

CONFUSION IN THE POLARITY OF MAJOR LOOPS

Mohammad T. Mojtahedzadeh

George P. Richardson

The Rockefeller College of Public Affairs and Policy

State University of New York at Albany, Albany, NY, 12222

Abstract:

There is an *a priori* in the system dynamics paradigm which implies that the behavior of a system is tightly related to its structure. A structure specifies the relationships between the elements of the system in terms of stocks and flows and feedback loops. The attempt is to explain the behavior of a system based upon the accumulations in a system and its positive and negative feedback loops. This note introduces the concept of implicit loops or pathways, as components of a major loop, to explain how changes in the behavior of a variable of the system occur over time. The note demonstrates that the presence of a negative loop (explicit or implicit) around a positive loop (explicit or implicit) is necessary in order for a system to oscillate. In particular, it explains how a pure positive major loop can oscillate.

There is an *a priori* in the system dynamics paradigm which implies that the behavior of a system is tightly related to its structure. A structure specifies the relationships between the elements of the system in terms of stocks and flows and feedback loops. The basic building block of dynamic models is the feedback loop. The feedback loop is a path coupling decision, action, level of the system, and information, with the path returning to the decision point (Forrester, 1968). The behavior of the system is then the result of the interaction between different feedback loops embodied within the system boundary. In control theory, the behavior of a system is usually characterized by eigenvalues, eigenvectors, poles, zeros, gain, phase, and so on. These concepts, however, help us to understand the system, but they are not relevant to the real process. Since the feedback loops are based on the real process in a phenomenon, the attempt is to explain the behavior of the system in terms of feedback loops. It is well known that a negative minor loop corresponds to a goal seeking behavior, and a positive minor loop generates exponential growth. There are no general rules about the behavior of negative and positive major feedback loops.

The more complicated behaviors, such as overshoot, S-shape, and oscillation, are

created by feedback loops with different polarity and most likely the possibility of shifts in their polarity. Literally, the growth phases of these behavior are the result of positive loops, and the goal seeking phases are generated by negative loops. Therefore, a necessary condition for oscillation is the presence of a negative loop around a positive loop. The growth phase and goal seeking phase of an oscillatory variable are shown in Figure 1.

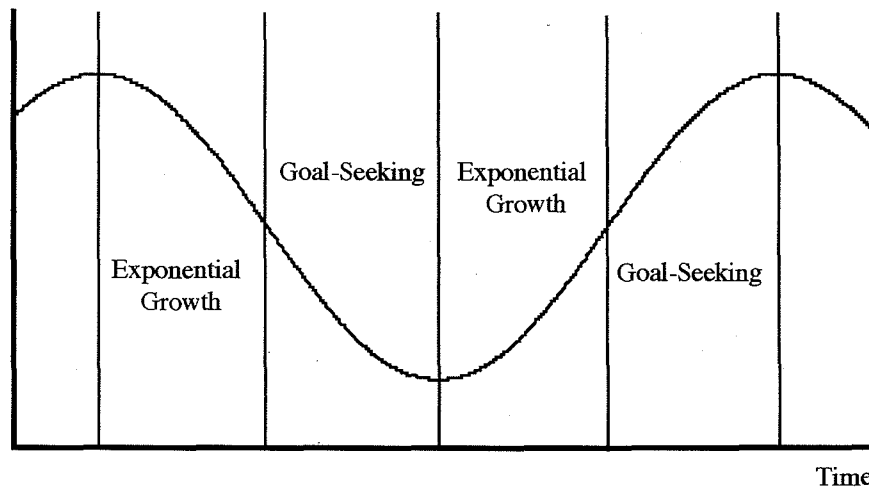


Figure 1: Growth Phase and Goal-seeking Phase of an Oscillation

A simple example which contradicts this theory is a single major loop with negative polarity that create oscillation. Another example is presented by Allan Graham (1977). His counterexample demonstrates that a pure positive loop can generate oscillations.

This note applies an extension of the formal definition of loop polarity to introduce the concept of implicit loops or pathways, as components of a major loop. It demonstrates that the presence of a negative loop (explicit or implicit) around a positive loop (explicit or implicit) is necessary in order for a system to oscillate. As will be defined later, the implicit loops help people to discern the behavioral characteristics of a dynamic system from its feedback structure. Although, the note deals with Graham's counterexample, the reader can apply this procedure to understand how a negative major loop with two level variables and no minor loop oscillates¹, or how a negative major loop with three level variables and

1. The equations of such an example are:

$$dX/dt = -Y$$

$$dY/dt = X$$

It is obvious that there is only one major loop with negative polarity in this system.

no minor loop generates exploding oscillation². Graham's counterexample uses a linear fourth order system with no minor loop and only one major loop. Since the parameters of the model are assumed to be positive (and equal to one), one may conclude that the polarity of the major loop is positive. A STELLA diagram of the model is depicted below.

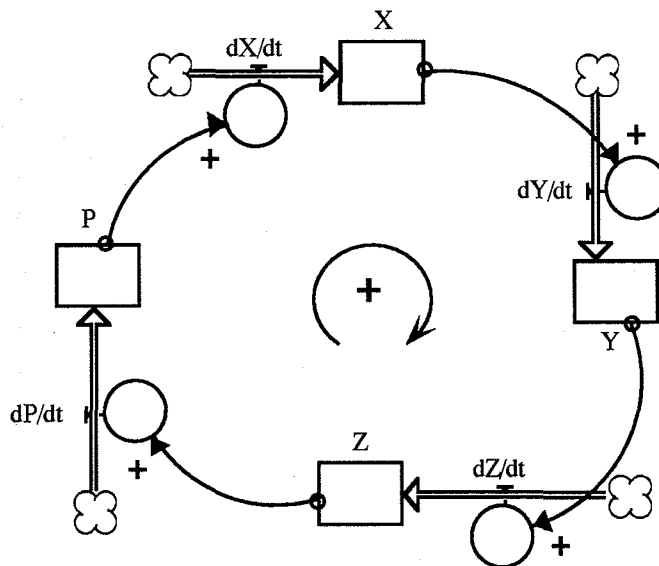


Figure 2: STELLA Diagram of a Major Positive Loop

The equations for the system are:

$$\begin{aligned} dX/dt &= P & [1] \\ dY/dt &= X & [2] \\ dZ/dt &= Y & [3] \\ dP/dt &= Z & [4] \end{aligned}$$

The eigenvalues of such a system are $+1, -1, +j, -j$, where $j = \sqrt{-1}$. The eigenvalue of $+1$ implies that the system grows exponentially. The eigenvalue of -1 indicates how quickly the system moves to its steady growth. The eigenvalues of $+j$ and $-j$ suggest that the system can oscillate. The trajectory of such a system looks like:

$$X(t) = a_1 * e^t + a_2 * e^{-t} + a_3 * \sin(t + \phi) \quad [5]$$

where a_1, a_2, a_3 , and ϕ depend on the right and left eigenvectors and initial conditions. Figure 3 shows the behavior of the model, when the initial values for X and Y are -1 and for Z and P are $+1$. Since no other loops are present, the oscillation then is presumably a

². The equations for a negative major loop with three level variables and no minor loop are:

$$\begin{aligned} dX/dt &= -Z \\ dY/dt &= X \\ dZ/dt &= Y \end{aligned}$$

result of the major positive loop. If all initial values are positive, the system will create exponential growth with a growth rate of one. In other initial values, the model will grow exponentially with the growth rate of one, after a transition period.

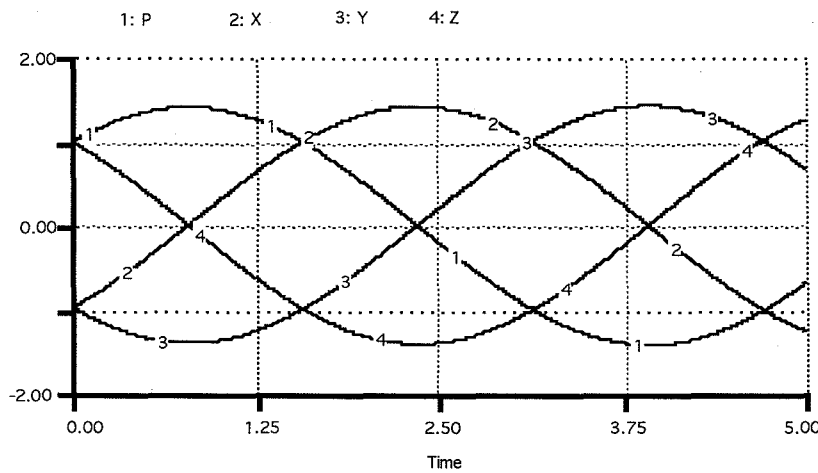


Figure 3: The Oscillatory Mode of a Fourth Order System with no Minor Loop

It is reasonable that the model generates pure exponential growth. However, because of the absence of a negative loop around the major loop shown in Figure 1, one does not expect the model to overshoot or oscillate.

The oscillation raises the question of whether it is enough to say that the loop shown in Figure 1 is positive. What do we mean by a positive loop? There are several definitions in the system dynamics literature of a positive loop. One of the more elaborated definitions is suggested by Graham (1977). He defines a loop as positive if its open-loop steady state gain is greater than one. This definition addresses the steady state characteristic of the system, while we may be interested in the transient regimes of system behavior rather than its long term behavior. Furthermore, what does a gain greater than one mean intuitively? Richardson (1995) suggests that a loop is positive if the sign of the partial derivative of the rates (inflow, outflow, or net flow) with respect to its corresponding level variable is positive. Therefore, a loop is positive if,

$$\text{Sign} \left[\frac{\partial(dX/dt)}{\partial X} \right] > 0 \quad [6]$$

where X is a level variable, and dX/dt is the rate of change in the level variable. A loop is

negative if;

$$\text{Sign} \left[\frac{\partial(dX/dt)}{\partial X} \right] < 0 \quad [7]$$

The last definition is consistent with the intuitive characterization. Richardson argues: " ∂X " can be thought of as "a small change in X" which is traced around the loop until it results in "a small change in dX/dt " in the inflow rate. If the change $\partial(dX/dt)$ is in the same direction as the change ∂X , then they have the same sign." Richardson uses this definition to determine the polarity of major loops as well. Suppose a major loop consists of n level variables; X_1, X_2, \dots, X_n . The polarity of the major loop is defined as:

$$\text{Sign} \left[\frac{\partial(dX_1/dt)}{\partial X_n} * \frac{\partial(dX_2/dt)}{\partial X_1} * \dots * \frac{\partial(dX_n/dt)}{\partial X_{n-1}} \right] \quad [8]$$

Form the above definition one can see that the polarity of the links coming from the major loop into a rate variable are determined, while the links coming from minor loops are ignored. The polarity of the major loop is then the sign of the production of the polarity of the links of the major loop. It is obvious that the links which connect rates to a level variable are missing. Thus, changes in inflows to outflows (or outflows to inflows) over time which is likely to occur in a dynamic system cannot be captured. This definition tells only part of the story about the behavior of the system. The limitation arises form the fact that some links coming into level variables are assumed to be constant. Relaxing this assumption eliminates the limitation mentioned above and leads to the extension of formal definition of loop polarity.

Although the extension of the formal definition of loop polarity is more complex, it provides a more precise statement about polarity of implicit and explicit loops. In this framework, explicit loops are equivalent to the conventional definition of major feedback loops. An implicit loop is a pathway that starts from a particular state variable, X_1 , and traces around the major loop until it reaches to the corresponding rate variable, dX_1/dt . Thus, in a linear system, the number of implicit loops, or pathways, in a major feedback loop is equal to the number of state variables in the major loop. It is important to note that pathways are not "intrinsic" but merely "descriptive". Their polarity is time dependent and

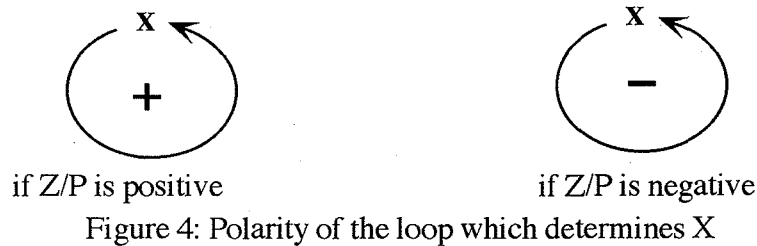
sensitive to initial values. They expose a connection between the behavioral characteristics of a system and its structural characteristics.

To understand the polarity of the major loop of the system shown in Figure 1, the extension of the formal definition of loop polarity is applied as follows:

Taking the derivative of [1] with respect to X yields:

$$\begin{aligned} \frac{d(dX/dt)}{dX} &= \frac{dP}{dX} \\ &= \frac{dP/dt}{dX/dt} \\ &= \frac{Z}{P} \end{aligned} \quad [9]$$

The polarity of the implicit loop which determines X (Figure 4) depends on the sign of Z/P . If both Z and P are positive or negative, then the implicit loop is positive. This means that a change in X leads to a change in the inflow rate of X in the same direction. Otherwise, if either X or P is negative (and the other one is positive) then the polarity will be negative. Thus, a change in the inflow rate of X as a result of a change in X is in the opposite direction. The polarity of the implicit loop or pathway, shown in Figure 4, shifts from positive to negative (or negative to positive) at point $Z=0$. The polarity of this implicit loop is not defined at $P=0$.



By taking the derivative of [2] with respect to Y we have:

$$\frac{d(dY/dt)}{dY} = \frac{P}{X} \quad [10]$$

The polarity of the implicit loop which determines Y is positive if both P and X are positive or negative, which means a change in the inflow rate of Y caused by a change in Y is in the same direction. If either X or P is negative (and the other is positive) then the polarity will be negative. This implies that a change in Y leads to a change in its inflow rate in the opposite direction. The polarity of the implicit loop, shown in Figure 5, shifts from

positive to negative (or negative to positive) at point $P=0$. The polarity of this loop is not defined at $X=0$.



Figure 5: Polarity of the loop which determines Y

By the same procedure, the polarity of the implicit loops which determine Z and P can be determined. Therefore,

$$\frac{d(dZ/dt)}{dZ} = \frac{X}{Y} \quad [11]$$



Figure 6: Polarity of the loop which determines Z

The polarity of the implicit loop which determines Z is positive if both X and Y are positive or negative. Thus a change in Z leads to a change in its inflow rate in the same direction. If either X or Y is negative (and the other is positive) then the polarity will be negative, which means a change in the inflow rate of Z as a result of a change in Z is in the opposite direction. The polarity of the implicit loop, shown in Figure 6, shifts from positive to negative (or negative to positive) at point $X=0$. The polarity of this implicit loop is not defined at $Y=0$. For P we have:

$$\frac{d(dP/dt)}{dP} = \frac{Y}{Z} \quad [12]$$

the polarity of the implicit loop which determines Z is positive if both X and Y are positive or negative. If either X or Y is negative (and the other is positive) then the polarity will be negative.

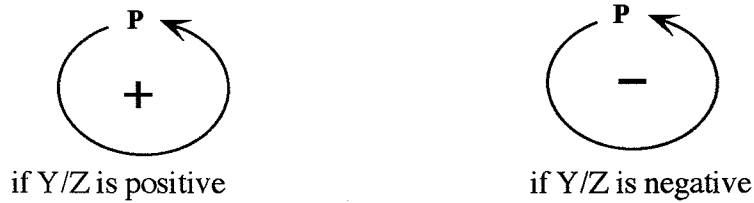


Figure 7: Polarity of the loop which determines P

The polarity of the implicit loop which determines P is positive if both Z and Y are positive or negative. This means a change in P leads to a change in the inflow rate of P in the same direction. If either Z or Y is negative (and the other is positive) then the polarity will be negative which implies a change in inflow rate of P as a result of a change in P is in the opposite direction. The polarity of the implicit loop, shown in Figure 7, shifts from positive to negative (or negative to positive) at point $Y=0$. The polarity of this implicit loop is not defined at $Z=0$.

As we have seen, a change in a state variable results in a change in its inflow rate in the same or opposite direction. This simply means that the implicit loop which determines a state variable can be positive or negative. Furthermore, the implicit loop that is responsible for generating the behavior of a state variable may or may not have the same polarity as an implicit loop which determines another state variable.

In the above example, when X, Y, Z, and P all are positive (or negative), the implicit loop which determines any state variable has positive polarity. This implies that all derivatives of each state variable are positive (or negative). For example for X we have:

$$\frac{dX}{dt} = P > 0$$

$$\frac{d^2X}{dt^2} = \frac{dP}{dt} = Z > 0$$

and in general, since X, Y, Z, and P are always positive, then

$$\frac{d^n X}{dt^n} > 0 ; \text{ for all } n$$

In such a condition, the system will produce exponential growth. When one or more of the state variables are negative and the rest of them are positive the system overshoots or oscillates, but the growth mode will be dominated in the long run, unless the absolute value

of all initial value problems are the same and sum to zero (see figure 3). It is worth noting that under this circumstances some derivatives of a state variable are negative and some are positive.

In any case, the polarity of the major loop in Figure 1 is positive. In this Figure, those implicit loops that are responsible for generating the behavior of different state variables (Figure 4 through 7) are only tacitly being taken into account. Indeed, they are collapsed into one loop as shown in Figure 1. The polarity of the implicit loop that is responsible for creating the behavior of a level variable over time can be thought of as one of the components of the polarity of the aggregated loop. The polarity of the components of the aggregated major loop (Figure 1) varies in terms of magnitude and sign as level variables change. The apparent changes in the strength and sign of the polarity of the components of the major loop may cause the system to oscillate.

The polarity of the aggregated major loop shown in Figure 1, is a function of the polarity of its components. A simple function form, but not complete, may be a multiplicative one:

$$LP = \frac{dP/dt}{dX/dt} * \frac{dX/dt}{dY/dt} * \frac{dY/dt}{dZ/dt} * \frac{dZ/dt}{dP/dt} \quad [13]$$

where LP is a rough measure of the polarity of the aggregated major loop. Substituting the equivalents of the elements of LP, which are presented in equations [9] through [12] gives:

$$LP = \frac{Z}{P} * \frac{P}{X} * \frac{X}{Y} * \frac{Y}{Z} = 1 \quad \text{for } X, Y, Z, \text{ and } P > 0 \quad [14]$$

From the above equation, one can see the polarity of the aggregated major loop is greater than zero. When any of the state variables (X, Y, Z, P) are zero, the polarity of the loop is not defined. The apparent discontinuity in the polarity of aggregated major loop is generated by the shifts in the polarity of its components. Notice that if somehow only one state variable, say P, becomes negative while the other state variables are positive, the loop which determines X becomes negative (Figure 4). At the same time, the polarity of the implicit loop which determines Y shifts from positive to negative. As a result, LP, a measure of the polarity of aggregated loop, remains positive. Because of the problem of cancellation, changes in the polarity of implicit loops shown in Figure 4 through 7, are not completely reflected on the measure of polarity of aggregated major loop.

In summary, the aggregate measure of the polarity of major feedback loops is not sufficient to characterize the behavior of complex systems. The application of the extended formal definition of loop polarity to identifying the polarity of major loop implies that the components of a major loop do not necessarily have a fixed and continuous polarity. However, the multiplicative function form used in [13], which is rough, abstract and incomplete, cannot reflect all changes in the polarity of the components of the major loop. The polarity of the components of the major loop may change, while the polarity of the major loop itself remains unchanged. The shifts in the polarity of the components of the aggregated major loop create discontinuity in the polarity of the major loop and lead the system to overshoot or oscillate. It should be noted that the extension of the formal definition of loop polarity only elaborates the identification of the polarity of the components of a major loop, not the major loop itself. The components of a major loop -- or implicit loops -- reveal the behavioral characteristic of the system. Therefore, they are merely descriptive, time variant and sensitive to initial conditions. The explanatory power of implicit loops enable us to understand how changes occur in the behavior of a variable of the system over time. In general, a positive (negative) implicit loop causes the corresponding level variable to grow exponentially (goal-seeking), unless it is dominated by other positive (negative) loops.

References:

- Forrester, Jay W. 1968. Principles of Systems. Cambridge, MA: The MIT Press, Wright-Allen Series.
- Forrester, Jay W. 1968. Market Growth as influenced by Capital Investment. Industrial Management Review, 9, 2. Cambridge, MA: The MIT Press, Wright-Allen Series.
- Graham, Alan K. 1977. Principles of the Relationship between Structure and Behavior of Dynamic Systems, Ph. D. dissertation, MIT, MA, 02139
- Richardson, G. P. 1995. Loop Polarity, Loop Dominance, and the concept of dominant polarity. *System Dynamics Review* 11(1) :67-88
- _____. 1986. Problems with causal-loop diagrams. *System Dynamics Review* 2(2) :158-170.
- _____. 1986. Dominant Structure. *System Dynamics Review* 2(1) :68-75.