda\theta}d\theta$$

c) $H_3(\lambda) = -\int_{-r}^{0} e^{\lambda\theta} d\eta(\theta)$.

Moreover $\det \Delta(\lambda) = \lambda^n h_1(\lambda) + h_2(\lambda)$ where $h_1(\lambda) = \det H_1(\lambda)$ and $h_2(\lambda) = \det \Delta(\lambda) - \lambda^n h_1(\lambda)$.

For any pair of real numbers α, β $(\alpha \leq \beta)$ let $[\alpha,\beta] = \{\lambda: \alpha \leq \text{Re}\,\lambda \leq \beta\}$. In any $[\alpha,\beta]$ the elements of $H_3(\lambda)$ are bounded and the elements of $H_2(\lambda)$ tend uniformly to zero as $|\lambda| \to \infty$. Thus $h_2(\lambda) = o(\lambda^n)$ as $|\lambda| \to \infty$ in $[\alpha,\beta]$.

LEMMA 1. *If $\{\lambda_k\}$ is a sequence of zeros of h_1 in $[\alpha + \delta, \beta-\delta]$, $\delta > 0$, with $|\lambda_n| \to \infty$, then there exists a sequence $\{\lambda_k'\}$ of zeros of* $\det \Delta(\lambda)$ *in* $[\alpha,\beta]$ *with the property that $|\lambda_k - \lambda_k'| \to 0$, as* $k \to \infty$.

LEMMA 2. *Let* a *be a real number such that only a finite number of zeros of* $\det \Delta(\lambda)$ *have real part greater than* $a-\varepsilon$ *for some* $\varepsilon > 0$. *Then there exists an* a^* *and a* $K > 0$ *such that* $a - \frac{1}{2}\varepsilon \leq a^* \leq a$ *and* $\|\Delta(a^*+i\xi)^{-1}\| \leq K/(1+|\xi|)$ for ξ real.

PROOFS. The function $h_1(\lambda)$ is an analytic almost periodic function for all λ. Then by a theorem in [14], page 351 there exists a number N such that the number of zeros of $h_1(\lambda)$ in the box $\mathcal{B}(\alpha + \delta, \beta-\delta, t^*) = \{\lambda: \alpha + \delta \leq \text{Re}\lambda \leq \beta-\delta,\ t^*-1/2 \leq \text{Im}\,\lambda \leq t^* + 1/2\}$ does not exceed N for any real t^*. Moreover for each $r > 0$ there exists an $m(r) > 0$ such that for all λ in $[\alpha,\beta]$ at a distance greater than r from a zero of $h_1(\lambda)$ the inequality $|h_1(\lambda)| \geq m(r)$ holds.

Thus Lemma 1 follows by applying Rouché's Theorem.

Now let a be as in Lemma 2. Since $h(\lambda)$ has only finitely many zeros with real part greater than $a-\varepsilon$ for some $\varepsilon > 0$ it follows from Lemma 1 that $h_1(\lambda)$ has only finitely many zeros with real part greater than $a-\varepsilon/2$. Therefore there exists an a^*, $a-\varepsilon/2 \leq a^* \leq a$, and a $K_2 > 0$ such that $|h_1(\lambda)| \geq K_2$ for all $\lambda = a^* + i\xi$, ξ, real. Thus $|h(a^* + i\xi)^{-1}| = O(\xi^{-n})$ as $|\xi| \to \infty$, ξ real.

Since $\Delta(\lambda)^{-1} = (h(\lambda)^{-1})\operatorname{adj} \Delta(\lambda)$ and $\|\operatorname{adj}\Delta(a^* + i\xi)\| = O(\xi^{n-1})$ as $|\xi| \to \infty$, ξ real, Lemma 2 follows.

With the aid of Lemma 2 one can now estimate the growth of the solutions on the space Q_Λ. Let Λ be a finite set of eigenvalues of A with the property that all other eigenvalues of A have real part less that $a-\varepsilon$ for a fixed real number a and some $\varepsilon > 0$.

Let $u(\cdot,\sigma,h)$ be the solution of the nonhomogeneous equation that satisfies $u_\sigma = 0$, i.e., the solution given by the integral in the Corollary 1 of Section III. Let u_t^Q be the projection of u_t on the space Q_Λ and $u(t)^Q = u_t^Q(0)$.

Let C^1 denote the set of continuously differentiable function from $[-r,0]$ into R^n with the norm $\|\varphi\|^1 = \sup_{\theta\in[-r,0]} \{|\varphi(\theta)| + |\dot{\varphi}(\theta)|\}$.

<u>THEOREM 1</u>. <u>Let</u> $\varphi \in C^1$. <u>Then there exist constants</u> M <u>and</u> N <u>such that</u>

(4) $$\| T(t)\varphi^Q \| \leq Me^{at} \|\varphi\|^1$$

and

(5) $$\| u_t^Q \| \leq N \{\int_\sigma^t | e^{a(t-\tau)} h(\tau)|^2 d\tau\}^{1/2}$$

PROOF: In the proof of this theorem the fact that the formulas

$$g(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ixy} f(y)dy \ ; \ f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixy} g(y)dy$$

define a unitary transformation of the space $L_2(-\infty,\infty)$ and its inverse will be used several times (see [10]). In the formulas $\int_{-\infty}^{\infty} = \lim_{T\to\infty} \int_{-T}^{T}$.

By standard Laplace transform methods

(6) $$u(t)^Q = \int_c e^{\lambda t} \Delta(\lambda)^{-1} \{\int_0^t e^{-\lambda\tau} h(\tau)d\tau\} d\lambda$$

where $\int_c = \lim_{T\to\infty} \int_{a^*-iT}^{a^*+iT}$ and a^* is in Lemma 2. Now (6) can be written

(7) $$u(t)^Q = i \int_{-\infty}^{\infty} e^{i\xi t} \Delta(a^*+i\xi)^{-1} \{\int_0^t e^{-i\xi\tau} [e^{a(t-\tau)} h(\tau)d\tau]\} d\xi \ .$$

The function in the braces is an L^2 function of ξ for each t and $\Delta(a^*+i\xi)^{-1}$ is an L_2 function of ξ by Lemma 2. Applying Schwartz's inequality yields

$$|u(t)^Q| \leq M_1\{\int_0^t |e^{a^*(t-\tau)}h(\tau)|^2 d\tau\}^{1/2}$$

from which the inequality (5) follows at once.

Let $\varphi \in \mathscr{D}(A)$. Then

$$(8) \qquad T(t)\varphi^Q = \int_c e^{\lambda t}[\Delta(\lambda)^{-1}\{-\varphi(0) + \int_{-r}^{0} d\mu(\theta)(\frac{d}{d\theta}\int_0^{\theta} e^{\lambda(\theta-\alpha)}\varphi(\alpha)d\alpha)$$

$$+ \int_{-r}^{0} d\eta(\theta) \int_0^{\theta} e^{\lambda(\theta-\alpha)}\varphi(\alpha)d\alpha\} + \int_0^{\theta} e^{\lambda(\theta-\alpha)}\varphi(\alpha)d\alpha]d\lambda$$

by [11].

The term containing $\int_0^{\theta} e^{\lambda(\theta-\alpha)}\varphi(\alpha)d\alpha$ contributes nothing since it is an entire function of λ and the contour can be shifted to $-\infty$.

Now

$$\int_{-r}^{0} d\mu(\theta)(\frac{d}{d\theta}\int_0^{\theta} e^{\lambda(\theta-\alpha)}\varphi(\alpha))d\alpha =$$

$$= \{\int_{-r}^{0} d\mu(\theta)e^{\lambda\theta}\}\varphi(0) + \int_{-r}^{0} d\mu(\theta)\int_0^{\theta} d^{\lambda(\theta-\alpha)}\dot{\varphi}(\alpha)d\alpha .$$

From the matrix identity $\{\lambda B + C\}^{-1}B = \lambda^{-1}\{I-(\lambda B+C)^{-1}B\}$ one obtains

$$\int_c e^{\lambda t}\Delta(\lambda)^{-1}\{-I + \int_{-r}^{o} d\mu(\theta)e^{\lambda\theta}\}\varphi(0) =$$

$$\{-\int_c \frac{e^{\lambda t}}{\lambda}.d\lambda + \int_c e^{\lambda t}\,\frac{\Delta(\lambda)^{-1}}{\lambda}\,[\int_{-r}^{o} d\eta(\theta)e^{\lambda\theta}]d\lambda\}\varphi(0).$$

The first integral is integrable and is known to admit an estimate of the form

$$|\int_c \frac{e^{\lambda t}}{\lambda}\,d\lambda| \leq M_2 e^{a^*t},$$

The second integral is absolutely convergent since $\lambda^{-1}\Delta(\lambda)^{-1}$ is like λ^{-2} on c, and thus

$$|\int_c e^{\lambda t}\,\frac{\Delta(\lambda)^{-1}}{\lambda}\,[\int_{-r}^{o} d\eta(\theta)e^{\lambda\theta}]d\lambda| \leq e^{a^*t}M_3 .$$

For $\lambda = a^*+i\xi$, we have

$$\int_{-r}^{o} d\mu(\theta) \int_{o}^{\theta} e^{\lambda(\theta-\alpha)}\phi(\alpha)d\alpha = \int_{o}^{-r} d\beta \int_{-r}^{\beta} d\mu(\theta)\{e^{\lambda\beta}\phi(\theta-\beta)\}$$

$$= \int_{o}^{-r} e^{i\xi\beta}[e^{a^*\beta} \int_{-r}^{\beta} d\mu(\theta)\phi(\theta-\beta)]d\beta .$$

As a function of ξ the above is an element of $L_2(-\infty,\infty)$ whose norm can be estimated by $M_4\|\varphi\|^1$. By applying Schwartz's inequality we obtain

$$|\int_c e^{\lambda t}\Delta(\lambda)^{-1}\{\int_{-r}^{0} d\mu(\theta)\int_{0}^{\theta} e^{\lambda(\theta-\alpha)}\dot{\phi}(\alpha)d\alpha\}| \leq M_5 e^{a^*t}\|\varphi\|^1.$$

In the same way as the above

$$|\int_c e^{\lambda t}\Delta(\lambda)^{-1}\{\int_{-r}^{0} d\eta(\theta)\int_{0}^{\theta} e^{\lambda(\theta-\alpha)}\varphi(\alpha)d\alpha\ |\} \leq M_5 e^{a^*t}\|\varphi\| .$$

Thus the estimate (4) is obtained for all $\varphi \in \mathscr{D}(A)$. The estimate (4) remains true for all continuously differentiable φ since $\mathscr{D}(A)$ is dense in C^1.

COROLLARY 1. If $g \equiv 0$ in III(1)b) then

$$\|T(t)\varphi\| \leq Me^{at}\|\varphi\| \quad \text{for all } \varphi \in C .$$

V. APPLICATIONS: STABILITY AND INTEGRAL MANIFOLDS.

In this section two applications are given to illustrate how the previously developed theory of linear equations can be used to study weakly nonlinear systems. It is hoped that this section will indicate the possibility of further extensions and applications. The first application is the analogue of a well known stability theorem by first approximation for ordinary differential equations. The second is an extension of the method of integral manifolds to this new class of equations.

The general outline of the proofs given below is the same as in the case of ordinary differential equations, but certain technical details are markedly different.

V.1. Stability

Our proof of the stability theorem is modeled on the standard proof using Gronwall's inequality (see [15] and [16]). For this we need the following:

LEMMA 1. *There exists a constant* $K > 0$ *independent of* $\alpha, \beta > 0$, *such that any function* u *that is continuous for all* $t \geq 0$ *and satisfies*

$$u(t) \leq \alpha + \beta\{\int_0^t u(s)^2 ds\}^{1/2} \quad \text{for} \quad t \geq 0$$

also satisfies the inequality $u(t) \leq \alpha K \exp \beta^2 t/2$.

PROOF. Note that there is no loss in generality by taking $\alpha = 1$. Consider the continuous linear operator I from $C([0,E],R)$ into itself for each $E > 0$ defined by

$$(Iu)(t) = 1 + \beta\{\int_0^t u(s)^2 ds\}^{1/2}$$

Observe that I has the following property: if $u(t) \leqq v(t)$ for $t \in [0,E]$, $E \geqq 0$, then $(Iu)(t) \leqq (Iv)(t)$ for $t \in [0,E]$. Hence by [1], p. 61, it follows that any function w continuous for $t \geqq 0$ will dominate functions satisfying (1) if $(Iw)(t) > w(t)$ for $t \geqq 0$. That is if w satisfies $(Iw)(t) > w(t)$ for all $t \geqq 0$ and $u(t)$ satisfies (1) then $u(t) \leqq w(t)$ for $t \geqq 0$.

Observe that if v satisfies $Iv = v$; $(Iv)(t) = v(t)$, $t \geqq 0$; then $w = Bv$, $B > 1$ satisfies $(Iw)(t) > w(t)$ for $t \geqq 0$. Hence we must only analyze the equation $Iv = v$.

By a simple application of the contracting mapping principle one finds that I has a fixed point in $C([0,E],R)$ for E sufficiently small. Denote this fixed point by u and then u satisfies the differential equation

$$\dot{u} = \frac{\beta^2}{2}\{\frac{u^2}{u-1}\} = \frac{\beta^2}{2}u\{\frac{1}{1-u^{-1}}\} \quad \text{for} \quad 0 < t < E. \tag{2}$$

Clearly u can be shown to exist for $t \geqq E$ and hence for $t \geqq 0$. Moreover it is clear from (2) that u admits an estimate of the form $u(t) \leqq K \exp(\beta^2 t/2)$.

Now consider the equation

$$(3) \qquad x(t) = \gamma(\varphi) + g(x_t) + \int_\sigma^t f(x_s)ds + \int_\sigma^t F(s,x_s)ds$$

$$x_\sigma = \varphi \ , \ \varphi \in C.$$

where F is a continuous mapping from $[\tau,\infty) \times S_E$ into R^n where $S_E = \{\varphi \in C: \|\varphi\| < E\}$ and also $\tau \leq \sigma$. Also assume F is Lipschitzian in the second argument on all of $[\tau,\infty) \times S_E$ and let $|F(t,\varphi)| = o(\|\varphi\|)$ uniformly in t as $\|\varphi\| \to 0$.

Furthermore let g be such that the estimates of section IV apply and let A be the infinitesimal generator of the semigroup generated by (3) with $F \equiv 0$.

THEOREM 1. Let all the eigenvalues of A have real parts less than $-a < 0$, let $\varphi \in C^1([-r,0],R^n)$, and let $x(\varphi)$ be the solution of (3) with $x_\sigma(\varphi) = \varphi$. Then for any $\varepsilon > 0$, $0 < \varepsilon < a$, there exists a pair of constants ρ and L such that

$$(4) \qquad \|x_t(\varphi)\| \leq L\|\varphi\|^1 e^{-(a-\varepsilon)(t-\sigma)} \ , \ t \geq \sigma$$

provided $\|\varphi\|^1 \leq \rho$.

REMARK. Existence and uniqueness of a solution to equation (3) can be established in a manner similar to that found in section II. The present problem is slightly more complicated since the application of the contracting mapping principle gives the existence of solutions over an interval whose length depends on the norm of the initial condition. This difficulty can be overcome by using a continuation argument as in ordinary differential equations. Indeed it can be shown that a solution of (3) can be extended either for all $t \geqq 0$ or until it reaches the boundary of S_E.

PROOF. Let x be the solution of (3) corresponding to the continuously differentiable inital function $\varphi \in S_E$. As long as $x(\varphi)$ satisfies (3) then

$$(5) \qquad x(t) = T(t)\varphi + \int_\sigma^t \{d_s W(t-s)\} F(s,x_s)$$

From the results of section IV there exist constants M and N such that

$$(6) \qquad \|x_t\| \leq M(\|\varphi\|^1)e^{-a(t-\sigma)} + N\{\int_\sigma^t e^{-2a(t-s)}|F(s,x_s)|^2 ds\}^{1/2}$$

and since $|F(s,\varphi)| = o(\|\varphi\|)$ we can choose a $\rho > 0$ such that $|F(s,\varphi)| \leq N^{-1}\sqrt{2\varepsilon}\,\|\varphi\|$ for all $\|\varphi\| < \rho$ and so

$$e^{a(t-\sigma)}\|x_t\| \leq M\|\varphi\|^1 + \sqrt{2\varepsilon}\ \{\int_\sigma^t [e^{a(s-\sigma)}\|x_s\|]^2 ds\}^{1/2}$$

and so by Lemma 1 $e^{a(t-\sigma)}\|x_t\| \leq KM\|\varphi\|^{1}e^{\varepsilon(t-\sigma)}$ or $\|x_t\| \leq KM\|\varphi\|^{1}e^{-(a-\varepsilon)(t-\sigma)}$ for $t \geq \sigma$. The last estimate holds for all $t \geq \sigma$ provided ρ is sufficiently small since the above estimate implies that the solution does not leave S_E.

V.2. <u>Averaging and integral manifolds.</u>

In this section, we shall show how the results of the previous pages together with generalizations of well known perturbational methods of ordinary differential equations can be used to discuss the existence and stability of periodic solutions and integral manifolds of perturbed linear systems where the nonlinear term is of a special type. The hypotheses are unnecessarily restrictive and the presentation is given in this way for simplicity only. Generalizations will be obvious to the reader acquainted with the theory of oscillations for ordinary differential equations.

Consider the linear system

$$\text{(7)} \quad \begin{aligned} &\text{a)} \quad x_\sigma = \varphi \quad \text{where} \quad \varphi \in C\,, \\ &\text{b)} \quad x(t) = \gamma(\varphi,\varepsilon)+g(x_t,\varepsilon)+\int_\sigma^t f(x_\tau,\varepsilon)d\tau \quad , \quad t \geq \sigma \end{aligned}$$

where $\varepsilon \geq 0$ is a parameter, $\gamma(\varphi,\varepsilon) = \varphi(0)-g(\varphi,\varepsilon)$, $g(\varphi,\varepsilon)$, $f(\varphi,\varepsilon)$ are linear in φ and continuous in φ, for all φ in C,

$0 \leq \varepsilon \leq \varepsilon_0$ with the continuity in φ being uniform in ε. Furthermore, suppose $g(\varphi,\varepsilon)$ has the nonatomic property I(5) uniformly in ε. The characteristic equation of (7) is

$$\text{(8)} \qquad \begin{aligned} &\det \Delta(\lambda,\varepsilon) = 0 \\ &\Delta(\lambda,\varepsilon) = \lambda[I-g(e^{\lambda\cdot},\varepsilon)] - f(e^{\lambda\cdot},\varepsilon). \end{aligned}$$

We shall always assume that equation (8) has two simple roots $\varepsilon\nu(\varepsilon) \pm i\omega(\varepsilon)$, $\omega(\varepsilon) = \omega_0 + \varepsilon\omega_1(\varepsilon)$, $\omega_0 > 0$, $\nu(\varepsilon)$, $\omega(\varepsilon)$ continuous in ε, $0 \leq \varepsilon \leq \varepsilon_0$, and the remaining roots have real parts $\leq -\delta < 0$. Notice that for $\varepsilon = 0$, this hypothesis implies that (7) has a two parameter family of periodic solutions of period $2\pi/\omega_0$ to which all other solutions (with smooth enough initial data) approach as $t \to \infty$. For $\varepsilon > 0$, there is a two parameter family of solutions [corresponding to the characteristic roots $\varepsilon\nu(\varepsilon) \pm i\omega(\varepsilon)$] which are exponentially stable. We shall let $\Phi_\varepsilon = (\varphi_{1\varepsilon},\varphi_{2\varepsilon})$ be a basis for the solutions in C generated by the roots $\Lambda = \{\varepsilon\nu(\varepsilon) \pm i\omega(\varepsilon)\}$ and $\Psi_\varepsilon = \text{col}\,(\psi_{1\varepsilon},\psi_{2\varepsilon})$ a corresponding basis for the solutions of the adjoint equation, $(\Psi_\varepsilon,\Phi_\varepsilon) = I$.

Suppose $F: R \times C \to R^n$ is continuous and $F(t,\varphi)$, $t \in R$, $\varphi \in C$ has continuous second derivatives with respect to φ and consider the nonlinear equation

$$\text{(9)} \qquad \begin{aligned} &\text{a)}\ x(t) = \varphi(t-\sigma),\ \sigma-r \leq t \leq \sigma, \\ &\text{b)}\ x(t) = \gamma(\varphi,\varepsilon)+g(x_t,\varepsilon)+ \int_\sigma^t f(x_\tau,\varepsilon)d\tau + \varepsilon \int_0^t F(\tau,x_\tau)d\tau, t \geq \sigma. \end{aligned}$$

Notice that formal differentiation of this equation with respect to t yields

$$(10) \qquad \dot{x}(t) = g(\dot{x}_t,\varepsilon) + f(x_t,\varepsilon) + \varepsilon F(t,x_t);$$

that is, an equation of neutral type where the nonlinearity does not involve the derivative of x. An equation of this type with $F(t,\varphi)$ independent of t was encountered by Miranker [17] and Brayton [13] in the theory of transmission lines. Similar equations have also been studied by Marchenko and Rubanik [18] in connection with some mechanical vibration problems.

If the space C is decomposed by $\Lambda = \{\varepsilon\nu(\varepsilon) \pm i\omega(\varepsilon)\}$, then the theory of section 3 shows that system (3) is equivalent to the system

$$(11) \qquad \begin{aligned} &\text{a)} \quad x_t = \Phi_\varepsilon y(t) + x_t^Q \;, \quad y(t) = (\Psi_\varepsilon, x_t) \\ &\text{b)} \quad \dot{y}(t) = B_\varepsilon y(t) + \varepsilon\Psi_\varepsilon(0)F(t,\Phi_\varepsilon y(t) + x_t^Q) \;, \\ &\text{c)} \quad x_t^Q = T_\varepsilon(t-\sigma)x_\sigma^Q + \varepsilon\int_\sigma^t [d_s W_{\varepsilon,t-s}^Q]F(s,\Phi_\varepsilon y(s) + x_s^Q)ds, \; t \geq \sigma, \end{aligned}$$

where the eigenvalues of B_ε are $\{\varepsilon\nu(\varepsilon) \pm i\omega(\varepsilon)\}$;$B_\varepsilon$ is determined by $\Phi_\varepsilon(\theta) = \Phi_\varepsilon(0)\exp B_\varepsilon\theta$, $-r \leq \theta \leq 0$, $T_\varepsilon(t)$, $t \geq 0$ designates the semigroup of transformations associated with (7) and $W_{\varepsilon,t}$ is the kernel function associated with the variation of constants formula II(3); that is, $W_\varepsilon(t)$ satisfies II(4) for $0 \leq \varepsilon \leq \varepsilon_0$. The matrix B_ε can actually be chosen as

$$B_\varepsilon = \begin{bmatrix} \varepsilon\nu(\varepsilon) & -\omega(\varepsilon) \\ \omega(\varepsilon) & \varepsilon\nu(\varepsilon) \end{bmatrix}$$

The above hypotheses on the characteristic equation (8) and the estimates of section 4 imply that there are positive constants K,c such that

$$(12)\qquad \begin{aligned} &\text{a) } \left|\int_\sigma^t [d_s W^Q_{\varepsilon,t-s}]h(s)ds\right| \le K\left(\int_\sigma^t (e^{-c(t-s)}|h(s)|)^2\right)^{1/2} \\ &\text{b) } \|T_\varepsilon(t)\varphi^Q\| \le Ke^{-ct}\|\varphi^Q\|^1 \quad , \quad t \ge 0, \end{aligned}$$

for all bounded functions $h(s)$ and $0 \le \varepsilon \le \varepsilon_0$.

If $y = \mathrm{col}(y_1,y_2)$, $y_1 = \rho\cos\zeta$, $y_2 = \rho\sin\zeta$, then equations (11b),(11c) are equivalent to

$$(13)\qquad \begin{aligned} &\text{a) } \dot\zeta = \omega(\varepsilon) + \varepsilon Z(t,\zeta,\rho,x^Q_t,\varepsilon) \\ &\text{b) } \dot\rho = \varepsilon R(t,\zeta,\rho,x^Q_t,\varepsilon) \\ &\text{c) } x^Q_t = T_\varepsilon(t-\sigma)x^Q + \varepsilon\int_\sigma^t [d_s W_{\varepsilon,t-s}]\tilde F(s,\zeta(s),\rho(s),x^Q_s,\varepsilon)ds, \quad t \ge \sigma, \end{aligned}$$

where

$$\text{(14)}\quad \begin{aligned} &\text{a)}\ \tilde{F}(t,\zeta,\rho,\varphi,\varepsilon) = F[t,\rho(\varphi_{1\varepsilon}\cos\zeta + \varphi_{2\varepsilon}\sin\zeta) + \varphi] \\ &\text{b)}\ Z(t,\zeta,\rho,\varphi,\varepsilon) = \frac{1}{\rho}[-\psi_{1\varepsilon}(0)\sin\zeta+\psi_{2\varepsilon}(0)\cos\zeta]\tilde{F}(t,\zeta,\rho,\varphi,\varepsilon) \\ &\text{c)}\ R(t,\zeta,\rho,\varphi,\varepsilon) = \nu(\varepsilon)\rho+[\psi_{1\varepsilon}(0)\cos\zeta+\psi_{2\varepsilon}(0)\sin\zeta]\tilde{F}(t,\zeta,\rho,\varphi,\varepsilon). \end{aligned}$$

Suppose that the functions $\tilde{F}$, Z, R are almost periodic in t uniformly with respect to the other variables [$F(t,\varphi)$ in (10) almost periodic in t uniformly with respect to φ will imply this] and suppose that

$$\text{(15)}\quad \begin{aligned} &\text{a)}\ Z_0(\rho,\varepsilon) \overset{\text{def}}{=} \lim_{T\to\infty} \frac{1}{T}\int_0^T Z(t+s,\ \zeta+s,\ \rho,0,\varepsilon)ds \\ &\text{b)}\ R_0(\rho,\varepsilon) \overset{\text{def}}{=} \lim_{T\to\infty} \frac{1}{T}\int_0^T R(t+s,\ \zeta+s,\ \rho,0,\varepsilon)ds \quad ; \end{aligned}$$

that is, the mean values of Z, R are independent of t, ζ. Notice that these mean values are computed slightly differently then in ordinary differential equations. As in [ ], we have put $x_t^Q = 0$ and this is the basic fact that allows the theory to go through in a simple way. On the other hand, it makes some estimates more delicate as we shall see below.

Following the same type of reasoning as in ordinary differential equations (see [6] or [19]), there is a transformation of variables

$$\text{(16)}\qquad \zeta \to \zeta + \varepsilon u(t,\zeta,\rho,\varepsilon), \quad \rho \to \rho + \varepsilon v(t,\zeta,\rho,\varepsilon)$$

such that system (13) is equivalent to the system

$$
(17)\qquad
\begin{aligned}
&\text{a)}\quad \dot{\zeta} = \omega(\varepsilon) + \varepsilon\, Z_o(\rho,\varepsilon) + \varepsilon\, Z_1(t,\zeta,\rho,x_t^Q,\varepsilon)\\
&\text{b)}\quad \dot{\rho} = \varepsilon\, R_o(\rho,\varepsilon) + \varepsilon\, R_1(t,\zeta,\rho,x_t^Q,\varepsilon)\\
&\text{c)}\quad x_t^Q = T_\varepsilon(t-\sigma)x_\sigma^Q + \varepsilon\int_\sigma^t [d_s W^Q_{\varepsilon,t-s}]F_1(s,\zeta(s),\rho(s),x_s^Q,\varepsilon)ds
\end{aligned}
$$

where $F_1(t,\zeta,\rho,\varphi,\varepsilon) = \tilde{F}(t,\zeta+\varepsilon u,\rho+\varepsilon v,\varphi,\varepsilon)$, the functions Z_1, R_1 have the same smoothness properties as Z, R, are almost periodic in t uniformly with respect to the other variables, periodic in ζ of period 2π, and the functions $Z_1(t,\zeta,\rho,0,\varepsilon)$, $R_1(t,\zeta,\rho,0,\varepsilon)$ as well as their lipschitz constants with respect to ζ,ρ approach zero as $\varepsilon \to 0$.

Equations of type (17) can arise from system (9) without the severe restrictions made above on the characteristic equation (8). In fact, there could be any number of roots of (8) with zero real parts for $\varepsilon = 0$. The main part of the assumption that we have used is the dependence of the roots on ε near $\varepsilon = 0$. In this case, various transformations on (11b) yield equation of the form (17) with ζ,ρ vectors of not necessarily the same dimension. Also, some roots (a finite number) of (8) could have positive real parts for $\varepsilon = 0$. This adds an extra equation to (17) which can be easily discussed. For the sake of generality in the applications, we will assume that ζ,ρ are vectors of dimension p,q, respectively, and the functions in (17) are 2π-periodic in the components of the vector $\zeta = (\zeta_1,\ldots,\zeta_p)$.

If $\alpha : R \times R^p \to R^q$, $\beta : R \times R^p \to C$ are given functions, we say that the set

(18) $$S(\alpha,\beta) = \{(t,\zeta,\rho,\varphi) : \rho = \alpha(t,\zeta),\ \varphi = \beta(t,\zeta),\ t \in R,\ \zeta \in R^p\}$$

is an <u>integral</u> <u>manifold</u> of (17) if for every θ in R^p, σ in R and $\zeta(t) = \zeta(t,\sigma,\theta)$, $\zeta(\sigma,\sigma,\theta) = \theta$, the solution of (11a) with ρ, x_t^Q replaced by $\alpha(t,\zeta)$, $\beta(t,\zeta)$, respectively, it follows that the triple $\zeta(t)$, $\rho(t) = \alpha(t,\zeta(t))$, $x_t^Q = \beta(t,\zeta(t))$ is a solution of (17).

<u>THEOREM 1. Suppose $W^Q_{\varepsilon,t}$, $T_\varepsilon(t)\varphi^Q$ satisfy (12) and there is a ρ_o such that $R_o(\rho_o,0) = 0$ and the eigenvalues of $\partial R_o(\rho_o,0)/\partial\rho$ have nonzero real parts. Then there is an $\varepsilon_o > 0$ and functions $\alpha_\varepsilon : R \times R^p \to R^q$, $\beta_\varepsilon : R \times R^p \to C$, $\alpha_\varepsilon(t,\zeta)$, $\beta_\varepsilon(t,\zeta)$ continuous in t,ζ,ε for $t \in R$, $\zeta \in R^p$, $0 \leq \varepsilon \leq \varepsilon_o$, almost periodic in t uniformly with respect to ζ, periodic in the components of ζ of period 2π, $\alpha_o = \rho_o$, $\beta_o = 0$ such that $S(\alpha_\varepsilon,\beta_\varepsilon)$ in (18) is an integral manifold of (17) for $0 \leq \varepsilon \leq \varepsilon_o$. Furthermore, if $\gamma_\varepsilon = (\alpha_\varepsilon,\beta_\varepsilon)$, then $\partial\gamma_\varepsilon(t,\zeta)/\partial t^{\beta_o}\partial\zeta_1^{\beta_1}\cdots\partial\zeta_p^{\beta_p}$ exists and is continuous for $\beta_o \leq k$, $\beta_o+\beta_1+\cdots+\beta_p \leq k+\ell$ if the functions in (17) have k lipschitz continuous derivatives with respect to t and $(k+\ell)$ lipschitz continuous derivatives with respect to (ζ,ρ,φ^Q). Finally, the manifold $S(\alpha_\varepsilon,\beta_\varepsilon)$ is asymptotically stable* if the matrix $\partial R_o(\rho_o)/\partial\rho$</u>

*The stability here is the same sense as in Section V.1; namely C^1 perturbations in the initial data.

<u>has all eigenvalues with negative real parts and unstable if there is one eigenvalue with a positive real part.</u>

<u>Sketch of the proof</u>: We only give the main elements of the proof of Theorem 2 since it is so similar to the usual ones in the theory of ordinary differential equations. Also, to avoid so many formulas, we assume all eigenvalues of $E \overset{\mathrm{def}}{=} \partial R_o(\rho_o, 0)/\partial\rho$ have negative real parts and $|\exp Et| \leq K\exp(-ct)$, $t \geq 0$. Letting $\rho \to \rho_o + \rho$, the equations (17) become

$$
\begin{aligned}
&\text{a)} \quad \dot{\zeta} = \omega_2(\varepsilon) + \varepsilon\, \bar{Z}_1(t,\zeta,\rho,x_t^Q,\varepsilon) \\
(19) \qquad &\text{b)} \quad \dot{\rho} = \varepsilon E\rho + \varepsilon\, \bar{R}_1(t,\zeta,\rho,x_t^Q,\varepsilon) \\
&\text{c)} \quad x_t^Q = T_\varepsilon(t-\sigma)x_\sigma^Q + \varepsilon \int_\sigma^t [d_s W^Q_{\varepsilon,t-s}]\bar{F}_1(s,\zeta(s),\rho(s),x_s^Q,\varepsilon)ds
\end{aligned}
$$

where $\omega_2(0) = \omega_o$, $\bar{F}_1(t,\zeta,\rho,\varphi,\varepsilon) = F_1(t,\zeta,\rho_o+\rho,\varphi,\varepsilon)$ and $\bar{Z}_1, \bar{R}_1$ satisfy the following properties. For any given $r > 0$, $\varepsilon_1 > 0$, $H > 0$, there exist a constant $K_1 > 0$ and a continuous nondecreasing function $\nu(\varepsilon)$, $0 \leq \varepsilon \leq \varepsilon_1$ such that $\nu(0) = 0$ and

$$
\begin{gathered}
|\bar{Z}_1(t,\zeta,0,0,\varepsilon)| \leq \nu(\varepsilon), \quad |\bar{R}_1(t,\zeta,0,0,\varepsilon)| \leq \nu(\varepsilon), \\
|\bar{F}_1(t,\zeta,0,0,\varepsilon)| \leq K_1, \\
|\bar{Z}_1(t,\zeta,\rho,\varphi,\varepsilon) - \bar{Z}_1(t,\zeta,\rho_1,\varphi_1,\varepsilon)| \leq \\
\leq [\nu(\varepsilon) + k_1 H][|\zeta-\zeta_1| + |\rho-\rho_1|] + K_1\|\varphi-\varphi_1\|,
\end{gathered}
$$

(20)

$$|\bar{R}_1(t,\zeta,\rho,\varphi,\varepsilon)-\bar{R}_1(t,\zeta_1,\rho_1,\varphi_1,\varepsilon)| \leqq$$

$$\leqq [\nu(\varepsilon)+K_1H]|\zeta-\zeta_1|+[\nu(r)+\nu(\varepsilon)+K_1H]|\rho-\rho_1|+K_1\|\varphi-\varphi_1\|,$$

$$|\bar{F}_1(t,\zeta,\rho,\varphi,\varepsilon)-\bar{F}_1(t,\zeta,\rho,\varphi,\varepsilon)| \leqq K_1[|\zeta-\zeta_1|+|\rho-\rho_1|+\|\varphi-\varphi_1\|]$$

for $t \in R$, ζ, $\zeta_1 \in R^p$, $\rho,\rho_1 \in R^q$, $|\rho|,|\rho_1| \leqq r$, $\varphi,\varphi_1 \in C$, $\|\varphi\|,\|\varphi_1\| \leqq H$ and $0 \leqq \varepsilon \leqq \varepsilon_1$. Of course, all functions are almost periodic in t and periodic in ζ.

Let $\mathscr{C}_1(\Delta_1,D_1)$ be the class of continuous functions $\alpha : R \times R^p \to R^q$ which are bounded by D_1 and have lipschitz constant Δ_1 with respect to the second variable. Similarly, let $\mathscr{C}_2(\Delta_2,D_2)$ be the class of $\beta : R \times R^p \to C$. We introduce the uniform norm in these spaces and designate the norm by $\|\alpha\|_{\mathscr{C}_1}$, $\|\beta\|_{\mathscr{C}_2}$, $\alpha \in \mathscr{C}_1(\Delta_1,D_1)$, $\beta \in \mathscr{C}_2(\Delta_2,D_2)$.

It is convenient to introduce some notation. For $\alpha \in \mathscr{C}_1(\Delta_1,D_1)$, $\beta \in \mathscr{C}_2(\Delta_2,D_2)$, we abbreviate the collection $(t,\zeta,\alpha(t,\zeta),\beta(t,\zeta),\varepsilon)$ by $(t,\zeta,\alpha,\beta,\varepsilon)$. Also let

$$a = (\nu(\varepsilon)+K_1D_2)(1+\Delta_1)+K_1\Delta_2$$

(21)

$$b = \nu(\varepsilon) + K_1D_2$$

and then it follows from (14) that

$$
(22)\quad
\begin{aligned}
|\bar{R}_1(t,\zeta,\alpha,\beta,\varepsilon)| &\leq \nu(\varepsilon) + [b+\nu(D_1)]D_1 + K_1D_2 \\
|\bar{F}_1(t,\zeta,\alpha,\beta,\varepsilon)| &\leq K_1(1+D_1+D_2) \\
|\bar{Z}_1(t,\zeta,\alpha,\beta,\varepsilon)-\bar{Z}_1(t,\zeta_1,\alpha_1,\beta_1,\varepsilon)| &\leq a|\zeta-\zeta_1| + b\|\alpha-\alpha_1\|_{\mathcal{C}_1} + K_1\|\beta-\beta_1\|_{\mathcal{C}_2} \\
|\bar{R}_1(t,\zeta,\alpha,\beta,\varepsilon)-\bar{R}_1(t,\zeta_1,\alpha_1,\beta_1,\varepsilon)| &\leq [a+\nu(D_1)\Delta_1]|\zeta-\zeta_1| + [b+\nu(D_1)]\|\alpha-\alpha_1\|_{\mathcal{C}_1} + \\
&\quad + K_1\|\beta-\beta_1\|_{\mathcal{C}_2} \\
|\bar{F}_1(t,\zeta,\alpha,\beta,\varepsilon)-\bar{F}_1(t,\zeta_1,\alpha_1,\beta_1,\varepsilon)| &\leq K_1\{(1+\Delta_1+\Delta_2)|\zeta-\zeta_1| + \|\alpha_1-\alpha_2\|_{\mathcal{C}_1} + \|\beta_1-\beta_2\|_{\mathcal{C}_2}\}
\end{aligned}
$$

With the constants defined as above choose $\varepsilon_1 > 0$ and continuous $\Delta_j(\varepsilon), D_j(\varepsilon)$, $0 \leq \varepsilon \leq \varepsilon_1$, $\Delta_j(\varepsilon)$, $D_j(\varepsilon) \to 0$ as $\varepsilon \to 0$ such that, for $0 \leq \varepsilon \leq \varepsilon_1$,

$$
(23)\quad
\begin{aligned}
&\nu(\varepsilon) + [b+\nu(D_1)]D_1 + K_1D_2 \leq D_1c/K ; \\
&\varepsilon K_1(1+D_1+D_2) \leq D_2\sqrt{2c}/K ; \\
&a+b\Delta_1+2\nu(D_1)\Delta_1+K_1\Delta_2 \leq \Delta_1 c/2K ; \\
&2\sqrt{3}\,\varepsilon KK_1/\sqrt{c} \leq \Delta_2 \leq \min\,[KK_1/4c,\ 1/4] ; \\
&c-a > c/2;\ c-\varepsilon a > c/2 ;\ \Delta_1+K < K/4;\ 2b^2+1 \leq 4 \\
&\Delta_1 ab+K(b+\nu(D_1)) < c/4;\ 1+8K_1^2\varepsilon^2/3c^2 \leq 4
\end{aligned}
$$

Let $\mathcal{C}(\Delta,D) = \mathcal{C}_1(\Delta_1,D_1) \times \mathcal{C}_2(\Delta_2,D_2)$ and for any γ in $\mathcal{C}(\Delta,D)$, $\gamma = (\alpha,\beta)$, define $\|\gamma\| = \|\alpha\|_{\mathcal{C}_1} + KK_1\|\beta\|_{\mathcal{C}_2}/4$. For any $\gamma = (\alpha,\beta)$ in $\mathcal{C}(\Delta,D)$, let $\zeta(t,\sigma,\theta,\gamma)$, $\zeta(\sigma,\sigma,\theta,\gamma) = \theta$, be the solution of

(19a) with (ρ, x_t) replaced by $\gamma(t,\zeta)$ and define a transformation $T\gamma$ by

$$
(24)\qquad
\begin{aligned}
&\text{a)}\quad T\gamma = (T_1\gamma, T_2\gamma) \\
&\text{b)}\quad (T_1\gamma)(\sigma,\theta) = \varepsilon \int_{-\infty}^{0} e^{-\varepsilon E u}\, \bar{R}_1(u+\sigma, \zeta(u+\sigma,\sigma,\theta,\gamma),\gamma,\varepsilon)\,du \\
&\text{c)}\quad (T_2\gamma)(\sigma,\theta) = \varepsilon \int_{-\infty}^{0} [d_u W^Q_{\varepsilon,-u}]\bar{F}_1(u+\sigma,\sigma,\theta,\gamma),\gamma,\varepsilon)\,du\,.
\end{aligned}
$$

We shall show that this equation has a unique solution in $\mathcal{C}(\Delta,D)$ for $0 \leq \varepsilon \leq \varepsilon_0$. This will prove the existence of an integral manifold. From (12),(22) and (23), we have $\|T_1\gamma\|_{\mathcal{C}_1} \leq D_1$, $\|T_2\gamma\|_{\mathcal{C}_2} \leq D_2$.

From the Lipschitz constant of $\bar{Z}_1$ in (22) and (19a), we obtain

$$
|\zeta(u+\sigma,\sigma,\theta_1,\gamma_1)-\zeta(u+\sigma,\sigma,\theta_2,\gamma_2)| \leq
$$

$$
\leq e^{-\varepsilon a u}|\theta_1-\theta_2| + (e^{-\varepsilon a u}-1)[b\|\alpha_1-\alpha_2\|_{\mathcal{C}_1} + \frac{K_1}{a}\|\beta_1-\beta_2\|_{\mathcal{C}_2}]
$$

for $-\infty < u \leq 0$.

Using this fact and the estimates (22),(23), we have

$$
|(T_1\gamma_1)(\sigma,\theta_1)-(T_1\gamma_2)(\sigma,\theta_2)| \leq \Delta_1|\theta_1-\theta_2| + \|\gamma_1-\gamma_2\|_{\mathcal{C}}/4,
$$

$$
|(T_2\gamma_1)(\sigma,\theta_1)-(T_2\gamma_2)(\sigma,\theta_2)| \leq \Delta_2|\theta_1-\theta_2| + \|\gamma_1-\gamma_2\|_{\mathcal{C}}/KK_1
$$

for $0 \leq \varepsilon \leq \varepsilon_1$. This implies $T : \mathcal{C}(\Delta,D) \to \mathcal{C}(\Delta,D)$ and is a contraction since $\|T\gamma_1 - T\gamma_2\| \leq \|\gamma_1 - \gamma_2\|_{\mathscr{C}}/4$ for $0 \leq \varepsilon \leq \varepsilon_1$. This completes the proof of the existence of an integral manifold and also shows that the integral manifold is lipschitzian in ζ.

To obtain the smoothness properties of the manifold $S(\alpha_\varepsilon, \beta_\varepsilon)$ one proceeds in exactly the same manner as above except making use of a different class of functions $\mathscr{C}_1, \mathscr{C}_2$. For example, to show that $\alpha_\varepsilon, \beta_\varepsilon$ have continuous first derivatives with respect to ζ if the functions in (17) have continuous first derivatives with respect to ζ, ρ, φ^Q one defines $\mathscr{C}_1(\Delta_1, D_1)$ to be a class of functions $\alpha : R \times R^p \to R^q$ such that $|\alpha(t,\zeta)| \leq D_1$, $|\partial\alpha(t,\zeta)/\partial\zeta| \leq D_2$ for all t, ζ. The class $\mathscr{C}_2(\Delta_2, D_2)$ is defined in the same manner. Using the same definition of T as in (24), one shows by a proper choice of $\Delta_j(\varepsilon), D_j(\varepsilon) \to 0$ as $\varepsilon \to 0$ that T has a unique fixed point in $\mathscr{C}_1 \times \mathscr{C}_2$. The other derivatives are analyzed in exactly the same manner.

We will not prove the stability result since it again involves complicated estimates of the above type and the reader can easily supply the details by following the standard procedure in the method of integral manifolds in ordinary differential equations together with the lemma 1 of section V.1.

It is clear that Theorem 2 has an interpretation in the original equation (9) at the beginning of this section. For simplicity, we state an important corollary for the special case when F in (9) is independent of t. The notations are the ones given at the beginning of this section.

<u>COROLLARY 1</u>. Suppose $F(t,\varphi) = \bar{F}(\varphi)$ for all t and let

$$(25) \quad G(\rho) = \nu(0)\rho + \frac{1}{2\pi}\int_0^{2\pi}[\psi_{1o}(0)\cos s+\psi_{2o}(0)\sin s]\bar{F}[\rho(\varphi_{1o}\cos s+\varphi_{2o}\sin s)]ds.$$

If there is a ρ_o such that $G(\rho_o) \neq 0$, $dG(\rho_o)/d\rho \neq 0$, then there is an $\varepsilon_1 > 0$, a constant $\omega^*(\varepsilon)$ and a function $x^*(t,\varepsilon)$, continuous in t,ε and having a continuous derivative with respect to t, $-\infty < t < \infty$, $0 \leq \varepsilon \leq \varepsilon_1$,

$$x^*(t,0) = \rho_o[\varphi_{1o}(0)\cos\omega_o t+\varphi_{2o}(0)\sin\omega_o t],$$

$\omega^*(0) = \omega_o$, $x^*(t+\omega^*(\varepsilon),\varepsilon) = x^*(t,\varepsilon)$ such that $x^*(t,\varepsilon)$ satisfies (9) and since it is differentiable satisfies (10). The periodic solution $x^*(\cdot,\varepsilon)$ is orbitally asymptotically stable* if $dG(\rho_o)/d\rho < 0$ and unstable if $dG(\rho_o)/d\rho > 0$.

As an example, consider the equation

$$(26) \quad \dot{x}(t) = \alpha\dot{x}(t-r)-\beta x(t) - \alpha\gamma x(t-r) + \varepsilon F(x_t)$$

where $\varepsilon \geq 0$, $r > 0$, $\gamma > \beta > 0$, $\alpha = \alpha(\varepsilon) = \alpha_o(1+\varepsilon)$, where α_o is

---

* A periodic solution $x(t)$ of (10) is called asymptotically orbitally stable if the orbit, $U_t x_t$, of x in C is asymptotically stable in the sense of C^1 perturbations.

the unique real number in $(0,1)$ such that the characteristic equation III(35) for the linear system III(33) has two purely imaginary roots $\pm i\omega_o$, $\omega_o > 0$, and the remaining roots have real parts $< -\delta < 0$. Brayton [13] has shown that such an α_o exists. This implies that there is an $\varepsilon_1 > 0$ such that the equation

$$\lambda - \alpha(\varepsilon)\lambda e^{-\lambda r} + \beta + \alpha(\varepsilon)\gamma e^{-\lambda r} = 0 \tag{27}$$

has two simple roots $\varepsilon\nu(\varepsilon) \pm i\omega(\varepsilon)$, $\omega(0) = \omega_o$, $\nu(\varepsilon),\omega(\varepsilon)$ continuous in $0 \leq \varepsilon \leq \varepsilon_1$, and the remaining roots have real parts $< -\delta < 0$ for $0 \leq \varepsilon \leq \varepsilon_1$. We are writing the equation (26) in differential form for simplicity in notation but it always understood that solutions are defined by means of the integrated form of this equation.

In the discussion of this example, we use the notations introduced at the end of section III. A straightforward computation on the characteristic equation (27) shows that

$$2\nu(0) = \frac{\beta C + \omega_o D}{C^2 + D^2} > 0 , \tag{28}$$

where C,D are defined in III(40). Using the formula for $\Psi(0)$ in III(40), it is easily seen that the function $G(\rho)$ in (25) is given by

$$G(\rho) = \frac{\beta C+\omega_0 D}{C^2+D^2} G^*(\rho)$$

(29)

$$G^*(\rho) = \frac{\rho}{2} + \frac{1}{\beta C+\omega_0 D} \cdot \frac{1}{2\pi} \int_0^{2\pi} (D\cos s+C\sin s)F(\rho(\varphi_{10}\cos s+\varphi_{20}\sin s))ds$$

From Corollary 1, we can now state the following result: equation (26) will have an asymptotically orbitally stable periodic solution if there exists a ρ_0 such that $G^*(\rho_0) = 0$, $dG^*(\rho_0)/d\rho < 0$ and an unstable one if $G^*(\rho_0) = 0$, $dG^*(\rho_0)/d\rho > 0$.

In the particular case where $F(x_t) = h(x(t))$ relation (29) yields

$$G^*(\rho) = \frac{\rho}{2} + \frac{C}{\beta C+\omega_0 D} \cdot \frac{1}{2\pi} \int_0^{2\pi} h(\rho\cos \zeta)\cos \zeta d\zeta$$

and the criterion for existence of a periodic solution is the same as the one obtained by Brayton [13]. However, we can also say something about the stability of the solution. In the particular case, when $h(x) = -x^3$, an easy computation yields $G^*(\rho) = (\rho/2)[1-3C\rho^2/4(\beta C+\omega_0 D)]$ and $G^*(\rho_0) = 0$, $dG^*(\rho_0)/d\rho = -1$ for $\rho_0^2 = 4(\beta C+\omega_0 D)/3C$. Thus, the equation has an asymptotically orbitally stable periodic solution.

As another illustration, suppose $F(x_t) = -x^3(t-s)$, $0 \leq s \leq r$. Then

$$(2/\rho)G^*(\rho) = 1 - \frac{3\rho^2}{4(\beta C + \omega_o D)}[C \cos \omega_o s - D \sin \omega_o s] .$$

As before, if $C \cos \omega_o s - D \sin \omega_o s > 0$, the we obtain an asymptotically orbitally stable periodic solution. To find the limitations on s for which this inequality remains valid is difficult since ω_o depends upon all parameters in the linear differential equation III(33).

This example illustrates the application of the general theory to autonomous systems, but it is clear that Theorem 2 is equally applicable to nonautonomous equation.

REFERENCES

[1] Bellman, R., and K. Cooke, Differential Difference Equations, Academic Press, New York, 1963.

[2] Halanay, A., Teoria Calitativa Equatilor Diferentiale, Editura Academici Republici Populare Romina, Bucharest, 1963. Translated by Academic Press, 1966.

[3] Shimanov, N., "On the vibration theory of quasilinear systems with lags, Prikl. Mat. Mek., 23(1959), 836-844.

[4] Shimanov, N., "On the theory of linear differential equations with retardations, Differentzialnie Uravneniya, 1(1965), 102-116.

[5] Hale, J.K., "Linear functional differential equations with constant coefficients",Contri. Dif. Eqs., II(1963),291-319.

[6] Hale, J.K., "Averaging methods for differential equations with retarded arguments and a small parameter, J. Dif. Eqs. 2(1966), 57-73.

[7] Hale, J.K., and C. Perelló, "The neighborhood of a singular point of functional differential equations, Contri. Dif. Eqs., III(1965), 351-375.

[8] Driver, R., "A functional-differential equation arising is a two body problem of electrodynamics", International Symposium of Nonlinear Differential Equations and Nonlinear Mechanics, 474-484, Academic Press, 1963.

[9] Driver, R., "Existence and continuous dependence of solutions of a neutral functional-differential equation", Arch. Rat. Mech. Ann., 19(1965), 149-166.

[10] Riesz, F., and B.Sz. Nagy, Functional Analysis, Frederick Ungar Publ., New York, 1952.

[11] Hille, E., and R. Phillips, Functional Analysis and Semi-Groups, Amer. Math. Soc. Colloq. Publ., 31(1957).

[12] Taylor, A.E., Introduction to Functional Analysis, John Wiley, New York, 1964.

[13] Brayton, R., "Bifurcation of periodic solutions in a nonlinear difference-differential equation", IBM Research paper, RC-1427, June, 1965.

[14] Levitan, B., Almost Periodic Functions, (Russian), Moscow, 1953.

[15] Coddington, E., and N. Levinson, Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.

[16] Miranker, W.L., "Existence, uniqueness, and stability of solutions of systems of nonlinear difference-differential equations", IBM Research Report RC-322, 1960.

[17] Miranker, W.L., "The wave equation with a nonlinear interface condition", IBM Journal Res. and Dev. 5(1961), 2-24.

[18] Marchenko, Y.I. and V.P. Rubanik, "On mutual synchronization of molecular generators", Izv. Vis. Uch. Zav. Radiophysica, 8(1965), 679-687.

[19] Hale, J.K., Oscillations in Nonlinear Systems, McGraw-Hill, New York, 1963.

Center for Dynamical Systems
Division of Applied Mathematics
Brown University
Providence, Rhode Island

PAPER [13]

To appear in the American Journal of Physics

A Note on Contact Transformations

by

K. R. Meyer*

Abstract

Not all contact transformations are of the form $p = W_q(q,P)$ and $Q = W_P(q,P)$ but this note shows that after a linear change of variables any contact transformation can be written in this form.

*This Research was supported by the National Aeronautics and Space Administration, Huntsville, under Contract No. NAS 8-11264.

It is well known that a transformation of the form

$$p_i = \frac{\partial W}{\partial q_i}(q,P), \quad Q_i = \frac{\partial W}{\partial P_i}(q,P) \tag{1}$$

defines a contact transformation from the q,p variables to the Q,P variables where q,p,Q,P are n-vectors, W is a scalar function with continuous second partial derivatives with respect to all arguments and subscripts denote components of the vectors. See for example [1], and [2]. It is not always true that any contact transformation can be written in the form 1) or even in one of the other three common variations of 1). This fact is pointed out in [2] and [3] and the author recommends [2] as a careful and readable source on contact transformations (see in particular, page 69-70 of [2]). This note will show, however, that <u>any contact transformation can be written as a composition of a linear orthogonal contact transformation and a contact transformation of the form 1)</u>. That is to say, given any contact transformation one can first make a change of variables that is linear, orthogonal and preserves Hamiltonian form and then write the transformation in the form 1). The above is to be taken as a local statement, that is, the above statement holds only in a sufficiently small neighborhood of a point. Also we shall assume that all functions are sufficiently differentiable that the indicated derivatives are continuous and that the implicit function theorem can be applied. The assumption that all functions considered have continuous second partial derivatives with respect to all arguments will suffice.

To avoid confusion a contact transformation will be taken in the sense of Whittaker [1], page 293. That is:

Definition: A transformation

(2) $$\mathcal{F}: \quad Q = \varphi(q,p), \quad P = \psi(q,p)$$

where q,p,Q,P are n-vectors and ψ and φ are n-vector valued used functions of q and p will be called a contact transformation if there exists a scalar valued function $S(q,p)$ such that

(3) $$dS(q,p) = \sum_{i=1}^{n} \{p_i dq_i + \varphi_i(q,p) d\psi_i(q,p)\}.$$

Observe that 3) is often written $dS = \sum_{i=1}^{n} \{p_i dq_i + Q_i dP_i\}$ and that this short notation is the cause of some of the confusion in the literature. The equality 3) states that S must be considered as a function of p and q only. Indeed the whole question of when a contact transformation 2) can be written in the form 1) rests on the question of when can S be written as a function of q,P. If the second equation in 2) can be solved for p in terms of P and q and the result substituted into S we would have the desired function W. But when can we solve the second equation in 2) for p in terms of q, and P? If the sub-Jacobian $\det\{\frac{\partial\psi_i}{\partial p_j}\}$ is non zero then we can solve this equation, but there is no reason to suppose that it is nonzero. At this point a result in [3] can be used to straighten things out.

The formal proof is as follows. Let 2) or $\mathcal{F}$ be a given contact transformation. Without loss of generality we can assume that $\mathcal{F}$ takes the origin into the origin since otherwise we would shift the origin by a translation. Let T be the Jacobian matrix of $\mathcal{F}$ evaluated at the origin, i.e.

$$T = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where $A = \{\frac{\partial\varphi_i}{\partial q_j}(0,0)\}$, $B = \{\frac{\partial\varphi_i}{\partial p_j}(0,0)\}$, $C = \{\frac{\partial\psi_i}{\partial q_j}(0,0)\}$ and $D = \{\frac{\partial\psi_i}{\partial p_j}(0,0)\}$.

Now by a result in [3], page 44 there exists nonsingular contact matrices O and R where O is orthogonal and R is positive definite symmetric such that T = RO. This result for contact matrices is the analog of the well known result in 3-dimensions that says that any matrix of a linear transformation is the product of a pure rotation (or rotation and reflection) and a pure dilation. It should be remarked that in [3] as in many other references a contact matrix is called symplectic and is sometimes given a different but equivalent definition (see [2]).

Let $\mathcal{O}$ be the transformation whose representation is the matrix O. Define a new transformation $\mathcal{G}$ by $\mathcal{G} = \mathcal{F} \circ \mathcal{O}^{-1}$

and so $\mathcal{F} = \mathcal{G} \circ \Theta$. Observe that we have "factored" the transformation $\mathcal{F}$ into two operations: first apply Θ and then $\mathcal{G} = \mathcal{F} \circ \Theta^{-1}$. Another way of looking at $\mathcal{G}$ is that we have changed coordinates by the linear transformation Θ and now $\mathcal{F}$ has the form $\mathcal{G}$ in the new coordinates. We now want to show that $\mathcal{G}$ can be written in the form 1).

$\mathcal{G}$ is a contact transformation since it is the composition of two contact transformations and moreover its Jacobian matrix at the origin is $TO^{-1} = (RO)O^{-1} = R$. Thus if $\mathcal{G}$ is given by $Q = a(q',p')$, $P = b(q',p')$ and $R = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}$ where $A' = \{\frac{\partial a_i}{\partial q'_j}(0,0)\}$ etc.

Now R is positive definite and symmetric and so by Sylvester's criterion [4], page 306, or [5] page 94 each principal subdeterminant of R is positive and so in particular $D' = \{\frac{\partial b_i}{\partial p'_i}(0,0)\}$ is nonsingular.

Thus, we can solve the equation $P = b(q',p')$ for p' to obtain $p' = h(q',P)$.

Since $\mathcal{G}$ is a contact transformation there exists a generating function $S'(q',p')$ such that

$$(4) \qquad dS'(q',p') = \sum_{i=1}^{n} \{p'_i dq'_i + b_i(q',p')da_i(q',p')\}.$$

Let $W(q',P) = S'(q',h(q',P))$ now

$$(5) \qquad dW(q',P) = \sum_{i=1}^{n} \{\frac{\partial W}{\partial q'_i} dq'_i + \frac{\partial W}{\partial P_i} dP_i\}$$

but $dW = dS$ at corresponding points and so

$$(6) \qquad dW(q',P) = \sum_{i=1}^{n} \{p_i' dq_i' + b_i(q',p')dP_i\}$$

where in 6) $p' = h(q',P)$.

Now since $dP_i = \sum_{j=1}^{n} \{\frac{\partial b_i}{\partial q_j'} dq_j' + \frac{\partial b_i}{\partial p_j'} dp_j'\}$ and since $\{\frac{\partial b_i}{\partial p_j}\}$ is nonsingular the differentials $dq_1', \ldots, dq_n', dP_1, \ldots, dP_n$ are linearly independent and so we can equate coefficients in 5) and 6) to obtain

$$(7) \qquad p_i' = \frac{\partial W}{\partial q_i'}(q',P) \quad \text{and} \quad Q_i = \frac{\partial W}{\partial P_i}(q',P).$$

Therefore $\mathcal{G}$ is of the form 1).

Observe that we can obtain one of the other common variations of 1) when any one of the other sub-Jacobian matrices is nonsingular. The procedure we have used gives that A' is nonsingular so this gives one variant. By changing variables again with the linear orthogonal contact matrix $\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ then the Jacobian of the new $\mathcal{G}$ is of the form $\begin{pmatrix} -B' & A' \\ -D' & C' \end{pmatrix}$ and so now the upper right and lower left sub-Jacobian matrices are nonsingular and by the same procedure you get the other two variants.

References

[1] Whittaker, E. T., A Treatise on the Analytic Dynamics of Particles and Rigid Bodies, Cambridge University Press, 4th edition, 1964.

[2] Pollard, H., Mathematical Introduction to Celestial Mechanics, Prentice Hall, 1966.

[3] Wintner, A., The Analytic Foundations of Celestial Mechanics, Princeton University Press, 1947.

[4] Gantmacher, F. R., The Theory of Matrices, Vol. 1, Chelsea Publishing Company, 1959, p. 306.

[5] Perlis, Theory of Matrices, Addison-Wesley Publishing Company, 1958, p. 94.

PAPER [14]

To appear in the Proceedings of the
American Mathematical Society

ON COMPUTING THE INDEX IN HIGHER DIMENSIONS

by

K. R. Meyer*

Poincaré [1] gave a simple geometric procedure for computing the index of a critical point of a vector field in the plane. If the trajectories induced by the field are tangent to a small circle at a finite number of points and i and e denote the number of internal and external tangents then Poincaré gave the following formula for the index

$$I = 1 + \frac{i-e}{2}$$

A generalization of this formula to higher dimensions will be discussed in this paper.

Now let ν be a smooth vector field defined in an open subset V of R^n, $n \geq 2$, where V contains the unit $n-1$ sphere S^{n-1} in its interior. That is to say ν is a smooth map from V into R^n.

*This research was supported by the National Aeronautics and Space Administration under Contract No. NAS8-11264.

Furthermore assume that $\|\nu(x)\| = 1$ for all $x \in S^{n-1}$ where $\|\ \|$ is the Euclidean norm. Fix a coordinate system in R^n and consider the n-frame $\Pi(x)$ at each $x \in V$ that is a parallel displacement of the original coordinate system. In this fixed coordinate system the point $(0,\ldots,0,1)$ will be called the north pole and other geographic terminology consistent with this convention will be used.

The degree of the map $\nu|S^{n-1} : S^{n-1} \to S^{n-1}$ is called the index of the vector field ν with respect to S^{n-1} and will be denoted by $I(\nu,S^{n-1})$. In terms of the framing Π the vector field ν has the form $\nu(x) = a_1(x)\Pi^1(x)+ \ldots +a_n(x)\Pi^n(x)$ and so

$$\nu|S^{n-1} : x \to (a_1(x),\ldots,a_n(x)).$$

A homotopy of the field ν will always be smooth and through fields ν_t such that $\|\nu_t(x)\| = 1$ for all $x \in S^{n-1}$. A homotopy of the framing will always be smooth and through rigid rotations about each point. Both the above operations leave the index unchanged.

Let η denote the unit outward normal vector field on S^{n-1}.

The following observation is due to M. M. Peixoto.

LEMMA: <u>There exists a smooth vector field μ that is ε-homotopic to ν such that the field μ is tangent to S^{n-1} only at a finite number of smooth closed connected n-2 submanifolds $M_1,\ldots,M_p$. These submanifolds are the boundaries of a finite number of smooth connected</u>

n-1 <u>submanifolds with boundary</u> $A_1,\dots,A_q$ <u>such that the field</u> μ <u>has a positive component in the direction of</u> η <u>at each interior point of</u> A_i, $i = 1,\dots,q$. <u>At each point of</u> $B = S^{n-1} - \bigcup_{i=1}^{q} A_i$ <u>the field</u> μ <u>has a negative component in the direction of</u> η.

Proof. Consider the map $g: S^{n-1} \to R$ defined by taking the component of ν in the direction of η. By applying Sard's theorem in the usual way one constructs μ from ν.

Henceforth it will be assumed that the vector field has been prepared in accordance with the above lemma. The M_i, $i = 1,\dots,p$ will be called manifolds of contact and any region where the field has a positive (negative) component in the direction of η will be called a region of egress (ingress). Thus the interior of A_i, $i = 1,\dots,q$, is a region of egress and B is a region of ingress.

Let the A_i, $i = 1,\dots,q$ be oriented as submanifolds of S^{n-1} and the M_i, $i = 1,\dots,p$ be oriented as boundaries of the appropriate A_j.

There exists a framing Σ^j of V that is smoothly homotopic to Π such that along M_j the last component of Σ^j is η. The field μ on M_j can be expressed as $b_1(x)\Sigma_1^j(x)+\dots+b_{n-1}(x)\Sigma_{n-1}^j(x)$. The degree of the map $h: M_j \to S^{n-2}$ defined by $h: x \to (b_1(x),\dots,b_{n-1}(x))$ will be called the index of μ with respect to M_j and will be denoted by $I(\mu,M_j)$. Clearly this index does not depend on the choice of Σ^j.

We can now state our main result.

Proposition: _The index of the field such that all of_ S^{n-1} _is a region of egress is_ +1. _In all other cases_

$$I(\mu,S^{n-1}) = (-1)^n + \sum_{j=1}^{p} I(\mu,M_j)$$

Proof. Clearly the theorem holds if all of S^{n-1} is a region of egress so we can assume that not all of S^{n-1} is a region of egress. The field μ can be deformed to the field μ' so that all the manifolds of contact of μ' lie north of the Tropic of Capricorn and so that the region south of the Tropic of Capricorn is a region of ingress. Moreover the deformation can be constructed so that there is a one to one correspondence between the manifolds of contact of μ and μ' and such that the corresponding indices are the same. The frame Π can be deformed to a frame Σ where Σ has the following properties (i) north of the Tropic of Capricorn the last component of Σ is η (ii) in the southern hemisphere Π and Σ agree (iii) between the equator and the Tropic of Capricorn the homotopy between Σ and Π can be accomplished by a rotation through an angle less than or equal to $3\pi/8$.

Now $\mu'(x) = b_1(x)\Sigma_1(x)+\ldots+b_n(x)\Sigma_n(x)$. The degree of the map $w\colon S^{n-1} \to S^{n-1}$ given by $w\colon x \to (b_1(x),\ldots,b_n(x))$ is $I(\mu,S^{n-1})$. Now w maps the manifolds of contact into the equator and the regions of egress into the northern hemisphere. Now we count the number of times the northern hemisphere is covered.

Let $F = M_1 \cup\ldots\cup M_{\ell_1}$ be the boundary of A_1, S^{n-2} the equator of S^{n-1} and N the northern hemisphere of S^{n-1}. From the following commutive diagram

$$\begin{tikzcd}
H_n(A_1,F) \arrow[r, "\partial"] \arrow[d, "w_*"] & H_{n-1}(F) \arrow[d, "w_*"] \\
H_n(S^{n-1},S^{n-2}) \arrow[r, leftrightarrow, "\partial"] & H_{n-1}(S^{n-2})
\end{tikzcd}$$

it follows that if $I(\mu',M_i) = k_i$, that is the generator of $H_{n-1}(M_i)$ is mapped by w_* onto k_i times the generator of $H_{n-1}(S^{n-2})$, then the generator of $H_n(A_1,F)$ is mapped by w_* onto $(k_1+\ldots+k_{\ell_1})$ times the generator of $H_n(S^{n-1},S^{n-2})$. In the above "the generator" is to be taken as the generator corresponding to the oriented manifold itself.

Now if μ_1 is the field obtained from μ' by changing the sign of the last component of μ' in the region A_1 then by the above

$$I(\mu',S^{n-1}) = I(\mu_1,S^2) + \sum_{i=1}^{\ell_1} I(\mu',M_i).$$

By repeating this process for each region of egress the theorem follows since in the last step

$$I(\mu',S^{n-1}) = I(\mu_q,S^{n-1}) + \sum_{i=1}^{q} I(\mu',M_i)$$

and since μ_q is a field such that all of S^{n-1} is a region of ingress and so $I(\mu_q, S^{n-1}) = (-1)^n$.

REMARK 1. One can see at once that this formula yields an effective geometric proceedure for computing the index in dimension 2 and 3. In dimension 2 it can readily be seen that this formula is essentially the same as that of Poincaré. For $n = 3$ the manifolds of contact are circles and so the formula can be applied to reduce the problem of computing the index on a 2 sphere to computing the index on several circles. Then one can apply either Poincare's formula of the above to compute the index on these circles.

REMARK 2. One can also use the above formula to compute the index in dimension 4. In this case the manifolds of contact are oriented 2-manifolds or spheres with handles. If the genus of a 2-manifold is g then manifold can be made into a sphere by making g cuts and adding 2q hemisphere. The cuts can be taken so as not to intersect any of the manifolds of contact. If a cut is in a region of egress (ingress) one can define a new field to be smooth and outward (inward) on the two hemispheres attached along the cuts. The index of the new field is increased by one for each cut in a region of egress and decreased by one for each cut in a region of ingress. Thus the problem of computing the index along a 2-manifold is reduced to computing the index along a 2-sphere and the above formula can be applied.

For example the index of the outward normal field 2-manifold embedded in R^3 is half the Euler characteristic. This is a well known result of Hopf [2].

REFERENCES

[1] H. Poincaré, Sur les courbes définies par une équation differentielle, Oeuvres, Vol. 1, Paris, 1892.

[2] H. Hopf, Vektorfelder in n-dimensionalen Mannigfaltigkeiten, Math. Annalen, 96, 1926.

PAPER [15]

To appear in the Journal of Differential Equations

PERIODIC SOLUTIONS OF DIFFERENTIAL EQUATIONS WITH TIME LAG CONTAINING A SMALL PARAMETER

Carlos Perelló *

* This research was supported in part by National Aeronautics and Space Administration under Grant No. NASA-NAS8-11264, in part by the United States Air Force through the Air Force Office of Scientific Research under Grant No. AF-AFOSR-693-65 and in part by the Instituto Nacional de la Investigación Cientifica, México, through a complementary grant.

PERIODIC SOLUTIONS OF DIFFERENTIAL EQUATIONS WITH TIME LAG CONTAINING A SMALL PARAMETER

By Carlos Perelló

Introduction.

In this paper we show that the method of Cesari and Hale for the determination of periodic solutions of ordinary differential equations can be extended to the case in which these equations contain a time lag.

An ordinary differential equation with time lag (also called functional differential equation) differs from those without lag in that the derivative of a solution function at a time t depends also on the values of this solution at times preceding t. We further restrict our equations by considering only time lags less than a fixed number r.

In the notation introduced by Hale [1] we consider equations of the form

$$(\alpha) \qquad \dot{x}(t) = F(t, x_t),$$

where F denotes a functional (real or complex) defined for each t and for the "segment of solution" x_t, of length r, preceeding t. Here x denotes an n-vector.

As a particular case we encounter the difference-differential equations

$$(\beta) \qquad \dot{x}(t) = f(t, x(t), x(t-\tau_1),\dots,x(t-\tau_\ell)).$$

We will consider here equations of the form

$$(\gamma) \qquad \dot{x}(t) = L(x_t) + N(t, x_t, \mu),$$

where L is linear in x_t (in a space to be defined) and $N(t, \varphi, \mu)$ tends to zero as both φ and the parameter μ tend to zero.

In analogy with what has been done for ordinary differential equations (see Hale [2]), we seek a method to determine the T-periodic solutions of (γ) when N is T-periodic in t.

Many of the methods which have been considered for ordinary differential equations are difficult to apply in the case of time lag as will be shown in the next paragraphs.

If in (α) F is T-periodic in t, we might assume that our solution has a trigonometric Fourier expansion of period T. We then reduce the problem of finding a T-periodic solution of (α) to that of solving the infinite number of equations obtained by equating coefficients. Making the above reduction and solving the equations which result is in general extremely difficult even when there is no lag present.

Cesari [3] shows that for nonlinear equations without lag it is not necessary to consider an infinity of solutions, but merely to see if some elements of a family of periodic functions, which are obtained as fixed points of a family of operators, satisfy a finite number $(2m+1)$ of "determining equations". Any of these fixed points which satisfies the determining equations is a periodic solution. The difficulty lies in finding the fixed points and verifying that they satisfy the determining equations. By means of an implicit function theorem, however, he succeeds in showing that under certain circumstances it is sufficient to consider the $(2m+1)$-parameter family of trigonometric polynomial containing the first m harmonics instead of the family of fixed points. The determining equations can then actually be used to calculate the $2m+1$ coefficients of the polynomials satisfying them. Further it can be shown that the functions thus obtained lie in the vicinity of the periodic solutions of the equation. This is nothing more than the justification of the Galerkin procedure. The method is still very difficult to apply, even in the most simple cases.

The generalization of the above method to equations with lag will not be attempted here and will be the subject of some further publication. Let us remark, however, that the method of Cesari in [3] relies on the use of L^2 spaces and these do not seem the most appropriate for equations like (β), which we want to be able to include in our theory. It looks as if the modification of the method introduced by Knoblock [4] using uniform norms would generalize without trouble to lag equations including the difference-differential type.

The basis of the perturbation procedure of Cesari and Hale for (γ) without lag, as it is shown in the last part of [3], is essentially the same as in the case above. Now, however, we look for periodic solutions of the perturbed system which tend to periodic solutions of the linear system as the parameter μ tends to zero.

The generalization of this procedure to lag equations is made possible by decomposing equation (γ) by means of the projection operators defined by Hale [1]. We then obtain an ordinary differential equation without time lag perturbed with a term containing some lag element which couples this equation with a second one. By neglecting this lag element we obtain an ordinary perturbation problem which can be dealt with by the methods mentioned above. For small μ the periodic solutions of the unperturbed equation yielded by the determining equations are close to periodic solutions of equation (γ). In a first approximation we want to find the periodic solution of the linear equation to which the periodic solutions of the perturbed one tend.

The basic idea behind the decomposition in [1] and the reduction of the problem to equations without lag is to consider a function space as our phase space. Notice that the initial value problem for equation (α) is well posed if we give as initial condition a function defined in an interval of length r. In fact there might be an infinity of functions which satisfy the equation and pass through a given point of the n-dimensional euclidean space.

In section I we give the required background material on equations with time lag. In section II we develop the method for (γ) nonautonomous. The reduction of (γ) autonomous to the previous case is treated in section III. In the next section we show how the basic procedure can be used to determine the asymptotic stability of a periodic solution. In order to do that we require some simple results from the theory of periodic linear equations with time lag that we borrow from Stokes [5] and Shimanov [6]. In section V we present a simple example arising from a control system with a delay in the feedback. Section VI is devoted to the procedure to be followed when we have to use higher order terms to ascertain the existence of periodic solutions and an example of the application of this procedure is given.

I. Preliminaries.

Let E^n be the n-dimensional complex euclidean space and consider the continuous function $x : [-r, \tau) \to E^n$, $\tau, r > 0$. Consider also the space $C([-r, 0], E^n) = C$ of the continuous functions defined in the closed interval $[-r, 0]$ with range in E^n, with the sup norm. We define the operator $\mathscr{A}_t$ associating an element of C to x for every t in $[0, \tau)$ by means of the rule

$$\mathscr{A}_t(x) = x(t + \theta), \quad \theta \text{ in } [-r, 0].$$

In order to simplify the notation we shall use $\mathscr{A}_t(x) = x_t$. (See Hale [1]).

Given a functional $F : R \times C \to E^n$ an letting $\dot{x}(t)$ represent the right hand derivative of x at t, we define a functional-differential equation as the relation

$$\dot{x}(t) = F(t, x_t). \tag{1}$$

The function F does not need to be defined on the whole of $R \times C$. In fact for our use in this work we shall suppose it defined for all R and in an open ball $C_H = \{\varphi \in C : \|\varphi\| < H\}$.

We say that $x(\sigma, \varphi) : [\sigma - r, \tau) \to E^n$ is a solution of (1) with initial value φ at σ if there exists $\tau > \sigma$ such that $x_t(\sigma, \varphi)$ is in C_H for t in $[\sigma, \tau)$, $x_\sigma(\sigma, \varphi) = \varphi$ and (1) is satisfied by $x(\sigma, \varphi)(t)$, t in $[\sigma, \tau)$.

If (1) is autonomous, i.e., F does not depend explicitly on t, and we choose $\sigma = 0$, we abbreviate $x_t(\sigma, \varphi)$ by $x_t(\varphi)$.

Consider the case in which (1) is autonomous and F is a continuous linear functional:

(2) $$\dot{u}(t) = L(u_t).$$

This case is particularly important to us, since most of the properties of our perturbed equations (γ) depend on the unperturbed ones.

In the next paragraphs we summarize the parts of the theory of (2) which are relevant to this work. For a more detailed exposition, with proofs, see [1].

The Riesz representation theorem tells us that we can write

(3) $$L(\varphi) = \int_{-r}^{0} [d\eta(\theta)]\varphi(\theta) ,$$

where $\eta(\theta)$ is an $n \times n$ matrix of function of bounded variation on $[-r, 0]$. On the other hand it is well known that (2) has a unique solution defined for t in $[0, \infty)$ for any initial value φ in C at zero (see Krasovskii [7], or Halanay [8]).

We define the semi-group of operator $U(t) : C \to C$ by

$$U(t)\varphi = u_t(\varphi) ,$$

where $u(\varphi)$ is the solution of (2) with initial value φ at zero. For each $t > 0, \tau > 0$, $U(t)$ is a bounded linear operator satisfying $U(t + \tau) = U(t)U(\tau)$.

In terms of the matrix η appearing on (3) we find that the characteristic values of (2) are given as the roots of the characteristic equation

(4) $$\det (\lambda I - \int_{-r}^{0} [d\eta(\theta)]e^{\lambda\theta}) = 0$$

There are only a finite number of roots of (4) in any half plane $\operatorname{Re} z \geq \gamma$, and each of these roots has finite multiplicity.

If λ has multiplicity k, then there are k, and no more than k linear independent solutions of (2) of the form $y(t) = p(t)e^{\lambda t}$, where $p(t)$ is a polynomial with coefficients in E^n of degree $\leq k-1$.

We observe that these solutions can be prolongated backwards, i.e., there is a function $y: R \to E^n$, such that

$$u(y_\tau)(t) = y(t-\tau) \quad \text{for} \quad t, \tau \in R.$$

Let Y denote the matrix having as columns the k linealy independent solution mentioned above. Then there exists a constant matrix B, with λ as only characteristic value, such that

$$Y(t) = Y(0)e^{Bt} \quad , \quad t \in R.$$

If we define $\Phi = Y_0$, i.e., the matrix whose columns are the elements of C corresponding to y_0, then we have:

$$Y_t = U(t)\Phi = \Phi e^{Bt}, \; \Phi(\theta) = \Phi(0)e^{B\theta}, \; \theta \epsilon [-r, 0].$$

This relation says that Φ is the basis of a finite dimensional subspace $P(\Lambda)$ of C which is invariant under $U(t)$. In this subspace we can extend the definition of $U(t)$ to negative values of t by taking $U(-t) = \Phi\, e^{-Bt}$.

Given any finite set $\Lambda = \{\lambda_i\}$ of characteristic values of (2) it is possible to obtain a set of functions of the form $y(t) = p(t)e^{\lambda_i t}$, $t \in R$, such that, if Y denotes the matrix whose columns are this basis, there exists a constant matrix B such that

(5) $$Y(t) = Y(0)e^{Bt}, \quad t \in R,$$

where B has as characteristic values the elements of Λ.

The linear subspace spanned by the columns of $Y_0 = \Phi$ is called the generalized eigenspace associated with Λ, and will be denoted by $P(\Lambda)$.

If φ is an element of $P(\Lambda)$ we have then

(6) $$u_t(\varphi) = U(t)\varphi = \Phi e^{Bt} b, \quad \varphi = \Phi b .$$

That shows that in $P(\Lambda)$ the behavior of the solution is the same as that of an ordinary differential equation with constant coefficients.

If L is a real functional $(L : C \to R^n)$, and we are only interested in the real part of u_t, then we know that both λ and $\bar{\lambda}$ are characteristic roots. By associating λ with $\bar{\lambda}$ we can choose Φ as a matrix whose elements are real functions and such that their columns form a basis for the real part of $R(\Lambda)$. In this case B will be a real constant matrix.

We will next characterize the space $Q(\Lambda)$ complementary to $P(\Lambda)$ which will be also invariant under the operator $U(t)$ for $t \geq 0$. Every element φ of C will then be uniquely expressible as the sum of an element of $P(\Lambda)$ and one of $Q(\Lambda)$. These elements are called the projections of φ on $P(\Lambda)$ and $Q(\Lambda)$ respectively. If p_P and p_Q designate the operators of projection we can write

(7) $$\varphi = p_P(\varphi) + p_Q(\varphi) .$$

To abbreviate we designate $p_P(\varphi)$ and $p_Q(\varphi)$ by φ^P and φ^Q respectively. We write then (7) as

$$\varphi = \varphi^P + \varphi^Q .$$

We obtain the characterization of $Q(\Lambda)$ with the help of the following equation, known as the <u>adjoint</u> to (2)

(8) $$\dot{v}(s) = -\int_{-r}^{0}[d\eta^T(\theta)]v(s-\theta) , \quad s \leq 0 ,$$

(η^T is the transpose of η), and its associated charactertistic equation

(9) $$\det\left(\lambda I - \int_{-r}^{0}[d\eta^T(\theta)]e^{\lambda\theta}\right) = 0 .$$

The solutions of (4) and (9) are the same. A solution of (8) is uniquely determined by giving an initial condition ψ in $C([0, r],E^n) \overset{def}{=} C^*$ at 0, and integrating (8) for $s \leq 0$. To any φ in C and ψ in C^* we associate the bilinear form (ψ, φ) defined by

(10) $$(\psi, \varphi) = \psi^T(0)\varphi(0) - \int_{-r}^{0}\int_{0}^{\theta} \psi^T(\xi - \theta)[d\eta(\theta)]\varphi(\xi)d\xi .$$

If Φ is a basis for $P(\Lambda)$ and Ψ is a basis for $P^*(\Lambda)$ (the generalized eigenspace of Λ in C^*), then $(\Psi, \Phi) = (\psi_j, \varphi_k))$ is non singular and, by changing the bases, can be taken as the identity matrix. Let us then assume

(11) $$(\Psi, \Phi) = I .$$

The space $Q(\Lambda)$ is characterized by

(12) $$Q(\Lambda) = \{\varphi \in c : (\Psi, \varphi) = 0\}.$$

If $\varphi \in Q(\Lambda)$, then $U(t)\varphi \in Q(\Lambda)$ for $t \geq 0$. In this case the solutions are not necessarily defined for negative t as in $P(\Lambda)$.

We have then that the projection operator p_P is defined by

$$\varphi^P = p_P\varphi = \Phi(\Psi, \varphi)$$

and

$$\varphi^Q = P_Q\varphi = \varphi - P_P\varphi .$$

Consider now the equation

$$\dot{x}(t) = L(x_t) + N(t, x_t). \tag{13}$$

We want an expression, similar to the variation of parameters formula, which will give the solution of (13) for a given initial value in terms of the solutions of (2).

Let $X(t)$ be the $n \times n$ matrix whose columns are the solutions for $t \geq -r$ of equation (2) with $X(t) = 0$ for t in $[-r, 0)$ and $X(0) = I$, the identity matrix. Then we have the following representation for the solutions of (13) with initial value φ at σ (see Halanay [8],[9] and Hale-Perelló [10]):

$$\begin{aligned} x(t) &= U(t-\sigma)\varphi(0) + \int_\sigma^t X(t-\tau)N(\tau, x_\tau)d\tau, && t \geq 0, \\ x(\sigma + \theta) &= \varphi(\theta) && \theta\in[-r, 0] \end{aligned} \tag{14}$$

It is shown in [9] that by projecting X_o on P and Q as indicated previously, that is, by taking

$$X_o^P = \Phi(\Psi, X_o) = \Phi\Psi^T(0)$$

$$X_o^Q = X_o - X_o^P \quad ,$$

the equation (14) can be decomposed as follows:

(15)
$$X_t^P(\theta) = U(t-\sigma)\varphi^P(\theta) + \int_\sigma^t U(t-\tau)X_0^P(\theta)N(\tau, x_\tau)d\tau\ , \quad t \in R$$

$$X_t^Q(\theta) = U(t-\sigma)\varphi^Q(\theta) + \int_\sigma^t U(t-\tau)X_0^Q(\theta)N(\tau, x_\tau)d\tau\ , \quad t \geq 0$$

From now on, in order to abbreviate, we will not write the θ when using these formulas.

II. The nonautonomous equation

Consider equation (2) and assume that Λ is the set of all of its characteristic roots of the form $i\frac{2\pi n}{T}$, n integer. We know there is only a finite number of such roots. Assume, moreover that the dimension of the eigenspaces spanned by these roots coincide with their multiplicity. Then $P(\Lambda)$ will consist of all those functions which are initial values of T-periodic solutions of (2).

According to (6) the orbits (or paths) of the equation in $P(\Lambda)$ are given by $u_t(\varphi) = \Phi e^{Bt}b$, where $\varphi = \Phi b$ and B is a $p \times p$ matrix which has the elements of Λ as eigenvalues and has simple elementary divisors. Notice that p and n are not related, and any can be larger than the other. If $w(t) = (\Psi, w_t)$ we have that for u_t in $P(\Lambda)$, $w(t)$ satisfies the linear equation

(16) $$\dot{w}(t) = Bw(t).$$

We introduce some more notation that we need in the next pages:

S^p denotes the space of T-periodic functions y from R into E^p with the norm $\|y\|_S = \sup\{|y(t)|, t \in R\}, |y|^2 = y^*y$, y^* the conjugate transpose of y.

Σ denotes the space of T-periodic functions x_t from R into C with the norm $\|x_t\| = \sup\{\|x_t\|, t \in R\}$, $\|x_t\| = \sup\{|x(\theta)|, \theta \in [-r, 0]\}, |x|$ as above.

$\mathscr{O}: S^p \to S^p$ denotes the operator defined by

$$\mathscr{O}(f) = \frac{1}{T}\int_0^T e^{B(t-\tau)} f(\tau)d\tau.$$

Notice that $\mathscr{O}(f)$ is of the form $e^{Bt}a$ and hence will correspond to some solution of (16).

By $\Omega : \Sigma \to \Sigma$ we denote the operator defined by

$$\Omega(x_t) = \Phi \mathcal{O}(\psi, x_t) .$$

Here we are using the notation $x_t : R \to C$ even if there is no $x : R \to E^n$ corresponding to it (see the definition of x_t at the beginning of section I). The use of this notation is similar to the abuse made when we write $x(t): R \to E^n$ which we do very frequently in order to use less symbols.

To begin with we will find necessary and sufficient conditions for the equation

(17) $$\dot{x}(t) = L(x_t) + f(t) ,$$

with f in S^n and L as above to have T-periodic solutions. Such conditions are given in a more general theorem in [8], but we prefer to include the proof for our case which is much simpler.

Lemma 1.

If $f \in S^p$, *then the equation*

(18) $$\dot{y}(t) = B\, y(t) + f(t) ,$$

B *as in* (16), *has a periodic solution if and only if* $\mathcal{O}(f) = 0$, *and in this case for every* $a \in E^p$ *there is a unique solution* $y^*(a)$ *of* (18) *such that* $\mathcal{O}(y^*(a) = e^{Bt}a = w(a)(t)$, i.e., $\mathcal{O}(y^*(a))$ *is the solution of* (16) *with initial value* a *at* $t = 0$.

Moreover the following estimate holds

$$\|y^*(a) - w(a)\|_S \leq K \int_0^T |f(\tau)|\, d\tau ,$$

where K *does not depend on* f *or* a.

Remark: $y^*(a)$ is not necessarily the solution of (18) with initial value a at $t = 0$.

Proof: The solution of (18) with initial value y_o at $t = 0$ is given by

$$(19) \qquad y(t) = e^{Bt}y_o + \int_o^t e^{B(t-\tau)}f(\tau)d\tau .$$

As $e^{Bt}y_o$ is T-periodic, in order to have $y(t)$ T-periodic it is necessary and sufficient that $\int_o^t e^{B(t-\tau)}f(\tau)d\tau$ be T-periodic, that is, we require the $\int_o^T e^{-B\tau}f(\tau)d\tau = 0$ or, using our notation, $\mathcal{O}(f) = 0$.

From (19) we have for $y \in S^p$ that $e^{-Bt}y(t) = a + g(t)$, where $a = y_o + \frac{1}{T}\int_o^T \int_o^t e^{-B\tau}f(\tau)d\tau = y_o + c$, and g is a function in S^p with mean value 0.

Applying the operator $\mathcal{O}$ to $y \in S^p$ we obtain

$$\mathcal{O}(y)(t) = e^{Bt}(y_o + \frac{1}{T}\int_o^T \int_o^\xi e^{-B\tau}f(\tau)d\tau \, d\,\xi) = e^{Bt}a = v(a)(t).$$

Hence $\mathcal{O}$ gives a 1-1 correspondence between the periodic solutions of (18) and those of (16).

From the fact that

$$\|g\|_S \leqq 2T \, \| e^{-Bt}\|_S \int_o^t |f(\tau)|d\tau = k \int_o^t |f(\tau)|d\tau$$

the last part of the lemma follows by taking $K = \|e^{Bt}\|_S k$. For the matrices e^{Bt} and e^{-Bt} we are using as S norm the supremum of the square root of the sum of the product of their elements by their conjugates for all t.

Lemma 2.

<u>If</u> h <u>is in</u> S^n, <u>then there exists a unique</u> $y \in Q(\Lambda)$ <u>such that</u>

$$x_t^{*Q} = U(t)\varphi + \int_0^t U(t-\tau)X_0^Q\, h(\tau)d\tau \tag{20}$$

is T-periodic.

Moreover we have

$$\|x_t^{*Q}\|_\Sigma \leq K' \int_0^T |h(\tau)|d\tau \quad ,$$

where K' is independent of the h chosen.

Proof: If x_t^{*Q} is T-periodic we have

$$\varphi = U(T)\varphi + \int_0^T U(T-\tau)X_0^Q\, h(\tau)d\tau, \quad \text{that is}$$

$$\varphi = (I-U(T))^{-1}\int_0^T U(T-\tau)X_0^Q\, h(\tau)d\tau \ .$$

We have that $I-U(T)$ has an inverse if $(I-U(T))\varphi = 0$ implies $\varphi = 0$. This is the case, since we have assumed that there are no T-periodic solutions of (2) in $Q(\Lambda)$ besides the identically zero. Hence φ is uniquely determined. Notice that $\int_0^T U(T-\tau)X_0^Q(\theta)h(\tau)d\tau$ is a continuous function in θ for θ in $[-r, o]$.

The expression for x_t^{*Q} is

$$x_t^{*Q} = (I-U(T))^{-1} \int_t^{t+T} U(t+T-\tau)X_0^Q\, h(\tau)d\tau \ .$$

The estimate on the Σ-norm of x_t^{*Q} is obtained as follows:

$$\|x_t^{*Q}\|_\Sigma = \|(I-U(T))^{-1}\| \sup_{\tau\in[t,t+T]} \|U(t+T-\tau)X_0^Q\| \int_0^T |h(\tau)|d\tau =$$

$$= \|(I-U(T))^{-1}\| \sup_{t\in[0,\ T]} \|U(t)X_0^Q\| \int_0^T |h(\tau)|d\tau = K' \int_0^T |h(\tau)|d\tau \ .$$

By using our decomposition (15) we obtain immediately the desired property concerning equation (17):

Theorem 1.

The equation

$$\dot{x}(t) = L(x_t) + f(t) \tag{17}$$

with $f \in S^n$ and L as in (2) has a T-periodic solution if and only if $\mathcal{O}(\Psi^T(0)f) = 0$, and in this case, for every Φa in $P(\Lambda)$ there exists a unique solution $x_t^*(a)$ such that $\mathcal{O}(\Psi, x_t(a)) = e^{Bt}a$.

Moreover the following estimate holds:

$$\|x_t^*(a) - u_t(\Phi a)\|_\Sigma \le K'' \int_0^T |f(\tau)|d\tau \quad , \tag{21}$$

where K'' does not depend on f.

Notice that the condition $\mathcal{O}(\Psi^T(0)f)$ is equivalent to

$$\int_0^T e^{-B\tau}\Psi^T(0)f(\tau)d\tau = 0 \text{ , or}$$
$$\int_0^T \Psi^T(\tau)f(\tau)d\tau = 0 \text{ ,} \tag{22}$$

that is, in order for (17) to have some T-periodic solution it is neccessary and sufficient that f be orthogonal, in the sense of (22), to the T-periodic solutions $\Psi^T(t)$ of the equation adjoint to (2) (See [8]),

In the case in which (2) has no T-periodic solutions besides the identically zero, then there is a unique T-periodic solution for every f in S^p.

The following two lemmas follow trivially from the ones above, but we prefer to state them explicitly for easier reference.

Lemma 3.

If x_t *is an element of* Σ, *then for every* $a \in E^p$ *the equation*

$$(23) \qquad \dot{y}(t) = B\, y(t) + \Psi^T(0)N(t, x_t) - \mathcal{O}(\Psi^T(0)N(t, x_t)) ,$$

where $N(t, \varphi)$ *is a functional of period* T *in* t, *continuous with respect to* (t, φ) *and uniformly lipschitzian in* φ *in* C_H, *has a unique solution* $y^*(a, x_t) \in S^p$ *such that* $\mathcal{O}(y^*(x_t)) = e^{Bt}a = w(a)(t)$.

To abbreviate we are going to write

$$e^{-Bt}(\Psi^T(0)N(t, x_t) - \mathcal{O}(\Psi^T(0)N(t, x_t)) = f(x_t)(t)$$

With this notation we have for the solution $y^*(a, x_t)$ of (23):

$$(24) \qquad y^*(a, x_t)(t) = e^{Bt}\left(a + \int_0^t f(x_t)(\tau)d\tau - \frac{1}{T}\int_0^T\int_0^\sigma f(x_t)(\tau)d\tau\, d\sigma\right) =$$

$$= e^{Bt}(a + g(t))$$

Here $g(t)$ stands for the unique T-periodic function with zero mean value whose derivative is $f(x_t)(t)$.

If we want to express $g(t)$ as an integral we have to deal with its components separately. In fact if the components of g are complex we have to deal separately with the real and imaginary part for each component. We can choose ξ_i , η_i in $[0, T]$, $i = 1,\ldots,p$, such that $\mathrm{Re}\, g_i(\xi_i) = \mathrm{Im}\, g_i(\eta_i) = 0$. We will have then that

$$(25) \qquad \mathrm{Re}\, g_i = \int_{\xi_i}^t f(x_t)(\tau)d\tau \quad \text{and} \quad \mathrm{Im}\, g_i(t) = \int_{\eta_i}^t f(x_t)(\tau)d\tau$$

have zero mean value. If $\bar{\xi}$ denotes the vector of E^p with components $(\zeta_1,\ldots,\zeta_p)$, $\zeta_i = \xi_i + i\eta_i$, we will write

(26) $$g(t) = \int_{\zeta(x_t)}^{t} f(x_t)(\tau)d\tau$$

for the vector function with components (25). Notice that $\zeta(x_t)$ is not necessarily uniquely determined.

Observe that if we take a new $x_t' \in \Sigma$, the following linear property holds for some $\zeta(x_t + x_t')$ with components with real and imaginary parts in $[0, T]$:

(27) $$\int_{\zeta(x_t)}^{t} f(x_t)(\tau)d\tau + \int_{\zeta(x_t')}^{t} f(x_t')(\tau)d\tau = \int_{\zeta(x_t+x_t')}^{t} f(x_t')(\tau) + f(x_t)(\tau))d\tau.$$

This follows because both terms of the first member have mean value zero and so must have their sum, $h(t)$ say. On the other hand $h'(t) = f(x_t)(t) + f(x_t')(t)$ and there exists $\zeta(x_t + x_t')$ in $[0, T]$ such that the second member of (27) is equal to $h(t)$.

Lemma 4.

If $x_t \in \Sigma$, then under the same hypothesis as above, there exists a unique φ in $Q(\Lambda)$ such that

$$V(t)\varphi + \int_0^t U(t-\tau)X_0^Q N(\tau, x_\tau)d\tau$$

is in Σ.

The main purpose of this section is to give conditions under which the following equation has T-periodic solutions:

(28) $$\dot{x}(t) = L(x_t) + N(t, x_t, \mu).$$

Here L is as before and $N(t, \varphi, \mu)$ fulfills the following conditions in the region $R \times C_H \times [-\mu_0, \mu_0]$ for some $H, \mu_0 > 0$:

i) $N(t, \varphi, \mu)$ is continuous in (t, φ, μ), $N(t, 0, 0) = 0$,

ii) $N(t, \varphi, \mu)$ is T-periodic in t,

iii) $|N(t, \varphi, \mu) - N(t, \varphi_2, \mu)| \leq \eta(|\mu|, H) \, \|\varphi_1 - \varphi_2\|$,

φ_1, φ_2 in C_H for some continuous function η defined in $[0,\mu_o] \times [0, H_o]$, η nondecreasing in $|\mu|$ and H and $\eta(0, 0) = 0$.

The above conditions are enough to insure locally the existence and uniqueness of solution for any μ in $[-\mu_o, \mu_o]$ and any initial condition φ in C_H at a time σ in R. If we do not leave C_H for any t, then the solution is defined for all $t \geq \sigma$, and if for some φ we have that $x_T(\varphi) = \varphi$, $x_t(\varphi) \in C_H$ for $t \in [0, T]$, then we can take $x_t(\varphi)$ T-periodic for every t in R. Notice here that it may happen that there is no uniqueness of solution going backwards in time. It may occur that two solutions with different initial conditions at σ coincide after some $t > \sigma$. For instance the equation $\dot{x} = Ax$ considered as a lag equation in the phase space $C([-r, 0], E^n)$, $r > 0$, is such that any solution with initial condition φ such that $\varphi(0) = 0$ will be zero for $t \geq 0$.

For any α, $0 < \alpha < 1$, and for any $a \in E^P$ fulfilling $\|\Phi \, e^{Bt}a\|_\Sigma \leq \alpha H$, we denote by $\Sigma_{a,H}$ the following subset of Σ:

$$(29) \qquad \Sigma_{a,H} = \{x_t \in \Sigma : \Omega(x_t) = \Phi \, e^{Bt}a, \ \|\Omega(x_t)\|_\Sigma \leq \alpha H, \ \|x_t\|_\Sigma \leq H\} \ ,$$

i.e. the set of those T-periodic solutions from R into C which never leave the ball C_H and such that their "average" Ω equals $\Phi \, e^{Bt}a$ and is contained in the smaller ball $C_{\alpha H}$.

We do not make explicit the choice of α, but we have to keep in mind that its value is fixed throughout the whole reasoning. Notice also that if η is independent of H our results will be valid for any H if μ is small enough.

Lemma 5.

There exist $\mu_1 > 0$, $H > 0$ such that for every $a \in E^P$ with $\|\Phi e^{Bt} a\|_\Sigma \leq \alpha H$ there exists a unique $x_t = x_t(a, \mu)$ in $\Sigma_{a,H}$ satisfying the relations

$$(30) \qquad \dot{y}(t) = By(t) + \Psi^T(0)\, N(t, x_t, \mu) - \mathcal{O}(\Psi^T(0) N(t, x_t, \mu)),$$

$$(31) \qquad x_t^Q = U(t)x_0^Q + \int_0^t U(t-\tau)X_0^Q\, N(\tau, x_\tau, \mu)d\tau$$

for every μ with $|\mu| \leq \mu_1$. Furthermore this $x_t(a, \mu)$ is continuous on (a, μ).

Proof:

We use the notation

$$(32) \qquad n(x_t, \mu)(t) = e^{-Bt}(\Psi^T(0)N(t, x_t, \mu) - \mathcal{O}(\Psi^T(0)N(t, x_t, \mu))\,,$$

for the function of t which results from substuting a given x_t in Σ in the right hand side of (32).

If we take z_t in Σ and substitute it in (30) and (31) we obtain two uncoupled equations:

$$(33) \qquad \dot{y}(t) = By(t) + e^{Bt}n(z_t, \mu)(t)$$

$$x_t^Q = U(t)x_o^Q + \int_o^t U(t-\tau)X_o^Q N(\tau,\ z_\tau,\ \mu)d\tau\ , \tag{34}$$

According to Lemma 3, for any a in E^p equation (33) has a unique T-periodic solution given by

$$y^*(t,\ a,\ z_t,\ \mu) = e^{Bt}(a + \int_{\overline{\zeta}(z_t,\mu)}^{t} n(z_t,\ \mu)(\tau)d\tau)\quad . \tag{35}$$

In a similar way, according to Lemma 4, equation (34) has a unique periodic solution x_t^Q given by

$$x_t^{*Q}(z_t,\ \mu) = (I-U(T))^{-1}\int_t^{t+T} U(t + T - \tau)X_o^Q N(\tau,\ z_\tau,\ \mu)d\tau. \tag{36}$$

Let's define the operator $\mathscr{F}(a,\ \mu)$ from Σ into Σ by

$$\begin{aligned}\mathscr{F}(a,\ \mu)(z_t) &= \Phi y^*(t,\ a,\ z_t,\ \mu) + x_t^{*Q}(z_t,\ \mu) = \\ &= \mathscr{F}^P(a,\ \mu)(z_t) + \mathscr{F}^Q(a,\ \mu)(z_t).\end{aligned} \tag{37}$$

We will show that for μ, H small enough $\mathscr{F}(a,\ \mu)$ maps $\Sigma_{a,H}$ into itself and that it is a contraction. Consequently there is a unique element in $\Sigma_{a,H}$ fixed under $\mathscr{F}(a,\ \mu)$.

The fact that $\Omega(\mathscr{F}(a,\ \mu)(z_t)) = \Phi\ e^{Bt}a$ is obvious. We have to show now that $\|\mathscr{F}(a,\ \mu)(z_t)\|_\Sigma \leq H$ if z_t is in $\Sigma_{a,H}$ and μ is sufficiently small.

From Lemma 1 we have the estimate

$$\|y^*(t,\ a,\ z_t,\ \mu) - e^{Bt}a\| \leq K \int_0^T |e^{B\tau} n(z_t,\ \mu)(\tau)|\, d\tau ,$$

for some K independent of n.

Since $\|\mathcal{O}(f)\|_s \leq \|f\|_s$, we get the estimate

$$\|\mathcal{F}^P(a,\ \mu)(z_t)\|_\Sigma = \|\Phi y^*(t,\ a,\ z_t,\ \mu)\|_\Sigma \leq$$

$$\leq \|\Phi\ e^{Bt}a\|_\Sigma + 2KT\ \|\ \Phi\ \|\|\Psi\|(\eta(|\mu|,H)H + \kappa(|\mu|)) \leq$$

$$\leq b + kK(\eta(|\mu|,H) + \kappa(|\mu|)) \leq$$

where κ is continous, increasing and $\kappa(0) = 0$.

By Lemma 2 we have

$$\|\mathcal{F}^Q(a,\ \mu)(z_t)\|_\Sigma \leq 2K'T\ \|\Psi\|(\eta(|\mu|,H)H + \kappa(|\mu|)) =$$

$$= k'K'(\eta(|\mu|,H)H + \kappa(|\mu|)) .$$

It is sufficient to take

$$(\eta(|\mu|,H) + \kappa(|\mu|))(Kk + K'k') \leq H - b$$

to have $\|\mathcal{F}(a,\ \mu)(z_t)\|_\Sigma \leq H$ and hence $\mathcal{F}(a,\ \mu)(z_t)$ in $\Sigma_{a,H}$. Due to the continuity of η and κ we can choose $\mu_1' > 0$, $H_1 > 0$ such that $\eta(\mu_1',H_1)H_1+\kappa(\mu_1') \leq \dfrac{H_1(1-\alpha)}{Kk + K'k'}$, and then $\mathcal{F}(a,\ \mu)$ maps Σ_{a,H_1} into Σ_{a,H_1} for $|\mu| \leq \mu_1'$.

We will now prove the contracting property of $\mathscr{F}(a, \mu)$, namely that for $|\mu|$ small enough then exists a $\delta_1 < 1$ such that for z_t and z_t' in $\Sigma_{a,H}$ the following holds:

$$\|\mathscr{F}(a, \mu)(z_t) - \mathscr{F}(a, \mu)(z_t')\|_\Sigma \leqq \delta_1 \|z_t - z_t'\|_\Sigma \quad . \tag{38}$$

According to (32) and (35) we have

$$\|\mathscr{F}^P(a, \mu)(z_t) - \mathscr{F}^P(a, \mu)(z_t')\|_\Sigma \leqq$$

$$\leqq \|\Phi\| \left\| \int_{\zeta(z_t - z_t')}^{t} (n(z_t, \mu)(\tau) - n(z_t', \mu)(\tau))d\tau \right\|_s \leqq$$

$$\leqq 2\, \|\Phi\|\ T\ \eta(|\mu|, H)\|\Psi\|\|z_t - z_t'\|_\Sigma = \eta(|\mu|)\ k\ \|z_t - z_t'\|_s \ .$$

Using (36) we get:

$$\|\mathscr{F}^Q(a, \mu)(z_t) - \mathscr{F}(a, \mu)(z_t')\| \leqq$$

$$\leqq K' \int_0^T |N(\tau, z_t, \mu) - N(\tau, z_t', \mu)| d\tau \leqq$$

$$\leqq K'T\ \eta(|\mu|, H)\|z_t - z_t'\|_\Sigma \ .$$

We can choose $\mu_1'' > 0$, $H_2 > 0$ such that $\eta(|\mu|, H_2)(k + K'T) < 1$ for $|\mu| \leqq \mu_1''$.

By choosing $\mu_1' = \min\{\mu_1', \mu_1''\}$ and $H = \min\{H_1, H_2\}$ we conclude that (38) holds for all $|\mu| \leqq \mu_1$.

Hence there is a unique element $x_t(a, \mu)$ of Σ_a such that

$$x_t(a, \mu) = \mathscr{F}(a, \mu)x_t(a, \mu) \tag{39}$$

Fron the continuity of $\mathscr{F}(a, \mu)$ and from the contracting property it follows that $x_t(a, \mu)$ is continuous on (a, μ) .

Theorem 2

If for some particular (a, μ) fulfilling the requirements of Lemma 5 it happens that $x_t(a, \mu)$, solution of (39), fulfills the relation

$$\mathscr{O}(\Psi^T(0)N(t, x_t(a, \mu), \mu)) = 0 \quad , \tag{40}$$

then $x(a, \mu)$ is a periodic solution of (28) and, conversely, if $\tilde{x}_t(\mu)$, $|\mu| < \mu_1$, is a periodic solution of (28) in Σ_a, then $\tilde{x}_t(\mu) = x_t(a, \mu)$ for some a.

Proof:

The first part is obvious, and the second follows from the fact that $\tilde{x}_t(\mu)$ fulfills (28) for every $t \in R$ and it has to fulfill (30),(31) and (40) according to the properties of $\mathscr{O}$. The results follow from the uniqueness of solution in $\Sigma_{a,H}$ of (30) and (31).

Equation (40) is generally known in the literature as "bifurcation equation" or "determining equation".

Notice that if Λ is empty, i.e., (2) has as only T-periodic solution the identically zero, then there is no relation (40) to fulfill and we conclude that equation (28) has a unique periodic solution $x_t(\mu)$ which depends continuously on μ and tends to 0 as $\mu \to 0$, i.e., $x_t(0) = 0$.

The method to determine T-periodic orbits of (28) for small μ is then to find $x_t(a, \mu)$ corresponding to (30), (31) for $|\mu|$ in some interval $[0, \mu]$, substitue this value in (40) and solve for a in terms of μ.

This method is too difficult to be practical. The main difficulty deriving from the fact that $x_t(a, \mu)$ is generally not known explicitly. On the other hand for any (a, μ) we can find a sequence $x_t^{(k)}(a, \mu)$, of T-periodic function converging uniformly to $x_t(a, \mu)$ due to the fact that it is the fixed point of a contracting mapping.

The sequence is given by:

$$x_t^{(0)}(a, \mu) = \Phi\, e^{Bt} a \tag{41}$$

$$x_t^{(k)}(a, \mu) = \mathscr{F}(a, \mu) x_t^{(k-1)}(a, \mu)$$

Notice that due to the form of $\mathscr{F}(a, \mu)$ we have $x_t(a, 0) = \Phi\, e^{Bt} a$.

If $\mathscr{O}(\Psi^T(0)N(t, x_t(a, \mu), \mu))$ is differentiable with respect to a we can apply the implicit function theorem and decide on the solvability of a as a function of μ in equation (40).

In order to insure this differentiablity we will ask for further restriction on N.

Lemma 6

If $N(t, \varphi, \mu)$ is as in Lemma 5 and moreover $D_\varphi N(t, \varphi, \mu)$ exists and is lipschitzian in φ with Lipschitz coefficient $\bar{\eta}(|\mu|, H)$, $\bar{\eta}$ with the same properties as in Lemma 5, then the fixed point $x_t(a, \mu)$ of $\mathcal{F}(a, \mu)$ and $\mathcal{O}(\Psi^T(0)N(t, x_t(a, \mu), \mu))$ are differentiable with respect to a for μ and H sufficiently small.

Remarks: The symbol D_φ stands for the Fréchet derivative and $D_a f$, with f a p-vector function, is a $p \times p$ matrix.

Notice that if $N(t, \varphi, \mu) = \mu N^*(t, \varphi)$, with N^* and $D_\varphi N^*$ lipschitzian the conditions of the lemma are fulfilled.

If η, $\bar{\eta}$ do not depend on H the results are valid independently of H.

Proof:

We use induction on the sequence (41). We have $D_a x_t^{(0)}(a, \mu) = \Phi e^{Bt}$. Assuming that $D_a x_t^{(k)}(a, \mu)$ exists we have:

$$D_a y^{(k+1)}(a, \mu)(t) = e^{Bt}(I + \int_0^t D_a\, n(x_t^{(k)}(a, \mu), \mu)(\tau)d\tau -$$

$$- \frac{1}{T} \int_0^T \int_0^t D_a n(x_t^{(k)}(a, \mu), \mu)(\tau)d\tau dt) ,$$

$$D_a x_t^{(k+1)Q}(a, \mu) = (I-U(T))^{-1} \int_t^{t+T} U(t+T-\tau)X_0^Q D_a N(\tau, x_\tau^{(k)}(a, \mu), \mu)d\tau .$$

Here $D_a n(x_t^{(k)}(a, \mu), \mu)$ and $D_a N(\tau, x_\tau^{(k)}(a, \mu), \mu)$ exist due to our hypothesis on N.

Notice that if the mean value of $x_t^{(k)}$ is zero, so is the mean value of $D_a x_t^{(k)}(a, \mu)$.

Hence $D_a x_t^{(k+1)}(a, \mu)$ exists and is continuous. Moreover if $\|e^{Bt}\|_s < M$, we can choose μ small enough as to have $\|D_a x_t^{(k)}(a, \mu)\|_\Sigma < M$ for all k. This can be proved by induction taking into account the Lipschitz property of $D_\varphi N(t, \varphi, \mu)$ in the same way as we proved in Lemma 5 that $\mathcal{F}(a, \mu)$ maps

$\Sigma_{a,H}$ into $\Sigma_{a,H}$.

We check next that $D_a x_t^{(k)}(a, \mu)$ converges uniformly in Σ_a to some function matrix, which is precisely $D_a x_t(a, \mu)$.

Notice that Σ is a complete space and that the sequence of function matrices $D_a x_t^{(k)}(a, \mu)$ is a Cauchy sequence, as we show in the next paragraphs:

(42) $$\|D_a x_t^{(k+1)}(a, \mu) - D_a x_t^{(k)}(a, \mu)\| \leqq$$

$$\leqq \bar{\eta}(|\mu|, H)K_1(\|x^{(k)} - x\|_\Sigma + \|x - x^{(k-1)}\|_\Sigma) +$$

$$+ \bar{\eta}(|\mu|, H)K_2\|D_a x^{(k)} - D_a x^{(k-1)}\|.$$

Here we are using as norms of the function matrices the supremum of the norms of its columns considered as vectors

The constant K_1 depends on M and K_2 on the upper bound of $D_\varphi N$ on C_H.

From (38) into (41) it follows that

$$\|x_t^{(k)}(a, \mu) - x_t(a, \mu)\|_\Sigma \leqq \frac{\delta_1^k}{1 - \delta_1} \| x_t^{(1)}(a, \mu) - x_t^{(0)}(a, \mu)\|_\Sigma$$

Denote by δ_2 the maximum of $\bar{\eta}(|\mu|, H)K_1$ and $\bar{\eta}(|\mu|, H)K_2$ and choose μ, H small enough to have $\delta_2 < 1$. Let δ be the maximum of δ_1 and δ_2. Then it follows from (42):

$$\|D_a x_t^{(k+1)}(a, \mu) - D_a x_t^{(k)}(a, \mu)\| \leqq$$

$$\leq \delta\left(\frac{\delta^k + \delta^{k-1}}{1 - \delta}\right) \| x_t^{(1)}(a, \mu) - x_t^{(0)}(a, \mu)\|_\Sigma +$$

$$+ \delta \| D_a x_t^{(k)}(a, \mu) - D_a x_t^{(k-1)}(a, \mu)\| \leq$$

$$\leq \delta(\Delta^k + \delta\Delta^{k-1} + \ldots + \delta^{k-1}\Delta^1) \|x_t^{(1)} - x_t^{(0)}\|_\Sigma +$$

$$+ \delta^k \| D_a x_t^{(1)} - D_a x_t^{(0)}\| \leq$$

$$\leq \delta^k \left(k \frac{1 + \delta}{1 - \delta} + L\right) \|x_t^{(1)} - x_t^{(0)} \|_\Sigma .$$

Here Δ^k stands for $(\delta^k + \delta^{k-1})/(1 - \delta)$, and L is a constant factor relating the norms of $\|x_t^{(1)} - x_t^{(0)}\|_\Sigma$ and $\|D_a x_t^{(1)} - D_a x_t^{(0)}\|$. As we have that $\sum_{k=1}^{\infty} k\delta^k$ converges, it follows that $\{D_a x_t^{(k)}(a, \mu)\}$ is a Cauchy sequence converging to some element of Σ which is $D_a x_t(a, \mu)$.

We are now in condition to state the following theorem which represents the most practical result of the method.

Theorem 3

If N *fulfills the conditions required for Lemma 6 besides i),ii),iii) above, with* η, $\bar{\eta}$ *depending only on* $|\mu|$ *and if*

(43) $$\mathcal{O}(\Psi^T(0)N(t, \Phi e^{Bt} a_o, 0)) = 0$$

$$\det (D_a \mathcal{O}(\Psi^T(0)N(t, \Phi e^{Bt} a_o, 0))) \neq 0 ,$$

then there exists $\mu_1 > 0$ *such that equation (28) has a T-periodic solution*

$x_t^*(a_o, \mu)$ <u>for</u> $|\mu| < \mu_1$. <u>This solution is continuous in</u> μ <u>and</u> $x_t^*(a_o, 0) = \Phi\, e^{Bt} a_o$.

Proof:

Notice that $\Phi\, e^{Bt} a_o = x_t(a_o, 0)$. From the continunity of $x_t(a, \mu)$ with respect to μ it follows, by applying the implicit function theorem to (40), that for $a = a_o$ and $\mu = 0$ we can express a as a function of μ such that $a(0) = a_o$.

The solution $x_t^*(a, \mu)$ is given by $x_t^*(a_o, \mu) = x_t(a(\mu), \mu)$.

Evidently $x_t^*(a_o, 0) = x_t(a_o, 0) = \Phi\, e^{Bt} a_o$, and this completes the proof.

Remark: The lemma will still be true even if $\eta, \bar{\eta}$ depend on H if $a_o = 0$, since we used the property only to check that $\Phi e^{Bt} a_o = x_t(a_o, 0)$.

Notice also that if H has a factor ϵ we can take it out and consider equations (43) divided by ϵ and we obtain the desired results.

III. The autonomous equation

We will apply here the results of the preceding section to some autonomous equations, in particular to those of the type

$$(44) \qquad \dot{z}(t) = L(z_t) + N(z_t, \mu) ,$$

L and N fulfilling the same conditions of the previous section.

In order to show how the things should be done in the real case we are going to assume that L and N are real functionals over the space $C = C([-r, 0], R^n)$ and we look for real solutions of (44). The complex case is alike but a little simpler because we can diagonalize B and with every eigenvalue we don't need the conjugate to be also an eigenvalue.

In the real case we can always choose Φ (see section I) in such a way that the matrix B is of the form

$$(45) \qquad B = \operatorname{diag}\,(0_q, C_1, \dots, C_r) ,$$

$$C = \begin{pmatrix} 0 & n_i\omega \\ -n_i\omega & 0 \end{pmatrix} .$$

Here 0_q stands for the $q \times q$ zero matrix, and $n_i\omega$ are the imaginary parts of the elements of Λ, n_i ranging in the positive integers. It may happen that two n_i have the same value for a finite number of indexes.

Contrasting with the nonautonomous case, we cannot expect to preserve the period $T = 2\pi/\omega$ under perturbation. However we do expect that if some periodic solution of (44) tends to some periodic solution of (2) as μ tends to zero, then its period is going to tend to T.

We are going to look for periodic solutions of period $T(\mu) = 2\pi/\omega(\mu)$, with $\omega(\mu) = \omega + \mu\eta$, where we have to determine η in function of μ and the particular solution of (2) to which we approach when μ tends to zero.

With the notation

(46)
$$B(\omega(\mu)) = \mathrm{diag}(0_q, C_1(\omega(\mu)), \ldots, C_r(\omega(\mu))),$$
$$C_i(\omega(\mu)) = \begin{pmatrix} 0 & n_i(\omega + \mu\eta) \\ -n_i(\omega + \mu\eta) & 0 \end{pmatrix}$$

we write (44) as

(47)
$$\dot{w}(t) = B(\omega(0))w(t) + \Psi^T(0)N(\Phi w(t) + z_t^Q, \mu)$$
$$z_t^Q = U(t)z_0^Q + \int_0^t U(t-\tau)X_0^Q N(\Phi w(\tau) + z_\tau^Q, \mu)d\tau$$

where $w(t) = (\psi, z_t)$.

If we apply the change of variables

(48) $$w(t) = e^{B(\omega(\mu))t}y(t), \quad z_t^Q = x_t^Q,$$

we obtain the systems

$$\dot{y}(t) = \mu\, e^{-B(\omega(\mu))t} B(\eta) e^{B(\omega(\mu))t} y(t) +$$

(49)

$$+\ e^{-B(\omega(\mu))t} \Psi^T(0) N(\Phi e^{B(\omega(\mu))t} y(t) + x_t^Q, \mu),$$

$$x_t^Q = U(t) x_0^Q + \mu \int_0^t U(t-\tau) X_0^Q N(\Phi e^{B(\omega(\mu))t} y(\tau) + x_\tau, \mu) d\tau \quad ,$$

which is of the form

$$\dot{y}(t) = Ay(t) + \quad F(t, y(t), x_t^Q, \mu, \eta)$$

(50)

$$x_t^Q = U(t) x_0^Q + \quad \int_0^t U(t-\tau) X_0^Q\, G(\tau, y(\tau), x_\tau^Q, \mu, \eta) d\tau \ ,$$

with $A = 0_q$ and F and G $T(\mu)$-periodic in t.

The functions F and G fulfill all of the conditions which are necessary to apply Lemma 5 and Theorems 2 and 3, even if in this case (50) does not correspond to any single equation like (28). Let us remark again that by x_t we are denoting a functional dependence of elements of C on R and we don't require the existence of $x(t)$ such that $x(t + \theta) = x_t(\theta)$.

If we take

(51) $$f(a, \eta, \mu) = \int_0^{T(\mu)} F(\tau, y(\tau, a, \eta, \mu), x_\tau^Q(a, \eta, \mu), \mu, \eta) d\tau \ ,$$

then we obtain that analogously as in Theorem 3

(52) $$f(a_0, \eta_0, 0) = 0 \ , \quad \operatorname{rank}\, (D_{(a,\eta)} f(a_0, \eta_0, 0)) = p$$

are sufficient conditions to insure the possibility of expressing a and η as functions of μ.

In this case we can determine η and $p-1$ components of a as functions of μ and the other component of a. The arbitrariness of one of the components of a is due to the autonomy of the system, in which a 1-parameter family of periodic solutions corresponds to every closed orbit.

IV. The stability of periodic solutions

The results of section II can also be used to determine the stability characteristics of periodic solutions of functional differential equations.

Consider, for instance, the equation

$$(44) \qquad \dot{x}(t) = L(x_t) + \mu N(x_t, \mu) .$$

Let x_t^* be a $T(\mu)$-periodic solution of (44). Take now $z = x - x^*$ and we obtain

$$(53) \qquad \dot{z}(t) = L(z_t) + \mu(N(x_t^* + z_t, \mu) - N(x_t^*, \mu)) =$$

$$= L(z_t) + \mu\, L^*(t, z_t, \mu) + \mu o(|z_t|).$$

Here the linear functional L^* is the Fréchet derivative of $N(x^* + \varphi, \mu)$ with respect to φ and is $T(\mu)$-periodic in t.

Equation (53) gives the behavior of the solutions of (44) with respect to x^*. If we are only interested in what happens in the vicinity of x^* it is sometimes enough to consider the first variational equation

$$(54) \qquad \dot{z}(t) = L(z_t) + \mu L^*(t, z_t, \mu) .$$

In the noncritical cases the stability properties of x_t^* can be decided by the knowledge of the characteristic exponents of (54). In fact, if all the characteristic exponents, except one which is zero, have negative real parts, then x_t^* is asymptotically stable with asymptotic phase.

For the general theory of periodic linear functional differential equations see Stokes [5] and Shimanov [6]. For the stability result mentioned above see Stokes [11].

We know that the characteristic multipliers of (54) are continuous in μ and we know their value for $\mu = 0$, namely, they are given by the expontentials of the roots of the characteristic equation (4).

Hence, if x_t^* is going to be at all stable we have to require that there are no roots of (4) with positive real parts. In fact we will require that all the characteristic values of (2) have negative real parts except those in Λ. In order to prove asymptotic stability of x_t^* in this case it is sufficient to show that for μ small enough all the elements of $\Lambda(\mu)$ are in the left hand plane with the exception of one which is at 0.

The decomposition of (54) by Λ yields the following equation for the orbits in $P(\Lambda)$:

$$\dot{w}(t) = Bw(t) + \mu\, \Psi^T(0)L^*(t, \Phi w(t) + z_t^Q, \mu) .$$

Notice now that L^* is $T(\mu)$-periodic and the change of variable $w(t) = e^{B(\omega(\mu))t}y(t)$, $z_t^Q = x_t^Q$ reduces it to the form

$$\dot{y}(t) = \mu(-e^{-B(\omega(\mu))t}B(\eta)e^{B(\omega(\mu))t}y(t) + \tag{55}$$

$$+ e^{-B(\omega(\mu))t}\Psi^T(0)L^*(t, \Phi e^{B(\omega(\mu))t}y(t) + x_t^Q, \mu)).$$

From the work of Stokes and Shimanov we know that corresponding to every characteristic exponent τ there exists a solution $y(t) = e^{\tau t}p(t)$, $x_t^Q = e^{\tau t}\tilde{x}_t^Q$, where $p(t) = p(t + T(\mu))$. (Just like in the case with no lag).

Substituting this value of $y(t)$ in (55) and taking $\tau = \mu\nu$ we obtain the following equation for $p(t)$:

$$
\begin{aligned}
\dot{p}(t) &= -\mu\nu p(t) + \mu(-e^{-B(\omega(\mu))t}B(\eta)e^{B(\omega(\mu))t}p(t) + \\
&+ e^{-B(\omega(\mu))t}\Psi^T(0)L^*(t, \Phi e^{B(\omega(\mu))t}p(t) + \tilde{x}_t^Q, \mu)) .
\end{aligned} \tag{56}
$$

This equation is of the type studied in section II, and we can find, by means of the determinimg equations, what are the values of ν for which we have $T(\mu)$-periodic solutions of (56). These values are the characteristic exponents and x_t^* is asymptotically stable if all but one have negative real parts.

In most cases we don't know what x_t^* is exactly, but we know its limit value when μ tends to zero, and this value is in general good enough to determine the stability conditions for small values of μ.

V. An example

Consider the equation

$$\dddot{z}(t) + a\ddot{z}(t) + b^2\dot{z}(t) + kz(t-r) + \mu\psi(z(t-r)) = 0 \;, \tag{57}$$

in which a, b^2, k, r and μ are positive constants and ψ is a real function of the real variable z such that, for any initial φ in C there is a unique solution of (57) with initial value φ at zero for all positive t.

Equation (57) arises from a control system with a nonlinearity and a delay of value r in the feedback.

For some values of the parameters and a special form of ψ we are going to determine the periodic solutions of (57) which tend to periodic solutions of

$$\dddot{v}(t) + a\ddot{v}(t) + b^2\dot{v}(t) + kv(t-r) = 0 \tag{58}$$

as μ tends to zero.

The characteristic equation of (58) is given by

$$\lambda^3 + a\lambda^2 + b^2\lambda + ke^{-r\lambda} = 0 \tag{59}$$

Using procedures similar to the ones used in Chapter 13 of [12] we find that for $r=2$, $a = (64-\pi)/8\pi$, $b=1$ and $k = a\pi^2\sqrt{2}/64$, equation (59) has exactly two purely imaginary roots $\pm i\omega$, $\omega = \pi/8$, and that the rest of the roots have negative real parts. (For the details see [13].) This means that

$P(\Lambda)$ is a plane in C where all the periodic orbits of (58) are contained.

We can write (58) as

(60) $$\dot{u}(t) = \int_{-r}^{0} d\eta(\theta)u(t+\theta),$$ u a vector with components

u_1, u_2, u_3 and

$$\eta(\theta) = \begin{pmatrix} 0 & u(\theta) & 0 \\ 0 & 0 & u(\theta) \\ -ku(\theta+r) & -b^2u(\theta) & -au(\theta) \end{pmatrix},$$

where

$$u(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t \geq 0 \end{cases}.$$

The matrix B and Φ are given, respectively by

$$B = \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}$$ and

$$\Phi = \begin{pmatrix} \cos\omega\theta & \sin\omega\theta \\ -\omega\sin\omega\theta & \omega\cos\omega\theta \\ -\omega^2\cos\omega\theta & -\omega^2\sin\omega\theta \end{pmatrix}, \quad \theta \in [-r, 0].$$

The value of $\Psi^T(0)$ turns out to be

$$\Psi^T(0) = \frac{1}{\delta}\begin{pmatrix} \beta(1-\omega^2)+a\gamma\omega & a\beta+\gamma\omega & \beta \\ \gamma(1-\omega^2)-a\alpha\omega & ar-\alpha\omega & \gamma \end{pmatrix} ,$$

where $\alpha \doteq 0.84$, $\beta \doteq -0.30$, $\gamma \doteq 1.60$ and $\delta \doteq 2.25$.

We write now equation (57) as

(61) $$\dot{x}(t) = \int_{-r}^{0} d\eta(\theta)x(t+\theta) + f(x_t) \quad , \qquad \eta(\theta) \text{ as}$$

above, x a vector with components x_1, x_2, x_3 and

$$f(x_t) = -\mu \begin{pmatrix} 0 \\ 0 \\ \psi(x_1(t-2)) \end{pmatrix}$$

With the decomposition

$$x_t = \Phi y(t) + x_t^Q \quad , \qquad y(t) = (\Psi, x_t) \quad ,$$

we obtain the equation

(62) $$\dot{y}(t) = By(t) + \Psi^T(0)f(\Phi y(t) + x_t^Q) \ .$$

After the substitutions are made we obtain

(63) $$\begin{aligned} \dot{y}_1(t) &= \frac{\pi}{8}\, y^2(t) - \mu\frac{\beta}{\delta}\, \psi(-y_2)(t) + (x_t^Q(-2))_1) \\ \dot{y}_2(t) &= \frac{\pi}{8}\, y_1(t) - \mu\, \frac{\gamma}{\delta}\, \psi(-y_2)(t) + (x_t^Q(-2))_1) \ . \end{aligned}$$

These equations are in the form (47) and we can apply the procedure explained there. We are going to take $\psi(x) = x - x^3$ in our example.

We apply the transformation (48) with

$$e^{B(\omega(\mu))t} = \begin{pmatrix} \cos\omega\ t & \sin\omega\ t \\ -\sin\omega\ t & \cos\omega\ t \end{pmatrix} .$$

We obtain for (52) with $a_o = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ the following:

$$f(a_o, \eta_o, 0) = \begin{pmatrix} -\eta a_2 \\ \eta a_1 \end{pmatrix} - \frac{1}{\delta} \begin{pmatrix} -\frac{\gamma}{2} a_1 + \frac{\beta}{2} a_2 \\ -\frac{\beta}{2} a_1 - \frac{\gamma}{2} a_2 \end{pmatrix} +$$

$$+ \frac{1}{\delta} \begin{pmatrix} -\frac{3}{8}\gamma a_1^3 + \frac{3}{8}\beta a_2^3 + \frac{3}{8}\beta a_1^2 a_2 - \frac{3}{8}\gamma a_1 a_2^2 \\ \frac{3}{8}\beta a_1^3 - \frac{3}{8}\gamma a_2^3 - \frac{3}{8}\gamma a_1^2 a_2 - \frac{3}{8}\beta a_1 a_2^2 \end{pmatrix} = 0 .$$

By taking $a_2 = 0$, which we can do due to the arbitrariness of one of the components of a_o, this equation reduces to

$$\frac{1}{2}\gamma a_1 - \frac{3}{8}\gamma a_1^3 = 0$$

$$\delta\eta a_1 + \frac{1}{2}\beta a_1 - \frac{3}{8}\beta a_1^3 = 0 ,$$

which yields $a_1 = 0$, η undetermined and $a_1^2 = \frac{4}{3}$ with $\eta = 0$.

This means that our equation has two periodic solutions (letting aside the phase) tending respectively to 0 and to the solution of $\dot{u} = Bu$ with "radius" $\sqrt{4/3}$ as μ tends to 0.

We apply now the procedure of the previous section to compute approximately the characteristic exponents of the first variational equation (54).

We take $x_t^* = \Phi\, e^{Bt} a_o$, $B = B(\omega)$, and we have

$$y_2(t) = (\Phi(-2)e^{Bt}a_o)_2 = a_1 \sin \omega\, t.$$

The value of $L^*(t, z_t)$ is given by

$$L^*(t, z_t) = \begin{pmatrix} 0 \\ 0 \\ -\psi'(a_1 \sin \omega\, t)\ z_1(t-2) \end{pmatrix} .$$

Decomposing $\underline{z}_t$ by Λ in order to have $z_t = \Phi\, w(t) + z_t^Q$ and performing the change of coordinates $w(t) = e^{B(\omega(\mu))t}y(t)$, $z_t^Q = x_t^Q$, we obtain equation (56) with

$$e^{-B(\omega(\mu))t}\Psi^T(0)L^*(t,\ \Phi e^{B(\omega(\mu))t}p(t)) =$$

$$= \frac{1}{8}\,\psi'(a_1 \sin \omega\, t)(-\sin \omega(\mu)t\ \cos \omega(\mu)t)p(t) \times$$

$$\times \begin{pmatrix} \beta \cos \omega(\mu)t - \gamma \sin \omega(\mu)t \\ \beta \sin \omega(\mu)t + \gamma \cos \omega(\mu)t \end{pmatrix} .$$

As $\psi'(a_1 \sin \omega t) = 1 - 3a_1^2 \sin^2 \omega t$ we obtain the determining equations

$$\begin{pmatrix} -\kappa + \frac{1}{\delta}(\frac{1}{2}\gamma - \frac{9}{8}\gamma a_1^2) & -\eta + \frac{1}{\delta}(\frac{1}{2}\beta - \frac{3}{8}\beta a_1^2) \\ \eta + \frac{1}{\delta}(-\frac{1}{2}\beta + \frac{9}{8}\beta a_1^2) & -\kappa + \frac{1}{\delta}(\frac{1}{2}\gamma - \frac{3}{8}\gamma a_1^2) \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = 0 .$$

As $a_1^2 = \frac{4}{3}$ and $\eta = 0$, we have κ given by the eigenvalues of the matrix

$$\begin{pmatrix} -\gamma & 0 \\ \beta & 0 \end{pmatrix} ,$$

which are 0 and $-\gamma$.

As $\gamma > 0$ we conclude that our solution is asymptotically stable.

VI. Higher order approximations and example.

Consider again the equation

$$\dot{x}(t) = L(x_t) + N(t, x_t, \mu) , \tag{28}$$

where N fulfills conditions i), ii) and iii) and moreover admits a large enough number of derivatives with respect to its arguments.

By means of the successive approximations given by (41) we can obtain in some cases the coefficients of the lower order terms in the expansion of

$$\mathcal{O}(\Psi^T(0)N(t, x_t(a, \mu), \mu) = F(a, \mu) \tag{64}$$

in terms of a and μ. Here (64) is the determining function for (28) and $x_t(a, \mu)$ represents the unique fixed point defined by (39).

We will show now how the knowledge of these terms may help us in determing the existence and the order of magnitude of a perioic solution of (28). This method may work even in the case in which the application of Theorem 3 has failed because $\det(D_a F(a_o, 0)) = 0$.

To simplify notation we will consider the scalar case with $a_o = 0$, i.e. we assume $F(0, 0) = 0$.

Suppose also that by means of (41) we have been able to obtain the lowest order terms in a and μ for $F(a, \mu)$. By this we mean that we can write

$$(67)\quad F(a, \mu) = \mu^{\nu}(k_o a^{m_o} + k_1 a^{m_1}\mu^{n_1} + \dots + k_p a^{m_p}\mu^{n_p}) + f(a, \mu) = \mu^{\nu} P(a,\mu) + f(a,\mu),$$

where $P(a, \mu)$ has been chosen in such a way that we take into account only the terms lying in the side of steepest slope of the Newton polygon, i.e., the terms for which $\nu n_j/(m_o - m_j)$ is a minimum. Let

$$(66)\qquad \lambda = \frac{n_j}{m_o - m_j}\ , \qquad j = 0, \dots, p\ .$$

If we now substitute

$$(67)\qquad a = \bar{a}\,\mu^{\lambda}$$

we obtain

$$\bar{F}(\bar{a}, \mu) = \mu^{\nu + \lambda m_o}(k_o \bar{a}^{m_o} + \dots + k_p \bar{a}^{m_p}) + f(\bar{a}, \mu) =$$

$$= \mu^{\nu + \lambda m_o}\,\bar{P}(\bar{a}) + f(\bar{a}, \mu).$$

where $f(a, \mu)$ is $o(\mu^{\nu+\lambda m_o})$ for a fixed $\bar{a}$.

If we want to find $\bar{a}(\mu)$ for μ sufficiently small such that $\bar{F}(\bar{a}(\mu), \mu) = 0$ we apply the implicit function theorem. Owing to the form of $\bar{F}(\bar{a}, \mu)$ what we have to do is solve for $\bar{a}$ in

(68) $$\overline{P}(\overline{a}) = 0 \quad \text{and check} \quad \frac{\partial}{\partial \overline{a}} \overline{P}(\overline{a}) \neq 0$$

at these values.

If we find such a value of $\overline{a}$ we get, by using (67), that there exists a solution of (28) which tends to 0 like

(69) $$a(\mu) = \overline{a}\,\mu^{\lambda}$$

as μ tends to zero.

In the case in which $F(a_o, 0) = 0$ for a_o different from zero the treatment is analogous, but expanding in terms of $a - a_o$. The same will apply for periodic solutions with amplitudes tending to ∞ when μ tends to zero. This corresponds to the case of negative λ. It can be treated by expanding in terms of the recipocal of a.

We present now an example due to J. K. Hale in which the above technique is utilized.

Consider

(66) $$\dot{x}(t) = (\frac{\pi}{2} + \mu)x(t-1)(1-x^2(t))$$

The unperturbed equation and its adjoint are given by

$$\dot{u}(t) = \frac{\pi}{2}\, u(t-1) \quad \text{and}$$
$$\dot{v}(s) = \frac{\pi}{2}\, v(s+1)$$

The bases for the generalized eigenvalues Φ and Ψ can be chosen as:

$$\Phi = (\varphi_1, \varphi_2),\ \varphi_1(\theta) = \sin\frac{\pi}{2}\,\theta,\ \varphi_2(\theta) = \cos\frac{\pi}{2}\,\theta,\ \theta \in [-1, 0]$$

$$\Psi^T = \frac{2}{\nu^2}\begin{pmatrix} \psi_1 - \frac{\pi}{2}\psi_2 \\ \frac{\pi}{2}\psi_1 + \psi_2 \end{pmatrix}$$

$$\nu^2 = \frac{1}{1+\frac{\pi^2}{4}}, \quad \psi_1(\theta) = \sin\frac{\pi}{2}\theta,\ \psi_2(\theta) = \cos\frac{\pi}{2}\theta,\ \theta \in [0, 1]$$

This choice has been made in order to have $(\Psi, \Phi) = I$, where here

$$(\psi, \varphi) = \psi(0)\varphi(0) - \frac{\pi}{2}\int_{-1}^{0} \psi(\xi+1)\varphi(\xi)d\xi$$

Equation (66) can then be written, by using

$x_t = \Phi y(t) + x_t^Q$, as

(67)
$$\begin{cases} y = By + \Psi^T(0)N(x_t, \mu) \\ x_t^Q = U(t)\varphi^Q + \int_0^t U(t-\tau)X_0^Q\, N(x_\tau, \mu)d\tau \end{cases},$$

where

$$B = \begin{pmatrix} 0 & -\pi/2 \\ \pi/2 & 0 \end{pmatrix}, \quad \Psi^T(0) = \begin{pmatrix} -\pi/\nu^2 \\ 2/\nu^2 \end{pmatrix}$$

$$N(x_t, \mu) = -\mu x(t-1)(1-x^2(t)) + \frac{\pi}{2}x(t-1)x^2(t) .$$

Let now $\omega(\mu) = -\frac{\pi}{2} + \mu\beta$ and

$$B(\omega(\mu)) = \begin{pmatrix} 0 & -\frac{\pi}{2} + \mu\beta \\ -\frac{\pi}{2} - \mu\beta & 0 \end{pmatrix} .$$

We preform the change of variables

$y = e^{B(\omega(\mu))t}z$ in (67) and we obtain:

$$(68) \qquad \dot{z}(t) = -\mu e^{-B(\omega(\mu))t}B(\beta)e^{B(\omega(\mu))t}z(t) + e^{-B(\omega(\mu))t}\Psi^T(0)N(x_t, \mu) .$$

Here we have

$$N(x_t, \mu) = \mu(z_1 \cos \omega t + z_2 \sin \omega t + x_t^Q(-1))(1-(-z_1 \sin \omega t + z_2 \cos \omega t + x_t^Q(0))^2) +$$
$$+ \frac{\pi}{2} (-z_1 \cos \omega t - z_2 \sin \omega t + x_t^Q (-1))(-z_1 \sin \omega t + z_2 \cos \omega t + x_t^Q(0))^2.$$

As the system is autonomous we can altogether forget about z_2, say, and we obtain for a vector with components $(a, 0)$ and for β the determining equations for $\mu = 0$

$$\frac{1}{\nu^2}\begin{pmatrix} \frac{\pi^2}{16} a^3 \\ -\frac{\pi}{8} a^3 \end{pmatrix} = 0$$

The only solutions is $a = 0$, but for this value the Jacobian with respect to a and β vanishes.

We look then for the lowest order terms.

In our determining equations we have terms like μa, μx_t^Q, a^3, $a^2(x_t^Q)^2$ $a(x_t^Q)^3$, etc.

We check first the order of x_t^Q. If x_t is periodic we have the representation

$$x_t^Q = \int_{-\infty}^{t} U(t-\tau)X_0^Q N(x_\tau, \mu)d\tau$$

As $N(x_\tau, \mu)$ has μa as its lowest order term it turns out that this is the order of x_t^Q. This means that the only terms to be considered are μa and a^3.

Taking these into account we obtain the determining equations:

$$\frac{\pi^2}{\nu^2\, 16}\; a^3 - \frac{\pi}{2\,\nu^2}\; a\mu = 0$$

$$\mu\beta a - \frac{\pi}{\nu^2\, 8}\; a^3 + \frac{1}{\nu^2}\; a\mu = 0$$

Hence

$$a = \sqrt{\frac{8\mu}{\pi}} \quad , \quad \beta = 0$$

We have for the jacobian with respect to a and β:

$$\begin{vmatrix} -\frac{3\pi^2 a^2}{16\nu^2} - \frac{\pi\mu}{2\nu^2} & \mu\,\beta - \frac{3\pi a^2}{8\nu^2} + \frac{\mu}{\nu^2} \\ 0 & \mu a \end{vmatrix}$$

which differs from zero for the value obtained for a.

We have then a solution close to

(69) $$x_t = \Phi e^{B(-\frac{\pi}{2})t} \begin{pmatrix} a \\ o \end{pmatrix} \;, \quad a = \sqrt{\frac{8\mu}{\pi}} \quad .$$

If we substitute x by $\sqrt{\mu}x$ in equations (66) and we get a problem which is solvable in the first approximation:

(70) $$x(t) = -\left(\frac{\pi}{2} + \mu\right)x(t-1)(1- \mu x^2(t))$$

The bifurcation equations turn out to be

$$\begin{pmatrix} 0 \\ \beta a \end{pmatrix} + \frac{1}{\nu^2} \begin{pmatrix} \frac{\pi^2}{16} a^3 - \frac{\pi}{2} a \\ - \frac{\pi}{8} a^3 + a \end{pmatrix} = 0$$

Hence $a = 0$ is a solution, the same as $a = \sqrt{\frac{8}{\pi}}$. For this last value the jacobian differs from zero and this proves that for μ small enough there is a periodic solution of (70) tending to (69) with $a = \sqrt{\frac{8}{\pi}}$, or, what it is the same, a solution of (66) tending to (69).

References

[1] Hale, J. K., "Linear Functional-differential Equations with Constant Coefficients", Contr. to Differential Equations, 2(1963), pp. 291-317.

[2] Hale, J. K., Oscillations in Nonlinear Systems, McGraw-Hill, New York, 1963.

[3] Cesari, L., "Functional Analysis and Periodic Solutions of Nonlinear Differential Equations", Contr. to Differential Equations, 1(1963), pp. 149-187.

[4] Knobloch, H. W., "Remarks on a paper of L. Cesari on Functional Analysis and Nonlinear Differential Equations", Michigan Math. J., 10(1963), pp. 417-430.

[5] Stokes, A., "A Floquet Theory for Functional Differential Equations", Proc. Nat. Acad. Sci., 48(1962), pp. 1330-1334.

[6] Shimanov, S. N., "On the Theory of Linear Differential Equations with Periodic Coefficients and Time Lag", Prikl. Mat. Meh. 27(1963), pp. 450-458 (Russian); translated as J. Appl. Math. Mech. 27(1963), pp. 674-687.

[7] Krasovskii, N. N., Stability of Motion, Stanford University Press, 1963.

[8] Halanay, A., Teoria Calitativa a ecuatiilor diferentiale , Editura Academiei Republicii Populare Romine, 1963.

[9] Halanay, A., "Periodic Solutions of Linear Systems with Lag", Revue de Math. Pures et Appl., 6(1961), pp. 141-158 (Russian).

[10] Hale, J. K. and Perelló, C., "The Neighborhood of a Singular Point of Functional-Differential Equations", Contr. to Differential Equations, 3(1964), pp. 351-375.

[11] Stokes, A., "On the Stability of a Limit Cycle of an Autonomous Functional Differential Equation", Contr. to Differential Equations, 3(1964), pp. 121-139.

[12] Bellman, R. and Cooke, K. L., Differential-Difference Equations, Academic Press, New York, 1963.

[13] Perelló, C., "Periodic Solutions of Ordinary Differential Equations with and without Time Lag", Thesis, Brown University, June 1965.

Carlos Perelló
Center for Dynamical Systems
Division of Applied Mathematics
Brown University

PAPER [16]

To appear in the SIAM Journal on Numerical Analysis
and appeared as CDS-TR 66-6.

SOME STABILITY THEOREMS FOR ORDINARY DIFFERENCE EQUATIONS*

by

James Hurt

*The author has received financial support from the National Aeronautics and Space Administration under their Traineeship Program and the preparation and reproduction of this paper was supported by NASA under Contract No. NAS8-11264.

SOME STABILITY THEOREMS FOR ORDINARY DIFFERENCE EQUATIONS

by

James Hurt

LaSalle [1,2,3] and others have developed a generalization of the "second method" of Liapunov which utilizes certain invariance properties of solutions of ordinary differential equations. Invariance properties of solutions of ordinary difference equations are utilized here to develop stability theorems similar to those in LaSalle [1]. As illustrations of the application of these theorems, a region of convergence is derived for the Newton-Raphson and Secant iteration methods. A modification of one of these theorems is given and applied to study the effect of round-off errors in the Newton-Raphson and Gauss-Seidel iteration methods.

I. INTRODUCTION. An ordinary difference equation is an equation of the type given in (1),

$$x(k+1) = f(k, x(k)) \tag{1}$$

where each x and $f(k,x)$ are elements of X, an n-dimensional vector space. Since the notation used in (1) can become very clumsy, the somewhat neater E notation is used. If E is defined as the operator where $Ex(k) = x(k+1)$, then equation (1) can be written as in (1*)

$$Ex = f(k,x) \tag{1*}$$

where the arguments of x and Ex are understood to be k.

A function $x(k;k_o,x_o)$ is called a <u>solution</u> of the difference equation (1) if it satisfies the following three conditions.

a) $x(k;k_o,x_o)$ is defined for $k_o \leqq k \leqq k_o + K$ for some integer $K > 0$.

b) $x(k_o;k_o,x_o) = x_o$, the initial vector.

c) $x(k+1;k_o,x_o) = f(k,x(k;k_o,x_o))$ for $k_o \leqq k \leqq k_o + K-1$.

Hereafter, it is assumed that a solution to (1) exists and is unique for all $k \geqq k_o$ and that this solution is continuous in the initial vector x_o. More specifically, if $\{x_n\}$ is a sequence of n-vectors with $x_n \to x_o$ as $n \to \infty$, then the solutions through x_n converge to the solution through x_o:

$$x(k;k_o,x_n) \to x(k;k_o,x_o) \quad \text{as} \quad n \to \infty$$

For all k on any compact interval, this convergence is assumed to be uniform.

For any n-vector x, let $|x|$ denote any vector norm of x. For any non-empty set of n-vectors A, denote the distance from

x to A by $d(x,A)$.

$$d(x,A) = \inf \{|x-y| : y \in A\}.$$

Introduce the vector ∞ to X and define $d(x,\infty) = |x|^{-1}$. Let $A^* = A \cup \{\infty\}$ and $d(x,A^*) = \min \{d(x,A), d(x,\infty)\}$.

A point $p \in X$ is a <u>positive limit point</u> of $x(k)$ if there is a sequence $\{k_n\}$ with $k_{n+1} > k_n \to \infty$, and $x(k_n) \to p$ as $n \to \infty$. The union of all the positive limit points of $x(k)$ is the <u>positive limit set</u> of $x(k)$.

II. <u>THE GENERAL STABILITY THEOREM</u>. Let G be any set in the vector space X. G may be unbounded. Let $V(k,x)$ and $W(x)$ be real valued functions defined for all $k \geqq k_0$ and all x in G. If $V(k,x)$ and $W(x)$ are continuous in x, $V(k,x)$ is bounded below, and

$$\Delta V(k,x) = V(k+1,f(k,x))-V(k,x) \leqq - W(x) \leqq 0$$

for all $k \geqq k_0$ and all x in G, then V is called a <u>Liapunov function</u> for (1) on G. Let $\overline{G}$ be the closure of G, including ∞ if G is unbounded, and define the set A by (2).

$$A = \{x \in \overline{G} : W(x) = 0\} \qquad (2)$$

The following result is the difference analog to Theorem 1 in LaSalle [1].

THEOREM 1. If there exists a Liapunov function V for (1) on G, then each solution of (1) which remains in G for all $k \geqq k_o$ approaches the set $A^* = A \cup \{\infty\}$ as $k \to \infty$.

PROOF: Let $x(k)$ be a solution to (1) which remains in G for all $k \geqq k_o$. Then, by assumption, $V(k,x(k))$ is a monotone non-increasing function which is bounded from below. Hence, $V(k,x(k))$ must approach a limit as $k \to \infty$, and $W(x(k))$ must approach zero as $k \to \infty$. From the definition of A^* and the continuity of $W(x)$, we get $d(x(k),A^*) \to 0$ as $k \to \infty$. Note that if G is unbounded and there exists a sequence $\{x_n\}$ such that $x_n \in G$, $|x_n| \to \infty$, and $W(x_n) \to 0$ as $n \to \infty$, then it is possible to have an unbounded solution under the conditions of the theorem. If G is bounded or if $W(x)$ is bounded away from zero for all sufficiently large x, then all solutions which remain in G are bounded and approach a closed, bounded set contained in A as $k \to \infty$.

This theorem can be used to easily prove all of the usual Liapunov stability theorems. See, for example, Hahn [1] and Kalman and Bertram [1]. For example, if G is the entire space X and $W(x)$ is positive definite, then $A = \{o\}$ and all solutions approach the origin as $k \to \infty$. However, as the following example shows, other considerations can be used to determine if a solution $x(k)$ will remain in G. The difference equation is given in equation (3).

$$Ex = x^{-2} \quad \text{for} \quad x > 0 \tag{3}$$

Let the set G be the set of positive numbers. Then, if $x > 0$, we get $Ex > 0$ from equation (3) and all solutions which start in G remain in G. The function $V(k,x) = V(x)$

$$V(x) = \frac{x}{1+x^2}$$

is a Liapunov function for (3) on G since $V(x) \geqq 0$ and

$$\Delta V(x) = \frac{x^{-2}}{1+x^{-4}} - \frac{x}{1+x^2} = \frac{x(1-x)(x^3-1)}{(1+x^2)(1+x^4)} = -W(x) \leqq 0$$

We have $W(x) = 0$ when $x = 0$, $x = 1$, and $W(x) \to 0$ as $x \to \infty$. Thus, the set A^* is the set $\{0, 1, \infty\}$. Each solution with $x_0 > 0$ approaches A^* as $k \to \infty$. A look at the solutions to (3)

$$x(k) = x_0^{(-2)^k}$$

shows that this is exactly the case. If $x_0 = 1$, then $x(k) = 1$ for all k. If $x_0 < 1$, then $x(k) \to 0$ for even k and $x(k) \to \infty$ for odd k.

Quite often the set G can be constructed so that all solutions which start in some smaller set G_1 remain in G. One such case is covered in the following corollary.

COROLLARY 1. Let $u(x)$ and $v(x)$ be continuous real-valued functions. Let $V(k,x)$ be such that

$$u(x) \leqq V(k,x) \leqq v(x)$$

for all $k \geqq k_o$. For some η, define the sets $G = G(\eta)$ and $G_1 = G_1(\eta)$ as

$$G(\eta) = \{x : u(x) < \eta\}$$

$$G_1(\eta) = \{x : v(x) < \eta\}$$

If V is a Liapunov function for (1) on $G(\eta)$, then all solutions which start in $G_1(\eta)$ remain in $G(\eta)$ and approach A as $k \to \infty$.

PROOF: Let $x(k)$ be a solution of (1) with $x(k_o) \in G_1(\eta)$. Then

$$u(x(k)) \leqq V(k,x(k)) \leqq V(k_o,x(k_o)) \leqq v(x(k_o)) < \eta$$

for all $k \geqq k_o$, implying that $x(k) \in G(\eta)$ for all $k \geqq k_o$.

Theorem 1 and Corollary 1 give sufficient conditons for the positive limit set of a solution $x(k)$ to be contained in A.

There is an art to finding the best V, W, u, and v, i.e., the functions V, W, u, and v which give the largest G, the largest G_1, and the smallest A. Often more information about the behavior of the solutions can be obtained by considering several different Liapunov functions and combining the results from each.

The following example is taken from Vidal and Laurent [1]. The sampled control systems covered in this paper are described by the difference equation (4).

$$Ex = M(k,x)x \tag{4}$$

where $M(k,x)$ is a matrix. For any vector norm, $|x|$, define the norm of the matrix $M(k,x)$ by

$$|M(k,x)| = \min \{b : |M(k,x)y| \leq b|y| \quad \text{for all} \quad y \neq 0\}$$

Then clearly, $|M(k,x)x| \leq |M(k,x)||x|$. For the difference equation (4), try the Liapunov function $V(k,x) = |x|$. Then

$$\begin{aligned}\Delta V(k,x) &= |M(k,x)x| - |x| \\ &\leq (|M(k,x)| - 1)|x| .\end{aligned}$$

Let $u(x) = v(x) = V(k,x) = |x|$, then $G_1(\eta) = G(\eta) = \{x:|x| < \eta\}$. For all x in $G(\eta)$ and all $k \geq k_0$ let $|M(k,x)| \leq a(x)$ and

$W(x) = +(1-a(x))|x|$. Then we have

$$\Delta V(k,x) \leqq -W(x) .$$

If $a(x) < 1$ for all x in $G(\eta)$, then $-W(x) \leqq 0$, the set A is the origin and possibly something on the boundary of $G(\eta)$. Since $V(k,x(k))$ is a non-increasing function of k and the boundary of $G(\eta)$ is a level surface of $V(k,x)$, the solutions cannot approach the boundary of $G(\eta)$. Hence, all solutions which start in $G(\eta)$ remain in $G(\eta)$ and approach the origin as $k \to \infty$. The set $G(\eta)$ is called a <u>domain of stability</u> for the system (4). The best $G(\eta)$ is chosen by picking η as large as possible without violating the inequality $a(x) < 1$ for all x in $G(\eta)$.

Various choices for the vector norm will result in various $a(x)$ and various domains of stability. Since each is sufficient, the union of all these domains of stability is also a domain of stability.

If $M(k,0)$ is a constant matrix, independent of k, and the spectral radius of $M(k,0)$ is less than one, then there is a vector norm such that $a(x)$ is continuous in x and $a(0) < 1$, indicating that there is a non-empty domain of stability (see the Appendix).

The following example illustrates that the results obtained in Theorem 1 and Corollary 1 are the best possible without further assumptions. The difference equation is (5).

$$\begin{aligned} Ex &= y \\ Ey &= a^2x + p(k)y \end{aligned} \tag{5}$$

where $0 < a < 1$ and $0 < \delta \leqq p(t) < 1-a^2$. If $p(k) = p$, a constant, then the conditons for stability are satisfied and all solutions approach the origin as $k \to \infty$.

Try the Liapunov function

$$V(k,x,y) = a^2x^2 + y^2$$

Then

$$\begin{aligned} \Delta V(k,x,y) &= -a^2p(k)(x-y)^2 + a^2(p(k)-(1-a^2))x^2 \\ &\quad + (p(k)+1)(p(k)-(1-a^2))y^2 \\ &\leqq -a^2p(k)(x-y)^2 \leqq -a^2\delta(x-y)^2 = -W(x,y) \leqq 0 \ . \end{aligned}$$

From Corollary 1, we see that all solutions are bounded and $x(k)-y(k) \to 0$ as $k \to \infty$.

If

$$p(k) = \frac{1-a^2}{1+a^{k+1}}$$

for all $k \geqq 0$, then this $p(k)$ satisfies the conditions given above and one solution of the difference equation (5) is

$$x(k) = 1 + a^k \to 1 \quad \text{as} \quad k \to \infty$$

$$y(k) = 1 + a^{k+1} \to 1 \quad \text{as} \quad k \to \infty\,.$$

The results obtained are the best possible. Notice, however, that this $p(k)$ approaches $1-a^2$ as $k \to \infty$. If, instead of $p(k) < 1-a^2$, we knew that $p(k) \leq 1-a^2 - \epsilon$ for some $\epsilon > 0$, then we get

$$\Delta V(k,x,y) \leq -a^2\delta(x-y)^2 - a^2\epsilon x^2 - (1+\delta)\epsilon\, y^2 = -W_1(x,y) \leq 0$$

and the only point where $W_1(x,y) = 0$ is $x = y = 0$. In this case, all solutions approach the origin as $k \to \infty$.

III. <u>AUTONOMOUS DIFFERENCE EQUATIONS</u>. If the function $f(k,x)$ in (1) is independent of k, then the difference equation is said to be <u>autonomous</u>, as in equation (6).

$$Ex = f(x)\,. \tag{6}$$

Just as is the case for autonomous differential equations, solutions to (6) are essentially independent of k_0 so we assume $k_0 = 0$ and write the solution as $x(k;x_0)$. A function $x^*(k)$ is said to be a solution for (6) on $(-\infty,\infty)$ if, for any k_0 in $(-\infty,\infty)$, we have for all $k \geqq k_0$

$$x(k-k_0;x^*(k_0)) = x^*(k).$$

A set B is an invariant set of (6) if $x_0 \in B$ implies that there is a solution $x^*(k)$ for (6) on $(-\infty,\infty)$ such that $x^*(k) \in B$ for all k and $x^*(0) = x_0$.

LEMMA 1. The positive limit set B of any bounded soltuion of (6) is a nonempty, compact, invariant set of (6).

PROOF: Let $x(k)$ be a bounded solution of (6) and B its positive limit set. For each $p \in B$, there is a monotone sequence of integers $\{k_n\}$ such that $k_n \to \infty$ and $x(k_n) \to p$ as $n \to \infty$. Then each function $y_n(k) = x(k+k_n)$ is a solution of (6) with $y_n(0) \to p$ as $n \to \infty$. From continuity in the initial conditions, these functions approach the solution $x(k;p)$ as $n \to \infty$. By extending each function $y_n(k)$ to $-k_n$, we can extend the solution $x(k;p)$ to $-\infty$. The simultaneous convergence to $x(k;p)$ and B implies that $x(k;p) \in B$ for all k, and so B is an invariant set. The fact that B is nonempty and compact is obtained from the definition of a positive limit set and the boundedness of $x(k)$.

For an autonomous equation, Theorem 1 can be strengthened as follows.

THEOREM 2. If there exists a Liapunov function $V(x)$ for (6) on some set G, then each solution $x(k)$ which remains in G is either unbounded or approaches some invariant set contained in A as $k \to \infty$.

PROOF. From Theorem 1, $x(k) \to A \cup \{\infty\}$ as $k \to \infty$. If $x(k)$ is unbounded, then Lemma 1 does not hold. If $x(k)$ is bounded, then its positive limit set is an invariant set.

If the set M is defined as the union of all the invariant sets contained in A, then $x(k) \to M$ as $k \to \infty$ whenever $x(k)$ remains in G and is bounded. The set M may be considerably smaller than the set A. Under the conditions of Theorem 2, an unbounded solution can exist only if G is unbounded and there is a sequence $\{x_n\}$, $x_n \in G$, $x_n \to \infty$ and $\Delta V(x_n) \to 0$ as $n \to \infty$.

Corollary 1 can be restated in a similar manner.

COROLLARY 2. If, in Theorem 2, the set G is of the form

$$G = G(\eta) = \{x : V(x) < \eta\}$$

for some $\eta > 0$, then all solutions which start in G remain in G and approach M as $k \to \infty$.

This corollary can be used to obtain regions of convergence for various iterative methods which can be described by an autonomous difference equation. A <u>region of convergence</u> is a set $G \subset X$ such that, if $x(0) \in G$, then $x(k) \in G$ for all $k \geqq 0$ and $x(k)$ approaches the desired vector as $k \to \infty$. The largest region of convergence is the union of all regions of convergence. The Secant and Newton-Raphson methods are treated as examples. For a derivation and discussion of these methods see, for example, Traub [1] or Ostrowski [1].

The Secant method for finding a root of $f(z) = 0$ ($f(z)$ and z are complex numbers) is given by assuming values for z_1 and z_2, then forming the sequence $\{z_k\}$ by repeated application of equation (7).

$$z_{k+2} = z_{k+1} - \frac{(z_{k+1}-z_k)f(z_{k+1})}{f(z_{k+1})-f(z_k)} \tag{7}$$

We assume that, for every k, $z_{k+1} \neq z_k$ and $f(z_{k+1}) \neq f(z_k)$, so this iteration formula is well defined for all k. Let α be the desired root of $f(z) = 0$ and let

$$f(\alpha + e) = f'(\alpha)e + g(\alpha, e)e^2 .$$

Then, letting $z_k = \alpha + e_k$ for each k, equation (7) becomes

$$e_{k+2} = M(\alpha, e_k, e_{k+1})e_k e_{k+1}$$

where

$$M(\alpha, e_k, e_{k+1}) = \frac{g(\alpha, e_{k+1})e_{k+1} - g(\alpha, e_k)e_k}{f(\alpha+e_{k+1}) - f(\alpha+e_k)} \quad .$$

With the assumption that α is a simple root of $f(z) = 0$ and $g(\alpha, e)$ is continuous and bounded in e, then $M(\alpha, e_k, e_{k+1})$ is continuous and bounded for e_k, e_{k+1} small enough.

The difference equation (8) is obtained by letting $x_1(k) = e_k$ and $x_2(k) = e_{k+1}$.

$$\begin{aligned} Ex_1 &= x_2 \\ Ex_2 &= M(\alpha, x_1, x_2)x_1 x_2 \end{aligned} \qquad (8)$$

Consider the Liapunov function $V_q(x_1, x_2) = |x_1|^q + |x_2|^q$ for some $q \geqq 1$. Then

$$\Delta V_q(x_1, x_2) = -(1 - |M(\alpha, x_1, x_2)x_2|^q)|x_1|^q$$

and $\Delta V_q(x_1, x_2) \leqq 0$ if $|M(\alpha, x_1, x_2)x_2| \leqq 1$. Let $G_q(\eta)$ be the set $G_q(\eta) = \{(x_1, x_2) : (|x_1|^q + |x_2|^q)^{1/q} < \eta\}$. Since $x_2 = 0$ implies $|M(\alpha, x_1, x_2)x_2| = 0 < 1$, there is some $\eta > 0$ such that $|M(\alpha, x_1, x_2)x_2| \leqq 1$ for all (x_1, x_2) in $G_q(\eta)$. From Corollary 2, this $G_q(\eta)$ is a region of convergence for the Secant method. If the initial guesses z_1 and z_2 are such that $(x_1, x_2) \in G_q(\eta)$ for some q, then (x_1, x_2) will remain in $G_q(\eta)$ for all k and approach an invariant set contained in the set

$A = \{(x_1,x_2) \in \overline{G}_q(\eta) : x_1 = 0\}$. The only invariant set of equation (8) with $x_1 = 0$ is the origin $x_1 = x_2 = 0$, so we get $(x_1,x_2) \to (0,0)$ as $k \to \infty$, and the method converges.

If, for $|e| \leq \eta_o$, we get

$$|f'(\alpha+e)| \geq F , \quad |g(\alpha,e)| \leq G$$

then we get that $|M(\alpha,x_1,x_2)x_2| < 1$ if $|x_2| < F/G$. Thus, η can be taken as the smaller of η_o and F/G. For the particular equation $f(z) = z^2-\alpha^2$, we get $|f'(\alpha+e)| \geq 2|\alpha|-2\eta_o$ and $|g(\alpha,e)| = 1$ for $|e| < \eta_o$. In this case, we can choose $\eta = \eta_o = \frac{2}{3}|\alpha|$.

It should be noted that the set $G_q(\eta)$, or even the union of these sets for all $q \geq 1$, is not always the largest region of convergence. For the simple equation $f(z) = z^2-\alpha^2$, almost any choice of z_1, z_2, provided only that $z_1 \neq z_2$ and $f(z_1) \neq f(z_2)$, will lead to a sequence which will converge either to $+\alpha$ or to $-\alpha$. However, if z_1 and z_2 are in the region defined by $G_q(\eta)$, then not only will the sequence converge to α but this convergence will be uniform in the sense that $|z_k-\alpha|^q + |z_{k+1}-\alpha|^q$ will be a decreasing function of k.

Corollary 2 can also be used to find a region of convergence for the Newton-Raphson method. The Newton-Raphson method for finding a root of $f(z) = 0$ ($f(z)$ and z are n-vectors) is given by assuming a value for z_1, then forming the sequence $\{z_k\}$

by repeated application of equation (9).

$$z_{k+1} = z_k - [\frac{\partial f}{\partial z}(z_k)]^{-1} f(z_k) \tag{9}$$

where $\frac{\partial f}{\partial z}(z_k)$ is the matrix of partial derivatives of f. Here, we assume that $[\frac{\partial f}{\partial z}(z_k)]$ always has an inverse. If the desired root is a simple root, then this is the case. By letting α be the desired root, expanding $f(\alpha+e)$ as

$$f(\alpha+e) = [\frac{\partial f}{\partial z}(\alpha)]e + f_o(e)$$

and letting $z_k = \alpha+e_k$, then the difference equation becomes

$$e_{k+1} = +M_1(e_k)[M_2(e_k)e_k - f_o(e_k)]$$

where $M_1(e) = [\frac{\partial f}{\partial z}(\alpha+e)]^{-1}$ and $M_2(e) = [\frac{\partial f}{\partial z}(\alpha+e) - \frac{\partial f}{\partial z}(\alpha)]$. Let $|e|$ be some vector norm (see the Appendix). If α is a simple root of $f(z) = 0$ and f is twice continuously differentiable at $z = \alpha$, then, for each $\eta > 0$, there exists a positive constant $k(\eta)$ such that, for all e with $|e| < \eta$, we have $|M_1(e)(M_2(e)e-f_o(e))| \leq k(\eta)|e|^2$. Then, letting $V(e) = |e|$, we get

$$\Delta V(e) \leq -(1-k(\eta)|e|)|e|$$

and $\Delta V(e) \leq 0$ if $k(\eta)|e| \leq 1$. Using Corollary 2, we get a region of convergence $G(\eta_o) = \{z : |z-\alpha| < \eta_o\}$ where $\eta_o = \min(\eta, \frac{1}{k(\eta)})$. We can choose η so as to maximize η_o, thus obtaining the best region of convergence obtainable with this Liapunov function.

For the case where z and $f(z)$ are complex numbers, if there is some $F > 0$ such that $|f_o(e)| \leq F|e|^2$ for all z where $|z-\alpha| \leq \eta$, some $\eta \geq \frac{2}{5}\frac{|f'(\alpha)|}{F}$, then we can get

$$k(\eta) = \frac{3}{2}\frac{F}{|f'(\alpha)|-F\eta}$$

and the best (with this $k(\eta)$) region of convergence is given by $G(\eta_o)$ where

$$\eta_o = \frac{2}{5}\frac{|f'(\alpha)|}{F}$$

For the simple case $f(z) = z^2-\alpha^2$, we get $\eta_o = \frac{2}{5}|\alpha|$. However, a sharper estimate may be used for $k(\eta)$ which results in $\eta_o = \frac{2}{3}|\alpha|$. This latter case is the best possible. Any disc centered at α with radius larger than $\frac{2}{3}|\alpha|$ will have points inside the disc which will map outside the disc on the next iteration and $\Delta V(x)$ is positive for some values of x.

It should be noted that the region of convergence $G(\eta_o)$ is not always the largest region of convergence. For the simple equation $f(z) = z^2-\alpha^2$, any initial guess $z_1 \neq 0$ will lead to a sequence $\{z_k\}$ which will converge either to $+\alpha$ or to $-\alpha$.

IV. PERIODIC DIFFERENCE EQUATIONS. If, in the difference equation (1), $f(k,x)$ is T-periodic for some integer $T \geqq 1$ and fixed x, i.e., $f(k+T,x) = f(k,x)$ for all k,x, then the difference equation is said to be a T-periodic difference equation. A function $x^*(k)$ is said to be a solution for (1) on $(-\infty,\infty)$ if, for any k_o in $(-\infty,\infty)$, we have for all $k \geqq k_o$

$$x(k;k_o,x^*(k_o)) = x^*(k)$$

A set B is an invariant set of (1) if $x_o \in B$ implies that there is a k_o and a solution $x^*(k)$ for (1) on $(-\infty,\infty)$ such that $x^*(k_o) = x_o$ and $x^*(k) \in B$ for all k.

LEMMA 2. Let $x(k)$ be a solution of (1) that is bounded for all $k \geqq k_o$. Then the positive limit set of $x(k)$ is an invariant set of (1).

PROOF: This lemma is proven in a manner very similar to that used in Lemma 1. The k_o used in the definition of an invariant set is obtained in the following manner. If $\{k_n\}$ is a monotone sequence such that $x(k_n) \to p \in B$, the positive limit set of $x(k)$, then there is a sequence of integers $\{M_n\}$ such that $k_n-M_nT \in [0,T)$ for all n. The set $[0,T)$

consists of a finite number of integers, so at least one of these integers, k_o, must satisfy $k_o = k_n - M_n T$ for an infinite number of n's. The solution $x(k;k_o,p)$ is the shown to be the limit of the functions $y_n(k) = x(k+k_n)$ and is in B for all k, thus demonstrating that B is an invariant set of (1).

Theorem 1 can now be restated for T-periodic difference equations.

THEOREM 3. Let $V(k,x)$ be a T-periodic, continuous function which is bounded below for all x in some set G. For $k \geqq k_o$ and x in G, let $\Delta V(k,x) \leqq 0$ and define the set A by $A = \{(k,x) : \Delta V(k,x) = 0,\ x \in \overline{G}\}$. Let M be the union of all solutions $x(k)$ of (1) such that $(k,x(k)) \in A$ for all k. Then each solution of (1) which remains bounded and in G for all $k \geqq k_o$ approaches some invariant set contained in M as $k \to \infty$.

PROOF. The function $V(k,x(k))$ is non-increasing and bounded below, hence $\Delta V(k,x(k)) \to 0$ as $k \to \infty$. The continuity of V and ΔV implies that $d((k,x(k)),A) \to 0$ as $k \to \infty$. Since $x(k)$ must approach an invariant set as $k \to \infty$, it must approach M as $k \to \infty$.

An unbounded solution is possible under the conditions of Theorem 3 only if G is unbounded and there exists a sequence $\{(k_n,x_n)\}$ with $|x_n| \to \infty$, and $\Delta V(k_n,x_n) \to 0$ as $n \to \infty$. If G is bounded or if $\Delta V(k,x)$ is bounded away from zero for all sufficiently large x, then all solutions of (1) which remain in G are bounded and approach M as $k \to \infty$.

V. ASYMPTOTICALLY AUTONOMOUS DIFFERENCE EQUATIONS. If the difference equation (1) can be written in the form of equation (10)

$$Ex = H(x) + F(k,x) \tag{10}$$

where $F(k,x) \to 0$ as $k \to \infty$ uniformly for all x in any compact set, the difference equation is said to be an asymptotically autonomous difference equation. With each asymptotically autonomous difference equation, (10), there is the associated autonomous difference equation (11).

$$Ex = H(x) \tag{11}$$

LEMMA 3. The positive limit set of any bounded solution of the asymptotically autonomous difference equation (10) is an invariant set of the autonomous difference equation (11).

This lemma is proven in the same manner as Lemma 1. Theorem 1 could now be restated in a manner similar to Theorem 2, but the following, more general statement has proven more useful in its applications.

THEOREM 4. If a solution $x(k)$ of the difference equation (1) approaches a closed, bounded set A as $k \to \infty$, and if $x(k)$ is also a solution of the asymptotically autonomous difference equation (10), then it approaches the largest invariant set of (11) contained in A as $k \to \infty$.

As an example of the application of Theorem 4, consider the difference equation (12),

$$\begin{aligned} Ex &= cx - s(1-p(k))y \\ Ey &= sx + c(1-p(k))y \end{aligned} \qquad (12)$$

where $c = \cos\omega$, $s = \sin\omega$, $0 < \omega < 2\pi$, $0 < \delta \leq p(k) \leq 2-\epsilon < 2$.

With the Liapunov function $V(x,y) = x^2+y^2$, we get

$$\Delta V(x,y) = -p(k)(2-p(k))y^2 \leq -\delta\,\epsilon\, y^2 \leq 0$$

Applying Corollary 1, we get that all solutions for (12) are bounded and $y(k) \to 0$ as $k \to \infty$.

Let $x_1(k)$, $y_1(k)$ be a solution for (12), then $y_1(k)$ is bounded and approaches 0 as $k \to \infty$. Also, $x_1(k)$, $y_1(k)$ is a solution of the difference equation (13).

$$\begin{aligned} Ex &= cx - sy + p(k)y_1(k) \\ Ey &= sx + cy - p(k)y_1(k) \end{aligned} \qquad (13)$$

This difference equation is asymptotically autonomous to the difference equation (14).

$$
\begin{aligned}
Ex &= cx - sy \\
Ey &= sx + cy
\end{aligned}
\tag{14}
$$

The only invariant set of (14) with $y = 0$ is the origin $x=y=0$ since $0 < \omega < 2\pi$. By Theorem 4, all solutions of (12) approach this invariant set, the origin, as $k \to \infty$.

VI. PRACTICAL STABILITY. For many difference equations a solution is considered a stable solution if it enters and remains in a sufficiently small set. For example, under the proper conditions all solutions of the Newton-Raphson equation (9) approach the desired solution as $k \to \infty$. But, when the effects of round-off errors are considered this is no longer the case. However, if all the solutions become and remain close to the desired solution, then the method is judged to be satisfactory. This type of stability is called practical stability. The following theorem and corollaries are concerned with practical stability for the difference equation (15).

$$Ex = f(k,x) \tag{15}$$

THEOREM 5. Given a set $G \subset X$, possibly unbounded. Let $V(x)$ and $W(x)$ be continuous, real valued functions defined on G

and such that, for all k and all x in G,

$$\text{(i)} \quad V(x) \geqq 0$$

$$\text{(ii)} \quad \Delta V(k,x) = V(f(k,x)) - V(x) \leqq W(x) \leqq a$$

for some constant $a \geqq 0$. Let the set S be the set

$$S = \{x \in \overline{G} : W(x) \geqq 0\}$$

Let $b = \sup\{V(x) : x \in S\}$ and the set A be the set

$$A = \{x \in \overline{G} : V(x) \leqq b + a\}$$

Then any solution $x(k)$ which remains in G and enters A when $k = k_1$ remains in A for all $k \geqq k_1$.

The properties of S, A, and $V(x)$ are used to show that, if $x(k)$ is in A, then $x(k+1)$ is in A. The theorem follows by induction.

COROLLARY 3. If $\delta = \sup\{-W(x) : x \in G-A\} > 0$, then each solution $x(k)$ of (15) which remains in G enters A in a finite number of steps.

If $x(k)$ does not enter A in a finite number of steps, then

$$V(x(k)) = V(x(k_o)) + \sum_{n=k_o}^{k-1} \Delta V(k, x(n))$$

$$\leqq V(x(k_o)) - (k-k_o)\delta$$

and $V(x(k)) \to -\infty$ as $k \to \infty$, a contradiction since $V(x) \geq b+a$ for all x in $G-A$.

COROLLARY 4. If G is of the form $G = G(\eta) = \{x : V(x) < \eta\}$ and the conditons of Theorem 5 and Corollary 3 are satisfied, then all solutions which start in G remain in G and enter A in a finite number of steps.

Corollary 4 can be used to study the effects of round-off errors in the Newton-Raphson method. Without errors, the Newton-Raphson method is given by equation (9). With errors, this method is given by equation (16).

$$z_{k+1} = z_k - [\frac{\partial f}{\partial z}(z_k)]^{-1} f(z_k) + h(k, z_k) \tag{16}$$

where all that is known about the error term $h(k, z_k)$ is its upper bound, say $|h(k, z_k)| \leq \epsilon$ for some vector norm and some $\epsilon > 0$. A value for ϵ can be obtained by assuming that z_k is known exactly and studying the steps of the computations in great detail to estimate the error in z_{k+1}. This error term includes the effects of errors in the functions $f(z)$ and $\frac{\partial f}{\partial z}(z)$, errors in evaluating $f(z)$ and $[\frac{\partial f}{\partial z}(z)]^{-1}$, and any other errors that may be encountered. Often it is not very difficult to find an estimate for ϵ, the problem is to determine the net effect of the term $h(k, z)$ on the positive limit set of a solution $z(k)$.

With the same assumptions on $f(z)$ and the same expansions used before, the difference equation (16) becomes

$$e_{k+1} = M_1(e_k)[M_2(e_k)e_k - f_o(e_k)] + h_1(k, e_k)$$

With $V(e) = |e|$, we get

$$\Delta V(e) \leq -(1-k(\eta)|e|)|e| + \epsilon = + W(e) \leq \epsilon$$

The set S becomes

$$S = \{e : W(e) \geq 0\} = \{e : |e| \leq b\}$$

where

$$b(\eta) = b = \frac{1 - \sqrt{1-4k(\eta)\epsilon}}{2k(\eta)} = \epsilon + 2k(\eta)\epsilon^2 + \ldots$$

provided that $4k(\eta)\epsilon < 1$. If $4k(\eta)\epsilon \geq 1$, then $W(e) \geq 0$ everywhere and the iterations may not converge. The set A is defined by

$$A = \{e : V(e) \leq b + \epsilon\} = \{e : |e| \leq b + \epsilon\}$$

We note that, for η small enough, we have

$$W(e) \leq -(b - k(\eta)(b+\epsilon)^2) = -\delta .$$

From Corollary 4, we have, if $\delta > 0$, then all solutions which start in $G(\eta_o)$ remain in $G(\eta_o)$, enter A in a finite number of interations, and remain in A thereafter. Here, η_o is not quite the same as before. We must choose η_o such that $k(\eta_o)\eta_o \leqq 1$, $4k(\eta_o)\epsilon < 1$, and $b(\eta_o)-k(\eta_o)(b(\eta_o)+\epsilon)^2 > 0$. If η_1 is the smallest positive solution of $\eta_1 k(\eta_1) = 1$, then choosing $\eta_o < \eta_1$ will satisfy both $k(\eta_o)\eta_o < 1$ and $b-k(\eta_o)(b+\epsilon)^2 > 0$. The condition $4k(\eta_o)\epsilon < 1$ becomes a condition on the precision or accuracy required in the computations.

Thus one effect of round-off errors is to reduce the region of convergence. Another effect of round-off errors is that the error of each z_k cannot generally be reduced much below the value $b+\epsilon = 2\epsilon + 2k(\eta)\epsilon^2 + \ldots$ no matter how many iterations are preformed. The value $b+\epsilon$ is called the <u>ultimate accuracy</u> obtainable with round-off errors. Notice that, for small ϵ, the ultimate accuracy is approximately 2ϵ, or about twice the round-off errors committed at each step.

If the ultimate accuracy is large, then the method is judged to be a poor since the effect of small round-off errors is a large error in the computed solution. If the ultimate accuracy is small, then the method is judged to be a good one since small round-off errors have a small effect on the computed solution. In this sense, the Newton-Raphson method is judged to be a good method.

For a nonsingular matrix A, many iteration methods for solving $Ax = b$ for the vector x are described by the difference equation (17)

$$x_{k+1} = Bx_k + c \tag{17}$$

where the matrix B and the vector c are determined in some fashion by A and b. For example, if $A = Q + R$, then $B = -Q^{-1}R$ and $c = Q^{-1}b$ would be a possibility. B and c must have the property that $x_o = Bx_o + c$ if and only if $Ax_o = b$. The iterations x_k will converge to the solution x_o if and only if $\rho(B)$, the spectral radius of B, is less than one. For a derivation of several of these methods, see, for example, Kunz [1] or Hildebrand [1]. Choose a vector norm $|x|$ such that $|B| = \lambda < 1$. Since $\rho(B) < 1$, this can always be done (see the Appendix).

Let x_o be the desired solution and let $x_k = x_o + e_k$. Then the e_k satisfy the difference equation

$$e_{k+1} = Be_k + h(k, e_k)$$

where the term $h(k, e_k)$ represents the round-off errors committed at step k. We assume that there exists positive constants η and ϵ such that $|h(k,e)| \leq \epsilon$ for all k and all e, $|e| < \eta$.

Try the Liapunov function $V(e) = |e|$. Then

$$\Delta V(e) \leqq -(1-\lambda)|e| + \epsilon = W(e) \leqq \epsilon$$

Then the set S is given by

$$S = \{e : W(e) \geqq 0\} = \{e : |e| \leqq \frac{\epsilon}{1-\lambda}\}$$

and $b = \frac{\epsilon}{1-\lambda}$. The set A is given by

$$A = \{e : V(e) \leqq b + \epsilon\} = \{e : |e| \leqq \frac{2-\lambda}{1-\lambda}\epsilon\}.$$

If $\eta > \frac{2-\lambda}{1-\lambda}\epsilon$, then we can choose

$$G = \{e : V(e) < \eta\} = \{e : |e| < \eta\}$$

and Corollary 4 holds. Thus, if e_1 is in the set G, then the solution will remain in G, will enter A after a finite number of iterations, and will remain in A for all following iterations.

By looking at the set A, we see that the ultimate accuracy is given by $b + \epsilon$.

$$b + \epsilon = \frac{2-\lambda}{1-\lambda}\epsilon$$

We note that, if λ is very nearly one, then this ultimate accuracy may be large even if ϵ is small. For example, if $\lambda = 1-\alpha$, then

$b + \epsilon = (\alpha^{-1} + 1)\epsilon \geqq \epsilon/\alpha$, and ϵ/α may be large. This indicates that these iteration methods will give acceptable results only if $\lambda = |B|$ is considerably less than one.

APPENDIX -- A THEOREM ON MATRIX NORMS. Let x be an n-vector and x^* its complex-conjugate transpose. Given some positive definite matrix B, let the norm of x, $|x|$, be defined by

$$|x|^2 = x^*Bx \tag{A1}$$

Other vector norms are possible, but vector norms of this type are all that are considered here.

Given a matrix A, the matrix norm of A, $|A|$, can be defined in terms of the vector norm by

$$|A| = \min\{b : |Ax| \leqq b|x| \text{ for all } x \neq 0\} \tag{A2}$$

In addition to the usual properties of a norm, this matrix norm satisfies the following.

a) $|Ax| \leqq |A|\ |x|$

b) $|\lambda| \leqq |A|$ for any eigenvalue λ of A.

c) $\rho(A) \leqq |A|$

where $\rho(A)$, the spectral radius of A, is the absolute value of the largest eigenvalue of A.

THEOREM: Let A_o be a matrix with spectral radius $a_o = \rho(A_o)$. For each $a > a_o$, there exists a vector norm such that

$$a_o \leq |A_o| \leq a \tag{A3}$$

PROOF: This theorem is proven by considering the equation

$$|A_o x|^2 - a^2 |x|^2 = x^*(A_o^* B A_o - a^2 B)x = -a^2 x^* C x \leq 0$$

where C is some positive definite matrix and

$$A_o^* B A_o - a^2 B = -a^2 C \tag{A4}$$

For any positive definite matrix C, let B be the positive definite matrix defined by

$$B = \sum_{k=0}^{\infty} a^{-2k} A_o^{*k} C A_o^k \tag{A5}$$

Since $a > a_o = \rho(A_o)$, this sum converges absolutely and B is perfectly well defined. Furthermore, this B satisfies equation (A4) and can be used to define a vector norm as in (A1). With this norm, we get

$$|A_o x|^2 - a^2|x|^2 \leq 0$$

or

$$|A_o x| \leq a|x|$$

From the definition of the matrix norm given in (A2), we get $|A_o| \leq a$. The other half of the inequality (A3) is a basic property of matrix norms.

The significance of this theorem is that a vector norm can be chosen so that the matrix norm of a matrix is made as close to the spectral radius of the matrix as desired. If $a_o < 1$, then letting $a = \frac{1}{2}(1+a_o) < 1$ leads immediately to the following corollary.

COROLLARY. A necessary and sufficient condition for the spectral radius a_o of a matrix A_o to be less than one is that there exist a vector norm such that the matrix norm of A_o satisfies $|A_o| < 1$.

It should be emphasized that the vector norm in the theorem and corollary depends quite heavily on the matrix under consideration. Given two different matrices A_1 and A_2 both with spectral radii less than one, there may not exist one vector norm so that both $|A_1| < 1$ and $|A_2| < 1$.

While the vector norm used satisfies all the requirements of a vector norm, it may be an "acceptable" norm. For example, the "unit sphere" $S = \{x : |x| = 1\}$ is an ellipsoid and the ratio of the longest axis to the shortest axis may be very high. Equation (A5) almost never can be used to compute the matrix B and resort must be made to solving (A4) directly for B. This may be a difficult task and it may be impossible to compute B to any desired degree of accuracy. This means that it may be very difficult to compute this norm of a vector.

This theorem and corollary are easily extended to cover continuous linear operators in a Hilbert space.

Acknowledgments

The author wishes to thank Professors J.K. Hale , J.P. LaSalle, and M.A. Feldstein of Brown University for their encouragement, advice, and criticism in the preparation of this article. Also, the author is grateful for the financial support rendered to him by the National Aeronautics and Space Administration under their Traineeship program.

REFERENCES

LaSalle, J. P.,

[1] Center for Dynamical Systems TR 66-1, "An Invariance Principle in the Theory of Stability".

[2] "Asymptotic Stability Criteria", Proc. Sym. in Appl. Math., 13(1962), 299-307.

[3] "Liapunov's Second Method", Stability Problems of Solutions of Differential Equations, Proc. NATO Advanced Study Institute, 1965, 95-106.

Vidal, P. and Laurent, F.,

[1] "Asymptotic Stability of Sampled Non-linear Systems with Variable Sampling Periods -- a Bionic Application", IFAC, London, 1966, session 30C.

Hahn, Wolfgang

[1] <u>Theory and Application of Liapunov's Direct Method</u>, translated by Hans Hosenthien and Siegfried Lehnigh, Prentice-Hall, Englewood Cliffs, New Jersey, 1963.

Kalman, R.E. and Bertram, J.E.

[1] "Control System Analysis and Design Via the 'Second Method' of Lyapunov, part II, Discrete-Time Systems", Journal of Basic Engineering, June, 1960, 394-499.

Ostrowski, A.

[1] <u>Solution of Equations and Systems of Equations</u>, Academic Press, New York, 1960.

Traub, J.F.

[1] <u>Iterative Methods for the Solution of Equations</u>, Prentice-Hall, Englewood Cliffs, New Jersey, 1964.

Kunz, K.S.

[1] <u>Numerical Analysis</u>, McGraw-Hill, New York, 1957.

Hildebrand, F. B.

[1] <u>Introduction to Numerical Analysis</u>, Prentice-Hall, Englewood Cliffs, New Jersey, 1952.

PAPER [17]

To be submitted to the Journal of Mathematics and Mechanics

EXISTENCE, UNIQUENESS AND CONTINUITY OF SOLUTIONS OF INTEGRAL EQUATIONS

by

Richard K. Miller*

Division of Applied Mathematics
Brown University
Providence, Rhode Island

George R. Sell**

School of Mathematics
University of Minnesota
Minneapolis, Minnesota

* This research was supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant No. AF-AFOSR-693-67, in part by the National Aeronautics and Space Administration under Contract No. NAS8-11264 and Grant No. NGR 40-002-015.

** This research supported by the National Science Foundation under Grant No. GP-3904.

EXISTENCE, UNIQUENESS AND CONTINUITY OF SOLUTIONS OF INTEGRAL EQUATIONS

by

Richard K. Miller and George R. Sell

1. INTORDUCTION. In this paper we shall be interested in the solutions $x(t)$ of a nonlinear integral equation of Volterra-type:

$$x(t) = f(t) + \int_0^t a(t,s)g(x(s),s)ds \; . \tag{1}$$

Our objective here is to present a number of theorems concerning the existence, uniqueness and continuity of solutions of (1). Existence and uniqueness theorems have been extensively studied. We note in particular, the following works: [2,3,6,7,8,10,11,14], as well as the bibliography in M.A. Krasnosel'skii's book [6]. The basic techniques for deriving existence and uniqueness criteria consist of certain fixed points theorems (for example, the Schauder-Tychonoff Fixed Point Theorem was used by C. Corduneau [2]) and comparison theorems, (for example J.A. Nohel [8] and T. Satō [11].) Our Theorem 1, is an existence theorem. It is proved with the Schauder-Tychonoff Theorem. Theorems 2 and 3 are existence and uniqueness theorems and they are proved by the contraction mapping theorem.

As is well known, equation (1) does include the initial-value problem for ordinary differential equations

$$x(0) = x_o \ , \quad x'(t) = g(x(t),t) \ . \tag{2}$$

So the theory of the solutions of (1) includes that of (2).

One question which seems to have been overlooked by the researchers in integral equations is: How does the solution $x(t)$ depend on the terms $f(t)$, $a(t,s)$ and $g(x,t)$? For ordinary differential equations, this question has been studied in an important paper [4] by E. Kamke and this theory for ordinary differential equations has been extended recently by Z. Opial [9]. We feel that the most significant results in this paper are Theorems 4 and 5 which say that the solutions $x(t)$ of (1) depend continuously on the terms f,g and a.

2. PRELIMINARIES. Let W be an open set in R^n and I an open interval in R containing 0. Let $|x|$ denote the Euclidean norm on R^n.

HYPOTHESIS A. The function f is a continuous function on I with values in W.

HYPOTHESIS B_p. Let p satisfy $1 \leqq p \leqq \infty$ and let $g(x,t)$ be a measurable function defined on $W \times I$ with values in R^n such that

(i) for each t, $g(x,t)$ is continuous in x, and

(ii) for each compact set $K \subset W$ and each compact set $J \subset I$ there is a measurable, real-valued function

$m(t)$ __with__

$$|g(x,t)| \leqq m(t) , \qquad (x \in K, \ t \in J)$$

__and__ $\int_J m(t)^p ds < \infty$.

A function $g(x,t)$ that satisfies Hypothesis B_p, $1 \leqq p < \infty$, is said to satisfy a __Lipschitz condition__ if for every pair of compact sets K, J $(K \subset W, \ J \subset I)$ there is a measureable, real-valued function $k(t)$ with

$$|g(x,t) - g(y,t)| \leqq k(t)|x-y| , \qquad (x,y \in K, \ t \in J) ,$$

and $\int_J k(t)^p dt < \infty$.

For each interval J we define the Banach space $B_p(J)$, $1 \leqq p < \infty$, by

$$B_p(J) = L_p(J,R^n) , \qquad (1 \leqq p < \infty) ,$$

where $L_p(J,R^n)$ is the Lebesgue space of all measurable functions x defined on J with values in R^n with $\int_J |x|^p dt < \infty$. We shall let $B_p^*(J)$ denote the adjoint spaces. By a well-known result one has $B_p^*(J) = B_q(J)$ if $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$.

HYPOTHESIS C_p. Let p satisfy $1 \leq p < \infty$ and let $a(t,s)$ be a mapping of $I \times I$ into the space M^n of linear operators on R^n such that

(i) for each compact interval $J \subset I$ and each t in I the mapping $S : B_p(J) \to R^n$ defined by

$$S : x \to \int_J a(t,s)x(s)ds$$

is a bounded linear mapping, and

(ii) the mapping $t \to a(t,\cdot)$ is continuous in the norm topology on $B_p^*(J)^n$.

We shall say that $a(t,s)$ satisfies Hypothesis C_p^*, $1 \leq p < \infty$, if the condition (ii) is replaced by:

(ii*) The mapping $t \to a(t,\cdot)$ is continuous in the weak*-topology on $B_p^*(J)^n$.

Hypothesis C_p and C_p^* needs some explanation. If we consider the points in R^n as column vectors and the points in M^n as square matrices, then (i) can be reformulated as: for every t in I, each row of $a(t,\cdot)$ is an element of $B_p^*(J)$. We then can view $a(t,\cdot)$ itself as an element of the direct sum

$$B_p^*(J)^n = B_p^*(J) + \ldots + B_p^*(J)$$

for every compact interval $J \subset I$. The weak*-topology, or the norm topology, on $B_p^*(J)^n$ is induced, respectively, by the weak*-topology, or the norm topology, on each component. It is clear that Hypothesis C_p implies Hypothesis C_p^*.

Hypothesis C_p or C_p^*, together with Hölders inequality, means that we can find norms on R^n and M^n so that

$$(2.p) \qquad \left|\int_J a(t,s)x(s)ds\right| \leq \{\int_J |a(t,s)|^q ds\}^{1/q} \cdot \{\int_J |x(s)|^p ds\}^{1/p},$$

if $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$; if $p = 1$,

$$(2.1) \qquad \left|\int_J a(t,s)x(s)ds\right| \leq \|a(t,\cdot)\|_\infty \cdot \int_J |x(s)|ds$$

where $\|a(t,\cdot)\|_\infty = \text{ess. sup}\ \{|a(t,s)| : s \in J\}$.

The continuity of the mapping $t \to a(t,\cdot)$ implies that if t is restricted to a compact set J' in I then the set

$$\{a(t,\cdot) : t \in J'\}$$

is a compact set in respectively, the norm topology, or the weak*-topology on $B_p^*(I)^n$. This means that

$$\sup_{t \in J'} \{\int |a(t,s)|^q ds\} < \infty$$

if $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$, with a similar statement holding for $q = \infty$.

Also, the continuity of the mapping $t \to a(t,\cdot)$ in the norm topology is equivalent to saying that

$$\int_J |a(t+h,s) - a(t,s)|^q ds \to 0 \quad \text{as} \quad h \to 0$$

where q is given as above and a similar statement holds for the case $q = \infty$. Continuity in the weak*-topology means that for each x in $B_p(\dot{J})$

$$\left|\int_J [a(t+h,s) - a(t,s)]x(s)ds\right| \to 0 \quad \text{as} \quad h \to 0 .$$

With p satisfying $1 \leq p < \infty$ we define $C = C(I,W)$ as the collection of all functions f that satisfy Hypothesis A, G_p as the collection of all functions g that satisfy Hypothesis B_p, and A_p, or A_p^*, as the collection of all functions a that satisfy Hypothesis C_p, or C_p^*, respectively.

We introduce topologies on C, G_p, A_p and A_p^* as follows:

On C we shall use the topology of uniform convergence on compact sets. This topology is metrizable. For a discussion of this cf. G.R. Sell [13].

On G_p we define two topologies T_c and T_b. We say that $g_n \to g$ in T_c if for each compact interval $J \subset I$ and each compact set $\mathcal{K} \subset C(J,W)$ the sequence $\{g_n(x(\cdot),\cdot)\}$ converges in $L_p(J,R^n)$ to $g(x(\cdot),\cdot)$ with convergence uniform for $x(\cdot) \in \mathcal{K}$.

We say that $g_n \to g$ in T_b if for every compact interval $J \subset I$ and every compact set $K \subset W$ the sequence $\{g_n(x(\cdot),\cdot)\}$ converges to $g(x(\cdot),\cdot)$ in $L_p(J,R^n)$ uniformly for $x \in \mathcal{K}$ where

$$\mathcal{K} = C(J,K) = \{x \in C(J,R^n); \quad x(t) \in K \text{ for all } t \in J\}.$$

The difference between the two topologies T_c and T_b can easily be seen in the case $W = R^n$. For both topologies we have the defining condition

$$\sup_{x \in \mathcal{K}} \int_J |g_n(x(t),t) - g(x(t),t)|^p dt \to 0 ,$$

as $n \to \infty$. For T_c the set $\mathcal{K}$ is required to be compact while for T_b the set $\mathcal{K}$ is required to be bounded. The topology T_b is meterizable and T_c is a uniform topology.

On A_p or A_p^* we say that a sequence $\{a_n\}$ converges to a limit a if for every compact interval $J \subset I$; the sequence $\{a_n(t,\cdot)\}$ converges to $a(t,\cdot)$ in the norm or respectively the weak*-topology on $B_p^*(J)^n$ uniformly for $t \in J$.

We are interested in the existence, uniqueness and continuity of solutions of $x(t)$ of the integral equation

$$(3.p) \qquad x(t) = f(t) + \int_0^t a(t,s)g(x(s),s)ds \ .$$

We say that $x(t)$ is a <u>solution</u> of (3.p) if it is measurable, satifies (3.p) on some interval $[0,\alpha)$ and is bounded on compact sub-intervals of $[0,\alpha)$. Before giving the main results, let us make note of a few lemmas.

LEMMA 1. <u>Let</u> $f \in C$, $g \in G_p$ <u>and</u> $a \in A_p^*$, $1 \leq p < \infty$. <u>If there exists a solution</u> x <u>of</u> (3.p) <u>on the interval</u> $[0,\alpha)$, <u>then</u> x <u>is a continuous function</u>.

PROOF. Consider the case where $1 < p < \infty$. (The proof for the case $p = 1$ differs only in the form of some of the equations.) Then if t and $t+h$ are in $[0,\alpha)$ one has

$$\begin{aligned} |x(t+h) - x(t)| &\leq |f(t+h) - f(t)| + |\int_0^t [a(t+h,s) - a(t,s)]g(x(s),s)ds| \\ &\quad + |\int_t^{t+h} a(t+h,s)g(x(s),s)ds| \\ &= I_1 + I_2 + I_3 \end{aligned}$$

$I_1 \to 0$ as $h \to 0$ since f is continuous. Also $I_2 \to 0$ as $h \to 0$ since $a(t,\cdot)$ is continuous in the weak*-topology on $B_p^*(I)^n$. By Hypothesis B_p, there is a function m such that

$$I_3 \leq \{\int_t^{t+h} |a(t+h,s)|^q ds\}^{\frac{1}{p}} \{\int_t^{t+h} m(s)^p ds\}^{\frac{1}{p}} \leq B\{\int_t^{t+h} m(s)^p ds\}^{\frac{1}{p}}$$

where

$$B = \sup_{0 \leqq t \leqq \beta} \{\int_0^\beta |a(t+h,s)|^q ds\}^{\frac{1}{q}}$$

for $0 \leqq t,\ t+h \leqq \beta < \alpha$. It follows from the integrability of m^p that $I_3 \to 0$ as $h \to 0$, which completes the proof.

LEMMA 2. Let $x(t)$, $0 \leqq t \leqq \alpha$, be a solution of (3.p) and let $\xi(t)$, $0 \leqq t \leqq \beta$, be a solution of

$$\xi(t) = \hat{f}(t) + \int_0^t a(t+\alpha, s+\alpha) g(\xi(s), s+\alpha) ds \, , \tag{4.p}$$

where

$$\hat{f}(t) = f(t+\alpha) + \int_0^\alpha a(t+\alpha, s) g(x(s), s) ds \, .$$

Then

$$x(t) = \begin{cases} x(t) \, , & 0 \leqq t \leqq \alpha \\ \xi(t-\alpha) \, , & \alpha \leqq t \leqq \alpha+\beta \end{cases}$$

is a solution of (3.p) on $0 \leqq t \leqq \alpha+\beta$.

The proof of this is straight forward and we omit the details.

3. EXISTENCE AND UNIQUENESS. In this section we state and prove our main results on the existence and uniqueness of solutions. The first theorem is an existence theorem. It also contains the generalization of H. Kneser's [5] theorem to integral equations. The question of continuous dependence on f, g and a is treated in the next section.

THEOREM 1. *Let* $f \in C$, $g \in G_p$ *and* $a \in A_p$, $1 \leq p < \infty$.

(A) *Then there exists an interval* $[0,\alpha)$, $\alpha > 0$, *and a continuous function* $x: [0,\alpha) \to W$ *such that* (3.p) *is satisfied for* $0 \leq t < \alpha$.

(B) *If* $[0,\alpha)$ *denotes the maximal interval of definition of* x (*which means that the solution* x *cannot be continued to the right of* α), *then either* α *is a boundary point of* I *or* $x(t) \to \text{bdy } W$ *as* $t \to \alpha$.

(C) *There is an* $\hat{\alpha} > 0$ *such that for each* t, $0 \leq t < \hat{\alpha}$, *the cross-section*

$$K_t = \{y \in W: y = x(t) \text{ where } x \text{ is some solution of } (3.p)\}$$

is compact. Moreover $\hat{\alpha}$ *can be chosen to be maximal in the sense that* $\hat{\alpha} = \alpha$, *where* α *is given by* (B) *for some solution* x.

PROOF. We shall give an argument for the case where $1 < p < \infty$. The proof for the case $p = 1$ differ from this only in the form of some of the equations.

The first part of the theorem is an existence theorem. We shall prove this by using the Schauder-Tychonoff Fixed Point Theorem.

Let $\beta > 0$ be fixed such that $[0,\beta] \subset I$. Define the operator T formally by $y = Tx$ where

$$y(t) = f(t) + \int_0^t a(t,s)g(x(s),s)ds$$

We want to show that T has a fixed point.

Since f is continuous it is bounded on $0 \leqq t \leqq \beta$, so there exist constants $M > 0$ and $\epsilon > 0$ so that

$$|f(t)| \leqq M, \qquad (0 \leqq t \leqq \beta),$$

and the compact set

$$K = \{y : |f(t)-y| \leqq \epsilon \text{ for some } t,\ 0 \leqq t \leqq \beta\}$$

lies in the open set W. Let $C([0,\beta],W)$ denote the space of continuous functions defined $[0,\beta]$ with values in W and let $D[0,\beta]$ denote the subset of $C([0,\beta],W)$ of those functions x

that satisfy

$$|f(t)-x(t)| \leqq \epsilon \,, \qquad (0 \leqq t \leqq \beta) \,.$$

$D[0,\beta]$ is a closed convex set in the Banach space $C([0,\beta],R^n)$. We shall now show that there is a β', $0 < \beta' \leqq \beta$, such that T maps $D[0,\beta']$ into itself.

First we define

$$B = \sup_{0\leqq t\leqq \beta} \left(\int_0^\beta |a(t,s)|^q ds\right)^{\frac{1}{q}}$$

where $p^{-1} + q^{-1} = 1$. By Hypothesis B_p there is a function m such that

$$|g(x,t)| \leqq m(t) \,, \qquad (x \in K,\ 0 \leqq t \leqq \beta),$$

$$\int_0^\beta m^p dt < \infty \,.$$

Now choose β', $0 < \beta' \leqq \beta$ so that

$$B\left(\int_0^{\beta'} m^p dt\right)^{\frac{1}{p}} \leqq \epsilon \,.$$

We then claim that T maps $D[0,\beta']$ into itself. Indeed, if $x \in D[0,\beta']$ and $0 \leqq t \leqq \beta'$ then by (2.p) we get

$$|y(t)-f(t)| = |\int_0^t a(t,s)g(x(s),s)ds|$$
$$\leq \{\int_0^t |a(t,s)|^q ds\}^{\frac{1}{q}}\{\int_0^t |g(x(s),s)|^p ds\}^{\frac{1}{p}}$$
$$\leq B\{\int_0^t m(s)^p ds\}^{\frac{1}{p}} \leq \epsilon ,$$

hence $y \in D[0,\beta']$.

Now we shall show that T is compact. For this purpose it suffices to show that the set of functions $T(D[0,\beta'])$ is equi-continuous. Let t be fixed with 0 t β' and let $\epsilon > 0$ be given. Then

$$|Tx(t+h)-Tx(t)| \leq |f(t+h)-f(t)| + |\int_t^{t+h} a(+h,s)g(x(s),s)ds|$$
$$+ |\int_0^t [a(t+h,s)-a(t,s)]g(x(s),s)ds|$$
$$\leq |f(t+h)-f(t)| + \{\int_t^{t+h} |a(t+h,s)|^q ds|\}^{\frac{1}{q}}\{|\int_t^{t+h} |g(x(s),s)|^p ds|\}^{\frac{1}{p}}$$
$$+ \{\int_0^t |a(t+h,s)-a(t,s)|^q ds\}^{\frac{1}{q}}\{\int_0^t |g(x(s),s)|^p ds\}^{\frac{1}{p}}$$
$$\leq |f(t+h)-f(t)| + B\{|\int_t^{t+h} m(s)^p ds|\}^{\frac{1}{p}}$$
$$+ \{\int_0^{\beta'} |a(t+h,s)-a(t,s)|^q ds\}^{\frac{1}{q}}\{\int_0^{\beta'} m(s)^p ds\}^{\frac{1}{p}} .$$

Now choose $\delta > 0$ so that if $|h| \leq \delta$ then

$$|f(t+h)-f(t)| \leqq \epsilon$$

$$B\left|\int_t^{t+h} m(s)^p ds\right|^{\frac{1}{p}} \leqq \epsilon$$

$$\left\{\int_0^{\beta'} |a(t+h),s)-a(t,s)|^q ds\right\}^{\frac{1}{q}}\left\{\int_0^{\beta'} m(s)^p ds\right\}^{\frac{1}{p}} \leqq \epsilon \ .$$

Note that δ depends on t and ϵ but it is independent of the function x. It follows then that

$$|Tx(t+h)-Tx(t)| \leqq 3\epsilon \qquad\qquad (|h| \leqq \delta) \ ,$$

which shows that T is compact.

Now we shall show that T is continuous. If $\{x_n\}$ is a sequence in $D[0,\beta']$ with limit x, then by the continuity of $g(x,t)$ in x we get

$$g(x_n(s),s) \to g(x(s),s)$$

for each s, $0 \leqq s \leqq \beta'$, and

$$a(t,s)g(x_n(s),s) \to A(t,s)g(x(s),s) \ .$$

Furthermore, the sequence $\{a(t,s)g(x_n(s),s)\}$ is bounded by an integrable function,

$$|a(t,s)g(x_n(s),s)| \leqq |a(t,s)|m(s) ,$$

hence by Lebesgue's Theorem

$$\int_0^t a(t,s)g(x_n(s),s)ds \to \int_0^t a(t,s)g(x(s),s)ds .$$

This implies that for each t, $0 \leqq t \leqq \beta'$, one has

$$(5) \qquad Tx_n(t) \to Tx(t) \quad \text{as} \quad n \to \infty .$$

In order to show that T is continuous we must show that the convergence in (5) is uniform. However, this follows easily from the fact that Tx is continuous and the set $\{Tx_n\}$ is equi-continuous.

We have thus shown that T is a compact, continuous operator, therefore by the Schauder-Tychonoff Fixed Point Theorem, cf. Cronin [3;p. 131], T has a fixed point x.

Let us now show that the maximal interval of definition $[0,\alpha)$ is characterized in the form described in the theorem. Proceeding by contradiction, assume that α is not a boundary point of J and that $x(t)$ remains in a compact set $K \subset W$ for $0 \leqq t < \alpha$. We will then show that there is a solution $\hat{x}(t)$ of (1) defined on an interval $[0,\alpha')$ where $\alpha < \alpha'$ and such that $\hat{x}(t) = x(t)$ for $0 \leqq t < \alpha$. This will contradict the maximality of $[0,\alpha)$.

The first step is to show that $\lim_{t \to \alpha} x(t)$ exists, we shall call this limit $x(\alpha)$. The limit exists if for every $\epsilon > 0$ there is a τ, $0 < \tau < \alpha$ such that $|x(t)-x(u)| \leqq \epsilon$ for all t and u with $\tau \leqq t < \alpha$, $t \leqq u < \alpha$.

By Hypothesis B_p, there is a function m such that

$$|g(x,t)| \leqq m(t) \qquad (x \in K,\ 0 \leqq t \leqq \alpha)\ ,$$

$$\int_0^\alpha m^p ds < \infty\ .$$

If $\tau \leqq t \leqq u < \alpha$, then

$$\begin{aligned} |x(u)-x(t)| &\leqq |f(u)-f(x)| + \left|\int_0^t \{a(t,s)-a(u,s)\} g(x(s),s)ds\right| \\ &\quad + \left|\int_t^u a(u,s)g(x(s),s)ds\right|\ , \\ &\leqq |f(u)-f(t)| \\ &\quad + \left\{\int_0^\alpha |a(t,s)-a(u,s)|^q ds\right\}^{\frac{1}{q}} \left\{\int_0^\alpha m(s)^p ds\right\}^{\frac{1}{p}} \\ &\quad + \left\{\int_0^\alpha |a(u,s)|^q ds\right\}^{\frac{1}{q}} \left\{\int_\tau^\alpha m(s)^p ds\right\}^{\frac{1}{p}}\ . \end{aligned}$$

Thus our assumptions clearly imply that if $\tau \leqq t \leqq u < \alpha$ and $\alpha-\tau$ is sufficiently small, then $|x(u)-x(t)| < \epsilon$. Thus we see that $x(t)$ is a solution of (3.p) on the closed interval $[0,\alpha]$.

Now by applying the previous existence proof with Lemma 2, we conclude that the solution x can be continued for $0 \leqq t \leqq \alpha+\beta$, $\beta > 0$, and this contradicts the maximality of $[0,\alpha)$.

The proof that the cross-sections are compact is simple modification of Kneser's Theorem for ordinary differential equations, of G.R. Sell [12, p. 373]. The critical thing to show is that if $\{x_n\}$ is a sequence of solutions of (3.p) that converges uniformly on compact sets to a function x, then x is a solution of (3.p). This, however, is a direct application of the Lebesgue Dominated Convergence Theorem, which completes the proof. The fact that $\hat{\alpha}$ is maximal in the sense indicated can also be proved with the same techniques, cf. [12; p. 382].

REMARKS. 1. As noted in the Introduction, the Schauder-Tychonoff Fixed Point Theorem has been used before to get existence criteria for integral equations. See, for example, [2,3,6]. The fact that the maximal interval of definition is characterized by Statement (B) has been proved by essentially the same argument but under more restrictive conditions by J.A. Nohel [8]. Finally a special case of (C), which generalizes Kneser's Theorem, has been proved by T. Satō [11].

2. Our argument does break down if we replace the Hypothesis C_p for $a(t,s)$ be the weaker Hypothesis C_p^*. The only place where the stronger hypothesis was used to show that the operator T is compact. We have no counter example to show

that Theorem 1 is false under the weaker hypothesis on $a(t,s)$.

In order to get uniqueness of solutions, we impose a Lipschitz condition on g.

THEOREM 2. *Let* $f \in C$, $g \in G_p$, $a \in A_p$, $1 \leq p < \infty$, *and assume that* g *satisfies a Lipschitz condition. Then the solution* x *of* (3.p) *is unique.*

Actually, if g satisfies a Lipschitz condition, then we can relax the assumption on the kernel a and ask that it satisfy Hypothesis C_p^*. Theorem 2 is then a special case of the following result.

THEOREM 3. *Let* $f \in C$, $g \in G_p$ *and* $a \in A_p^*$, $1 \leq p < \infty$, *and assume that* g *satisfies a Lipschitz condition. Then there exists one and only one solution of* (3.p). *Furthermore, the maximal interval of definition is characterized by* (B) *of Theorem* 1.

The proof of this is completely straightforward. One proves, by the usual arguments, cf. [8], that the operator T is a contraction on some set $D[0,\beta'']$. We omit the details.

REMARK. One can replace the Lipschitz condition on g with a weaker statement. For example, one could replace it with the Osgood condition:

$$|g(x,t)-g(y,t)| \leq k(t)\psi(|x-y|)$$

where $\int_0^\alpha \frac{dr}{\psi(r)} = +\infty$. Comparison theorems of this type are well-known for differential equations and they have been used for integral equations, cf. [8,11].

4. CONTINUITY OF SOLUTIONS. In this section we investigate the dependence of the solutions x on the three terms f,g and a.

THEOREM 4. Let $\{f_n\}$, $\{g_n\}$ and $\{a_n\}$ be sequences in C, G_p and A_p respectively where $1 \leq p < \infty$. Assume that these sequences have limits $f_n \to f$, $g_n \to g$ (in T_c) and $a_n \to a$ in the respective spaces. Let $\{x_n\}$ be a sequence of solutions of

$$x_n(t) = f_n(t) + \int_0^t a_n(t,s)g_n(x_n(s),s)ds\ , \tag{5.p}$$

on the maximal intervals $[0,\alpha_n)$. Then the sequence $\{x_n\}$ has a uniformly convergent subsequence on some interval $0 \leq t \leq \sigma$, $\sigma > 0$. The limit function x is a solution of the limiting equation

$$x(t) = f(t) + \int_0^t a(t,s)g(x(s),s)ds\ . \tag{6.p}$$

Moreover, the subsequence $\{x_{n_j}\}$ *of* $\{x_n\}$ *may be chosen so that* $x_{n_j}(t) \to x(t)$ *as* $j \to \infty$ *uniformly on compact subsets of* $[0,\hat{\alpha})$, *where the interval* $[0,\hat{\alpha})$ *is the maximal interval on which the cross sections* K_t *of* (6.p) *are compact* (*see Theorem* 1(C)) *and* $[0,\hat{\alpha}) \subset \liminf [0,\alpha_n)$.

PROOF. We will show that for any β with $0 < \beta < \hat{\alpha}$, one has:

1. $[0,\beta] \subseteq [0,\alpha_n)$ for n sufficiently large;

2. on the interval $[0,\beta]$ the sequence of functions $\{x_n\}$ is bounded and equicontinuous; and

3. if $\{x_{n_j}\}$ is a convergent subsequence of $\{x_n\}$ with limit $x(t)$ on $[0,\beta]$, then $x(t)$ is a solution of (6.p) on $[0,\beta]$.

Again we shall prove this for the case $1 < 0 < \infty$. The proof for the case $p = 1$ is similar.

Let β, $0 \leqq \beta < \hat{\alpha}$, be given where $\hat{\alpha}$ is given by the hypothesis. Then the cross section

$$K_t = \{y \in W;\ y = x(t) \text{ for some solution } x \text{ of } (6.p)\}$$

is a compact subset of W for $0 \leqq t \leqq \beta$. It is easily show that

$$\hat{K} = \cup\{K_t;\ 0 \leqq t \leqq \beta\}$$

is a compact subset of W, cf. e.g. [12, p. 378]. Let K be a compact set in W that contains $\hat{K}$ in its interior. By hypothesis B_p there is a function $m \in L_p[0,\beta]$ such that

$$|g(x,t)| \leq m(t) . \qquad (x \in K,\ 0 \leq t \leq \beta)$$

The convergence $g_n \to g$ in T_c implies that $\lim \varepsilon_n = 0$ as $n \to \infty$ where

$$\varepsilon_n = \sup_{x \in K} \int_0^\beta |g_n(x,s)-g(x,s)|^p ds .$$

Furthermore if $0 \leq \sigma \leq \beta$, then

$$\{\int_0^\sigma |g_n(x,s)|^p ds\}^{\frac{1}{p}} \leq \{\int_0^\sigma |g_n(x,s)-g(x,s)|^p ds\}^{\frac{1}{p}} + \{\int_0^\sigma m(s)^p ds\}^{\frac{1}{p}}$$

$$\leq \varepsilon_n + \{\int_0^\sigma m(s)^p ds\}^{\frac{1}{p}} .$$

For σ in the interval $[0,\beta]$ set

(7) $$M(\sigma,n) = \varepsilon_n + \{\int_0^\sigma m(s)^p ds\}^{\frac{1}{p}} .$$

Similarly we can find a common bound for the sequence $\{a_n\}$, that is

(8) $$B = \sup_n [\sup_{0 \leq t \leq \beta} \{\int_\theta^\beta |a_n(t,s)|^q ds\}^{\frac{1}{q}}] < \infty ,$$

where $p^{-1} + q^{-1} = 1$. Note that the bound B in (8) also holds for the limiting function $a(t,s)$.

By the choice of the set K, there is an $\varepsilon > 0$ such that if $x(t)$ is any solution of (6.p), $0 \leqq t \leqq \beta$, and $|y-x(t)| \leqq \varepsilon$ then $y \in K$. Fix δ so that $0 < \delta < \varepsilon/2$ and fix $N_1 \geqq 1$ so that if $n \geqq N_1$ then $2B\varepsilon_n < \delta$ and

$$|f_n(t)-f(t)| < \delta . \qquad (0 \leqq t \leqq \beta,\ n \geqq N_1)$$

Now choose σ so that $0 < \sigma \leqq \beta$ and

$$2B\{\int_0^\sigma m(s)^p ds\}^{\frac{1}{p}} = \varepsilon - 2\delta .$$

If this equality cannot be satisfied for $0 \leqq \sigma \leqq \beta$, then choose $\sigma = \beta$.

We will now show that on the interval $[0,\sigma]$ one has $x_n(t) \in K$ for all $n \geqq N_1$. Let $x(t)$ be any solution of (6.p) defined on $[0,\sigma]$. We shall show that $|x_n(t)-x(t)| \leqq \varepsilon$ for $0 \leqq t \leqq \sigma$ and $n \geqq N_1$. For $t = 0$ we have

$$|x_n(0)-x(0)| = |f_n(0)-f(0)| < \delta < \varepsilon .$$

Suppose $|x_n(t)-x(t)| \leqq \varepsilon$ for $0 \leqq t < \xi \leqq \sigma$. Then

$$|x_n(\xi)-x(\xi)| \leqq |f_n(\xi)-f(\xi)| + |\int_0^\xi a_n(t,s)g_n(x_n(s),s)ds|$$

$$+ |\int_0^\xi a(t,s)g(x(s),s)ds| .$$

Applying (2.p), (7) and (8) we see that

$$|x_n(\xi)-x(\xi)| \leqq \delta + 2B\,M(\xi,n)$$

$$\leqq \delta + 2B\,M(\sigma,n) < \varepsilon .$$

Hence the maximal interval $[0,\xi]$ for which $x_n(t) \in K$ if $0 \leqq t < \xi$ must include $[0,\sigma]$.

We shall now show that on the interval $[0,\sigma]$, the sequence of functions $\{x_n;\ n \geqq N_1\}$ is equicontinuous. If $0 \leqq t,\ t+h \leqq \sigma$, then

$$|x_n(t+h)-x_n(t)| \leqq |f_n(t+h)-f_n(t)|$$

$$+ |\int_0^t \{a_n(t+h,s)-a_n(t,s)\}g_n(x_n(s),s)ds|$$

$$+ |\int_t^{t+h} a_n(t+h,s)g_n(x_n(s),s)ds| ,$$

$$\leqq |f_n(t+h)-f_n(t)| + B\{\int_t^{t+h} m(s)^p ds\}^{\frac{1}{p}}$$

$$+ \{\int_0^\sigma |a_n(t+h,s)-a_n(t,s)|^q ds\}^{\frac{1}{q}} M(\sigma,n) .$$

Since $f_n \to f$ uniformly on $[0,\sigma]$, the sequence $\{f_n\}$ is equicontinuous. Since $a_n(t,\cdot) \to a(t,\cdot)$ uniformly in $t \in [0,\sigma]$, it follows that the sequence $\{a_n\}$ is equicontinuous as functions of t with values in $B_q[0,\sigma]^n$. Since B is a fixed constant and $M(\sigma,n)$ is bounded in n, we see that $\{x_n; n \geq N_1\}$ is equicontinuous on $[0,\sigma]$.

Now choose any convergent subsequence of $\{x_n\}$. To simplify the notation we shall write $\{x_n\}$ for this subsequence. Then there is a function x such that $x_n(t) \to x(t)$ uniformly on $[0,\sigma]$. Since

$$\mathcal{K} = \{x, x_1, x_2, x_3, \dots\},$$

is a compact set in $C([0,\sigma],W)$ and $g_n \to g$ in T_c it follows that

$$g_n(x_n(\cdot),\cdot) \to g(x(\cdot),\cdot)$$

in $L_p([0,\sigma],R^n)$. Also we have

$$a_n(t,\cdot)g_n(x_n(\cdot),\cdot) \to a(t,\cdot)g(x(\cdot),\cdot)$$

in $L_1([0,\sigma],R^n)$ for $0 \leq t \leq \sigma$. Thus

$$\int_0^t a_n(t,s)g_n(x_n(s),s)ds \to \int_0^t a(t,s)g(x(s),s)ds .$$

It follows that $x(t)$ satisfies (6.p) on $[0,\sigma]$.

We now want to show that the interval $[0,\sigma]$ can be extended to $[0,\beta]$. This extension can be performed in a finite number of repetition of the above argument. That is, consider the translation of (6.p) given by

$$(9.p) \qquad X(t) = \hat{f}(t) + \int_0^t a(t+\sigma\,, s+\sigma)g(X(s), s+\sigma)ds$$

where $\hat{f}(t) = f(t+\sigma) + \int_0 a(t+\sigma, s)g(x(s), s)ds$. Equation (5.p) is similarly translated. By the above argument one can find a $\tau > 0$ such that a subsequence of the solutions $\{X_n(t)\}$ converge to a solution $X(t)$ of (9.p) on $[0,\tau]$. By Lemma 2, we see that

$$x(t) = \begin{cases} x(t) & 0 \leqq t \leqq \sigma \\ X(t-\sigma) & 0 \leqq t \leqq \sigma+\tau \end{cases}$$

is a solution of (6.p) on $[0,\sigma+\tau]$, and it is the limit of a subsequence of $\{x_n(t)\}$ on $0 \leqq t \leqq \sigma+\tau$.

This process can now be repeated. In order to show that one can extend to $[0,\beta]$ in a finite number of steps, it is necessary to keep track of the size of each step. This is governed by the function $M(\sigma,n)$ defined above. That is, the number τ can be chosen so that $0 < \tau \leqq \beta-\sigma$ and

$$2B\{\int_\sigma^{\sigma+\tau} m(s)^p ds\}^{\frac{1}{p}} = \varepsilon - 2\delta \quad ,$$

or if this last equality cannot be satisfied then set $\tau = \beta - \sigma$. Since the integral $\int_0^\beta m(s)^p ds$ is finite, it is clear that one can extend $[0,\beta]$ to $[0,\beta]$ in a finite number of steps. This completes the proof of Theorem 4.

In the last theorem above we assumed that the kernals $\{a_n(t,s)\}$ and the limiting kernal $a(t,s)$ satisfy Hypothesis C_p and that $a_n \to a$ in the norm topology, uniformly for t on compact sets $J \subset I$. One can ask whether the weaker convergence would suffice. The answer is yes if one strengthens the convergence on $\{g_n\}$. More precisely we prove the following result:

THEOREM 5. Let $\{f_n\},\{g_n\}$ and $\{a_n\}$ be sequences in C, G_p and A_p^* respectively where $1 \leqq p < \infty$. Assume that the sequences have limits $f_n \to f$, $g_n \to g$ (in T_b) and $a_n \to a$ (in A_p^*) with f,g and a in the respective spaces. Assume further that g_n and g satisfy Lipschitz conditions. Let x_n be the solutions of

$$x_n(t) = f_n(t) + \int_0^t a_n(t,s)g_n(x_n(s),s)ds, \tag{5.p}$$

on the maximal intervals $[0,\alpha_n)$. Then the sequence $\{x_n\}$ converges uniformly on compact subsets of $[0,\alpha)$ to a function $x(t)$. The function $x(t)$ is the unique solution of the limiting equation

$$(6.p) \qquad x(t) = f(t) + \int_0^t a(t,s)g(x(s),s)ds$$

<u>defined on the maximal interval</u> $[0,\alpha)$. <u>Moreover,</u>
$\alpha \leqq \liminf \alpha_n \quad (n \to \infty)$.

PROOF. For any β, $0 < \beta < \alpha$, we shall show that $\alpha_n \geqq \beta$ for n sufficiently large and that $x_n(t) \to x(t)$ uniformly on $[0,\beta]$. This will prove the theorem.

Fix and $\beta \in [0,\alpha)$. Let K be a compact subset of W that contains the curve $x(t)$: $0 \leqq t \leqq \beta$ in its interior. Let $m \in L_p([0,\beta],R^n)$ with

$$|g(x,t)| \leqq m(t) . \qquad (x \in K,\ 0 \leqq t \leqq \beta)$$

Since the kernals $a_n(t,\cdot)$ converge to $a(t,\cdot)$ in the weak*-topology on $B_q[0,\beta]^n$, they are bounded in the norm topology. Furthermore since the convergence in uniform for t on compact sets, the number B defined by

$$B = \sup_n \{ \sup_{0\leqq t\leqq\beta} \{\int_0^\beta |a_n(t,s)|^q ds\}^{\frac{1}{q}} \}$$

is finite. Let $M(\sigma,n)$, ε, δ, N_1 and σ be defined as in the first part of the proof of Theorem 4.

Instead of showing the equicontinuity of $\{x_n\}$ on $[0,\sigma]$ we proceed directly to estimate $|x_n(t)-x(t)|$. Define $R_n(t)$ by

$$R_n(t) = |f_n(t)-f(t)| + \int_0^t |a_n(t,s)||g_n(x_n(s),s)-g(x_n(s),s)|ds$$
$$+ |\int_0^t \{a_n(t,s)-a(t,s)\}g(x(s),s)ds|$$

and let $\varepsilon_n = \sup\{R_n(t);\ 0 \leqq t \leqq \sigma\}$. Because of the convergence assumptions on $\{f_n\},\{a_n\}$ and $\{g_n\}$ and the fact that $x_n(t) \in K$ for $0 \leqq t \leqq \sigma$, $n \geqq N_1$ one has

$$\varepsilon_n \to 0 \quad \text{as} \quad n \to \infty \ .$$

Since g satisfies a Lipschitz condition, there is a function $k(t) \in L_p([0,\beta],R^n)$ such that

$$|g(x,t)-g(y,t)| \leqq k(t)|x-y| \ . \qquad (x,y \in K,\ 0 \leqq t \leqq \beta)$$

By a straight forward computation we get

$$|x_n(t)-x(t)| \leqq R_n(t) + \int_0^t |a_n(t,s)||g(x_n(s),s)-g(x(s),s)|ds$$
$$\leqq \varepsilon_n + \int_0^t k(s)|a_n(t,s)||x_n(s)-x(s)|ds \ .$$

By Gronwall's inequality [1] we get for all $n \geqq N_1$ and for $0 \leqq t \leqq \sigma$

$$|x_n(t)-x(t)| \leqq \varepsilon_n + \varepsilon_n \int_0^t k(s)|a_n(t,s)|(\exp \int_s^t k(r)|a_n(t,r)|dr)ds \ .$$

If we define

$$K_o = \{\int_o^\beta k(s)ds\}^{\frac{1}{p}},$$

then for $0 \leqq t \leqq \sigma$

$$|x_n(t)-x(t) \leqq \varepsilon_n(1+K_oB \exp(K_oB)) \to 0$$

as $n \to \infty$. This shows that $x_n(t) \to x(t)$ uniformly for $0 \leqq t \leqq \sigma$. One can extend $[0,\sigma]$ to $[0,\beta]$ by the same reasoning process used in the proof of Theorem 4. This completes the proof of Theorem 5.

REMARKS.

1. The assumption that the limit function $g(x,t)$ satisfies a Lipschitz condition can be weakened. One could use an Osgood condition or a comparison theorem used by J. Nohel [8] or T. Satō [11]. However, it does not appear that in Theorem 5 one can drop this type of analytical criterion, which implies uniqueness, and assume directly that the solutions are unique.

2. It should be noted that E. Kamke's Theorem [4] on the continuity of solutions of ordinary differential equations, as well as Z. Opial's generalization [9] are included as special cases of Theorem 4. In these papers the kernal $a(t,s)$ reduces to the identity matrix. Kamke assumed that the functions $g_n(x,t)$ and $g(x,t)$ were con-

tinuous and that $g_n \to g$ uniformly on compact sets. This convergence implies $g_n \to g$ in T_b for every space G_p, $1 \leq p < \infty$. Opial assumed that the functions g_n and g satisfied Hypothesis B_1 and $g_n \to g$ in T_b for $p = 1$.

3. Many variations of our theorems are possible. For example suppose we set $p = \infty$ and $q = 1$. Here we assume g to be continuous in (t,x) and $g_n \to g$ means uniform convergence on compact sets. Suppose now that a satisfies the following conditions:

(i) for each compact iterval $J \subset I$ and each $t \in I$ the map $S: C(J,W) \to R^n$ defined by

$$S : x \to \int_J a(t,s)x(s)ds$$

is a bounded linear functional,

(ii) the mapping $t \to a(t,\cdot)$ is continuous in the norm topology on $B_1(J)$, and

(iii) for any compact set $J \subset I$,

$$\lim_{h \to 0} \int_t^{t+h} |a(t,s)|\, ds = 0$$

uniformly for $t \in J$.

Under these conditions on g and a, the obvious variations of Theorems 1 through 5 are true. We omit a formal statement.

4. Continuity results of the type given by Theorem 5 have been obtained by Levin and Nohel [15] in a special, scalar example.

REFERENCES.

[1] L. Cesari, Asymptotic behavior and stability problems in ordinary differential equations, Second edition, Academic Press, New York, 1963.

[2] C. Corduneau, "Problemes globeaux dans la theorie des equations integrales de Volterra", Ann. Mat. Pura Appl. (4), 67 (1965), 349-363.

[3] J. Cronin, Fixed points and topological degree in nonlinear analysis, Amer. Math. Soc., Providence, R.I., 1964.

[4] E. Kamke, "Zur Theorie der Systeme gewohnlicher Differential-gleichungen II", Acta. Math. 58(1932), 57-85.

[5] H. Kneser, "Uber die Losungen eines Lipschitzschen Bedingung nicht genugt", Sitzber. Preuss. Akad. Wiss. Phys.-Math. Kl. (1932), 171-174.

[6] M. A. Krasonsel'skii, Topological Methods in the Theory of Nonlinear Integral Equations, Perga on Press, New York, 1964.

[7] L. A. Ladyzenskii, "Conditions for the complete continuity of P. S. Uryson's integral operator in the space of continuous functions", Dokl. Akad. Nauk SSSR 97(1954), 1105-1108.

[8] J. A. Nohel, "Some problems in nonlinear Volterra integral equations", Bull. Amer. Mat. Soc. 68(1962), 323-329.

[9] Z. Opial, "Sur la dependance des solutions d'un systeme d'equations differentielles de leurs seconds membres. Applications aux systemes presque antonomous", Ann. Polon. Math. 8(1960), 75-89.

[10] M. Reichert, "Eindeutigkeits- und Iterationsfragen bei Volterra-Hammersteinschen Integralgleichungen ", J. Reine Angew. Math., 220(1965), 74-87.

[11] T. Satō, "Sur l'equation integrale non lineaire de Volterra", Composito Math., 11(1953), 271-290.

[12] G. R. Sell, "On the fundamental theory of ordinary differential equations", J. Diff. Equations, 1(1965), 370-392.

[13] G. R. Sell, "Nonantonomous differential equations and topological dynamics I. The basic theory", Trans. Amer. Math. Soc. (to appear).

[14] D. Willett, "Nonlinear vector integral equations as contraction mappings", Arch. Rational Mech. Anal., 15(1964), 79-86.

[15] J. J. Levin and J. A. Nohel, "On a system of integrodifferential equations occuring in reactor dynamics, II, Archive Rat. Mech. Anal., 11(1962), 210-243.