COWLES COMMISSION DISCUSSION PAPER: ECONOMICS NO. 288

AN EXPLORATION INTO THE USE OF SERVOMECHANISM THEORY IN THE STUDY OF PRODUCTION CONTROL

by Herbert A. Simon

August 17, 1950

AN EXPLORATION INTO THE USE OF SERVOMECHANISM THEORY IN THE STUDY OF PRODUCTION CONTROL

By Herbert A. Simon

August 17, 1950

This paper is of an exploratory character. Powerful, and extremely general, techniques have been developed in the past decade for the analysis of electrical and mechanical control systems and servomechanisms. There are obvious analogies between such systems and the human systems, usually called production control systems, that are used to plan and schedule production in business concerns. The depth or superficiality of these analogies can be tested by subjecting a fairly simple, but relatively concrete, example of a production control system to the techniques of analysis usually employed for servomechanisms.

It might be pointed out that the notion of a servo-mechanism incorporating human links is by no means novel. In particular, many gun-sighting servos involve such a link. The idea of social, as distinguished from purely physiological, links is relatively new. However, Richard M. Goodwin has independently arrived at the same idea as a means for studying price mechanisms and market behavior. All such systems would be included in Wiener's general program for cybernetics.

1. Introduction

Some preliminary remarks are necessary, first to characterize servomechanism theory, and second, to describe the production control system we will study.

1.1. Servomechanism Theory

Many of the systems, with which electrical and mechanical engineering deal, are described, at least approximately, by systems of linear differential or integro-differential equations with constant coefficients. Included are electrical networks with lumped constants, and included among these (or their mechanical analogues) are many of the systems known as "controllers," "regulators", and "servomechanisms." We will not attempt here to distinguish these terms precisely, but instead to set forth an example of such a system.

Consider a system consisting of a house, or other enclosed space, a gas furnace, and a thermostat that controls the rate of gas flow in the furnace.[1]

1. In order to preserve as close an analogy as possible with the production control system to be described later, we will assume a thermostat that is continuous in operation, instead of the more familiar on-off thermostat. The system described here is analysed in reference [4] (see bibliography at end of paper), pp. 298-303.

The temperature setting of the thermostat (that is, the desired house temperature) will be referred to as the <u>input</u> of the system, and designated by θ_I; the actual house temperature will be referred to as the <u>output</u> (θ_O); their difference ($\theta_I - \theta_O$) as the <u>error</u> ($\mathcal{E}$); and the outside temperature as the load (θ_L).[2] All variables are functions of time.

2. In our example, the input is generally fixed, while the load is variable. This is the typical case of the controller; and the input is frequently referred to as the <u>standard</u>. The term "servomechanism" more commonly refers to a system in which the input is variable and the variable load absent. However, in many important engineering systems both variable input and variable load are present.

The system is so constructed that the rate of gas flow in the furnace, and hence the rate at which heat is supplied to the house, is a function of the error (in the simplest case proportional to the error). Further, this function relating the error to the output is selected so that the error will tend to be reduced, whatever the load imposed on the system.

The system just described is shown in Figure 1. The equations that describe the system are:

(1) $$\frac{d\theta_o}{dt} = f(\varepsilon) - k(\theta_o - \theta_L)$$

(1.2) $$\varepsilon = \theta_I - \theta_o$$

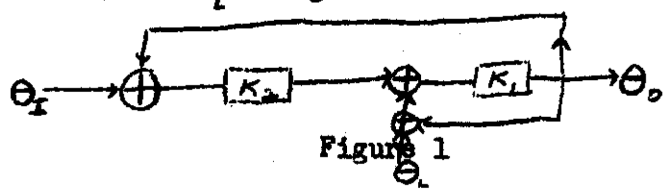

Figure 1

The symbols $\oplus$ indicate differential devices (subtraction); The box K_2 corresponds to $f(\varepsilon)$; and the box K_1 to the integration (from equation 1) which gives θ_o as a function of $f(\varepsilon)$ and $(\theta_o - \theta_L)$.

Two important features of this system should be noted. The first is the control loop or feedback loop (the upper loop in Figure 1), by means of which (a) the output is compared with the input and (b) their difference is fed back into the system to alter the output in the direction of reducing the difference.

The second important feature is shown by the directional arrows. The input and load affect the behavior of the system (and, in particular the output and error), but are themselves unaffected by it. Hence, variables not included in the loop may be regarded as independent, and may be assumed to have any arbitrary time paths. This kind of relationship is sometimes referred to as unilateral coupling, or cascading. A reciprocal relation must be represented by a closed loop (as the lower loop in Figure 1).

In a physical servomechanism, cascading is made possible by the fact that the closed portion of the system involves very little energy in comparison with the energy of the independent variables (as in a solar system with large central sun and small planets), or, more generally, draws its energy from an independent power source (as in an amplifier). It is this characteristic of the system which permits the output to follow the input without disturbing the path of the input. A servomechanism, then, is a system (1) unilaterally coupled to an input and a load, (2) with one or more feedback loops whereby the output is compared with the input, and (3) with a source of energy controlled by the error that tends to bring the output in line with the input. If the load is bilaterally coupled with the output, then the former must be included in the system and cannot be treated as an independent variable.

The most powerful technique for treating servomechanisms employs the Laplace transform. (See [2], Chapter 2, [3]). The Laplace transform of the input may be interpreted as its frequency spectrum (i.e., it is very closely related to the Fourier integral). The Laplace transform of the entire servo-system connecting input with output describes its behavior in filtering (altering amplitude and phase) the frequencies in the spectrum. The Laplace transform of the output, which is the product of the two previous transforms, is the frequency spectrum of the output. (In this statement we disregard the load, which enters as another input.) The system is studied by determining the Laplace transform of the servo, multiplying this by various input transforms, and analysing the resulting behaviors of the output. We are interested in the stability of the output (which is related to its transient response), and its steady-state behavior for various inputs. By defining a criterion function (a function of the output) we may compare the merits of alternative servos for controlling an output under specified conditions.

1.2. Production Control

In this paper we shall consider the control of the rate of production of a single item. The item is supposed to be manufactured to standard specifications, placed in stock, and shipped out on order of customers. The item is manufactured continuously, and control consists in issuing instructions that vary the quantity to be manufactured per day (or other unit of time).

The aim of the control system is to minimize the cost of manufacture over a period of time. This cost, or the variable part of it, is assumed to depend upon (1) the variations in the manufacturing rate (i.e., it costs more to make 1000 items if the manufacturing rate fluctuates than if it is constant); and (2) the inventory of finished goods (i.e., increase in this inventory involves interest costs, decrease in the inventory below a certain point involves delay in filling customers' orders). Hence the criterion by which we will judge the system will be some function of the magnitudes of the fluctuations in manufacturing rate and in inventory of finished goods.

We will take as our input the optimum inventory (θ_I). Since this will be assumed constant throughout our problem, it may be taken as zero. The actual inventory of finished goods will be taken as the output (θ_O). The error ($\mathcal{E}$) will then be the deficiency (positive or negative) of inventory ($\theta_I - \theta_O$). Customers' orders will be treated as the load (θ_L). We need two additional variables, the actual production rate (μ), and the rate of planned production or, more accurately, planned new production (η).

Assume that, on the basis of information about orders and the inventory excess or deficiency, instructions are issued daily (in our model, continuously) for the manufacture of a certain number of units. At some later date, the lag being determined by the time required for production, these units are actually produced and added to inventory. Meanwhile, customers' orders have been daily (continuously) withdrawn from inventory. Information regarding the inventory level is in turn fed back to be compared daily (continuously) with the optimum inventory, and the calculated error employed, in turn, to redetermine the planned production rate.

This system obviously possesses the characteristics of a servomechanism. It is unilaterally coupled to the load and input (customers' orders and optimum inventory). It has a feed-back loop: error→planned production→actual production→inventory→error. The error initiates a change in planned production in such a direction as to reduce the error.

In succeeding sections systems will be described for accomplishing the functions just listed. We will start with some highly simplified structures, and add complications as we proceed.

2. A Simple System for Inventory Control

We will consider a series of three systems. In the first we shall be concerned only with the control of inventories; we will base production decisions only on information about inventories (and will ignore information about orders); and we will assume a zero time-lag for production. In the second system, the second restriction will be removed. In the third system, all three restrictions will be removed.

2.1 Description of the System

The first system is shown in Figure 2, which is identical with Figure 1 except for the absence of the lower loop that appears in the former. In

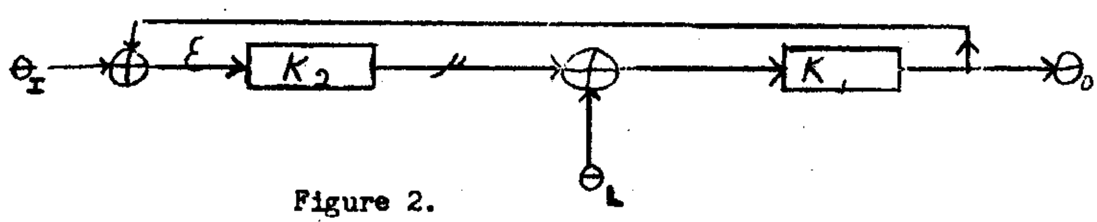

Figure 2.

this system, it is assumed that μ is identically equal to η. That is, the rate of production at time t is equal to the rate at which new production is scheduled at time t. This implies that production plans are carried out without an appreciable time lag. The equations of the system are:

(2.1) $\Theta_o(t) = K_1[\mu(t) - \theta_L(t)]$
(2.2) $\mu(t) = K_2[\varepsilon(t)]$
(2.3) $\varepsilon(t) = \Theta_I(t) - \Theta_o(t)$

K_1 and K_2 are operators whose form will be specified. Equation (2.3) is a definition. Equation (2.2) represents a rule of decision--it specifies the rate of production that will be scheduled (and achieved) as a function of the excess and deficiency of inventory. The precise form of (2.1) is determined by the conditions of the problem since, by definition:

(2.4) $$\frac{d\theta_o}{dt} = \mu - \theta_L$$

Hence, if we wish to design a servomechanism of the class described that meets some criterion of optimality, we have at our disposal only the operator K_2 --the decision rule.

Our equations (2.1)- (2.4) can be restated in terms of the Laplace transforms of the quantities involved. The real variable t is replaced by the complex variable p. The Laplace transform of $y(t)$, written $\mathcal{L}[y(t)]=y(p)$ is defined by:

(2.5) $$y(p) = \int_0^\infty y(t)e^{-pt}dt$$

This integral exists for a wide class of functions, although in some cases it must be defined as an improper Lebesque integral rather than a Riemann integral. The inverse transformation is:

(2.6) $$y(t) = \frac{1}{2\pi}\int_{b-j\infty}^{b+j\infty} y(p)\, e^{pt}dp$$

where the path of integration runs parallel to the imaginary axis to the right of some real constant b.[3] The operator K_1 corresponding to (2.4) is $\frac{1}{p}$.

3. No attempt will be made in this paper at mathematical rigor. Virtually all the mathematical tools employed here will be found in [2], Chapter 2; and [3] The latter also contains (pp.332-357) a very useful table of Laplace transform pairs.

We then have:

(2.7) $\theta_0(p) = \frac{1}{p}(\mu(p) - \theta_L(p))$

(2.8) $\mu(p) = K_2(p) \cdot \mathcal{E}(p)$

(2.9) $\mathcal{E}(p) = \theta_I(p) - \theta_0(p)$

where $\theta_0(p)$, etc., represent the Laplace transforms of $\theta_0(t)$, etc. respectively. If now we assume $\theta_I \equiv 0$ and introduce the system transform

(2.10) $Y(p) = \theta_0(p)/\theta_L(p)$

we derive from (2.5) - (2.7)

(2.11) $Y(p) = \dfrac{-1/p}{1+\frac{K_2}{p}} = \dfrac{-1}{p+K_2(p)}$

2.2 Theorems About the Laplace Transform

The behavior of the system under varying load can be discussed in terms of the properties of the system transform, $Y(p)$. As a basis for this discussion we will outline some results from Laplace transform theory without attempting proofs or complete rigor in their statement.

A system will be termed stable if the output remains bounded (in the t domain) provided the input remains bounded (in the t domain). The equation obtained by setting the denominator of $Y(p)$ equal to zero we call the characteristic equation of $Y(p)$. In our particular system we have:

(2.12) $p + K_2(p) = 0$

Provided that the numerator of $Y(p)$ has no finite poles, the system will be stable if and only if all the roots of the characteristic equation have negative real parts.

Let $W(t)$ be the inverse transform of $Y(p)$, as defined in (2.6). We call $W(t)$ the weighting function of the system. Multiplication in the p domain corresponds to convolution in the t domain. Hence, we have from (2.10):

(2.13) $\theta_0(t) = \int_0^{t} W(\tau)\,\theta_L(t-\tau)\,d\tau$

Equation (2.13) relates the time path of the output to the weighting function of the system and the time path of the load. Generally, however, we do not employ this relationship. Instead, we multiply the Laplace transform of $\theta_L(t)$ by the system transform and then take the inverse transform of this product, obtaining $\theta(t)$ directly. Indeed, it is this procedure, together with the availability of tables of transform pairs, that makes the Laplace transform method particularly powerful.

Two additional theorems, which hold when the indicated limits exist, will prove useful:

(2.14) $\lim_{t \to \infty} y(t) = \lim_{p \to 0} p\,y(p)$

(2.15) $\lim_{t \to 0} y(t) = \lim_{p \to \infty} p\,y(p)$

In particular, (2.14) enables us to calculate immediately the steady state output for given load and system transform without transforming again to the t domain.

(2.3) Steady-State and Transient Behavior.

We return now to the task of discovering for our particular system a system transform, $K_2(p)$, that will induce appropriate behavior of $\theta_\delta(t)$. By "appropriate" behavior we mean that we wish $\theta_\delta(t)$ to be as small as possible. We consider first the steady state behavior which we will study by means of (2.14).

(2.16) $\lim_{t \to \infty} \theta_\delta(t) = \lim_{p \to 0} p\,\theta_\delta(p) = \lim_{p \to 0} - \frac{p\,\theta_L(p)}{p + K_2(p)}$

A. Suppose that up to the time t = o, orders have been zero, and that after that time they are received at the rate of 1 order per unit of time:

(2.17) $\theta_L(t) = 0$ for $t < 0$; $\theta_L(t) = 1$ for $t \geq 0$

We have:

(2.18) $\mathcal{L}[\theta_L(t)] = \theta_L(p) = \frac{1}{p}$

Hence:

(2.19) $\lim_{t\to\infty} \theta_0(t) = \lim_{p\to 0} -\frac{1}{p+K_2(p)}$

Therefore we wish the denominator of the right hand side of (2.19) to become very large as p approaches zero. This can be accomplished, for example, by setting:

(2.20) $K_2 = \frac{1}{p^k}(a+bp)$ with $k \geq 1,\ a>0,\ b>0$

Rapid convergence is assured by making a large.

B. Suppose now, that up to the time t =o, orders have been zero, and that after that time they are received at the rate of t^n orders per unit of time:

(2.21) $\theta_L(t) = 0$ for $t<0$; $\theta_L(t) = t^n$ for $t \geq 0$.

We have:

(2.22) $\mathcal{L}[\theta_L(t)] = \vartheta_L(p) = \frac{n!}{p^{n+1}}$

In this case we can assure a zero steady-state error with K_2 of the same form as (2.20), but with $k \geq (n+1)$

C. Suppose that $\theta_L(t)$ is sinusoidal:

(2.21) $\vartheta_L(t) = 0$ for $t<0$, $\theta_L(t) = \sin \omega t$ for $t \geq 0$

(2.22) $\theta_L(t) = A[e^{i\omega t} + e^{-i\omega t}] = A \cos \omega t$

Here we cannot use the method of the previous two cases since it can be shown that $\lim_{t\to\infty} \theta_0(t)$ does not exist. Instead we use the result that if θ_L (t) is sinusoidal, θ_0 (t) will be sinusoidal, with the same frequency but altered amplitude and phase. That is,

(2.23) $\lim_{t\to\infty} \theta_0(t) = B\cos(\omega t + \psi)$

The amplitude, B, of the output is given by:

(2.24) $B = A[Y(i\omega)\, Y(-i\omega)]^{\frac{1}{2}}$

If, for example, $K_2 = \frac{a}{p} + b$, we have

$$(2.25)\quad Y(p) = \frac{-1}{p + K_2(p)} = \frac{-p}{p^2 + bp + a} = \frac{-p}{(p-p_1)(p-p_2)}$$

where p_1 and p_2 are the roots of the characteristic equation.

$$(2.26)\quad Y(i\omega)\,Y(-i\omega) = \frac{-i\omega}{(i\omega - p_1)(i\omega - p_2)} \cdot \frac{i\omega}{(-i\omega - p_1)(-i\omega - p_2)}$$
$$= \frac{\omega^2}{(\omega^2 + p_1^2)(\omega^2 + p_2^2)}$$

$$(2.27)\quad B = A\omega\left[(\omega^2 + p_1^2)(\omega^2 + p_2^2)\right]^{-\frac{1}{2}}$$

For given p_1, p_2, as ω approaches zero, B approaches zero; as ω grows large, B approaches A/ω^2. When p_1 and p_2 are equal B approaches its maximum for $\omega = p_1$. This maximum is $B = A/4p_1^2$. Hence we see that by selecting K_2 so that the characteristic function has large roots we guarantee rapid damping of θ_o for sinusoidal loads.

We have now indicated the properties that our decision rules (the operator K_2) must possess to assure small or vanishing steady-state inventory excesses and deficiencies for various loads.

Next, let us interpret these results in the t domain. Suppose that $Y(p)$ is an algebraic expression:

$$(2.28)\quad Y(p) = \frac{b_m p^m + b_{m-1} p^{m-1} + \ldots + b_o}{a_n p^n + a_{n-1} p^{n-1} + \ldots + a_o}$$

Then the equation in the t domain obtained by transforming (2.10) is:

$$(2.29)\quad a_n \frac{d^n \theta_o}{dt_n} + a_{n-1} \frac{d^{n-1}\theta_o}{dt_{n-1}} + \ldots a_o\theta_o = b_m \frac{d^m \theta_L}{dt^m} + b_{m-1}\frac{d^{m-1}\theta_L}{dt^{m-1}} + \ldots + b_o\theta_L$$

If, for example, $Y(p)$ is defined by (2.25) we have:

$$(2.30)\quad \frac{d^2\theta_o}{dt^2} + b\frac{d\theta_o}{dt} + a\,\theta_o = \frac{-d\,\theta_L}{dt}$$

From this equation we can verify the conclusions already reached above. For example, in A, we had $\frac{d\theta_L}{dt} = 0$ for $t \geq 0$. Here the general solution of (2.30) is

$$(2.31) \quad \theta_o = M e^{p_1 t} + N e^{p_2 t}$$

where p_1 and p_2 are the roots of the characteristic equation.

$$(2.32) \quad p = \frac{-b \pm \sqrt{b^2 - 4a}}{2}$$

Since $a > 0$, $b > 0$, p_1 and p_2 will be real and negative or complex with negative real part, and hence (2.31) will convey to zero as t increases.

If, on the contrary, we had $\frac{d\theta_L}{dt} = 1$ for $t \geq 0$, the general solution of (2.30) would be

$$(2.33) \quad \theta_o = M e^{p_1 t} + N e^{p_2 t} + \frac{1}{a}$$

Then the system transform (2.25) would yield a steady-state error of $\frac{1}{a}$ as $t \to \infty$. Again this result can be obtained directly by substitution of (2.25) and $\theta_L(p) = \frac{1}{p^2}$ in (2.16). Moreover the transient part of θ_o will be rapidly damped if the negative real parts of p_1 and p_2 are large.

Returning to the case of a sinusoidal load,

$$(2.34) \quad \theta_L(p) = \mathcal{L}[\cos \omega t] = \frac{\omega}{p^2 + \omega^2}$$

with the system function of (2.25), we have:

$$(2.35) \quad \theta_o(p) = \frac{-p\omega}{(p-p_1)(p-p_2)(p^2+\omega^2)}$$

$$= \frac{A}{(p-p_1)} + \frac{B}{(p-p_2)} + \frac{Cp+D}{p^2+\omega^2}$$

The transform of this is

$$(2.36) \quad \theta_o(t) = A e^{p_1 t} + B e^{p_2 t} + E \cos(\omega t + \psi)$$

The final term of (2.36) we have already encountered in (2.23)-- it is the steady-state response to the sinusoidal load. The first two terms represent the transient response which, again, will be rapidly damped if p_1 and p_2 have large negative real parts.

2.4 Stability of the System

We stated in Section (2.2) that a system will be stable if the roots of the characteristic equation of the system transform have negative real parts. In Section (2.3) we noticed that the transient response of the system is independent of the load, and is determined by the roots of the characteristic equation. If the roots have large negative real parts the transient will be strongly damped. These results suggest that many properties of the system can be determined directly by examination of the roots of the characteristic equation. We next carry out this program for various choices of K_2[4].

---

4. We will not employ in this paper some of the procedures, such as Nyquist's rule, widely used in servomechanism analysis to determine whether a system has any roots with positive real parts. For Nyquist's rule, see [2], pp. 67-75; [5], Chapter 5.

A. Let $K_2 = a/p$, with a real. Then $p_0 = \pm\sqrt{a}$, and the system is unstable since at least one of the roots has a non-negative real part.

B. Let $K_2 = \frac{a}{p} + b$, with a and b real. Then $p_0 = \frac{-b \pm \sqrt{b^2 - 4a}}{2}$, and the system is stable if $a > 0$, $b > 0$; otherwise unstable. This result has already appeared from (2.31) and (2.32).

C. Let $K_2 = \frac{a}{p} + b + cp$ with a, b, c real. Then $p_0 = \frac{-b \pm \sqrt{b^2 - 4a(c+1)}}{2(c+1)}$, ~~and~~ the system is stable if a, b, and (c +1) all have the same sign, otherwise unstable.

2.5 Interpretation of the Decision Operator

The operator K_2 represents a rule of decision. Since $\mu(p) = K_2(p) \cdot \varepsilon(p)$, this rule determines, on the basis of information as to the current deficit or excess of inventory $(\varepsilon(p))$, at what rate $(\mu(p))$ manufacture should be carried on. Among the operators that have been examined that possess satisfactory properties is

$K_2 = \frac{a}{p} + b$, with a and b large positive constants. With this operator, equation (2.2) becomes:'

$$(2.37)\quad \frac{d\mu(t)}{dt} = a\varepsilon(t) + b\frac{d\varepsilon(t)}{dt}$$

which interpreted means: the rate of production should be increased or decreased by an amount proportional to the deficiency or excess of inventory plus an amount proportional to the rate at which the inventory is decreasing. The constants of proportionality, a and b, should be large if it is desired to keep the inventory within narrow bounds. The relation $b^2 > 4a$ should be preserved if hunting is to be avoided.

All of this is obvious to common sense. What is perhaps not obvious is that the derivative control (the final term in 2.37) is essential to the stability of the system. To base changes in production rate only upon the size of inventory would introduce undamped fluctuations in the system. [Compare A and B of Section 2.4]

3. Decision Rule Involving Orders and Inventories

The system thus far discussed is unduly restricted in that the decision rule is based entirely on the inventory situation and takes no account of the rate at which new orders are flowing in. The system is easily modified to remedy this defect. Figure 3 shows the modified system.

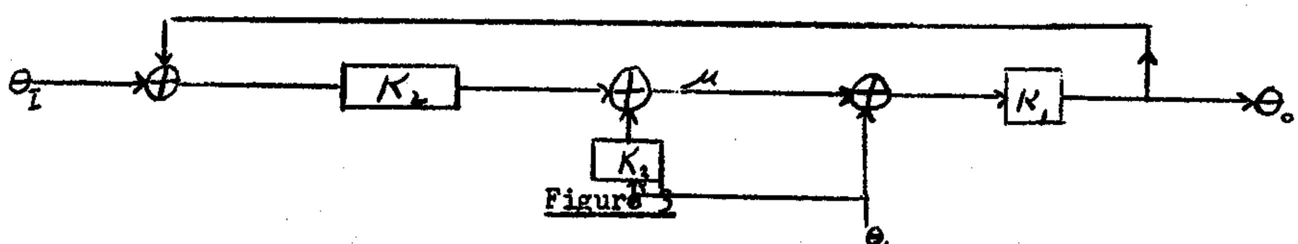

Figure 3

The equations of the system in the p domain are:

$$(3.1)\quad \theta_0(p) = K_1(p)[\mu(p) - \theta_L(p)]$$

$$(3.2)\quad \mu(p) = K_2(p)\varepsilon(p) + K_3(p)\theta_L(p)$$

$$(3.3)\quad \varepsilon(p) = \theta_I(p) - \theta_0(p)$$

Only the second equation, the decision rule, has been modified. The rate of production will now be determined by the inventory level ($\mathcal{E}$) and the rate of new orders (θ_L). As before, we will have $K_1 = 1/p$. Again, without loss of generality we may set $\theta_i = 0$ and derive the system transform:

(3.4) $$Y(p) = \theta_o/\theta_L = \frac{K_3(p)-1}{p+K_2(p)}$$

The system transform of (3.4) has exactly the same characteristic equation as the transform of (2.11). Hence our analysis of the appropriate form for K_2 holds in this system also. Moreover, the performance of the system for any load will be improved if the absolute value of the numerator is made as small as possible. In particular, if $K_3 = 1$, we will have:

(3.5) $\theta_o \equiv 0$ identically in θ_L

The interpretation of this is that if we set the rate of production just equal to the rate of orders, if there is no production lag, and if $\mathcal{E}$ is initially zero, it will remain zero. In this case the problem of control is trivial. The decision rule is simply:

(3.6) $\mu(t) = \theta_L(t)$

4. System With Production Lag

With this preliminary analysis of two simple systems we are ready to study a system that approximates more closely to the problems we would expect to encounter in actual situations. The most important feature that is missing from the previous systems is a production lag. In actual cases a period of time will elapse from the moment instructions are issued to increase the rate of production to the moment when the increased flow of goods is actually produced.

4.1 Description of the System

In Figure 4 is shown a system with a production lag.

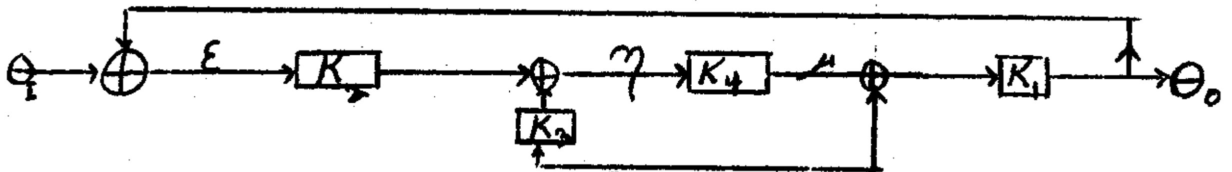

Figure 4.

The equations of this system are:

(4.1) $\theta_O = K_1 (\mu - \theta_L)$

(4.2) $\mu = K_4 \eta$

(4.3) $\eta = K_2 \varepsilon + K_3 \theta_L$

(4.4) $\varepsilon = \theta_I - \theta_O$

The new variable $\eta(t)$ represents the instructions at time t as to the production rate, μ (t) now represents the actual rate of production of finished goods at time t. $\mu(p)$ and $\eta(p)$ are connected (equation 4.2) by the "production lag operator", K_4. As before, we will have by definition $K_1 = 1/p$. The operators K_2 and K_3 again correspond to the decision rule and are at our disposal in seeking an optimal scheduling rule. It remains to find a plausible form for K_4.

The simplest assumption is that the production lag is a fixed time period, τ. That is:

(4.5) $\mu(t) = \eta(t - \tau)$

This means that if a given rate of production is decided upon at a certain time, this rate of production of finished items will be realized τ units of time later. The operator transform K_4(p) corresponding to (4.5) is:

(4.6) $K_4(p) = e^{-\tau p}$

Substituting the known functions K_1(p) and K_4(p) in the system equations, and solving for the system transform, we get:

(4.7) $Y(p) = \theta_O(p)/\theta_L(p) = \dfrac{(e^{-\tau p}K_3 - 1)}{p + e^{-\tau p}K_2}$

A comparison of (4.7) with (3.4) reveals that both numerator and denominator have been affected by the introduction of the production lag. Hence we shall have to re-examine the entire situation.

4.2 Feedback of Information About New Orders

Consideration of the numerator of (4.7) shows that the control of inventory is no longer a trivial problem. If we set $K_3 = 1$ the numerator becomes $(\varepsilon^{-\tau p} - 1)$ which approaches zero only as p approaches zero. Hence this procedure would stabilize the inventory perfectly only for a sinusoidal load of very long (infinite) period. As $p \to \infty (\varepsilon^{-\tau p}) \to -1$, and hence the numerator provides no attenuation for a high-frequency sinusoidal load. At best we can say that the system will perform better with the operator $K_3 = 1$, than without any information about orders, but the improvement in performance will be important only for long period fluctuations in load.

Why not set $K_3 = \varepsilon^{\tau p}$? Then we would have $(\varepsilon^{-\tau p} K_3 - 1) \equiv 0$. Define the variable ϕ so that:

(4.8) $$\phi(p) = \varepsilon^{\tau p} \theta_L(p)$$

Taking the inverse transforms of both sides, we find:

(4.9) $$\phi(t) = \theta_L(t + \tau)$$

Hence, setting $K_3 = \varepsilon^{\tau p}$ corresponds to predicting the value of θ_L for τ units of time in advance of the actual receipt of orders. Again the result is intuitively obvious. If we could predict orders over the time interval τ, we could schedule production in anticipation of the actual receipt of these orders and avoid any inventory fluctuation whatsoever.

Various methods of forecasting $\theta_L(t+\tau)$ are available to us. One procedure is to use the autocorrelation function of $\theta_L(t)$. This is the method devised by Wiener [6] which has been successfully applied to electrical servomechanisms such as radar tracking systems. Another is to expand $\varepsilon^{\tau p}$ into a convergent series:

(4.10) $$e^{\tau p} = 1 + p\tau + \tfrac{1}{2}p^2\tau^2 + \ldots$$

This series will converge rapidly for small values of p --hence can be used to damp out the long-period components of Θ_L. If we retain, for example, the first two terms, (4.7) becomes:

(4.11) $$Y(p) = \frac{p\tau}{p + e^{-\tau p}K_2}$$

The behavior rule corresponding to this form of K_3 is found by transforming:

(4.12) $$\phi(p) = (1 + p\tau)\,\theta_L(p)$$

into

(4.13) $$\phi(t) = \theta_L(t) + \tau\frac{d\theta_L(t)}{dt}$$

This amounts to straight-line extrapolation of θ_L at the rate of change holding at time t. Comparison of the decision rule represented by (4.12) with the rule $K_3 = 1$ shows that the former is advantageous for sinusoidal loads with $|p| < \frac{1}{\tau}$, while the latter is advantageous for $|p| > \frac{1}{\tau}$. That is, extrapolation of the rate of change of orders will further damp fluctuations whose period is greater than 2π times the production lag, but will accentuate fluctuations of shorter period.

4.3 Feedback of information About Inventories

We consider next the denominator of (4.7). For small values of p, this approaches $(p + K_2)$. Hence, the system will behave for long-period oscillations of Θ_L in the same manner as the system analysed in Section 2. Moreover, because of the theorem of equation (2.14) we may expect the same general behavior in the steady state of this system as of the system of Section 2.

The roots of the characteristic equation:

(4.14) $p^2 + (a + bp)e^{-\tau p} = 0$

are not easily evaluated, but we can reach some qualitative conclusion. We assume $a > 0$, $b > 0$, $b^2 > 4a$. First, it can be shown that (4.14) has an infinite number of roots. Unfortunately, examination of the roots by Nyquist's criterion reveals that not all the roots have negative real parts, hence the operator K_2 will have to be modified to produce stability. We will not investigate further here what form this modification should take. In the next section, however, we will suggest a method of replacing the fixed lag, with operator $e^{-p\tau}$, by a distributed lag, with operator $\frac{a^2}{(a+p)^2}$, which retains the algebraic character of the system transform and avoids the difficulties encountered in handling (4.14).

5. Problems Raised by the Analysis

The analysis of two simple systems of production control has been carried far enough to indicate the way in which servomechanism theory can be applied to the problems of system design. We may now ask what lines of further investigation are suggested by the analysis.

1. One such has already been indicated. Performance of the system can be improved by proper design of the operator K_3--prediction of new orders.

2. A second direction of inquiry is the form of the operator K_4. The assumption of a fixed production lag seems not implausible as a first approximation when the rate of production varies within narrow limits. When the system undergoes large shifts in production rate it is more realistic to suppose that a rapid increase in scheduled production will increase the lag.

It is not easy to see, however, how a variable lag can be introduced without destroying the linearity of the system-- a property on which the applicability of the Laplace transform method depends. For K_4 must have the property that, if $K_4[\eta(t)] = \mu(t)$ then $K_4[k\eta(t)] = k\mu(t)$ for any real k.

Some realism can be gained by replacing the fixed lag τ, with a <u>distributed</u> lag. In place of $\mu(t) = \eta(t-\tau)$ we write:

(5.1) $\mu(t) = \int_0^t P(\tau)\,\eta(t-\tau)\,d\tau$, where $\int_0^\infty P(\tau)\,d\tau = 1$

$P(\tau)$ may be regarded, then, as the probability that the lag in producing a particular scheduled item will be of length τ. For large values of τ, we would expect $P(\tau)$ to be zero, or at least very small. If (5.1) holds, we have, by (2.13),

(5.2) $\mu(p) = P(p)\,\eta(p)$

For example, suppose $P(\tau) = a^2\tau e^{-a\tau}$. Then $P(p) = \frac{a^2}{(a+p)^2}$, and

(5.3) $$\theta_0 = \frac{\left(\frac{a^2}{(a+p)^2}K_3 - 1\right)\theta_L}{p + \frac{a^2}{(a+p)^2}K_2}$$

If we define $\bar{\tau} = \int_0^\infty \tau P(\tau)\,d\tau = \frac{2}{a}$ as the <u>mean lag</u>, we see that the mean lag is still independent of η. The system transform defined by (5.3) can be analysed by the methods previously employed to determine suitable forms for K_3 and K_2.

For example, if we take $K_3 = 1$, $K_2 = (b + cp^2)$

(5.4) $$Y(p) = \frac{-p^2}{p^3 + a^2cp^2 + a^2p + a^2b}$$

This has zero, steady-state error for $\theta_L = \frac{1}{p^2}$, ie. for $\theta_L = t\ (t \geq 0)$. The parameters <u>b</u> and <u>c</u> can now be given such values that the real parts of the roots of the characteristic equation will be negative, and the system consequently stable. The necessary and sufficient conditions for this are: $c > b > 0$.

3. A third direction of inquiry arises from the possibility that actual production may deviate from scheduled production by some error which itself is an independent variable. That is, we suppose:

(5.5) $\mu = K_4\eta + \theta_W$

where θ_W is the deviation of actual from intended production. This leads to the new system equation:

(5.6) $$\theta_0 = \frac{(K_4K_3 - 1)\theta_L + \theta_W}{p + K_4K_2}$$

Since the characteristic equation is unchanged, this assumption does not lead to any essentially new problems with regard to the control of the system.

4. A fourth direction of inquiry will be explored in the next section. The criterion that we wish to keep inventories low is incomplete. We wish to keep inventories low, but we also wish to stabilize the production rate. What can be said about the optimal forms of K_3 and K_2 when this more complex criterion is introduced?

6. Control of Inventories and Production-Rate Fluctuations.

The general criterion for the optimality of a production control system of the sort we are analysing is that cost of production, in some sense, be minimized.

Large inventories involve interest costs, possible costs through physical depreciation in storage, warehousing costs, etc. An inventory deficiency, on the other hand, involves a "cost" in the sense of delay in filling orders, and consequent customer ill-will. It appears reasonable to include, therefore, in the cost of production an element that represents the cost of excess or deficiency in inventories, say $\xi(\theta_0)$. In first approximation we may take ξ_1 proportionate to $|\theta_0|$, or to θ_0^2.

It also appears reasonable to assume that the cost of producing a given quantity of output over a period of time is minimised if output is constant during that time. If we represent the output as a constant plus an oscillating function with zero mean--$\mu(t) = M + \mu(t)$--then we may assume that the rate at which cost is being incurred is a function of M and of the frequency and amplitude of $\mu(t)$.

Now, from equation (4.1), we know that

$$\theta_0(p) = \frac{1}{p}\mu(p) - \frac{1}{p}\theta_L(p) \tag{6.1}$$

that is,

$$\frac{d\theta_0(t)}{dt} = \mu(t) - \theta_L(t) \tag{6.2}$$

Hence, if we succeed in stabilizing θ_o at $\theta_o = Q$ μ will not be constant, but will follow $\theta_L(t)$ Conversely, if we stabilize μ, θ_o will not be constant, but will follow the integral of $\theta_L(t)$ We cannot devise a system that will simultaneously eliminate inventory and production fluctuations, but must, instead, establish a criterion that is some weighted average of these.

6.1 Analysis of a Specific Criterion

To be specific, we consider the steady-state of the system under sinusoidal inputs and outputs. This assumption is consistent with the system (6.2) In fact, in the steady state, if θ_L is sinusoidal, θ_o and μ will be sinusoidal with the same period. We assume that the cost associated with μ is proportional to the square of the amplitude of its oscillation, i.e., is of the form $\rho|\beta|^2$, where $|\beta|$ is the amplitude. Similarly, we assume that the cost of holding inventories is $\sigma|C|^2$ where $|C|$ is the amplitude of θ_o.

We let:

(6.3) $\theta_L(t) = a \cos \omega t$

(6.4) $\mu(t) = b \cos \omega t + \beta \sin \omega t$

(6.5) $\theta_o(t) = c \cos \omega t + \gamma \sin \omega t$

with a, b, β, c, γ real

whence,

(6.6) $\omega\gamma = b - a \, ; \; -\omega c = \beta$

We wish now to minimize

(6.7) $\rho(b^2+\beta^2) + \sigma(c^2+\gamma^2) = \xi$

subject to (6.6) Substituting for c and γ from (6.6) into (6.7), taking derivatives of ξ with respect to b and B and setting these equal to zero, we find:

(6.8) $b = \dfrac{a\sigma}{\rho\omega^2+\sigma} \, ; \; \beta = 0$

(6.9) $c = 0 ; \; \gamma = \dfrac{a\rho\omega}{(\rho\omega^2+\sigma)}$

For small ω, b $\to$ a; $\gamma \to 0$. For large ω, b $\to$ 0; $\gamma \to 0$, $\omega\gamma \to a$.

Interpreting these results, we find that the optimal decision-rule will adjust the production rate and hold inventories down for long-period fluctuations in orders, but will stabilize production and permit inventories to fluctuate for rapid fluctuations in orders. In the latter case, the inventory excess or deficiency will remain small ($\gamma \to 0$) because the period of oscillation is short. The amplitude of manufacturing fluctuations (b) will vary inversely with ω. The magnitude of inventory fluctuations (γ) will have a maximum for $\omega^2 = \sigma/\rho$.

6.1½ Appendix to Section 6.1

In the previous section we used the quadratic cost function (6.7). Highly interesting results are obtained by using the linear function:

(a) $\xi = \rho\left|\sqrt{b^2+\beta^2}\right| + \sigma\left|\sqrt{c^2+\gamma^2}\right|$

Minimizing ξ after substitution for c and γ from (6.6), we find as optimum values:

(b) $\beta = 0,\ c = 0$

But, for b we find:

(c) b = o for $\omega > \frac{\sigma}{\rho}$

b = a for $\omega < \frac{\sigma}{\rho}$

Correspondingly, for γ:

(d) $\gamma = -a$ for $\omega > \frac{\sigma}{\rho}$

$\gamma = 0$ for $\omega < \frac{\sigma}{\rho}$

Writing $z(p) = \mu(p)/\epsilon_0(p)$ we see that for optimum results, $z(p)$ should have the characteristics of an ideal low-pass filter: it should transmit without distortion all frequencies below $\frac{2\pi\sigma}{\rho}$, and should filter all frequencies above $\frac{2\pi\sigma}{\rho}$. The meaning of this requirement in terms of a decision rule can be interpreted by the same methods as those used for the quadratic cost functions in succeeding sections.

6.2 Requirements for the System Transform

We must now determine what kind of a system transform will satisfy (6.4) - (6.5) with b, β, c, γ given by (6.8) - (6.9). For brévity we write $\theta_o(p)/\theta_L(p) = Y(p)$; $\mu(p)/\theta_L(p) = Z(p)$. Then from (6.1),

(6.9) $Z(p) = 1 + p\,Y(p)$

The optimum transform $Z(p)$ is found readily as follows. Recalling (2.13) we can write the output $\mu(t)$ for a sinusoidal load $e^{i\omega t}$:

(6.10) $\mu(t) = \int_0^\infty \omega(\tau) e^{i\omega(t-\tau)} d\tau = e^{i\omega t} \int_0^\infty \omega(\tau) e^{-i\omega\tau} d\tau$

But the factor under the integral sign of the right-hand side of (6.9) is, by definition, $Z(i\omega)$. That is, for a sinusoidal load with period $2\pi/\omega$, we will have:

(6.11) $\mu(t) = Z(i\omega)\,\theta_L(t)$

We see immediately that

(6.12) $Z(i\omega) = \frac{b}{a} = \frac{\sigma}{\rho\omega^2 + \sigma}$

Hence:

(6.13) $Z(p) = \frac{\sigma}{\rho p^2 + \sigma}$

and, from (6.9)

(6.14) $Y(p) = \frac{-\rho p}{\rho p^2 + \sigma}$

The characteristic equations of $Z(p)$ and $Y(p)$, however, have imaginary roots $p = \pm i\sqrt{\frac{\sigma}{\rho}}$. Hence a system with these transforms would be unstable. The transient output would be an undamped sinusoidal oscillation.

The reason for this somewhat unpleasant result is that we have designed the transform to minimize costs for steady-state operation. This will not in general minimize cost when the system is passing from one steady-state to another. Clearly, for $\theta_L(t) = T_K$, a constant $(t \geq 0)$, we want

(6.15) $\mu(t) = T +$ . . a transient term

(6.16) $\theta_o(t) = a +$ a transient term

The transient term in (6.15) should be such that $\mu(t)$ will not overshoot --that is, the system should be over-damped. This implies that the roots of the characteristic equation of $Z(p)$ should be negative and real.

To get the desired steady state behavior of $\mu(t)$ for the indicated load we require that $\lim_{p\to 0} p\,Z(p)\,\frac{T}{p} = \lim_{p\to 0} Z(p)T = T$. From (6.16) we infer that $\theta_o(t)$ should be heavily damped, with $\lim_{p\to 0} Y(p)T = 0$. From (6.9) we see that this latter condition is a sufficient condition that $\lim_{p\to 0} Z(p)T = T$.

As can be seen from inspection, these limiting conditions are satisfied by the transforms of (6.13) and (6.14) although, because of instability, the conditions on the transients are not. To remedy this situation, we replace the denominator of $Y(p)$ by $(\sqrt{e}\,p + \sqrt{\sigma})^2$. The resulting transform:

(6.15) $$Y(p) = \frac{-ep}{(e^{\frac{1}{2}}p + \sigma^{\frac{1}{2}})^2}$$

is critically damped, and approaches the transform of (6.14) for large p. The characteristic equation has the two equal negative real roots: $p_0 = -\left(\frac{\sigma}{e}\right)^{\frac{1}{2}}$ and the transient term in $\mu(t)$ will be of the form $Ate^{-p_0 t}$.

6.3 Construction of the Decision Rule

We return now to the problem of finding a K_3 and K_2 that will realize the $Y(p)$ of (6.15) We will first explore the simple case where $K_4 = 1$ (no production lag). In this case:

(6.16) $$Y(p) = \frac{K_3 - 1}{p + K_2}$$

If we now set $K_3 = 1 - \rho p$, $K_2 = \rho p^2 + [2(\rho\sigma)^{\frac{1}{2}} - 1] + \sigma$ and substitute in (6.16), the result is (6.15). Moreover, we will have for $Z(p)$:

(6.17) $$Z(p) = \frac{2\sqrt{\rho\sigma}\,p + \sigma}{(\sqrt{\rho}\,p + \sqrt{\sigma})^2}$$

Since, in the case of zero time-lag, $\mu(t) = \eta(t)$, (6.17) gives the following rule for determining $\eta(t)$:

(6.18) $$\rho\frac{d^2\eta}{dt^2} + 2\sqrt{\rho\sigma}\frac{d\eta}{dt} + \sigma\eta(t) = 2\sqrt{\rho\sigma}\frac{d\theta_2}{dt} + \sigma\theta_2(t)$$

In the case where there is a fixed production lag: $K_4 = e^{-\tau p}$, we have:

(6.19) $$Y(p) = \frac{e^{-\tau p}K_3 - 1}{p + e^{-\tau p}K_2}\;;\quad Z(p) = \frac{e^{-\tau p}(K_2 + pK_3)}{p + e^{-\tau p}K_2}$$

If we define $X(p) = \frac{\eta(p)}{\theta_2(p)}$, we obtain from (4.2) and (6.19):

(6.20) $$X(p) = \frac{K_2 + pK_3}{p + e^{-\tau p}K_2}$$

Giving K_2 and K_3 the same values as in the previous case, we get:

(6.21) $$X(p) = \frac{2\sqrt{\rho\sigma}\,p + \sigma}{p(1 - e^{-\tau p}) + e^{-\tau p}(\sqrt{\rho}\,p + \sqrt{\sigma})^2}$$

The corresponding decision rule is:

(6.22) $$\frac{d\eta(t)}{dt} = (1 - 2\sqrt{\rho\sigma})\frac{d\eta(t-\tau)}{dt} - \rho\frac{d^2\eta(t-\tau)}{dt} - \sigma\eta(t-\tau) + 2\sqrt{\rho\sigma}\frac{d\theta_2(t)}{dt} + \sigma\theta_2(t)$$

This rule we may take as a realizable approximation to the rule that would minimize costs. For the limiting cases as $p \to \infty$ and $p \to 0$, it has the same properties as the rule derived from (6.17).

7. Further Consideration of the Cost Criterion

The cost criterion developed in Section (6.1) is undoubtedly greatly over-simplified. In this Section we will consider possible methods of constructing a more realistic criterion. In particular, we wish to introduce a more complete analysis of that part of the cost function that depends on rate of manufacture.

We suppose that the cost of manufacture is the sum of three components.

1. Variable costs proportional to the rate of manufacture (e.g. cost of materials). Since these costs are determined by the number of orders to be filled, and hence are independent of the control system, we may continue to ignore them.

2. Fixed costs proportional to plant capacity, i.e. to the maximum rate of manufacturing activity. The previous section indicates how these costs can be handled in designing the control system.

3. Sticky costs proportional to the rate of manufacture when this is constant, but not capable of being reduced immediately as the rate of manufacture declines. In first approximation we may assume that as the rate of manufacture increases from a stable level, sticky costs will increase proportionately, but that if the rate of manufacture decreases, there is a fixed upper limit to the rate at which sticky cost will decrease.

Suppose (Figure 5) that $\mu(t)$ is subject to oscillation of period $\frac{2}{\alpha}$ and amplitude A. The slope of μ will be $\pm A\alpha$. Suppose, further that sticky costs, $\xi(t)$ can only decrease with a slope of $-\beta$, but can increase as rapidly as μ.

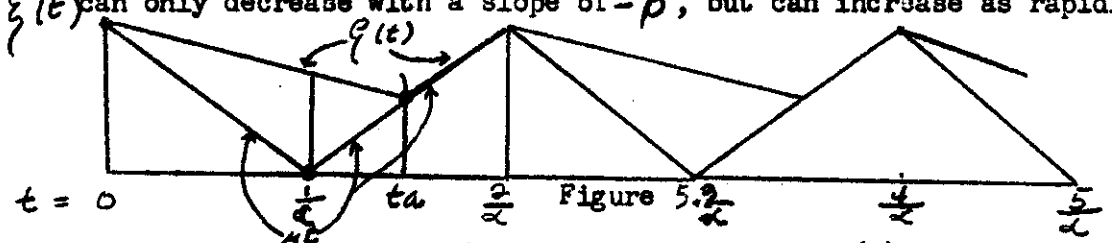

Figure 5.

The ordinate of $\mu(0)$ is A; the slope of $\mu(t)$ is $-A\alpha$ in the interval $0 \leq t < \frac{1}{\alpha}$, $A\alpha$ in the interval $\frac{1}{\alpha} \leq t < \frac{2}{\alpha}$. The ordinate of $\xi(0)$ is A; the slope of $\xi(t)$ is $-\beta$ in the interval $0 \leq t < t_a$, $A\alpha$ in the interval $t_a \leq t < \frac{2}{\alpha}$. The integral of $\xi(t) - \mu(t)$ over the interval $0 \leq t < \frac{2}{\alpha}$ is then the area of the triangle whose vertices are indicated by dots. This area is 1/2 the ordinate of $\xi(\frac{1}{\alpha})$ times t_a. $\xi(\frac{1}{\alpha}) = A - \frac{\beta}{\alpha}$, The value of t_a is given by

(7.1) $A\alpha(t_a - \frac{1}{\alpha}) = A - \beta t_a$ or

(7.2) $t_a = \frac{2A}{A\alpha + \beta}$

Hence:

(7.3) $$\int_0^{2/\alpha} [\xi - \mu]\, dt = \frac{A\alpha - \beta}{2\alpha} \cdot \frac{2A}{A\alpha + \beta} = \frac{A}{\alpha} \frac{(A\alpha - \beta)}{(A\alpha + \beta)}$$

If 1 is an integral multiple of $2/\alpha$, then

(7.4) $$\int_0^1 (\xi - \mu)\, dt = \frac{\alpha}{2} \int_0^{2/\alpha} (\xi - \mu) = \frac{A}{2} \frac{(A\alpha - \beta)}{A\alpha + \beta} = C$$

But, for this same interval $\int \mu\, dt = \frac{A}{2}$, hence, we have for the ratio of sticky costs to production:

(7.5) $$C \cdot \frac{2}{A} = \frac{(A\alpha - \beta)}{(A\alpha + \beta)}$$

It follows that sticky costs will be increased by an increase in the amplitude of the oscillations of $\mu(t)$--behaving, in this respect, like fixed costs--and will also be increased by an increase in the frequency of μ . In designing our optimal criterion, we disregarded this latter consideration. Hence, the design of the decision rule can be improved by decreasing the response of $\mu(t)$ to $\theta_L(t)$ for high frequencies of the latter at the expense of increasing somewhat the response of $\theta_0(t)$. Again it is reassuring that our results coincide with common sense.

Assumption of a sinusoidal oscillation of μ and of ξ leads to the same kind of results. Finally, fixed costs and variable costs can be subsumed as a limiting case of sticky costs by considering a range of different cost categories, each with its characteristic β. The β for variable costs would be infinite, for fixed costs 0. If we can define some kind of an average β, this can be used as a basis for our criterion of manufacturing costs.

8. Conclusion

The general conclusion to be drawn from our explorations, however tentative these have been, is that the basic approach and fundamental techniques of servomechanism theory can indeed be applied fruitfully to the analysis and design of decisional procedures for controlling the rate of manufacturing activity. To be sure, most of the conclusions we have reached could, at least in a qualitative sense, be reached intuitively. But even here intuition has been aided by the frame of reference that servomechanism theory provides. Moreover, the more exact procedures permit statement of our results with a degree of precision that could not be attained without them. Even in this very early stage the theory permits actual numbers to be inserted for the construction of specific decision rules that would apply, with a considerable degree of realism, to actual situations.

In the body of the paper, and particularly in Sections 4.2, 5, and 7, we have suggested a number of lines of further inquiry. From a formal standpoint, it would also be desirable to take the basic elements of the system developed in this paper, and to treat the system in terms of difference equations rather than differential equations. The actual results should not be much changed, but it is obvious that in the application of the theory it would in any event be necessary to approximate differential equations, if these are used, by difference equations.

9. References

The following list of references covers some of the more systematic and lucid introductions to servomechanism theory. They are listed in order from the relatively elementary treatments to the more advanced or specialized.

1. H. Lauer, R. Lesnick, and L. E. Matson, Servomechanism Fundamentals (McGraw-Hill, 1947). An introductory treatment employing differential equations rather than the Laplace transform. Analyses in detail the behavior of a few

very simple engineering servos, and gives a clear picture of the servo concept.

2. H. M. James, N. B. Nichols, and R. S. Phillips (eds), Theory of Servomechanisms (McGraw-Hill, 1947). Chapter 1, "Servo Systems", gives a good introduction to basic concepts. Chapter 2, "Mathematical Background", gives an excellent introduction to the Laplace transform, its physical meaning, and its relation to the weighting function. General design principles and techniques are discussed in Chapter 4, and more advanced topics in other chapters.

3. M. F. Gardner, and J. L. Barnes, Transients in Linear Systems, Vol. I (Wiley, 1942). A clear, systematic exposition of Laplace transform theory and methods.

4. G. S. Brown, and D. P. Campbell, Principles of Servomechanisms. Parallel to [3], but with more emphasis on system design, and less on analysis.

5. L. A. MacColl, Fundamental Theory of Servomechanisms (Van Nostrand, 1945). An elegant brief treatment stressing fundamental concepts and employing Laplace transform methods.

6. N. Wiener, The Extrapolation, Interpolation, and Smoothing of Stationary Time Series (Wiley, 1949). An approach to the problem of forecasting a stochastic input or load. (See also [2] Chapters 6-8).

7. R. C. Oldenbourg and H. Sartorius, The Dynamics of Automatic Controls, (ASME, 1948). A systematic treatment of controllers, using Laplace transform methods. Includes extensive discussion of fixed lags, non-linearities, and discontinuous regulation. (On the last point, see also [2], Chapter 5; [5], Chapter 10 and Appendix; and [3], Chapter 9.

Excellent bibliographies will be found in [3], [4], [5], and [7].