PHASE-PLANE ANALYSIS OF NONLINEAR SAMPLED-DATA SERVOMECHANISMS

by

Rudolf Emil Kalman

B.S., Massachusetts Institute of Technology
(1953)

SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June, 1954

Signature of Author ______________________
Department of Electrical Engineering, May 24, 1954

Certified by ______________________
Thesis Supervisor

Accepted by ______________________
Chairman, Departmental Committee on Graduate Students

PHASE-PLANE ANALYSIS OF NONLINEAR SAMPLED DATA SERVOMECHANISMS

by

Rudolf E. Kalman

Submitted to the Department of Electrical Engineering on 24 May 1954, in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering.

ABSTRACT

This thesis is an attempt to formulate an exact theory of nonlinear difference equations which describe the behavior of second-order control systems operating on sampled data and containing nonlinear gain elements. The problem was previously investigated by C. K. Chow and F. A. Russell using the describing-function method. From their work, it is known that relay-controlled servomechanisms may become unstable when sampling is introduced in the loop, and may oscillate autonomously at any one of several different amplitudes and periods.

By the direct use of difference equations defining the system behavior, the present investigation rigorously establishes the following results:

(1) In the case of relay servos, the oscillatory solutions predicted by the describing-function method are stable if and only if the open-loop difference equation relating error and output is stable.

(2) In addition to those oscillations predicted by the describing-function method, other oscillations of distinctly non-sinusoidal character may also exist.

(3) The sets of initial conditions in the phase plane giving rise to the various autonomous oscillations are in general disjoint. The transient behavior of such systems, especially in regard to the final steady-state periodic solution, is therefore extremely involved.

(4) Standard techniques of linear sampled-data systems may be utilized in the prevention of oscillations.

Exact examples of transient behavior have been worked out which illustrate the extreme complexity of the problem. In order to achieve further and better understanding of the phenomena involved, advanced mathematical techniques appear indispensable.

Thesis Supervisor: Samuel J. Mason
Title: Associate Professor of Electrical Engineering

ACKNOWLEDGMENTS

The investigation carried out in this thesis was originally suggested by Mr. M. Goldman, of the U. S. Naval Underwater Ordnance Station, Newport, R. I., and was sponsored by the U. S. Navy under Contract No. NOrd-11799 with the Servomechanisms Laboratory, Massachusetts Institute of Technology.

The writer has benefited much from various conversations with members of the Servomechanisms Laboratory. He is especially grateful to this Laboratory for having made available personnel to perform the numerical computations.

The author is much indebted to Professor S. J. Mason, thesis supervisor, whose help in connection with certain scientific, educational, and administrative aspects of the present work was invaluable.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

1. INTRODUCTION

In many automatic control applications, especially those involving difficult measurement problems, input or error information may be available only in the form of discrete, periodically repeated signals. Such cases are usually referred to as sampled-data systems. The best-known example of a sampled-data system is tracking by pulsed radar, where the target position is measured by means of the reflection of periodically emitted high-frequency pulses. Somewhat less obvious but equally typical further examples are digital data handling equipment, chemical process control based on batchwise analysis, and numerical computations.

It has been recognized very early that when sampling is introduced into an otherwise linear automatic control system, the linearity is not affected. However, since the loop is effectively closed only at the sampling instants, we cannot describe the closed-loop input-output relationships in terms of a differential equation, but have to settle for a linear difference equation with constant coefficients, which is a dynamic description of the system behavior at the sampling instants.

Because of linearity, and in analogy with conventional servo theory, the frequency-transform viewpoint appears as the natural method of analysis. Such a frequency-transform technique was indeed developed following World War II; it has now attained a considerable degree of maturity so that questions concerning analysis may be regarded as satisfactorily disposed of, even though some difficulties do remain in regard to simplified computational procedures, design methods, etc. It should be borne in mind, however, that this frequency transform technique is merely a convenient way of "mechanizing" the solution of linear difference equations of the system in a manner that is completely analogous to the use of Laplace transforms in solving time-invariant linear differential equations.

So much for the linear theory.

We now ask the question: "What new problems and phenomena, if any, are created in nonlinear control systems by the addition of sampling in the loop?" This is not purely an academic topic: A control system using a periodically operated error detection device which (perhaps because of noise) can distinguish only two different levels of signal, is essentially a sampled-data system with an ideal relay. If the device is capable of differentiating between three signal levels, say, +, o, or -, then we have a relay with a deadzone as the nonlinear element.

Surprisingly enough such systems have not been studied until very recently: two doctorate dissertations at Cornell[1a-b] and Columbia[2], completed simultaneously and independently in June 1953, plus a report by the Westinghouse Research Laboratories in September 1953[3], make up the totality of the accessible literature on this topic at present (May 1954). All of these investigations are concerned exclusively with relay servos.

The pure mathematical theory of nonlinear difference equations pertinent to the above problem is very incompletely developed. In particular, the existence of complicated autonomous periodic solutions of difference equations, reminiscent of the limit cycles of continuous nonlinear systems, has apparently received no attention from the mathematicians* and in a sense

*With the exception of an investigation by Picard in the 1890's (cf. Enzyk. math. Wiss., II 3, 2 (1922) 705.) who showed that certain higher transcendental functions may be represented as periodic solutions of certain difference equations. Picard did not, however, investigate difference equations as such for periodic solutions.

these phenomena seem to have been "discovered" by the electrical engineers whose work has been cited above.

This thesis will be concerned primarily with the investigation of the properties of self-sustained oscillations in nonlinear difference equations; although the mathematical difficulties at the present time forbid exploring more than a very restricted section of this baby-new field (much of which is probably of very questionable engineering relevance), it is believed that the present work constitutes the first systematic attempt to formulate an exact theory of nonlinear difference equations, along lines similar to those introduced by Poincaré and others before the turn of the present century in connection with differential equations. This attempt is largely a failure by the yardstick of mathematical elagance or generality; it does, however, set up a workable computational machinery for exploring a given problem. This should open to way to collecting precise information about particular cases; once an appreciable amount of such information is available, the success of attempts to generalize will appear more probable. Thus the second fundamental contribution of this work is contained in two complete, exact examples which reveal much that has at best been suspected heretofore.

In outline, the following points will be investigated in turn, after a brief "historical" survey of previous work in the area:

(1) Mathematical representation of nonlinear sampled-data systems.

(2) Existence and stability of self-sustained oscillations.

(3) Structure of transient response in the phase plane.

(4) Elimination of self-sustained oscillations.

Even though the mathematical formalism and arguments presented in the following pages are applicable to second-order sampled-data systems containing any types of frequency-independent nonlinearities, almost all of our treatment will be limited to the relay-servo case in order to come to grips with the problem in its simplest possible setting.

2. SURVEY OF PREVIOUS WORK

2.1 Results Obtained by the Describing-Function Method

In his doctorate thesis, Chow[1a] attacked the problem shown in Fig. 1, which represents a conventional single-loop servo with a linear output element $G(s)$, a relay with a deadzone Δ, and an impulse-modulator-clamper combination which is the usual way to describe the sampling operation when the value of the sampled signal is held constant until the next sampling instant. Chow uses the linearization procedure known by the name "describing-function method"[4 and treats the combination of impulse modulator, clamper, and relay as a single nonlinear element. This describing function N will, in general, depend on the the frequency $j\omega$, the amplitude A, and the phase ϕ (with respect to the impulse modulator) of the hypothetical input sinusoid, plus the normalized deadzone Δ of the relay. It is easily seen that in order for the oscillations to be self-maintaining they must be in synchronism with the impulse

---

*This method calls for computing the amplitude and phase shift of the fundamental at the output of the nonlinear element, when there is a hypothetical sine wave applied at the input. The resulting complex expression relating input to output is then used in the same manner as the usual transfer function $G(s)$ in connection with Nyquist plots. It should be noted that the describing-function method does no more than detect the possibility or otherwise of self-sustained oscillations in the loop, under the basic assumption that they are approximately sinusoidal.

modulator, that is, their periods must be <u>integer</u> multiples nT $(n \geq 2)$ of the sampling period T. Thus self-sustained oscillations are possible if there exist discrete values of $s = j\omega = j2\pi/nT$ which satisfy the well-known condition:

$$1 + N(j2\pi/nT, A, \phi; \Delta)G(j\omega) = 0 \tag{2.1}$$

which is the denominator of the linearized closed-loop transfer function of the system Fig. 1. Condition (2.1) is usually written as:

$$-\frac{1}{N(j2\pi/nT, A, \phi; \Delta)} = G(j\omega) \tag{2.2}$$

and the describing function is plotted in the form $-1/N$.

Chow has shown that $-1/N$ may be defined by a series of inequalities in the complex $G(j\omega)$ plane, which give rise to overlapping polygonal regions; thus self-sustained oscillations are possible if a point on the $G(j\omega)$ locus corresponding to some $\omega = 2\pi/nT$, falls in the appropriate region $-1/N(n)$. A further complication arises due to the dependence of the describing function on Δ, as a result of which is is possible to have several oscillations of the same period but of different amplitudes.* For the reader's convenience, we give in Appendix A a derivation of the describing function for the simple case $\Delta = 0$. (This is contained in Chow's thesis[1a] but not in the AIEE paper[1b].)

---

*The reason is this: If the relay had no deadzone, it would always be actuated in one sense or the other during a half period of sinusoidal oscillation. If there is a deadzone, however, the relay may or may not be actuated when the sampled sinusoidal signal into it goes through zero; thus the total length of time t_o for which the relay is closed during a half period may be anywhere in the range $T \leq t \leq nT/2$. Obviously the amplitude of the oscillation is larger if the relay is closed longer during a half period.

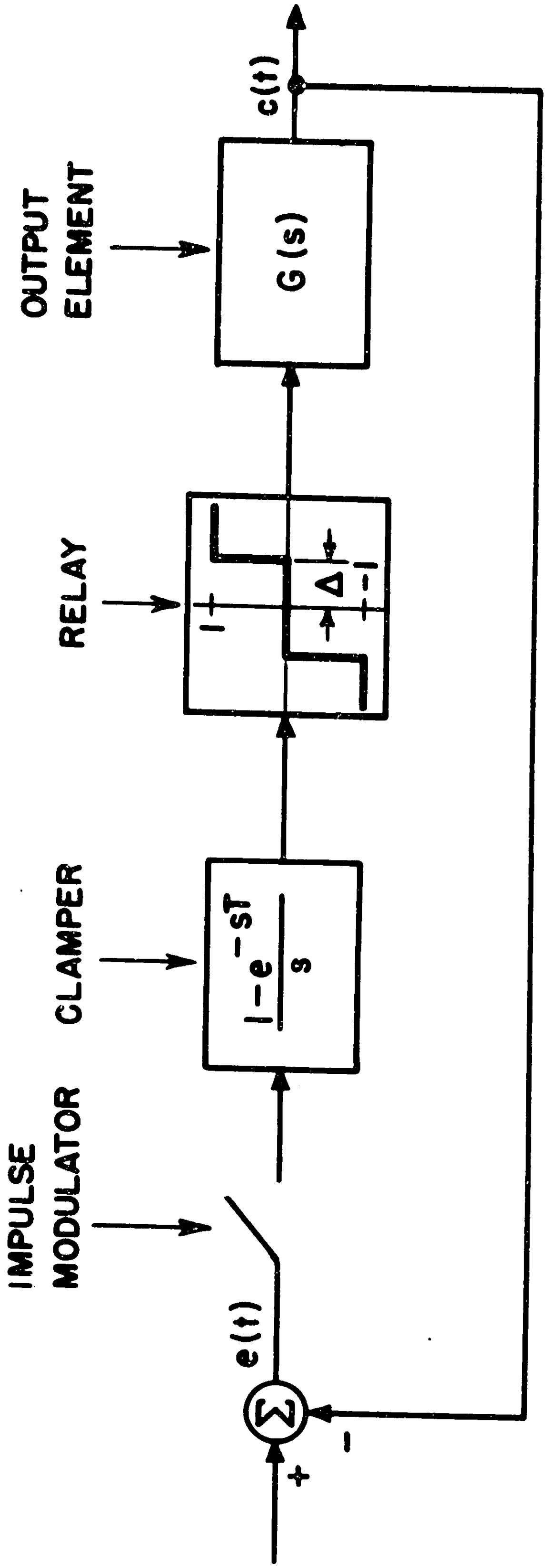

IMPULSE MODULATOR
CLAMPER
RELAY
OUTPUT ELEMENT
e(t)
c(t)
G (s)
$\frac{1-e^{-sT}}{s}$
Δ
1
-1
+
-

FIG. 1

Chow's analytical conclusions were checked on a real-time analog computer and they stand verified in every instance. His results point generally to the conclusion that a relay servo with small deadzone and conventional types of $G(s)$ may possess, as a rule, several oscillating modes. For instance, if $G(s) = 1/s(s + 1)$ and $\Delta = 0.1$ $(T = 1)$, then there are two modes with period $4T$, one with $5T$, and one with $6T$.

A similar investigation carried out by Russell at Columbia[2] uses a different representation for the describing function but apparently arrives at entirely similar conclusions. Further discussion is impossible since the writer has only a fragmentary knowledge of Russell's thesis.

2.2 Critique of the Above Results

Let us now turn to a skeptical examination of the above work.

The describing-function method is an approximate linearization and has at least two major shortcomings. First, it is known from recent studies, especially those carried out in France (see, for instance, an article by J. Loeb[5]) that condition (2.1) while necessary is not sufficient to insure the existence of self-sustained oscillations because a value $j\omega$ satisfying (2.1) may represent an unstable oscillation. Chow has been apparently unaware of this ramification inherent in his method and has been criticized for it by Russell in the discussion of the AIEE paper[1b].

A further difficulty at this point arises regarding the deeper nature of the steady-state oscillations observed. In the classical Poincaré theory of second-order continuous differential systems[6], self-sustained oscillations, i.e. limit cycles in the dignified terminology, always occur in such a manner that stable limit cicles are separated from one another by unstable ones, and the initial conditions leading to the various stable limit cycles always form a connected set as shown in Fig. 2 by the various cross-hatched regions (heavy circles: stable limit cycles; dashed circled: unstable limit cycles.) Russell[1b] calls "each mode of operation satisfying the criterion [(2.1)] a <u>'sort of'</u> limit cycle of sustained oscillations", because he did not observe the regular phase-plane structure of Fig. 2 in the course of his high-speed analog computer investigation of systems of the type Fig. 1.

Both of these above questions, namely (1) stability of the solutions of (2.1), and (2) the nature of the limit cycles in the phase plane, will be fully answered by the present investigation. This is done by using an <u>exact</u> mathematical method; the writer feels that such questions simply cannot be satisfactorily resolved at the present time by the use of the describing-function method.

The <u>second</u>, and even more fundamental shortcoming of the describing-function method is the assumption that self-sustained oscillations in the loop are approximately <u>sinusoidal</u>. In phase-plane terms, this is equivalent to the assumption that the limit cycles are approximately elliptical in shape. This assumption is not too serious when we are dealing with second-order continuous systems since then any closed trajectory in the phase plane must necessarily be a <u>simple</u> closed curve, i. e. have an index* of +1, by virtue of the uniqueness of solutions (one and only one trajectory through every noncritical point in the phase plane). In the sampled case, however,

---

*The index of a closed curve in the phase plane is obtained by drawing small vectors at each point of the curve in the direction of the phase-plane trajectories. The total rotation of these vectors as the curve is traversed in the clockwise sense is then $2\pi i$, where i is the index. Thus a simple closed curve which is itself a trajectory has index +1; if the curve crosses itself once, the index is +2, etc. All this is used here only for convenience of terminology.

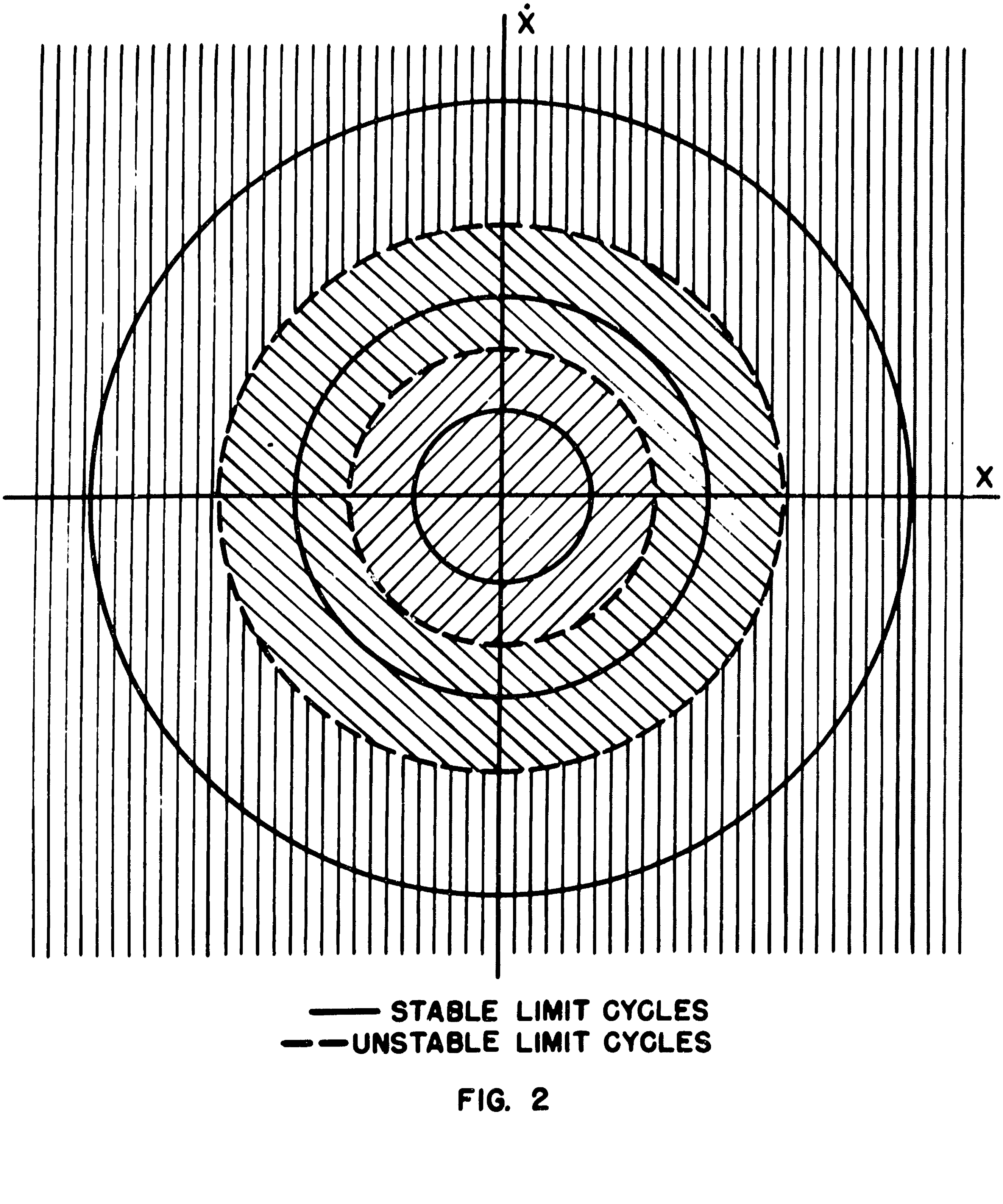


FIG. 2

the proper phase space is actually three-dimensional, so that the closed curve corresponding to the limit cycle, when projected into the phase plane, may cross itself, and the limit cycle in the phase plane may appear as a multiply connected curve of index +2 or even higher. Such an oscillation is distinctly nonsinusoidal, and has not and could not be detected by the describing-function method because of the ex-ante assumption of approximately sinusoidal oscillations used in deriving the describing function. The existence of limit cycles of index higher than +1 is not a mere hypothesis: it has been demonstrated by the writer (Example B) as well as by Altar and Helstrom[3] whose work will be discussed next.

2.3 Results Obtained by the Phase-Plane Method

An entirely different, and in many ways more profound, analysis of the same type of problem has been carried out by Altar and Helstrom[3], using a graphical method based on the phase plane. They investigate the transient response by subdividing the phase plane into finite closed regions such that initial conditions contained within any one region give rise to the same pattern of relay output sequences over time. This provides a relatively simple picture of how a servo synchronizes to a steady-state oscillation and will be considerably elaborated upon in Sect. 6 in connection with the examples.

Further, on the basis of an analytical technique, based also on phase-plane notions, Altar and Helstrom were able to devise a procedure for computing points on the limit cycles which correspond to the sampling instants. By this approach, they show the existence of a limit cycle of index +3!

The writer has greatly benefited from the conceptual content of the work of Altar and Helstrom; in fact, much of this thesis is directed toward a more general formulation of the problem from the phase-plane point of view. The great advantage of the phase-plane method, which so far has not been claimed by any other, is the possibility of establishing exact results. The importance of this, the writer feels, can hardly be overemphasized in the present state of our knowledge of nonlinear systems.

3. OUTLINE OF PRESENT WORK

3.1 General

In what follows, we will be intimately motivated by phase-plane concepts; in particular, our approach to the study of transient response will be methodologically identical with that of Altar and Helstrom. In addition, we will make full use of the powerful analytical tool of difference equations. The latter approach has not been utilized in any of the three investigations referred to above; it helps to streamline the mathematical analysis a great deal in view of the fact that linear difference equations are the standard formalism of studying linear sampled-data systems.

Unfortunately and unavoidably, the ensuing discussion will make fairly strenuous demands on most of the "sophistication" the automatic control engineer has acquired since the end of World War II. It is not felt, however, that the various mathematical concepts could be, without injustice to the material, included in appendices or thoroughly discussed each time they arise in the text; hence we will be compelled to refer to the existing literature for terminology and mathematical details, much of which, to be sure, is rapidly becoming standardized.

3.2 Difference Equations

As pointed out in the Introduction, the standard approach to the analysis of linear sampled-data systems consists in formulating the input-output relationships in terms of difference equations which are valid at the sampling instants only. The actual labor in doing this is greatly lessened by employing frequency-domain techniques, which are now widely known under the name of z-transforms, which enables one to convert continuous transfer functions to discrete (also called pulsed, starred) transfer functions which correspond to the difference equations. It is assumed that the reader is well familiar with the elementary aspects of this material.*

3.3 Phase-Plane Representation

The standard tool of nonlinear system analysis is the phase-plane representation of solutions of the differential equations governing the system. Once again, it must be supposed that the reader is thoroughly familiar with this important theory.** Actually, we will make only scant use of the more advanced mathematical concepts connected with phase-plane theory and will emphasize qualitative arguments having to do with the geometry of trajectories of simple linear systems.

3.4 Mathematical Framework

We are now in a position to begin a precise discussion of the problem at hand. Assume that in the system Fig. 1, G(s) represents a second-order

---

*The reader unacquainted with this theory may find the writer's survey[7] or various other recent papers[8-9] particularly helpful. Specifically, the writer's paper seems to be the only one at present where the difference-equatic and z-transform methods are discussed side-by-side from the engineer's viewpoi

**We refer to Minorsky's classic monograph[6] or the writer's recent paper for additional information.

transfer function. Then, by the application of the z-transform method which will be elaborated upon in Sect. 5.1, and disregarding for the moment the relay, the open-loop differential equation is transformed into a difference equation relating error and output which may be written in general as:

$$a_1c_k + a_2c_{k-1} + a_3c_{k-2} = b_2e_{k-1} + b_3e_{k-2} \tag{3.1}$$

Further, it is easy to see that nothing is altered if the positions of the impulse modulator-clamper combination and the relay (or any other nonlinear element) are interchanged; this enables us to establish a relationship between error and output and thus close the loop (assuming the input to be identically zero in all cases for the rest of this investigation):

$$e_k = f(-c_k) = -f(c_k) \tag{3.2}$$

where $f(x)$ or simply f denotes the functional relationship characterizing the nonlinear element in question(10); for practical reasons, it is always assumed that f is an odd function.

Equations (3.2) and (3.1) together represent the complete mathematical description of the dynamic behavior of the closed loop; the next logical step is to convert them into the standard form of two coupled, first-order equations that are used in phase-plane work, with the sole distinction that now the difference operator E defined by $Ex_k = x_{k+1}$ will appear in the place of the differential operator D $(= d/dt)$:

$$\left.\begin{aligned} Ex_k &= y_k \\ Ey_k &= -((a_2 + b_2f)y_k + (a_3 + b_3f)x_k)/a_1 \end{aligned}\right\} \tag{3.3}$$

where we have made the substitutions $x_k = c_{k-2}$ and $y_k = c_{k-1}$; the expressions $(a_2 + b_2f)$ and $(a_3 + b_3f)$ are to be regarded as new functional operators acting on whichever variable follows them.

It is important to thoroughly understand the significance of equations (3.3). On the one hand, (3.3) may be regarded as defining a nonlinear transformation T of the (x, y) plane into itself, i.e. mapping every point (x_k, y_k) of the plane into another point (x_{k+1}, y_{k+1}) in the same plane.

On the other hand, since x_k and y_k together represent two parameters which specify the state of the system at any one sampling instant, and because the output of the servo is, for physical reasons, always continuous, even though only its sampled values $c(kT) \equiv c_k$ appear in (3.3), it is clear that the plane (x, y) may also be regarded as a phase plane (which should, however, not be confused with the phase plane (c, dc/dt)) in which we have the continuous trajectory $c(t) = F(c(t - T))$, that is, a curve y(x). The points (x_k, y_k), (x_{k+1}, y_{k+1}), ... (x_{k+n}, y_{k+n}) will all lie along this continuous trajectory as illustrated in Fig. 3. This latter viewpoint is particularly pleasing since it reestablishes in a way the continuity of the physical aspects of the problem which "somehow" got lost when the difference equations were set up. In other words, it fills the gap concerning the behavior of a system between the sampling instants, which is a serious problem even in the analysis of linear sampled-data systems.

All the heuristic beauty of the reinterpretation of nonlinear sampled-data systems in terms of the conventional phase-plane representation should not obscure the fact that it is the nonlinear transformation (3.3) and it alone which has to be understood to obtain insight into the problem. The quantitative aspects of systems dynamics are all "buried" in (3.3)!

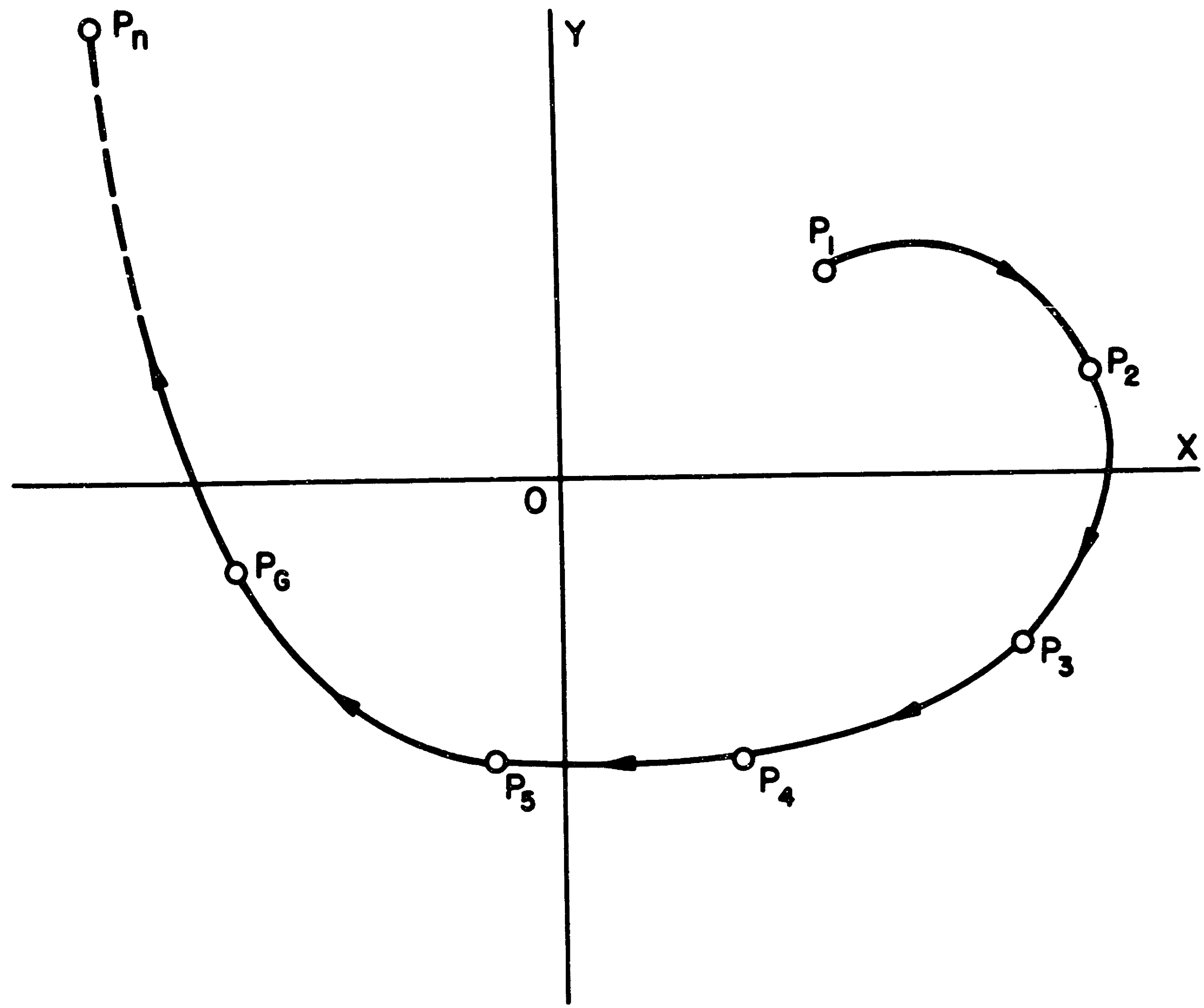


FIG. 3

Since (3.3) is an exact mathematical characterization of system dynamics, our job would be essentially completed if we could now proceed to set up a general theory of second-order nonlinear difference equations. Alas, this appears to be at present hopelessly difficult, at least from the vantage point of this thesis.

Although the succeeding pages will do much to further clarify the behavior of nonlinear sampled-data systems for any practical calculations, the reader must be asked to rely on (3.3) and compute the transient response step-by-step. Similarly, the computation of limit cycles must also be a stepwise trial and error process as will be explained in Sect. 5.2.

4. QUALITATIVE DISCUSSION OF PERIODIC SOLUTIONS

As stated in the Introduction, our main objective is to better understand the nature of the periodic solutions of the difference equations (3.3). Before narrowing down the discussion to purely formal terms, it is highly instructive to say a few words about the fundamental reason for the existence of steady-state oscillations, using simple qualitative arguments.

Consider a system as in Fig. 1, with $G(s) = 1/s(s+1)$. If there is no sampling in the loop, and control is effected by means of an ideal relay without deadzone, with output levels of ± 1 as indicated in Fig. 1, it is clear that the system is governed by the following set of linear differential equations:

$$\left.\begin{aligned} \ddot{c} + \dot{c} &= -1 \qquad c > 0 \qquad \text{(a)} \\ \ddot{c} + \dot{c} &= +1 \qquad c < 0 \qquad \text{(b)} \end{aligned}\right\} \qquad (4.1)$$

A transient solution of (4.1) is sketched in the phase-plane diagram Fig. 4a. It is easy to see that all trajectories, regardless of initial conditions, will ultimately spiral into the origin, in such a manner that the relay

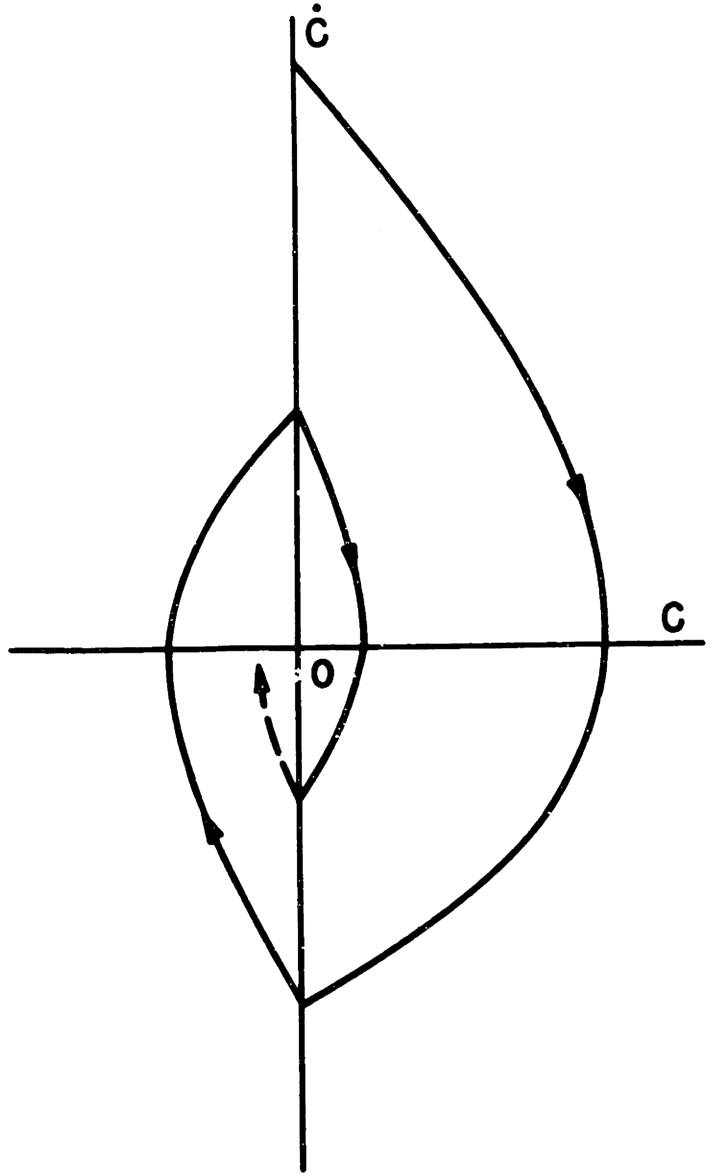


FIG. 4a

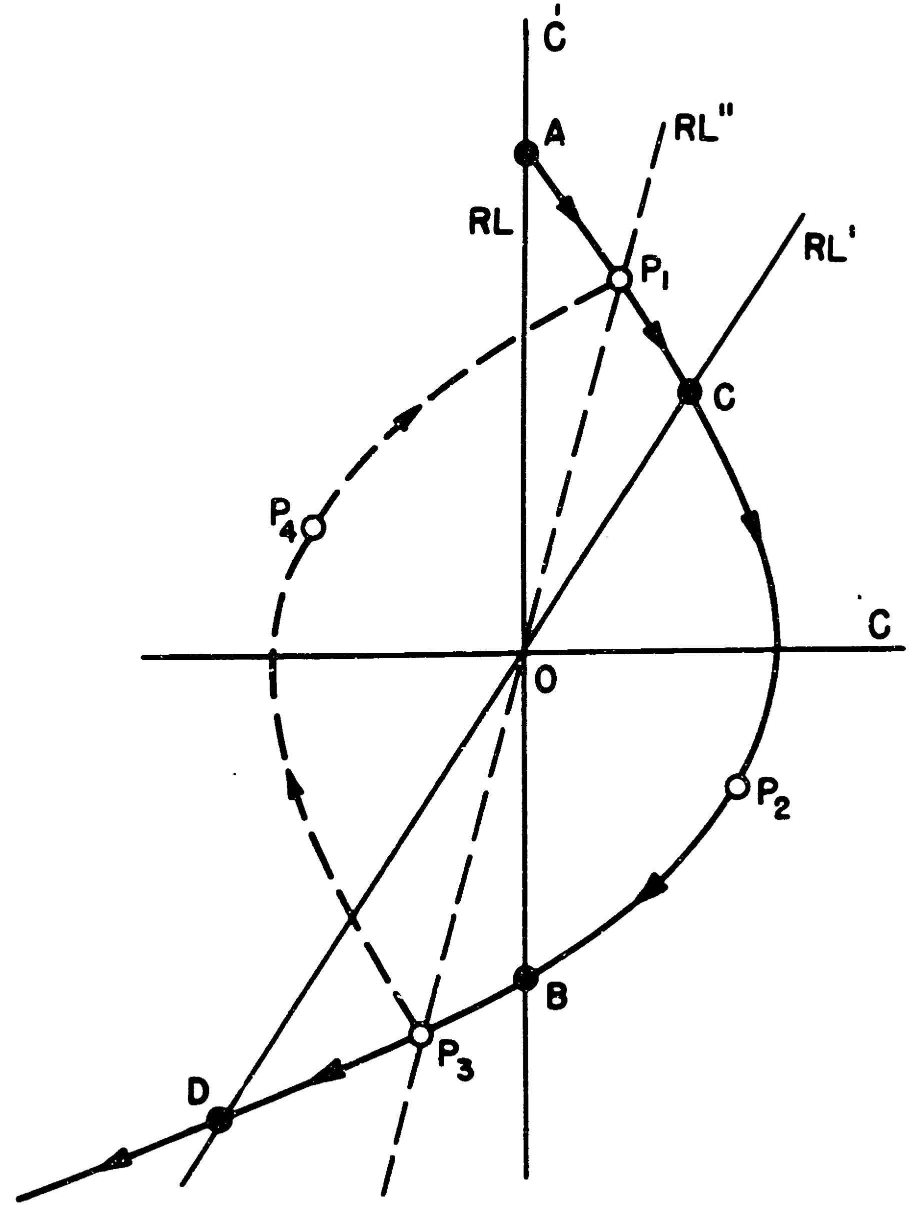


FIG. 4b

reverses its polarity every time that trajectory crosses the $\dot{c}$ axis.

If sampling is now introduced in the loop, the crossing of the "reversal line", i.e. the $\dot{c}$ axis, may not be "noticed" by the system until $0 \leq t < T$ seconds later. If the time delay is sufficiently large, a limit cycle may result.

This may be explained further by reference to Fig. 4b, disregarding at first the presence of sampling. If the reversal line (RL) is the $\dot{c}$ axis, then successive crossings will take place closer and closer to the origin, i.e. $\overline{OA} > \overline{OB}$. If the reversal line is sufficiently oblique (RL'), however, then the reverse may hold: $\overline{OC} < \overline{OD}$. Therefore, by continuity, there must exist some reversal line RL" between RL and RL' such that the crossings are equidistant from the origin: $\overline{OP_1} = \overline{OP_3}$. The arc $P_1P_2P_3$ belongs to the system (4.1a); by symmetry then there must also be an arc $P_3P_4P_1$ belonging to the system (4.1b) which has the same property, namely $\overline{OP_3} = \overline{OP_1}$. Therefore if the reversal line is RL", a closed trajectory or limit cycle $P_1P_2P_3P_4P_1$ will result By an extension of the above arguments, it is easy to show further that if a trajectory starts in the area bounded by the limit cycle, it will grow and converge to the limit cycle monotonically from the inside, and if a trajectory starts anywhere else in the phase plane, it will converge to the limit cycle from the outside. This proves that the assumed limit cycle is stable.

If now the time delay in reversing the relay after passing through the $\dot{c}$ axis is sufficiently large, then the points P may be identified with the sampling instants, and we have a limit cycle of period 4T.

The period of any possible limit cycle must obviously be 2nT (n = positiv integer); moreover, a little reflection show that no matter how high the sampling frequency may be raised, a solution of period $2T$ will always exist, since no matter how small the time delay in crossing the $\dot{c}$ may become one can always find an amplitude which is sufficiently small so that a closed trajectory may be constructed.* Thus, strictly speaking, sampling introduces an essentially novel physical element in the loop which cannot be neutralized by raising the sampling frequency arbitrarily high (although, of course, the amplitude of the limit cycle approaches zero and its practical significance soon vanishes).

Since the lowest mode of oscillation cannot be suppressed by reducing the sampling period, the question naturally arises as to what determines the highest possible mode of oscillations? Imagine that we have formed an infinite continuous family of closed contours in the phase plane, composed of arcs belonging to (4.1a) and (4.1b); we are not now considering sampling. Connect the vertices of these contours by a smooth curve, called the critical reversal line (CRL). The situation is sketched in Fig. 4c. We now construct another curve representing the locus of all end-points of trajectories T seconds after they have crossed the $\dot{c}$ axis and call it the delayed reversal line (DRL); both curves are shown in Fig. 4d. It is seen that there exists some closed contour A whose vertices are the intersections of CRL and DRL. Any trajectory originating outside of A will have its relay polarity reversed before reaching CRL, and con-

---

*It is perhaps fitting to formulate this argument with rigor: If the discontinuity in the ideal relay characteristic at the origin is replaced by an arbitrarily high but finite slope, an examination of the difference equations of the system shows that the origin is a point of unstable equilibrium. (This is merely a consequence of the well-known fact that a linear second-order control system whose open-loop transfer function is stable cannot be made unstable by any arbitrarily high gain, but if the system contains sampling, it can be made unstable by increasing the gain.) Since the system obviously possesses stability in-the-large, the presence of an unstable equilibrium point implies the existence of at least one stable limit cycle by Bendixson's theorem.

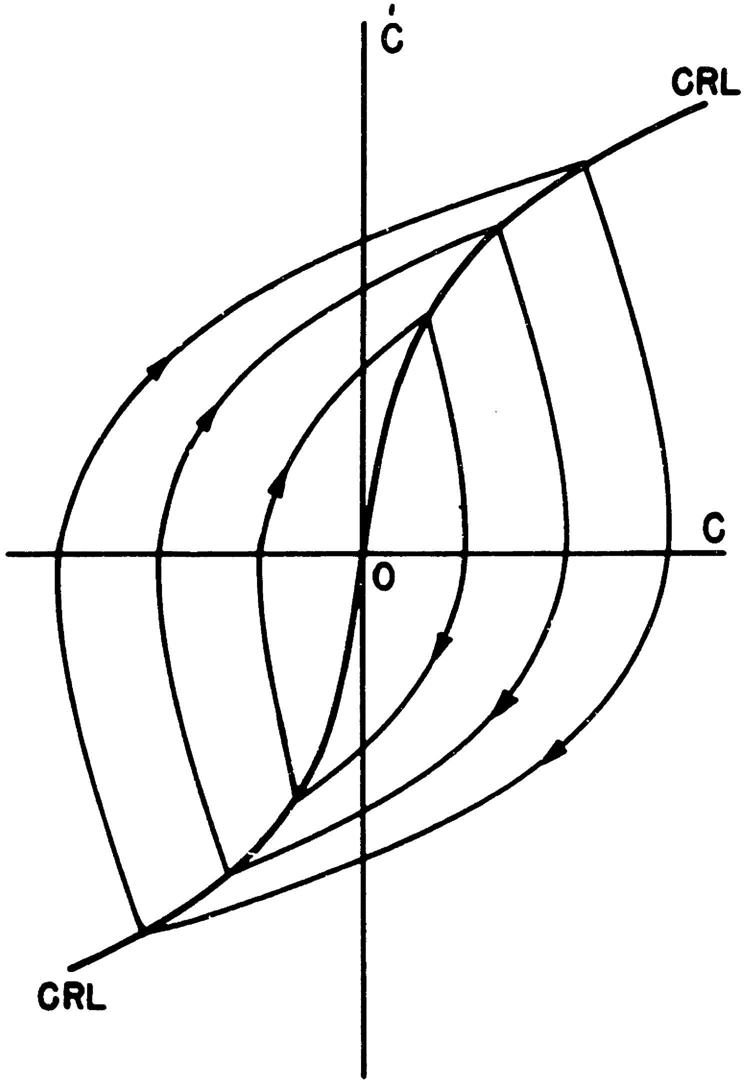


FIG. 4c

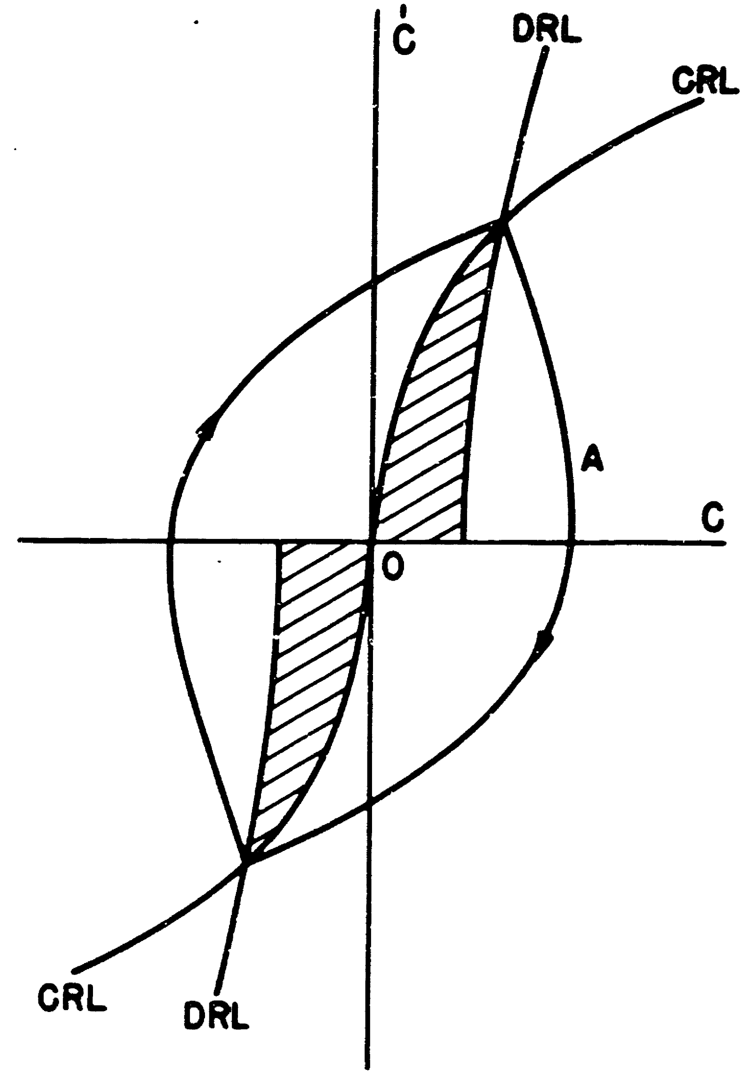


FIG. 4d

sequently it will converge toward the origin, until it enters the cross-hatched region, after which its behavior cannot be reliably predicted qualitatively but requires a more exact analysis. Of course, A does not constitute a limit cycle (except if by sheer accident its period is exactly 2nT), but we can be sure that there exists at least one stable limit cycle with vertices inside the cross-hatched area. The reasoning used here, which is quite general, guarantees that the maximum possible amplitude of oscillations is limited, which is a pleasing and plausible conclusion from the physical point of view.

The qualitative but not entirely heuristic explanations given above should help in visualizing the phenonenon of limit cycles in sampled-data systems. Moreover, what was said agrees in all points with the conclusions drawn from the describing-function method as the reader may verify by consulting Appendix A. In fact, the two viewpoints, time delay here and phase shift there, have much in common without, however, being identical.

Unfortunately, a qualitative explanation is not sufficient in the present case. The cross-hatched areas of Fig. 4d may contain a large number of periodic solutions; if we are interested in the "majestic" question of how the transients from outside the cross-hatched regions will synchronize to one of the periodic oscillations, it is clear that the details of transient behavior far away from the origin are of little interest but everything depends on <u>how</u> the trajectories enter the cross-hatched regions. At this point, the problem becomes distinctly <u>quantitative</u>. Accordingly, the rest of this investigation will be concerned with quantitative problems.

These are the topics of the next several sections.

5. QUANTITATIVE ASPECTS OF PERIODIC SOLUTIONS

5.1 Derivation of Closed-Loop Difference Equations

We now show the details of derivation of the closed-loop difference equations (3.3). Consider a system as in Fig. 1, and assume that the transfer function $G(s)$ has two poles and at most one zero. Expand $G(s)$ into a partial fraction with each term containing one of the poles:

$$G(s) = G_1(s) + G_2(s) \tag{5.1}$$

and rearrange the block diagram accordingly as shown in Fig. 5. In accordance with the usual notation(7-8), let x^* denote the sampled values $x(kT) = x_k$ of any variable $x(t)$. Then the problem is to derive a difference equation relating u^* and v^* ($u^* - v^* = c^*$ from Fig. 5) to

$$f(e^*) = - f(c^*) = - f(u^* - v^*) \tag{5.2}$$

at the sampling instants. In the frequency or z-domain, this relationship is called the pulse (or "starred") transfer function $H(z)$, written symbolically as:

$$\left.\begin{aligned} \frac{U^*(z)}{f(E^*(z))} &= \left[\frac{1-z}{s} G_1(s)\right]^* = H_1^*(x) \\ \frac{V^*(z)}{f(E^*(z))} &= \left[\frac{1-z}{s} G_2(s)\right]^* = H_2^*(z) \end{aligned}\right\} \tag{5.3}$$

The complex variable z of the pulse transfer function is related to the difference operator E and the usual Laplace-transform variable s $(= d/dt)$ by the operational identity

$$z = E^{-1} = \epsilon^{-sT} \tag{5.4}$$

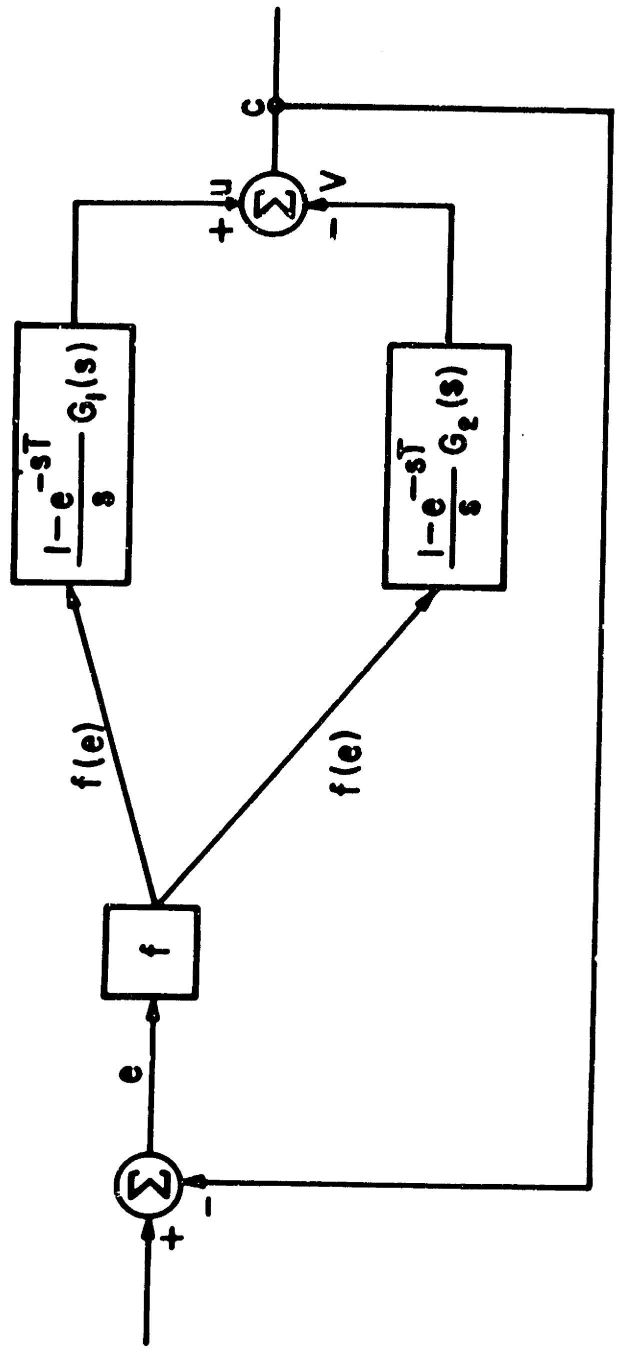


FIG. 5

A word of caution concerning (5.4): Sometimes[8-9] z is defined in the opposite fashion: $z = E = \epsilon^{sT}$. This, of course, is just a simple but confusing change of variable which should be kept in mind while reading the literature. -- The computation of $G^*(z)$ from $G(s)$ involves a summation process which we cannot discuss here; for the reader's convenience a short list of transforms is included in Appendix C.

Returning now to (5.3), and computing the pulse transfer functions $H_1^*(z)$ and $H_2^*(z)$ explicitly, we can write the closed-loop difference equations governing $u^* \equiv u_k$ and $v^* \equiv v_k$ by virtue of (5.2) and (5.4) in general as:

$$\left.\begin{aligned} Eu_k &= \lambda_1 u_k + af(u_k - v_k) \\ Ev_k &= \lambda_2 u_k + bf(u_k - v_k) \end{aligned}\right\} \qquad (5.5)$$

By redefining variables, namely $u_k - v_k = c_k = y_k$ and $y_{k-1} = x_k$, we would obtain the form (3.3) discussed previously; however, for most purposes, especially for step-by-step-computations, the form (5.5) is the most convenient one.* The numerical examples illustrating the computation of (5.5) from a given open-loop transfer function $G(s)$ are presented in Sect. 6.

For completeness, it should be added that the computational approach as outlined above is not at all limited to second-order transfer functions. By a partial-fraction expansion, we can always find a set of n simultaneous

---

*When $G(s)$ has complex poles, the coefficients of the terms in (5.5) will be likewise complex; in this case it is more convenient <u>not</u> to break $G(s)$ down in terms of its poles, but compute the pulse transfer function $((1 - z)G(s)/s)^*$ <u>directly</u>, which leads at once to the form (3.3).

coupled nonlinear difference equations of the first order in $u_k, v_k, w_k, \ldots$ analogously to (5.5). The phase-space representation of such systems is not sufficiently far advanced, however, to permit any meaningful discussion at the present time.

5.2 Existence of Periodic Solutions

If we regard (3.3) or (5.5) as a nonlinear transformation T operating on the vector $\vec{Q}$ with components X, Y or U, V then the existence of a periodic solution implies the existence of a vector $\vec{Q}$ such that

$$T^n \vec{Q} = \vec{Q} \tag{5.6}$$

In other words, $\vec{Q}$ is a fixed point of the n-th iterate of T ($\vec{Q}$ is mapped into itself under T^n). The study of fixed points of transformations of manifolds (in the present case, the two-dimensional Euclidean space) is one of the important tasks of topology. Unfortunately, the writer was unable to uncover any considerations of such problems in textbooks of elementary combinatorial topology, and therefore the above viewpoint did not lead to a useful result. Perhaps this situation will change in the future.

If we regard (3.3) as defining points along continuous phase-plane trajectories, it must be realized at once that knowledge of the values x_k and y_k does not specify the state of the system completely, but there must be added an additional piece of information which shows the state of the system as far as its relative position in time between the sampling instants is concerned. In other words, the addition to x and y there must be a third phase-space dimension $0 \leq t < T$. This is a reassuring fact for its explains the possibility of limit cycles of index higher than +1 which cannot occur in two-dimensional systems. While this viewpoint places into evidence the "causality" of the system (namely that all of its future behavior is completely determined once three initial conditions are specified), it is of no help whatsoever in the absence of a good theory of dynamic systems in three-dimensional phase-space.*

For these reasons, one is forced to adopt an entirely unglamorous but practically fairly convenient method for computing the fixed points of (5.6) by a trial and error procedure. In the following, f will be restricted to be a relay without deadzone denoted by Γ_k and described by:

$$\Gamma_k \begin{array}{ll} = +1 & v_k > u_k \\ = -1 & v_k < u_k \end{array} \Bigg\} \qquad (5.7)$$

Here it is especially convenient to work with the system in the form (5.5). Using (5.5a), set:

$$u_{n+1} = u_1 = \lambda_1^n u_1 + \sum_{k=1}^{k=n} b\lambda_1^{n-k}\Gamma_k \qquad (5.8)$$

as the condition for periodicity, and then assume that the output of the relay is given by a certain sequence of Γ_k's, e.g. + + - - . This makes it possible to evaluate the constant term in (5.8) given by the summation of the Γ_k's, and thereby compute u_1. Then substitute u_1 into (5.5a) and compute u_2, u_3, ... u_n by a stepwise process using the assumed Γ_k's Apply the same procedure to (5.5b) and find v_1, v_2, ... v_n. Then check whether the u's and v's thus obtained satisfy the inequalities (5.7) corresponding to all the assumed values of Γ_k's. If so, the assumed periodic solution exists and the u's and v's represent points along the limit cycle in the phase plane--they are also fixed points of the transformation (5.6).

If the limit cycle is assumed to be of index +1, then (5.8) may be

---

*The phase space involved here is the surface of the four-dimensional cylinder; or, making the substitutions: $r = \exp(x)$, $z = y$, $2\pi t/T = \phi$, the punctured solid cylinder (r, ϕ, z). No published treatment of such three-dimensional dynamic systems exists at the present time in the writer's knowledge.

replaced by the somewhat simpler condition:

$$u_{\frac{n}{2}+1} = -u_1 \qquad (5.8')$$

since half-periods of such a limit cycle must be symmetric.

It is apparent that the inequalities (5.7) will be satisfied, if at all, with some leeway; hence small perturbations of the values of the u's and v's will not immediately destroy the nature of the limit cycle. This is the problem of local stability of periodic solutions which is discussed next.

5.3 Local Stability of Limit Cycles

As pointed out in Sect. 2, this question gives rise to considerable concern in conjunction with the describing-function method. In the case the problem is formulated in terms of difference equations, however, this question is very easily and simply disposed of.

Consider (5.5). Assume further that the nonlinear function $f \equiv \Gamma$ representing the relay can only assume three discrete values: +, o, and -. Therefore the n-th iterate of the nonlinear transformation (5.5) may be written as:

$$\left.\begin{aligned} E^n u_k &= \lambda_1^n u_k + \text{constant} \\ E^n v_k &= \lambda_2^n v_k + \text{constant} \end{aligned}\right\} \qquad (5.9)$$

If (5.5) possesses a periodic solution of period n, i.e.

$$\left.\begin{aligned} E^n u_k &= u_k \\ E^n v_k &= v_k \end{aligned} \quad \text{for some} \quad \begin{aligned} u &= U \\ v &= V \end{aligned}\right\} \qquad (5.10)$$

defining small perturbations around the periodic solution:

$$\left.\begin{aligned} u_k &= \delta u_k + U \\ v_k &= \delta v_k + V \end{aligned}\right\} \qquad (5.11)$$

and substituting into (5.9) leads to the linear perturbation equations:

$$\left.\begin{aligned} E\,\delta u_k &= \lambda_1 \delta u_k \\ E\,\delta v_k &= \lambda_2 \delta v_k \end{aligned}\right\} \qquad (5.12)$$

It is well known that the solutions of (5.12) are stable if

$$|\lambda_1| < 1 \qquad \text{and} \qquad |\lambda_2| < 1 \qquad (5.13)$$

The conditions (5.13) are always satisfied if the poles of the open-loop transfer function all lie in the left-half s-plane. Thus we conclude that all the limit cycles of the system (5.5) are stable if G(s) is stable. The method used here is identical with the Liapounoff theory of stability of autonomous periodic solutions[6]. Equations (5.12) govern the local stability of the limit cycles; the much more difficult question of in-the-large stability is discussed in Sect. 5.42.

It is very interesting to observe that we have reached very general conclusions regarding local stability without any knowledge whatsoever concerning the number or indices of the limit cycles or even their existence. These simple conclusions would not apply, however, if $df(x)/dx \neq 0$ in which case (5.12) would contain other linear terms in δu_k and δv_k which could not be evaluated without an explicit knowledge of the actual magnitudes of U and V in (5.11). Thus the relay servo is surely the simplest possible case.

A peculiar situation arises when $G(s)$ contains an integration. In this case, $\lambda_1 = 0$. Hence along the u axis the limit cycles are neither stable nor unstable. In other words, a small perturbation δu will shift the limit cycle in the (u, v) phase plane by an amount δu without affecting the general shape of the limit cycle but changing its time-average (DC) value. This phenomenon is similar to, but not identical with, the notion of a center which characterizes a linear or nonlinear conservative oscillator. There are an infinite number of approximately elliptical trajectories surrounding a center; any small perturbation will cause the system to oscillate with a slightly different amplitude (and period, in the nonlinear case). Thus a center may be said to be neutral from the stability point of view—just as (5.12) with $\lambda_1 = 0$. The "intermarriage" of the limit cycle of the usual type with a center is unknown in continuous nonlinear systems.

More discussion on this point in conjunction with the examples.

5.4 Transient Behavior

5.41 Generalities. As is well known, solutions of difference equations may always be found step-by-step. Thus one may calculate the transient response of the systems considered here for any initial conditions (u_0, v_0) by substituting into (5.5), finding (u_1, v_1), resubstituting into (5.5), and repeating the process. Once the system response at the sampling instants has been determined, it is not difficult (making use of the knowledge of the step response of $G(s)$) to connect these by a smooth continuous curve to obtain approximately the continuous output of the system. This has been done in connection with Example B of Sect. 6 (cf. Figs. 8-10).

5.42 Prediction of Transients in the Phase Plane. The step-by-step calculation of transients affords virtually no insight into the nature of transient behavior over the whole phase plane. A considerable amount of information may be obtained, however, by the following procedure proposed by Altar and Helstrom[3], who concentrate on determining the relay output sequences (i.e., some combination of + 1, 0, and - 1 which eventially becomes repetitive if the transient synchronizes to one of the modes of self-sustained oscillations.) We assume here that $\Delta = 0$ and will deal with the closed-loop equations in the form (5.5). Then we may proceed as follows:

The condition that the first output of the relay be + 1 (adopting the usual convention that the first sampling instant is at $t = 0$) is given by the inequality: (cf. equations (5.7))

$$+ \qquad v_1 > u_1 \tag{5.14}$$

In general, the condition that the output of the relay be positive at the k-th sampling instant is

$$+ \qquad v_k > u_k \tag{5.15}$$

Now if the output of the relay is known for the preceding k-1 sampling instants, then the inequality (5.15) may be evaluated in terms of u_1 and v_1 by making use of (5.5) as:

$$+ \quad v_k = \lambda_2^{k-1} v_1 + \sum_{i=1}^{i=k-1} b \lambda_2^{-k+1+i} \Gamma_i > u_k = \lambda_1^{k-1} v_1 + \sum_{i=1}^{i=k-1} a \lambda_2^{-k+1+i} \Gamma_i \tag{5.16}$$

Therefore we may proceed in a step-by-step fashion as follows: At the first step, (5.14) divides the phase plane into two regions: + and -, separated by the line $u = v$. In each of the regions, the magnitude of relay output,

i. e. Γ_1 is known; hence computing (5.16) for the two values of Γ_1 gives the four regions $++$, $+-$, $-+$, $--$. At this stage, we know the values of Γ_1 and Γ_2 in each of the four regions, and therefore (5.16) for $k = 3$ gives rise to four new inequalities, each valid for one of the four regions obtained in step 2. Thus at each step, each region is subdivided into two parts, adding a $+$ and $-$ to the already existing sequence of the region in question. Lest the reader jump to the conclusion that the number of regions grows as 2^n, we remark:

(1) In carrying out the above computations, we naturally have to restrict our attention to a finite portion of the phase plane, centered around the origin.

(2) Only sequences <u>starting</u> with $+$ need be considered because of the symmetry of the phase-plane trajectories around the line $v - u = -c = 0$.

(3) Not every inequality, defined by (5.16), will fall inside the region defined by the previous relay output sequence. This is a consequence of the existence of periodic solutions and the fact that they are locally stable for small perturbations.

In spite of the above factors, the resulting number of regions is still extremely large in most practical cases. This prohibits detailed consideration of the nature of the transients; but it does enable one to say a great deal about the final steady-state solutions.

5.43 <u>Stability In-The-Large</u>. For reasons explained in Sect. 4 on the basis of Fig. 4d, the most important question concerning transient response is not the behavior of trajectories at $e \gg 0$ which is qualitatively the same as in the case when there is no sampling, but the events which take place after the transient has entered the cross-hatched region of Fig. 4d.

In other words, our main interest lies in predicting to which limit cycle a trajectory starting far away from the origin will finally synchronize to.

The procedure outlined in Sect. 5.42 may be considered as a convergent iterative method for computing the boundary curves (which must necessarily be straight lines because of the form of the inequality (5.16)) which separate certain regions in the phase plane so that all trajectories originating in a given region converge to one limit cycle. These boundary lines reveal the structure of in-the-large stability of the limit cycles.

It should be added that the computational process of Sect. 5.42 must be carried out to a large number of steps in many cases before it can be decided with certainty that the transient will synchronize to some limit cycle. The boundary curves separating the regions, which appear as broken straight lines at the earlier stages of the process, become more and more regular as the iteration progresses.

6. EXAMPLES

We now present the principal quantitative results of this investigation in the form of two examples for which a detailed transient analysis has been made in the phase plane. The forbiddingly large amount of labor required to carry out the pertinent iterative calculations as outlined in Sect. 5.42 is responsible for the lack of further or more complicated examples.

6.1 Example A

Consider a system as defined in Fig. 1 $(\Delta = 0)$, with

$$G(s) = \frac{1}{s(s+1)} \tag{6.1}$$

whose z-transform is, setting $T = 1$,

$$H^*(z) = (1 - z)\left[\frac{1}{s^2} - \frac{1}{s(s+1)}\right]^*$$

$$= \frac{z}{1 - z} - \frac{(1 - \epsilon^{-1})z}{1 - \epsilon^{-1}z} \qquad \epsilon = 2.718\ldots \tag{6.2}$$

We identify the first term in the partial-fraction of $G^*(z)$ with u, and the second term with v as in Fig. 5. Then the dynamic behavior of the closed-loop system in the (u, v) phase plane is governed by the difference equations:

$$\left.\begin{aligned} Eu_{k+1} &= u_k + \Gamma_k \\ Ev_{k+1} &= \epsilon^{-1}v_k + (1 - \epsilon^{-1})\,\Gamma_k \end{aligned}\right\} \tag{6.3}$$

where Γ_k is defined as in (5.7) is the mathematical description of an ideal relay without deadzone.

(Page [illegible] follows)

There are three different periodic solutions (all of index +1) as listed in Table I. They have been found by trial and error as outlined in Sect. 5.2 Since the damping is zero for the u coordinate, the periodic values of u are neither stable nor unstable locally; a small perturbation changing u will not affect the mode of oscillation. The values of u shown in Table I are average values and were computed by the use of (5.8'). We see again that only even integer periods are possible, by virtue of $\sum_{k}^{n} \Gamma_k = 0$ from (6.3a). From the discussion of Sect. 4, it is clear that <u>all</u> limit cycles are <u>locally</u> stable since G(s) is stable.

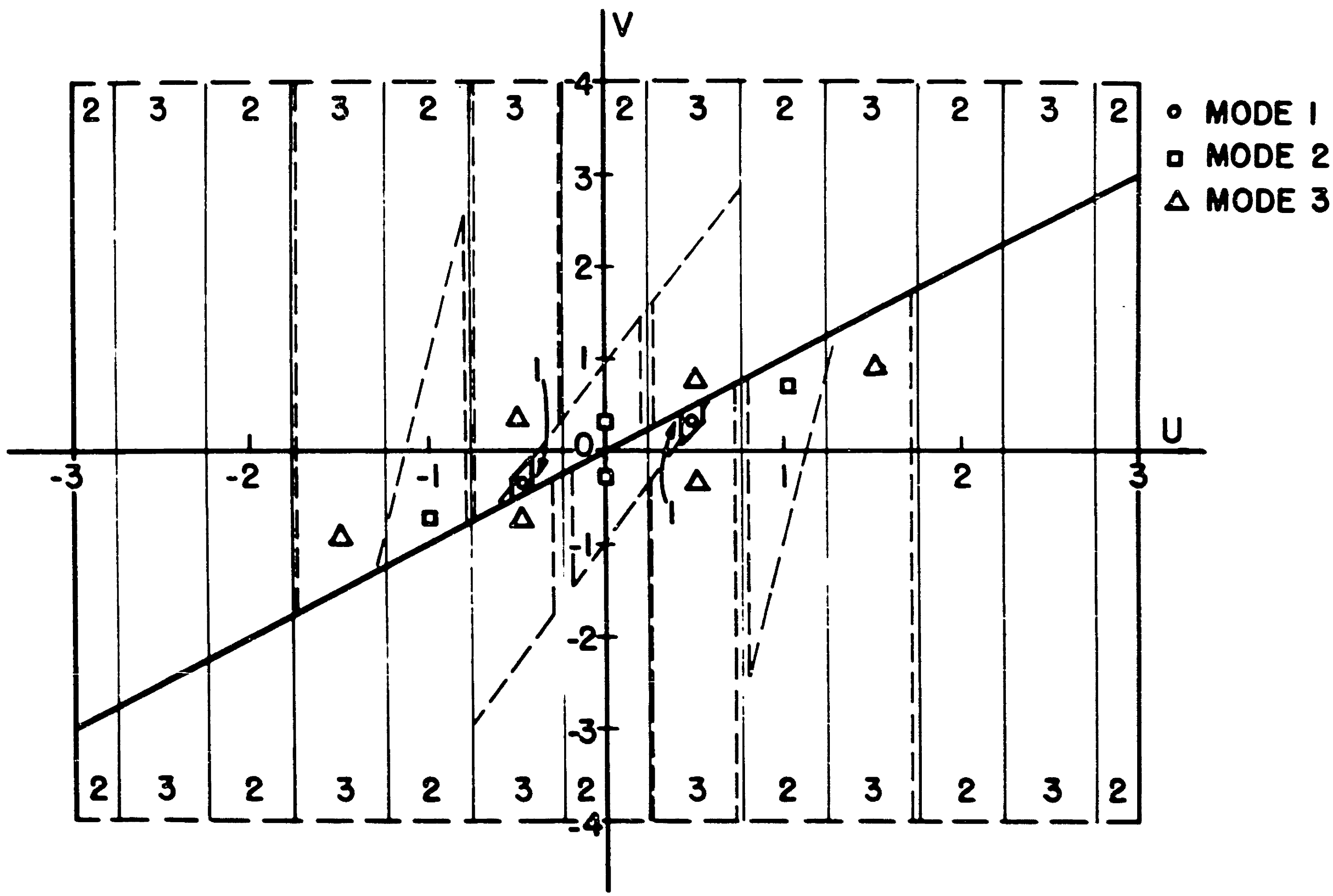
The question of stability in-the-large, as well as the possibility of the existence of periodic solutions not detected by the trial and error method, is resolved by the calculation of relay output sequences by the method outlined in Sect. 5.4. This computation has been carried out to 15 steps in this example, giving rise to roughly 160 regions of different relay output sequences contained in a 6 x 8 rectangle centered at the origin of the (u, v) phase plane. The detailed diagram cannot be reproduced here because of its complexity. However, Fig. 6 shows the in-the-large stability structure of the phase plane, i.e. regions of initial conditions are labeled as to which of the three possible limit cycles the transient will ultimately converge to. Fig. 6 also shows the regions of local stability (i.e. regions where the relay output sequences are periodic for any initial condition) as "small" closed regions (bordered by dashed lines) around the sampled points of the various limit cycles.

A profound analysis of the features of Fig. 6 is not possible because of the lack of a satisfactory theory of nonlinear difference equations. The following, however, is apparent by inspection:

(1) All transients within the rectangle lead to one of the three periodic solutions uncovered by trial and error (Table I). Since the

TABLE I

<table>
<tr><td rowspan="3">Mode 1
(2T)
Index = +1</td><td>Relay Output</td><td colspan="3">+</td><td colspan="3">-</td></tr>
<tr><td>u</td><td colspan="3">-0.500</td><td colspan="3">+0.500</td></tr>
<tr><td>v</td><td colspan="3">-0.462</td><td colspan="3">+0.462</td></tr>
<tr><td rowspan="3">Mode 2
(4T)
Index = +1</td><td>Relay Output</td><td colspan="2">+</td><td>+</td><td>-</td><td>-</td><td></td></tr>
<tr><td>u</td><td colspan="2">-1.000</td><td>0</td><td>+1.000</td><td>0</td><td></td></tr>
<tr><td>v</td><td colspan="2">-0.762</td><td>+0.352</td><td>+0.762</td><td>-0.352</td><td></td></tr>
<tr><td rowspan="3">Mode 3
(6T)
Index = +1</td><td>Relay Output</td><td>+</td><td>+</td><td>+</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td>u</td><td>-1.500</td><td>-0.500</td><td>+0.500</td><td>+1.500</td><td>+0.500</td><td>-0.500</td></tr>
<tr><td>v</td><td>-0.905</td><td>+0.299</td><td>+0.742</td><td>+0.905</td><td>-0.299</td><td>-0.742</td></tr>
</table>



PHASE-PLANE STRUCTURE, EXAMPLE "A"
FIG. 6

rectangle considered is quite large relative to the amplitude of the largest limit cycle present, we may conclude (recalling the discussion in Sect. 4 on the maximum possible amplitude of the limit cycles) that no other limit cycles exist.

(2) The vast majority of initial conditions inside the rectangle lead to either Mode 2 or 3; only two very small regions of initial conditions lead to Mode 1. This fact is quite unexpected but can be explained by a qualitative argument. See Appendix D.

(3) If we disregard the regions leading to Mode 1, it is seen that the two other kinds of regions alternate in a periodic fashion in vertical strips across the entire rectangle. On closer examination, it is seen that the total width of two adjacent regions is unity. This is a consequence of the form of (6.3a): namely, the u-coordinate of any point is mapped one unit to the right or left at each sampling instant, hence any vertical strip of unit width $\alpha < u < \alpha + 1$ is mapped into the adjacent strip of unit width. The moment that any damping is introduced into (6.3a), this phenomenon would disappear.

6.2 <u>Example B</u>

This example was picked solely for the sake of computational convenience: by accident it happens to exhibit a number of important features and offers an interesting contrast to the preceding case.

Here the open-loop transfer function is*

$$G(s) = \frac{0.305s + 0.695}{s(s + 0.695)} \qquad (6.4)$$

By a procedure which is entirely analogous to that followed in Example A, the closed-loop difference equations governing the system dynamics are found to be:

$$\left.\begin{aligned} Eu_{k+1} &= u_k + \Gamma_k \\ Ev_{k+1} &= \frac{1}{2} v_k + \frac{1}{2} \Gamma_k \end{aligned}\right\} \qquad (6.5)$$

where Γ_k is defined as in (5.7).

By a trial and error procedure, it is established that five different limit cycles exist; three of index +1, and two of index +2. Numerical information is given in Table II where the values of u are again averages.

The relay output sequences have been calculated up to at least 15 steps inside a 5×6 rectangle in the phase plane. In some cases, as many as 18-20 steps (!) had to be checked in order to ascertain the mode to which the transients in certain regions converge. The entire calculation yielded well over 200 (!) distinct regions of different relay output sequences within the rectangle considered. The representation of these regions with a

---

* The presence of a zero in $G(s)$ is <u>not</u> the same thing as an idealized lead network frequently used in the discussion of relay servos. Such a lead network, or rather a left-half-plane zero is inserted in <u>front</u> of the relay, and would shift the reversal line so that the switching of the relay would occur earlier in time; this, in general, tends to improve stability. In the present case, the zero occurs <u>after</u> the relay, and introduces a discontinuity in the phase-plane trajectories whenever the relay polarity is reversed; this is apparently detrimental to stability in the present sampled-data case. Thus putting a zero before or after the nonlinear element gives rise to essentially different phase-plane situations — a good example of some of the subtle differences between linear and nonlinear systems.

TABLE II

Mode 1 (2T) Index = +1	Relay Output	+			-		
	u	-0.500			+0.500		
	v	-0.333			+0.333		
Mode 2 (4T) Index = +1	Relay Output		+	+	-	-	
	u		-1.000	0	+1.000	0	
	v		-0.600	+0.200	+0.600	-0.200	
Mode 3 (6T) Index = +1	Relay Output	+	+	+	-	-	-
	u	-1.500	-0.500	+0.500	+1.500	+0.500	-0.500
	v	-0.778	+0.111	+0.556	+0.778	-0.111	-0.556
Mode 4 (6T) Index = +2	Relay Output	+	+	-	+	-	-
	u	-1.250	-0.250	+0.750	-0.250	+0.750	-0.250
	v	-0.651	+0.175	+0.587	-0.206	+0.397	-0.302
Mode 5 (6T) Index = +2	Relay Output	+	+	-	-	+	-
	u	-0.750	+0.250	+1.250	+0.250	-0.750	-0.250
	v	-0.397	+0.302	+0.651	-0.175	-0.587	+0.206

reasonable degree of accuracy and clarity required a surface of approx. 50 square feet (!) and for obvious reasons is not included. The phase-plane stability structure analogous to Fig. 6 is, however, "reasonably" simple and regular: it is shown in Fig. 7. We wish to make the following comments on it:

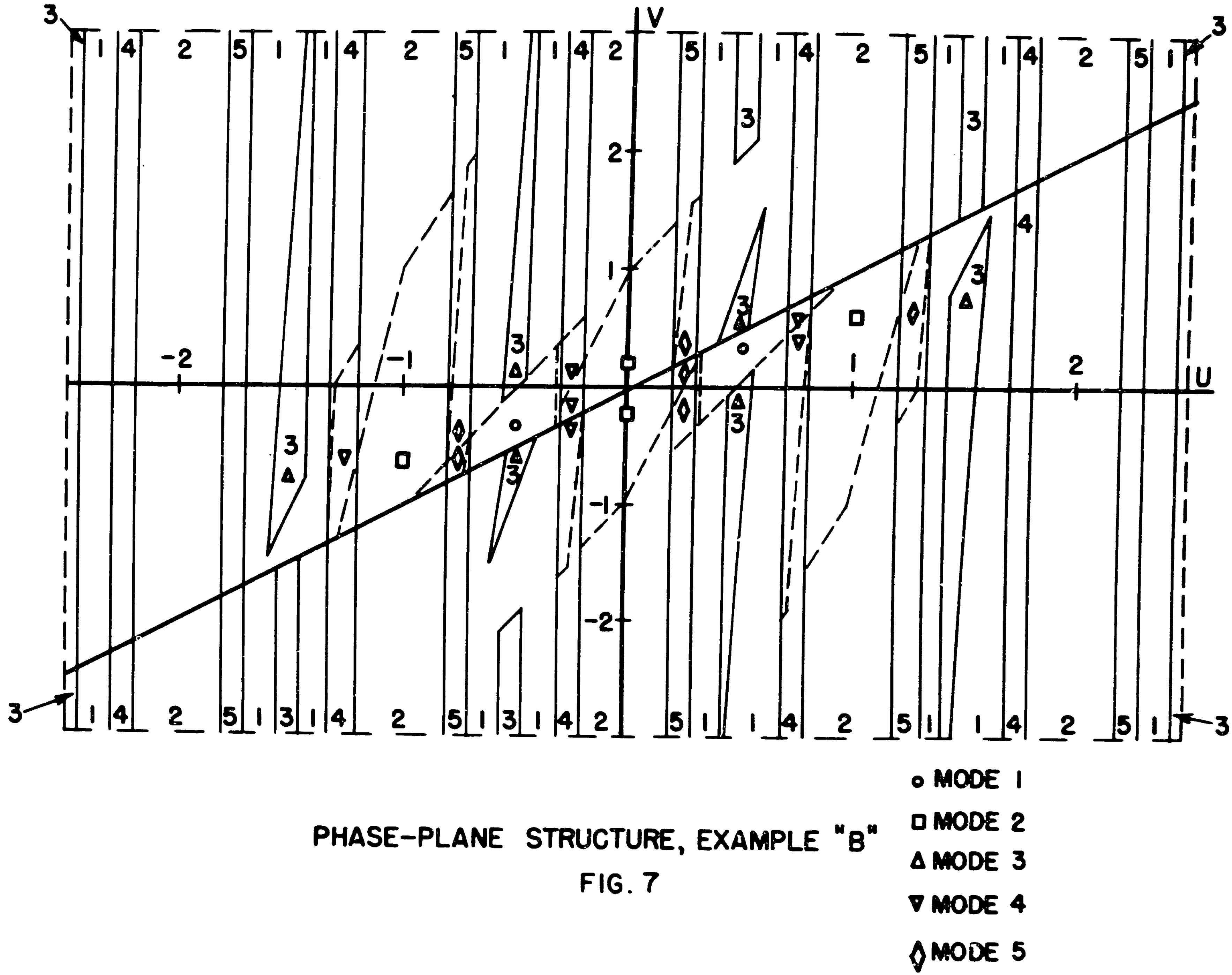
(1) The five limit cycles shown in Table II are manifestly the only ones existing.

(2) In contrast to Fig. 6, there is now a large area of initial conditions leading to the Mode 1. On the other hand, we see that regions leading to Mode 3 are relatively small and do not "out across" the phase-plane but they are "embedded" in the regions leading to Mode 1. A qualitative explanation appears in Appendix D.

(3) A periodicity of regions, similar to that observed in Fig. 6 is again apparent; the explanation is the same as before.

To further illustrate the situation, the transient responses corresponding to three slightly different initial conditions have been calculated step-by-step, using the difference equations (6.7) as a recursion formula. The continuous outputs $c(t) = u(t) - v(t)$ have been plotted in Figs. 8-10, making use of careful graphical interpolation between the sampling instants which are indicated by small circles.

It is interesting to verify by reference to these figures the wisdom of remarks made earlier in Sect. 4 on the basis of qualitative notions, namely, that the transient behavior is relatively uninteresting until the trajectories get very close to the origin, but as soon as they enter a region corresponding to the cross-hatched area of Fig. 4d, it is not possible



PHASE-PLANE STRUCTURE, EXAMPLE "B"

FIG. 7

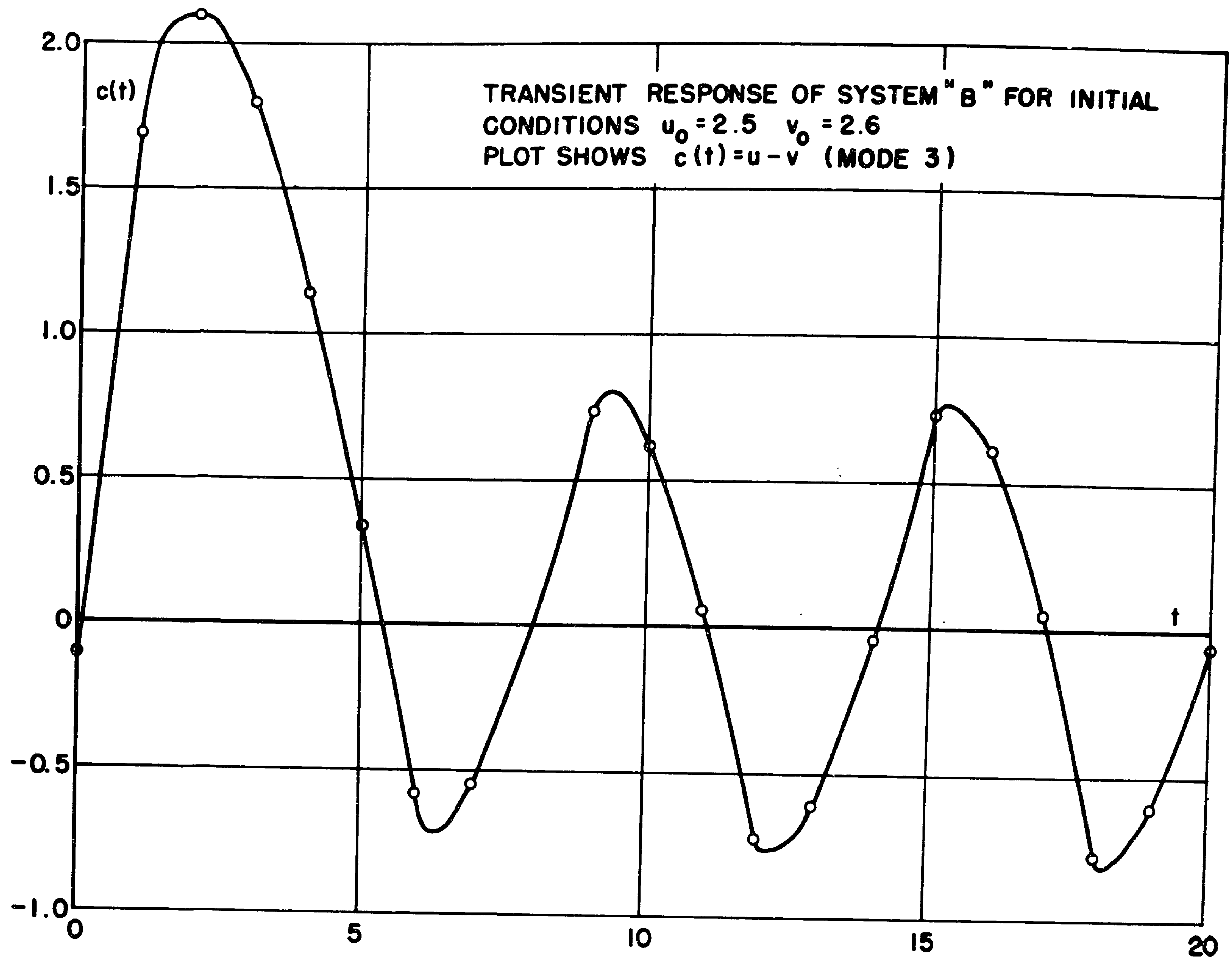

TRANSIENT RESPONSE OF SYSTEM "B" FOR INITIAL
CONDITIONS $u_0 = 2.5$ $v_0 = 2.6$
PLOT SHOWS $c(t) = u - v$ (MODE 3)
c(t)
t
2.0
1.5
1.0
0.5
0
-0.5
-1.0
0
5
10
15
20

FIG. 8

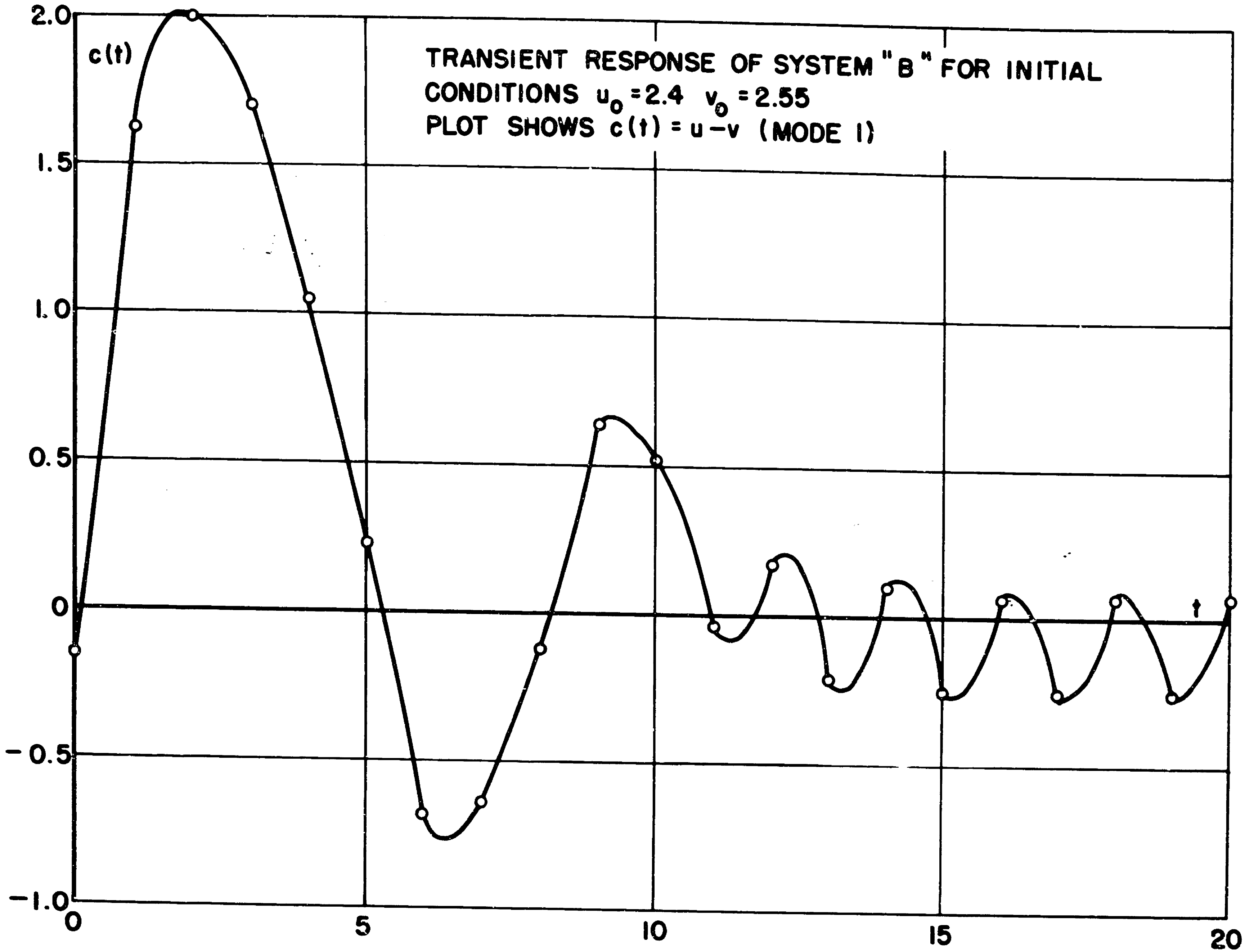

TRANSIENT RESPONSE OF SYSTEM "B" FOR INITIAL
CONDITIONS $u_0 = 2.4$ $v_0 = 2.55$
PLOT SHOWS $c(t) = u - v$ (MODE I)
c(t)
t
2.0
1.5
1.0
0.5
0
-0.5
-1.0
0
5
10
15
20

FIG. 9

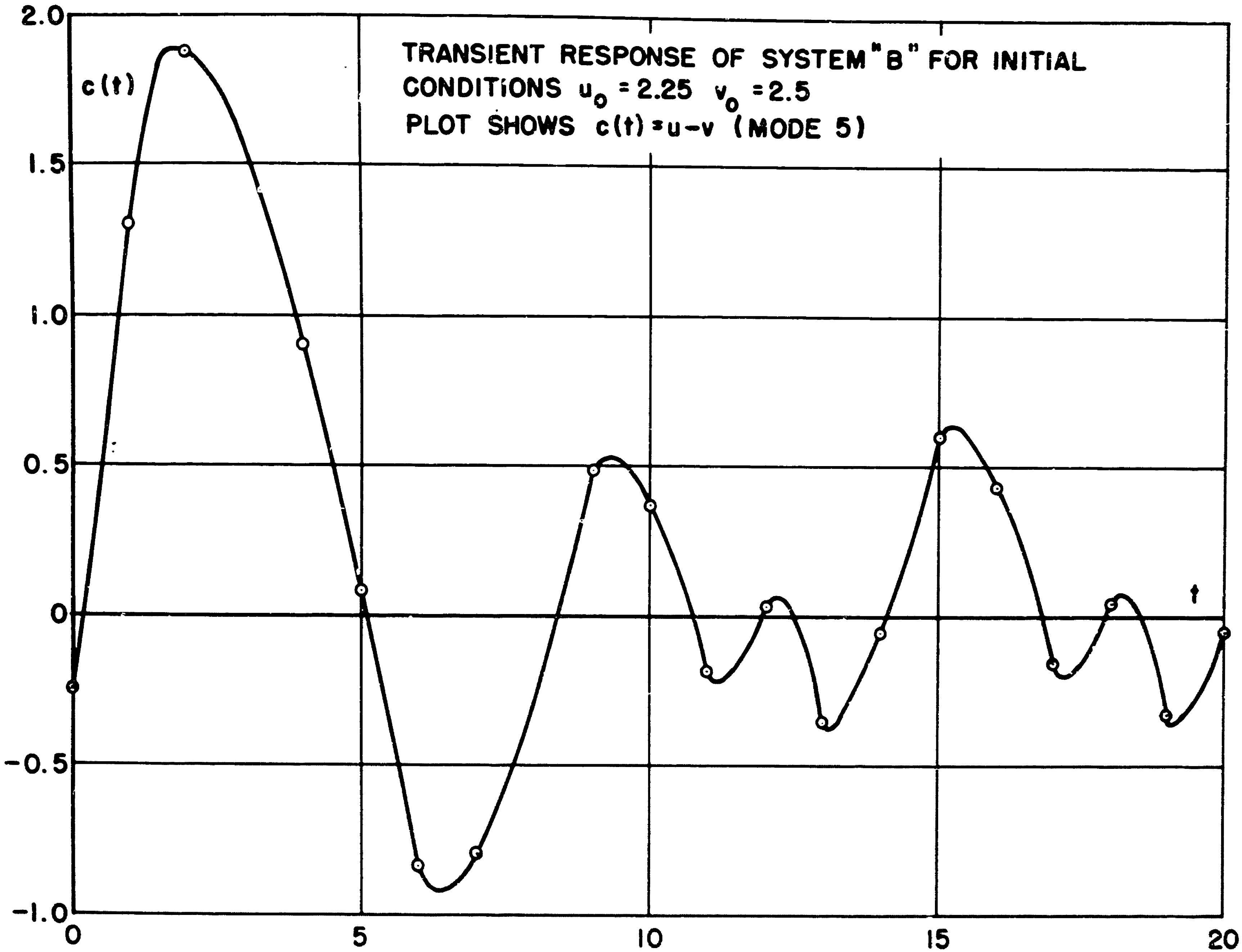

TRANSIENT RESPONSE OF SYSTEM "B" FOR INITIAL
CONDITIONS $u_0 = 2.25$ $v_0 = 2.5$
PLOT SHOWS $c(t) = u - v$ (MODE 5)
c(t)
t
2.0
1.5
1.0
0.5
0
-0.5
-1.0
0
5
10
15
20

FIG. 10

to predict further behavior without quantitative means. In fact, the relay output sequence for all three of Figs. 8-10 is identical during the first eleven (!) sampling periods but after that time the situation changes drastically and the transients synchronize to very different types of steady-state behavior (Modes 3, 1, 5). For further illustration, the limit cycle of index +2 corresponding to Mode 5 has been plotted in the phase plane by means of a real-time analog computer; it is reproduced in Fig. 11. It is also instructive to observe that the limit cycle in Fig. 11 may be shifted along the c axis until either P_4 or P_6 reaches the $\dot{c}$ axis, because the shape of the trajectories defined by (6.5) does not depend on the c coordinate. This is merely another way of visualizing the results of the stability analysis presented in Sect. 5.3.

The complexity of transient behavior of such a simple-appearing system as (6.6) is truly remarkable; it is a sure indication that we are dealing with a rather deep-seated problem.

6.3 Generalizations

The above examples have demonstrated the complex nature of the in-the-large stability of nonlinear sampled-data systems. In particular, the regions of initial conditions leading to the same mode of oscillation do not constitute a connected set, in sharp contrast to the situation in continuous nonlinear systems (cf. Fig. 2).

In the absence of a satisfactory mathematical apparatus for the treatment of nonlinear difference equations such as (6.3) or (6.5), we could at best make vague remarks of questionable validity concerning the general picture (which we know from only two examples); this would be pointless. That the situation is certainly not amenable to an intuitive semi-quantitative

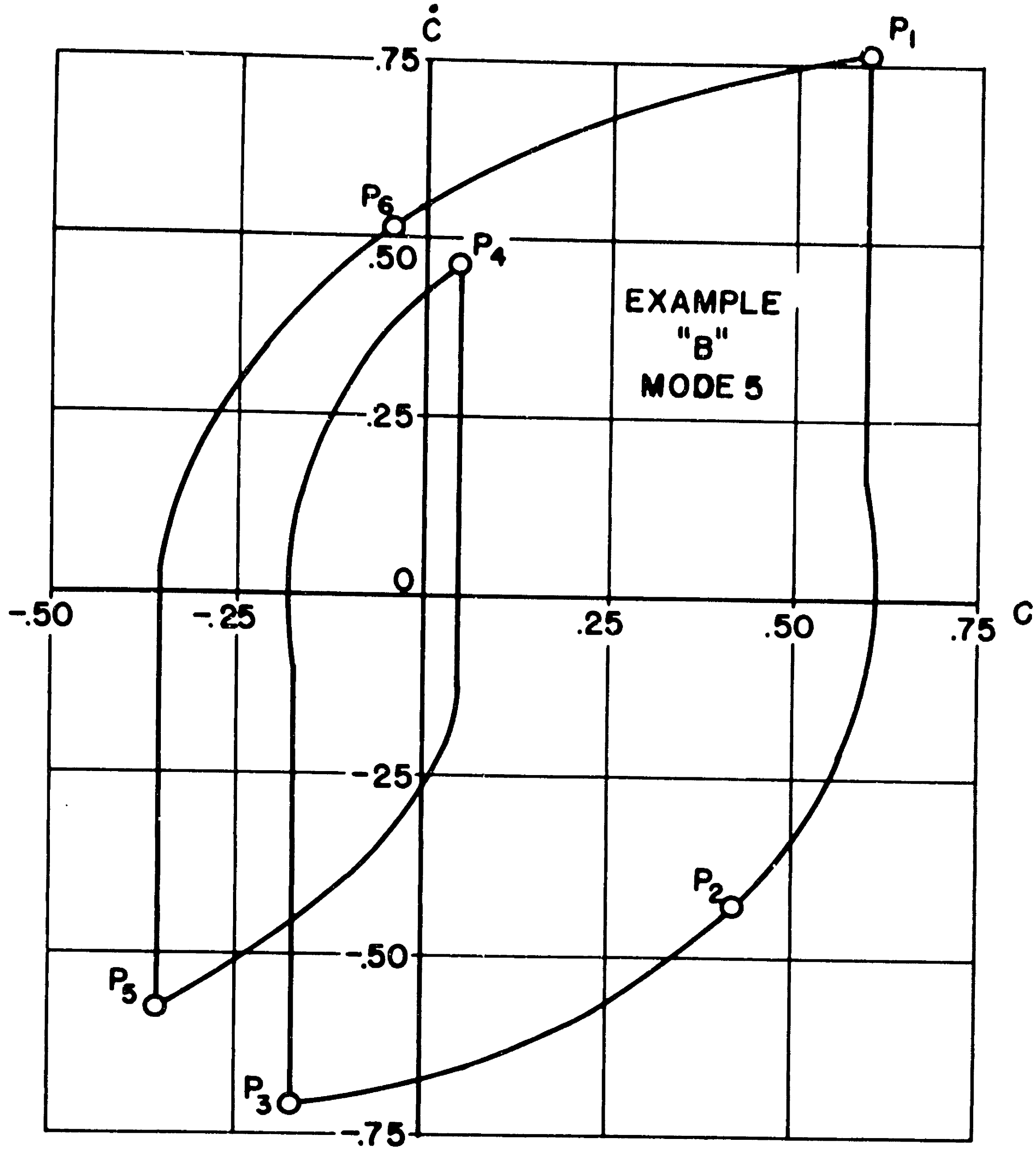

Ċ
P1
.75
P6
.50
P4
EXAMPLE
"B"
MODE 5
.25
0
-.50
-.25
.25
.50
.75
C
-25
P2
-.50
P5
P3
-.75

FIG. 11

understanding at this stage is quite clear from the quantitative and qualitative differences between Figs. 6 and 7, in spite of the fact that the impulse responses of the transfer functions (6.1) and (6.4) are quantitatively fairly similar.

Because of the forbidding complexity of the phase-plane situation even without deadzone, no detailed investigation was attempted of the case when $\Delta \neq 0$ — this is known from the work of Chow and Russell to lead to extremely involved phenomena. We indicate, however, what are believed to be the various topological possibilities for different modes up to period 6T inclusive, to aid in the trial and error computation of limit cycles. Table III contains many more entries than have been considered by Russell and Chow because of the possibility of limit cycles of index +2 or +3. Modes joined by brackets in Table III will always occur in pairs as in Example B because of symmetry in the phase plane; further, it may be shown that if the system contains at least one integration, the modes marked by asterisks do not exist.* It should be added that Table III is based on general (but nonrigorous) topological considerations and therefore is valid regardless of the order of G(s), i.e. we are not necessarily confined to second-order systems. Probably only a few of the modes of Table III can be present in a given system, but there is no concrete information available on this point.

---

* In all of the cases referred to, the trajectories enter the 0 region from the + or - region and then go back to the same region, i.e. there are subsequences of the type + 0 + or - 0 - . If the +, 0, and - regions are separated by straight lines, this behavior is only possible if the system has a center, focus, or saddle point in the 0 region. If there is an integration in the system, then trajectories in the 0 region will be of the form $\dot{x}$ = const, and + 0 + or - 0 - sequences are ruled out. The above arguments carry over to higher-order systems as well.

TABLE III

Period	Mode	Index
2T	+ −	+1
3T	+ − 0 } + 0 − }	+1
4T	+ 0 − 0 + + − −	+1
5T	+ 0 0 − 0 } + 0 − 0 0 } + + 0 − − } + + − − 0 }	+1
	* + 0 + − − } * + + − 0 − }	+2
6T	+ 0 0 − 0 0 + + 0 − − 0 + + + − − −	+1
	+ 0 − + − 0 } + 0 + − + 0 } + + − + − − } + + − − + − } * + + 0 − 0 − } * + 0 + − − 0 } * + 0 + 0 − − } * + + − 0 − 0 }	+2
	* + 0 − − 0 −	+3

7. PREVENTION OF OSCILLATIONS

So far no mention has been made of the effect on the limit cycles of continuously varying the width of the relay deadzone simply because the complexity of problems introduced by it is outside the scope of this thesis. It is apparent from the stability discussion of Sect. 5, however, that a small deadzone will not affect the existence of the periodic solutions already present in a system without deadzone (although it may perhaps create new ones) if the inequalities $v_k \gtrless u_k$ are satisfied with a sufficiently wide margin by all u's and v's corresponding to the periodic solutions. Further, it seems intuitively clear that widening the deadzone will slowly destroy <u>all</u> the modes, but it is strictly a quantitative question as to <u>which</u> mode will disappear first or last.

Both Chow and Russell discuss quantitative conditions on the minimum deadzone required to eliminate <u>all</u> modes. Their treatment is based primarily on geometrical aspects of the describing functions. In what follows, we shall approach the problem from a different view, which, although somewhat heuristic, is conceptually far simpler.

We will make use of a known theorem of Perron[11]. Consider the system of difference equations defined in the phase-space $(x_1, x_2, \ldots x_n)$.

$$x_i(t+1) = \sum_j^n a_{ij} x_j(t) + f_i(t, x_1(t), x_2(t), \ldots x_n(t)) \qquad i = 1, 2, \ldots n \tag{7.1}$$

where $\| a_{ij} \|$ is the matrix of the linear part of the system; the nonlinear functions f_i are essentially small perturbations. The theorem states that the critical point of (7.1) at the origin is <u>stable</u>, if the following conditions are met:

(a) The characteristic roots of the linear system, given by

$$\left| \| a_{ij} \| - \lambda I \right| = 0 \qquad I = \text{unit matrix} \quad (7.2)$$

are all less than unity in absolute value (the well-known stability requirement for the linear system);

(b) Further, the function f_i must be bounded, i.e.

$$|f_i| < K \cdot \|x\| \quad \text{for some} \quad \|x\| < a \qquad (7.3)$$

and vanish at least quadratically at the origin

$$\frac{f_i}{\|x\|} \rightarrow 0 \quad \text{as} \quad \|x\| \rightarrow 0 \quad \text{when} \quad t \rightarrow \infty \qquad (7.4)$$

In the above expressions, $\| x \|$ is the norm in the phase space $(x_1, x_2, \ldots x_n)$ defined by $\| x \| = \sum_i^n |x_i|$.

As a corollary of this theorem, it may be shown also that all solutions of the system (7.1) will tend to the origin if the f_i remain sufficiently small, or when the roots of

$$\left| \| a_{ij} \| + \left\| \frac{\partial f_i}{\partial x_j} \right\| - \lambda I \right| = 0 \qquad (7.5)$$

are also within the unit circle. This condition is clearly more restrictive than (7.2).

Condition (7.5) is capable of a very simple interpretation in engineering terms: If a frequency-independent nonlinear element $f(x)$ in the loop is regarded as giving rise to a variable effective gain (hence the term "gain element"), defined for any value of x by $f(x) = K(x).x$, where K is a real constant depending on x, then it is clear that (7.5) is equivalent to the requirement that the root locus, namely the paths described by the roots of

$$1 + KG^*(z) = 0 \qquad (7.6)$$

in the complex z-plane, do not cross the unit circle from the outside for any permissible value of K. (For stability, the roots of (7.6) must lie outside the unit circle in the complex z-plane. See definition of z, Equation (5.4).)

If we now define the maximum effective gain of a relay by $K = 1/\Delta$, then the question of how much deadzone should be used to get rid of all oscillations is reduced to the routine computational problem of adjusting the loop gain so that the system is just on the verge of instability!*

It is interesting to observe that this reasoning, which is just another way to approximately linearize the system, is quantitatively consistent with the describing-function method: Kochenburger[4] states that the maximum magnitude of the describing function of a relay with deadzone (but without hysteresis) is $4/\pi$ which occurs when the input sinusoid to the relay has the amplitude $\Delta\sqrt{2}$. Thus:

$$K' = \frac{4/\pi}{\Delta\sqrt{2}} = \frac{4}{\pi\sqrt{2}} \cdot \frac{1}{\Delta} \approx \frac{1}{\Delta}$$

By relating the question of deadzone width to the system stability as a function of gain, we have in effect shown that there are two ways of avoiding oscillations: (1) by keeping the effective gain down, namely by using large deadzone; (2) by shaping the root locus (or the open-loop frequency response curve) so as to be able to use a higher value of gain without provoking instability. The design of a nonlinear sampled-data system is thus largely a compromise between these two alternatives, with (2) being preferable from the standpoint of dynamic response, etc.

---

* This argument is not rigorous. Since there is always a discontinuity in the relay, one should actually assume arbitrarily high gain for a very small range of x around the discontinuity, with arbitrarily small gain for the rest of the deadzone. However, the simplifying assumption $K = 1/\Delta$ is probably valid for most practical cases.

On the other side of the picture, it must be pointed out that one or two of the lowest modes of oscillations may be entirely tolerable in practical applications where fast transient response rather than high steady-state accuracy are the determining requirements. The author cannot offer any meaningful analytical approach to the evaluation of this last suggestion; all in all it is regretfully but shamelessly admitted that the design must be carried out by trial and error at this stage in the art.

The pertinent information for the construction of root loci of second-order systems is summarized in Appendix B, which also contains a numerical comparison of the above method with one of the results of Chow concerning the minimum deadzone required for the system of Example A. Agreement is excellent.

It would be unfair to leave the reader without a warning that the application of the usual Nyquist stability criterion to sampled-data systems is rather awkward and laborious, and that therefore the root-locus method, which depends only on the location of poles and zeros but not on where they "came from", is especially useful. (See reference (7).)

8. SUMMARY AND CONCLUSIONS

After the delicate and often tortuous argumentation which has preceded, certain points of the general picture regarding nonlinear sampled-data systems emerge in relative simplicity:

(1) The phase-plane representation may be satisfactorily extended to sampled-data system, utilizing the difference-equation technique. However, to obtain real insight into the quantitative aspects of the problem,

one has to deal with the nonlinear difference equation directly, and this is a very difficult and substantially unsolved problem.

(2) The existence of numerous periodic solutions, already "discovered" by the describing-function method, is clearly demonstrated; in addition, more complicated limit cycles have been shown to exist by Altar and Helstrom as well as the author. Limit cycles of index higher than +1 arise as a direct consequence of the three-dimensionality of the "causal" description of the sampled-data problem; by no stretch of the imagination can they be considered as sinusoidal.

(3) The stability of the limit cycles is very easily investigated via the difference equations. A new phenomenon arises here: the limit cycles may be locally stable along one of the coordinates of the normal (diagonalized) system, but may behave analogously to a center (which is neither stable nor unstable) along the other. For this reason, the oscillations need not have zero average value; in fact, the average value may be influenced by perturbations.

(4) Means of investigating the transient response are available through iterative operations performed on the difference equations. While the phenomena are extremely complicated quantitatively, some light has been thrown on the phase-plane structure as far as the regions of in-the-large stability of the limit cycles are concerned. By contrast, the question as to how the phase-plane structure is affected by changes in the open-loop parameter has by no means been answered, although the complexity of the situation has been revealed by Examples A and B.

(5) The prevention of oscillations may be rather simply treated by considering the root-locus in the complex z-plane; no detailed study has been made since this is a standard problem of linear theory.

In addition, the author would like to reaffirm his convictions concerning methodology of attack on nonlinear systems which has been already stated in the introduction and served as a guiding principle during the entire investigation:

The problem treated here is a very difficult and novel one. It concerns not only a nonlinear dynamic system, but a time-variant nonlinear system — an area where our theoretical (and even practical) knowledge at present is decidedly in a prehistoric stage. We need first of all know how things actually are. If a solution by exact means appears feasible, it should certainly be attempted; only when sufficiently extensive precise and reliable knowledge concerning nonlinear phenomena has become available, can there be a hope for substantial further progress.

The reader is asked to view the present contribution in that light.

9. REFERENCES

(1a) C. K. Chow, "Contactor Servomechanisms Employing Sampled Data", Ph. D. Thesis, School of Electrical Engineering, Cornell University, June 1953, 319 pp.

(1b) Same author and title, Trans. AIEE, 73 II (1954) 51-64.

(2) F. A. Russell, "Design Criterion for Stability of Sampled-Data On-Off Servomechanisms", Ph. D. Thesis, Columbia University, June 1953.

(3) W. Altar and C. W. Helstrom, "Phase-Plane Representation of Sampling Servomechanisms", Westinghouse Research Report R-94410-14-B, Sept. 1953.

(4) R. J. Kochenburger, "A Frequency-Response Method for Analyzing and Synthesizing Contactor Servomechanisms", Trans. AIEE, 69 (1950) 270-284.

(5) J. Loeb, "Recent Advances in Nonlinear Servo Theory", ASME Paper 53-A-19.

(6) N. Minorsky, INTRODUCTION TO NONLINEAR MECHANICS (book), Edwards Bros., Ann Arbor, Mich., 1947.

(7) R. E. Kalman, "A Critical Survey of Analysis and Design Methods for Sampled-Data Control Systems", Notes for MIT Summer Course 6.604 (revised March, 1954).

(8) J. R. Ragazzini and L. A. Zadeh, "The Analysis of Sampled-Data Systems", Trans. AIEE, 71 II (1952) 225-234.

(9) R. H. Barker, "The Pulse Transfer Function and Its Application to Sampling Servo Systems", Proc. IEE, 99 IV (1952), Monograph 43.

(10) R. E. Kalman, "Phase-Plane Analysis of Automatic Control Systems Containing Nonlinear Gain Elements", to be published in Trans. AIEE, 1954.

(11) O. Perron, "Über die Stabilität und asymptotisches Verhalten des Lösungen eines Systems endlicher Differenzengleichungen", J. für reine u. angew. Math., 161 (1929) 41-64 (Theorem 5).

(12) Y. C. M. Yeh, "The Study of Transients in Linear Feedback Systems by Conformal Mapping and the Generalized Root Locus Method", Sc. D. Thesis, Mechanical Engineering Department, Massachusetts Institute of Technology, 1952.

APPENDIX A

DERIVATION OF DESCRIBING FUNCTION FOR IMPULSE MODULATOR-CLAMPER-IDEAL RELAY

The following derivation is adapted from Chow[1a]. We wish to determine the describing function for the impulse-modulator, clamper and relay combination, for the case of an ideal relay ($\Delta = 0$), i.e., one described by the functional relationship:

$$\begin{aligned} f(x) &= 1 \quad \text{if} \quad x > 0 \\ f(x) &= -1 \quad \text{if} \quad x < 0 \end{aligned} \qquad \text{(A.1)}$$

Assume that the input to the combined element defined above is a sinusoid given by

$$A \sin(\Omega t/n + \phi) = \text{Im}\left[A\, e^{j\phi}\, e^{j\Omega t/n}\right] \qquad \text{(A.2)}$$

where n is a positive even integer (cf. the discussion on p. 5 and p. 18 relative to this point) and ϕ the phase-shift between the sinusoid and the impulse modulator. It is clear that the output of the idealized relay will be a square wave of period $nT = 2n\pi/\Omega$, and amplitude $4/\pi$. Further, it may be seen by reference to Fig. A.1 that the phase of the incoming sinusoid (A.2) may be shifted relative to the phase of the impulse modulator by an amount

$$0 < \phi < 2\pi/n$$

without in any way affecting the output of the relay. (This phase shift is the source of all evil.) The describing function may now be written as

$$-\frac{1}{N(n, A, \phi;\ \Delta = 0)} = -\frac{A\pi}{4} e^{j\phi} \qquad 0 < \phi < 2\pi/n \quad \text{(A.3)}$$

The necessary (but not sufficient, cf. p. 7) condition for the existence of stable steady-state oscillations is

$$-1/N = G(j\Omega/n) \qquad \text{(A.4)}$$

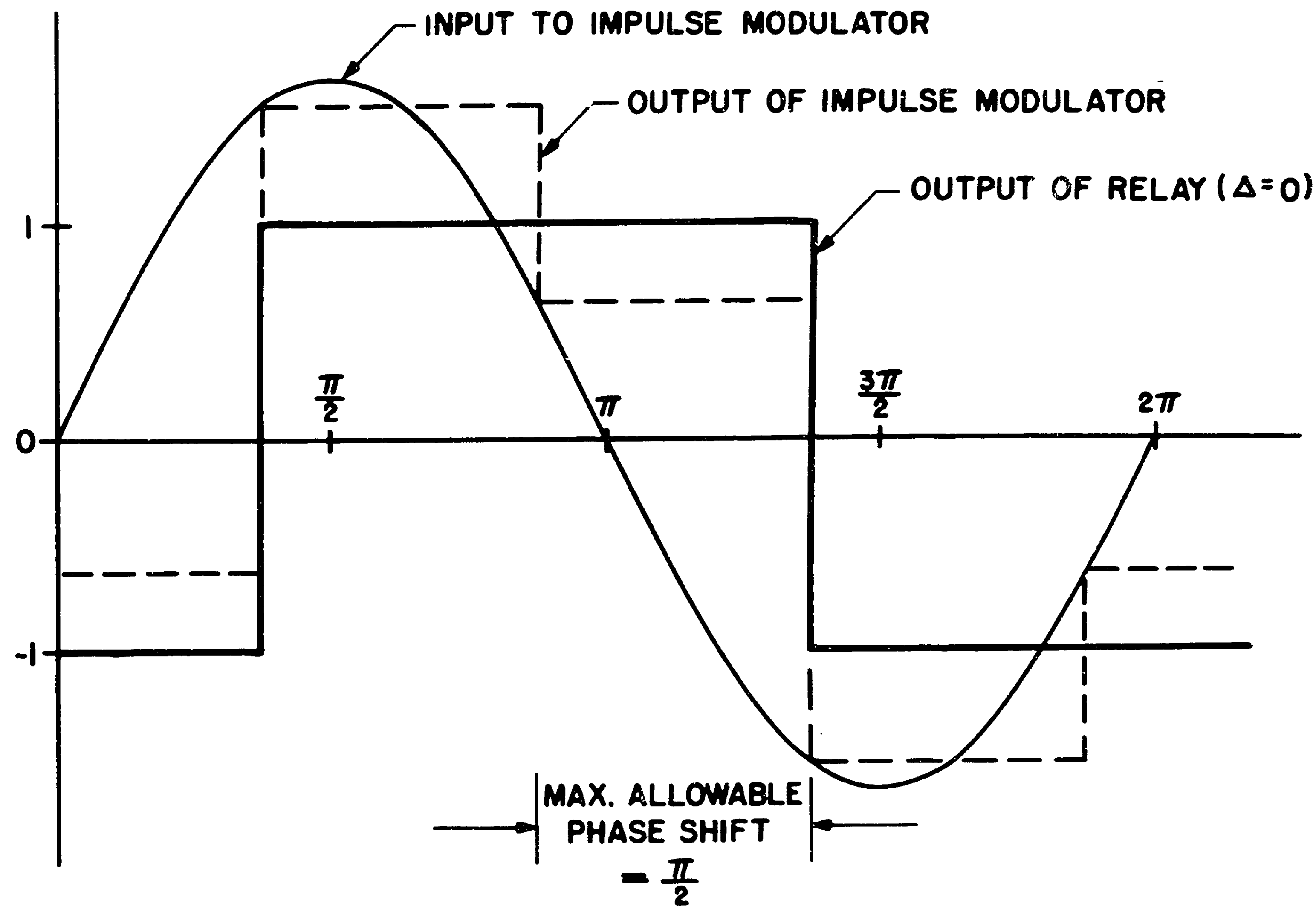

INPUT TO IMPULSE MODULATOR
OUTPUT OF IMPULSE MODULATOR
OUTPUT OF RELAY (Δ=0)
1
0
-1
π/2
π
3π/2
2π
MAX. ALLOWABLE
PHASE SHIFT
= π/2

FIG. A1

which is satisfied whenever $\arg\left[G(j\Omega/n)\right]$ for some n lies within the range of ϕ specified by (A.3). It should be noted that (A.4) is actually purely a <u>phase condition</u>, because the corresponding amplitude constraint

$$\frac{4}{A\pi}\left|G(s)\right| = 1$$

can always be satisfied by some real value of A (the approximate amplitude of the resulting steady-state oscillation).

APPENDIX B

ROOT LOCUS

B.1 Properties of the Root Locus in Second-Order Systems

In examining the stability of a second-order linear sampled-data system, it is useful to have quantitative information about the geometry of the root locus. For a single-loop system, the closed-loop poles given by the roots of the equation

$$1 + KG^*(z) = 0 \qquad 0 < K < \infty$$

lie either on straight lines or on a circle in the z-plane. Using the letter p for poles and r for zeros of $G^*(z)$, the various possibilities for the location of closed-loop poles are as follows:

(1) Poles and zeros alternate on the real axis: real axis

(2) Two real poles, no zero: real axis, plus line $(p_1 + p_2)/2 = \text{const.}$

(2a) Two complex poles: line $(p_1 + p_2)/2 = \text{const.}$

(3) Two poles, one zero: real axis, plus circle centered at r_1, with radius

$$R = \sqrt{(p_1 - r_1)(p_2 - r_1)}$$

(4) Two poles, two zeros: real axis, plus circle centered at $\dfrac{p_1p_2 - r_1r_2}{p_1 + p_2 - r_1 - r_2}$ with radius

$$R = \frac{\sqrt{(p_1 - r_1)(p_2 - r_1)(p_1 - r_2)(p_2 - r_2)}}{p_1 + p_2 - r_1 - r_2}$$

The above results have been adapted from Yeh[12].

B.2 Quantitative Check on the Method of Sect. 7

In discussing the sampled-data system of Fig. 1 with

$$G(s) = 1/s(s + 1),$$

Chow states(1a) that the computer result for the minimum nondimensionalized relay deadzone which eliminates all oscillations is

$$\Delta \approx 0.5 \ .$$

For reasons pointed out in Sect. 7, this corresponds to an effective maximum gain $K = 2$. Let us compute the closed-loop poles with this value of K. The open-loop pulse transfer function is

$$G^*(z) = \frac{z}{1 - z} - \frac{0.632}{1 - 0.368z} = \frac{z(0.264z + 0.368)}{(1 - z)(1 - 0.368z)},$$

so that

$$1 + KG^*(z) = 0.896z^2 - 0.632z + 1 = 0 \ ,$$

and the roots are located at $z = 0.352 \pm j\,0.989$. Therefore $|z| = 1.05$, i.e. the closed-loop poles are barely outside the unit circle. For convenience, the root locus in this case is depicted in Fig. B.1; it obviously belongs to category (4) above.

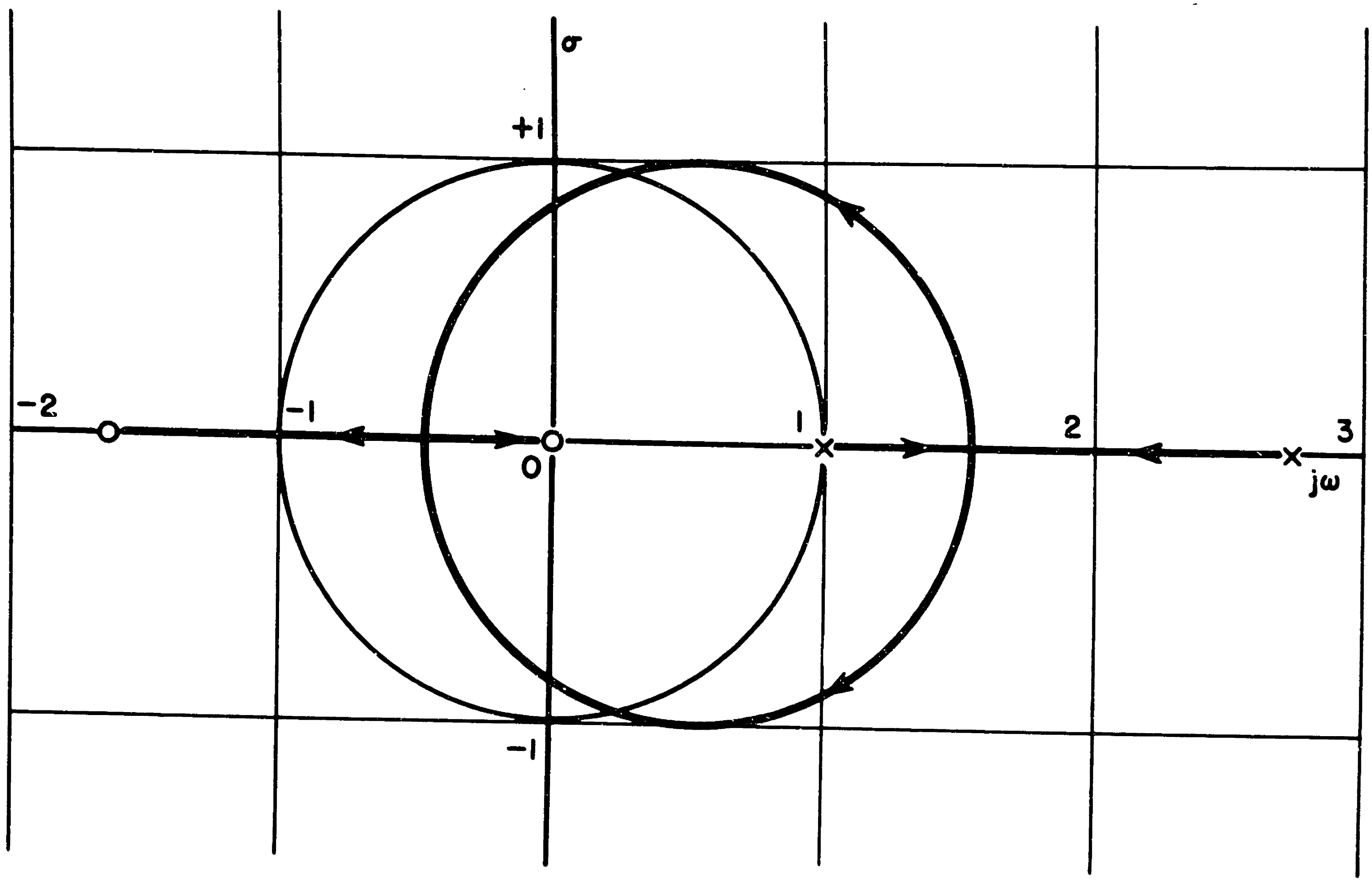


FIG. B.1

APPENDIX C

LAPLACE TRANSFORMS, PULSE TRANSFORMS AND THEIR TIME-DOMAIN EQUIVALENTS

$f(t)$	f_k	$F(s)$	$F^*(z)$
(1) Impulse $\delta(t)$	$f_k = 0 \quad k \neq 0$; $= 1 \quad k = 0$	1	1
(2) Step 1	1	$\frac{1}{s}$	$\frac{1}{1 - z}$
(3) Ramp t	k	$\frac{1}{s^2}$	$\frac{Tz}{(1 - z)^2}$
(4) Pole e^{-at}	$z_a^k = e^{-aTk}$	$\frac{1}{s + a}$	$\frac{1}{1 - z_a z}$
(5) $1 - e^{-at}$	$1 - z_a^k$	$\frac{a}{s(s + a)}$	$\frac{(1 - z_a)z}{(1 - z_a z)(1 - z)}$
(6) $\sin \omega t$	$\sin \omega kT$	$\frac{\omega}{s^2 + \omega^2}$	$\frac{z \sin \omega T}{1 - 2z \cos \omega T + z^2}$
(7) $\cos \omega t$	$\cos \omega kT$	$\frac{s}{s^2 + \omega^2}$	$\frac{1 - z \cos \omega T}{1 - 2z \cos \omega T + z^2}$
(8) $e^{-\alpha t} \sin \omega t$	$z_a^k \sin \omega kT$	$\frac{\omega}{(s + \alpha)^2 + \omega^2}$	$\frac{z_\alpha z \sin \omega T}{1 - 2z_\alpha z \cos \omega T + z_\alpha^2 z^2}$
(9) $e^{-\alpha t} \cos \omega t$	$z_\alpha^k \sin \omega kT$	$\frac{s + \alpha}{(s + \alpha)^2 + \omega^2}$	$\frac{1 - z_\alpha z \cos \omega T}{1 - 2z_\alpha z \cos \omega T + z_\alpha^2 z^2}$

T = sampling period

$z = e^{-sT}$

APPENDIX D

REMARKS RELATIVE TO FIGS. 6 AND 7

Discovered too late to be included in the main text, the following qualitative argument offers good insight into the problem of why the regions corresponding to Mode 1 are entirely different in Figs. 6 and 7.

Consider Fig. D.1 which shows a limit cycle of period 2 (Mode 1). It is known from stability considerations that the system will remain in this mode for small perturbations. In other words, there must be small regions of local stability around P_1 and P_2. Is it possible, however, to find a trajectory starting far away from the origin which will converge to this mode? There are two topological possibilities to consider: the assumed trajectory reaching P_1 from the outside may have resulted from a + or − relay output at the preceding sampling instant. (For simplicity, we assume that $G(s)$ is of the form $(s + a)/s(s + b)$ with a, b positive or zero):

(a) Previous Relay Output −. In this case the trajectory is indicated by the dotted line in Fig. D.1. If the preceding sampled point, P_o, removed by a distance T (measured in time along the trajectory) from P_1 lies in the left-hand plane, then the assumed trajectory clearly cannot exist, because at P_o the relay output must be +.

If, however, the ratio of sampling period to the system time constants is decreased, then P_o may move into the position P_o' (we assume that Mode 1 cannot be destroyed by this, in accordance with the discussion of Sect. 4). At P_o' the relay output is − and therefore the assumed trajectory exists. But then it must also be possible to find a trajectory which preceded P_o'. Continuing this process of describing the trajectory backwards in time it is easy to show using continuity and symmetry considerations that this trajectory will grow indefinitely as t

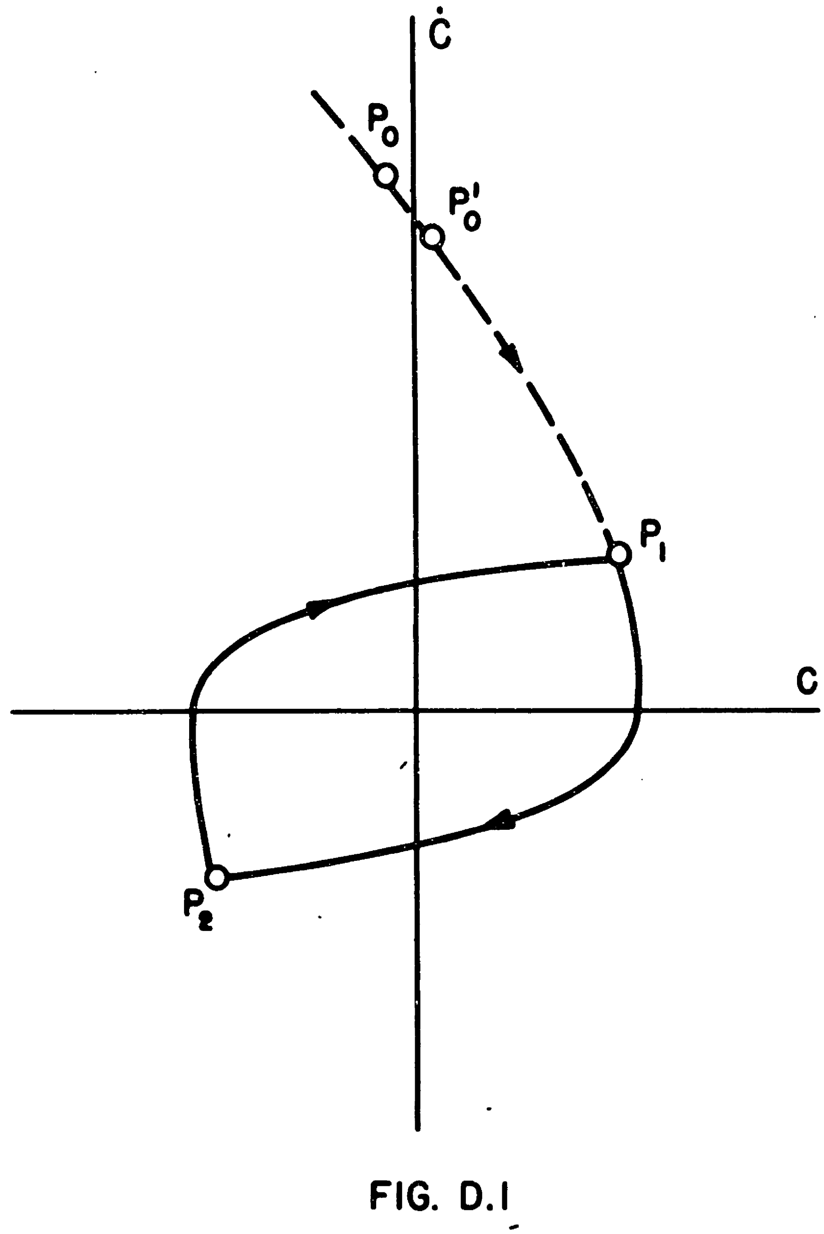


FIG. D.1

Thus if the sampled point preceding P_1 is P_0, then there will be no points in the phase plane (aside from small regions of local stability around P_1 and P_2) from which the transient can synchronize to Mode 1. If the preceding sampling instant is P_0', then, by sharp contrast, there exists no finite closed contour of index +1 in the phase plane outside of which the transient will <u>not</u> synchronize to Mode 1.

(b) <u>Previous Relay Output +</u>. In this case, the trajectory is an arc of the limit cycle, leading from P_1 to the previous sampling point, which must be P_2. By symmetry, the situation at P_2 is the same as at P_1, which we have already explored in (a) above. Hence this possibility adds nothing new.

These conclusions are entirely in accordance with Fig. 6 and 7. In Fig. 6, there are trajectories which may reach Modes 2 and 3 not originating in regions of local stability while this is not the case for Mode 1. The fact that the situation is different for modes of higher period, is apparent from Fig. 4c. In Fig. 7, where the system time constant is slightly longer, regions of initial conditions leading to Mode 1 are scattered all over the finite portion of the phase plane which has been studied. In the case of both figures, regions of synchronization for the higher-period modes are always scattered over the phase plane.

A weak point of our preceding and otherwise rigorous considerations is that we have assumed that the trajectory synchronizing to Mode 1 ends at P_1 or P_2. This assumption is too restrictive and clearly not necessary since instead of P_1 we may have considered any point within the region of local stability. Hence we may restate the preceding in rigorous mathematical form as follows:

Theorem. <u>There exist regions in the phase plane, other than those of local stability, from which the transients converge to a given mode, if and only if the inverse of the transformation (5.5) exists for at least one point within the regions of local stability (but not necessarily a vertex point* on the limit cycle) of the mode in question.</u>

Corollary. <u>The regions referred to in the theorem are not confined within any finite simple Jordan curve in the phase plane.</u>

Our theorem is thus a strong sufficiency condition relating to the structure of in-the-large stability in the phase plane. Unfortunately, it relies on the knowledge of regions of local stability. The question of whether the foundaries of these regions can be computed in a simpler fashion than the application of the iterative method of Sect. 5.42 has not been explored.

To illustrate the theorem, we write down the inverse transformation (6.3) of the system in Example A as:

$$E^{-1}u_k = u_k - \Gamma_{k-1}$$
$$E^{-1}v_k = \epsilon v_k - \epsilon(1 - \epsilon^{-1})\Gamma_{k-1} \qquad (6.3')$$

A vertex of the limit cycle of Mode 1 is $u_1 = 0.500$ and $v_1 = 0.462$. The preceding relay output for a trajectory reaching this point from the outside must have been −, which implies that $v_0 < u_0$. But from (6.3')

$$v_0 = 0.462\epsilon + \epsilon(1 - \epsilon^{-1}) = 2.97$$
$$u_0 = 0.5 + 1.0 = 1.5$$

Thus $v_0 > u_0$ which violates the assumption. In fact, the discrepancy is quite large. Therefore, no outside trajectory can reach Mode 1.

In the case of Example B, a vertex of Mode 1 is $u_1 = 0.500$ and

---

*A point where the relay output changes sign.

$v_1 = 0.333$; using (6.5), we get

$$u_o = u_1 - \Gamma_1 = 1.5 \qquad \text{assuming } \Gamma_1 = -1$$
$$v_o = 2v_1 - \Gamma_1 = 1.666$$

Hence again $v_o > u_o$ but the discrepancy is quite small. If we now choose a point $u_1 = 0.5$ $v_1 = 0.2$ (which still lies in the region of local stability as may be easily verified by computing the transient with the aid of (6.5)) we get

$$u_o = u_1 - \Gamma_1 = 1.5 \qquad \text{assuming } \Gamma_1 = -1$$
$$v_o = 2v_1 - \Gamma_1 = 1.4$$

hence $u_o > v_o$ and the assumed trajectory <u>exists</u>.

Thus while a trajectory originating outside of the regions of local stability cannot reach the vertex points it may nevertheless ultimately converge to Mode 1.

These considerations are another step forward in helping to explain the nature transient behavior and synchronization.