

Stability of Closed-loop Systems

1. Introduction

A feedback control system must be stable as a prerequisite for satisfactory control. Consequently, it is of considerable practical importance to be able to determine under which conditions a control system becomes unstable. For example, what values of the PID controller parameters K_c , τ_I , and τ_D keep the controller process stable?

Definition of stability

Before we proceed, we introduce the following definition for unconstrained linear systems. Notice that the term “unconstrained” is used to refer to the ideal situation where there are no physical limits on the output variable.

Definition of stability. An unconstrained linear system is said to be stable if the output response is bounded for all bounded inputs. Otherwise, it is said to be unstable.

Characteristic equation

Consider the general block diagram, which is discussed in the previous chapter. Using block diagram algebra that was developed in the previous chapter, we obtain

$$y = \frac{G_c G_v G_p}{1 + G_m G_c G_v G_p} y_{sp} + \frac{G_L}{1 + G_m G_c G_v G_p} d \quad (1)$$

or equivalently,

$$C = G_{SP} R + G_{Load} L \quad (2)$$

The stability characteristics of the closed-loop response will be determined by the poles of the transfer functions G_{SP} and G_{Load} . These poles are common for both transfer functions (because they have common denominator) and are given by the solution of the equation

$$1 + G_c G_m G_v G_p = 0 \quad (3)$$

Equation (3) is called the *characteristic equation* for the generalized feedback system.

Let p_1, p_2, \dots, p_n be the n roots of the characteristic Equation (3):

$$1 + G_c G_m G_v G_p = (s - p_1)(s - p_2) \dots (s - p_n) \quad (4)$$

Then we can state the following criterion for the stability of a closed-loop system:

A feedback control system is stable if all the roots of its characteristic equation have negative real parts (i.e. are to the left of the imaginary axis).

If any root of the characteristic equation is on or to the right of the imaginary axis (i.e. it has real part zero or positive), the feedback system is unstable. Figure 1 provides graphical interpretation of this stability criterion. The qualitative effects of these roots on the transient response of the closed-loop system are shown in Figure 2. The left portion of each part of this figure shows representative root locations in the complex plane. The corresponding figure on the right shows the contributions these poles make to the closed-loop response to a step change in the set point. Similar responses would occur for a step change in load.

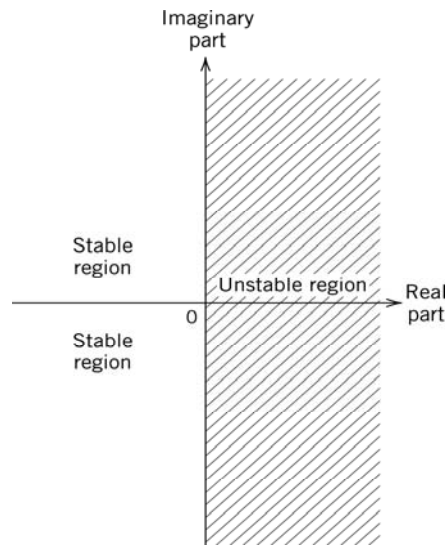


Figure 1 Stability regions in the complex plane for roots of the characteristic equation.

The root locations also provide an indication of how rapid the transient response will be. A real root at $s = p_1$ corresponds to a closed-loop time constant of $\tau_1 = 1/p_1$. Thus, real roots close to the imaginary axis result in slow responses. Similarly, complex roots near the imaginary axis correspond to slow response modes. The further the complex roots are away from the real axis, the more oscillatory the transient response will be.

Remarks

The product $G_{OL} = G_c G_m G_v G_p$ is called *open-loop transfer function* because it relates the measurement indication y_m to the set point if the feedback loop is broken just before the comparator

$$y_m = G_{OL} y_{sp} \quad (5)$$

Note that the same characteristic equation occurs for both load and set-point changes since the term, $1+G_{OL}$, appears in the denominator of both terms in Equation (1). Thus, if the closed-loop system is stable for load disturbances, it will also be stable for set-point changes.

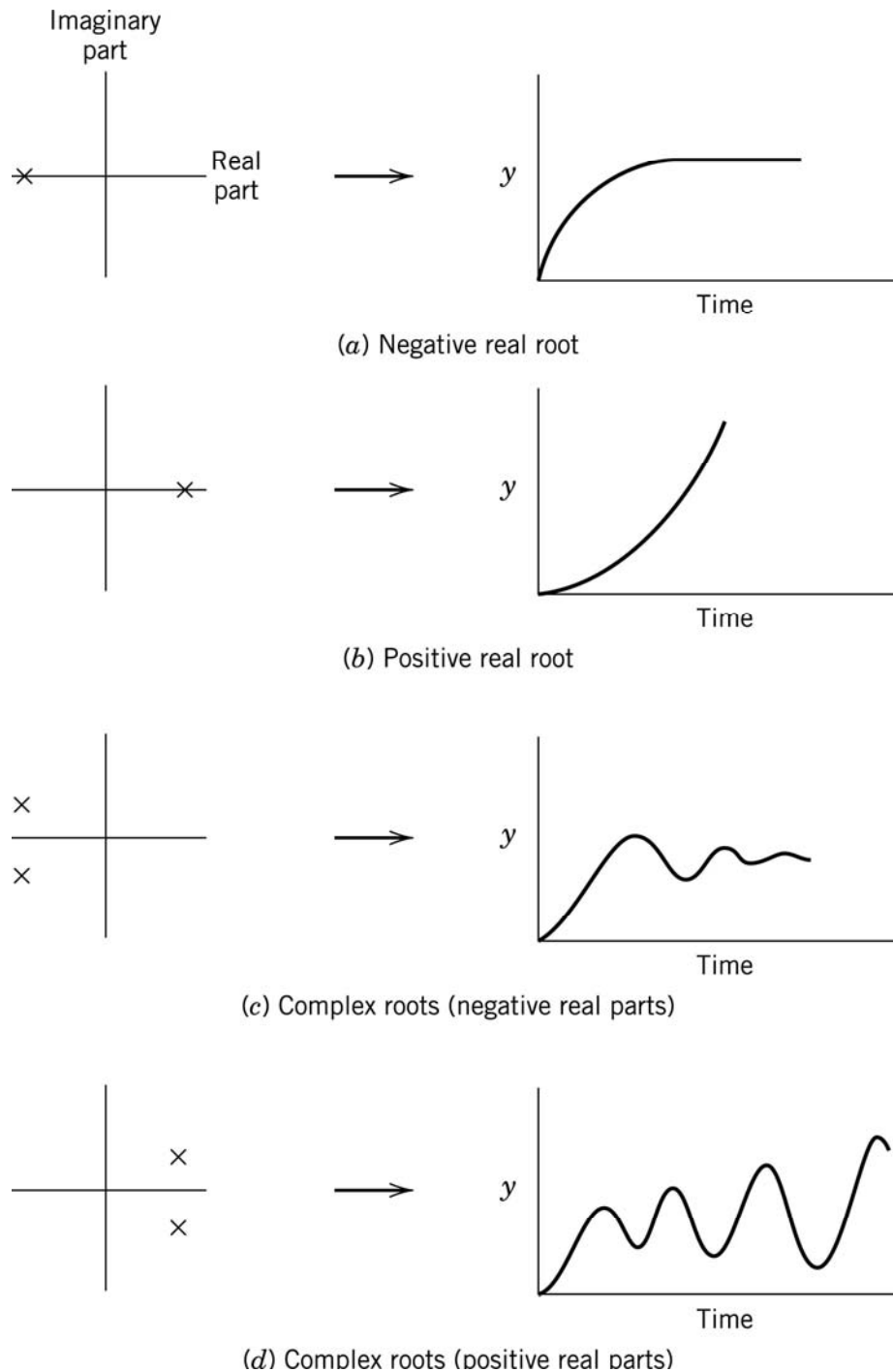


Figure 2 Contributions of characteristic equation roots to closed-loop response.

Example 1

Consider a process with the following transfer functions:

$$G_p = \frac{10}{s-1}, \quad G_f = G_m = 1, \quad G_c = K_c$$

Determine the range of K_c values that result in a stable closed-loop system.

Solution

The corresponding characteristic equation is

$$\begin{aligned} 1 + G_f G_m G_c G_p &= 1 + \frac{10}{s-1} \cdot 1 \cdot K_c \cdot 1 = 0 \\ &= \frac{s-1+10K_c}{s-1} = 0 \end{aligned}$$

which has the root

$$p = 1 - 10K_c$$

The system is stable if $p < 0$ (i.e. $K_c > 1/10$).

Example 2

Consider a process with the following transfer functions:

$$G_p = \frac{1}{s^2 + 2s + 1}, \quad G_f = G_m = 1, \quad G_c = 100 \left(1 + \frac{1}{0.1s} \right)$$

Determine whether the PI controller can stabilize the system.

Solution

The corresponding characteristic equation is

$$1 + G_f G_m G_c G_p = 1 + \frac{1}{s^2 + 2s + 1} \cdot 1 \cdot K_c \left(1 + \frac{10}{s} \right) \cdot 1 = 0$$

The equation above yields

$$s^3 + 2s^2 + 102s + 100 = 0$$

with roots -7.185 , $2.59 + 11.5j$ and $2.59 - 11.5j$. The closed-loop system is unstable because two roots of the characteristic equation have positive real parts.

2. Routh Stability Criterion

The criterion of stability for closed-loop systems does not require calculation of the actual values of the roots of the characteristic polynomial. It only requires that we know if any root is to the right of the imaginary axis. Routh Stability Criterion is an analytical technique for determining whether any roots of the polynomial have positive real parts. This approach can be applied only to systems whose characteristic equations are

polynomials in s . Thus, the Routh Stability Criterion is not *directly* applicable to systems containing time delays, since an $e^{-\theta s}$ term appears in the characteristic equation.

The Routh Stability Criterion is based on a characteristic equation that has the form

$$a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0 = 0 \quad (6)$$

We arbitrarily assume that $a_n > 0$. If $a_n < 0$, we multiply Equation 6 by -1 to generate a new equation that satisfies this condition.

First test. A necessary (but not sufficient) condition for stability is that *all* of the coefficients $(a_n, a_{n-1}, \dots, a_1, a_0)$ in the characteristic equation must be positive. If any coefficient is negative or zero, then at least one root of the characteristic equation lies on the right of, or on, the imaginary axis, and the system is unstable.

Second test. If all of the coefficients are positive, we can construct the following Routh array:

Row1	a_n	a_{n-2}	a_{n-4}	\dots
2	a_{n-1}	a_{n-3}	a_{n-5}	\dots
3	b_1	b_2	a_3	\dots
4	c_1	c_2	c_3	\dots
\vdots				
$n+1$	z_1			

where

$$b_1 = \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}}$$

$$b_2 = \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}}$$

$$\vdots$$

$$c_1 = \frac{b_1 a_{n-3} - a_{n-1} b_2}{b_1}$$

$$c_2 = \frac{b_1 a_{n-5} - a_{n-1} b_3}{b_1}$$

$$\vdots$$

Routh Stability Criterion. A necessary and sufficient condition for all roots of the characteristic equation to have negative real parts is that all of the elements in the left column of the Routh array are positive.

Remarks

The number of sign changes in the elements of the first column is equal to the number of roots to the right of the imaginary axis.

Example 3

Determine the stability of a system that has the characteristic equation

$$s^4 + 5s^3 + 3s^2 + 1 = 0$$

Solution

Since the s term is missing, its coefficient is zero. Thus, the system is unstable (*First test*).

Example 4

Find the values of controller gain K_c that make the following feedback control system stable.

$$G_p = G_L = \frac{1}{5s+1}, \quad G_f = \frac{1}{2s+1}$$

$$G_m = \frac{1}{s+1}, \quad G_c = K_c$$

Solution

The characteristic equation is

$$10s^3 + 17s^2 + 8s + 1 + K_c = 0 \quad (7)$$

All coefficients are positive provided that $1+K_c > 0$ or $K_c > -1$. The Routh array is

$$\begin{array}{cc} 10 & 8 \\ 17 & 1+K_c \\ b_1 & b_2 \\ c_1 & \end{array}$$

where

$$b_1 = \frac{17(8) - 10(1+K_c)}{17} = 7.41 - 0.588K_c$$

$$b_2 = 0$$

$$c_1 = \frac{b_1(1+K_c) - 17(0)}{b_1} = (1+K_c)$$

To have a stable system, each element in the left column of the Routh array must be positive. Element b_1 will be positive if $K_c > 7.41/0.588 = 12.6$. Similarly, c_1 will be positive if $K_c > -1$. Thus, we conclude that the system will be stable if

$$-1 < K_c < 12.6$$

This example illustrates that stability limits for controller parameters can be derived analytically using the Routh array; that is, it is not necessary to compute the roots of the characteristic equation nor specify a numerical value for K_c before performing the stability analysis.

3. Direct substitution method

Direct substitution method is a convenient method for determining the range of controller parameters for which the closed-loop response is stable. The method is based on the fact that the roots of the characteristic equation vary continuously with the loop parameters. Consequently, at the point of instability, at least one of the roots must lie on the imaginary axis of the complex plane as they cross from the left-half plane to the right. *This means that the roots are pure imaginary numbers at the verge of instability.* At this point the loop is said be *marginally stable*. This means that, at this point, the characteristic equation must have a pair of pure imaginary roots at $s_{1,2} = \pm j\omega_u$. The frequency ω_u with which the loop oscillates is the *ultimate frequency*. The controller gain at which this point of marginal instability is reached is called the *ultimate gain*, K_u . At a gain just below the ultimate, the loop oscillates with decaying amplitude, while at a gain just above the ultimate gain, the amplitude of oscillations increases with time. At the point of marginal stability, the amplitude of oscillation remains constant with time (Figure 3).

The ultimate period of oscillation T_u is related to the ultimate frequency, ω_u , in rad/s, by

$$T_u = 2\pi/\omega_u \quad (8)$$

The method of direct substitution consists of substituting $s = j\omega_u$ in the characteristic equation. This results in a complex equation that can be converted into two simultaneous equations

$$\text{Real part} = 0$$

$$\text{Imaginary part} = 0$$

From these we can solve for two unknowns: one is the ultimate frequency ω_u , the other is any of the parameters of the loop, usually the controller gain at the point of marginal stability or ultimate gain. *Generally, the closed-loop response is unstable when the controller gain is greater than the ultimate gain.*

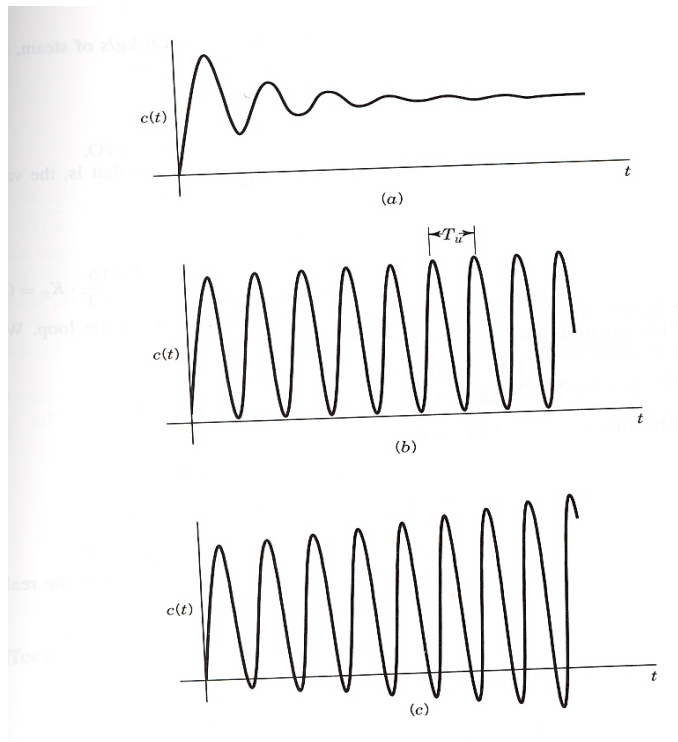


Figure 3 Responses of closed-loop with the controller gain less than (a), equal to (b), and greater than (c) the ultimate gain.

Example 5

Use the direct substitution method to determine K_u for the system described in Example 4.

Solution

Substitute $s = j\omega_u$ and $K_c = K_u$ into Equation (7):

$$-10j\omega^3 - 17\omega^2 + 8j\omega + 1 + K_u = 0$$

or

$$(1 + K_u - 17\omega^2) + j(8\omega - 10\omega^3) = 0$$

Solve both equations

$$1 + K_u - 17\omega^2 = 0$$

$$8\omega - 10\omega^3 = \omega(8 - 10\omega^2) = 0$$

$$\omega^2 = 0.8 \Rightarrow \omega = \pm 0.894$$

$$\Rightarrow K_u = 12.6$$

(9)

Thus, we conclude that $K_c < 12.6$ for stability. Equation (9c) indicates that at the stability limit, $K_c = K_u = 12.6$, a sustained oscillation occurs that has a frequency of $\omega = 0.894$ rad/min., if the time constants have units of minutes. The corresponding period = 7.03 min.

Performance criteria for closed-loop systems

Introduction

The function of a feedback control system is to ensure that the closed-loop system has desirable dynamic and steady state response characteristics. Ideally, we would like the closed-loop system to satisfy the following performance criteria:

1. The closed-loop system must be stable.
2. The effects of disturbances are minimized.
3. Rapid, smooth responses to set point changes are obtained.
4. Offset is eliminated.
5. Excessive control action is avoided.
6. The control system is robust.

In typical control problems, it is not possible to achieve all of these goals since they involve inherent conflicts and trade-offs. *For example*, PID controller settings that minimize the effects of disturbances tend to produce large overshoots for set point changes. *On the other hand*, if the controller is adjusted to provide a rapid, smooth response to a set point change, it usually results in sluggish control for disturbances. Thus, the trade-off is required in selecting controller settings that are satisfactory for both load and set point changes.

Design relations for PID controllers

In this section we consider some well-known controller design relations that are based on a specific model, namely the first-order plus time-delay model, Equation (10).

$$G(s) = \frac{K_p e^{-\theta s}}{\tau_p s + 1} \quad (10)$$

Design relations based on integral error criteria

A controller design relations, which is based on performance index that considers the entire closed-loop response, are developed. Three performance indices will be considered in this course:

Integral of the absolute value of the error (IAE)

$$IAE = \int_0^{\infty} |e(t)| dt \quad (11)$$

where the error signal $e(t)$ is the difference between the set point and the measurement. Notice that for P controller, where offset occurs, the integral given by Equation does not converge. In these cases, one can use a modified integrand, which replaces the error by $y(\infty) - y(t)$, since this term does approach zero as t goes to infinity.

Integral of squared error (ISE)

$$ISE = \int_0^{\infty} [e(t)]^2 dt \quad (12)$$

Integral of time-weighted absolute error

$$ITAE = \int_0^{\infty} t|e(t)|dt \quad (13)$$

A graphical interpretation of the IAE performance index is shown in Figure 4.

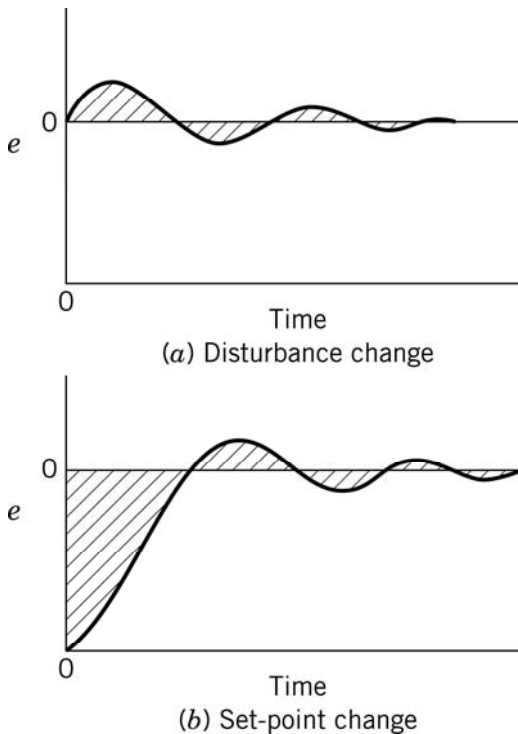


Figure 4 Graphical interpretation of IAE.

The ISE will penalize the response that has large errors, which usually occur at the beginning of a response because the error is squared. The ITAE will penalize a response which has errors that persist for a long time. The IAE will treat all errors in a uniform manner; thus, it allows larger deviation than ISE. In general, ITAE is the preferred integral error criterion since it results in the most conservative controller settings.

Design relations that minimize the ITAE performance index are shown in Table 1. These relations are based on the first-order plus time-delay model, i.e. Equation (10), and the ideal PID controller. Note that the optimal controller settings are different depending on whether step responses to load or set point are considered. For load changes, the load and process transfer functions are assumed identical.

Table 1. Controller design relations based on ITAE performance index and a first-order plus time delay model.

Type of Input	Type of Controller	Mode	A	B
Load	PI	P	0.859	-0.977
		I	0.674	-0.680
Load	PID	P	1.357	-0.947
		I	0.842	-0.738
		D	0.381	0.995
Set point	PI	P	0.586	-0.916
		I	1.03 ^b	-0.165 ^b
Set point	PID	P	0.965	-0.85
		I	0.796 ^b	-0.1465 ^b
		D	0.308	0.929

^aDesign relation: $Y = A(\theta/\tau)^B$ where $Y = KK_c$ for the proportional mode, τ/τ_I for the integral mode, and τ_D/τ for the derivative mode.

^bFor set-point changes, the design relation for the integral mode is $\tau/\tau_I = A + B(\theta/\tau)$. [8]

Example 6

For the process model,

$$G(s) = \frac{4e^{-3.5s}}{7s + 1}$$

Compare PI and PID controller settings based on ITAE tuning relations for both load and set point changes.

Solution

$$KK_c = 0.859 \cdot \left(\frac{3.5}{7}\right)^{-0.977}$$

$$\Rightarrow K_c = 0.423$$

$$\tau/\tau_I = 0.647 \cdot \left(\frac{3.5}{7}\right)^{-0.68}$$

$$\Rightarrow \tau_I = 6.48$$

The ITAE settings are shown below:

Controller/Design method	K_c	τ_I	τ_D
PI / load	0.423	6.48	-
PI / set point	0.276	7.39	-
PID / load	0.654	4.98	1.34
PID / set point	0.435	9.69	1.13

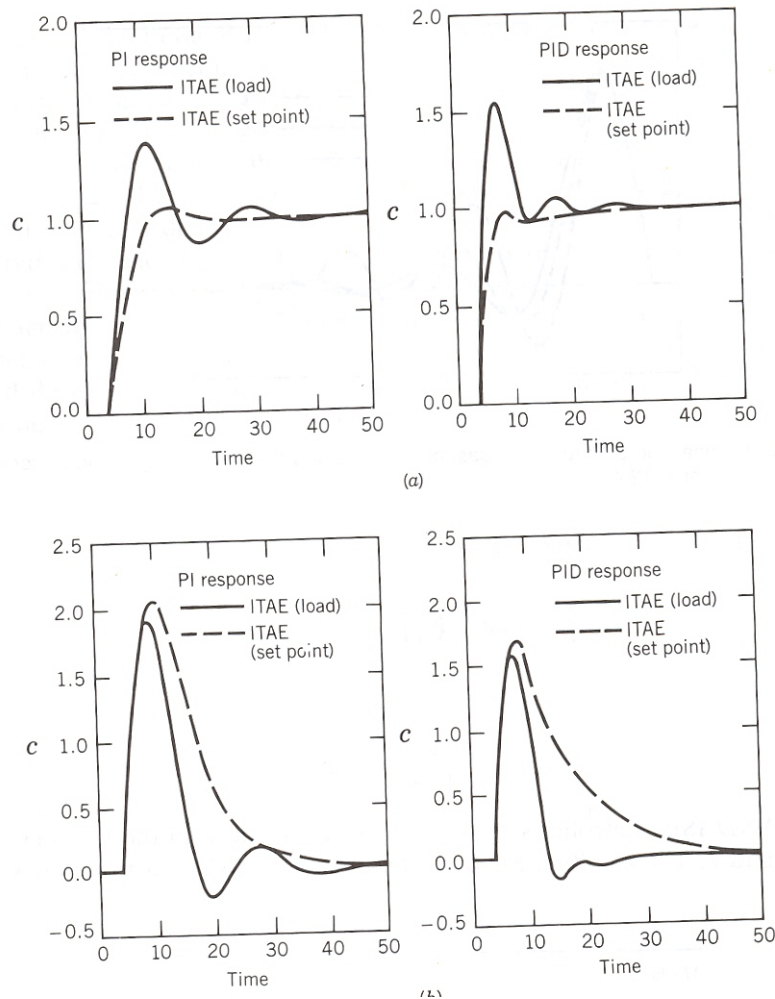


Figure 5 Comparison of controllers design using ITAE criteria for (a) set point and (b) load changes.

Figure 5 compares the ITAE controllers. Design for load changes results in large overshoots for set-point changes, while set-point design produces sluggish responses to load disturbances. *If set-point changes and load disturbances are both likely to occur, then a compromise in the controller settings should be employed.*

This example has demonstrated that, in general, integral error criteria for set point changes results in more conservative controller settings than for load changes.

Controller tuning

After controller installation, the controller settings must usually be adjusted until control performance is considered satisfactory. This activity is referred to as *controller tuning* or *field tuning* of the controller.

In order to save time and effort, it is desirable to have preliminary estimates of satisfactory controller settings. A good first guess may be available from experience with similar control loops. Alternatively, if a process model is available, integral error methods can be employed to calculate controller settings. Field tuning, may still required to fine tune the controller, especially if the available information is incomplete or not very accurate.

1. Continuous cycling method

This approach has also been referred to as *loop tuning* or the *ultimate gain method*. This method is described as a *closed-loop method* because the controller remains in the loop as an active controller in automatic mode. A typical *experimental* approach for PID controllers can be summarized as follows:

1. After the process reaches steady state, remove the integral and derivative modes of the controller, leaving only proportional control. On some PID controllers, this requires that the integral time (τ_I) be set to its maximum value and the derivative time (τ_D) to its minimum value.
2. Select a value of proportional gain (K_c), disturb the system, and observe the transient response. If the response decays, select a higher value of K_c and again observe the response of the system. Continue increasing the gain in small steps until the response first exhibits a sustained oscillation, Figure 6. The value of gain and the period of oscillation that correspond to the sustained oscillation are the ultimate gain (K_{cu}) and the ultimate period (P_u). In performing the experimental test, it is important that the controller output does not saturate. If saturation does occur, then a sustained oscillation can result even though $K_c > K_{cu}$. Typical results are shown in Figure 7.

Definition 1. The ultimate gain K_{cu} is the largest value of the controller gain K_c that results in closed-loop stability when proportional only controller is used.

Definition 2. The ultimate period P_u is defined as the period of sustained cycling that would occur if a proportional controller with gain K_{cu} were used.

3. From the values of K_{cu} and P_u found in the previous step, use the Ziegler-Nichols rules given in Table 19.1 to determine controller settings (K_c , τ_I , τ_D). These tuning relations were empirically developed to provide a $1/4$ decay ratio, Figure 8.

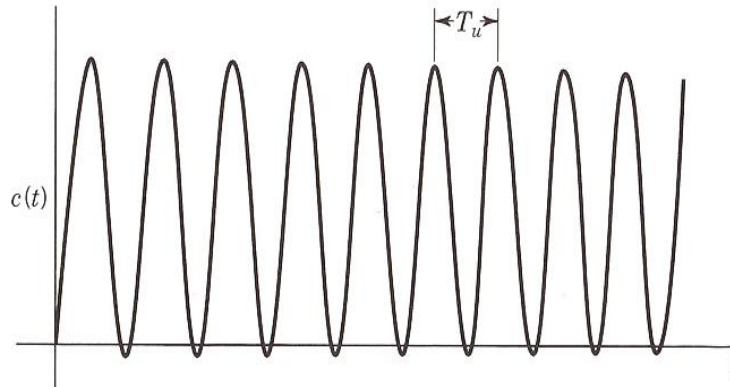


Figure 6 Response of the loop with the controller gain set equal to the ultimate gain K_{cu} . T_u is the ultimate period.

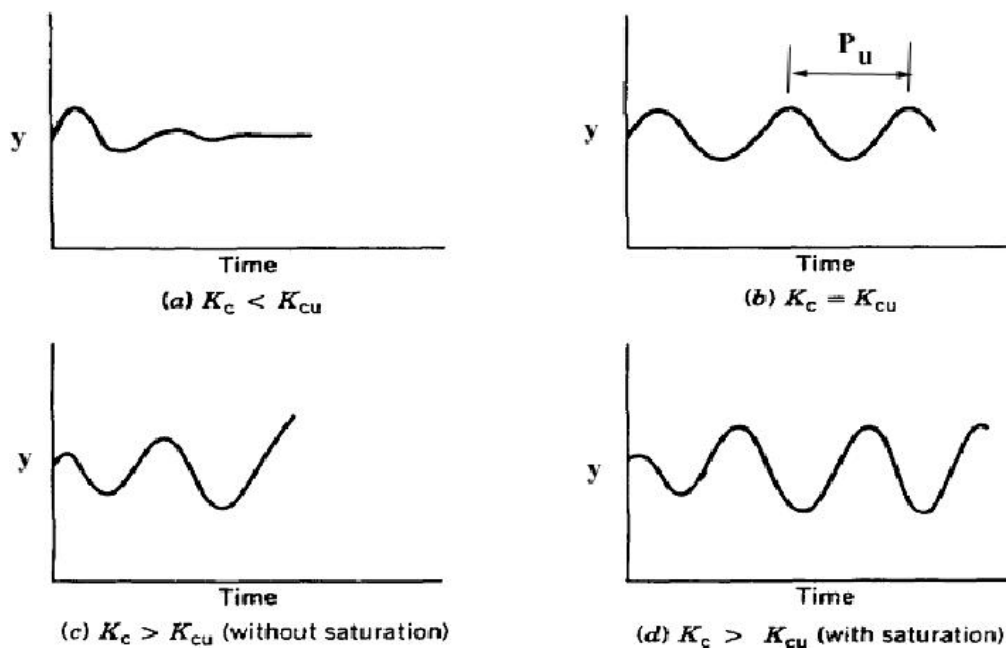


Figure 7 Determination of K_{cu} using Continuous Cycling Method.

Table 2. Ziegler-Nichols controller settings based on the Continuous Cycling Method.

controller	K_c	τ_I	τ_D
P	$0.5 K_{CU}$	-	-
PI	$0.45 K_{CU}$	$P_U/1.2$	-
PID	$0.6 K_{CU}$	$P_U/2$	$P_U/8$

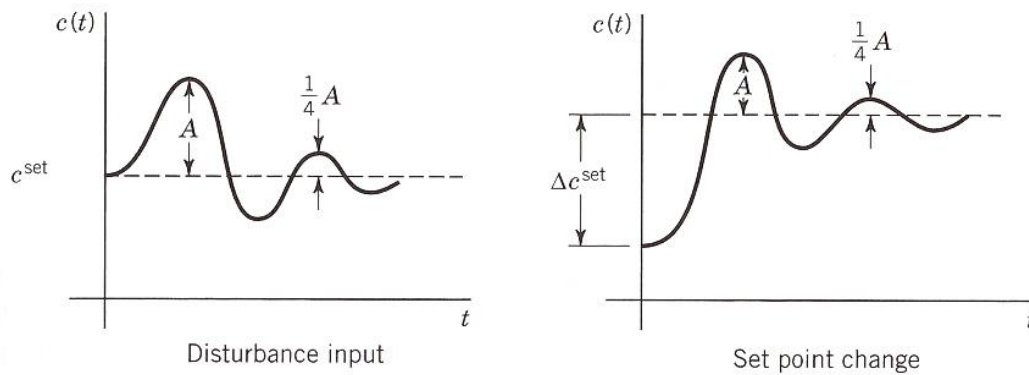


Figure 8 Quarter decay ratio response to disturbance input and to change in set point.

Remarks

1. K_{cu} and P_u can be determined by the direct substitution method if the transfer functions of all of the components of the loop are known qualitatively.
2. The quarter decay ratio response is very desirable for disturbance inputs because it prevents a large initial deviation from the set point without being too oscillatory. However, it is not as desirable for step changes in set point, because it causes a 50 % overshoot. This is because the maximum deviation from the new set point in each direction is one-half the preceding maximum deviation in the opposite direction, Figure 8. This difficulty can easily be corrected by reducing the proportional gain from the value predicted by the formulas of Table 2. In fact, the decay ratio is a direct function of the controller gain, and can be adjusted at any time by simply changing the gain. In other words, if for a given process the quarter decay ratio response is too oscillatory, a reduction of the gain will smooth out the response.
3. Shortcomings:
 - 3.1. It may be objectionable because the process is pushed to stability limit. Consequently, if external disturbances or a change in the process occurs during tuning, unstable operation or hazardous situation could occur.
 - 3.2. This tuning procedure is not applicable to processes that are open-loop unstable because such processes are unstable at both high and low values of K_c , but stable for intermediate range of values.
 - 3.3. Some processes do not have an ultimate gain; for example, first-order and second-order processes *without time delay*.
 - 3.4. The set of tuning parameters necessary to obtain the quarter decay ratio response is not unique, except for the case of P controller.

- 3.5. Controller settings in Table 3 should be regarded as first estimates. Subsequent fine tuning via trial and error is often required.

2. Process reaction curve method

This is an open-loop method; it is used as an alternative to Z-N method. Figure 9 shows a typical control loop in which the control action is removed and the loop opened for the purpose of introduction a step change (M/s) to the valve. The step response is recorded at the output of the measuring element. The step change to the valve is conveniently provided by the output from the controller, which is in manual mode. The response of the system (including the valve, process, and measuring element) is called the *process reaction curve*. Two different types of process reaction curves are shown in Figure 10 for a step change occurring at $t = 0$.

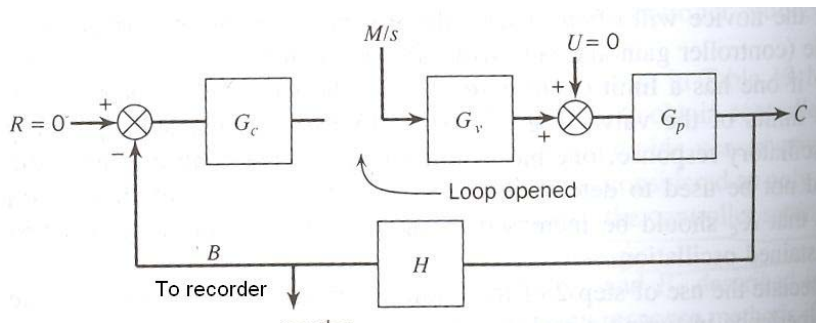


Figure 9 Block diagram of a control loop for measurement of the process reaction curve.

This method is summarized in the following steps:

1. After the process reaches steady state at the normal level of operation, switch the controller to manual.
2. With the controller in manual, introduce a small step change in the controller output that goes to the valve and record the transient, which is the process reaction curve (Figure 10).
3. Draw a straight line tangent to the curve at the point of inflection, as shown in Figure 10. The intersection of the tangent line with the time axis is the apparent transport lag (θ); the apparent first-order time constant (τ) is obtained from:

$$\tau = B_u / S \quad (14)$$

where B_u is the steady state “ultimate” value of B and S is the slope of the tangent line. The steady state gain that relates B to M in Figure 9 is given by:

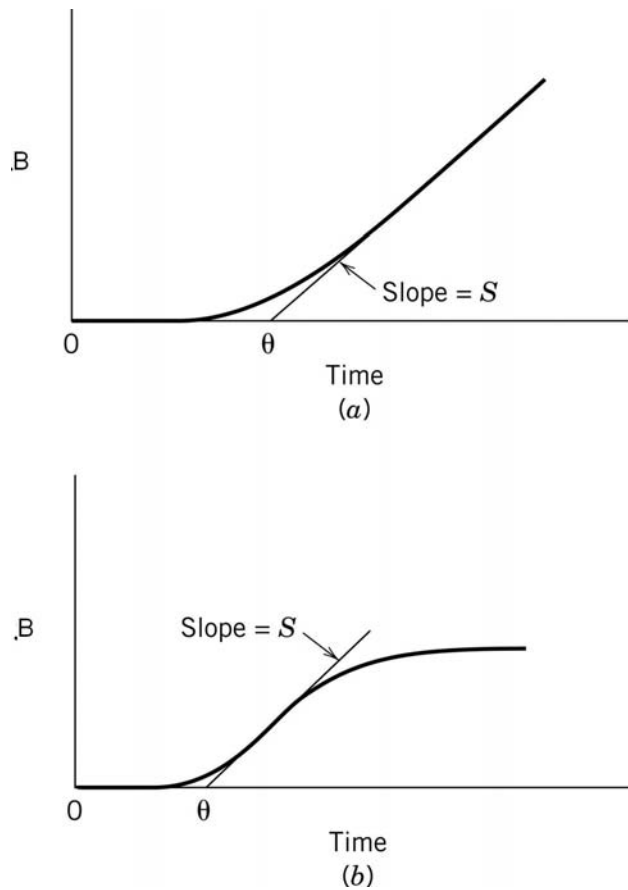


Figure 10 Typical process reaction curves: (a) non-self regulating process, (b) self-regulating process.

$$K_p = B_u / M \quad (15)$$

Note that if the process reaction curve has the typical sigmoidal shape shown in Case b of Figure 10, the following model usually provides a satisfactory fit:

$$\frac{B(s)}{M(s)} = G_v G_p G_m = \frac{K_p e^{-\theta s}}{\tau_p s + 1} \quad (16)$$

4. Using the values of K_p , τ_p and θ from step 3, the controller settings are found from the relations given in Table 3.

Notice that the settings given in this table were developed to provide closed-loop responses with a decay ratio of $1/4$.

Table 3 Cohen and Coon Controller design relations.

Controller	Settings	Cohen-Coon
P	K_c	$\frac{1}{K} \frac{\tau}{\theta} [1 + \theta/3\tau]$
PI	K_c	$\frac{1}{K} \frac{\tau}{\theta} [0.9 + \theta/12\tau]$
	τ_I	$\frac{\theta[30 + 3(\theta/\tau)]}{9 + 20(\theta/\tau)}$
PID	K_c	$\frac{1}{K} \frac{\tau}{\theta} \left[\frac{16\tau + 3\theta}{12\tau} \right]$
	τ_I	$\frac{\theta[32 + 6(\theta/\tau)]}{13 + 8(\theta/\tau)}$
	τ_D	$\frac{4\theta}{11 + 2(\theta/\tau)}$

Remarks

The advantages of PRC method:

1. Only a single experimental test is necessary.
2. It does not require trial and error.
3. The controller settings are easily calculated.

However, it has several disadvantages:

1. The experimental test is performed under open-loop conditions. Thus, if a significant load change occurs during the test, no corrective action is taken and the test results may be significantly distorted.
2. It may be difficult to determine the slope at the inflection point accurately, especially if the measurement is noisy and a small recorder chart is used.
3. The recommended settings in Tables 3 tend to result in oscillatory responses since they were developed to provide a 1/4 decay ratio.
4. The method is not recommended for processes that have oscillatory open-loop responses since the process model in Equation (16) will be quite inaccurate.

Example 7