Contents

Chapter 5

Controller design

5.1 The Design of Feedback Control Systems

..from (Dorf and Bishop, 2008)

The performance of a feedback control system is of primary importance. A suitable control system is stable, and it results in an acceptable response to input commands, is less sensitive to system parameter changes, results in a minimum steady-state error for input commands and is able to reduce the effect of undesirable disturbances.

A feedback control system that provides an optimum performance without any necessary adjustments is rare indeed. Usually we find it necessary to compromise among many conflicting and demanding specifications and adjust the system parameters to provide a suitable and acceptable performance when is not possible to obtain all the desirable optimum specifications.

The design of a control system is concerned with the arrangement, or the plan, of the system structure and the selection of suitable components and parameters., (Dorf and Bishop, 2008)

In redesigning a control system to alter the system response, an additional component is inserted within the structure of the feedback system. It is this additional component or device that equalizers or compensates for the performance deficiency. The compensating device may be an electric, a mechanical, a hydraulic, a pneumatic or other type of device or network and is often called a *compensator*. Commonly an electric circuit serves as a compensator in many control systems.

5.1.1 Approaches to System Design

The performance of a control system can be described in terms of time-domain performance measures or the frequency-domain performance measures. The performance of a system can be specified by requiring a certain peak time, T_p, a maximum overshoot and settling time for a step input. It is usually necessary to specify the maximum allowable steady-state error for several test signal inputs and disturbance inputs. These performance specifications can be defined in terms of the desirable location of the poles and zeros of the closed-loop transfer function.

As we found in a previous chapter, the locus of the roots of the closed-loop system can be readily obtained for the variation of one system parameter. When the locus of the roots does not result in a suitable root configuration, we must add a compensating network transfer function so that the resultant root locus results in the desired closed-loop configuration.

Alternatively, we can describe the performance of a feedback control system in terms of frequency performance measures. Then, a system can be described in terms of the peak of the closed-loop frequency response, $M_{p\omega}$, the resonant frequency, ω_r, the bandwidth, and the phase margin of the system. We can add a suitable compensation network, if necessary, in order to satisfy the system specifications. Thus, in this case, we use compensation networks to alter and reshape the system characteristics represented on the Bode diagram.

In the following sections we will assume that the process has been improved as much as possible and the transfer function representing the process is unalterable.

If a compensator is needed to meet the performance specifications, the designer must realize a physical device that has the prescribed transfer function of the compensator.

Numerous physical devices have been used for such purposes. In fact, many useful ideas for physically constructing compensators may be found in the literature.

Among many kinds of compensators, widely employed compensators are the lead compensators, lag compensators and lead-lag compensators. In this chapter we shall limit the discussion mostly to these types. Lead, lag and lead-lag compensators may be electronic devices (such as circuits using operational amplifiers) or RC networks (electrical, mechanical, hydraulic, pneumatic or combinations) and amplifiers.

In the trial-and-error approach to system design, we set up a mathematical model of the control system and adjust the parameters of a compensator. The most time-consuming part of such work is the checking of the performance specifications by analysis with each adjustment of the parameters. The designer should use a digital computer to avoid much of the numerical problems necessary for this checking.

Once a mathematical model has been obtained, the designer must construct a prototype and test the open-loop system. If the absolute stability of the closed-loop is assured, the performance of the closed-loop system will be tested. Because of neglected loading effects among the components, nonlinearities, and so on, which were not taken into consideration in the original design work, the actual performance of the prototype system will probably differ from the theoretical predictions. By trial and error, the designer must make changes in the prototype until the system meets the specifications, is reliable and economical.

5.1.2 Root-locus approach to control system design

The root-locus method is a graphical method for determining the locations of all closed-loop poles from knowledge of all locations of the open-loop poles and zeros as some parameter (usually the gain) is varied from zero to infinity.

The method yields a clear indication of the effects of parameter adjustment. An advantage of the root-locus method is that we find that it is possible to obtain information on the transient response as well as on the frequency response from the pole-zero configuration of the system in the s-plane.

In practice, the root-locus plot of a system may indicate that the desired performance cannot be achieved just by the adjustment of gain. In fact, in some cases, the system may not be stable for all values of gain. Then, it is necessary to reshape the root loci to meet the performance specifications.

In designing a control system, if other than a gain adjustment is required, we must modify the original root-locus by inserting a suitable compensator. Once the effects on the root locus of the addition of poles and/or zeros are fully understood, we can readily determine the locations of the pole(s) and zeros(s) of the compensator that will reshape the root locus as desired.

In essence, in the design via the root-locus method the root loci of the system are reshaped through the use of a compensator so that a pair of dominant closed-loop poles can be placed at the desired location. (Often the damping ratio and undamped natural frequency of a pair of dominant closed-loop poles are specified).

5.1.2.1 Effects of addition of poles

The addition of a pole to the open-loop transfer function has the effect of pulling the root-locus to the right, tending to lower the system's relative stability and to slow down the settling of the response.

The addition of integral control adds a pole at the origin, making the system less stable.

Figure 5.1 shows examples of root loci illustrating effects of the addition of a pole to a single-pole system, and the addition of two poles to a single-pole system.

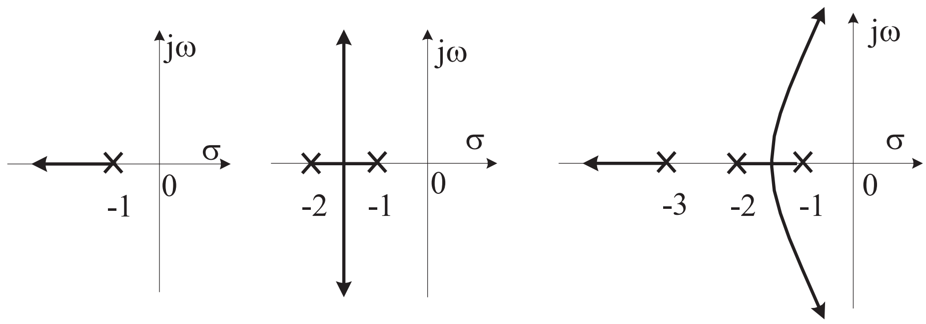

Figure 5.1: Effects of addition of poles

Obs. The settling time of the dominant complex poles is 4/(real part of pole). The closer the poles are to the imaginary axis, the larger the settling time.

5.1.2.2 Effects of addition of zeros

The addition of a zero to the open-loop transfer function has the effect of pulling the root-locus to the left, tending to make the system more stable and to speed up the settling of the response. Physically, the addition of a zero in the feed-forward transfer function means the addition of derivative control to the system. The effect of such control is to introduce a degree of anticipation into the system and speed up the transient response.

Figure 5.2 shows the root loci for a system that is stable for all small gains but unstable for large gain, and the root-locus plots for the system when a zero is added in the open-loop transfer function. Notice that when a zero is added to the system in figure 5.2 it becomes stable for all values of gain.

5.1.3 Cascade compensation networks

The compensation network $G_c(s)$ is cascaded with the unalterable process $G(s)$ in order to provide a suitable loop transfer function $G_c(s)G(s)H(s)$ as shown in figure 5.3.

The compensator $G_c(s)$ can be chosen to alter the shape of the root locus or the frequency response. The network may be chosen to have a transfer function:

$$G_c(s) = \frac{k \prod_{i=1}^{M}(s+z_i)}{\prod_{j=1}^{N}(s+p_j)}$$

Then, the problem reduces to the selection of the poles and zeros of the compensator. To illustrate the properties of the compensation network we will consider a first-order compensator. The compensation approach developed on the basis of a first-order compensator can be extended to higher-order compensators, for example by cascading several first-order compensators. A compensator $G_c(s)$ is used with a plant $G(s)$ so that the overall loop gain can be set to satisfy the steady-state error requirement and then $G_c(s)$ is used to adjust the system dynamics favorably without affecting the steady-state error.

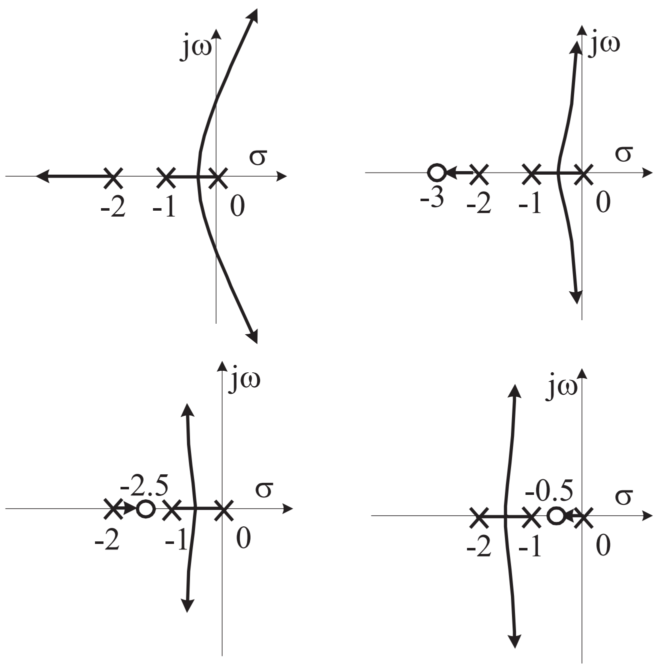

Figure 5.2: Effects of addition of zeros

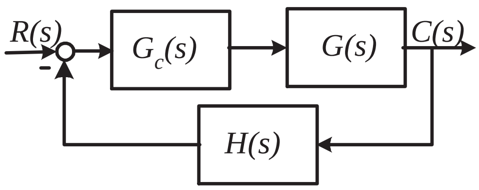

Figure 5.3: Closed-loop control system

Consider the first-order compensator with the transfer function:

$$G_c(s) = \frac{k(s+z)}{s+p}$$

When $|z| < |p|$ the network is called a **phase-lead network** and has a pole-zero configuration as shown in Figure 5.4.

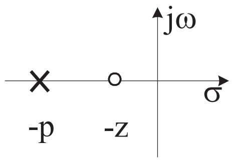

Figure 5.4: Pole-zero location of a phase-lead compensator

The frequency response is:

$$G_c(j\omega) = \frac{k(j\omega+z)}{j\omega+p}$$

and the Bode diagram of a phase-lead compensator is shown in figure 5.5.

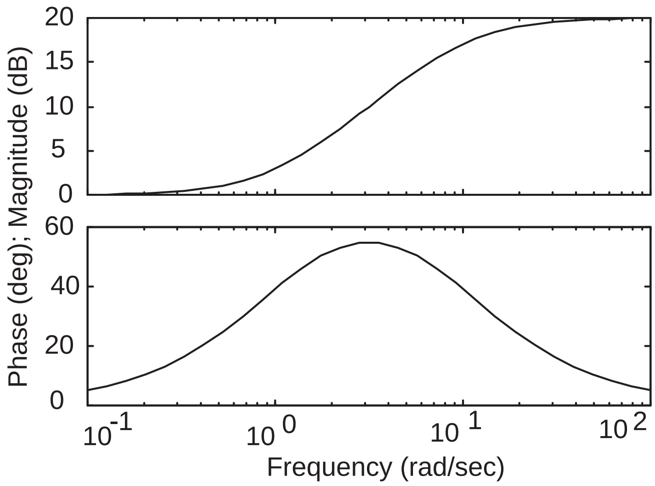

Figure 5.5: Bode diagram of a phase-lead compensator

If the pole was neglijible, that is $|p| >> |z|$ and the zero occurred at the origin of the s-plane we have a differentiator so that

$$G_c(s) = \frac{k}{p}s$$

It is often useful to add a cascade compensation network that provides a phase-lag characteristic. The **phase-lag network** has a transfer function:

$$G_c(s) = \frac{k(s+z)}{s+p}$$

where $|z| > |p|$. The poles-zero location is shown in figure 5.6. The Bode diagram of a phase-lag compensator is shown in Figure 5.7.

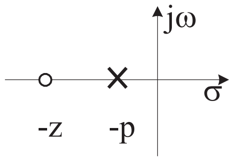

Figure 5.6: Pole-zero location of a phase-lag compensator

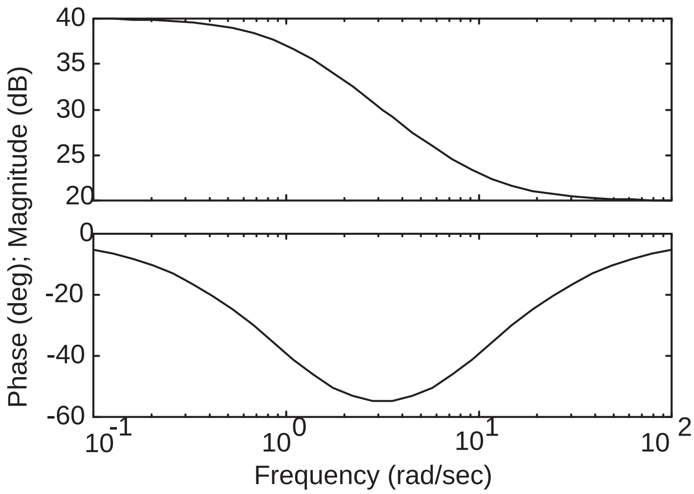

Figure 5.7: Bode diagram of a phase-lag compensator

5.1.4 Phase-lag compensation based on root-locus

Consider the problem of finding a suitable compensation network for the case where the system exhibits satisfactory transient-response characteristics but unsatisfactory steady-state characteristics (see section 3.5.3 in Chapter 3).

Compensation in this case essentially consists of increasing the open-loop gain without appreciably changing the transient response characteristics. This can be accomplished if a lag compensator is put in cascade with the given feed-forward transfer function.

To avoid an appreciable change in the root loci the angle contribution of the lag network should be limited to a small amount, say 5^o. To assure this, we place the pole and zero of the lag network relatively close together and near the origin of the s-plane. Then the closed-loop poles of the compensated system will be shifted only slightly from their original locations. Hence, the transient response characteristics will be essentially unchanged.

The transfer function of a phase-lag network is of the form:

$$G_c(s) = \frac{s+z}{s+p}$$

The steady-state error of un uncompensated system with the open-loop transfer function $G(s)$ (Figure 5.8) is:

$$\epsilon_{ss} = \lim_{s \to 0} \frac{sR(s)}{1+G(s)}$$

Then, for example, the velocity constant of a type-one system ($N = 1$) is

$$K_v = \lim_{s \to 0} sG(s)$$

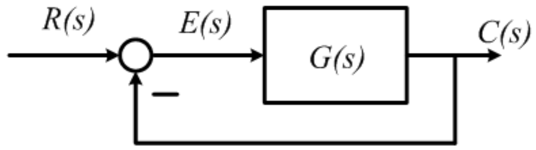

Figure 5.8: Closed-loop system

Therefore, if $G(s)$ is written as:

$$G(s) = \frac{k\prod_{i=1}^{M}(s+z_i)}{s\prod_{i=1}^{Q}(s+p_i)}$$

we obtain the velocity constant of the uncompensated system:

$$K_v = \frac{k\prod_{i=1}^{M}(z_i)}{\prod_{i=1}^{Q}(p_i)}$$

We will now add the phase-lag network as a compensator and determine the compensated velocity constant. If the velocity constant of the compensated system is designated as K_{vcomp} we have:

$$K_{vcomp} = \lim_{s\to 0} sG_c(s)G(s) = \lim_{s\to 0}(G_c(s))K_v = \frac{k\cdot z}{p}\frac{\prod z_i}{\prod p_j}$$

The gain of the compensated root locus at the desired root location will be kz/p. Now, if the pole and zero of the compensator are chosen so that $|z| = \alpha|p| < 1$, the resultant K_{vcomp} will be increased at the desired root location by the ratio $z/p = \alpha$. Then, for example, $z = 0.1$ and $p = 0.01$, the velocity constant at the desired root location will be increased by a factor of 10. However, if the compensator pole and zero appear relatively close together on the s-plane, their effect on the location of the desired root will be negligible. Therefore, the compensator pole-zero combination near the origin of the s-plane, comparing to ω_n can be used to increase the error constant of a feedback system by the factor α, while altering the root location very slightly.

The steps necessary for the design of a phase-lag compensator on the s-plane are as follows:

1. Obtain the root locus of the uncompensated system
2. Determine the transient performance specifications for the system and locate suitable dominant root locations on the uncompensated root locus that will satisfy the specifications.
3. Calculate the loop gain at the desired root location and thus the system error constant.
4. Compare the uncompensated error constant with the desired error constant, and calculate the necessary increase that must result from the pole-zero ratio of the compensator.
5. With the known ratio of the pole-zero combination of the compensator, determine a suitable location of the pole and zero of the compensator so that the compensated root locus will still pass through the desired root location. Locate the pole and zero near the origin of the s-plane in comparison with ω_n.

The fifth requirement can be satisfied if the magnitude of the pole and zero is significantly less than ω_n of the dominant roots and they appear to merge as measured from the desired root location. The pole and zero will appear to merge at the root location if the angles from the compensator pole and zero are essentially equal as measured to the root location. One method of locating the zero and pole of the compensator is based on the requirement that the difference between the angle of the pole and the angle of the zero as measured at the desired root is less than 2^o. An example will illustrate this approach to the design of a phase-lag compensator.

Example 5.1.1 *Consider an uncompensated system with the open-loop transfer function:*

$$G(s) = \frac{k}{s(s+2)}$$

It is required that the damping ratio of the dominant complex roots be 0.45, while a system velocity constant equal to 20 is attained. The uncompensated root locus is a vertical line at $s = -1$ *and results in a root on the* $\zeta = 0.45$ *line at* $s = -1 \pm j2$*, as shown in Figure 5.9. Measuring the gain at this root, we have*

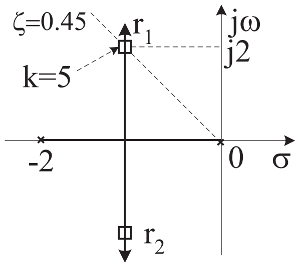

Figure 5.9: Root locus of the uncompensated system

$k = (2.24)^2 = 5$*. Therefore the velocity constant of the uncompensated system is:*

$$K_v = \frac{k}{2} = 2.5$$

Thus the required ratio of the zero to the pole of the compensator is:

$$\left|\frac{z}{p}\right| = \alpha = \frac{K_{vcomp}}{K_v} = \frac{20}{2.5} = 8$$

Examining Figure 5.10 we find that we might set $-z = -0.1$ *and then* $-p = -0.1/8$*. The difference of the*

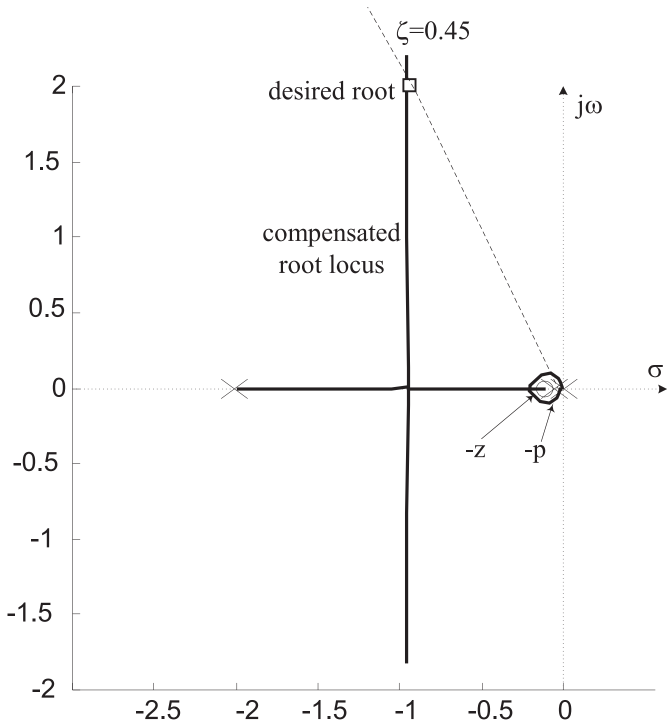

Figure 5.10: Root locus of the compensated system

angles from p *and* z *at the desired root is approximately* 1^o*, and therefore* $s = -1 \pm j2$ *is still the location* *of the dominant roots. A sketch of the compensated root locus is shown also in Figure 5.10. Therefore the compensated system transfer function is:*

$$G_c(s)G(s) = \frac{5(s+0.1)}{s(s+2)(s+0.0125)}$$

where $k/\alpha = 5$ *or* $k = 40$ *in order to account for the attenuation of the lag network.*

5.1.5 Phase-lead compensation based on root-locus

The root-locus approach to design is very powerful when the specifications are given in terms of time domain quantities, such as the damping ratio and undamped natural frequency of the desired dominant closed-loop poles, maximum overshoot, rise time and settling time.

Consider a design problem in which the original system either is unstable for all values of gain or it is stable but has undesirable transient response characteristics. In such a case, the reshaping of the root locus is necessary in the broad neighborhood of the $j\omega$ axis and the origin in order that the dominant closed-loop poles be at desired locations in the complex plane. This problem may be solved by inserting an appropriate lead compensator in cascade with the feed-forward transfer function.

The procedure for designing a lead compensator may be stated as follows:

1. List the system specifications and translate them into a desired root location for the dominant roots.
2. Sketch the uncompensated root locus and determine whether the desired root locations can be realized with an uncompensated system.
3. If a compensator is necessary, place the zero of the phase-lead network directly below the desired root location (or to the left of the first two real poles).
4. Determine the pole location so that the total angle at the desired root location is -180^o and therefore is on the compensated root locus.
5. Evaluate the total system gain at the desired root location and then calculate the steady-state error.
6. Repeat steps if the error is not satisfactory.

Therefore, we first locate our desired dominant root locations so that the dominant roots satisfy the specifications in terms of ζ and ω_n, as shown in Figure 5.11.

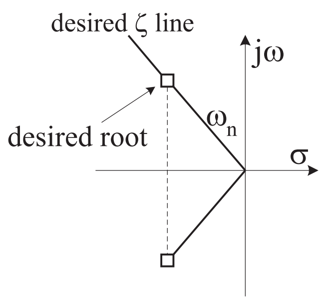

Figure 5.11: Desired poles location in the complex plane

The root locus of the uncompensated system is sketched as illustrated in Figure 5.12.

Then a zero is added to provide a phase lead by placing it to the left of the first two real poles. Some cautions must be maintained because the zero must not alter the dominance of the desired roots; that is,

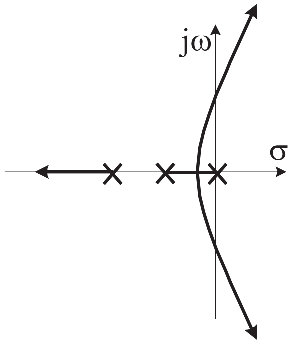

Figure 5.12: Root locus of an uncompensated system

the zero should not be placed nearer the origin than the second pole on the real axis, or a real root near the origin will result and will dominate the system response.

Thus, in Figure 5.13, we note that the desired root is directly above the second pole, and we place the zero z somewhat to the left of the second real pole.

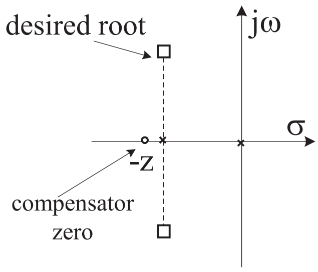

Figure 5.13: Location of the compensator zero

Consequently, the real root may be near the real zero and the coefficient of this term of the partial fraction expansion may be relatively small. Thus, the response due to this real root may have very little effect on the overall system response. Nevertheless, the designer must be aware that the dominant roots will not by themselves dictate the response. It is usually wise to allow for some margin of error in the design and to test the compensated system using a computer simulation.

Because the desired root is a point on the root locus, when the final compensation is accomplished, we expect the algebraic sum of the vector angles to be -180^o at that point.

Thus we calculate the angle from the pole of the compensator, θ_p, in order to result in a total angle of -180^o. Then, locating a line at an angle θ_p intersecting the desired root, we are able to evaluate the compensator pole, p, as shown in Figure 5.14.

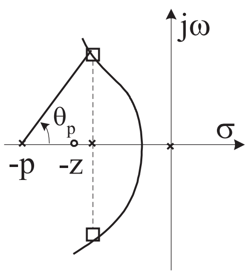

Figure 5.14: Location of the compensator pole

After the design is completed, one evaluates the gain of the system at root location, which depends on p

and z, and then calculates the steady-state error for the compensated system. If the error is not satisfactory, one must repeat the design steps and alter the location of the desired root, as well as the location of the compensator pole and zero.

Example 5.1.2 *Lead compensation using root locus. Let us consider a single-loop feedback control system with the open-loop transfer function*

$$G(s) = \frac{k}{s^2}$$

The response of the uncompensated system is an undamped oscillation because

$$T(s) = \frac{G(s)}{1+G(s)} = \frac{k}{s^2+k}$$

Therefore, the compensation network is added so that the loop transfer function is $G_c(s)G(s)$.

The specifications for the system are:

- *Settling time, $t_s \leq 4$ seconds*
- *Percent overshoot for a step input $M_p \leq 35\%$.*

We desire to compensate this system with a network $G_c(s)$, where

$$G_c(s) = \frac{s+z}{s+p}$$

and $|z| < |p|$.

The damping ratio of the compensated system ζ is obtained from:

$$M_p = e^{-\pi\zeta/\sqrt{1-\zeta^2}} \leq 0.35$$

We obtain, $\zeta \geq 0.32$.

The settling time requirement is

$$t_s = \frac{4}{\zeta\omega_n} \leq 4$$

and therefore, $\zeta\omega_n \geq 1$.

Thus we choose the real part of the dominant closed-loop poles equal to -1 and they will be placed in the left-hand s-plane at an angle of maximum $\alpha = arccos\zeta = 71.3^o$, measured from the negative real axis.

We will choose a desired dominant root location as $r_{1,2} = -1 \pm j2$, as shown in Figure 5.15.

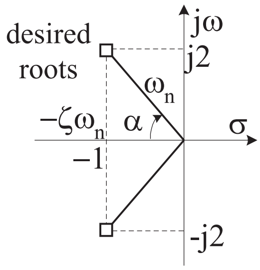

Figure 5.15: Location of the desired roots

Thus, $\zeta = 0.45$ and $\alpha \simeq 64^o$. Now we place the zero of the compensator below the desired location at $s = -z = -1$ as shown in Figure 5.16.

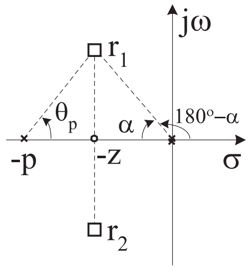

Figure 5.16: Location of the compensator pole and zero

The total angle at the desired root location is:

$$\begin{aligned}\angle(G_c(s)G(s))|_{s=r_1} &= \angle(r_1+z) - \angle(r_1+0) - \angle(r_1+0) - \angle(r_1+p)\\ &= 90^o - 2(180^o - \alpha) - \theta_p\\ &= -142^o - \theta_p\end{aligned}$$

To have a total of -180^o at the desired root (such that $r_{1,2}$ are on the compensated root locus) we evaluate the angle from the undetermined pole θ_p as:

$$-180^o = -142^o - \theta_p$$

or

$$\theta_p = 38^o$$

Then, a line is drawn at an angle $\theta_p = 38^o$ intersecting the desired root location and the real axis as shown in Figure 5.17. The point of intersection with the real axis is then $s = -p = -3.6$.

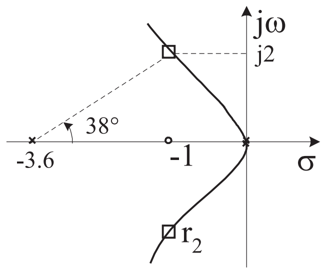

Figure 5.17: Location of the compensator pole and zero

Therefore the compensator is

$$G_c(s) = \frac{s+1}{s+3.6}$$

and the compensated transfer function for the system is:

$$G_c(s)G(s) = \frac{k(s+1)}{s^2(s+3.6)}$$

The gain k is evaluated from the magnitude condition

$$|G_c(s)G(s)|_{s=r_1} = 1$$

$$k = \left|\frac{s^2(s+3.6)}{s+1}\right|_{s=-1+j2} = \frac{2.23^2 \cdot 3.25}{2} = 8.1$$

Example 5.1.3 *Lead compensation using root locus. Let us consider a single-loop feedback control system as shown in Figure 5.18 with the open-loop transfer function*

$$G(s) = \frac{k}{s(s+2)}$$

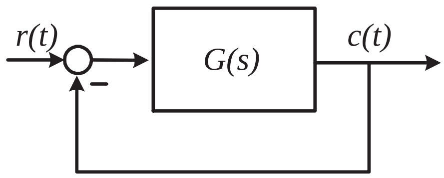

Figure 5.18: Control system

It is desired that the damping ratio of the dominant roots be $\zeta = 0.45$ and the steady-state error for a ramp input be 0.05.

The error $\epsilon(s)$ is:

$$\epsilon(s) = R(s) - C(s) = R(s) - G(s)\epsilon(s)$$

or

$$\epsilon(s) = \frac{R(s)}{1 + G(s)}$$

The steady-state error for a ramp input $R(s) = 1/s^2$ can be calculated with the final value theorem

$$\epsilon_{ss} = \lim_{s\to 0} sR(s)\frac{1}{1+G(s)}$$

Thus,

$$0.05 = \lim_{s\to 0} s\frac{1}{s^2}\frac{1}{1 + k/s(s+1)} = \lim_{s\to 0}\frac{s+2}{s^2+2s+k} = \frac{2}{k}$$

To satisfy the error requirement, the gain of the uncompensated system must be $k = 2/0.05 = 40$. when $k = 40$, toe roots of the closed-loop uncompensated system are given by:

$$s^2 + 2s + 40 = 0$$

and

$$s_{1,2} = -1 \pm j6.25 = -\zeta\omega_n \pm \sqrt{1-\zeta^2}$$

The damping ratio of the uncompensated roots is approximately 0.16, and therefore a compensation network must be added.

To achieve a rapid settling time, we will select the real part of the desired roots as $\zeta\omega_n = 4$ and therefore $t_s = 1$ second. The desired root location is shown in Figure 5.19, for $\zeta\omega_n = 4$, $\zeta = 0.45$ and $\omega_n = 9$. The desired roots are then $r_{1,2} = -4 \pm j8.03$.

The zero of the compensator is placed at $s = -z = 4$ directly below the desired root location. Then the angle criterion at the root r_1 is:

$$\begin{aligned} -180^o &= \angle(r_1 + z) - \angle(r_1 + 0) - \angle(r_1 + 2) = \\ &= 90^o - (180^o - arccos\frac{4}{9}) - (180^o - arctan\frac{8.03}{2}) - \theta_p \\ &= 90^o - 116^0 - 104^0 - \theta_p \end{aligned}$$

The angle from the undetermined pole is:

$$\theta_p = 50^o$$

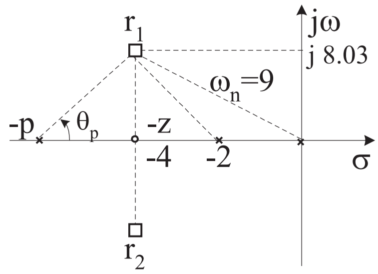

Figure 5.19: Desired roots of the compensated system

This angle is drawn to intersect the desired root location, and p is evaluated as $s = -p = 10.6$, *(for example from* $tan\theta_p = 8.03/(p-4)$*). The compensator is:*

$$G_c(s) = \frac{s+4}{s+10.4}$$

The gain of the compensated system:

$$G_c(s)G(s) = \frac{k(s+4)}{s(s+2)(s+10.4)}$$

is then:

$$k = \left|\frac{s(s+2)(s+10.4)}{s+4}\right|_{s=-4+8.03j} = 96.5$$

The steady-state error for a ramp input is:

$$\epsilon_{ss} = \lim_{s\to 0} s\frac{1}{s^2}\frac{1}{1+96.5(s+4)/(s(s+2)(s+10.4))} = \lim_{s\to 0}\frac{(s+2)(s+10.4)}{s(s+2)(s+10.4)+96.5(s+4)}$$

$$\epsilon_{ss} = \frac{21.2}{386} = 0.054$$

The steady-state error of the compensated system is greater than the desired value 0.05. Therefore we must repeat the design procedure for a second choice of a desired root. If we choose $\omega_n = 10$ *(and* $t_s = 0.88$*) the process can be repeated and the resulting gain k will be increased. The compensator pole and zero location will also be altered. Then the steady-state error can be again evaluated.*

Exercise. show that for $\omega_n = 10$ *the steady-state error for a ramp input is 0.4 when* $z = 4.5$ *and* $p = 11.6$.

The addition of an integration as $G_c(s) = k_2 + k_3/s$ *can also be used to reduce the steady-state error for a ramp input* $r(t) = t, t \geq 0$. *For example, if the uncompensated system* $G(s)$ *possessed one integration, the additional integration due to* $G_c(s)$ *would result in a zero steady-state error for a ramp input.*

Example 5.1.4 *The uncompensated loop transfer function of a control system is:*

$$G(s) = \frac{k_1}{(2s+1)(0.5s+1)}$$

where k_1 *can be adjusted. To maintain zero steady-state error for a step input, we will add a PI compensation network*

$$G_c(s) = k_2 + \frac{k_3}{s} = k_2\frac{s+k_3/k_2}{s}$$

Furthermore, the transient response of the system is required to have an overshoot less than or equal to 10%.

Therefore the dominant complex roots must be on, or below the $\zeta = 0.6$ *line (or at an angle* $\alpha < 53^o$ *measured from the origin) as shown in Figure 5.20.*

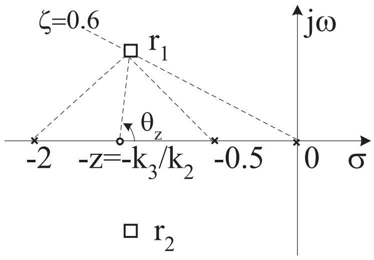

Figure 5.20: Desired roots of the compensated system

We will adjust the compensator zero so that the real part of the complex roots is $\zeta\omega_n = 0.75$ *and thus the settling time is* $t_s = 4/(\zeta\omega_n) = 16/3$ *seconds. Now, we will determine the location of the zero* $z = -k_3/k_2$ *by ensuring that the angle at the desired root is* -180^o*. Therefore, the sum of angles at the desired root is:*

$$-180^o = -127^o - 104^o - 38^o + \theta_z$$

where θ_z *is the angle from the undetermined zero. We find that* $\theta_z = 89^o$ *and the location of the zero is* $-z = -0.75$*. To determine the gain at the desired root we evaluate the vector lengths from the poles and zeros and obtain:*

$$k = k_1 k_2 = \frac{1.25(1.03)1.6}{1} = 2.08$$

It should be noted that the zero $-k_3/k_2$ *should be placed to the left of the pole at* $s = -0.5$ *to ensure that the complex roots dominate the transient response. In fact, the third root of the compensated system can be determined as* $s = -1$ *and therefore this real root is only 4/3 times the real part of the complex roots. Thus, although complex roots dominate the response of the system, the equivalent damping of the system is somewhat less than* $\zeta = 0.6$ *due to the real root and zero. The closed-loop transfer function of the system is:*

$$T(s) = \frac{G_c(s)G(s)}{1 + G_c(s)G(s)} = \frac{2.08(s + 0.75)}{(s + 1)(s + r_1)(s + r_2)}$$

where $r_{1,2} = -0.75 \pm j1$*. The effect of the zero is to increase the overshoot to a step input. If we wish to attain an overshoot of* 5%*, we may use a pre-filter (an element placed in series with T(s)) so that the zero is eliminated in T(s) by setting*

$$G_p(s) = \frac{0.75}{s + 0.75}$$

The overshoot without pre-filter is 17.6% *and with the pre-filter it is* 2%*.*

5.2 PID - The Basic Technique for Feedback Control

A feedback controller is designed to generate an *output* that causes some corrective effort to be applied to a *process* so as to drive a measurable *process variable* towards a desired value known as the *setpoint*. The *controller* uses an actuator to affect the process and a *sensor* to measure the results. Figure 5.21 shows a typical feedback control system with blocks representing the dynamic elements of the system and arrows representing the flow of information, generally in the form of electrical signals. Virtually all feedback controllers determine their output by observing the *error* between the setpoint and the actual process variable measurement, (VanDoren, 1997).

Errors occur when an operator changes the setpoint intentionally or when a process load changes the process variable accidentally.

Example 5.2.1 *In warm weather, a home thermostat is a familiar controller that attempts to correct temperature of the air inside a house. It measures the room temperature with a thermocouple and activates the*

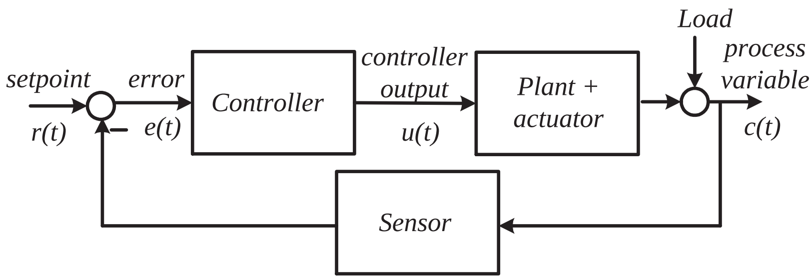

Figure 5.21: Control system

air conditioner whenever an occupant lowers the desired room temperature or a random heat source raises the actual room temperature. In this example, the house is the process, the actual room temperature inside the house is the process variable, the desired room temperature is the setpoint, the thermocouple is the sensor, the activation signal to the air conditioner is the controller output, the air conditioner itself is the actuator, and the random heat sources (such as sunshine and warm bodies) constitute the loads on the process.

5.2.1 PID Control

PID (proportional-integral-derivative) is the control algorithm most often used in industrial control. Despite the abundance of sophisticated tools, including advanced controllers, the PID controller is still the most widely used in modern industry,controlling more than 95 % of closed-loop industrial processes. It is implemented in industrial single loop controllers, distributed control systems and programmable logic controllers (PLC).

Figure 5.22: Industrial controllers

A PID controller performs much the same function as a thermostat but with a more elaborate algorithm for determining its output. It looks at the current value of the error, the integral of the error over a recent time interval, and the current derivative of the error signal to determine not only how much of a correction to apply, but for how long. Those three quantities are each multiplied by a *tuning constant* and added together to produce the current controller output $u(t)$ as in equation 5.1:

$$u(t) = K_P e(t) + K_I \int_0^t e(\tau) d\tau + K_D \frac{de(t)}{dt} \tag{5.1}$$

In this equation:

- K_P is the *proportional tuning constant*,
- K_I is the *integral tuning constant*,
- K_D is the *derivative tuning constant*,

- the error $e(t)$ is the difference between the setpoint $r(t)$ and the process variable $c(t)$ at time t (see Figure 5.21).

If the current error is large or the error has been sustained for some time or the error is changing rapidly, the controller will attempt to make a large correction by generating a large output. Conversely, if the process variable has matched the setpoint for some time, the controller will leave well enough alone.

The parallel structure described by the equation (5.1) is presented as a block diagram in Figure 5.23.

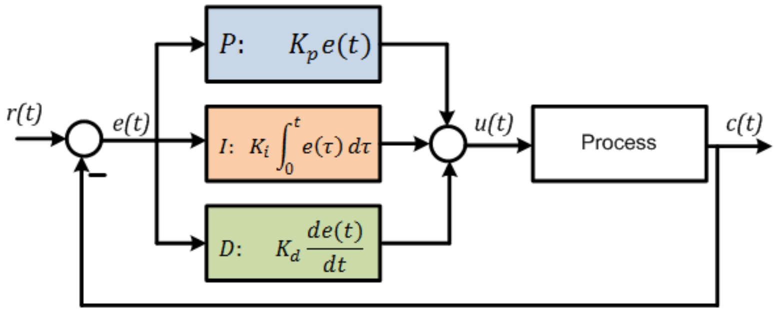

Figure 5.23: PID controller: parallel structure

If we apply the Laplace transform to relation (5.1), and take the error $e(t)$ as the input variable and the command $u(t)$ the output variable of the PID controller, the controller transfer function will be determined from:

$$U(s) = K_P E(s) + K_I \frac{1}{s} E(s) + K_D s E(s) \tag{5.2}$$

and will result as:

$$G_{PID}(s) = \frac{U(s)}{E(s)} = K_P + K_I \frac{1}{s} + K_D s \tag{5.3}$$

The block diagram for this structure is presented in Figure 5.24.

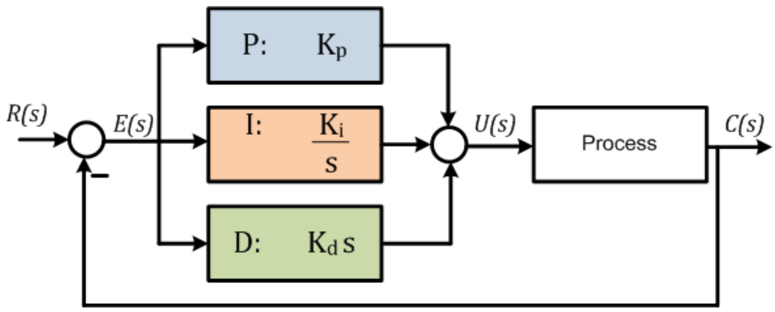

Figure 5.24: PID controller: parallel structure

5.2.2 PID controller actions

Consider a unity-feedback closed loop system with a process having the transfer function $G(s)$ and a PID controller as shown in Figure 5.25. The effect of each term will be presented in the following sections.

5.2.2.1 P action

A **P** controller is obtained from (5.1) by setting the integral and derivative tuning constants to zero: $K_D = 0$, $K_I = 0$. The output of a P controller is then:

$$u(t) = K_P e(t)$$

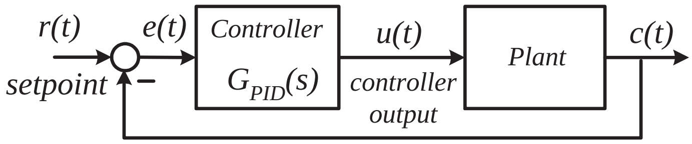

Figure 5.25: Control system

and the transfer function:

$$G_P(s) = K_P$$

The control action (command) is proportional to the error. A typical example of proportional control is presented in Figure 5.26. The figure shows the error signal and the controller output, as well as the process output for different values of K_P, after a step change in the setpoint.

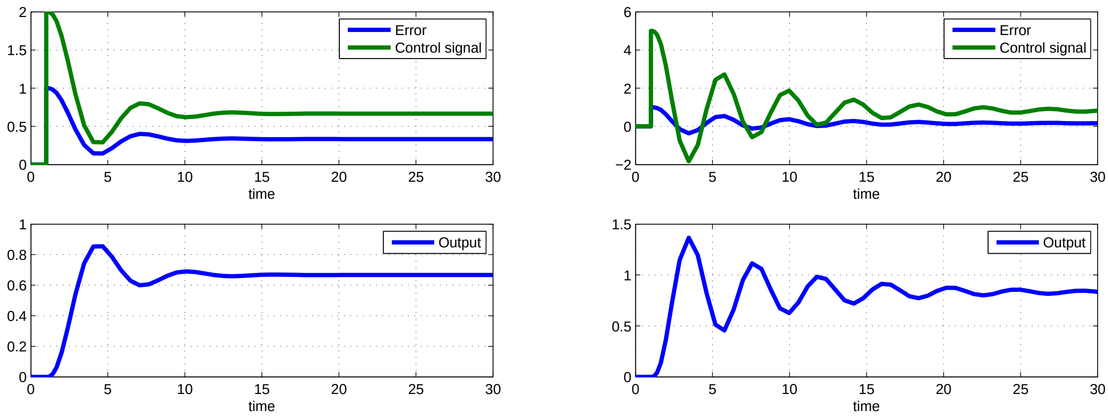

Figure 5.26: P action. (left) $K_P = 2$, (right) $K_P = 5$

Notice that:

- the steady-state error decreases with increasing controller gain
- the response becomes more oscillatory with increasing controller gain

5.2.2.2 I action

A **PI** controller is obtained from (5.1) by setting the derivative tuning constants to zero: $K_D = 0$. The output of a PI controller is then:

$$u(t) = K_P e(t) + \int_0^t e(\tau) d\tau$$

and the transfer function:

$$G_{PI}(s) = K_P + \frac{K_I}{s}$$

The value of the control variable at each moment of time is given by the current value of the error and the area under the error signal, both weighted with the constants K_P and K_I, respectively.

A typical example of proportional-integral control is presented in Figure 5.27. The figure shows the error signal and the controller output, as well as the process output for different values of K_I, after a step change in the setpoint.The value of K_P was kept constant, $K_P = 1$.

Notice that:

- the steady-state error is zero for any value of the integral gain K_I

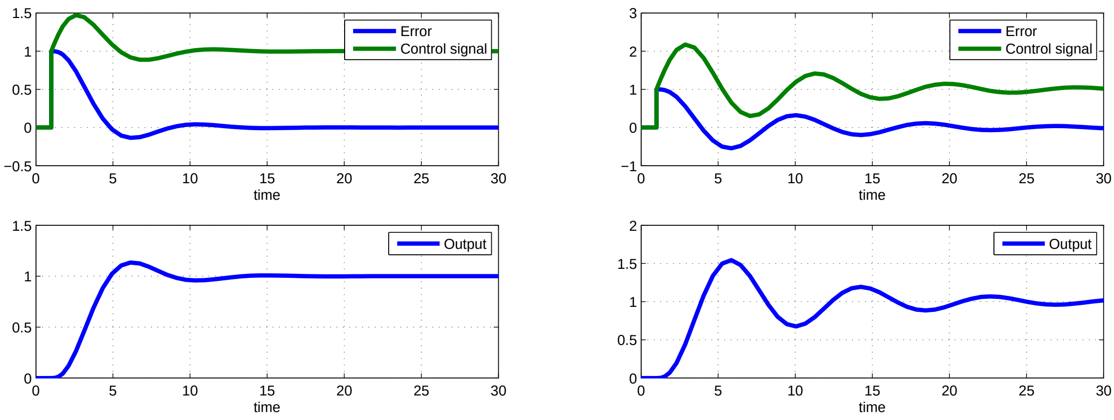

Figure 5.27: I action - PI control. (left) $K_P = 1, K_I = 0.5$, (right) $K_P = 1, K_I = 1$

- the response is faster (smaller rise time) but more oscillatory with increasing integral gain K_I

The main function of the integral action is to make sure that the error is zero at steady-state. With proportional control only, there is normally a steady-state error. With integral action, a small positive error will always increase the control signal, and a negative error will give a decreasing control signal, no matter how small the error is.

5.2.2.3 D action

A **PD** controller is obtained from (5.1) by setting the integral tuning constants to zero: $K_I = 0$. The output of a PD controller is then:

$$u(t) = K_P e(t) + K_D \frac{de(t)}{dt}$$

and the transfer function:

$$G_{PI}(s) = K_P + K_D s$$

The value of the control variable at each moment of time is given by the current value of the error and the derivative of the error signal, both weighted with the constants K_P and K_D, respectively.

The purpose of the derivative action is to improve the closed-loop stability. The instability can be described as follows. Because of the process dynamics, it will take some time before a change in the control variable is noticeable in the process output. Thus, the control system will be late in correcting for an error. The action of a controller with a proportional and derivative action may be interpreted as if the control is made proportional to the *predicted* process output, where the prediction is made by extrapolating the error by the tangent to the error curve, as shown in Figure 5.28, (Astrom and Hagglund, 1995).

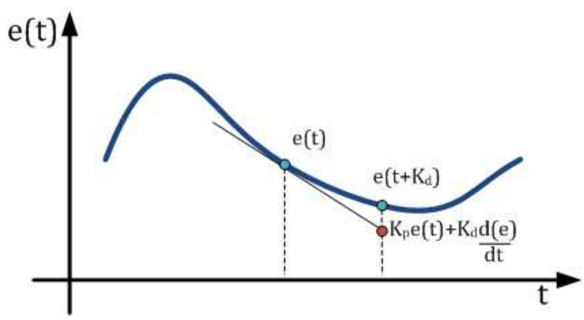

Figure 5.28: Derivative as prediction obtained by linear extrapolation

The properties of the derivative action are illustrated in Figure 5.29 which shows a simulation of a system with a PID controller. The proportional and the integral gains are kept constant, $K_P = 1$ and $K_I = 1$, and the derivative constant K_D is changed. For $K_D = 0$ PI control and the closed-loop system is oscillatory (see Figure 5.27 - right). When $K_D = 1$ or $K_D = 3$, damping increases with K_D.

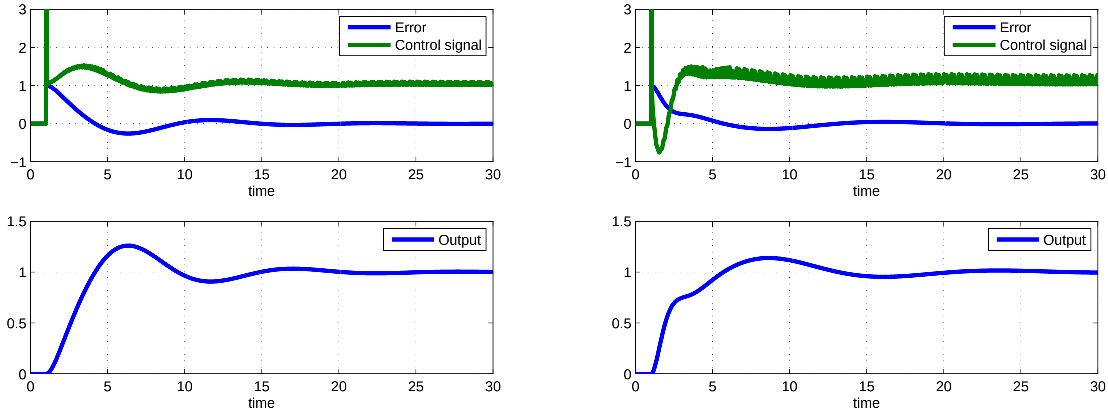

Figure 5.29: D action - PID control. (left) $K_P = 1, K_I = 1, K_D = 1$, (right) $K_P = 1, K_I = 1, K_D = 3$

Notice that:

- the steady-state error is zero because of the integral term in the controller
- the overshoot decreases with K_D
- because of the sudden change in error, the derivative term has a large value at the initial time, thus producing a very large initial control signal.

5.2.2.4 Other forms of a PID controller

The general transfer function of a PID controller is:

$$G_{PID}(s) = \frac{U(s)}{E(s)} = K_P + K_I\frac{1}{s} + K_D s \tag{5.4}$$

This relation can be rearranged to give:

$$G_{PID}(s) = K_P\left(1 + \frac{K_I}{K_P}\frac{1}{s} + \frac{K_D}{K_P}s\right) \tag{5.5}$$

or

$$G_{PID}(s) = K_P\left(1 + \frac{1}{T_i s} + T_d s\right) \tag{5.6}$$

The controller parameters are now:

- K_P the controller gain
- $T_i = \frac{K_I}{K_P}$ the integral time
- $T_d = \frac{K_D}{K_P}$ the derivative time

We have assumed that a D behavior can be realized by a PID controller. This is an ideal assumption and in reality the ideal D element cannot be realized. In real PID controllers a lag in included in the D behavior. An element with the transfer function

$$G_D(s) = \frac{T_d s}{\frac{T_d}{N}s + 1}$$

is introduced in the block diagram in Figure 5.3 instead of the D element, where N has a large value such that the time constant at the denominator is a small. The transfer function of the *real* PID controller is then:

$$G_{PID} = K_P\left(1 + \frac{1}{T_i s} + \frac{T_d s}{\frac{T_d}{N}s + 1}\right)$$

5.2.3 Tuning a PID controller

(from (VanDoren, 1998))

Tuning is setting the K_P, K_I, and K_D tuning constants so that the weighted sum of the proportional, integral, and derivative terms produces a controller output that steadily drives the process variable in the direction required to eliminate the error.

How to best tune a PID controller depends upon how the process responds to the controller's corrective efforts. Consider a sluggish process that tends to respond slowly. If an error is introduced abruptly (as when the setpoint is changed), the controller's initial reaction will be determined primarily by the derivative term in equation 5.1. This will cause the controller to initiate a burst of corrective efforts the instant the error changes from zero. The proportional term will then come in to play to keep the controller's output going until the error is eliminated.

After a while, the integral term will also begin to contribute to the controller's output as the error accumulates over time. In fact, the integral term will eventually come to dominate the output signal, since the error decreases so slowly in a sluggish process. Even after the error has been eliminated, the controller will continue to generate an output based on the history of errors that have been accumulating in the controller's integrator. The process variable may then overshoot the setpoint, causing an error in the opposite direction.

If the integral tuning constant is not too large, this subsequent error will be smaller than the original, and the integral term will begin to diminish as negative errors are added to the history of positive ones. This whole operation may then repeat several times until both the error and the accumulated error are eliminated. Meanwhile, the derivative term will continue to add its share to the controller output based on the derivative of the oscillating error signal. The proportional term too will come and go as the error increases or decreases.

Now suppose the process responds quickly to the controller's efforts. The integral term in equation 5.1 will not play as dominant a role in the controller's output since the errors will be so short lived. On the other hand, the derivative term will tend to be larger since the error will change rapidly.

Clearly, the relative importance of each term in the controller's output depends on the behavior of the controlled process. Determining the best mix suitable for a particular application is the essence of controller tuning. For the sluggish process, a large value for the derivative tuning constant K_D might be advisable to accelerate the controller's reaction to an error that appears suddenly. For the fast-acting process, however, an equally large value for K_D might cause the controller's output to fluctuate as every change in the error (including extraneous changes caused by measurement noise) is amplified by the controller's derivative action.

Hundreds of mathematical and heuristic techniques for selecting appropriate values for the tuning constants have been developed over the last 50 years.

There are basically three schools of thought on how to select K_P, K_I, and K_D values to achieve an acceptable level of controller performance.

1. The first method is the simple trial-and-error approach. Experienced control engineers seem to know just how much proportional, integral, and derivative action to add or subtract to correct the performance of a poorly tuned controller. Unfortunately, intuitive tuning procedures can be difficult to develop since a change in one tuning constant tends to affect the performance of all three terms in the controller's output.

   For example, turning down the integral action reduces overshoot. This in turn slows the rate of change of the error and thus reduces the derivative action as well.

2. The analytical approach to the tuning problem is more rigorous. It involves a mathematical model of the process that relates the current value of the process variable to its current rate of change plus a history of the controller's output.

   There are hundreds of analytical techniques for translating model parameters into tuning constants. Each approach uses a different model, different controller objectives, and different mathematical tools.

3. The third approach to the tuning problem is a compromise between purely heuristic trial-and-error techniques and the more rigorous analytical techniques. It was originally proposed in 1942 by John G. Ziegler and Nathaniel B. Nichols of Taylor Instruments and remains popular today because of its simplicity and its applicability to many real-life processes.

5.2.4 Ziegler-Nichols methods for tuning PID controllers

(from (VanDoren, 1998))

The block diagram of a simplified control system is shown in Figure 5.30. In practice the output of a

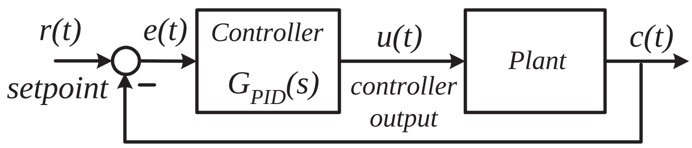

Figure 5.30: Control system

PID controller is given by:

$$u(t) = K_p \left[e(t) + \frac{1}{T_i} \int_0^t e(\tau) d\tau + T_d \frac{de(t)}{dt} \right] \tag{5.7}$$

The transfer function of a PID controller is:

$$G_{PID}(s) = \frac{U(s)}{E(s)} = K_p \left(1 + \frac{1}{T_i s} + T_d s \right)$$

where K_p = proportional gain
T_i = integral time
T_d = derivative time

If a mathematical model of the plant can be derived, then it is possible to apply various design techniques for determining parameters of the controller that will meet the transient and steady-state specifications of the closed-loop system.

If the plant is so complicated and its mathematical model cannot be easily obtained, then analytical approach to the design of a PID controller is not possible. Then we must resort to experimental approaches to the design of PID controllers.

The process of selecting the controller parameters to meet given performance specifications is known as controller tuning.

Ziegler and Nichols (1942) proposed rules for tuning PID controllers (for determining values of K_p, T_i and T_d) based on the transient response characteristics of a given plant. Such determination of the parameters of PID controller can be made by engineers on site by experiments on the plant.

There are two methods called Ziegler-Nichols tuning rules. In both they aimed at obtaining 25% maximum overshoot in step response.

5.2.4.1 The Ultimate Sensitivity Method

The goal is to achieve a marginally stable controller response. A P-controller is used first to control the system as shown in Figure 5.31. Set $T_i = \infty$, and $T_d = 0$.

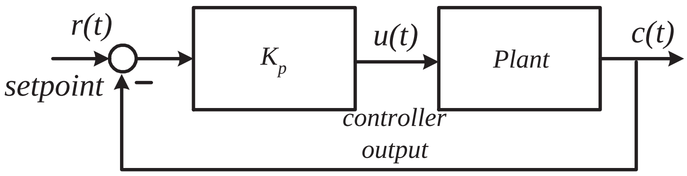

Figure 5.31: Control system with proportional control

Using the proportional control action only, increase K_p from 0 to a critical value K_0 where the output, $c(t)$, first exhibits sustained oscillations. (If the output does not exhibits sustained oscillations for whatever value K_p may take, then this method does not apply.)

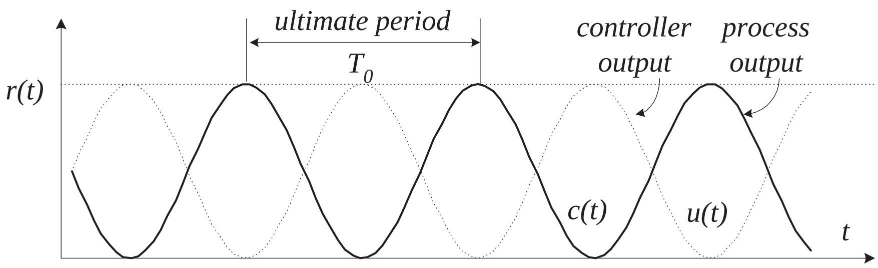

Figure 5.32: Ultimate system response

When the response shown in Figure 5.32 is obtained, the result is termed the *ultimate gain* setting that causes a continuous sinusoidal response in the process output.

Thus, the critical gain and the corresponding period T_0 are experimentally determined.

Ziegler and Nichols suggested to set the values of the parameters K_p, T_i and T_d according to the formula shown in Table 5.1.

Type of controller	K_p	T_i	T_d
P	$0.5K_0$	∞	0
PI	$0.45K_0$	$1/1.2T_0$	0
PID	$0.6K_0$	$0.5T_0$	$0.125T_0$

Table 5.1: PID Parameters

Notice that the PID controller tuned by the ultimate sensitivity method of Ziegler-Nichols gives:

$$G_{PID}(s) = K_p\left(1+\frac{1}{T_i s}+T_d s\right) = 0.6K_0\left(1+\frac{1}{0.5T_0 s}+0.125T_0 s\right)$$
$$= 0.075K_0T_0\frac{(s+4/T_o)^2}{s}$$

Thus, the PID controller has a pole at the origin and double zeros at $s=-4/T_0$.

Derivative cautions. Derivative action is applied only one time when the process output moves away from the setpoint. Derivative works on rate of change. If the process output rate of change is caused by noise, the derivative may cause over-corrections. Control loops likely to have significant noise are pressure and flow. Level can also be noisy when stirring/aggitating or splashing occurs.

Ziegler-Nichols tuning rules have been widely used to tune PID controllers in process control where the plant dynamics are not precisely known. Over many years such tuning rules proved to be very useful.

Ziegler-Nichols tuning rules can, of course, be applied to plants where dynamics are known.

Example 5.2.2 *Consider the control system shown in Figure 5.33 in which a PID controller is used to control a system with the transfer function $G(s)$. The PID controller has the transfer function*

$$G_{PID}(s) = K_p\left(1+\frac{1}{T_i s}+T_d s\right)$$

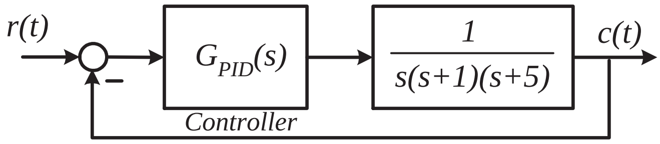

Figure 5.33: PID-controlled system

Although many analytical methods are available for the design of a PID controller for the present system let us apply a Ziegler Nichols tuning rule for determination of the values of parameters K_p, T_i, T_d. Then obtain a unit-step response curve and check to see if the designed system exhibits approximately 25% maximum overshoot. If the maximum overshoot is excessive, make a fine tuning and reduce the amount of the maximum overshoot to approximately 25%.

By setting $T_i=\infty$ and $T_d=0$, we obtain the closed-loop transfer function as follows:

$$\frac{C(s)}{R(s)} = \frac{K_p}{s(s+1)(s+5)+K_p}$$

The value of K_p that makes the system marginally stable so that sustained oscillation occurs can be obtained by use of Routh's stability criterion. Since the characteristic equation for the closed-loop system is:

$$s^3+6s^2+5s+K_p=0$$

the Routh array becomes as follows:

$$\begin{array}{lccc} s^3 & : & 1 & 5 \\ s^2 & : & 6 & K_p \\ s^1 & : & (30-K_p)/6 & \\ s^0 & : & K_p & \end{array}$$

Examining the coefficients of the first column of the Routh table, we find that sustained oscillation will occur if $K_p = 30$. Thus, the critical gain is $K_0 = 30$.

With the gain K_p set equal to $K_0(= 30)$, the characteristic equation becomes:

$$s^3 + 6s^2 + 5s + 30 = 0$$

or

$$(s + 1)(s^2 + 5) = (s + 1)(s + \omega_n^2) = 0$$

from which we find the frequency of the sustained oscillation to be $\omega_n = \sqrt{5}$. Hence, the period of the sustained oscillation is:

$$T_0 = \frac{2\pi}{\omega_n} = \frac{2\pi}{\sqrt{5}} = 2.81$$

Referring to Table 5.1, we determine K_p, T_i, T_d as follows:

$$K_p = 0.6K_0 = 18$$
$$T_i = 0.5T_0 = 1.405$$
$$T_d = 0.125T_0 = 0.35$$

The transfer function of the PID controller is thus:

$$G_{PID}(s) = K_p\left(1 + \frac{1}{T_i s} + T_d s\right) = 18\left(1 + \frac{1}{1.405s} + 0.35s\right) = \frac{6.32(s + 1.42)^2}{s}$$

The PID controller has a pole at the origin and a zero at $s = -1.42$. The closed-loop transfer function is:

$$\frac{C(s)}{R(s)} = \frac{G_{PID}(s)G(s)}{1 + G_{PID}(s)G(s)} = \frac{6.32s^2 + 18s + 12.81}{s^4 + 6s^3 + 11.32s^2 + 18s + 12.81}$$

The unit step response of this system can be obtained easily by using a computer simulation as shown in Figure 5.34. The maximum overshoot in the unit-step response is approximately 60%. The amount of

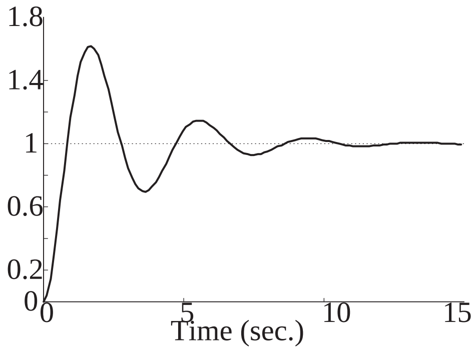

Figure 5.34: Unit-step response of PID-controlled system

maximum overshoot is excessive. It can be reduced by fine tuning the control parameters. Such fine tuning can be made on the computer. We find that keeping $K_p = 18$ and by moving the double zero of the PID controller to $s = -0.65$, that is using the PID-controller:

$$G_{PID}(s) = 18\left(1 + \frac{1}{3.07s} + 0.76s\right) = \frac{13.84(s + 0.65)^2}{s}$$

the maximum overshoot in the unit step response ca be reduced to approximately 18% (see Figure 5.35 - left). If the proportional gain K_p is increased to 39.42 without changing the new location of the double zero, that is using the PID controller:

$$G_{PID}(s) = 39.42\left(1 + \frac{1}{3.07s} + 0.76s\right) = \frac{30.322(s + 0.65)^2}{s}$$

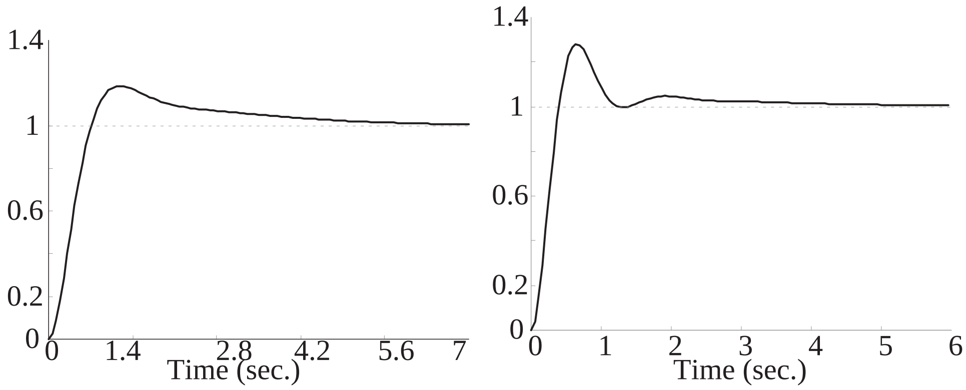

Figure 5.35: Unit-step response of PID-controlled system

then the speed of response in increased, but the maximum overshoot is also increased to approximately 28%, as shown in Figure 5.35 - right. Since the maximum overshoot in this case is fairly close to 25% and the response is faster we may consider that the last controller designed is acceptable. Then the tuned values of K_p, T_i, T_d are:

$$K_p = 39.42, \quad T_i = 3.07, \quad T_d = 0.76$$

The important thing to note here is that the values suggested by the Ziegler-Nichols tuning rule has provided a starting point for fine tuning.

5.2.4.2 Ziegler-Nichols transient response method

The method is also known as *reaction curve (open-loop) method.* The philosophy of open loop testing is to begin with a steady-state process, make a step change to the final control element and record the results of the process output, as shown in Figure 5.36.

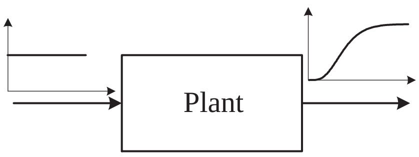

Figure 5.36: Open-loop step response)

Zeigler-Nichols' transient response method will work on any system that has an open-loop step response that is an essentially critically damped or overdamped character like that shown in Figure 5.37. Information produced by the open-loop test is the open-loop gain K, the loop apparent deadtime L, and the loop time constant, T. The transfer function of the plant may then be approximated by a first-order system with a transport lag:

$$\frac{C(s)}{U(s)} = \frac{Ke^{-Ls}}{Ts+1}$$

Ziegler and Nichols suggested to set the values of K_p, T_i and T_d according to the formula shown in Table 5.2.

Notice that the PID controller tuned by the transient response method gives:

$$G_{PID}(s) = K_p\left(1 + \frac{1}{T_i s} + T_d s\right) = 1.2\frac{T}{L}\left(1 + \frac{1}{2Ls} + 0.5Ls\right) = 0.6T\frac{(s+1/L)^2}{s}$$

Thus the PID controller has a pole at the origin and double zeros at $s = -1/L$.

Ziegler-Nichols tuning methods, however, tend to produce systems whose transient response is rather oscillatory and so will need to be tuned further prior to putting the system into closed-loop operation.

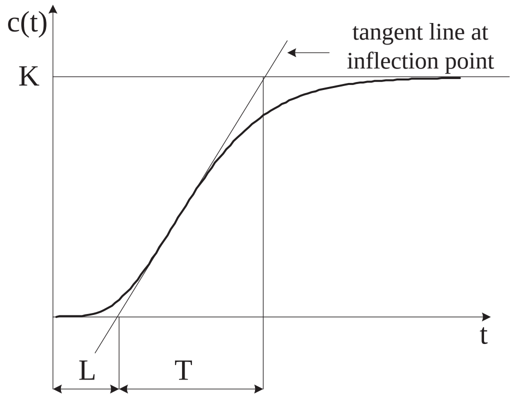

Figure 5.37: Open-loop step response (S-shaped response curve)

Type of controller	K_p	T_i	T_d
P	T/L	∞	0
PI	$0.9T/L$	$L/0.3$	0
PID	$1.2T/L$	$2L$	$0.5L$

Table 5.2: PID Parameters

Bibliography

Astrom, K. and Hagglund, T. (1995). *PID Controllers: Theory, Design and Tuning.* ISA: The Instrumentation, Systems, and Automation Society; 2 Sub edition.

Dorf, R. C. and Bishop, R. H. (2008). *Modern Control Systems.* Pearson Prentice Hall, 11th edition.

VanDoren, V. (1997). `PID`–the basic technique for feedback control. www.controleng.com.

VanDoren, V. (1998). Tuning fundamentals: Basics of proportional-integral-derivative control. www.controleng.com.