# Chapter 4
# Stability Criteria and Steady-State Error

A general control system is made up of several interacting components : the plant to be controlled, the controller, the measurement systems, the actuators, etc. This is the reason why a control system may be very complex. Despite this, any closed-loop control system may be reduced to being represented by a single transfer function that relates the output to be controlled and the reference or the desired output. This closed-loop transfer function can be studied as in Chap. 3. In the present chapter, the concept of block diagrams (see [1] for a historical perspective) is introduced to represent how a closed-loop control system is constituted and how to manipulate them to obtain the closed-loop transfer function is also explained.

On the other hand, a control system is always designed such that: *a*) it is stable, *b*) the desired transient response specifications are satisfied, and *c*) the desired steady-state response specifications are satisfied. According to Chap. 3, a transfer function is stable if all its poles have a negative real part. However, checking this property analytically may be difficult and, because of that, it is necessary to find simple methods to solve this problem. Two alternatives are presented in this chapter: 1) *the rule of signs* to determine the sign of the real part of the roots of a polynomial and, 2) *Routh's stability criterion* (see [1] for a historical perspective).

The transient response of a control system depends on the location of poles (and zeros) of the closed-loop transfer function. There are two different methods of designing a controller, assigning poles and zeros of the closed-loop transfer function at suitable locations: the *root locus method* (see Chap. 5) and the *frequency response method* (see Chap. 6).

Finally, satisfying the desired specifications for the steady-state response is concerned with forcing the closed-loop system response to reach the reference or desired output once the natural response vanishes.[1] This is accomplished by

[1] When time is large enough or in a steady state.

 
V. M. Hernández-Guzmán, R. Silva-Ortigoza, *Automatic Control with Experiments*,
Advanced Textbooks in Control and Signal Processing,
https://doi.org/10.1007/978-3-319-75804-6_4

designing a controller that renders the closed-loop system forced response equal to, or at least close to, the reference or desired output. This subject is also studied in the present chapter.

## 4.1 Block Diagrams

Some examples are presented in the following to show how a block diagram can be simplified to obtain a single transfer function representing the whole block diagram. We also present some examples of block diagrams where a single output is affected by two different inputs.

*Example 4.1* Consider two systems such that the input of one of them is the output of the other, as depicted in Fig. 4.1. It is said that these systems are cascade-connected. Notice that:

$$Y_1(s) = G_1(s)U_1(s), \quad Y_2(s) = G_2(s)U_2(s), \quad U_1(s) = Y_2(s),$$

to obtain:

$$\begin{aligned} Y_1(s) &= G_1(s)G_2(s)U_2(s), \quad G(s) = G_1(s)G_2(s), \\ Y_1(s) &= G(s)U_2(s). \end{aligned}$$

This means that cascade-connected systems, such as those in Fig. 4.1, can be represented by a single transfer function $G(s)$ that can be computed as the product of the individual transfer functions of systems in the connection. The reader can verify that this is true no matter how many cascade-connected systems there are.

*Example 4.2* Suppose that we have two parallel-connected systems, as depicted in Fig. 4.2. Notice that:

$$Y_1(s) = G_1(s)U(s), \quad Y_2(s) = G_2(s)U(s),$$

to write:

$$\begin{aligned} Y_1(s) + Y_2(s) &= (G_1(s) + G_2(s))U(s), \quad G(s) = G_1(s) + G_2(s), \\ Y_1(s) + Y_2(s) &= G(s)U(s). \end{aligned}$$

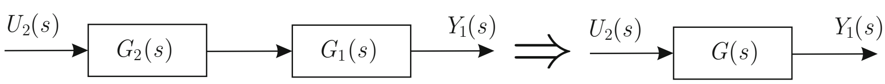



**Fig. 4.1** Cascade-connected systems

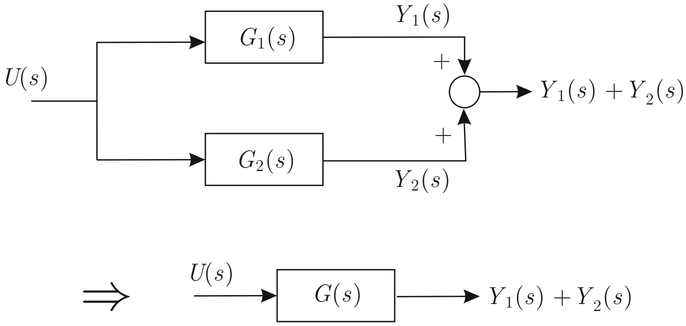
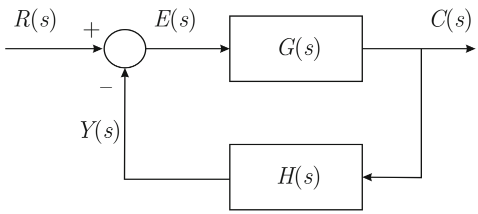



**Fig. 4.2** Parallel-connected systems

**Fig. 4.3** A closed-loop or feedback control system




This means that parallel-connected systems, such as those in Fig. 4.2, can be represented by a single transfer function $G(s)$ that is computed as the addition of the transfer functions of each single system in the parallel connection. The reader can verify that this is true no matter how many systems are in the parallel connection.

*Example 4.3* Consider the closed-loop control system depicted in Fig. 4.3. Blocks $G(s)$ and $H(s)$ stand for the transfer functions of the several components of a control system. In fact, $G(s)$ and $H(s)$ can be the result of the combination of several transfer functions of several components of the control system, such as those in the previous examples. According to the definition of transfer function, from Fig. 4.3 it is found that:

$$C(s) = G(s)E(s), \quad E(s) = R(s) - Y(s), \quad Y(s) = H(s)C(s).$$

Replacing these expressions in each other, the following is obtained:

$$\begin{aligned}
C(s) &= G(s)[R(s) - Y(s)],\\
&= G(s)[R(s) - H(s)C(s)],\\
&= G(s)R(s) - G(s)H(s)C(s),\\
C(s)[1 + G(s)H(s)] &= G(s)R(s),\\
C(s) &= \frac{G(s)}{1 + G(s)H(s)}R(s),
\end{aligned}$$

where:

$$M(s) = \frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)}, \tag{4.1}$$

is known as the closed-loop transfer function.

*Example 4.4* Consider the block diagram shown in Fig. 4.4a. To simplify this block diagram, it is convenient to shift the subtraction point placed at the right of $U(s)$ to the point where $I^*(s)$ appears. This must be accomplished by keeping the definition of $I(s)$ in Fig. 4.4a without change. This means that, according to Fig. 4.4a:

$$I(s) = \frac{1}{Ls + R}[KA_p(I^*(s) - I(s)) - nk_e s\theta(s)].$$

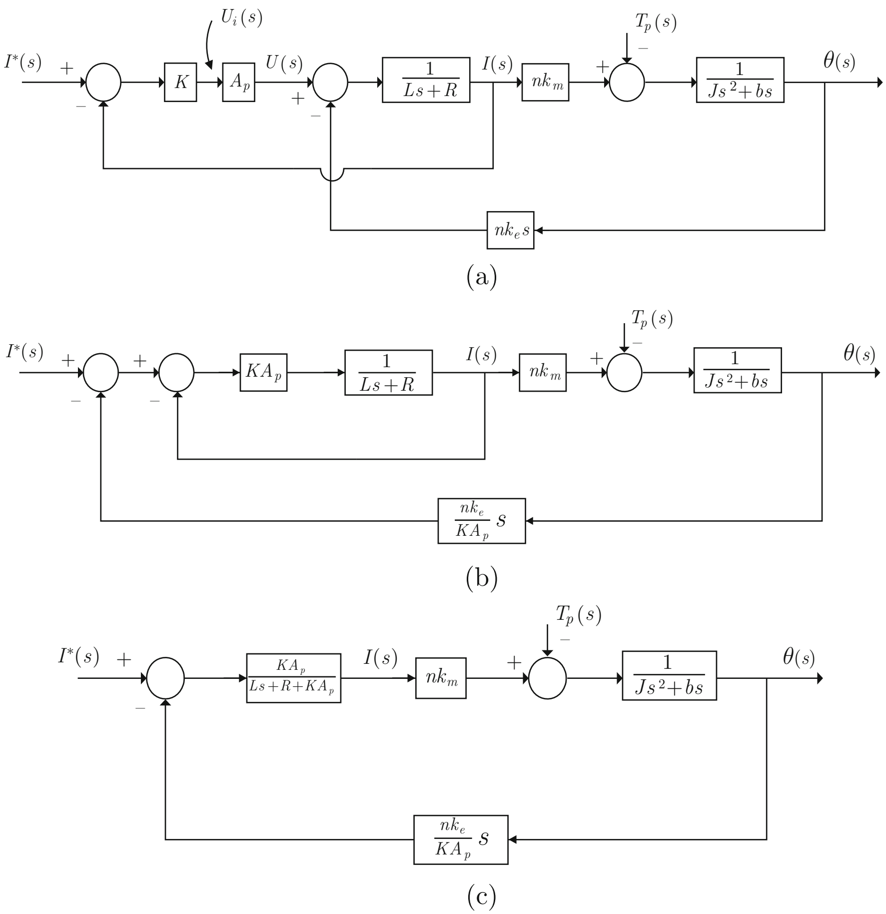



**Fig. 4.4** Block diagram simplification

Notice that this expression is also valid in the block diagram of Fig. 4.4b; hence, this block diagram is equivalent to that in Fig. 4.4a. It is clear that the loop between the second subtraction point in Fig. 4.4b and $I(s)$ is identical to the closed-loop system shown in Fig. 4.3 with:

$$G(s) = \frac{KA_p}{Ls + R}, \quad H(s) = 1.$$

It is clear that the main result in Example 4.1 has been employed to compute the transfer function of two cascaded systems as the product of the individual transfer functions of the cascaded systems. Hence, using (4.1), the following transfer function is found, which must be placed before $I(s)$:

$$\frac{KA_p}{Ls + R + KA_p}.$$

This is shown in Fig. 4.4c. The block diagram in Fig. 4.4c has two inputs. To represent the output $\theta(s)$ as a function of both inputs, the superposition principle has to be employed (see Sect. 3.7), i.e.,

$$\theta(s) = G_1(s)I^*(s) + G_2(s)T_p(s). \tag{4.2}$$

$G_1(s)$ is the transfer function computed when $I^*(s)$ is the input and $\theta(s)$ is the output, and it is assumed that $T_p(s) = 0$ in the block diagram in Fig. 4.4c, i.e., when the block diagram is depicted as in Fig. 4.5a. Thus, using (4.1), the following is defined:

$$M(s) = \frac{G(s)}{1 + G(s)H(s)} = \frac{\theta(s)}{I^*(s)} = G_1(s),$$

with:

$$G(s) = \frac{KA_p n k_m}{(Ls + R + KA_p)(Js^2 + bs)}, \quad H(s) = \frac{nk_e s}{KA_p},$$

to find:

$$G_1(s) = \frac{nk_m}{\left[\left(\frac{sL+R}{KA_p} + 1\right)(sJ + b) + \frac{n^2 k_m k_e}{KA_p}\right]s}. \tag{4.3}$$

On the other hand, $G_2(s)$ is the transfer function obtained when $T_p(s)$ is the input, $\theta(s)$ is the output and it is assumed that $I^*(s) = 0$ in the block diagram in Fig. 4.4c, i.e., when block diagram is depicted as in Fig. 4.5b. Hence, using (4.1), the following is defined:

$$M(s) = \frac{-G(s)}{1 + G(s)H(s)} = \frac{\theta(s)}{T_p(s)} = G_2(s),$$

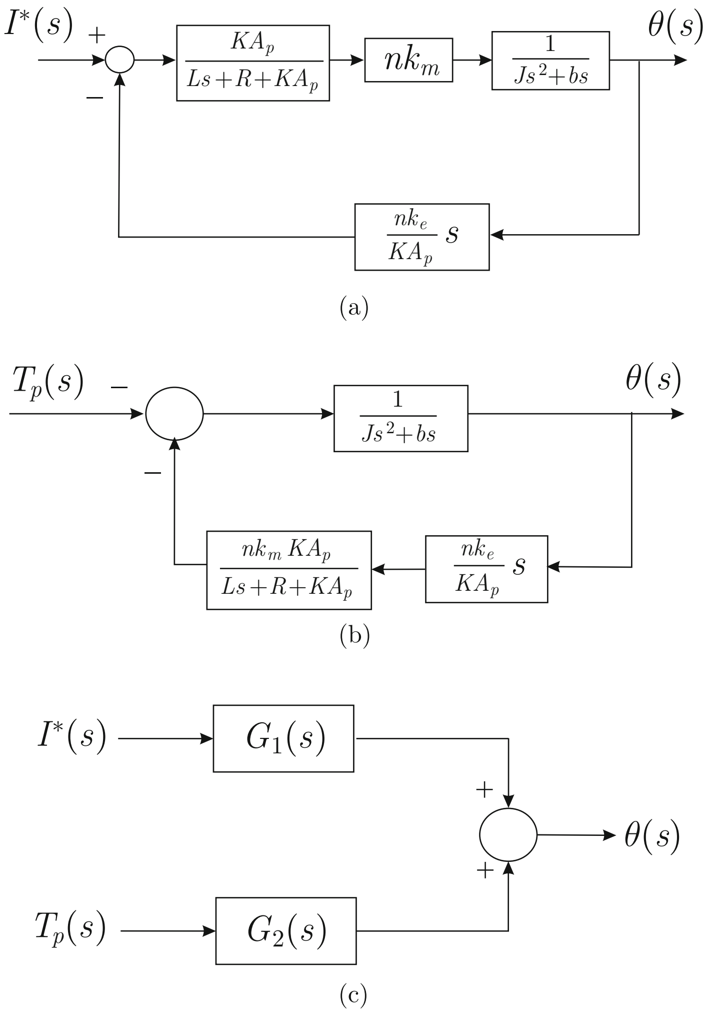



**Fig. 4.5** Simplifying the block diagram in Fig. 4.4c . (**a**) Block diagram when $T_p(s) = 0$. (**b**) Block diagram when $I^*(s) = 0$. (**c**) Block diagram equivalent to any of the block diagrams in Fig. 4.4

with:

$$G(s) = \frac{1}{Js^2 + bs}, \quad H(s) = \frac{n^2 k_m k_e s}{Ls + R + KA_p},$$

to obtain:

$$G_2(s) = \frac{-\left(\frac{sL+R}{KA_p} + 1\right)}{\left[\left(\frac{sL+R}{KA_p} + 1\right)(sJ + b) + \frac{n^2 k_m k_e}{KA_p}\right] s}. \tag{4.4}$$

Thus, any of block diagrams in Figs. 4.4a, b or c can be represented as block diagram in Fig. 4.5c with $G_1(s)$ and $G_2(s)$ given in (4.3) and (4.4).

*Example 4.5* Consider the closed-loop system shown in Fig. 4.6a. This block diagram represents a control scheme known as *two degrees of freedom control*. This is a closed-loop system with two inputs and one output. Hence, we can proceed as in the previous example to write, using the superposition principle:

$$C(s) = G_1(s)R(s) + G_2(s)D(s).$$

$G_1(s)$ is the transfer function obtained when using $R(s)$ as the input, $C(s)$ as the output and $D(s) = 0$, i.e., when the block diagram in Fig. 4.6b is employed. On the other hand, $G_2(s)$ is the transfer function obtained when using $D(s)$ as the input, $C(s)$ as the output and $R(s) = 0$, i.e., when the block diagram in Fig. 4.6c is employed.

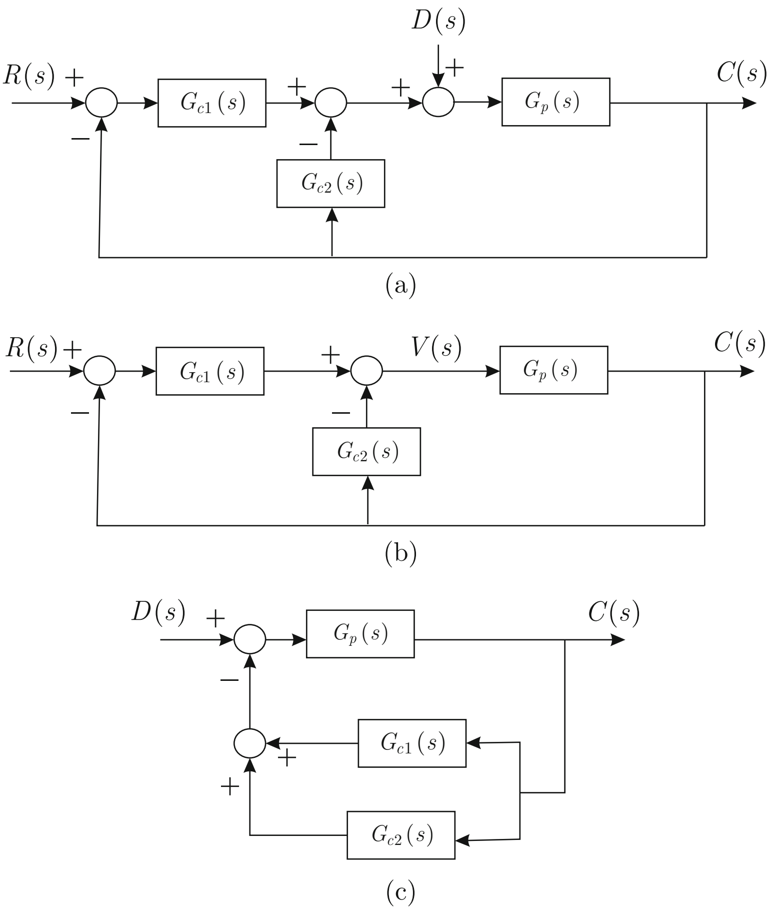



**Fig. 4.6** A two-degree-of-freedom control system. (**a**) Closed-loop system. (**b**) Case when $D(s) = 0$. (**c**) Case when $R(s) = 0$

First, proceed to obtain $G_1(s)$. To simplify the block diagram in Fig. 4.6b notice that:

$$
\begin{aligned}
V(s) &= G_{c1}(s)(R(s) - C(s)) - G_{c2}(s)C(s),\\
&= G_{c1}(s)(R(s) - C(s)) + G_{c2}(s)(R(s) - C(s)) - G_{c2}(s)R(s),\\
&= (G_{c1}(s) + G_{c2}(s))(R(s) - C(s)) - G_{c2}(s)R(s).
\end{aligned}
$$

Then, the block diagrams in Figs. 4.7a, b, and c are successively obtained. Notice that, according to the Example 4.2, the following can be written:

$$
\begin{aligned}
F(s) &= \left[1 - \frac{G_{c2}(s)}{G_{c1}(s) + G_{c2}(s)}\right] R(s),\\
&= \frac{G_{c1}(s) + G_{c2}(s) - G_{c2}(s)}{G_{c1}(s) + G_{c2}(s)} R(s),\\
&= \frac{G_{c1}(s)}{G_{c1}(s) + G_{c2}(s)} R(s).
\end{aligned}
\tag{4.5}
$$

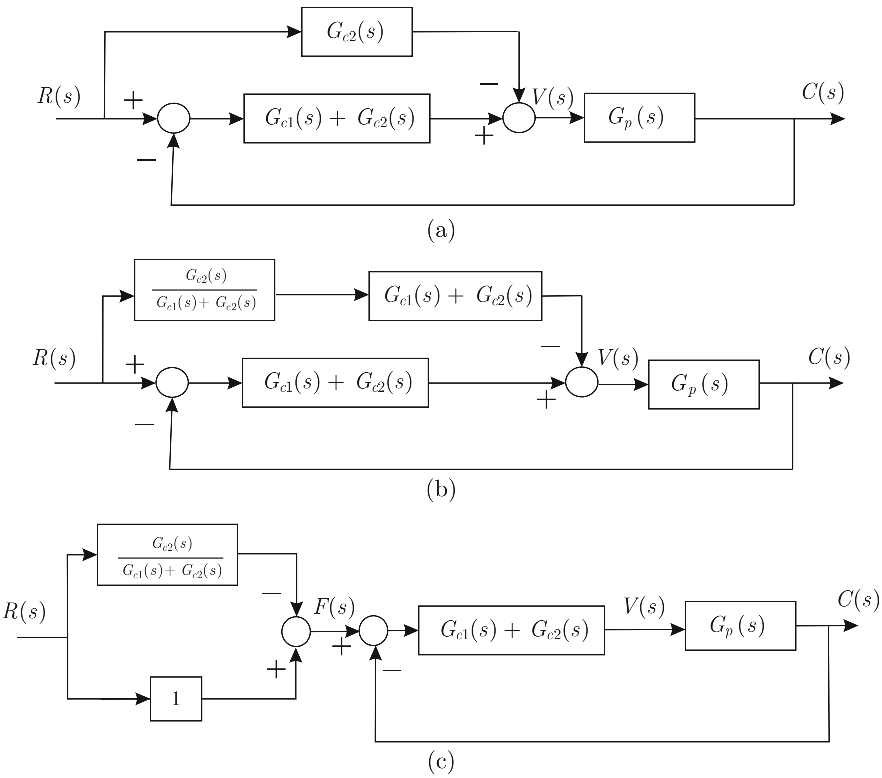



**Fig. 4.7** Simplifying the block diagram in Fig. 4.6b

On the other hand, (4.1) can be used together with:

$$G(s) = (G_{c1}(s) + G_{c2}(s))G_p(s), \quad H(s) = 1,$$

to find:

$$C(s) = \frac{(G_{c1}(s) + G_{c2}(s))G_p(s)}{1 + (G_{c1}(s) + G_{c2}(s))G_p(s)} F(s). \tag{4.6}$$

Replacing (4.5) in (4.6), the following is found:

$$C(s) = \frac{G_{c1}(s)G_p(s)}{1 + (G_{c1}(s) + G_{c2}(s))G_p(s)} R(s),$$

hence:

$$G_1(s) = \frac{G_{c1}(s)G_p(s)}{1 + (G_{c1}(s) + G_{c2}(s))G_p(s)}. \tag{4.7}$$

On the other hand, from the block diagram in Fig. 4.6c and according to Example 4.2, the block diagram in Fig. 4.8 is obtained. Using (4.1) and:

$$G(s) = G_p(s), \quad H(s) = G_{c1}(s) + G_{c2}(s),$$

yields:

$$C(s) = \frac{G_p(s)}{1 + (G_{c1}(s) + G_{c2}(s))G_p(s)} D(s),$$

i.e.,

$$G_2(s) = \frac{G_p(s)}{1 + (G_{c1}(s) + G_{c2}(s))G_p(s)}. \tag{4.8}$$

Hence, the block diagram in Fig. 4.6a can be represented as the block diagram in Fig. 4.9 with $G_1(s)$ and $G_2(s)$ given in (4.7) and (4.8).

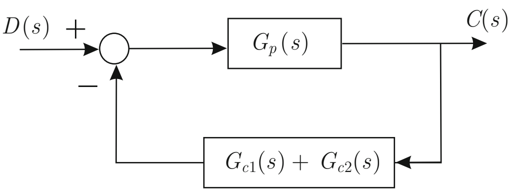



**Fig. 4.8** Simplifying block diagram in Fig. 4.6c

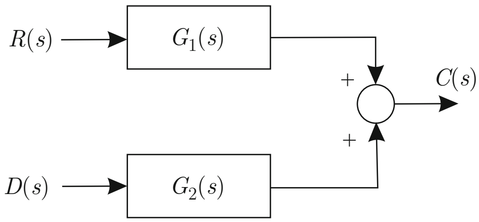



**Fig. 4.9** Equivalent block diagram to that in Fig. 4.6a

## 4.2 The Rule of Signs

According to Sect. 3.4, the stability of a transfer function is ensured if all its poles have a negative real part, i.e., if all the roots of its characteristic polynomial have a negative real part. However, sometimes, it is difficult to compute the exact value of the poles and this is especially true when the polynomial has a degree that is greater than or equal to 3. The reason for this is that the procedure to compute the roots of third- or fourth-degree polynomials is complex. Moreover, for polynomials with a degree greater than or equal to 5, analytical procedures do not exist. Even for second-degree polynomials the analytical procedure may be tedious. Thus, it is desirable to have a simple method for determining whether all roots of a polynomial have a negative real part or whether there are some roots with a positive real part. Some simple criteria for solving this problem are introduced in this section. These criteria are based on the study of signs of the polynomial coefficients.

### *4.2.1 Second-Degree Polynomials*

**Criterion 4.1** *If all the coefficients of a second-degree polynomial have the same sign, then all its roots have a negative real part.*

*Proof* Consider the following polynomial:

$$p(s) = s^2 + cs + d,$$

where $c > 0$, $d > 0$. If all the coefficients of the polynomial had a negative sign, then this sign can be factorized to proceed as shown in the following, i.e., when all the coefficients are positive. A second-degree polynomial $p(s)$ has two roots, which can be computed using the general formula:

$$s_{1,2} = \frac{-c \pm \sqrt{c^2 - 4d}}{2}. \tag{4.9}$$

- Case $(i)$: $c^2 - 4d < 0$.

In this case, both roots are complex conjugate with a negative real part because $c > 0$:

$$s_{1,2} = \frac{-c}{2} \pm \frac{\sqrt{4d - c^2}}{2} j.$$

- Case $(ii)$: $c^2 - 4d > 0$.
  In this case,

$$c^2 > c^2 - 4d > 0,$$

because $d > 0$. Applying the square root on both sides yields:

$$c > \sqrt{c^2 - 4d},$$

because the square root is a strictly increasing function. Then:

$$c - \sqrt{c^2 - 4d} > 0.$$

According to (4.9), this means that both roots are real, different, and negative in this case:

$$s_1 = -\frac{c + \sqrt{c^2 - 4d}}{2} < 0,$$
$$s_2 = -\frac{c - \sqrt{c^2 - 4d}}{2} < 0.$$

- Case $(iii)$: $c^2 - 4d = 0$.
  In this case, both roots are real, repeated, and negative:

$$s_{1,2} = \frac{-c}{2}.$$

*Example 4.6* Notice that the polynomial $s^2 + s + 1$ satisfies the case $(i)$; hence, its roots are complex conjugate with a negative real part. In fact, it is not difficult to verify that these roots are $-0.5 + j0.866$ and $-0.5 - j0.866$. On the other hand, the polynomial $s^2 + 2.5s + 1$ satisfies the case $(ii)$; hence, its roots are real, different, and negative. It is not difficult to verify that these roots are $-2$ and $-0.5$. Finally, the polynomial $s^2 + 2s + 1$ satisfies the case $(iii)$; hence, its roots are real, repeated, and negative. It is not difficult to verify that these roots are $-1$ and $-1$.

**Criterion 4.2** *If some coefficients of a second-degree polynomial have a sign that is different to signs of the other coefficients, then at least one root has a positive real part.*

*Proof* Consider the following polynomial:

$$p(s) = s^2 + cs + d.$$

There are three possibilities:

1) $c < 0, d > 0$.
2) $c > 0, d < 0$.
3) $c < 0, d < 0$.

Both roots are computed using the general formula:

$$s_{1,2} = \frac{-c \pm \sqrt{c^2 - 4d}}{2}.$$

- Case $(i)$: $c^2 - 4d < 0$.
  In this case, both roots are complex conjugate with a positive real part if 1) is satisfied. Cases 2) and 3) are not possible because $d < 0$ implies $c^2 - 4d > 0$.
- Case $(ii)$: $c^2 - 4d > 0$.
  This case is possible for 2) and 3) and some small values for $d > 0$ in 1). For larger values of $d > 0$, $i)$ is retrieved. Consider cases 2) and 3), then:

  $$c^2 < c^2 - 4d.$$

  Applying the square root on both sides of this inequality yields:

  $$abs(c) < \sqrt{c^2 - 4d},$$

  because the square root is a strictly increasing function. Then:

  $$abs(c) - \sqrt{c^2 - 4d} < 0, \quad \text{or} \quad 0 < -abs(c) + \sqrt{c^2 - 4d}.$$

  This means that both roots are real and different, with one of them positive:

  $$s_1 = -\frac{c + \sqrt{c^2 - 4d}}{2} < 0,$$
  $$s_2 = -\frac{c - \sqrt{c^2 - 4d}}{2} > 0.$$

  Now consider case 1):

  $$c^2 > c^2 - 4d.$$

Applying the square root on both sides of this inequality yields:

$$abs(c) > \sqrt{c^2 - 4d},$$

i.e.,

$$abs(c) - \sqrt{c^2 - 4d} > 0, \quad \text{or} \quad -abs(c) + \sqrt{c^2 - 4d} < 0.$$

As $c < 0$ and $d > 0$, then there are two positive real roots:

$$s_1 = -\frac{c + \sqrt{c^2 - 4d}}{2} > 0,$$
$$s_2 = -\frac{c - \sqrt{c^2 - 4d}}{2} > 0.$$

- Case $(iii)$: $c^2 - 4d = 0$.
  In this case $d = c^2/4 > 0$, only if 1) is satisfied, i.e., $c < 0$. This implies that both roots are real, repeated, and positive:

$$s_{1,2} = \frac{-c}{2} > 0.$$

*Example 4.7* Several second-degree polynomials and their corresponding roots are presented in what follows. It is left as an exercise for the reader to verify these results and to relate them to the cases analyzed above.

- $s^2 + 2s - 1$, roots: $-2.4142$ and $0.4142$.
- $s^2 - 2.5s + 1$, roots: 2 and 0.5.
- $s^2 - s - 1$, roots: 1.618 and $-0.618$.
- $s^2 - s + 1$, roots: $0.5 + j0.866$ and $0.5 - j0.866$.
- $s^2 - 2s + 1$, roots: 1 and 1.

### 4.2.2 First-Degree Polynomials

In this case, is very easy to prove that the same conditions stand as for second-degree polynomials:

**Criterion 4.3** *If both coefficients of a first-order polynomial have the same sign, then its only root is real and negative. If one coefficient has the opposite sign to the other, then the only root is real and positive.*

### 4.2.3 Polynomials with a Degree Greater Than or Equal to 3

**Criterion 4.4** *If at least one coefficient has the opposite sign to the other coefficients, then there is at least one root with a positive real part.*

*Proof* Consider the following polynomial where $n \geq 3$:

$$p(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 = (s - p_1)(s - p_2)\cdots(s - p_n). \tag{4.10}$$

According to the study of second-degree polynomials, it is known that the product $(s - p_i)(s - p_j) = s^2 + cs + d$ has $c < 0$ and/or $d < 0$ if at least one of the roots $p_i$ or $p_j$ has a positive real part. On the other hand, the first-degree polynomial $(s - p_k)$ has a negative coefficient if its root $p_k$ is positive. Also notice that the coefficients $a_j$, $j = 0, \ldots, n-1$, on the left-hand side of (4.10) are obtained as the algebraic sum of the product of coefficients of polynomials of the form $s^2 + cs + d$ and $(s - p_k)$ existing on the right-hand side of (4.10). Then, the only way for a coefficient $a_j < 0$ to exist on the left-hand side of (4.10) is that at least one root with a positive real part exists on the right-hand side of (4.10), i.e., that some of the polynomials $s^2 + cs + d$ or $(s - p_k)$ have a negative coefficient.

**Criterion 4.5** *Even when all the coefficients have the same sign, it is not a given that all roots have a negative real part.*

This can be explained using the same arguments as for the proof of the previous criterion, recalling that the product of two negative coefficients results in a positive number. Then, even when all the coefficients of the polynomial on the left-hand side of (4.10) are positive, there is the possibility that two negative coefficients (owing to roots with positive real parts) on the right-hand side of (4.10) multiply to give a positive coefficient on the left-hand side of (4.10).

*Example 4.8* To illustrate what happens in a polynomial with a degree greater than 2, some polynomials and their roots are presented in the following. The reader can corroborate these results by checking that $s^3 + a_2 s^2 + a_1 s + a_0 = (s - p_1)(s - p_2)(s - p_3)$, where $p_1$, $p_2$, $p_3$ are the roots of the corresponding polynomial.

- $s^3 + s^2 + s + 1.5$, roots: $-1.2041$, $0.102 + j1.1115$ and $0.102 - j1.1115$. There are two roots with a positive real part, despite all the polynomial coefficients having the same sign.
- $s^3 - s^2 + s + 1$, roots: $0.7718 + j1.1151$, $0.7718 - j1.1151$ and $-0.5437$. There are two roots with a positive real part because one coefficient of the polynomial has the opposite sign to the other coefficients.
- $s^3 + s^2 + 1$, roots: $-1.4656$, $0.2328 + j0.7926$, $0.2328 - j0.7926$. There are two roots with a positive real part because one coefficient has a zero value, despite the other coefficients having the same sign.

- $s^3 + s^2 + 3s + 1$, roots: $-0.3194 + j1.6332$, $-0.3194 - j1.6332$ and $-0.3611$. All roots have a negative real part and all the polynomial coefficients have the same sign.

## 4.3 Routh's Stability Criterion

As explained in the previous section, although the rule of signs is very easy to apply, its main drawback appears when analyzing polynomials with a degree greater than or equal to 3: when all coefficients have the same sign, nothing can be concluded from the sign of the real part of the roots. The solution to this problem is provided by Routh's stability criterion, which, given a polynomial with an arbitrary degree, establishes necessary and sufficient conditions to ensure that all its roots have a negative real part.

Given an arbitrary polynomial, order its terms by descending powers of its variable:

$$a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0. \tag{4.11}$$

Routh's stability criterion is applied in two steps [2–5]:

1. Fill Table 4.1. Notice that the first two rows of the table are filled by direct substitution of the polynomial coefficients. Also notice that the last entries in the first two rows are zero, as that is the value of the coefficients of powers not appearing in the polynomial in (4.11). The entries of the remaining rows are computed using the following formulae:

$$b_1 = \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}}, \quad b_2 = \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}},$$
$$b_3 = \frac{a_{n-1}a_{n-6} - a_n a_{n-7}}{a_{n-1}},$$
$$\vdots$$
$$c_1 = \frac{b_1 a_{n-3} - a_{n-1} b_2}{b_1}, \quad c_2 = \frac{b_1 a_{n-5} - a_{n-1} b_3}{b_1}, \quad c_3 = \frac{b_1 a_{n-7} - a_{n-1} b_4}{b_1},$$
$$\vdots$$
$$d_1 = \frac{c_1 b_2 - b_1 c_2}{c_1}, \quad d_2 = \frac{c_1 b_3 - b_1 c_3}{c_1}, \quad d_3 = \frac{c_1 b_4 - b_1 c_4}{c_1},$$
$$\vdots$$

   Notice that, according to these formulae, the last entry in each row is zero and the table is triangular.

**Table 4.1** Application of Routh's stability criterion

| | | | | | | |
|---|---|---|---|---|---|---|
| $s^n$ | $a_n$ | $a_{n-2}$ | $a_{n-4}$ | $a_{n-6}$ | $\dots$ | 0 |
| $s^{n-1}$ | $a_{n-1}$ | $a_{n-3}$ | $a_{n-5}$ | $a_{n-7}$ | $\dots$ | 0 |
| $s^{n-2}$ | $b_1$ | $b_2$ | $b_3$ | $b_4$ | $\dots$ | 0 |
| $s^{n-3}$ | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $\dots$ | 0 |
| $s^{n-4}$ | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $\dots$ | 0 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\dots$ | 0 |
| $s^2$ | $p_1$ | $p_2$ | 0 | | | |
| $s^1$ | $q_1$ | 0 | | | | |
| $s^0$ | $p_2$ | | | | | |

**Table 4.2** Application of Routh's criterion to the polynomial $s^2 + as + b$

| | | | |
|---|---|---|---|
| $s^2$ | 1 | $b$ | 0 |
| $s^1$ | $a$ | 0 | |
| $s^0$ | $b$ | | |

2. Only analyze the first column in Table 4.1. The number of sign changes found from top to bottom in the first column of the table is equal to to the number of roots with a positive real part.

Several examples are presented in the following.

*Example 4.9* Consider the following second-degree polynomial:

$$s^2 + as + b,$$

with $a$ and $b$ two real constants. The rule of signs (Sect. 4.2) can be employed to conclude that both roots have a negative real part if:

$$a > 0, \quad b > 0,$$

as the coefficient of $s^2$ is $+1$. Routh's criterion is now used to corroborate this result. First, Table 4.2 is filled. Routh's criterion establishes that the number of sign changes in the first column of the table is equal to the number of roots with a positive real part. Hence, if there are no sign changes, both roots have a negative real part. As the first entry of the first column is $+1$ then:

$$a > 0, \quad b > 0,$$

must be true to ensure that both roots have a negative real part. Thus, Routh's criterion and the rule of signs in Sect. 4.2 have the same conclusion, as expected.

*Example 4.10* Consider the following polynomial:

$$s^3 + 5s^2 + 2s - 8.$$

**Table 4.3** Application of Routh's criterion to the polynomial $s^3+5s^2+2s-8$

| | | | |
|---|---|---|---|
| $s^3$ | 1 | 2 | 0 |
| $s^2$ | 5 | $-8$ | 0 |
| $s^1$ | $\frac{5\times 2-1\times(-8)}{5}=3.6$ | 0 | |
| $s^0$ | $-8$ | | |

**Table 4.4** Application of Routh's criterion to the polynomial $s^3+1.8s^2+0.61s+2.02$

| | | | |
|---|---|---|---|
| $s^3$ | 1 | 0.61 | 0 |
| $s^2$ | 1.8 | 2.02 | 0 |
| $s^1$ | $\frac{1.8\times 0.61-1\times 2.02}{1.8}=-0.512$ | 0 | |
| $s^0$ | 2.02 | | |

Because of coefficient $-8$, i.e., with a negative sign when all the other coefficients are positive, it is known from the rule of signs in Sect. 4.2 that there is at least one root with a positive real part. Routh's criterion is now used to corroborate this result. First, Table 4.3 is filled. Routh's criterion establishes that the number of sign changes in the first column is equal to the number of roots with a positive real part. Hence, it is concluded that the polynomial $s^3+5s^2+2s-8$ has one root with a positive real part. In fact, the use of the MATLAB command:

```
roots([1 5 2 -8])
```

allows us to find that:

$$s=-2,\quad s=1,\quad s=-4,$$

are the roots of the polynomial.

*Example 4.11* Consider the following polynomial:

$$s^3+1.8s^2+0.61s+2.02.$$

As all the coefficients have the same sign and the polynomial is third-degree, the rule of signs in Sect. 4.2 is not useful in this problem. In this case, Routh's stability criterion is a suitable alternative. First, Table 4.4 is filled. Routh's criterion establishes that the number of sign changes in the first column is equal to the number of roots with a positive real part. Hence, it is concluded that the polynomial $s^3+1.8s^2+0.61s+2.02$ has two roots with a positive real part. In fact, use of the MATLAB command:

```
roots([1 1.8 0.61 2.02])
```

allows us to find that:

$$s=-2,\quad s=0.1+j,\quad s=0.1-j,$$

are the roots of the above polynomial.

**Table 4.5** Application of Routh's criterion to the polynomial $s^3+as^2+bs+c$

| | | | |
|---|---|---|---|
| $s^3$ | 1 | $b$ | 0 |
| $s^2$ | $a$ | $c$ | 0 |
| $s^1$ | $\frac{ab-1\times c}{a}$ | 0 | |
| $s^0$ | $c$ | | |

*Example 4.12* Consider the following polynomial:

$$s^3+as^2+bs+c,$$

where $a$, $b$ and $c$ are real unknown constants. This type of situation is very common in control systems where coefficients of the characteristic polynomial depend on the controller gains, which are not known and must be found such that closed-loop stability is ensured. This is achieved by forcing all the roots of the closed-loop characteristic polynomial to have a negative real part. However, the computation of the roots of a third-degree polynomial using analytical methods is a very tedious process; hence, that is not a practical solution. In this type of problem, Routh's stability criterion is, again, very useful. First, fill Table 4.5. Routh's criterion establishes that the number of sign changes in the first column of the table is equal to the number of roots with a positive real part. Thus, as the first element in the first column is $+1$, then if:

$$a>0, \quad \frac{ab-c}{a}>0 \quad c>0, \tag{4.12}$$

it is ensured that all three roots have a negative real part. Notice that conditions in (4.12) are equivalent to:

$$a>0, \quad b>\frac{c}{a}>0 \quad c>0.$$

This means that, although the coefficients $a$ and $c$ are only required to be positive, in the case of the coefficient $b$, this is not enough and a little more is demanded: $b>\frac{c}{a}>0$.

*Example 4.13 (Special Case: Only the Entry at the First Column of a Row Is Equal to Zero)* Consider the following polynomial:

$$s^5+2s^4+3s^3+6s^2+5s+3.$$

First fill Table 4.6. However, the procedure stops at the row corresponding to $s^3$ because the entry in the first column in that row is zero. As a consequence, some divisions by zero would appear when continuing filling the remaining rows. In this case, it is advised to replace the zero in the first column with a small $\varepsilon>0$ to continue filling the table as shown in Table 4.7.

**Table 4.6** Application of Routh's criterion to the polynomial $s^5+2s^4+3s^3+6s^2+5s+3$

| | | | |
|---|---|---|---|
| $s^5$ | 1 | 3 | 5 |
| $s^4$ | 2 | 6 | 3 |
| $s^3$ | $\frac{2\times 3-1\times 6}{2}=0$ | $\frac{2\times 5-1\times 3}{2}=3.5$ | 0 |
| $s^2$ | | | |
| $s^1$ | | | |
| $s^0$ | | | |

**Table 4.7** Application of Routh's criterion to the polynomial $s^5+2s^4+3s^3+6s^2+5s+3$ (cont.)

| | | | |
|---|---|---|---|
| $s^5$ | 1 | 3 | 5 |
| $s^4$ | 2 | 6 | 3 |
| $s^3$ | $\varepsilon$ | 3.5 | 0 |
| $s^2$ | $\frac{\varepsilon\times 6-2\times 3.5}{\varepsilon}$ | $\frac{\varepsilon\times 3-2\times 0}{\varepsilon}=3$ | 0 |
| $s^1$ | $\frac{42\varepsilon-49-6\varepsilon^2}{12\varepsilon-14}$ | 0 | |
| $s^0$ | 3 | | |

Now, determine the number of sign changes in the first column as $\varepsilon>0$ tends toward zero. Notice that in this example there are two sign changes. This means that the polynomial under study has two roots with a positive real part. In fact, use of the MATLAB command:

```
roots([1 2 3 6 5 3])
```

allows us to find that the roots of the polynomial $s^5+2s^4+3s^3+6s^2+5s+3$ are:

$$s=0.3429+j1.5083,\quad s=0.3429-j1.5083,\quad s=-1.6681,$$
$$s=-0.5088+j0.7020,\quad s=-0.5088-j0.7020.$$

If after $\varepsilon>0$ is employed to fill the table, there were no sign changes in the first column, then the system would be marginally stable, i.e., there would be some roots on the imaginary axis.

*Example 4.14 (Special Case: A Row Is Filled Exclusively with Zeros or a Row only Has One Entry, Which Is Zero)* Consider the following polynomial:

$$s^3+3s^2+s+3. \tag{4.13}$$

First fill Table 4.8. The process stops at the row corresponding to $s^1$ because that row is filled exclusively with zeros. This is a special case that is indicative of three possibilities [3], pp. 336,[6], pp. 136:

1. There are real roots that are placed symmetrically with respect to origin.
2. There are imaginary roots that are placed symmetrically with respect to origin.
3. There are four complex conjugate roots placed at the vertices of a rectangle centered at the origin.

**Table 4.8** Application of Routh's criterion to the polynomial $s^3+3s^2+s+3$

| | | | |
|---|---|---|---|
| $s^3$ | 1 | 1 | 0 |
| $s^2$ | 3 | 3 | 0 |
| $s^1$ | $\frac{3\times 1-1\times 3}{3}=0$ | 0 | |
| $s^0$ | | | |

**Table 4.9** Application of Routh's criterion to the polynomial $s^3+3s^2+s+3$ (cont.)

| | | | |
|---|---|---|---|
| $s^3$ | 1 | 1 | 0 |
| $s^2$ | 3 | 3 | 0 |
| $s^1$ | 6 | 0 | |
| $s^0$ | 3 | | |

To continue, use the entries in the row $s^2$ to form a new polynomial, i.e.,

$$P(s)=3s^2+3,$$

compute its derivative:

$$\frac{dP(s)}{ds}=6s,$$

and replace these coefficients in the row $s^1$, as shown in Table 4.9, to continue filling the table. As no sign changes exist in the first column, it is concluded that no roots exist with a positive real part; hence, only the case 2 above is possible: there are two imaginary roots that are placed symmetrically with respect to origin and the remaining root has a negative real part. Furthermore, another important feature of this special case is the following:

"Form a polynomial with entries of the row above the row filled exclusively with zeros. The roots of such a polynomial are also the roots of the polynomial in (4.13)."

This means that the roots of polynomial $P(s)=3s^2+3$ are also roots of $s^3+3s^2+s+3$. These roots can be obtained as:

$$3s^2+3=0, \quad \Rightarrow s=\pm j.$$

Hence, the polynomial $s^3+3s^2+s+3$ has one root at $s=j$ and another at $s=-j$. In fact, use of the MATLAB command:

```
roots([1 3 1 3])
```

allows us to find that the roots of $s^3+3s^2+s+3$ are:

$$s=-3, \quad s=j, \quad s=-j.$$

*Example 4.15 (Special Case: A Row Is Filled Exclusively with Zeros or a Row only Has One Entry, Which Is Zero)* Consider the following polynomial:

$$s^4+s^2+1.$$

**Table 4.10** Application of Routh's criterion to the polynomial $s^4 + s^2 + 1$

| | | | | |
|---|---|---|---|---|
| $s^4$ | 1 | 1 | 1 | 0 |
| $s^3$ | 0 | 0 | 0 | |
| $s^2$ | | | | |
| $s^1$ | | | | |
| $s^0$ | | | | |

**Table 4.11** Application of Routh's criterion to the polynomial $s^4 + s^2 + 1$ (cont.)

| | | | | |
|---|---|---|---|---|
| $s^4$ | 1 | 1 | 1 | 0 |
| $s^3$ | 4 | 2 | 0 | |
| $s^2$ | $\frac{4\times1-1\times2}{4} = 0.5$ | $\frac{4\times1-1\times0}{4} = 1$ | | |
| $s^1$ | $\frac{0.5\times2-4\times1}{0.5} = -1.5$ | 0 | | |
| $s^0$ | 1 | | | |

Fill Table 4.10. As there is a row filled exclusively with zeros, proceed as in the previous example. The polynomial $P(s) = s^4 + s^2 + 1$ is differentiated with respect to $s$:

$$\frac{dP(s)}{ds} = 4s^3 + 2s.$$

These coefficients are substituted in the row filled exclusively with zeros and continue filling Table 4.11. As two sign changes exist in the first column of this table, it is concluded that two roots with a positive real part exist. This means that only the following case is possible: there are four complex conjugate roots placed on vertices of a rectangle centered at the origin. In fact, use of the MATLAB command:

```
roots([1 0 1 0 1])
```

allows us to find that the roots of the polynomial $s^4 + s^2 + 1$ are:

$$s = -0.5 \pm j0.8660, \quad s = 0.5 \pm j0.8660.$$

*Example 4.16 (Repeated Roots on the Imaginary Axis)* When a characteristic polynomial has repeated roots on the imaginary axis, the transfer function is unstable. However, this class of instability is not detected by Routh's criterion. For instance, consider the following polynomial:

$$s^4 + 2s^2 + 1 = [(s + j)(s - j)]^2.$$

Notice that there are two repeated roots on the imaginary axis. Fill Table 4.12. As there is a row exclusively filled with zeros, the polynomial $P(s) = s^4 + 2s^2 + 1$ is differentiated with respect to $s$:

$$\frac{dP(s)}{ds} = 4s^3 + 4s,$$

and continue filling Table 4.13.

**Table 4.12** Application of Routh's criterion to the polynomial $s^4+2s^2+1$

| | | | | |
|---|---|---|---|---|
| $s^4$ | 1 | 2 | 1 | 0 |
| $s^3$ | 0 | 0 | 0 | |
| $s^2$ | | | | |
| $s^1$ | | | | |
| $s^0$ | | | | |

**Table 4.13** Application of Routh's criterion to the polynomial $s^4+2s^2+1$ (cont.)

| | | | | |
|---|---|---|---|---|
| $s^4$ | 1 | 2 | 1 | 0 |
| $s^3$ | 4 | 4 | 0 | |
| $s^2$ | $\frac{4\times 2-4\times 1}{4}=1$ | $\frac{4\times 1-1\times 0}{4}=1$ | 0 | |
| $s^1$ | 0 | 0 | | |
| $s^0$ | | | | |

**Table 4.14** Application of Routh's criterion to the polynomial $s^4+2s^2+1$ (cont.)

| | | | | |
|---|---|---|---|---|
| $s^4$ | 1 | 2 | 1 | 0 |
| $s^3$ | 4 | 4 | 0 | |
| $s^2$ | 1 | 1 | 0 | |
| $s^1$ | 2 | 0 | | |
| $s^0$ | 1 | | | |

As there is another row exclusively filled with zeros, the polynomial $P_1(s) = s^2+1$ is differentiated with respect to $s$:

$$\frac{dP_1(s)}{ds} = 2s,$$

and continue filling Table 4.14. Notice that no sign changes exist in the first column of Table 4.14; hence, the method correctly indicates that no roots exist with a positive real part. Also notice that the roots of $P_1(s) = s^2+1$ are $s = \pm j$, i.e., marginal stability is concluded. However, this is incorrect because the polynomial $s^4+2s^2+1$ has two imaginary roots, which are repeated twice, which implies instability. Thus, attention must be paid to these cases.

## 4.4 Steady-State Error

The solution of a differential equation is given as the addition of the natural response and the forced response. If the differential equation is stable or, equivalently, if the transfer function is stable then the natural response disappears as time increases and the complete solution reaches the forced response. Recall that the forced response depends on (or is similar to) the applied input. According to these ideas, in a closed-loop control system, we have the following. 1) The closed-loop system input is a signal that stands for the desired closed-loop output. This is why such a signal is known as the reference or desired output. 2) The controller is designed such that the closed-loop system is stable and the forced response is equal to or is close to the reference or desired output.

The properties that a closed-loop control system must possess to ensure that the forced response is equal to or close enough to the closed-loop system input are studied in this section. Notice that this is equivalent to requiring the system error, defined as the difference between the system output and the reference, to be zero or close to zero in a steady state, i.e., when time is large. Thus, the behavior of the steady-state error is studied in the following.

Consider the closed-loop system with unit feedback shown in Fig. 4.10. An important concept in the study of the steady-state error is the *system type*, which is defined next. It is assumed that any $n-$order open-loop transfer function can be written as:

$$G(s) = \frac{k(s - z_1)(s - z_2)\cdots(s - z_m)}{s^N(s - p_1)(s - p_2)\cdots(s - p_{n-N})},$$

where $n$ is the number of open-loop poles and $m$ is the number of open-loop zeros with $n > m$, $z_j \neq 0$, $j = 1, \ldots, m$ and $p_i \neq 0$, $i = 1, \ldots, n - N$. Hence, $N$ is the number of poles that the open-loop transfer function has at the origin, i.e., after cancelling with any zeros that the open-loop transfer function might have at the origin. The *system type* is defined as $N$, i.e., it is the number of poles at the origin or, equivalently, the number of integrators that the open-loop transfer function possesses.

The error signal is given as the difference between the closed-loop system output and the reference or desired output:

$$E(s) = R(s) - C(s) \quad \Rightarrow C(s) = R(s) - E(s),$$

and, on the other hand:

$$C(s) = G(s)E(s) = R(s) - E(s),$$

hence:

$$E(s)[1 + G(s)] = R(s),$$

$$E(s) = \frac{1}{1 + G(s)} R(s). \tag{4.14}$$

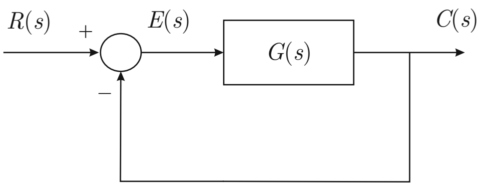



**Fig. 4.10** Closed-loop system with unitary feedback

The steady-state error, $e_{ss}$, is obtained using (4.14) and the final value theorem in (3.4), i.e.,

$$e_{ss} = \lim_{t\to\infty} e(t) = \lim_{s\to 0} sE(s) = \lim_{s\to 0} \frac{s}{1+G(s)} R(s). \tag{4.15}$$

Notice that the steady-state error is the difference between the system output $c(t)$ and the reference, or desired output, $r(t)$, in a steady state. This explains why $e_{ss}$ depends on the reference $R(s)$. Hence, to continue this study, it is necessary to know $R(s)$, which motivates the definition of the so-called *test signals*. A test signal is a function of time with two features: 1) It must make sense in real applications, and 2) It must be mathematically simple. Some important test signals, or test inputs, in classical control are defined in the following.

Consider the problem of a video camera tracking a target. The control objective in this problem is to force the video camera to aim at a target that moves very fast. Some situations arising in this problem are the following.

- **Step test signal**. Suppose that the target approaches the video camera directly along a constant direction represented by $A$. If video camera is at the beginning aiming in another direction and it is desired that it aims in the direction from which the target is approaching, then the reference or desired direction has the shape depicted in Fig. 4.11. This test signal is known as a *step* and it represents an abrupt change in the desired output. The system output is the direction in which the video camera is aiming. It is desired that the difference between these signals is zero or close to zero in a steady state. If $r(t)$ is a step:

$$r(t) = \begin{cases} 0, & t < 0 \\ A, & t \geq 0 \end{cases},$$

  where $A$ is a constant, then $R(s) = \mathcal{L}\{r(t)\} = \frac{A}{s}$ is a function that is simple enough to be mathematically handled.
- **Ramp test signal**. Suppose that the target passes in front of the video camera with a constant velocity $A$, approaching another point, and that the video camera is aiming at the target at the beginning. Then, the reference or desired direction has the shape shown in Fig. 4.12. This test signal is known as a *ramp* and it indicates

**Fig. 4.11** Step test signal

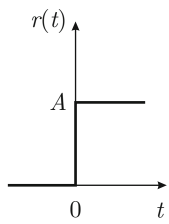


**Fig. 4.12** Ramp test signal

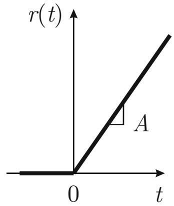
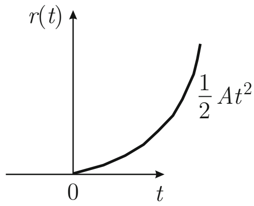



**Fig. 4.13** Parabola test signal




that the desired output changes at a constant rate $A$, i.e., the target velocity. If $r(t)$ is a ramp:

$$r(t) = \begin{cases} 0, & t < 0 \\ At, & t \geq 0 \end{cases},$$

where $A = \frac{dr(t)}{dt}$ is a constant standing for the rate of change of $r(t)$, then $R(s) = \mathcal{L}\{r(t)\} = \frac{A}{s^2}$ is a function that is simple enough to be mathematically handled.

- **Parabola test signal**. Suppose that the target passes in front of the video camera with a constant acceleration $A$, approaching another point, and that the video camera is aiming at the target at the beginning. Then, the reference or desired output has the shape shown in Fig. 4.13. This signal is known as a *parabola* and it indicates that the desired output changes with a constant acceleration $A$, i.e., the target acceleration. If $r(t)$ is a parabola:

$$r(t) = \begin{cases} 0, & t < 0 \\ \frac{1}{2}At^2, & t \geq 0 \end{cases},$$

where $A = \frac{d^2r(t)}{dt^2}$ is a constant standing for the acceleration of $r(t)$, then $R(s) = \mathcal{L}\{r(t)\} = \frac{A}{s^3}$ is a function that is simple enough to be mathematically handled.

Any of these three situations may appear under the normal operating conditions of a recording video camera control system. Thus, the design of a controller must ensure that the steady-state error is zero or close to zero for any of these desired outputs. The conditions required to achieve this control objective are studied in the following.

### 4.4.1 Step Desired Output

Suppose that $R(s) = A/s$ is the step desired output. Using (4.15) the following is found:

$$e_{ss} = \lim_{s\to 0} \frac{s}{1+G(s)} \frac{A}{s},$$
$$= \frac{A}{1+k_p}, \quad k_p = \lim_{s\to 0} G(s),$$

where $k_p$ is known as the *position constant* and is related to the system type. Hence, the following cases are considered:

- System type 0 ($N = 0$). In this case, the open-loop transfer function has the form:

  $$G(s) = \frac{k(s-z_1)(s-z_2)\cdots(s-z_m)}{(s-p_1)(s-p_2)\cdots(s-p_n)},$$

  i.e.,

  $$k_p = \lim_{s\to 0} G(s) = \frac{k(-z_1)(-z_2)\cdots(-z_m)}{(-p_1)(-p_2)\cdots(-p_n)},$$

  $k_p$ is finite. This means that the steady-state error is constant and different from zero:

  $$e_{ss} = \frac{A}{1+k_p} \neq 0.$$

- System type greater than or equal to 1 ($N \geq 1$). In this case, the open-loop transfer function has the form:

  $$G(s) = \frac{k(s-z_1)(s-z_2)\cdots(s-z_m)}{s^N(s-p_1)(s-p_2)\cdots(s-p_{n-N})},$$

  i.e.,

  $$k_p = \lim_{s\to 0} G(s) = \frac{k(-z_1)(-z_2)\cdots(-z_m)}{s^N(-p_1)(-p_2)\cdots(-p_{n-N})} \to \infty.$$

  This means that the steady-state error is zero:

  $$e_{ss} = \frac{A}{1+k_p} = 0.$$

### 4.4.2 Ramp Desired Output

Suppose that $R(s) = A/s^2$ is the desired output. Using (4.15), the following is found:

$$e_{ss} = \lim_{s \to 0} \frac{s}{1 + G(s)} \frac{A}{s^2},$$
$$= \frac{A}{k_v}, \quad k_v = \lim_{s \to 0} sG(s),$$

where $k_v$ is known as the *velocity constant* and it is related to the system type. Hence, consider the following cases:

- System type 0 ($N = 0$). In this case, the open-loop transfer function has the form:

$$G(s) = \frac{k(s - z_1)(s - z_2) \cdots (s - z_m)}{(s - p_1)(s - p_2) \cdots (s - p_n)},$$

  i.e.,

$$k_v = \lim_{s \to 0} sG(s) = (0) \frac{k(-z_1)(-z_2) \cdots (-z_m)}{(-p_1)(-p_2) \cdots (-p_n)} = 0.$$

  This means that the steady-state error grows without a limit:

$$e_{ss} = \frac{A}{k_v} \to \infty.$$

- System type 1 ($N = 1$). The open-loop transfer function has the form:

$$G(s) = \frac{k(s - z_1)(s - z_2) \cdots (s - z_m)}{s(s - p_1)(s - p_2) \cdots (s - p_{n-1})},$$

  i.e.,

$$k_v = \lim_{s \to 0} sG(s) = \frac{k(-z_1)(-z_2) \cdots (-z_m)}{(-p_1)(-p_2) \cdots (-p_{n-1})},$$

  $k_v$ is finite and different from zero. This means that the steady-state error is constant and different from zero:

$$e_{ss} = \frac{A}{k_v} \neq 0.$$

- System type greater than or equal to 2 ($N \geq 2$). The open-loop transfer function can be written as:

$$G(s) = \frac{k(s - z_1)(s - z_2) \cdots (s - z_m)}{s^N(s - p_1)(s - p_2) \cdots (s - p_{n-N})},$$

i.e.,

$$k_v = \lim_{s\to 0} sG(s) = \frac{k(-z_1)(-z_2)\cdots(-z_m)}{s^{N-1}(-p_1)(-p_2)\cdots(-p_{n-N})} \to \infty,$$

because $N-1 \geq 1$. This means that the steady-state error is zero:

$$e_{ss} = \frac{A}{k_v} = 0.$$

### 4.4.3 Parabola Desired Output

Suppose that $R(s) = A/s^3$ is the desired output. Using (4.15), the following is obtained:

$$\begin{aligned} e_{ss} &= \lim_{s\to 0} \frac{s}{1+G(s)} \frac{A}{s^3}, \\ &= \frac{A}{k_a}, \quad k_a = \lim_{s\to 0} s^2 G(s), \end{aligned}$$

where $k_a$ is known as the *acceleration constant* and is related to the system type. The open-loop transfer function is given as:

$$G(s) = \frac{k(s-z_1)(s-z_2)\cdots(s-z_m)}{s^N(s-p_1)(s-p_2)\cdots(s-p_{n-N})},$$

i.e.,

$$k_a = \lim_{s\to 0} s^2 G(s) = \frac{k(-z_1)(-z_2)\cdots(-z_m)}{s^{N-2}(-p_1)(-p_2)\cdots(-p_{n-N})}.$$

This means that:

- The steady-state error grows without limit (to infinity) when the system type is 0 or 1:

$$e_{ss} = \frac{A}{k_a} \to \infty, \quad N \leq 1.$$

- The steady-state error is constant and different from zero when the system type is 2:

$$e_{ss} = \frac{A}{k_a} \neq 0, \quad N = 2.$$

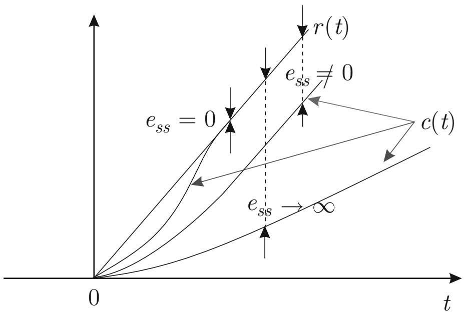



**Fig. 4.14** Behavior of the output $c(t)$ when the reference $r(t)$ is a ramp. Notice that $e_{ss} = 0$ when the system type is greater than or equal to 2, $e_{ss} \neq 0$ when the system type is 1 and $e_{ss} \to \infty$ when the system type is 0

**Table 4.15** Steady-state error

| | Step | Ramp | Parabola |
|---|---|---|---|
| Type 0 | $\frac{A}{1+k_p}$ | $\infty$ | $\infty$ |
| Type 1 | 0 | $\frac{A}{k_v}$ | $\infty$ |
| Type 2 | 0 | 0 | $\frac{A}{k_a}$ |
| Type 3 | 0 | 0 | 0 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |

- The steady-state error is zero when the system type is greater than or equal to 3:

$$e_{ss} = \frac{A}{k_a} = 0, \quad N \geq 3.$$

Notice that, in the case of the three references, or test signals, that have been considered, the steady-state error becomes zero if the number of open-loop integrators are suitably increased. This explains why some controllers include an integrator, i.e., proportional–integral (PI) or proportional–integral–derivative (PID) control. On the other hand, it is important to note that all the results presented above are true only if it is ensured that the closed-loop system is stable. This is because the final value theorem assumes that the natural response vanishes as time increases.

Finally, in Fig. 4.14, the behavior of the output is depicted $c(t)$ with respect to the reference $r(t)$, when the latter is a ramp for different system types. The results found above are summarized in Table 4.15.

*Example 4.17* A velocity control system for a permanent magnet brushed DC motor is depicted in Fig. 4.15 (see Chap. 10). The motor model is given by the transfer function $G_m(s) = \frac{k}{s+a}$, with $k > 0$ and $a > 0$, whereas $k_p$ is a positive constant standing for the transfer function of a proportional controller. This means that the electric current used as the control signal is computed as:

$$i^* = k_p(\omega_d - \omega),$$

where $\omega$ is the measured velocity, $\omega_d$ is the desired velocity and $I^*(s)$ is Laplace transform of $i^*$. The open-loop transfer function is:

$$G(s)H(s) = \frac{k_p k}{s + a}.$$

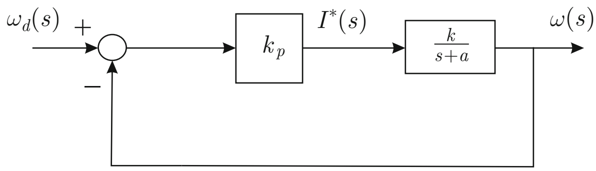



**Fig. 4.15** Proportional control of velocity

Notice that, as $a > 0$, the system type is 0 because the open-loop transfer function has no pole at the origin. Hence, according to Sect. 4.4.1, if the desired velocity is constant, i.e., a step signal, then the steady-state error $e_{ss} = \omega_d - \lim_{t\to\infty} \omega(t)$ is constant and different from zero, i.e., a proportional velocity controller is not useful for motor velocity to reach the desired velocity. See Fig. 3.32 for a simulation in which these observations are corroborated. However, it is important to stress that the steady-state error analysis that is presented in this chapter is useful for describing the situation in Fig. 3.32 only as long as no external disturbance $T_p$ is present. Notice that, according to Sects. 4.4.2 and 4.4.3, the steady-state error would be greater if the desired velocity was either a ramp or a parabola.

A PI controller solves this problem. A PI controller performs the following operation on the velocity error:

$$i^* = k_p e + k_i \int_0^t e(r)dr, \quad e = \omega_d - \omega,$$

where $k_p$ and $k_i$ are positive constants. Use of the Laplace transform yields:

$$\begin{aligned}
I^*(s) &= k_p E(s) + k_i \frac{E(s)}{s}, \\
&= \left(k_p + \frac{k_i}{s}\right) E(s), \\
&= \left(\frac{k_p s + k_i}{s}\right) E(s), \\
&= k_p \left(\frac{s + \frac{k_i}{k_p}}{s}\right) E(s),
\end{aligned}$$

where $E(s)$ is the Laplace transform of the velocity error. Hence, the block diagram in Fig. 4.16 is obtained. Notice that, now, the system type is 1 because the open-loop transfer function:

$$G(s)H(s) = k_p \left(\frac{s + \frac{k_i}{k_p}}{s}\right) \frac{k}{s+a},$$

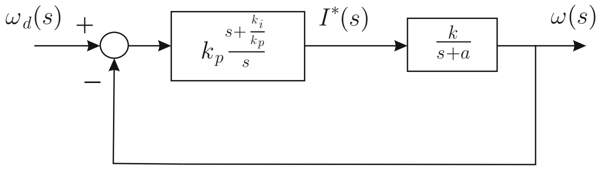



**Fig. 4.16** Proportional–integral control of velocity

has one pole at $s = 0$ and, thus, according to Sect. 4.4.1, the steady-state error is zero, i.e., the motor velocity reaches the desired velocity when time is large enough if $\omega_d$ is constant, or a step signal. We conclude that the integral term of a PI controller is introduced to render the steady-state error zero. It must be stressed, however, that this result is absolutely true only if it is ensured that the closed-loop system is stable, i.e., if all the closed-loop poles have a negative real part.

Some simulation results are shown in Fig. 3.39 where these ideas are corroborated. Notice that in those simulations the steady-state error is still zero, even when a constant external disturbance is applied. However, it is important to say that this property is not described by the steady-state error analysis presented in this chapter. Recall that this analysis is only valid for a closed-loop system such as that represented in Fig. 4.10, i.e., when no external disturbance is present.

Finally, according to Sects. 4.4.2 and 4.4.3, a PI velocity controller is only useful for step desired outputs as the steady-state error is still different from zero if the desired velocity is either a ramp or a parabola.

*Example 4.18* Consider the position control system of a permanent magnet brushed DC motor depicted in Fig. 4.17. Notice that the motor transfer function has one pole at $s = 0$, when the output to be controlled is in the position. Hence, the system type is 1 and the steady-state error is zero if the desired position is a constant, i.e., a step (see Sect. 4.4.1). It is important to realize that this result stands if any of the following controllers are employed (see Chap. 11): *a*) A PD controller, i.e., Fig. 4.18, *b*) A proportional position controller with velocity feedback, i.e., Fig. 4.19, or *c*) A lead-compensator, i.e., Fig. 4.20. The purpose of these controllers is to ensure that the closed-loop system is stable, introducing suitable damping and accomplishing a suitable response time. It is important to stress that closed-loop stability must be ensured to really achieve a zero steady-state error.

Some simulation results are presented in Figs. 3.34 and 3.36, which corroborate the above-mentioned ideas. Notice, however, that the steady-state error analysis presented in this chapter is not intended to explain the different from zero steady-state error observed in those simulations when a constant external disturbance appears. Recall that this analysis is only valid for a closed-loop system, such as that represented in Fig. 4.10, i.e., when no external disturbance is present.

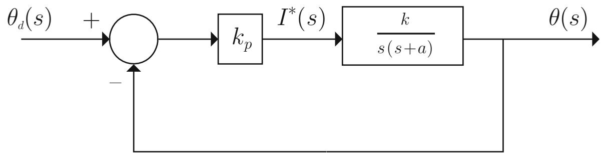



**Fig. 4.17** Proportional control of the position

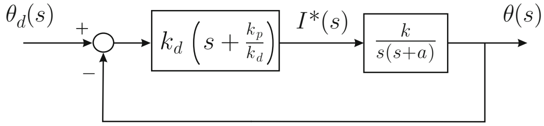



**Fig. 4.18** Proportional–derivative control of the position

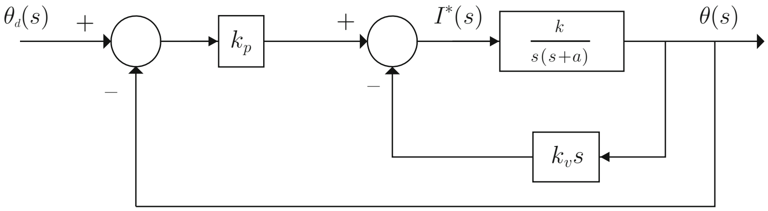



**Fig. 4.19** Proportional control with velocity feedback

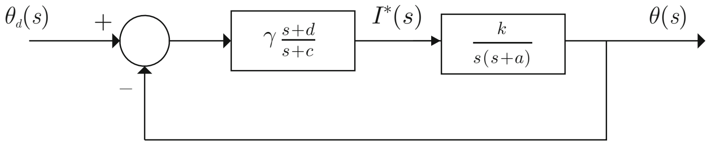



**Fig. 4.20** Lead-compensator for position control

*Example 4.19* Consider the control of a permanent magnet brushed DC motor when constant references are employed. How can it be explained that the proportional control of velocity cannot achieve a zero steady-state error, but a proportional control of the position does achieve a zero steady-state error?

The answer to this question is the following. When the error is zero in position control, then the commanded current $i^* = k_p(\theta_d - \theta)$ is zero; hence, the motor stops and the condition $\theta_d = \theta$ can stand forever.

On the other hand, when the error is zero in velocity control, then the commanded current $i^* = k_p(\omega_d - \omega)$ is zero; hence, the motor tends to stop. This means that the condition $\omega_d = \omega$ cannot stand forever. As a result the steady-state error is such that the difference between $\omega$ and $\omega_d$ is large enough to command an electric current $i^* = k_p(\omega_d - \omega)$, which maintains the motor rotating at that velocity. When a PI velocity controller is employed, then the integral of the error $\int_0^t (\omega_d - \omega(r))dr$ is constant when $\omega_d = \omega$. This constant value, multiplied by $k_i$ is enough to command a suitable electric current to maintain the motor rotating at the desired velocity.

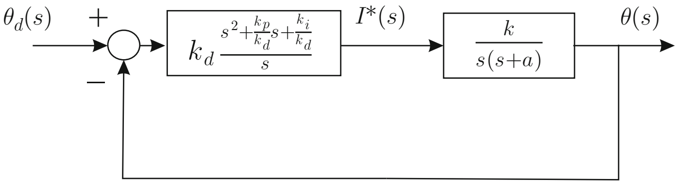



**Fig. 4.21** Proportional–integral–derivative (PID) control of the position

*Example 4.20* Consider the position control problem of a permanent magnet brushed DC motor where a PID controller is employed, i.e.,

$$i^* = k_p e + k_d \frac{de}{dt} + k_i \int_0^t e(r)dr, \quad e = \theta_d - \theta.$$

Use of the Laplace transform yields:

$$\begin{aligned} I^*(s) &= k_p E(s) + k_d s E(s) + k_i \frac{E(s)}{s}, \\ &= \left(k_p + k_d s + \frac{k_i}{s}\right) E(s), \\ &= k_d \frac{s^2 + \frac{k_p}{k_d} s + \frac{k_i}{k_d}}{s} E(s). \end{aligned}$$

Hence, the block diagram in Fig. 4.21 is obtained. Notice that the motor has one pole at $s = 0$ and the controller has another pole at $s = 0$, i.e., the system type is 2 because the open-loop transfer function has two poles at the origin. This means that, according to Sects. 4.4.1 and 4.4.2, the position reaches the desired position $\theta_d$, if this is either a step or a ramp. A steady-state error different from zero is obtained if the desired position is a parabola, see Sect. 4.4.3.

To verify the above conclusions, some simulations have been performed using the MATLAB/Simulink diagrams shown in Fig. 4.22. The top simulation diagram uses a ramp as the reference with a slope of 10 and an initial value of $-10$. The PID controller gains are $k_d = 1$, $k_p = 2$, $k_i = 1$. It is observed in Fig. 4.23 that the steady-state error is zero, which corroborates the steady-state error analysis presented above. The bottom simulation diagram in Fig. 4.22 uses a ramp with slope of 1 and a zero initial value, which passes through an integrator with an initial value of $-10$. This results in a parabola reference given as $\frac{1}{2}t^2 - 10$, i.e., $A = 1$. The PID controller gains are the same as in the previous simulation, i.e., $k_d = 1$, $k_p = 2$, $k_i = 1$. According to Sect. 4.4.3 and Figs. 4.21, 4.22:

$$e_{ss} = \frac{A}{k_a}, \quad k_a = \lim_{s\to 0} s^2 G(s), \quad G(s) = k_d \frac{s^2 + \frac{k_p}{k_d} s + \frac{k_i}{k_d}}{s} \frac{k}{s(s+a)},$$

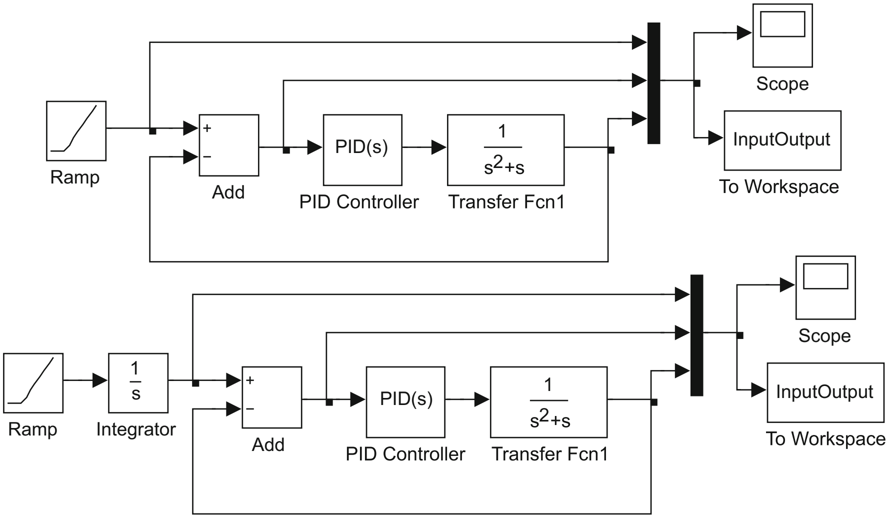



**Fig. 4.22** MATLAB/Simulink diagrams used to obtain results in Figs. 4.23 and 4.24

i.e., $k_a = \frac{kk_i}{a}$, and $e_{ss} = \frac{a}{kk_i} = \frac{1}{1\times 1} = 1$, because $A = 1$. The reader can verify in Fig. 4.24 that the system error is unity at $t = 20$, i.e., in a steady state. This corroborates the steady-state error analysis presented above.

Once simulations in Fig. 4.22 stop, the following MATLAB code is executed in an m-file to draw Figs. 4.23 and 4.24:

```
nn=length(InputOutput(:,1));
n=nn-1;
Ts=20/n;
t=0:Ts:20;
subplot(2,1,1)
plot(t,InputOutput(:,1),'k-.',t,InputOutput(:,3),
'k-');
axis([-1 20 -30 200])
subplot(2,1,2)
plot(t,InputOutput(:,2),'k-');
axis([-1 20 -12 8])
xlabel('time [s]')
```

On the other hand, consider the system depicted in Fig. 4.25. It represents the closed-loop control system of a ball and beam (see Chap. 14). Notice that the open-loop transfer function has two poles at $s = 0$. In this case, these two open-loop poles at the origin are part of the plant model, i.e., the plant naturally possesses those poles and it is not necessary to use a controller with an integral part to introduce them. Hence, the steady-state error is zero if the desired position is either a step or a ramp. However, if it is a parabola, then the steady-state error is different from zero.

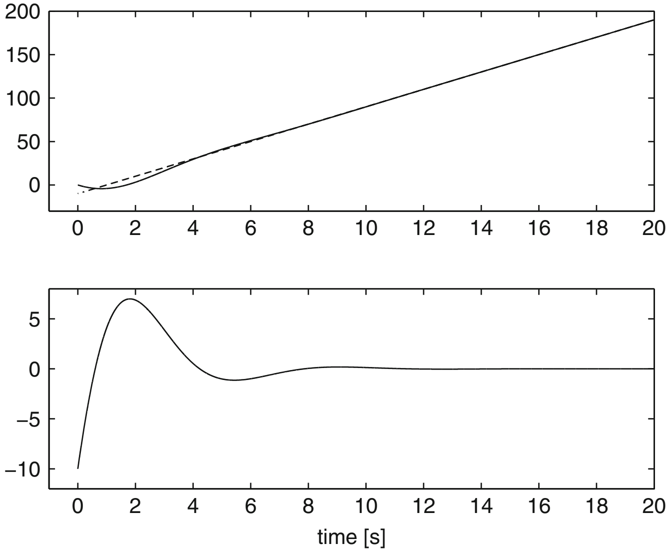



**Fig. 4.23** Time response obtained from the top simulation diagram in Fig. 4.22, which is equivalent to the system in Fig. 4.21, when a ramp reference is commanded. Top subfigure: system response (continuous), ramp (dashed) Bottom subfigure: system error

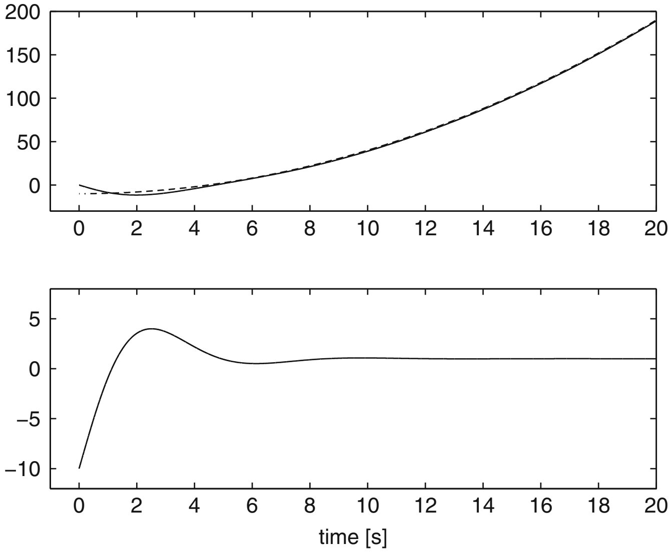



**Fig. 4.24** Time response obtained from the bottom simulation diagram in Fig. 4.22, which is equivalent to the system in Fig. 4.21, when a parabola reference is commanded. Top subfigure: system response (continuous), parabola (dashed) Bottom subfigure: system error

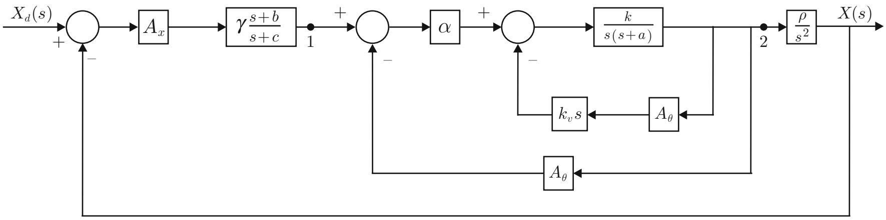



**Fig. 4.25** Control of a ball and beam system

*Example 4.21* Consider a permanent magnet brushed DC motor with transfer function $G_m(s) = \frac{\theta(s)}{I^*(s)} = \frac{k}{s(s+a)}$. Suppose that it is desired to achieve a zero steady-state error when the position reference $\theta_d$ is a parabola. In this case, the system type is required to be 3. As the plant has one pole at $s = 0$, it is necessary to use a controller introducing two poles at $s = 0$. Hence, the following controller is proposed:

$$i^* = k_p e + k_d \frac{de}{dt} + k_i \int_0^t e(r)dr + k_{ii} \int_0^t \int_0^r e(\tau)d\tau dr, \quad e = \theta_d - \theta,$$

which can be rewritten as:

$$I^*(s) = \frac{k_d s^3 + k_p s^2 + k_i s + k_{ii}}{s^2} E(s), \tag{4.16}$$

where $E(s)$ stands for the Laplace transform of the position error. This shows that the controller has two poles at $s = 0$ and, thus, it is useful for solving the problem. Notice that not only the double-integral term is included, with gain $k_{ii}$, but also the single-integral term with gain $k_i$. This is done for closed-loop stability reasons. If the single- integral term is not considered, then closed-loop instability is observed. This can be verified by the reader obtaining the closed-loop characteristic polynomial and applying Routh's stability criterion. As a general rule, if a controller has to introduce an integral term of order $k$, then the terms of integrals of order 1 to $k - 1$ must also be included.

Consider the simulation diagram in Fig. 4.26. This is a system with type 3, which is equivalent to a plant $G_m(s) = \frac{1}{s(s+1)}$ with a controller such as that in (4.16) where $k_d = 0.5$, $k_p = 3$, $k_i = 2$, $k_{ii} = 1$. A ramp with an unit slope and a zero initial value is passed through an integrator with $-10$ as initial value. This corresponds to a parabola reference given as $\frac{1}{2}t^2 - 10$, i.e., $A = 1$. In Fig. 4.27 the steady-state error is corroborated to be zero, which verifies the above predictions.

Once the simulation in Fig. 4.26 stops, the following MATLAB code is executed in an m-file to draw Fig. 4.27:

```
nn=length(InputOutput(:,1));
n=nn-1;
```

```
Ts=20/n;
t=0:Ts:20;
subplot(2,1,1)
plot(t,InputOutput(:,1),'k--',t,InputOutput(:,3),
'k-');
axis([-1 20 -30 200])
subplot(2,1,2)
plot(t,InputOutput(:,2),'k-');
axis([-1 20 -12 8])
xlabel('time [s]')
```

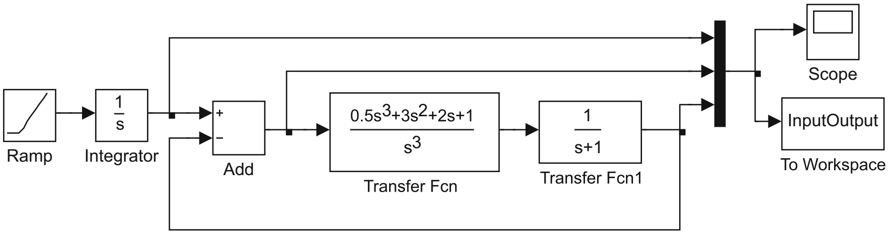



**Fig. 4.26** MATLAB/Simulink diagram used to obtain results in Fig. 4.27

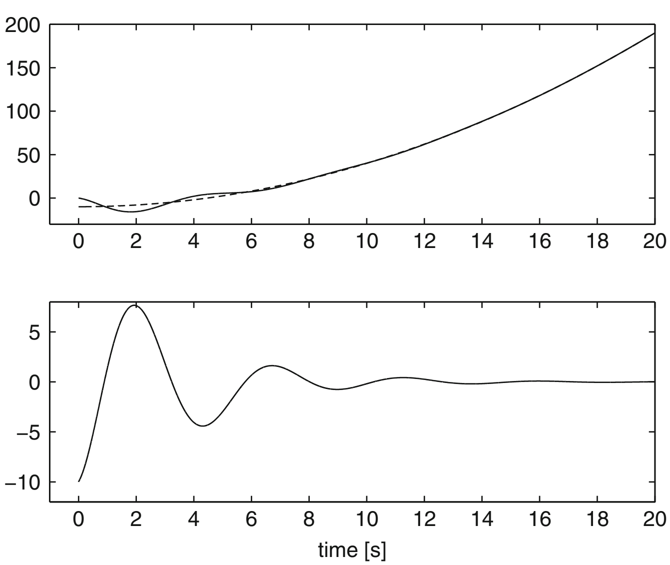



**Fig. 4.27** Time response obtained from the simulation diagram in Fig. 4.26, i.e., a type 3 system, when a parabola reference is commanded. Top subfigure: system response (continuous), parabola (dashed) Bottom subfigure: system error

## 4.5 Summary

The basic preliminary tools for the analysis and design of arbitrary order control systems have been presented in this chapter. It must be understood that the response of any closed-loop control system, no matter how complex, is determined by the equivalent closed-loop transfer function.

The control system representation by means of block diagrams, and their simplification, are instrumental in obtaining the corresponding closed-loop transfer function. From this transfer function it is possible to determine: $i$) The closed-loop stability and the transient response, i.e., the closed-loop poles, and $ii$) The steady-state response.

It is explained in Chap. 3 that a necessary and sufficient condition for a transfer function to be stable is that all its poles have a negative real part. However, checking this condition through the exact computation of the poles is a complex problem. This is especially true when the numerical values of the characteristic polynomial coefficients are not known. This situation is common in control systems design as the characteristic polynomial coefficients depend on the controller gains, which are initially unknown. Moreover, it is desirable that the controller gains can be chosen within a range to render the design flexible. These features require the use of analytical tools, instead of numerical tools, useful to determine when a closed-loop control system is stable. This fact represents an important obstacle because the analytical formulas existing to compute the roots of higher-degree polynomials are complex. Furthermore, no analytical solution exists for polynomials with a degree greater than or equal to 5. This problem is successfully solved by Routh's stability criterion, which, however, only verifies whether all the polynomial roots have a negative real part, but does not compute the exact values of the roots. On the other hand, although the rule of signs presented in Sect. 4.2 has some limitations, it may be simpler to use than Routh's stability criterion in some applications. This is the reason for including such a methodology.

As a result of the study of the steady-state error, some criteria are established for suitable selection of a controller such that the steady-state response reaches the desired value; in other words, to render the forced response equal or very close to the close-loop system input, i.e., the desired response. Finally, the controller selection must be also performed such that the desired transient response is achieved through the suitable location of the closed-loop poles. The solution to this latter problem is presented in the next chapter.

## 4.6 Review Questions

1. What does system type mean?
2. How can the system type be increased in a control system?

3. How does Routh's stability criterion ensure that the real part of all roots of the characteristic polynomial is negative?
4. What are the advantages and the disadvantages or the rule of signs presented in Sect. 4.2?
5. Suppose that analysis of the steady-state error concludes that the output final value reaches the desired output value. Why is it still necessary for the closed-loop system to be stable?
6. Suppose that a controller possessing a five times iterated integral is required to render the steady-state error zero. Why is it also necessary to include terms with a four times integrated integral, a three times iterated integral, a two times iterated integral and a simple integral? Illustrate your answer using a permanent magnet brushed DC motor as a plant and Routh's stability criterion.
7. What is the relationship between the study of the steady-state error in this chapter and the requirement that the forced response reaches the closed-loop system input, i.e., the desired output?
8. The study of the steady-state error that has been presented in this chapter only considers test inputs such as step, ramp, and parabola. What would happen in applications, such as the video camera tracking control system, where the desired output is not exactly known in advance?

## 4.7 Exercises

1. Based on the knowledge of the system type required to ensure a zero steady-state error for references such as the step, the ramp and the parabola, show that the corresponding closed-loop transfer function (see Fig. 4.10) must possess the following features:

   - The terms independent of $s$ must be equal in the polynomials at the numerator and the denominator to ensure a zero steady-state error when the reference is a step.
   - The terms independent of $s$ in addition to the first-order terms in $s$ must be equal in the polynomials at the numerator and the denominator to ensure a zero steady-state error when the reference is a ramp.
   - The terms independent of $s$ in addition to the first-order and the second-order terms in $s$ must be equal in the polynomials at the numerator and the denominator to ensure a zero steady-state error when the reference is a parabola.

   This means that the controller also has to suitably assign the closed-loop transfer function zeros, which supports the arguments presented in Example 4.21 (see (4.16)).
2. In Example 3.18, in Chap. 3, everyday experience-based arguments are presented to explain the steady-state error achieved by a proportional control of a mass-spring-damper system when $x_d$ is constant. Now, explain this result using the system type as the argument.

3. Consider the mass-spring-damper system presented in Example 3.8, Chap. 3. If after a constant force $A$ is applied, the mass reaches $\dot{x} = 0$ and $\ddot{x} = 0$, then replacing this in the corresponding differential equation, it is found that $x = \frac{A}{K}$ in a steady state. Now, use the final value theorem to compute the final value of $x$ when $f = A$. What is the relationship among conditions $\dot{x} = 0$, $\ddot{x} = 0$, and $s \to 0$ in the final value theorem?
4. Consider the rotative mass-spring-damper system:

$$J\ddot{\theta} + f\dot{\theta} + K\theta = T,$$

where $J$ is the inertia, $f$ is the viscous friction coefficient, $K$ is the spring stiffness constant, $\theta$ is the body angular position, and $T$ the applied torque, which is given by the following PID controller:

$$T = k_p e + k_d \frac{de}{dt} + k_i \int_0^t e(r)dr, \quad e = \theta_d - \theta.$$

Use Routh's stability criterion to show that the following conditions:

$$J > 0 \quad k_i > 0, \quad K + k_p > 0, \quad \frac{f + k_d}{J} > \frac{k_i}{K + k_p} > 0,$$

ensure closed-loop stability. Notice that a large integral gain $k_i$ tends to produce instability and this effect can be compensated for using large values for either $k_p$ or $k_d$. Also notice that a small inertia $J$ allows the use of larger integral gains before instability appears and something similar occurs if the stiffness constant $K$ is large. Can you find an explanation for this from the point of view of physics (mechanics)?
5. Proceeding as in the Example 4.3 in this chapter, show that the closed-loop transfer function shown in Fig. 4.3 is:

$$M(s) = \frac{C(s)}{R(s)} = \frac{G(s)}{1 - G(s)H(s)},$$

when the system has positive feedback, i.e., when the feedback path adds (instead of subtracts) in Fig. 4.3.
6. Consider the mechanical system depicted in Fig. 4.28. In the Example 2.3, Chap. 2, it was found that the corresponding mathematical model is given as:

$$\frac{d^2 x_{m1}}{dt^2} + \frac{b_1}{m_1}\frac{dx_{m1}}{dt} + \frac{K}{m_1}(x_{m1} - x_{m2}) = \frac{1}{m_1}F(t),$$
$$\frac{d^2 x_{m2}}{dt^2} + \frac{b_2}{m_2}\frac{dx_{m2}}{dt} - \frac{K}{m_2}(x_{m1} - x_{m2}) = 0,$$

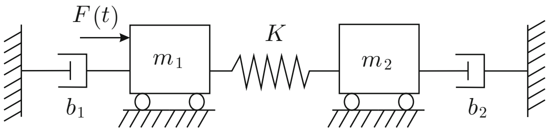



**Fig. 4.28** Two bodies connected through a spring

where $x_{m1}$ is the position of $m_1$ and $x_{m2}$ is the position of $m_2$. Apply the Laplace transform to each one of these equations to express $X_{m1}(s)$ and $X_{m2}(s)$ as the outputs of two transfer functions. Draw the corresponding block diagram suitably connecting these two transfer functions according to their inputs and outputs. Using the result in the previous exercise, simplify this block diagram to verify that:

$$X_{m2}(s) = \frac{G_1(s)G_2(s)/m_1}{1 - G_1(s)G_2(s)K/m_1} F(s),$$

where $F(s)$ is the Laplace transform of $F(t)$ and:

$$G_1(s) = \frac{1}{s^2 + \frac{b_1}{m_1}s + \frac{K}{m_1}},$$

$$G_2(s) = \frac{\frac{K}{m_2}}{s^2 + \frac{b_2}{m_2}s + \frac{K}{m_2}}.$$

- Using this result, verify that the transfer function $\frac{X_{m2}(s)}{F(s)}$ has one pole at $s = 0$. What does this mean? Suppose that a constant force $F(t)$ is applied. What happens with position of mass $m_2$ under the effect of this force? What happens with the position of the mass $m_1$? It is suggested that the transfer function between $X_{m1}(s)$ and $X_{m2}(s)$ should be found.
- Use Routh's criterion to find the conditions for the transfer function $\frac{X_{m2}(s)}{F(s)}$ to be stable.

Notice that one may proceed similarly in the other examples including springs in Chap. 2.

Some other exercises are proposed at the end of Chap. 5 whose solution involves concepts and tools introduced in the present chapter.

## References


1. S. Bennett, A brief history of automatic control, *IEEE Control Systems Magazine*, pp. 17–25, June 1996.
2. K. Ogata, *Modern control engineering*, 4th edition, Prentice-Hall, Upper Saddle river, 2002.

3. N. S. Nise, *Control systems engineering*, 5th. edition, John Wiley and Sons, New Jersey, 2008.
4. R. C. Dorf and R. H. Bishop, *Modern control systems*, 10th. edition, Pearson Prentice-Hall, 2005.
5. B. C. Kuo, *Automatic control systems*, Prentice-Hall, 1995.
6. G. H. Hostetter, C. J. Savant and R. T. Stefani, *Design of feedback control systems*, Holt, Rinehart and Winston, 1982.