1957 IRE WESCON Convention Record

PART 4

Sessions Sponsored by

IRE Professional Groups on

Automatic Control

Electronic Computers

at

the Western Electronic Show and Convention

San Francisco, Calif.

August 20-23, 1957

The Institute of Radio Engineers

1957 IRE WESCON CONVENTION RECORD

PART 4 - COMPUTERS AND AUTOMATIC CONTROL

TABLE OF CONTENTS

Page

Session 21: Statistical Methods in Feedback Control
(Sponsored by the Professional Group on Automatic Control)

Session 24: Data Handling Devices
(Sponsored by the Professional Group on Electronic Computers)

Session 28: Computer Circuit and Logical Design
(Sponsored by the Professional Group on Electronic Computers)

Session 40: Analog and Digital Computer Devices
(Sponsored by the Professional Group on Electronic Computers)

ON THE DESIGN AND COMPARISON OF CONTACTOR CONTROL SYSTEMS

By I. Flügge-Lotz and H. E. Lindberg
Stanford University, Stanford, California

Summary

This paper points out some of the similarities of systems synthesized by phase plane and frequency response techniques. A method is given for comparing contactor control systems with each other, and a logical argument is presented for designing contactor systems on the basis of their step response. The last section gives a new viewpoint and a simple method for finding the error of a contactor system operating in the presence of high frequency relay "chatter" oscillations. The range of inputs for which chatter operation will occur is defined. All of the theoretical investigations are supported by analog computer simulations.

Introduction

A contactor or relay control system is, as the name implies, any control system which employs some sort of switching device as an essential component. Its major advantage is its ability to control large amounts of power by rather simple means. It is nonlinear and the accompanying diffifulty encountered in analyzing it is a very real disadvantage. Also, in some applications it has the further disadvantage of being inferior to a linear system designed to do the same job. There is a large class of applications, however, where a contactor system can be shown to be superior to a linear system of the same power handling capacity. These are applications where the sudden accelerations and high frequency oscillations caused by the presence of a switching device are not objectionable. If such oscillations are objectionable, they can be eliminated by the use of a dual mode scheme[1], but many times such an oscillation furnishes "dynamic lubrication" which diminishes the effects of static friction, backlash, and other parasitic nonlinearities.

In recent years, roughly since the early 1940's, more and more effort has been made to break down the analytical "barrier" which has limited the use of contactor systems. Many methods have been used to synthesize contactor systems and various types of contactor systems have been synthesized. In what follows, some of the similarities of systems designed by very different techniques will be brought out. Also, a method for comparing contactor systems with each other will be given and a logical argument for designing contactor systems on the basis of their step response will be presented. The last section gives a new viewpoint and a simple method for finding the error of a contactor system operating in the presence of the high frequency "chatter" oscillations mentioned earlier.

Similarities of Contactor Control Systems

A Phase Plane Approach

One method of studying a contactor control system is to look directly at its transient response in the phase space. Consider, for example, the second order system studied by Flügge-Lotz[2]. This is a zero-seeking device whose differential equation in a de-dimensionalized form is

$$\psi'' + 2D\psi' + \psi = -\operatorname{sgn}(\psi + k\psi') = \phi \qquad (1)$$

where ψ is the controlled variable, ϕ is the switching function, D is a fixed parameter of the controlled process, and k is a parameter of the switching function to be determined to give good response. Notice that in this system, the switching function is simply the sign of $\psi + k\psi'$; that is, it takes on the values + 1 or − 1 depending on whether $\psi + k\psi'$ is plus or minus respectively. There are systems with more complicated switching functions that give "optimum" response[3] but for the moment we will study only linear switching functions because they are more easily realized in an actual application.

Being a second order system, the phase space becomes a phase plane as shown in Figure (1). Notice that the ψ axis is inclined to the ψ' axis by the angle $\sigma = \arccos(-D)$. If this is done, the phase plane trajectories become logarithmic spirals of the form

$$r = r_o e^{-\frac{D}{v} v\tau} \qquad (2)$$

where r is measured from the point (0, ± 1) depending on the sign of ϕ; $v\tau$ is the angle between two radii extending from (0, ± 1); and $v = \sqrt{1 - D^2}$. The curve shown is a typical response to an initial error and error rate represented by point P_1.

Notice in Figure (2), which is essentially an enlarged view of part of Figure (1), that for the ideal system of this type (k > 0) no motion is defined beyond point A. That is, only two trajectories pass through this point, namely: AB for $\phi = +1$ and CD for $\phi = -1$; and the only paths leading out of point A are those towards B and D. Motion cannot proceed along the path towards B because above the switching line ($\psi + k\psi' = 0$), Eq. (1) dictates that the control function ϕ is − 1; also, motion cannot proceed along the path towards D because below the switching line, $\phi = +1$. A point such as A is called an "end point." An end point occurs when a trajectory such as PAB intersects the switching line twice in succession on the same side of the origin. Every path such as shown in Figure (1) has an end point.

If the device which forms the switching function ϕ is allowed to have an imperfection such as a time delay, then the motion does not stay at point A but proceeds to point F. Here ϕ can change to -1 etc., and the motion proceeds toward the origin as shown by the solid path. The average path is along the switching line and the differential equation of this average motion is

$$\psi + k\psi' = 0 \qquad (3)$$

whence

$$\psi = \psi_A e^{-\frac{1}{k}\tau} \qquad (4)$$

Eq. (4) shows that in the "end motion," ψ tends toward zero exponentially. A consideration of this motion alone would indicate that k should be chosen small and positive. However, if motion is to proceed from a point such as P_1 in Figure (1) to an end point quickly, then it is advantageous to increase k. Depending on the position of P_1, a compromise selection of k can be made.

For study of response to step inputs, much of the theory developed for Eq. (1) can be extended to a follow-up system as shown in Figure (3). The equation of this system is

$$y'' + 2Dy' + y = \operatorname{sgn}(e + ke') \qquad (5)$$

which can be written by substituting $y = x - e$ as

$$e'' + 2De' + e = x - \operatorname{sgn}(e + ke') \qquad (6)$$

for a step input where $x'' = x' = 0$. The presence of x affects the position of the foci of the phase plane spirals indicated in Figure (1).

The Method of Kochenburger

Another method of approaching non-optimum relay systems is presented by Kochenburger[4]. This technique is entirely different from the phase plane method. Whereas the study of the phase plane gives a pictorial view of the exact transient response, the method presented by Kochenburger provides an extension of the frequency-response-stability criterion of Nyquist. The basic assumption is that if a sinusoidal signal is impressed on the relay coil, the periodic square wave output of the relay can be replaced by its first harmonic. This assumption becomes more and more acceptable as the complexity (degree of the differential equation) of the system being controlled increases [G_S in Figure (4)] because of the filtering action of these components.

With this assumption, the action of the relay can be expressed in terms of a "describing function" which gives the amplitude and phase shift of this harmonic in terms of the amplitude a of the impressed sinusoidal input and the relay characteristics. The block diagram is shown in Figure (4), where the describing function is G_D, the transfer function of the controlled process is G_S, and G_C is a compensation network to be determined to give good response and stability. In order to determine the compensation network, Kochenburger uses the standard frequency response techniques that are used for linear systems with the exception that the system must be examined for each value of G_D since it is amplitude dependent. Absolute stability means, for example, that the polar plot of $1/G_C(j\omega)G_S(j\omega)$ must completely enclose the locus of $-G_D(a)$ rather than merely the point -1. See Figure (5).

If $G_S = 1/(p^2 + 2Dp)$, Kochenburger found that the compensation network should be of the form $G_C = (1 + \tau p)/(1 + \frac{\tau}{\alpha} p)$ where α is made as large as possible without allowing too much noise to be transmitted. As an interesting comparison between this result and the conclusions of Flugge-Lotz, consider the output c of the compensating network

$$c = e\left[\frac{1 + \tau p}{1 + \frac{\tau}{\alpha} p}\right] = e\left[1 + \frac{(1 - \frac{1}{\alpha})\tau p}{1 + \frac{\tau}{\alpha} p}\right] \qquad (7)$$

In the time domain, this can be written as

$$c = e + \tau^* e'_f \qquad (8)$$

where e'_f is the time derivative of e filtered by the network whose transfer function is $1/(1 + \frac{\tau}{\alpha} p)$. If we reduce the system equations to a non-dimensional form, the output of an ideal relay is simply

$$R = \operatorname{sgn} c = \operatorname{sgn}(e + \tau^* e'_f) \qquad (9)$$

and the differential equation of the complete system is

$$y'' + 2Dy' = \operatorname{sgn}(e + \tau^* e'_f) \qquad (10)$$

A direct comparison of this equation with Eq. (5) studied by Flügge-Lotz shows that what Kochenburger refers to as a "series compensation network" is referred to as a "control function" by Flugge-Lotz. The only difference between them is that e' appears in one and e'_f appears in the other. But this is only an academic difference because in any physical application of Flugge-Lotz' work, the error derivative e' must be filtered by some means. The left hand sides of the equations differ because Flügge-Lotz studied the more general case where the output itself appears in the equation.

If one examines other examples given by Kochenburger, an extension of this comparison can be made. For $G_S = 1/p(\tau p + 1)^2$ he finds that the lead network required for compensation must be quadratic. The network he used had a transfer function of

$$G_C = \frac{1 + ap + bp^2}{1 + dp + fp^2} \qquad (11)$$

For the required phase lead, he makes d and f as small as possible, again without transmitting too much noise. In the ideal case where d and f are both zero, the compensator-relay combination gives an output of

$$R = \operatorname{sgn}(e + ae' + be'') \qquad (12)$$

which is the same form of control function which Flügge-Lotz found necessary for good performance with a differential equation of third order.

A Varied Coefficient Scheme

What at first appears to be a completely different scheme for synthesizing a contactor system is presented by Flügge-Lotz and Taylor[5]. The block diagram for this system is shown in Figure (6). The differential equation of this system in a de-dimensionalized form is

$$y'' + 2D(1 + \beta_m)y' + (1 + \gamma_n)y = x \qquad (13)$$

where

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

$$\gamma_n = - {}_1\gamma \operatorname{sgn}(ye) - {}_2\gamma \operatorname{sgn}(ye') \qquad (14)$$

with ${}_1\beta$, ${}_2\beta$, ${}_1\gamma$, ${}_2\gamma$ being constant. Notice that the coefficients of the differential equation are varied rather than a lumped saturation quantity as was the case in the two previous examples.

The switching functions for β_m and γ_n are each broken up into two parts, one which changes with e and one which changes with e'. It is shown in Reference (6) that it is more advantageous to switch single quantities β and γ according to the following equations

$$\beta_m = - \beta \operatorname{sgn}\left[y'(e + ke')\right]$$

$$\gamma_n = - \gamma \operatorname{sgn}\left[y(e + ke')\right] \qquad (15)$$

where β and γ are positive constants. If Eqs. (15) are substituted into (13), the complete equation becomes

$$y'' + 2Dy' + y$$

$$= x + \left(2D\beta\ |y'| + \gamma\ |y|\right) \operatorname{sgn}(e + ke') \qquad (16)$$

Notice that this differential equation is somewhat similar to Eq. (5) studied by Flügge-Lotz and Eq. (10) studied by Kochenburger. The main difference is that in Eq. (16) the coefficient of the "switching function" is a variable whereas in Eqs. (5) and (10) it is constant. The input x is also present in Eq. (16) but it will be shown later that this has the same sort of effect as the varying switching function coefficient.

Comparison of Contactor Control Systems

Although these systems are similar, there are differences that warrant investigation. Can it be said, for example, that Eq. (16) will give better response to a random input than Eq. (5)? To begin to answer this, let us first write Eq. (16) in terms of the error e by substituting $y = x - e$

$$e'' + 2De' + e$$

$$= x'' + 2Dx' - \left(2D\beta\ |y'| + \gamma\ |y|\right) \operatorname{sgn}(e + ke') \qquad (17)$$

Consider now that this system is operating in a region where input velocities and accelerations are small, or $|x'' + 2Dx'| < 2D\beta\ |y'| + \gamma\ |y|$. In fact, if this inequality is not true, the output will soon begin to diverge from the input x. If the inequality does hold, then the terms governed by the control function $\operatorname{sgn}(e + ke')$ predominate and determine the sign of the entire right hand side of the differential equation. Under these conditions the error is soon driven to zero and it has been shown analytically by Flügge-Lotz[2] and experimentally by Lindberg that the error "chatters" about zero with a small amplitude and very high frequency relative to x or y due to relay imperfections. Because the frequency is high and the amplitude very small, quantities such as x', x'', and y may be regarded as constants during a few error cycles, $2D\beta\ |y'|$ can be considered small relative to $\gamma\ |y|$, and the equation of motion becomes

$$e'' + 2De' + e = - a_1 \quad \text{if} \quad e + ke' > 0$$

$$e'' + 2De' + e = + b_1 \quad \text{if} \quad e + ke' < 0 \qquad (18)$$

where a_1 and b_1 are positive constants.

Similarly, we can transform Eq. (5) into

$$e'' + 2De' + e = x'' + 2Dx' + x - \text{sgn}(e + ke') \quad (19)$$

Again, if the error is not to diverge,

$+ 2Dx' + x| < 1$ and we can write

$$e'' + 2De' + e = -a_2 \quad \text{if} \quad e + ke' > 0$$

$$e'' + 2De' + e = +b_2 \quad \text{if} \quad e + ke' < 0 \quad (20)$$

Notice that these equations are identical in form to Eqs. (18). The only difference between these equations is the values of a_1, a_2, b_1, and b_2 and perhaps the value of k. In fact, these equations would result even if the switching function was something more complicated such as $\text{sgn}(e + k_1 e' |e'|)$ because we are studying chatter response which is close enough to the origin of the e, e' phase plane that any complicated switching function can be approximated by a straight line, and hence the switching function reduces to $\text{sgn}(e + ke')$.

The parameters of these equations could conceivably be very different between the two systmes and within each system itself at different times depending on the instantaneous values of x, y, and their derivatives. It is shown in Reference (6) that the amplitudes of the chatter errors for Eqs. (18) and (20) are approximated by

$$e_{max_\varepsilon} = \frac{1}{2a} \left(\frac{\varepsilon}{k}\right)^2 \quad (21)$$

for a relay threshold imperfection of ε, and

$$e_{max_T} = \frac{aT^2}{2} \quad (22)$$

for a relay time delay imperfection of T, if each occurs separately. In each of these expressions, $\underline{a}$ represents the driving force amplitude appearing on the right-hand sides of Eqs. (18) or (20), such as a_1, for example.

In an actual system, both threshold and time delay imperfections occur together and since $\underline{a}$ appears in the denominator of Eq. (21) and in the numerator of Eq. (22), the error amplitude tends to depend only slightly on $\underline{a}$. The error depends a little more on k but even this dependence is secondary to the dependence on relay characteristics because k appears in only one of the two expressions for error. We must then conclude that during chatter response, that is, when the input is varying slowly as indicated by the inequalities that were necessary for the formulation of Eqs. (18) and (20), the system response depends much more strongly on the relay characterisitics than on the design of the control system equations.

A Design Criterion

How, then, should one compare two different system equations? First, recall that the errors during chatter operation are very small. As was mentioned earlier, there is a large class of applications where this type of response is very satisfactory, and we might call the range of inputs for which chatter occurs the "region of satisfactory response." Hence, one measure of merit for a systme is that it have the largest region of satisfactory response consistent with its saturation values. From this standpoint we can say that Eq. (5) is a better system than Eq. (16). To see this, first recall that the region of satisfactory response for Eq. (5) is defined by

$$|x'' + 2Dx' + x| < 1 \quad (23)$$

where the saturation value of this system is unity for the units chosen.

To find the region of satisfactory response including saturation effects for Eq. (16) we must first gather together all of the variables that are produced by the component whose saturation we are considering. For Eq. (16) this refers to the quantities that exist at point m in Figure (6) where a motor or amplifier, for example, would saturate. Eq. (17) is now written

$$e'' + 2De' + e = x'' + 2Dx' + x$$

$$- \left[x + (2D\beta|y'| + \gamma \; |y|) \, \text{sgn}(e + ke')\right] \quad (24)$$

where the terms in the brackets are the same terms as on the RHS of Eq.(16) and are the quantities at point m. The inequality to insure chatter operation is now defined by

$$|x'' + 2Dx' + x|$$

$$< |x + (2D\beta|y'| + \gamma|y|) \, \text{sgn}(e + ke')| \quad (25)$$

This must necessarily give the same or a smaller region of satisfactory response than that defined by Eq. (23) because the RHS of Eq. (25) is a complicated expression which must be equal or less than unity if the systems of Eq. (5) and Eq. (16) are to have the same saturation values.

Nothing has been said yet about what control function would give better response to a random input. During chatter operation the choice of different control functions is equivalent to changing k. In our discussion of Eqs. (21) and (22), however, it was seen that changing k had a secondary effect on an already small error. Thus if we are to compare or design a control function, we must examine responses to inputs which fall in part, al least, outside the "region of satisfactory response." Also, the test input must fall largely in the region of satisfactory response in order to insure that the test will not degenerate into a simple examination of saturation operation. The most easily produced inputs that satisfy these requirements are discontinuities, the simple step function being the most commonly used.

Therefore, after a system is designed to take the full benefit of its saturation values, a rational approach to designing it to follow random inputs is to design it to follow a step input in spite of the fact that superposition does not hold!

Chatter Response of a Higher Order System

It has been shown that for a second order system, the error during chatter response is very small and depends mainly on the relay characteristics. There may be another error that occurs simultaneously, but it does not become noticeably large until one considers higher order systems. This is an error which arises due to the delays encountered in forming the quantities in the switching function. To demonstrate a method of finding these errors, let us consider a third order system with a linear switching function. In a de-dimensionalized form the differential equation is

$$y''' + 2\zeta\Omega y'' + \Omega^2 y' = -N\,\mathrm{sgn}(e + k_1 e' + k_2 e'') = -N\phi \qquad (26)$$

In forming the derivatives e' and e'', filtering delays are invariably incurred in a physical system. Recall also that $e' = x' - y'$. If in obtaining this difference more delay is encountered in obtaining x' than y' or vice versa, then a noticeable error will incur. Suppose, for example, that y' is available with no appreciable delay but that filtering is necessary on x'. If we call $\tilde{x}'$ the filtered x', then the transfer function for x' is of the form

$$\frac{\tilde{x}'}{x} = \frac{p}{(1 + T_{11}p)(1 + T_{12}p)\cdots(1 + T_{1n}p)} \qquad (27)$$

If we consider sinusoidal inputs $x = A \sin \omega\tau$ and assume that the time constants T_{1i} are small, the difference between $\tilde{x}'$ and x' is found to be

$$\tilde{x}' - x' = A\omega^2 \sum_n [T_{1i}] \cos(\omega\tau - \frac{\pi}{2} - \frac{\omega}{2}\sum_n T_{1i}) \qquad (28)$$

Similarly, if we define $\tilde{x}''$ by

$$\frac{\tilde{x}''}{x} = \frac{p^2}{(1 + T_{21}p)(1 + T_{22}p)\cdots(1 + T_{2m}p)} \qquad (29)$$

for small T_{2i} we can write

$$\tilde{x}'' - x'' = -A\omega^3 \sum_m [T_{2i}] \sin(\omega\tau - \frac{\pi}{2} - \frac{\omega}{2}\sum_m T_{2i}) \qquad (30)$$

Defining $\tilde{e}'$ and $\tilde{e}''$ by

$$\tilde{e}' = \tilde{x}' - y'$$

and

$$\tilde{e}'' = \tilde{x}'' - y'' \qquad (31)$$

Eq. (26) for a system with delays becomes

$$y''' + 2\zeta\Omega y'' + \Omega^2 y' = -N\,\mathrm{sgn}(e + k_1\tilde{e}' + k_2\tilde{e}'') \qquad (32)$$

Recall that we are considering operation of the system for sinusoidal inputs with low enough frequency that chatter operation is insured. Under this condition the argument of the switching function is driven to zero and we can investigate the average motion in a way similar to that of Eq. (3) by considering

$$e + k_1\tilde{e}' + k_2\tilde{e}'' = 0 \qquad (33)$$

In terms of the undelayed error derivatives, Eq. (33) becomes a differential equation for the actual error e in which we are interested

$$e + k_1 e' + k_2 e'' = -k_1(\tilde{x}' - x') - k_2(\tilde{x}'' - x'') \qquad (34)$$

This is simply the differential equation of a linear forced vibration problem, where the driving term is a function of the input x and the time delays T_{1i} and T_{2i}.

Verification of the Theory for Sinusoidal Inputs

Experimental verification of this theory was made using the analog computer circuit of Figure (7). The sinusoidal input was generated by solving the equation

$$x'' + \omega^2 x = 0$$

on amplifiers 7, 8, and 9. This was done rather than using an external function generator in order to reduce the noise on x and its derivatives so that the noise problems could be studied separately. Generating these sine waves in the computer gave the additional advantage that x' and x'' were available as outputs of integrators. This allowed the study of the effect of x' and x'' filtering delays separately because these integrator outputs could be used directly with no filtering.

To study the effect of x' filtering delay alone, $\tilde{x}'$ was derived by actually differentiating x in amplifier 17 and filtering it as shown, while x'' was taken directly from amplifier 7. Solving Eq. (34) with $\tilde{x}'' - x'' = 0$ and taking $x' - x'$ from Eq. (28) the error (neglecting chatter) is

$$e = \frac{-Ak_1\omega^2 \sum_i T_{1i}}{\sqrt{(1 - k_2\omega^2)^2 + k_1^2\omega^2}} \times$$

$$\times \cos\left(\omega\tau - \frac{\pi}{2} - \frac{\omega}{2}\sum_i T_{1i} - \theta\right) \qquad (35)$$

where

$$\theta = \tan^{-1}\left[\frac{k_1\omega}{1 - k_2\omega^2}\right]$$

Experiments were run using various values for A, k_1, k_2, ω, and T_{1i} and the error amplitude was measured and compared with the values given by Eq. (35). A typical computer response is shown in Figure (8a) and the results are tabulated in Table (1). The agreement between theory and experimental results is quite good.

Further experiments were run with x'' filtering delays alone and then with x' and x'' delays occurring simultaneously and the results again compared quite well with the theory. The theoretical expression for the error with x'' delays alone is

$$e = \frac{-Ak_2\omega^3 \sum_i T_{2i}}{\sqrt{(1 - k_2\omega^2)^2 + k_1^2\omega^2}} \times$$

$$\times \sin\left(\omega\tau - \frac{\pi}{2} - \frac{\omega}{2}\sum_i T_{2i} - \theta\right) \qquad (36)$$

and for x' and x'' delays together is

$$e = \frac{A\omega^2\sqrt{k_1^2\left(\sum_i T_{1i}\right)^2 + k_2^2\omega^2\left(\sum_i T_{2i}\right)^2}}{\sqrt{(1 - k_2\omega^2)^2 + k_1^2\omega^2}} \times$$

$$\times \cos(\omega\tau - \mu) \qquad (37)$$

In deriving Eq. (37), the RHS of Eq. (34) became

$$-k_1A\omega^2\sum_i\left[T_{1i}\right]\cos\left(\omega\tau - \frac{\pi}{2} - \frac{\omega}{2}\sum_i T_{1i}\right)$$

$$-k_2A\omega^3\sum_i\left[T_{2i}\right]\sin\left(\omega\tau - \frac{\pi}{2} - \frac{\omega}{2}\sum_i T_{2i}\right) \qquad (38)$$

Since both $\sum_i T_{1i}\omega$ and $\sum_i T_{2i}\omega$ are small, they were taken equal for phase angle purposes so that the two terms in Eq. (38) could be taken as orthogonal to give the simple expression of Eq. (37).

Response to Other Inputs

We have demonstrated that for sinusoidal inputs of low enough frequency that chatter occurs, the resulting error, exclusive of the chatter error, may be calculated by linear theory using Eq. (34). It is now postulated that Eq. (34) can be used to find the error for any input that varies slowly enough to insure chatter operation. As an experimental verification of this proposition, tests were run with an input that consisted of two sinusoids at different frequencies. It was observed that the error was merely the sum of the errors that would result if each sinusoid were applied separately. That is, the law of linear superposition holds and hence the simple linear theory can be used for any input where chatter occurs.

The input that was used for this test was $x = 15\cos(0.446\tau) + 1.00\cos(2.83\tau)$. The response for this input and for each of the sinusoids separately is shown in Figure (8). For simplicity only an x' filtering delay was used. Experimental error amplitudes were 0.14 volts and 0.27 volts respectively and a computation using Eq. (35) gave 0.15 volts and 0.27 volts for these amplitudes. The error for the combined input is merely the sum of these errors as can be seen by inspection of Figure (8c).

The errors in all of these experiments are quite large because the filtering lags were made intentionally large.

Reduction of Errors Due to Filtering Lags

It was mentioned at the beginning of this chapter that the reason for the errors discussed here is that the lags encountered in finding x' and x'' are not the same as those encountered in finding y' and y'' respectively. The validity of this statement can be seen if we look at the lags in x' and y' for example. The delayed e' used in the switching function is

$$\tilde{e}' = x'(t - T_{x'}) - y'(t - T_{y'}) \qquad (39)$$

This can also be written

$$\tilde{e}' = x'(t - T_{y'}) - x'(t - T_{y'})$$

$$+ x'(t - T_{x'}) - y'(T - T_{y'}) \qquad (40)$$

or

$$\tilde{e}' = e'(t - T_{y'}) - x'(t - T_{y'}) + x'(t - T_{x'}) \qquad (41)$$

Substituting Eq. (41) into Eq. (33) ssuming for the moment that $\tilde{e}'' = e''$ we obtain

$$e(t) + k_1 e'(t - T_{y'}) + k_2 e''(t)$$

$$= k_1 x'(t - T_{y'}) - k_1 x'(t - T_{x'}) \qquad (42)$$

or

$$e(t) + k_1 e'(t) + k_2 e''(t)$$

$$= k_1 \left[x'(t - T_{y'}) - x'(t - T_{x'})\right]$$

$$+ k_1 \left[e'(t) - e'(t - T_{y'})\right] \qquad (43)$$

But even in the case when we do get undesirable errors, they are always quite small relative to the amplitude of the input and we can neglect the second term on the RHS of Eq. (43) as compared to the first. If this is done, it can be seen by inspection that the magnitude of the driving term in Eq. (43) depends very strongly on the difference between $T_{y'}$ and $T_{x'}$ and disappears if $T_{y'} = T_{x'}$. In this case we must reconsider the second term on the RHS, but this would still give a very small error as compared to the errors when $T_{y'} \neq T_{x'}$.

In any relay system such as discussed here, therefore, in order to reduce the errors incurred during chatter operation, the designer should make an effort to provide circuitry to make the filtering delays for x' and x'' equal to the delays for y' and y'' respectively.

Input Limits to Give Chatter Response

It is rather curious to notice that so far in this discussion of response to "slowly varying" inputs, no mention has been made about the coefficients of the differential equation of the controlled process. The entire response during chatter operation has been determined by the coefficients in the switching function argument and the delays required to measure input and output derivatives. In considering the limits for the inputs that will give chatter response, however, the reverse situation exists. These limits depend only on the coefficients of the differential equation of the controlled process and the magnitude of the driving term N, and are found, as for the second order system, by simple saturation considerations. For Eq. (26), the expression for x which defines the limits of chatter operation is

$$N > \left[x''' + 2\zeta\Omega x'' + \Omega^2 x'\right]_{max} \qquad (44)$$

For $x = A \sin \omega\tau$ this expression becomes

$$N > A\sqrt{(\omega^3 - \omega\Omega^2)^2 + 4\zeta^2\Omega^2\omega^4} \qquad (45)$$

This expression, in the limit when it becomes an equality, defines the breakdown frequency ω_{cr} where chatter stops and the error becomes very large. The values found from Eq. (45) were only slightly less than those found in analog computer experiments.

Acknowledgements

This work has been supported by the National Advisory Committee for Aeronautics (NACA). The authors also wish to thank Dr. A. M. Peterson of the Stanford Electrical Engineering Department for his interest and most helpful advice on the electronic problems which were encountered during this investigation.

References

1. McDonald, D. C., Multiple Mode Operation of Servomechanisms, Review of Scientific Instruments, Vol. 23, No. 1, pp. 22-30, Jan. 1952.

2. Flügge-Lotz, I., Discontinuous Automatic Control, Princeton University Press, Princeton, N. J., 1953.

3. Bushaw, D. W., Differential Equations with a Discontinuous Forcing Term, Experimental Towing Tank, Stevens Institute of Technology Report 469, Hoboken, N. J., 1953.

4. Kochenburger, R. J., A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, AIEE Trans., Vol. 69, pp. 270-84, 1950.

5. Flügge-Lotz, I. and Taylor, C. F., Synthesis of a Nonlinear Control System, IRE Trans., PGAC-1, pp. 3-9, 1956, and Investigation of a Nonlinear Control System, NACA TN 3826, 1957.

6. Flügge-Lotz, I. and Lindberg, H. E., Nonlinear Control Systems of Second Order, Div. of Eng. Mech., Stanford Univ., TR No. 104, Dec. 1956.

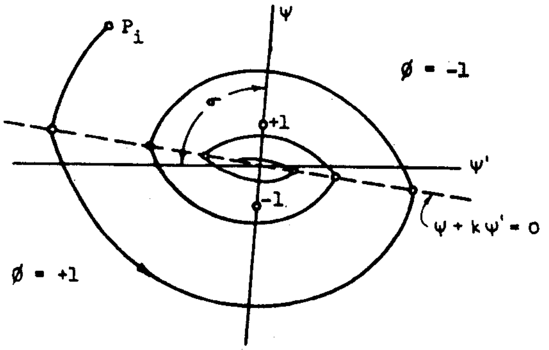

Fig. 1 Transient response in the phase plane.

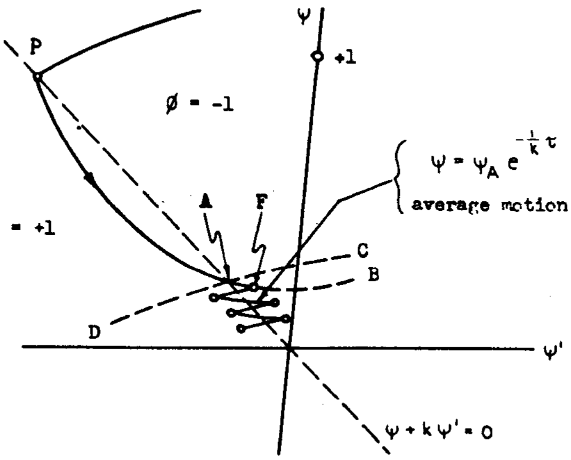

Fig. 2 End motion for k>O.

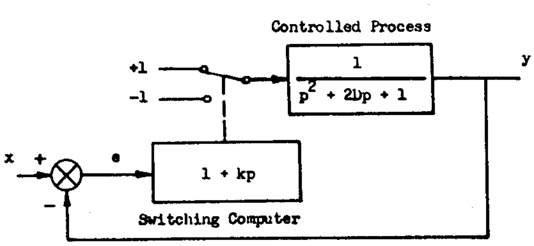

Fig. 3 A second order follow-up system.

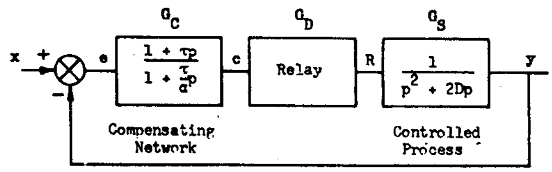

Fig. 4 Block diagram showing the series transfer functions used by Kochenburger.

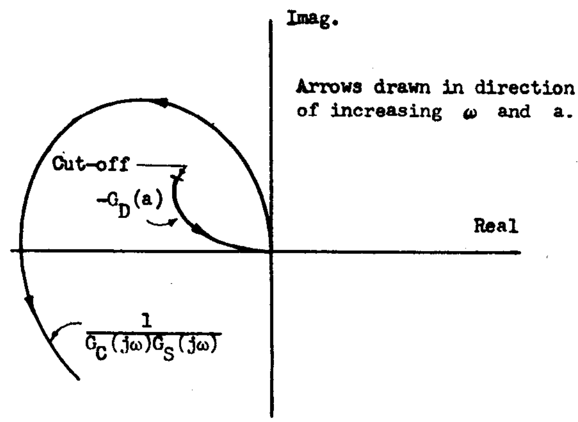

Fig. 5 Frequency response in the complex plane.

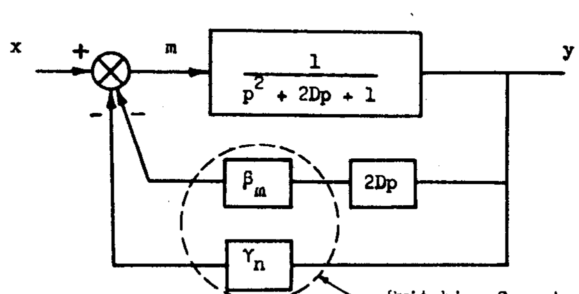

Fig. 6 A varied coefficient system.

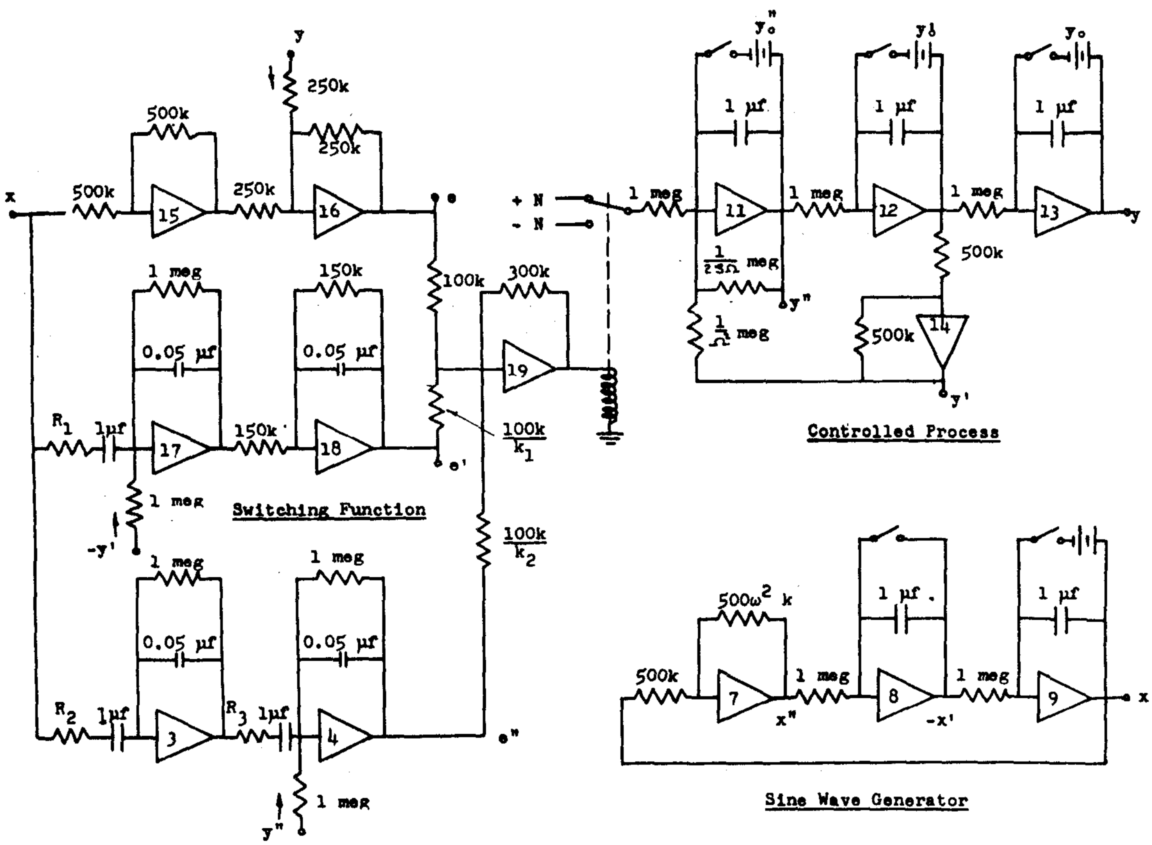

Fig. 7 Analog computer circuit for third order contactor system.

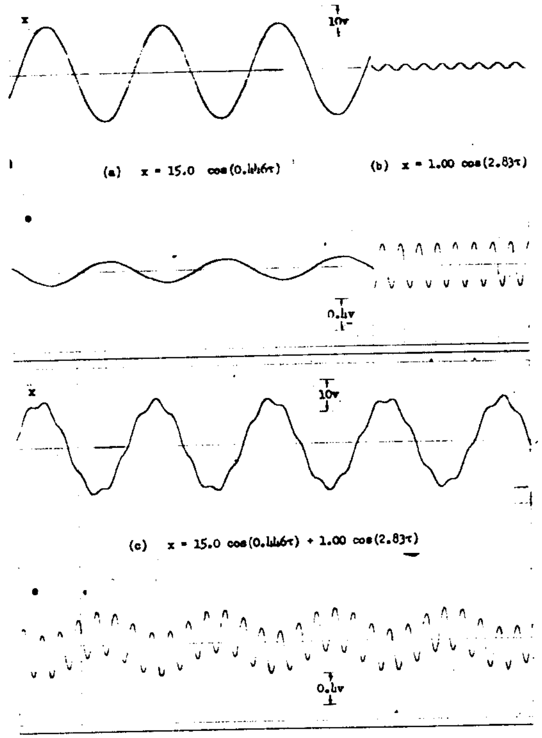

Fig. 8 Chatter response of a third order system. In all cases, $\sum_n T_{1i} = 0.100$; $k_1 = 0.527$; $k_2 = 0.100$.

$\Omega^2 = 2.79$
$\zeta = 0.6$
$N = 47$
$k_1 = 0.527$
$k_2 = 0.079$
$\omega = 0.815$
$A = 20$
$\sum_n T_{1i} = 0.052$

The parameter values as listed at the left are defined as "normal". For other listings in the table below, all of the parameters are "normal" except the one listed.

$$Q = \frac{A k_1 \omega^2 \sum_n T_{1i}}{\sqrt{(1 - k_2\omega^2)^2 + k_1^2\omega^2}}$$

Q_{ex} = experimental error amplitude

	Condition	Q_{ex}	Q	$Q - Q_{ex}$	% Diff.
1	Normal	0.35	0.350	0	0
2	$\sum_n T_{1i} = 0.102$	0.65	0.686	0.036	5 %
3	$k_1 = 1.05$	0.54	0.570	0.03	5 %
4	$k_1 = 0.263$	0.184	0.188	0.004	2 %
5	$\omega = 0.407$	0.08	0.0898	0.0098	11 %
6	$\omega = 0.628$	0.206	0.211	0.005	2 %
7	$A = 10$	0.17	0.175	0.005	3 %
8	$k_2 = 0.213$	0.362	0.380	0.018	5 %
9	$k_2 = 0.588$	0.436	0.488	0.052	11 %

Table 1 Comparison of error amplitudes found in an analog computer simulation with those given by Eq. (35) for $x = A\sin\omega t$.

PHASE-PLANE TRAJECTORIES AS A TOOL IN ANALYZING NON-LINEAR ATTITUDE STABILIZATION FOR SPACE MISSILE APPLICATION

Jack L. Halvorsen
Lockheed Missile Systems Division
Palo Alto, California

Summary

In a variety of missile-control problems, a steady-state oscillation in displacement angle is not objectionable. If this oscillation can be tolerated, the on-off servo is extremely helpful in conserving control energy in the presence of system noise. An attitude-stabilization system using a two-way relay servo, reaction jets, HIG gyros, and lead networks, to give an equivalent rate and position feedback, provides the type of steady-state response previously mentioned.

This paper presents phase-plane techniques to determine the relationship between frequency, required impulse, maximum rates and maximum displacement as a function of lead network parameters, inherent hysteresis, jet force, and time lag. Examples are verified using analogue computer mechanization.

If the transfer function for the lead network is appropriately expanded, the problem of analysis lends itself quite readily to phase-plane solution. In cases where the initial conditions are known, the phase plane gives the transient and steady-state response within the accuracy of graphical solution. The effects of inherent hysteresis, time lags, and extraneous torques are easily determined.

Introduction

Perhaps the most important consideration in the design of the control system is that the system be as simple, inexpensive, and reliable as practicable and yet require as little angular control impulse as possible. If a steady-state oscillation in displacement angle is not objectionable, the on-off servo can be used in lieu of a proportional system, since the on-off servo can be designed to use much less control energy in the presence of system noise. In addition the on-off servo system can be built with suitable reliability and at much less expense than can the proportional system*.

Since this paper is concerned with stabilization it will be assumed that no missile maneuver commands will be introduced into this mode**. The paper will be restricted to roll stabilization by a relatively simple and reliable autopilot system.

*The proportional system is normally more expensive owing to the additional cost of the proportional gas valve.

**The introduction of certain simple types of maneuver command can be handled very readily in the phase-plane analysis, while more complex types require a more complex phase-plane approach. See references 1 and 2.

The phase-plane method, however, is readily adaptable to either of the other two modes of operation.

The attitude stabilization system to be discussed consists in part of a two-way relay servo (bang-bang servo) driving a reaction-jet nozzle valve. The reaction jets are located on the periphery of the missile in a plane perpendicular to the missile body axis. When energized, the jets produce constant control torques in either of opposite directions. These control torques produce a continuous steady-state dither about the missile body axis.

The stabilizing control signal is derived from a HIG gyro. The rate signal used for damping is obtained from either a conventional rate gyro or a suitable lead network.

Missile Roll Dynamics

The equation for the summation of torques is written

$$\Sigma L = \dot{P}I_x + QR(I_z - I_y) - I_{xz}(\dot{R} + PQ) \quad (1)$$

Because of symmetry, the inertial cross-coupling term and the product-of-inertia term are equal to zero, since

$$I_{xz} = 0 \text{ and } I_z = I_y$$

(See figure 1)

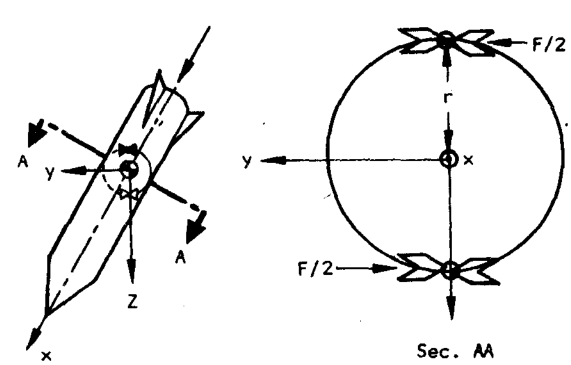

Figure 1. Missile Roll Control Configurations

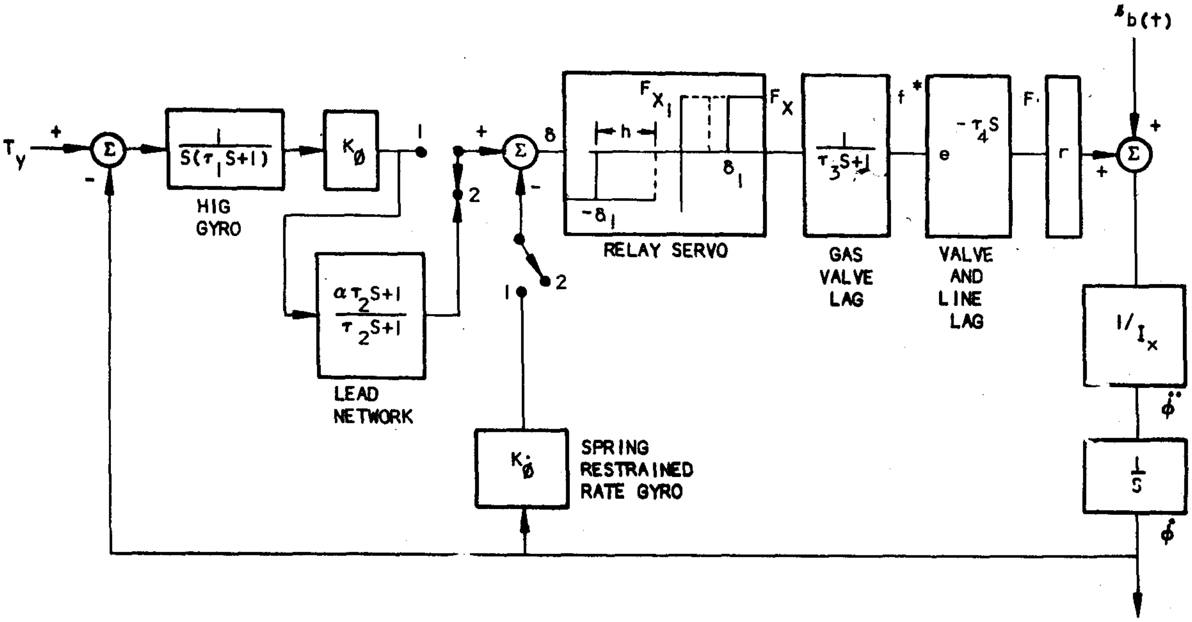

Figure 2. Roll Control System and Missile Dynamics, Block Diagram

Therefore, equation (1) reduces to

$$\Sigma L = \dot{P}I_x = I_x\ddot{\Phi} \qquad (2)$$

where

$$\Sigma L = (F/2 + F/2)r + \ell_{b(t)}$$

F/2 = the force exerted by one nozzle

r = length of the moment arm

The term $\ell_{b(t)}$ represents extraneous torques due to misalignments or accelerating rotational machinery. The $\ell_{b(t)}$ can be a function of time, but should be a step function. (This is not a very severe restriction, since most of the torques encountered can be closely approximated by a step function.) Equation (2) is rewritten

$$\ddot{\Phi} = \frac{Fr}{I_x} + \frac{\ell_{b(t)}}{I_x} \qquad (3)$$

Equation (3) is the equation of motion and will be used to develop the phase-plane trajectories.

The Roll Control System

The roll control system is made up of the sensing device, the servo system, and the missile roll dynamics. It is best described in block diagram form. Two configurations are shown in figure 2. When S_1 is in position 2, the derivative signal is derived from a lead network.

Autopilot Parameters

In analyzing the system, a number of simplifications can be used which will lead to a good approximation for system performance. The HIG gyro time constant (T_1) is typically 0.003 second and is usually negligible in this type of system. A typical "bang-bang" control valve characteristic can be approximated from an actual valve characteristic shown in figure 3. The valve time constant (T_3) is negligible. Consequently, the valve transfer function can be represented by a transport lag of seven

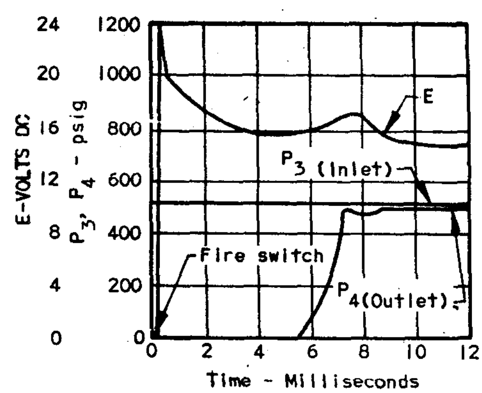

Figure 3. Response Test for the Normally Closed Three-way Futurecraft Solenoid Valve No. 20090

milliseconds if the valve is physically located adjacent to the jet nozzle. Usually an attempt is made to keep the line between the valve and nozzle very short so that additional transport lags are not introduced. All other parameters are considered design parameters and will be controlled to obtain the desired system performance.

Phase-Plane Analysis

The equations for the phase-plane trajectories will be obtained from equation (3). The time constants T_1 and T_3 are considered negligible. The equation of motion is rewritten

$$\ddot{\Phi} = \frac{Fr}{I_x} + \frac{\ell_{b(t)}}{I_x} \qquad \delta_1 < \delta < -\delta_1 \qquad (4)$$

$$\ddot{\Phi} = \frac{\ell_{b(t)}}{I_x} \qquad -\delta_1 + h < \delta < \delta_1 - h \qquad (5)$$

where

$$\frac{d\Phi}{dt} = \dot{\Phi} \qquad (6)$$

and

$$\frac{d\dot{\Phi}}{dt} = \ddot{\Phi} \qquad (7)$$

Dividing equations (4) and (5) by equation (6) and integrating yields the equations of the phase-plane trajectories. They are written

$$\dot{\Phi}^2 = \frac{2}{I_x}[Fr + \ell_{b(t)}]\Phi + C_1 \qquad \delta_1 < \delta < -\delta_1 \qquad (8)$$

$$\dot{\Phi}^2 = \frac{2\ell_{b(t)}}{I_x}\Phi + C_2 \qquad -\delta_1 + h < \delta < \delta_1 - h \qquad (9)$$

where C_1 and C_2 are constants of integration.

The trajectory is obviously parabolic in nature, and the foci always lie on the Φ axis but are translated and reversed in sign as the motion continues. Consequently, a template that will aid in plotting the trajectories is easily made.

The equation for the control quantity δ is written

$$-\delta_{(s)} = \left[\frac{K_{\dot{\Phi}}}{K_{\Phi}} S + 1\right] K_{\Phi}\, \Phi_{(s)}$$

or

$$-\delta = \left[\frac{K_{\dot{\Phi}}}{K_{\Phi}} \dot{\Phi} + \Phi\right] K_{\Phi}$$

Whenever $\delta_1 < \delta < -\delta_1$ the reaction jets are thrusting. The jet-on line is therefore defined by equation (10)

$$\pm\delta_1 = \delta = \left[\frac{K_{\dot{\Phi}}}{K_{\Phi}} \dot{\Phi} + \Phi\right] K_{\Phi} \qquad (10)$$

Equation (10) can be rewritten by solving for $\dot{\Phi}$

$$\dot{\Phi} = -\frac{K_{\Phi}}{K_{\dot{\Phi}}}\Phi \pm \frac{\delta_1}{K_{\dot{\Phi}}} \qquad (11)$$

The jet-off lines are derived from the conditions that the jets are off for $-\delta_1 + h < \delta < \delta_1 - h$

The jet-off equations are written

$$\dot{\Phi} = -\frac{K_{\Phi}}{K_{\dot{\Phi}}}\Phi + \left(\frac{\delta_1 - h}{K_{\dot{\Phi}}}\right) \qquad (12)$$

and

$$\dot{\Phi} = -\frac{K_{\Phi}}{K_{\dot{\Phi}}}\Phi - \left(\frac{\delta_1 - h}{K_{\dot{\Phi}}}\right) \qquad (13)$$

Transport Lags and Time Approximations from the Phase Plane. There are a number of ways of determining the elapsed time from the phase-plane trajectories. The two methods found most convenient for the purpose of this paper are shown below.

In method one, by definition

$$\ddot{\Phi} = \frac{d\dot{\Phi}}{dt}$$

$$dt = \frac{d\dot{\Phi}}{\ddot{\Phi}}$$

then

$$t_1 - t_0 = \int_{\dot{\Phi}_0}^{\dot{\Phi}_1} \frac{d\dot{\Phi}}{\ddot{\Phi}} \approx \frac{\Delta\dot{\Phi}}{\ddot{\Phi}_{(1)}} \qquad (14)$$

if $\ddot{\Phi}_{(1)}$ is a constant.

Method two uses the definition

$$\dot{\Phi} = \frac{d\Phi}{dt}$$

then

$$t_1 - t_0 = \int_{\Phi_0}^{\Phi_1} \frac{d\Phi}{\dot{\Phi}} \approx \frac{\Delta\Phi}{\dot{\Phi}(\text{average})} \qquad (15)$$

The transportation lag is now easily handled. The motion proceeds along the trajectory until the jet-on or jet-off line is traversed, at which time (t_0) the control quantity calls for a torque. Because of the transport lag, the motion continues along the same trajectory until time (t_1) when the force is applied at the nozzle and the body changes its direction of motion. Since $t_1 - t_0$ is specified by the time constant of the transport lag, the transition point on the trajectory is easily calculated from equation (14) or (15).

Lead network case. The equation for the control quantity δ can be written

$$-\delta(s) = \left(\frac{\alpha\tau_2 s+1}{\tau_2 s+1}\right) K_\Phi \Phi(s) \qquad (16)$$

Applying the well-worn but useful technique of expanding equation (16) into a power series and then obtaining the equation for δ in the time domain yields equation (17)

$$-\delta = K_\Phi\left[\Phi + \tau_2(\alpha-1)\dot{\Phi} - \tau_2^{2}(\alpha-1)\ddot{\Phi} + \tau_2^{3}(\alpha-1)\dddot{\Phi} + \ldots\ldots \qquad (17)\right.$$

When the network time constant (τ_2) is small, the higher-order terms become negligible and the equation for δ is written

$$-\delta \approx K_\Phi\left[\Phi + \tau_2(\alpha-1)\dot{\Phi} - \tau_2^{2}(\alpha-1)\ddot{\Phi}\right] \qquad (18)$$

Within the dead band, the acceleration is zero and the equation for the jet-on line is written from equation (18)

$$\dot{\Phi} = \frac{-1}{\tau_2(\alpha-1)}\Phi \mp \frac{\delta_1}{K_\Phi[\tau_2(\alpha-1)]} \qquad (19)$$

The equation for the jet-off line is also written from equation (18)

$$\dot{\Phi} = -\frac{1}{\tau_2(\alpha-1)}\Phi \mp \frac{\delta_1 \pm h}{K_\Phi\tau_2(\alpha-1)} \pm \tau\ddot{\Phi}_{(1)} \qquad (20)$$

Thus the transition line is a function of any torques that might accelerate the body when lead compensation is used.

When a larger time constant is required in the compensating network, or if an exact expression is required for the control quantity δ, equations (19) and (20) do not give satisfactory results. An exact expression for δ for a step input in the second derivative of the input displacement is derived in Appendix A. Although the expression is expanded for step inputs in acceleration to the lead network, the expansion has been done in general terms for step inputs in the nth derivative by the author and the end result is contained in Appendix A.

Substituting into equation (F) and rearranging gives the jet-off equation

$$\dot{\Phi} = \frac{1}{\tau_2(\alpha-1)}\Phi \pm \frac{\delta_{(1)} \mp h}{K_\Phi\tau_2(\alpha-1)}$$

$$\mp \tau_2\ddot{\Phi}_{(1)}\left(1 - e^{-\frac{t}{\tau_2}}\right) \qquad (21)$$

It is seen that equation (21) differs from equation (20) only by the transient exponential term. Now, since the elapsed time (t) in the transient term can be determined from equation (14), the jet-off line can be plotted on the phase portrait after obtaining a few points by trial and error. The jet-on lines are obtained by allowing the acceleration term and hysteresis term (h) to go to zero. Under these conditions, equation (21) reduces to equation (19). In cases where extraneous torques exist, the jet-on line also shifts, since the extraneous torques cause the body to accelerate. Consequently, equation (21) must be used with the proper magnitude and direction of the acceleration that results from the extraneous torques.

Angular Impulse Calculations

The required control energy in this type of autopilot usually must be kept to a minimum because of weight and space limitations. This requires that angular-impulse calculations must be made to determine the amount of pressurized gas that must be carried. The required angular impulse is easily calculated from the phase portrait. The equation for angular impulse is written

$$r\,|F|\,dt = I_x d\dot{\Phi}$$

Integrating over the elapsed time gives

$$r\int_{t_o}^{t_1}|F|\,dt = I_x\int_{\dot{\Phi}_o}^{\dot{\Phi}_1} d\dot{\Phi} \approx I_x(\Delta\dot{\Phi}) \qquad (22)$$

Usually the system will fly in the steady-state dither for nearly the entire flight. Consequently, the impulse required during steady-state will be of the greatest import. If $t_1 - t_o$ is the period of one cycle of the fundamental steady-state dither, then $\Delta\dot{\Phi}$ must be the change in angular rate per cycle. Multiplying both sides of equation (22) by the fundamental frequency (f) of the steady-state oscillation yields an expression for the angular impulse required per second. It can be seen that, as $(\Delta\dot{\Phi})f$ approaches a minimum, the total steady-state impulse does also.

Discussion

It should be kept in mind at this point that the objective of this paper was to show that this type of a control system problem can be nicely handled using phase-plane portraits. Consequently, no attempt was made to come up with an optimized system. The following discussion indicates the general trend of system performance as a function of the control parameters. The optimized system will then depend first upon the designer's definition of an optimized control system and, secondly, upon the ingenuity and skill of the designer in using the phase-plane method of analysis.

Three examples have been selected to demonstrate the technique. The results are shown in figures 4, 5, and 6. Phase portraits plotted directly from analogue-computer simulation are included to verify analytical results. The system constants and parameters to be used are shown in table I. Substituting the assigned quantities