ON VARIOUS METHODS OF OPTIMALIZING CONTROL

SEIZO FUJII

Automatic Control Laboratory

(Received October 31, 1961)

CONTENTS

Nomenclature

The following nomenclature is used in this paper:

A, A' = amplitudes of rectangular and sinusoidal waves as the correlate signal, respectively,

a, a_1, a_2 = incremental step size of the input (in Chap. III), amplitude of triangular input (in Chap. IV),

a' = amplitude of sinusoidal wave approximating to the input,

c = allowable difference of the output (defined in Chap. IV),

E = amplitude of sinusoidal wave as the test signal,

E_1, E_2 = errors in y_s/y_m^* due to y_0 and $\dot{y}_0$, respectively,

$f(z)$ = probability density function of the statistical variable z,

$G_i(s)$, $G_0(s)$ = transfer functions of the input and output linear groups, respectively,

$g_i(t)$, $g_0(t)$ = weighting functions corresponding to $G_i(s)$ and $G_0(s)$, respectively,

H = hunting loss,

H_s, H_t, H_r, H_n = hunting losses of sinusoidal wave, triangular wave, rectangular wave and random signal, respectively,

H_s', H_t', H_r', H_n' = dimensionless hunting losses H_s/kE^2, H_t/kE^2, H_r/kE^2 and $H_n/k\phi_{11}(0)$, respectively,

$j = \sqrt{-1}$,

K_p = proportional sensitivity,

L = dead time,

k, k_1, k_2 = characteristic constants of the controlled system,

m = number of input cycles over which the correlation process is carried out,

$m' = T/\tau_{11} = m$ corresponding to the case of random signal,

N = number of moves which is required for the output to reach within ε%, when the input starts from the maximum value x_m,

$n(t)$ = noise evaluated in the output of the controlled system,

$\overline{n(t)}$ = time average of $n(t)$,

n_1, n_2 = random signal as the test input and noise, respectively,

$R = R_s + R_n$,

R_r = cross-correlated value in case where the test input is a sinusoidal wave and the correlate signal is a rectangular wave,

R_{n1} = cross-correlated value in case where both the test signal and correlate signal are random signals,

R_n = cross-correlated value between $n(t)$ and $r(t)$,

R_s = cross-correlated value between $y(t)$ except $n(t)$ and $r(t)$,

$r(t)$ = correlate signal,

$S = a/T$ (in Chap. III), input drive spped $2a/T$ (in Chap. IV),

s = operator of Laplace transformation,
T = hunting period of $y(t)$ (in Chap. IV), period of $x^*(t)$ (in Chap. V),
T_s = length of time between steps, or sampling period,
$T_{sum} = \tau_{01} + \tau_{02}$,
t = time,
t^* = time instant when y becomes the maximum value,
x = normalized pseudo-input of the controlled system, or normalized output of $G_i(s)$,
x^* = normalized input to $G_i(s)$,
x_n = normalized input applied to the controlled system at the sampling insant nT_s,
x_m = maximum value of the normalized input,
x_{max} = value of the normalized input corresponding to the extreme value,
$x_\infty = \lim_{n\to\infty} x_n$,
Δx_l = saturated value of variation of the normalized input,
$\Delta x_n = x_{n-1} - x_n$,
δx = test input with constant step size,
$Y(s)$ = Laplace transform of $y(t)$,
y = normalized output of $G_0(s)$,
$y^* = kx^2$ = normalized input to $G_0(s)$, or normalized pseudo-output of the controlled system,
y_n, $\dot{y}_n$ = value and differential value of y at the sampling instant $(n+1)T_s$, respectively,
y_m = maximum value of $|y|$,
$y_m^* = kx_m^2$ = maximum value of y^*,
y_{max} = extreme value of y,
y_s = value of y at the instant x_0/S,
$\Delta y_n = y_{n-1} - y_n$,
δy = variation of y corresponding to δx,
z = statistical variable (see Eq. (5.15)),
$\bar{z}$ = time average of z,
$\beta = \omega_n x_0/S$,
$\beta' = x_0/T_{sum}S$,
Δ = hunting zone,
ε = magnitude of allowable error represented by a percentage of y_m,
ζ = damping ratio,
λ, $\lambda_n = x_0/x_m$ and x_n/x_m respectively,
$\mu = a/x_m$ or $\delta x/x_m$,
$\nu = \tau_{02}/\tau_{01}$,
σ_z^2 = variance of z,
τ = time constant of $\phi(t)$,
τ_{11}, τ_{22} = time constants of $\phi_{11}(t)$ and $\phi_{22}(t)$, respectively,
τ_i, τ_{i1}, τ_{i2} = time constants of $G_i(s)$,
τ_0, τ_{01}, τ_{02} = time constants of $G_0(s)$,
$\Phi(\omega)$ = power spectrum of $n(t)$,
$\phi(t)$ = auto-correlation function of $n(t)$,
$\phi_{11}(t)$, $\phi_{22}(t)$ = auto-correlation functions of $n_1(t)$ and $n_2(t)$, respectively,
$\phi_{12}(0) = \overline{n_1(t)\cdot n_2(t)} = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} n_1(t)n_2(t)dt$,
φ_i, φ_0, φ_0', φ_1, φ_2, φ_3 = phase angles (rad),

ω_i, ω_0 = angular frequencies of $x^*(t)$ and $y(t)$ respectively; ω_i is always related to ω_0 by $\omega_0 = 2\,\omega_i$,

ω_n = natural angular frequency.

One prime and two primes added to x and y denote the dimensionless and actual values of input and output, respectively; the follwoing relations hold between the quantities: $y' = x'^2$, $x' = K_p kx$, $y' = K_p^2 ky$, $y'' = y''_{\max} + y$ and $x'' = x''_{\max} + x$.

Introduction

In order to maintain the output of any engineering system, such as the efficiency of boiler plant, at a desired extreme value in spite of natural changes in the environment of the system, it is necessary to apply an optimalizing control.

There are two scientific methods in the schemes of such an optimalizing control. One is "model method", which is to determine an extreme value by solving the mathematical model of the controlled system with a computor, and the other is "direct method", which is to find out an extreme value and hold the output at it without using any mathematical model, but basing only on the knowledge that the output of the controlled system has an extreme value. The one is less practical than the other, since it is generally difficult to obtain an exact analytical expression, or a mathematical model, relating the output to the input of the controlled system in practice.

For these ten years, various kinds of control methods belonging to the class of the direct method have been conceived and developed by several investigators. In 1951, Draper and Li,[1] who proposed the original idea of the optimalizing control, devised "peak-holding method" which is little affected by noise interference, and applied it to the optimum control of the internal combustion engine, in which the consumption of fuel to produce the load torque at the specified speed was minimized by adjusting the ignition timing. After that, the peak-holding method was analyzed by Tien and his coworker.[2][3] Another example of applications was the cruise control of an airplane[6]; there, under the restriction of engine cruising r.p.m. and assigned altitude, there is an optimum combination of trim setting and engine throttle for maximum miles per gallon of fuel. Moreover, on the basis of such an idea, several simple optimalizing controllers consisting of the elements such as logic circuits have been constructed,[4][5] but their applications to the actual system have not been reported.

Draper and Li also tried "cross-correlation method", which is to find out an extreme value by intentionally injecting a test signal, such as a sinusoidal wave, into the controlled system. The application of this method to the optimum control of the fuel consumption for maximum output in a Jet engine was discussed by Vasu.[7]

Opcon has been developed by Westinghouse Co.,[8] on the basis of another operating principle of an optimalizing control. This is temporarily called "cut-and-try method", which is to find out an extreme value of the output by the incremental change in the input. This method will be of great use in searching out an extreme value of the output as a function of many inputs.

These three mentioned above are regarded as the practical optimalizing control schemes in the class of the direct method. These methods have been applied to a few engineering systems as mentioned above, but little has been discussed theoretically as to the analysis and synthesis of these systems.

In this paper, we shall analize in detail the optimalizing control system with one input in its theoretical aspect, choosing several of the practical methods from the class of the direct method. As the result, we shall present many data and clearer insight to the design of such optimalizing control systems.[9)-13)]

Chapter I. General Views of the Optimalizing Control System

1. Fundamental problems

In order to design the optimalizing control system, which is also known as "extremum control system", we must, in general, take into account the following items:

(1) The output of the controlled system has an extreme value, but its magnitude can not be exactly predicted in advance.

(2) There are the effects of the dynamics, usually existing between the input and output of the controlled system, on the performance of the system.

(3) The performance of the system is also affected by noises in the system or from the outside of the system.

(4) The extreme value of the output tends to change in its magnitude with natural changes in the environment of the system.

(5) The optimalizing control methods which meet these requirements should be invented, and the controllers which implement the methods are required to be simple in construction and justifiable economically.

These items are somewhat interrelated each other. In short, the ultimate aim of the optimalizing control is that the output of the system reaches an extreme value as quickly as possible, and the system operates always as close to the optimum state as possible or has the smallest hunting loss, and is not misled by the noise interference.

2. Generalized optimalizing control system

In the engineering control system, for example, in the boiler plant, the efficiency of boiler as the output of the controlled system has an extreme value or maximum value at a certain value of the excess air ratio as a single controlled input, hereafter referred to as the input. Furthermore, the maximum value changes in its magnitude with the steam consumption as a disturbance.

For simplicity of the analytical treatment of the optimalizing control system, it is assumed that the steady state characteristics relating the output to the input can be approximated by a parabolic curve in the neighbourhood of the extreme point, and that the disturbance is so slow to occur that we may neglect its effect on the characteristics during the period in which a series of control actions takes place. This fact permits us to normalize the characteristic curve, that is; the input and output can be measured from their optimum point. Suppose the extreme value exists at the input x_e^* and the output y_e''. This point is then chosen as the origin of a new coordinate or normalized coordinate (x, y) and thus the extreme value occures at the point $x=y=0$. The input and output in the normalized coordinate are related to those in the actual coordinate by $x''^* = x^* + x_e''^*$ and $y'' = y + y_e''$, respectively.

It is to be noted that the fundamental operation of the optimalizing control should not be affected by such a transformation of the coordinate system; in other words, the control operation by the direct method does not depend upon the absolute values of the input and output, as will be shown in the subsequent sections.

Figure 1.1 shows the block diagram of a generalized optimalizing control system with a single controlled input. The relation that $y^*=kx^2$ represents the static performance characteristics of the controlled system with the normalized input and output, as mentioned above. $G_i(s)$ and $G_0(s)$ are the transfer functions of the input and output linear groups of the controlled system respectively, and $G_i(0)=G_0(0)=1$ is assumed to hold. Here, we shall not touch upon what these transfer functions represent in the practical engineering systems. x^* and y will be refered to as the (actual) normalized input and output, and x and y^*, as the normalized pseudo-input and -output respectively, if necessary. When the dynamic effects are neglected, it is readily seen to be $x^*=x$ and $y^*=y$.

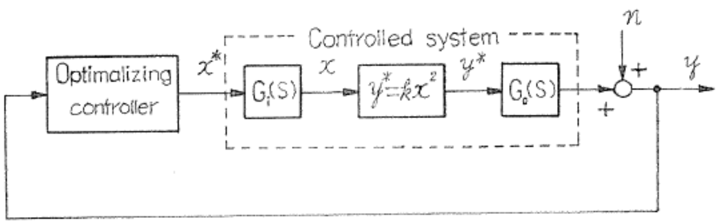


FIG. 1.1. Block diagram of a general optimalizing control system with a normalized controlled system.

Chapter II. Cut-and-Try Method

This control method is to find out an extreme value in a stepwise manner: an incremental step in the input is given to the controlled system at constant sampling periods, and the resulting variation of the output is measured, and then the direction and size of a new incremental step in the input are determined on the basis of these informations. In this method there are three main types, that is, proportional-difference type, on-off type, and proportional-gradient type.

Part I. Proportional-difference type

1. Principle of operation

Figure 2.1 shows a typical control behavior for the proportional-differnce type of the cut-and-try method. Figure 2.2 is a functional diagram of the optimalizing controller for this control type at the sampling instant $(n+1)T_s$. The upper left part of Fig. 2.1 illustrates a static performance characteristics of the controlled system, in which the output has different maxima for various levels of the disturbance.

Suppose the output is at its maximum value with the disturbance of the state I, then if the disturbance changes from the state I to II in Fig. 2.1, the output will be decreased from y''_{-1} to y''_0 according to this static characteristics, where the input is assumed to remain unchanged. Here, the optimalizing controller of this control type is made to operate with a constant sampling period T_s, and con-

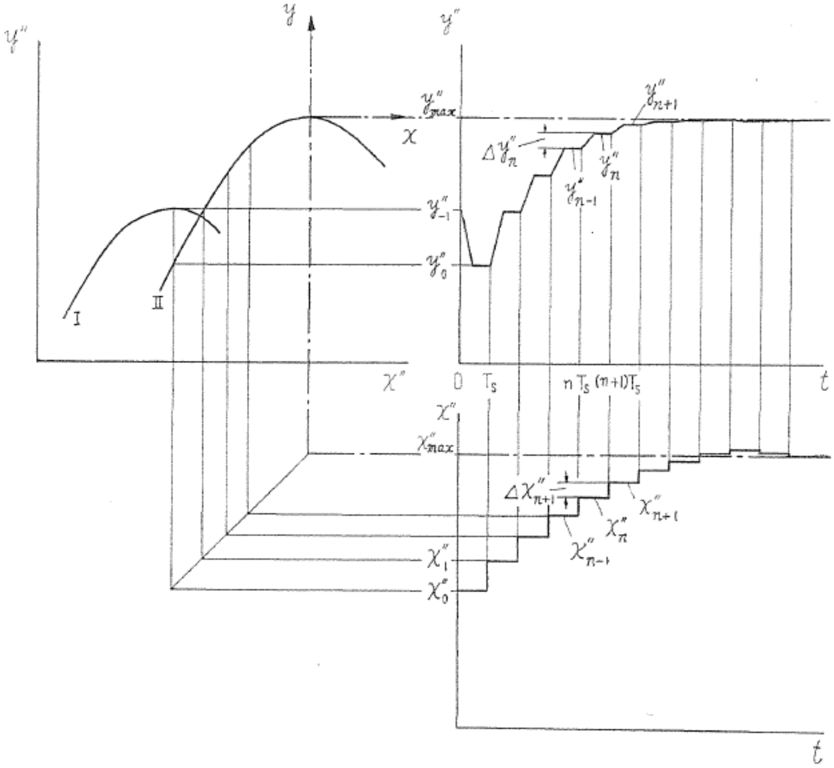


FIG. 2.1. Typical control behavior by the proportional-difference type of the cut-and-try method.

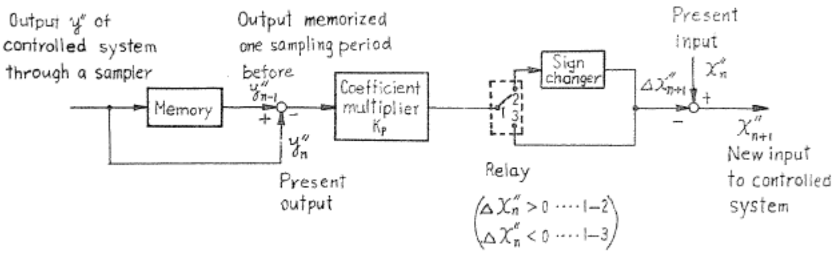


FIG. 2.2. Functional diagram of the optimalizing controller for the proportional-difference type of the cut-and-try method, at the sampling instant $(n+1)T_s$.

sists of a measuring device of the output, a memory unit and a simple logic circuit and so on. The roll of the logic circuit is to decide the direction of the input change, that is; whether it should be increaased or decreased. Now, the output y_0'' is subtracted from the previous output y_{-1}'', which was memorized one sampling period before, to give the variation $\Delta y_0''(=y_{-1}''-y_0'')$ of the output. Thus, with the variation $\Delta y_0''$ and the value x_0'' of the input at that time as the initial conditions, the control operation starts to find out a new extreme value (x_{max}'', y_{max}'') at the state *II* of the disturbance, as shown in Fig. 2.1. The subsequent control operation can be explained as follows, referring to Figs. 2.1 and 2.2. At the sampling instant nT_s, the input x''_n is applied to the controlled system, and the resulting steady state value y''_n of the output is measured after a sufficiently long time interval T_s for the transient to die out. The value y''_n is compared with the value y''_{n-1}, which corresponds to the input x''_{n-1} and has been stored in a memory unit. The result gives the difference $\Delta y''_n = y''_{n-1} - y''_n$ of the output. The difference $\Delta y''_n$ is multiplied by K_p at the coefficient multiplier: $K_p \Delta y''_n$. The sign of $K_p \Delta y''_n$, that is, the sign of the new variation $\Delta x''_{n+1} (= x''_n - x''_{n+1})$ is decided, as shown in Table 2.1, depending on the sign of the preceding input variation $\Delta x''_n$ and the sign of the present difference $\Delta y''_n$. The new variation $\Delta x''_{n+1}$ with the sign is subtracted from the present input x''_n to give the new input x''_{n+1}. Once the new input x''_{n+1} is determined, the value y''_{n-1} which has been stored during this period is eliminated to store newly the present value y''_n of the output. The similar control operation is repeated till the output reaches the maximum value with the disturbance at the state *II*.

TABLE 2.1. Sign of New Input Increment x''_{n+1}

Sign of x''_n	Sign of y''_n	Sign of x''_{n+1}
+	+	−
+	−	+
−	+	+
−	−	−

2. Analysis of the control system

As is seen in the previous section, the proportional-difference control type depends upon the variations of the input and output, but not upon their absolute values. Accordingly, the optimalizing system of this control type can be generally analyzed on the normalized controlled system shown in Fig. 1.1. Moreover, if we choose the sampling period T_s so long as we can neglect the dynamic effects, then we can deduce this control system to the form, shown in Fig. 2.3, with the dimensionless input x' and output y', which satisfy the relation that $y' = x'^2$.

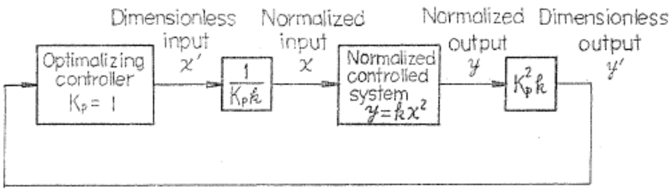


FIG. 2.3. Block diagram of the optimalizing control system with the proportional-difference type of the cut-and-try method, represented by use of the dimensionless input and output.

Even when the dynamic effects are absent, the stability of the control behavior, the steady state deviation of the output from the extreme value, or offset, and the time required for the output to reach the extreme value or quickness of the response are affected by K_p, the proportional sensitivity, k, the characteristic constant, and $(x'_0, \Delta y'_0)$, the initial conditions, which must be included because of the non-linearity of the system.

2.1. Stability of the control behavior

Among the problems proposed in the last paragraph of Sec. 1 only the effect of the initial condition can be discussed here because if the analysis of the controlled system is treated with the dimensionless input x' and output y' shown in Fig. 2.3, the effects of K_p and k become implicit.

Figures 2.4 (a) and (b) show the typical examples of stable control behavior and unstable one due to the effects of the initial conditions. In general, it is difficult to determine what initial conditions make the control behavior stable or not. It is seen that the control behavior tends to be unstable with increase of the value of the initial condition. In other words, even if the control behavior is stable for a large value of initial condition, the output will oscillate many times with large amplitudes around the extreme value before the output settles to it.

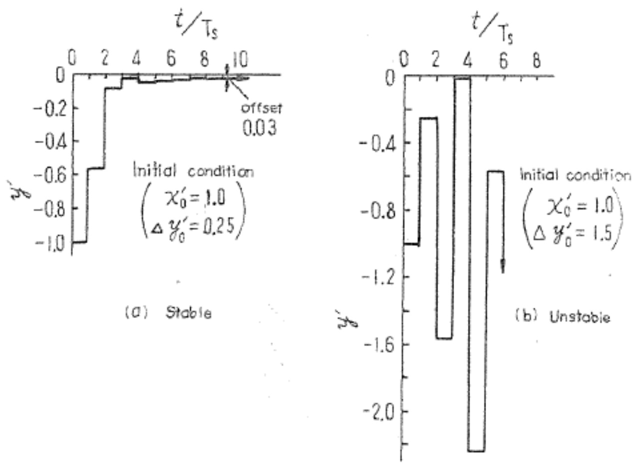


FIG. 2.4. Typical examples of control behavior depending on the initial conditions.

Here, we are going to explain the fundamental control behaviors necessary for the later analysis of the system. Starting from the assumption that the first step in the input is positive and towards the optimum value $(x'=0)$ corresponding to the extreme value $(y'=0)$, three different control behaviors are produced by two following steps, as shown in Figs. 2.5 (1), (2) and (3).

The control behavior (1) is that when the input x' starts from an initial value

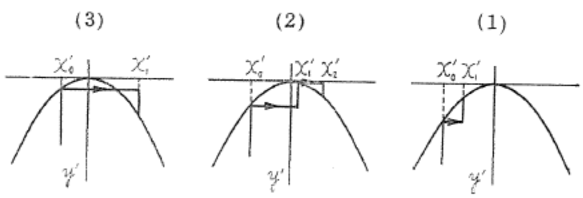


FIG. 2.5. Three fundamental control behaviors.

x'_0, the new input x'_1 has the same sign as the input x'_0, and the absolute value of x'_1 is less than that of x'_0, and then the next step in the input is moved towards the optimum value. The control behavior (2) is that since the input x' goes beyond the optimum value, the input x'_1 has an opposite sign to the input x'_0, and $|x'_1|$ is less than $|x'_0|$, and thus the new input x'_2 goes further away from the optimum value according to the principle of the control operation mentioned previously. The control behavior (3) is that in the similar manner to (2) the input x'_1 has an opposite sign to the input x'_0, but $|x'_1|$ is more than $|x'_0|$, and thus the new input x'_2 comes closer to the optimum value. A careful inspection of the control behavior shows that there exists an oscillation in which the output y' has the amplitude 1 and the period $2\,T_s$, and the input x' has the amplitude 1 and the period $4\,T_s$, as is readily obtained from the relation that $y'=x'^2$. Therefore, in order to determine the initial conditions giving the stable control behaviors, it is convenient to consider the plane of the initial condition $(x'_0, \Delta y'_0)$, as shown in Fig. 2.6, and to divide the plane into two regions: one is the region in which both x'_0 and $\Delta y'_0$ are less than 1, and the other is that in which both x'_0 and $\Delta y'_0$ are more than 1. (Since $y'=x'^2$ is symmetric for the axis $x'_0=0$, we shall hereafter consider only the case where both x'_0 and $\Delta y'_0$ are positive, and assume that the first step in the input starts towards the optimum value.)

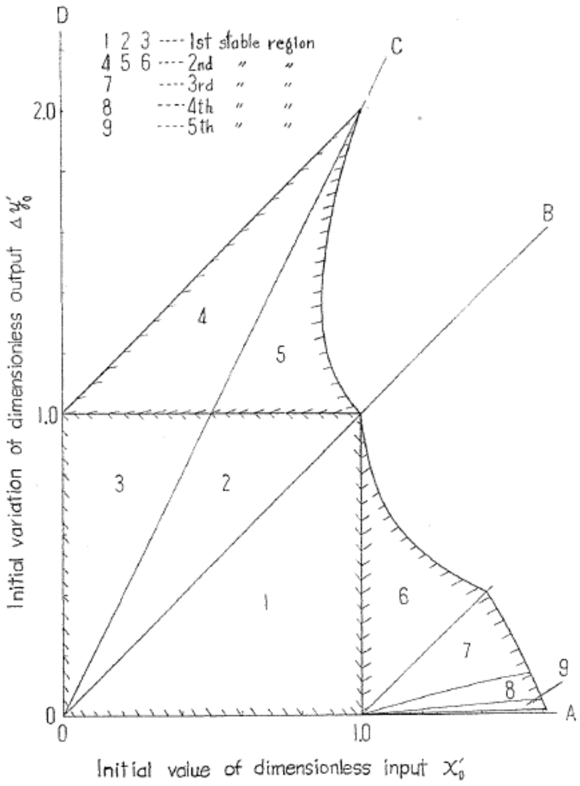


Fig. 2.6. Stable regions depending on the initial conditions.

When the control behavior starts from any point (x'_0, $\Delta y'_0$) in the region that both x'_0 and $\Delta y'_0$ are less than 1 in Fig. 2.6, it is understood that at the successive instants T_s, $2\,T_s$, . . . , the inputs x'_1, x'_2, . . . and the corresponding variations of the output $\Delta y'_1$, $\Delta y'_2$, . . . are all less than 1. Accordingly, the output may approach, nonoscillatory or oscillatory, the extreme value: the control behavior is stable for the initial conditions in this region. This region is temporarily called "the first stable region". Let us next consider the region that the initial condition (x'_0, $\Delta y'_0$), are both more than 1. Even if the control behavior starts from a point (x'_0, $\Delta y'_0$) in this region, it is obviously stable if the next point (x'_1, $\Delta y'_1$), [or the third point (x'_2, $\Delta y'_2$) for the case (2) in Fig. 2.5] falls within the first region, or in other words if both x'_1 and $\Delta y'_1$ are less than unity. Thus, the stable control behaviors can also be attained in the region consisting of the initial conditions (x'_0, $\Delta y'_0$) entering into the 1st stable region at the following sampling instant or at the third sampling instant for the case of Fig. 2.5 (2). This region is called the 2nd stable region.

We can again look for the initial conditions which are followed by a point in the 2nd stable region in the region that x'_0, $\Delta y'_0 > 1$. This region is also temporarely called the 3rd stable region; the control behavior, starting from the 3rd stable region, enters into the 1st stable region through the 2nd stable region.

On the basis of such an idea, we can extend the stable regions on the plane (x'_0, $\Delta y'_0$). For example, Fig. 2.6 shows several stable regions determined by the similar procedure. However, as will be seen in the subsequent sections, it is not necessary for the design of the control system to determine all stable regions.

2.2. Deviation of the output from the extreme value (offset)

When the control behavior starts with any initial condition (x'_0, $\Delta y'_0$) in the stable region determined in the previous section, questions arise as to what couse the initial conditions (x'_1, $\Delta y'_1$), (x'_2, $\Delta y'_2$), . . . at the successive sampling instants are following on the plane of initial condition, before the output comes into the vicinity of the extreme value, and also as to whether the output converges to the extreme value or some other value, in other words whether the output has an offset or not.

In order to track the control behavior easily, it is convenient to divide the 1st stable region where $x'_0 < 1$, $\Delta y'_0 < 0$ according to the fundamental control behaviors. Figure 2.7 shows only the 1st stable region. In this figure, the regions corresponding to three fundamental control behaviors (1), (2) and (3) are given by Δaeb, Δabc and Δacj, respectively.

Now, assuming the output converges to some value deviated from the extreme value with no overshoot, let us determine the relations to be satisfied among the successive initial conditions. Both x'_0 and $\Delta y'_0$ are assumed to be positive. Then, the

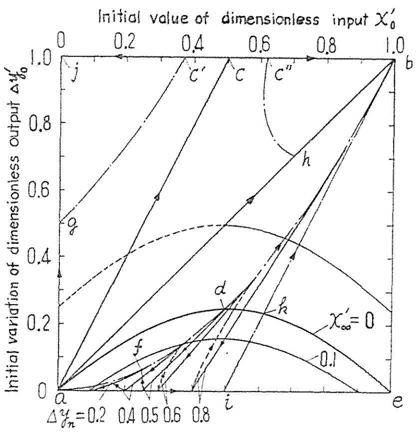


FIG. 2.7. Diagram for estimating the control behavior and offset dependent on the initial conditions.

relations among the successive initial conditions $(x'_1, \Delta y'_1)$, $(x'_2, \Delta y'_2)$, . . . , $(x'_n, \Delta y'_n)$, according to the principle of operation, are given by

$$x'_1 = x'_0 - \Delta y'_0,\ \ x'_2 = x'_1 - \Delta y'_1,\ \ldots,\ x'_n = x_{n-1} - \Delta y'_{n-1},\ \ldots, \tag{2.1}$$

where $\quad \Delta y'_i = y'_{i-1} - y'_i \quad (i \geq 1), \qquad y'_i = x_i^2 \quad (i \geq 0).$

By summing up each side of the Eq's 2.1, we obtain the following relation between x'_0 and $\Delta y'_0$:

$$\Delta y'_0 = -\left(x'_0 - \frac{1}{2} \right)^2 + \left(x'_\infty - \frac{1}{2} \right)^2, \tag{2.2}$$

where x'_∞ is the final value to which the input converges, i.e., $\lim_{n\to\infty} x'_{n-1},\ x'_n \to x'_\infty$.
Especially when the output converges to the extreme value, we have

$$\Delta y'_0 = -\left(x'_0 - \frac{1}{2} \right)^2 + \frac{1}{4} \tag{2.3}$$

by letting $x'_\infty \to 0$ in Eq. (2.2).
Whenever the control behavior starts with the initial condition $(x'_0, \Delta y'_0)$ satisfying Eq. (2.3), the output converges to the extreme value $(y'=0)$. If x'_∞ is chosen not zero but a finite positive value, the parabolic curve, Eq. (2.3) is shifted down vertically along $\Delta y'_0$ axis. For example, Fig. 2.7 shows the curve, *ade*, for $x'_\infty = 0$ and the curve for $x'_\infty = 0.1$. It is seen from this figure that the output, starting with any initial conditions between two curves, has an offset of less than 0.1 as evaluated in the input.

Now, the 1st stable region can be divided into two parts by the parabolic curve *ade*. In the part under the curve *ade*, the control behavior has the successive initial conditions along a parabolic curve which is formed by shifting down the curve *ade* along $\Delta y'_0$ axis until it goes through the first initial condition $(x'_0, \Delta y'_0)$. Then, the output converges, without overshoot, to the square of x' at the intersection of the new parabolic curve with x'_0 axis. As the output starts with the initial conditions underpositioned further from the curve *ade*, the output has a larger offset. It is also seen from Eq. (2.2) that when $x'_\infty = 1/2$, the parabolic curve disappears from the plane in Fig. 2.7. Therefore, together with the considerations mentioned later, it may be concluded that the offset is within 0.5 as evaluated in the input.

Let us next consider the upper part of the curve *ade*. Starting with any initial condition in the upper part, the output moves along a new parabolic curve, which is formed by the similar procedure as before. After some sampling instant, the output takes the initial conditions on the extension of the new curve either in region (2) or (3), as is shown by the chain curve in Fig. 2.7.

Let us next consider the case where the control behavior starts with any initial condition either in region (2) or (3). To what regions in the plane does the following initial conditon, $(x'_2, \Delta y'_2)$ for the region (2) and $(x'_1, \Delta y'_1)$ for the region (3), move respectively? The detailed analysis is not shown here, but the computation has revealed the following facts: the boundary line between the regions (2) and (3) corresponds to the line *ai* in the region (1); the boundary line *ab* between the regions (1) and (2) corresponds to the parabolic curve *afdb* in the region (1); the line *cj* in the region (3) or the line *cb* in the region (2) corresponds to the line *ib* in the region (1), where the direction of the arrow corresponds to each other. From this result, it is seen that at the following sampling instant any initial condition in the regions (2) and (3) goes into the sub-region *afdbi* in the region (1) but does not go into the region (3) from (2) and *vice versa*. In order to inspect further the region, in the region *afdbi* we shall show the curves (chain line and fine line) corresponding to $\Delta y'_0 = 0.2$, 0.4, 0.5, 0.6 and 0.8, where the direction of an arrow corresponds to that of increasing x'_0, and the region *afdki* corresponds to the region *ahc''c* in (2) and *acc'g* in (3), in which the input converges to any value between 0 and 0.5 with one over shoot. It is also seen that when the output starts with other initial conditions in the stable regions, the output may overshoot the extreme value more than once, as the initial condition deviates larger towards the upper right part of the parabolic curve *ade*.

In order to make the offset smaller, we shall be able to devise the following counterplans from the inspection of the control behavior on the plane of the initial condition $(x'_0, \Delta y'_0)$.

(1) One of the counterplans is to introduce the saturation to the dimensionless input variation, $\Delta x'_n\ (=x'_{n-1} - x'_n)$, as shown in Fig. 2.8. In other words, the input must not be changed to more than a certain size. Here it is to be noted that when $K_p = 1$ as shown in Fig. 2.3, the variation $\Delta x'_n$ is equal to the difference $\Delta y''_{n-1}$. By this counterplan, it may be possible that the control behavior is stabilized and that the output converges to some point near the extreme value, that is; the output has a smaller offset.

Another advantage is found in this scheme in connection with the fact that the approximation of the static characteristics by a parabolic curve $y = kx^2$ is more likely to be valid around the extreme value than far away from it. In the system with the limiter shown in Fig. 2.8, the input variation is proportional to the output variation only in the neighbourhood of the extreme point, and it saturates whenever the system is away off from the extreme point. Hence, it will be seen that the control operation is little affected by the fact that the parabolic approximation deviates from the real static performance characteristics at the position far away from the extreme value.

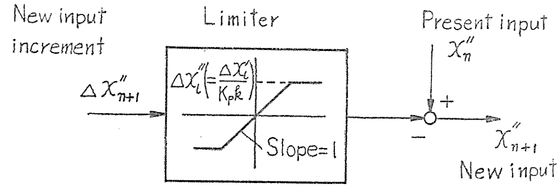


Fig. 2.8. Limiter to add into the optimalizing controller shown in Fig. 2.2, in order to obtain the stable control behavior and little offset.

(2) From the previously-mentioned fact that the offset is less than 0.5 as evaluated in the dimensionless input x', the next counterplan is to extend the initial condition to the unknown region, which may be an unstable region, on the plane of the initial conditions by choosing the maximum value x'_m of the dimension-

less input such that the allowable assigned error is above 0.5 in the dimensionless input. Then, the offset is less than the allowable assigned error, *i.e.*, the output has no apparent value of offset. At first sight, it seems to be possible that the maximum value of the normalized input, x_m, or the maximum value of the actual input, x''_m, is chosen freely. However, it is to be noted that free choice of x'_m does not correspond to that of x''_m or x_m but to that of K_p, because of the relation that $K_p k x_m = x'_m$. Of couse, for this case, the requirement (1) must be met in order that the control behavior is stable, even when starting with the initial conditions in the unstable region.

(3) When x'_m is chosen to be a small value near 1, and the control behavior starts with near $x'_0 = 0.5$ and smaller $\Delta y'_0$, the output has larger offset, that is, it stops at the position far away from the extreme value. Then, it will be necessary to apply another step change in the input in order that the output converges to the extreme value whenever $\Delta y'_n$ becomes less than a certain specified value. Althuogh such additional step change may cause an desired oscillation around the extreme value, but it is necessary for the counterplan to eliminate the offset.

The following problems for these counterplans should be resolved under the considerations of quickness of the response, described in the next section:

What value should be chosen as the saturation value $\Delta x'_l$ of the dimensionless input variation?

How much extention to unknown region should be permitted?

What would be the relation between the extention and the saturation value $\Delta x'_l$?

Whether is it necessary to apply additional step change in the input?

2. 3. *Quick response problem*

In this section, we shall consider how to choose the saturation value of $\Delta x'$ if the control is satisfactory both in the offset and the quickness of the response. Here, it is assumed that the deviation from the extreme value or the offset is allowed up to ε % of the maximum value of the dimensionless output, $y'_m (= x'^2_m)$. Then, we shall determine the relation between $\Delta x'_l$, the dimensionless saturation value, and N, the number of moves required for the output to converge to a point within the allowable error. It is seen that the value of the allowable error becomes larger with increase in y'_m, and the allowable error evaluated in the dimensionless input x' is $\sqrt{y'_m \varepsilon/100} = x'_m \sqrt{\varepsilon}/10$.

In order that the input converges to a point within the allowable error $x'_m \sqrt{\varepsilon}/10$ as quickly as possible, we must adopt $\Delta x'_l$ determined by the following procedure.

(1) The quickest control behavior is that in which the input enters within the region of allowable error without overshoot and all the steps from the initial point to the final are saturated in size.

There are two kinds of such a control behavior.

(a) The last computed step size of the input before it enters within the region of allowable error must be, for the first time, less than the saturation value.

(b) If the above mentioned input variation is equal to or greater than the saturated value, it may be possible that the input overshoots and goes out of the region. So the saturation value must be chosen such that this does not occur.

Now, suppose that the input enters first within the allowable error at the sampling instant nT_s. According to the case (a) where the input x'_n at nT_s is within it, we obtain $x'_n = x'_{n-1} - (x'^2_{n-2} - x'^2_{n-1}) \leqq x'_n \cdot \sqrt{\varepsilon}/10$. It is also assumed that the input, starting from the maximum value x'_m, reaches first the optimum value in N steps, when the saturated value $\Delta x'_l$ is used all the way: $\Delta x'_l = x'_m/N$. From this assumption, the above expression becomes $\Delta x'_l - [(2\,\Delta x'_l)^2 - (\Delta x'_l)^2] \leqq x'_m \cdot \sqrt{\varepsilon}/10$. Consequently,

$$\Delta x'_l \geqq \frac{1}{3}\left(1 - \frac{N\sqrt{\varepsilon}}{10}\right). \tag{2.4}$$

Also, in this case, by the requirement that $\Delta y'_{n-2}$ must be more than $\Delta x'_l$, we obtain $\Delta y'_{n-2} = x'^2_{n-3} - x'^2_{n-2} \geqq \Delta x'_l$.

Similarly, simplifying the above equation gives

$$\Delta x'_l \geqq 0.2. \tag{2.5}$$

According to the case (b), the first input variation after the input enters within the allowable error, must be more than $\Delta x'_l$, we obtain $\Delta y'_{n-1} \geqq \Delta x'_l$. Accordingly,

$$\Delta x'_l \geqq \frac{1}{3}. \tag{2.6}$$

It is clear that Eq. (2.6) is limited by Eq. (2.5). The value $\Delta x'^2_l$ is also required to be less than the allowable error, since the input x'_n at nT_s is zero, and x'_{n+1} at $(n+1)T_s$ is equal to $\Delta x'_l$. Hence, we obtain $\Delta x'^2_l \leqq x'_m \sqrt{\varepsilon}/10$. Consequently,

$$\Delta x'_l \leqq N\frac{\sqrt{\varepsilon}}{10}. \tag{2.7}$$

Then, the saturation value $\Delta x'_l$ satisfying Eqs. (2.4), (2.5) and (2.7) at the same time is to give the control behavior in which the output, starting from y'_m, converges to a point within the error of ε % of y'_m in N steps.

(2) When the input starts from an arbitrary initial condition, it is possible that the input has the offset, even if the saturation value meeting the above requirement (1) is used. As mentioned in the previous section, the offset is always less than 0.5 as evaluated in the dimensionless input x'. Accordingly, if $\Delta x'_l$ is so chosen that the allowable error is more than 0.5, *i.e.*, $x'_m\sqrt{\varepsilon}/10 \geqq 0.5$, then the output has apparently no offset. Hence,

$$\Delta x'_l \geqq 5/N\sqrt{\varepsilon}. \tag{2.8}$$

(3) The requirement described by Eq. (2.4) is to be satisfied only when the input starts from the maximum value. In order to satisfy the requirement (1), even when the input starts from any value x'_0, it is sufficient that $x'_0 \geqq \Delta x'_l$. Accordingly,

$$N \geqq \frac{x'_m}{x'_0} \tag{2.9}$$

(4) As x'_m becomes larger or $\Delta x'_l$ is chosen smaller, $\Delta x'_l$ becemes less than the allowable error assigned independently. In such a case, the controller does not

need to have the coefficient multiplier shown in Fig. 2.2, but may have only an on-off element. Thus, in order that the control mode is proportional-difference type, the equation that $\Delta x_l' \geqq x_m'\sqrt{\varepsilon}/10$ must be satisfied. Hence,

$$N\sqrt{\varepsilon} \leqq 10 \qquad (2.10)$$

The shaded region satisfying Eqs. (2.4), (2.5), (2.7), (2.8), (2.9) and (2.10), as shown in Fig. 2.9 (where x_0'/x_m' is chosen 0.1 in Eq. (2.9)), is that in which the offset is apparently zero whenever the input x_0' starts from the value less than one-tenth of x_m', and that the input does not overshoot the allowable error, and that the control operation is the proportional-difference type with the saturation value in the input variation. In this region, the quickest response time is obtained by the point A, where $\Delta x_l' = 0.706$ and $N = 10$. If we consider the region where all before-named Eq's except Eq. (2.8) are satisfyed, a quicker response time is obtained by the points on the line DA. If Eq. (2.9) is also omitted, a quicker response time is obtained by the points on the line BAE.

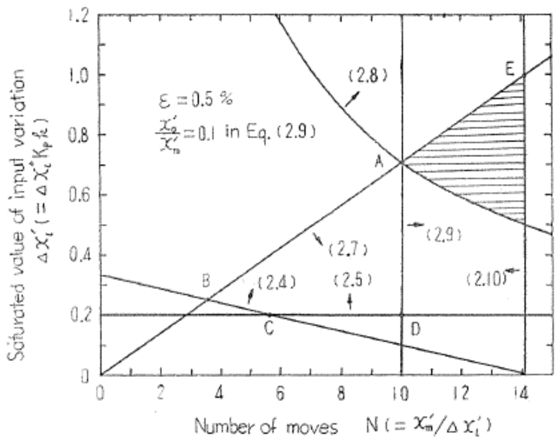


Fig. 2.9. Regions satisfying various kinds of design conditions.

3. Design principle of the control system

According to the analysis discussed above, K_p, the proportional sensitivity, and $\Delta x_l''$ $(= \Delta x_l)$, the saturation value of the actual input can be determined by the aid of the expressions that $x_m' = \Delta x_l' N = K_p k x_m$ and $\Delta x_l' = K_p k \Delta x_l''$ (see Fig. 2.3), when x_m, the maximum value of the normalized input, k, the characteristic constant, and $(\Delta x_l', N)$, a desirable point in Fig. 2.9 are specified.

Here we list a couple of remarks on how to select the system constants involved in this design method.

(1) Maximum value of the normalized input, x_m

x_m may be determined from the static performance characteristics of the controlled system. If the value x_n is chosen larger, the maximum value of the normalized output, y_m $(= kx_m^2)$, corresponding to x_n becomes larger, and then the number of steps, N, in which the output enters within the allowable error, becomes larger. Then the proportional sensitivity, K_p $(= x_m'/kx_m)$, becomes smaller, whereas Δx_l ($= \Delta x'_l/K_p k$) becomes larger. Accordingly, the offset may be larger, since the frequencies of having the normalized input variation greater than $\Delta x'_l$ get smaller.

(2) *Characteristic constant, k*

The static performance characteristics of the controlled system have many curves corresponding to the states of the disturbance, as shown in Fig. 2.1. According to this design principle, those curves must be represented by a single parabolic curve with a common characteristic constant k around each extreme value. Except the case where the forms of those curves around each extreme value, are similar to each other, what value should be chosen as the characteristic constant k?

Let us consider its effect for the case where the characteristic constant k is chosen to be either of the following extremities: (a) the least value k_1 and (b) the largest value k_2 of all those curves. In this case (b), on the curves with smaller k the apparent value of the allowable error becomes larger, and hence it appears as if the system had a quicker response, but the control behavior turns out to have offset, since the proportional sensitivity K_p becomes smaller. On the contrary, in the case (a), the apparent value of the allowable error becomes smaller, and then the response becomes slower on the curves with larger k. In general, the control behavior before the output reaches the extreme value suffers little change, but after that it may have a few overshoots. Occasionally, it happens that the output oscillates around the extreme value. This phenomenon results from the fact that the apparent value of K_p becomes larger. Supposing the case where the proportional sensitivity K_p and the saturation value Δx_l are designed based on the characteristic constant k_1 and the system is trying to search out the extreme value of the characteristic curve having the constant k_2, let us inspect the control behavior near the extreme value, and then determine the requirement under which the oscillation does not occur.

Soon after the output overshoots, for the first time, the extreme value, the normalized output variation is $k_2 \Delta x_l^2$, since the previous output is zero.

In order for the output not to oscillate, the following equation must then hold:

$$K_p k_2 \Delta x_l^2 < \Delta x_l. \tag{2.11}$$

By using the relation that $\Delta x_l = \Delta x'_l/K_p k_1$ in Eq. (2.11), we obtain

$$\frac{k_2}{k_1} < \frac{1}{\Delta x'_l}. \tag{2.12}$$

According to Eq. (2.12), the ratio k_2/k_1 of the characteristic constant is larger as $\Delta x'_l$ is chosen smaller. This means that there is less danger of the oscillation. For example, if $\Delta x'_l = 0.2$, then k_2/k_1 is equal to 5. In other words, even if the characteristic constant is five times as large as that used in design, the oscillation will not occur, although the response time may be longer.

(3) *Maximum value of the dimensionless input, x'_m ($= N\Delta x'_l$)*

As is clearly explained in the previous section, once x'_m is chosen more than 7, the apparent value of the offset becomes zero, though the response time is worse.

Accordingly, a little excess of x'_m may be allowed over 7. On the other hand, once x'_m is chosen less than 2 or between 2 and 7 giving priority to the quickness of response, there is a danger of more overshoots, especially when the input starts from smaller x'_0. If the input starts always with smaller output variation near small x_0, we shall have less overshoots, smaller offset, and quicker response time.

Part 2. On-off type

1. Principle of operation

In on-off control type, the size of input variation is constant irrespective of the output variation. This type is a special case of the proportional-difference type, and requires the largest number of moves of its type, as mentioned previously. Here, in order to improve the response time, we shall consider the on-off type with several different incremental step sizes: larger sizes are used when the output deviates largely from the extreme value and smaller sizes, when the output approches it.

2. Design principle of the control system

The problems as to what value should be used as the size of different steps and when they should be switched from one to the other will be considered for an on-off type with two sizes. It is assumed that the smaller size is a_1 and the larger one is a_2. First, in order that the error is within ε % of the output, a_1 must satisfy the following relation:

$$a_1 \leqq x_m \cdot \frac{\sqrt{\varepsilon}}{10}. \tag{2.13}$$

If a_1 is used within the range of γa_1 (γ: positive integer) from the optimum value ($x=0$), then a_2 should be equal to γa_1, because $a_2 < \gamma a_1$ causes slower response while $a_2 > \gamma a_1$, more overshoots. Then, the problems are reduced to choosing the best γ so that the response time is minimum.

The improvement of the response time may be evaluated by the decrease of the number of moves required for the output to converge to the extreme value. The number of moves, N, may be then given by

$$N \propto \frac{\gamma a_1}{a_1} + \frac{x_m}{a_2} = \gamma + \frac{x_m}{\gamma a_1}. \tag{2.14}$$

We may use the condition that $dN/d\gamma = 0$ in order to determine γ such that N is as small as possible. Then, Eq. (2.14) becomes

$$\gamma = \left(\frac{10}{\sqrt{\varepsilon}}\right)^{1/2} \tag{2.15}$$

by use of Eq. (2.13).

Further, it is required in this type of control to have the control condition by which the switching-over between a_1 and a_2 is performed. If, for safety's sake, the control condition is determined by the variation Δy for larger a_2, then we obtain

$$\frac{\Delta y}{a_2} \geqq \frac{k\{(2\gamma a_1)^2 - (\gamma a_1)^2\}}{\gamma a_1} = 3k\gamma a_1.$$

Consequently, the control conditions evaluated in terms of a_1 are as follows:

$$\begin{aligned} &\text{if } \Delta y \geqq 3k\gamma a_1^2, \quad \text{then } a_2 \text{ is used,} \\ &\text{if } \Delta y < 3k\gamma a_1^2, \quad \text{then } a_1 \text{ is used.} \end{aligned} \tag{2.16}$$

Example: If $\varepsilon=0.5\%$, then $a_1/x_m=0.07$ from Eq. (2.13), and $\gamma=3.7$ from Eq. (2.15). The control condition can be then determined by Eq. (2.16). If γ is approximately chosen 4, the number of moves N, by reference to Fig. 2.10, is 4 when the input starts from x_m, and the maximum number N_{max} is 6, which occures when the input starts from the point near m in Fig. 2.10. In both cases, it is assumed that the smaller size a_1 is chosen as the first step of the input for safety's sake and that the number of moves is counted till the output first reaches the region within the allowable error. When the number of moves is compared between this type with two sizes and the case of the point A in the proportional-difference type (see Fig. 2.9), the difference due to the initial input value is very small when x is less than $x_m/2$, but for the initial input from $x_m/2$ to x_m, the number of moves N of the one is smaller than that of the other. In this control type, however, it is a disadvantage that the controller becomes somewhat complicated, since the additional control device is required to switch the step size.

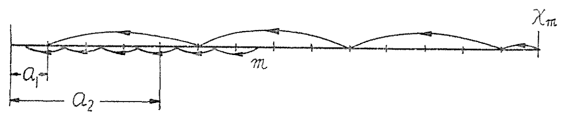


FIG. 2.10. Control behavior by the on-off type of the cut-and-try method, with $\varepsilon=0.5\%$.

Part 3. Proportional-gradient type

1. Principle of operation

The control operation by proportional-gradient type is as follows. After a test input δx of a constant size is applied to the controlled system, the resulting output variation δy is measured, and then the gradient at the state may be approximately calculated as $\delta y/\delta x$. The input variation Δx_n, which is applied from the controller to the controlled system, is chosen in proportion to the gradient; $\Delta x_n = K_p(\delta y/\delta x)$. Such a control operation is repeated at constant periods, making the output converge to the extreme value.

In this control type, it is of interest to note that in contrast to the proportional-difference type the sign of both Δx_n and Δy_n are not necessary to determine the direction of Δx_{n+1}, since the input δx is always applied in the same direction, and the resulting output variation is measured and thus $\delta y/\delta x$ can indicate a right direction of Δx_{n+1}. Because of no logic circuit, its controller becomes simpler as compared with the proportional-difference type. However, the output oscillates around the extreme value or has the hunting loss, since this operation is constantly repeated even after the output converges to the extreme value. The magnitude of the hunting loss is thought to be about $\delta x/2$ as evaluated in the input.

In a different proportional-gradient type, once the gradient at some state is determined, the input variation is applied in the same direction as determined then without checking the gradient at every sampling instant, until the output variation changes its sign. When the sign of the output variation changes, the gradient is measured afresh by injecting the test input δx, and then the above-mentioned control operation is continued. In such a type of control, the response time is about a half in comparison with the previously mentioned type of control, but the controller becomes more complicated. So, it will not be discussed further in this part.

2. Analysis of the control system

2.1. The first construction

As shown in Fig. 2.11, let us consider an analog computor circuit of the optimalizing control system with the proportional-gradient type. In this figure, r_1 is the contact of a relay operating once in every constant sampling period T_s. Here, T_s is assumed to be sufficiently long as compared with the time constants of the controlled system, so the dynamics effects can be neglected.

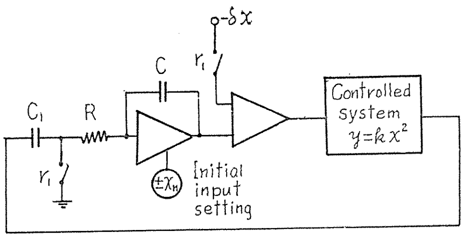


FIG. 2.11. Analog computor circuit diagram for a simple but incomplete optimalizing controller with the input increment proportional to an approximate gradient ($=\delta y/\delta x$) of the output, or with the proportional-gradient type of the cut-and-try method.

At some initial input value x_n, the test input δx is applied to the controlled system ($y=kx^2$) when r_1 is closed, and then the resulting output is charged into the condenser C_1. The voltage of C_1 is equal to $-k(-x_n+\delta x)^2$ calculated in term of the voltage in the input side of the integrator. At the following operation, once r_1 is opened, δx is removed and the input is returned to the initial value x_n. At the same time, the charge in C_1 is transfered in the condenser C of the integrator through the resistor R, and thus the output of the integrator is equal to x_n subtracted the value proportional to the difference between the output voltage of the controlled system and the voltage charged in C_1, the value approximately proportional to the gradient at x_n. Accordingly, if the decrement from x_n is denoted as $-\Delta x_n$, the following relation holds at steady state for the cases of the initial values $\mp x_n$. (Hereafter the signs follow in the same order).

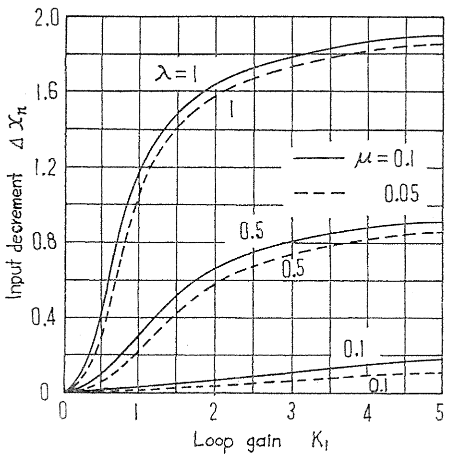
$$\Delta x_n = k\cdot\frac{C_1}{C}[\pm(\pm x_n+\delta x)^2 \mp (\mp x_n \mp \Delta x_n)^2]. \tag{2.17}$$

By using the dimensionless quantities, $\Delta x_n/x_m \equiv \Delta x_n$, $x_0/x_m \equiv \lambda$, $\delta x/x_m \equiv \mu$, $kx_m\frac{C_1}{C} \equiv K_1$, we can also write Eq. (2.17) as follows:

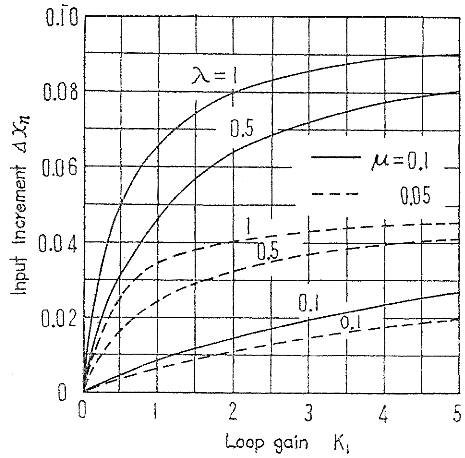
$$\Delta x_n = \frac{1}{2}\left[\left(2\lambda \mp \frac{1}{K_1}\right) \pm \sqrt{\left(2\lambda \mp \frac{1}{K_1}\right)^2 \pm 4(2\lambda\mu \pm \mu^2)}\right]. \tag{2.18}$$

Figures 2.12 (a) and (b) illustrate Eq. (2.18) for the cases of $\mp x_n$, respectively. Here it is to be noted that different units are being used for the ordinates of

Figs, 2.12 (a) and (b).



(a) at high-level side



(b) at low-level side

FIG. 2.12. Sizes of the input increment and decrement versus the loop gain in the cases where the input starts from a low-level side and a high-level side of the optimum point, respectively, in the system shown in Fig. 2.11.

2.2. The second construction

Let us next consider the second analog computor circuit shown in Fig. 2.13. In this figure, r_1, r_2 and r_3 are the contacts of the relays, operating in this order at a constant sampling period T_s. Although this circuit is more complicated than that of Fig. 2.10, the control operation is more complete and clear to be understood.

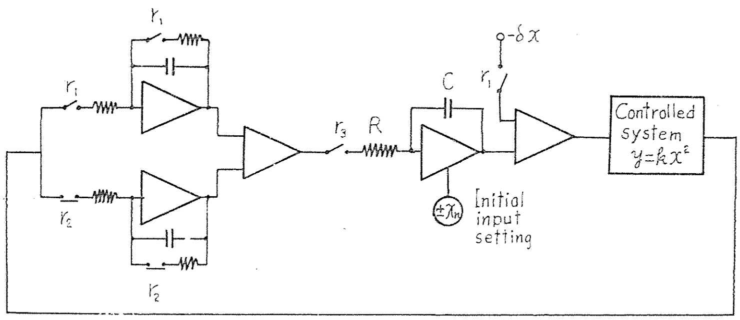


FIG. 2.13. Analog computor circuit diagram for a complicated but complete optimalizing controller as compared with that shown in Fig. 2.11.

Since the time constant of the integrator is represented by RC, the input variation Δx_n is given by

$$\Delta x_n = \frac{1}{RC}\int_0^{T_s} \Delta y_n dt, \tag{2.19}$$

where, for $\mp x$,

$$\Delta y_n = \pm k(\pm x_n + \delta x)^2 \mp kx_n^2. \tag{2.20}$$

By substituting Eq. (2.19) into Eq. (2.20) and using the previously mentioned dimensionless quantities, we obtain

$$\Delta x_n = K_2(2\lambda\mu \pm \mu^2), \tag{2.21}$$

for $\mp x$, where $K_2 = \dfrac{kx_mT_s}{RC}$.

3. Design principle of the control system

Two control circuits have been proposed and analized theoretically. From the result, it is seen that in the circuit of Fig. 2.11 the control behaviors are quite different according as the input starts from the positive side or the negative side, since in Fig. 2.12 (b), the input variation Δx_n can always be no more than δx, while in (a) it attains to several tens of δx. Accordingly, if K_1 is chosen suitably small such that the output does not overshoot in the latter case, then the output does not reach the extreme value in short time in the former case. Therefore, K_1 should be chosen larger so that the output reaches the extreme value as quickly as possible for the former case, then, for the latter case the size of Δx_n should be limited as in the proportional-difference type. Because of the simplicity of the construction, this circuit may be applicable to the controlled system, especially in which the input starts more often from either of the two sides.

The circuit of Fig. 2.13 compensates the unsymmetrical character of Fig. 2.11, but requires rather complicated programing and construction. When the input starts from larger λ as compared with μ, and has no overshoots, the number of moves, N, in the case of Fig. 2.13 is given by

$$\left.\begin{aligned} &\Delta x_0 = 2K_2\mu\lambda_0,\ \Delta x_1 = 2K_2\mu\lambda_1 = 2K_2\mu\lambda_0(1-2K_2\mu),\ \ldots \\ &\Delta x_n = 2K_2\mu\lambda_n = 2K_2\mu\lambda_0(1-2K_2\mu)^n, \end{aligned}\right\} \tag{2.22}$$

by neglecting the terms of μ in Eq. (2.21). From Eq. (2.22), the number of moves N before λ_n $(=x_n/x_m)$ reaches a point within $\mu/2$ starting from any λ_0 becomes

$$N = \log_{10}(\mu/2\lambda_0)/\log_{10}(1-2K_2\mu). \tag{2.23}$$

Figure 2.14 shows N as a function of loop gain K_2 for $\lambda_0 = 1.0$. In this figure, it is clear that the values of K_2 for $N<1$ are meaningless.

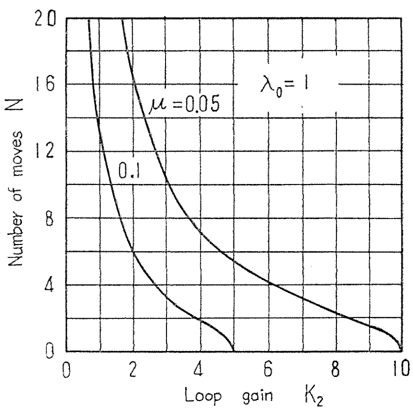


FIG. 2.14. Relation between the number of moves and the loop gain in the system shown in Fig. 2.13.

Chapter III. Effects of Dynamics of the Controlled System on the Cut-and-Try Method

When the cut-and-try method is adopted in an optimalizing system, the waiting interval or sampling period for the measurement of the resulting output after the introduction of a change in the input is usually assumed to be sufficiently long for the transient caused by the dynamics of the controlled system to die out. Thus, it takes very long time for the system to find an optimum state by this method, especially when the controlled system has large time constants. So, the sampling period should be chosen as short as possible for the output to reach an extreme value quickly.

How much shorter we can make the sampling period will be discussed under various kinds of design conditions in this chapter.

1. Various kinds of design conditions

When the dynamics of the controlled system is not negligible, several design conditions to be taken further into account are the following:

(1) When the input is at a desired value which corresponds to the extreme value of the output at steady state, the overshoot or undershoot of the output should be limited in size. Otherwise, the system will search largely back and forth through an extreme value, and have much hunting loss due to such an undesirable behavior.

(2) Initial conditions x_0, y_0 and $\dot{y}_0$ are to be taken into account in order to satisfy the condition (1), and if necessary, $\dot{y}_n$ should be limited.

(3) By the noise interference, the control action is not to be misled near an extreme value or on the way to the value.

(4) When the on-off type of the cut-and-try method is adopted, the hunting loss exists inevitably due to the oscillation about an extreme value. From the economical point of views, this value should be as small as possible.

(5) Other conditions: Once the reasonable size of the overshoot is determined, the incremental size of the input should be chosen large enough to cancel the overshoot and the initial velocity. Otherwise, unwanted subharmonic oscillations are inclined to take place in the system.

Let us determine the relation between the incremental size a of the input and the sampling period T_s to satisfy five design conditions mentioned above.

2. Effects of the dynamics of the controlled system

In this chapter, the effects of the dynamics of the controlled system are discussed only for the case of the on-off type of the cut-and-try method, in which the size of the incremental input is constant independently of the resulting variation of the output as shown in Fig. 3.8.

For simplicity of computation, it is assumed that the dynamics of the controlled system exists only in $G_0(s)$, the transfer function of the output linear group in the controlled system (see Fig. 1.1). The effect of $G_i(s)$, the transfer fnnction of the input linear group in the controlled system, will be discussed in a later section.

The input x changes incrementally towards a desired value of the input corresponding to an extreme value of the output. Since it generally requires a great deal of calculation to determine the output y for the incremental input x, the incremental input is approximated by a ramp input with a constant slope $S(=a/T_s)$, and the approximate solution will be compared with an exact one in the subsequent section. Now, in view of the condition (1) above, let us derive the relation which must hold between the parameters of the dynamics and the slope S so that the output $y(=y_s)$ does not exceed the assigned allowable overshoot or undershoot at $t=x_0/S$, the moment when the input reaches the desired value, which is, under the normalized coordinate (x, y), equal to zero.

If x, starting with negative initial value $-x_0$, goes to zero giving an extreme value with a positive slope S, an equation relating between x and y^* is

$$y^* = k(St - x_0)^2. \tag{3.1}$$

$G_0(s)$ is assumed to take either of the two forms:

$$G_0(s) = \frac{\omega_n^2 e^{-Ls}}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \tag{3.2 a}$$

$$= \frac{e^{-Ls}}{(\tau_{01}s+1)(\tau_{02}s+1)}. \tag{3.2 b}$$

If the initial conditions of $y(t)$ are specified as y_0 and $\dot{y}_0$, Laplace transform $Y(s)$ of $y(t)$ for Eq. (3.2 a) becomes

$$\begin{aligned} Y(s) = {} & \frac{\omega_n^2 e^{-Ls}}{s^2 + 2\zeta\omega_n s + \omega_n^2}\left(\frac{1}{s}y_0^* - 2\,Skx_0\frac{1}{s^2} + kS^2\frac{1}{s^3}\right) \\ & + \frac{(s + 2\zeta\omega_n)e^{-Ls}}{S^2 + 2\zeta\omega_n s + \omega_n^2}y_0 + \frac{e^{-Ls}}{s^2 + 2\zeta\omega_n s + \omega_n^2}\dot{y}_0 . \end{aligned} \tag{3.3}$$

Then, the inverse Laplace transformation of Eq. (3.3) yields

$$y(t) = kS^2 \cdot A(t) - 2\,Skx_0 \cdot B(t) + \{1 - C(t)\}y_0^* + C(t) \cdot y_0 + D(t) \cdot \dot{y}_0, \tag{3.4}$$

where

$$A(t) = \frac{-2}{\omega_n^2\sqrt{1-\zeta^2}}e^{-\zeta\omega_n(t-L)}\sin(\sqrt{1-\zeta^2}\omega_n(t-L)+\varphi_1)+\frac{2}{\omega_n^2}(4\zeta^2-1)-\frac{4\zeta(t-L)}{\omega_n}+(t-L)^2,$$

$$B(t) = \frac{e^{-\zeta\omega_n(t-L)}}{\omega_n\sqrt{1-\zeta^2}}\sin(\sqrt{1-\zeta^2}\omega_n(t-L)+\varphi_2)+(t-L)-\frac{2\zeta}{\omega_n},$$

$$C(t) = \frac{e^{-\zeta\omega_n(t-L)}}{\sqrt{1-\zeta^2}}\sin(\sqrt{1-\zeta^2}\omega_n(t-L)+\varphi_3),$$

$$D(t) = \frac{1}{\omega_n\sqrt{1-\zeta^2}}e^{-\zeta\omega_n(t-L)}\sin(\sqrt{1-\zeta^2}\,\omega_n(t-L)),$$

and

$$\varphi_1 = \tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}\frac{4\zeta^2-1}{4\zeta^2-3}, \qquad \varphi_2 = \tan^{-1}\frac{2\zeta\sqrt{1-\zeta^2}}{2\zeta^2-1}, \qquad \varphi_3 = \tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}.$$

Before proceeding further it is convenient in the present problem to introduce the following dimensionless quantities:

$$\lambda = x_0/x_m, \qquad \mu = a/x_m, \qquad \beta = \omega_n(x_0/S - L).$$

By using these dimensionless quantities, from Eq. (3.4) we have

$$\frac{y_s}{\lambda^2 y_m^*} = \frac{1}{\beta^2}\cdot A'(\beta) - \frac{2}{\beta}\cdot B'(\beta) + \{1 + C'(\beta)\} + C'(\beta)\frac{y_0}{\lambda^2 y_m^*} + \frac{1}{\lambda^2}\frac{\dot{y}_0}{kx_m S}\frac{\mu}{\omega_n T_s}\cdot D'(\beta), \tag{3.5}$$

where

$$A'(\beta)=\frac{-2}{\sqrt{1-\zeta^2}}e^{-\zeta\beta}\sin(\sqrt{1-\zeta^2}\beta+\varphi_1)+2(4\zeta^2-1)-4\zeta\beta+\beta^2,$$

$$B'(\beta)=\frac{e^{-\zeta\beta}}{\sqrt{1-\zeta^2}}\sin(\sqrt{1-\zeta^2}\beta+\varphi_2)+\beta-2,$$

$$C'(\beta)=\frac{e^{-\zeta\beta}}{\sqrt{1-\zeta^2}}\sin(\sqrt{1-\zeta^2}\beta+\varphi_3),$$

$$D'(\beta)=\frac{e^{-\zeta\beta}}{\sqrt{1-\zeta^2}}\sin\sqrt{1-\zeta^2}\beta.$$

Similarly, for Eq. (3.2 b),

$$A'(\beta')=\frac{\nu^2}{(1+\nu)^2}\Big\{\frac{2\nu}{\nu-1}e^{-\frac{(\nu+1)}{\nu}\beta'}+\frac{2}{\nu^2(\nu-1)}e^{-(\nu+1)\beta'}+\frac{2(\nu^2+\nu+1)}{\nu^2}$$
$$-\frac{2(\nu+1)^2}{\nu^2}\beta'+\frac{(\nu+1)^2}{\nu^2}\beta'^2\Big\},$$

$$B'(\beta')=\frac{\nu^2}{\nu^2-1}\Big\{e^{-\frac{(1+\nu)}{\nu}\beta'}-\frac{1}{\nu^2}e^{-(1+\nu)\beta'}\Big\}-1+\beta',$$

$$C'(\beta')=\frac{\nu}{\nu-1}\Big\{e^{-\frac{(1+\nu)}{\nu}\beta'}-\frac{1}{\nu}e^{-(1+\nu)\beta'}\Big\},$$

$$D'(\beta')=\frac{\nu}{\nu^2-1}\Big\{e^{-\frac{(1+\nu)}{\nu}\beta'}-e^{-(1+\nu)\beta'}\Big\},$$

where

$$\nu=\tau_{02}/\tau_{01},\qquad \beta'=(x_0/S-L)/T_{\text{sum}},\qquad T_{\text{sum}}=\tau_{01}+\tau_{02}.$$

Thus, from Eq. (3.5), β or β' can be calculated as a function of ζ or ν respectively with $y_s/\lambda^2 y_m^*$ as parameter. When the initial conditions are $y_0=y_0^*$ and $\dot{y}_0=0$, then β or β' is shown in Fig. 3.1 (a) or (b) respectively, where $y_s/\lambda^2 y_m^*$ is specified as 0.5, 1 or 5%.

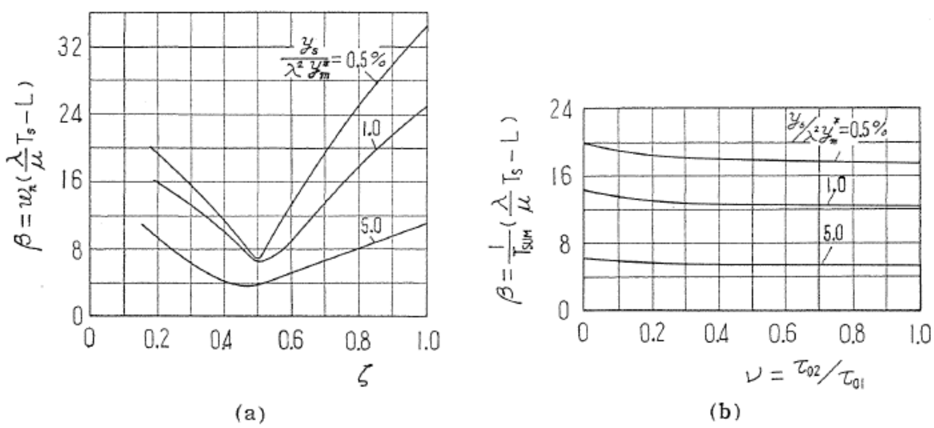


FIG. 3.1. β and β' for different values of ζ and ν, respectively, with $y_s/\lambda^2 y^*=0.5$, 1.0 and 5.0% as parameters.

3. Effect of initial conditions

When the dynamics of the controlled system is negligible, only x_0 and Δy_0 have been considered as the initial conditions, but further conditions y_0 and $\dot{y}_0$ must be taken into account if the dynamics of the controlled system is to be approximated by a 2nd order lag system.

3.1) Effect of initial position y_0

According to the fourth term in Eq. (3.5), for $\zeta<1$, the error E_1 in y_s/y_m due to the initial position is less than the following value:

$$\frac{E_1}{\frac{y_0^* - y_0}{y_0^*}} = \frac{\lambda^2}{\sqrt{1-\zeta^2}}\, e^{-\zeta\beta}. \qquad (3.6)$$

Figure 3.2 gives E_1 in terms of ζ with λ as parameter. It is to be noted that E_1 has a peak near $\zeta=0.5$. On the other hand, E_1 is negligible for $\zeta \geqq 1$, because β' for assigned value of y_s/y_m^* is sufficiently large. If the error due to the initial velocity is taken into design conditions as given in the following part, then E_1 may be negligible even for $\zeta<1$.

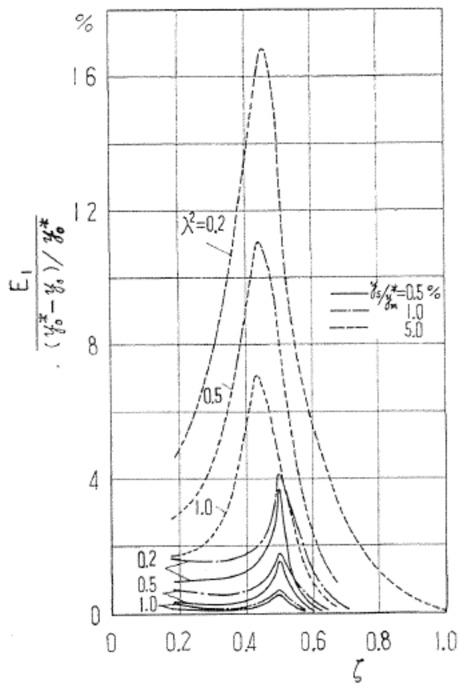


FIG. 3.2. Error E_1 in y_s/y_m^* due to the initial position y_0.

3.2) Effect of initial velocity $\dot{y}_0$

The effect of the initial velocity $\dot{y}_0$ is evaluated from the fifth term in Eq. (3.5). For $\zeta \leqq 1$, the error E_2 in y_s/y_m^* due to the initial velocity is less than the following value:

$$\frac{E_2}{\dot{y}_0/kx_mS} = \frac{\lambda}{\beta}\frac{e^{-\zeta\beta}}{\sqrt{1-\zeta^2}} \qquad (\zeta<1),$$
$$= \frac{\lambda}{\beta}\,\beta e^{-\beta} \qquad (\zeta=1). \qquad (3.7)$$

Similarly to the case of the initial position y_0, the error E_2 for $\zeta>1$ may be neglected for the value of β' in Fig. 3.1 (b). As is clearly seen in Fig. 3.3, which is obtained from Eq. (3.7), there is a peak near $\zeta=0.5$. Figure 3.4 gives $\omega_n(T_s/\mu - L)$ in terms of ζ and λ for the cases where the effect of $\dot{y}_0$ on y_s/y_m^* is equal to 0.5, 1 and 5% respectively under the condition that $\dot{y}_0$ equals kx_mS. It is to be noted that if this condition has to be met even for smaller λ, it imposes a more severe restriction on the system design than that discussed in Sec. 2, especially when ζ is also small (see Fig. 3.1 (a)). If T_s/μ in Figure 3.4 is taken as the designed value, it is necessary that the control action is stopped whenever $\dot{y}_0$ exceeds kx_mS.

4. Effect of noise

The control action is misled by the interference of the noise existing in the system or entering from the outside of the system in the sign of the input rather

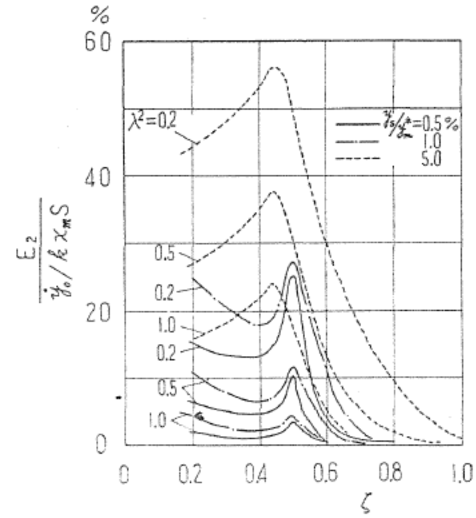


FIG. 3.3. Error E_2 in y_s/y_m^* due to the initial velocity $\dot{y}_0$.

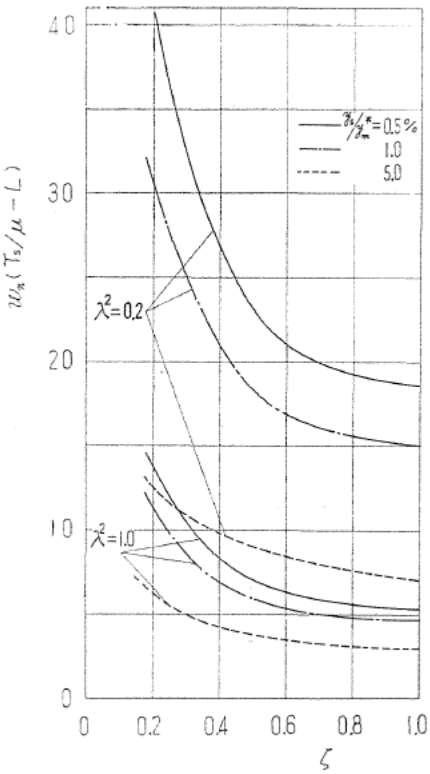


FIG. 3.4. Relation between $\omega_n\left(\frac{T_s}{\mu}-L\right)$ and ζ obtained from Fig. 3.3, for the cases where the effect of $\dot{y}_0$ on y_s/y_m^* is equal to 0.5, 1.0 and 5.0% of y_s/y_m^* under the condition that $\dot{y}_0=kx_mS$.

than its size, so that the performance of the system may be unstabilized. As the variation of the output to an input change gets smaller, the control action may make mistakes by the noise.

The smallest variation of the output takes place in the oscillation around an extreme value. So, at first, let us determine the magnitude of the smallest one. Figure 3.5 shows a typical oscillating behavior around an extreme value. According to the condition of the oscillation, the following relations are to be satisfied among the initial conditions at the points 0, 1, 2 and 3:

$$\left.\begin{aligned} y_0(T_s) &= y_1(0) = y_2(T_s) = y_3(0),\\ y_1(T_s) &= y_2(0) = y_3(T_s) = y_0(0),\\ \dot{y}_0(T_s) &= \dot{y}_1(0) = \dot{y}_2(T_s) = \dot{y}_3(0),\\ \dot{y}_1(T_s) &= \dot{y}_2(0) - \dot{y}_3(T_s) = \dot{y}_0(0). \end{aligned}\right\} \quad (3.8)$$

Also, at the intervals 0-1 and 0-2,

$$\left.\begin{aligned} y_{01}(t) &= \{1-C(t)\}\frac{ka^2}{2} + D(t)\dot{y}_0(0) + C(t)y_0(0),\\ y_{12}(t) &= -\{1-C(t)\}\frac{ka^2}{2} + D(t)\dot{y}_1(0) + C(t)y_1(0). \end{aligned}\right\} \quad (3.9)$$

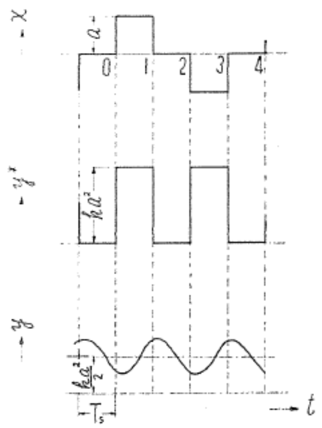


FIG. 3.5. Typical control behavior around an extreme value.

From Eqs. (3.8) and (3.9), the variation $\Delta y_{\min}=y_0(T_s)-y_1(T_s)$ is determined as follows:

$$\Delta y_{\min}=ka^2\left\{\frac{2}{1+C(T_s)-\dot{C}(T_s)\dfrac{D(T_s)}{1+\dot{D}(T_s)}}-1\right\}, \tag{3.10}$$

where $\dot{C}(T_s)=\left[\dfrac{dC(t)}{dt}\right]_{t=T_s}$, $\dot{D}(T_s)=\left[\dfrac{dD(t)}{dt}\right]_{t=T_s}$:

for Eq. (3.2 a),

$$\dot{C}(T_s)=\frac{\omega_n e^{-\zeta\omega_n(T_s-L)}}{\sqrt{1-\zeta^2}}\sin\{\sqrt{1-\zeta^2}\,\omega_n(T_s-L)\},$$

$$\dot{D}(T_s)=-\frac{e^{-\zeta\omega_n(T_s-L)}}{\sqrt{1-\zeta^2}}\sin\{\sqrt{1-\zeta^2}\,\omega_n(T_s-L-\varphi_3)\},$$

for Eq. (3.2 b),

$$\dot{C}(T_s)=\frac{1+\nu}{\nu-1}\cdot\frac{1}{T_{\text{sum}}}\{e^{-(1+\nu)((T_s-L)/T_{\text{sum}})}-e^{-\frac{(1+\nu)}{\nu}\cdot\frac{(T_s-L)}{T\text{sum}}}\},$$

$$\dot{D}(T_s)=\frac{\nu}{\nu-1}\left\{e^{-(1+\nu)\cdot\frac{(T_s-L)}{T\text{sum}}}-\frac{1}{\nu}e^{-\frac{(1+\nu)}{\nu}\cdot\frac{(T_s-L)}{T\text{sum}}}\right\}.$$

Figure 3.6 (a) or (b) gives $\Delta y_{\min}$ in terms of $\omega_n T_s$ or T_s/T_{sum} with ζ or ν as parameter, respectively.

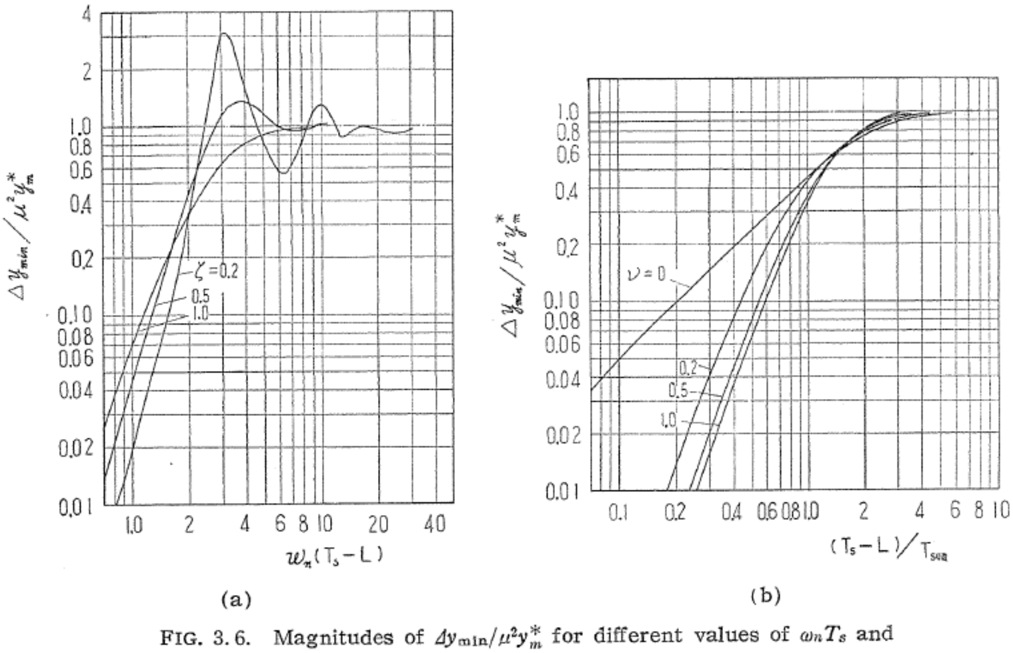


FIG. 3.6. Magnitudes of $\Delta y_{\min}/\mu^2 y_m^*$ for different values of $\omega_n T_s$ and T_s/T_{sum} with ζ and ν as parameters, respectively.

Let us next consider the effect of the noise on the control action. The probability with which the sign of $\Delta y(t)$, defined by $y(t)-y(t-T_s)$, is misled by the noise $n(t)$ is equivalent to the probability of making $z=\Delta y(t)\{\Delta y(t)+n(t)-n(t-T_s)\}$ negative. When the minimum value $\Delta y_{\min}$ around an extreme value is taken as Δy in the above expression, an average value $\bar{z}$ and variance σ_z^2 of a statistical variable z are easily determined as follows:

$$\bar{z}=\Delta y_{\min}^2, \tag{3.11}$$

$$\sigma_z^2=\overline{(z-\Delta y_{\min}^2)^2}=2\,\Delta y_{\min}^2\{\phi(0)-\phi(T_s)\}, \tag{3.12}$$

where $\phi(t)$ is an autocorrelation function of $n(t)$ and an average value $\overline{n(t)}$ of $n(t)$ is assumed to be zero.
From Eqs. (3.11) and (3.12), we can deduce the following relation:

$$\frac{\Delta y_{\min}}{y_m^*}=\sqrt{2(1-e^{-T_s/\tau})}\cdot\left(\frac{\bar{z}}{\sigma_z}\right)\cdot\frac{\sqrt{\phi(0)}}{y_m^*}, \tag{3.13}$$

where $\phi(t)=\phi(0)e^{-|t|/\tau}$ in Eq. (3.12).

Provided that we fix the quantities τ and $\sqrt{\phi(0)}$ representing the statistical property, an allowable probable error, and then $\bar{z}/\sigma_z$ assuming the probability distribution of z to be Gaussian, we can calculate the minimum allowable value of $\Delta y_{\min}/y_m^*$ from Eq. (3.13). Figure 3.7 shows the dimensionless noise level $\sqrt{\phi(0)}/y_m^*$ versus $\Delta y_{\min}/y_m^*$ with T_s/τ as parameter when an allowable probable error is taken as 0.5, 1 or 5%. By using Figs. 3.6 (a), (b) and 3.7, we can determine the minimum value of $\omega_n(T_s-L)$ and $(T_s-L)/T_{\text{sum}}$ for μ under the interference of the noise, respectively.

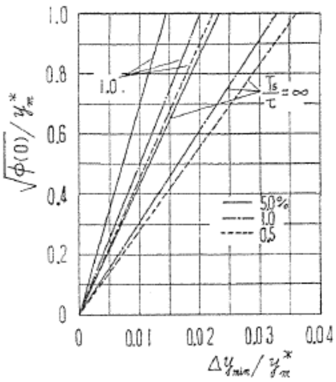


FIG. 3.7. Diagram for determining a limited value of $\Delta y_{\min}/y_m^*$.

5. Magnitude of hunting loss

When the on-off type is adopted, an oscillation about an extreme value inevitably takes place. The magnitude of hunting loss is dependent on only $G_i(s)$, regardless of $G_0(s)$. If $G_i(s)$ is assumed to be negligible as mentioned above, the hunting loss becomes simply $ka^2/2$. The magnitude of hunting loss for the case where $G_i(s)$ is not to be neglected is presented in Sec. 2 of Chap. IV.

6. Comparison between an approximate solution and the exact solution

Since the solution calculated hitherto in this chapter is an approximate one with replacement of an incremental input by a ramp input for ease of calculation, it is necessary to compare an approximate solution with the exact one.

Although the exact solution is complicated and troublesome in the numerical calculation, it will be determined as follows.
The expressions corresponding to Eqs. (3.3) and (3.5) are

$$Y(s)=\sum_{p=0}^{\infty}k(-x_0+pa)^2\cdot\frac{1-e^{-sT_s}}{s}\cdot\frac{\omega_n^2e^{-Ls}}{s^3+2\zeta\omega_ns+\omega_n^2}$$
$$+\frac{(s+2\zeta\omega_n)e^{-Ls}}{s^2+2\zeta\omega_ns+\omega_n^2}y_0+\frac{e^{-Ls}}{s^2+2\zeta\omega_ns+\omega_n^2}\dot{y}_0, \quad (3.3)'$$

$$\frac{y(nT_s)}{\lambda^2y_m^*}=-C(nT_s-L)+\frac{1}{(n-1)^2}+\frac{2n-1}{(n-1)^2}\sum_{\nu=1}^{n}C(nT_s-L-pT_s)$$
$$-\frac{2}{(n-1)^2}\sum_{\nu=1}^{n}pC(nT_s-L-pT_s)+C(nT_s-L)\frac{y_0}{\lambda^2y_m^*}$$
$$+\frac{1}{\lambda^2}\cdot\frac{\dot{y}_0}{kx_m(a/T_s)}\cdot\frac{\mu}{\omega_nT_s}D'(nT_s-L), \quad (3.5)'$$

where $x_0/a=n-1$.
For $G_0(s)$ of Eq. (3.2 a), the calculation of Eq. (3.5)′ yields

$$\frac{y(nT_s)}{\lambda^2y_m^*}=-C(nT_s-L)+\frac{1}{(n-1)^2}+C(nT_s-L)\frac{y_0}{\lambda^2y_m^*}+\frac{1}{\lambda^2}\cdot\frac{\dot{y}_0}{kx_m(a/T_s)}\cdot\frac{\mu}{\omega_nT_s}D'(nT_s-L)$$
$$+\frac{2n-1}{(n-1)^2}\cdot\frac{K}{(1-2w\cos A+w^2)}[w\sin(A+B)-w^{n+1}\sin\{(n+1)A+B\}$$
$$+w^{n+2}\sin(nA+B)-w^2\sin B]-\frac{2}{(n-1)^2}\cdot\frac{K}{(1-2w\cos A+w^2)^2}$$
$$[w\sin(A+B)-w^3\sin(A-B)-(n+1)w^{n+1}\sin\{(n+1)A+B\}+2(n+1)w^{n+2}$$
$$\times\sin(nA+B)+nw^{n+2}\sin\{(n+2)A+B\}-(n+1)w^{n+3}\sin(n-1)A\cos B$$
$$+(n-1)w^{n+3}\cos(n-1)A\cdot\sin B-2nw^{n+3}\sin(n+1)A\cos B$$
$$+nw^{n+4}\sin(nA+B)-2w^2\sin B], \quad (3.5\,a)'$$

where $K=e^{-\zeta\omega_n(nT_s-L)}/\sqrt{1-\zeta^2}$, $\quad w=e^{-\zeta\omega_n(T_s-L)}$,
$A=-\sqrt{1-\zeta^2}\,\omega_n(T_s-L)$, $\quad B=\sqrt{1-\zeta^2}\,\omega_n(nT_s-L)+\varphi_3$.

Similarly, for Eq. (3.2 b),

$$\frac{y(nT_s)}{\lambda^2y_m^*}=-C(nT_s-L)+\frac{1}{(n-1)^2}+C(nT_s-L)\cdot\frac{y_0}{\lambda^2y_m^*}+\frac{1}{\lambda^2}\cdot\frac{\dot{y}_0}{kx_m(a/T_s)}$$
$$\times\frac{T_{\text{sum}}\cdot\mu}{T_s}\cdot D'(nT_s-L)+\frac{2n-1}{(n-1)^2}\left[\frac{K_1v(1-v^n)}{1-v}-\frac{K_2u(1-u^n)}{1-u}\right]$$
$$-\frac{2}{(n-1)}\left[K_1\left\{\frac{w(1-v^n)}{(1-v)^2}-\frac{nv^{n+1}}{1-v}\right\}-K_2\left\{\frac{u(1-u^n)}{(1-u)^2}-\frac{nu^{n+1}}{1-u}\right\}\right]. \quad (3.5\,b)'$$

where

$$K_1=\frac{\nu}{\nu-1}e^{-\frac{(1+\nu)}{\nu}-\frac{(nT_s-L)}{T_{\text{sum}}}},\qquad K_2=\frac{1}{\nu-1}e^{-(1+\nu)\frac{(nT_s-L)}{T_{\text{sum}}}},$$
$$v=e^{\frac{(1+\nu)}{\nu}\cdot\frac{(T_s-L)}{T_{\text{sum}}}},\qquad u=e^{(1+\nu)\cdot\frac{T_s-L}{T_{\text{sum}}}}.$$

Table 3.1 shows several examples of comparison between an approximate solution

TABLE 3.1

μ	y_s/y_m^* →		$\zeta=1.0$: 0.5	$\zeta=1.0$: 1.0	$\zeta=1.0$: 5.0	$\zeta=0.5$: 1.0	$\nu=0$: 1.0
0.1	$\left(\frac{y_s}{y_m^*}\right)_e$	($\lambda=1$)	0.15	0.44		0.90	0.50
		($\lambda=.2$)	0.05	0.40	1/3 Subharmonic oscillation	0.25	0.43
	$\Delta y_{min}/y_m^*$		0.75	0.50		0.40	0.60
	$\dot{y}_m/kx_mS$		67	104		100	—
	$\omega_n T_s$		3.5	2.5	1.1	1.8	(T_s/T_{sum}) 1.4
0.2	$\left(\frac{y_s}{y_m^*}\right)_e$	($\lambda=1$)	0.03	0.17	2.5	0.40	0.29
		($\lambda=.2$)	0.03	0.16	1.4	0.60	0.24
	$\Delta y_{min}/y_m^*$		4.0	3.6	1.6	3.8	4.0
	$\dot{y}_m/kx_mS$		10	31	120	3.0	—
	$\omega_n T_s$		7.0	5.0	2.2	3.6	(T_s/T_{sum}) 2.8

and the exact one for $\zeta=1.0$, $\zeta=0.5$ and $\nu=0$ under the initial conditions $y_0=y_0^*$ and $y_0=0$, when $L=0$. Figure 3.8 gives a few typical control behaviors. It is to be noted that an unwanted oscillation is occuring about an extreme value in Fig. 3.8 (b), and that the control action is stopping when $\dot{y}_n > kx_mS$ in Fig. 3.8 (c).

General remarks from these comparisons are given as follows:

(1) The magnitudes of the overshoot or undershoot according to the exact solution are smaller than those according to an approximate one. This trend becomes more obvious with increase of the incremental size μ, and justifies the design based on the approximate solution.

(2) Not being able to cancel the overshoot, smaller μ is apt to cause an unwanted subharmonic oscillation, when a large overshoot is allowed. Therefore, an allowable overshoot or undershoot is to be determined under consideration of the size of μ.

(3) The maximum value $\dot{y}_m$ of $\dot{y}$ on the way to an extreme value becomes smaller with decrease of μ. This results from the increase in the apparent value of $\omega_n T_s$.

7. Effect of $G_i(s)$

In this section the effect of $G_i(s)$ on the design of the system will be considered.

Laplace transform $X(s)$ of $x(t)$ for an incremental input $x^*(t)$ is

$$X(s)=G_i(s)\cdot\frac{1-e^{-sT_s}}{s}\sum_{p=0}^{\infty}(-x_0^*+pa)e^{-pT_s\cdot s}. \tag{3.14}$$

If $G_i(s)$ is assumed to be approximately represented by a first order lag system $1/(1+\tau_i s)$, from Eq. (3.14) with the initial condition $x_0^*=x_0$, the value $x(nT_s)$ of $x(t)$ at the instant nT_s is

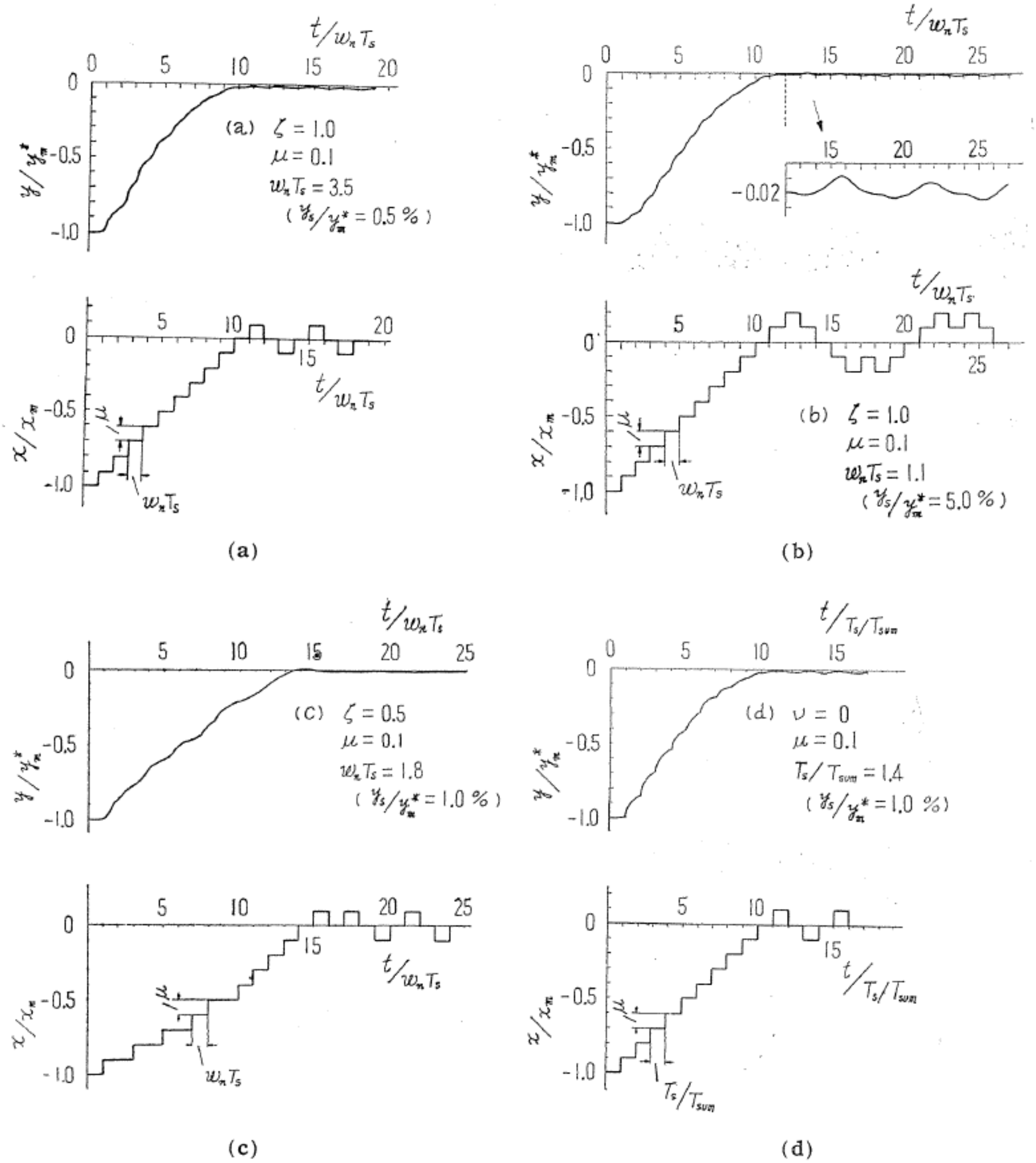


FIG. 3. 8. Actual examples of control behavior with $L=0$.

$$\frac{x(nT_s)}{x_m^* \cdot \lambda} = \frac{\mu}{\lambda} \cdot \frac{1}{1-e^{T_s/\tau_i}} [1 - e^{-\frac{\lambda T_s}{\mu \tau_i}}], \tag{3.15}$$

where $\frac{x_0^*}{a} = n - 1 = \frac{\lambda}{\mu}$, $\lambda = \frac{x_0^*}{x_m^*}$ and $\mu = \frac{a}{x_m^*}$.

Figure 3.9 shows the relation between τ_i/T_s and the deviation of $\frac{x(nT_s)}{x_m^* \lambda}$ from the desired value $x(nT_s) = 0$ for different values of λ/μ. For instance, if $\lambda = 1$ and

$\mu=0.1$, the undershoot in $x(nT_s)$ reaches 10% of the maximum value x_m^* of x^* at $\tau_i/T=1.44$. Then, because $x(t)$ is squared at nonlinear element, which is assumed a parabolic form, the size of $y^*(t)$ becomes 1% of y_m^*. So, if an allowable undershoot is 1% of y_m^*, then τ_i/T must be chosen, under consideration of the additional effect of the time lag element $G_0(s)$ which follows, so as to make τ_i/T less than 1.44.

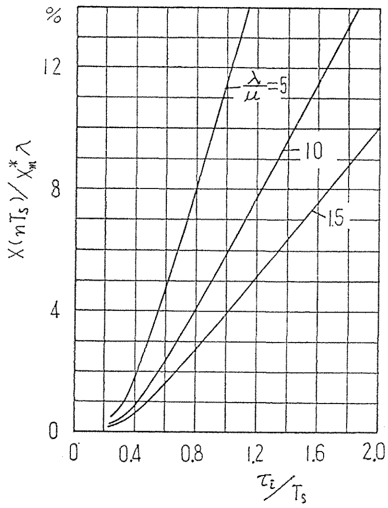


FIG. 3.9. $x(nT_s)/x_m^*\lambda$ as a function of τ_i/T_s with λ/μ as parameter.

8. Example of design

According to the design conditions discussed above, when the on-off type of the cut-and-try method is adopted, a size μ of incremental input and a sampling period T_s can be determined. And, when the proportional type is adopted, only a sampling period T_s can be determined after the determination of a size μ, since the size μ is restrained by the conditions described in Chap. II in this case.

An example of design is given to explain the design principle for determing both T_s and μ so as to satisfy all design conditions. For instance, let us determine both T_s and μ, under the design conditions that ζ is 0.5, L is 0, the hunting loss is below 1%, the allowable overshoot is below 0.5% and the effect of the noise is 0.01 as evaluated in terms of $\Delta y_{\min}/y_m^*$. The shaded region in Fig. 3.10 satisfies all design conditions, and the number bracketted corresponds to the number of the design conditions described above, and the arrows show the favorable direction. In the design where μ can be taken as large as possible, point 2 in the desired region is supposed to have the quickest response time.

Intersecting point 1 of the straight line (4) and the chain line in Fig. 3.10 offers a dimensionless time $\omega_n T_s$ which is required for an indicial response to reach 99% of the steady state value with the size μ corresponding to point 1. If point 1 is assumed as a design point in a conventional technique, it is possible for the response time to be speeded up by the factors of approximately 3 in the present design method.

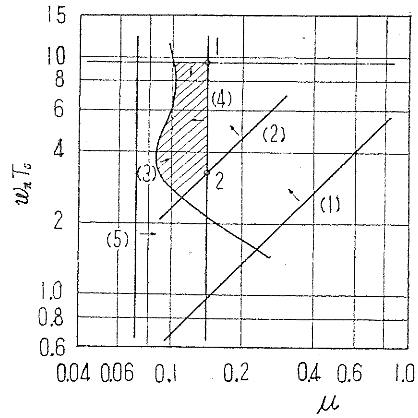


FIG. 3.10. Example of design satisfying various kinds of design conditions that $y_s/y_m^*=0.5\%$, $\Delta y_{\min}/y_m^*=0.01$ and $H/y_m^*=1\%$, with $\zeta=0.5$ and $L=0$.

Chapter IV. Peak-Holding Method

The peak-holding method is characterized by its quicker response and by its simpler construction of the controller than other methods.

1. Principle of operation

A typical control behavior by the peak-holding method, when the dynamic effects of the controlled system are neglected, and the functional diagram of the optimalizing controller for this method are shown in Figs. 4.1 and 4.2, respectively. In Fig. 4.1, suppose the input x'' is below the desired value x''_{max}, which corresponds to the maximum value y_{max} of the output y''. The input driving device is then set to increase the input at a constant rate. At the time instant 1 (Fig. 4.1) the input passes the desired value x''_{max}. The output y'' is thus the maximum value y''_{max} at the time instant 1 and decreases after the instant 1. Now if the peak-holding circuit, shown in Fig. 4.2, is so designed as to follow the output exactly when the output is increasing, but hold to the maximum value after the maximum is passed and the output starts to decrease; then there will be a difference $\Delta y''$ between the output y''_p of this peak-holding circuit and the output y'' itself after the time instants 1. This difference is shown in Fig. 4.1 (d). When this difference amounts to an assigned allowable value, denoted by c in Fig. 4.1 (d), at the time instant 2, the direction of the input drive is reversed, while keeping

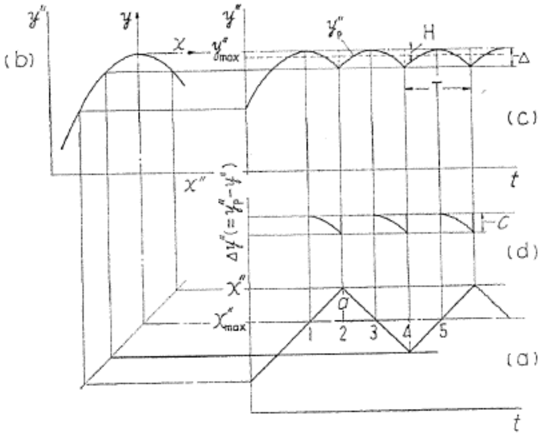


FIG. 4.1. Typical control behavior by the peak-holding method.

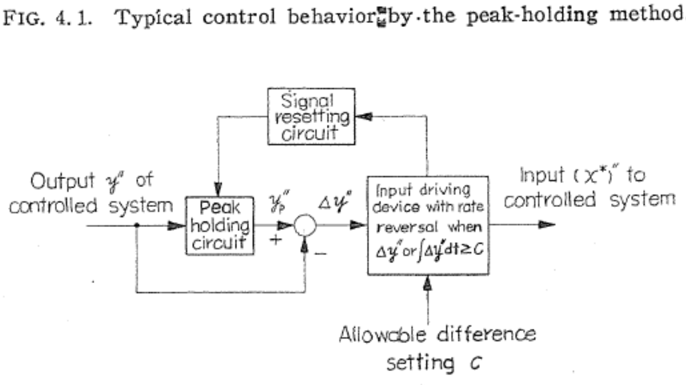


FIG. 4.2. Functional diagram of the optimalizing controller for the peak-holding method.

the driving speed the same as before. At the same time, the signal resetting circuit shown in Fig. 4.2 is triggered and then the output y_p'' of the peak-holding circuit is reset to the present output of the controlled system.

Here, instead of the difference $\Delta y''$, we may use the time integral of the difference, $\int \Delta y'' dt$, to generate the drive-reversal signal. In this case, it is expected that the direction of the input drive is less likely to be misled by the noise interference.

After the instant 2, the input decreases and the maximum output reaches again at the time instant 3. From this instant, the difference between the output of the peak-holding circuit and the output of the controlled system again starts to increase. At the time instant 4, the difference reaches the allowable limit c again, and the direction of the input drive is again reversed. At the time instant 5, the input becomes x''_{max} again and the output reaches y''_{max}.

This way of generating the drive-reversal signal is less subject to the noise interference as compared with the way of using the time derivative of output to generate it.

The period of output variation is the time interval from the instant 1 to the instant 3, and is called the hunting period T. The period of input variation is then $2T$. The extreme variation Δ of the output (Fig. 4.1 (c)) is called the hunting zone. If a is the amplitude of the triangular input (Fig. 4.1 (a)), then $\Delta = ka^2$. The difference between the maximum value and average value of the output is called the hunting loss H (Fig. 4.1 (c)), because of the fact that the output changes along a series of parabolic arcs, $H = \left(\frac{1}{3}\right)\Delta = \left(\frac{1}{3}\right)ka^2$.

It is then clear from this discussion that in order to reduce the hunting loss for better efficiency of the system, we must try to reduce the amplitude of input variation a or the allowable difference c. However, in practical applications, these values are limited by the undesirable effects of the dynamics of the controlled system and the noise interference on the control operation of the input drive.

2. Effects of the dynamics of the controlled system

In the operation principle of the peak-holding method, it was indicated that the control action was based only on the difference of the output. Accordingly, in the analysis of the system, it is also convenient to refer the output and the input to the optimum point, as discussed in Chap. II.

So, referring to the normalized optimalizing control system shown in Fig. 1.1, we shall first determine the general relation between the normalized input x^* and output y for the peak-holding method.

For the input x^* specified as a triangular wave with the period $2T$ and the amplitude a, the normalized pseudo input x is given by

$$x = \frac{8a}{\pi^2}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2(2j)}\left[G_i\left(\frac{2n+1}{2}\cdot j\omega_0\right)e^{[(2n+1)/2]j\omega_0 t} - G_i\left(-\frac{2n+1}{2}j\omega_0\right)e^{-[(2n+1)/2]j\omega_0 t}\right], \tag{4.1}$$

where $\omega_0 = 2\pi/T$.

Then, the output y is given as

$$y=\frac{16a^2k}{\pi^4}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{(-1)^{n+m}}{(2n+1)^2(2m+1)^2}\Big[G_0\{(n+m+1)j\omega_0\}\cdot G_i\Big(\frac{2n+1}{2}j\omega_0\Big)\cdot G_i\Big(\frac{2m+1}{2}j\omega_0\Big)$$

$$\times e^{(n+m+1)j\omega_0 t}-G_0\{(n-m)j\omega_0\}\cdot G_i\Big(\frac{2n+1}{2}j\omega_0\Big)\cdot G_i\Big(-\frac{2m+1}{2}j\omega_0\Big)\cdot e^{(n-m)j\omega_0 t}$$

$$-G_0\{(m-n)j\omega_0\}\cdot G_i\Big(-\frac{2n+1}{2}j\omega_0\Big)\cdot G_i\Big(\frac{2m+1}{2}j\omega_0\Big)e^{-(n-m)j\omega_0 t}$$

$$+G_0\{-(n+m+1)j\omega_0\}\cdot G_i\Big(-\frac{2n+1}{2}j\omega_0\Big)\cdot G_i\Big(-\frac{2m+1}{2}j\omega_0\Big)e^{-(n+m+1)j\omega_0 t}\Big]. \qquad (4.2)$$

The average of the output with respect to time, being here referred to the optimum output (=0), gives directly the hunting loss H. Equation (4.2) shows that this average value is the sum of terms with $n=m$ from the second and the third terms of that equation. Therefore, using $G_0(0)=1$, we have

$$H=\frac{32a^2k}{\pi^4}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^4}G_i\Big(\frac{2n+1}{2}j\omega_0\Big)G_i\Big(-\frac{(2n+1)}{2}j\omega_0\Big). \qquad (4.3)$$

This equation can be easily checked by observing that when the dynamic effects are absent, $G_i(s)=1$, then the series can be easily summed $H=\left(\frac{1}{3}\right)ka^2$ as determined before. Equation (4.3) also shows that the average output and hence the hunting loss are independent of the transfer function of the output linear group.

Equations (4.1) to (4.3) fully determine the performance of the optimalizing control system once the values of a, k and ω_0 are specified and the transfer functions $G_i(j\omega)$ and $G_0(j\omega)$ are given. However, for various forms of $G_i(s)$ and $G_0(s)$, it is not always possible to carry out the summantion in these equations. From the practical considerations that the hunting period is usually small and these transfer functions can be closely approximated by the first order or second order lag system, we shall only consider the cases where these transfer functions are combined as follows:

$$\text{(a)}\quad G_i(s)=1/(\tau_{i1}s+1)(\tau_{i2}s+1),\ G_0(s)=1, \qquad (4.4)$$

$$\text{(b)}\quad G_i(s)=1,\ G_0(s)=1/(\tau_{01}s+1)(\tau_{02}s+1), \qquad (4.5)$$

$$\text{(c)}\quad G_i(s)=1/(\tau_i s+1),\ G_0(s)=1/(\tau_0 s+1). \qquad (4.6)$$

Case (a):

By substituting Eq. (4.4) into Eq. (4.1), we obtain

$$x=\frac{8a}{\pi^2}\sum_{n=0}^{\infty}\frac{(-1)^n}{2j(2n+1)^2}\Big[\frac{e^{[(2n+1)/2]j\omega_0 t}}{\{1+(2n+1)j(\omega_0\tau_{i1}/2)\}\cdot\{1+(2n+1)j(\omega_0\tau_{i2}/2)\}}$$

$$-\frac{e^{-[(2n+1)/2]j\omega_0 t}}{\{1-(2n+1)j(\omega_0\tau_{i1}/2)\}\cdot\{1-(2n+1)j(\omega_0\tau_{i2}/2)\}}\Big]. \qquad (4.7)$$

The result of resolving Eq. (4.7) into partial fractions, carrying out the summation and simplifying the expressions is

$$x=(-1)^i ST\Big\{\Big(\frac{t}{T}-i\Big)-\Big(\frac{\tau_{i1}}{T}+\frac{\tau_{i2}}{T}\Big)+\frac{(\tau_{i1}/T)^2}{(\tau_{i1}/T)-(\tau_{i2}/T)}\cdot\frac{e^{-(T/\tau_{i1})((t/T)-i)}}{\cosh(T/2\,\tau_{i1})}$$
$$-\frac{(\tau_{i2}/T)^2}{(\tau_{i1}/T)-(\tau_{i2}/T)}\cdot\frac{e^{-(T/\tau_{i2})((t/T)-i)}}{\cosh(T/2\,\tau_{i2})}\Big\}, \tag{4.8}$$

when $-\frac{1}{2}\le\Big(\frac{t}{T}-i\Big)\le\frac{1}{2}$,

where $i=0,\ 1,\ 2,\ \ldots.$

$S(=2a/T)$ in this equation denotes the input drive speed.
The hunting loss given by Eq. (4.3) becomes

$$H=\frac{kS^2T^2}{12}\Big[1-12\Big\{\Big(\frac{\tau_{i1}}{T}\Big)^2+\Big(\frac{\tau_{i2}}{T}\Big)^2\Big\}$$
$$+24\Big\{\frac{(\tau_{i1}/T)^5}{(\tau_{i1}/T)^2-(\tau_{i2}/T)^2}\tanh\frac{T}{2\,\tau_{i1}}-\frac{(\tau_{i2}/T)^5}{(\tau_{i1}/T)^2-(\tau_{i2}/T)^2}\tanh\frac{T}{2\,\tau_{i2}}\Big\}\Big]. \tag{4.9}$$

Similarly, the output y is calculated as follows:

$$y=2\,kS^2T^2\Big[\frac{1}{2}\Big(\frac{t}{T}-i\Big)^2-\Big(\frac{\tau_{i1}}{T}+\frac{\tau_{i2}}{T}\Big)\Big(\frac{t}{T}-i\Big)$$
$$+\frac{1}{2}\Big(\frac{\tau_{i1}}{T}+\frac{\tau_{i2}}{T}\Big)^2+\frac{1}{(\tau_{i1}/T)-(\tau_{i2}/T)}\Big\{\Big(\frac{t}{T}-i\Big)-\Big(\frac{\tau_{i1}}{T}+\frac{\tau_{i2}}{T}\Big)\Big\}$$
$$\times\Big\{\Big(\frac{\tau_{i1}}{T}\Big)^2\cdot\frac{e^{-(T/\tau_{i1})((t/T)-i)}}{\cosh(T/2\,\tau_{i1})}-\Big(\frac{\tau_{i2}}{T}\Big)^2\cdot\frac{e^{-(T/\tau_{i2})((t/T)-i)}}{\cosh(T/2\,\tau_{i2})}\Big\}$$
$$+\frac{1}{2\{(\tau_{i1}/T)-(\tau_{i2}/T)\}^2}\Big\{\Big(\frac{\tau_{i1}}{T}\Big)^4\cdot\frac{e^{-(2T/\tau_{i1})((t/T)-i)}}{\cosh^2(T/2\,\tau_{i1})}+\Big(\frac{\tau_{i2}}{T}\Big)^4\cdot\frac{e^{-(2T/\tau_{i2})((t/T)-i)}}{\cosh^2(T/2\,\tau_{i2})}$$
$$-2\Big(\frac{\tau_{i1}}{T}\Big)^2\cdot\Big(\frac{\tau_{i2}}{T}\Big)^2\frac{e^{-((T/\tau_{i1})+(T/\tau_{i2}))((t/T)-i)}}{\cosh(T/2\,\tau_{i1})\cosh(T/2\,\tau_{i2})}\Big\}\Big], \tag{4.10}$$

when $-\frac{1}{2}\le\Big(\frac{t}{T}-i\Big)\le\frac{1}{2}$,

where $i=0,\ 1,\ 2,\ \ldots.$

From the principle of operation, it is seen that the most important quantity to be assigned for its design is the allowable difference c. By definition, c is the difference of the maximum value of the output y and the value of y at $t/T=1/2$, the time instant at which the input drive is reversed. If the instant corresponding to the maximum value of the output is t^*, then the allowable differrence c is calculated as

$$c=y(1/2)-y(t^*/T) \tag{4.11}$$

by use of Eq. (4.10).

To determine t^*, we may use the condition $dy/dt=0$ at Eq. (4.10). Then, noting that the instant of the input drive reversal must come after the instant of the maximum output, that is, $t^*/T<1/2$, we obtain

$$\frac{t^*}{T}-\Big[\Big(\frac{\tau_{i1}}{T}+\frac{\tau_{i2}}{T}\Big)+\frac{(\tau_{i1}/T)^2}{(\tau_{i1}/T)-(\tau_{i2}/T)}\cdot\frac{e^{-(T/\tau_{i1})(t^*/T)}}{\cosh(T/2\,\tau_{i1})}-\frac{(\tau_{i2}/T)^2}{(\tau_{i1}/T)-(\tau_{i2}/T)}\cdot\frac{e^{-(T/\tau_{i2})(t^*/T)}}{\cosh(T/2\,\tau_{i2})}\Big)\Big]=0. \tag{4.12}$$

This transcendental equation for t^*/T may be solved by graph or iteration.

When t^*/T is determined by Eq. (4.12), Eq. (4.11) gives c. Here, the specified quantities of the system are k, the characteristic constant of the controlled system, and τ_{i1}, τ_{i2}, the time constants of $G_i(s)$. In consideration of the noise interference and the quickness of the responce, c and T must be chosen properly. Once the values of k, τ_{i1}, τ_{i2}, T and c are known, S, the input drive speed, and H, the hunting loss, can be determined. From Eq. (4.11),

$$S=\frac{1}{T}\sqrt{\frac{c}{k}}\Big[\frac{1}{2}-\Big(\frac{\tau_{i1}}{T}+\frac{\tau_{i2}}{T}\Big)+\frac{(\tau_{i1}/T)^2}{(\tau_{i1}/T)-(\tau_{i2}/T)}\cdot\frac{e^{-T/2\tau_{i1}}}{\cosh(T/2\,\tau_{i1})}-\frac{(\tau_{i2}/T)^2}{(\tau_{i1}/T)-(\tau_{i2}/T)}\cdot\frac{e^{-T/2\,\tau_{i2}}}{\cosh(T/2\,\tau_{i2})}\Big]^{-1/2}. \tag{4.13}$$

When S is determined, Eq. (4.9) then gives the hunting loss H.

Case (b):

By substituting Eq. (4.5) into Eq. (4.2), we obtain

$$y=-\frac{4\,T^2S^2k}{\pi^4}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{(-1)^{n+m}}{(2\,n+1)^2(2\,m+1)^2}\Big\{\frac{e^{j(n+m+1)\omega_0 t}}{\{1+(n+m+1)j\omega_0\tau_{01}\}\{1+(n+m+1)j\omega_0\tau_{02}\}}-\frac{e^{j(n-m)\omega_0 t}}{\{1+(n-m)j\omega_0\tau_{01}\}\{1+(n-m)j\omega_0\tau_{02}\}}-\frac{e^{-j(n-m)\omega_0 t}}{\{1-(n-m)j\omega_0\tau_{01}\}\{1-(n-m)j\omega_0\tau_{02}\}}+\frac{e^{-j(n+m+1)\omega_0 t}}{\{1-(n+m+1)j\omega_0\tau_{01}\}\{1-(n+m+1)j\omega_0\tau_{02}\}}\Big\}. \tag{4.14}$$

By carrying out the summation in Eq. (4.14) and simplifying the expressions, we have

$$y=2\,T^2S^2k\Big[\Big\{\frac{1}{2}\Big(\frac{t}{T}-i\Big)^2-\Big(\frac{\tau_{01}}{T}+\frac{\tau_{02}}{T}\Big)\Big(\frac{t}{T}-i\Big)+\Big(\frac{\tau_{01}}{T}\Big)^2+\Big(\frac{\tau_{01}\tau_{02}}{T}\Big)+\Big(\frac{\tau_{02}}{T}\Big)^2\Big\}-\frac{1}{2}\Big\{\frac{(\tau_{01}/T)^2}{(\tau_{01}/T)-(\tau_{02}/T)}\cdot\frac{e^{-(T/\tau_{01})((t/T)-i)}}{\sinh(T/2\,\tau_{01})}-\frac{(\tau_{02}/T)^2}{(\tau_{01}/T)-(\tau_{02}/T)}\cdot\frac{e^{(T/\tau_{02})((t/T)-i)}}{\sinh(T/2\,\tau_{02})}\Big\}\Big], \tag{4.15}$$

when $$-\frac{1}{2}\le\Big(\frac{t}{T}-i\Big)\le\frac{1}{2},$$

where $$i=0,\ 1,\ 2,\ \ldots .$$

Through a similar procedure to the case (a), the instant t^* can be determined by the following equation:

$$\Big(\frac{t^*}{T}\Big)-\Big(\frac{\tau_{01}}{T}+\frac{\tau_{02}}{T}\Big)+\frac{1}{2}\Big\{\frac{(\tau_{01}/T)}{(\tau_{01}/T)-(\tau_{02}/T)}\cdot\frac{e^{-(T/\tau_{01})(t^*/T)}}{\sinh(T/2\,\tau_{01})}-\frac{(\tau_{02}/T)}{(\tau_{01}/T)-(\tau_{02}/T)}\cdot\frac{e^{-(T/\tau_{02})(t^*/T)}}{\sinh(T/2\,\tau_{02})}\Big\}=0. \tag{4.16}$$

And, the input drive speed S is given by

$$
\begin{aligned}
S=\frac{1}{T}\sqrt{\frac{c}{k}}\Big[&-\Big(\frac{t^*}{T}\Big)^2+\frac{1}{4}+2\Big(\frac{t^*}{T}-\frac{1}{2}\Big)\Big(\frac{\tau_{01}}{T}+\frac{\tau_{02}}{T}\Big)\\
&+\frac{(\tau_{01}/T)^2}{(\tau_{01}/T)-(\tau_{02}/T)}\cdot\frac{1}{\sinh(T/2\,\tau_{01})}\{e^{-(T/\tau_{01})(t^*/T)}-e^{-(T/2\tau_{01})}\}\\
&-\frac{(\tau_{02}/T)^2}{(\tau_{01}/T)-(\tau_{02}/T)}\cdot\frac{1}{\sinh(T/2\,\tau_{02})}\{e^{-(T/\tau_{02})(t^*/T)}-e^{-(T/2\tau_{02})}\}.
\end{aligned}
\tag{4.17}
$$

When S is determined by Eq. (4. 17), a simple formula $H=\frac{1}{12}kS^2T^2$ will give the hunting loss H for this case.

Case (*c*):

In this case, we can readily obtain the input x and the hunting loss H by putting $\tau_{i2}=0$ in Eq. (4.8) and Eq. (4.9), respectively.

The output y can be calculated by the insertion of Eq. (4.6) into Eq. (4.2).

$$
\begin{aligned}
y=2\,kT^2S^2\Big\{&-\Big[\frac{1}{2}\Big(\frac{t}{T}-i\Big)^2-\Big(\frac{\tau_0}{T}+\frac{\tau_i}{T}\Big)\Big(\frac{t}{T}-i\Big)+\frac{1}{2}\Big(\frac{\tau_i}{T}\Big)^2+\frac{\tau_i\tau_0}{T}+\Big(\frac{\tau_0}{T}\Big)^2\Big]\\
&-\frac{1}{2}\Big[\frac{(\tau_0/T)^2}{(\tau_0/T)-(\tau_i/T)}\left\{\frac{2\Big(\frac{\tau_i}{T}\Big)^3\tanh\frac{1}{2}\Big(\frac{T}{\tau_i}\Big)}{\Big(\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)\Big(2\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)}-1\right\}\frac{e^{-(T/\tau_0)((t/T)-i)}}{\sinh(T/2\,\tau_0)}\\
&-\Big\{\Big(\frac{t}{T}-i\Big)+\frac{(\tau_i/T)^2}{(\tau_0/T)-(\tau_i/T)}\Big\}\frac{2\Big(\frac{\tau_i}{T}\Big)^2e^{-(T/\tau_i)((t/T)-i)}}{\Big(\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)\cosh(T/2\,\tau_i)}\\
&-\frac{(\tau_i/T)^3}{\Big(2\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)}\cdot\frac{e^{-(T/\tau_i)2((t/T)-i)}}{\cosh^2(T/2\,\tau_i)}\Big]\Big\},
\end{aligned}
\tag{4.18}
$$

when $\qquad -\frac{1}{2}\leq\Big(\frac{t}{T}-i\Big)\leq\frac{1}{2},$

where $\qquad i=0,\ 1,\ 2,\ \ldots\ .$

Similarly to the previous case, t^* is determined by the following equation:

$$
\begin{aligned}
&\Big(\frac{t^*}{T}\Big)-\Big(\frac{\tau_i}{T}+\frac{\tau_0}{T}\Big)-\frac{1}{2}\cdot\frac{(\tau_0/T)}{(\tau_0/T-\tau_i/T)}\left[\frac{2\Big(\frac{\tau_i}{T}\Big)^3\tanh\frac{T}{2\,\tau_i}}{\Big(\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)\Big(2\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)}-1\right]\cdot\frac{e^{-(T/\tau_0)(t^*/T)}}{\sinh(T/2\,\tau_0)}\\
&\quad-\Big[1-\Big(\frac{t^*}{T}\Big)\Big(\frac{T}{\tau_i}\Big)-\frac{(\tau_i/T)}{(\tau_0/T)-(\tau_i/T)}\Big]\cdot\frac{\Big(\frac{\tau_i}{T}\Big)^2e^{-(T/\tau_i)(t^*/T)}}{\Big(\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)\cosh\Big(\frac{T}{2\,\tau_i}\Big)}\\
&\quad+\frac{\Big(\frac{\tau_i}{T}\Big)^2e^{-(2T/\tau_i)(t^*/T)}}{\Big(2\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)\cosh^2\Big(\frac{T}{2\,\tau_i}\Big)}=0.
\end{aligned}
\tag{4.19}
$$

And, the input drive speed S is given by

$$S=\frac{1}{T}\sqrt{\frac{c}{k}}\Big[-\Big(\frac{t^*}{T}\Big)^2+\frac{1}{4}+2\Big(\frac{t^*}{T}-\frac{1}{2}\Big)\Big(\frac{\tau_i}{T}+\frac{\tau_0}{T}\Big)-\frac{(\tau_0/T)^2}{\Big(\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)\sinh(T/2\,\tau_0)}$$

$$\times\left\{\frac{2\Big(\frac{\tau_i}{T}\Big)^3\tanh(T/2\,\tau_i)}{\Big(\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)\Big(\frac{2\,\tau_0}{T}-\frac{\tau_i}{T}\Big)}-1\right\}\cdot\{e^{-(T/\tau_0)\cdot(t^*/T)}-e^{-(T/\tau_0)\cdot 1/2}\}$$

$$+\frac{2\Big(\frac{\tau_i}{T}\Big)^2}{\Big(\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)\cosh(T/2\,\tau_i)}\left\{\Bigg(\frac{t^*}{T}+\frac{(\tau_i/T)^2}{\Big(\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)}\Bigg)e^{-(T/\tau_i)(t^*/T)}-\Bigg(\frac{1}{2}+\frac{(\tau_i/T)^2}{\Big(\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)}\Bigg)\right.$$

$$\left.\times e^{-(T/\tau_i)\cdot 1/2}\right\}+\frac{(\tau_i/T)^3}{\Big(2\frac{\tau_0}{T}-\frac{\tau_i}{T}\Big)\cosh^2(T/2\,\tau_i)}\{e^{-(T/\tau_i)(2\,t^*/T)}-e^{-(T/\tau_i)}\}\Big]^{-1/2}. \tag{4.20}$$

3. Comparison of an approximate solution and the exact solution

For engineering analysis and synthesis it is far more preferable to work with the simple equations which may be derived with the input approximated by a sinusoid. So, it is assumed that the input x^* can be approximated by a sinusoidal wave with the amplitude a' and the period T as follows:

$$x^*=a'\sin\omega_i t, \tag{4.21}$$

where $\omega_i=\pi/T$.

Then, referring to Fig. 1.1, we obtain the output y corresponding to the input of Eq. (4.21),

$$y=\frac{ka'^2}{2}|G_i(j\omega_i)|^2\{1-|G_0(j\,2\,\omega_i)|\cos(2\,\omega_i t+2\,\varphi_i+\varphi_0')\}, \tag{4.22}$$

where $\varphi_i=\angle G_i(j\omega_i),\quad \varphi_0'=\angle G_0(j\,2\,\omega_i)$.

Through a similar procedure to the case of the exact solution, t^*, the time instant when the maximum output occurs, S, the input drive speed, and H, the hunting loss, for the case of the approximate solution are found to be

$$\frac{t^*}{T}=-\frac{1}{\pi}\Big(\varphi_i+\frac{1}{2}\varphi_0'\Big), \tag{4.23}$$

$$S=\frac{1}{T}\sqrt{\frac{c}{k}}\cdot\sqrt{6}\left\{|G_i(j\omega_i)|^2\cdot|G_0(2\,\omega_i)|\cos^2\pi\Big(\frac{t^*}{T}\Big)\right\}^{-1/2}, \tag{4.24}$$

$$H=\frac{1}{2}ka'^2|G_i(j\omega_i)|^2=\frac{kT^2S^2}{12}\cdot|G_i(j\omega_i)|^2. \tag{4.25}$$

By use of Eqs. (4.24) to (4.26), we obtain the following results for the cases of (a), (b) and (c):

for the case (a);

$$\frac{t^*}{T} = -\frac{1}{\pi}\varphi_i, \quad S = \frac{1}{T}\sqrt{\frac{c}{k}}\cdot\sqrt{6}\left\{\frac{1}{\sqrt{1+\left(\pi\frac{\tau_{i1}}{T}\right)^2}}\cdot\frac{1}{\sqrt{1+\left(\pi\frac{\tau_{i2}}{T}\right)^2}}\cos\varphi_i\right\}^{-1},$$

$$H = \frac{kT^2S^2}{12}\cdot\frac{1}{1+\left(\pi\frac{\tau_{i1}}{T}\right)^2}\cdot\frac{1}{1+\left(\pi\frac{\tau_{i2}}{T}\right)^2},$$

where
$$\varphi_i = -\left\{\tan^{-1}\left(\pi\frac{\tau_{i1}}{T}\right)+\tan^{-1}\left(\pi\frac{\tau_{i2}}{T}\right)\right\},$$

for the case (b):

$$\frac{t^*}{T} = -\frac{1}{2\pi}\varphi_0', \quad S = \frac{1}{T}\sqrt{\frac{c}{k}}\cdot\sqrt{6}\left\{\frac{1}{\sqrt{1+\left(2\pi\frac{\tau_{01}}{T}\right)^2}}\cdot\frac{1}{\sqrt{1+\left(2\pi\frac{\tau_{02}}{T}\right)^2}}\right.$$

$$\left.\times\cos^2\left(\frac{1}{2}\varphi_0'\right)\right\}^{-1/2}, \quad H = \frac{kT^2S^2}{12},$$

where
$$\varphi_0' = -\left\{\tan^{-1}\left(2\pi\frac{\tau_{01}}{T}\right)+\tan^{-1}\left(2\pi\frac{\tau_{02}}{T}\right)\right\},$$

for the case (c):

$$\frac{t^*}{T} = -\frac{1}{\pi}\left(\varphi_i+\frac{1}{2}\varphi_0'\right), \quad S = \frac{1}{T}\sqrt{\frac{c}{k}}\sqrt{6}\left\{\frac{1}{1+\left(\pi\frac{\tau_i}{T}\right)^2}\cdot\frac{1}{\sqrt{1+\left(2\pi\frac{\tau_0}{T}\right)^2}}\right.$$

$$\left.\times\cos^2\left(\varphi_i+\frac{1}{2}\varphi_0'\right)\right\}^{-1/2}, \quad H = \frac{kT^2S^2}{12}\cdot\frac{1}{1+\left(\pi\frac{\tau_i}{T}\right)^2}.$$

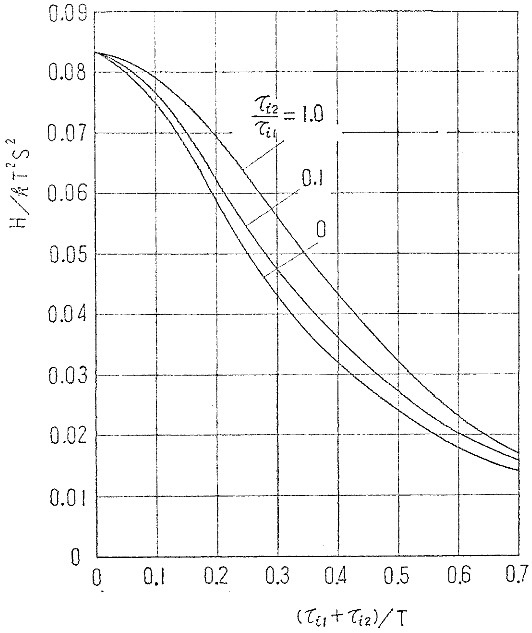


FIG. 4.3. Dimensionless hunting loss H/kT^2S^2 in case of $G_i(s)=1/(\tau_{i1}s+1)(\tau_{i2}s+1)$.

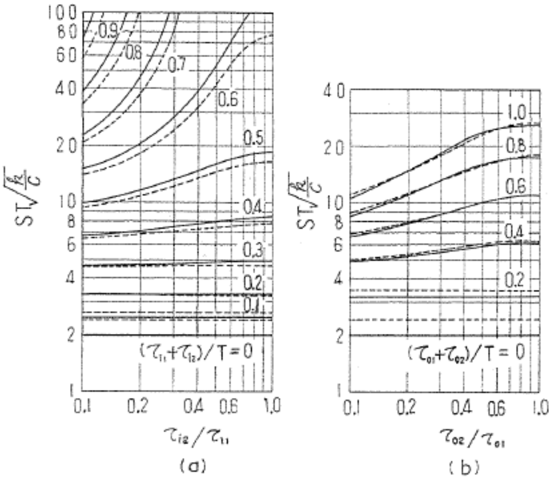


FIG. 4.4. Dimensionless input drive speed $ST\sqrt{\frac{k}{c}}$ in case of $G_i(s)=1/(\tau_{i1}s+1)\times(\tau_{i2}s+1)$ or $G_0(s)=1/(\tau_{01}s+1)(\tau_{02}s+1)$ (—— exact solution, ----- approximate one).

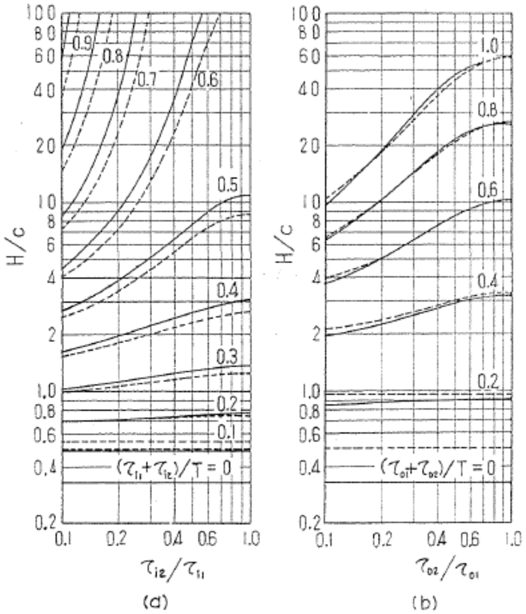


FIG. 4.5. Relative hunting loss H/c in case of $G_i(s)=1/(\tau_{i1}s+1)(\tau_{i2}s+1)$ or $G_0(s)=1/(\tau_{01}s+1)(\tau_{02}s+1)$ (—— exact solution, ----- approximate one).

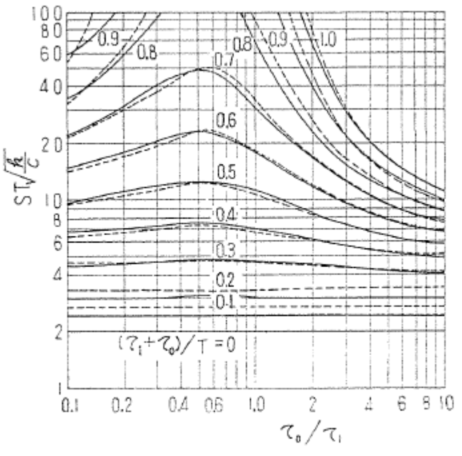


FIG. 4.6. Dimensionless input drive speed $ST\sqrt{\frac{k}{c}}$ in case of $G_i(s)=1/(\tau_i s+1)$ and $G_0(s)=1/(\tau_0 s+1)$ (—— exact solution, ----- approximate one).

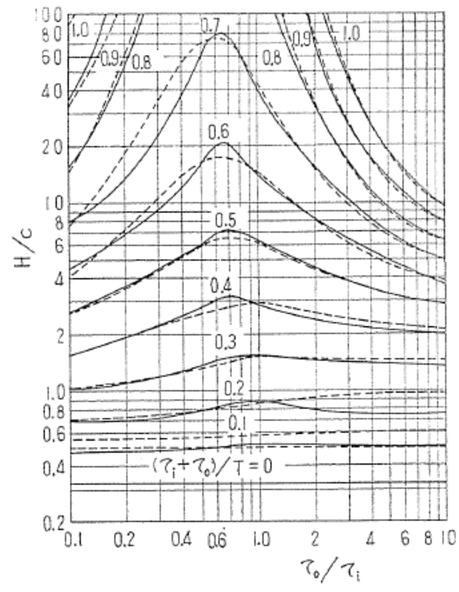


FIG. 4.7. Relative hunting loss H/c in case of $G_i(s)=1/(\tau_i s+1)$ and $G_0(s)=1/(\tau_0 s+1)$ (—— exact solution, ----- approximate one).

Now, we compare the approximate solution with the exact one under the condition that the root mean squares, or effective values, of the inputs are equal between these two cases. Accordingly, the amplitudes of a triangular wave a and that of a sinusoidal wave a' must satisfy the following relation:

$$a' = \sqrt{\frac{2}{3}}\,a. \tag{4.26}$$

As the result, it can be shown that the magnitudes of the hunting loss of the two solutions agree with each other (Fig. 4.3). Figures 4.4 to 4.7 illustrate the comparisons of the two solutions for the input drive speed S and the relative hunting loss H/c, respectively, of the cases (a), (b) and (c).

4. Design charts and discussions

Figures 4.3 to 4.7 are the design charts derived from the present analysis for the peak-holding method. From Fig. 4.3, it is observed that the dimensionless hunting loss becomes smaller with the increase of $(\tau_{i1}+\tau_{i2})/T$, regardless of the value of τ_{i2}/τ_{i1}. Figures 4.6 and 4.7 show that the values of $ST\sqrt{k/c}$ and H/c are almost constant for smaller $(\tau_i+\tau_0)/T$, but have peaks for larger $(\tau_i+\tau_0)/T$ near $\tau_0/\tau_i=1$. Similar tendencies are observed in Figs. 4.4 and 4.5, too.

According to the present analysis, the necessary input drive speed S and the hunting loss H are given for any specified hunting period T, the assigned allowable difference c, and the time constants for the input linear group and the output linear group. c is fixed by considerations on the noise interference. As is clearly shown in Figs. 4.5 and 4.7, whenever the hunting period T is relatively short with respect to the time constants, the relative hunting loss H/c will be large, especially when the time constants are nearly equal. This situation is quite unfavorable, because even if c is chosen sufficiently small for a better efficiency of the system, the increase in H will prevent the desired purpose from being attained. In order to avoid such a situation, we may choose T to be very large and reduce H/c at the cost of the deteriorating response time, or alternatively we can adjust the time constant of either input or output linear group by means of compensation technique and make the two time constants differ sufficiently from each other as required.

Chapter V. Cross-Correlation Method

The cross-correlation method is to search out an extreme value by use of a test signal with relatively small amplitude, which is superimposed intentionally on the controlled input. The distinct advantage of this control method is that the control action is little affected by the noise interference, though its controller is apt to be rather complicated.

1. Principle of operation

The principle of the cross-correlation method can be illustrated by Figs. 5.1 and 5.2 as follows. For simplicity, it is assumed that the test signal is a sinusoidal wave and the effects of the dynamics are absent.

As is shown in Fig. 5.1, if the test signal is superimposed on the controlled input x_1'' which is located on the low-input side of the optimum point, the com-

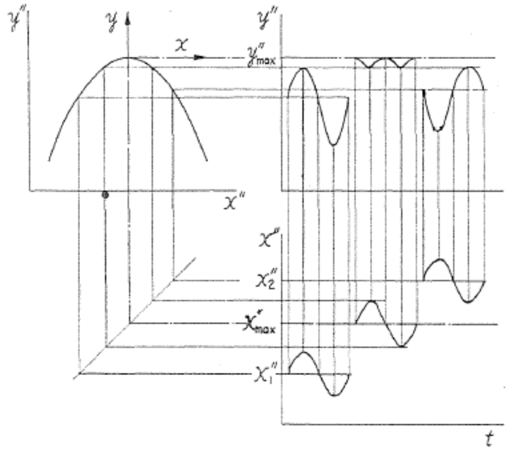


FIG. 5.1. Steady-state output response to a sinusoidal test signal on each level of the controlled input.

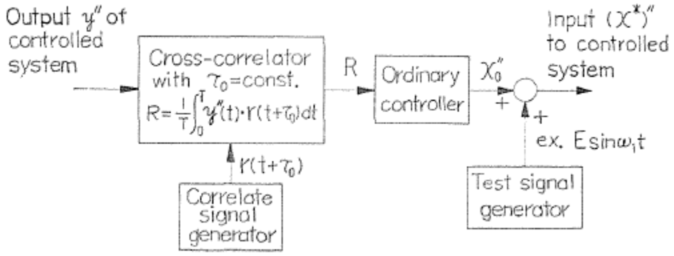


FIG. 5.2. Functional diagram of the optimalizing controller for the cross-correlation method.

ponent of the output produced by this test signal will contain the fundamental sinusoidal component which will always be in phase with the test signal, whereas if the test signal is superimposed on the high-input-level side of the optimum point, the fundamental output component will always be out of phase by 180 degrees with the test signal. Accordingly, phase relationships between the test signal and the fundamental output component indicates whether the controlled input is above or below the optimum value. In Fig. 5.1, the test signal is assumed to be superimposed on the input at the optimum value. The corresponding output is determined by the parabolic characteristic curve, which squares the sinusoidal test signal. As the result, the output completes two cycles while the test signal does one. In other words, the component having the same frequency as the test signal will disappear from the output.

In priciple, the cross-correlation method is based on this simple fact. That is, the output is cross-correlated with a sinusoidal wave, or correlate signal, having the same frequency and phase as those of the test signal.

Here the following relations hold in general:

$$\left.\begin{aligned}&\int_{-\pi}^{\pi}\sin\tau\cdot\cos m\tau d\tau=0\\&\int_{-\pi}^{\pi}\sin\tau\cdot\sin m\tau d\tau=0\qquad(m\neq1)\end{aligned}\right\}\tag{5.1}$$

Accordingly only if the controlled input is at its optimum value, the cross-correlation vanishes, otherwise it takes some non-zero value, either positive or negative, and it is only the fundamental component of the output that does contribution to the value.

From the above discussion, we can use any periodic signal as the test signal. So, for easy generation in practice, the wave forms shown in Table 5.1 are considered as the test signal. Table 5.1 also gives an appropriate correlate signal for each test signal. Here, the major reason why the random signal is selected as the convenient test signal is that the random signal is not necessarily required to be generated intentionally, but the noise which is inevitably introduced into the system and disturbs it, can be utilized for the test purpose. In a similar manner to the case of other test signals, we can detect the deviation of the controlled input from the optimum value in this case, too.

TABLE 5.1

Test signal	Correlate signal
Sinusoidal wave	Sinusoidal or rectangular wave
Rectangular or triangular wave	Sinusoidal wave
Random signal	Random signal with the same statistical properties as the test signal

2. Analysis of the control system

2.1. The case where the test signal is a sinusoid

When a sinusoidal wave is chosen as the test signal, we can use, as the correlate signal, any periodic signal with the same frequency as that of the test signal. Since a square (or rectangular) wave used as the correlate signal permits a simple construction of the controller, we are going to calculate the cross-correlation for this case.

Now, let it be assumed that the controlled input is at the value of x_0'' and the sinusoidal test signal has the amplitude E and the angular frequency ω_i, then the total input to the controlled system is given by

$$(x^*)'' = E\sin\omega_i t + x_0''. \tag{5.2}$$

From this expression together with the consideration of the dynamic effects of the controlled system, the output y'' of the controlled system is found to be

$$\begin{aligned} y'' = y''_{\max} + k\Big[& (x_0'' - x''_{\max})^2 + \frac{E^2}{2}|G_i(j\omega_i)|^2 \\ & + 2E(x_0'' - x''_{\max})|G_i(j\omega_i)|\cdot|G_0(j\omega_i)|\sin(\omega_i t + \varphi_i + \varphi_0) \\ & - \frac{E^2}{2}|G_i(j\omega_i)|^2\cdot|G_0(j2\omega_i)|\cos(2\omega_i t + 2\varphi_i + \varphi_0)\Big], \end{aligned} \tag{5.3}$$

where $\qquad \varphi_i = \angle G_i(j\omega_i), \quad \varphi_0 = \angle G_0(j\omega_i), \quad \varphi_0' = \angle G_0(j2\omega_i).$

The third term in Eq. (5.3) gives the time average of the deviation of the output from its extreme, or the hunting loss. Figure 5.3 gives the magnitude of the hunting loss for various test signals when $G_i(j\omega_i) = 1/(1+j\omega\tau_i)$. Here, H_s, H_r and

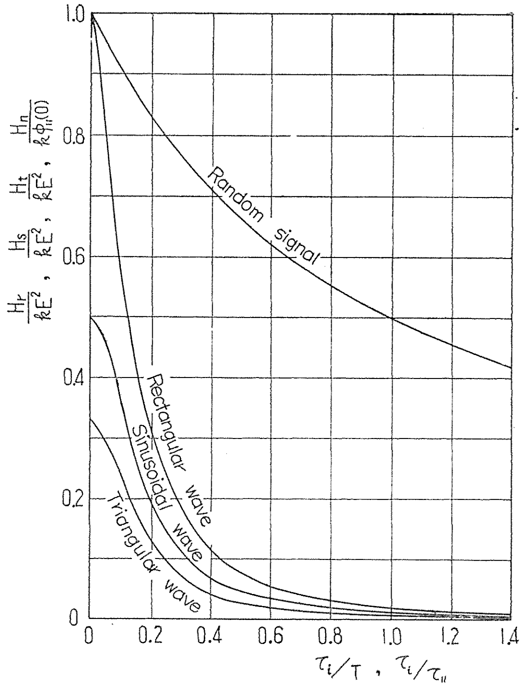


FIG. 5. 3. Dimensionless hunting loss for different test signals in case of $G_i(s)=1/(\tau_i s+1)$.

H_t are the hunting losses corresponding to the sinusoidal, rectangular and triangular wave respectively. They are evaluated by use of the following expressions obtained from Eq. (4.3) in Chap. IV:

$$H_s=\frac{1}{2}kE^2\cdot\frac{1}{1+4\pi^2\left(\frac{\tau_i}{T}\right)^2},$$

$$H_r=kE^2\left\{1-4\left(\frac{\tau_i}{T}\right)\tanh\left(\frac{T}{4\tau_i}\right)\right\},$$

$$H_t=kE^2\left\{\frac{1}{3}-16\left(\frac{\tau_i}{T}\right)^2+64\left(\frac{\tau_i}{T}\right)^3\tanh\left(\frac{T}{4\tau_i}\right)\right\},$$

where $$T=2\pi/\omega_i.$$

At cross-correlator, the output y'' is correlated with a rectangular wave with the amplitude A and the frequency ω_i. According to Eq. (5.1), only the fundamental component of the rectangular wave comes into the picture in the calculation of the cross-correlation. Acordingly,

$$R_r=\frac{1}{2\pi}\int_{-\pi}^{\pi}y''(\tau)\cdot\frac{4A}{\pi}\sin(\tau+\theta_0)d\tau, \tag{5.4}$$

where $\tau\equiv\omega_i t$ and θ_0 is the phase shift between the test signal and the fundamental component of the correlate signal.

By substituting Eq. (5.3) into Eq. (5.4) and simplifying the expressions, we obtain

$$R_r = \frac{4kA(x_0'' - x_{max}'')E}{\pi} |G_i(j\omega_i)| \cdot |G_0(j\omega_i)| \cos(\varphi_i + \varphi_0 - \theta_0). \tag{5.5}$$

Here, if θ_0 is chosen to be $\varphi_i + \varphi_0$, then Eq. (5.5) becomes

$$R_r = \frac{4kA(x_0'' - x_{max}'')E}{\pi} |G_i(j\omega_i)| \cdot |G_0(j\omega_i)|. \tag{5.6}$$

It is observed from Eq. (5.6) that the cross-correlated value R_r is proportional to the deviation of the input from the desired value x_{max}''. From this fact it follows that any conventional controller, such as P.I.D. controllers, can be used in tandem with the cross-correlator. Consequently, the output can be brought to the extreme value by use of a cross-correlator combined with signal generators, and a conventional controller.

In design of the system, the amplitude E and the angular frequency ω_i of the test signal must be determined so as to give a large signal to noise ratio in the output of the cross-correlator. To do so, we must have sufficient knowledge about the power spectrum of the noise. Moreover, E and ω_i are related directly to the hunting loss, which we desire to make small. Namely, as E or ω_i is chosen larger, the hunting loss becomes larger, too. Consequently, we must make a good compromise between the two requirements.

2.2. The case where the test signal is random

A random signal is here assumed to be a stationary stochastic signal with Gaussian distribution. For ease of expression, its time average is assumed to be zero. To proceed with the discussion concisely, we shall consider hereafter the normalized control system illustrated in Fig. 1.1.

If the weighting functions corresponding to $G_i(s)$ and $G_0(s)$ are specified as $g_i(t)$ and $g_0(t)$ respectively, then the output $x(t)$ for the input $x^*(t)$ is given as

$$x(t) = \int_0^{8} g_i(\tau) \cdot x^*(t-\tau) d\tau. \tag{5.7}$$

Since the input $x^*(t)$ is equal to the sum of the random signal $n_1(t)$ as the test signal and the controlled input x_0 from the controller, Eq. (5.7) can be written as

$$x(t) = \int_0^{\infty} g_i(\tau) \cdot \{x_0 + n_1(t-\tau)\} d\tau.$$

By use of the relations that $\int_0^{\infty} g_i(t) dt = G_i(0)$ and $G_i(0) = G_0(0) = 1$, the above expression is simplified as follows:

$$x(t) = x_0 + \int_0^{\infty} g_i(\tau) \cdot n_1(t-\tau) d\tau. \tag{5.8}$$

Then, the output $y(t)$ is given by

$$y(t)=k\Big\{x_0^2\int_0^\infty g_0(\tau)d\tau+2x_0\int_0^\infty\int_0^\infty d\tau_1 d\tau_2\cdot g_i(\tau_1)g_0(\tau_2)\cdot n_1(t-\tau_1-\tau_2)$$

$$+\int_0^\infty\int_0^\infty\int_0^\infty d\tau_i d\tau_2 d\tau_3\cdot g_i(\tau_1)g_i(\tau_2)g_0(\tau_3)\cdot n_1(t-\tau_1-\tau_3)\cdot n_1(t-\tau_2-\tau_3)\Big\}. \quad (5.9)$$

At cross-correlator, the output $y(t)$ is cross-correlated with the correlate signal $n_1(t)$ as follows:

$$R_{n1}=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}y(t)\cdot n_1(t)dt. \quad (5.10)$$

By substituting Eq. (5.9) into Eq. (5.10) and using the assumption that $\overline{n_1(t)}=0$, we have

$$R_{n1}=2x_0k\int_0^\infty\int_0^\infty d\tau_1 d\tau_2\cdot g_i(\tau_1)\cdot g_0(\tau_2)\cdot\phi_{11}(\tau_1+\tau_2), \quad (5.11)$$

where $\phi_{11}(\tau_1+\tau_2)\left(=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}n_1(t-\tau_1-\tau_2)\cdot n_1(t)dt\right)$ is the auto-correlation function of $n_1(t)$.

The hunting loss is derived from Eq. (5.9) as follows:

$$H_n=\overline{y(t)}=k\int_0^\infty\int_0^\infty d\tau_i d\tau_2\cdot g_i(\tau_1)\cdot g_i(\tau_2)\phi_{11}(\tau_1-\tau_2). \quad (5.12)$$

It is of interest to note that Eq. (5.12) corresponds to Eq. (4.3).

As an example, when $G_i(s)=1/(1+\tau_i s)$, $G_0(s)=1/(1+\tau_0 s)$ and $\phi_{11}(\tau)=\phi_{11}(0)e^{-|\tau|/\tau_{11}}$, we can calculate R_{n1} and H_n from Eqs. (5.11) and (5.12) as follows:

$$R_{n1}=\frac{2kx_0\phi_{11}(0)}{\left(1+\frac{\tau_i}{\tau_{11}}\right)\left(1+\frac{\tau_0}{\tau_{11}}\right)},\qquad H_n=\frac{k\phi_{11}(0)}{1+\frac{\tau_i}{\tau_{11}}} \quad \text{(Fig. 5.3)}.$$

3. A method for eliminating the noise interference

When the noise $n(t)$, evaluated on the output of the controlled system as shown in Fig. 1.1, is assumed to be superimposed on the output $y(t)$ which corresponds to the input $x(t)$, the cross-correlated value, defined in the previous section, consists of the following two parts:

$$R=R_s+R_n, \quad (5.13)$$

where R_s = cross-correlated value between the output $y(t)$ and the correlate signal $r(t)$,

R_n = cross-correlated value between the noise $n(t)$ and the correlate signal $r(t)$.

If the correlation process is carried out over a sufficiently long time interval, it is clear that R_n tends to zero and thus R becomes equal to R_s. In view of quick response, however, it is desired to reduce the time required to compute the cross-correlated value. Therefore, R becomes inevitably an cross-correlated value containing unwanted cross-correlated value R_n due to the noise interference. Thus, for small value of R the control action may be misled with a certain probable error associated with the value of R.

It is conceivable that the error occurs both in the magnitude and the algebraic sign of R. Error in the algebraic sign is more serious than in the magnitude, since the controlled input is moved in the wrong direction. So, whenever R is less than certain limited value, it is preferable that the control action is stopped so that the controlled input is left unchanged. In the following sections, the limited value of R for a given value of the probable error will be determined for different correlate signals, as shown in Table 5.1.

3.1. The case where the correlate signal is a rectangular wave

As was mentioned in the previous section, R_n in Eq. (5.13) can be approximately written as

$$R_n = \frac{A}{mT}\Big[\int_{0-}^{T/2-\theta} n(t)dt - \int_{T/2-\theta}^{T-\theta} n(t)dt + \cdots + \int_{(m-1)T-\theta}^{\{(2m-1)/2\}T-\theta} n(t)dt - \int_{\{(2m-1)/2\}T-\theta}^{mT-\theta} n(t)dt\Big],$$

where m is a positive integer and indicates the number of input cycles over which the correlation process is carried out.

By changing the integration variable to $t' = t + \theta - \left\{\frac{(2m-1)}{2}\right\}T$, and simplifying the expression, we obtain

$$R_n = \frac{A}{mT}\sum_{i=1}^{m}\int_0^{T/2}\left\{n[t+(i-1)T-\theta] - n\left[t+\frac{2i-1}{2}T-\theta\right]\right\}dt. \tag{5.14}$$

On the other hand, the probability with which the controlled input is moved in the wrong direction is equivalent to the probability with which $R \cdot R_s$ becomes negative. To compute the probability that $R \cdot R_s < 0$, the following statistical variable is considered:

$$\begin{aligned} Z &= R \cdot R_s = R^2 - R_n \cdot R \\ &= R^2 - R\left[\frac{A}{mT}\sum_{i=1}^{\infty}\int_0^{T/2}\left\{n[t+(i-1)T-\theta] - n\left[t+\frac{2i-1}{2}T-\theta\right]\right\}dt\right]. \end{aligned} \tag{5.15}$$

From the assumption that the noise considered in this analysis is a stationary random signal, the mean value $\bar{z}$ and the variance σ_z^2 of the statistical variable z are given as

$$\bar{z} = R^2 - R\left[\frac{A}{mT}\sum_{i=1}^{m}\int_0^{T/2}\overline{\left\{n[t+(i-1)T-\theta] - n\left[t+\frac{(2i-1)}{2}T-\theta\right]\right\}}dt\right], \tag{5.16}$$

$$\begin{aligned} \sigma_z = \frac{RA}{mT}\Bigg\{&\Bigg[\sum_{i=1}^{m}\sum_{j=1}^{m}\int_0^{T/2}\int_0^{T/2} dt_1 \cdot dt_2 \,\overline{\left\{n[t_1+(i-1)T-\theta] - n\left[t_1+\frac{(2i-1)}{2}T-\theta\right]\right\}} \\ &\overline{\times\left\{n[t_2+(j-1)T-\theta] - n\left[t_2+\frac{(2j-1)}{2}T-\theta\right]\right\}}\Bigg] \\ &- \Bigg[\sum_{i=1}^{m}\int_0^{T/2} dt\,\overline{\left\{n[t+(i-1)T-\theta] - n\left[t+\frac{(2i-1)}{2}T-\theta\right]\right\}}^{2}\Bigg]\Bigg\}^{1/2}. \end{aligned} \tag{5.17}$$

It can be expected that the time average of the noise $n(t)$ will vanish. Thus, letting $\overline{n(t)}=0$ in Eqs. (5.16) and (5.17), we obtain

$$\bar{z}=R^2, \tag{5.18}$$

$$\sigma_z=\frac{RA}{mT}\Big[\sum_{i=1}^{m}\sum_{j=1}^{m}\int_0^{T/2}\int_0^{T/2}dt_1\cdot dt_2\,\overline{\Big\{n[t_1+(i-1)T-\theta]-n\Big[t_1+\frac{(2i-1)}{2}T-\theta\Big]\Big\}}$$
$$\overline{\times\Big\{n[t_2+(j-1)T-\theta]-n\Big[t_2+\frac{(2j-1)}{2}T-\theta\Big]\Big\}}\Big]^{1/2}. \tag{5.19}$$

The cross-correlation functions of noise in the above expression are related to the power spectrum of the noise, $\Phi(\omega)$, as follows:

$$\left.\begin{aligned}
&\overline{n\{t_1+(i-1)T-\theta\}\cdot n\{t_2+(j-1)T-\theta\}}=\int_0^\infty \Phi(\omega)\cos\omega\{t_1-t_2+(i-j)T\}d\omega,\\
&\overline{n\Big\{t_1+\frac{(2i-1)}{2}T-\theta\Big\}\cdot n\{t_2+(j-1)T-\theta\}}=\int_0^\infty \Phi(\omega)\cos\omega\Big\{t_1-t_2+\Big(i-j+\frac{1}{2}\Big)T\Big\}d\omega,\\
&\overline{n\{t_1+(i-1)T-\theta\}\cdot n\Big\{t_2+\frac{(2j-1)}{2}T-\theta\Big\}}=\int_0^\infty \Phi(\omega)\cos\omega\Big\{t_1-t_2+\Big(i-j-\frac{1}{2}\Big)T\Big\}d\omega,\\
&\overline{n\Big\{t_1+\frac{(2i-1)}{2}T-\theta\Big\}\cdot n\Big\{t_2+\frac{(2j-1)}{2}T-\theta\Big\}}=\int_0^\infty \Phi(\omega)\cos\omega\{t_1-t_2+(i-j)T\}d\omega.
\end{aligned}\right\} \tag{5.20}$$

By substituting Eq. (5.20) into Eq. (5.19), and carrying out the integrations with respect to t_1 and t_2, we have

$$\sigma_z=\frac{4AR}{mT}\Big[\sum_{i=1}^{m}\sum_{j=1}^{m}\int_0^\infty \Phi(\omega)\cdot\frac{1}{\omega^2}\cos\omega(i-j)T\cdot\sin^4\Big(\frac{\omega T}{4}\Big)d\omega\Big]^{1/2}=ARg_m, \tag{5.21}$$

where $$g_m=\frac{4}{mT}\Big[\sum_{i=1}^{m}\sum_{j=1}^{m}\int_0^\infty \Phi(\omega)\cdot\frac{1}{\omega^2}\cos\omega(i-j)T\cdot\sin^4\Big(\frac{\omega T}{4}\Big)d\omega\Big]^{1/2}.$$

Now, the statistical variable z is expected to be a stationary random signal. If the probability distribution of z is assumed to be Gaussian, the probability density function of z is given as

$$f(z)=\frac{1}{\sqrt{2\pi}\cdot\sigma_z}e^{-((z-\bar{z})/\sigma_z)^2/2}. \tag{5.22}$$

Then, the probability that $z<0$ is simply given by

$$\varepsilon_z=\int_{-\infty}^0 f(z)dz. \tag{5.23}$$

By the substitution of Eq. (5.22) into Eq. (5.23), the probability ε_z is computed as follows:

$$\varepsilon_z=\frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^{\bar{z}/\sigma_z}e^{-z^2/2}dz. \tag{5.24}$$

Inspection of Eq. (5.24) shows that ε_z is a function of the quotient $\bar{z}/\sigma_z$ only as given in Fig. 5.4.

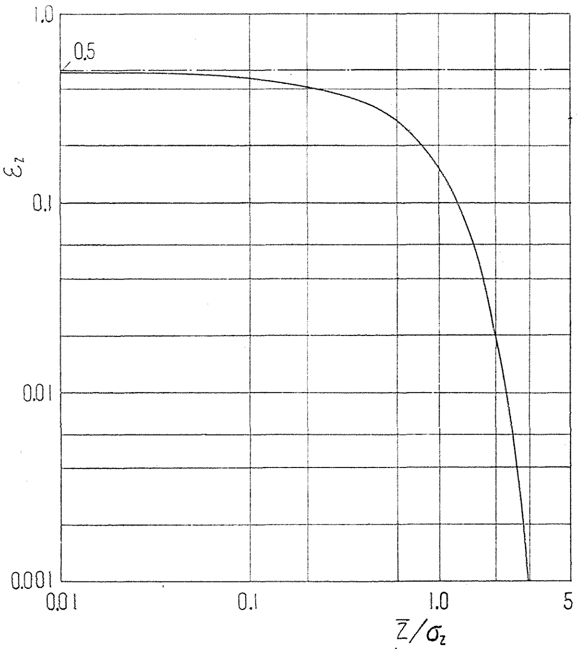


FIG. 5.4. Probability that z is less than zero, when probability distribution of the statistical variable z is Gaussian.

In order to determine the limited value of R when the probable error is assigned and the statistical properties of the noise are known, we derive the following expression from Eqs. (5.8) and (5.21).

$$\frac{R}{kAE^2} = \left(\frac{\bar{z}}{\sigma_z}\right) \cdot g'_m \cdot H'_s \cdot \frac{\sqrt{\phi(0)}}{H_s}, \tag{5.25}$$

where $g'_m = g_m/\sqrt{\phi(0)}$ and $H'_s = H_s/kE^2$.

The limited value of R for a given value of the probable error can be calculated from Eq. (5.25) when the following quantities are known: the amplitudes A and E of the correlate signal and the test signal respectively, the dynamics of the controlled system, and the power spectrum $\Phi(\omega)$, or auto-correlation function, of the noise.

For example, if $\phi(t) = \phi(0)e^{-|t|/\tau}$, $m = 1$ and $G_i(j\omega) = 1/(1 + j\omega\tau_i)$, then $g'_1 = \frac{\sqrt{2}}{2\pi/\omega_i\tau}(2\pi/\omega_i\tau + 4e^{-\pi/\omega_i\tau} - 3 - e^{-2\pi/\omega_i\tau})^{1/2}$ and $H'_s = 1/2\{1 + 4\pi^2(\tau_i/T)^2\}$, where the test signal is assumed to be a sinusoidal wave. The quotient $\bar{z}/\sigma_z$ is obtained from Fig. 5.4 when the value of ε_z is assigned. For $\varepsilon_z = 0.5$, 1 and 5%, the dimensionless limited value R/kE^2A as a function of the dimensionless noise level $\sqrt{\phi(0)}/H_s$ is shown in Fig. 5.5, with m, the number of input cycles over which the correlation process is carried out, as a parameter.

From this figure, it is observed that the limited value of R is linearly related to the noise level; and that with increase in m the limited value R decreases, and thus the controlled input is less likely to move in the wrong direction; and that when a larger value is assigned for the probable error, the limited value of R is allowed to be smaller, as expected.

3. 2. The case where the correlate signal is a sinusoid

Through a similar procedure to the previous case, we can determine the limited value of R for the case where the correlate signal is a sinusoid. In this case, however, we can attempt to derive the variance σ_z^2 from the autocorrelation function of the noise instead of its power spectrum.

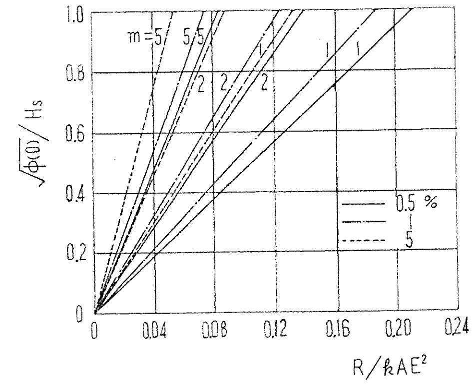


FIG. 5. 5. Diagram for giving the limited value of R for $\omega_0\tau=1$, $\tau_i/T=0.2$ and $\phi(\tau)=\phi(0)e^{-|t|/\tau}$, when the correlate signal is a rectangular wave and the test signal is a sinusoid.

The expressions corresponding to Eqs. (5.4) and (5.20) can be written as

$$R_n = \frac{A'}{mT}\int_0^{mT} n(t)\cdot \sin \omega_0(t+\theta)dt, \tag{5.26}$$

$$\sigma_z = \frac{RA'}{mT}\left[\int_0^{mT}\int_0^{mT} dt_1\cdot dt_2\cdot \overline{n(t_1)n(t_2)}\sin \omega_i(t_1+\theta)\sin \omega_i(t_2+\theta)\right]^{1/2}, \tag{5.27}$$

where A' denotes the amplitude of the sinusoidal correlate signal. Since $\overline{n(t_1)n(t_2)}$ is defined as $\phi(t_1-t_2)$, we can rewrite Eq. (5.27) as follows:

$$\sigma_z = \frac{RA'}{mT}\left[\int_0^{mT}\int_0^{mT} dt_1\cdot dt_2\cdot \phi(t_1-t_2)\sin \omega_i(t_1+\theta)\sin \omega_i(t_2+\theta)\right]^{1/2}. \tag{5.28}$$

For example, if $\phi(t)=\phi(0)e^{-|t|/\tau}$ in Eq. (5.28), then g_m, defined by Eq. (5.21), is calculated as

$$g_m = \frac{\sqrt{\phi(0)}}{1+(\omega_i\tau)^2}\left[\frac{\omega_i\tau}{2\pi m}\left\{1+(\omega_i\tau)^2-\frac{1}{m\pi}\left(\frac{1}{\omega_i\tau}\right)(\sin^2\omega_i\theta-(\omega_i\tau)^2\cos^2\omega_i\theta)(1-e^{-2\pi m/\omega_i\tau})\right\}\right]^{1/2}. \tag{5.29}$$

Similarly to the previous case, Fig. 5.6 gives the relation of R/kE^2A' and $\sqrt{\phi(0)}/H_s$, where the test signal is also a sinusoidal wave.

When the limited value of R in Fig. 5.6 is compared with that in Fig. 5.5 under the condition that the effective value of the rectangular wave is equal to that of the sinusoidal wave, i.e., $A'=\sqrt{\frac{2}{3}}A$; then it is indicated that the latter is about 1.7 times as large as the former irrespective of the values of m and the assigned probable error. This difference is considered due to the correlated value between the noise and the higher-order components of the rectangular wave.

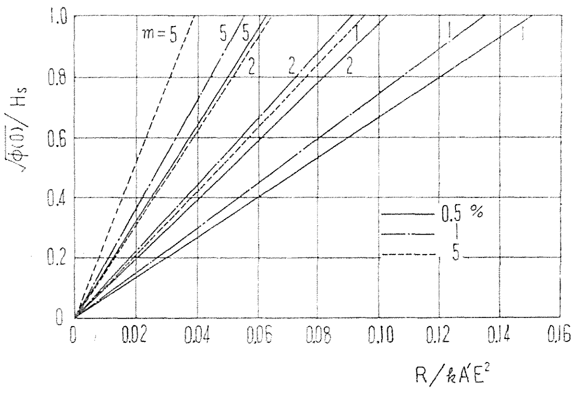


FIG. 5.6. Diagram for giving the limited value of R for $\omega_0\tau=1$, $\tau_i/T=0.2$ and $\phi(\tau)=\phi(0)e^{-|t|/\tau}$, when both the correlate signal and the test signal are sinusoidal waves.

3.3. *The case where the correlate signal is random*

When the correlate signal is a random signal, the expression corresponding to Eq. (5.14) is given by

$$R_n=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}n_1(t)\cdot n_2(t)dt=\overline{n_1(t)n_2(t)}=\phi_{12}(0), \tag{5.30}$$

where $n_1(t)$ and $n_2(t)$ denote the correlate signal and the noise, respectively.

In general, it is expected that there is no correlation between $n_1(t)$ and $n_2(t)$; then $\phi_{12}(0)$ reduces to zero. In practical applications, however, we can not carry out the correlation process over an infinite time interval, as shown in Eq. (5.30). This fact is applied to the computation of R_s, where the time average of $n_1(t)$ is assumed to be zero. When the time average of $n_1(t)$ is taken over the time interval from 0 to T, it can be proved that in order to make its standard deviation from the mean value $(=0)$ less than 10%, we must choose T more than $200\,\tau_{11}$, where the auto-correlation function of the random signal is assumed to be $\phi_{11}(0)e^{-|t|/\tau_{11}}$. Accordingly, under the assumption that the correlation process is carried out over a period larger than $200\,\tau_{11}$, we proceed with the discussion as follows. Equation (5.30) must be rewritten as

$$R_n=\frac{1}{T}\int_0^T n_1(t)\cdot n_2(t)dt. \tag{5.31}$$

The noise $n_2(t)$ is assumed to be a stationary stochastic signal, and its time average $\overline{n_2(t)}$ is expected to vanish. Then, the mean value $\bar{z}$ and variance σ_z^2 of the statistical variable z are given by

$$\left.\begin{aligned}&\bar{z}=R^2,\\&\sigma_z^2=\frac{R}{T}\Big[\int_0^T\int_0^T dt_1\cdot dt_2\,\overline{n_1(t_1)\,n_1(t_2)\,n_2(t_1)\,n_2(t_2)}\,\Big]\end{aligned}\right\} \tag{5.32}$$

By the definition that $\overline{n_1(t_1)n_1(t_2)}=\phi_{11}(t_1-t_2)$ and $\overline{n_2(t_1)\,n_2(t_2)}=\phi_{22}(t_1-t_2)$, the auto-correlation function of the noise $n_2(t)$, Eq. (5.32) can also be written as

$$\left.\begin{aligned} \bar{z} &= R^2, \\ \sigma_z &= \frac{R}{T}\left[\int_0^T\int_0^T dt_1 dt_2 \phi_{11}(t_1-t_2)\cdot\phi_{22}(t_1-t_2)\right]^{1/2} \\ &= \frac{R}{T}\left[2\int_0^T (T-t)\phi_{11}(t)\phi_{22}(t)dt\right]^{1/2}. \end{aligned}\right\} \tag{5.33}$$

Similarly to the previous case, the expression corresponding to Eq. (5.25) can be derived as follows:

$$\frac{R}{k\{\phi_{11}(0)\}^{3/2}} = \left(\frac{\bar{z}}{\sigma_z}\right)\cdot g_m \cdot H'_n \cdot \frac{\sqrt{\phi_{22}(0)}}{H_n}. \tag{5.34}$$

For example, choose that $\phi_{11}(t) = \phi_{11}(0)e^{-|t|/\tau_{11}}$, and $\phi_{22}(t) = \phi_{22}(0)e^{-|t|/\tau_{22}}$, then $g_m = \{2b(1-b+be^{-1/b})\}^{1/2}$ and $H'_n = H_n/k\phi_{11}(0)$, where $b = \dfrac{\tau_{11}\tau_{22}}{T(\tau_{11}+\tau_{22})}$. Figure 5.7 illustrates the relation of Eq. (5.34) for $\tau_{22}/\tau_{11}=1$ and $\tau_i/\tau_{11}=100$. From this figure, it is also observed that the limited value of R decreases with increase in $m'(=T/\tau_{11})$. In this case, we can expect that the effects of the noise interference is eliminated, since m' is chosen sufficiently large such that Eq. (5.11) holds, or the correlation process is carried out over a sufficiently long time interval.

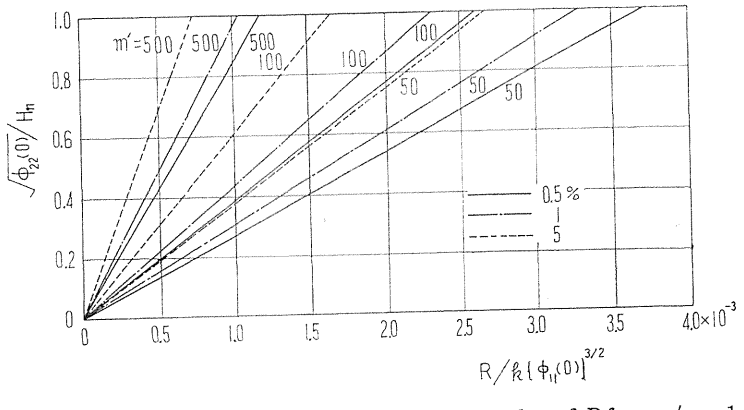


FIG. 5.7. Diagram for giving the limited value of R for $\tau_{22}/\tau_{11}=1$, $\tau_i/\tau_{11}=100$, $\phi_{11}(\tau)=\phi_{11}(0)e^{-|t|/\tau_{11}}$ and $\phi_{22}(\tau)=\phi_{22}(0)e^{-|t|/\tau_{22}}$, when both the correlate signal and the test signal are random.

Conclusions

In regard to the optimalizing control system with one input, various practical control methods belonging to a class of the so-called "direct method" have been proposed and then analyzed in detail in its theoretical aspect. As the result, many data and clearer insight useful for the design of such systems have been obtained.

The control methods proposed in this paper are also applicable to the optimalizing system with many inputs. For example, in the cut-and-try method and the peak-holding method, if each input is optimalized in turn, an extreme value as a function of many inputs will be found after all inputs are scanned; and in the cross-correlation method if as many number of test signals as that of the inputs are available and if none of their frequency is a round multiple of others, then the deviation of each input from its optimum value will be determined simultaneously.

The merit or demerit of these control methods can be decided considering such factors as the difficulty in the construction of the controller, the quickness of the response of the system, and the susceptibility to the noise interference. There is, however, no absolutely best control method, since each method has both good points and weak points. Therefore, the ultimate control method would be decided in views of the operating conditions of the particular system, the properties of the disturbance, and the economical demand for the system performance and so on.

The following problems are left for future researches: the extension of this theory to the optimalizing system with many inputs, the practical applications of these presented control methods to the engineering system, and the solution of various problems occuring there.

Acknowledgements

In concluding this paper, the author acknowleges his debt to Prof. Dr. T. Koga of Univ. of Southen California, former Prof. of Nagoya Univ., for introducing him to the present field of study. The author wishes to express his sincere gratitude to Prof. Dr. T. Suga of Nagoya Univ., Prof. Dr. Y. Sawaragi of Kyoto Univ., and Prof. Dr. T. Yamamoto of Nagoya Univ. for their guidance and encouragement, to Ass't. Prof. K. Nakamura, Ass't. Prof. I. Sugiura and Mr. S. Minagawa of Automatic Control Lab. of Nagoya Univ. for their invaluable criticisms. He also thanks Mr. T. Naruse for his assistance in the final compilation of this paper.

This research is indebted to the Automatic Control Laboratory, Faculty of Engineering, Nagoya University for use of the laboratory facilities.

Finally, the author expresses thanks for Ministry of Education of Japan for giving him the Fund for Research.

References

1) C. S. Draper and Y. T. Li, "Principles of Optimalizing Control Systems and an Application to the Internal Combustion Engine", ASME Publication, Sept. (1951).
2) H. S. Tsien, Engineering Cybernetics, McGraw-Hill (1954), pp. 214–230.
3) H. S. Tsien and S. Serdengecti, "Analysis of Peak-Holding Optimalizing Control", J. Aeron. Sci., Vol. **22**, No. 8 (1955), pp. 561–570.
4) R. L. Cosgriff, "Servos that Use Logic Can Optimize", Control Engng., Vol. **2**, Sept. (1955), pp. 135–135.
5) R. L. Cosgriff and R. A. Emerling, "Optimizing Control Systems", Appl. and Ind. AIEE., No. **35**, March (1958), pp. 13–16.
6) R. Jr. Shull, "An Automatic Cruise Control Computor for Long Range Aircraft", Trans. IRE., Professional Group on Electronic Computers, Dec. (1952), pp. 47-51.
7) Vasu George, "Experiments with Optimalizing Controls Applied to Rapid Control of Engine Pressures with High Amplitude Noise Signal", Trans. ASME, April (1957), pp. 481–488.

8) R. Hooke and P. L. Van Nice, "Optimizing Control by Automatic Experimentation", ISA Journal, Vol. **6**, No. 7 (1959), pp. 78–79.

9) S. Fujii, "Some General Considerations for Various Modes of Optimalizing Control", Seigyokogaku (in Japanese), Vol. **5**, No. 1 (1961), pp. 35–45.

10) S. Fujii, "Optimalizing Control by a Cut-and-Try Method Applied to One Variable System", Trans. of Mech. Engrs. (in Japanese), Vol. **27**, No. 176 (1961), pp. 530–542.

11) S. Fujii, "Effects of Process Dynamics on a Cut-and-Try method in the Optimalizing Control System", Proc. of the 11th Japan National Congress for App. Mech., (1961).

12) S. Fujii, "Analysis of Ramp Type Optimalizing Control", Seigyokogaku (in Japanese), Vol. **4**, No. 5 (1960), pp. 299–307.

13) S. Fujii and T. Koga, Trans. of Soc. of Mech. Engrs. (in Japanese), Vol. **26**, No. 161 (1960), pp. 129–138.