Controller Tuning

The sort of closed loop response that can be obtained from a PID controller with appropriate settings was seen in Chapter 23. This gives rise to two issues. First, what is the best sort of response? This is normally characterised in terms of the response to a step input, the ideal being a fast response with no overshoot or offset. This is not physically possible because of a plant's dynamics. There is therefore a need to compromise between the speed of response and the amount of overshoot. For certain critical applications an overdamped response with no overshoot is essential, the penalty being a slow response. However, for most purposes, an underdamped response with a decay ratio of 1/4 and no offset is good enough, as depicted in Figure 24.1. Often referred to as the optimal response, this has no mathematical justification: it is just accepted good practice.

Second, what are the appropriate settings and how do you find them? The process of finding the optimum settings is generally referred to as loop tuning. They may be found by trial and error. However, given that each of the settings for K_C, T_R and T_D can typically be varied from 0.01 to 100, and that the various lags and delays associated with process plant are often large, this could be rather tedious. The settings may be predicted theoretically. However, this requires a model of the system that is reasonably accurate, and any realistic model usually needs simulation to provide the settings. The cost and time involved in modelling and simulation can seldom be justified. There is therefore a need for a practical approach.

This chapter outlines two practical methodologies; one is empirically based and the other theoretical, for establishing the so called optimum settings. An informed account of these methods is given by Coughanowr (1991) and a comprehensive coverage of both these and many other methods of tuning is given by Astrom (1995). However, first, an insight into the nature of stability is required.

24.1 Stability

Stability is a fundamental consideration in the design of control systems. Many open loop systems are stable. For example, the level control system of Figure 22.1 is open loop stable. Suppose the controller is in its manual mode and the valve opening is fixed at its normal value. Following a step increase in inlet flow, the level will rise until a new equilibrium is established at which there is sufficient head for the flow out to balance the flow in. Such a system is said to be self regulating and is relatively easy to control. However, with inappropriate controller settings, a system which is open loop stable can become closed loop unstable.

Some systems are open loop unstable. The classic example of an item of plant that is inherently unstable is the exothermic reactor. A slight increase in temperature will make the reaction go faster.

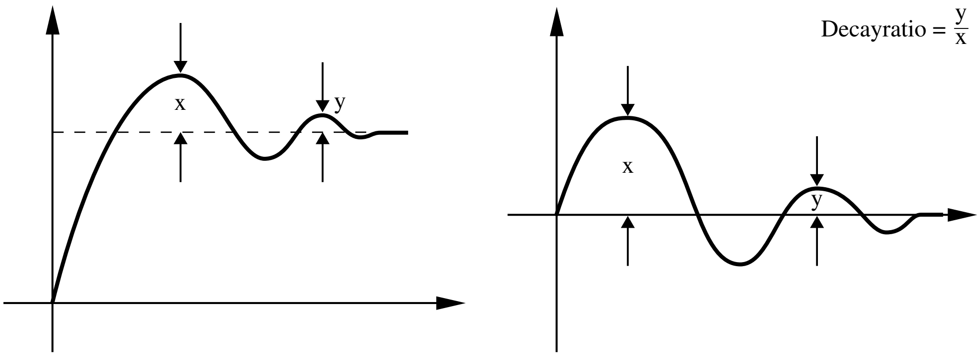

Fig. 24.1 The optimal response

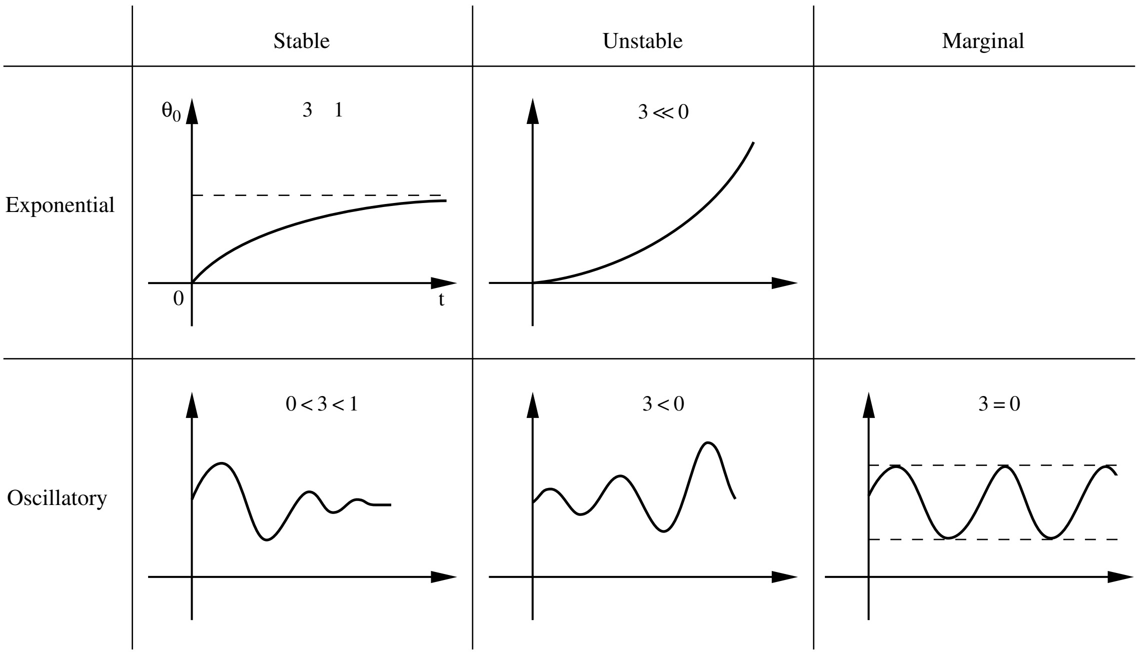

Fig. 24.2 Categorisation of signals on the basis of stability

This produces more heat which increases the temperature further, and so on. However, by applying feedback, such an open loop unstable system can be made closed loop stable.

From a control engineering point of view, stability manifests itself in the form of a system's signals. These may be categorised, depending on whether the system is stable or unstable, as being exponential or oscillatory as depicted in Figure 24.2. An important point to appreciate is that it is the system that is stable or otherwise, not its signals. Also, stability is a function of the system as a whole, not just of some parts of it. It follows that all the signals of a system must be of the same form. It is not possible, for example, for the controller output of the level control system of Figure 22.1 to be stable but for the level to be unstable. Similarly, it is not possible for the response of the level to be exponential whilst that of the valve opening is oscillatory.

As will be seen in Chapters 71 and 72, stability may be characterised by a so-called damping factor ζ in relation to a second order system. Table 24.1 relates the categories of stability to values of the damping factor.

Table 24.1 Stability as a function of damping factor

Factor	Damping	Nature of stability
$\zeta > 1.0$	Overdamped	Stable exponential
$\zeta = 1.0$	Critically damped	Limiting case
$0 < \zeta < 1$	Under-damped	Stable oscillatory
$\zeta = 0$	Undamped	Marginally stable
$\zeta < 0$	Self excited	Unstable oscillatory
$\zeta \ll 0$	Over excited	Unstable exponential

24.2 Marginal Stability

The case of marginal stability is of particular importance for design purposes. The most common design philosophy is to establish the conditions under which a system is marginally stable. An appropriate safety factor is then specified which ensures that the system's operation is stable for all foreseeable circumstances.

Marginal stability corresponds to the situation when all the signals in a system are sinusoidal with constant amplitude, *i.e.* the oscillations are neither growing nor decaying. Consider the feedback system of Figure 24.3 in which all the elements, other than the controller, have been lumped in with the process. The controller has P action only.

Suppose that the error signal is a sine wave of constant amplitude:

$$e = A.\sin(\omega_C t)$$

Now suppose that the frequency w_C, known as the critical frequency, is such that the effect of the process on the error is to produce a measured value whose phase is shifted by 180°. Also suppose that

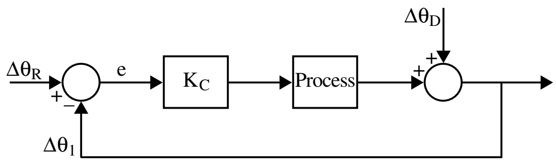

Fig. 24.3 Feedback system with P controller and lumped process

the controller gain has a value K_{CM} such that the controller and the process have no net effect on the amplitude of the signal. Thus, assuming that the signals are in deviation form:

$$\Delta\theta_1 = A.\sin(\omega_C t - 180)$$

So $\Delta\theta_1$ is the mirror image of e as shown in Figure 24.4. But

$$e = \Delta\theta_R - \Delta\theta_1$$

If the set point is constant, *i.e.* $\Delta\theta_R = 0$, then

$$e = -\Delta\theta_1 = -A.\sin(\omega_C t - 180) = A.\sin(\omega_C t)$$

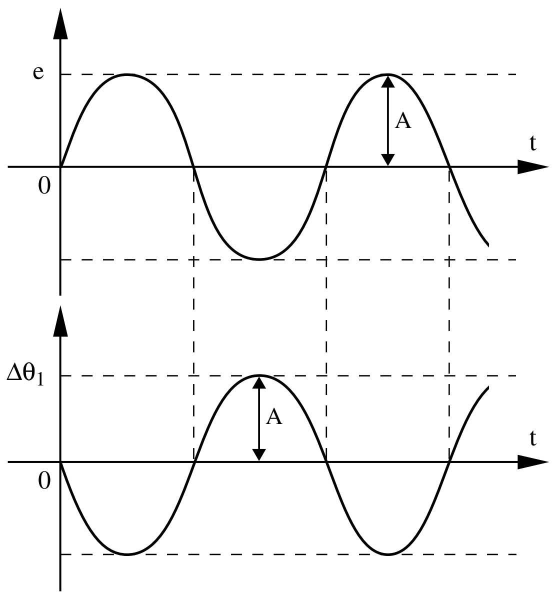

Fig. 24.4 Sinusoidal error and measured value signals

The comparator introduces a further 180° phase shift which converts the measured value back into the error. This is a self sustaining sine wave of constant amplitude. Thus the twin criteria necessary for marginal stability are

$$\text{Open loop gain} = 1.0$$
$$\text{Open loop phase shift} = -180^\circ$$

These are known as the Bode stability criteria and will be considered in more detail in Chapter 73.

24.3 Continuous Cycling Method

The continuous cycling method is a simple practical method that tunes a control loop as installed, rather than as designed. The method is carried out at the controller, in its automatic mode, and takes advantage of the faceplate for observing the signals. The procedure for carrying out the method is shown in flow chart form in Figure 24.5.

In essence the process consists of changing the controller's gain K_C incrementally and observing the loop's response to small step changes in set point. If the oscillations decay then K_C is too low and if the oscillations grow then K_C is too high. The value of the marginal gain K_{CM} that forces the loop into self sustained oscillation of constant amplitude is noted. So too is the period P_U of that oscillation, sometimes referred to as the ultimate period. The critical frequency is given by

$$\omega_C = \frac{2\pi}{P_U}$$

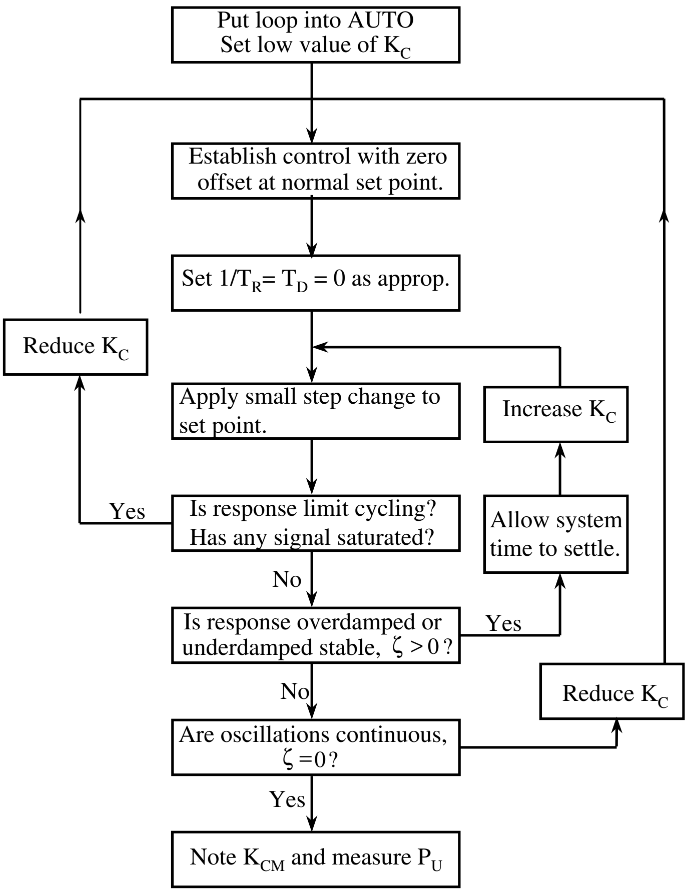

Fig. 24.5 Procedure for the continuous cycling method

Knowing the values of K_{CM} and ω_C, the optimum settings can be determined from the Zeigler and Nichols formulae given in Table 24.2.

Table 24.2 The Zeigler and Nichols formulae

	K_C	T_R	T_D
P	$K_{CM}/2.0$	–	–
PI	$K_{CM}/2.2$	$P_U/1.2$	–
PID	$K_{CM}/1.7$	$P_U/2.0$	$P_U/8.0$

The amplitude of the oscillations under conditions of marginal stability is a function of the system and cannot be controlled. Therefore, before carrying out the continuous cycling method on a plant, it should be established whether it is acceptable to do so. Whilst it may well be acceptable to force the plant into oscillation during commissioning or periods of shut-down, approval to do so is unlikely to be forthcoming during production!

Care should be taken to protect the system from external disturbances whilst the tests are being carried out so as not to distort the results. The most common source of disturbance is due to changes in the supply pressure of utilities such as steam and cooling water. Also, the control loop being tuned may interact with other loops as, for example, in cascade control. In such cases it is usually necessary to put the other loops into their manual mode to prevent them from trying to compensate for the oscillations in the loop being tuned.

Whether the controller has P, P&I or P, I&D actions, the continuous cycling method must be carried out with the controller set for P action only. The I and D actions are switched off by setting T_R = max and T_D = min respectively. Once the procedure has been completed, the optimum settings for the P, I and D actions can then be set as appropriate.

It is important to establish oscillations of constant amplitude that are sinusoidal. These can easily be confused with limit cycles which are of constant amplitude but non-sinusoidal. Limit cycles occur when a system is in oscillation and at least one signal is saturated. The most likely signal to saturate, and the easiest to observe, is the controller output. This may saturate at either the top and/or the bottom of its range, resulting in the valve opening being more square than sinusoidal in form. The process will filter out this squareness resulting in a measured value that may appear to be sinusoidal. Reducing the controller gain should stop the saturation and prevent limit cycles from occurring.

The optimum settings for a control loop will vary across the range of its measured value if there are any non-linearities present. It follows that a loop must be tuned for its normal operating conditions. The oscillations should therefore be established about the normal value of each signal. In particular, the normal set point should be used, and the step changes applied in alternate directions about it to ensure that the system stays close to normal.

The nature of the Zeigler and Nichols formulae needs some explanation. First published in 1941, they are used extensively in industry and have stood the test of time. The formulae are empirical, although they do have a rational theoretical explanation. They predict settings that are optimum on the basis of a decay ratio of 1/4. However, because the formulae are empirical, they do not predict the optimum settings precisely, and further tuning of a trial and error nature may be required. This might not seem to be very satisfactory but, noting that each of the settings typically has a range of 0.01 to 100, *i.e.* a rangeability of some 10^4, a method that predicts settings to within even 50% of the optimum as a first estimate is extremely useful. In practice the predictions are often to within 10% of the optimum.

24.4 Reaction Curve Method

Whereas forcing a plant into oscillation with no control over amplitude may be unacceptable, introducing a small step change of known size, which is the basis of the reaction curve method of predicting optimum settings, is another matter altogether.

This is an open loop method of controller tuning and is depicted in Figure 24.6. With the con-

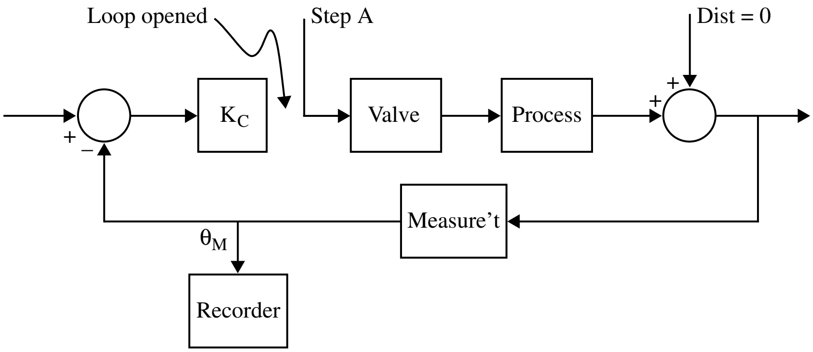

Fig. 24.6 The open loop reaction curve method

troller in its manual mode, the output is adjusted to its normal value and the system allowed to reach equilibrium. Then a small step change of known magnitude A is applied to the controller output. The system is allowed to respond and the measured value recorded for an appropriate period, as shown in deviation form in Figure 24.7.

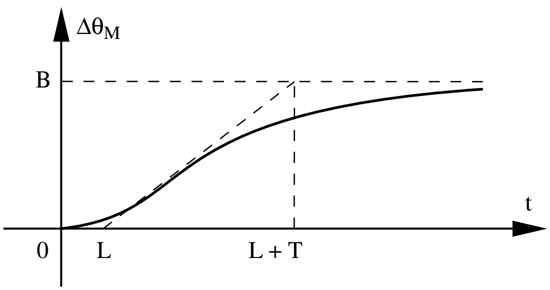

Fig. 24.7 Characteristic "S-shaped" reaction curve

The reaction curve shown, often referred to as being "S shaped" for some unknown reason, is characteristic of most process control systems. Indeed, as will be seen in Chapter 72 in relation to higher order systems, many systems can be approximated by a combination of a steady state gain K, a first-order system with a time constant of T min and a time delay of L min. The values of K, T and L can be estimated from the reaction curve.

First, the steady state asymptote B of θ_M is established. Since B is simply the steady state effect of the open loop elements operating on the step A, the gain can be obtained from the ratio K=B/A. Second, by drawing a tangent to the reaction curve at the point of inflexion, and finding its intersection with the asymptote and the time axis as shown, the values of T and L are found. Knowing the values of K, T and L, the optimum settings can be determined from the Cohen and Coon formulae given in Table 24.3.

These formulae are theoretically derived on the assumption that the plant consists solely of a gain K, time constant T and time delay L. They predict settings that are optimum on the basis of a decay ratio of 1/4.

There are a number of important precautions. Care should be taken to protect the system from external disturbances whilst the tests are being carried out so as not to distort the results. The reaction curve method is much more susceptible to distortion by disturbances than the continuous cycling method. The step input A applied should be small enough for the response to stay within the bounds of linearity. The response of the recording system must be fast enough not to distort the reaction curve. It is normal to apply the step change at the controller output. However, it may be applied anywhere in the loop, provided that the reaction curve is the open loop response of all the elements of the loop. If the controller is included then it must have no effect, *i.e.* set $K_C = 1$, $T_R = \max$ and $T_D = \min$.

There are two major problems in using the reaction curve method. First, it is often difficult to insulate the plant from disturbances long enough to obtain a true reaction curve. And second, given

Table 24.3 The Cohen and Coon formulae

	K_C	T_R	T_D
P	$\frac{1}{K}\frac{T}{L}\left(1+\frac{L}{3T}\right)$		
PI	$\frac{1}{K}\frac{T}{L}\left(\frac{9}{10}+\frac{L}{12T}\right)$	$L\left(\frac{30+3L/T}{9+20L/T}\right)$	
PD	$\frac{1}{K}\frac{T}{L}\left(\frac{5}{4}+\frac{L}{6T}\right)$		$L\left(\frac{6-2L/T}{22+3L/T}\right)$
PID	$\frac{1}{K}\frac{T}{L}\left(\frac{4}{3}+\frac{L}{4T}\right)$	$L\left(\frac{32+6L/T}{13+8L/T}\right)$	$L\left(\frac{4}{11+2L/T}\right)$

the constraints on linearity, it is difficult to obtain a large enough change in output to predict with confidence the value of the asymptote B and the position of the point of inflexion.