9

Transfer Functions and Block Diagrams

Background

Feedback control systems are made up of components that are reactive by nature. This means that each one has an input (sometimes more than one) and the means to generate an output. The inputs and outputs have a variety of forms, but in process control the most common are process variables and instrument signals.

To be usable, the output of a component must exhibit a consistent relationship to its input. Output relationships are not necessarily neat and tidy, but the same input must consistently produce the same output; otherwise the component is unacceptable.

Control systems experts need techniques to determine and describe how the components of a control system will perform. If the behavior of the individual components that make up the system can be identified, then the behavior of the overall system can be evaluated.

This leads to the question: What kind of behavior are we interested in? There are two factors:

1. The gain of the component. If the input to the component is changed a known amount, how much does its output change? The gain factor will be the ratio of the change in output to the change in the input that created it. The output change is measured as it goes from the initial steady state value to the final steady state value. Time is not a factor. The output is permitted all the time that is necessary to assume its new value.

A further consideration is whether or not the gain of the component remains the same over the whole range of operation of the component. If it doesn't, then this factor has to be allowed for in the control system design.

2. The dynamic response of the component. This does not mean how much the output of the component responds, but how quickly and on what pattern it assumes its new steady state output after its input is changed. Components that react quickly make for better overall control system performance.

Transfer Functions

While frequency response analysis uses a graphical method to describe the gain and dynamic response of components, transfer functions do the same thing using mathematical expressions. By using transfer functions, it is often possible to describe both the gain and the dynamic response of a component in a single mathematical function.

In general, the transfer function of a component is the ratio of the change in its output to the change in its input, but herein lies a problem. While the gain of the component can be identified by a simple number, the dynamic character (how it varies with time) of both the input and output can be described only by differential equations in which time is the independent variable, and then only if the input or output function is continuous, which it may not be. Obviously, a transfer function that consisted of the ratio of two differential equations would be of little practical use. The situation can be made workable, however, not by using the differential equations of the input and output, but by using their Laplace transforms.

The bottom line is that in feedback control systems, the transfer function of a component is defined as the ratio of the Laplace transform of its output to the Laplace transform of its input. What is now required are some examples of transfer functions, and then a study of just what performance information can be gained from them.

The Step Input Function

Frequency response analysis of a component or system is based on a sine wave input signal. A sine wave input results in a sine wave output. It is obvious, however, that if the input were anything other than a sine wave, then the output would be different, even though it emerged from the same component. For transfer functions to have practical value, therefore, it was necessary to standardize the form of the input signal.

In field test work, the most practical test signal is a step change from one input level to another, for at least two reasons. First, it is an easy test signal to devise, and second, the output that results from this relatively simple input change will yield all of the dynamic information that is of any real value. It is not surprising, therefore, that the step change was selected as the standard input for transfer functions. A further refinement was that the step input should have a unit value.

In the chapter on Laplace transforms, it was shown that the transfer function F(s), which results from a step change from zero to a value of C, is F (s) = C/s. If C = 1, then F(s) becomes 1/s. The simplicity of this function is another plus for an input consisting of a unit step change.

In real life process control systems, two particular components, time constants and dead time, predominate over all others. The transfer functions for these two components should now be worked out.

Time Constants

In Chapter 6 (Differential Equations), Example 1 described a component whose rate of change, in response to a step change input, is proportional to the distance remaining for the output to attain its ultimate value. The output in this case is the variable x. In the simplest case, the input and the output have the same value at t = 0, and the gain of the component is 1. Then the equation x = f(t) for the output is:

$$x = \left(1 - e^{-kt}\right).$$

The exponent of the exponential e is required to be dimensionless, and the units of t are time units. Accordingly, it is more realistic to set k = 1/T. T will now be in time units. T is, in fact, the time constant of the component. The differential equation thus becomes

$$x = \left(1 - e^{-\frac{1}{T}t}\right).$$

The table of Laplace transforms shows that the function

$$f(t) = \left(1 - e^{-\frac{1}{T}t}\right)$$

has the transform

$$F(s) = \frac{1}{s(Ts + 1)}.$$

Consequently, the transfer function for a time constant element will be

$$\frac{\text{Laplace transform of the output}}{\text{Laplace transform of the input}} = \frac{\dfrac{1}{s(Ts+1)}}{\dfrac{1}{s}}$$

$$= \frac{1}{Ts+1}.$$

Dead Time

If a component has dead time, then the output function will duplicate the input function f(t), but only after a delay of L time units. L is usually referred to as the *dead time*. The output function will consequently be f (t – L). The initial value of t must be zero (a Laplace transforms requirement), after which t increases positively.

The table of Laplace transforms shows that the transform for f (t – L) is e^{-Ls} F(s), where F(s) is the transform for f(t). Thus, for whatever form f(t) may have, the transfer function for a dead time element will be

$$\frac{\text{Transform of the output}}{\text{Transform of the input}} = \frac{e^{-Ls}F(s)}{F(s)} = e^{-Ls}.$$

The Value of the Transfer Function

The goal in developing transfer functions was to devise a mathematical expression that would incorporate both the steady state gain and dynamic characteristics of a control system component. It now remains to be shown that this goal has been achieved.

A useful attribute of the transfer function is that by applying the appropriate procedure, the transfer function will yield the frequency response data of its component. More specifically, from the transfer function, other functions will evolve from which the frequency response magnitude ratio and phase angle can be determined at any desired frequency. The procedure is as follows:

1. In the transfer function, replace the operator s with jω, where j is the imaginary quantity $\sqrt{-1}$, and ω is the angular velocity (radians per second).

2. Using the standard techniques for complex numbers, separate the resulting expression into its real part (RP) and imaginary part (IP). This will lead to the expressions for the frequency response magnitude ratio (MR) and the frequency response phase angle (ϕ). These expressions will be functions of the angular velocity ω, which is directly related to the frequency by the relation $\omega = 2\pi$ f.

3. The magnitude ratio at the specified value of ω will then be

$$\mathrm{MR} = \sqrt{(\mathrm{RP})^2 + (\mathrm{IP})^2}.$$

4. The phase angle at the specified value of ω will be

$$\phi = \text{angle whose tangent is } \frac{\mathrm{IP}}{\mathrm{RP}}, \text{ that is, } \phi = \tan^{-1}\left(\frac{\mathrm{IP}}{\mathrm{RP}}\right).$$

Example 1: Time Constant

The transfer function for a time constant component is

$$\frac{1}{\mathrm{T}s + 1}.$$

Converting to the frequency response domain, this expression becomes

$$\frac{1}{j\omega\mathrm{T} + 1}.$$

To separate the real and imaginary parts:

$$\frac{1}{j\omega\mathrm{T}+1} = \frac{1}{j\omega\mathrm{T}+1} \times \frac{j\omega\mathrm{T}-1}{j\omega\mathrm{T}-1} = \frac{j\omega\mathrm{T}-1}{-\omega^2\mathrm{T}^2-1} = \frac{1-j\omega\mathrm{T}}{-\omega^2\mathrm{T}^2+1}$$

$$= \frac{1}{\omega^2\mathrm{T}^2+1} - j\frac{\omega\mathrm{T}}{\omega^2\mathrm{T}^2+1}.$$

The real part $\mathrm{RP} = \dfrac{1}{\omega^2\mathrm{T}^2+1}$.

The imaginary part $\mathrm{IP} = \dfrac{-\omega\mathrm{T}}{\omega^2\mathrm{T}^2+1}$.

$$\text{The magnitude ratio} = \sqrt{(RP)^2 + (IP)^2} = \sqrt{\left(\frac{1}{\omega^2T^2+1}\right)^2 + \left(\frac{-\omega T}{\omega^2T^2+1}\right)^2}$$

$$= \frac{1}{\omega^2T^2+1}\sqrt{1+\omega^2T^2} = \frac{1}{\sqrt{\omega^2T^2+1}}.$$

$$\text{The phase angle} = \tan^{-1}\frac{IP}{RP} = \tan^{-1}\frac{\left(\frac{-\omega T}{\omega^2T^2+1}\right)}{\left(\frac{1}{\omega^2T^2+1}\right)} = \tan^{-1}(-\omega T).$$

It should be noted that in the development of the magnitude ratio and phase relationships above, the transfer function for the time constant was

$$\frac{1}{Ts+1}.$$

This would indicate a steady state gain of 1, which is not necessarily the case. If the time constant component contributes a steady state gain (magnitude k), as well as dynamics, then the transfer function would be

$$k\frac{1}{Ts+1}$$

and the magnitude ratio would be

$$\frac{k}{\sqrt{\omega^2T^2+1}}.$$

Example 2: Dead Time

The transfer function for a dead time element is $F(s) = e^{-Ls}$. In the frequency response domain, this becomes $e^{-j\omega L}$. Applying the work done in Chapter 5 (Complex Quantities), $e^{-j\omega L} = \cos \omega L - j \sin \omega L$.

Therefore, $RP = \cos \omega L$, and $IP = -\sin \omega L$.

The magnitude ratio will be $\sqrt{\cos^2 \omega L + \sin^2 \omega L} = \sqrt{1} = 1$.

In control system analysis, it is considered that there is only one dead time element, if any. If dead time is present in more than one place in a control system, all of the dead times can be summed together to form a single dead time component, without any loss of accuracy in the analysis.

The expression for the magnitude ratio shows that since the magnitude ratio is 1, dead time does not contribute any gain or attenuation to the control system, irrespective of the frequency. Increases or decreases in the overall system gain will be contributed by components of some other type, most likely by time constants.

The phase angle will be $\phi = \tan^{-1}\left(\dfrac{-\sin\omega L}{\cos\omega L}\right) = \tan^{-1}(-\tan\omega L) = -\omega L.$

Block Diagrams

A fact of life that sometimes eludes control system theorists is that process control earns its keep in industry, not in the lab. Furthermore, the brand of process control that is dominant in industry, and which will continue to dominate, is feedback control. There are two reasons for this.

First, feedback control requires the minimum investment. All that is required is a controlling device (analog or digital), a sensor from which the controller can get information about the variable it is supposed to control, and a final device, such as a control valve, which can manipulate some other variable whose value affects the value of the variable under control.

Second, to create a feedback control system, there is no requirement for any extensive engineering study. As far as the process to be controlled is concerned, it is only necessary to know that the same input to the process consistently produces the same reaction from the process.

It helps in the study of what goes on in a feedback control system if the system is diagrammed in block form, with each of the major components in the system represented by one block. The major components in a feedback system are the process, the controller, the sensor from which the controller gets its measurement information about what it is controlling, and the final device, which the controller uses to effect appropriate changes in the controlled variable. In a block diagram, a feedback control system appears in Figure 9-1.

The first (circular) block is the comparison block, in which the measured value of the controlled variable c is compared with the desired value r. The difference, ε, is equal to c – r. If the control system is on control, then $\varepsilon = 0$. If ε is not zero, corrective action by the controller is required. The comparison block is actually built into the controller but is shown separately in the

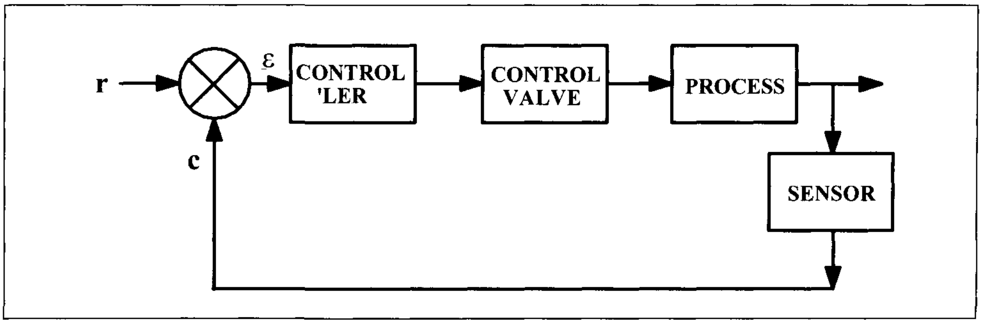

Figure 9-1. The Feedback Control System Block Diagram.

block diagram so that the location of the variables r, c, and ε can be identified.

The block diagram shows how the output of one block becomes the input of the next. Each block may contain one or more time constants, dead time, or other behavioral characteristics. The input to every block will be modified by whatever gain and dynamics the previous component contributes to the system.

The output of the process is the variable that is being controlled. In real life it is more likely to be a process operating condition such as pressure or temperature, rather than an actual product.

The diagram also shows the measured value of the controlled variable being fed back to the input of the controller. Hence the name *feedback control*. In addition, having the feedback path means that the system comprises a closed loop. Accordingly, the term *closed loop control* is also used for systems of this kind.

In any system, when some or all of the output of the system is fed back as an input to the system, this creates conditions under which the system can oscillate. Oscillations do occur in process control systems, and as such, are an important characteristic of the system.

The overall behavior of the control system can be determined if the characteristics of each of the components represented by its block can be established. This may be achieved in either of two ways.

1. If the output of the component in response to a step change input can be expressed as a differential equation, then a transfer function for the component can be written, and the frequency response magnitude ratio and phase angle data can be worked out.

2. The steady state and dynamic characteristics of the component can be determined by conducting an appropriate test. This is the method that usually has to be used to determine the characteristic of real processes.

In some texts, the control valve and sensor blocks are lumped into the process block, as in Figure 9-2.

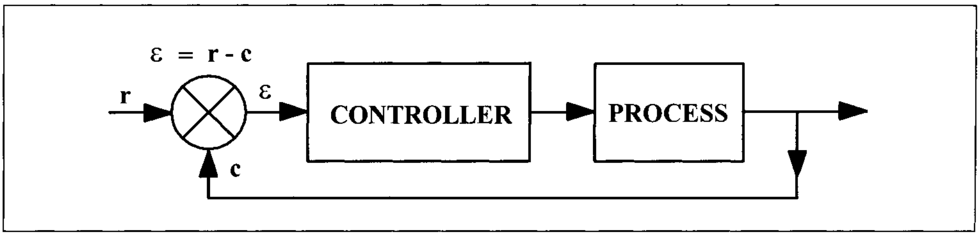

Figure 9-2. The Abbreviated Feedback Control System Block Diagram.

Since the controller does not know how many components are downstream of it, only the results that come back to it, it is considered satisfactory to show all three of these components as a single block.

An apparent anomaly may be present if the expression for the control error ε is written as (ε = r – c). From the logical point of view, the control error should be positive if c is greater than r, but this expression says the opposite. This point requires clarification, which will be forthcoming after one further matter has been considered.

Conditions for Continuous Oscillation

The feedback nature of the control system makes it possible for oscillations to occur. This raises the question: What conditions are necessary if the oscillations in the system are to be self sustaining and therefore continuous? Suppose that by using a signal generator or by some other artificial means, a sine wave input is introduced into the system via the r input. The oscillations would proceed around the loop and back to the comparator block as the c input. If the applied sine wave were then withdrawn, under what conditions would the oscillations continue on their own? Actually, two conditions would have to be met.

1. The magnitude of the sine wave that returned via the c input would have to be as great as the magnitude of the sine wave that was applied at the r input. If the c input magnitude were less than that of the r input, then the oscillations would die out.

2. The oscillations returning at the c input must be in phase with the sinusoidal r input. If the magnitude of the c sine wave input were

the same as that of the r input, but the two wave trains were out of phase, then the interference between the two out-of-phase waves would cause the oscillations to die out.

In real process control systems, the output from the process block will inevitably emerge lagging behind the input, owing to the dynamic delays that are inherent in the control valve, the process, and the measurement sensor. Thus, for the returning wave at the c input to be in phase with the incoming wave r, the returning wave must actually be 360° or one complete cycle out of phase with the input r. This effectively puts it back in phase with the r wave.

In Chapter 8 (Frequency Response Analysis), it was shown that the phase lag will increase with increasing frequency of the oscillations. From this it might be deduced that continuous oscillations in the system can occur only at the frequency that causes a phase lag of 360° in the process block.

However, at this point a second anomaly occurs. Control theory texts say that continuous oscillations occur at the frequency at which there will be 180°, not 360°, of phase shift in the process. How can this be reconciled?

The answer lies in the expression $\varepsilon = r - c$. We have already noted that this does not appear to be logical, but the difficulty lies in the fact that the expression has been abbreviated. The actual expression is

$$\varepsilon = (-1) \times (c - r).$$

This clears up the two apparent inconsistencies. First, the difference term is really $(c - r)$, not $(r - c)$, so logic prevails. The second is that multiplying a sine wave by (-1) inverts the wave, which is equivalent to shifting the phase by 180°. This inversion occurs inside the circular block. Consequently, 180° of phase lag in the process, plus the 180° contributed by the inversion factor -1, provides the overall 360° phase lag that is required to make continuous oscillations possible. The key factor to remember is that *continuous oscillation in a feedback control system will occur at the frequency that causes 180° of phase lag in the process.*

It is theoretically possible that all of the dynamic elements in the process block, in combination, will not produce a phase lag of 180°, no matter how great the frequency. In this case, oscillations will not be sustained in the loop, irrespective of the steady state gain in the loop. Such a hypothetical control system would have to have zero dead time and no more than two time constants, even if they are small. Since in actual control systems there will be at least one time constant in each of the control valve, the process, and the sensor, a real process control system that cannot oscillate does not exist.

The Transfer Function of a Closed Loop

The diagram below is the abbreviated closed loop diagram except for the different symbols. The transfer functions are identified by the letter G, so that G_c is the transfer function of the controller, and G_p is the transfer function of the process. The input and output, which will be functions of time, are identified by the letter Z, with Z_i being the input to the system, and Z_o being the output.

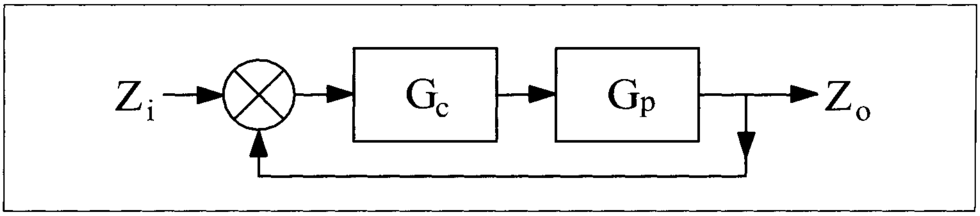

Figure 9-3. Block Diagram with Transfer Functions.

For the two blocks, the input to the second block is the output from the first, so the transfer function for the two blocks in tandem will be the product G_cG_p. At first glance, it seems that the ratio of the system output to the input should be

$$\frac{Z_o}{Z_i} = G_cG_p.$$

This would be the case if there were no feedback. However, the feedback path that directs the output back to the input of the loop, adds a complication. The input to the G_c block will not be Z_i, but $Z_i - Z_o$. Consequently,

$$\frac{Z_o}{Z_i - Z_o} = G_cG_p. \text{ So, } Z_o = G_cG_pZ_i - G_cG_pZ_o$$

$$Z_o\left(1 + G_cG_p\right) = G_cG_pZ_i \text{ and } \frac{Z_o}{Z_i} = \frac{G_cG_p}{1 + G_cG_p}$$

which is the transfer function for the closed loop.

Evaluating the Closed Loop Transfer Function

The frequency response gain and phase lag for the closed loop might be obtained through the rather onerous application of mathematics, but Peter Harriott, in his excellent text (*Process Control*, McGraw-Hill, New York City, 1964), suggests an easier approach.

Since any component represented by a transfer function has a gain and a phase lag at every frequency, these two characteristics can be represented by a vector, and the overall gain and phase lag can then be determined by adding vectors to produce the vector for the closed loop.

To simplify matters, let the product G_cG_p be replaced by G, so that the closed loop transfer function becomes

$$\frac{G}{1+G}.$$

In Figure 9-4, the vector G has a length equivalent to the open loop gain, and it is orientated at an angle (a) equal to the open loop phase lag, at the selected frequency.

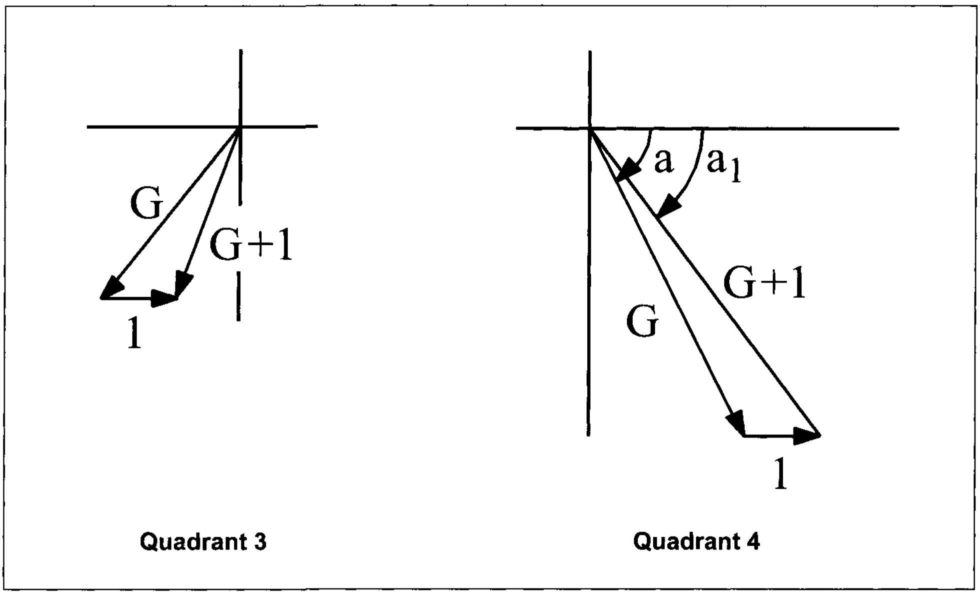

Figure 9-4. Using Vectors to Determine the Closed Loop Gain and Phase Angle.

The 1 vector, as a constant, has unit magnitude but no phase angle. Its direction is parallel to the horizontal axis and in the positive direction of the horizontal axis. The result of adding the G vector and the 1 vector diagrammatically is the vector representing G + 1.

At a somewhat higher frequency than that represented by the quadrant 4 diagram, the greater phase lag shifts the vector diagram into quadrant 3. The gain will likely be lower also, as indicated by the shorter G vector.

To find the closed loop gain and phase lag, it is necessary to divide the G vector by the G + 1 vector, since the form of the transfer function for the closed loop is a quotient. The procedure for dividing vectors of this type calls for *dividing* the magnitude of the vector in the numerator by the magnitude of the vector in the denominator, and *subtracting* the phase angle of the vector in the denominator from the phase angle of the vector in the numerator. Following this procedure yields the closed loop gain and phase lag at the particular frequency at which the open loop gain and phase lag were determined.

Specifically, this means that the closed loop gain will be equal to

$$\frac{\text{Magnitude of the G vector}}{\text{Magnitude of the (G+1) vector}}$$

and the closed loop phase lag will be equal to $(a - a_1)$.

Figure 9-5 appears confusing but it contains an important fact. It represents the case in which the open loop phase lag is –180°. At this particular frequency, it is quite possible that the open loop gain will have fallen off to a value less than 1, which is why the G vector in Figure 9-5 is shown considerably shorter than in Figure 9-4. At a phase angle of –180°, all three vectors are going to lie on the horizontal axis.

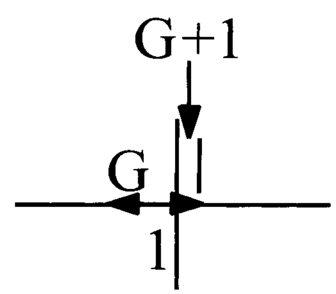

Figure 9-5. Both Vectors Have the Same Phase Lag at 180°.

The G vector, which represent the open loop, will point in the negative direction, in accordance with the –180° phase lag. If the open loop gain is less than 1, then when the unit vector is added, the resultant G + 1 vector will end on the positive side of the vertical axis, pointing in the positive direction. This indicates a phase angle of zero. Thus, when the phase angle of the G + 1 vector is subtracted from the open loop phase angle, the closed loop phase angle turns out to be –180° – 0°, or –180°, the same as the open loop phase angle.

The all important point here is that the frequency that creates a phase lag of $-180°$ in the open loop also creates a $-180°$ phase lag in the closed loop, and *this is the crucial frequency at which the closed loop is going to oscillate.* No other frequency will cause an identical phase lag in both the open and closed loops.