THE ANALYSIS OF FEEDBACK SYSTEMS

by

ERIK VIGGO PEDERSEN

S.B., Massachusetts Institute of Technology
(1962)

SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF

SCIENCE

at the

MASSACHUSETTS INSTITUTE OF

TECHNOLOGY

September, 1967

Signature of Author ....
Alfred P. Sloan School of Management, June 30, 1967

Certified by ....
Thesis Supervisor

Accepted by ....
Chairman, Departmental Committee on Graduate Students

THE ANALYSIS OF FEEDBACK SYSTEMS

by

ERIK VIGGO PEDERSEN

Submitted to the Alfred P. Sloan School of Management on June 30th, 1967 in partial fulfillment of the requirements for the degree of Master of Science in Management.

This thesis is concdrned with developing an analytic approach to systems analysis that will serve as a suitable basis for an Industrial Dynamics education. The material contained in the thesis indicates new ways to present material in a management curriculum that is more appropriate than the approaches normally followed in the engineering curricula. Several specific areas in which innovations were made are included in addition to a critical commentary of the material and an indication of what should next be accomplished in developing this course.

Thesis Advisor: Jay W. Forrester
Title: Professor of Management

Professor Edward N. Hartley
Secretary of the Faculty
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Dear Professor Hartley:

In accordance with the requirements for graduation, I herewith submit a thesis entitled, "The Analysis of Feedback Systems". I would also like to take this opportunity to express my appreciation to Professor J. W. Forrester for his advice in the direction and development of this thesis.

Sincerely,

Erik Viggo Pedersen

TABLE OF CONTENTS

LIST OF FIGURES

List of Figures Continued

CHAPTER I

INTRODUCTION

There are many opportunities in modern management for new ways to solve complex problems and the Industrial Dynamics[1] approach developed at M.I.T.'s Sloan School of Management shows promise of being one of the best new approaches. The strength of Industrial Dynamics lies in its ability to understand the behavior of large social systems in terms of feedback processes, processes in which the variables of a system interact with one another over time. Much of the successful work that has been accomplished in Industrial Dynamics to date has been performed by men with engineering backgrounds and it is felt, although not without some reservation, that a familiarity with engineering system analysis is important for a man to be successful in the field. To transfer this capability to the student with a more general background has been the objective of a portion of the group in the last few years and, in particular, to develop educational material that could be used to establish a better foundation in Industrial Dynamics in all types of management schools. This effort is currently following two parallel paths, one in which the student is exposed to many system models and develops experience through experimentation with the models and the other, which is based on an analytic approach to system analysis that is not quite the same as the normal engineering approach.

This thesis is concerned with the latter alternative of developing an analytic approach as a foundation for Industrial Dynamics study and represents

---

[1] Forrester, Jay W., Industrial Dynamics (M.I.T. Press, Cambridge, Mass.) 1961.

an initial effort to develop the type of material that will be most successful for the task at hand. This task is defined as teaching a student with little or no background in calculus (or engineering analysis) the basic mathematics of linear feedback systems. Along with the basic material it is also intended to provide the student with some tools that will allow him to extent his ability to analyze more complex systems by relying on his understanding of the basics. To what extent this approach to the problem will need to be separate from or to interact with the alternative experimental approach has yet to be determined and it is expected that this thesis will assist in defining further goals. The thesis does not contain material in a form that is directly useable by the student. As it stands it presents a development of ideas which could become the structure for a course. Material that could be borrowed directly from other sources has not been included as little modification would be required.

Comparative Analysis of the Material Development

The material that is being developed for the analytic approach to feedback systems relies very heavily on its related engineering courses. The development is not exactly the same, however, and the material is much more strongly related to the engineering curriculum in some areas than in others. The material for the management course emphasizes the analysis of low order (first and second) linear systems with particular attention paid to what parts of the system structure are responsible for what parts of the behavior. For higher order systems a graphical approach is introduced which provides an approximate solution only. This is in contrast to an engineering curriculum which typically starts with the analysis of low order systems but quickly builds a basis in the frequency domain through the use of the Laplace transform.

Students are thus able to analyze linear systems of arbitrary order and complexity. Engineering is by its very nature, however, a field devoted to dealing with linear systems. In that the engineer creates the system to satisfy a particular function he controls the structure that he is attempting to analyze. As he wishes to be highly certain of system behavior, he uses system elements whose linear characteristics will be most important in determining that behavior. If a system element does not satisfy such needs, it will not be used. Managers concerned with social systems, however, are rarely in such direct control of the structure of the system. Social systems tend to evolve over time as a result of many conflicting pressures and the uncertainties of history. The highly complex systems that result contain many nonlinearities and the highly accurate linear analysis that is provided by the Laplace transform is of relatively little value. The emphasis in the material that is presented here is much more on the strengthening of the student's ability to understand the basic systems. In addition tools are provided which allow him to evaluate the relative importance of one basic system element over another in a more complex system. A detailed examination of how this material has been developed will serve to assist in the appreciation of its merits. The following three sections compare the material contained in Chapters 2 through 4 to existing approaches.

1. The Analysis of First Order Systems

The initial development of first order systems concentrates on describing the response of a very simple system with which the average student would have some experience. The physical model of a water reservoir system is used to demonstrate the effect of simple control actions and the basic exponential response of first order systems. Several approaches to justifying the exponential as the appropriate time function for a dynamic system response are used. Normally mathematical treatments simply demonstrate that the exponential is an acceptable solution and it is later proved that it is a unique solution, a process that is satisfying only to the mathematically inclined. In addition to demonstrating the feedback operation of the system that results in the exponential solution, the text offers a logarithmic derivation as a final confirmation.

The remainder of the Chapter 2 is devoted to an analytic derivation of the response of a first order system and the examination of the response of the system to a variety of driving functions. The mathematical formulation of the first order system is presented in integral equation form which, in contrast to the differential equation format normally used, explicitly includes the initial conditions. The student is also better able to relate the conceptual model based on levels and rates to its analytic equivalent. For example, it is easier to conceptualize the level as resulting from an accumulation of a rate of flow over time than to work with an equation that identifies the rate of flow as being a function of the rate of change of the level.

As in traditional approaches the solution to the equation describing the system is developed by components except the single integral equation is used to derive all the unknown coefficients. In using a differential equation it is necessary to resort to a truncated version (with the right hand side set to zero) to solve for the homogeneous solution. The separate components of the solution are identified in a way which more directly relates them to specific system processes. The homogeneous solution (which is referred to herein as the structural solution) is explicitly identified as being associated with the changes in the levels within the system. Its two components, one due to the initial condition of the level and the other due to the equilibrium of the level necessary to balance the driving function, are combined to produce a net structural response. This is in contrast to the normal approach in which the amplitude of the structural approach is determined as a last step to ensure consistency of the equation at time zero. In addition, the approach taken here for the driven response (usually referred to as the particular solution) emphasizes a unique feature of feedback systems. The most common types of feedback systems are those in which a system variable is being fed back and compared with some desired value or goal for that variable. Assuming the system is able to, it will move the system variable in a direction to match the goal and if the goal changes continually, the system will force the system variable to follow. The driven component of the system response is defined as being that component which represents the system's attempt to follow the driving function. As such it is defined as having the same time behavior as the driving function, and its amplitude is obtained by substituting the time behavior of the driving function into the complete integral equation. For a step input this amplitude is typically unity but for a sinusoidal input the gain involves a complex polynomial in terms of the period of the oscillation. This polynomial is the characteristic equation that is typically developed separately in other treatments.

The analysis of a first order positive feedback system is carried out in exactly the same way as in the case of the negative feedback systems already discussed. The difference between the two types that is emphasized is that the structural component dominates the solution in the case of positive feedback. The driven response is still identified as a separate component and simply results in an offset to the structural response, opposite in polarity to the change in the driving function. As positive feedback is not of interest for engineering design purposes there is little if any literature with which to compare this approach.

2. Analysis of Second Order Systems

Chapter 3 develops the derivations for second order systems using the integral equation formulation in the same manner as was used for first order. The structural solutions are able to exhibit more complex behavior, however, and the manner in which the type of behavior exhibited depends on the system's structural parameters is discussed. One of the important distinctions in the approach used here is the recognition of the initial conditions of the two levels in a second order system as being an important factor in determining the structural response of the system. Conventional treatments use the initial value of the variable of interest plus the initial value of the rate of change of that variable. When the stability of that rate is considered, a very poor intuitive feel for the starting condition of a system is obtained. For example, the impression is usually given that the rate of change of a variable can change very rapidly and that an error in estimating the derivative at time zero will be of little import because of its tendency to move rapidly. The approach used in this thesis identifies this second initial condition as the initial condition of a level and explicitly includes in the integral equation for the system.

Variations in the structure of second order systems are also included in Chapter 3, including the explict analysis of several forms of nested feedback systems using both positive and negative feedback. The addition of a single term to the left hand side of the general integral equation results in a variety of behavior including the exponential and oscillatory forms of both growth and decay, depending upon the polarity of each of the feedback loops. The material that is presented on the analysis of two loop systems where one loop is positive and the other is negative is not found in any other literature.

3. Analysis of High Order Systems

Chapter 4 presents a graphical analysis technique which may be used to qualitatively evaluate the performance of higher order linear systems. Initially the tool is demonstrated by developing the same results for first and second order systems that had been developed in earlier chapters. Then its use with higher order systems is demonstrated with extensions of the examples used for the low order systems.

This graphical approach contrasts with the Laplace transform analysis normally developed in engineering curricula. Although the transform results in an analytic expression that is accurate for higher order systems, the user rarely realizes that many of the components of his solution have very little influence on the total response. The mathematical operations that must be understood for its derivation leave the user with a very poor feeling for the reasons for the system's behavior. Although the graphical approach developed in Chapter 4 depends on a frequency domain analysis, it is not expressed as such. Instead the curves are derived from examining the same integral equation that was used in the earlier chapters. In addition the sequential nature of the approach forces the student to examine the characteristics of each group of system elements and to consider the effects of adding feedback loops one at a time around the components. The nature of the problem solving process is thus forcing the student to be aware of the role each part of the system plays in developing the final system characteristics. When the characteristic is completed it will be readily apparent that the transition in the gain curve resulting from some parts of the structure will be relatively unimportant compared to others and that the response can be approximated by the response of a much simpler structure. It is important that a student be taught to reduce a system structure to its simplest form as the subtleties resulting from a more sophisticated model are very often obscured by a noisy environment. Similarly, when a large feedback system involving nonlinear elements is analyzed using ideas about loop dominance it is more important to understand the interactions of the simple structures within the system than to study a complete linear analysis of the entire system.

Additional Material Sources

Not all the material required for this course is included in the thesis and apologies are offered to the reader for the lack of continuity that may be apparent. Chapters 2 through 4 include the major innovations required to successfully teach a course in social systems analysis. In addition to the material included in the thesis, other sources need to be drawn on to supply the missing pieces of the course. Among the more important are:

1) the concept of a system. The student must be introduced the idea of interdependence of variables that have some common association. The recent article in

Management Science by Forrester[1] offers the framework for such material and there is much in introductory engineering literature that can be drawn on for support.

2) The development and use of models. The student should understand that any model is an approximation and should be taught to evaluate its effectiveness for a particular problem situation. Some excellent material on engineering models is contained in the new system dynamics book by Shearer, Murphy, and Richardson.[2]

3) The concepts of linearity and superposition. These ideas are important in that they underly many of the mathematical operations that are used to derive solutions in the material in the thesis. The student should be exposed to material such as that contained in Introductory Network Theory by Bose and Stevens.[3]

4) Complex algebra. The student will need to be familiar with basic complex number operations at the level of the presentation in the book by Shearer, Murphy, Richardson.[2]

In addition to the above, the course will need effort in developing course notes from the thesis material that are directly useable by the student, as well as a complete set of exercises.

---

[1]Forrester, Jay W., "Industrial Dynamics--After the First Decade", to be published Management Science, Spring, 1967.

[2]Shearer, J.L., Murphy, A.T., Richardson, H.H., Introduction to System Dynamics (Addison-Wesley, Reading, Mass.), 1967.

[3]Bose, A.G., Stevens, K.N., Introductory Network Theory (Harper and Row, New York, N.Y.), 1965.

CHAPTER II

FIRST ORDER FEEDBACK SYSTEMS

The types of behavior that can be expected from simple first order feedback systems will be explored in this chapter. A water reservoir system will be considered in which the interdependence of variables that characterizes feedback will be introduced. Reservoirs are built for a purpose and in the simplest case they are used to store rainfall in order that a community may have an ample water supply during dry periods. This implies that the level of water is important and that when it is low people will be cautioned to decrease their water consumption, to stop watering lawns, etc. Similarly, those in charge of the water supply will become concerned when the water level is too high for fear that it will overflow and damage the reservoir, perhaps even flooding a nearby community. Action will be taken to ensure that the water stays near some desired level, either by reducing consumption or by draining off the excess water depending on whether it is too low or too high. Action is a direct function of the water level and the water level itself is affected by the action. As the water level changes, this in turn has a further affect on the control actions, decreasing them if the water level is moving closer to its desired level. This interdependence of cause and effect relationships is identified as feedback. Those variables that are involved in the feedback loop constitute a closed system and those that are not, such as the rainfall are considered to be external variables.

The type of behavior that a feedback system can exhibit is considerably more complex than that of a simple open system. Because of their interdependence the variables are not only changing in magnitude but the rate at which they are changing is also affected. For example, if the man at the reservoir is instructed to drain off excess water at a rate in proportion to the excess water in the reservoir, his maximum outflow rate will be set at the beginning. As the outflow has an effect on the water level, however, he will soon begin to cut back the rate at which he is draining it off as there will be less excess water. In fact, as the corrective action takes effect he will continually decrease the outflow rate until it is down to a trickle when the water level is very close to the desired level. In such a case the response of the system will be such that it moves from its present state to a new one at an ever decreasing rate, until just before it gets to the desired state its rate of change is zero. The exponential function is one possibility to fulfill the requirements of such a function of time. Figure 2-1 illustrates the exponential function and the fact that its slope (rate of change) is also decreasing as the magnitude of the function decreases. In addition to its starting value the only parameter that is needed to define an exponential function is its time constant, that time it takes the exponential to attain 1/e of its transition between the two values (e is approximately 2.718). This time constant can also be considered as the time it would have taken the exponential to reach its final value if it had continued at its initial rate of change. In the illustration this constant has been selected as 20 time units.

In the example of the reservoir system, the time constant of an exponential response would be controlled by the fraction of the excess water that is being used to set the outflow rate. This fraction is set for some very practical reasons. Because the stream below the reservoir has a limited capacity it

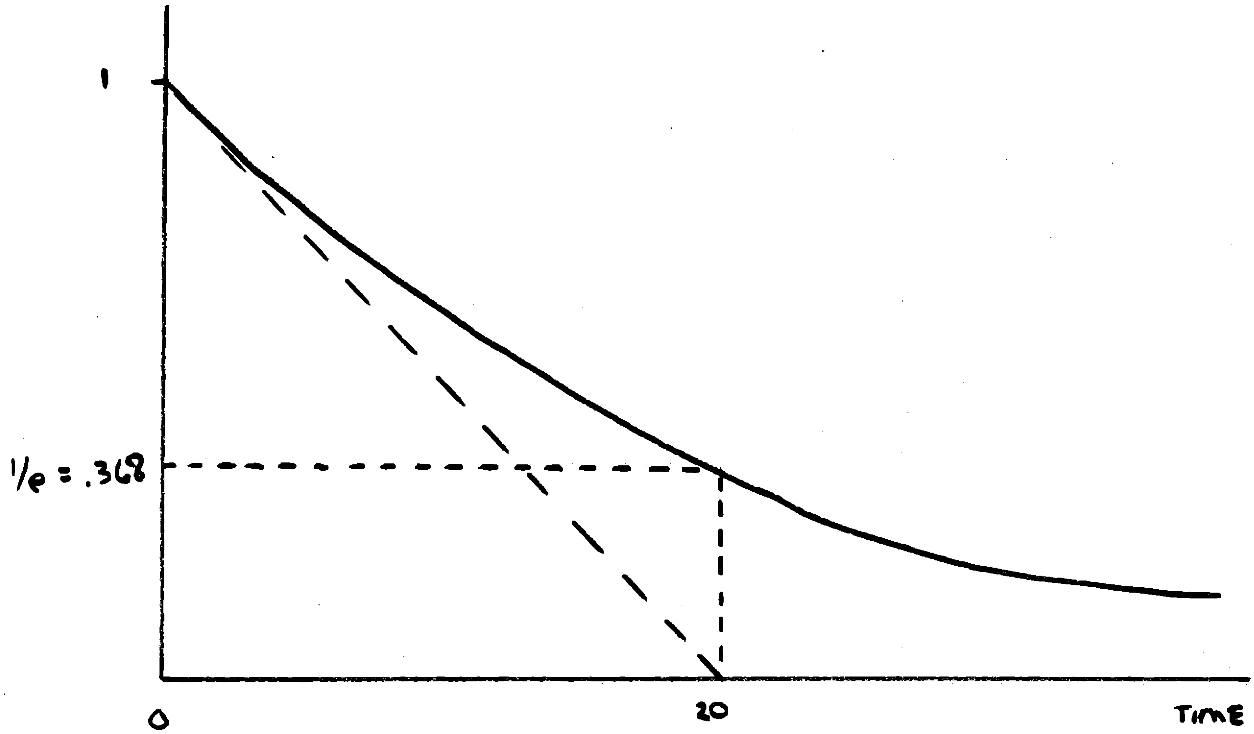


FIGURE 2-1: THE EXPONENTIAL FUNCTION

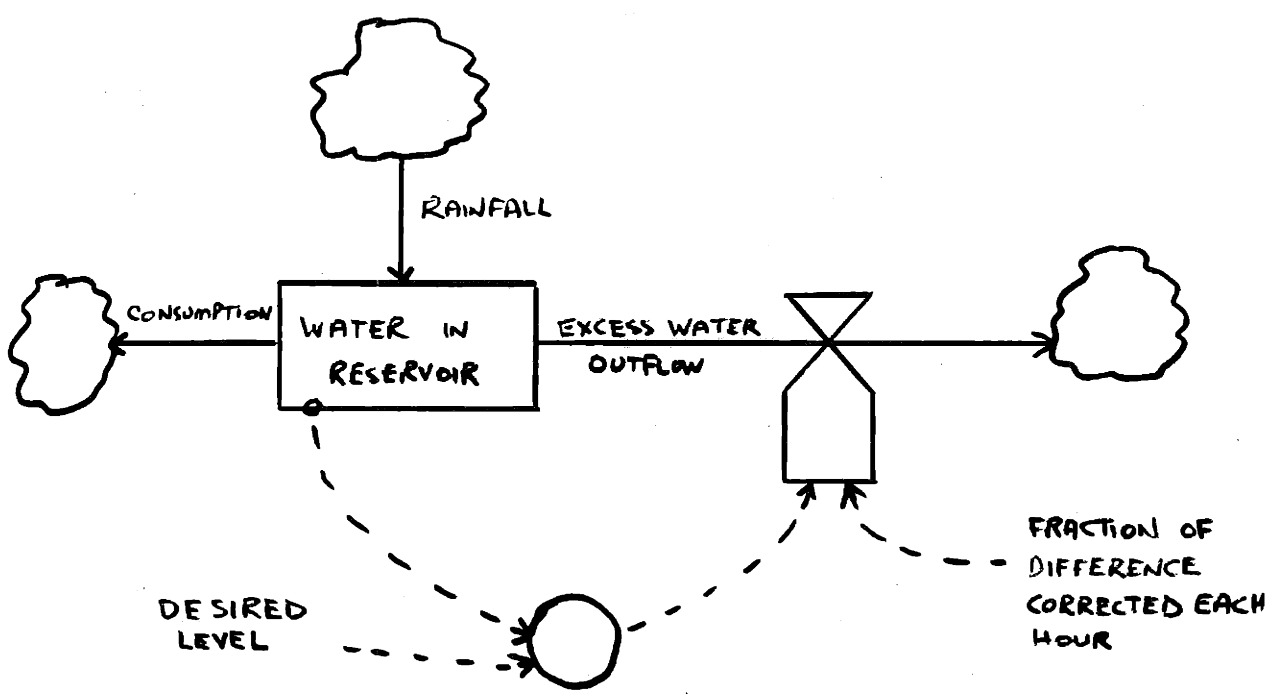


FIGURE 2-2: THE WATER RESERVOIR SYSTEM

would be harmful to attempt to remove all the excess water from the reservoir at once. Perhaps more important, however, the man at the reservoir shouldn't be overresponding to short term fluctuations in the water level. If three wet days are followed by two dry days, then the water level would deviate more from the desired level when the excess is completely removed each day than when just a fraction of it is removed. The removal of a fraction of the water at a time allows the reservoir to average some of the changes in rainfall conditions, having a stabilizing influence on the level of water. However, this policy on removing a fraction of the excess water at a time does slow down the system's response. For example, a large increase in rainfall could create an overflow if the fraction corrected for were too small. Thus, the selection of this parameter of the system will directly affect the time performance of the system variables. To understand how dynamic performance and the system parameters are related exactly some analytic solutions to the behavior of the systems will be considered. But first the nature of the exponential solution will be examined in a little more detail.

The Exponential Solution of the Integral Equation

A system such as the water reservoir can be represented analytically by an integral equation. Figure 2-2 shows the relationship among the variables in the reservoir diagrammatically. The rectangle represents the level of water in the reservoir and the flows represent the various water flows in and out of the system. The flow with the control valve represents the process whereby the man in charge makes sure the water level does not get dangerously low or high. As is shown by the interconnection of the variables this flow is set as a fraction of the difference between the actual water level and the desired level. Assume that there is no rainfall initially. Then the time behavior of the system can be represented by the following integral equation.

$$W(t) = -\int_o^t f(t)dt + W(o)$$

where $$f(t) = a[W(t)-d]$$

W(t) = water level at time t in kilogallons
f(t) = outflow rate at time t in kilogallons/hour
t = time in hours
d = desired water level in kilogallons
a = fraction of the difference to be corrected each hour - dimensionless

Substituting the equation for f(t) into the one for W(t) results in

$$W(t) = -a\int_o^t [W(t)-d]dt + W(o)$$

If the water level is currently 110 kilogallons and the desired level is only 100, then the difference term within the brackets on the right hand side of the equation will be +10 kilogallons. When this is integrated there is no change of sign and the negative sign before the integral operator shows that the net effect of this term will be negative. Thus, the change in water level will be negative, driving it toward the desired 100 kilogallon level and the system, as represented, functions correctly. If the operator only adjusts the outflow rate once every hour instead of continuously and he is told that the fraction of the difference to be adjusted each hour is 0.1, the behavior of the water level can be calculated directly. Figure 2-3 indicates the time behavior that results. Figure 2-3(a) shows the outflow rate that decreases with time as the corrective action takes effect. Figure 2-3(b) shows the effect of integrating this outflow over time, summing all the outflows that

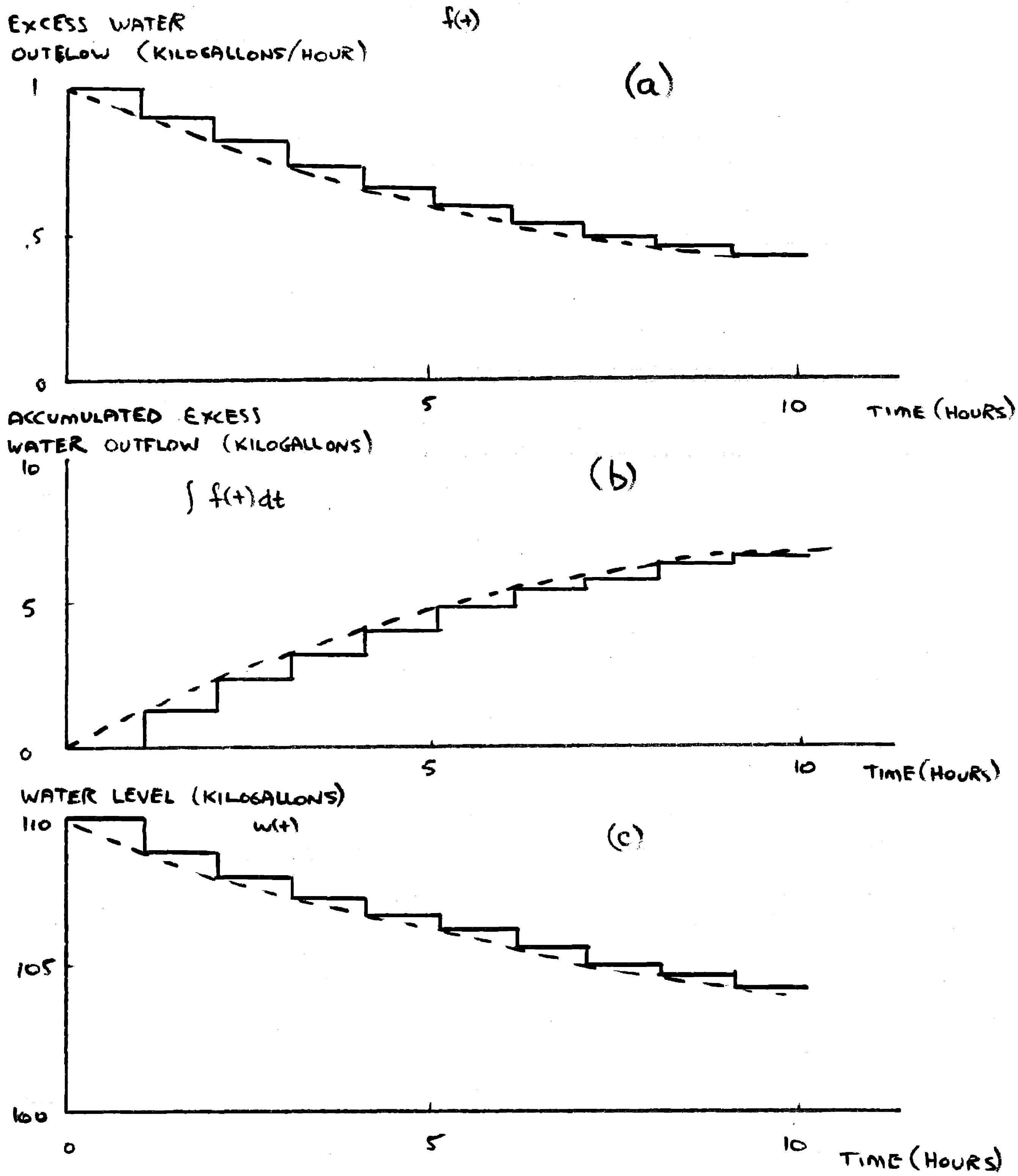


FIGURE 2-3: THE DYNAMIC RESPONSE OF THE WATER RESERVOIR SYSTEM

have occurred up to each point in time, and Figure 2-3(c) shows the resulting behavior of the water level when the curve in (b) is subtracted from the initial value of the water level. If the man at the reservoir were to adjust the outflow rate continuously instead of once per hour, the curves would look more like the broken line that is also on each of the graphs. Note that all three waveforms have the same time function, although changed in sign or in scaling. It is also apparent that the outflow is heading toward an equilibrium value of zero, while the sum of the outflow is tending towards a value of ten kilogallons. This is, of course, the amount of water that has to be removed from the reservoir and the third curve shows the water level itself moving toward the desired value of one hundred kilogallons.

The fact that the excess water outflow function illustrated in Figure 2-3(a) decreases with time in the same manner as the water level in the reservoir approaches its desired value indicates that it has at least one of the properties of the exponential function described earlier. By measuring the time it takes for either of these two functions to reach 1/e of their initial value, it can be determined that the time constants of both functions are approximately equal to ten hours. Assume that the water level can be represented by a constant plus an exponential function with a ten hour time constant. By substituting the values that are apparent in Figure 2-3(c), it is possible to test our result in the integral equation developed previously.

$$W(t) = 100 + 10e^{-t/10}$$

Substituting this expressing into the integral equation,

$$W(t) = -a\int_o^t [W(t) - d] + W(o)$$

results in

$$100 + 10e^{-t/10} = -0.1\int_o^t [100 + 10e^{-t/10} - 100]dt + 110$$

$$= -0.1\int_o^t 10e^{-t/10} dt + 110$$

$$= -0.1[-100e^{-t/10}]_o^t + 100$$

$$= -0.1[-100e^{-t/10} + 100] + 110$$

$$= 10e^{-t/10} - 10 + 110$$

$$= 100 + 10e^{-t/10}$$

As the integral equation describing the system is satisfied, the expression obtained from the graphical solution assuming an exponential form is an acceptable solution for the integral equation. The exponential function is a solution to many feedback system responses as it has the property of being able to retain its characteristic form even when integrated or differentiated an arbitrary number of times. The scaling constants and other constants in the equation will vary but the exponential will not. It will retain its basic shape (the time constant is unchanged) as was illustrated in the plots in Figure 2-3. This important property is necessary to satisfy equations that contain integrals of a function in the same equation as the function itself. The system equation that has just been evaluated demonstrated the fact that the levels and rates in the system had the same time function. Other functions do not have this property as can be readily demonstrated by substituting them into the system equation. Suppose $W(t) = 110 - Zt$, where Z is an arbitrary constant.

$$110 - Zt = -0.1\int_o^t [110 - Zt - 100] + 110$$

$$= -0.1[10t - \frac{Zt^2}{2}]_o^t + 110$$

Without going any further it is apparent that there is no term on the left hand side of the equation (the expression for W(t) itself) that can be equated with the $t^2/2$ term on the right hand side. Similar results will be obtained for other functions which have the general characteristic needed to satisfy the system response of declining at an ever decreasing rate. In fact, mathematicians have shown that the exponential function is a unique solution to such an integral equation.[1] It can also be derived directly by operating with differentials, instead of the integral form of the system equation.[2] Having thus shown the nature of the solution to the integral equation, it is now appropriate to proceed with the full derivation of the general first order feedback equation.

---

[1]Kaplan, Wilfred, Ordinary Differential Equations (Addison-Wesley, Reading, Mass., 1958), p. 23.

[2]Derivation of the solution to the integral equation

$$W(t) = -\int_o^t f(t)dt + W(o)$$

$$= -a\int_o^t [W(t) - d]dt + W(o)$$

Take the derivative of both sides

$$\frac{dw(t)}{dt} = -a[W(t)-d]$$

Reorganize the equation to contain the same variables on one side of the equation (cross-multiply).

$$\frac{dw(t)}{[W(t)-d]} = -adt$$

Integrate both sides of the equation with respect to the range over which the variables will change. The term which represents the excess water will vary from +10 kilogallons initially to some value X while time varies from 0 to t.

$$\int_{10}^{x} \frac{d[w(t)-d]}{[w(t)-d]} = -a\int_o^t dt \qquad d[w(t)-d] = dw(t)$$

$$\ln[w(t)-d]_{10}^{0} = -at]_o^t$$

Footnote continued from page 19

$$\ln X - \ln 10 = -at$$

$$\ln[X/10] = -at$$

Raise both sides of the equation to the exponential power and we have

$$X/10 = e^{-at}$$

or

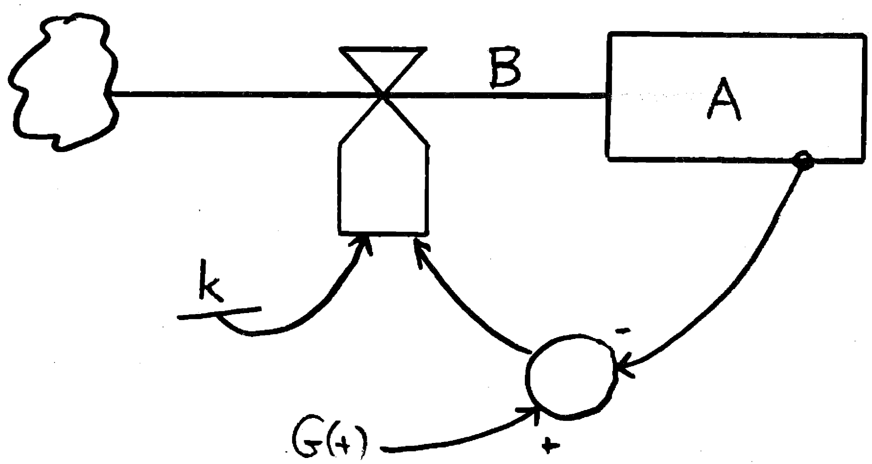
$$X = 10e^{-at}$$

Thus the difference between the water level and the desired level is the function we have identified in Figure 2-3(c) plus a constant.

Derivation of the First Order Feedback System Response

The preceding section demonstrated the solution to the time behavior of a system that resulted only from the system structure and its initial state and did not include the effect of any external variables. Such a solution can be suitably identified as the structural solution of the system although mathematicians normally refer to it as the homogeneous solution. External variables will introduce additional components to the system response. These components can be defined as divided into two groups, the behavior of the system variables that results directly from the external inputs and the structural response or change in the internal state of the system that results from the external variables. The first of these will be called the driven response although it is also known as the particular solution, and it is the attempt by the system variables to reproduce the external variable. For example, if we consider the desired water level in the reservoir to be an external variable that changes with time, then the actual water level would follow the desired level provided it did not change too fast. If the desired level were held constant, the actual level would attempt to remain constant. The driven response is then that part of the behavior which has the same form as a particular external variable. The structural response consists of two components, one which is a function of the initial conditions and the other, the response of the system to the disequilibrium caused by the external variable. When the external variable is suddenly applied at time zero, the system may have to move to another state from the one it is in initially before it can reproduce the form of the external variable. For example, if the desired water level is suddenly increased, the reservoir will not be able to respond immediately to reproduce the constant water level of the new increased value. The response of the system in attempting to restore the balance is the second component of the structural response.

The analytic derivation of the total response to a general first order feedback system will serve to identify these components more clearly. The system illustrated in Figure 2-4 shows the same general structure that was used earlier except that we have now chosen to control the input flow into the level rather than the outflow. This only changes a sign.



FLOW DIAGRAM OF A FIRST ORDER FEEDBACK SYSTEM

FIGURE 2-4

The general integral equation which defines this system is:

$$A = \int_o^t B dt + A(o)$$

Substituting B = K(G-A), we have

$$A = K\int_o^t (G-A)dt + A(o)$$

which can be regrouped to provide

$$A + K\int_o^t Adt = K\int_o^t Gdt + A(o)$$

From our definition earlier, the solution for A is of the form

$$A = A_s + A_d = Qe^{qt} + R$$

where Q, q, and R are as yet undetermined coefficients. In that this is a linear system, the terms are separate and additive. The first term represents the structural solution of the system and the second, the driven response. If G is the input to the system and is some arbitrary function of time, then R is defined to be equal to XG, X being some as yet undetermined coefficient.

By evaluating these coefficients the complete solution to the first order feedback system can be obtained. Again, because the system is linear, it is possible to evaluate one term of the solution at a time. It is convenient to start by determining R, substituting it in the general equation in place of A.

$$R + K\int_o^t Rdt = K\int_o^t Gdt + A(o)$$

What the evaluation of the integral terms is will depend on what function of time G describes. Assume that the time function is a constant G in this case but keep in mind that the equation must be evaluated specifically for each different time function. To evaluate X for the case when G is a constant, XG must be substituted for R in the integral equation.

$$XG + K\int_o^t XGdt = K\int_o^t Gdt + A(o)$$

Integrating the constant time functions results in

$$XG + KXGt\Big]_o^t = KGt\Big]_o^t + A(o)$$

$$XG + KXGt = KGt + A(o)$$

To solve for X we must take the derivative of both sides to eliminate the constant terms.

$$KXG = KG$$

$$X = 1$$

Therefore, the driven response R is G itself.

Next, it is useful to obtain the time constant, q, of the exponential in the structural response. This can be accomplished by considering the effect of the initial conditions only, setting the external variable G to zero.

$$A + K\int_o^t Adt = A(o)$$

Substituting in the structural component of the solution, Qe^{qt}, we obtain

$$Qe^{qt} + K\int_o^t Qe^{qt}\, dt = A(o)$$

$$Qe^{qt} + k/q[Qe^{qt}]_o^t = A(o)$$

$$Qe^{qt} + KQ/q[e^{qt}-1] = A(o)$$

To simplify the expression it is again useful to take the derivative of both sides.

$$qQe^{qt} + KQe^{qt} = 0$$

Now Qe^{qt} can be factored out, leaving

$$q + K = 0$$

$$q = -K$$

Now the amplitude of the structural response can be obtained by substituting the total expression for A in the general equation:

$$A + K\int_o^t Adt = K\int_o^t Gdt + A(o)$$

$$Qe^{qt} + R + K\int_o^t]qe^{qt}+R]dt = K\int_o^t Gdt + A(o)$$

By evaluating the expression at time zero, we see that the integral terms disappear and the one remaining exponential becomes unity.

$$Q + R = A(o)$$

As we have already evaluated R, we can readily obtain Q.

$$Q = A(o) - R = A(o) - G$$

Substituting the values of the coefficients back into the expression for the complete solution results in

$$A = Qe^{qt} + R = [A(o)-G]e^{-kt} + G$$

Note that there are two terms in the structural response, one due to the initial condition and the other resulting from the need to compensate for the sudden application of the input G at time zero. Note that in the simple case where A(o) = G, there is no structural response component and the system continues to operate with A = G.

The solution for A can be confirmed by substituting it into the original general integral equation.

$$A + K\int_o^t Adt = K\int_o^t Gdt + A(o)$$

$$[A(o)-G]e^{-kt} + G + K\int_o^t [A(o)-G]e^{-kt}dt + K\int_o^t Gdt = K\int_o^t Gdt + A(o)$$

The terms containing the integral of Gdt cancel each other and the evaluation of the remaining integrals results in

$$[A(o)-G]e^{-kt} + G - k/k[A(o)-G][e^{-kt}]_o^t + A(o)$$

$$[A(o)-G]e^{-kt} + G - [A(o)-G][e^{-kt}-1] = A(o)$$

$$G + A(o) - G = A(o)$$

$$A(o) \quad = A(o)$$

The expression derived as the solution does indeed satisfy the equation for the system.

A Management System Example

At this point it will prove useful to examine the various types of behavior that a first order system can exhibit by looking at a particular management system. Although it is not usually identified as a unit in a typical firm, the personnel department and the plant work force can be thought of as a simple first order feedback system. The interdependence comes in that the personnel department does the hiring and firing that affects the size of the work force and yet these changes are a direct function of the present number of employees compared to the number needed. In the first order system in Figure 2-4, the men in the work force corresponds to the level (A) while the net hiring rate is the rate (B) that flows into the level. The action of the personnel department is based on the difference G-A, where G in this case is the desired number of employees. In the typical industrial situation this goal would be set by top management in accordance with anticipated needs. In the defined system, the goal is an external variable as it is independent of how many employees the firm has currently and how many are presently being hired by the personnel department. Finally management has set a policy for the personnel department limiting the number of employees hired in any one week to a fraction of the total that is actually needed. In addition to preventing an overload on the firm's ability to train the new people, the policy allows the firm's work force to be more stable in situations where management's plans for needed production capacity are changing rapidly. The extra capacity that is needed in the short run is made up with overtime. The number that management establishes as a fraction of the work force expansion that is the policy limited is represented in Figure 2-4 as the structural parameter K. Thus all the elements of the personnel - work force system can be described by the simple first order feedback model for which a general solution has been developed. It is now useful to examine the behavior observed when numbers representing the various situations that could arise are substituted.

The Step Input

Many times a sudden change will require management to set a new level of desired employees very suddenly. If unexpected orders or loss of orders occur, there will be an immediate shortage or surplus of production capacity in the respective situations. A sudden change in an input is commonly identified as a step and our personnel system's response is determined by substituting some specific numbers into the equation for the solution. Assume that management policy limits the number of people that can be hired or fired in one week to 1/10 that of the total number that need to be hired or fired. Assume also that the work force has 1000 people in it initially, and management suddenly changes the desired level to 800. Then,

$$G = 800$$

$$A(o) = 1000$$

$$K = 1/10$$

and the solution for the general forst order system is

$$\begin{aligned} A &= [A(o)-G]e^{-kt} + G \\ &= [1000-800]e^{-t/10} + 800 \\ &= +200e^{-t/10} + 800 \end{aligned}$$

This function of time can be represented graphically as in Figure 2-5.

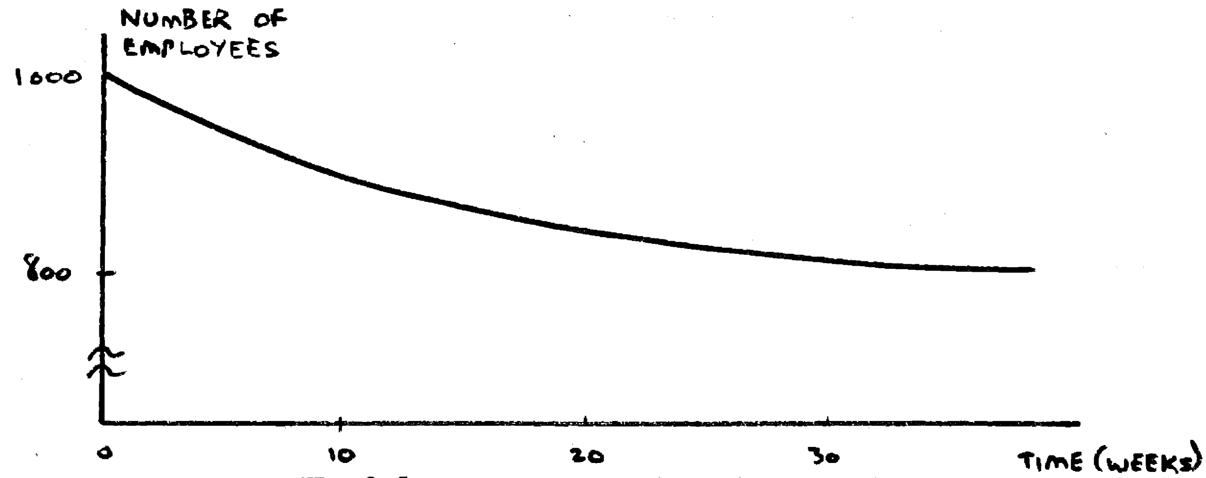


FIGURE 2-5: RESPONSE TO A NEGATIVE STEP INPUT

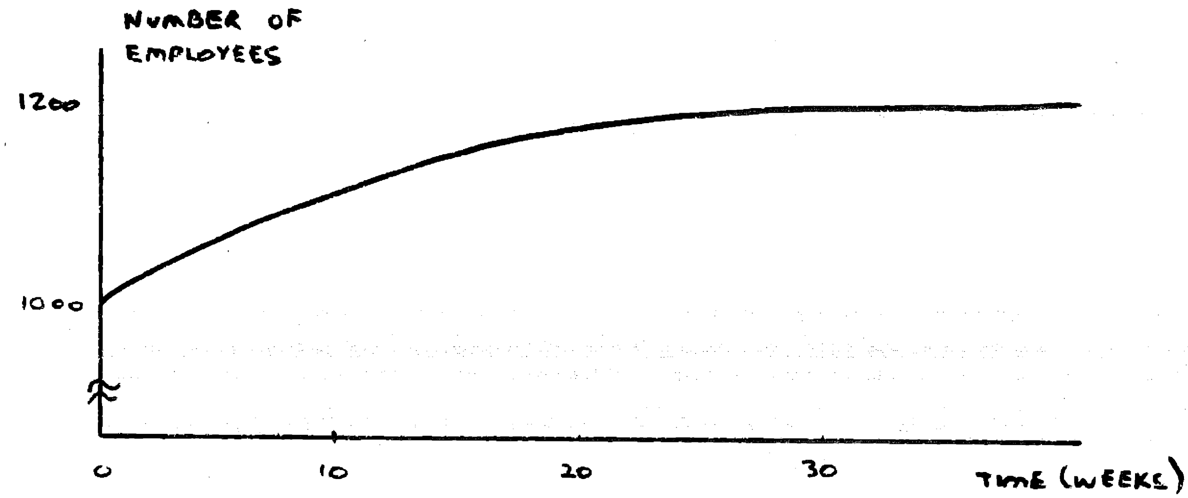


FIGURE 2-6: RESPONSE TO A POSITIVE STEP INPUT

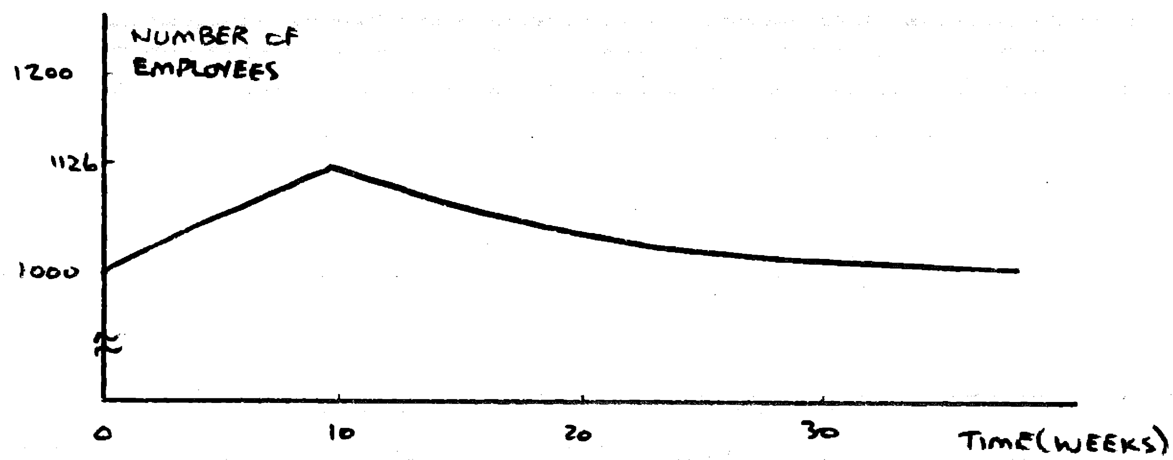


FIGURE 2-7: RESPONSE TO A PULSE INPUT

Note how the two components of the structural response are combined to create the net effect of the system moving from one equilibrium to another. Separately, the two components would move the size of the work force in opposite directions. If there were no desired level at all, the first term would discharge the 1000 people that were initially in the work force. Similarly, if there were no people in it initially, the second term would hire the entire 800. Taken together there is a net change of minus 200.

Similarly, it can be shown the effect of a sudden increase in the desired number of people in the work force.

$$G = 1200$$

$$A(o) = 1000$$

$$K = 1/10$$

$$A = [A(o)-G]e^{-kt} + G$$

$$= [1000-1200]e^{-t/10} + 1200$$

The number of employees as a function of time is shown graphically in Figure 2-6.

In this situation the net change in the number of men in the work force is positive, represented by a net structural response that is negative. As the magnitude of the structural response decreases with time, its sum with the driven response produces a total that increases with time. Note that it also fulfills the requirements of the system that the largest rate of hiring employees is in the beginning.

The Pulse Input

A pulse input implies a change that is temporary, a sudden change returning the desired number of employees to its original value after a step had originally increased it. If the return step occurs a long time after the first one and the structural component of the response has gone to zero, the problem can simply be handled by treating the pulse as a positive step followed by a negative step. If the time is relatively short between the changes, however, the actual value of the number of employees at the time of the second step must be determined. The state of the system at this time is then treated as an initial condition in evaluating the response of the system to the second step. For example, suppose management had increased the desired level of employment to 1200 as was shown in the last example. The actual level of employment was found to be:

$$A = -200^{-t/10}+1200$$

Suppose that management changes the desired level back to 1000 after 10 weeks. The first thing to do is to evaluate the above expression after 10 weeks.

$$A = -200e^{-10/10} + 1200$$

$$= -200\ (.368) + 1200$$

$$= 1126$$

This value of A is then used as the initial condition and the response after 10 weeks is determined as it would have for a negative step at time zero.

$$G \quad = 1000$$

$$A(o) = 1126$$

$$k \quad = 1/10$$

$$A \quad = [A(10)-G]e^{-kt} + G$$

$$= [1126-1000]e^{-t/10} + 1000$$

$$= 126e^{-t/10} + 1000 \qquad \text{for time} > 10 \text{ weeks}$$

The number of employees during both time periods is shown in Figure 2-7.

<u>The Ramp Input</u>

Business conditions might be such that the need for production capacity would increase or decline at a steady rate. Such an input to a system is called a ramp input, whether it is decreasing or increasing. If 10 additional employees were needed each week and at time zero we had 1000, then the driving function of the system would be represented by

$$G(t) = G_0 + G_1 t$$

$$= 1000 + 10t$$

The ramp function is the integral of the constant, the rate of change of the ramp being equivalent to the magnitude of the constant. As an integration is basically a summation operation, the superposition property of linear systems tells us that the application of an input that is an integration of another function will produce an output that is the integration of the output resulting from the other function. In other words, the response of the system to a ramp input can be obtained by integrating the response of a constant or step input with a magnitude equal to the slope of the ramp.

Assume that it is necessary to determine the response of the system when management sets a desired work force level that increases by 10 each week. By investigating the response when only an increase of ten employees is requested, how the structural response of the system affects the behavior of the employment level can be observed. Assume there are no employees initially. From the previous sections, the solution for A is

$$A = [A(o)-G]e^{-kt} + G$$

$$= [0-10]e^{-t/10} + 10$$

The response will start quickly and, although slowing down as it increases, will eventually reach the desired level of 10 employees. Now it is necessary to find the system response when the desired level is increased continuously at the rate of ten a week. The solution could be found by adding all the individual responses to each 10 employee increment, but to add the individual responses on a continuous basic the above function is simply integrated.

$$A_0 = \int_o^t \{[A(o)-G]e^{-kt} + G\}dt$$

$$= -1/k[A(o)-G]e^{-kt} + G]_o^t$$

$$= -1/k[A(o)-G](e^{-kt}-1) + Gt$$

$$= -10[0-10](e^{-kt}-1) + 10t$$

$$= 100e^{-t/10} - 100 + 10t$$

If the response were only the 10t term, the actual level of employees would be exactly equal to the desired level at all times. However, the combination of the exponential term and the negative constant means that the actual work force level lags behind, worsening rapidly initially and then stabilizing at a constant deficit of 100 workers. No matter how long the system attempts to catch up, its built-in decision policies will always keep a lag of 100 workers. This deficiency is understandable as the policy explicitly limits the personnel department to hiring only a fraction (in this case, 1/10th) of the deficiency in the current time period. In a continuously changing situation, the policy results in a number of employees that need to be hired, stabilizing as was shown after the structural response of the system has become negligible.

By separately evaluating the response of the constant component of the desired employment function and including the actual initial conditions of employment, the total response can be obtained. The response to a constant of 1000 with an initial employment of 1000 is

$$A_1 = [A(o)-G]e^{-kt} + G$$

$$= [1000-1000]e^{-t/10} + 1000$$

$$= 1000$$

There is no structural response component because the initial system equilibrium is not changing. The total system response is given by

$$A = A_0 + A_1$$

$$= 100e^{-t/10} - 100 + 10t + 1000$$

$$= 100e^{-t/10} + 10t + 900$$

Figure 2-8 shows the total system response along with the ramp input which the system is never able to overtake. A ramp input that stops after a certain length of time can be treated in the same way as the pulse input. By determining the actual level of employment at the time the desired level stops increasing, one can determine the system response to the new driving function using the actual level as an initial condition.

The Sinusoidal Input

There are some industries in which business conditions fluctuate with a fairly regular pattern, cycling through good and then bad conditions every few years. Such a situation can be approximated by using a sinusoidal function to describe management's desired level of employment.

Assume the desired level of employment is of the form

$$G(t) = G_0 + G_1 \cos (mt + n)$$

where m and n are arbitrary constants representing the coefficients controlling the length of the cycle and the value of the phase shift respectively. The

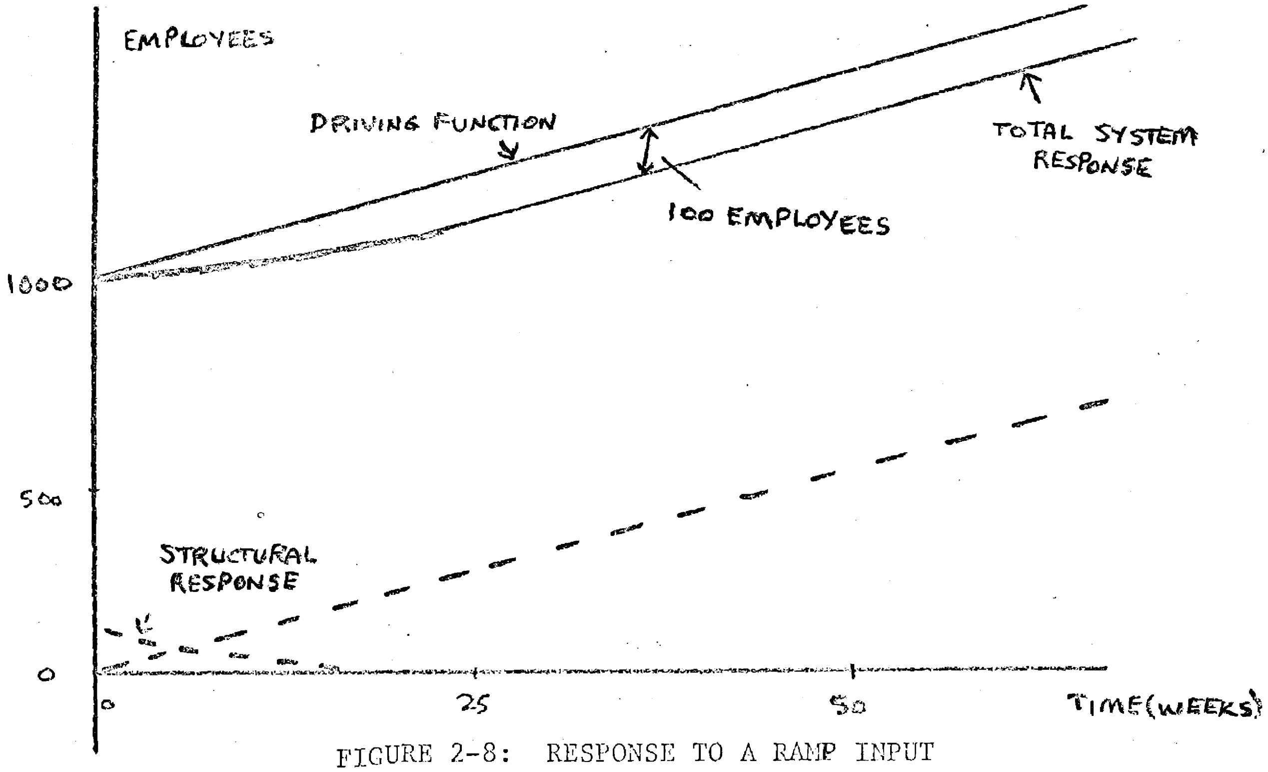


FIGURE 2-8: RESPONSE TO A RAMP INPUT

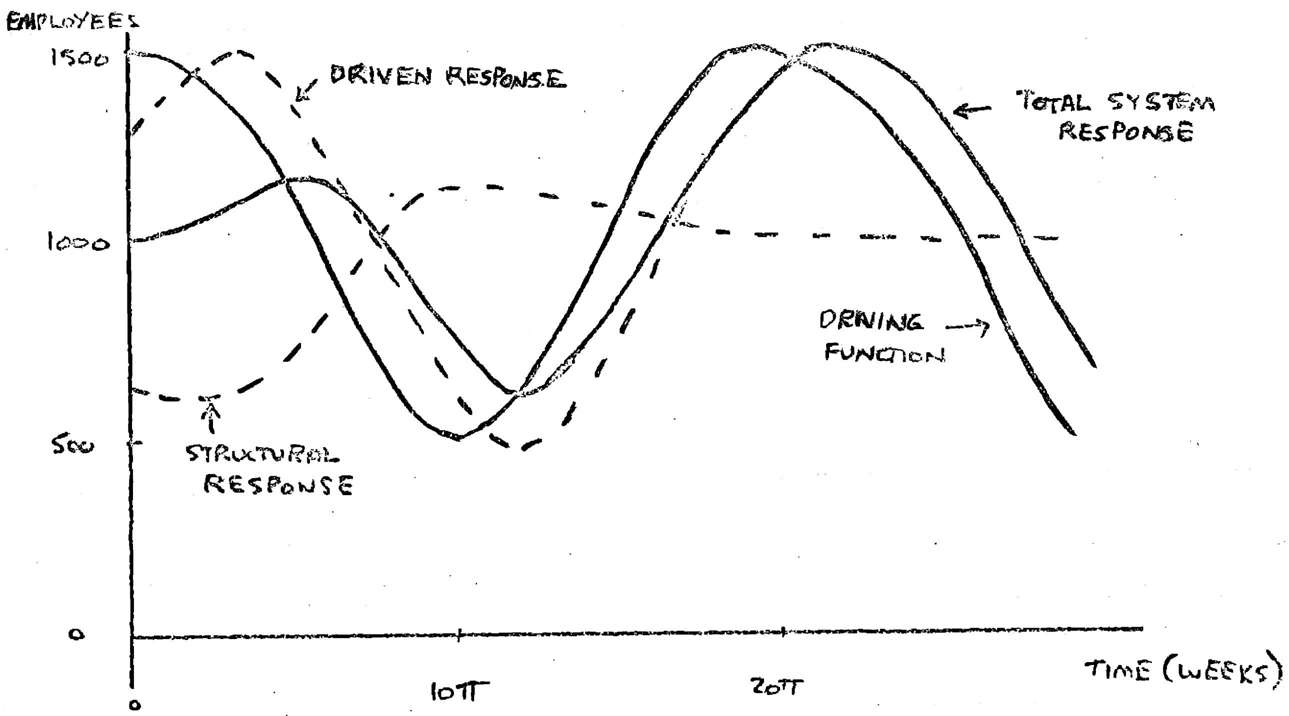


FIGURE 2-9: RESPONSE TO A SINUSOIDAL INPUT

constant m is actually equal to the inverse of the period multiplied by 2Π while n is the phase shift measured in radians, 2Π radians or 360 degrees being the phase shift that corresponds to a full cycle. The magnitude of the total function can vary between $G_0 + G_1$ and $G_0 - G_1$.

Because the cosine function is not composed of functions already evaluated, it will be necessary to determine the response of the system by direct evaluation of the general integral equation that describes a first order feedback system. The cosine function can be represented by complex exponentials, however, as follows.

$$G_1 \cos\ (mt + n) = G/2\ e^{j(mt + n)} + e^{-j(mt + n)}$$

This expression is readily derived from Euler's formula.[1] By considering one term of the cosine expression at a time the amplitude coefficient of the driven response component can be obtained. In the previous derivation R, representing the driven response, was substituted for A in the general integral equation giving

$$R + k\int_o^t Rdt = k\int_o^t Gdt + A(o)$$

R was assumed to be equivalent to XG, where X was an arbitrary coefficient. G in this case will be $G_1/2\ e^{j(mt + n)}$ and substituting these expressions, we have

---

[1]Readers unfamiliar with complex algebra are advised to review a summary such as Shearer, J. Lowen, Murphy, Arthur T., Richardson, Herbert H., Introduction to System Dynamics, (Reading, Mass, Addison-Wesley, 1967), Chapter 9.

$$XG_{1/2}e^{j(mt + n)} + k\int_o^t XG_{1/2}e^{j(mt + n)}\,dt = k\int_o^t G_{1/2}e^{j(mt + n)}dt + A(o)$$

which can be reorganized and integrated to provide

$$XG_{1/2}e^{jn}e^{jmt} + kXG_{1/2}e^{jn}\int_o^t e^{jmt}dt = kG_{1/2}e^{jn}\int_o^t e^{jmt}dt + A(o)$$

$$XG_{1/2}e^{jn}e^{jmt} + kXG_{1/2jm}e^{jn}e^{jmt}]_o^t = kG_{1/2jm}e^{jn}e^{jmt}]_o^t + A(o)$$

$$XG_{1/2}e^{jn}e^{jmt} + kG_{1/2\ jm}e^{jn}[e^{jmt}-1] = kG_{1/2jn}e^{jn}[e^{jmt}-1] + A(o)$$

By taking the derivative of both sides with respect to time,constant terms can be removed.

$$jmXG_{1/2}e^{jn}e^{jmt} + kXG_{1/2}e^{jn}w^{jmt} = kG_{1/2}e^{jn}e^{jmt}$$

Factoring out $G_{1/2}e^{jn}e^{jmt}$ results in

$$jmX + kX = k$$

$$X(jm+k) = k$$

$$X = \frac{k}{jm+k}$$

Therefore, the driven component of the response for this particular input is

$$R = XG_{1/2}e^{j(mt + n)}$$

$$= [\frac{k}{jm+k}]\ G_{1/2}e^{j(mt + n)}$$

Similarly, it can be shown that the response to the second exponential term is

$$R = [\frac{k}{-jm+k}]\ G_{1/2}e^{-j(mt + n)}$$

Together with the response to the constant term, the full driven response is represented by

$$R = G_0 + [\frac{k}{jm+k}] G_{1/2} e^{j(mt + n)} + [\frac{k}{-jm+k}] G_{1/2} e^{-j(mt + n)}$$

The coefficients in front of the exponential terms are complex numbers and, as such, have both magnitude and phase shift. By converting these coefficients to polar form,[1] we can reduce the exponentials to a cosine function once again.

$$R = G_0 + \frac{G_1}{2}[\frac{k}{\sqrt{m^2+k^2}} e^{j\tan^{-1}(m/k)}] e^{j(mt + n)} + \frac{G_1}{2} [\frac{k}{\sqrt{m^2+k^2}} e^{-j\tan^{-1}(m/k)}]$$

$$= G_0 + \frac{KG_1}{2\sqrt{m^2+k^2}} \{e^{j[mt+n+\tan^{-1}(m/k)]} + e^{-j[mt+n+\tan^{-1}(m/k)]}\}$$

$$= G_0 + \frac{kG_1}{\sqrt{m^2+k^2}} \cos[mt+n+\tan^{-1}(m/k)]$$

The above expression shows that the system has attenuated the magnitude of the driving function and introduced additional phase shift (or lag). Note that both the attenuation and phase shift depend on the relative magnitude of the structural parameter k to the coefficient m that is inversely proportional to the period of the driving function. As m gets larger, or as the period gets smaller, theamplitude of the response decreases and the phase shift increases until they reach limits of zero and $\Pi/_2$ radians respectively. The same result occurs, of course, if k becomes smaller relative to m. In the example, if k becomes smaller the fraction of the people hired is so small that it is impossible to keep up with the rapidly changing desired work force figure. As k

---

[1]These concepts will have to be developed as they have been in Introduction to System Design, Ibid.

becomes very small, the desired level will have gone through a cycle before a significant number of people have been hired. The change in direction of the driving function will then have the personnel department firing some people but not enough to follow the rapid change in direction in desired level once again. Thus, the component of the total system response that has the same time behavior as the driving function will be very small under such conditions.

A particular example will be examined using the values for the personnel-work force system that was used for other inputs. Suppose the desired level of employment is given by

$$G = 1000 + 500 \cos(t/_{10})$$

Using the value of k as 1/10 results in a driven response component of

$$R = 1000 + \frac{.1(500)}{\sqrt{(.1)^2+(.1)^2}} \cos[t/_{10} + \tan^{-1}(.1/.1)]$$

$$= 1000 + (500/\sqrt{2}) \cos(t/_{10} + \Pi/4)$$

Using the equation for the total response to a first order feedback system results in the following solution.

$$A = [A(o) - R]e^{-kt} + R$$

$$= [1000 - 1000 - (500/\sqrt{2}) \cos(t/_{10} + \Pi/4)]e^{-t/_{10}}+1000+(500/\sqrt{2}) \cos(t/_{10} + \Pi/4)$$

$$= -(500/\sqrt{2}) \cos(t/_{10} + \Pi/4)e^{-t/_{10}}+1000+(500/\sqrt{2}) \cos(t/_{10} + \Pi/4)$$

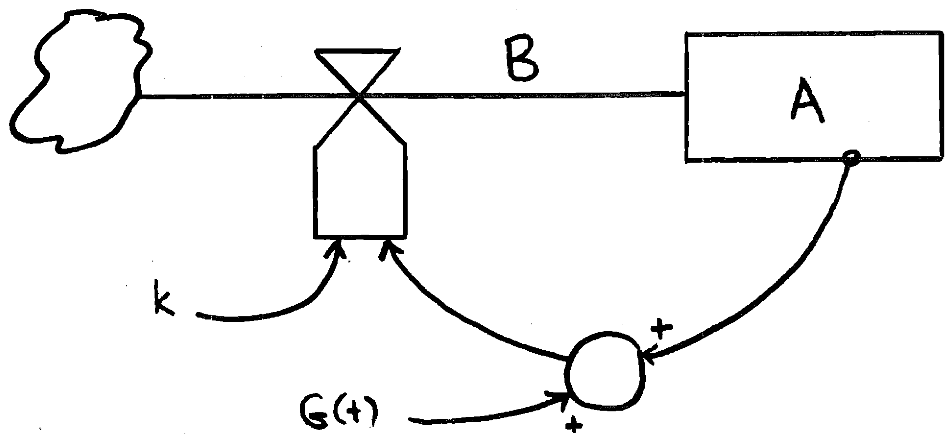
There is a structural component of the cosine function that completely cancels the driven response at time zero and retards its amplitude until the system has had time to establish a new equilibrium. This is shown clearly in Figure 2-9.

The envelope of the cosine response is modified by the exponential response of the system structure, due to the fact that the cosine did not exist before time zero. As in the case of the ramp input the actual level of employment is never able to overtake the desired level, lagging behind with a phase shift of 1/8th of a period ($\Pi/4$ radians) and a corresponding reduction in amplitude. Again it is the policy limiting the number of people hired and fired to a fraction of the amount that is deficient or in excess at any particular point in time that causes this behavior. As can be seen in the general expression for the response, the fraction k is the significant system parameter.

Feedback Polarity

The interdependence of variables that we have examined in the first order feedback systems so far in this chapter has been such that the level has been compared to an external variable to determine a corrective action. In doing this the level that has been fed back has had a negative sign associated with it, in order that the corrective action would be such as to move it in the direction of the desired level. This is known as negative feedback and is a common structure in systems where an external variable is being used to control the system variables. The structural response tends to be short term in nature and the driving function usually produces the dominant component of the total system response.

Another class of systems use positive feedback, resulting in behavior which is dominated by the structural response instead of the external variable. Figure 2-10 illustrates the general first order feedback system once again, although this time the sign associated with the feedback path from the level is positive. The general solution for such a system can be derived in much the same manner as for the negative one.



FLOW DIAGRAM OF A FIRST ORDER POSITIVE FEEDBACK SYSTEM

FIGURE 2-10

$$A = \int_0^t Bdt + A(o)$$

This time B = G + A, the sum of the feedback term and the external variable. Therefore,

$$A = k\int_0^t (G+A)dt + A(o)$$

which can be regrouped to provide

$$A - k\int_0^t Adt = k\int_0^t Gdt + A(o)$$

If the solution is again assumed to be the form of

$$A = A_s + A_d = Qe^{qt} + R$$

the unknown coefficients can be determined. Initially, R will be substituted and then replaced with XG to determine the gain coefficient X of the driven response of the system. Assume that G is a constant.

$$R - k\int_o^t Rdt = k\int_o^t Gdt + A(o)$$

$$XG - k\int_o^t XGdt = k\int_o^t Gdt + A(o)$$

$$XG - kXGt]_o^t = kGT]_o^t + A(o)$$

$$XG - kXGt = kGt + A(o)$$

Taking the derivative of both sides, we have

$$-kXG = kG$$

$$X = -1$$

Therefore,

$$R = -G$$

indicating that the driven response of the positive feedback system is the inverse of the driving function at the input.

To determine the structural response, the general integral equation is used with G at zero.

$$A - k\int_o^t Adt = A(o)$$

Substituting $A_s = Qe^{qt}$ results in

$$Qe^{qt} - k\int_o^t Qe^{qt}\,dt = A(o)$$

$$Qe^{qt} - Qk/q[e^{qt}-1] = A(o)$$

Taking the derivative of both sides removes the constant terms.

$$qQe^{qt} - Qke^{qt} = 0$$

Now Qe^{qt} can be factored out, leaving

$$q - k = 0$$

$$q = k$$

The amplitude of the structural response can be obtained by substituting the total expression for A in the general integral equation.

$$A - k\int_o^t Adt = k\int_o^t Gdt + A(o)$$

$$Qe^{qt} + R - k\int_o^t [Qe^{qt}+R]dt = k\int_o^t Gdt + A(o)$$

By evaluating this at time zero, the integral terms disappear and the one remaining exponential becomes zero.

$$Q + R = A(o)$$

Substituting the value for R

$$Q - G = A(o)$$

$$Q = A(o) + G$$

The complete solution is then given as

$$A = Qe^{qt} + R$$

$$= [A(o)+G]e^{kt}-G$$

Note that the structural component due to the driving function adds its effect to the initial condition portion of the structural response, as contrasted with its subtraction in the negative feedback case. The external variable thus reinforces the structural response which, because of the positive coefficient of time in the exponential, is the dominant component of the total system response.

The behavior of this system can be examined in more detail by considering a specific example. The relationship between the number of salesmen a company can afford to hire and the sales volume the firm is handling can be approximated by such a system. The level can be used to represent the number of salesmen that work for the firm at a particular point in time. In that sales volume will be proportional to the number of salesmen covering the market, the level will be one of the terms fed back to affect the company's total sales. The other termwill be the market conditions that arise external to the system, factors such as the state of the economy. The two of these will be added together to produce the company sales volume and from this management policy will determine how many new salesmen can be hired per month. Assume that management limits this hiring rate to a fraction of the number that can be afforded from the profits of the current sales volume. Referring to Figure 2-10, the level of salesmen will be A, the hiring rate B, the market conditions G, and the fraction that can be afforded k.

Assume that we start initially with 1000 salesmen, that each salesman can average sales of 1000 units per month and that the profits from 1000 units could pay for hiring another salesman. Thus each salesman can sell enough in one month to justify hiring one additional salesman. In addition assume that management decides to hire only 1/10th as many salesmen as they could afford to, and that market conditions have no effect on the system in this first case. The general solution tells us that

$$A = [A(o)+G]e^{kt} - G$$

$$= [1000+0]e^{t/10} - 0$$

$$= 1000e^{t/10}$$

Thus, the number of salesmen employed by the firm will increase at an ever-increasing rate, assuming no future effects from market conditions. After 10 months, the number of salesmen will be equal to $1000e^{10/10} = 2700$. The behavior of the system for this situation is shown in Figure 2-11(a). Consider also the case where the market conditions suddenly provide sales which allow for hiring an extra 500 salesmen. Then

$$G = 500$$

$$A(o) = 1000$$

$$k = 1/10$$

and

$$A = [A(o)+G]e^{kt} - G$$
$$= (1000+500)e^{kt} - 500$$
$$= 1500e^{kt} - 500$$

The behavior of the system in this case is illustrated in Figure 2-11(b). Note that the sudden but constant shift in market conditions at time zero results in an increased rate of growth over the entire time period, not just a shift in the number of salesmen.

Needless to say, market conditions can also have a negative influence on the growth momentum of the firm's sales organization. If a recession were to cause a drop in volume that necessitated cutting back 500 salesmen, the growth rate would be depressed.

$$G = -500$$
$$A(o) = 1000$$
$$k = 1/10$$

and

$$A = [A(o)+G]e^{kt} - G$$
$$= (1000-500)e^{t/10} + 500$$
$$= 500e^{t/10} + 500$$

Figure 2-12(a) shows the resulting decline in growth rate. If the recession indicated that the present sales force should be cut back to zero, the pressure from the sales force attempting to grow would exactly balance the pressure from the market condition to keep the number of salesmen constant. In the equation the structural response would simply disappear. The balance would

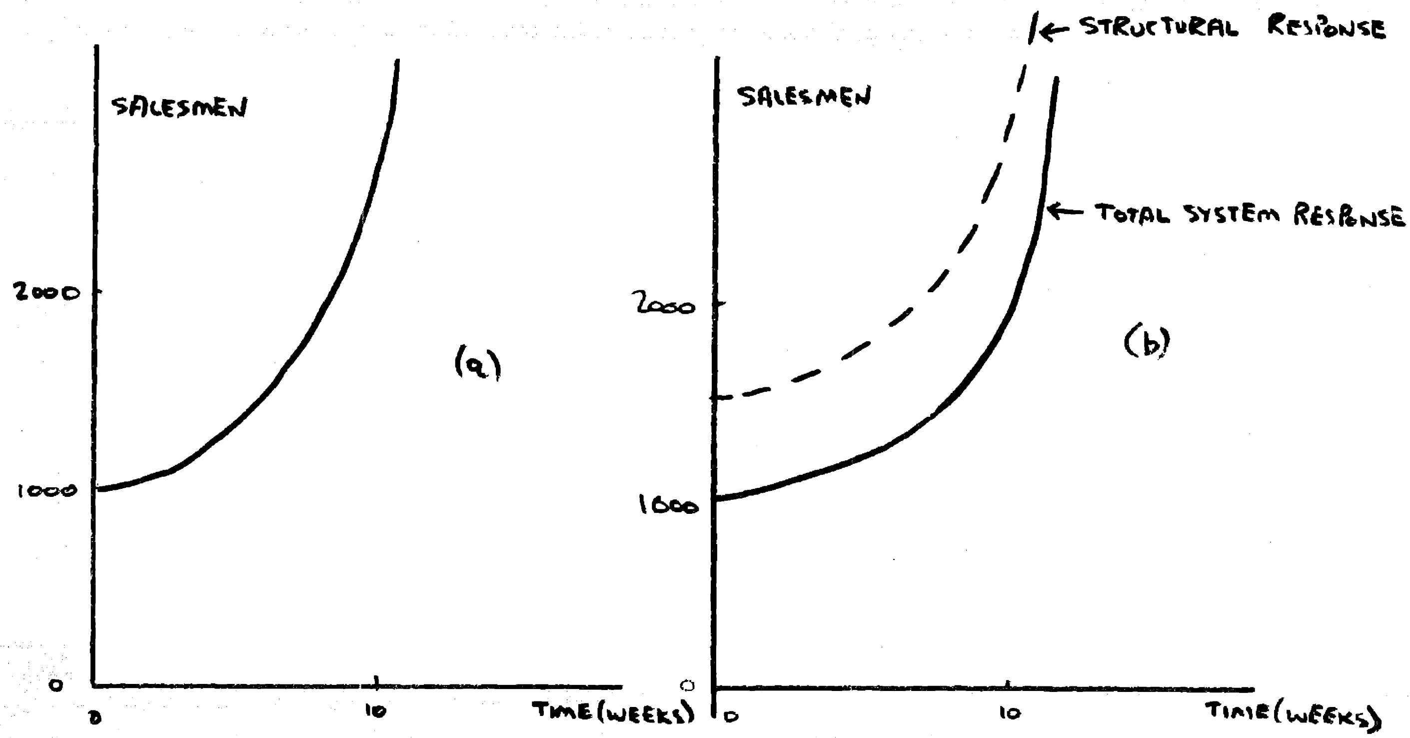


FIGURE 2-11: RESPONSE OF A POSITIVE FEEDBACK SYSTEM TO A POSITIVE STEP INPUT

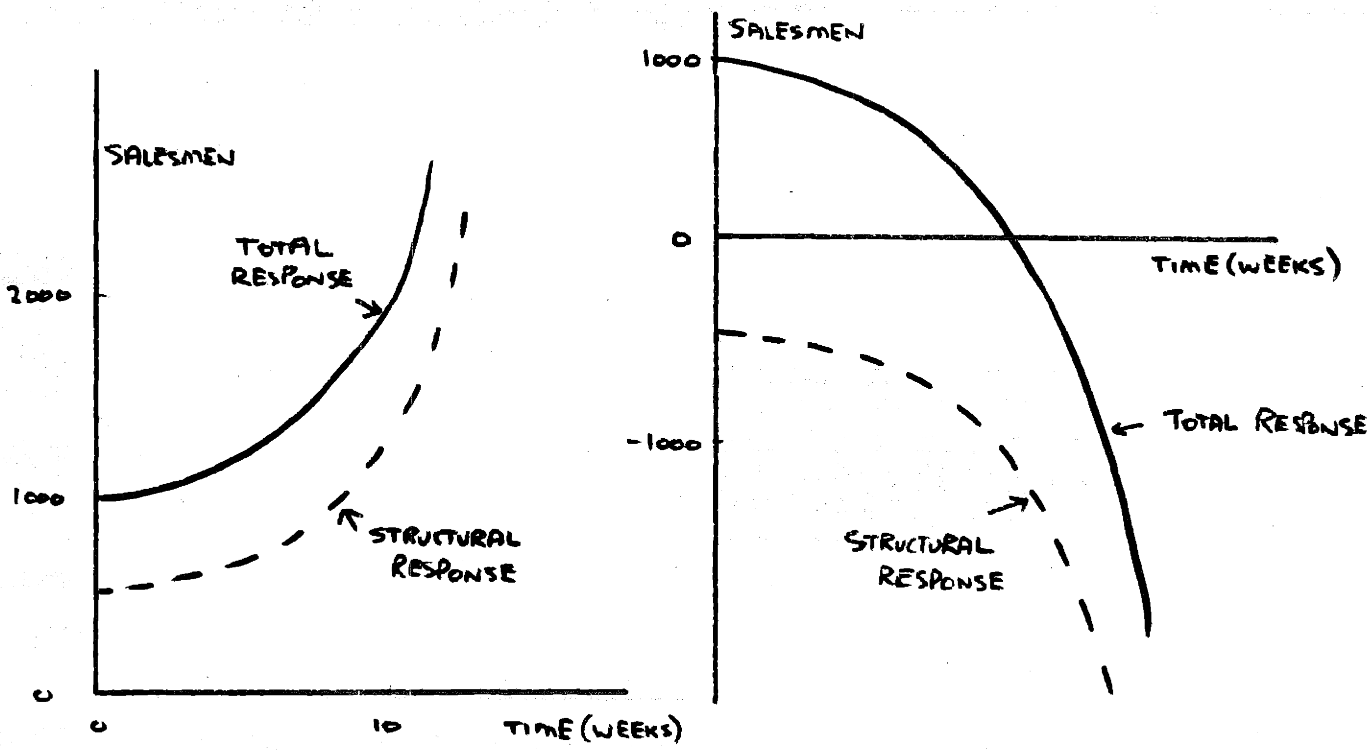


FIGURE 2-12: RESPONSE OF A POSITIVE FEEDBACK SYSTEM TO A NEGATIVE STEP INPUT

be a tenuous one, however, as any small additional changes in the market condition would cause the system to start growing once again. As an absurd example consider a recession that is severe enough to indicate that more salesmen should be cut back than the company currently is employing. This by itself, of course, unrealistic but when balanced against the growth pressures exerted by the salesmen, it causes a number of salesmen to be laid off. Suppose market conditions were such that

$$G = -1500$$

$$A(o) = 1000$$

$$k = 1/10$$

and

$$A = [A(o)+G]e^{kt} - G$$

$$(1000-1500)e^{t/10} + 1500$$

$$-500e^{t/10} + 1500$$

The behavior of the system in this instance is shown in Figure 2-12(b). The unusual result of this situation is that the growth process is operating in reverse. As the number of salesmen decreases, the ones that are remaining are less able to resist the adverse market condition and company sales volume begins to decline at an even greater rate. Thus, systems with positive feedback structure can provide structural responses that either increase or decrease the system variables, depending entirely on the current state of the system variables relative to the external variables. One can begin to appreciate the importance of resisting temporary external forces while a growth process is in its early stages. At a later point in time the system will have grown to such a level that the external disturbances will be small relative to the internal variables and the growth pattern cannot be affected significantly.

CHAPTER III

SECOND ORDER FEEDBACK SYSTEMS

Second order feedback systems are systems that contain two levels or integrations, as contrasted to the one level (first order) systems studied in the previous chapter. The increase in the number of levels introduces an increase in the complexity of the behavior that can be exhibited. The structural response can perform in an oscillatory manner as well as with the exponential behavior shown by the first order systems, depending upon the values of the structural parameters selected. In addition second order systems also introduce the possibility of variations in system structure itself. Configurations will be examined which have two feedback loops, one nested within the other. In that the two loops share one level, the entire system is still only second order. Examples of such systems including variations in polarity of the two loops will be considered in this chapter. Initially the general solution for a one-loop, second order system will be developed.

Derivation of the Second Order Feedback System Response

As in the case of first order systems, there are three components in the system response of a second order system. There is a structural response consisting of two components and a driven response. The structural response is divided into a response arising from the initial values of the two levels in the system and the shift in system equilibrium that is required to compensate for the change in the external variable or driving function. The driving function also results in a component of the response which, in a unity feedback system, reproduces the driving function in form. This latter component is what has been defined as the driven response. The procedure for deriving the analytic solution of a second order system is basically the same as for the first order system, introducing some extra complexity and a few extra steps in some of the mathematical operations.

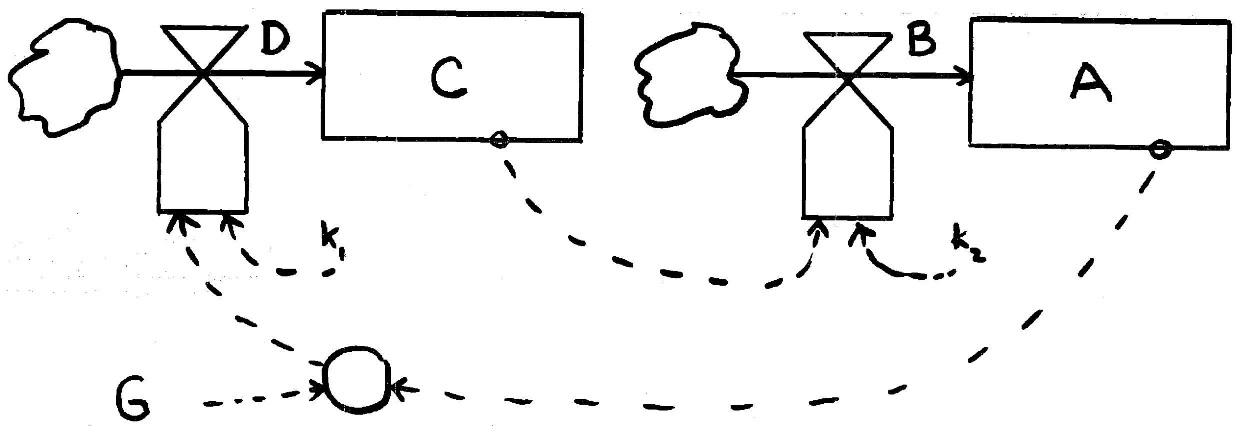


FIGURE 3-1: FLOW DIAGRAM OF A SECOND ORDER FEEDBACK SYSTEM

FIGURE 3-1

The simplest form of second order system, utilizing a single feedback loop, is illustrated in Figure 3-1. The information fed back from level A is compared to the external variable and is now used to control the flow into level C, instead of into level A itself as in the first order case. The value of level C is now used to directly control the flow into level A. Each controlled flow has its own value of the system parameter k associated with it, resulting in a net gain around the entire feedback loop that is dependent on both values. That is, if both values of k are 1/10th, then the total fractional adjustment that will be made around the loop is 1/100th.

The general integral equation for this system can be developed as follows:

$$A = \int_o^t B dt + A(o)$$

Substituting

$$B = k_1 C$$

we have

$$A = \int_o^t k_1 C \, dt + A(o)$$

The level C itself is given by an equation that is exactly the same as the one for level A.

$$C = \int_o^t D dt + C(o)$$

Substituting this into the equation for A results in

$$A = k_1 \int_o^t \int^t D dt + k_1 \int_o^t C(o) dt + A(o)$$

The rate D is in turn defined by

$$D = k_2 (G-A)$$

which when substituted into the equation for A provides

$$A = k_1 k_2 \int_o^t \int (G-A) dt + k_1 \int_o^t C(o) dt + A(o)$$

Regrouping to place all the terms containing A on the left hand side results in

$$A + k_1 k_2 \int_o^t \int A dt = k_1 k_2 \int_o^t \int G dt + k_1 \int_o^t C(o) dt + A(o)$$

The initial value of level C is explicitly integrated in this equation even though it is a constant. This is consistent in that level C directly controls

B, the inflow rate to level A. If C has an initial value at time zero, it will result in an initial inflow rate which, although it has no immediate effect on A, will be present as a separate component of B along with the component that is due to the changes in C after time zero.

Although the left hand side of the equation for A contains a double integral the procedure for developing the solution is the same as in the case of a first order system. Specifically, the solution is still of the form

$$A = A_s + A_d = Q^{qt} + R$$

where Q, q, R are coefficients that need to be determined. As in the previous chapter the first term represents the total structural response and the second, the driven response. Again, because of superposition, the terms of the solution are able to be considered for A separately and the evaluation is begun by substituting R for A.

$$R + k_1 k_2 \int_o^t \int R dt = k_1 k_2 \int_o^t \int G dt + k_1 \int_o^t C(o) dt + A(o)$$

Again, because direct feedback systems are being dealt with and the driven response has been defined to be of the same form as the driving function, XG can be substituted for R, X being a coefficient.

$$XG + k_1 k_2 \int_o^t XG dt = k_1 k_2 \int_o^t \int G dt + k_1 \int_o^t C(o) dt + A(o)$$

The integral of XGdt will depend on what function of time is described by the driving function, G. Assume that it is a constant for this evaluation but keep in mind that the integral must be reevaluated for other functions of time. The first integration of a constant provides

$$XG + k_1 k_2 \int_o^t XGt]_o^t \, dt = k_1 k_2 \int_o^t Gt]_o^t \, dt + k_1 C(o) \; t]_o^t + A(o)$$

which must be evaluated at time t and time zero to obtain

$$XG + k_1 k_2 \int_o^t XG(t-0)dt = k_1 k_2 \int_o^t G(t-0)dt + k_1 C(o)(t-0)$$

Integrating a second time and evaluating at the boundaries, we obtain

$$XG + k_1 k_2 XGt^{2/2} = k_1 k_2 Gt^{2/2} + k_1 C(o)t + A(o)$$

To determine the coefficient X, the derivative must be taken with respect to time of both sides of the equation to eliminate the constants.

$$k_1 k_2 XGt = k_1 k_2 Gt + k_1 C(o)$$

As one constant is still remaining, it is necessary to take the derivative once again

$$k_1 k_2 XG = k_1 k_2 G$$

which can be reduced to

$$X = 1$$

As in the case of the first order systems the driven response to a constant (or step input) is the input itself. That is,

$$R = G$$

The next step is to determine the time constant, q, of the exponential in the structural response. This can be accomplished by considering the effect of the initial condition only, setting the driving function G to zero. The general equation becomes

$$A + k_1 k_2 \int_o^t Adt = k_1 \int_o^t C(o)dt + A(o)$$

Substituting the structural component of the solution Qe^{qt} results in

$$Qe^{qt} + k_1 k_2 \int_o^t \int Qe^{qt} = k_1 \int_o^t C(o)dt + A(o)$$

The first integration results in

$$Qe^{qt} + k_1k_2\int_o^t Q[e^{qt}]_o^t\ dt = k_1C(o)\ t]_o^t + A(o)$$

which is evaluated at time t and time zero to obtain

$$Qe^{qt} + k_1k_2\int_o^t Q/q(e^{qt}-1)\ dt = k_1C(o)t + A(o)$$

Now the second integration must be taken on two terms, one of which is the constant resulting from evaluating the exponential at time zero. The result of this is

$$Qe^{qt} + k_1k_2Q/q^2[e^{qt}]_o^t - k_1k_2Q/q\ t]_o^t = k_1C(o) + A(o)$$

which is evaluated to provide

$$Qe^{qt} + k_1k_2Q/q^2(e^{qt}-1) - k_1k_2(Q/q)t = k_1C(o) + A(o)$$

The constants may be removed from the expression by taking the derivative of both sides twice

$$qQe^{qt} + k_1k_2Q/q\ e^{qt} - k_1k_2Q/q = k_1C(o)$$

$$q^2\ Qe^{qt} + k_1k_2\ Qe^{qt} = 0$$

Qe^{qt} can be factored out leaving

$$q^2 + k_1k_2 = 0$$

$$q^2 = -k_1k_2$$

$$q = \pm\sqrt{-k_1k_2}$$

$$= \pm j\sqrt{k_1k_2}$$

There are two solutions to the equation for q, a fact characteristic of second order equations containing double integrals. The structural component of the solution is then given by

$$A_s = Q_1e^{qt} + Q_2e^{q_2t}$$

instead of a single exponential. As the coefficients are imaginary numbers and are different only in sign in this negative feedback case, the sum of the two exponentials will result in oscillatory behavior in the structural response. In fact, as there is no real component in the exponential coefficients, the oscillation will neither grow nor decay but will continue on indefinitely. This will be explored in more detail in the example given later in this chapter.

By substituting into the general equation the expression for A containing the constants that have been determined so far, we can determine Q_1 and Q_2.

$$A + k_1 k_2 \int_o^t \int Adt = k_1 k_2 \int_o^t \int Gdt + k_1 \int_o^t C(o)dt + A(o)$$

$$Q_1 e^{qt} + Q_2 e^{q_2 t} + R + k_1 k_2 \int_o^t \int (Q_1 e^{q_1 t} + Q_2 e^{q_2 t} + R)dt = k_1 k_2 \int_o^t \int Gdt + k_1 \int_o^t C(o) + A(o)$$

By evaluating the expression at time zero, the integral terms become zero and the two remaining exponential terms become unity, resulting in

$$Q_1 + Q_2 + R = A(o)$$

However, there are still two unknowns, so a second equation that can be evaluated at time zero must be obtained. Taking the derivative of the general equation with the expression for A substituted provides

$$q_1 Q_1 e^{q_1 t} + q_2 Q_2 e^{q_2 t} + k_1 k_2 \int_o^t (Q_1 e^{q_1 t} + Q_2 e^{q_2 t} + R)dt = k_1 k_2 \int_o^t Gdt + k_1 C(o)$$

Evaluating this at time zero results in

$$q_1 Q_1 + q_2 Q_2 = k_1 C(o)$$

These two equations must be solved simultaneously to determine Q_1 and Q_2.

Multiplying the first equation by q_2 results in

$$q_2 Q_1 + q_2 Q_2 + q_2 R = q_2 A(o)$$

Subtracting this from the second equation obtains

$$q_1 Q_1 - q_2 Q_1 - q_2 R = k_1 C(o) - q_2 A(o)$$

regrouping results in

$$Q_1 \, (q_1 - q_2) = k_1 C(o) - q_2 A(o) + q_2 R$$

and substituting the values for the coefficients already obtained provides

$$Q_1 [j \sqrt{k_1 k_2} - (-j \sqrt{k_1 k_2})] = k_1 C(o) + j \sqrt{k_1 k_2} \, A(o) - j \sqrt{k_1 k_2} \, G$$

$$Q = \frac{-k_1}{2j\sqrt{k_1 k_2}} C(o) + \frac{A(o)}{2} - \frac{G}{2} = \frac{-1}{2j} \sqrt{\frac{k_1}{k_2}} \, C(o) + \frac{A(o)}{2} - \frac{G}{2}$$

From the second equation, Q_2 can be shown to be

$$Q_2 = 1/q_2 [k_1 \, C(o) - q_1 Q_1]$$

the value of Q_1 and the other coefficients are substituted in the second equation providing

$$Q_2 = 1/-j\sqrt{k_1 k_2} \, [k_1 C(o) - j\sqrt{k_1 k_2} \, (\frac{k_1}{2j\sqrt{k_1 k_2}} \, C(o) + \frac{A(o)}{2} - \frac{G}{2})]$$

$$= -1/j\sqrt{k_1 k_2} \, [k_1 C(o) - \frac{K_1 C(o)}{2} - \frac{j\sqrt{k_1 k_2}}{2} A(o) + \frac{j\sqrt{k_1 k_2}}{2} \, G]$$

$$= \frac{1}{2j} \sqrt{\frac{k_1}{k_2}} \, [C(o) + \frac{A(o)}{2} - \frac{G}{2}]$$

Substituting the expressions obtained for the coefficients into the expression for the solution results in

$$A = Q_1 e^{q_1 t} + Q_2 e^{q_2 t} + R$$

$$= [\frac{1}{2j}\sqrt{\frac{k_1}{k_2}}\ C(o) + \frac{A(o)}{2} - \frac{G}{2}]e^{j\sqrt{k_1 k_2}} + [\frac{-1}{2j}\sqrt{\frac{k_1}{k_2}}\ C(o) + \frac{A(o)}{2} - \frac{G}{2}]e^{-j\sqrt{k_1 k_2}}$$

$$+ G$$

The total solution in this case contains seven components, six of which are part of the structural response. These in turn can be subdivided into those which are related to the initial conditions of the two levels and those which are the structural response required to readjust the system equilibrium for the driving function G. If C(o) were equal to zero, the solution could be quite simply reduced to

$$A = [A(o) - G] \cos\sqrt{k_1 k_2}\ t + G$$

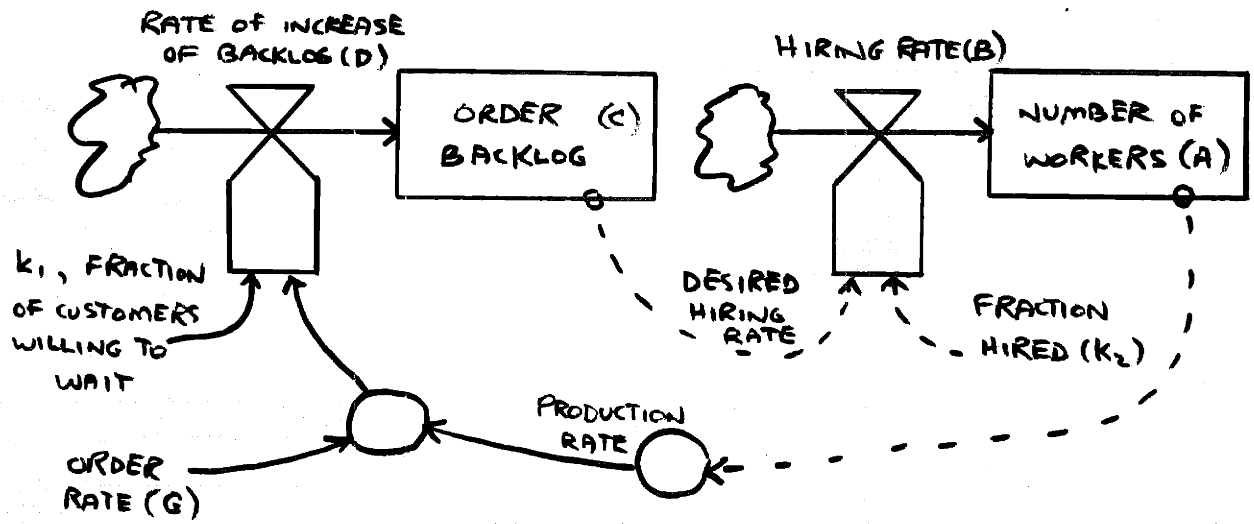
through the use of the formula

$$\cos xt = 1/2\ (e^{jxt} + e^{-jxt})$$

This simplification shows the structural response as performing in a purely oscillatory manner, with no growth or decay. When C(o) assumes non-zero values the behavior is basically the same except that phase shift is introduced into the cosine term and the amplitude is a more complex expression.

A Management System Example

The personnel department - work force example that was used to illustrate a first order negative feedback system in the previous chapter can be expanded to a second order feedback system conveniently. The two levels in the expanded system will be the number of men in the work force and the order backlog, which in turn will be used to control the hiring rate for the work force. Figure 3-2 illustrates the management system, the number of workers corresponding to level A and the size of the order backlog corresponding to level C in the general response for a second order system. The driving function G will be the order rate which will be determined entirely as a function of external market conditions. The rate at which the order backlog (D) is increased (or decreased) will be a function of what the difference is between the order rate and the plant production rate. When production is lower than the order rate, backlog will increase. The production rate is a direct function of the number of men in



FLOW DIAGRAM OF A PERSONNEL DEPARTMENT-WORK FORCE SYSTEM

<u>FIGURE 3-2</u>

the work force in this particular company. Not all of the excess orders will be used to increase the backlog, however, as some of the customers will buy from another firm if they cannot get immediate delivery.[1] The fraction of the excess orders that are added to the backlog is the structural parameter k_1. The size of the backlog at any point in time is used in this firm to

[1] The realism of the example has been sacrificed to provide an opportunity to examine this particular system structure. More work is needed in developing appropriate examples.

determine the desired hiring rate for the plant work force. Not all of the workers are hired in the current time period, however, as the personnel department has a policy of only hiring a portion of those needed to limit the strain on the training channel in the firm. This fraction of the desired hiring rate that is used to control the actual hiring rate (B) is the structural parameter k_2. As workers are hired the firm's production rate goes up and, if the order rate stays constant, the backlog will be decreased, also decreasing the hiring rate itself.

Consider a specific example with the following values:

$A(o)$	$= 1000$	the initial number of workers
$C(o)$	$= 0$	the initial backlog level
G	$= 800$	the number of workers required to satisfy the order rate
k_1	$= 1/2$	the fraction of customers willing to wait for delivery
k_2	$= 1/4$	the fraction of the desired hiring rate that is the current hiring rate

From our general solution for the second order negative feedback system, the plant work force is the following time function.

$$A = [A(o)-G] \cos\sqrt{k_1 k_2} + G$$

$$= (1000-800) \cos (t\sqrt{8}) + 800$$

$$= 200 \cos .354t + 800$$

Note that the oscillation is entirely a function of the two system parameters, k_1 and k_2, and that either one can be increased to decrease the period with which the system oscillates. Note also that the presence of the oscillation depends on A(o), the initial number of workers, and G, the order rate, having different values, but it makes no difference which is greater except to determine the initial direction the system variables move. If the order rate remains constant the oscillation will continue indefinitely. As the production rate and the order rate come into balance the backlog will be nonzero and when the backlog is zero, the production rate and the order rate will be different. The only way in which the oscillation can be controlled within the existing policies in the system is for the order rate to change to exactly equal the production rate at the point where the backlog becomes zero. However, it would be a marginal situation in that any further change in the order rate would be enough to start the oscillation once again.

This example is not intended to demonstrate a realistic situation. Obviously management would not continue to operate with policies that exhibited behavior of this type, but would tend to introduce tighter control of individual variables. The system that has been analyzed uses a hiring rate that is only a function of the order backlog and is very indirectly a function of the actual work force. In that an increase in hiring will increase the number of men in the work force, the production rate will increase. Over a period of time this will decrease the backlog of orders, resulting in a reduction of the hiring rate. The delay of the feedback in passing through the additional level, order backlog, is sufficient to cause the instability. Typically, company management would control the level of employees in the work force more directly by hiring at a rate that was a direct function of the difference between the present level and some desired level determined by other factors. The simple first order system for control of the level of work force that was illustrated in the previous chapter is most appropriate for use within the enlarged system boundary required by the example here. Management models in which some of the feedback loops are contained within others are most representative of the real situation. Such a configuration can be identified as a nested feedback loop system and will be discussed in detail in the following sections.

Nested Second Order Feedback Systems

Many real control systems contain one feedback loop within the other. This is most apparent in management hierarchy where each manager measures and evaluates the performance of his subordinates and each subordinate does the same for those who are further subordinated to him. Each level of control can be conceptualized as a negative feedback loop and the firm as a whole is a series of loops, one nested within the other. The number of loops that must be included in a model to provide an accurate representation of the process will depend on the particular problem. In this section the general solution to a two loop second order system will be developed and several of the examples which have been discussed so far will be investigated in terms of the more accurate nested, two-loop model.

The general structure for a nested loop with the second loop controlling the inflow to level A is shown in Figure 33. The general integral equation for this system can be developed as follows.

$$A = \int_o^t Bdt + A(o)$$

In this system B is given as $k_2(C-A)$. Therefore,

$$A = \int_o^t k_2 (C-A)dt + A(o)$$

Again, $C = \int_o^t Ddt + C(o)$

Substituting this into the equation for A provides

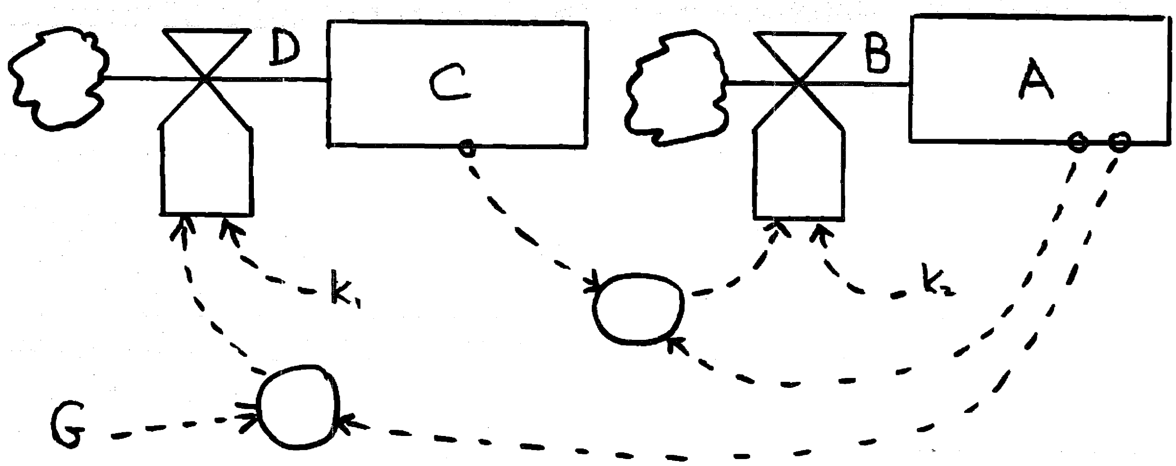
$$A = k_2 \int_o^t \int Ddt + k_2 \int_o^t C(o)dt - k_2 \int_o^t Adt + A(o)$$

The rate D is defined by $D = k_1 (G-A)$. Substituting this into the equation results in

$$A = k_1 k_2 \int_o^t \int (G-A)dt + k_2 \int_o^t C(o)dt - k_2 \int_o^t Adt + A(o)$$

Regrouping provides

$$A + k_2 \int_o^t Adt + k_1 k_2 \int_o^t \int Adt = \int_o^t \int Gdt + k_2 \int_o^t C(o)dt + A(o)$$



FLOW DIAGRAM OF A NESTED SECOND ORDER FEEDBACK SYSTEM

FIGURE 3-3

The addition of the inner feedback loop has added one term on the left hand side of the equation. Note that the single integral of A is the term that is created by the presence of the inner loop and the double integral term results from the outer loop. The driven response of this system is the same as in the single loop second order system studied previously, R = G. As this is a unity feedback system (the system variable A is compared directly to the driving function G) it is understandable that the driven response is unchanged. No matter what responses occur within the inner loop, the outer loop will continue to move A in the direction of G.

Using the same solution development as in the single loop second order system, we will obtain the remainder of the coefficients of the solution. In solving for the exponential coefficient, the general equation is reduced to

$$q^2 + k_2 + k_1k_2 = 0$$

By using the binominal formula, the values of q which satisfy the equation are

$$q_{1,2} = \frac{-k_2 \pm \sqrt{k_2^2 - 4k_1k_2}}{2}$$

Depending upon the values of k_1 and k_2 used in a particular system, the structural response will exhibit either decaying exponential or oscillatory behavior. When the expression within the square root sign is greater than zero, the square root operation will result in a real number. In this case, the structural response will consist of two exponentials that decline in value after time zero. For the expression within the square root sign to be positive

$$k_2^2 > 4k_1k_2$$

or $$k_2 > 4k_1$$

When $k_2 < 4k_1$, the expression within the square root sign is negative and the square root operation will produce two imaginery numbers resulting in exponential coefficients that are complex. In fact, as the two complex numbers differ only in the sign of the imaginery component, the two coefficients are complex conjugates. In this case the structural response will exhibit oscillatory behavior with the period of the oscillation a direct function of the imaginery portion of q_1 and q_2. The real portion of the coefficients determines the time constant of the exponential decay which the oscillatory behavior exhibits.

For example, if $k_1 = 1/10$, $k_2 = 1/2$ then the coefficients would be

$$q_{1,2} = \frac{-1/2 + \sqrt{1/4 - 2/10}}{2}$$

$$= -.25 \pm .112$$

$$= -.138, -.362$$

The general form of this structural response is shown in Figure 3-4. The solution consists of two decaying exponentials as k_2 is more than 4 times as large as k_1. Both exponentials are negative also as the term under the square root sign is the smaller of the two terms. This is true in general as k_2 is always greater than the square root of $(k_2^2 - 4k_1k_2)$ except in the limiting case of $k_1 = 0$ when the two terms are equal. Note also that the larger k_1 is the greater the difference between the coefficients of the two exponentials in the structural response.

In another situation, the values of the two structural parameters might be reversed. If we have $k_1 = 1/2$, $k_2 = 1/10$ then the coefficients would be

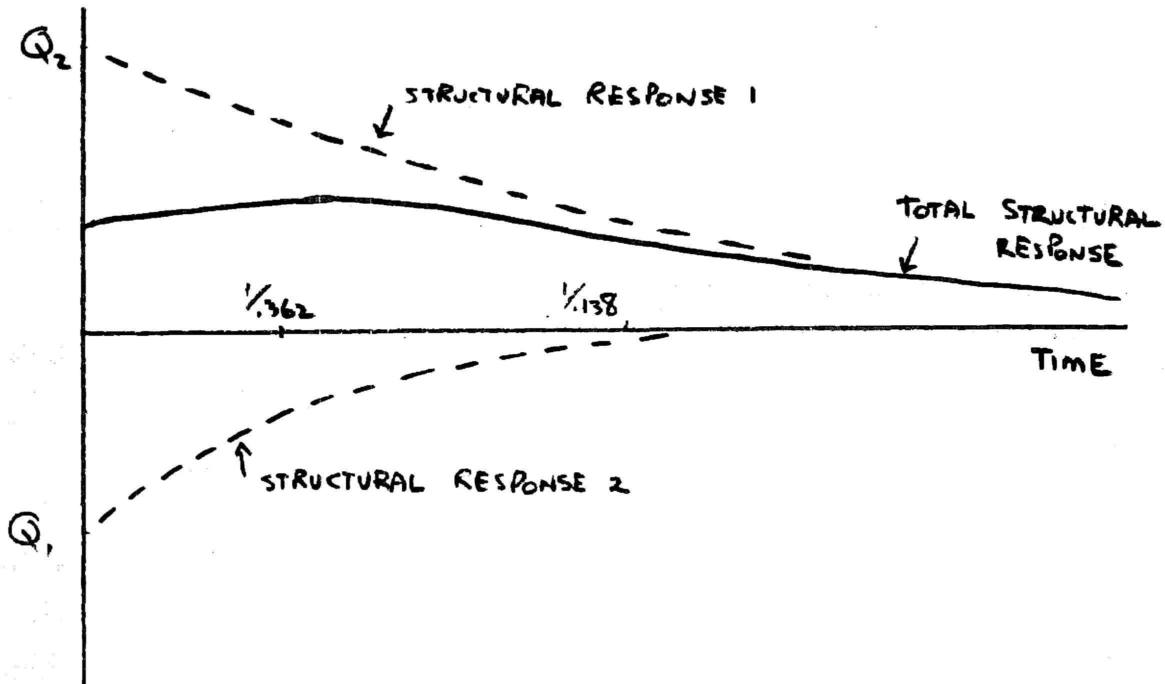


FIGURE 3-4: REAL EXPONENTIAL STRUCTURAL RESPONSE

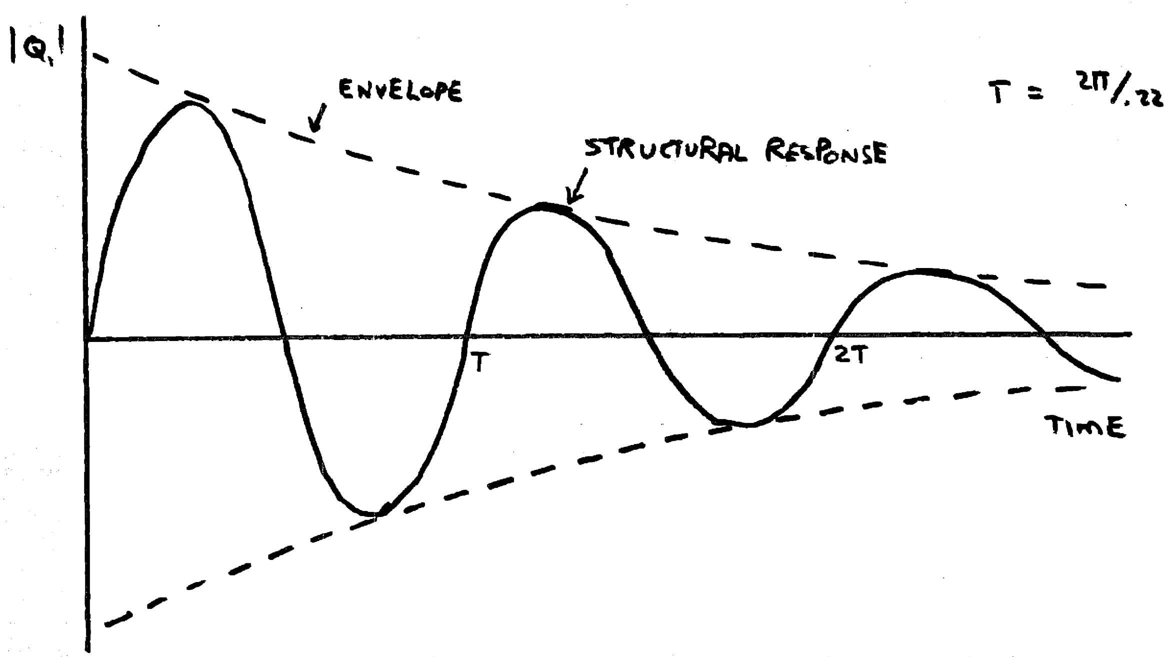


FIGURE 3-5: COMPLEX EXPONENTIAL STRUCTURAL RESPONSE

$$q_{1,2} = \frac{-1/10 + \sqrt{1/100-2/10}}{2}$$

$$= -.05 \pm \sqrt{-.047}$$

$$= -.05 \pm j\ .22$$

The general behavior of a system with complex conjugate exponential coefficients is shown in Figure 3-5. In this situation the system parameter k_2 alone controls the decay rate of the oscillatory response, while the term within the square root sign controls the period of oscillation.

The final coefficients to be obtained are the amplitude terms of the components of the structural response, Q_1 and Q_2. The addition of the extra single integral term in the left hand side of the general equation necessary to determine the values of Q_1 and Q_2. As in the single-loop, second-order system the first equation is

$$Q_1 + Q_2 + R = A(o)$$

The second equation is obtained by taking the derivative of the general equation and substituting the general solution of A at time zero. This results in

$$q_1Q_1 + q_2Q_2 + k_2Q_1 + k_2Q_2 + k_2R = k_2C(o)$$

Regrouping this equation gives us

$$(q_1+k_2)Q_1 + (q_2+k_2)Q_2 = k_2C(o) - k_2R$$

Multiplying the first equation by (q_2+k_2) and subtracting the second one from it results in

$$(q_2-q_1)Q_1 = (q_2+k_2)A(o) - q_2R - k_2C(o)$$

$$Q = \frac{1}{(q_2-q_1)}\cdot[(q_2+k_2)A(o) - q_2R - k_2C(o)]$$

The first equation can be used to determine the value of Q_2.

$$Q_2 = A(o) - R - Q_1$$

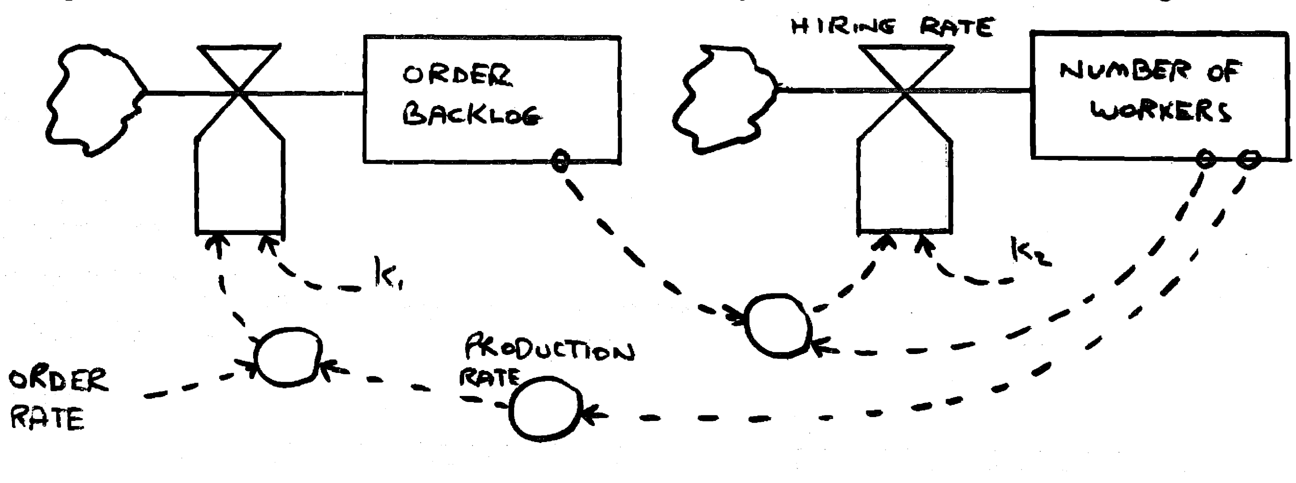
$$= A(o) - R - \frac{1}{(q_2-q_1)} [(q_2+k_2)A(o)-q_2R-k_2C(o)$$

$$= \frac{1}{(q_2-q_1)} [(-q_1-k_2)A(o) + q_1R + k_2C(o)$$

From the expressions that determine the amplitude of the structural response terms, it can be seen that they are a function not only of the relative magnitudes of the two initial conditions and the driving function, but also of the relative magnitude of the system parameters k_1 and k_2.

The Plant Work Force System

At this point it is appropriate to reevaluate the personnel department-work force system in terms of a more elaborate nested feedback loop system. It was suggested that the single loop second order system demonstrated earlier in the chapter was somewhat unrealistic in its control of the hiring rate as a direct function of the backlog. Assume as an alternative that the backlog is used to establish the desired number of workers and that the hiring decision is a function of the difference between the desired number of the present level of the work force. Such a system is illustrated in Figure 3-6.



FLOW DIAGRAM OF THE PLANT WORK FORCE SYSTEM

FIGURE 3-6.

Assume that in this business each worker can produce one unit per week, and that it is management policy to keep a backlog of one week's work. Then the units of backlog directly control the desired number of workers in the plant work force. When the workers are producing less than the company is selling, the backlog will increase and raise the desired number of workers. As workers are hired, the production rate will overtake the customer order rate and decrease backlog until an equilibrium is reached. A similar corrective action results when the production rate is greater than the customer order rate.

Assume the initial state of the system and the system structure is as follows:

$A(o)$	$= 1000$	the number of workers in the work force
$C(o)$	$= 1000$	the number of orders in the backlog
k_1	$= 1$	the fraction of customers willing to wait for delivery of their orders
k_2	$= 1/2$	the fraction of the shortage of workers that is hired each week
G	$= 1200$	the new order rate at time zero

From the general derivation of the nested second order feedback system, the solution is of the form

$$A = Q_1 e^{q_1 t} + Q_2 e^{q_2 t} + R$$

where $R = G$

and $$q_{1,2} = \frac{-k_2 \pm \sqrt{k_2^2 - 4k_1 k_2}}{2}$$

$$= \frac{-1/2 \pm \sqrt{1/4-2}}{2}$$

$$= -.25 \pm \sqrt{-.435}$$

$$= -.25 \pm j\ .66$$

and

$$Q_1 = \frac{1}{(q_2 - q_1)}[(q_2 + k_2)A(o) - q_2 R - k_2 C(o)]$$

$$= \frac{1}{-2j(.66)}[(-.25-j.66+.5)1000-(-.25-j.66)1200-(.5)1000]$$

$$= \frac{j}{1.32}[(.25+j.66)200]$$

$$= -100 + j38$$

$$Q_2 = A(o) - R - Q_1$$

$$= 1000 - 1200 + 100 - j38$$

$$= -100 - j38$$

The complete analytic solution is thus given by

$$A = (-100 + j38)e^{(-.25+j.66)t} + (-100-j38)e^{(-.25-j.66)t}$$

By converting the complex amplitudes to the exponentials to polar form, the expression can be simplified to

$$A = 107\ [e^{j159.2°}e^{(-.25+j.66)t} + e^{-j159.2°}e^{(-.25-j.66)t}] + 1200$$

$$= 107e^{-.25t}[e^{j(.66+159.2°)} + e^{-j(.66+159.2°)}] + 1200$$

$$= 214e^{-.25t}\cos(.66t + 159.2°) + 1200$$

The graphical representation of this expression is shown in Figure 3-7. As k_1 has been selected as being greater than k_2, the system exhibits the decaying oscillatory response that was described as characteristic of second order systems in which $k_2 < 4k_1$. If the opposite condition is considered in which $k_2 > 4k_1$ a non-oscillatory response can be observed in the personnel department system. For example, assume the following values for the system parameters and the initial state of the system:

$$A(o) = 1000$$

$$C(o) = 1000$$

$$k_1 = 1/10$$

$$k_2 = 1/2$$

$$G = 1200$$

The solution is again of the form

$$A = Q_1 e^{q_1 t} + Q_2 e^{q_2 t} + R$$

Substituting into the equations for the various coefficients determines that

$$R = G$$

$$q_{1,2} = \frac{-1/2 \pm\sqrt{1/4 - 2/10}}{2}$$

$$= -.25 \pm .112$$

$$= -.138, -.362$$

and $$Q_1 = \frac{1}{(-.362+.138)} [(-.362+.5)1000 - (-.362)1200 - (.5)1000]$$

$$= \frac{-1}{.244}[.362(200)]$$

$$= -340$$

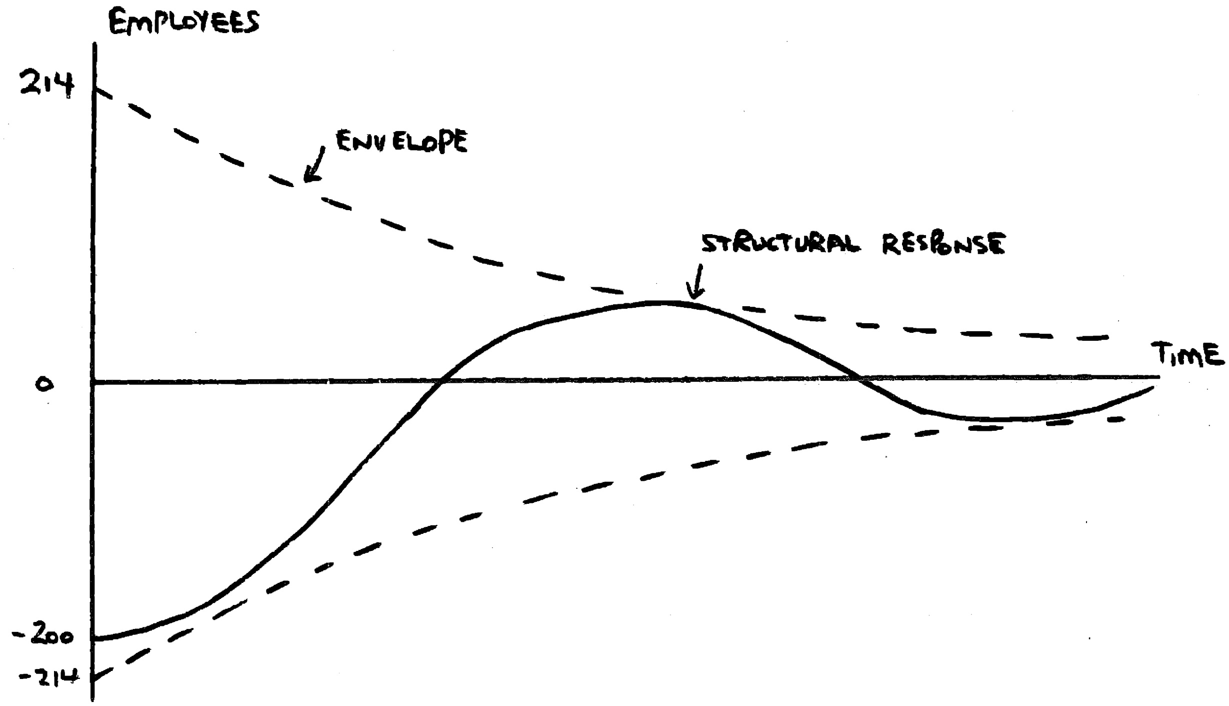


FIGURE 3-7: COMPLEX EXPONENTIAL RESPONSE OF THE PLANT WORK FORCE SYSTEM

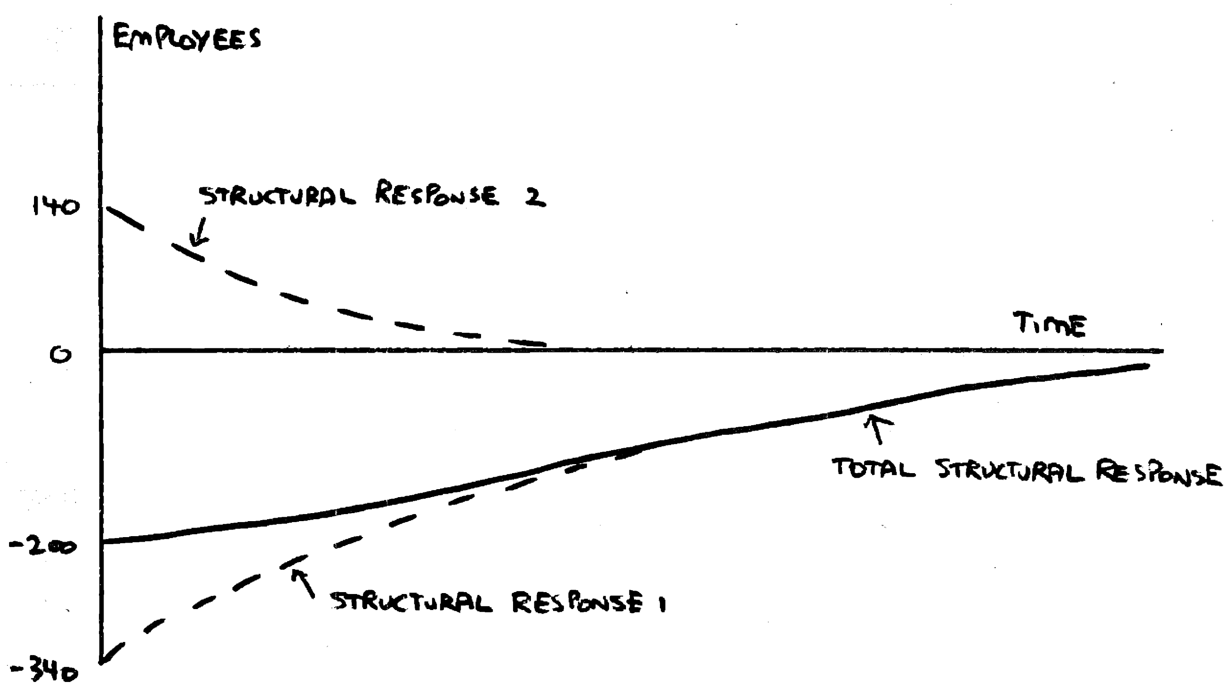


FIGURE 3-8: REAL EXPONENTIAL RESPONSE OF THE PLANT WORK FORCE SYSTEM

$$Q_2 = A(o) - R - Q_1$$

$$= 1000 - 1200 + 340$$

$$= 140$$

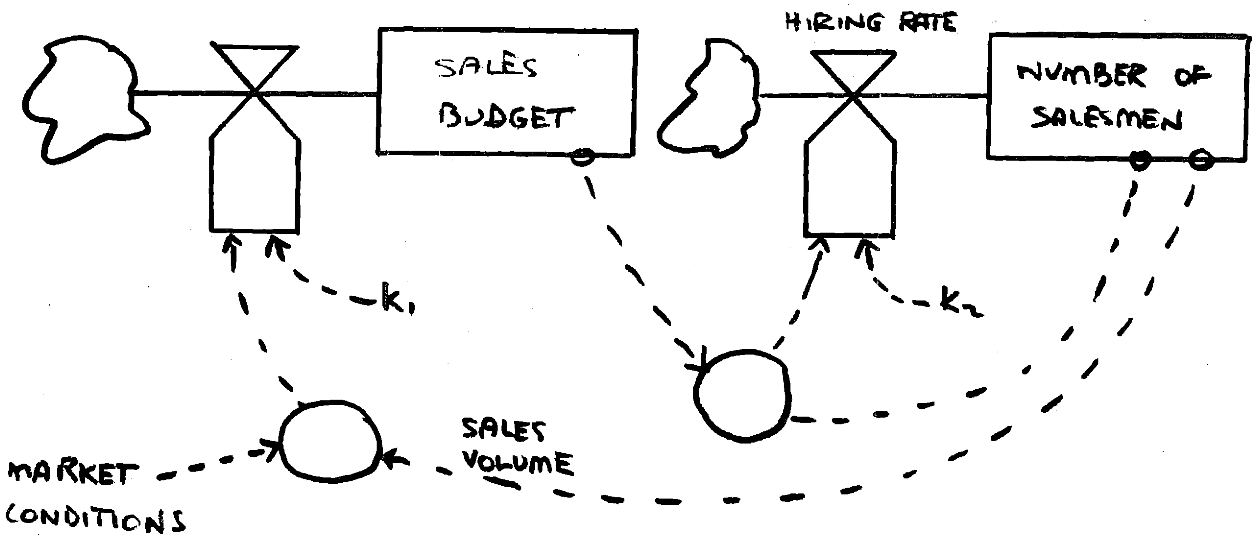
The complete analytic solution is thus given by

$$A = -340e^{-.138t} + 140e^{-.362t} + 1200$$

which is illustrated in Figure 3-8. Note that the two exponential components of the structural response cancel one another initially and the movement of the employment level away from its initial value of 1000 is very slow. After the cancelling effect of the positive exponential has decayed the total system response moves quickly to the equilbrium.

The Salesman Hiring Model

The system which was used to illustrate a first order positive feedback system in the previous chapter was limited in its ability to represent realistic behavior. To provide a more realistic model, a nested feedbackloop structure is required in which the inner negative loop controls the size of the sales force and the outer positive loop contributes to the growth momentum. Such a system is illustrated in Figure 3-9. The structure is identical with the personnel - work force system already discussed with the exception that the feedback in the outer loop is positive. The external variable G is again the market conditions that result from factors outside the control of the firm's salesmen. The feedback term provides the sales volume that is directly a function of the salesmen in terms of units sold. Assume that one unit of revenue is realized from the selling effort of each salesman each month.



FLOW DIAGRAM OF THE SALESMAN HIRING SYSTEM

FIGURE 3-9

The system parameter k_1 is then the fraction of the revenue realized from total sales (G+A) that is used to support the sales department budget. The rate of change of the budget (D) thus is accumulated in the total sales budget represented by level C. The total budget determines the desired number of men in the sales force. The feedback from the level of the sales force produces a difference between the desired and actual levels and system parameter k_2 determines the fraction of the difference that is hired in the current month. The rate at which salesmen are hired directly affects the level of the sales force.

The general equation for the system is the same as for the general nested feedback loop system except the sign of one term is changed, the double integral term that represents the feedback loop that was changed from negative to positive. The equation now is

$$A + k_2\int_o^t Adt - k_1k_2\int_o^t\int Adt = \int_o^t\int Gdt + k_2\int_o^t C(o)dt + A(o)$$

The general solution to the equation is still of the same form as for other second order systems and the equations for the coefficients have changed slightly to become

$$R = -G$$

and

$$q_{1,2} = \frac{-k_2 \pm\sqrt{k_2^2+4k_1k_2}}{2}$$

The equation for Q_1 and Q_2 are the same as for the nested feedback system with both loops negative.

Consider a specific example with the following parameters

$A(o)$	$= 500$	number of salesmen
$C(o)$	$= 1000$	number of salesmen budgeted for
k_1	$= 1/2$	fraction of revenue assigned to sales budget
k_2	$= 1/2$	fraction of needed salesmen hired per month
G	$= 500$	sales revenue resulting from market conditions

Then

$$R = -500$$

and

$$q_{1,2} = \frac{-1/2 \pm \sqrt{1/4 + 1}}{2}$$

$$= -.25 \pm .56$$

$$= .31, -.81$$

and

$$Q_1 = \frac{1}{(-.81-.31)} [(-.81+.5)500 - (.31)500 - .5(1000)]$$

$$= \frac{-81}{-1.12}$$

$$= 723$$

$$Q_2 = 500 + 500 - 723$$

$$= 277$$

Therefore the complete solution is given by

$$A = 723e^{.31t} + 277e^{-.81t} - 500$$

and is shown in Figure 3-10. The negative exponential dies out very quickly, introducing a lag because of the inability of the inner negative feedback loop to immediately follow the growth resulting from the outer positive feedback loop. The positive loop dominates the system performance from the beginning, however, and the total system response becomes very quickly the positive exponential component.

Consider the case now in which a severe adverse market condition causes a setback in the organization's growth. Assume that

$$A(o) = 500$$

$$C(o) = 500$$

$$k_1 = 1/2$$

$$k_2 = 1/2$$

$$G = -1500$$

Then $$Q_1 = \frac{1}{(-.81-.31)} [(-.81+.5)500 - .31(-1500) - .5(500)]$$

$$= \frac{.60}{-1.12}$$

$$= -54$$

$$Q_2 = 500 - 1500 + 54$$

$$= -946$$

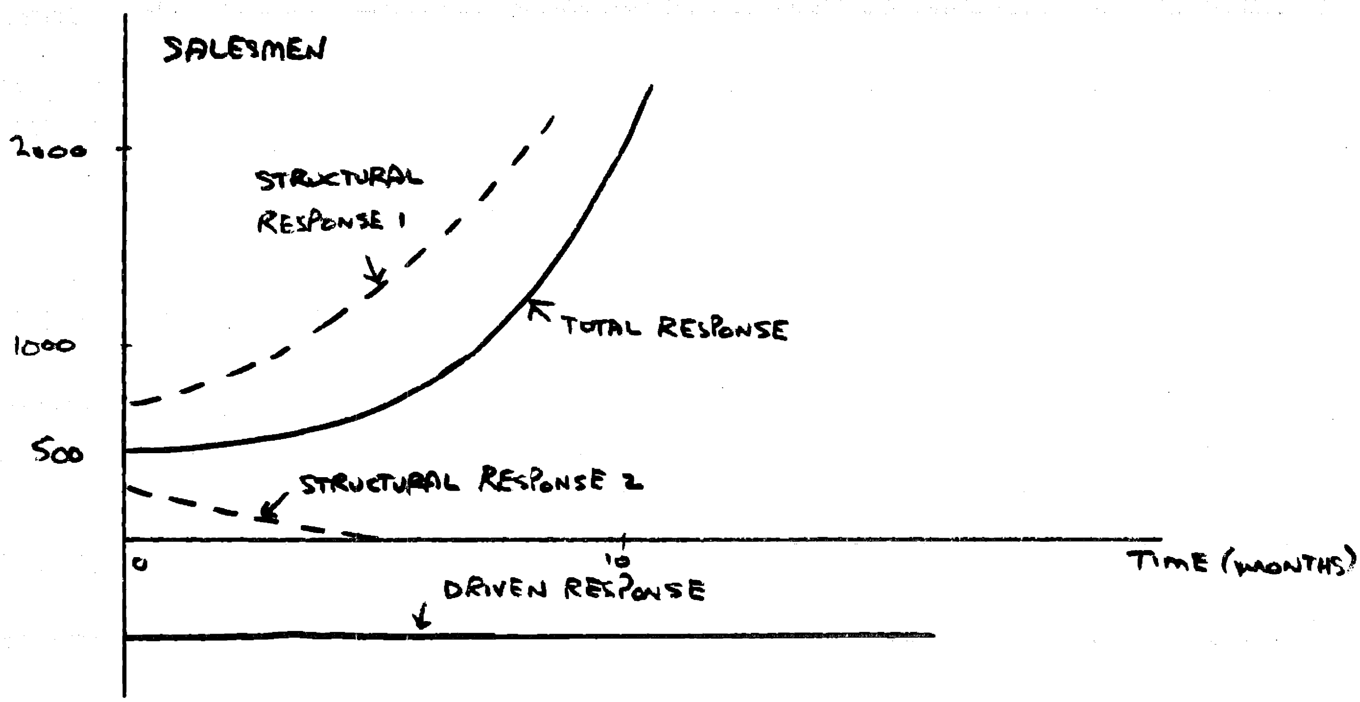


FIGURE 3-10: RESPONSE OF THE SALESMAN HIRING SYSTEM TO A POSITIVE STEP

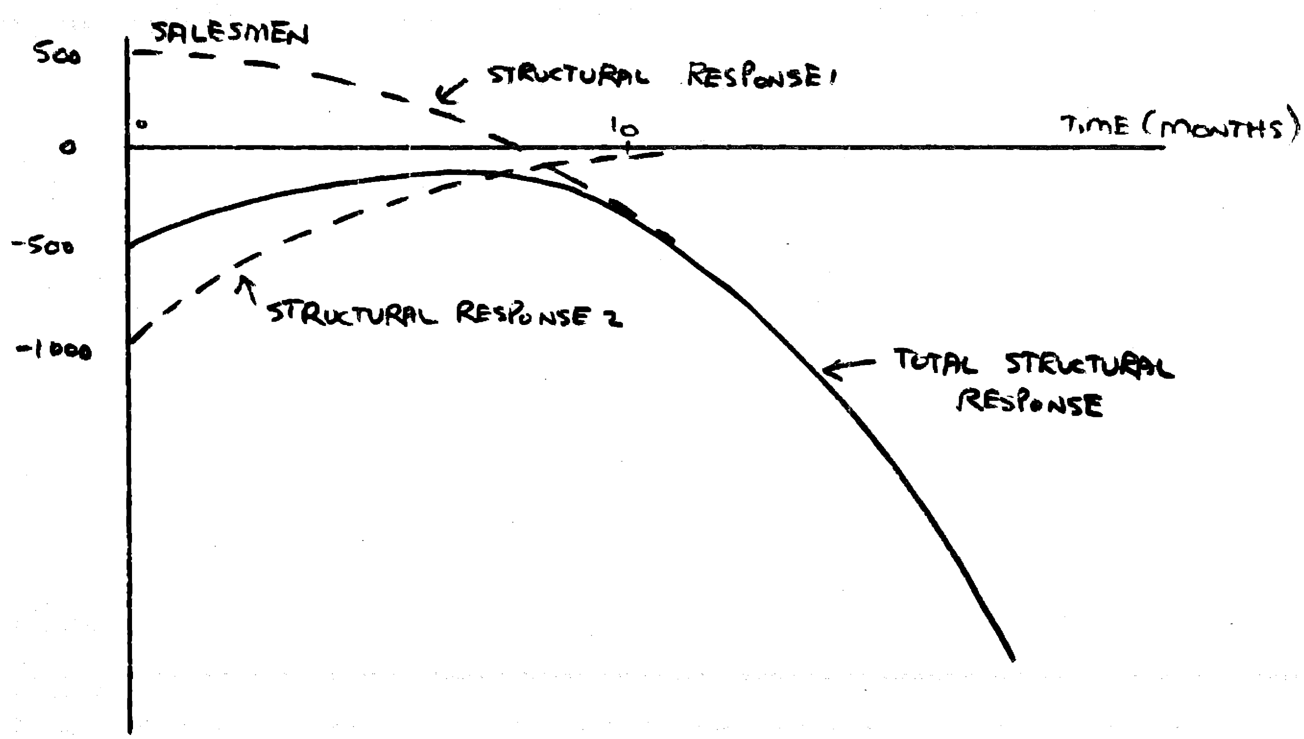


FIGURE 3-11: RESPONSE OF THE SALESMAN HIRING SYSTEM TO A NEGATIVE STEP

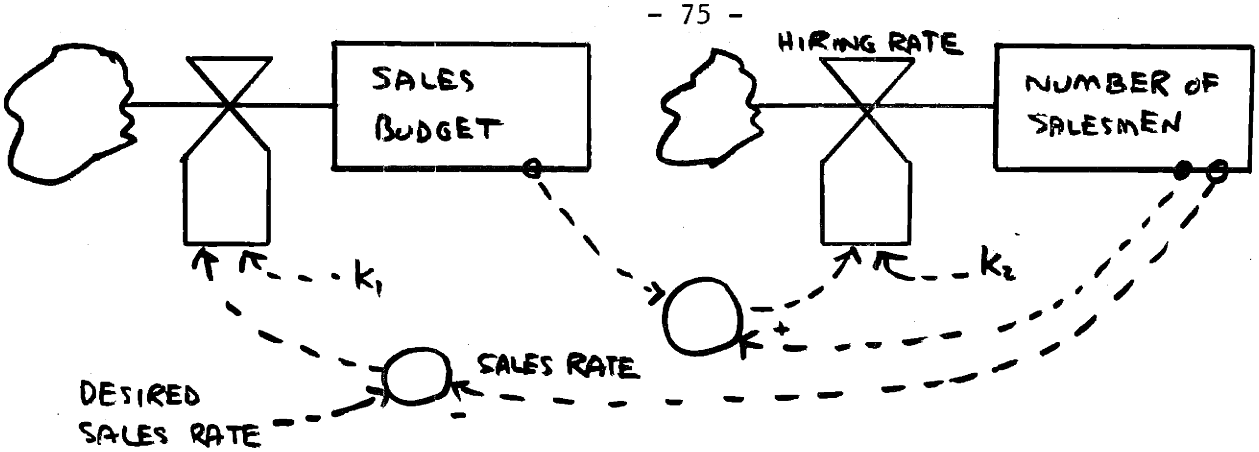
The complete solution is given by

$$A = -54e^{.31t} - 946e^{-.81t} + 1500$$

As in the previous example, the negative exponential decays quickly as the negative feedback loop moves to follow the declining number of salesmen in the sales force. Figure 3-11 illustrates the behavior of this system.

Constrained Sales Growth

In many situations it is possible that the growth exhibited by a positive feedback process such as the salesman hiring loop just described can be constrained by other feedback loops outside it. For instance, a negative feedback loop which controls the rate of growth to some desired rate is one possibility of such an arrangement. Each time the inner positive loop attempts to grow faster than management's desired rate, pressure would be exerted to restrain selling effort. One way in which such a system can be represented is shown in Figure 3-12. The external variable G is the desired sales rate which is compared to the actual sales rate generated by the number of salesmen currently with the firm to determine the rate at which the sales budget should be increased or decreased. Assume that each salesman can sell one thousand units per month. The level C represents the sales budget at any point in time and it has a direct affect on the rate at which salesmen are hired, being added to the feedback term in the inner loop. System parameter k_2 is a fraction of salesmen that need to be hired that are actually hired per unit time and k_1 is the fraction of the deviation of actual sales from desired sales that is considered permanent and is used to control the sales budget.



FLOW DIAGRAM OF THE CONSTRAINED SALESMAN HIRING SYSTEM
FIGURE 3-12

The presence of positive feedback in the inner loop in this case means that the single integral term on the left hand side of the general equation now has a negative sign.

$$A - k_2\int_o^t Adt + k_1k_2\int_o^t\int Adt = \int_o^t\int Gdt + k_2\int_o^t C(o)dt + A(o)$$

The equations for the coefficients reflect the change in sign in several places.

$$R = G$$

$$q_{1,2} = \frac{+k_2 \pm \sqrt{k_2^2-4k_1k_2}}{2}$$

$$Q_1 = \frac{1}{(q_2-q_1)}\left[(q_2-k_2)A(o) - q_2G - k_2C(o)\right]$$

$$Q_2 = A(o) - G - Q_1$$

By considering some specific values for the system parameters and the initial state of the system, a system response can be determined. Assume

$A(o) = 500$

$C(o) = 500$

$k_1 = 1/2$

$k_2 = 1/2$

$G = 500$

In other words, at time zero management changes the sales budget to a level that justifies 500 salesmen resulting in immediate pressure on the firm to cut back.

$$R = 500$$

$$q_{1,2} = \frac{1/2 \pm \sqrt{1/4-4/4}}{2}$$

$$= .25 \pm j.43$$

$$Q_1 = \frac{1}{2(-j.43)} [(.25-j.43-.5)500 - (.25-j.43)500 - .5(500)]$$

$$= \frac{-500}{-j.86}$$

$$= -j580$$

$$Q_2 = A(o) - G - Q_1$$

$$= 500 - 500 - Q_1$$

$$= j580$$

The complete solution is thus given by

$$A = -j580e^{(.25+j.43)t} + j580e^{(.25-j.43)t} + 500$$

$$= 580e^{j-90°} e(.25+j.43)t + 580e^{j90°} e^{(.25-j.43)t} + 500$$

$$= 580e^{.25t}[e^{j(.43t-90°)} + e^{j(-.43+90°)}] + 500$$

$$= 1160e^{.25t} \cos(.43t-90°) + 500$$

The behavior of the system is illustrated in Figure 3-13. Basically the sales force oscillates about the desired value which management has established as the desired level. The positive feedback action of the inner loop causes the change in sales force to reinforce itself in such a way as to cause an ever-increasing oscillation. The negative feedback loop is only able to control the system to the extent that the point at which the oscillation is centered about is the desired operating value. If the system parameters have a relationship where $k_2 > 4k_1$, we observe another type of behavior; for example, suppose

$$A(o) = 500$$

$$C(o) = 500$$

$$k_1 = 1/10$$

$$k_2 = 1/2$$

$$G = 500$$

Then

$$R = 500$$

$$q_{1,2} = \frac{+1/2 \pm \sqrt{1/4-2/10}}{2}$$

$$= .25 \pm .112$$

$$= .362, .138$$

$$Q_1 = \frac{1}{-.224} [(.138-.5)500 - (.138)500 - (.5)500]$$

$$= \frac{-500}{-.224}$$

$$= 2230$$

$$Q_2 = A(o) - R - Q_1$$

$$= 500-500-2230$$

$$= -2230$$

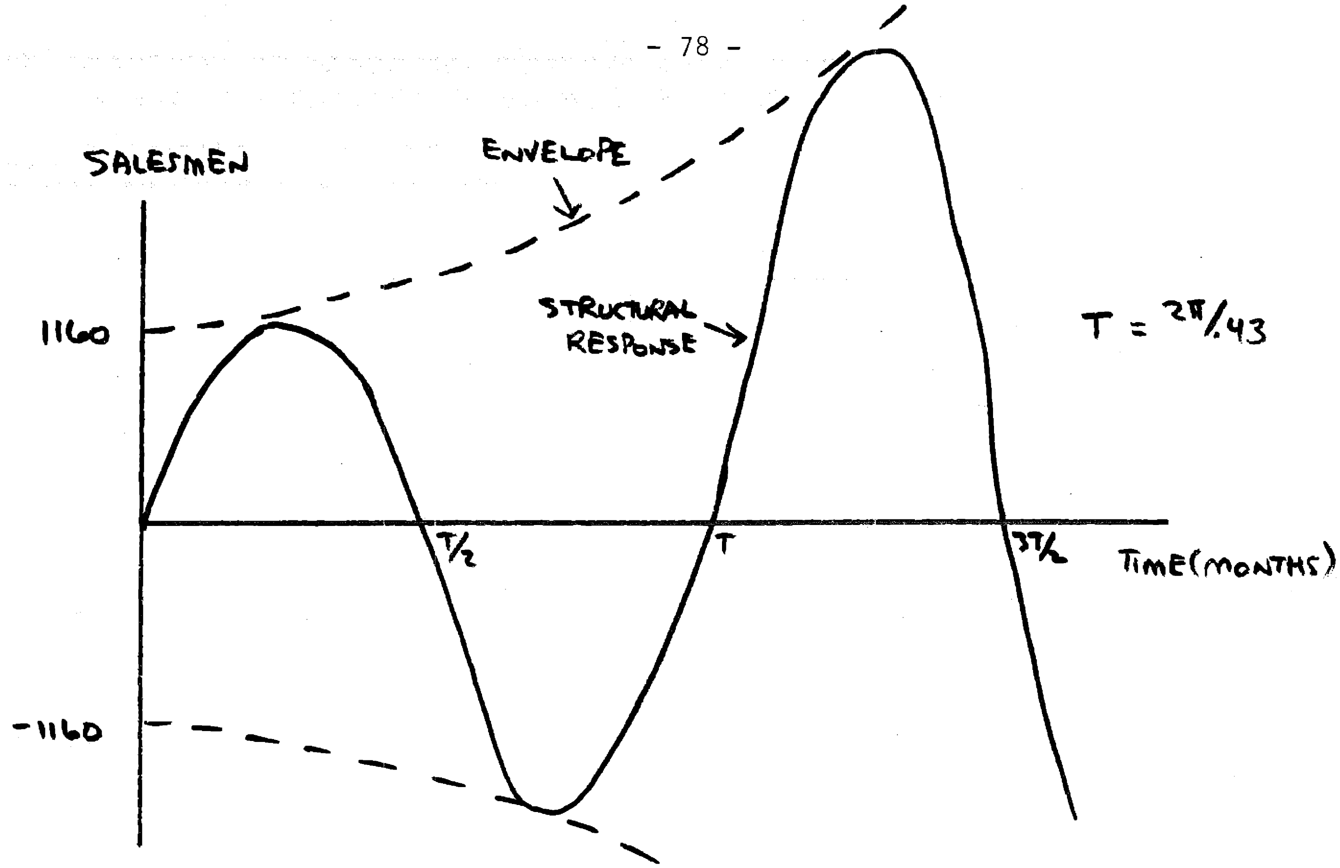


FIGURE 3-13: COMPLEX EXPONENTIAL RESPONSE OF THE CONSTRAINED SALESMAN HIRING SYSTEM

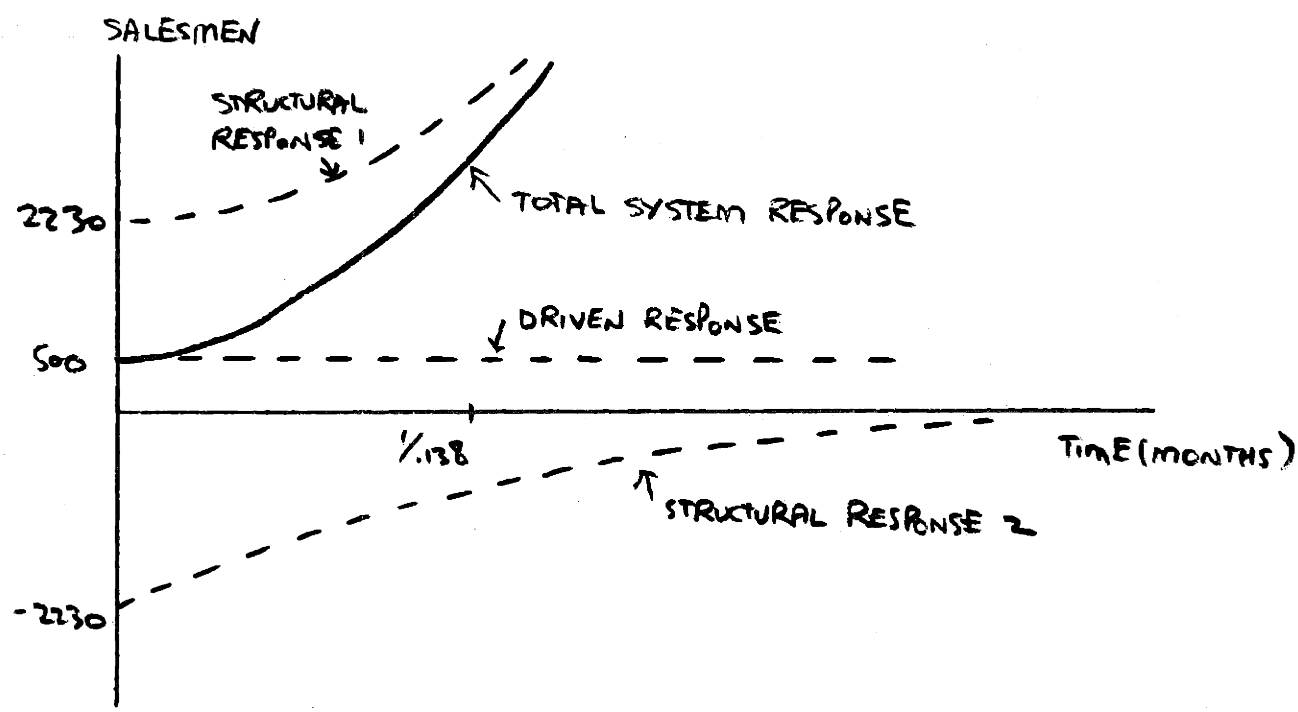


FIGURE 3-14: REAL EXPONENTIAL RESPONSE OF THE CONSTRAINED SALESMAN HIRING SYSTEM

The complete solution is then given by

$$A = 2230e^{.362t} - 2230e^{.138t} + 500$$

The behavior of this system is illustrated in Figure 3-14. Note that the growth momentum dominates the total system response. It is interesting to observe that the exponential with the shortest time constant (highest initial rate of change) dominates the solution in the positive feedback case (growing exponentials, positive coefficients) while exactly the opposite is true in the case of negative feedback systems.[1] This is an example of a situation where the controlling mechanism (the outer loop) does not have enough gain (k_1 is too small) to overcome the effect of the inner loop.

---

[1]This chapter has provided an analytic treatment of a small fraction of the number of configurations that exist for second order feedback systems. The approach used in developing these solutions can be applied to additional configurations including systems with delays in the feedback loops. The basic forms of behavior that are possible with second order systems have been demonstrated, however, and the analysis of additional configurations will produce patterns of behavior that are the same as have been analyzed here. Work needs to be done in developing additional examples to illustrate the performance of other types of second order systems.

CHAPTER IV

ANALYSIS OF HIGH ORDER FEEDBACK SYSTEMS

The previous chapters have shown explicit solutions to both first and second order feedback systems. For systems of third order and higher, however, analytic solutions become increasingly difficult. In addition, the effort involved in obtaining such a solution is not justified as it is commonly the case that some portions of the system structure are more important in determining the system response than others. What is needed is a convenient way to test the system to determine which portions dominate the response. One such procedure is developed in this chapter along with its application to several of the systems that have been analyzed previously and several higher order systems.

In the examples studied earlier it has been observed that feedback systems have a tendency to respond slowly to a sudden change in the driving function or external variable. As was discussed, this is because feedback systems typically have memory within their structure in the form of one or more levels and the system response must adjust the state of these levels before a balanced condition can be achieved with the new input. Depending on the characteristics of the system, the response will be either a simple exponential, sum of exponentials or perhaps a decaying oscillation. With any type of behavior a feedback system is introducing delay in the change of the system variables relative to the change in the external variable. If the external variable cycles in a sinusoidal fashion, the system variables will attempt to do so also but their ability to keep up will be limited by this same delay. A sinusoidal driving function never gives the system an opportunity to reach equilibrium as it is continually changing. However, the system variables will reach an equilibrium in which they will move in a sinusoid with the same period but will always be a fraction of cycle behind the oscillation of the driving function. In addition, when the period of the cycle is short, the delay experienced by the system can be long enough that the system variables will not have time to reach the peak value of the oscillation before the external variable is forcing them down once again. The ratio of the peak value of the system variable to that of the external variable is referred to as the gain of the system. A general characteristic of the driven response of feedback systems is that the gain will, after a particular period has been reached, tend to decrease as the period of the cycle of the driving function does. The value of the period at which this transition in the system's performance occurs is dependent upon the system structure and the values of the system parameters. Thus the delay of the system response due to the levels becomes more important as more rapidly changing external variables are applied to the system.[1]

Characteristics of First Order Systems

It is useful to construct the gain characteristics of a first order feedback system by examining the gain characteristics of the individual components of the system and the way in which they are related. Consider the effect of a sinusoidal inflow on a level. The level variable A can

---

[1]The ideas in this paragraph must be more completely developed, for use in teaching the material in the thesis. It has only been summarized here in that these ideas have been more fully developed in the engineering literature.

be determined analytically from

$$A = \int_0^t G dt + A(o)$$

If $A(o) = 0$ and $G(t) = Q \cos xt$, then

$$A = \int_0^t Q \cos xt \, dt$$

$$= (Q/x) \sin xt \,\Big]_0^t$$

$$= (Q/x) \sin xt$$

$$= (Q/x) \cos (xt - 90°)$$

Thus the level variable reproduces the original cosine function in time, although shifted by 90 degrees (1/4 cycle), and changed in gain by a factor of 1/x. Because x is equivalent to the number of cycles per time period (in radians/second), we can evaluate x in terms of T, the period of the cycle. For one cycle of oscillation

$$x = 2\Pi/T \quad \text{and} \quad t = T$$

Thus, when t progresses from zero to T, the cosine function progresses from 0 to 2Π, reproducing one complete cycle of oscillation. The gain of a level is thus given as $1/x = T/2\Pi$ and it is observed that the gain increases linearly with T, the period of the cycle. Figure 4-1 illustrates this relationship.

Note that the gain of a level is less than one for all periods less than 2Π time units and greater than one for all periods greater than this value. The analytic solution also shows that the level introduces a 90° lag, regardless of the period T of the cycle. This is understandable if we examine the relative values of the level and its inflow at two points in time. When the inflow is at its peak positive value, the level is increasing at its highest rate. The cosine function has reached zero one quarter cycle later and the

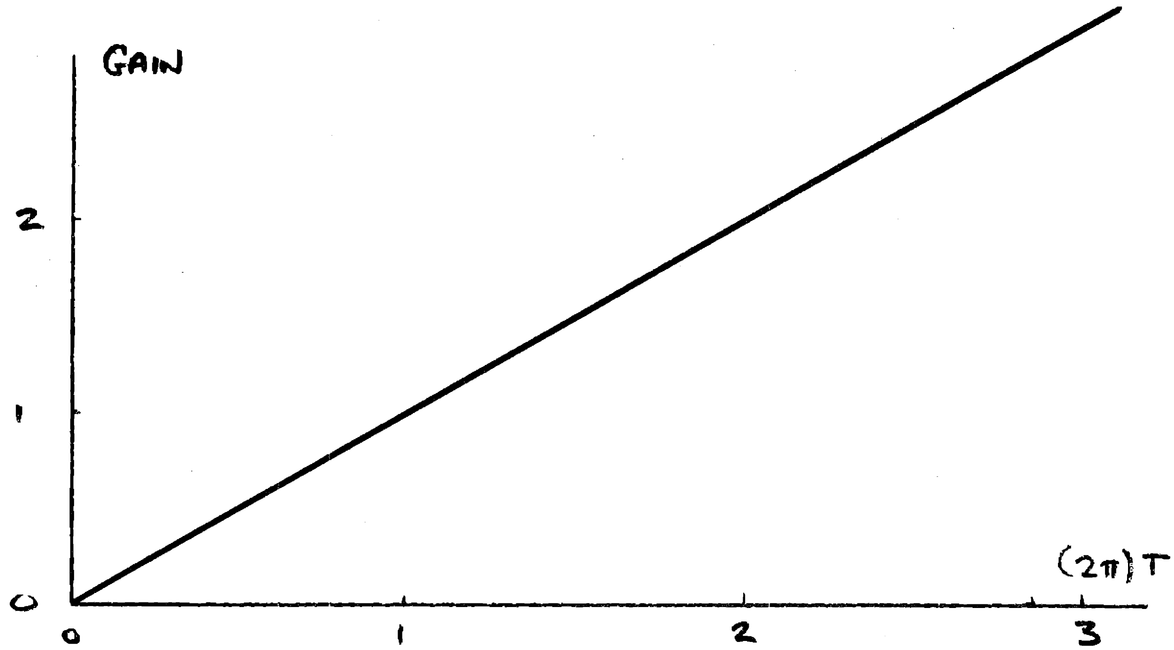


FIGURE 4-1: THE LINEAR GAIN CHARACTERISTIC OF A LEVEL

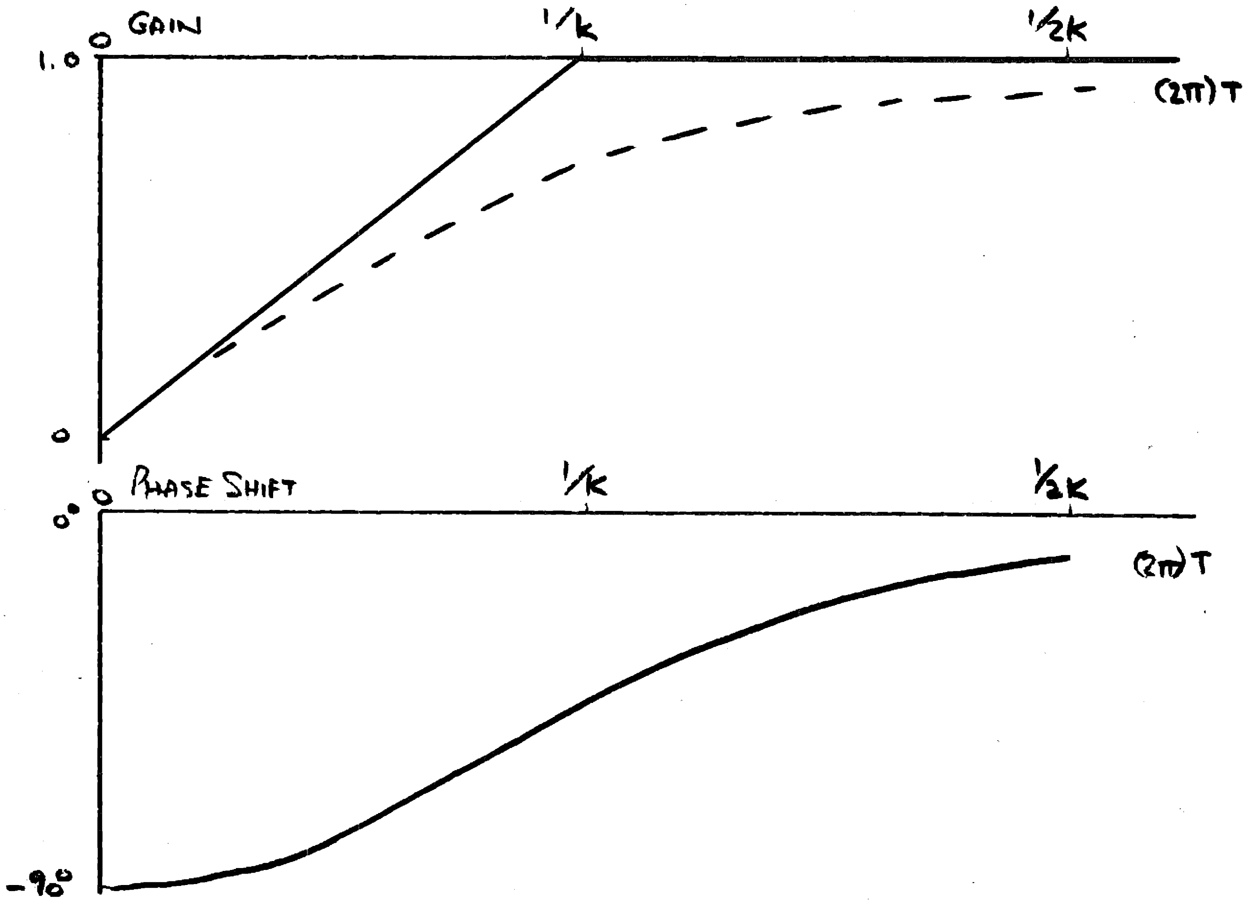


FIGURE 4-2: THE GAIN-PHASE CHARACTERISTIC OF A LEVEL WITH FEEDBACK

level stops increasing at this point and begins to decrease when the inflow swings negative. Thus the peak in the value of the level occurs one quarter cycle after the peak in the inflow.

A gain element can be considered as any part of a system in which the change in a flow or auxilliary does result in a change in another flow or auxilliary that is not one for one. The fractional adjustments (k) that have been used as management policy in the examples used earlier in the thesis are gain elements.

If a gain element is considered in combination with a level, the gain curve in Figure 4-1 is shifted by the amount of the gain k and there is no change in the phase shift relationship. This can be tested readily by adding an additional gain factor to the analytic derivation of the response of a level by itself.

To complete a simple first order negative feedback system, it is necessary to feed back the value of the level variable to the inflow and produce a new inflow which is the differences between the external variable and the actual value of the level. The implication of such structure is that the level should equal the inflow at all times and the difference that is now used for an inflow to the level is for correction purposes only. It is useful to examine the change in the system characteristics that this introduces. From the analytic derivation in Chapter 2, it was learned that the driven response of a first order feedback system was given by

$$R = \frac{k}{k \pm jm}$$

where G is the amplitude of the sinusoidal driving function and m is equal to

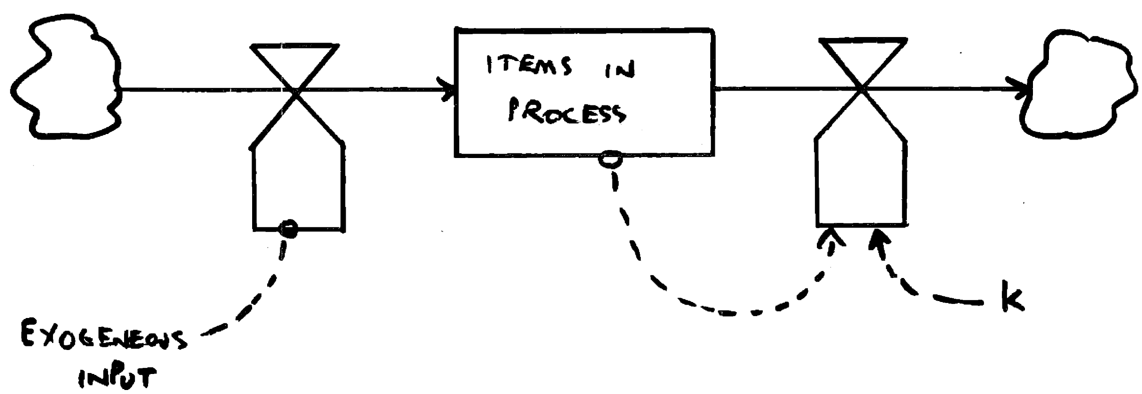
$2\Pi/T$. For values of T such that the imaginery number in the denominator is much larger than k, that is, $2\Pi/T >> k$ or $T << 2\Pi/k$, the gain characteristic can be approximated by that of a level and a gain element k alone. That is

$$R = kG/jm = (kT/j2\Pi)G$$

For values of T such that the imaginery number which is smaller than k, that is, $2\Pi/T >> k$ or $T << 2\Pi/k$, the gain characteristic is approximately unity, independent of k. This is represented by

$$R = kG/K = G$$

Thus the feedback affects the characteristics of the system only for cycles with long periods, and the characteristics for short period cycles are the same as if the feedback did not exist. Note also that the imaginery component of the gain expression is very small for long period oscillations. This means that there is little phase shift in such situations and that the level variable reproduces the driving function with a small lag. Changes in the driving function as it moves through its sinusoidal cycle are compared to the current value of the level and a correction signal is able to be generated to drive the level in the same direction with a small delay relative to the length of the cycle. For the range of values of T where the two components of the denominator are approximately equal, the characteristic undergoes a smooth transition between the two characteristics discussed. Figure 4-2(a) illustrates the true gain characteristic in dotted line and the approximate (asymptotic) one drawn in solid lines by simply extrapolating the characteristics at both extremes. The phase shift curve shown in Figure 4-2(b) also transitions between the two values of 0° and -90° in the same range of values of T. Note that the choice of system parameter k will determine the basic responsiveness of the system. Using the approximate gain curve the decline in gain can be considered to start when T = 2Π/k. Thus the higher the k the smaller the period of oscillation that can be reproduced by the feedback system without any loss in amplitude. This also means that a system will respond faster to a sudden change in equilibrium that requires a readjustment of the value of the one or more levels a system may contain.



FLOW DIAGRAM OF A FIRST ORDER DELAY

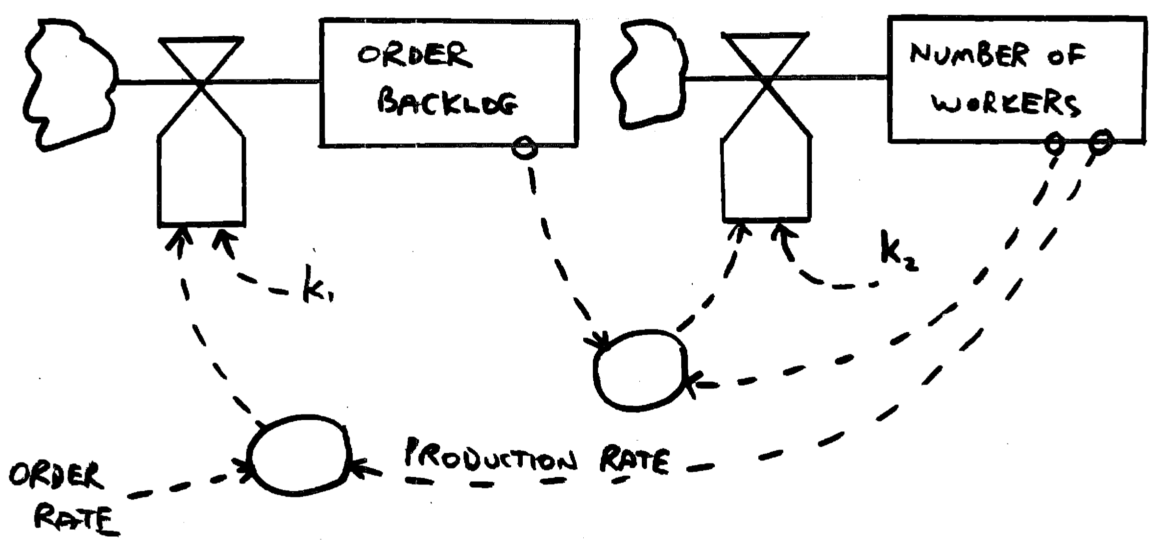
<u>FIGURE 4-3</u>

A first order negative feedback system that controls the outflow as a function of the level instead of the inflow is known as a first order delay and has exactly the same characteristics that are shown in Figure 4-2. Such a system is shown in Figure 4-3 where k is once again the gain element contained within the feedback loop. If the level represents the number of items that are undergoing some process and k the fraction of that number that are completed in each time period, then 1/k time units would be the time required to remove all the items if they continued to be removed at that rate. If there are an equal number of items being added to the level as are being removed, then 1/k would be the average time delay of each item in the process. Changing k in this system would also change the responsiveness of the system to sudden changes in the driving function.

The property of reproducing the long period oscillations more accurately and the short oscillations with every decreasing gain is a general property of all negative feedback systems regardless of order. The levels within systems introduce the delay in the response that causes the decreasing gain for faster changes in the driving function. Thus, feedback systems tend to pass only the long slowly-changing time functions and to suppress fast changes. Because of this property many feedback systems serve a function of averaging or filtering data by suppressing spurious short term changes and responding only to significant changes in trends. Similarly delays of physical flows such as in Figure 4-3 exhibit characteristics wherein a transition in the output shows up much more slowly than it did in the originating flow.

Characteristics of Second Order Systems

The gain-phase characteristics of second order feedback systems can be developed in much the same manner as for first order systems. In addition, nested feedback systems can be handled in a building block fashion which will allow the effect on the characteristics of each additional feedback loop to be investigated as the total characteristics is built up. For example, Figure 4-4 shows the nested order backlog-plant work force system that was discussed in Chapter 3. Considering the plant work force level and its own feedback loop, the characteristics that have already been obtained for the simple first order negative feedback system apply and are shown in



FLOW DIAGRAM OF A PLANT WORK FORCE SYSTEM

FIGURE 4-4

Figure 4-5. The system parameter k_2 controls the responsiveness in the inner loop but because it is first order the maximum phase shift is 90° and there are no conditions under which the structural response can dominate.[1] The maximum phase shift of any feedback loop is simply the number of levels in the loop multipled by the 90° shift that each one of them can attain. This loop has the same characteristics of a simple first order delay. The gain-phase characteristics of the order backlog and its associated system parameter k_1 are shown in Figure 4-6 as isolated components. Note that the gain increases indefinitely for long period oscillations as there is no feedback connected around the level as yet. The work force level and its associated feedback loop can now be considered as a component of the outer loop along with the

[1]Whether the structural or the driven response dominates a particular system's response depends upon the relationship of the system's gain and phase shift curves. In particular, a system is unstable (the structural response dominates) if for a length of period for which the gain is unity or greater and the phase shift is greater than 180°. Material covering stability is well developed in the engineering literature and should be included in this course at this point.

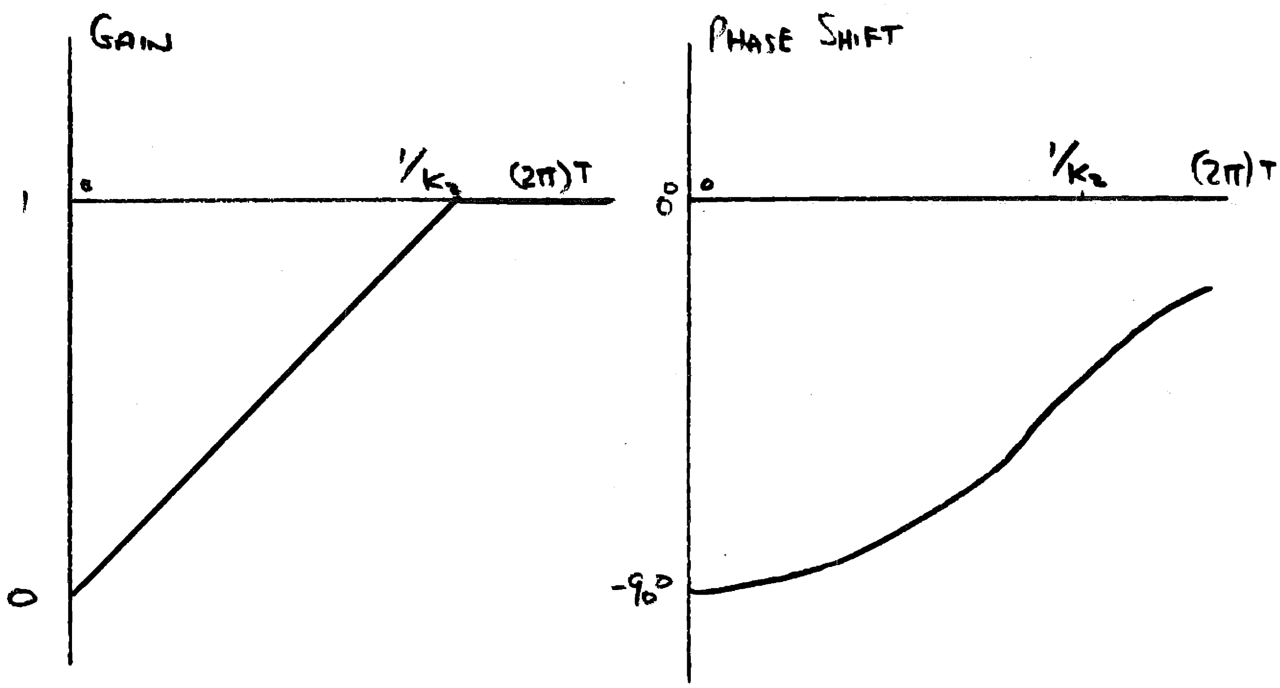


FIGURE 4-5: THE GAIN-PHASE CHARACTERISTIC OF THE INNER LOOP

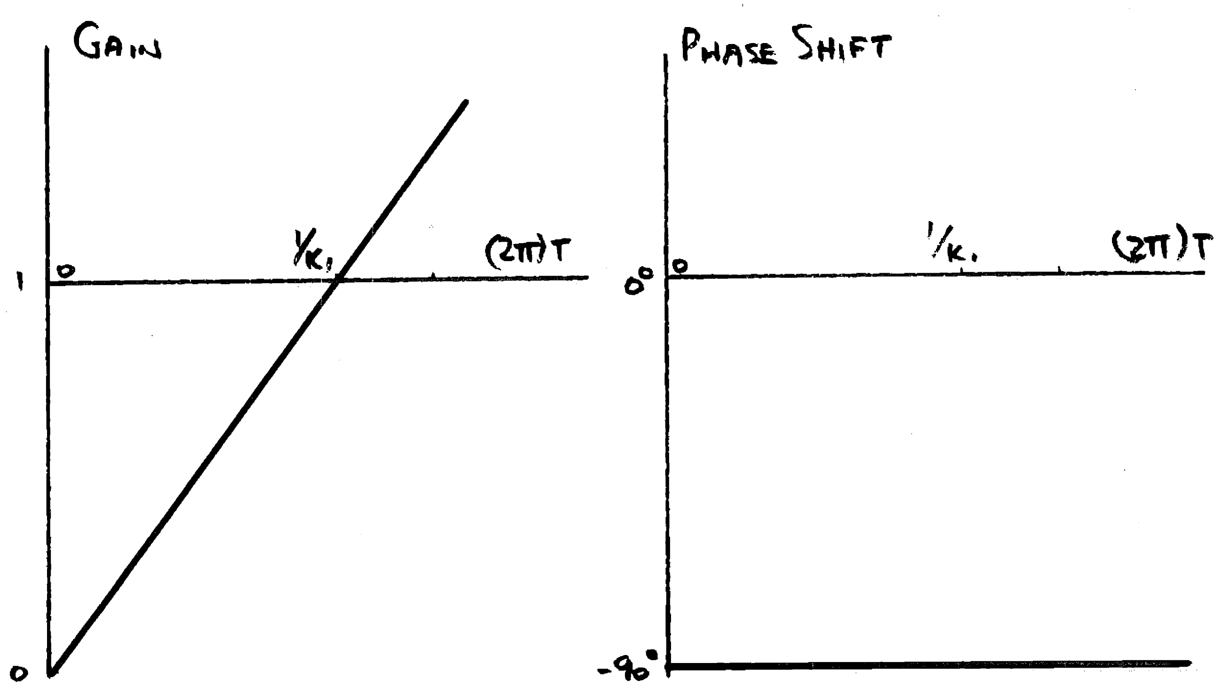


FIGURE 4-6: THE GAIN PHASE CHARACTERISTIC OF THE OUTER LOOP COMPONENTS

order backlog and parameter k_1. To obtain the overall gain-phase characteristic of these elements, the gain characteristics must be multiplied together and the phase characteristics added.[1] In that the graph scales become difficult to handle after several multiply operations for a nested loop, a logarithmic plot will be introduced. By introducing a scale that is the logarithm to the base ten for both the horizontal and vertical axes, the gain characteristics can be represented as shown in Figure 4-7. Note that equal spaces now result in an order of magnitude (factor of ten) change on each axis. The horizontal axis of the phase characteristic will be changed but the vertical need not be. Multiplying these two gain characteristics together can be accomplished by simply adding the two curves graphically. The resultant gain and phase characteristics of the nested feedback loop system with the outer feedback loop open are shown as the solid line curves in Figure 4-8. Figure 4-8(a) shows the case where $k_1 > k_2$ and Figure 4-8(b), the case where $k_1 > k_2$. Note that the characteristics for neither of the two situations is such that unity gain is achieved for values of T in which the phase shift is greater than 180°. When the outside feedback loop is closed all gains greater than unity are removed and the characteristics are changed only as indicated by the dotted lines in Figure 4-9. The two slopes of the gain characteristics in Figure 4-8(a) represent a structural solution with two components. Each of the components is represented by an exponential with a negative coefficient as the dominance of the driven response in the long

[1]Material introducing and supporting the ideas about cascaded systems and the isolation of one element from the ones before or after it should be included at this point.

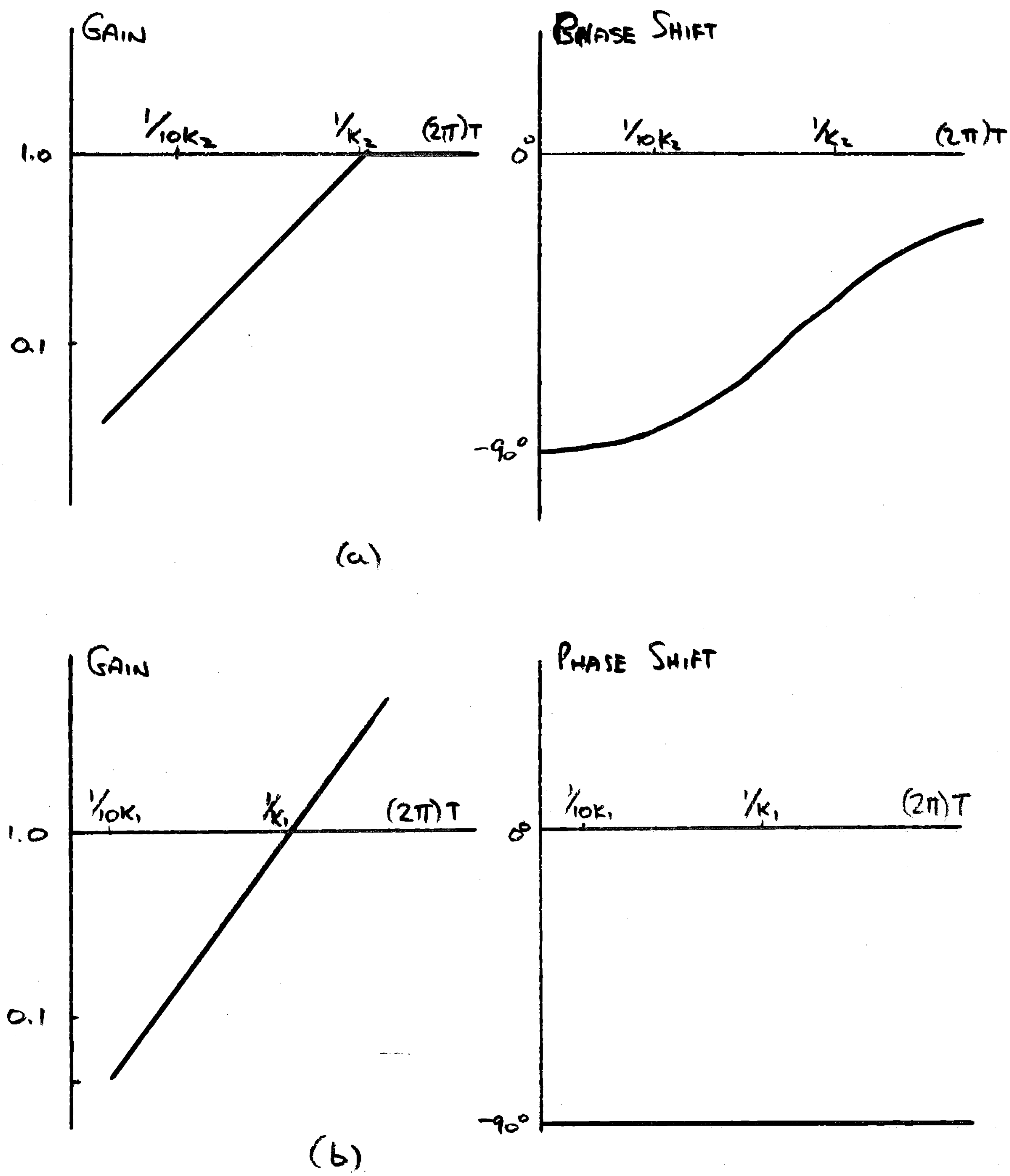


FIGURE 4-7: LOGARITHMIC GAIN PHASE CHARACTERISTICS

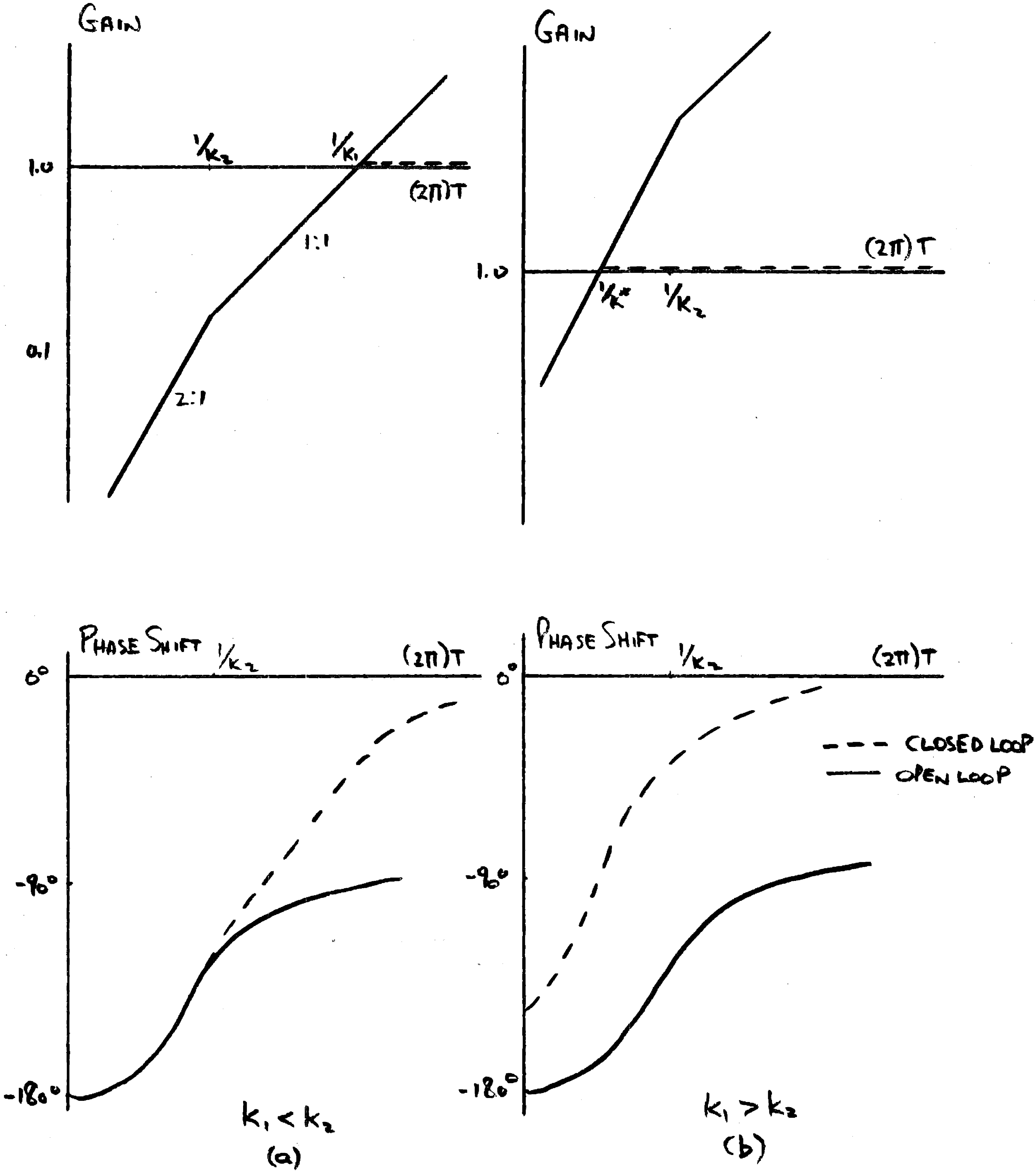
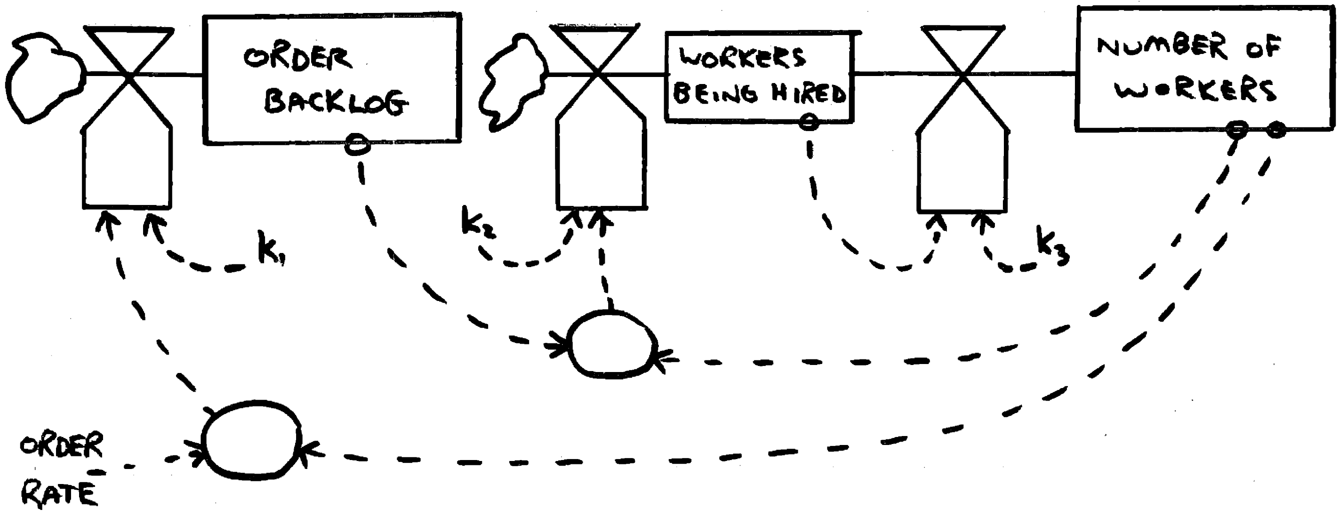


FIGURE 4-8: THE GAIN-PHASE CHARACTERISTICS OF THE PLANT WORK FORCE SYSTEM

run has already been demonstrated. The resultant single slope of 2:1 in Figure 4-8(b) indicates that the structural response will be oscillatory as our analytic derivation in Chapter 3 has indicated that an oscillatory response will result for all values of $4k_1 > k_2$. The solution will again be dominated in the long run, by the driven response, however. Now that a qualitative measure of performance has been developed for systems for which analytic derivations have been made it is possible to proceed with the analysis of a higher order system.

Characteristics of Higher Order Systems

As has been discussed earlier, an explicit analysis of higher than second order systems is usually not warranted both because of its complexity and the fact that a first or second order approximation provides a useful substitute in many cases. Consider the same system as above with the addition of a first order delay in the worker hiring process. Such a system is illustrated in Figure 4-9. If it is assumed that $k_3 < k_2$, then the characteristics shown in Figure 4-8(a) will represent the complete inner feedback loop including the delay. By itself this loop exhibits a structural response consisting of two exponentials (or a slight oscillatory tendency if $4k_3 > k_2$). By adding the gain-phase characteristics of the additional components in the outer loop the curves shown in Figure 4-10 are obtained. Figure 4-10(a) shows the resultant gain-phase curves of all the components in the outer loop for values of $k_1 < k_3$. If k_3 is much less than k_2 then the gain effects of k_2 can be ignored and the system response can be approximated as a function of k_1 and k_3 only. It has already been shown that



FLOW DIAGRAM OF THE PLANT WORK FORCE SYSTEM WITH HIRING DELAY

FIGURE 4-9

a gain characteristic of this type would produce a structural response consisting of two exponentials. When $k_3 < k_1 < k_2$ as shown in Figure 4-10(b) care must be taken to ensure that k* is not too close to k_2 as the phase shift moves from 180° to 270° centered about k_2. In fact, if k* is more than an order of magnitude less than k_3 the phase shift will be 180° and the structural response of the solution will dominate. k* can be determined from the relationship

$$2k_1/k_3 = k_3/k^*$$

and $$k^* = k_3^2/{}_2k_1$$

which is derived from comparing the gains at k_3 and k* to the change in T given that the slope is 2:1 in this portion of the characteristic. If k_2 is much greater than k* than the characteristic approximates that of a second order system with a pair of complex conjugate exponential coefficients at k*.

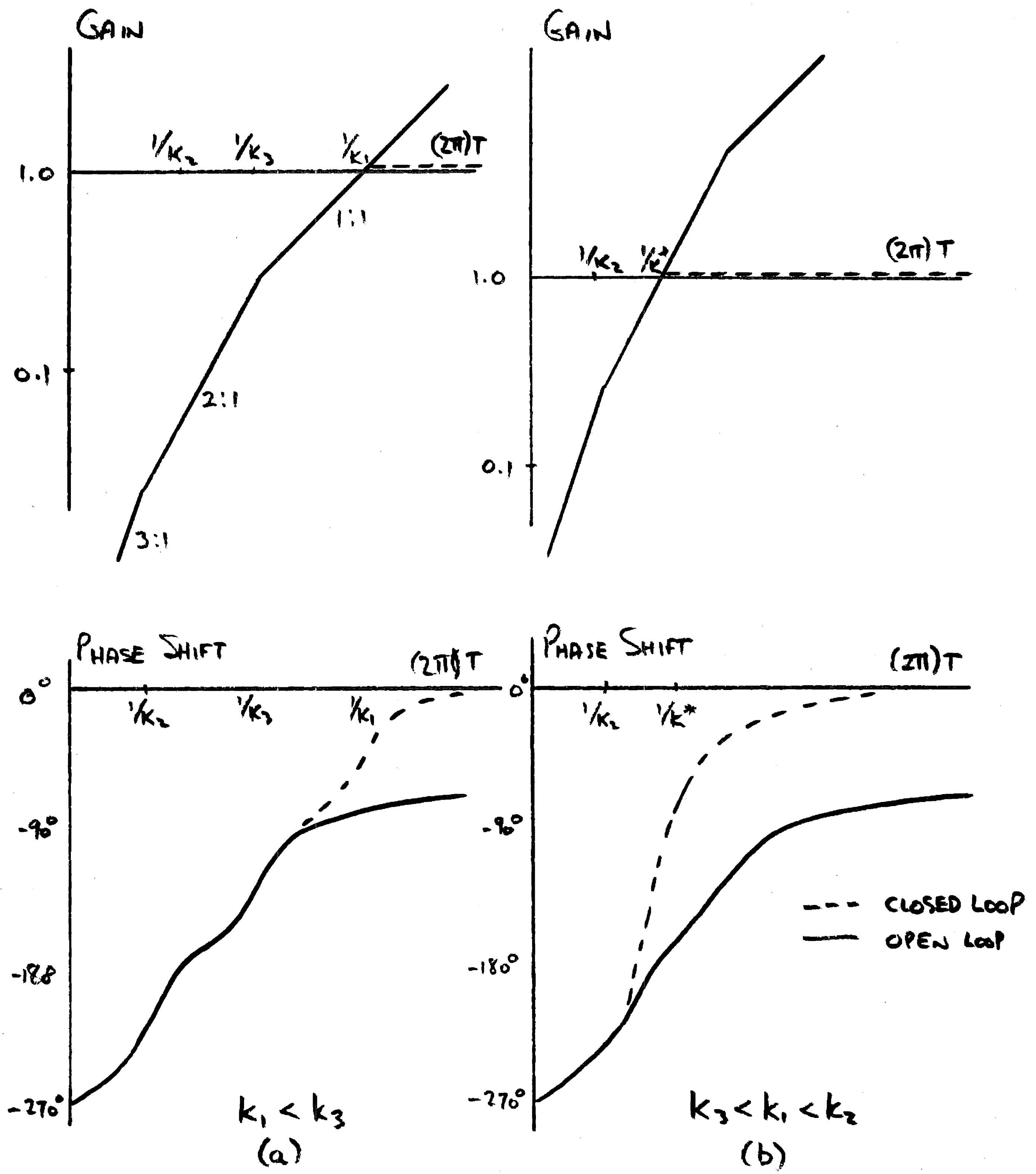


FIGURE 4-10: THE GAIN-PHASE CHARACTERISTICS OF THE PLANT WORK FORCE SYSTEM WITH A FIRST ORDER HIRING DELAY

From our analytic derivation in Chapter 3, it was observed that such a system produced an oscillatory structural response that decayed at a rate of $k^*/2$. From this type of analysis it is possible to determine the approximate range of values over which satisfactory system performance can be obtained. The sensitivity of the system to a higher order delay than the first order chosen initially can also be tested. Assume that the hiring delay in the inner feedback loop is better represented by a third order delay. If it is also assumed that $k_3 > k_2$ then the only change that is needed is to increase the slope after k_3 to 3:1 from 1:1 and to change the phase shift transition from 0 to -90° to 0 to -270° (90° for each level in the delay). This increased phase shift transition, however, makes the system much more sensitive to the value of k_1 selected. The resultant gain phase characteristics in Figure 4-11 for cases when $k_1 > k_3$ and $k_3 < k_1 < k_2$ can be used to estimate the system response. The phase shift transition due to the third order delay now moves from -90° to -360° and is centered at -225°. Thus any value of $k_1 > k_3$ will result in a dominant structural solution and to achieve normal operation k_1 will probably have to be an order of magnitude less than k_3. Under such conditions the total system behavior can best be approximated by a single first order negative feedback system of delay $1/k_1$. If k_2 had been less than k_3 the approximate solution might also include its effects in the form of a second component to a non-oscillatory second order response.

This type of analysis can be applied to linear systems of arbitrary order to determine both the gross behavior of the system (whether the structural or driven response dominates) and the best approximation that can be used to develop a specific system response. It is also particularly helpful in showing the effect of each additional feedback loop and its relative value in the system structure.

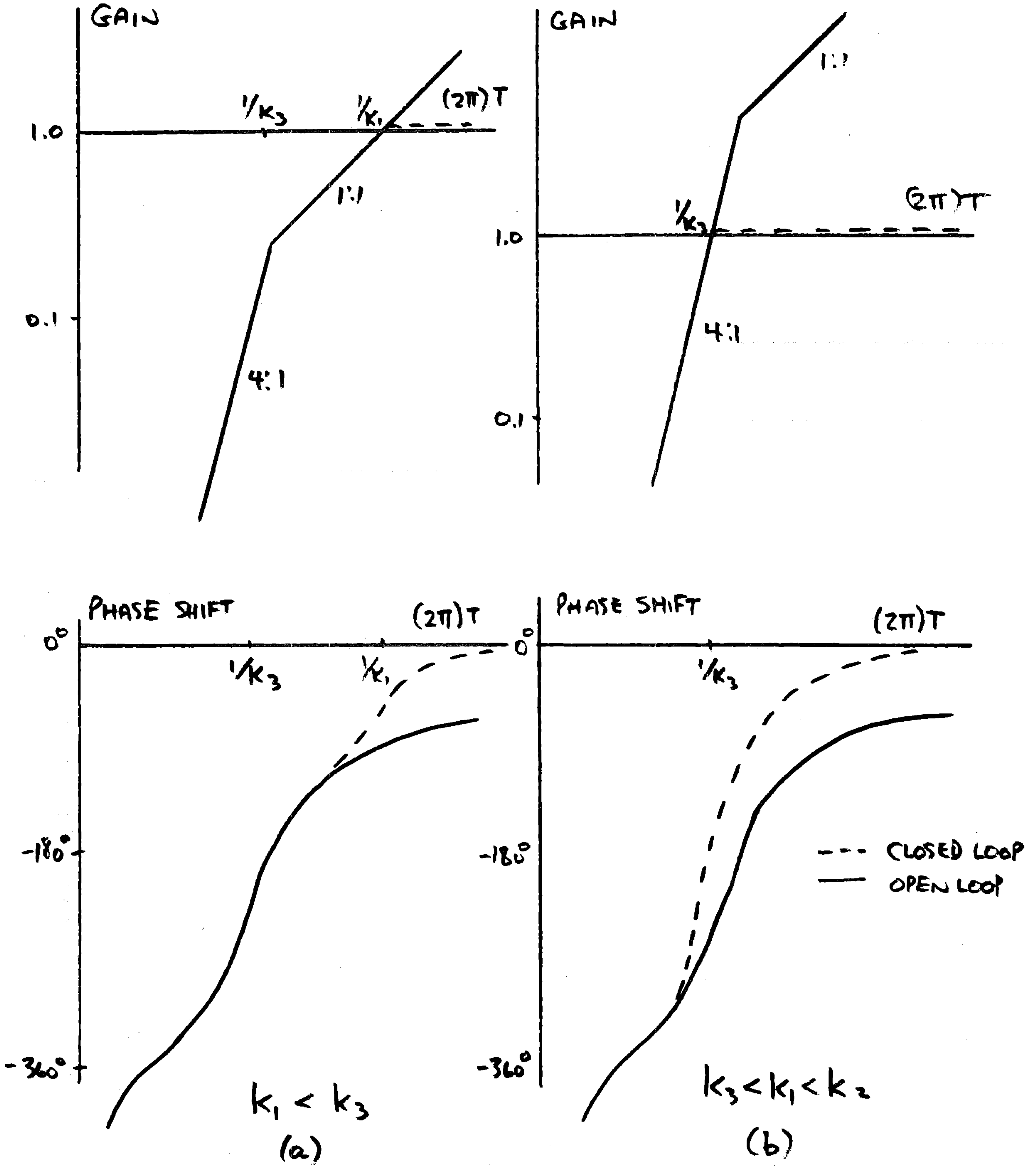


FIGURE 4-11: THE GAIN-PHASE CHARACTERISTICS OF THE PLANT WORK FORCE SYSTEM WITH A THIRD ORDER HIRING DELAY

CHAPTER V

FUTURE DIRECTIONS

The body of this thesis has outlined a method of teaching basic system dynamics that proceeds through two distinct stages, an analytic foundation for both first and second order systems and a graphical approach for higher order systems that allows approximate solutions to be developed. As it stands the material is not ready for use in a course directly, but instead should be considered as a reference guide. The pieces of the material that are missing, as was mentioned in the Introduction, can be taken from the engineering curricula in their present form. There are, however, several areas that need strengthening before the material will be useable in a regular management course.

One of the weaknesses of the present treatment is the still limited nature of the development of second order systems. Although the structure characterized by two feedback loops, one within the other, is analyzed thoroughly, it is only handled for one configuration with the inner loop about the second level. Additional work needs to be done which handles situations in which the inner feedback loop is about the first level as it is in the case of a first order delay. Also those systems in which there are loops about both levels as well as a feedback loop from the second level back to the first level must be studied. Such analyses are possible using the same approach as has been demonstrated in this case. In its final form, the course should provide the student with the ability to set up the integral equations for any second order system and obtain an analytic solution.

Another portion of the material that requires some additional effort is the development of a meaningful interpretation of the interaction of the system parameters in a second order system. In a simple first order system the system parameter k that represents the fractional adjustment of the difference or error function in a negative feedback loop can be directly related to the equivalent or average delay of a system. In fact, it is simply the inverse of this fraction. The situation in second order systems is more complex, however. In the nested feedback loop system, the time coefficients of the exponentials in the structural response were shown to be a complex function of the two system parameters associated with the two levels. This function is commonly represented by the quadratic formula. Although "...Every student of algebra is familar with the formula ...",[1] very few understand its implications in terms of the effects of one part of a system on another. How do the individual elements respond to the flows that surge through the system? What are the points at which k_2 relative to k_1 cause the system to overshoot its desired equilibrium and exhibit oscillatory behavior instead of the damped exponential transition between system states. To develop a more complete understanding on the part of the student, it would be important to develop these ideas more fully.

Another addition to the material that would be possible once the additional second order system configurations were analyzed would be a complete catalog of first and second order configurations and their characteristic behavior patterns.

---

[1]Thomas, George P., Calculus and Analytic Geometry, (Addison-Wesley Company, Reading, Mass.), 1958, p. 620.

Also, exercises could be developed which would drill the student until his responses to a system model would become more or less automatic. In addition the student needs exposure to many more examples than have presented here and the extent to which they are realistic and intuitively acceptable could be greatly improved upon. The best type are those which can be used continually throughout the material, initially being used to demonstrate simple structures and gradually being increased in sophistication and their ability to handle more of the modes of behavior exhibited by the real system. Also, the ability to carry across these ideas in terms of real management systems will greatly enhance the student's ability to carry over into the productive Industrial Dynamics work at a later time.

Short Term Needs

The most critical current requirement is manpower to begin to experimentally teach the material and to develop exercises to support the basic course material. It seems most appropriate to teach experimentally at first in that the students will find many flaws as they attempt to assimilate the material. Rather than expend more effort to refine the material from its present state, it would be more effective to work with a small group of students. Ideally a man will be found that has a mathematical or technical background sufficient to understand the depth of presentation in this thesis and who is able to expand and develop some of the material as he prepares to make a specific course offering. If possible he should be supported by a research assistant who, while extending areas such as second order system analysis can be contributing to the course by developing some of the exercises as operating a computer laboratory period in which the students can evaluate the models that they will be treating analytically.

If a faculty member or instructor cannot be obtained to teach the course immediately (and this should be given first priority), it will still be most important to gain some teaching experience with the material immediately. The next best possibility is to introduce portions of the material into the course that is presently being taught emphasizing the development of intuitive abilities through the experimental study of models. The students themselves could assist in the material evaluation and possibly even some of the exercise development. It is not inconceivable that some combination of these two basic approaches will in the long run provide the best foundation for Industrial Dynamics work. After the course has been taught at least three times and well-tested material has resulted, including exercises, it will be time to publish the material in book form. By this time research assistants and possibly even a doctoral student will have been able to make progress in the areas that need extra work and their contributions can be added also. At this point material will be able to flow to other management schools and the task will be well on its way to completion.