NATIONAL ADVISORY COMMITTEE FOR AERONAUTICS

REPORT 1391

INVESTIGATION OF A NONLINEAR CONTROL SYSTEM

By I. FLÜGGE-LOTZ, C. F. TAYLOR, and H. E. LINDBERG

1958

For sale by the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C. Yearly subscription, $10; foreign, $13; single copy price varies according to size - - - - - - - - - Price 35 cents

REPORT 1391

INVESTIGATION OF A NONLINEAR CONTROL SYSTEM

By I. FLÜGGE-LOTZ, C. F. TAYLOR, and H. E. LINDBERG

Stanford University

National Advisory Committee for Aeronautics

Headquarters, 1512 H Street NW., Washington 25, D. C.

Created by Act of Congress approved March 3, 1915, for the supervision and direction of the scientific study of the problems of flight (U. S. Code, title 50, sec. 151). Its membership was increased from 12 to 15 by act approved March 2, 1929, and to 17 by act approved May 25, 1948. The members are appointed by the President and serve as such without compensation.

(Rev. 6-14-56)

REPORT 1391

INVESTIGATION OF A NONLINEAR CONTROL SYSTEM [1]

By I. Flügge-Lotz, C. F. Taylor, and H. E. Lindberg

SUMMARY

Nonlinear elements are sometimes added to linear control systems in order to improve the response of the system to an arbitrary input. This can be done in different ways, one of them being the variation of the coefficients of the differential equation describing the system before the nonlinear elements are added. This variation of the coefficients may be done in a continuous or in a discontinuous way. In the present paper a discontinuous variation of the coefficients is studied in detail and investigated for practical use.

The nonlinear feedback is applied to a second-order system. From former analytical considerations the process of control is visualized as establishing an ensemble of linear second-order differential equations (some with stable and some with unstable homogeneous solutions) and switching from one equation to another so as to maintain small instantaneous error for relatively arbitrary inputs. Physically, this control process is realized with a linear second-order control system to which have been added possible discrete combinations of proportional and derivative feedback. The particular combination of feedback employed at any instant is determined by a feedback switching circuit which is in turn operated by sensed binary information obtained from the output, output derivative, error, and error derivative (namely, the signs of these variables). Techniques that are common to the digital computer field are used to implement this switching circuit.

Once physical realization is completed, simulation techniques are used to study and evaluate the performance of the nonlinear control system and to compare it with a linear system for a wide variety of inputs. A detailed quantitative study of the influence of relay delays and of a transport delay is also given. In addition, the effects of physical imperfections that are likely to be encountered in any application of the control theory are considered (e. g., velocity and acceleration limits).

An analysis of the experimental results shows that this type of nonlinear control system performs better than a linear control system having a natural frequency 15 times greater. For this comparison, performance is evaluated in terms of the average value of the magnitude of the instantaneous error for band-limited inputs. Further, in contrast with the linear system, the nonlinear system performance is virtually independent of variation in the damping factor of the system.

A preliminary extension of this type of nonlinear control concept to higher order systems is presented. Experimental results are given for a third-order system. These results show that just as in the second-order case the nonlinear system performance is better than that of a comparable linear system.

INTRODUCTION

With the demand for more exacting performance, more emphasis has been placed on nonlinear aspects of control systems. The term "control systems" can be interpreted to include active networks and feedback amplifiers as well as servomechanisms. From the standpoint of analysis, unintentional nonlinearities have to be taken into account to explain performance. From the standpoint of synthesis, intentional nonlinearities have been introduced to improve performance. However, up to this date only in special cases have advancements been obtained in the field of nonlinear control systems.

The designs of nonlinear control systems have inherent advantages. One advantage is that the response of a nonlinear system at a certain time can be made less dependent upon past response than can a linear system of comparable power-handling capability.[2] This means that the nonlinear system can be made to follow more arbitrary classes of inputs with less dynamic error than the comparable linear system. Another advantage is that the mathematical difficulties encountered may actually be conducive to consideration of more realistic criteria of performance. In the nonlinear realm it is essentially as easy to invoke a criterion such as the minimization of instantaneous error for nonstationary random inputs as it is to invoke the largest possible flat amplitude response for sinusoidal inputs.

In the present paper a control system of second order, which was first suggested by Flügge-Lotz and Wunch on the basis of analytical studies (ref. 1), is investigated. The physical realization of this system and its performance are studied in great detail by Dr. I. Flügge-Lotz and Dr. C. F. Taylor. A detailed quantitative study of the influence of relay delays and of a transport delay was made by Dr. H. E. Lindberg and is given in appendix A.

This investigation was conducted at Stanford University under the sponsorship and with the financial assistance of the National Advisory Committee for Aeronautics. The authors wish to thank Dr. A. M. Peterson of the Electrical Engineering Department of Stanford University for his continued interest and his most helpful advice on the

[1] Supersedes NACA TN 3826, "Investigation of a Nonlinear Control System" by I. Flügge-Lotz and C. F. Taylor, 1957.

[2] In linear theory, the impulse response or the autocorrelation function of the system gives an indication of how past response is weighed.

electronic problems which were encountered during this investigation. They also wish to thank Dr. G. S. Bahrs for his useful suggestions for a special transistor switching circuit.

SYMBOLS

A	peak-to-peak amplitude of input
a,b,c	parameters defining a system
a^+,b^+,c^+	constants used in appendix B
$\bar{a},\bar{b},\bar{c}$,	parameters in differential equation for control servo
B	viscous damping of motor and reflected load referred to motor shaft
D	linear damping factor
$E=y-x=-e$	
e	instantaneous error, $x-y$
f	function
g_i,g_o	input and output of Padé circuit, respectively
$H(p)$	transfer function, $1/(\alpha p+1)$
I	inertia of motor rotor, gears, and reflected load
$i=\sqrt{-1}$	
K_1,K_2	gain constants
k_m	constant of proportionality between output velocity and back electromotive force
M	constant depending on initial conditions
p	operator, indicating differentiation with respect to independent variable; that is, d/dt if real time variable is used and $d/d\tau$ if nondimensional time variable is used
T	repetition rate or period
t	time
t_d	switching delay
t_{md}	maximum allowable switching delay
V	input voltage
x	input into system
y	output from system
$\bar{y}$	approximation of output
α	symbol used to denote different constants
$\beta_m=-{}_1\beta \text{ sgn } (y'e)-{}_2\beta \text{ sgn } (y'e')$	
$(\beta_m,\gamma_n)_{min}$	smallest values of parameters giving good nonlinear system performance
${}_1\beta,{}_2\beta,{}_1\gamma,{}_2\gamma$	positive constants
$\gamma_n=-{}_1\gamma \text{ sgn } (ye)-{}_2\gamma \text{ sgn } (ye')$	
δ	positive constant
ϵ	small positive quantity
μ,ν	coordinates introduced in appendix E
ρ	radius of curvature
σ	real part of complex frequency variable $\sigma+i\omega$
τ	nondimensional time variable normalized with respect to ω_ν, $\omega_\nu t$
Ω	nondimensional frequency, ω/ω_ν
ω	frequency
ω_ν	natural frequency of undamped linear system
$\text{sgn } (f)=f/\lvert f\rvert$	
$\overline{(\ )}$	time average
$(\ )'$	$d(\ )/d\tau$
$(\ )''$	$d^2(\ )/d\tau^2$
$\overset{o}{=}$	equality sign in equations which describe operations (see eqs. (2) and (3))

Subscripts:

as	at an actual switch point
d	ideal or desirable
e	error
i	initial or input
im	image
L	limit
lin	linear
$m,n=0,1,2,3$	
max	maximum or upper bound
min	minimum or lower bound
$nonlin$	nonlinear
o	optimum, output

REVIEW OF LINEAR CONTROL THEORY

It is desirable to obtain from linear control theory some useful concepts that can be generalized to the nonlinear case. These concepts are:

(1) Operational notation
  (a) Transfer functions
  (b) Block-diagram representations
(2) Control criteria
(3) Control through parameters

At the onset second-order systems are considered. However, there is no difficulty in extending these concepts to higher order systems.

OPERATIONAL NOTATION, TRANSFER FUNCTIONS, AND BLOCK DIAGRAMS

Consider a physical process or situation in which the output is described in terms of the input as

$$a\frac{d^2y}{dt^2}+b\frac{dy}{dt}+cy=x(t) \tag{1}$$

where a, b, and c are constants, $y=y(t)$ is the output, and $x=x(t)$ is the input.

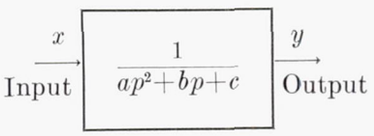
Utilizing the operator $p=d/dt$, equation (1) may be written

$$(ap^2+bp+c)y\overset{o}{=}x \tag{2}$$

(Eq. (2) reads "(ap^2+bp+c) operating on y equals operationally x.") Formal solution of equation (2) for the ratio of output over input yields by definition the operational transfer function for the system. Thus

$$\frac{y}{x}\overset{o}{=}\frac{1}{ap^2+bp+c} \tag{3}$$

The operational block diagram for the system is obtained by placing inside a box the transfer function, equation (3). Coming into the box is the input; going out of the box is the output (see sketch a).



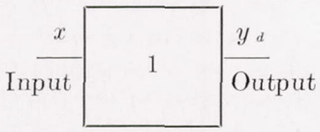
Sketch a

Here transfer functions and block diagrams are utilized merely as shorthand operational notations for differential equations. This is opposed to the Laplace transformation

viewpoint where transfer functions (and thus block diagrams) have the properties of functions of the complex frequency variable, $p=\sigma+i\omega$. The reason for stressing this interpretation is that shorthand (operational) notation has proven useful in the transition to nonlinear control whereas the Laplace transformation viewpoint (e. g., synthesis in the complex frequency plane in terms of poles and zeros) has not.

CONTROL CRITERIA

To gage the performance of an actual system an ideal or desirable system is usually established as a straight through connection (i. e., $y_d=x$ as denoted in sketch b, a block diagram of an ideal system).



Sketch b

Comparison between the desired output y_d and actual output y is accomplished by utilizing the instantaneous error:

$$e=(y_d-y)=(x-y)$$

A control criterion or criterion of performance is defined as the minimization of some property of the instantaneous error e for a given class of inputs. The minimizing process can be exact (i. e., resulting from a variational formulation of the problem) or approximate.

CONTROL THROUGH PARAMETERS

In linear systems the process of control is usually physically obtained by applying feedback and/or compensation to the system that is to be controlled. A control criterion is realized (as closely as possible) by adjustment of these applied quantities. The concept of control through parameters is an interpretation of this control process in terms of the differential equation describing the process. A simple example illustrates this concept.

Consider the position control servo shown in figure 1. The uncontrolled (open-loop) system consists of an amplifier, armature-controlled motor, gear train, and load. Closed-loop operation is obtained by utilizing proportional and derivative feedback. The gain constants K_1 and K_2 are adjustable. Armature inductance has been neglected.

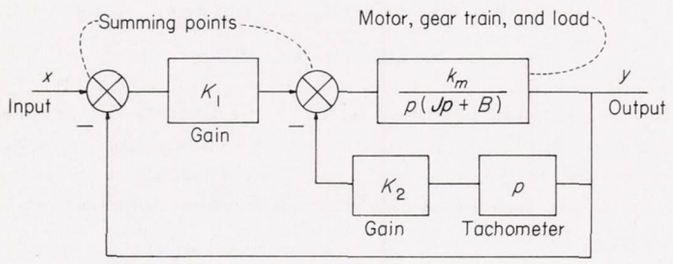


FIGURE 1.—Block diagram of simple positional servo. J, inertia of motor rotor, gears, and reflected load; B, viscous damping of motor and reflected load referred to motor shaft; k_m, constant of proportionality between output velocity and back electromotive force (it includes armature resistance and gear ratio from motor shaft to load).

From the block diagram the differential equation for the open-loop system may be written:

$$\left(\frac{I}{K_1k_m}\right)\frac{d^2y}{dt^2}+\left(\frac{B}{K_1k_m}\right)\frac{dy}{dt}=x(t) \tag{4}$$

Similarly, the closed-loop differential equation is

$$\left(\frac{I}{K_1k_m}\right)\frac{d^2y}{dt^2}+\left(\frac{B+K_2k_m}{K_1k_m}\right)\frac{dy}{dt}+(1)y=x(t) \tag{5}$$

In either case the differential equation is of the form

$$\bar{a}\,\frac{d^2y}{dt^2}+\bar{b}\,\frac{dy}{dt}+\bar{c}y=x(t) \tag{6}$$

where a set of three parameters $\bar{a}$, $\bar{b}$, and $\bar{c}$ completely characterizes the system. It is possible, then, to view the process of control in terms of these parameters. One starts with a parameter set (a,b,c) defining the uncontrolled system. A control criterion yields an optimum parameter set $(a,b,c)_o$. Control (feedback and/or compensation) is introduced ideally making it possible to adjust (a,b,c) to $(a,b,c)_o$. In the above example the gain constants K_1 and K_2 afford this adjustment.

This adjustment of the coefficients may be done in a continuous or a discontinuous way. In reference 2 Schmid and Triplett have described an interesting and efficient way to vary the coefficients of a basically linear system continuously.

NONLINEAR CONTROL

TRANSITION TO NONLINEAR CONTROL

In the preceding section it has been mentioned that the process of linear control of second-order systems may be visualized as the adjustment of the parameter set (a,b,c) to the set $(a,b,c)_o$. The term optimum was used in the sense that some criterion of performance was approached as closely as possible.

It still seems logical in the transition to nonlinear control to hypothesize control through parameters. The transition is obtained by allowing the parameters to become functions of the output $y(t)$ and the input $x(t)$; that is,

$$a\rightarrow a(x,y)$$
$$b\rightarrow b(x,y)$$
$$c\rightarrow c(x,y)$$

The mathematical description of the system is now

$$a(x,y)\,\frac{d^2y}{dt^2}+b(x,y)\,\frac{dy}{dt}+c(x,y)y=x(t) \tag{7}$$

Without knowing the specific nature of the functions $a(x,y)$, $b(x,y)$, and $c(x,y)$ it may be seen that equation (7) is a nonlinear, inhomogeneous, and/or nonautonomous differential equation. Mathematically, little in general can be said about the solution of equation (7) given the function set $a(x,y)$, $b(x,y)$, and $c(x,y)$. It seems, then, even more hopeless to attempt a synthesis problem which involves both finding the function set $[a(x,y), b(x,y), c(x,y)]_o$ for a specified control criterion and then physically realizing the system described mathematically.

NONLINEAR CONTROL THEORY

One analytical attack on the nonlinear control-system-synthesis problem has been made by Flügge-Lotz and Wunch (refs. 1 and 3 to 5). They suggested varying the coefficients a, b, and c, not continuously, but discontinuously. That means that for $t_1<t<t_2$ there is one set of coefficients, for $t_2<t<t_3$ there is another set of coefficients, and so on. The different sets of coefficients are chosen in advance, but the times t_i for change from one set to another are determined by the value and the decrease or increase of the deviation $(x\text{-}y)$. In other words, the system is linear in any interval $t_i<t<t_{i+1}$, but is nonlinear in the whole. The transition at any switching time t_i occurs with continuous values of $y(t)$ and dy/dt, but discontinuous values of (d^2y/dt^2).

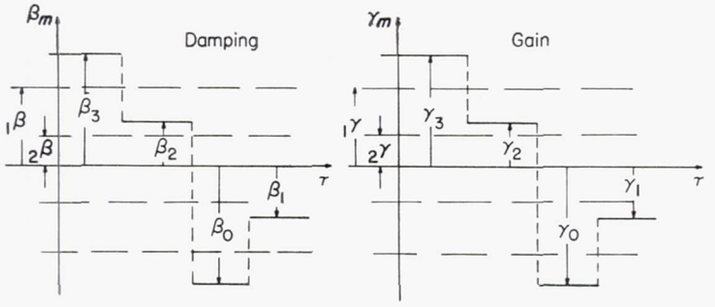


FIGURE 2.—Illustration of stepwise nature of parameters β_m and γ_n.

Phase-plane techniques were used for studying appropriate sets of coefficients and the appropriate dependence of the switching times on the deviations.[3] The authors of references 1 and 3 to 5 succeeded in finding a switching rule which assures good performance in a multitude of cases. Their control system is mathematically described in the following way:

$$\frac{d^2y}{d\tau^2}+2D(1+\beta_m)\frac{dy}{d\tau}+(1+\gamma_n)y=x(\tau) \tag{8}$$

where

$x(\tau)$ input

$y(\tau)$ output

D linear damping factor (when $\beta_m=\gamma_n=0$)

τ nondimensional time variable normalized with respect to ω_ν; that is, $\tau=\omega_\nu t$

ω_ν natural frequency of undamped linear system, $D=\beta_m=\gamma_n=0$

$\beta_m=-{}_1\beta \operatorname{sgn}(y'e)-{}_2\beta \operatorname{sgn}(y'e')$

$m=0, 1, 2, 3$

$\gamma_n=-{}_1\gamma \operatorname{sgn}(ye)-{}_2\gamma \operatorname{sgn}(ye')$

$n=0, 1, 2, 3,$

$_1\beta,{}_2\beta,{}_1\gamma,{}_2\gamma$ positive constants

$$\operatorname{sgn}(f)=\frac{f}{|f|}=\begin{cases}+1 \text{ for } f>0\\ -1 \text{ for } f<0\end{cases}$$

[3] For details the reader is referred to references 1 and 3 to 5. Reference 1 contains the ideas but is so condensed that the inquisitive reader will find it useful to read references 3 to 5, of which reference 5 is probably the most accessible. Figure 3, p. 12, and figure 30, p. 70, of reference 5 will help in getting acquainted with the phase-plane trajectory of an output. Some of the original studies are described again later in the present paper when the performance of the system is discussed.

e instantaneous error, $e=(y_d-y)=(x-y)$

$(\ )'=d(\ )/d\tau$

Actually, equation (8) is a normalized form of the control equation derived by Flügge-Lotz and Wunch. However, the notation has been somewhat changed. See appendix B for a comparison of notations and the normalization involved.

The subscript convention is

$$\left.\begin{aligned}
&\beta_3={}_1\beta+{}_2\beta && \gamma_3={}_1\gamma+{}_2\gamma\\
&\beta_2={}_1\beta-{}_2\beta && \gamma_2={}_1\gamma-{}_2\gamma\\
&\beta_1=-{}_1\beta+{}_2\beta=-\beta_2 && \gamma_1=-{}_1\gamma+{}_2\gamma=-\gamma_2\\
&\beta_0=-{}_1\beta-{}_2\beta=-\beta_3 && \gamma_0=-{}_1\gamma-{}_2\gamma=-\gamma_3
\end{aligned}\right\} \tag{9}$$

PROPERTIES OF EQUATION (8)

The following properties of equation (8) are noteworthy:

(1) Equation (8) is a piecewise linear but overall nonlinear differential equation.

(2) The parameters β_m and γ_n are stepwise switching functions of time (their implicit variable). This property is illustrated in figure 2.

(3) The time of switching and the particular combination of the parameters β_m and γ_n employed at any instant are explicit functions of the output y and the input x. Specifically, they are determined by quantized information derived from the output, output derivative, error, and error derivative, namely, the sign of the products sgn $(y'e)$, sgn $(y'e')$, sgn (ye), and sgn (ye').

(4) There are basically 16 m,n subscript combinations and thus 16 β_m,γ_n parameter combinations. However, a detailed study of property (3) shows that only 8 are allowed. The allowed combinations may be

$$m=n=0, 1, 2, 3$$

or

$$m\neq n$$

$$m+n=3$$

The reason for the "exclusion principle" on coefficient combinations stems from a desire to obtain mirrored-image outputs for mirrored-image inputs; that is,

$$y_{im}(\tau)=-y(\tau)$$

is desired when

$$x_{im}(\tau)=-x(\tau)$$

(5) The control criterion that was employed in obtaining the functional dependence of β_m and γ_n was

$$|y_d-y|=|x-y|=|e|<\epsilon$$

where ϵ is a small positive quantity. This criterion of maintenance of small instantaneous error between desired output and actual output enabled reduction of equation (8) to the approximate autonomous differential equation (see refs. and 3 to 5)

$$y''+2D(1+\beta_m)y'+\gamma_n y=(x-y)\approx 0 \tag{10}$$

Thus phase-plane techniques could be employed to find the functional dependence of β_m and γ_n.

(6) Once the linear damping factor D is fixed the process of control is obtained by switching parameters β_m and γ_n. Equation (8) consists of an ensemble of eight (see property (4)) linear differential equations with constant coefficients. The process of control may be visualized as the switching from one member of the ensemble to another. This switching is determined by quantized information derived from the input and output (see property (3)). From another point of view (consistent with the approximation described in property (5)), the output y is to be forced to satisfy two conditions simultaneously, that is, both sides of equation (10). This is approximately possible by switching to various β_m and γ_n parameter combinations and can be visualized as the process of switching to various phase trajectories of equation (10) in the phase plane of y' against y.

DISCUSSION OF NONLINEAR CONTROL

In the section entitled "Transition to Nonlinear Control" a logical transition to nonlinear control systems utilizing the concept of control through parameters is suggested. However, mathematical difficulty hampers the development of this approach. In the next two sections a particular nonlinear control theory is presented. This theory constitutes the first step in the synthesis of a nonlinear control system which obtains control through parameters. Since the functional dependence of the parameters has been established, the problem is reduced to finding a set of five (constant) parameters $(D,_1\beta,_2\beta,_1\gamma,_2\gamma)_o$. It should be appreciated, however, that even the optimization of this five-parameter set cannot in general be accomplished analytically because of the overall nonlinear nature of the problem.

Aside from questions on the analytical optimization of parameters in equation (8), there are equally important practical questions such as:

(1) Can a useful control system that is described by the nonlinear differential equation (eq. (8)) be realized?

(2) If the system is realizable, what is its physical nature?

(3) If the system is realizable, how does it compare in performance and complexity with a "good" second-order linear control system?

There are then mathematical difficulties on the one hand and physical difficulties on the other. The mathematical difficulties could be handled by numerical methods of integration of the differential equation (e. g., utilizing a digital computer). However, this would give little insight into the nature of a system that is controlled through discontinuous variation of the parameters β_m and γ_n. It has been found advantageous to investigate the physical questions first and then to utilize simulation techniques (analog computer) to investigate the analytical properties of this type of control.

PHYSICAL REALIZATION

PHYSICAL MODEL

It is desirable to study the nature of a physical control system that is described by an ensemble of eight linear differential equations with control being accomplished by switching from one member of the ensemble to another. To do this, equation (8) is rearranged as shown below:

$$\frac{d^2y}{d\tau^2}+2D\frac{dy}{d\tau}+y=x(\tau)-\left(2D\beta_m\frac{dy}{d\tau}+\gamma_n y\right) \tag{11}$$

or in operational notation

$$(p^2+2Dp+1)y\overset{o}{=}x-(2D\beta_m p+\gamma_n)y \tag{12}$$

Forgetting for the time being that the parameters β_m and γ_n are actually functions, one interpretation of this operational equation and thus of equation (8) is shown in figure 3.

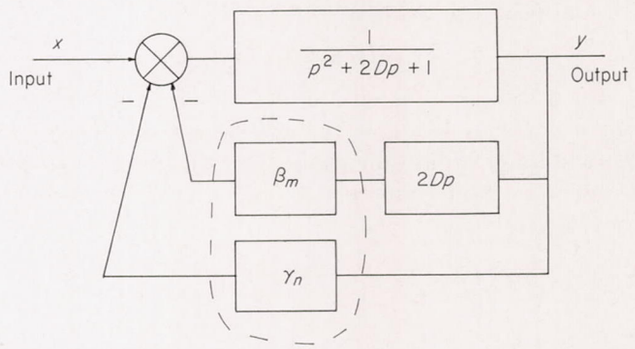


FIGURE 3.—Block diagram of equation (8) assuming that β_m and γ_n are constant (denoted by encircling dotted line).

Figure 3 can be modified to take into account the fact that β_m and γ_n are stepwise switching functions of time, their implicit variable (i. e., β_m and γ_n can each take on four discrete values). This is shown symbolically in figure 4. The explicit functional dependence of the parameters β_m and γ_n has not yet been given and is thus indicated as a switching logic of undefined character.

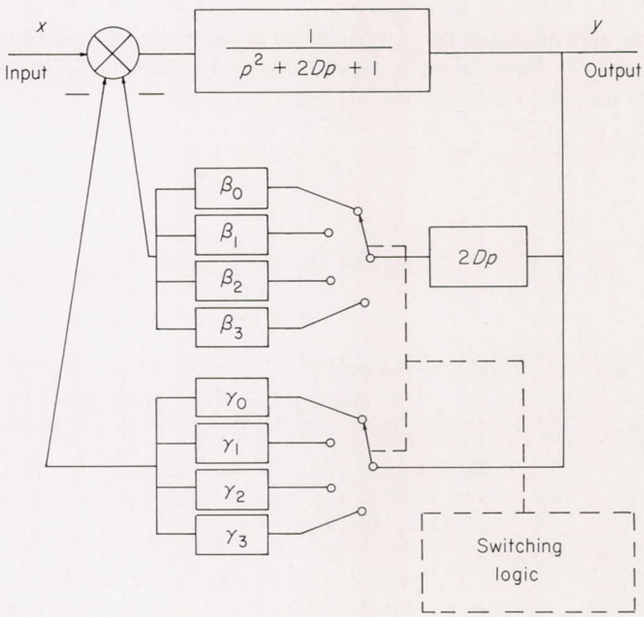


FIGURE 4.—Block diagram of equation (8) taking into account stepwise switching nature of β_m and γ_n.

Utilizing the block diagram of figure 4, the physical interpretation of the nonlinear control system described by equation (8) is quite straightforward. This system consists of:

(1) A linear feedforward portion. This portion could be a linear control system in itself (e. g., the simple position control servo of figure 1 and equation (5)).

(2) A feedback switching circuit comprised of:

(a) Four discrete values of proportional feedback γ_n (two positive and two negative as shown in fig. 2)

(b) Four discrete values of derivative feedback $2D\beta_m$ (two positive and two negative as shown in fig. 2)

(c) A switching logic which at any instant determines the particular combination of derivative and proportional feedback $2D\beta_m,\gamma_n$ employed

SWITCHING LOGIC

In this section it is shown that digital-computer techniques can be utilized to establish the switching logic for the feedback switching circuit mentioned in the previous section.

Recall that the parameters β_m and γ_n have been defined as functions; that is,

$$\left.\begin{aligned} \beta_m &= -{}_1\beta \operatorname{sgn}(y'e) - {}_2\beta \operatorname{sgn}(y'e') \\ & m=0,1,2,3 \\ \gamma_n &= -{}_1\gamma \operatorname{sgn}(ye) - {}_2\gamma \operatorname{sgn}(ye') \\ & n=0,1,2,3 \end{aligned}\right\} \tag{13}$$

where the subscript convention has been given by equations (9). Equations (13) determine the switching logic. Thus, for example, β_3 is chosen when $(y'e)<0$ and $(y'e')<0$ and γ_0 is chosen when $(ye)>0$ and $(ye')>0$, so that the combination $\beta_3\gamma_0$ is chosen when $(y'e)<0$, $(y'e')<0$, $(ye)>0$, and $(ye')>0$. At this point it appears necessary to form the products ye, ye', $y'e$, and $y'e'$ and then to find the sign of these products in order to establish the switching logic. Physically, however, the process of multiplication is to be avoided if possible. That there is a possibility of avoiding multiplication may be gleaned by realizing that

$$\operatorname{sgn}(ab)=\operatorname{sgn}(a)\operatorname{sgn}(b)$$

since

$$\frac{ab}{|ab|}=\frac{a}{|a|}\frac{b}{|b|}$$

Thus equations (13) may be rewritten as

$$\left.\begin{aligned} \beta_m &= -\operatorname{sgn}(y')[{}_1\beta \operatorname{sgn}(e) + {}_2\beta \operatorname{sgn}(e')] \\ & m=0,1,2,3 \\ \gamma_n &= -\operatorname{sgn}(y)[{}_1\gamma \operatorname{sgn}(e) + {}_2\gamma \operatorname{sgn}(e')] \\ & n=0,1,2,3 \end{aligned}\right\} \tag{14}$$

Again the subscript convention is defined by equations (9).

Thus, utilizing equations (13), for example, β_3 is chosen when

$$y'>0,\ e<0,\ e'<0$$

or when

$$y'<0,\ e>0,\ e'>0$$

γ_0 is chosen when

$$y>0,\ e>0,\ e'>0$$

or when

$$y<0,\ e<0,\ e'<0$$

so that the combination (β_3,γ_0) is chosen when

$$y>0,\ y'<0,\ e>0,\ e'>0$$

or when

$$y<0,\ y'>0,\ e<0,\ e'<0$$

From this example it can be seen that it is not necessary to find the signs of products but rather that it is sufficient to find separately the signs of y, y', e, and e'.

Since the sign of a variable is quantized binary information of the variable, it is convenient to utilize digital-computer techniques to further the switching logic. This may be done as follows:

Let the convention be adopted that $y>0$ be represented by 0 (binary zero), $y<0$ be represented by 1 (binary one), and similarly for y', e, and e'. If the ordered sequence is now established as

$$(y,\ y',\ e,\ e')$$

four-digit binary logic may be employed to encode equation (14). In particular, a binary coded decimal may be used (see table 1).

It was mentioned in property (4) of the section "Properties of Equation (8)" that not all of the 16 possible β_m,γ_n parameter combinations were allowed under their definition. This was termed an exclusion principle on the allowed coefficient combinations. One of the advantages of the suggested binary coding scheme of table 1 is that this exclusion principle is built into the code. To understand this, consider the example of the combination $\beta_3\gamma_0$ given previously. In the code language $\beta_3\gamma_0$ is chosen when the binary number 0100 or 1011 occurs. What is implied by this example is that a binary number and its complement must be identical (i. e., $0000=1111$, $0111=1000$) as far as the switching logic is concerned. Thus out of the 16 possible four-digit binary numbers only the first 8 are unique. That is, in counting from 0 to 7 in a binary coded decimal, if complements are included then so are the other 8 possibilities, 8 to 15 (see table 1).

The allowed β_m,γ_n parameter combinations along with the encoded logic of table 1 are summarized in matrix form in table 2. Examples are given to illustrate the meaning of the table. In general the allowed subscript combinations are

$$m=n=0,\ 1,\ 2,\ 3$$

or

$$m\neq n$$

$$m+n=3$$

REALIZATION COMPLETED

Now that equation (14) has been successfully interpreted (encoded) in binary-logic form (table 2), the realization of a feedback switching circuit utilizing this encoded logic is a typical digital-computer switching-circuit problem. As is characteristic of any synthesis process there will, in general, be many ways to design this feedback switching circuit.

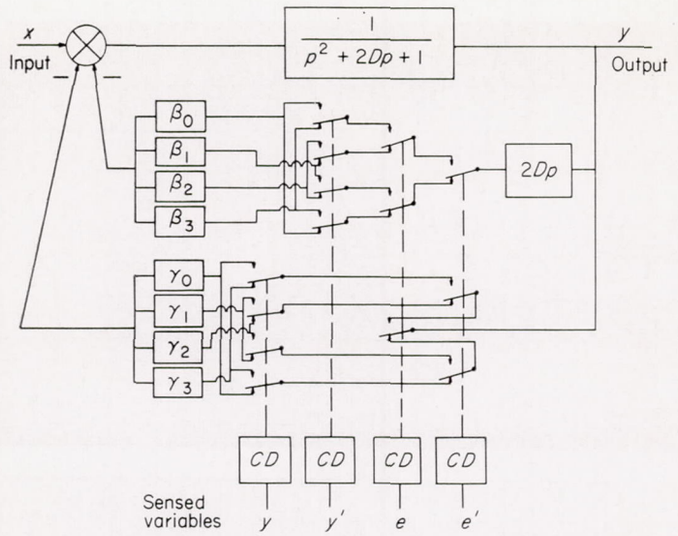


FIGURE 5.—Block diagram showing complete physical interpretation of equation (8). *CD*, zero-coincidence detectors.

The block diagram of figure 5 shows one design that completes the physical interpretation of equation (8) along the lines started in figures 3 and 4.

In figure 5 the feedback switching circuit consists of:

(1) The four discrete values of both derivative and proportional feedback $2D\beta_m,\gamma_n$

(2) A relay switching circuit that connects the proper feedback combination

(3) Zero-coincidence detectors *CD* that drive the banks of relays to one position or the other depending upon the signs of the sensed variables

It should be noted that, depending upon the application, other forms of sign-sensing devices and other switching devices such as diodes, transistors, electronic switches, and/or magnetic amplifiers could be employed to obtain other realizations of equation (8). In any case the following properties are basic to any realization:

(1) The signs of the four variables y, y', e, and e' are sensed. This may be thought of as the process of "reading in" the four-digit binary logic of table 1.

(2) On the basis of the 2^4 possible binary decisions the required feedback combinations $2D\beta_m,\gamma_n$ as defined in table 2 are connected around a linear second-order member.

It is important to stress that the only types of nonlinear operations required in the realization of this nonlinear control system are switching-type operations. In addition, all the switching is to be performed in feedback paths, which means that the switching can be done at low electronic power levels. These practical features are definite design advantages. Thus, in summary, it can be said that this type of nonlinear control system is not only physically realizable but also practical from an instrumentation standpoint.

EXPERIMENTAL VERIFICATION OF NONLINEAR CONTROL THEORY

DISCUSSION OF SIMULATION TECHNIQUES

Simulation techniques were chosen as an experimental mode of investigation of performance of the nonlinear control system realized from equation (8). These techniques offer the following advantages:

(1) Proximity to the actual control system. This means that the same practical features with regard to instrumentation (see the section entitled "Realization Completed") are exploited to fullest advantage. Thus, just as in the actual system, the only nonlinear device required for the simulated model is a binary logic feedback switching circuit (see appendix C for details). The linear portion of the system is simulated on an analog computer. Here the only operations required are summations, two integrations, and one differentiation. These are all operations which an analog computer does well. It can be said then that the accuracy to which the simulated model simulates equation (8) depends primarily upon the realized feedback switching circuit. The most essential type of imperfection to be expected in this switching circuit is time delay in switching. Exactly the same type of imperfection will be met in the physical control system. Thus there will be more nearly a one-to-one correspondence between the simulated model and the actual system than between either and equation (8).

(2) Convenience in experimental investigation. In order to characterize the output y of the nonlinear system completely, a set of five parameters

$$(D,\ {}_1\beta,\ {}_2\beta,\ {}_1\gamma,\ {}_2\gamma)$$

and the input x must be specified. In the performance evaluation of the system it is necessary to be able to vary these characterizing quantities conveniently. Simulation techniques allow this.

PRESENTATION OF EXPERIMENTAL RESULTS

Figures 6 to 17 present experimental results obtained from the simulation studies of equation (8). Briefly, the results are presented as follows:

Figures 6 to 12 compare the responses (output y and error e) of the nonlinear system with that of a linear system for various classes of inputs x. (In comparing the linear and nonlinear responses it will be noted that there is not exact synchronism of events because, with the available experimental facilities, it was necessary to obtain the two responses separately.) The linear system utilized is that which constitutes the feedforward member of the nonlinear system corresponding to the case where $\beta_m=\gamma_n=0$. The nonlinear system for figures 6 to 15 is

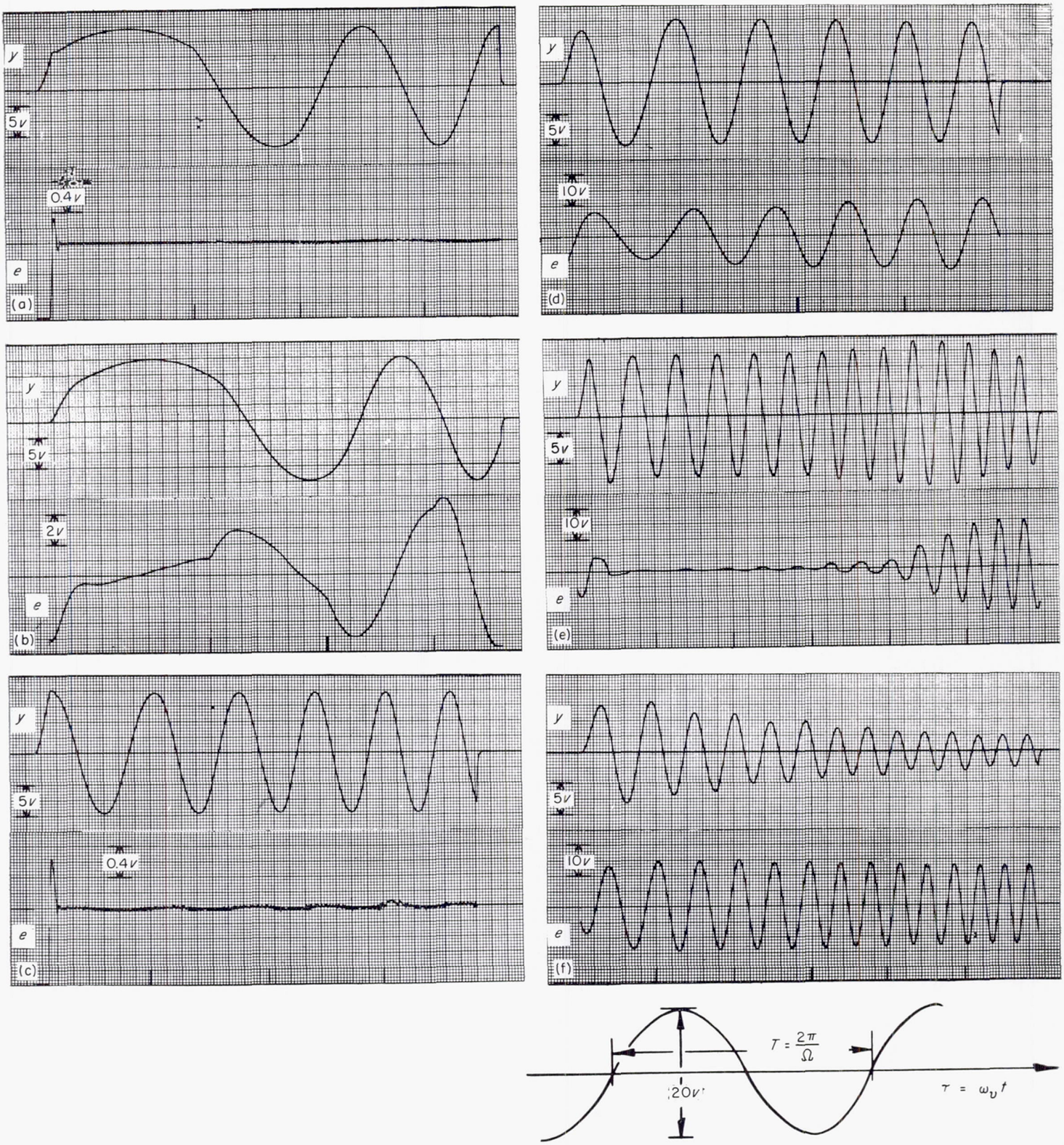
$$\beta_3=-\beta_0=2$$

$$\beta_2=-\beta_1=0.5$$

$$\gamma_3=-\gamma_0=2$$

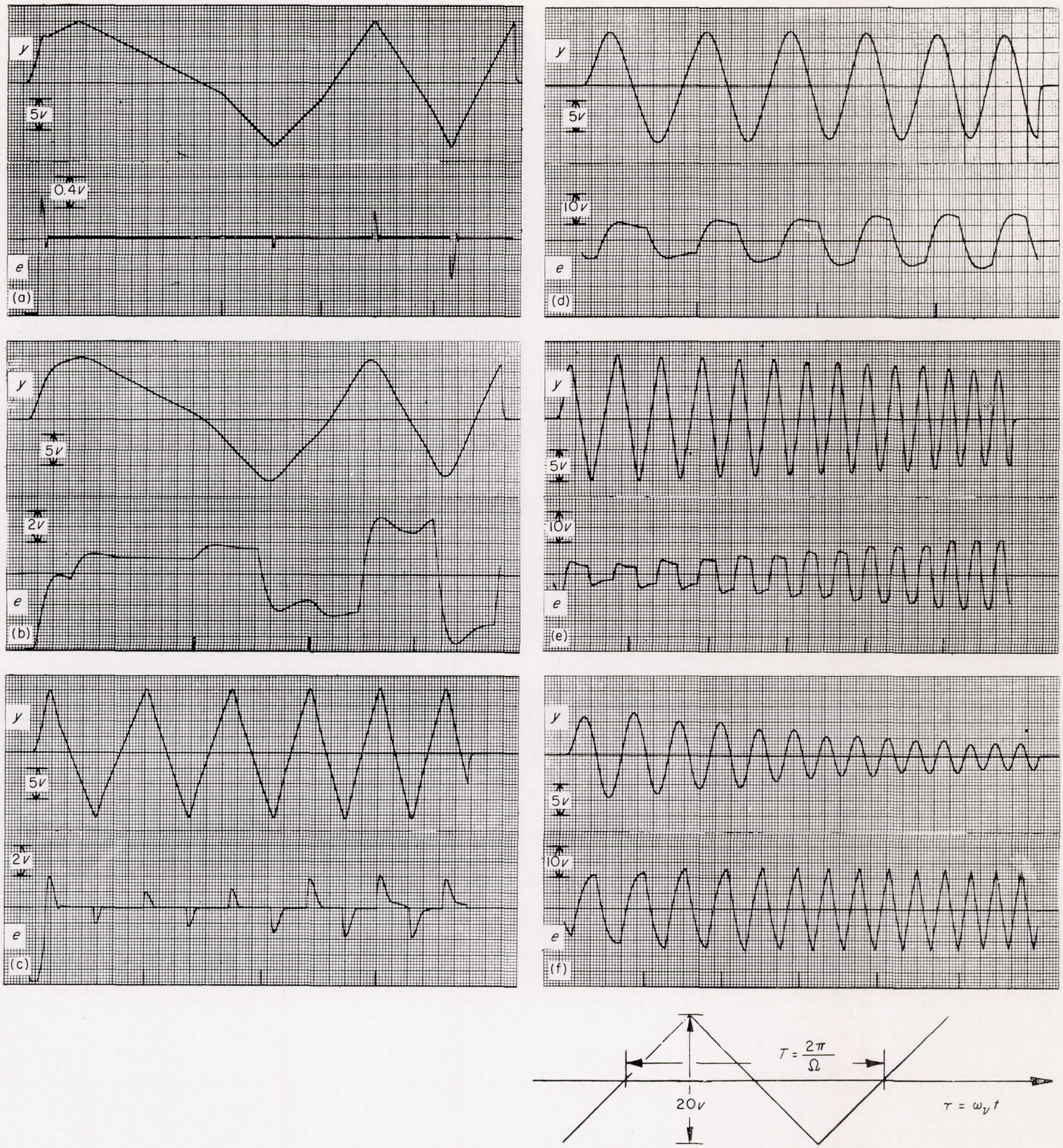
$$\gamma_2=-\gamma_1=0.5$$

Figure 6 compares the system responses to sinusoidal inputs and figure 7 shows the responses to triangular-wave inputs. A partially integrated square wave $x(\tau)$ is defined as the output of a first-order linear system characterized by the transfer function $H(p)=1/(\alpha p+1)$, when the input $x_1(\tau)$ is a square wave. The responses of linear and nonlinear systems to this type of input are given in figure 8.



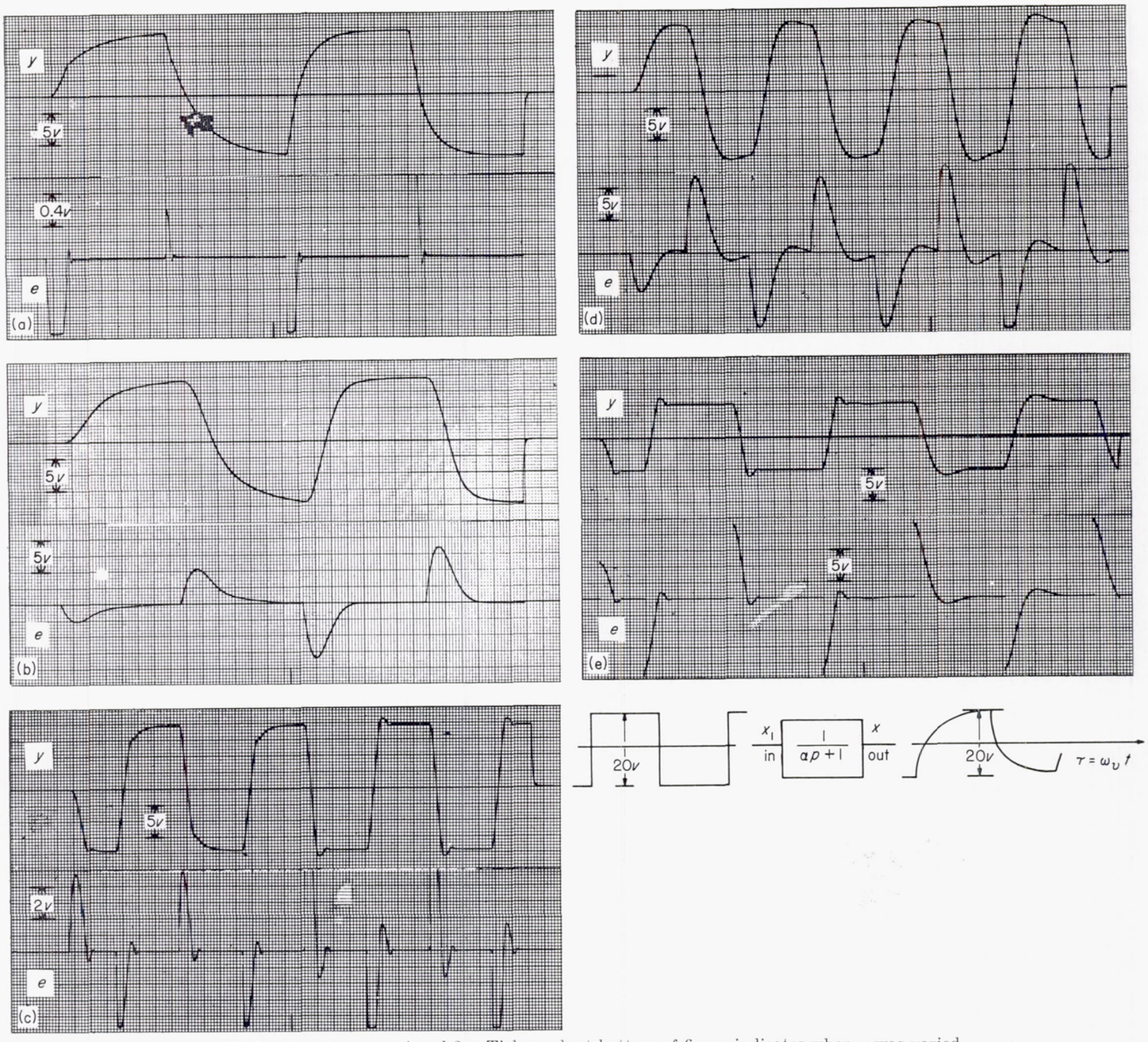
(a) Nonlinear system; $\Omega=\omega/\omega_v=0.1$ to 0.4 in 0.1 steps.
(b) Linear system; $\Omega=\omega/\omega_v=0.1$ to 0.4 in 0.1 steps.
(c) Nonlinear system; $\Omega=\omega/\omega_v=0.5$ to 0.8 in 0.1 steps.
(d) Linear system; $\Omega=\omega/\omega_v=0.5$ to 0.8 in 0.1 steps.
(e) Nonlinear system; $\Omega=\omega/\omega_v=1.0$ to 2.0 in 0.2 steps.
(f) Linear system; $\Omega=\omega/\omega_v=1.0$ to 2.0 in 0.2 steps.

FIGURE 6.—Linear and nonlinear system responses for 20-volt peak-to-peak sinusoidal inputs with frequency Ω varied. $D=0.6$; 2.5 small divisions on time scale=1 normalized time unit; tick marks at bottom of figures indicate when frequency was varied.



(a) Nonlinear system; $\Omega = 2\pi/T = 0.1$ to 0.4 in 0.1 steps.
(b) Linear system; $\Omega = 2\pi/T = 0.1$ to 0.4 in 0.1 steps.
(c) Nonlinear system; $\Omega = 2\pi/T = 0.5$ to 0.8 in 0.1 steps.
(d) Linear system; $\Omega = 2\pi/T = 0.5$ to 0.8 in 0.1 steps.
(e) Nonlinear system; $\Omega = 2\pi/T = 1.0$ to 2.0 in 0.2 steps.
(f) Linear system; $\Omega = 2\pi/T = 1.0$ to 2.0 in 0.2 steps.

FIGURE 7.—Linear and nonlinear system responses to 20-volt peak-to-peak triangular-wave inputs with period T varied. $D = 0.6$; 2.5 small divisions on time scale = 1 normalized time unit; tick marks at bottom of figures indicate when periods were varied.



(a) Nonlinear system; $\alpha=4$ and 2. Tick mark at bottom of figure indicates when α was varied.
(b) Linear system; $\alpha=4$ and 2. Tick mark at bottom of figure indicates when α was varied.
(c) Nonlinear system; $\alpha=1$ and 0.5. Tick mark at bottom of figure indicates when α was varied.
(d) Linear system; $\alpha=1$ and 0.5. Tick mark at bottom of figure indicates when α was varied.
(e) Nonlinear and linear systems; $\alpha=0$. Tick mark at bottom of figure indicates demarcation between nonlinear and linear systems.

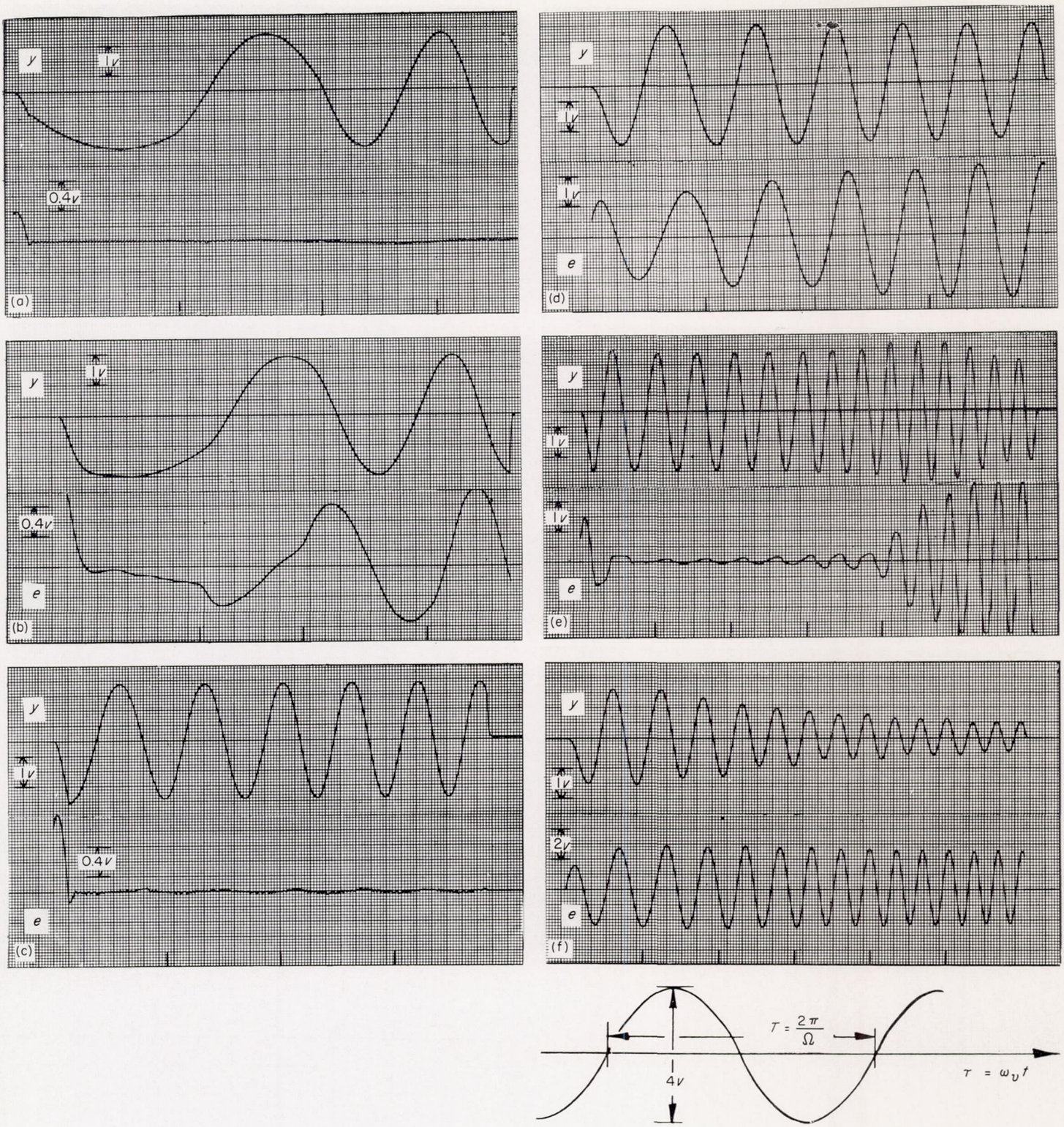
FIGURE 8.—Linear and nonlinear system responses to 20-volt peak-to-peak partially integrated square-wave input. $D=0.6$; 2.5 small divisions on time scale=1 normalized time unit.

The responses of the systems to small sinusoidal inputs are given in figure 9. Figure 10 gives the responses to clipped sinusoidal inputs. Figure 11 shows the response of the nonlinear system to sinusoidal inputs that have been displaced with a direct-current component. Figure 12 shows the responses of linear and nonlinear systems to a triangular-wave input whose periods and amplitudes are randomly modulated.

Figures 13 to 15 deal with the effects of imperfections that are likely to be encountered in the actual control system. Figure 13(a) gives the results and data of an experimental investigation on the effects of switching delays due to threshold in sensing the sign of the error, sgn (e) for a triangular-wave input. The experimental results for a constant 9-volt input (see fig. 13(b)) are given below:

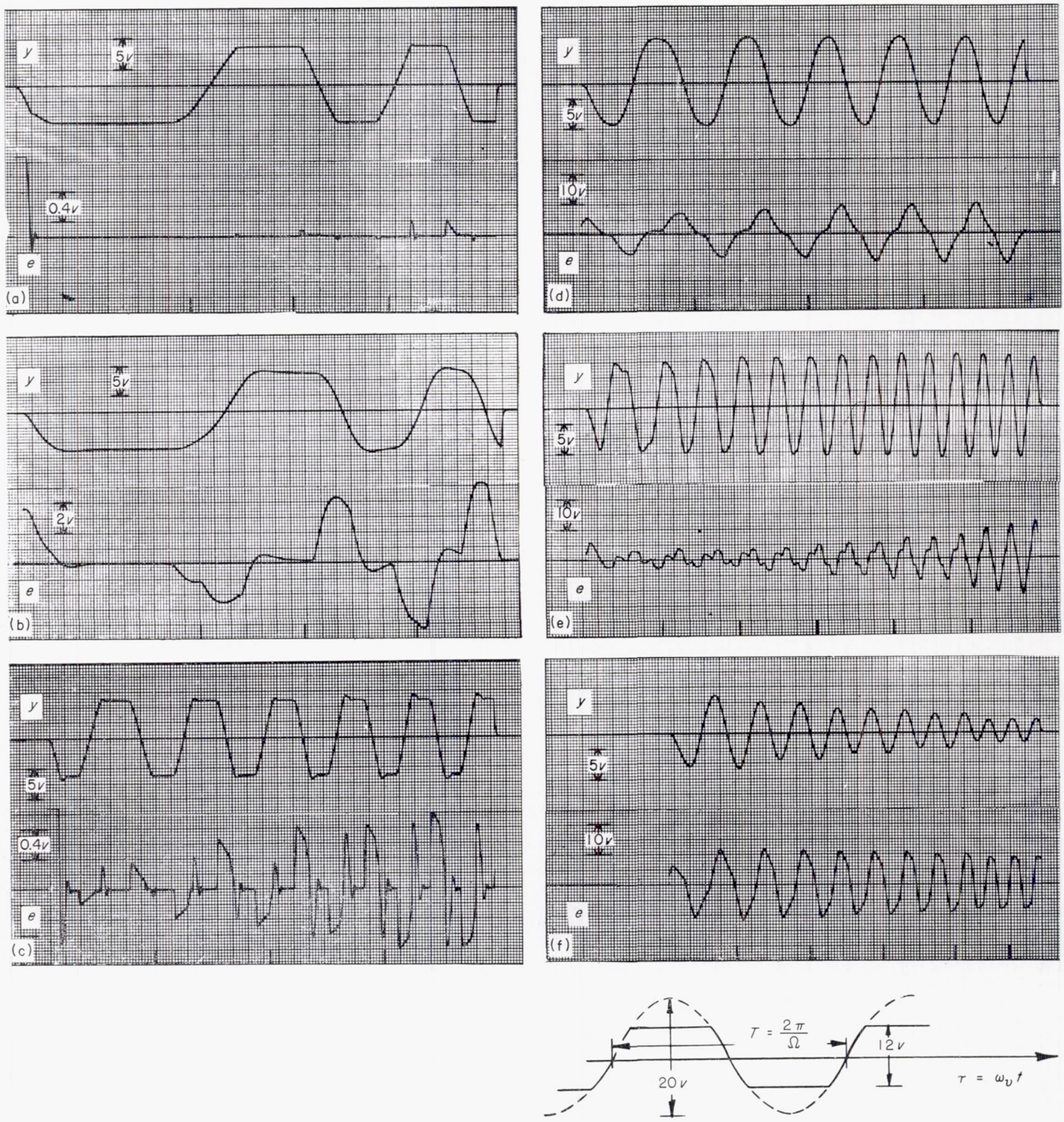
Threshold, mv	14	26	36	44
Peak-to-peak error, mv	44	124	220	290

Figure 14 shows the effects of placing progressively smaller limits on the acceleration of the nonlinear control system. For each value of y'' limit considered, the output, output derivative, output acceleration, and instantaneous error are shown. The effects of a velocity limit on performance of the nonlinear and linear systems are compared in figure 15.



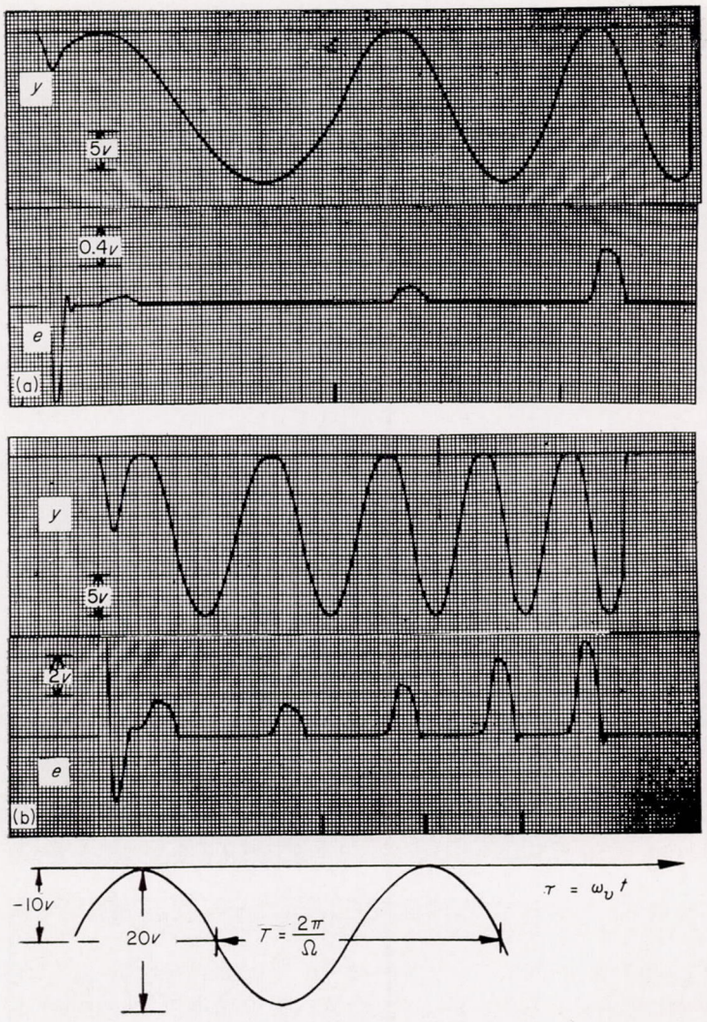
(a) Nonlinear system; $\Omega = \omega/\omega_v = 0.1$ to 0.4 in 0.1 steps.
(b) Linear system; $\Omega = \omega/\omega_v = 0.1$ to 0.4 in 0.1 steps.
(c) Nonlinear system; $\Omega = \omega/\omega_v = 0.5$ to 0.8 in 0.1 steps.
(d) Linear system; $\Omega = \omega/\omega_v = 0.5$ to 0.8 in 0.1 steps.
(e) Nonlinear system; $\Omega = \omega/\omega_v = 1.0$ to 2.0 in 0.2 steps.
(f) Linear system; $\Omega = \omega/\omega_v = 1.0$ to 2.0 in 0.2 steps.

FIGURE 9.—Linear and nonlinear system responses to 4-volt peak-to-peak sinusoidal inputs with frequency Ω varied. $D = 0.6$; 2.5 small division on time scale = 1 normalized time unit; tick marks at bottom of figures indicate when frequency was varied.



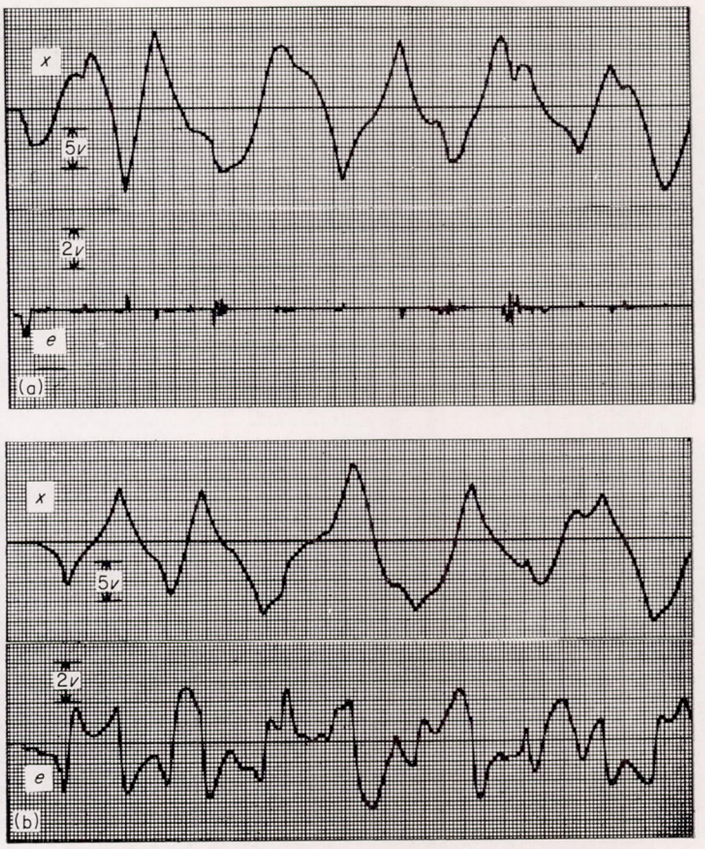
(a) Nonlinear system; $\Omega=\omega/\omega_v=0.1$ to 0.4 in 0.1 steps.
(b) Linear system; $\Omega=\omega/\omega_v=0.1$ to 0.4 in 0.1 steps.
(c) Nonlinear system; $\Omega=\omega/\omega_v=0.5$ to 0.8 in 0.1 steps.
(d) Linear system; $\Omega=\omega/\omega_v=0.5$ to 0.8 in 0.1 steps.
(e) Nonlinear system; $\Omega=\omega/\omega_v=1.0$ to 2.0 in 0.2 steps.
(f) Linear system; $\Omega=\omega/\omega_v=1.0$ to 2.0 in 0.2 steps.

FIGURE 10.—Linear and nonlinear system responses to 20-volt peak-to-peak sinusoidal inputs with frequency Ω varied that are clipped symmetrically at ± 6 volts. $D=0.6$; 2.5 small divisions on time scale = 1 normalized time unit; tick marks at bottom of figures indicate when frequency was varied.



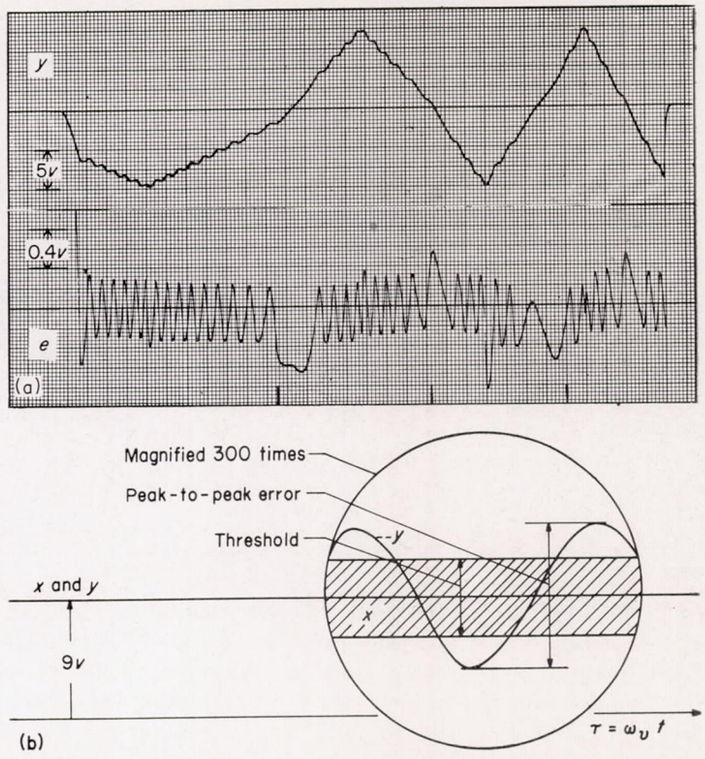
(a) $\Omega = \omega/\omega_v = 0.1$ to 0.4 in 0.1 steps.
(b) $\Omega = \omega/\omega_v = 0.5$ to 0.8 in 0.1 steps.

FIGURE 11.—Nonlinear system responses of 20-volt peak-to-peak sinusoidal input with frequency Ω varied that has been displaced by −10 volts (direct-current component) for obtaining osculation. $D = 0.6$; 2.5 small divisions on time scale = 1 normalized time unit; tick marks at bottom of figures indicate when frequency was varied.



(a) Nonlinear system. (b) Linear system.

FIGURE 12.—Linear and nonlinear system responses for triangular-wave input whose period and amplitude are randomly modulated. $D = 0.6$; 1 small division on time scale = 1 normalized time unit.



(a) Nonlinear system response to triangular-wave input with period varied as in figure 7 (a). $D = 0.6$; tick marks at bottom of figure indicate when period was varied. Threshold in sensing sgn e was 44 millivolts.

(b) Constant input $x(\tau)$ was 9 volts.

FIGURE 13.—Effects of switching delays due to threshold in sensing sgn e.

Figure 16 presents the responses of some special cases of the nonlinear system (special with respect to the choice of the β_m and γ_n parameter values). In this figure the system responses to a triangular-wave input are given for four different parameter sets. The parameters pertinent to these results are listed in table 3 for easy reference.

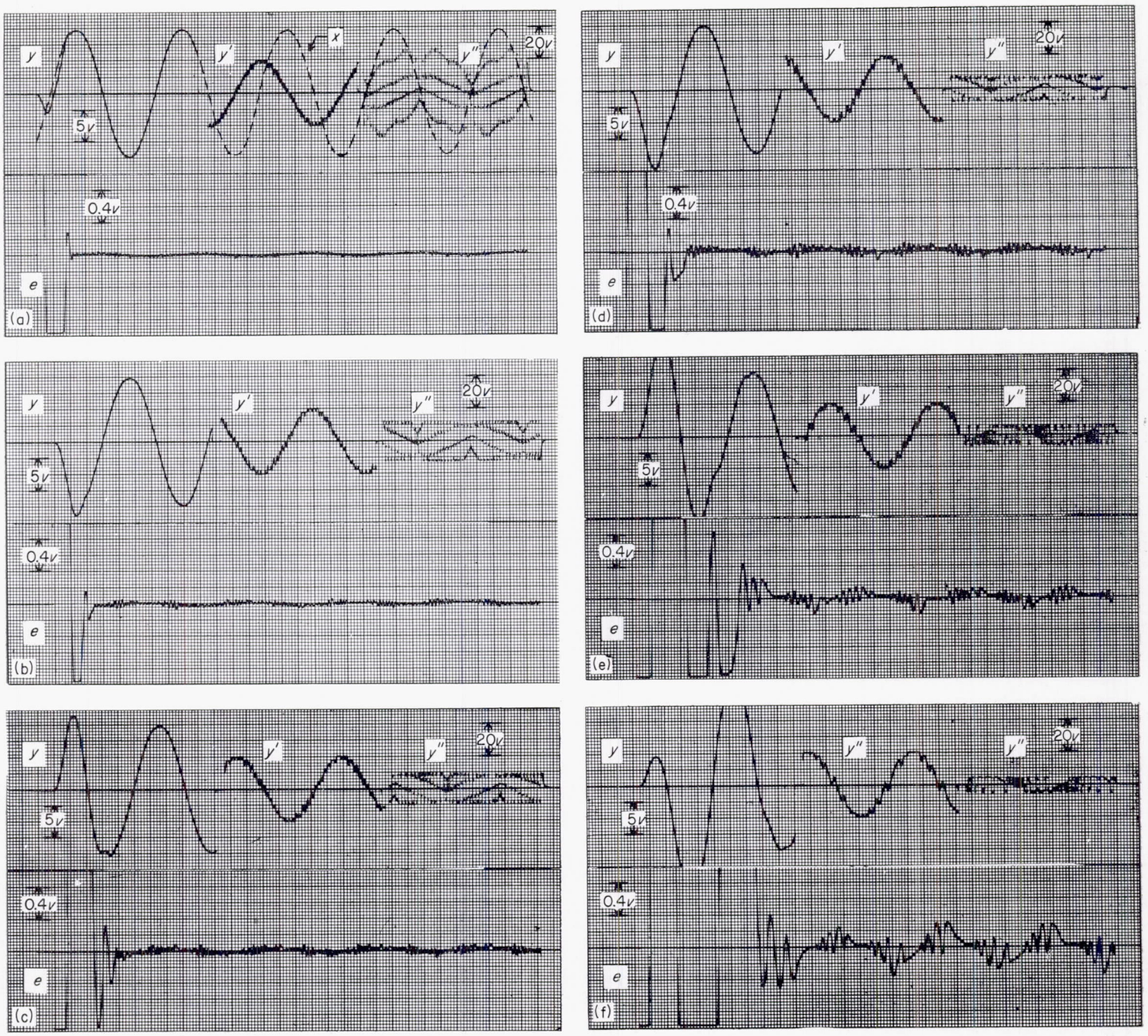
Figure 17 gives the response of the nonlinear system possessing a low linear damping factor $D = 0.1$. The responses are for a triangular-wave input whose frequency was varied in the same manner as that of figures 7(a) and 7(c).

Detailed discussions of these results are given in the section entitled "Discussion of Results."

DISCUSSION OF RESULTS

PERFORMANCE EVALUATION OF SINUSOIDAL INPUTS

To complete the synthesis of the nonlinear control system that has been derived from equation (8), it is necessary to choose the magnitudes of the parameters that characterize the system, that is, $(D,_1\beta,_2\beta,_1\gamma,_2\gamma)$ or $(D,\beta_3,\beta_2,\gamma_3,\gamma_2)$. It is expected that the performance of the system depends on



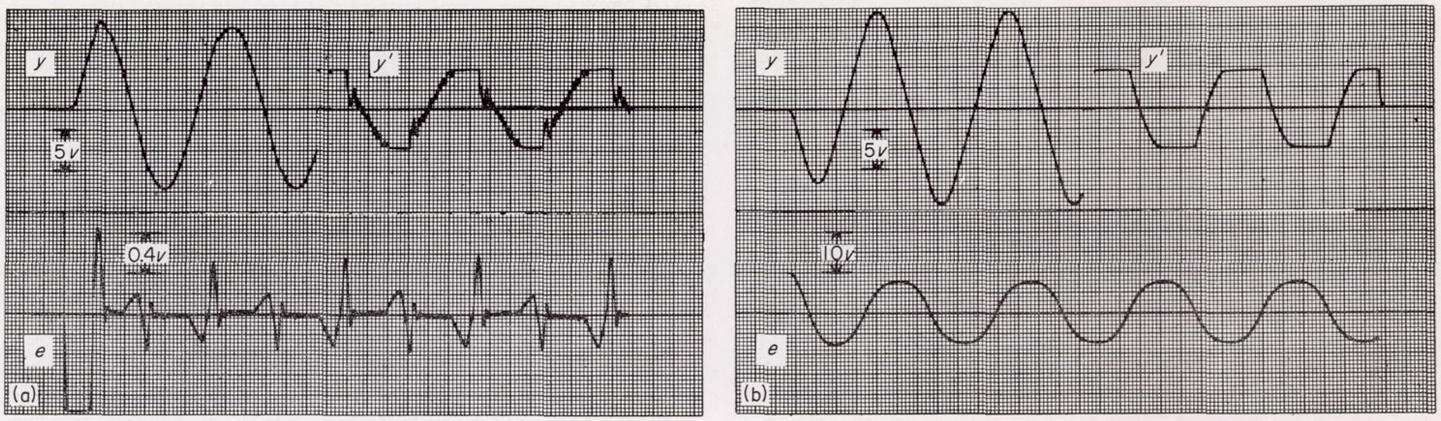
(a) Nonlinear system; no y'' limit.
(b) Nonlinear system; ± 12-volt y'' limit.
(c) Nonlinear system; ± 9-volt y'' limit.
(d) Nonlinear system; $\pm 7\frac{1}{2}$-volt y'' limit.
(e) Nonlinear system; ± 6-volt y'' limit.
(f) Nonlinear system; $\pm 4\frac{1}{2}$-volt y'' limit.

FIGURE 14.—Nonlinear system responses to 20-volt peak-to-peak sinusoidal input with varying acceleration limits. $D=0.6$; $\Omega=\omega/\omega_v=0.5$; 2.5 small divisions on time scale = 1 normalized time unit.

the choice of these parameters. For studying their influence, simulation techniques proved to be very convenient. Experimental results were given in figures 6 to 12 where the response (output y and error e) of the nonlinear system was compared with the response of a linear system for a variety of inputs. This gives the possibility of establishing the properties of the nonlinear system not only by itself but also with respect to a linear standard. The linear system employed for this purpose was that which constitutes the ieedforward member of the nonlinear system (i.e., $\beta_m=\gamma_n=0$).

These experiments allow parameter values for good performance of the nonlinear system to be found. Analytical and practical considerations that aid in the optimization are treated later in the section entitled "Choice of Parameter Values."

The sinusoidal responses of the nonlinear and linear systems are compared in figure 6. Here, the frequency range considered was $0.1 \leqq \Omega=\omega/\omega_y \leqq 2$. The peak-to-peak input amplitude was 20 volts. These results show that the nonlinear system reproduced the sinusoidal inputs up to the frequency



(a) Nonlinear system.
(b) Linear system.

FIGURE 15.—Linear and nonlinear system responses to 20-volt peak-to-peak sinusoidal input with $\pm 4\frac{1}{2}$-volt velocity limit. $D=0.6$; $\Omega=2\pi/T=0.5$; 2.5 small divisions on time scale = 1 normalized time unit.

$\Omega=1.4$ with virtually no instantaneous error when compared with that of the linear system. For higher frequencies the error for the nonlinear system increased rapidly to the same order of magnitude as that of the linear system.

Substantially the same comparative performance was displayed by the two systems when the input was a smaller (4-volt peak-to-peak) sinusoid as is shown in figure 9.

Figure 11 gives the response of the nonlinear system to a 20-volt peak-to-peak osculating sinusoid over the frequency range $0.1 \leqq \Omega \leqq 0.8$. This is a severe type of input for the nonlinear system since both x and x' simultaneously go to zero. This implies that y and y' are also small so that in equation (8) the discontinuous variations of the parameters β_m, γ_n cannot be so effective in determining the acceleration y'', since

$$y''=x-[2D(1+\beta_m)y'+(1+\gamma_n)y] \tag{15}$$

From the figures it is seen that the nonlinear system did have some difficulties near the osculating regions; further, the peak error increased as the frequency was increased (see appendix D). As might be expected from equation (15) this error for small values of y and y' can be reduced by increasing the magnitudes of the parameters (this will be discussed in more detail in the section entitled "Choice of Parameters"). At any rate, by comparison, the error for the present system is always less than that of the linear system of figure 6. (In the linear realm, error is independent of a shift in the direct-current level of the input.)

Taken collectively, the 20-volt, the 4-volt, and the osculating 20-volt sinusoidal inputs tend to form a more realistic appraisal of the nonlinear system performance than does a single input amplitude. There is still much that can be learned from a detailed study of these three responses, but first it is desirable to obtain some sort of a quantitative comparison between the nonlinear and the linear system performance. One such comparison can be made as follows:

(1) Assume that the input to both systems $x(\tau)$ is, and has been for a long time, a 4-volt peak-to-peak sinusoid of frequency

$$\Omega=\omega/\omega_\nu=0.1$$

so that as far as the linear system is concerned this is a steady-state alternating-current input.

(2) Determine how much the band width or the natural frequency ω_ν of the linear system must be increased in order that the time average of the magnitude of the instantaneous error

$$\overline{\overline{|e|}}=\frac{\Omega}{2\pi}\int_0^{2\pi/\Omega}|e|d\tau$$

for the linear system be reduced to that value given by the nonlinear system.

Here it is easy to show that for the low-frequency steady-state alternating-current case

$$\overline{\overline{|e|}}_{lin}\approx\frac{4D}{\pi}\,\Omega|x|_{max} \tag{16}$$

For an input $x=|x|_{max}e^{i\Omega\tau}$ the steady-state error is given by

$$e\triangleq x-y=|x|_{max}\left(1-\frac{1}{1+i\Omega 2D-\Omega^2}\right)e^{i\Omega\tau}$$

For small values of Ω this equation yields

$$e\approx|x|_{max}[1-(1-i\Omega 2D+\ldots)]e^{i\Omega\tau}$$
$$\approx|x|_{max}(i\Omega 2D)e^{i\Omega\tau}$$

or

$$|e|_{max}\approx 2D\Omega|x|_{max}$$

but

$$\overline{\overline{|e|}}_{lin}=\frac{1}{\pi}\int_0^{\pi}|e|_{max}\sin\Omega\tau\,d\Omega\tau=\frac{2}{\pi}|e|_{max}$$

therefore

$$\overline{\overline{|e|}}_{lin}\approx\frac{4D}{\pi}\,\Omega|x|_{max} \tag{17}$$

For the given system and input

$$\overline{\overline{|e|}}_{lin}\approx 0.15$$

For the nonlinear system and the same input (approximate calculation from the much larger and clearer original of fig. 9 (a))

$$\overline{\overline{|e|}}_{nonlin}\approx 0.01$$

Thus,

$$\overline{\overline{|e|}}_{nonlin}\approx\frac{1}{15}\,\overline{\overline{|e|}}_{lin}$$

Now from equation (17), since $\Omega=\omega/\omega_\nu$, it is seen that in order to reduce $\overline{\overline{|e|}}_{lin}$ by this factor of 15 that ω_ν must be increased by the same factor.

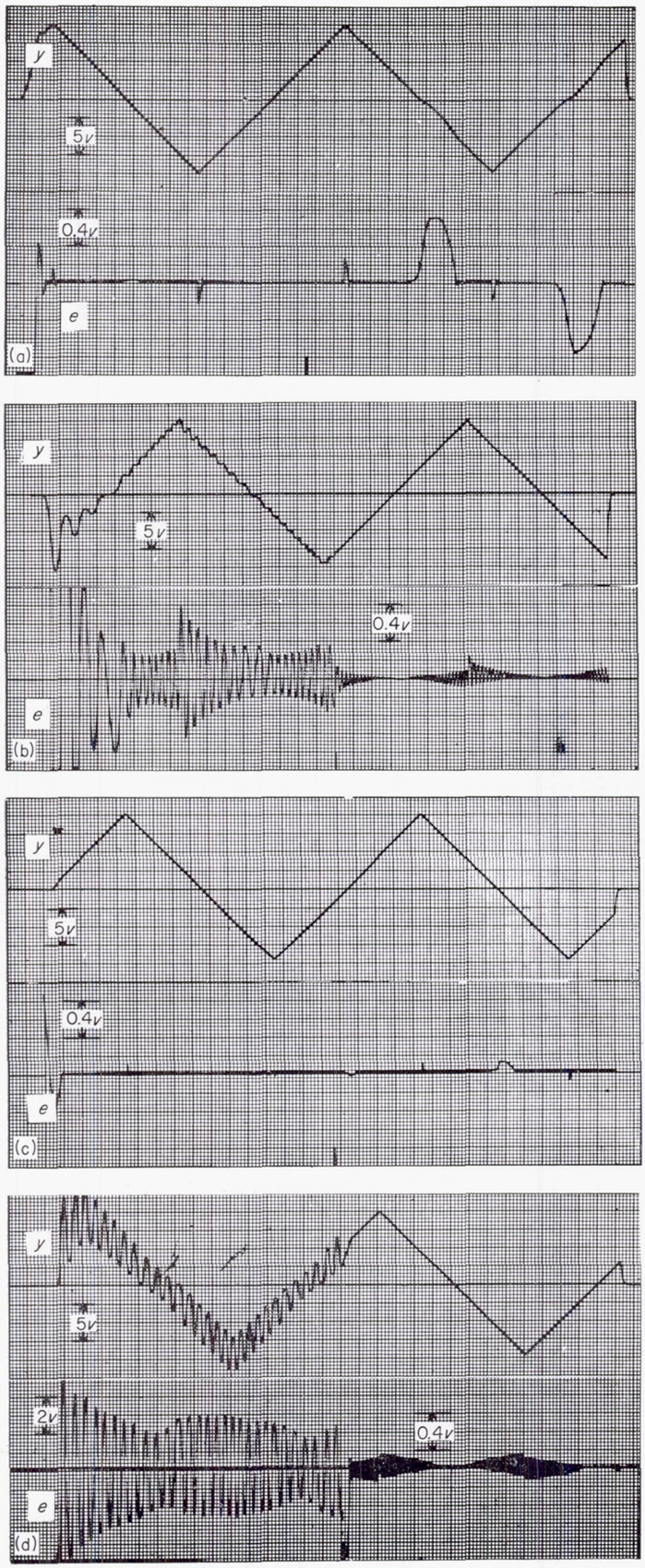


FIGURE 16.—Responses of nonlinear system with four different parameter sets to a 20-volt peak-to-peak triangular-wave input with fixed period $T=10\pi$. $D=0.6$; 2.5 small divisions on time scale=1 normalized time unit; tick marks at bottom of figures indicate demarcation between systems. For systems' numbering see table 3.

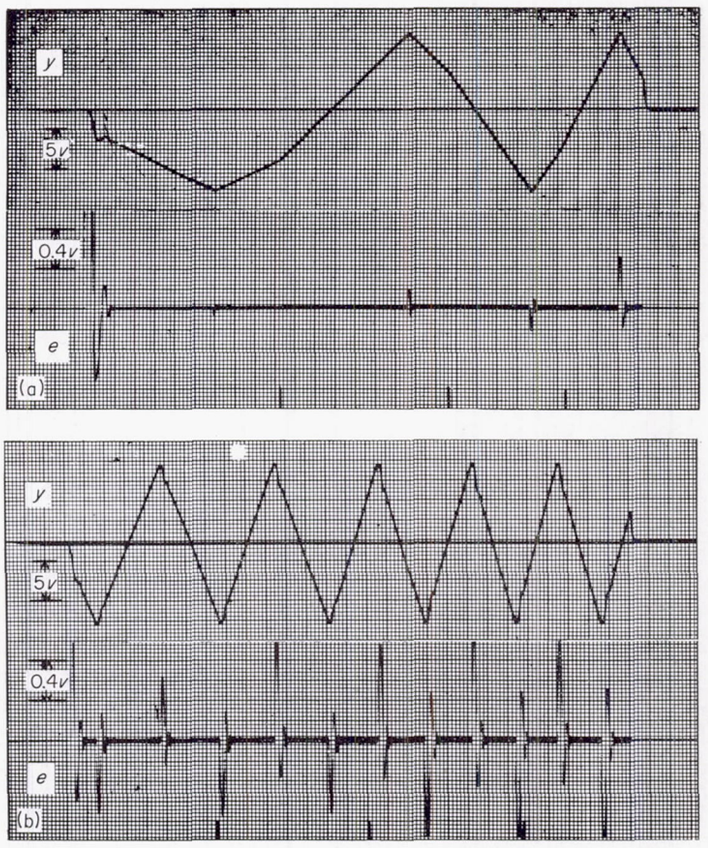
(a) Complete system 1 and system with no derivative feedback 2.
System 1: $\beta_3=-\beta_0=2$; $\beta_2=-\beta_1=0.5$; $\gamma_3=-\gamma_0=2$; $\gamma_2=-\gamma_1=0.5$.
System 2: $\beta_3=-\beta_0=0$; $\beta_2=-\beta_1=0$; $\gamma_3=-\gamma_0=2$; $\gamma_2=-\gamma_1=0.5$.
(b) System with no e' sensing 3 and system with no e sensing in derivative feedback loop and no e' sensing in proportional feedback loop 4.
System 3: $\beta_3=-\beta_0=2$; $\beta_2=-\beta_1=2$; $\gamma_3=-\gamma_0=2$; $\gamma_2=-\gamma_1=2$.
System 4: $\beta_3=-\beta_0=2$; $\beta_2=-\beta_1=-2$; $\gamma_3=-\gamma_0=2$; $\gamma_2=-\gamma_1=2$.
(c) Complete system 5 and system with no derivative feedback 6.
System 5: $\beta_3=-\beta_0=10$; $\beta_2=-\beta_1=1$; $\gamma_3=-\gamma_0=10$; $\gamma_2=-\gamma_1=1$.
System 6: $\beta_3=-\beta_0=0$; $\beta_2=-\beta_1=0$; $\gamma_3=-\gamma_0=10$; $\gamma_2=-\gamma_1=1$.
(d) System with no e' sensing 7 and system with no e sensing in derivative feedback loop and no e' sensing in proportional feedback loop 8.
System 7: $\beta_3=-\beta_0=10$; $\beta_2=-\beta_1=10$; $\gamma_3=-\gamma_0=10$; $\gamma_2=-\gamma_1=10$.
System 8: $\beta_3=-\beta_0=10$; $\beta_2=-\beta_1=-10$; $\gamma_3=-\gamma_0=10$; $\gamma_2=-\gamma_1=10$.

FIGURE 16.—Continued.

In the example of the linear control system of figure 1 and equation (5) this increase in ω_ν by a factor of 15 would mean that the gain K_1 must be increased by the order of magnitude of $(15)^2$ since

$$\frac{K_1 k_m}{I}=(\omega_\nu)^2$$

The gain K_2 must be increased by an order of magnitude of 15 if the value of D shall not change. Such an increase in the loop gains is frequently not at all physically possible.



(a) $\Omega=2\pi/T=0.1$ to 0.4 in 0.1 steps.
(b) $\Omega=2\pi/T=0.5$ to 0.8 in 0.1 steps.

FIGURE 17.—Nonlinear system responses to 20-volt peak-to-peak triangular-wave input with period T varied. $D=0.1$; $\beta_3=-\beta_0=12$; $\beta_2=-\beta_1=3$; $\gamma_3=-\gamma_0=2$; $\gamma_2=-\gamma_1=0.5$; 2.5 small divisions on time scale=1 normalized time unit; tick marks at bottom of figures indicate when period was varied.

Up to this point little attention has been given to the detailed nature of the nonlinear system response. Closer inspection of, for example, figures 6 (a) and 6 (c) shows that the output y is a function that links serpentine fashion (oscillates) at a very high rate about the input x, but still the magnitude of the error is small. In fact, it is necessary to inspect the error at a scale 20 times larger than that of the output even to notice this phenomenon. Mathematically this means that the functions x and y approach one another closely but that their derivatives differ appreciably. Physically, however, this is not at all undesirable as long as the magnitude of the error is small. (Actually for mechanical systems this property would be useful in preventing static friction.) This fine-grained oscillating character of y is the very essence of the nonlinear control theory. Every time the error or error derivative goes through zero the parameter set β_m,γ_n of equation (8) changes discretely as defined in equations (13) or by the binary logic of table 2. The discrete changes in the parameters cause discontinuities in the second derivative y'', which when integrated twice give y its serpentine character. To illustrate this point, the sketch of figure 18 shows samples of the superposed input and output of the nonlinear system. The input in this case could be that of figure 6 (a) or 9 (a).

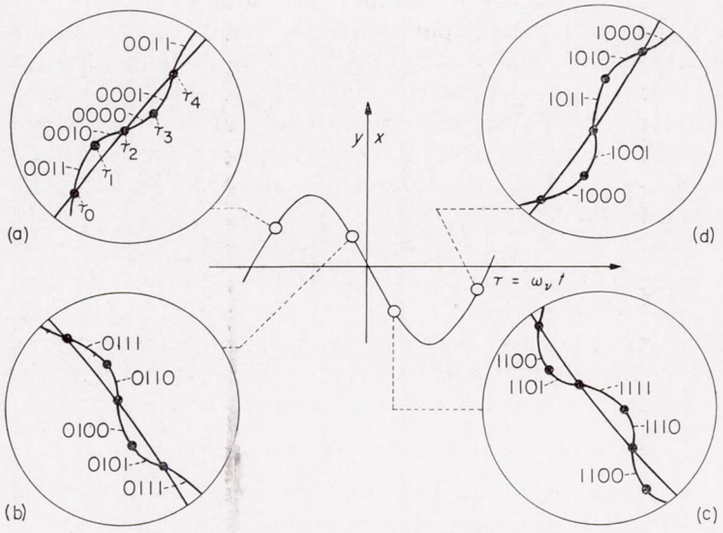


FIGURE 18.—Sketch showing portions of superposed input and output of nonlinear system magnified approximately 100 times.

Circle (a) in figure 18 is a typical cycle of the oscillating character of the output y. Commencing at τ_0, the error changes sign at τ_0, the binary number 0011 is "read into" the feedback switching circuit of figure 5, and using the notation of table 2 the parameter combination $\beta_3\gamma_3$ is switched into the circuit. This causes an immediate reversal in the sign of the output acceleration y'', so that at time τ_1 the error derivative changes sign, the binary number 0010 is read in, and the parameter combination $\beta_2\gamma_2$ is switched into the circuit; acceleration is still in the same direction but weaker. At time τ_2 the error again changes sign, the binary number 0000 occurs, and the combination $\beta_0\gamma_0$ is switched into the circuit; acceleration is in the opposite direction. At time τ_3 the error derivative changes sign, the binary number 0001 occurs, and $\beta_1\gamma_1$ is switched in; this reduces the acceleration until at time τ_4 the error again changes sign, 0011 occurs, $\beta_3\gamma_3$ is again switched in; the cycle is complete. Although it was not mentioned at the time, figure 2 shows this sequence of β_m,γ_n combinations. The nature of the acceleration resulting from the switching can be seen in figure 14 (a).

The comparable switching cycle for $y>0$ and $y'<0$ as shown in circle (b) of figure 18 is:

Logic	Parameters
0111	$\beta_0\gamma_3$
0110	$\beta_1\gamma_2$
0100	$\beta_3\gamma_0$
0101	$\beta_2\gamma_1$
0111	$\beta_0\gamma_3$

Similarly, for $y<0$ and $y'<0$ as shown in circle (c) of figure 18:

Logic	Parameters
1100	$\beta_3\gamma_3$
1101	$\beta_2\gamma_2$
1111	$\beta_0\gamma_0$
1110	$\beta_1\gamma_1$
1100	$\beta_3\gamma_3$

This is the mirror-image switching cycle for circle (a) (see property (4) in the section "Properties of Equation (8)" and also the discussion in the section "Switching Logic").

Finally, for $y<0$ and $y'>0$ as in circle (d) of figure 18:

Logic	Parameters
1000	$\beta_0\gamma_3$
1001	$\beta_1\gamma_2$
1011	$\beta_3\gamma_0$
1010	$\beta_2\gamma_1$
1000	$\beta_0\gamma_3$

This is the mirror-image switching cycle for circle (b).

With this insight into the detailed behavior of the nonlinear system more information can be obtained from the experimental sinusoidal responses of figures 6, 9, and 11 that have up until now been treated from a macroscopic rather than microscopic viewpoint. Along these lines, the following experimentally observed facts are noteworthy:

(1) In comparing the errors for the 4-volt and the 20-volt peak-to-peak sinusoidal inputs in the frequency range of good reproduction $0.1 \leqq \Omega \leqq 0.8$, it is seen that:

(a) The magnitudes of the errors are nearly the same (see figs. 6 (a) and 6 (c) and 9 (a) and 9 (c)).

(b) The period of the error is generally smaller in the larger amplitude case. This is even more pronounced if the 4-volt peak-to-peak case is compared with figure 11 for y in the region of -20 volts.

(2) Good reproduction is characterized by many e and e' switchings (see fig. 18) per cycle of the input. As input frequency is increased and the upper limit of small error is reached (fig. 6 (e) and 9 (e)) the e and e' switchings become more infrequent until there are finally only two of each per cycle of the input.

The fact that the period of the error is smaller for larger inputs can be gleaned from equation (8) when it is rearranged as

$$y''=-[2D(1+\beta_m)y'+\gamma_n y]+(x-y)$$

or roughly

$$y''\approx-[2D(1+\beta_m)x'+\gamma_n x] \tag{18}$$

For a system with specified β_m,γ_n parameter values, the larger input and input derivatives will give stronger discontinuities in y'' as the β_m,γ_n combinations change. This implies that β_m,γ_n will change more often making the period of the error smaller. See figure 14 (a) for the justification of the approximations in equation (18) since y, y', y'', and x are shown in this figure. (The input x was sketched in by hand.)

In the frequency range of good reproduction the reason that the magnitude of the error is virtually independent of the input amplitude cannot be explained from equation (8) since this fact is intimately tied in with the imperfections in the feedback switching circuit (see section entitled "Effects of Switching Imperfections" for details). As measured from the experimental sinusoidal responses, $|e|_{max}$ is of the order of 15 to 20 millivolts. It might be noted that earlier in this section the smaller 4-volt peak-to-peak input sinusoid was employed in the comparison when it was determined that ω_v should be increased by a factor of 15 in order to obtain the same $\overline{\overline{|e|}}_{lin}$. This choice of the smaller input was decidedly in favor of the linear system since the linear error increases linearly with input amplitude. Thus, if the 20-volt peak-to-peak input were used, an increase in ω_v by a factor of $5\times15=75$ would be required to obtain the same $\overline{\overline{|e|}}_{lin}$.

The fact that there is a correlation between good reproduction (small error) and the existence of many error and error-derivative switchings is very important since it is the key to understanding the upper limits of good performance of the nonlinear system. One approximate way to investigate this matter is to establish a deficiency between the output acceleration required for good reproduction (many switchings) and the available output acceleration. Equation (18) gives an approximation to the output acceleration when the system is functioning well. As an example of this approach consider the input x to be a sinusoid of frequency Ω (fig. 6 or 9). Since

$$|x''|_{max}=\Omega^2|x|_{max}$$

it is to be expected that the nonlinear system will have the greatest difficulty in the vicinity of $|x|_{max}$ where, from equation (18),

$$|y''|_{max}\approx|\gamma_n x|_{max}$$

Thus, if in this region the output y is going to interweave the input x (as is characteristic of good reproduction), then

$$|y''|_{max}>|x''|_{max}$$

This inequality then places an upper limit on good performance of the system in response to sinusoids

$$\Omega<\sqrt{|\gamma_n|_{max}} \tag{19}$$

For the system used in obtaining the experimental results of figures 6 and 9, $|\gamma_n|_{max}=2$; therefore,

$$\Omega<\sqrt{2}$$

Inspection of both figures 6 (e) and 9 (e) tends to substantiate the above result. For example, in figure 9 (e) it is seen for $\Omega=1.2$ (after the transient caused by turning on the input has been absorbed) that as the input goes through its maximum, frequent e and e' switchings stop and do not occur again as the frequency is increased.

PERFORMANCE EVALUATION OF INPUTS OTHER THAN SINUSOIDAL

In the preceding section only the sinusoidal response of the nonlinear control system was discussed in studying performance. However, it is easy to see that the nonlinear system will cope with any input in the same manner as it does with sinusoids. Thus, as long as the magnitude of input acceleration does not continuously exceed the available magnitude of the output acceleration the switching process will commence and excellent reproduction will result. This type includes inputs with discontinuous derivatives and discontinuous inputs. The experimental sinusoidal responses of figures 6, 9, and 11 themselves give some indication of these facts. For example, in figure 6 (a) there was an initial discontinuity in the input and there were discontinuities in the input derivative when the frequency was changed. The results in figures 7, 8, 10, and 12 prove further that the nonlinear system response is not dependent upon any specific type of input. Given in these figures are triangular-wave, partially integrated square-wave (including square-wave), clipped sinusoidal, and random[4] inputs, respectively.

It should be noted that in the literature (refs. 1 and 3 to 5) a marginal-type input that would present a case of indecision to the switching circuit of the nonlinear system is discussed. This case has never been encountered experimentally even when the attempt was to produce this case. Thus, the marginal-type input is not considered practically important.

USE OF PHASE-PLANE METHODS TO STUDY PERFORMANCE

The phase-plane methods that were used in the original analytical development of equation (8) (refs. 1 and 3 to 5) can also be gainfully employed in studying the performance of the nonlinear control system derived from equation (8) once the β_m and γ_n parameter values have been specified. (In this section it is important to distinguish between β_m and γ_n parameter values, i. e., $\beta_3=2$, $\beta_2=1$, etc., and $\beta_m\gamma_n$ parameter combinations, i. e., $\beta_3\gamma_0$, etc.)

---

[4] In the strict statistical sense the probability that the inputs shown in figures 12 (a) and 12 (b) are not samples of random stationary time series is admittedly high because of existing experimental facilities. These inputs were obtained by random manual modulation (both frequency and amplitude) of a triangular-wave input of peak-to-peak amplitude A and period T where

$$0\leqq A\leqq20 \text{ volts}$$

$$0.1\leqq\Omega=2\pi/T\leqq0.8$$

Random manual modulation means that the operator varied by hand both the frequency and amplitude controls of the input generator as randomly as possible. In the present investigation the comparative results of the linear and nonlinear responses to what appear to be band-limited random inputs are felt to be more important than the exact statistical properties of the inputs.

If in equation (8) the error $e=(x-y)$ is assumed small, then the output y can be approximated by portions of curves that satisfy the autonomous differential equation

$$\frac{d^2\overline{y}}{d\tau^2}+2D(1+\beta_m)\frac{d\overline{y}}{d\tau}+\gamma_n\overline{y}=0 \tag{20}$$

where D, β_m, and γ_n are defined in the section "Nonlinear Control Theory." Further, if $d\overline{y}/d\tau=\overline{y}'$, these approximating curves are defined in the $\overline{y}'\overline{y}$ phase plane by

$$(\overline{y}\lambda_2-\overline{y}')^{\lambda_2}=M(\overline{y}\lambda_1-\overline{y}')^{\lambda_1} \tag{21}$$

where M is a constant depending on initial conditions and

$$\lambda_{1,2}=-D(1+\beta_m)\pm\sqrt{D^2(1+\beta_m)^2-\gamma_n}$$

Equation (21) comes from integration of the first-order differential equation

$$\frac{d\overline{y}'}{d\overline{y}}=\frac{-2D(1+\beta_m)\overline{y}'-\gamma_n\overline{y}}{\overline{y}'_1}$$

$$=-2D(1+\beta_m)-\gamma_n\frac{\overline{y}}{\overline{y}'} \tag{22}$$

Since the particular β_m, γ_n combination employed in each point of the phase plane depends upon $\overline{y}$, $\overline{y}'$, e, and e', four approximating curves go through each point (see appendix E). The tangents to these curves (eq. (22)) indicate four directions which lie in an angular sector (see fig. 19). This angular sector is defined by the two extreme directions which apply if e and e' have the same sign. The two inner direction values apply if e and e' have opposite signs. By superposing the input $x'x$ phase plane on this $\overline{y}'\overline{y}$ plane it can be stated that at any point the tangent to the input phase curve must

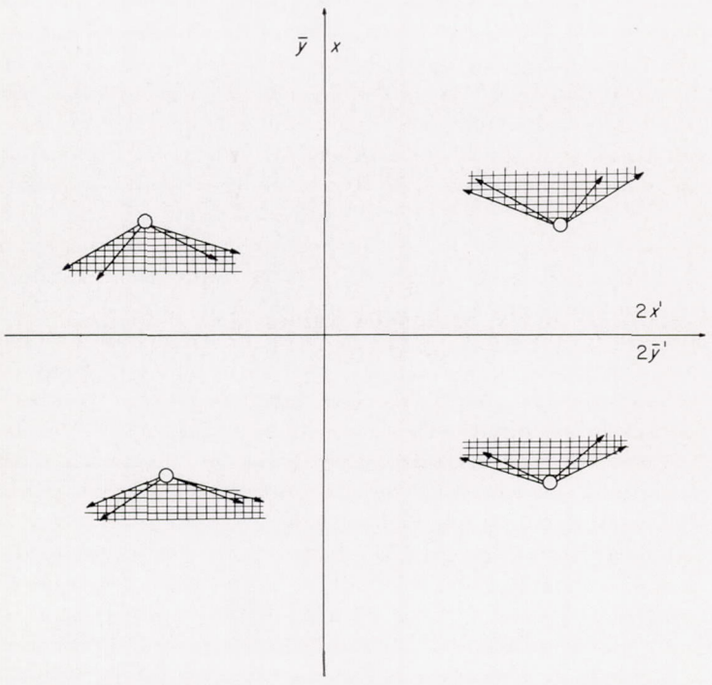


FIGURE 19.—Phase-plane angular sectors defined by equation (22) for the given parameter values. ${}_1\beta=1.5$; ${}_2\beta=0.5$; ${}_1\gamma=0.45$; ${}_2\gamma=0.05$; $2D=0.25$; $\omega_v=2$.

be included within the shaded angular sector (discussed above) if small error is to be obtained. Thus, equation (22) can be used to study performance. The width of the angular sector changes with the ratio $\overline{y}/\overline{y}'$ for given β_m and γ_n parameter values. It is largest (180°) for $\overline{y}'_1\to 0$ and smallest for $\overline{y}=0$ if $(1+\beta_m)$ and γ_n have the same sign. However, the width for $\overline{y}'\to 0$ is not significant since all phase curves (including the input) have infinite slope there. Thus, along the line $\overline{y}'=0$ the curvature of the input phase curve must be used to determine limits on good performance. The radius of curvature of the input must be smaller than that of the flattest approximating curve.

In order to illustrate the use of these phase-plane methods in predicting limits of good performance consider the example of a sinusoidal input $x=\sin\Omega\tau$. In this case the input is represented by an ellipse in the phase plane and

$$\frac{d(x')}{dx}=-\Omega^2\cdot\frac{x}{x'}$$

In the superposed xx' and $\overline{y}\overline{y}'$ phase planes of figure 20 are shown three sinusoidal-input phase curves (i. e., three different frequencies) and the families of phase trajectories of equation (22) for the β_m, γ_n combinations where e and e' have the same sign. (In order to avoid extensive computations the parameter values indicated in fig. 20 are those of an earlier investigation (see ref. 1).) Thus, the tangents to two intersecting phase curves define the angular sector at that point, as has been discussed. By tracing the inputs with different values of Ω it can be understood that good performance for the presented system can be obtained only for $\Omega<\Omega_o$ where Ω_o is the parameter belonging to that ellipse which has the same radius of curvature as the curve through $(\overline{y},\overline{y}')=(1,0)$. (Note that the radius of curvature of the approximating curves jumps at $\overline{y}'=0$ and is smaller for $\overline{y}'<0$.) In the first quadrant (i. e., approaching $\overline{y}'=0$ from $\overline{y}'>0$) this radius of curvature is given by

$$\rho_{max}=|\rho|_{\overline{y}'=+0}=\gamma_n\overline{y}$$

Since at this point the ellipse radius of curvature is $\Omega^2/1$,

$$\Omega^2<\gamma_n\overline{y}$$

Therefore, for the first quadrant

$$\Omega<\sqrt{\gamma_3} \tag{23}$$

Here it should be noted that this is the same result as that obtained in the section entitled "Performance Evaluation of Sinusoidal Inputs" even though the parameter values are different. For the system of figure 20, $\gamma_3=0.5$; thus $\Omega<\sqrt{0.5}$.

Another example is given by the input $x=1-e^{-\alpha\tau}$ with $d(x')/dx=-\alpha$. Since the smallest angular sector is at $\overline{y}=0$, $\alpha=2D(1+\beta_{max})$ determines the limits on performance. This means that for the system represented in figure 20 good control can be expected for a value of α slightly smaller than 0.75.

A step input is represented by $x=1-e^{-\alpha\tau}$ with $\alpha\to\infty$. The picture in the phase plane is $x=1-(1/\alpha)x'$ with $\alpha\to\infty$. For very large values of α this is a straight line which forms a

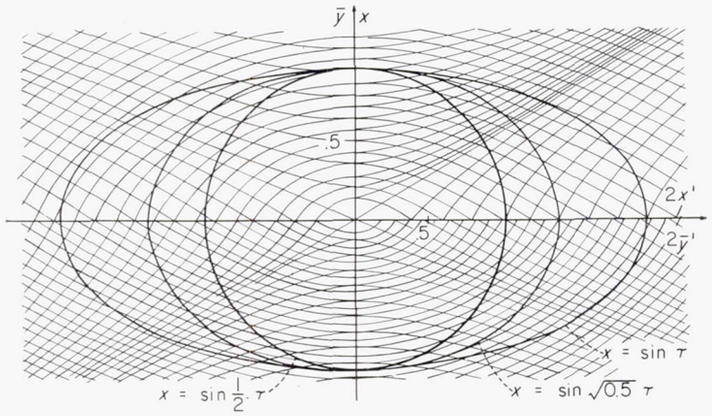


FIGURE 20.—Superposed input and output phase planes showing available angular sectors for sinusoidal inputs. ${}_1\beta=1.5$; ${}_1\gamma=0.45$; ${}_2\beta=0.5$; ${}_2\gamma=0.05$; $2D=0.25$; $\omega_\nu=2$; $|\beta|_{max}=2$; $|\gamma|_{max}=0.5$; $|\beta|_{min}=1$; $|\gamma|_{min}=0.4$.

small angle with $x'=0$. For $\alpha\to\infty$ the curve degenerates to a point $(x=1, x'=0)$. There is no doubt that a perfect followup of a step is not possible because for practical reasons the line for large values of α does not lie in the allowed angular sector at any point it is passing through.

Related to the step input is the square-wave input. A portion of the output phase trajectory for a square-wave input to the system of figure 8 (e) is shown in figure 21. Figure 21 is computed with the help of the differential equation

$$\frac{dy}{dy'_1}=-\frac{y'_1}{2D(1+\beta_m)y'_1+(1+\gamma_n)y-x}$$

This equation is obtained from equation (8) by replacing $d^2y/d\tau^2$ by $y'_1(dy'_1/dy)$. In this example one cannot immediately use approximating curves for designing the output because the error is too big at the start of the motion. The computed diagram is in good agreement with the test run shown in figure 8 (e).

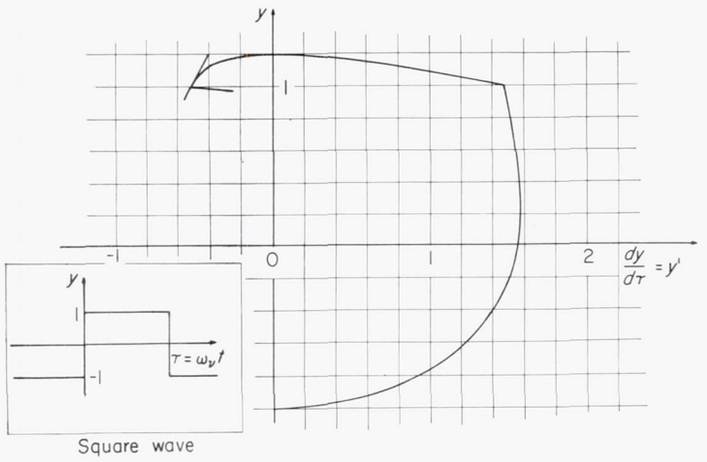


FIGURE 21.—Portion of output phase diagram for square-wave input. $\beta_3=-\beta_0=2$; $\beta_2=-\beta_1=0.5$; $\gamma_3=-\gamma_0=2$; $\gamma_2=-\gamma_1=0.5$; $D=0.6$; $\omega_\nu=1$.

EFFECTS OF SWITCHING IMPERFECTIONS

In equation (8) it is assumed that the parameters β_m and γ_n change upon exact zero coincidence of any one or more of the variables y, y', e, and e' as defined in equation (13) or by the logic of table 2. Physical imperfections, however, preclude this possibility. Thus, in the simulated model of equation (8), exact zero coincidence cannot be detected because of threshold effects, and relays are subject to time lag (both mechanical and electrical), dead zone, and chatter.

In the present section an attempt is made to evaluate the effects that these switching imperfections had on the experimental result in order to obtain some practical design criteria for specifying switching requirements for good performance in other applications of this method of nonlinear control.

The experimental results of figures 6, 9, and 11 can be used to demonstrate that relay imperfections were not important in the simulated model. Comparison of the sinusoidal responses in these figures has shown that the period of the error became smaller as the magnitude of the input amplitude was increased. (This was discussed in the section "Performance Evaluation of Sinusoidal Inputs.") From figure 11 (a) in the region of $|y|\approx 20$ volts (the largest magnitude considered in all the experimental studies) the period of the error T_e is measured as

$$T_e=\omega_\nu t_e\approx 0.2$$

There are four parameter switchings per error cycle (see fig. 18). Assuming these to be approximately uniformly spaced, the minimum time between parameter switchings is approximately $T_e/4$. Now assuming that the relays must be capable of closure in at least one-fifth of this time, the maximum allowable (real time) switching delay t_{md} is

$$\omega_\nu t_{md}\approx T_e/20=10^{-2} \tag{23}$$

In this simulated model $\omega_\nu=1$ radian per second. Therefore,

$$t_{md}\approx 10 \text{ millisec}$$

As given in table 4, the relays employed were capable of closure in 3 milliseconds or less so that they were entirely adequate for the experimental studies.

(After the investigation reported here in the main text was finished, Dr. H. E. Lindberg performed a number of experiments to study quantitatively the influence of time delays on the performance of this control system (see appendix A). He investigated the influence of switching delays in the individual relays and additionally that of a delay in the output line of all relays (represented by an analog-computer approximation of $e^{-\tau_d p}$). The results indicate that for sinusoidal inputs with frequencies up to 0.8 of the natural frequency of the undamped linear member ω_ν, for example, the maximum error can be held within 5 percent of the maximum value of the input with relay delays as large as $47/\omega_\nu$ milliseconds. For a natural frequency of 10 cps this would be a delay of 0.75 millisecond. After this relay delay is selected, the maximum error per input cycle is considerably less for lower input frequencies. For instance, at $\omega=0.1\omega_\nu$ the maximum error per input cycle would be slightly more than 1 percent. Also, the average error would be less than this (see fig. 31).)

The ability to disregard the relays in the evaluation of the effects of switching imperfection on performance leaves only threshold effect in sensing the sign of the variables y, y', e, and e' to be considered. As has been observed and discussed in the section "Performance Evaluation of Sinusoidal Inputs" the amplitude of the error for the nonlinear system was relatively independent of the magnitude of the input (i. e., when the system is operating in the rapid e to e' switching sequence

so that this is the minimum-error case). This constancy of the lower limit on the error magnitude is caused by switching imperfections and thus threshold in sensing the signs of variables. Since under a normal switching sequence there are many more e and e' switchings than y or y' switchings and since the error is of an order of magnitude less than the error derivative, the primary cause of the lower limit on error is localized as threshold in sensing the sign of the error. Figure 13 shows the results of an investigation of this threshold effect. In all system-response figures except figure 13 the peak-to-peak threshold was approximately 14 millivolts. In figure 13 (a) the peak-to-peak threshold was 44 millivolts so that figures 7 (a) and 13 give a good comparison of the effects of these two threshold values.

CHOICE OF PARAMETER VALUES

The performance of a completely specified nonlinear system has been discussed. The parameter values for this system, that is,

$$D=0.6$$
$$\beta_3=-\beta_0=2$$
$$\beta_2=-\beta_1=0.5$$
$$\gamma_3=-\gamma_0=2$$
$$\gamma_2=-\gamma_1=0.5$$

were initially chosen in the following manner: D was first selected to give good linear system performance; then the smallest values of the β_m and γ_n parameters giving good nonlinear system performance were chosen experimentally from a systematic variation of parameters utilizing the simulated system. This particular set of β_m and γ_n parameter values can thus be denoted as $(\beta_m,\gamma_n)_{min}$ since they establish the lower bound on parameter values for good nonlinear system performance. The physical significance of $(\beta_m,\gamma_n)_{min}$ is that loop gains and acceleration requirements of the linear member are minimized since β_m and γ_n are feedback gain constants.

From the discussion in the section "Performance Evaluation of Sinusoidal Inputs" centering about equation (18) or from the phase-plane methods of the preceding section it is to be expected that a general increase in the parameter values over $(\beta_m,\gamma_n)_{min}$ will result in improved nonlinear system performance by increasing the available acceleration of the system or increasing the angular sectors in the phase plane. Inspection of figures 16 (a), system 1, and 16 (c), system 5, shows this to be true. In figure 16 (a) the parameter values were $(\beta_m,\gamma_n)_{min}$ as given above, and in figure 16 (c), system 5, they were

$$\beta_3=-\beta_0=10$$
$$\beta_2=-\beta_1=1$$
$$\gamma_3=-\gamma_0=10$$
$$\gamma_2=-\gamma_1=1$$

On comparing the performance of these two systems it is noted that the corners of the input triangular wave are reproduced with less error by the system of figure 16 (c). Thus far then it would seem that there is no upper bound on the parameter values; that is

$$(\beta_m,\gamma_n)_{max}\rightarrow\infty$$

Invariably, however, there will be upper bounds on the values of the parameters because of acceleration limits in the physical system. Figure 14 shows the effects of placing progressively smaller acceleration limits on the system. Here it is seen that for $|y''|_L>0.3|y''|_{max}$ performance is not appreciably affected but for values less than this good performance is no longer obtained so that acceleration limits definitely tend to determine $(\beta_m,\gamma_n)_{max}$. In general, then, there will be a whole range of values of β_m and γ_n; that is,

$$(\beta_m,\gamma_n)_{max}>(\beta_m,\gamma_n)>(\beta_m,\gamma_n)_{min}$$

for which good nonlinear system performance results. The final choice must depend upon the particular application and can easily be found experimentally.

There are certain special cases of the β_m and γ_n parameter values that lead to simplified feedback switching circuits and thus lead to nonlinear systems that are simpler to realize. In table 3 three of these are denoted as case (1) no derivative feedback, case (2) no e' sensing, and case (3) no e sensing in derivative feedback loop and no e' sensing in proportional feedback loop.

By making inoperative the appropriate relays in figure 5 the simplified switching circuits for these cases are easily obtained. It is desirable then to inspect the performance of these special cases (in comparison with that of the complete system) to ascertain whether or not as good performance can be obtained with less complexity. Figure 16 shows the response of these cases in comparison with that of the complete system. Figures 16 (a) and 16 (b) differ from figures 16 (c) and 16 (d) in that the general magnitude of the parameters was increased in figures 16 (c) and 16 (d).

These results show that case (2) is not worthy of much consideration since the magnitude of the oscillating error is large. Cases (1) and (3), however, should be considered for certain applications. For example, if it were known that the amplitude distribution of the input was relatively void near zero, then case (1) would serve as well as the case of the complete system. Case (3) shows nearly constant percentage error so that it could be useful in cases where accuracy was not so important as economy in components.

To this point the parameter D has received little attention mainly because its value (within limits) is not particularly important. It has been observed experimentally that D may be anywhere in the range $0\leqq D\leqq 1$ even for $(\beta_m,\gamma_n)_{min}$ and performance of the nonlinear system is not affected. Inspection of the block diagram of the nonlinear system (fig. 4 or 5) shows that the physical significance of the variation of D for given values of β_m,γ_n is that the damping factor of the linear member and the derivative feedback around the linear member change simultaneously. A case of more practical importance such as might arise in aerodynamic applications of this type of control system is the variation of D for given values of $D\beta_m,\gamma_n$. That is, the damping factor of the linear member alone varies while the

feedback values around this member remain unaffected. Even in this case it has been found that the nonlinear system performs well. For example, figure 17 shows the response of the nonlinear system

$$D=0.1$$
$$\beta_3=-\beta_0=12$$
$$\beta_2=-\beta_1=3$$
$$\gamma_3=-\gamma_0=12$$
$$\gamma_2=-\gamma_0=0.5$$

to triangular-wave inputs, while figure 7 shows the response of the system

$$D=0.6$$
$$\beta_3=-\beta_0=2$$
$$\beta_2=-\beta_1=0.5$$
$$\gamma_3=-\gamma_1=2$$
$$\gamma_3=-\gamma_0=0.5$$

to the same inputs. Comparing these results it is seen that performance is nearly the same for both systems.

PRELIMINARY EXTENSION TO HIGHER ORDER SYSTEMS

From a practical standpoint, limitations in the applicability of the nonlinear control system described by equation (8) do not stem from the inability to realize the feedback switching circuit but rather from assuming the linear member to be of second order. In many cases a more realistic approach is to consider the linear member to be of higher order but still predominantly second order. Figure 22 shows the block diagram of a third-order system of this nature. The linear feedforward member could be, for example, the servo of figure 1 including the effects of armature inductance. The same second-order feedback switching circuit was still employed. However, it could not be

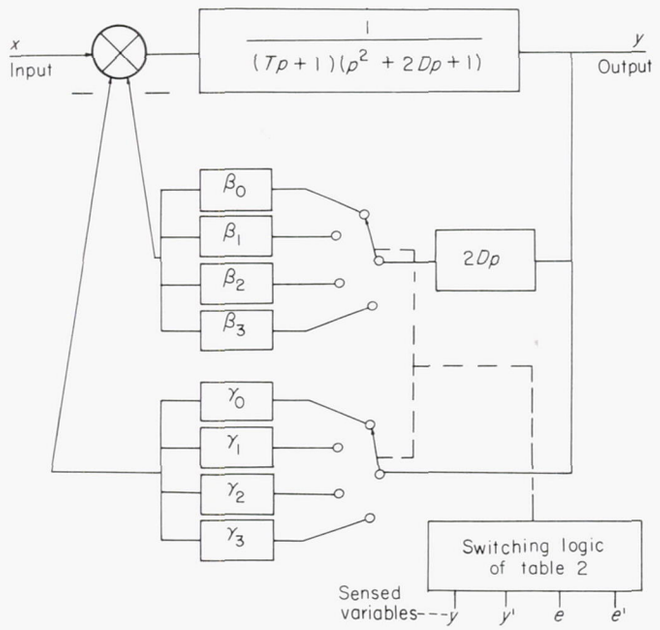
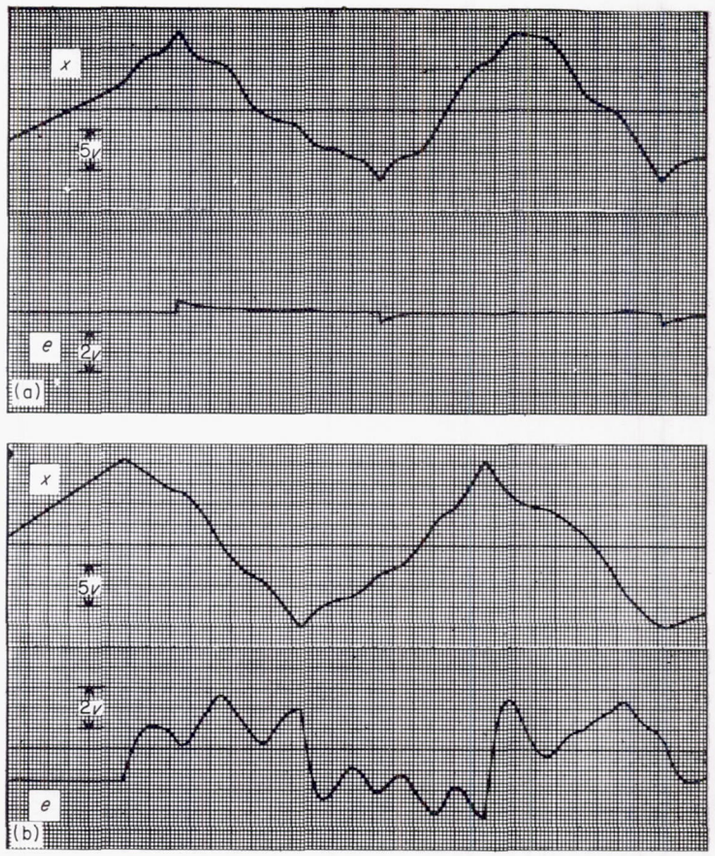


FIGURE 22.—Block diagram of third-order nonlinear control system.



(a) Nonlinear system. $\beta_3=-\beta_0=2$; $\beta_2=-\beta_1=0.1$; $\gamma_3=-\gamma_0=2$; $\gamma_2=-\gamma_1=-0.5$.
(b) Linear system.

FIGURE 23.—Linear and nonlinear third-order system responses to random input. $T=0.1$; $D=0.6$; 2.5 small divisions on time scale=1 normalized time unit.

expected that the β_m,γ_n parameter values remain the same. Figure 23 compares the response of a third-order linear and nonlinear system. Here it is seen that the nonlinear system still responded with much less error than the linear system.

On the basis of the comparative performance of these two systems it seems important to further studies toward control of general higher order systems using techniques similar to those developed in this investigation.

CONCLUDING REMARKS

From control equation (8) a second-order nonlinear control system that tends to maintain small instantaneous error for relatively arbitrary inputs has been synthesized using digital-computer techniques. The only type of nonlinear operations required in the realization are switching-type operations (zero-coincidence detection and parameter switching). The switching requirements are severe from the aspect of detector sensitivity and switching time delay but not impractical since all the switching is done in feedback paths at low power levels and may thus be performed electronically.

The system demands sensing of error and error rate of change. Since general noise in a system of this type has to be expected, both error and error rate are smoothed. The influence of time delay in the relays and of a general transport delay on the performance of the system has been investigated

and diagrams have been developed which allow one to estimate the maximum error due to imperfections.

It is a difficult task to compare the performance of a nonlinear system with that of a linear system because no general criterion for comparison is available. Since for nonlinear systems the law of superposition does not hold, it is not adequate to choose the response to a certain input (e. g., the step input) as a criterion for performance comparisons. A number of different inputs therefore have been chosen for exhibiting the qualities of the nonlinear system.

Experimental results indicate that this type of nonlinear control system performs better than a linear control system having a normalized frequency 15 times greater. Performance is evaluated in terms of the average value of the magnitude of the instantaneous error for band-limited inputs. Further, the nonlinear system performance is virtually independent of variations in the damping factor of the system.

STANFORD UNIVERSITY,
STANFORD, CALIF., *October 18, 1957.*

APPENDIX A

INFLUENCE OF TIME DELAYS

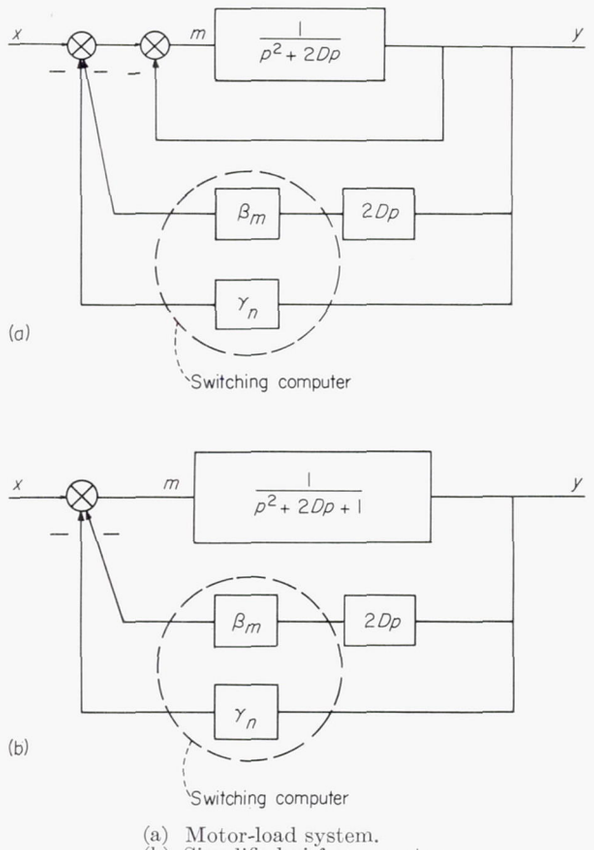
By H. E. Lindberg

In the experiments reported in the main text of this report, the relays were carefully selected in order to keep the bad effects of finite closure time to a minimum and no really quantitative study of such closure delays was made at that time. Later, studies of this particular system were continued and quantitative information about the influence of imperfections on the performance of the control system is presented here.

One numerical and two experimental methods of finding response with relay delays are used. Plots of maximum error incurred per input cycle are given for sinusoidal inputs of various frequencies and for various relay delays. Effects of a transport time delay are also discussed.

INFLUENCE ZONES AND TYPES OF DELAYS

The system described by equation (8) can be interpreted as describing either of the systems shown in figure 24 if there are no imperfections to be considered. However, there are certainly relay imperfections and there may be transport delays. Relay imperfections would affect directly those parts of the systems that are encircled by dotted lines in figure 24, that is, only the discontinuous feedback paths. If no further imperfections are considered, the results of such a study apply equally well to both configurations shown in figure 24 because the switching computer sees the same linear member in either case. Other studies, however, can conceivably be made on these systems where the difference between the two configurations becomes significant. A study of transport delays introduced at points m, for instance, would certainly give different results for figures 24 (a) and 24 (b) because in figure 24 (a) the output y is fed directly through this point while in figure 24 (b) no such feedback is present because the equivalent of such a feedback is built into the differential equation of the mechanism being controlled (equivalent, i. e., when there are no imperfections).



(a) Motor-load system.
(b) Simplified airframe system.

FIGURE 24.—Block diagrams of two physical systems giving same differential equation providing no imperfections exist. m, point transport delay is introduced into circuit.

One of the most serious imperfections to be considered in the design of a relay servomechanism is the time lag between the application of a signal to the relay and the actual closing or opening of the relay. Associated very closely with this imperfection are other phenomena, such as signal threshold, relay dead time (the period of time in which the relay arm is in contact with neither terminal), and contact bounce. The only imperfection studied here is relay time lag, because in most relays dead time is very small as compared with the relay lag time. Contact bounce is also a secondary effect and is better studied in an actual installation. By appropriate interpretation, the results of this time-lag study can be extended to include the effect of signal threshold which is, of course, a variable time delay.

Rather than increasing the natural frequency of the system being studied until the effect of inherent relay lag became noticeable, an artificial lag was introduced while the system was allowed to operate in a conveniently low frequency range. Two techniques were used to provide the time lag. The first was a direct delay of each relay signal, effected by providing an additional relay in series with each original relay and a resistance-capacitance delay circuit in between (this will be referred to as the electromechanical simulation). The second method consisted of inserting an analog-computer approximation of $e^{-t_d p}$ in the total output line of all of the relays. The two methods gave consistent results.

ELECTROMECHANICAL SIMULATION OF TIME DELAY

The armature of each logic relay in the control system with no intentional delays was energized by the plate current of a zero-coincidence detector circuit. In the delayed system, the plate current is made to flow through an added relay. This relay energizes an adjustable resistance-capacitance time-delay circuit which in turn feeds the armature of the logic relay. Two of the four circuits are shown in figure 25.

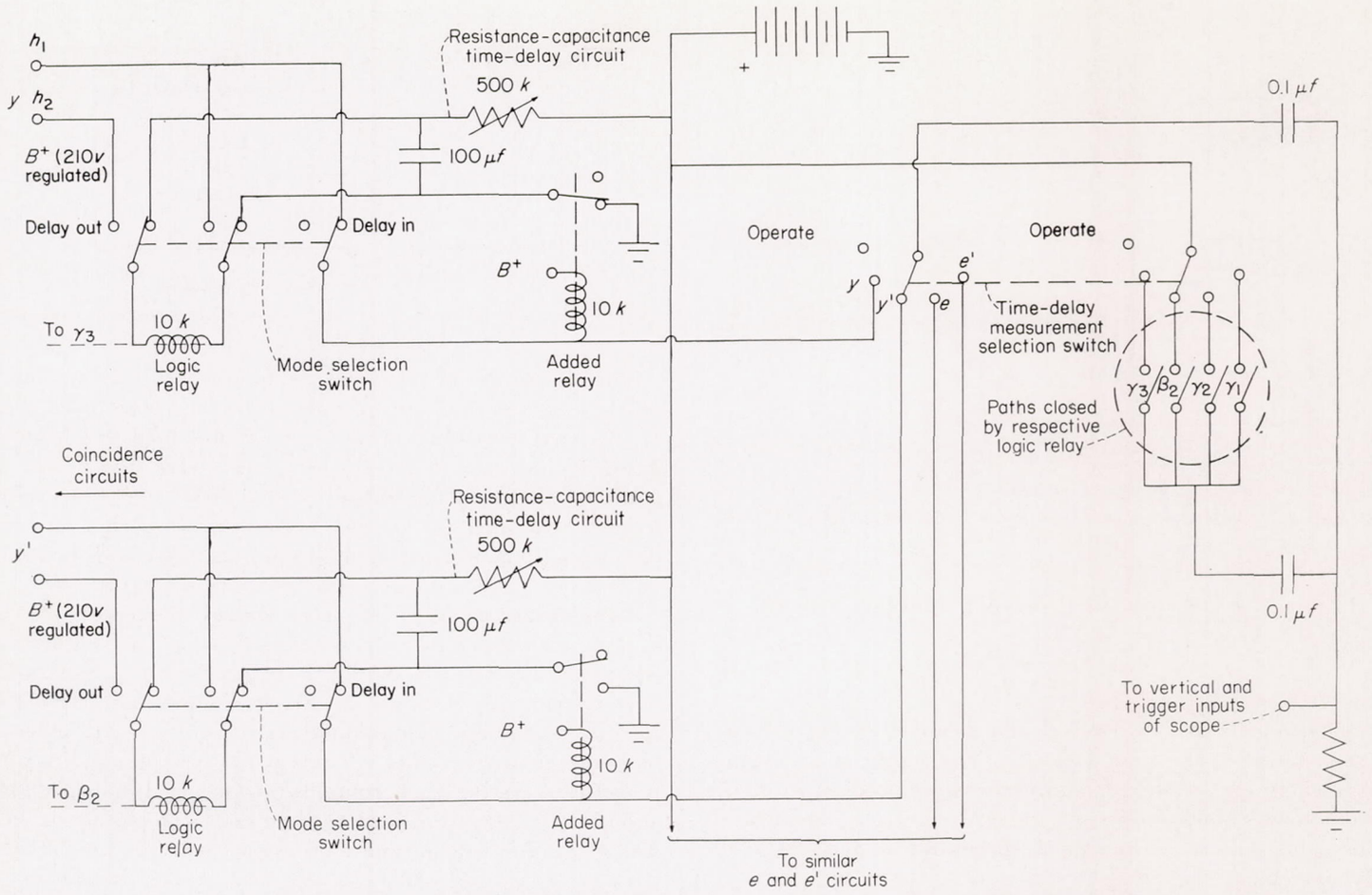


FIGURE 25.—Schematic diagram of electromechanical switching-delay simulation.

The amount of delay between activation of the zero-coincidence circuit and the closing or opening of the logic relay was measured by observing the coincidence pulse and the logic-relay contact pulse with a long-persistence oscilloscope whose beam traversed the scope at a known rate and was triggered by the coincidence pulse. The voltage supply in the time-delay circuit was adjusted so that the closing and opening delays were roughly equal. It is estimated that the accuracy of time-delay measurement was of the order of 10 percent. Delays of less than 25 milliseconds ($t=\tau$ with the analog-computer wiring used) were difficult to measure by this method, and, more important, could not be easily simulated by this particular circuit because of high contact current in the added relay which caused the contacts to weld.

The delayed system was operated using sinusoidal inputs of 20-volt peak-to-peak amplitude at frequencies from 1/10 the natural frequency of the undamped linear system to the full value of the natural frequency. Nondimensional time delays of $\omega_\nu t_d=0.05, 0.075, 0.100$, and 0.150 were used. Knowing that the error caused by these delays should vary linearly with the applied voltage (see the section "Response to sinusoidal inputs" which follows), runs were also made at 10- and 40-volt peak-to-peak amplitudes as a check of the self-consistency of the data.

Analog-computer simulation of time delay.—An analog-computer approximation of $e^{-t_d p}$ was wired using a circuit suggested by Morrill (ref. 6). The circuit is based on the Padé second-order approximation of $e^{-t_d p}$

$$\frac{12-6t_dp+t_d^2p^2}{12+6t_dp+t_d^2p^2}=1-t_dp+\frac{t_d^2p^2}{2}-\frac{t_d^3p^3}{6}+\frac{t_d^4p^4}{24}-\frac{t_d^5p^5}{144}+0t_d^6p^6-\frac{t_d^7p^7}{1{,}728}+\cdots \quad \text{(A1a)}$$

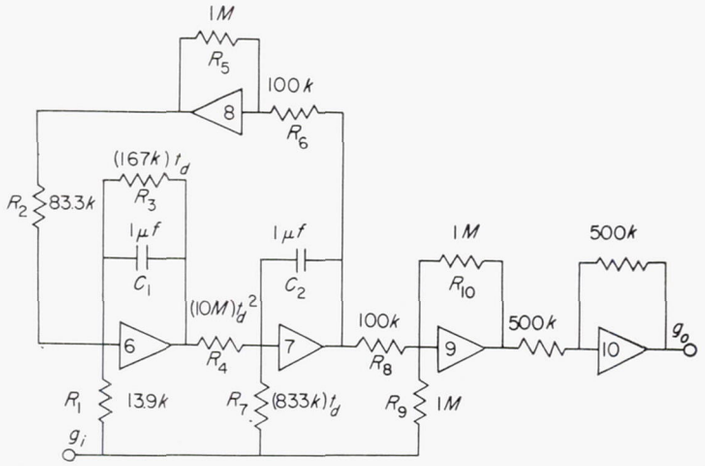
which is roughly equivalent to the first six terms of the Taylor expansion of $e^{-t_d p}$ about $t_dp=0$:

$$e^{-t_dp}=1-t_dp+\frac{t_d^2p^2}{2}-\frac{t_d^3p^3}{6}+\frac{t_d^4p^4}{24}-\frac{t_d^5p^5}{120}+\frac{t_d^6p^6}{720}-\cdots+\cdots \quad \text{(A1b)}$$

The circuit is shown in figure 26. The element values were found by comparing the transfer functions of the circuit with the desired Padé polynomial fraction and equating coefficients of powers of p. The transfer function of the circuit in a nondimensional form is:

$$\frac{g_o}{g_i}=\frac{R_{10}}{R_9}-\frac{\dfrac{R_{10}}{R_1}-\dfrac{R_4R_{10}}{R_3R_7}-\dfrac{R_4R_{10}C_1}{R_7}\,p}{\dfrac{R_8R_5}{R_2R_6}+\dfrac{R_4R_8C_2}{R_3}\,p+(R_4R_8C_1C_2)p^2} \quad \text{(A2)}$$

where the resistances R_n are as shown in figure 26.



$$\frac{g_o}{g_i}=\frac{12-6t_dp+t_d^2p^2}{12+6t_dp+t_d^2p^2}\approx e^{-t_dp}$$

FIGURE 26.—Analog-computer circuit for Padé second-order approximation for e^{-t_dp}.

Comparing this with the Padé approximation written

$$e^{-t_dp}\approx 1-\frac{12t_dp}{12+6t_dp+t_d^2p^2} \tag{A3}$$

6 equations for the 10 resistors and 2 capacitors can be written in terms of the time delay t_d. The resulting six degrees of freedom on the circuit parameters were used to select the convenient values shown. Figure 27 shows the location of this delay circuit in the overall simulation of the relay control system.

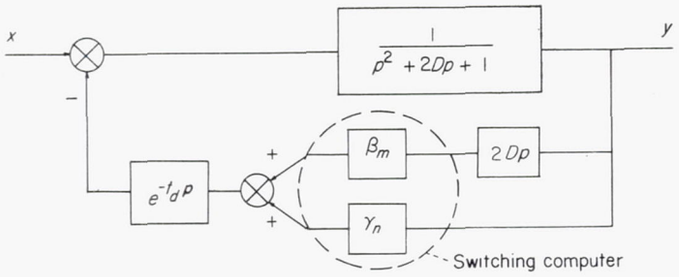


FIGURE 27.—Insertion of time-delay circuit of figure 26 into control-system analog.

Since this method was used only as a check on the electromechanical method, the only input studied was a constant of 10 volts at delays of $t_d=0.025, 0.050, 0.075$, and 0.100 second.

Simulation of a time delay by this method allowed greater flexibility of the location of the time delay. The delay was inserted as shown in figure 28 to simulate the effect of a transport time delay in the system. The output for this configuration was essentially the same as that found for relay delays. This result was to be expected because the essential difference between a transport delay and relay delay in this system is that relay delays affect only the feedback terms which go through the relays while transport delays affect the input x as well. But for the great majority of operation, x changes very slowly as compared with the changes in the feedback terms and hence, it makes very little difference in the response whether the input x is delayed.

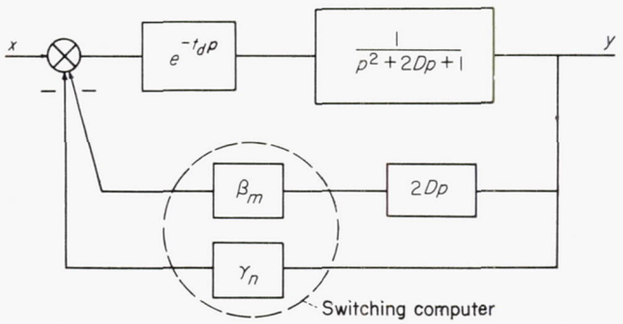


FIGURE 28.—Insertion of time-delay circuit for study of transport delays.

The responses of the two configurations become more nearly identical as the magnitude of the time delay decreases. If the basic controlled element is a motor load as shown in figure 24 (a) and if a transport delay were introduced at point m, the resulting output would differ somewhat more from a relay-delayed output because then both x and the direct y path as well as the discontinuous relay signals would be delayed. No tests were run with this configuration.

Computational study of relay delay.—Response of the system with relay lag was computed using the differential equations directly. To simplify the amount of calculations, only response to constant inputs were computed. Calculations were based on approximating the output with a series of parabolic arcs. This is equivalent to assuming that y'' is constant during each computation interval, a reasonable assumption except for large time delays. The input studied was $x=10$ and calculations were made until the limit cycles on a plot of E against $\frac{dE}{d\tau}$ were approached within the accuracy of computation. Computations were made such that the limit cycles were approached from the inside in some cases and from the outside in others. Limit cycles for various time delays are shown in figure 29. Computation intervals were such that two parabolic arcs were used in each quadrant.

The computation details are as follows: With the assumption that the output is a series of parabolic arcs, the equation for the error during the nth interval between switch points is:

$$E_n=E_{ni}+E_{ni}'\tau_n+\tfrac{1}{2}E_{ni}''\tau_n^2 \tag{A4}$$

where $\tau_n=0$ at the start of any interval and E_{ni}, E_{ni}', and E_{ni}'' are the initial error, error velocity, and error acceleration, respectively, for the nth interval. Upon differentiating, the velocity during any interval becomes:

$$E_n'=E_{ni}'+E_{ni}''\tau_n \tag{A5}$$

Eliminating the time τ_n between these two equations, the equation of the phase-plane trajectories becomes:

$$E_n-\left[E_{ni}-\frac{1}{2}\frac{(E_{ni}')^2}{E_{ni}''}\right]=\frac{1}{2E_{ni}''}(E_n')^2 \tag{A6}$$

Equations (A4) and (A5) are used to find E_{as} and E_{as}' at an actual switch point by letting $\tau_n=\omega_v t_d$ (the delay time) and using as initial conditions E at $E'=0$ or E' at $E=0$ (the conditions at a theoretical switch point). When this point is found, equation (A6) is used with $E_{ni}=E_{as}$ and $E_{ni}'=E_{as}'$ and the next theoretical switch point is computed

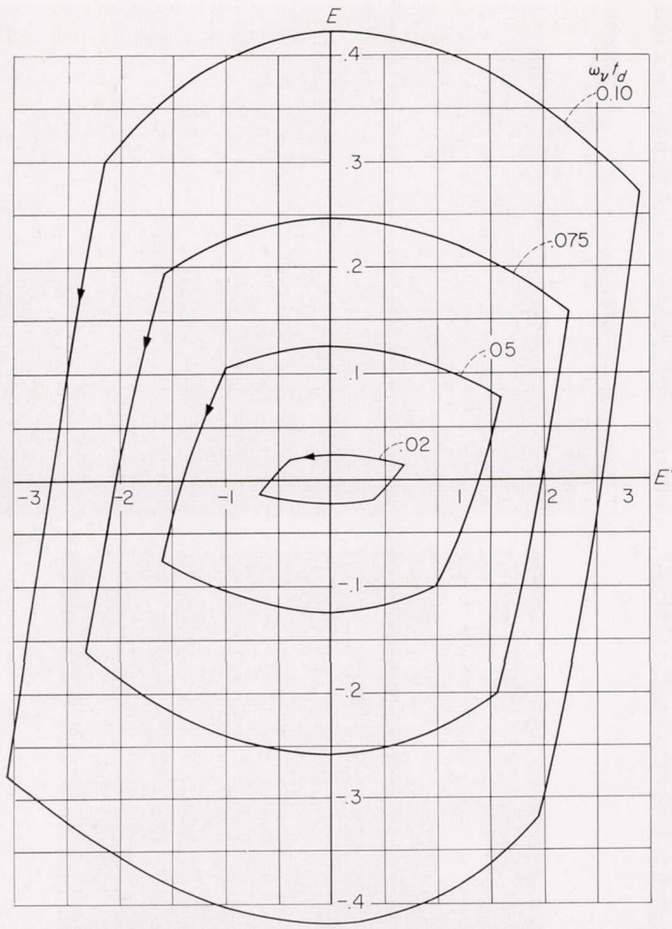


FIGURE 29.—Error limit cycles due to switching delays for a constant input of $x=10$ volts.

by setting either $E_n=0$ or $E_n'=0$ depending on which type of switch point is being approached. In all cases the acceleration E_{ni}'' is found from the differential equation of the control system, equation (8). For the special case where $x=10$, equation (8) can be rewritten using $y=E+x$ as:

$$E''+2D(1+\beta_m)E'+(1+\gamma_n)E=-10\gamma_n \qquad \text{(A7)}$$

An effort was made to determine some general method for predicting the limit cycles as an analytic function of t_d, D, β_m, and γ_n but, even with this very simple input, the limit cycles are approached in such a complicated way that finding such a function would be a formidable task at best. However, a systematic series of calculations as shown here could be made and charts plotted to indicate the effect of these various parameters.

Comparison of the three methods for constant inputs.—For a constant input of $x=10$ a good comparison of results from the three methods of studying relay delay was made by constructing a logarithmic plot of error versus time delay. Data from the two experimental methods had to be modified slightly to account for actual relays delays and thresholds of zero coincidence. The modification consisted of measuring the system error with no intentional delay, finding the delay required to produce this error from the calculated plot of error versus delay, and adding this delay to each of the experimental points. (The effective unintentional relay delay was found to be 20 milliseconds. This is considerably greater than the 3-millisecond delay specified by the relay manufacturer. The additional effective delay is due to the inherent thresholds in the switching circuit which can be considered as a variable time delay. Also, the actual delay of the relays may have exceeded 3 milliseconds.)

After this modification, the analog-computer data were indistinguishable from the calculated data. Data from the electromechanical method also agreed quite well but differed from the computed results by as much as 10 percent at a time delay of 0.12. (See fig. 30.)

Response to sinusoidal inputs.—As mentioned before, the error resulting from relay lags varies linearly with the magnitude of the input x. That this is true is easily seen by examining equation (8). If x is replaced by Kx and y by Ky where K is a constant, the differential equation remains unchanged and the error $e=x-y$ is now $e^*=Kx-Ky=Ke$.

The reader should be warned that although this is a well-known property of linear differential equations, this system remains essentially nonlinear. Using the definitions of β_m and γ_n given in the text following it, equation (8) can be written:

$$y''+2D(\text{sgn } y'-{}_1\beta \text{ sgn } e-{}_2\beta \text{ sgn } e')|y'|+$$
$$(\text{sgn } y-{}_1\gamma \text{ sgn } e-{}_2\gamma \text{ sgn } e')|y|=x \qquad \text{(A8)}$$

With the equation in this form, the reader can easily verify that if y_1 and y_2 are solutions to equation (A8) in its homogeneous form ($x=0$) then y_1+y_2 is not necessarily a solution

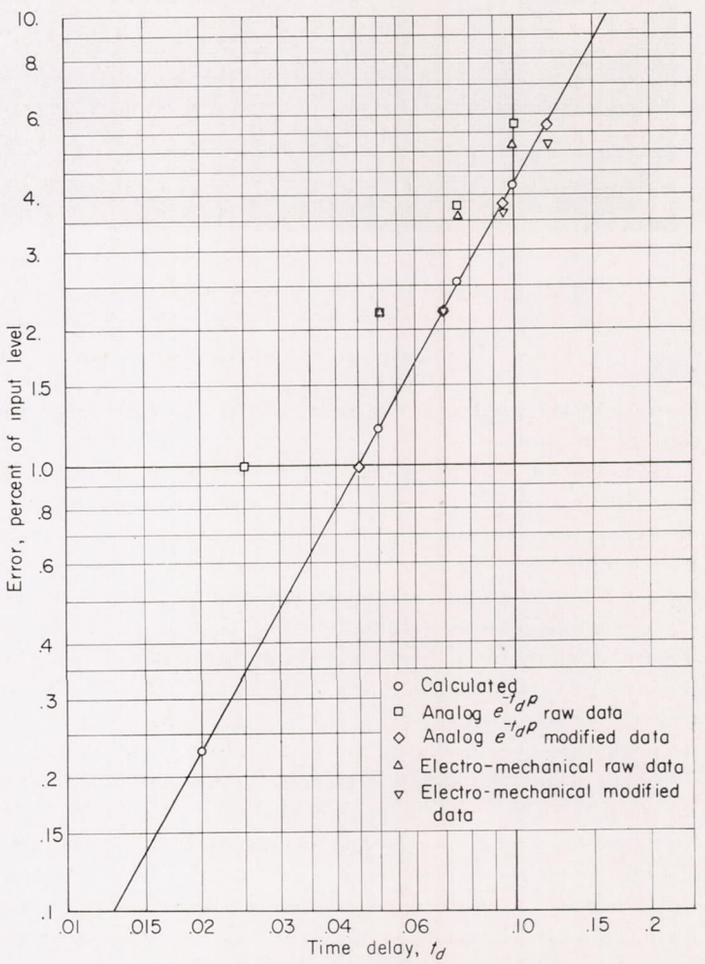
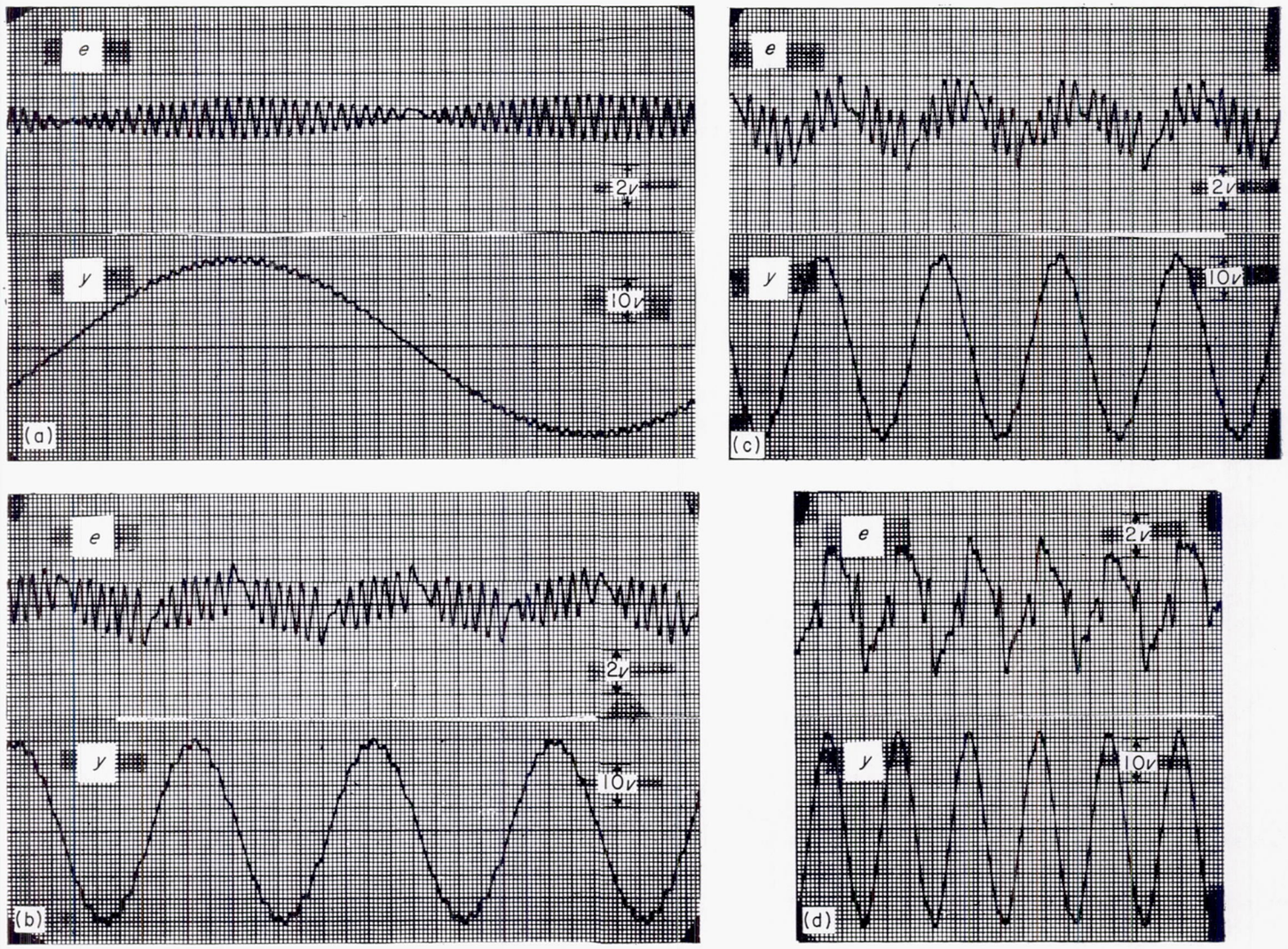


FIGURE 30.—Comparison of methods of switching-delay study with a constant input of $x=10$ volts.



(a) $\omega=0.1\omega_\nu$.
(b) $\omega=0.4\omega_\nu$.
(c) $\omega=0.6\omega_\nu$.
(d) $\omega=\omega_\nu$.

FIGURE 31.—Responses for inputs of various frequencies with relay delays of $\omega_\nu t_d=0.10$.

because of the nonlinear character of the absolute value and signum functions.

However, it can be concluded from the linear property that if the error encountered for any constant input is obtained, the error for any other constant input can be found by a simple scale change. For very low frequency sinusoidal inputs where x' and x'' are relatively small it was found that the error was governed almost entirely by this consideration. This is demonstrated by figure 31 (a) for a 40-volt peak-to-peak input with $\omega=0.1\omega_\nu$ and a delay of $\omega_\nu t_d=0.100$. Notice that the envelope of the error is very nearly a sine wave with its peak value occurring at the peak value of the input.

As the input frequency is increased, the point of greatest error shifts away from the peak of the input. In figure 31 (b) for $\omega=0.4\omega_\nu$ the greatest error occurs almost ¼ cycle after the input reaches its peak in either direction. This is because the $(1+\beta_m)y'$ term becomes sizable in this frequency range during the acceleration period (i. e., as $|x'|$ increases) and reaches a maximum at ¼ cycle past the input peak. This large feedback term is applied in an undesired direction for 0.10 unit of nondimensional time and causes the large lag error.

At still higher frequencies the y'' term coupled with the $(1+\gamma_n)y$ term both reach their maximum at the peak value of the input and the greatest error again occurs near this time. An example of this is given in figure 31 (d) for an input frequency of $\omega=1.0\omega_\nu$. For intermediate frequencies the peak shifts around depending on initial conditions as shown for $\omega=0.6\omega_\nu$ in figure 31 (c).

From an overall standpoint a plot of the greatest error versus frequency showed an approximately linear increase of error with frequency (see fig. 32). Similarly, a plot of greatest error versus time delay is linear up to about $\omega_\nu t_d=0.10$ for frequencies up to $\omega=0.8\omega_\nu$ (see fig. 33). An interesting comparison of these results and the results found for the system with no intentional delays is shown in figure 34. Also plotted in this figure is the maximum error for the linear system alone $(y''+2Dy'+y=x)$.

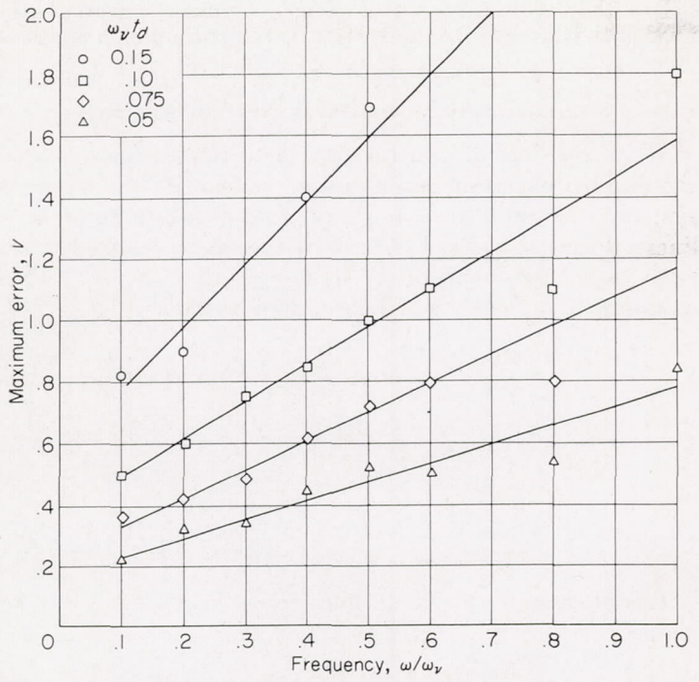


FIGURE 32.—Maximum error occurring during each input cycle plotted against input frequency with switching delay as a parameter.

$$x=10 \sin \omega t=10 \sin\left(\frac{\omega}{\omega_\nu}\right)\tau \text{ volts; } \beta_3=-\beta_0=2;$$

$$\beta_2=-\beta_1=0.5;\ \gamma_3=-\gamma_2=2;$$

$$\gamma_2=-\gamma_1=0.5;\ D=0.6;\ \omega_\nu=1.0.$$

Specification of time-delay tolerances.—If an actual system were to be set up such as that described by equation (8), figure 32 would provide an indication to the designer of how to specify relay-delay tolerances. Take, for example, a system which has an undamped natural frequency $f_\nu=10$ cps. If the error of this system is to remain below 5 percent of the maximum value of the input for frequencies up to $\omega=0.8\omega_\nu$, then the relay delays must be held below 0.75 millisecond. This result is found by first finding that value of $\omega_\nu t_d$ on the $\omega=0.8\omega_\nu$ curve of figure 33 at which the maximum error is 5 percent. This value of $\omega_\nu t_d$ is 0.047. The actual time delay t_d is:

$$t_d=\frac{0.047}{\omega_\nu}=\frac{0.047}{2\pi f_\nu}=\frac{0.047}{20\pi}=0.00075 \text{ sec} \qquad \text{(A9)}$$

This value of time delay will not necessarily assure that the error will always remain within 5 percent of the input for other types of inputs which are band-limited to $0.8\omega_\nu$ because of the nonlinear nature of the system. However, it is known from the special type of inputs studied that if the error is to be as specified the time delay can be at most 0.75 millisecond.

For sinusoidal inputs with $\omega<0.8\omega_\nu$ the maximum error will be less than 5 percent of the maximum value of the input. For example, at $\omega=0.1\omega_\nu$ the maximum error will be slightly more than 1 percent.

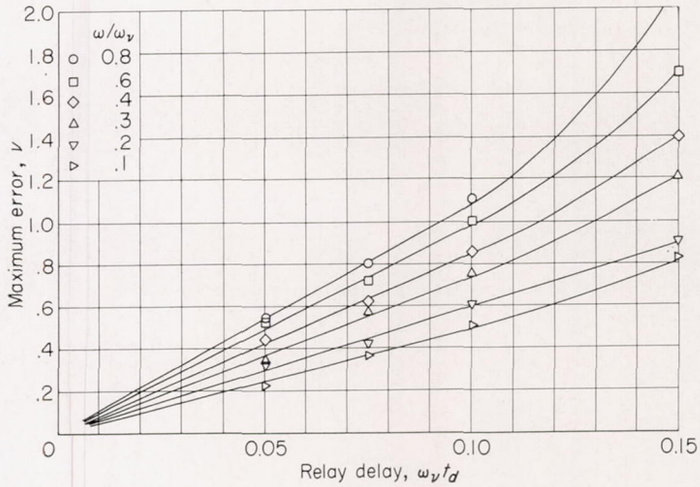


FIGURE 33.—Maximum error occurring during each input cycle plotted against switching delay with input frequency as a parameter (same data as for fig. 32).

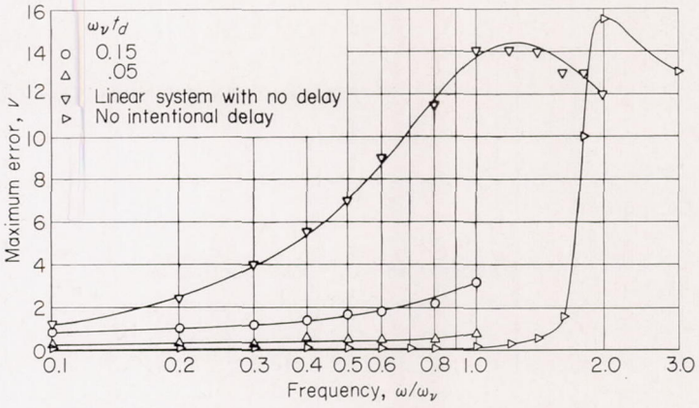


FIGURE 34.—Maximum error plotted against input frequency for several time delays and for linear member only.

APPENDIX B

COMPARISON OF EQUATION (8) AND NONLINEAR CONTROL EQUATION OF REFERENCES 1 AND 3 TO 5

COMPARISON OF NOTATIONS

In original notation the nonlinear control equation was given as (refs. 1 and 3 to 5)

$$a^+\frac{d^2y_o}{dt^2}+b\frac{dy_o}{dt}+cy_o=y_i(t) \qquad \text{(B1)}$$

where

$y_o(t)$	output
$y_i(t)$	input

$$b=b^+\left[1+\frac{\Delta b_1}{b^+}\operatorname{sgn}(y_o'E)+\frac{\Delta b_2}{b^+}\operatorname{sgn}(y_o'E')\right]$$

$$c=c^+\left[1+\frac{\Delta c_1}{c^+}\operatorname{sgn}(y_oE)+\frac{\Delta c_2}{c^+}\operatorname{sgn}(y_oE')\right]$$

$E=(y_o-y_i)$

$(\ )'=d(\ )/dt$

$\Delta b_1,\Delta b_2,\Delta c_1,\Delta c_2$	constants, $\Delta b_1>\Delta b_2,\Delta c_1>\Delta c_2$
a^+,b^+,c^+	constants

In terms of the notation of equation (8) the above equation is written

$$a^+\frac{d^2y}{dt^2}+b^+(1+\beta_m)\frac{dy}{dt}+c^+(1+\gamma_n)y=x(t) \qquad \text{(B2)}$$

where

$\beta_m=-{}_1\beta\,\mathrm{sgn}\,(y'e)-{}_2\beta\,\mathrm{sgn}\,(y'e')$; $m=0, 1, 2, 3$

$\gamma_n=-{}_1\gamma\,\mathrm{sgn}\,(ye)-{}_2\gamma\,\mathrm{sgn}\,(ye')$; $n=0, 1, 2, 3$

$e=(x-y)$

${}_1\beta,{}_2\beta,{}_1\gamma,{}_2\gamma$ constants, ${}_1\beta>{}_2\beta$, ${}_1\gamma>{}_2\gamma$

a^+,b^+,c^+ constants

The subscript convention is

$$\beta_3={}_1\beta+{}_2\beta \qquad \gamma_3={}_1\gamma+{}_2\gamma$$

$$\beta_2={}_1\beta-{}_2\beta \qquad \gamma_2={}_1\gamma-{}_2\gamma$$

$$\beta_1=-\beta_2 \qquad \gamma_1=-\gamma_2$$

$$\beta_0=-\beta_3 \qquad \gamma_0=-\gamma_3$$

Evidently, the correlation between equations (B1) and (B2) is then

$$y=y_o$$

$$x=y_i$$

$${}_1\beta=\frac{\Delta b_1}{b^+},\ {}_2\beta=\frac{\Delta b_2}{b^+},\ {}_1\gamma=\frac{\Delta c_1}{c^+},\ {}_2\gamma=\frac{\Delta c_2}{c^+}$$

$$e=-E$$

and a^+, b^+, and c^+ are the same.

In either notation a set of seven (constant) parameters is needed to characterize the system. For example, in equation (B2) the set $(a^+,b^+,c^+,{}_1\beta,{}_2\beta,{}_1\gamma,{}_2\gamma)$ is sufficient.

NORMALIZATION OF NONLINEAR CONTROL EQUATION

When considering control systems it is possible to reduce the number of parameters necessary to specify the nonlinear system. To do this a form of normalization familiar to linear theory is utilized. First, it should be realized that if error is to be defined as $e=(x-y)$ then $c^+=1$. Thus, assuming $c^+=1$ equation (B2) may be written

$$\frac{1}{\omega_\nu^2}\frac{d^2y}{dt^2}+\frac{2D}{\omega_\nu}(1+\beta_m)\frac{dy}{dt}+(1+\gamma_n)y=x(t) \qquad \text{(B3)}$$

where

$$\omega_\nu^2=\frac{1}{a^+}$$

$$\frac{2D}{\omega_\nu}=b^+$$

Introducing normalized time $\tau=\omega_\nu t$, equation (B3) becomes

$$\frac{d^2y}{d\tau^2}+2D(1+\beta_m)\frac{dy}{d\tau}+(1+\gamma_n)y=x(\tau) \qquad \text{(B4)}$$

Thus, knowing the natural frequency of the undamped linear system, ω_ν, the number of parameters necessary to specify performance in the nonlinear case is five, that is,

$$(D,\ {}_1\beta,\ {}_2\beta,\ {}_1\gamma,\ {}_2\gamma)$$

or alternatively

$$(D,\ \beta_3,\ \beta_2,\ \gamma_3,\ \gamma_2)$$

APPENDIX C

SIMULATION OF SECOND-ORDER NONLINEAR CONTROL SYSTEM

EQUIPMENT

Experimental studies were carried out with the following equipment:

(1) General equipment:
- (a) Analog computer, Beckman Ease
- (b) Low-frequency function generator, hp 202 A (input device)
- (c) Pen recorder, Sanborn Twin-Viso (output device)
- (d) Vacuum-tube voltmeter, RCA WV–97A
- (e) Oscilloscope, Dumont 304–A

(2) Special equipment:
- (a) Binary-logic switching circuit employed in conjunction with (a), described in detail in section "Binary-Logic Relay Switching Circuit."

COMPUTER SETUP

The computer diagram for the differential equation

$$\frac{d^2y}{d\tau^2}+2D(1+\beta_m)\frac{dy}{d\tau}+(1+\gamma_n)y=x(\tau) \qquad \text{(C1)}$$

is given in figure 35. The correspondence between this simulated system and the block diagram of figure 4 or 5 is straightforward (see ref. 7). Operational amplifiers 1 to 6 are used in simulating the linear member of the physical nonlinear system. Resistor R_1 provides adjustment of the linear damping factor D. The input to this simulated linear member is $x-(2D\beta_m y'+\gamma_n y)$ where x is obtained from the input device, while the values of feedback $2D\beta_m y'$ and $\gamma_n y$ are obtained with resistors R_2 through R_9 (see table 5) connected to y' and y through a binary-logic relay switching circuit derived in the following section. Amplifiers 7 to 10 are sign changers. The four variables y, y', e, and e' whose signs are to be sensed are made available as shown in the lower right of figure 35.

BINARY-LOGIC RELAY SWITCHING CIRCUIT

To complete the simulation of equation (8) it is necessary to implement the binary logic of table 2. Figure 5 shows a relay switching circuit (designed on an "and/or" basis) that realizes the necessary logic.

In order to preclude the possibility of time delay in switching, "fast" relays have been employed. The average properties of the 14 double-pole single-throw relays comprising the switching circuit are given in table 4.

The necessary synchronism in relays connected by dashed lines (fig. 5) is obtained by series connection of field coils.

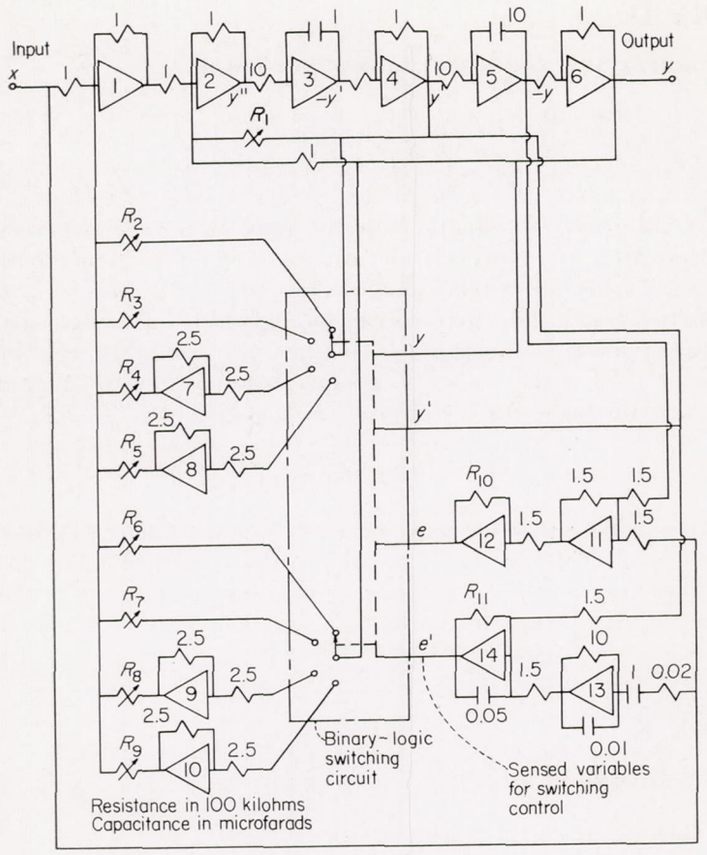


FIGURE 35.—Computer diagram for simulation of equation (8).

Into the four channels of field coils thus formed is read the four-digit binary logic of table 2 where now

$$0 = \text{Function} > 0 = \text{No coil current}$$

$$1 = \text{Function} < 0 = \text{Coil current}$$

This process of reading in the binary logic may be done by monitoring the variable (y, y', e, e') with four separate zero-coincidence detectors or amplitude selectors (denoted *CD* in fig. 5 and described in detail in the following section). The output of these coincidence detectors then drives the respective channels of relay coils.

SENSITIVE TRANSISTORIZED ZERO-COINCIDENCE DETECTOR

One of the four identical amplitude selectors designed for use with the switching circuit of figure 5 is shown in figure 36. This circuit basically consists of a grounded emitter PNP transistor stage T_1 driving a pentode pulse amplifier T_2. (It should be possible to utilize a transistor in place of the pentode if desired.) Four relay coils connected in series form the plate load of T_2. Positive feedback R_3 has been incorporated for regenerative switching.

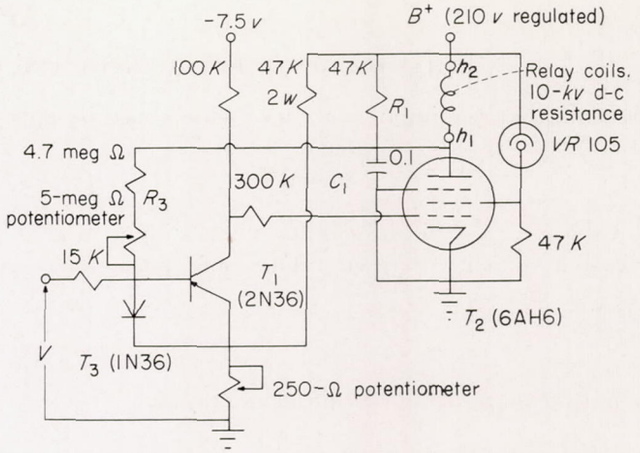


FIGURE 36.—Transistorized zero-coincidence detector.

The operation of this circuit is as follows: When the input voltage V becomes more negative than −50 millivolts, current starts to flow in the base-emitter region of T_1 This initiates collector-emitter current which effectively grounds the collector and thus the grid of T_2. Plate current flows in T_2 so that the relays which form this plate load are switched to the up position (fig. 5). Positive feedback R_3 has been incorporated to make the switching regenerative. Stages T_1 and T_2 will continue to conduct until V goes positive by 50 millivolts, at which time current is cut off in T_1 and thus T_2; the relays return to the normally closed positions (down in fig. 5).

Refinements that have been included in this circuit are:

(1) Adjustable positive feedback R_3. This feedback gives control over the zero-sensing threshold of the detector. The greater the positive feedback the greater the threshold. This adjustment is desirable for studying the effects of threshold in sensing discussed in the section "Effects of Switching Imperfections." The measured peak-to-peak threshold values obtainable with this arrangement were a maximum of 400 millivolts and a minimum of 60 millivolts. (In the computer setup it was necessary to amplify error e that formed the input to one of these detectors five times in order to bring the switching threshold down to 15 millivolts, peak to peak.)

(2) Adjustable bias for emitter of T_1. This is necessary to compensate for the slight positive bias (approximately equal to 100 millivolts) given to the base of T_1 by the positive feedback R_3.

(3) A clamping or clipping diode T_3 to protect the transistor T_1 from excessive base-emitter inverse voltages. Thus it can be seen that the input impedance of this coincidence detector is 15 kilohms since the base of T_1 is effectively always grounded.

APPENDIX D

BEHAVIOR OF OUTPUT IF INPUT AND INPUT DERIVATIVE ARE SIMULTANEOUSLY SMALL

Assume that the input is a curve which may be approximated by

$$x=\alpha\tau^2$$

for values $\tau<\delta$. Since $x'=2\alpha\tau$, both input and input derivative vanish with $\tau\to 0$. The input phase curve is

$$x=\frac{1}{4\alpha}(x')^2$$

The output depends on the initial conditions:

(1) For $y(0)=0$ and $y'(0)=0$

$$y=\alpha\left[\frac{\tau^4}{12}+\frac{D(1+\beta_m)}{30}\tau^5+\ldots\right]$$

(2) For $y(0)=0$ and $y'(0)=\epsilon_1$

$$y=\epsilon_1[\tau-D(1+\beta_m)\tau^2+\ldots]$$

(3) For $y(0)=\epsilon_2$ and $y'(0)=0$

$$y=\epsilon_2\left[1+\left(\frac{1+\gamma_n}{2}\right)\tau^2+\ldots\right]$$

In all cases reproduction of the input is not perfect very close to $\tau=0$; however, in cases (2) and (3) switch points may occur for rather small values of τ. In case (1) no switch point close to $\tau=0$ can be expected. This case will rarely occur; in most cases neither $y(0)$ nor $y'(0)$ will be zero. Then the output is a superposition of cases (2) and (3). In this event the error $e=x-y$ is given by

$$e=-\epsilon_2-\epsilon_1\tau+\left[\alpha+\epsilon_1 D(1+\beta_m)-\epsilon_2\left(\frac{1+\gamma_n}{2}\right)+\ldots\right]\tau^2$$

It is evident that for small values of ϵ_1 and ϵ_2 the error grows with α. This can easily be seen in the results of figure 11. The input was $x=A(1-\sin\Omega\tau)$; that means, near $x=0$, the input may be approximated by a parabola with $\alpha=1/2(\Omega^2)A$ or the errors near $x=0$ grow with Ω^2.

APPENDIX E

APPROXIMATING CURVES

The approximating curves which can be used to trace the output for a given input form a network in the phase plane. It has proved practical to present the network for e and e' with the same sign in one sheet (e. g., see fig. 20) and the network for e and e' with opposite signs on another sheet. Superposition of both sheets allows establishing of the four approximating curves through each point of the phase plane.

The approximating phase curves are determined by equation (21). Since the values λ_1 and λ_2 change from quadrant to quadrant the curves are composed of portions of different analytical curves which are patched at $y'=0$ and $y=0$.

The roots λ_1 and λ_2 depend on D, β_m, and γ_n; they may be complex or real. If the roots are complex, the approximating curves are curves of the spiral type which wind around the origin of the phase plane. In the case of real roots the curves have a quite different character. This can easily be understood by transforming the equation of the approximating curves (see refs. 1 and 3 to 5 and ch. V of ref. 8). For real roots the new coordinates μ and ν are introduced:

$$\bar{y}\lambda_1-\bar{y}'=\mu$$
$$\bar{y}\lambda_2-\bar{y}'=\nu$$

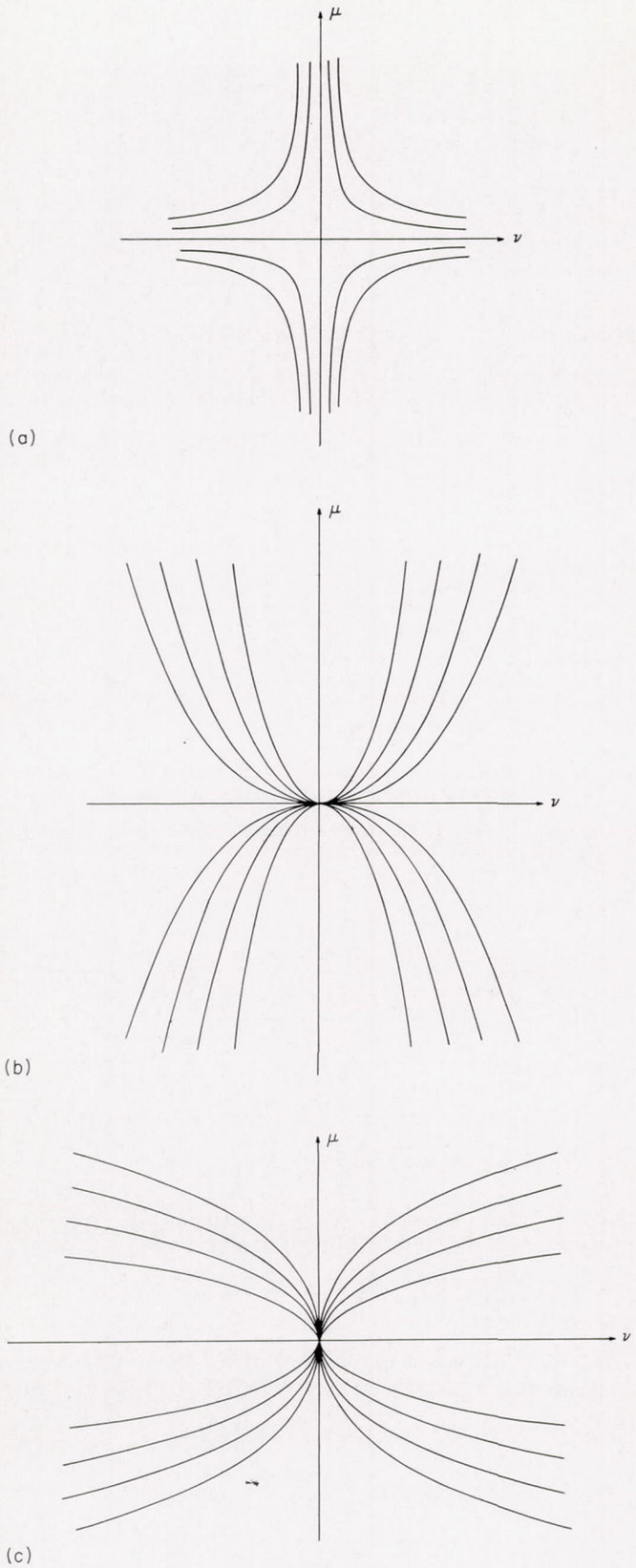
Then equation (21) yields

$$\nu^{\lambda_2}=M\mu^{\lambda_1}$$

If λ_1 and λ_2 are real and of opposite sign, the curves in the $\mu\nu$ plane have a hyperbolic character with $(\mu,\nu)=(0,0)$ as saddle point (see fig. 37 (a)). If λ_1 and λ_2 are of equal sign, the curves have a nodal point (see figs. 37 (b) and 37 (c)). In figures 38 and 39 phase curves with saddle point and stable and unstable node are shown in the original $\bar{y}\,\bar{y}'$ plane.

The approximating curves are composed of portions of these different types (see fig. 40).

There is no need of avoiding β_m,γ_n combinations which lead to node-type approximating curves because only portions of these curves are used. In the earlier publications it appeared as if (in either one of the networks of approximating curves) one set of curves had to be formed by curves of spiral character (complex roots λ). However, this has proved to be an unnecessary restriction. There might be some trouble with node-type curves if large delays in switching should occur (e. g., delays in $y'=0$ switching in fig. 40).



(a) $\frac{\lambda_2}{\lambda_1} < 0.$

(b) $\frac{\lambda_2}{\lambda_1} > 1.$

(c) $0 < \frac{\lambda_2}{\lambda_1} < 1.$

FIGURE 37.—Curves in $\mu\nu$ plane.

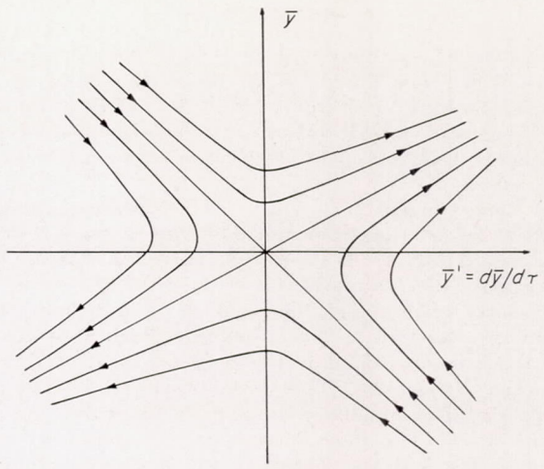
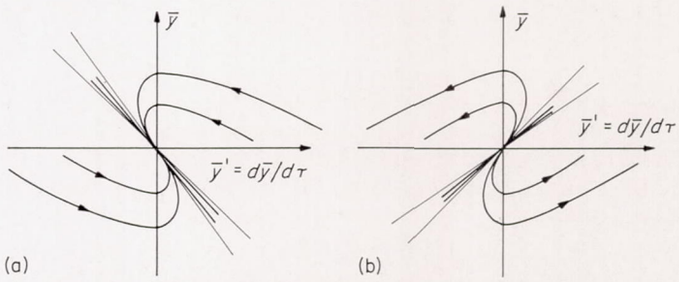


FIGURE 38.—Phase curves with saddle point.



(a) Stable node.
(b) Unstable node.

FIGURE 39.—Phase curves with nodal point.

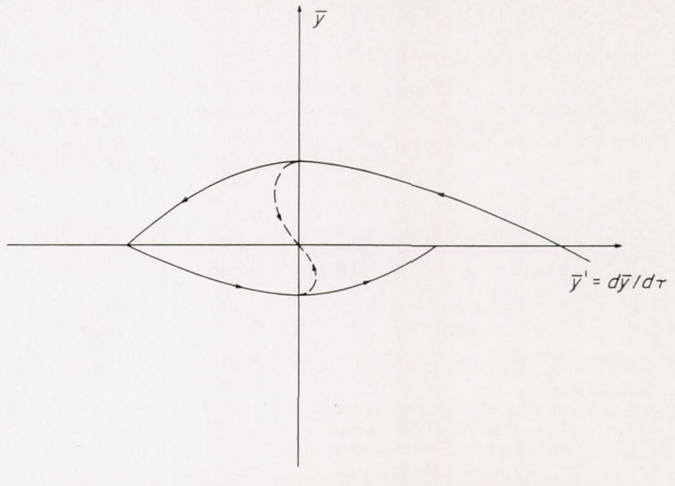


FIGURE 40.—Composition of an approximating curve.

REFERENCES

1. Flügge-Lotz, I., and Wunch, W. S.: On a Nonlinear Transfer System. Jour. Appl. Phys., vol. 26, no. 4, Apr. 1955, pp. 484–488.
2. Schmidt, Stanley F., and Triplett, William C.: Use of Nonlinearities to Compensate for the Effects of a Rate-Limited Servo on the Response of an Automatically Controlled Aircraft. NACA TN 3387, 1955.
3. Flügge-Lotz, Irmgard, and Wunch, W. S.: Mechanical Reproduction of an Arbitrary Function of Time. TR-No. 19 (Pt. I), Contract N6–onr–251, Task Order 2, Office of Nav. Res. and Div. Eng. Mech., Stanford Univ., Mar. 1, 1952.
4. Flügge-Lotz, Irmgard, and Wunch, W. S.: The Reproduction of an Arbitrary Function of Time (Reducing Transfer Function Dynamic Errors by Nonlinear Components). TR-No. 19 (Pt. II), Contract N6–onr–251, Task Order 2, Office of Nav. Res. and Div. Eng. Mech., Stanford Univ., Sept. 15, 1952.
5. Wunch, W. S.: The Reproduction of an Arbitrary Function of Time by Discontinuous Control. Ph. D. Thesis, Stanford Univ., May 1953.
6. Morrill, C. D.: A Sub-Audio Time Delay Circuit. Trans. Inst. Radio Eng., PGEC, vol. EC–3, no. 2, June 1954, pp. 45–49.
7. Korn, Granino A., and Korn, Theresa M.: Analog Computers. First ed., McGraw-Hill Book Co., Inc., 1952.
8. Andronow, A. A., and Chaikin, C. E.: Theory of Oscillations. Princeton Univ. Press, 1949.

TABLE 1.—CODING SCHEME FOR SWITCHING LOGIC FROM EQUATION (14)

[0, function$>$0; 1, function$<$0]

Binary coded decimal				Decimal
y	y'	e	e'	0
0	0	0	0	0
0	0	0	1	1
0	0	1	0	2
0	0	1	1	3
0	1	0	0	4
0	1	0	1	5
0	1	1	0	6
0	1	1	1	7
1	0	0	0	8
1	0	0	1	9
1	0	1	0	10
1	0	1	1	11
1	1	0	0	12
1	1	0	1	13
1	1	1	0	14
1	1	1	1	15

TABLE 2.—MATRIX OF ALLOWED β_m, γ_n COMBINATIONS AS DETERMINED BY ENCODED SWITCHING LOGIC OBTAINED FROM EQUATION (14)[a]

	β_0	β_1	β_2	β_3
γ_0	0000 1111			0100 1011
γ_1		0001 1110	0101 1010	
γ_2		0110 1001	0010 1101	
γ_3	0111 1000			0011 1100

[a] Examples:
(1) $\beta_3\gamma_0$ is chosen when 0100 occurs, i. e., ($y>0$, $y'<0$, $e>0$, $e'>0$), or when 1011 occurs, i. e., ($y<0$, $y'>0$, $e<0$, $e'<0$);
(2) $\beta_2\gamma_1$ is chosen when 0101 occurs, i. e., ($y>0$, $y'<0$, $e>0$, $e'<0$), or when 1010 occurs, i. e., ($y<0$, $y'>0$, $e<0$, $e'>0$);
(3) $\beta_1\gamma_3$ is not possible.

TABLE 3.—PARAMETERS FOR FIGURE 16

Figs. 16(a) and 16(b)	Figs. 16(c) and 16(d)
Complete system	
System 1: $\beta_3=-\beta_0=2$ $\gamma_3=-\gamma_0=2$ $\beta_2=-\beta_1=0.5$ $\gamma_2=-\gamma_1=0.5$	System 5: $\beta_3=-\beta_0=10$ $\gamma_3=-\gamma_0=10$ $\beta_2=-\beta_1=1$ $\gamma_2=-\gamma_1=1$
Case (1) No derivative feedback	
System 2: $\beta_3=-\beta_0=0$ $\gamma_3=-\gamma_0=2$ $\beta_2=-\beta_1=0$ $\gamma_2=-\gamma_1=0.5$	System 6: $\beta_3=-\beta_0=0$ $\gamma_3=-\gamma_0=10$ $\beta_2=-\beta_1=0$ $\gamma_2=-\gamma_1=1$
Case (2) No e' sensing	
System 3: $\beta_3=-\beta_0=2$ $\gamma_3=-\gamma_0=2$ $\beta_2=-\beta_1=2$ $\gamma_2=-\gamma_1=2$	System 7: $\beta_3=-\beta_0=10$ $\gamma_3=-\gamma_0=10$ $\beta_2=-\beta_1=10$ $\gamma_2=-\gamma_1=10$
Case (3) No e sensing in derivative feedback loop and no e' sensing in proportional feedback loop	
System 4: $\beta_3=-\beta_0=2$ $\gamma_3=-\gamma_0=2$ $\beta_2=-\beta_1=-2$ $\gamma_2=-\gamma_1=2$	System 8: $\beta_3=-\beta_0=10$ $\gamma_3=-\gamma_0=10$ $\beta_2=-\beta_1=-10$ $\gamma_2=-\gamma_1=10$

TABLE 4.—AVERAGE PROPERTIES OF RELAYS FOR SWITCHING CIRCUIT OF FIGURE 5

Description: General Electric CR–2791 double-pole single-throw relay mounted on 5-prong Amphenol base with permanent aluminum dust cover

Connections:

Electrical Properties	Values
Coil resistance	2.5 kilohms
Coil current for positive action	5 to 10 ma
Pull-in time	2 millisec
Drop-out time	3 millisec

TABLE 5.—CORRESPONDENCE BETWEEN ADJUSTABLE COMPUTER ELEMENTS AND PARAMETERS OF EQUATION (8)

Element	Corresponds to	Units, kilohms
R_1	$(2D)^{-1}$	100
R_2	$(2D\beta_0)^{-1}$	100
R_3	$(2D\beta_1)^{-1}$	100
R_4	$(2D\beta_2)^{-1}$	100
R_5	$(2D\beta_3)^{-1}$	100
R_6	$(\gamma_0)^{-1}$	100
R_7	$(\gamma_1)^{-1}$	100
R_8	$(\gamma_2)^{-1}$	100
R_9	$(\gamma_3)^{-1}$	100

U. S. GOVERNMENT PRINTING OFFICE: 1959