IFAC

WARSZAWA 1969

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Nonlinear Systems

Fourth Congress of the International Federation of Automatic Control
Warszawa 16–21 June 1969

TECHNICAL SESSION

34

Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Nonlinear Systems

TECHNICAL SESSION No 34

FOURTH CONGRESS OF THE INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL
WARSZAWA 16 – 21 JUNE 1969

Organized by
Naczelna Organizacja Techniczna w Polsce

Contents

Wydawnictwa Czasopism Technicznych NOT
Warszawa, ul. Czackiego 3/5 — Polska

Zakład Poligraficzny WCT NOT. Zam. 31/69.

ON THE CONTROLLABILITY OF NONLINEAR SYSTEMS

H. Tokumaru and N. Adachi ; Kyoto Univ., Kyoto, Japan

1. Introduction

The concept of controllability of linear systems was introduced by R. E. Kalman. It is admitted that the concept plays a fundamental role in the modern control theory. Kalman's discussion is based on the linear algebra, and essentially restricted to linear control systems[1,2].

A few authors studied the controllability of nonlinear systems. E. Roxin studied the controllability of special types of nonlinear systems[3]. He introduced the concept of the reachable zone and discussed the relations between optimal controls and reachable zones[4]. L. Markus studied the local controllability of nonlinear systems, controllability in the neighborhood of the critical point. He also showed that it is possible to apply global stability theories to the controllability analysis[5,6]. The generalization of the concept of controllability of linear systems to nonlinear systems was also tried by H. Hermes[7]. He reduced the problem of controllability to the problem of nonintegrability of some Pfaffian form, and discussed the relation between controllability and singular problems which appear in the theory of optimal control.

In this paper, we discuss the controllability of nonlinear systems with controls appearing linearly, by reducing the controllability of the given system to that of the auxiliary lower dimensional control system. We introduce the concept of quasi-controllability, and show sufficient conditions for them. Sufficient conditions for the controllability can be obtained by connecting the conditions for quasi-controllability and local controllability.

2. Definitions

The motion of the given control system is described by a system of ordinary differential equations,

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_r) \quad (i=1, 2, \ldots, n) \tag{1}$$

or in a vector form

$$\frac{dx}{dt} = f(x,u), \tag{2}$$

where x is a state vector and u is a control vector. The functions

$$f_i(x,u) \quad , \quad \frac{\partial f_i(x,u)}{\partial x_j} \quad (i,\ j = 1,\ 2,\ \ldots,\ n)$$

are defined and continuous on the product space $R^n \times R^r$. In general, the function f(x,u) is nonlinear with respect to both x and u. In the case when the function f(x,u) is linear with respect to control u, the system (2) is called a system with controls appearing linearly and expressed as follows,

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

where G(x) is n x r matrix with elements $g_{ij}(x)$. The functions

$$\frac{\partial f_i(x)}{\partial x_j} \ , \ \frac{\partial g_{ik}(x)}{\partial x_j} \quad (i,\ j=1,2,\ldots,\ n,\ k=1,2,\ldots,\ r)$$

are continuous functions of x.

In this paper we say that a control u(t) is admissible if it is continuous for all t under consideration, with exception of a finite number of t at which u(t) may have discontinuity of the first kind. If a certain admissible control u(t) is given, the equation (2) takes the form

$$\frac{dx}{dt} = f(x,\ u(t)) \tag{4}$$

For any initial condition $x(t_0)=x^0$, the solution of the equation (4) is uniquely determined. This solution x(t) will be called the solution of the system (2) corresponding to the control u=u(t) for the initial condition $x(t_0)=x^0$. If the solution of the system (2) corresponding to the control u(t) for the initial condition $x(t_0)=x^0$ satisties the condition $x(t_1)=x^1$ at time $t=t_1$, then we say that the admissible control u(t) transfers the initial state x^0 to the final state x^1. Since the systems under consideration is time-invariant we can set always t = 0.

We define several concepts with respect to the given system.
<u>Definition 1.</u> For the two given state x^0 and x^1, if there

exists a finite time $t_1 > 0$ and an admissible control which transfers the initial state x^o given at the time $t=0$, to the state x^1 at time $t=t_1$, we say that the state x^o is "controllable" to x^1.

Definition 2. The state x^o is said to be "quasi-controllable" to x_1, if in every neighborhood of x^1 there is a state to which x^o is controllable.

Definition 3. If the properties mentioned in Definition 1 and Definition 2 hold for all $x^o \in R^n$, the system is said to be "controllable" to x^1 or "quasi-controllable" to x^1 respectively.

Definition 4. If there exists a neighborhood U of the origin of R^n and every $x \in U$ is controllable to the origin, the system is said to be locally controllable.

Remark 1. In all of the above definitions if x^1 is the origin we say only "controllable" or "quasi-controllable" for simplicity.

Remark 2. A sufficient condition for local controllability is obtained by L. Markus[6]. From these definitions, if the given system is quasi-controllable and locally controllable, then the system is controllable (to the origin).

3. Quasi-Controllability of Nonlinear Systems with Controls Appearing Linearly

In this section we discuss the quasi-controllability of control systems with controls appearing linearly. Such system is described by the equation (3)

$$\dot{x} = f(x) + G(x)u \qquad (3)$$

where f(x) and G(x) have properties mentioned in the preceeding section. Moreover, we assume that the column vectors $g_1(x), \ldots, g_r(x)$ of the matrix G(x) is linearly independent for all $x \in R^n$. Define the matrix D(x) as

$$D(x) \equiv (g_{ij}(x)) \quad (i, j=1,2,\ldots, r),$$

then we assume, for simplicity, that D(x) is nonsingular for all $x \in R^n$.

Now, we state a simple necessary condition for controllability of the system (3).

Theorem 1.

If the system (3) is controllable (to the origin), a system of linear partial differential equations

$$\begin{cases} \frac{\partial P}{\partial x_1} f_1(x) + \dots, + \frac{\partial P}{\partial x_n} f_n(x) = 0 \\ \frac{\partial P}{\partial x_1} g_{1j}(x) + \dots, + \frac{\partial P}{\partial x_n} g_{nj}(x) = 0 \end{cases} \tag{5}$$

has no solutions which are independent at the origin.

Proof. Assume that the equation (5) has m ($m \leq n-(r+1)$) solutions $P_1, \dots, P_m$, which are independent each other. Then the transformation

$$y_i = P_i(x) \quad (i=1,\dots, m)$$
$$= x_i \quad (i=m+1,\dots, n)$$

is nonsingular at the origin. By this transformation the equation (3) becomes

$$\dot{y} = F(y,u)$$

with an appropriate function $F(y,u)$. Here, by definition of y_i $\dot{y}_i=0$ ($i=1,2,\dots, m$). Q.E.D.

We show a lemma which is essentially due to E. Roxin[3].

Lemma 1.

If a state x^0 is controllable to x^1 with respect to the system

$$\dot{x} = G(x)u, \tag{6}$$

then the state is quasi-controllable to x^1 with respect to the original system (3).

We shall now transform the equation (3) into a simple form. Corresponding to the matrix $G(x)$ consider the following system of linear partial differential equations

$$\frac{\partial P}{\partial x_1} g_{1j}(x) + \frac{\partial P}{\partial x_2} g_{2j}(x) + \frac{\partial P}{\partial x_n} g_{nj}(x) = 0 \tag{7}$$

$$(j=1,2,\dots, r).$$

In general, the number of the independent solutions of this equation is less than or equal to $(n-r)$. Here, we regard the system (7) as a complete system, so that the equation has $(n-r)$ independent solutions. Let the solutions be $P_{r+1}(x), \dots P_n(x)$.

Now, a transformation from x to (y,z)

$$\begin{cases} y_i = x_i & (i=1,2,\dots, r) \\ z_i = P_i(x) & (i=r+1,\dots, n) \end{cases} \tag{8}$$

is defined and assumed to give a one-to-one correspondence on the whole space. The equation (3) transformed by (8) is expressed as follows with suitable functions $\varphi(y,z)$ and $\psi(y,z)$.

$$\begin{cases} \dot{y} = \varphi(y,z) + H(y,z)u \\ \dot{z} = \psi(y,z), \end{cases} \tag{9}$$

where H(y,z) is an r x r matrix which is nonsingular by the assumption on G(x).

Applying Lemma 1 to the rewritten system (9) we have the following theorem.

Theorem 2.

Let $(y^o,z^o) \in R^n$ be a given initial state, then (y^o,z^o) is quasi-controllable to the state (y^1,z^o) where $y^1 \in R^r$ is an arbitrary fixed point in R^r.

Proof. Corresponding to the system, consider a control system

$$\begin{cases} \dot{y} = H(y,z)u \\ \dot{z} = 0 \end{cases}$$

Since the matrix H(y,z) is nonsingular, a initial state (y^o,z^o) is controllable to (y^1,z^o). Then, the assertion of the theorem holds by Lemma 1. Q.E.D.

Corresponding to the system (9) we define a (n-r)- dimensional control system

$$\dot{z} = \psi(v,z) \tag{10}$$

where $z \in R^{n-r}$ is an (n-r)- dimensional state vector, and $v \in R^r$ is an r-dimensional control vector. Between controllability of the system (3) and that of the system (10) there exist some relations.

Theorem 3.

If the given system (3) is controllable, then the system (10) is controllable. Conversely, if the system (10) is quasi-controllable with continuously differentiable controls, then the original system (3) is quasi-controllable.

Proof. Assuming that the original system (3) is controllable, we shall show that the system (10) is controllable. Since the system (3) and the system (9) is equivalent, the system (9) is controllable. Hence, there exists an admissible control $u^o(t)$ which transfers a given initial state (y^o, z^o) to the origin in a finite time. Let $(y(t;u^o), z(t;u^o))$ be the solution of the equation (9) corresponding to $u^o(t)$ with the initial state (y^o, z^o). In the system (10) we take the function $v^o(t) = y(t;u^o)$ as a control function. Then $v^o(t)$ transfers the initial state z^o of the system (10) to the origin of R^{n-r}. Since z^o is arbitrarily given, we conclude that the system is controllable. Now, we assume that the system (10) is quasi-controllable with continuously differentiable controls, and shall show that the system (3) becomes quasi-controllable. From the assumption there exists a continuously differentiable control $v^o(t)$ which transfers the initial state $z^o \in R^{n-r}$ to a given neighborhood of the origin at some finite time $t=t_1$. In the control system (9) determine the control law by

$$u^o(t) = H^{-1}(v^o(t), z(t;v^o))\left\{ \frac{dv^o}{dt} - \varphi(v^o(t), z(t,v^o)) \right\},$$

then this control law transfers the initial state $\tilde{x}^o = (v^o(0), z^o)$ to the $\tilde{x}^1 = (v^o(t_1), z(t_1;v^o))$. By Theorem 2 it will be shown that the given initial state $x^o = (y^o, z^o)$ is quasi-controllable to x^o, and x^1 is also quasi-controllable to the state $(0, z(t_1;v^o))$. Since the solutions of the differential equations continuously depend on initial conditions, it is easily proved that (y^o, z^o) is quasi-controllable to the origin. Q.E.D.

<u>Remark</u> : Under suitable conditions on the equation (9) we can prove that a state is quasi-controllable with a continuously differentiable control if the state is quasi-controllable with an admissble control. So, in that case, we may assert that quasi-controllability of the system (10) is a necessary and sufficient condition for quasi-controllability of the system (3).

4. Quasi-Controllability of some Special Types of Nonlinear Systems.

In this section we shall apply the general theory in the preceeding section to some special types of nonlinear systems.

(A) Linear Systems

In the case of a linear time-invariant system, a transformed system corresponding to the expression (9) is expressed as follows.

$$\begin{cases} \dot{y} = F_{11}y + F_{12}z + E_r u \\ \dot{z} = F_{21}y + F_{22}z, \end{cases} \tag{11}$$

where $F_{ij}(i,j=1,2)$ are constant matrices with compatible dimensions, and E_r is an r x r dimentional unit matrix. Consider the following (n-r)- dimensional control system ;

$$\dot{z} = F_{22}z + F_{21}v \, , \tag{12}$$

where z is an (n-r)- dimensional state vector, and v is an r-dimensional control vector.

Theorem 4.

The linear time-invariant system (11) is controllable if and only if the system (12) is controllable.

Proof. Define n x nr matrix M and (n-r) x (n-r)r matrix N as follows,

$$M \equiv (H, FH, \ldots, F^{n-1}H)$$
$$N \equiv (F_{21}, F_{22}F_{21}, \ldots, F_{22}^{n-r-1}F_{21})$$

where

$$F \equiv \begin{bmatrix} F_{11}, & F_{12} \\ F_{21}, & F_{22} \end{bmatrix}, \qquad H \equiv \begin{bmatrix} E_r \\ 0 \end{bmatrix}.$$

Then, we shall show that the rank of the matrix M is n if and only if the rank of the matrix N is n-r. With simple calculations, the matrix M is expressed as

$$M = \begin{bmatrix} E_r & P \\ 0 & \overline{N}R \end{bmatrix},$$

where P is a r x (n-1)r matrix and

$$\overline{N} = (F_{21}, F_{22}F_{21}, \ldots, F_{22}^{n-2}F_{21})$$

and R is a (n-1)r x (n-1)r nonsingular matrix. Since rank$N$ = rank$\overline{N}$ and R is nonsingular, rank$M$ = n if and only if rank$N$=n-r. Q.E.D.

(B) Systems with (n-1) controls

The transformed system is described in this case as follows.

$$\begin{cases} \dot{y} = \varphi(y,z) + H(y,z)u \\ \dot{z} = \psi(y,z), \end{cases} \tag{13}$$

where z is one dimensional vector.

Theorem 5.

If there exists a continuously differentiable (n-1)-dimensional vector function v(z) such that

$$z\,\psi(v(z),z) < 0 \quad \text{for all} \quad z \in R^1, \quad z \neq 0$$

then the original system is quasi-controllable.

Proof. With Theorem 3 and simple stability considerations it is clear.

5. Examples

Example 1.

Consider the case when $\psi(y,z)$ in the system (9) is linear. The system equation becomes

$$\begin{cases} \dot{y} = \varphi(y,z) + H(y,z)u \\ \dot{z} = F_1 y + F_2 z \end{cases} \tag{14}$$

Here, $\varphi(0,0)=0$, $H(0,0)=E_r$ are assumed. Then, from theorem 3 and well-known controllability criterion[2] for linear time invariant system, this system is quasi-controllable if

$$\operatorname{rank}\left(F_1,\ F_2F_1, \ldots,\ F_2^{n-r-1}F_1 \right) = n-r. \tag{15}$$

On the other hand, consider the linear time-invariant system

$$\begin{cases} \dot{y} = Ay + Bz + E_r u \\ \dot{z} = F_1 y + F_2 z\ . \end{cases} \tag{16}$$

Here the matrices A and B are defined by

$$A \equiv \frac{\partial \varphi(0,0)}{\partial y}\ , \quad B \equiv \frac{\partial \varphi(0,0)}{\partial z}\ .$$

If the system (16) is controllable, then the critical point of the system (14) is locally controllable[6]. From Theorem 4 the linear system (16) is controllable if and only if the condition (15) is satisfied. Here, if the condition (15) holds the given system (14) is quasi-controllable and locally controllable, so that the system is controllable.

Example 2.

Consider a higher order system

$$x^{(n)} + a_1 x^{(n-1)} + \dots + a_n x = u\ , \tag{17}$$

where $x^{(i)} = \frac{d^i x}{dt^i}$ and a_i is a function of $x, \dot{x}, \dots, x^{(n-1)}$, $(i=1,2,\dots, n)$.

If we set $x = x_1$, $\dot{x} = x_2, \dots, x^{(n-1)} = x_n$, then the system (17) is equivalent to the system

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \vdots \\ \dot{x}_n = -(a_1(x)x_n + a_2(x)x_{n-1} + \dots + a_n(x)x_1) + u. \end{cases} \tag{18}$$

Since the linear system with control v

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \vdots \\ \dot{x}_{n-1} = v \end{cases}$$

is controllable clearly, the original system (17) is quasi-controllable by Theorem 3. Moreover it is easily verified that the system (18) is locally controllable. Hence, the higher order system (17) is controllable.

Example 3.

$$\begin{cases} \dot{x}_1 = g_1(x_1, x_2, x_3) + u \\ \dot{x}_2 = x_1 + g_2(x_2, x_3) \\ \dot{x}_3 = x_2 + g_3(x_3) \end{cases} \tag{19}$$

where $g_i(0) = 0 \quad (i=1,2,3)$.

This system is quasi-controllable if the system

$$\begin{cases} \dot{x}_2 = g_2(x_2, x_3) + v \\ \dot{x}_3 = x_2 + g_3(x_3) \end{cases} \tag{20}$$

is quasi-controllable with continuously differentiable control. This condition is satisfied since one dimensional system

$$\dot{x}_3 = g_3(x_3) + w$$

is controllable with sufficiently smooth control.

Define matrices A and B as

$$A \equiv \frac{\partial f(0)}{\partial x} \equiv \begin{bmatrix} a_{11}, & a_{12}, & a_{13} \\ 1 & a_{22}, & a_{23} \\ 0 & 1 & a_{33} \end{bmatrix}, \quad B \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

where $a_{ij} = \dfrac{\partial g_i(0.0)}{\partial x_j}$, $(i, j = 1,2,3)$. Since rank of $[B, AB, A^2B] = 3$, the system (19) is locally controllable. Since the system (19) is quasi-controllable and locally controllable, the system is controllable.

6. Conclusion

The concepts of controllability, quasi-controllability, etc. for nonlinear control systems are introduced, and sufficient conditions for them are obtained. A global discussion of controllability for general nonlinear system is very difficult. A known technique for them is an application of stability theory. But systems to which such a method is applicable are restricted. In most cases we cannot discuss directly the controlollability of general nonlinear systems. So, we treated some special types of nonlinear systems ; systems which are nonlinear with respect to x but linear with respect to control u. In such a system, it is possible to reduce the discussion for the original system to that of some corresponding lower dimensional system.

In the case when the origin expresses the stationary state of the controlled object, the concept of local controllability is important. Connecting quasi-controllability and local controllability conditions for controllability are obtained for several types of nonlinear system.

References

1. R. E. Kalman, "On the general theory of control systems," Proc. 1st International Congress on Automatic Control, Butter worths Scientific Publications. London, England, Vol. 1, pp. 481-492;1962

2. R. E. Kalman, Y. C. Ho and K. S. Narendra, " Controllability of linear dynamical systems" Contrib. to Differential Equations, Vol. 1, pp. 189-213;1962

3. E. Roxin and V. Spinadel, " Reachable Zones in Autonomous Differential Systems," Contrib. to Differential Equations, Vol. 1, No. 3.

4. E. Roxin, " A Geometric Interpretation of Pontryagin's Maximum Principle", International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics, Academic Press New York and London,1963

5. L. Markus, E. B. Lee, "On the Existence of Optimal Controls" Transaction of the ASME. Vol. 84, 1962

6. L. Markus "Controllability of Nonlinear Processes", J. SIAM Control Ser. A, Vol. 3, No. 1, Printed in U.S.A., 1965.

7. H. Hermes "Controllability and the Singular Problems", J. SIAM Control Ser. A, Vol. 2, No. 2, Printed in U.S.A., 1965

ANALYSIS OF RELAY SAMPLED-DATA SYSTEMS WITH A NONLINEAR PLANT

H.L. Burmeister

Institut für Regelungs- und Steuerungstechnik der Deutschen Akademie der Wissenschaften zu Berlin

Dresden / GDR

1. Introduction

Sampled-data systems containing a relay as the only nonlinearity have been studied in detail during the past years, major attention given to the determination of self-oscillations sustained in the closed loop.

Besides the describing function method (harmonic balance) [1-9], which gives approximate solutions and basically applies only to oscillations with a dominating harmonic, several analytic methods to exactly determine the oscillations whose period is an integer multiple of the sampling period have been suggested. They are based more or less obviously on the principle of assuming the relay output by trial as a periodic sequence of pulses, calculating the corresponding steady-state response of the linear element and finally checking by means of the relay equation, whether the assumed pulse sequence is sustained in the closed loop.

The methods differ by the description of the pulse sequence (sequence of amplitudes [10, 11], z-transform [9, 12-14], finite Fourier series [8, 15, 16] or finite series of more general orthogonal functions [15]) and the characteristics of the linear part (pulse transfer function [8, 9, 12, 13], transmission matrix [11], difference equation [14-16] or state equations [10]).

Second-order systems have also been studied in the phase plane [19] or a state plane [20], combined with analytic methods. A procedure of determining all the oscillations occurring in the system on account of different initial conditions, by means of a finite algorithm, has not been known till now. In trying to find an upper bound for the period of all oscillations, tacitly simple oscillations were assumed [13, 16, 18].

As already emphasized in [21, 22], the state equations are the natural mathematical tool for treating such systems. This holds all the more for systems with additional nonlinearities, as e.g. extremum control systems, in which the plant has a nonlinear steady-state characteristic with an extremum and the switching function of the relay may be nonlinear as well. Applying transform methods [23, 24] in this case leads to hardly manageable relations, whereas the method of harmonic balance [25] is subject to the limitations mentioned before.

In the following the state equations are taken as a basis for analysing relay sampled-data systems with a nonlinear plant. In this way exact results suitable for an easy computation are successfully derived under rather general assumptions on the plant structure, the pulse element and the switching function of the relay.

2. The complete system equations

The block diagram of the system under consideration is shown in Fig. 1. Its elements are subject to the following assumptions:

a) The nonlinear plant can be represented as a series-connection of an r-th order linear element, a static nonlinearity with parabolic characteristics

$$v = -ax^2 \quad (a > 0) \tag{1}$$

and another s-th order linear element. The rational transfer functions F(p) and G(p) of the linear elements have simple, negative poles p_k (k = 1, 2, ..., r) resp. q_l (l = 1, 2, ..., s) and arbitrary zeros. After formally including the regulating unit, a pure integrator, in the first linear element, the following partial fraction expansions hold (with $p_o = 0$):

$$\frac{1}{p} F(p) = \sum_{k=0}^{r} \frac{c_k}{p-p_k} \; ; \qquad G(p) = \sum_{l=1}^{s} \frac{d_l}{p-q_l} \tag{2}$$

b) For i, k = 0, 1, 2, ... , r; l = 1, 2, ... , s:

$$p_i + p_k \neq q_l \tag{3}$$

This condition rules out a kind of resonance between the two linear elements.

c) The pulse element consisting of a sampler and a shaping unit generates the control rate as a sequence of pulses of equal shape and different signs

$$u(t) = \sum_{n=0}^{\infty} u[n]\, h(t-nT) \qquad (4)$$

$$\text{with} \quad u[n] = \pm 1 \quad \text{and} \quad h(t) \begin{cases} \equiv 0 \text{ in } t \leqq 0,\ t \geqq T \\ \geqq 0 \text{ in } 0 \leqq t < T \end{cases} \qquad (5)$$

The transfer function of the shaping unit

$$H(p) = \int_0^T e^{-pt}\, h(t)\, dt \qquad (6)$$

is a regular analytic function in the finite p-plane.
As special cases are included herein:

$H(p) = \dfrac{1 - e^{-p\tau}}{p}$ with $0 < \tau \leqq T$: constant control rate,

$H(p) = e^{-p\tau}$ with $0 < \tau < T$: pure delay up to a sampling period.

d) The relay switching function $\Psi[n]$ is a single-valued, continuous function of the sampled values $y(nT) = y[n]$ and $y[n-1]$ of the plant output:

$$\Psi[n] = \Psi(y[n],\ y[n-1]) \ , \qquad (7)$$

controlling, according to the switching condition

$$u[n+1] = u[n] \operatorname{sgn} \Psi(y[n+1],\ y[n]) \ , \qquad (8)$$

the sign changes of the control rate. x)
By introducing canonical state variables (normal coordinates) [20, 26, 27], the differential equations of the linear elements corresponding to the transfer functions (2) are transformed into systems of uncoupled first order equations

$$\dot{x}_k(t) = p_k\, x_k(t) + c_k\, u(t) \qquad (k = 0,1,2,\ldots,r) \qquad (9)$$

$$\dot{y}_l(t) = q_l\, y_l(t) + d_l\, v(t) \qquad (l = 1,2,\ \ldots\ ,s) \qquad (10)$$

$$x(t) = \sum_{k=0}^{r} x_k(t) \ ; \qquad y(t) = \sum_{l=1}^{s} y_l(t) \qquad (11)$$

---

x) Unlike the usual definition, the following one is assumed for convenience:

$$\operatorname{sgn} x = \begin{cases} 1 & \text{for } x > 0 \\ -1 & \text{"} \quad x \leqq 0 \end{cases}$$

Eqs. (1), (4), (8) - (11) describe the relay sampled-data system for all t.

3. The difference equations of the system

The values of the system variables between the sampling instants do not influence the relay switchings and consequently not the system performance. Therefore one is led to describe the essential features of the system behaviour by difference equations for the sampled variables, while expressing the intermediate values, if necessary, by the preceding sampled values.

For this purpose, Eqs. (9) and (10) are integrated with initial values $x_k[n]$ and $y_l[n]$; this gives in $nT \leqq t \leqq (n+1)T$:

$$x_k(t) = x_k[n]\, e^{p_k(t-nT)} + u[n]\, \gamma_k(t-nT) \qquad (12)$$

$$x(t) = \sum_{k=0}^{r} x_k[n]\, e^{p_k(t-nT)} + u[n]\, \gamma(t-nT) \qquad (13)$$

with $\gamma_k(t) = c_k \int_0^t e^{p_k(t-\tau)} h(\tau)\, d\tau \; ; \; \gamma(t) = \sum_{k=0}^{r} \gamma_k(t)$ (14)

and $y_l(t) = y_l[n]\, e^{q_l(t-nT)} - a\, d_l \int_{nT}^{t} e^{q_l(t-\tau)} x^2(\tau)\, d\tau$ (15)

After inserting (13) in (15), the integration can be carried out explicitly as in [28] for relay systems without sampling. In this way the state variables between the sampling instants are expressed by their discrete values and known functions, defined in $0 \leqq t \leqq T$.

Putting $t = (n+1)T$ in (12) and (15) and using the notations

$$\alpha_k = e^{p_k T} \text{ with } \alpha_o = 1 \text{ and } 0 < \alpha_k < 1 \quad (k = 1,2,\ldots,r) \qquad (16)$$

$$\beta_l = e^{q_l T} \text{ with } 0 < \beta_l < 1 \quad (l = 1,2,\ldots,s) \qquad (17)$$

$$\gamma_k = \gamma_k(T) = c_k \alpha_k H(p_k) \qquad (18)$$

$$\omega_{lik} = -a\, d_l \beta_l \int_0^T e^{(p_i + p_k - q_l)t}\, dt = -a\, d_l \frac{\alpha_i \alpha_k - \beta_l}{p_i + p_k - q_l} \qquad (19)$$

$$\varrho_{lk} = -2\,a\,d_l\beta_l \int_0^T \gamma(t)\,e^{(p_k-q_l)t}\,dt = \tag{20}$$

$$= -2\,a\,d_l\left[\alpha_k \sum_{i=0}^{r} \frac{c_i\alpha_i\,H(p_i)}{p_i + p_k - q_l} + \beta_l\,\frac{H(q_l-p_k)\,F(q_l-p_k)}{q_l - p_k}\right]$$

$$\sigma_l = -\,a\,d_l\,\beta_l \int_0^T \gamma^2(t)\,e^{-q_l t}\,dt\,, \tag{21}$$

the following system of linear and quadratic difference equations for the discrete values of the state variables is obtained:

$$x_k[n+1] = \alpha_k\,x_k[n] + \gamma_k\,u[n] \tag{22}$$

$$y_l[n+1] = \beta_l\,y_l[n] + \sum_{i=0}^{r}\sum_{k=0}^{r} \omega_{lik}\,x_i[n]\,x_k[n] + \sum_{k=0}^{r}\varrho_{lk}\,x_k[n]u[n]+\sigma_l$$

In contrast with [29], where multidimensional z-transforms are used for setting up the difference equations of an LNL-chain, the method presented here assumes that there is no additional sampling between the linear elements. It applies just as well to nonlinearities with a polynomial characteristics.

4. <u>Linearization of the quadratic difference equations</u>

Introducing new state variables $z_l[n]$ instead of $y_l[n]$ by the substitution

$$y_l[n] = z_l[n] + \sum_{i=0}^{r}\sum_{k=0}^{r} c_{lik}\,x_i[n]\,x_k[n] + \delta_l \quad (l = 1,2,\ldots,s) \tag{23}$$

and inserting in (22), the quadratic terms and the absolute term are eliminated, if

$$c_{lik} = -\frac{a d_l}{p_i + p_k - q_l} \tag{24}$$

and

$$\delta_l = \frac{1}{1-\beta_l}\left(\sigma_l - \sum_{i=0}^{r}\sum_{k=0}^{r} c_{lik}\,\gamma_i\,\gamma_k\right) \tag{25}$$

can be chosen. For this it is necessary and sufficient that conditions (3) be fulfilled.

Define

$\underline{x}[n]$, $\underline{z}[n]$ and $\underline{\gamma}$ - the column vectors with components

$$x_k[n]\, u[n],\ z_l[n] \text{ and } \gamma_k ,$$

A and B - the diagonal matrices with elements α_k and β_l,

Ω - the (s, r + 1) matrix with elements

$$\omega_{lk} = \varrho_{lk} - 2\alpha_k \sum_{i=0}^{r} c_{lik}\, \gamma_i = -\, 2a\, d_l \beta_l \frac{H(q_l-p_k)\, F(q_l-p_k)}{q_l - p_k} \qquad (26)$$

and $$u^*[n] = u[n]\, u[n+1] \ . \qquad (27)$$

Then the linearized equations read:

$$\begin{aligned} \underline{x}[n+1] &= (A\, \underline{x}[n] + \underline{\gamma})\, u^*[n] \\ \underline{z}[n+1] &= B\, \underline{z}[n] + \Omega \underline{x}[n] \end{aligned} \qquad (n = 0,1,2, \ldots) \qquad (28)$$

It is worth noting that this system of equations can be written down immediately knowing the transfer functions and their poles without any necessity of setting up and transforming by (23) the equations (22).

The linear equations (28) are joined by the nonlinear switching condition (8) as the condition of closing the loop, which by (11), (22), (23) and (27) becomes

$$u^*[n] = \operatorname{sgn} \Psi^*(\underline{x}[n], \underline{z}[n]) \qquad (29)$$

Eqs. (28) and (29) give the desired simplified description of the system performance by restricting to the sampled state variables. They represent a system of recurrence formulae, which, given the initial values $\underline{x}[0]$, $\underline{z}[0]$ resp. $x_k[0]$, $y_l[0]$, $u[0]$, render possible an easy computation of transients, just as well forming the basis for determining the steady-state oscillations.

5. Steady-state oscillations

The method mentioned in the introduction gives the most natural and general approach for calculating "commensurable" oscillations. It consists in

a) determining, given a periodic sequence $u[n]$ $(n = 0,1,2,\ldots)$ with

$$u[n] \equiv u[n+N] \quad (NT \ - \ \text{period of the oscillation}), \qquad (30)$$

the initial values $\underline{x}[0]$, $\underline{z}[0]$ causing a transient-free, periodic motion in the open-loop chain between u[n] and y[n],

b) calculating by (28) the complete oscillation $\underline{x}[n]$, $\underline{z}[n]$ and

c) checking, whether the values of $u^*[n]$ calculated from the switching condition (29) correspond to the presupposed values of u[n].

Together with u[n], $u^*[n]$ is given by (27) as a periodic sequence. After applying (28) N times the following conditions for the initial values result from the periodicity conditions $\underline{x}[N] = \underline{x}[0]$, $\underline{z}[N] = \underline{z}[0]$:

$$\underline{x}[0] = A^N\underline{x}[0] + u[0]\sum_{i=0}^{N-1} u[N-1-i]\, A^i\underline{\gamma} \qquad (31)$$

$$\underline{z}[0] = B^N\underline{z}[0] + \sum_{j=0}^{N-1} B^{N-1-j}\,\Omega\,\underline{x}[j]\,. \qquad (32)$$

Because of (17), Eq. (32) can be uniquely solved for $\underline{z}[0]$:

$$\underline{z}[0] = (I_s - B^N)^{-1}\sum_{j=0}^{N-1} B^j\,\Omega\,\underline{x}[N-1-j]\,. \qquad (33)$$

The matrix $I_{r+1} - A^N$, however, is singular as $\alpha_o = 1$, hence (31) must be solved by components:

$$x_k[0] = \alpha_k^N x_k[0] + \sum_{i=0}^{N-1} u[N-1-i]\alpha_k^i\gamma_k \quad (k=0,1,2,\ldots,r) \qquad (34)$$

These equations have a unique solution for k = 1,2,...,r. They are solvable for k = 0 if and only if

$$\sum_{i=0}^{N-1} u[i] = 0\;; \qquad (35)$$

in this case $x_o[0]$ remains undetermined. From (35) can be concluded that N = 2M is an even integer.
Hence (31) has a one-parameter family of solutions

$$\underline{x}[0] = \underline{x}^s[0] + \varepsilon u[0]\,\underline{e}_o \quad (-\infty < \varepsilon < \infty) \qquad (36)$$

with $\underline{e}_o^T = (1,0,0,\ldots,0)$, the particular solution $\underline{x}^s[0]$ obtained by passing to the limit $\alpha_o \longrightarrow 1$:

$$\underline{x}^s[0] = u[0]\lim_{\alpha_o\to 1}(I_{r+1}-A^N)^{-1}\sum_{i=0}^{N-1} u[N-1-i]A^i\underline{\gamma} \qquad (37)$$

Further follows from (28) and (36)

$$\underline{x}[n] = u[n]\left[\lim_{\alpha_0 \to 1} (I_{r+1}-A^N)^{-1} \sum_{i=0}^{N-1} u[n-i-1]A^i \underline{\gamma} + \varepsilon \underline{e}_0\right] \qquad (38)$$

and by combining with (33):

$$\underline{z}[0] = \lim_{\alpha_0 \to 1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} u[N-1-j]\ u[2N-2-i-j]\ \Omega^{(ij)} \underline{\gamma} + \varepsilon (I_s - B^N)^{-1} \sum_{j=0}^{N-1} u[N-1-j]\ B^j \Omega \underline{e}_0 \qquad (39)$$

and finally

$$\underline{z}[n] = \lim_{\alpha_0 \to 1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} u[n-j-1]\ u[n-i-j-2]\ \Omega^{(ij)} \underline{\gamma} + \varepsilon (I_s - B^N)^{-1} \sum_{j=0}^{N-1} u[n-j-1]\ B^j \Omega \underline{e}_0 \quad (n = 0,1,2,..) \qquad (40)$$

where the (s, r+1) matrix

$$\Omega^{(ij)} = (I_s - B^N)^{-1}\ B^j \Omega A^i\ (I_{r+1} - A^N)^{-1} \qquad (41)$$

$$(i,j = 0,1,2,\ \dots\ , N-1)$$

has the elements

$$\omega_{lk}^{(ij)} = \frac{\alpha_k^i\ \beta_l^j}{(1-\alpha_k^N)(1-\beta_l^N)}\ \omega_{lk} \qquad \begin{pmatrix} k = 0,1,2,\dots,r \\ l = 1,2,\ \dots,\ s \end{pmatrix} \qquad (42)$$

Eqs. (38) and (40) give the most general expression for any oscillation sustained in the open-loop chain by a periodic excitation.

In spite of the double sum their numerical evaluation is not too laborious, since all its coefficients are equal to ± 1, and the matrix elements (42) can be easily computed too. An alternative way consists in determining $\underline{x}[0]$ and $\underline{z}[0]$ from (34), (39) and then the complete oscillation by recurrence from (28).

Finally it must be checked by inserting in (29) whether the oscillation continues to exist in the closed-loop system.

In special cases Eqs. (38) and (40) simplify.

a) Let $\qquad u[n+M] = -u[n] \quad (2M = N) \qquad (43)$

for all n. Then

$$\underline{x}[n] = u[n]\left[(I_{r+1}+A^M)^{-1} \sum_{i=0}^{M-1} u[n-i-1]\ A^i \underline{\gamma} + \varepsilon \underline{e}_0\right] \qquad (44)$$

$$\underline{z}[n] = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} u[n-j-1]\, u[n-i-j-2]\, \tilde{\Omega}^{(ij)} \underline{\gamma} + \\ + \varepsilon (I_s+B^M)^{-1} \sum_{j=0}^{M-1} u[n-j-1]\, B^j \Omega\, \underline{e}_o \tag{45}$$

with
$$\tilde{\Omega}^{(ij)} = (I_s-B^M)^{-1} B^j \Omega A^i (I_{r+1}+A^M)^{-1}$$

$$\tilde{\omega}_{lk}^{(ij)} = \frac{\alpha_k^i \beta_l^j}{(1+\alpha_k^M)(1-\beta_l^M)} \omega_{lk} \tag{46}$$

For $\varepsilon = 0$ the oscillation is symmetrical:

$$x_k[n+M] = -x_k[n] \; ; \quad z_l[n+M] = z_l[n]$$

b) Let the oscillation be simple:

$$u[n] = \begin{cases} 1 & \text{for} \quad n = 0, 1, 2, \ldots, M-1, \\ -1 & \text{"} \quad n = M, M+1, \ldots, 2M-1 \end{cases} \tag{47}$$

Carrying out the double summation, it is found that

$$x_o[0] = \varepsilon - \frac{M}{2}\gamma_o \; ; \quad x_k[0] = -\varphi(\alpha_k)\gamma_k \quad (k = 1,2,\ldots,r)$$
$$z_l[0] = \frac{1+\beta_l^M}{1-\beta_l^M} \sum_{k=0}^{r} \frac{\varphi(\alpha_k)-\varphi(\beta_l)}{\alpha_k - \beta_l} \omega_{lk}\gamma_k - \varepsilon\varphi(\beta_l)\omega_{lo} \quad (l = 1,2,\ldots,s) \tag{48}$$

with
$$\varphi(\alpha) = \frac{1-\alpha^M}{(1-\alpha)(1+\alpha^M)} \; ; \quad \varphi(1) = \lim_{\alpha \to 1} \varphi(\alpha) = \frac{M}{2} \tag{49}$$

6. Stability of steady-state oscillations

Using canonical state variables renders it possible to solve the stability problem in an almost trivial way. The following stability assertion holds:

The steady-state oscillation $\underline{x}[n]$, $\underline{z}[n]$ is stable in the sense of Lyapunov if for all n

$$\Psi^*(\underline{x}[n], \underline{z}[n]) \neq 0 \; ; \tag{50}$$

it is unstable i.s.L., if for at least one $n = n_o$

$$\Psi^*(\underline{x}[n_o], \underline{z}[n_o]) = 0 \quad \text{and} \quad \underline{x}[n_o+1] \neq 0 \, . \tag{51}$$

The stability proof is based on an extension of ideas in [20, 27].

In the (r+s+1)-dimensional ($\underline{x}$, $\underline{z}$) state space of the discrete system the points ($\underline{x}[n]$, $\underline{z}[n]$) ($n = 0,1,2,\ldots,N-1$) constitute the discrete "trajectory" of the oscillation. Around each of these phase points a neighbourhood

$$K[n] : \begin{cases} |x_k - x_k[n]| \leq \lambda_k \\ |z_l - z_l[n]| \leq \frac{1}{1-\beta_l} \sum_{k=0}^{r} |\omega_{lk}| \lambda_k \end{cases} \qquad (52)$$

is defined; the union of these K[n] is a neighbourhood U of the "trajectory". If (50) holds, the constants $\lambda_k > 0$ can be chosen so small that for all points in K[n]

$$\operatorname{sgn} \Psi^*(\underline{x}, \underline{z}) = \operatorname{sgn} \Psi^*(\underline{x}[n], \underline{z}[n])$$

If ($\underline{x}'$, $\underline{z}'$) denotes the image of ($\underline{x}$, $\underline{z}$) when mapping the state space into itself by (28), the following estimation holds on account of (28) and (52):

$$|x_k' - x_k[n+1]| = \alpha_k |x_k - x_k[n]| \leq |x_k - x_k[n]| \leq \lambda_k$$

$$|z_l' - z_l[n+1]| \leq \beta_l \; |z_l - z_l[n]| + \sum_{k=0}^{r} |\omega_{lk}| \, |x_k - x_k[n]| \leq$$

$$\leq \frac{\beta_l}{1-\beta_l} \sum_{k=0}^{r} |\omega_{lk}| \lambda_k + \sum_{k=0}^{r} |\omega_{lk}| \lambda_k = \frac{1}{1-\beta_l} \sum_{k=0}^{r} |\omega_{lk}| \lambda_k \; ;$$

this means, that K[n] is mapped by (28) into K[n+1]. Each "trajectory" originating in U does not leave U. By this the stability is proved since U can be made arbitrarily small by reducing the λ_k's.

If, however, (51) holds, any arbitrarily small neighbourhood of ($\underline{x}[n_o]$, $\underline{z}[n_o]$) will contain points ($\underline{x}$, $\underline{z}$) with $\Psi^*(\underline{x}, \underline{z}) > 0$, for which

$$|x_k' - x_k[n_o+1]| = |\alpha_k(x_k - x_k[n_o]) - 2x_k[n_o+1]| \geq 2|x_k[n_o+1]| - \alpha_k \lambda_k \, .$$

For at least one k the right hand side does not tend to zero for $\lambda_k \to 0$; this proves the instability.

If for all n_o, for which $\Psi^*(\underline{x}[n_o], \underline{z}[n_o]) = 0$, at the same time holds $\underline{x}[n_o+1] = 0$, the oscillation is stable indeed, but it becomes unstable when slightly varying the system parameters (structural instability).

7. Mean value of a steady-state oscillation

The mean value of the plant output, in extremum control systems the so-called hunting loss, defined by

$$M(y) = \frac{1}{N}\sum_{n=0}^{N-1} y[n] , \tag{53}$$

can be expressed by the system parameters without calculating the oscillation itself. The appropriate tool is the Fourier expansion of all periodic sequences [8, 15, 16]:

$$u[n] = \sum_{\varrho=0}^{N-1} \vartheta_\varrho \, \varepsilon_\varrho^{\,n} \quad \text{with} \quad \vartheta_\varrho = \frac{1}{N}\sum_{n=0}^{N-1} u[n]\, \varepsilon_n^{-\varrho} , \quad \varepsilon_\varrho = e^{\frac{2\pi j \varrho}{N}} \tag{54}$$

$$x_k[n] = \sum_{\varrho=0}^{N-1} \xi_{k\varrho}\, \varepsilon_\varrho^{\,n} \; ; \quad y_l[n] = \sum_{\varrho=0}^{N-1} \eta_{l\varrho}\, \varepsilon_\varrho^{\,n} \; ; \quad z_l[n] = \sum_{\varrho=0}^{N-1} \zeta_{l\varrho}\, \varepsilon_\varrho^{\,n} \tag{55}$$

From (35) follows $\vartheta_o = 0$.

Combining (55) with (28) and equating the coefficients leads to

$$\begin{aligned} &\xi_{k\varrho} = \frac{\vartheta_\varrho}{\varepsilon_\varrho - \alpha_k}\,\gamma_k \quad (\varrho = 1, 2, \ldots , N-1) \\ &\xi_{ko} = 0 \quad (k = 1, 2, \ldots , r) \; ; \quad \xi_{oo} = \varepsilon \text{ (arbitrary)} \\ &M(z_l[n]) = \zeta_{lo} = \frac{1}{1-\beta_l}\sum_{\varrho=1}^{N-1} |\vartheta_\varrho|^2 \sum_{k=0}^{r} \frac{\omega_{lk}\,\gamma_k}{\varepsilon_\varrho - \alpha_k} \end{aligned} \tag{56}$$

From (55) and (56) is obtained

$$M\Big(\sum_{i=0}^{r}\sum_{k=0}^{r} c_{lik}\, x_i[n]\, x_k[n]\Big) = c_{loo}\,\varepsilon^2 + \sum_{\varrho=1}^{N-1} |\vartheta_\varrho|^2 \sum_{i=0}^{r}\sum_{k=0}^{r} \frac{c_{lik}\,\gamma_i\,\gamma_k}{(\varepsilon_\varrho - \alpha_i)(\varepsilon_{N-\varrho} - \alpha_k)} \tag{57}$$

With regard to

$$\sum_{l=1}^{s} c_{lik} = -\,a \sum_{l=1}^{s} \frac{d_l}{p_i + p_k - q_l} = -\,a\, G(p_i + p_k)$$

using the notations

$$\omega_k = \sum_{l=1}^{s} \frac{\omega_{lk}}{1-\beta_l} \; ; \quad \delta = \sum_{l=1}^{s} \delta_l$$

$$v_\varrho(\alpha) = \frac{\cos\frac{2\pi\varrho}{N} - \alpha}{1-2\alpha\cos\frac{2\pi\varrho}{N}+\alpha^2} \;; \quad w_\varrho(\alpha) = \frac{\sin\frac{2\pi\varrho}{N}}{1-2\alpha\cos\frac{2\pi\varrho}{N}+\alpha^2}$$

the final solution reads:

$$M(y) = -a\varepsilon^2 G(0) - a\sum_{\varrho=1}^{N-1}|\vartheta_\varrho|^2 \sum_{i=0}^{r}\sum_{k=0}^{r} G(p_i+p_k)[v_\varrho(\alpha_i)v_\varrho(\alpha_k) + w_\varrho(\alpha_i)w_\varrho(\alpha_k)]\,\gamma_i\gamma_k + \sum_{\varrho=1}^{N-1}|\vartheta_\varrho|^2\sum_{k=0}^{r} v_\varrho(\alpha_k)\omega_k\gamma_k + \delta \tag{58}$$

Special cases:

a) LN-chain; $G(p) \equiv 1$.
By passing to the limit $q_1 \to -\infty$, $d_1 \to \infty$, $-\frac{d_1}{q_1} \to 1$, $\beta_1 \to 0$ or directly from (57) with $c_{1ik} = -a$ it is found that

$$M(y) = -a\,[\varepsilon^2 + \sum_{\varrho=1}^{N-1}|\vartheta_\varrho F^*(\varepsilon_\varrho)|^2]\;, \tag{59}$$

where

$$F^*(z) = \sum_{k=0}^{r}\frac{\gamma_k z}{z-\alpha_k} = \mathcal{Z}\left\{\frac{1}{p}F(p)\,H(p)\right\}$$

is the pulse transfer function of the linear chain. Eq. (59) can as well easily be obtained by z-transforms.

b) NL-chain; $F(p) \equiv 1$.
In this case Eq. (58) becomes with $r = 0$

$$M(y) = -aG(0)\,[\varepsilon^2 + \frac{1}{4}\gamma_o^2\sum_{\varrho=0}^{N-1}\frac{|\vartheta_\varrho|^2}{\sin^2\frac{\pi\varrho}{N}}] - \frac{1}{2}\omega_o\gamma_o + \delta \tag{60}$$

8. Applications

The method outlined above was applied to several types of extremum control systems using the switching condition (8)

$$u[n+1] = u[n]\ \mathrm{sgn}\ (y[n+1] - y[n]) \tag{61}$$

or slight modifications, which describe one of the simplest extremum controllers. By specializing the general formulae, smooth and well manageable results concerning steady-state oscillations and their existence regions as well as transient responses were found in the following cases:

a) if δ-impulses, i.e. $H(p) = 1$, are assumed;

b) if simple oscillations, in particular of least non-trivial period 4 T, are considered;

c) for systems with L_rN-plant [27, 30] and NL_s-plant;

d) for second- and third-order systems with L_1N-, L_2N-, L_1NL_1-, NL_1- and NL_2-plants respectively [27, 30];

e) for second-order systems with rectangular pulses and pure delay.

With certain modifications the method can be extended to systems with the input or output of the nonlinearity drifting with constant rate, a case important in practice.

References

1. Chow, C.K.; Appl. and Ind. 73 (1954/55), 51-64
2. Kazakov, V.P.; Avtom. i Telem. 18 (1957), 34-46
3. Simkin, M.M.; Dokl. AN SSSR 131 (1960), 1323-1326
4. Simkin, M.M.; Dokl. AN SSSR 149 (1963), 586-588
5. Kuo, B.C.; Proc. IRE 48 (1960), 941-942
6. Kuo, B.C.; IRE Intern. Conv. Rec. 10 (1962), pt. 2, 99-111
7. Kuo, B.C.; IEEE Trans. AC - 11 (1966), 290-292
8. Tsypkin, Ya.Z.; Int. Symp. Nonlin. Vibr., vol. 2, 433-439; Kiev 1963
9. Kondo, B. and Iwai, S.; Proc. 2nd Congr. IFAC, Basle 1963; vol. Theory, 154-164
10. Nelson, W.L.; J. Bas. Engg. 83 (1961), 65-76
11. Boudarel, R., Sënouillet, G. and Guichet, P.; Automatisme 8 (1963), 247-254
12. Bergen, A.R.; Discussion of 15.
13. Jury, E.J. and Nishimura, T.; J. Bas. Engg. 84 (1962), 71-84
14. Pai, M.A.; Appl. and Ind. 81 (1962/63), 350-355
15. Torng, H.C. and Meserve, W.E.; IRE Trans. AC - 5 (1960), 298-305
16. Torng, H.C.; Appl. and Ind. 81 (1962/63), 355-363
17. Mullin, F.J. and Jury, E.I.; Appl. and Ind. 77 (1958/59), 517-524

18. Izawa, K. and Weaver, L.E.; Appl. and Ind. 78 (1959/60), 49-54

19. Morozov, A.A.; Avtom. i Telem. 1967, No. 7, 64-74

20. Kalman, R.E.; Proc. Symp. Nonl. Circ. Anal., Brooklyn (1956), 276-313

21. Gibson, J.E. and Schultz, D.G.; Discussion of 13.

22. Kalman, R.E.; Discussion of 9.

23. Popkov, Yu.S.: Avtom. i Telem. 24 (1963), 472-481, 1487-1500

24. Benua, E.Yu.; Sb. rabot po voprosam elektromekhaniki, vyp. 5 (1961), 84-94

25. Mańczak, K.; Archivum Autom. i Telemech. 7 (1962) 71-88

26. Lurje, A.I.; Einige nichtlineare Probleme aus der Theorie der selbsttätigen Regelung; Berlin 1957

27. Burmeister, H.L.; messen, steuern, regeln 9 (1966) 258-268

28. Dolgolenko, Yu.V.; Teoriya i primenenie diskretnykh avtomaticheskikh sistem; Moskva 1960, p. 399-412

29. Alper, P.; IEEE Trans. AC-10 (1965), 322-327

30. Burmeister, H.L.; Theory of self-adaptive control system, Proc. 2nd Symp. IFAC (Teddington); New York 1966, p. 353-360

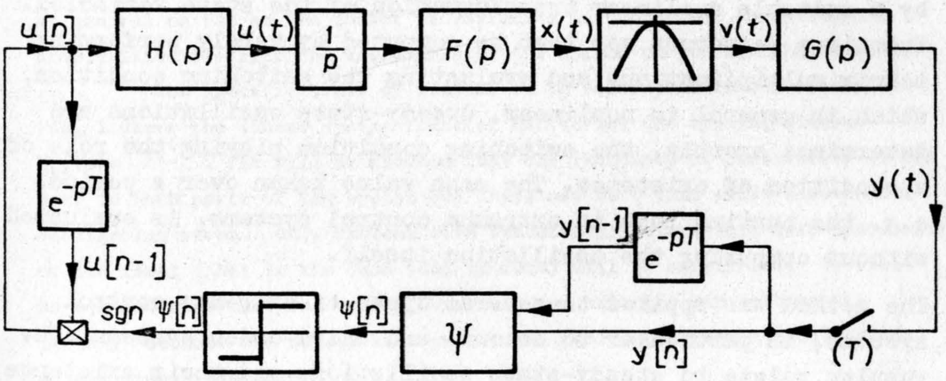

Fig. 1

Abstract

Analysis of relay sampled-data systems with a nonlinear plant

H.L. Burmeister, Dresden / GDR

An exact method of analysing a class of relay sampled-data systems with additional nonlinearities, occurring e.g. in the field of extremum control of plants with parabolic characteristics, is suggested. It is simpler and applicable under more general conditions than the existing methods and proves to be practicable for numerical computation.

The method applies to plants that can be represented as $L_r N L_s$-chains, consisting of stable linear elements L_r and L_s of any order and a parabolic static nonlinearity N. The pulse shape is arbitrary; rectangular pulses and pure delay are included as special cases.

The sampled state variables (normal coordinates) satisfy a system of nonlinear difference equations, which is linearized by a suitable nonlinear transformation of the state variables. Transient responses may then be computed by merely performing matrix multiplications and evaluating the switching condition, which in general is nonlinear. Steady-state oscillations are determined exactly, the switching condition playing the rôle of a condition of existence. The mean value taken over a period, e.g. the hunting loss in extremum control systems, is evaluated without computing the oscillation itself.

The method was applied to several types of extremum control systems, in particular to second- and third-order systems. The results relate to steady-state oscillations and their existence regions as well as to the boundedness or divergence of transients.

SUBHARMONIC OSCILLATIONS IN COUPLED RELAY CONTROL SYSTEMS

S. T. Nugent
Division of Engineering Physics
Dalhousie University
Halifax, Canada

R. J. Kavanagh
Department of Electrical Engineering
University of New Brunswick
Fredericton, Canada

1. Introduction

In single variable relay control systems, the phenomenon of subharmonic oscillations, when the system is subjected to certain periodic inputs, is well known. A number of investigators including Sakawa[1], Gille and Paquet[2] and Gille, Paquet and Pouliquen[3] have given methods of predicting this phenomenon using the approach of Tsypkin[4] and Hamel.

With the increasing importance of multivariable control systems, and in particular, relay systems, it is necessary to study all aspects of the behaviour of such systems so that designs may be optimized. It is therefore the purpose of this paper to extend the use of Tsypkin's method of analysis to the specific problem of predicting whether subharmonic oscillations may occur in certain multivariable relay control systems. While the approach is general as far as the number of variables is concerned, computational complexities restrict the usefulness of the method to two-variable systems.

The specific class of systems to be considered is that shown in Fig. 1 where the linear system transfer matrix has the typical element $W_{ij}(\omega)\varepsilon^{j\phi_{ij}(\omega)}$. It will be assumed that the frequency of oscillation is the same in both parts of the system but there may be a time shift between the oscillating waves. Only systems with symmetrical relays that have hysteresis or are ideal (that is the dead band is zero) will be considered. It will be assumed also that the relays have only two switches per subharmonic period. Because the relays are symmetrical, only subharmonic oscillations of odd orders can occur.

2. Forced Oscillations in Relay Systems with Hysteresis but without Dead Band

The conditions for a forced oscillation at frequency ω_f in this system have been given by Nugent and Kavanagh[5]. Let the system inputs be

$$r_1(t) = A_1 f_1(\omega_f t + \beta_1 - \delta_1)$$

$$r_2(t) = A_2 f_2(\omega_f t + \beta_1 - \delta_1 - \sigma)$$

where A_1 and A_2 are the respective maximum values of r_1 (t) and $r_2(t)$, β_1 is the angle by which the output $m_1(t)$ of relay 1 lags the error $e_1(t)$, δ_1 is the angle by which $r_1(t)$ lags $e_1(t)$ and σ is the angle by which $r_2(t)$ lags $r_1(t)$. In order for the system to exhibit forced oscillations, it was shown that the following conditions are necessary

$$\begin{aligned} &\text{Im}\,\{R_1(\pi + \beta_1 - \delta_1) + \Lambda_1(\omega_f)\} = -h_1/2 \\ &\text{Re}\,\{R_1(\pi + \beta_1 - \delta_1) + \Lambda_1(\omega_f)\} < 0 \end{aligned} \tag{1}$$

and

$$\begin{aligned} &\text{Im}\,\{R_2(\pi + \beta_1 - \delta_2) + \Lambda_2(\omega_f)\} = -h_2/2 \\ &\text{Re}\,\{R_2(\pi + \beta_1 - \delta_2) + \Lambda_2(\omega_f)\} < 0 \end{aligned} \tag{2}$$

where for $\omega = \omega_f$ and $t = \pi/\omega_f$

$$\Lambda_1(\omega_f) = -\frac{1}{\omega_f}\frac{d\,c_1(t)}{dt}\bigg|_{t=\pi/\omega_f} - j\,c_1(\pi/\omega_f)$$

$$R_1(\pi + \beta_1 - \delta_1) = A_1[f_1'(\pi + \beta_1 - \delta_1) + j\,f_1(\pi + \beta_1 - \delta_1)]$$

and for $t = 2\pi(1/2-\tau)/\omega_f$

$$\Lambda_2(\omega_f) = -\frac{1}{\omega_f}\frac{d\,c_2(t)}{dt}\bigg|_{t=2\pi(1/2-\tau)/\omega_f} - j\,c_2(2\pi(1/2-\tau)/\omega_f)$$

$$R_2(\pi + \beta_1 - \delta_2) = A_2[f_2'(\pi + \beta_1 - \delta_2) + j\,f_2(\pi + \beta_1 - \delta_2)]$$

where f_i' is the derivative of $f_i(\omega_f t + \beta_1 - \delta_i)$, $i = 1,2$ with respect to $\omega_f t$, $2\pi\tau/\omega_f$ is the phase shift between $m_1(t)$ and $m_2(t)$, and $\delta_2 = \delta_1 + 2\pi\tau + \sigma$.

Also, $c_1(t)$ and $c_2(t)$ are the system outputs which are given by

$$c_1(t) = \frac{4}{\pi}\Big[\sum_{n=1,3,\ldots}\frac{1}{n}\{M_1W_{11}(n\omega)\sin(n\omega t + \phi_{11}(n\omega)) + M_2W_{12}(n\omega)\sin(n\omega t + n2\pi\tau + \phi_{12}(n\omega))\}\Big] \tag{3}$$

$$c_2(t) = \frac{4}{\pi}\Big[\sum_{n=1,3,\ldots}\frac{1}{n}\{M_1W_{21}(n\omega)\sin(n\omega t + \phi_{21}(n\omega)) + M_2W_{22}(n\omega)\sin(n\omega t + n2\pi\tau + \phi_{22}(n\omega))\}\Big] \tag{4}$$

Simultaneous satisfaction of the two sets of conditions (1) and (2) for some specified τ and $\beta_1 - \delta_1$ indicates a possible forced oscillation. In addition, the following conditions on the number of switches of the relays per period must be satisfied:

$e_1(t) = r_1(t) - c_1(t) > -h_1/2 \qquad (0 \leq t < \pi/\omega_f)$

and

$e_2(t) = r_2(t) - c_2(t) > -h_2/2 \qquad (-2\pi\tau/\omega_f \leq t < 2\pi(1/2-\tau)/\omega_f).$

The $R_1(\pi + \beta_1 - \delta_1)$ and $R_2(\pi + \beta_1 - \delta_2)$ loci are closed curves centered at the $\omega = \omega_f$ points on the $\Lambda_1(\omega)$ and $\Lambda_2(\omega)$ loci respectively. They are circles for the particular case of the sinusoidal inputs $r_1(t) = A_1 \sin(\omega_f t + \beta_1 - \delta_1)$ and $r_2(t) = A_2 \sin(\omega_f t + \beta_1 - \delta_1 - \sigma)$.

At frequency ω_f there exists critical values of A_1 and A_2 (A_{1K} and A_{2K}) since these values must be large enough to ensure that the $R_1(\pi + \beta_1 - \delta_1)$ and $R_2(\pi + \beta_1 - \delta_2)$ loci intersect the $-h_1/2$ and $-h_2/2$ lines respectively. If the critical values A_{1K} and A_{2K} are plotted against ω_f for a specific value of τ, the curves will have the general shape shown in Fig. 2. In Fig. 2(a), A^1_{1K} is the critical amplitude for fundamental oscillation and ω_{o1} corresponds to the frequency at which the $\Lambda_1(\omega)$ locus intersects the $-h_1/2$ line when $\tau=\tau_1$. Similarly, in Fig. 2(b), A^1_{2K} is the critical amplitude for fundamental oscillation and ω_{o2} corresponds to the frequency at which the $\Lambda_2(\omega)$ locus intersects the $-h_2/2$ line when $\tau=\tau_1$.

These curves which are conveniently obtained from the Λ loci divide the A_{1K}, ω_f and the A_{2K}, ω_f planes into two zones: the $A_1 > A_{1K}$ and $A_2 > A_{2K}$ zones in which a forced oscillation is possible and the $A_1 < A_{1K}$ and $A_2 < A_{2K}$ zones in which a forced oscillation cannot occur. For the two-variable system being considered, there will be a pair of curves similar to those of Fig. 2 for each value of τ.

3. Conditions for the Existence of Subharmonic Oscillations

The conditions given in Section 2 can be generalized for the study of subharmonic oscillations. A subharmonic oscillation of order μ is a periodic oscillation of the outputs $c_1(t)$ and $c_2(t)$, the frequency of which is an exact submultiple $1/\mu$ of the input frequency ω_f (assuming that both inputs have the same frequency). That is, the period of the subharmonic oscillation is $T_\mu = \mu T_f = 2\mu\pi/\omega_f$. The outputs of both relays are assumed to have the same frequency of oscillation with a possible time shift of $2\mu\pi\tau$ radians (with reference to the input period) between the oscillating waveforms. The new conditions for periodicity are obtained from the conditions given in Secti n 2 by replacing T_f by μT_f and ω_f by ω_f/μ. The resulting conditions for periodicity are

$$\text{Im}\,\{R_1(\mu\pi + \beta_1 - \delta_1) + \Lambda_1(\omega_f/\mu)\} = -h_1/2$$
$$\text{Re}\,\{R_1(\mu\pi + \beta_1 - \delta_1) + \Lambda_1(\omega_f/\mu)\} < 0 \qquad (5)$$

and

$$\text{Im}\,\{R_2(\mu\pi + \beta_1 - \delta_2) + \Lambda_2(\omega_f/\mu)\} = -h_2/2$$
$$\text{Re}\,\{R_2(\mu\pi + \beta_1 - \delta_2) + \Lambda_2(\omega_f/\mu)\} < 0 \qquad (6)$$

where

$$\delta_2 = \delta_1 + 2\mu\pi\tau + \sigma. \qquad (7)$$

For a possible periodic solution, these conditions must be satisfied simultaneously. In addition, there must be only two switches of the relays per subharmonic period. That is, the following conditions must be satisfied:

$$e_1(t) > -h_1/2 \qquad (0 \leq t < \mu\pi/\omega_f)$$

and (8)

$$e_2(t) > -h_2/2 \qquad (-2\mu\pi\tau/\omega_f \leq t < 2\mu\pi(1/2-\tau)/\omega_f)$$

Conditions (5) and (6) lead to a consideration of the intersections, in the left half plane, of the $R_1(\mu\pi + \beta_1 - \delta_1)$ and $R_2(\mu\pi + \beta_1 - \delta_2)$ loci with the $-h_1/2$ and $-h_2/2$ lines respectively. The $R_1(\mu\pi + \beta_1 - \delta_1)$ and $R_2(\mu\pi + \beta_1 - \delta_2)$ loci are identical to the considered in the case of the fundamental oscillation but are centered at the $\omega_\mu = \omega_f/\mu$ points on the Λ loci (and not at the ω_f points as in the former case).

At frequency ω_f/μ there exists critical values of A_1 and A_2 (A^μ_{1K} and A^μ_{2K}) since these values must be large enough to ensure that the $R_1(\mu\pi + \beta_1 - \delta_1)$ and $R_2(\mu\pi + \beta_1 - \delta_2)$ loci intersect the $-h_1/2$ and $-h_2/2$ lines respectively. These critical values can be plotted in the A_{1K}, ω_f and A_{2K}, ω_f planes. The critical values for possible subharmonic oscillations can be obtained from the A^1_{1K} and A^1_{2K} curves by simple translation towards the right by an appropriate amount. Typical curves are shown in Fig. 3 for third and fifth order subharmonics. In general, a μ-th order subharmonic cannot occur if the values of A_1 and A_2, in the A_{1K}, ω_f and A_{2K}, ω_f planes, lie below the A^μ_{1K} and A^μ_{2K} curves respectively. A subharmonic oscillation is possible if the values of A_1 and A_2 are above the respective A^μ_{1K} and A^μ_{2K} curves. Fig. 3 shows that there are regions in which only third order subharmonics are possible, and regions where only fifth order subharmonics are possible. There are also regions where both the fundamental and subharmonics are possible. The conditions represented by Figs. 3(a) and (b) must be satisfied simultaneously for an oscillation to occur. The oscillation that actually occurs will depend upon the specific problem being considered. It should be noted that a pair of curves similar to those shown in Fig. 3 will be obtained for each value of τ considered.

The curves of Fig. 3 correspond to the conditions given by (5) and (6). The conditions given by (8) must be satisfied also. If a system exhibits a certain mode of oscillation, it will continue to do so unless one of the necessary conditions for this mode to occur is violated. For example, consider a two-variable system that can have a third order subharmonic oscillation. Suppose the input frequency to the system is fixed at a certain value close to $3\omega_o$, where ω_o is the self-oscillating frequency of the two-variable system. Fig. 4 shows the critical values of the inputs plotted against ω_f when $\tau=\tau_1$. A pair of curves similar to the ones shown will result for each value of τ considered. Now suppose that the values of A_1 and A_2 are such that the system is oscillating at the frequency of the inputs. Upon decreasing A_1 and A_2, the system will remain oscillating at the input frequency until one of the necessary conditions for this oscillation is violated. The conditions represented in Figs. 4(a) and (b) must be satisfied simultaneously. Thus, if the input amplitudes are decreased until either $A_1 = A_{1Kb}$ or $A_2 = A_{2Kb}$, then any further decrease results in a third order subharmonic. If the input amplitudes are decreased until either $A_1 < A_{1Kc}$ or $A_2 < A_{2Kc}$, then the system cannot have a third order subharmonic. For this case, the system exhibits almost periodic oscillations.

If now the amplitudes are increased, third order subharmonics will be obtained when both A_1 and A_2 are within the shaded areas shown in Figs. 4(a) and (b). These subharmonic oscillations will exist as long as the necessary conditions given by (5), (6) and (8) are fulfilled. This could result in an area in the A_{1K}, ω_f and A_{2K}, ω_f planes where both fundamental and subharmonics can occur. The critical necessary conditions are the ones given by (8). That is, when either

$$\min [e_1(t)] = -h_1/2 \qquad (0 < t < 3\pi/\omega_f)$$

or (9)

$$\min [e_2(t)] = -h_2/2 \qquad (-6\pi\tau/\omega_f < t < 6\pi(1/2 - \tau)/\omega_f)$$

the system will go from the third order subharmonic to the fundamental oscillation (see Fig. 5). The boundaries represented by (9) are shown by the dashed lines in Fig. 4. Thus, if the input amplitudes are increased until either $A_1 > A_{1Ka}$ or $A_2 > A_{2Ka}$, then the system will go from the third order subharmonic to the fundamental oscillation.

The boundaries represented by (9) can be found by first obtaining the output waveforms of $c_1(t)$ and $c_2(t)$, as given by (3) and (4), for some specified value of ω_3 and τ. The inputs $r_1(t)$ and $r_2(t)$ have known frequency (in this case three times the output frequency) and shape. Thus, $e_1(t)$ and $e_2(t)$ can be found since $e_1(t) = r_1(t) - c_1(t)$ and $e_2(t) = r_2(t) - c_2(t)$. In addition, if a μ-th order subharmonic exists, the following conditions at $t=0$, $-2\mu\pi\tau/\omega_f$, $\mu\pi/\omega_f$ and $2\mu\pi(1/2-\tau)/\omega_f$ are true:

$$\begin{aligned} e_1(0) &= h_1/2 \quad &, \quad \dot{e}_1(0) > 0 \\ e_2(-2\mu\pi\tau/\omega_f) &= h_2/2 \quad &, \quad \dot{e}_2(-2\mu\pi\tau/\omega_f) > 0 \\ e_1(\mu\pi/\omega_f) &= -h_1/2 \quad &, \quad \dot{e}_1(\mu\pi/\omega_f) < 0 \\ e_2(2\mu\pi(1/2-\tau)/\omega_f) &= -h_2/2 \quad &, \quad \dot{e}_2(2\mu\pi(1/2-\tau)/\omega_f) < 0 \, . \end{aligned} \tag{10}$$

Points on the dashed boundaries in Fig. 4 are obtained by changing the amplitude of $r_i(t)$ until $e_i(t) = -h_i/2$, $i=1,2$ at some point within the half period. This procedure is repeated for different values of frequency over the desired range. Since Fig. 4 shows the conditions for only one value of τ, the whole procedure must be repeated for a number of values of τ over its range. This means that a considerable amount of computation is involved. However, the whole procedure may be conveniently carried out by using a digital computer.

Once a family of curves similar to those of Fig. 4 are obtained for a number of values of τ over its range, a plot of the critical values of the input amplitudes versus τ can be made for a specified ω_f. That is, for each pair of curves similar to the ones shown in Fig. 4, the values corresponding to A_{1Ka}, A_{1Kb} and A_{1Kc} are plotted as functions of τ, and those corresponding to A_{2Ka}, A_{2Kb} and A_{2Kc} are plotted on the same plane. The common areas give the values of A_1, A_2 and τ for which third order subharmonics are possible (see Fig. 6). Fig. 6 shows a number of possibilities. If the values of A_1 and A_2 fall within the shaded area, the system can have third order subharmonics with τ somewhere between τ_d and τ_c. If now the amplitudes are increased, the system will continue to have third order subharmonics until one of the values of the amplitude goes outside the area a b c f e d a. The system will then go over to the fundamental oscillation. Once the system starts oscillating at fundamental frequency, the parameter τ assumes the value necessary to allow the oscillation to occur. That is, the parameter τ is not subject to direct external control. If the amplitudes of the inputs are decreased, the system will continue to oscillate at the fundamental frequency until the amplitudes approach the c d, d e lines of Fig. 6. Because the parameter τ is changing in some unknown manner, it is difficult to say where the system will start oscillating with a frequency of one third the input frequency.

For each point on the boundaries shown in Fig. 6, there is a value of $\delta_1 - \beta_1$ and $\delta_2 - \beta_1$. The values of $\delta_1 - \beta_1$ and $\delta_2 - \beta_1$ for the A_{Ka} boundaries are obtained from (10) with $\mu = 3$, while the values for the A_{Kb} and A_{Kc} boundaries are obtained from the intersections of the $R_1(3\pi + \beta_1 - \delta_1)$ and $R_2(3\pi + \beta_1 - \delta_2)$ loci with the $-h_1/2$ and $-h_2/2$ lines respectively. Then, using (7) with $\mu = 3$, a plot of $\delta_1 - \beta_1$ versus τ can be made. For each intersection of the $R_1(3\pi + \beta_1 - \delta_1)$ and $R_2(3\pi + \beta_1 - \delta_2)$ loci with the $-h_1/2$ and $-h_2/2$ lines, two values of $\delta_1 - \beta_1$ and $\delta_2 - \beta_1$ may be obtained. However, only one value of each need be considered as it has been shown[4,6] that only the values which result from the intersections to the left of Re $\Lambda_1(\omega_f/\mu)$ and Re $\Lambda_2(\omega_f/\mu)$ are stable. In order for a third order subharmonic to exist the value of σ must be such that the range of possible τ so obtained must correspond, at least in part, to the range of possible τ shown in Fig. 6.

Upon examining Fig. 6, it can be concluded that in order to avoid the possibility of third order subharmonics occurring, the amplitudes of the inputs should be larger than A_a. On the other hand, if the amplitudes of the inputs are smaller than A_a, the value of $\delta_1 - \beta_1$, $\delta_2 - \beta_1$ and τ for the possible subharmonic oscillation can be determined as follows. From the intersections of the $R_1(3\pi + \beta_1 - \delta_1)$ and $R_2(3\pi + \beta_1 - \delta_2)$ loci (which are centered at the $\omega = \omega_f/3$ points on the $\Lambda_1(\omega)$ and $\Lambda_2(\omega)$ loci respectively) with the $-h_1/2$ and $-h_2/2$ lines, obtain the values of $\delta_1 - \beta_1$ and $\delta_2 - \beta_1$. Next, using (7), plot $\delta_1 - \beta_1$ versus τ. Two values of $\delta_1 - \beta_1$ and τ are possible but as was shown by Nugent and Kavanagh[5], only one value will correspond to a stable oscillation. The example being discussed deals only with the determination of possible third order subharmonics. In a similar manner, possible fifth, seventh or higher order subharmonics can be predicted.

4. Illustrative Example

Consider the system shown in Fig. 1 where $W_{11}(s) = W_{22}(s) = 10/(1+s)^2$, $W_{12}(s) = -W_{21}(s) = -10\alpha/(1+s)^2$ and with relay parameters $M_1 = M_2 = 1$, $h_1 = h_2 = 2$ and $\Delta_1 = \Delta_2 = 0$. Let the system inputs be given by $r_1(t) = A_1 \sin(3.7t + \beta_1 - \delta_1)$ and $r_2(t) = A_2 \sin(3.7t + \beta_1 - \delta_1 - \sigma)$ where $r_2(t)$ leads $r_1(t)$ by 0.524 radians (30^o) and the cross coupling gain $\alpha = 4.0$.

The Λ_1 and Λ_2 loci are plotted in Fig. 7 for a number of values of τ. From these curves, the critical amplitudes A_{1K} and A_{2K} are obtained for different values of frequency. Typical plots of these critical amplitudes versus ω_f are shown in Fig. 8 for τ = 0.16 and 0.34. The A^3_{1K} and A^3_{2K} loci are obtained by moving the A^1_{1K} and A^1_{2K} loci to the right as shown. The boundaries where the system goes from the third order subharmonic to the fundamental oscillation (shown by the dashed lines in Fig. 8) are obtained by the method given in Section 3. When ω_f = 3.7 rad/sec, the A_{1Ka}, A_{1Kb}, A_{1Kc}, A_{2Ka}, A_{2Kb}, A_{2Kc} loci, as obtained from a series of plots of the type given in Fig. 8, are shown in Fig. 9. Fig. 9 shows that third order subharmonics are possible inside the area a b c d e f g h a. Inside the area g h f g only third order subharmonics can occur whereas, inside area a b c d e f h a either third order subharmonics or the fundamental oscillation can occur, depending on the initial conditions of the system. Fig. 10 shows the critical boundaries in the $\delta_1 - \beta_1$, τ plane. For the value of σ specified, third order subharmonics are possible inside the areas a b c d e f g a and p q r p. Note that the range of τ in Fig. 10 is included in the range of τ given by Fig. 9. Since the inputs to the system are sinusoidal, the $\delta_1 - \beta_1(A_{1Kb})$, $\delta_1 - \beta_1(A_{1Kc})$, $\delta_2 - \beta_1(A_{2Kb})$, $\delta_2 - \beta_1(A_{2Kc})$ boundaries where found by using the relationship

$$\delta_i - \beta_1 = \sin^{-1} A_{iK}/A_i \qquad (i=1,2). \qquad (11)$$

The $\delta_1 - \beta_1(A_{1Ka})$, $\delta_2 - \beta_1(A_{2Ka})$ boundaries were found by using (10). That is

$$\delta_1 - \beta_1 = -\sin^{-1}\left(\frac{c_1(0) + h_1/2}{A_1}\right)$$

and

$$\delta_2 - \beta_1 = -\sin^{-1}\left(\frac{c_2(-6\pi\tau/\omega_f) + h_2/2}{A_2}\right).$$

Before plotting the $\delta_2 - \beta_1$ values, $6\pi\tau + \sigma$ must be subtracted from each value as required by (7). Only the boundaries of the possible stable oscillations are plotted.

It can be concluded that since the parameter τ cannot be directly controlled externally, it would be necessary to keep either A_1 or A_2 greater than 22.0 (see Fig. 9) in order to be certain that third order subharmonics do not occur when ω_f = 3.7 rad/sec. However, this may not be possible and if third order subharmonics are undesirable, then some other frequency may have to be used or perhaps the linear transfer matrix may have to be changed in some manner so that the possibility of third order subharmonics occurring at this frequency would be reduced.

Now suppose that $A_1 = 3.0$ and $A_2 = 2.0$. With these values of input amplitudes, the system would be expected to exhibit third order subharmonics (see Fig. 9). It is of interest to determine the values of $\delta_1 - \beta_1$, $\delta_1 - \beta_1 + \sigma$ and τ. From Fig. 10, it is seen that the value of $\delta_1 - \beta_1$ and τ must lie within the area a b h e f g a. This reduces the amount of computation considerably. The procedure is to obtain the $\delta_1^a - \beta_1^a = f_1^a(\tau)$, $\delta_1^a - \beta_1^a = f_2^a(\tau) - 6\pi\tau - \sigma$ curves and look for intersections. These curves were obtained by using (11) and (7) where the values of A_{1K} and A_{2K} were obtained from Fig. 9. The result is shown in Fig. 11. In this figure, $\delta_1^a - \beta_1^a$ corresponds to the stable oscillation. It is seen that the system could have a subharmonic oscillation with frequency 1.23 rad/sec when $\delta_1 - \beta_1 = 1.39$ radians and $\tau = 0.267$. The value of $\delta_1 - \beta_1 + \sigma$, which is the time shift of $r_2(t)$ referred to the time origin, is 0.866 radians.

When the inputs $r_1(t) = 3\sin(3.7t + \beta_1 - \delta_1)$ and $r_2(t) = 2\sin(3.7t + \beta_1 - \delta_1 - \sigma)$ were applied to an analogue simulation with $\alpha = 4.0$ and $\sigma = -0.524$ radians, the system was observed to oscillate at a frequency of 1.23 rad/sec. Measured and predicted results are compared in Table 1.

TABLE 1. Measured and Predicted Results

Measured Values			
ω_3(rad/sec)	σ(rad)	$\delta_1 - \beta_1$(rad)	τ
1.23	-0.524	1.4	0.27
Predicted Values			
ω_3(rad/sec)	σ(rad)	$\delta_1 - \beta_1$(rad)	τ
1.23	-0.524	1.39	0.267

5. Conclusions

Tsypkin's method has been used to predict the existence of subharmonic oscillations in two-variable relay control systems. The validity of the approach has been confirmed by experimental investigations. The proposed method will lead to a better understanding of the behaviour of multivariable relay control systems and is of potential value in the design of multivariable oscillators.

6. Acknowledgements

This research was supported in part by the National Research Council of Canada (Grant Number A-1068 and a scholarship to the first author) and by the Defence Research Board of Canada (Grant Number 4003-05).

7. References

1. Sakawa, Y., "Subharmonic Oscillations in Relay Control Systems", Automation and Remote Control, Butterworth's, London, Vol. 1, pp. 404-409, 1961.
2. Gille, J.C. and Paquet, J.G., "Subharmonic Oscillations in On-Off Control Systems", Trans. of A.I.E.E., Vol. 81, Pt. 2, pp. 210-216, 1962.
3. Gille, J.C., Paquet, J.G. and Pouliquen, R., "Sur l'existence des Oscillations Forcees dans les Systems Asservis a Relais", Automatica, Vol. 2, No. 2, pp. 107-116, 1964.
4. Tsypkin, J.A., "Theorie der Relaissysteme der Automatischen Regelung", R. Oldenbourg-Verlag, Munich, 1958.
5. Nugent, S.T. and Kavanagh, R.J., "Self and Forced Oscillations in Multivariable Relay Control Systems", IFAC Symposium on Multivariable Control Systems, Duesseldorf, 1968.
6. Nugent, S.T., "Steady State Oscillations in Multivariable Relay Control Systems", Ph.D. Thesis, Electrical Engineering Department, University of New Brunswick, Fredericton, Canada, 1967.

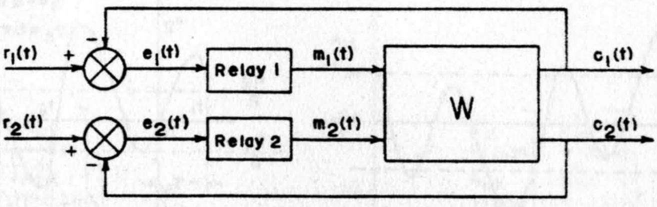

FIGURE 1. Block diagram of a two-variable relay control system

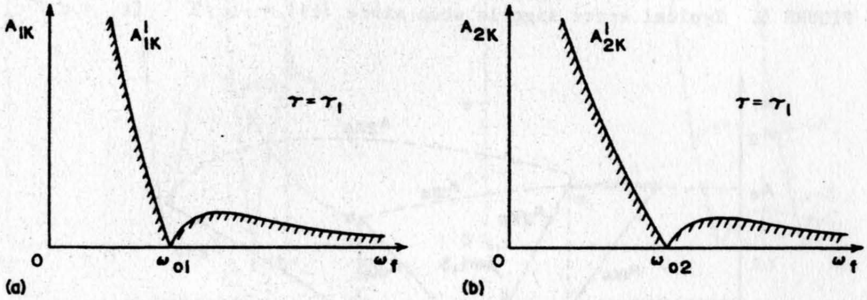

FIGURE 2. (a) A_{1K}, ω_f plane; (b) A_{2K}, ω_f plane

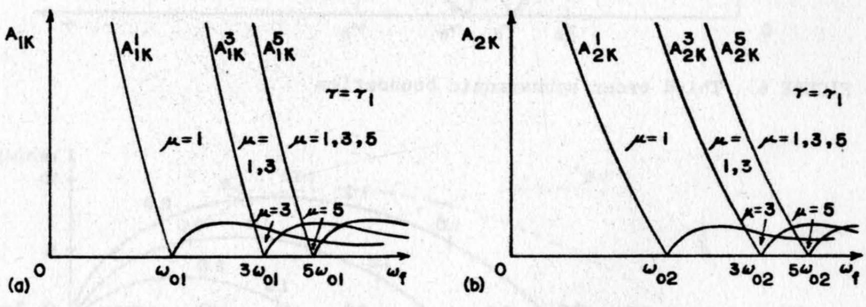

FIGURE 3. Subharmonic oscillations

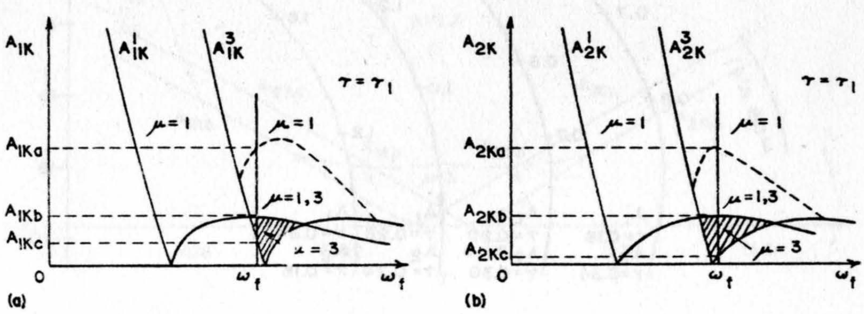

FIGURE 4. Third order subharmonic oscillations

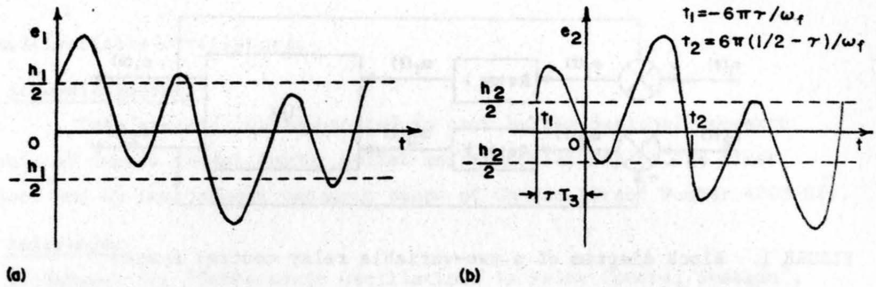

FIGURE 5. Typical error signals when $\min[e_2(t)] = -h_2/2 \quad (t_1 < t < t_2)$

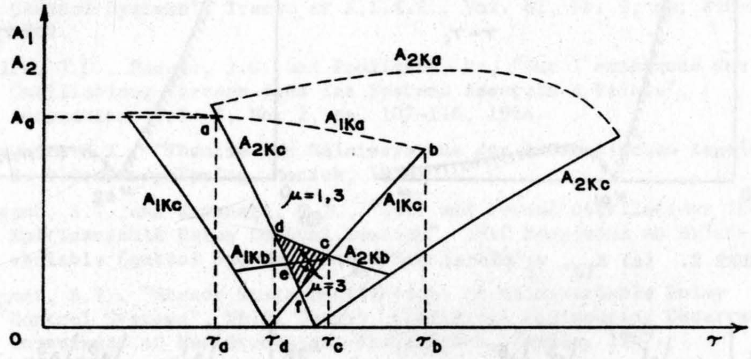

FIGURE 6. Third order subharmonic boundaries

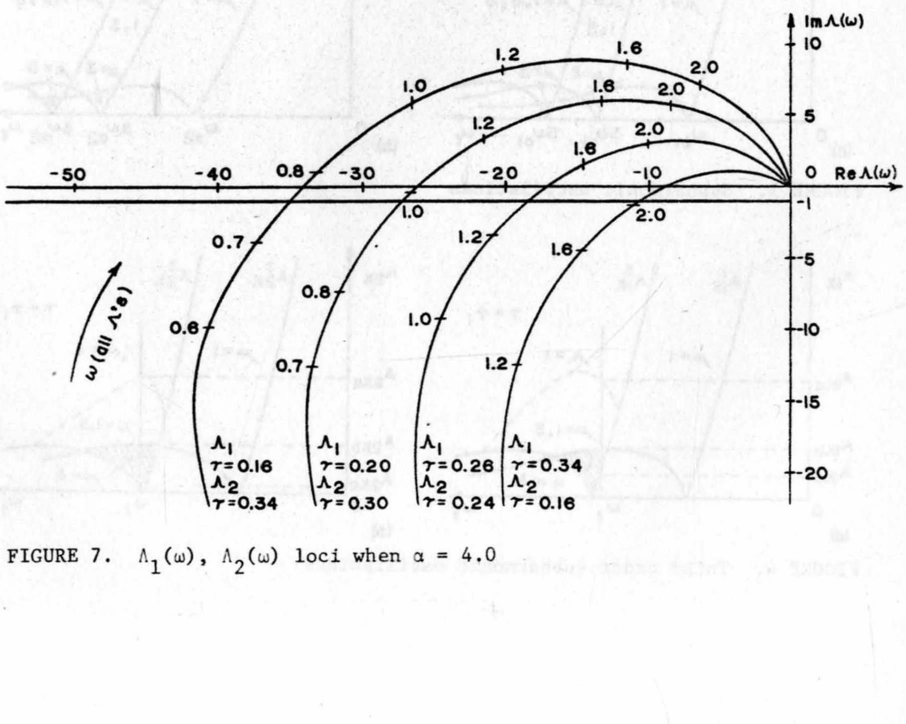

FIGURE 7. $\Lambda_1(\omega)$, $\Lambda_2(\omega)$ loci when $\alpha = 4.0$

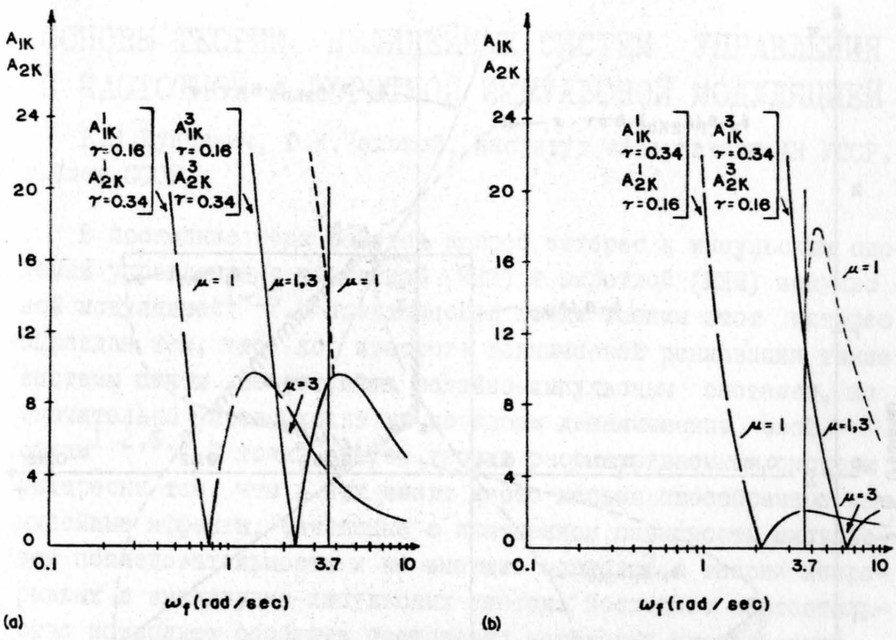

FIGURE 8. Critical input amplitudes A_{1K}, A_{2K} versus ω_f

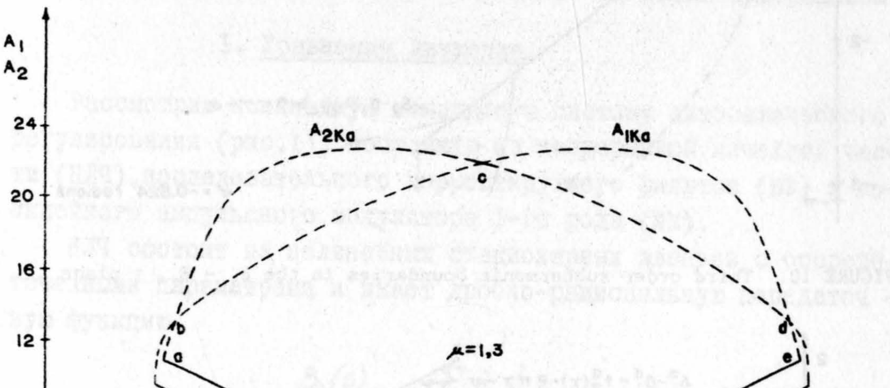

FIGURE 9. The third order subharmonic boundaries when ω_f = 3.7 rad/sec

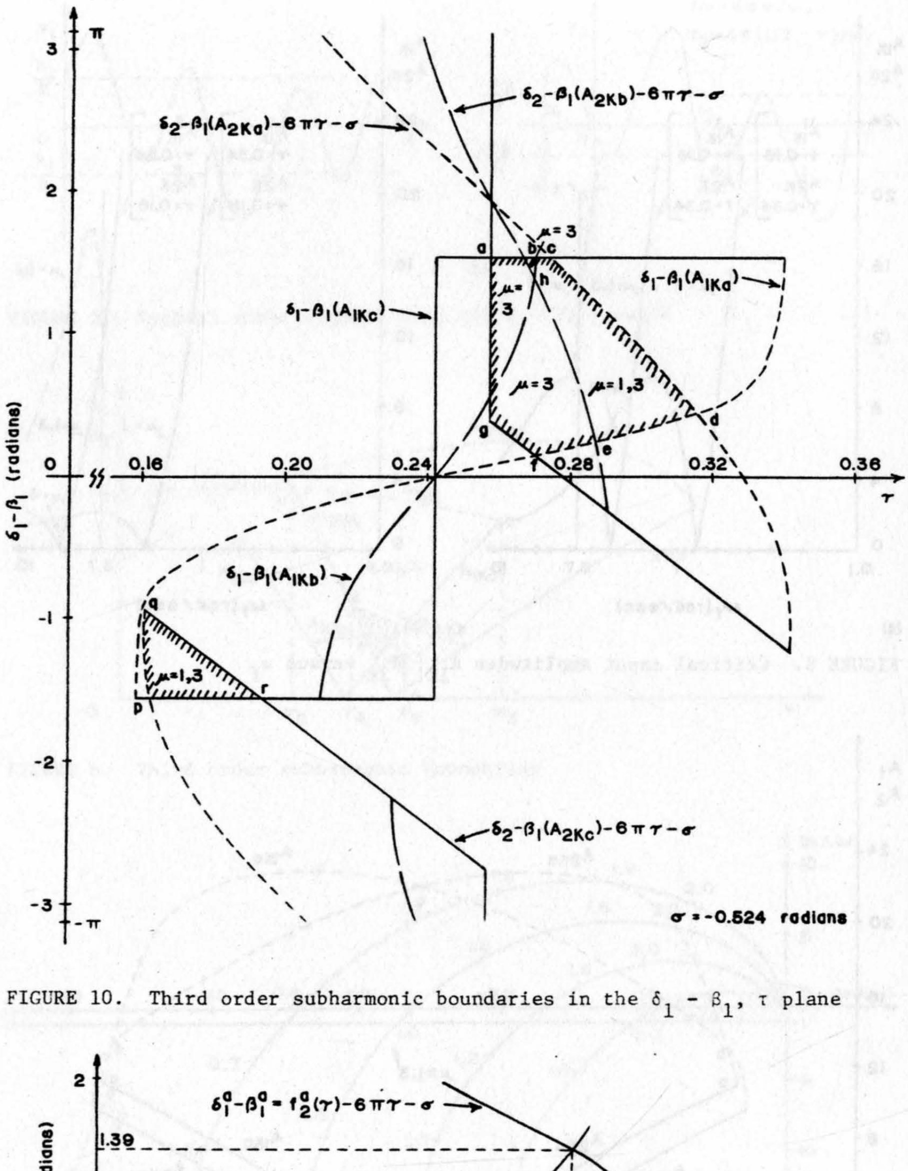

FIGURE 10. Third order subharmonic boundaries in the $\delta_1 - \beta_1$, τ plane

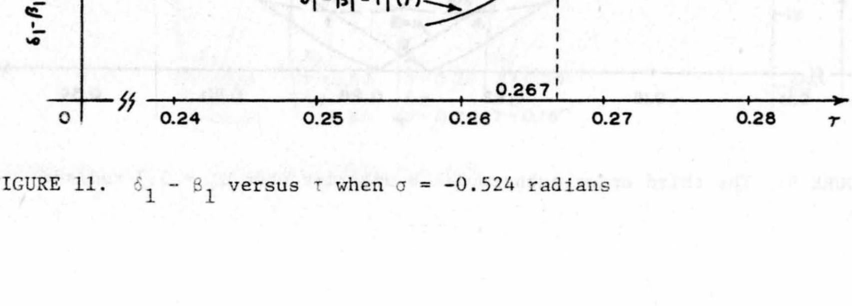

FIGURE 11. $\delta_1 - \beta_1$ versus τ when $\sigma = -0.524$ radians

ОСНОВЫ ТЕОРИИ НЕЛИНЕЙНЫХ СИСТЕМ УПРАВЛЕНИЯ С ЧАСТОТНОЙ И ШИРОТНОЙ ИМПУЛЬСНОЙ МОДУЛЯЦИЕЙ

В.М.Кунцевич, Ю.Н.Чеховой, Институт кибернетики АН УССР, г.Киев,СССР.

В последние годы заметно возрос интерес к импульсным системам управления с частотной (ЧИМ) и широтной (ШИМ) импульсной модуляцией[I-I6].С практической точки зрения этот интерес оправдан тем, что по простоте технической реализации такие системы почти не уступают релейно-импульсным системам, но значительно превосходят их по своим динамическим свойствам[I,2,6,9].С точки зрения теории рассматриваемые системы интересны тем, что в них имеют место весьма своеобразные нелинейные эффекты, связанные с изменением скважности импульсной последовательности и не имеющие аналогов в теории непрерывных и амплитудно-импульсных систем. Последнее обстоятельство позволяет обобщить постановки некоторых классических задач теории автоматического регулирования (например, задачи об абсолютной устойчивости) и обогащает их новым содержанием.

I. Уравнения движения

Рассмотрим нелинейную импульсную систему автоматического регулирования (рис.I), состоящую из непрерывной линейной части (НЛЧ),последовательного корректирующего фильтра (КФ) и нелинейного импульсного модулятора I-го рода (ИМ).

НЛЧ состоит из нелинейных стационарных звеньев с сосредоточенными параметрами и имеет дробно-рациональную передаточную функцию

$$W(s)=\frac{B_\ell(s)}{A_m(s)}=\frac{\sum\limits_{i=0}^{\ell} b_i s^i}{s^m+\sum\limits_{i=0}^{m-1} a_i s^i} \quad (\ell<m). \qquad (I.I)$$

КФ представляет собой линейный корректирующий фильтр, описываемый уравнением

$$\sigma = C^T(U-X), \qquad (I.2)$$

где $X=(x, x', \dots, x^{(m-1)})$ - вектор-столбец фазовых координат системы; $U=(u,u',\dots,u^{(m-1)})$; u - задающее воздействие;

$C=(c_1, c_2, \ldots, c_m)$ -числовой вектор-столбец; $c_i=0$ при $i>k$ $(1\leq k\leq m-l)$; символ "т" обозначает операцию транспонирования.

Структурные схемы возможных вариантов ИМ (осуществляющих различные виды импульсной модуляции) изображены на рис.2. Здесь ИЭ - идеальный импульсный элемент (амплитудно-импульсный модулятор); "звездочкой" обозначена операция квантования по времени, осуществляемая импульсным элементом: $\sigma^*=\sum_{n=0}^{\infty}\sigma_n\delta(t-t_n)$; $\delta(t)$ -единичная δ-функция; $t_n=\sum_{i=0}^{n-1}T_i$ -момент появления n-го импульса $(t_0=0)$; T_n -интервал между n-м и $(n+1)$-м импульсами; $\sigma_n=\lim_{0<\varepsilon\to 0}\sigma(t_n-\varepsilon)$; Ф - фиксатор нулевого порядка (с переменным или постоянным интервалом фиксации); РЭ-релейный элемент; f и F -время-задающие элементы, управляющие фиксаторами Ф и импульсными элементами ИЭ. Идеальный частотно-импульсный модулятор (рис.2,а) модулирует по частоте и знаку последовательность $z^*(t)$ единичных δ-импульсов. Реальный частотно-импульсный модулятор (рис.2,б) модулирует по частоте и знаку последовательность $y(t)$ прямоугольных импульсов,которые имеют постоянную длительность τ и единичную амплитуду.Широтно-импульсный модулятор (рис.2,в) модулирует по знаку и длительности последовательность $y(t)$ прямоугольных импульсов,следующих с постоянной частотой $1/T$. Наконец,частотно-широтный импульсный модулятор (рис.2,г) осуществляет модуляцию последовательности $y(t)$ по знаку, частоте и длительности.

Для всех вариантов ИМ модуляция по знаку определяется релейной функцией (характеристикой РЭ)

$$z_n=z(\sigma_n)=\begin{cases}\mathrm{sign}\,\sigma_n & \text{при } |\sigma_n|>\Delta;\\ 0 & \text{при } |\sigma_n|\leq\Delta;\end{cases} \qquad (\text{I}.3)$$

модуляция по частоте определяется законом ЧИМ

$$T_n=F(\sigma_n) \qquad (\text{I}.4)$$

и модуляция по длительности - законом ШИМ

$$0\leq\tau_n=f(\sigma_n)=\begin{cases}<T_n & \text{при } |\sigma_n|<\Delta_0;\\ =T_n=const & \text{при } |\sigma_n|\geq\Delta_0.\end{cases} \qquad (\text{I}.5)$$

Здесь $F(\sigma)$ и $f(\sigma)$ -чётные однозначные функции, определенные при всех σ ; $F(\sigma)>0, f(\sigma)\geq 0$, причем $f(\sigma)$ может обращаться в нуль только при $|\sigma| \le \Delta$; $\Delta_0 > \Delta$ -порог насыщения ИМ. Разностные уравнения движения рассматриваемых систем приводятся к следующему виду[10,11,17,18]:

$$X_{n+1} = H_n (X_n + K_n), \tag{I.6}$$

где $X_n = (x_n, x_n', ..., x_n^{(m-1)})$; $x_n^{(i)} = \lim\limits_{0<\varepsilon\to 0} x^{(i)}(t_n - \varepsilon)$; $H_n = H[F(\sigma_n)] = \exp AF(\sigma_n)$ - переходная матрица НЛЧ; A -сопровождающая матрица характеристического полинома НЛЧ; $K_n = K[f(\sigma_n)] \geq (\sigma_n)$ -вектор смены состояний НЛЧ. Функция $K(f)$ зависит от характера импульсной модуляции. При идеальной ЧИМ (рис.2,а)

$$K(f) = G = (g, g', ..., g^{(m-1)}); \; g^{(i)} = g^{(i)}(0); \; g(t) = \mathcal{L}^{-1}[W(s)]; \tag{I.7}$$

при реальной ЧИМ (рис.2,б)

$$K(f) = H(-\tau)R(\tau); \quad R(\tau) = (r(\tau), r'(\tau), ..., r^{(m-1)}(\tau)); \quad r(t) = \mathcal{L}^{-1}[\tfrac{1}{s}W(s)]; \tag{I.8}$$

наконец, при ШИМ (рис.2,в) и при двойной импульсной модуляции (ЧИМ и ШИМ, рис.2,г)

$$K(f) = H(-\tau_n)R(\tau_n) = H[-f(\sigma_n)]R[f(\sigma_n)]. \tag{I.9}$$

Матричное уравнение (I.6) описывает движение системы (рис.I) в естественном фазовом пространстве $E^m = \{X_n\}$. Покажем, что от (I.6) всегда можно перейти к уравнению

$$\check{X}_{n+1} = \check{H}_n (\check{X}_n + \check{K}_n) \tag{I.10}$$

в разностном фазовом пространстве $\mathcal{D}^m = \{\check{X}_n\}$, $\check{X}_n = (x_n, x_{n+1}, ..., x_{n+m-1})$. Составим следующую систему уравнений[10,11]:

$$\begin{Vmatrix} -H_n & I & 0 & \cdots & 0 & 0 \\ 0 & -H_{n+1} & I & \cdots & 0 & 0 \\ 0 & 0 & -H_{n+2} & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -H_{n+m-1} & I \end{Vmatrix} \cdot \begin{Vmatrix} X_n \\ X_{n+1} \\ X_{n+2} \\ \cdots \\ X_{n+m-1} \\ X_{n+m} \end{Vmatrix} = \begin{Vmatrix} H_n K_n \\ H_{n+1} K_{n+1} \\ H_{n+2} K_{n+2} \\ \cdots \\ H_{n+m-1} K_{n+m-1} \end{Vmatrix}, \tag{I.11}$$

где I -единичная матрица. Система (I.11) имеет прямоугольную матрицу размера $m^2 \times m(m+1)$ и ранга m^2. Поэтому можно, положив переменные x_{n+i} $(i = 0, 1, ..., m-1)$ известными, разрешить (I.11) относительно x_{n+m}. Этот результат, представленный в матричной форме, дает уравнение (I.10).

В дальнейшем мы будем пользоваться уравнениями движения только в форме (I.6), так как она проще связана с параметрами системы и удобнее для исследования.

2. Устойчивость равновесных состояний

Воспользуемся дискретными аналогами теорем прямого метода Ляпунова[II,19,20]. Рассмотрим основной случай, когда НЛЧ системы устойчива, и простейший критический случай, когда НЛЧ нейтральна. Для того чтобы система (рис.I) имела равновесное состояние, положим $u = const$.

Основной случай. Согласно (I.6) вектор-столбец X_∞ координат равновесной точки должен удовлетворять равенству

$$X_\infty = \left(H_\infty^{-1} - I\right)^{-1} K_\infty, \tag{2.I}$$

где H_∞ и K_∞ – матрица H_n и вектор K_n при $X_n = X_\infty$.

При $|u| < \Delta/c_1$ ($c_1 > 0$ –элемент вектора C) уравнение (2.I) имеет нулевое решение $X_\infty = 0$, т.е. система имеет равновесную точку в начале координат. Этот факт легко проверяется простой подстановкой. При $|u| > \Delta/c_1$ решение $X_\infty \neq 0$, причем в общем случае решений может быть несколько[21].

С помощью подстановки $X_n = E_n^0 + X_\infty$ переместим начало координат фазового пространства E^m в равновесную точку системы. В новых координатах вместо (I.6) получим

$$E_{n+1}^0 = H_n\left(E_n^0 + K_n^0\right), \tag{2.2}$$

где

$$K_n^0 = \left(I - H_n^{-1}\right)X_\infty + K_n; \quad \sigma_n = C^T\left(E_\infty - E_n^0\right); \quad E_\infty = U - X_\infty. \tag{2.3}$$

Функцию Ляпунова выберем в виде положительно определенной квадратичной формы

$$v_n = \left(E_n^0\right)^T P E_n^0, \qquad P > 0. \tag{2.4}$$

Первая разность функции (2.4) в силу (2.2) равна

$$\Delta v_n = -\left(E_n^0\right)^T\left(P - M_n\right)E_n^0 + 2\left(E_n^0\right)^T M_n K_n^0 + \left(K_n^0\right)^T M_n K_n^0, \qquad M_n = H_n^T P H_n. \tag{2.5}$$

Система (2.2) асимптотически устойчива в целом, если при всех $E_n^0 \neq 0$ функция (2.5) отрицательна, т.е. если выполнены неравенства[11,19,20]:

$$P-M_n>0; \qquad (2.6)$$

$$(E_n^o)^T(P-M_n)E_n^o-2(E_n^o)^T M_n K_n^o>(K_n^o)^T M_n K_n^o. \qquad (2.7)$$

Показано,что $\lim\limits_{F\to\infty} M[F(\sigma_n)]=0$,поэтому существует класс достаточно больших функций (1.4), при которых условие (2.6) выполняется[22].Допустим,что (1.4) принадлежит этому пока еще неизвестному нам классу, и обратимся к условию (2.7).

Рассмотрим уравнение поверхности, на которой функция (2.5) обращается в нуль:

$$(E_n^o)^T(P-M_n)E_n^o-2(E_n^o)^T M_n K_n^o=(K_n^o)^T M_n K_n^o. \qquad (2.8)$$

Условие (2.7) будет выполнено,если при всех $E_n^o\neq 0$ эта поверхность не существует. Подставим в (2.8) вместо σ_n ,определяемого из (2.3),выражение $\sigma_\infty-\sigma$,где $\sigma_\infty=C^T E_\infty$, a σ -произвольный вещественный параметр,не зависящий от E_n^o :

$$(E_n^o)^T(P-M)E_n^o-2(E_n^o)^T MK^o=(K^o)^T MK^o; \quad M=M_n|_{\sigma_n=\sigma_\infty-\sigma}; \quad K^o=K_n^o|_{\sigma_n=\sigma_\infty-\sigma}. \qquad (2.9)$$

Уравнение (2.9) описывает семейство m-мерных эллипсоидов,зависящих от параметра σ , причем поверхность (2.8) не существует при $E_n^o\neq 0$, если при всяком $\sigma\neq 0$ эллипсоид (2.9) не соприкасается с плоскостью $C^T E_n^o=\sigma$ (рис.3,а)[11,20]. Построим плоскость

$$C^T E_n^o=\rho(\sigma), \qquad (2.10)$$

касательную к эллипсоиду (2.9). Тогда условие (2.7) транс - формируется в неравенство

$$|\sigma|>|\rho(\sigma)|, \quad \sigma\neq 0, \quad |\sigma|\leq\Delta_o. \qquad (2.11)$$

Для определения функции $\rho(\sigma)$ запишем общее уравнение плоскости, касательной к эллипсоиду (2.9) в некоторой точке A[21-23]:

$$(E_n^o)^T(P-M)E^o(A)-[E_n^o+E^o(A)]^T MK^o=(K^o)^T MK^o, \qquad (2.12)$$

где $E^o(A)$ -радиус-вектор точки касания (рис.3,а). Плоскости (2.10) и (2.12) совпадают, если при некотором $\alpha>0$

$$(P-M)E^o(A)-MK^o=\alpha C; \qquad (2.13)$$

$$(K^o)^T MK^o+[E^o(A)]^T MK^o=\alpha\rho(\sigma). \qquad (2.14)$$

Совместное решение (2.9),(2.13) и (2.14) дает[21,22]:

$$\rho(\sigma)=\left\{(K^{\circ})^{T}\left[M+M(P-M)^{-1}M\right]K^{\circ}C^{T}(P-M)^{-1}C\right\}^{\frac{1}{2}}\operatorname{sign}\sigma+C^{T}(P-M)^{-1}MK^{\circ}. \qquad (2.15)$$

Заметим, что в тех случаях, когда $|u|\leq \Delta/c_1$, функция (2.15) обращается в нуль при $|\sigma|\leq\Delta$ и неравенство (2.11) достаточно проверить лишь при $|\sigma|>\Delta$.

Условие (2.11) и формула (2.15) получены в предположении, что при всех $\sigma\neq 0$ имеет место (2.6), однако проверять выполнение этого условия на всей оси σ нет необходимости. В самом деле, допустим, что при $\sigma=\sigma_1$ неравенство (2.6) выполнено, а при $\sigma=\sigma_2\neq\sigma_1$ нарушено; тогда обязательно найдется такое $\sigma_3\in(\sigma_1,\sigma_2)$, при котором матрица $(P-M)$ вырождена и функция (2.15) претерпевает разрыв непрерывности. Это означает, что проверку (2.6) достаточно произвести всего в одной (любой) точке каждого интервала непрерывности функции (2.15). Теперь окончательный результат можно сформулировать следующим образом: в основном случае равновесное состояние X_∞ системы (1.6) асимптотически устойчиво в целом, если выполнено условие (2.11) и внутри каждого интервала непрерывности функции (2.15) можно указать хотя бы одно значение σ, при котором выполняется неравенство (2.6).

<u>Простейший критический случай</u>. В простейшем критическом случае один корень характеристического уравнения $A_m(s)=0$ равен нулю $(a_0=0)$, матрица A вырождена и имеет ранг $m-1$. Преобразуем уравнение (1.6), умножив его слева на матрицу[22]

$$R=\begin{Vmatrix} I & \frac{a_2}{a_1} & \cdots & \frac{a_{m-1}}{a_1} & \frac{1}{a_1}\\ 0 & I & \cdots & 0 & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & I & 0\\ 0 & 0 & \cdots & 0 & I \end{Vmatrix}. \qquad (2.16)$$

После несложных преобразований получим:

$$\tilde{E}^{\circ}_{n+1}=\tilde{H}_n\left(\tilde{E}^{\circ}_n+\tilde{K}_n\right),\qquad \sigma_n=\tilde{C}^{T}(U-\tilde{X}_n)=-\tilde{C}^{T}\tilde{E}^{\circ}_n, \qquad (2.17)$$

где $\tilde{E}^{\circ}_n=\tilde{X}_n-U$; $\tilde{X}_n=RX_n$; $\tilde{K}_n=RK_n$; $\tilde{C}=(R^{T})^{-1}C$; $\tilde{H}_n=RH_nR^{-1}=$ $=\operatorname{diag}\{1,(\tilde{H}_n)_{1,1}\}$; $(\tilde{H}_n)_{1,1}$ – матрица порядка $m-1$, полученная из $\tilde{H}_n$ вычеркиванием 1-й строки и 1-го столбца. Уравнению

(2.I7) соответствует множество $\mathfrak{E}$ положений равновесия, на котором

$$(\tilde{X}_n)_1 = (X_n)_1 = 0; \quad |u - \tilde{x}_{1,n}| = |u - x_n| \leq \frac{\Delta}{\tilde{c}_1} = \frac{\Delta}{c_1}. \qquad (2.18)$$

Здесь $(X)_1$ -вектор-столбец,полученный из X вычеркиванием I-го элемента; $\tilde{x}_{1,n}$ и $\tilde{c}_1$ -элементы векторов $\tilde{X}_n$ и $\tilde{C}$.

Функцию Ляпунова выберем в виде

$$v_n = (\tilde{E}_n^{\circ})^T P \tilde{E}_n^{\circ}, \qquad (2.19)$$

потребовав дополнительно от матрицы P ,чтобы $P = diag\{1, (P)_{1,1}\}$; $(P)_{1,1} > 0$. Тогда

$$\Delta v_n = -(\tilde{E}_n^{\circ})_1^T (P - \tilde{M}_n)_{1,1} (\tilde{E}_n^{\circ})_1 + 2(\tilde{E}_n^{\circ})^T \tilde{M}_n \tilde{K}_n + \tilde{K}_n^T \tilde{M}_n \tilde{K}_n;$$
$$\tilde{M}_n = \tilde{H}_n^T P \tilde{H}_n. \qquad (2.20)$$

В соответствии с дискретным аналогом теоремы Ж.Ла-Салля множество $\mathfrak{E}$ равновесных состояний системы (2.I7) асимптотически устойчиво в целом, если [11,20]

$$\Delta v_n = 0, \quad \tilde{E}_n^{\circ} \in \mathfrak{E}; \quad \Delta v_n < 0, \quad \tilde{E}_n^{\circ} \notin \mathfrak{E}. \qquad (2.21)$$

Из (2.20),(2.I7) и определения $\mathfrak{E}$ следует, что первое из двух условий (2.2I) всегда выполняется. Проверка второго условия (2.2I) аналогично предыдущему приводит нас к геометрической задаче: найти условия, при которых m-мерный параболоид

$$(\tilde{E}_n^{\circ})_1^T (P - \tilde{M})_{1,1} (\tilde{E}_n^{\circ})_1 - 2(\tilde{E}_n^{\circ})^T \tilde{M} \tilde{K} = \tilde{K}^T \tilde{M} \tilde{K}; \quad \begin{matrix} \tilde{M} = \tilde{M}_n |_{\sigma_n = -\sigma}; \\ \tilde{K} = \tilde{K}_n |_{\sigma_n = -\sigma} \end{matrix} \qquad (2.22)$$

не соприкасается с плоскостью $\tilde{C}^T \tilde{E}_n^{\circ} = \sigma$ (рис.3,б).Как и ранее, решение этой задачи приводит к неравенству (2.II). Опуская промежуточные выкладки (аналогичные выполненным в основном случае), запишем выражение для функции $\rho(\sigma)$ [22]:

$$\rho(\sigma) = -\frac{c_1}{2\tilde{k}_1} \Big\{ \tilde{k}_1^2 + (\tilde{K})_1^T \big[(\tilde{M})_{1,1} + (\tilde{M})_{1,1} (P - \tilde{M})_{1,1}^{-1} (\tilde{M})_{1,1} \big] (\tilde{K})_1 + \qquad (2.23)$$
$$+ \frac{\tilde{k}_1^2}{c_1^2} (\tilde{C})_1^T (P - \tilde{M})_{1,1}^{-1} (\tilde{C})_1 \Big\} + (\tilde{C})_1^T (P - \tilde{M})_{1,1}^{-1} (\tilde{M})_{1,1} (\tilde{K})_1,$$

где $\tilde{k}_1$ -элемент вектора $\tilde{K}$.

Функция (2.23) обращается в нуль при $|\sigma| \leq \Delta$,поэтому в простейшем критическом случае (независимо от величины u) неравенство (2.II) достаточно проверить лишь при $|\sigma| > \Delta$.

Дополнительное условие (2.6) в простейшем критическом случае принимает вид

$$(P-\tilde{M})_{11} > 0. \qquad (2.24)$$

С учётом последних замечаний сформулируем окончательный результат: в простейшем критическом случае множество $\mathfrak{S}$ равновесных состояний системы (I.6) асимптотически устойчиво в целом, если при $|\sigma|>\Delta$ выполнено условие (2.II) и внутри каждого интервала непрерывности функции (2.23) можно указать хотя бы одно значение σ, при котором выполняется неравенство (2.24).

Критический коэффициент передачи НЛЧ. Представим (I.I) в виде $W(s)=kW_0(s)$, где постоянная k (назовем ее коэффициентом передачи НЛЧ) равна $\lim_{s\to 0} W(s)$ в основном и $\lim_{s\to 0} sW(s)$ в простейшем критическом случаях. Тогда неравенство (2.II) приводится к виду

$$\frac{|\sigma|}{k} > |\rho_0(\sigma)|, \qquad (2.26)$$

где $\rho_0(\sigma)$ находится по формулам (2.I5) или (2.23), если вместо (I.I) подставить $W_0(s)$. Из (2.26) следует, что критический коэффициент передачи, (т.е. наименьшее k, при котором не соблюдено условие устойчивости) соответствует наибольшему числу k, при котором функция $\rho_0(\sigma)$ содержится в секторе $[0, \frac{1}{k}]$ (рис.4).

Абсолютная устойчивость равновесных состояний. Система, состоящая из нелинейного элемента (НЭ) и линейной части (ЛЧ), называется абсолютно устойчивой, если она асимптотически устойчива в целом при всех характеристиках НЭ, принадлежащих некоторому классу[24,25]. В исследуемой системе (рис.I) нелинейным элементом является ИМ, свойства которого полностью определяются тремя характеристиками: (I.3), (I.4) и (I.5). Поэтому задачу об абсолютной устойчивости здесь можно рассматривать в трех различных постановках: I) известны (I.4) и (I.5), требуется найти класс допустимых функций (I.3); 2) известны (I.3) и (I.5), требуется найти класс допустимых законов ЧИМ (I.4); 3) известны (I.3) и (I.4), требуется найти класс допустимых законов ШИМ (I.5)[26].

Для упрощения задачи положим в основном случае $u=0$. Тогда условие устойчивости в целом примет вид:

$$|\sigma| > |\rho(\sigma)|, \qquad \Delta < |\sigma| \le \Delta_0. \qquad (2.27)$$

I-я постановка задачи об абсолютной устойчивости тривиальна и полностью решается условием (2.27). В самом деле, если при некотором $\Delta=\Delta_*$ условие (2.27) выполнено, то оно бу -

дет выполнено и при всех $\Delta \geq \Delta_*$.Последнее неравенство полностью решает задачу, т.к. оно определяет искомый класс функций (I.3).

Для решения задачи во 2-й и 3-й постановках представим функцию $\rho(\sigma)$ при $\sigma > 0$ в следующей форме:

$$\rho(\sigma) = \rho_\Delta[F(\sigma), f(\sigma)] = \rho_\Delta(F, f), \quad \sigma > 0. \qquad (2.28)$$

При $F \to \infty$ и при $f \to 0$ исследуемая система размыкается.По условию НЛЧ устойчива (или предельно устойчива[24]), поэтому допустимый класс функций $F(\sigma)$ ограничен снизу,а $f(\sigma)$ -сверху. Рассмотрим уравнение

$$\sigma = \rho_\Delta(F, f), \quad \sigma > 0, \qquad (2.29)$$

соответствующее границе области устойчивости,обеспечиваемой неравенством (2.27).Если функция (I.5) задана,то уравнение (2.29) неявно задает функцию

$$\sigma = \rho_{\Delta, f}(F), \quad \sigma > 0. \qquad (2.30)$$

Функция (2.30) определена и положительна на интервале $F \in (F_0, \infty)$,где $F_0 \geq 0$ -наибольшее значение F ,при котором нарушаются условия (2.6) или (2.24). Поэтому существует положительная обратная функция $F_*(\sigma) = \rho^{-1}_{\Delta, f}(\sigma)$, с помощью которой условие (2.27) можно привести к виду:

$$F(\sigma) > F_*(\sigma) = \rho^{-1}_{\Delta, f}(\sigma), \quad \sigma \in (\Delta, \Delta_0]. \qquad (2.31)$$

Полученное неравенство определяет искомый класс функций (I.4).Аналогично класс допустимых функций (I.5) определяется неравенством

$$f(\sigma) < f_*(\sigma) = \rho^{-1}_{\Delta, F}(\sigma), \quad \sigma \in (\Delta, \Delta_0]. \qquad (2.32)$$

Обратные функции $\rho^{-1}_{\Delta, f}(\sigma)$ и $\rho^{-1}_{\Delta, F}(\sigma)$ в общем случае не удается найти в аналитической форме,однако они легко находятся графически[26].

Область асимптотической устойчивости.Если условие (2.27) нарушается при $\sigma = \Delta_1 \in (\Delta, \Delta_0]$,то исследуемая система не обладает асимптотической устойчивостью в целом, но она имеет область (в пространстве E^m) асимптотической устойчивости.Оценкой этой области является наибольшая из открытых областей,ограниченных поверхностями $v_n = \beta = const$ и удовлетворяющих нера -

венству $|\sigma| < \Delta_1$.Нетрудно показать, что в основном и простейшем критическом случаях, соответственно, величина β равняется[26]:

$$\beta = \frac{\Delta_1}{C^T P^{-1} C}; \quad \beta = \frac{\Delta_1}{\bar{C}^T P^{-1} \bar{C}}. \qquad (2,33)$$

Замечание относительно выбора функции Ляпунова.Из изложенного следует, что неравенства (2.3I) и (2.32) обеспечивают область устойчивости при любых $P>0$ и $(P)_{1,1}>0$,соответственно. Неплохие результаты и существенное упрощение выкладок можно получить, если выбрать

$$P = S^* S, \quad P = S^* \mathcal{D} S, \quad (P)_{1,1} = S^* S \text{ или } (P)_{1,1} = S^* \mathcal{D} S, \qquad (2.34)$$

где S -матрица, преобразующая матрицу A или $(A)_{1,1}$,соответственно,к нормальной жордановой форме; S^* -матрица,эрмитово сопряженная с S; $\mathcal{D}$ -диагональная матрица с положительными элементами[II,I7,I8,20-22,26].

3. Устойчивость стационарного вынужденного режима

Рассмотрим систему (I,6) в простейшем критическом случае при $u(t) = \omega t$ (режим слежения за линейно нарастающим задающим сигналом). С помощью подстановки $X_n = U_n - E_n$ приведем (I.6) к следующему виду:

$$E_{n+1} = H_n(E_n + L_n), \quad L_n = H_n^{-1} U_{n+1} - U_n - K_n. \qquad (3.I)$$

Можно показать,что при $u(t) = \omega t$ это уравнение не зависит явно от t_n и для него существует предельное равенство $\lim_{n\to\infty} E_n = E_\infty$, где E_∞ -числовой вектор[27].Согласно (3.I) вектор E_∞ должен удовлетворять равенству:

$$(H_\infty^{-1} - I) E_\infty = L_\infty, \quad L_\infty = \lim_{n\to\infty} L_n. \qquad (3.2)$$

Можно показать,что всегда существует интервал $\omega \in (\omega_1, \omega_2)$ $(\omega_2 > \omega_1 \geq 0)$, при котором уравнение (3.2) имеет по крайней мере одно решение[27]. Путем подстановки $E_n = E_\infty - E_n^\circ$ переместим начало координат фазового пространства в точку, соответствующую исследуемому стационарному режиму.В новых координатах вместо (3.I) получим:

$$E_{n+1}^\circ = H_n(E_n^\circ + L_n^\circ), \quad L_n^\circ = -(I - H_n^{-1}) E_\infty - L_n, \quad \sigma_n = C^T(E_\infty - E_n^\circ). \qquad (3.3)$$

Аналогично предыдущему преобразуем уравнение (3.3), умножив его слева на матрицу (2.I6):

$$\tilde{E}^{\circ}_{n+1} = \tilde{H}_n\left(\tilde{E}^{\circ}_n + \tilde{L}^{\circ}_n\right), \quad \sigma_n = \tilde{C}^T\left(\tilde{E}_\infty - \tilde{E}^{\circ}_n\right) = \sigma_\infty - \tilde{C}^T\tilde{E}^{\circ}_n, \qquad (3.4)$$

где $\tilde{L}^{\circ}_n = RL^{\circ}_n$, $\tilde{E}_\infty = RE_\infty$ (остальные обозначения соответствуют принятым ранее). Уравнение (3.4) аналогично уравнению (2.I7), однако вместо множества $\mathfrak{E}$ равновесных состояний система (3.4) имеет одну равновесную точку $\tilde{E}^{\circ}_n = 0$. При выборе функции Ляпунова в виде (2.I9) условие устойчивости (аналогично предыдущему) получим в виде (2.II), где

$$\rho(\sigma) = -\frac{c_1}{2\tilde{\ell}^{\circ}_1}\left\{(\tilde{\ell}^{\circ}_1)^2 + (\tilde{L}^{\circ})^T_1\left[(\tilde{M})_{11} + (\tilde{M})_{1}(P-\tilde{M})^{-1}_{11}(\tilde{M})_{11}\right](\tilde{L}^{\circ})_1 + \frac{(\tilde{\ell}^{\circ}_1)^2}{c_1^2}(\tilde{C})^T_1(P-\tilde{M})^{-1}_{11}(\tilde{C})_1\right\} + (\tilde{C})^T_1(P-\tilde{M})^{-1}_{11}(\tilde{M})_{11}(\tilde{L}^{\circ})_1 . \qquad (3.5)$$

Здесь $\tilde{L}^{\circ} = \tilde{L}^{\circ}_n\big|_{\sigma_n = \sigma_\infty - \sigma}$; $\tilde{M} = \tilde{M}_n\big|_{\sigma_n = \sigma_\infty - \sigma}$; $\tilde{\ell}^{\circ}_1$ –элемент вектора $\tilde{L}^{\circ}$.

4. Предельная ограниченность (диссипативность)

В тех случаях, когда условие асимптотической устойчивости в целом не выполняется, в системе (2.2) возможно существование стационарных периодических режимов. Точная и приближенная методика анализа таких режимов на практике мало эффективна, так как она громоздка и требует априорных сведений о форме периодического процесса (число импульсов на полупериод, порядок их чередования и т.п.) или перебора всех возможных вариантов[10,11,28]. Лучшие результаты даёт методика исследования предельной ограниченности (диссипативности) автоматических систем.

Система (2.2) называется предельно ограниченной (диссипативной) в целом, если существует такое компактное множество $\mathfrak{S}$ (асимптотически устойчивое множество), что при любых начальных условиях $X_n \to \mathfrak{S}$ при $n \to \infty$. На основании дискретного аналога теоремы Т.Иошизавы система (2.2) предельно ограничена, если ограничено множество $\mathfrak{P}$, на котором функция (2.5) неотрицательна[11,29]. Оценкой асимптотически устойчивого множества $\mathfrak{S}$ является замкнутая область, ограниченная поверхностью $v_n = \mu$ $(0 < \mu = const)$, описывающей $\mathfrak{P}$[11,29,30].

Представим матрицу P квадратичной формы (2.4) в виде $P=\tilde{Q}Q$ и преобразуем уравнение (2.2), умножив его слева на Q:

$$Y_{n+1}=QH_nQ^{-1}(Y_n+QK_n^{\circ}), \quad Y_n=QE_n^{\circ}. \qquad (4.I)$$

В пространстве $E_Q^m=\{Y_n\}$ поверхность $v_n=\rho$ представляет собой сферу и, следовательно, границей асимптотически устойчивого множества 𝔖 является сфера, описанная вокруг множества 𝔅 [II,29,30]. Для систем 2-го порядка этот вывод приводит к простой графической процедуре (рис.5), что существенно упрощает исследование.

Примеры исследования устойчивости и предельной ограниченности конкретных систем читатель найдет в работах [II,I7,I8,20-22,26-30].

Литература

I. Цыпкин Я.З. Дискретные автоматические системы, их теория и перспективы развития. В сб. "Теория и применение дискретных автоматических систем". Изд-во АН СССР, I960.

2. Кунцевич В.М. Исследование переходных и установившихся процессов в релейно-импульсных и экстремальных системах с постоянным и переменным периодом регулирования. Изв. АН СССР, ОТН, Энергетика и автоматика, № 5, I96I.

3. Dorf R.C., Farren M.C., Phillips C.A. Adaptive sampling frequency for sampled-data control systems. „IRE Transactions", v. AC-7, No1, 1962.

4. Murphy G., West R.L. The use of pulse-frequency modulation for adaptive control. Proc. of the National Electronics Conference, v.XVIII, ptI, Chicago, 1962.

5. Ли С.С., Джонс Р.В. Интегральные системы управления с частотно-импульсной модуляцией. Доклад на 2-м Международном конгрессе ИФАК, М., I963.

6. Кунцевич В.М. Исследование импульсных самонастраивающихся систем и способов повышения их качества. Дисс. на соискание уч.ст. доктора техн.наук, Ин-т кибернетики АН УССР, Киев, I964.

7. Murphy G.J., Wu S.H. A stability criterion for pulse-width-modulated feedback control systems. „IEEE Transactions", v. AC-10, No1, 1965.

8.Pavlidis T., Jury E.I. Analysis of a new class of pulse-frequency modulated feedback control systems. „IEEE Tr",AC-10,1.

9.Кунцевич В.М.Импульсные экстремальные и самонастраивающиеся системы. Изд-во "Технiка",Киев,1966.

10.Чеховой Ю.Н.Динамика импульсных систем с частотно-импульсной модуляцией. Точные методы анализа,ч.I и II.В сб."Теория автоматического управления", вып.I,изд-во "Наукова думка", Киев,1966.

11.Чеховой Ю.Н.Динамика нелинейных импульсных систем управления с частотно-импульсной модуляцией. Дисс.на соискание уч. ст.канд.техн.наук,Институт кибернетики АН УССР,Киев,1966.

12.Pavlidis T. Stability of a class of discontinuous dynamical systems. „Information and Control", v.9, No3, 1966.

13.Clark J.P.C., Noges E. The stability of pulse frequency modulated closed loop control systems. „IEEE International Conventional Record", v.14, No6, 1966.

14.Чеховой Ю.Н.Нелинейный частотно-импульсный регулятор. Энергетика и электрификация,№ I,Киев,1967.

15.Гелиг А.Х.Стабилизация нелинейных систем с частотно-импульсной модуляцией. Автоматика и телемеханика, № 6,1967.

16.Дымков В.И. Об абсолютной устойчивости частотно-импульсных систем.Автоматика и телемеханика,№ 10,1967.

17.Кунцевич В.М.,Чеховой Ю.Н.Устойчивость систем с двойной (частотной и широтной) импульсной модуляцией.Автоматика и телемеханика, № 7,1967.

18.Кунцевич В.М.,Чеховой Ю.Н. Исследование устойчивости систем управления с двойной (частотной и широтной) импульсной модуляцией прямым методом Ляпунова.В сб."Сложные системы управления",вып.3,изд-во "Наукова думка",Киев,1967.

19.Бромберг П.В.Матричные методы в теории релейного и импульсного регулирования. Изд-во "Наука",М.,1967.

20.Кунцевич В.М.,Чеховой Ю.Н. Исследование устойчивости импульсных систем управления с частотно-импульсной модуляцией прямым методом Ляпунова. Автоматика и телемеханика, № 2, 1967.

21.Чеховой Ю.Н.Устойчивость систем управления с частотно-импульсной модуляцией при постоянных внешних воздействиях. В сб."Дискретные системы управления",вып.I.Изд-во "Наукова думка",Киев,1968,в печати.

22.Чеховой Ю.Н.К задаче об устойчивости в целом импульсных систем управления с частотно-импульсной модуляцией,Автоматика и телемеханика, № 6,1968.

23.Розенфельд Б.А. Многомерные пространства. Изд-во "Наука",М.,1966.

24.Айзерман М.А.,Гантмахер Ф.Р. Абсолютная устойчивость регулируемых систем.Изд-во АН СССР,М.,1963.

25.Гантмахер Ф.Р.,Якубович В.А. Абсолютная устойчивость нелинейных регулируемых систем. В трудах 2-го всесоюзного съезда по теор.и прикл.механике,вып.I,изд-во "Наука",М.1965.

26.Чеховой Ю.Н.Абсолютная устойчивость импульсных систем управления с частотной и широтной импульсной модуляцией. В сб."Теория автоматического управления",вып.I,Киев,1968.

27.Кунцевич В.М.Исследование устойчивости стационарного вынужденного режима в системах управления с частотно-импульсной модуляцией.В сб."Дискретные системы управления", вып.I, изд-во "Наукова думка",Киев,1969,в печати.

28.Чеховой Ю.Н.Динамика импульсных систем с частотно-импульсной модуляцией.Ч.Ш,В сб."Теория автоматического управления",вып.3,Киев,1967.

29.Чеховой Ю.Н.Об одной форме устойчивости нелинейных импульсных систем. Изв.АН СССР.Техническая кибернетика,№ 4, 1967.

30.Чеховой Ю.Н.Предельная ограниченность нелинейных импульсных систем. В сб."Теория автоматического управления", вып.2,Киев,1967.

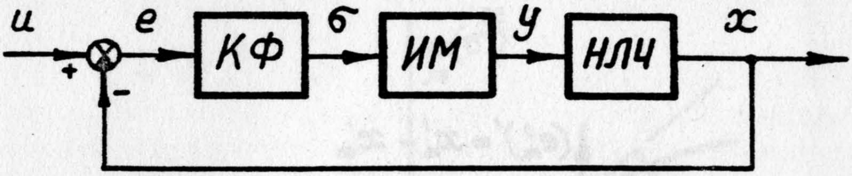

Рис.I. Структурная схема нелинейной импульсной системы автоматического управления.

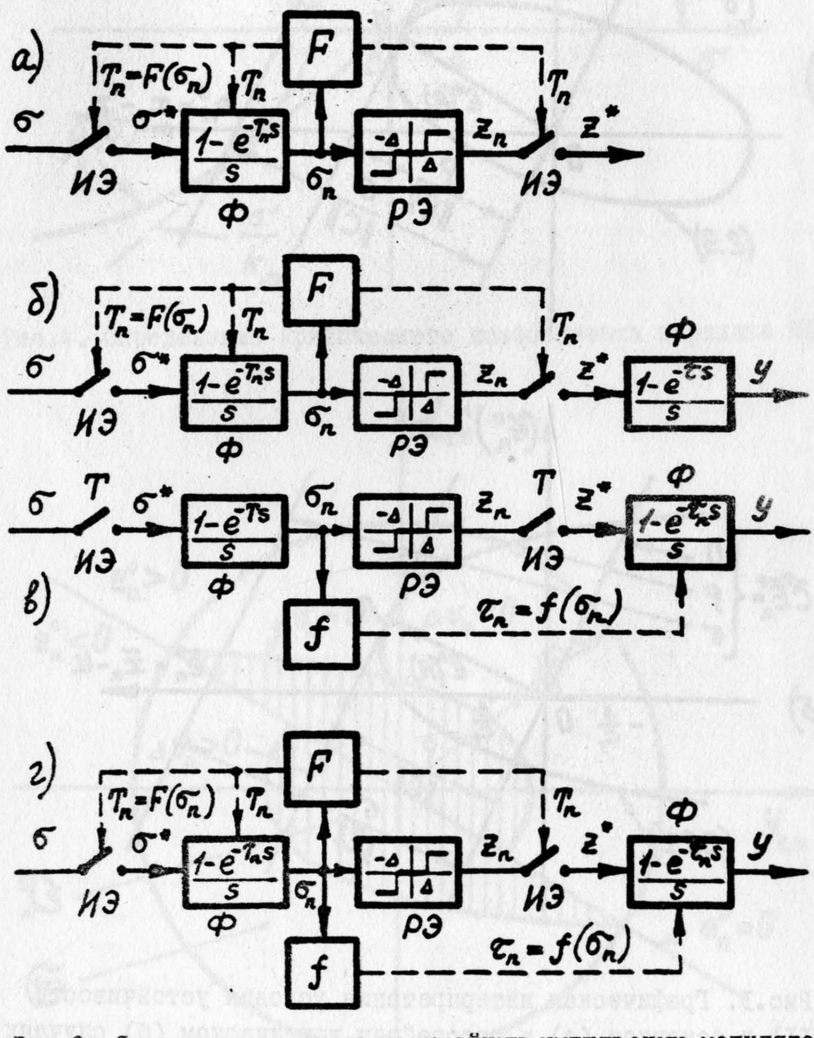

Рис.2. Структурные схемы нелинейных импульсных модуляторов: а - идеальный частотно-импульсный модулятор; б - реальный частотно-импульсный модулятор; в - широтно-импульсный модулятор; г - частотно-широтный импульсный модулятор.

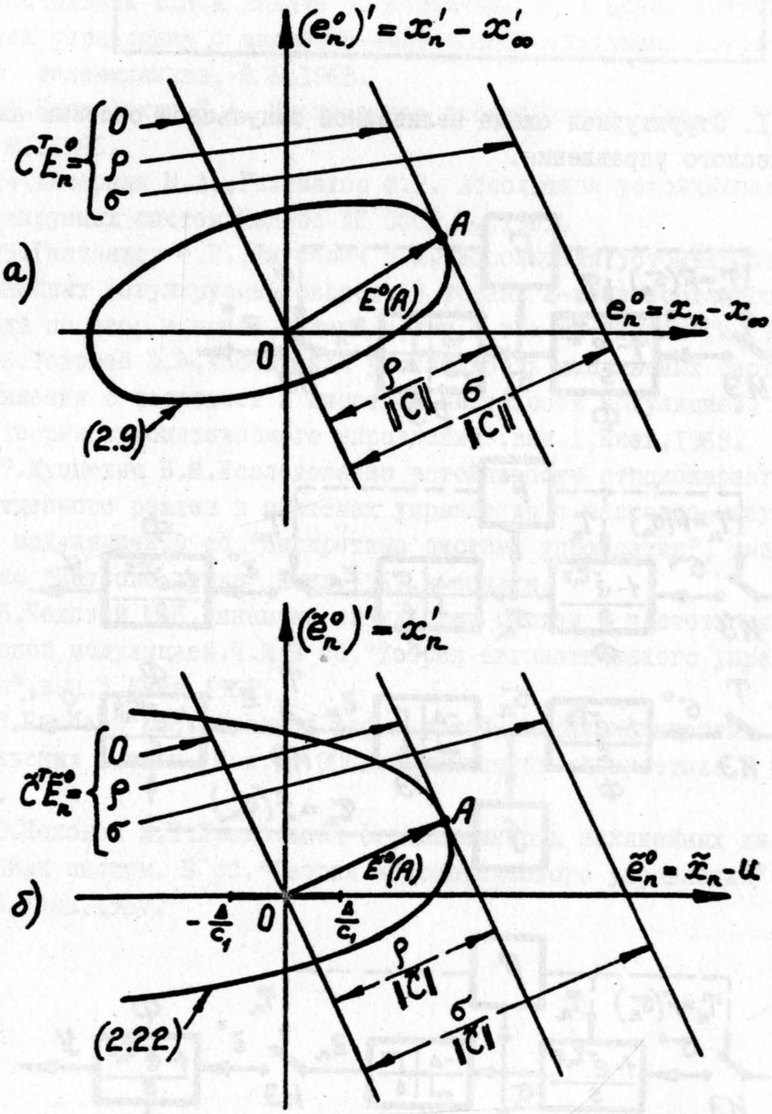

Рис.3. Графическая интерпретация условия устойчивости (2.II) в основном (а) и простейшем критическом (б) случаях.

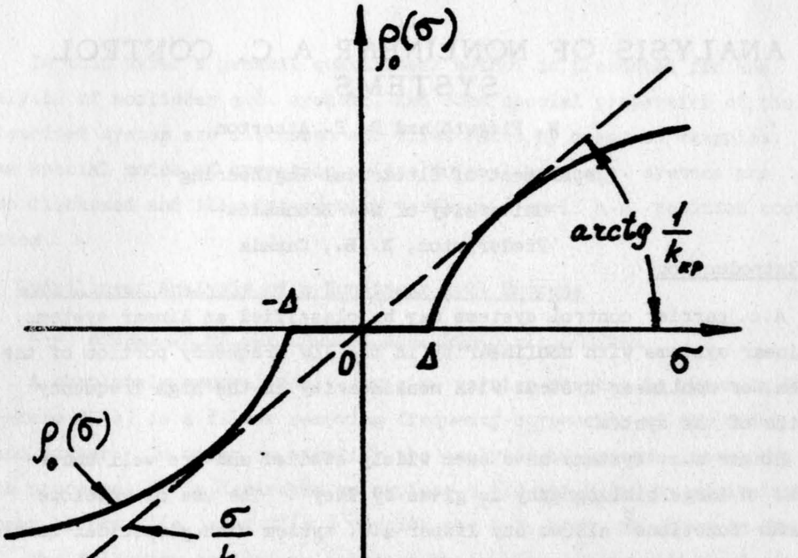

Рис.4. Определение критического коэффициента передачи НЛЧ.

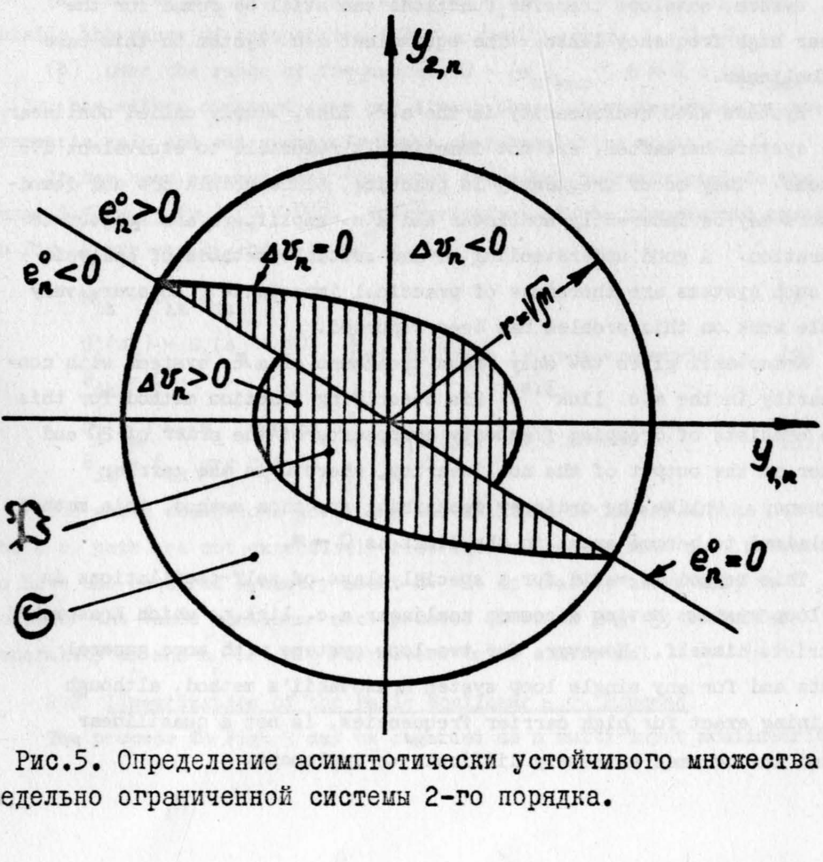

Рис.5. Определение асимптотически устойчивого множества предельно ограниченной системы 2-го порядка.

ANALYSIS OF NONLINEAR A.C. CONTROL SYSTEMS

W. Fieguth and D. P. Atherton

Department of Electrical Engineering

University of New Brunswick

Fredericton, N. B., Canada

1. Introduction

A.c. carrier control systems may be classified as linear systems, nonlinear systems with nonlinearity in the low frequency portion of the system, or nonlinear systems with nonlinearity in the high frequency portion of the system.

Linear a.c. systems have been widely studied and are well understood. A large bibliography is given by Ivey[1]. The use of envelope transfer functions[2] allows any linear a.c. system with sinusoidal carrier signals to be transformed into an equivalent d.c. system to which any of the numerous methods of linear analysis can be applied.

If nonlinearity is restricted to the low frequency portions of an a.c. system, envelope transfer functions can still be found for the linear high frequency links. The equivalent d.c. system in this case is nonlinear.

Systems with nonlinearity in the a.c. link, simply called nonlinear a.c. systems hereafter, are not immediately reducible to equivalent d.c. systems. They occur frequently in practice, since modulators and demodulators may be inherently nonlinear and a.c. amplifiers are subject to saturation. A good understanding of and suitable methods of analysis for such systems are therefore of practical importance. However, very little work on this problem has been reported.

Krasovskii gives the only known treatment of a.c. systems with nonlinearity in the a.c. link[3,4]. His describing function method for this case consists of dropping frequency components of the order of 2Ω and higher at the output of the nonlinearity, where Ω is the carrier frequency. Unlike the ordinary describing function method, this method is claimed to become exact in the limit as $\Omega \rightarrow \infty$.

This method is valid for a special class of self-oscillations in two-loop systems having a common nonlinear a.c. link to which Krasovskii restricts himself. However, for two-loop systems with more general inputs and for any single loop system Krasovskii's method, although remaining exact for high carrier frequencies, is not a quasilinear method and becomes extremely difficult to implement.

In this paper a general quasilinear method is presented for the analysis of nonlinear a.c. systems, and some special properties of the linearized system are discussed and illustrated by means of examples. Some special modes of operation unique to nonlinear a.c. systems are also discussed and illustrated with tests on a small a.c. position control system.

2. Quasilinear Analysis of a Nonlinear a.c. Process

2.1 Reduction of the General Nonlinear a.c. Process

A completely general nonlinear a.c. carrier process is shown in Fig. 1, where $G_f(s)$ is a filter removing frequency components of the order of Ω and higher. In order to simplify the analysis of systems incorporating such processes it is desirable to replace $G_1(s)$ and $G_2(s)$ by equivalent linear blocks $G_1'(s_m)$ and $G_2'(s_m)$ outside the a.c. path, as shown in Fig. 2.

The following conditions obtained by setting $a(t) = a'(t)$ and $y(t) = y'(t)$ in Figs. 1 and 2, govern the transformation and shifting of the linear blocks:

(a) Either $M_{ik} = 0$ for all $k \neq 1$ or $G_i(j\omega)$ is essentially zero outside the range of frequencies $\Omega/2 < \omega < 3\Omega/2$, where $i = 1, 2$.

(b) Over the range of frequencies $\Omega - (\omega_m)_{max} < \omega < \Omega + (\omega_m)_{max}$, $G_i(j\omega)$ has either constant gain and linear phase characteristics or even symmetric gain and odd symmetric phase characteristics about $\omega = \Omega$.

It has been assumed that the input frequency is restricted to the range $0 \leq \omega_m \leq (\omega_m)_{max} < \Omega/2$. The parameters of the transformed system in Fig. 2 are easily shown to be[5]

$$M'_{ik} = \delta_{1k} M_{ik} \qquad (1)$$

$$\left.\begin{aligned} G_i'(s_m) &= G_i(s_m + j\Omega) \\ \theta_{ik} &= 0 \end{aligned}\right\} \text{if } |G_i(j\omega)| \text{ is even symmetric} \qquad (2)$$

$$\left.\begin{aligned} G_i'(s_m) &= G_i \ \exp(ms_m) \\ \theta_{ik} &= \delta_{1k} \arg G_i(j\Omega) \end{aligned}\right\} \text{if } |G_i(j\omega)| \text{ is constant} \qquad (3)$$

The above conditions governing the shifting of linear blocks out of the a.c. path are not excessively restrictive because most a.c. networks do have the required symmetry about Ω. It is useful, therefore, to consider the basic nonlinear a.c. process shown in Fig. 3, where for generality m_1 and m_2 are not restricted to be sinusoidal.

2.2 Linearization of the Basic Nonlinear a.c. Process

The process in Fig. 3 may be regarded as a multi-input nonlinearity

$$y = g(x, m_1, m_2) = m_2\, f(xm_1) \tag{4}$$

This nonlinearity can be characterized by its equivalent gain to the input signal, x, which is given by

$$K_x = \overline{xy} / \overline{x^2} = \overline{xm_2\, f(xm_1)} / \overline{x^2} \tag{5}$$

where the bars denote time averaging. Writing (5) as an ensemble average,

$$K_x = \sigma_x^{-2} \iiint xm_2\, f(xm_1)\, p(x, m_1, m_2)\, dx\, dm_1\, dm_2 \tag{6}$$

and making the reasonable assumption that the input, x, is unrelated to the carrier signals, gives

$$K_x = \sigma_x^{-2} \int_{-\infty}^{\infty} x\, N(x)\, p(x)\, dx \tag{7}$$

where

$$N(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m_2\, f(xm_1)\, p(m_1, m_2)\, dm_1\, dm_2 \tag{8}$$

N(x) is a modified nonlinearity equivalent to the entire process in Fig. 3 and allows any nonlinear a.c. system to be reduced to an equivalent nonlinear d.c. system.

Although this completes the formulation in principle of a method of analysis of nonlinear a.c. systems, it is interesting to consider such systems in greater depth. They possess a number of properties not found in ordinary nonlinear systems, which may be of practical importance.

3. The Modified Nonlinearity

In this section some of the properties of the modified nonlinearity, given by (8), are considered.

Although it is possible to formulate expressions for the joint probability density, $p(m_1, m_2)$, of common carrier signals and to use these to evaluate N(x) from (8), it is often simpler to obtain N(x) using the time average equivalent of (8),

$$N(x) = \overline{m_2\, f(xm_1)} \tag{9}$$

where x is considered fixed during the averaging process.

Fairly simple expressions for N(x) in terms of the original nonlinearity or its equivalent gains result for common carrier signals.

For example, if the carriers are square waves,

$$m_i(t) = M_i\, \mathrm{sgn}[\cos(\Omega t + \varphi_i)] \tag{10}$$

averaging (9) over one cycle of the carrier gives

$$N(x) = M_2[(\pi + 2\varphi_1 - 2\varphi_2)/2\pi][f(M_1 x) - f(-M_1 x)] \tag{11}$$

If the nonlinearity is odd and $\varphi_1 = \varphi_2$, (11) becomes

$$N(x) = M_2 \; f(M_1 x) \tag{12}$$

That is, the form of the original nonlinearity has remained unchanged, but it has been scaled along its input and output axes.

In the much more common case of sinusoidal carriers

$$m_i(t) = M_i \; \cos(\Omega t + \varphi_i) \tag{13}$$

(9) becomes

$$\begin{aligned} N(x) &= M_2 \; \cos(\varphi_2 - \varphi_1) \; \overline{\cos \Omega t \; f(xM_1 \; \cos \Omega t)} \\ &\qquad - M_2 \; \sin(\varphi_2 - \varphi_1) \; \overline{\sin \Omega t \; f(xM_1 \; \cos \Omega t)} \\ &= M_1 M_2 x \cdot [\cos(\varphi_2 - \varphi_1) \; \mathrm{Re}\; K(xM_1) + \sin(\varphi_2 - \varphi_1) \; \mathrm{Im}\; K(xM_1)]/2 \end{aligned} \tag{14}$$

where Re K(A) and Im K(A) are the in-phase and quadrature portions of the equivalent gain of the nonlinearity f(a) to an input $a = A \cos \omega t$.

Equations (11) and (14) make the determination of N(x) trivial for sinusoidal or square wave carrier signals. Two very general properties of N(x) are discussed and illustrated below.

3.1 Odd Symmetry

The modified nonlinearity is almost always an odd function. This is seen by inspection from (11) and (14) for square wave or sinusoidal carrier signals, but can be shown to be true more generally. From (8)

$$N(-x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m_2 \; f(-xm_1) \; p(m_1, m_2) \; dm_1 \; dm_2 \tag{15}$$

Setting $\mu = -m_1$ gives

$$N(-x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m_2 \; f(x\mu) \; p(-\mu, m_2) \; d(-\mu) \; dm_2 \tag{16}$$

Comparison of (16) and (8) shows that N(x) is an odd function, that is, $N(-x) = -N(x)$, if

$$\int m_2 \; p(m_1, m_2) \; dm_2 = \int m_2 \; p(-m_1, m_2) \; dm_2 \tag{17}$$

Since m_1 and m_2 are periodic signals with the same time period, the point (m_1, m_2) traces out a closed trajectory in the m_1-m_2 plane. The condition given by (17) is equivalent to requiring this trajectory to be symmetric through the origin or about the m_2 axis. This follows because $p(m_1, m_2)$ is a line density of probability along the trajectory.

It is easily shown that a sufficient, but not necessary, condition for (17) to hold is that the Fourier series of m_1 and m_2 contain no even terms. Most carrier signals satisfy this condition, resulting in odd symmetry for N(x) regardless of the nature of the nonlinearity f(a) in Fig. 3.

This symmetry of N(x) affects the type of subharmonic oscillation which is likely to occur in a nonlinear a.c. system. Although there is no rigid rule relating nonlinearity symmetry to possible orders of subharmonic oscillations in a nonlinear d.c. system, there are strong tendencies. A system with an odd symmetric nonlinearity tends to have odd order subharmonics, and a system with a strongly asymmetric nonlinearity tends to have even order subharmonics. As a qualitative test of the symmetry of nonlinear a.c. systems an analogue computer simulation was carried out of the system shown in Fig. 4 in both its a.c. and d.c. versions, that is, with and without modulators. The d.c. system exhibits very strong, spontaneously starting second order subharmonic oscillations over a wide range of input frequencies and amplitudes. For the a.c. system no second order subharmonic could be found for any operating conditions.

3.2 Single-Valuedness

If the modulator input is sufficiently slowly varying with respect to the carrier, the modified nonlinearity, N(x), is single-valued. This is fairly obvious. A double valued nonlinearity with a modulated carrier input will impart a phase shift to the carrier, not to its slowly varying amplitude.

In analogy to (6) the quadrature portion of the equivalent gain of the basic nonlinear a.c. process in Fig. 3 is given by

$$K_{\dot{x}} = \sigma_{\dot{x}}^{-2} \iiint \dot{x}\, m_2\, f(xm_1)\, p(\dot{x}, m_1, m_2)\, d\dot{x}\, dm_1\, dm_2 = \sigma_{\dot{x}}^{-2} \int_{-\infty}^{\infty} \dot{x}\, N(x)\, p(\dot{x})\, d\dot{x} \qquad (18)$$

if $\dot{x}$ is independent of m_1 and m_2.

For sinusoidal and square wave carrier signals N(x) is given by (11) and (14), and is independent of $\dot{x}$. Therefore $K_{\dot{x}} = 0$. The same result holds for other carrier signals.

Although the only condition on x in deriving (7) and (18) was that x be independent of m_1 and m_2, it is also necessary, in the case of double valued nonlinearities, to restrict x to be slowly varying. Otherwise $f(xm_1)$ in (11) cannot be described analytically, and it is improper to consider x fixed during the averaging process in (9).

In order to determine how large the ratio of carrier to modulator input frequency must be to ensure a purely real equivalent gain of a basic a.c. process having a double valued nonlinearity, a series of direct analogue computer measurements was carried out for the case of a relay with hysteresis. The results of these measurements are shown in Fig. 5. It is seen that a frequency ratio $\frac{\Omega}{\omega_m}$ of the order of 30 is necessary to reduce the gain angle to around 10.

The theoretical equivalent gain expression for this relay computed from (7) using N(x) given by (14) is

$$K_x = 8[E'(1/X) - K'(1/X)]/\pi^2 \tag{19}$$

where E' and K' are complete elliptic integrals. This curve is also plotted in Fig. 5 and agrees well with the measured magnitudes for $\frac{\Omega}{\omega_m} \geq 10$.

It can be shown that the sinusoidal equivalent gain of an a.c. process with sinusoidal carriers and a piecewise linear nonlinearity can always be expressed as a finite series of elliptic integrals. However no such closed form solution appears possible for Gaussian input signals.

The fact that the high frequency multiplicative carrier signals modify the nonlinearity so as to make the low frequency input see only a single-valued characteristic is an interesting general property, analogous to an identical effect for additive high and low frequency inputs to a double valued nonlinearity.[6]

4. Closed Loop Behaviour of Nonlinear a.c. Systems

Predictions of the behaviour of nonlinear a.c. systems can confidently be made using the equivalent nonlinear d.c. system described above provided one can assume that the modulating signal is relatively slowly varying with respect to, and is independent of, the carrier signals.

However, as in the case of nonlinear d.c. systems, which may exhibit effects such as subharmonic oscillations, which are not predicted in a simple describing function analysis, nonlinear a.c. systems may have modes of behaviour not predicted by a quasilinear analysis of the equivalent nonlinear d.c. system if either of the above assumptions is not valid. Two such special effects which are unique to a.c. systems having nonlinearity in the a.c. link are discussed and illustrated below with tests on a small a.c. position control system.

4.1 Self-Oscillations at Submultiples of the Carrier Frequency

If the modulator input in a nonlinear a.c. system has the form

$$x(t) = X \cos[(\Omega/s)t + \theta] \tag{20}$$

then for a sinusoidal carrier the modulator output, which is the nonlinearity input, is given by

$$a(t) = M_{11}X \cos\left\{[(s + 1)/s]\ \Omega t + \theta\right\}$$
$$+ M_{11}X \cos\left\{[(s - 1)/s]\ \Omega t - \theta\right\} \qquad (21)$$

Thus the nonlinearity has two sinusoidal inputs whose frequencies are related by the ratio $(s - 1)/(s + 1)$. In general, therefore, the nonlinearity output components at the frequencies $(s + 1)\ \Omega/s$ and $(s - 1)\ \Omega/s$ are not in phase with the corresponding input terms. For the relevant large values of s these phase angles are small, but they do cause self oscillations to occur at submultiples of the carrier frequency where the phase shift through the linear elements is slightly different from 180°. Since this deviation is small the error involved in finding the frequency and amplitude of oscillation assuming unsynchronized behaviour is slight. These carrier-synchronized oscillations have been observed in analogue computer simulations with values of s as high as 40. They are similar to the ripple-instability oscillations which have been observed in pulse modulation systems[7], but are of much higher order.

4.2 <u>Low Frequency Output Response Produced by High Frequency Inputs</u>

In a linear a.c. system with no dynamic elements in the a.c. path the modulation and demodulation together are equivalent to a multiplicative sampling signal

$$s(t) = m_1(t)\ m_2(t) = M_1M_2[\cos\varphi + \cos(2\Omega t + \varphi)]/2 \qquad (22)$$

where the carrier signals are given by (13) with $\varphi_1 = 0$ and $\varphi_2 = \varphi$. In analogy to a sampled data system[8] a low frequency output signal is obtained for inputs ω_m near 2Ω.

This effect extends to lower input frequencies in a nonlinear a.c. system due to the production by the nonlinearity of harmonics and intermodulation frequencies. In particular if $x(t)$ and $m_1(t)$ are assumed sinusoidal the nonlinearity input is the sum of two sinusoids and its output $b(t)$ can be expressed[6] in a series form. Inspection of the series reveals that $b(t)$ contains output frequencies near Ω, which yield low frequency terms on demodulation, for input frequencies

$$\omega_m \approx \left|\frac{s \pm k - 1}{- s \pm k}\right| \Omega \qquad (23)$$

where α_{sk} is the output coefficient for the nonlinearity. To clarify the above, suppose a hypothetical nonlinearity is such that $\alpha_{10} = \alpha_{01} = 1$ and α_{11} is the only other non-zero output coefficient. Then putting $s = k = 1$ in (23) gives $\omega_m \approx \Omega/2$. That is, input frequencies near $\Omega/2$

to an a.c. system containing the above nonlinearity will give rise to low frequency outputs with amplitude dependent on α_{11}. Figure 6 shows a sketch of the form of the frequency response magnitude of such a system with an arbitrary linear plant.

The above effect has been clearly demonstrated by analogue computer simulations. For common hard spring nonlinearities low frequency outputs are prevalent for input frequencies ω_m near $2\Omega/(2n + 1)$ and $\Omega/2n$ for integers n up to 4 or 5.

5. Tests on an Acutal a.c. System

A block diagram of the small a.c. position control system is shown in Fig. 7. The angles denote the phases of the carrier signals, whilst the nonlinear element and the parameters α and k_T were open to choice. The transfer function G(s) was added to provide an additional time constant so that self oscillations occurred for low values of k_T. No attempt was made to determine a sophisticated nonlinear model for the servomotor as most of the tests performed were of a qualitative nature. The linear transfer function given for the motor in Fig. 7 was determined from a frequency response test in which a control phase voltage comparable to that occurring in the self oscillation experiment was used. A simplified block diagram of the system, with a saturation nonlinearity as used in all the tests, is shown in Fig. 8.

5.1 Step Response with an Asymmetrical Nonlinearity

The response of the system of Fig. 8, with parameters of $b_1 = 1.5$, $b_2 = 0.5$ and $K = 5$ for the asymmetrical saturation and $\alpha = 50$, $k_T = 1.0$ for the linear elements, is given in Fig. 9 for step inputs of approximately 80° in either direction. This symmetrical output may be compared with the step response in Fig. 10 of an approximately equivalent nonlinear d.c. system obtained by analogue simulation.

5.2 Self Oscillations

The system was set up with the parameters $b_1 = b_2 = 0.03$, $K = 200$, $\alpha = 24$ and $k_T = 0$. The resulting self oscillation was found to have an amplitude of 34° at a frequency of 3.0 Hz, and was shown to be phase locked to the carrier at a frequency $\Omega/20$. This compares with a theoretical solution of 28° at 3.34 Hz obtained using the theory presented earlier which assumes the self oscillation and carrier are unrelated. As the conditions for shifting given in section 2.1 are satisfied by G(s) in this example it was replaced by its equivalent transfer function

$$G'(s_m) = \frac{\alpha}{s_m + \alpha} \qquad (25)$$

preceding the modulator, whilst the equivalent gain of the resulting nonlinear a.c. process was calculated from (14) and (7).

The parameter α was varied in order to study the locking phenomenon in more detail. It was found that the oscillation remained at the frequency $\Omega/20$ for α in the range 22.9 to 24.9, which corresponds to a change in phase of about 2.5° for a signal of frequency $\Omega/20$ through the linear transfer function of the system. As a first approximation this can be taken to be the maximum possible angle of the describing function for $s = 20$.

The relatively large discrepancy between the predicted and measured frequencies of oscillation given above is probably due to inaccuracies in the transfer functions used to model the system components. This is supported by analogue computer studies where differences between measured and computed values are found to be compatible with a describing function angle of 2 or 3 degrees.

5.3 Low Frequency Responses to High Frequency Inputs

To demonstrate the behaviour discussed in section 4.2 an electrical demodulator and modulator were added to the system immediately preceding the saturation nonlinearity. A narrow band Gaussian noise near 40 Hz, that is $2\Omega/(2n + 1)$ with $n = 1$, was applied to the input of the electrical modulator. Figs. 11 and 12 show the measured low frequency random outputs of the system with parameters $b_1 = b_2 = 2$, $K = 4$, $\alpha = 100$ and $k_T = 1$ for different noise levels.

For low levels of noise there is little or no output. This is expected for a system with a saturation characteristic because the effect is nonlinear and an input large enough to cause strong saturation is required to generate the harmonics and intermodulation components which cause the low frequency output.

6. Conclusions

A method has been presented for the analysis of a.c. systems with nonlinearity in the a.c. link. Dynamic elements in the a.c. link are transformed and shifted to the low frequency part of the system. The remaining nonlinear a.c. process is equivalent to a single nonlinearity. Thus a nonlinear a.c. system can be reduced to a nonlinear d.c. system, allowing exact or quasilinear analyses by known methods.

A number of unique properties of nonlinear a.c. systems have been discussed and illustrated. These include inherent symmetry of the system, a tendency of self oscillations to synchronize with the carrier, and the low frequency response of such a system to high frequency inputs, which can render it less immune to noise than expected.

7. Acknowledgements

The authors with to acknowledge the partial support of this work by the National Research Council of Canada through Grant No. A1646.

8. References

1. Ivey, K. A., "A-C Carrier Control Systems", Wiley, New York; 1964.
2. Weiss, G., "A Network Theory for Carrier Suppressed Modulated Systems", I.R.E. Trans. on Automatic Control, Vol. AC-6, pp. 54-56; February 1961.
3. Krasovskii, A. A., "On the Theory of Two-Channel Follow Up Systems with a Relay in the A-C Circuit", Automation and Remote Control, Vol. 21, No. 9, pp. 911-919; September, 1960.
4. Krasovskii, A. A., "A Two-Channel Servo System with Antisymmetric Feedback in the Case of Random Disturbances", Automation and Remote Control, Vol. 22, No. 2, pp. 122-133; February, 1961.
5. Fieguth, W., "Multi-input Quasilinearisation", Ph.D. Thesis, University of New Brunswick, June 1968.
6. Atherton, D. P. and Turnbull, G. F., "Response of Nonlinear Characteristics to Several Inputs and the Use of the Modified Nonlinearity Concept in Control Systems", Proc. I.E.E., Vol. 111, No. 1, pp. 157-164; January, 1964.
7. Fallside, F., "Ripple Instability in Closed-Loop Pulse-Modulation Systems Including Invertor Drives", Proc. I.E.E., Vol. 15, No. 1, pp. 218-228; January, 1968.
8. Kuo, B. C., "Analysis and Synthesis of Sampled-Data Control Systems", Prentice-Hall, Englewood Cliffs, N. J., 1963; p. 222.

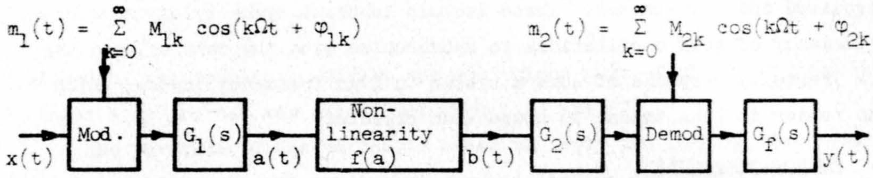

Figure 1 General Nonlinear a.c. Carrier Process.

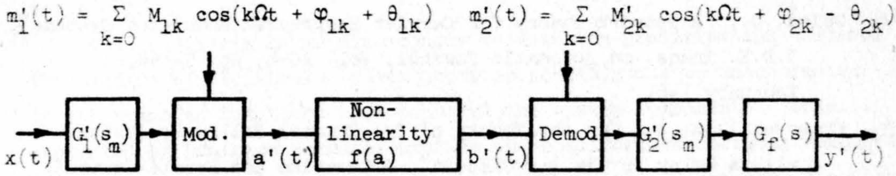

Figure 2 Nonlinear a.c. Carrier Process Equivalent to that in Fig. 1.

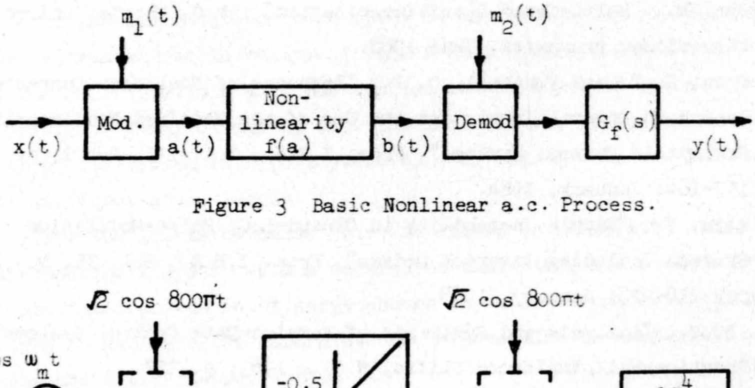

Figure 3 Basic Nonlinear a.c. Process.

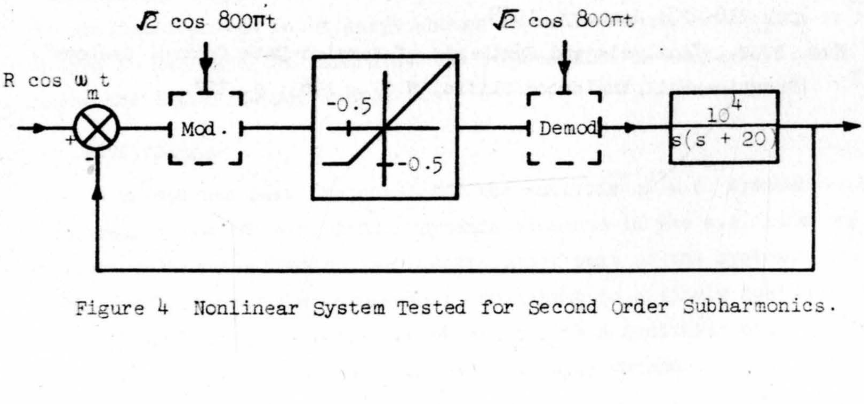

Figure 4 Nonlinear System Tested for Second Order Subharmonics.

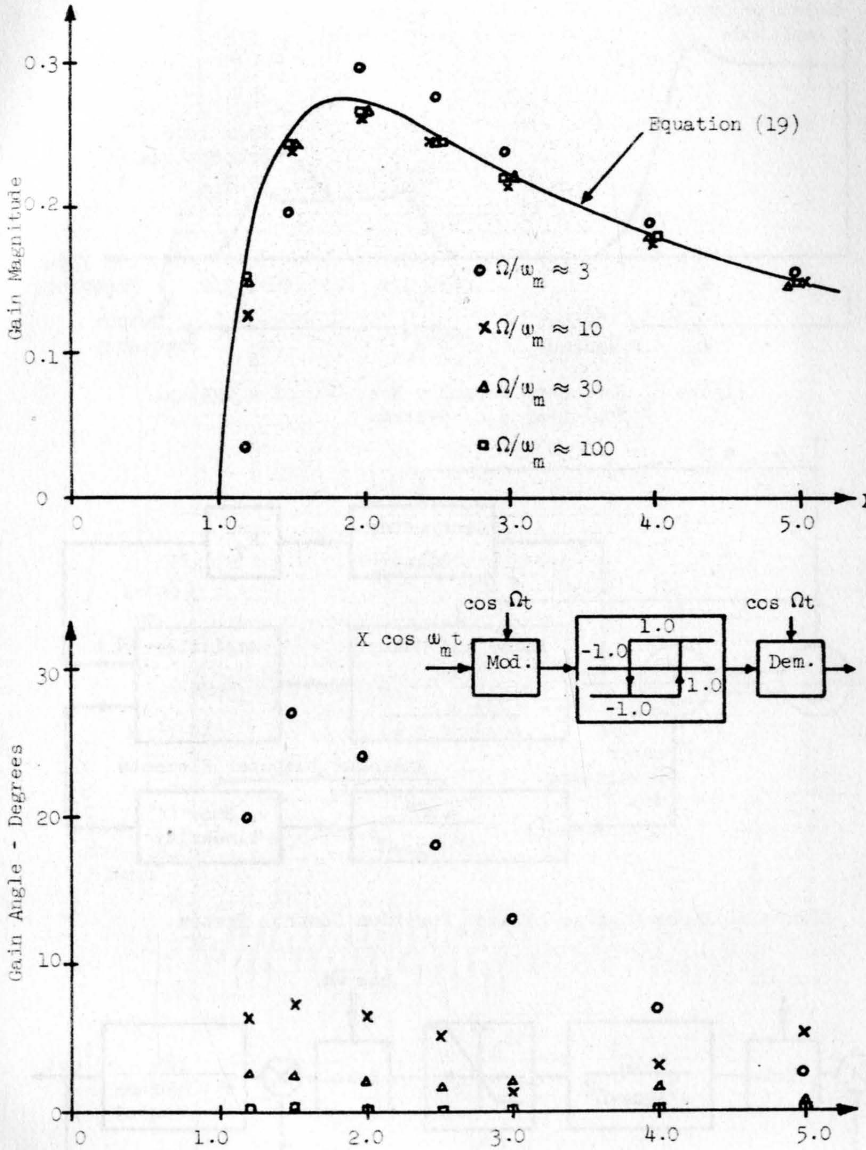

Figure 5 Measured and Computed Equivalent Gains of an a.c. Process Having a Double Valued Nonlinearity.

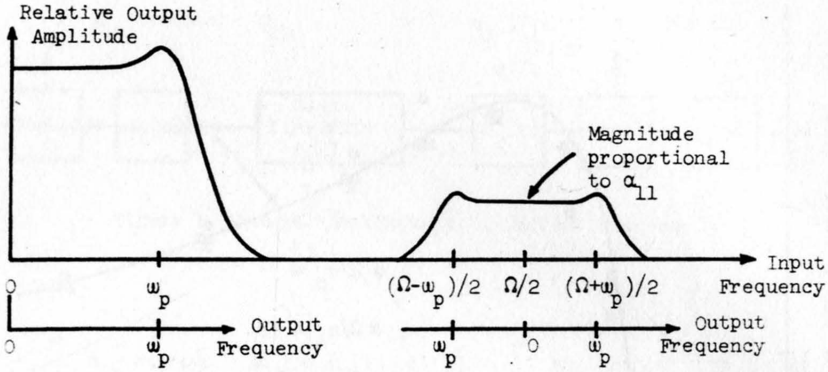

Figure 6 Frequency Response Near $\Omega/2$ of a Typical Nonlinear a.c. System.

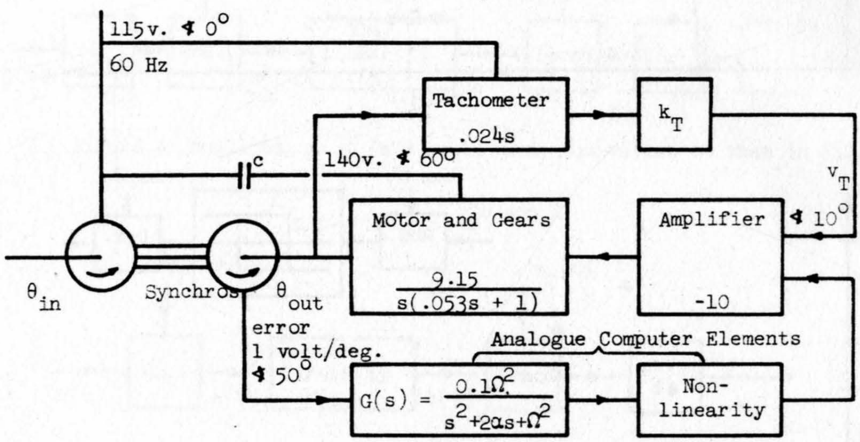

Figure 7 Block Diagram of a.c. Position Control System.

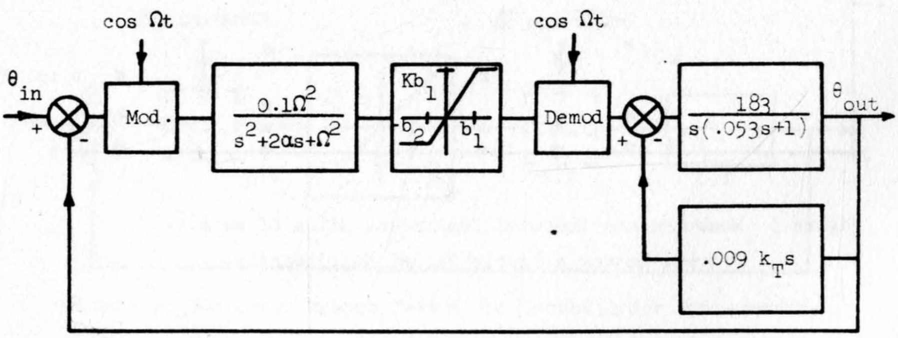

Figure 8 Simplified Block Diagram of the a.c. Position Control System in Fig. 7.

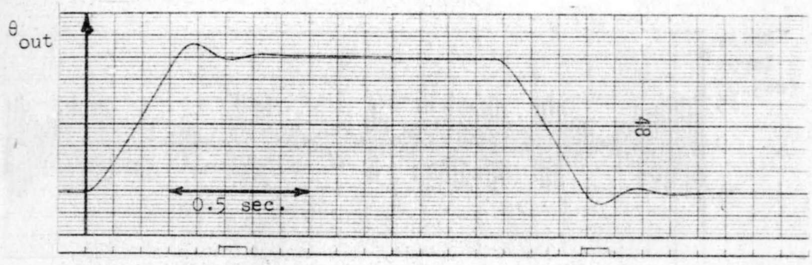

Figure 9 Step Response of Asymmetric Nonlinear a.c. System.

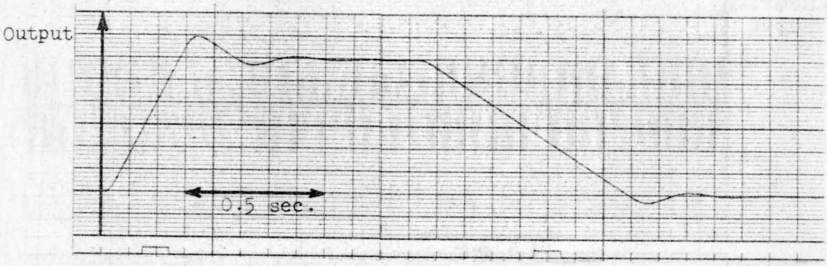

Figure 10 Step Response of Asymmetric Nonlinear d.c. System.

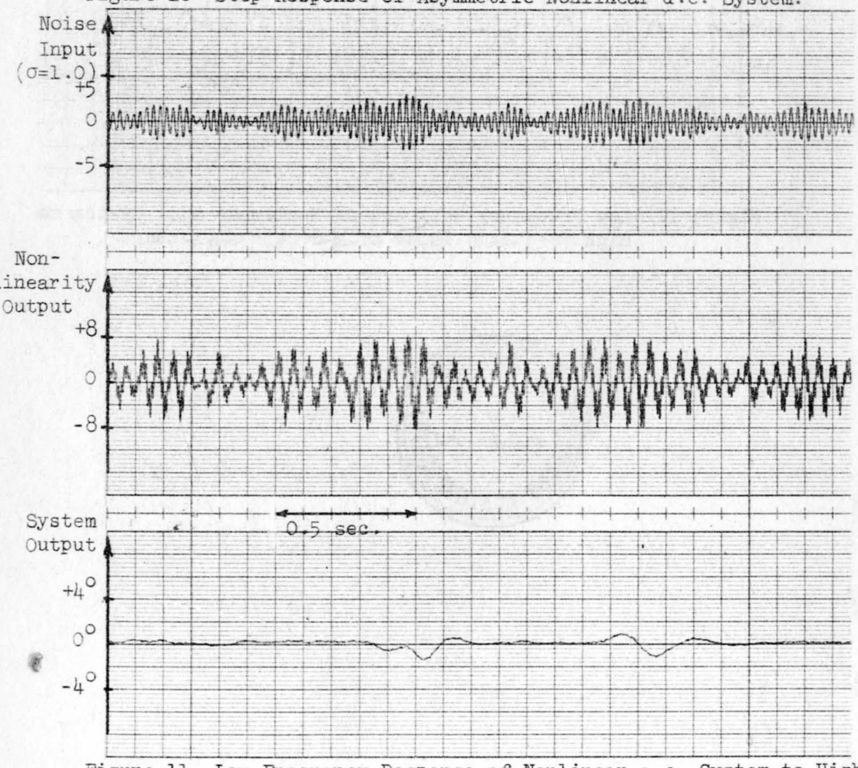

Figure 11 Low Frequency Response of Nonlinear a.c. System to High Frequency Noise of Low Amplitude.

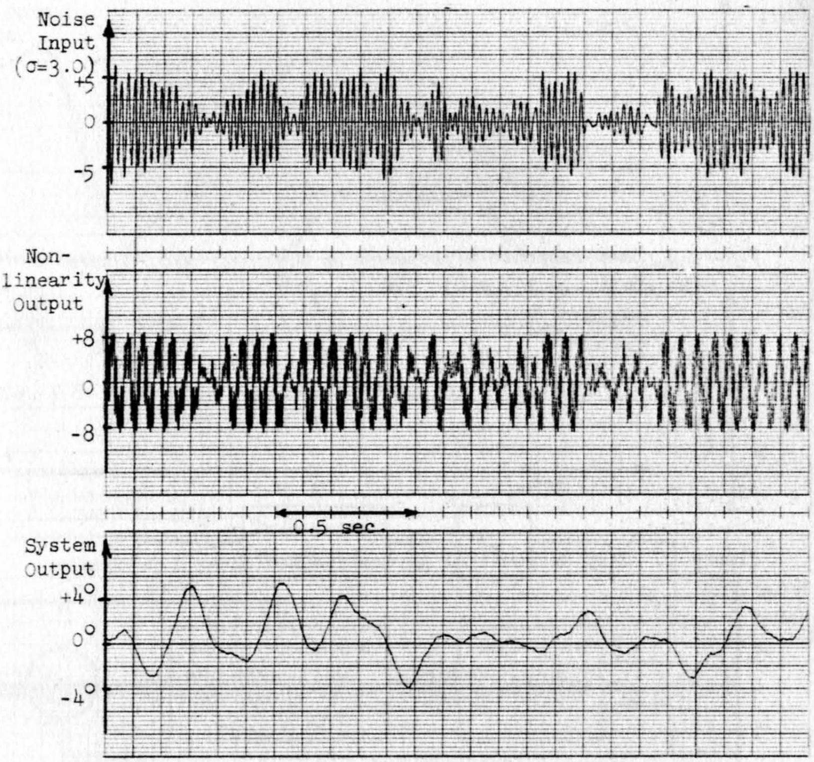

Figure 12 Low Frequency Response of Nonlinear a.c. System to High Frequency Noise of Moderate Amplitude.