Chapter 5
Stability of Linear Control Systems

In practical applications, the stability of the control system is an important requirement. A control task can not be realized with an unstable control loop. The stability of a control system has to be distinguished from the stability of the process itself. There are cases when an unstable process has to be stabilized and controlled with a closed-loop control system. There are processes which would not even operate without control, the closed-loop control stabilizes the process. The best known examples of such systems are the control of an airplane, or in everyday life, riding a bicycle.

Closed-loop control circuits may present surprising phenomena. These phenomena are due to the process dynamics, inertia and time-delays. Therefore the processes can not follow immediately the commands acting on their input. Generally, the time of the response of a process is not within the time scale of human reactions (sometimes it is much slower, and sometimes much faster). In some cases the short time response does not agree with what we would experience waiting for a bit longer time (e.g., non-minimum phase processes). Therefore experimental investigation of the stability is not acceptable in operating and controlling real processes, which are generally very expensive. Precise mathematical methods are needed to analyze the stability of control systems.

5.1 The Concept of Stability

If a system has the property that it will get back into the equilibrium state again after moving away from its equilibrium state, then it is stable.

If the system is non-linear, its stability depends on the input signal and also on the operating point. In this case, stability is a characteristic of a state of the system, and not of system as a whole. In case of a linear system, stability is characteristic for

L. Keviczky et al., *Control Engineering*, Advanced Textbooks in Control and Signal Processing, https://doi.org/10.1007/978-981-10-8297-9_5

the system. Stability depends on the system's structure and parameters, but does not depend on the input signal. As far as stability is concerned, a number of various formulations exist.

Stability of an un-excited system

A system is stable if after removing it from its equilibrium state and allowing it to move freely, it returns to its original state. If the system moves away from its original state, its behavior is unstable. The system is on the boundary of stability and unstability if after removal from the equilibrium state it does not return to it, but remains in its close vicinity, which depends on the extent of the removal (e.g. it makes un-damped oscillations with bounded amplitude around the initial state). In non-linear systems, the system is also considered stable if in the boundary case removing it from the steady state it returns to an arbitrarily prescribed small vicinity of the steady state. The system is *asymptotically stable* if after removing it from its equilibrium state it returns to its original state. A stable linear system is asymptotically stable. In the case of asymptotic stability the weighting function $w(t)$ of a linear system is decreasing in the following sense:

$$\lim_{t\to\infty} w(t) = 0 \tag{5.1}$$

and furthermore $w(t)$ is absolutely integrable, i.e.

$$\int_0^\infty |w(t)|\mathrm{d}t < \infty \tag{5.2}$$

Stability of an excited system

A system is stable if it responds to any bounded input signal with a bounded output signal, from any initial condition. Stability of the excited system is called *Bounded-Input–Bounded-Output* (*BIBO*) stability.

For linear systems, stability is a system property. Stability does not depend on the magnitude of the excitation. Additionally, for linear systems, if the un-excited system is stable, then the excited system is also stable. Stability can be checked unambiguously from the system response to a simple input signal.

Internal stability

A closed-loop control system fulfills the requirement of *internal stability* if its output signal and all of its inner signals respond in a stable way to any outer excitation signal. Let us investigate the control system shown in Fig. 5.1. Besides tracking the reference signal r the rejection of the effect of disturbance y_{ni} and y_{no} acting at the input and the output of the plant P, respectively and the effect of the measurement noise y_{z} on the output are also investigated. The system is stable if for

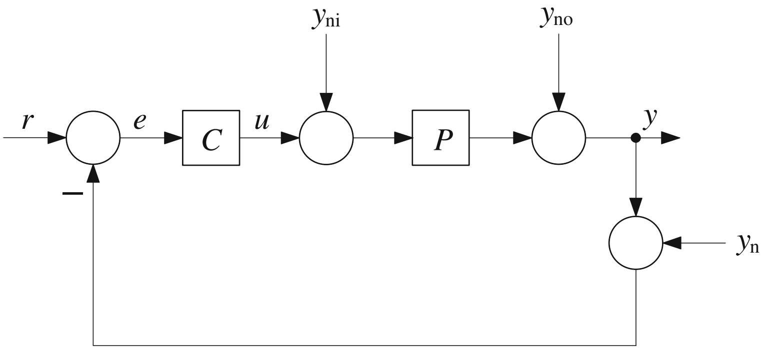


Fig. 5.1 Block diagram of a closed-loop control system

all the considered bounded input signals the controlled output signal y, the manipulated control variable u and the error signal e are bounded. It can be shown that in the structure of Fig. 5.1 it is always sufficient to choose two arbitrary external and two arbitrary internal signals. Internal stability requires the investigation of the stability of the following four overall transfer functions: $CP/(1+CP)$, $1/(1+CP)$, $P/(1+CP)$, $C/(1+CP)$. This can be characterized by the *transfer matrix* of the closed-loop control circuit

$$\boldsymbol{T}_{\mathrm{t}} = \begin{bmatrix} \frac{CP}{1+CP} & \frac{P}{1+CP} \\ \frac{C}{1+CP} & \frac{1}{1+CP} \end{bmatrix} \tag{5.3}$$

A closed-loop control system has the property of internal stability if $\boldsymbol{T}_{\mathrm{t}}$ is stable, i.e. all its elements are stable. Internal stability is equivalent to the stability of the excited system if the open-loop system has no non-observable or non-controllable right side poles (i.e. the zeros of the controller do not cancel the pole in the right half-planes of the plant). (It has to be emphasized that it is not allowed to cancel an unstable pole of the plant with a right side zero of the controller, as the unstable pole would become invisible only in the relationship between the output signal and the reference signal, but would remain in the relationship between the output signal and the disturbance acting at the input of the plant.)

Lyapunov stability

According to the Lagrange energy theorem a system is in balance if its potential energy is minimal. Lyapunov prescribes the determination of a scalar function of energy property (the so called Lyapunov function) belonging to the differential equation or state equation of a general nonlinear system with constant coefficients. If in the considered range of the state variables this function is positive and its derivative is negative, the system is asymptotically stable. The methods of Lyapunov provide sufficient conditions for the determination of the stability properties of nonlinear systems. Choosing a Lyapunov function is not always a simple task. Lyapunov suggests first the investigation of the stability of the linearized system at individual operating points. Of course, the method of Lyapunov can also be applied to investigate the stability of a linear system. But for linear systems, it is expedient to use simpler direct methods.

5.2 Stability of the Closed-Loop System

Negative feedback, which is the basic structure of a closed-loop control system, also involves the risk of instability. To demonstrate this let us consider the control loop shown in Fig. 5.2. Assuming a step-like abrupt change of the reference signal, the output signal starts to grow from zero. Then the error signal decreases starting from an initial value of 1. If the gain of the controller is high, first a large input signal appears at the plant input, which results in a sharp rise in the output signal. The dynamics of this change is determined by the dynamics of the process P and the controller C, i.e., by the gains and the time constants of the corresponding transfer functions. When the output signal reaches its required value, viz., the one prescribed by the reference signal, the error signal reaches zero. But because of the inertia of the system, the signals will not be settled immediately at their required values, but maintaining their trend they will continue changing further, according to their actual slope. If the output signal exceeds its prescribed value, the error signal becomes negative, and after a while the output signal will start decreasing. With large time constants of the process and high gains of the controller, the overshoot may be significant. Steady or increasing oscillations may appear in the control system. The problem of stability emerges because the system uses the information supplied by the error signal in a delayed manner, and if the gains are high, during the delay time the output signal "runs away" so much that the control system will not be able to bring it back to its required value. The parameters of the controller always have to be chosen in such a way that the control system is stable.

The instability of a feedback control system is caused by large time constants and high gains. This phenomenon is illustrated by the behavior of the control system shown in Fig. 5.3. The process is represented by a pure dead-time with unit gain, given by its step response. The dead-time element follows its input signal u after a time specified by T_d. The controller is a pure proportional element with gain A, thus the loop gain is $K = A$. Let us investigate the signals in the control circuit for $K = 0.5$, 1 and 2. The evolution of the signals can be easily followed.

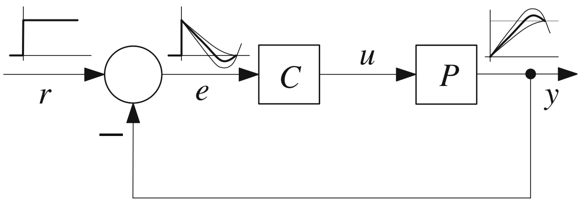


Fig. 5.2 Dynamics of a control system

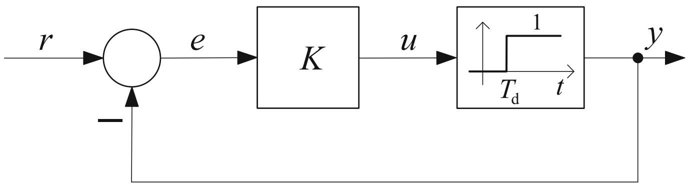


Fig. 5.3 Control system with dead-time

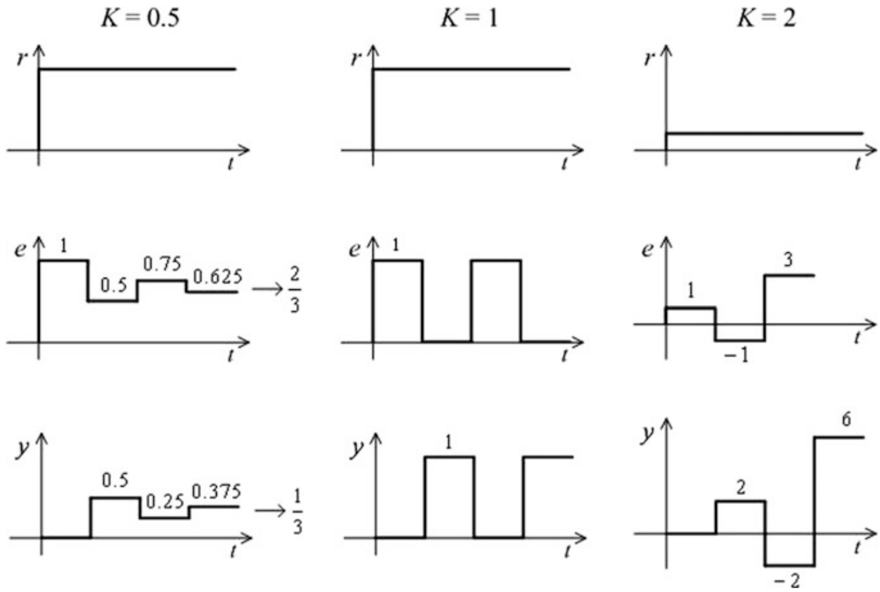


Fig. 5.4 Signals in a control circuit with dead-time

Figure 5.4 shows the reference signal, the error signal, and the controlled output signal. With $K = 0.5$ the control system is stable (but it is inaccurate: the output signal settles at 1/3 instead of the required value 1). In the case of $K = 1$, steady oscillations appear: the system is on the borderline between stability and instability. With $K = 2$, the system is unstable.

The values of the individual signals can also be given analytically in the considered time ranges according to Table 5.1.

Table 5.1 Signal values in a closed-loop control system with dead-time

Time range t	Error signal e	Output signal y
$0–T_d$	1	0
$T_d–2T_d$	$1 - K$	K
$2T_d–3T_d$	$1 - K(1 - K)$	$K(1 - K)$
$3T_d–4T_d$	$1 - K[1 - K(1 - K)]$	$K[1 - K(1 - K)]$

It is seen that with the progress of time the error signal e can be given by a geometrical series with quotient $-K$. If $K<1$, the series converges to $\lim_{t\to\infty} e(t) = 1/(1+K)$, and the limit value of the output signal is $\lim_{t\to\infty} y(t) = K/(1+K)$. Thus the stability limit is $K = 1$. The higher the value of K, the smaller the steady error in the control circuit, but the requirement of stability sets a limit for increasing K. Stability and static accuracy are often contradictory requirements. In the design of a control system, an appropriate compromise has to be realized to ensure both stability and the required static accuracy.

Stability is an important property for a linear system. In the case of instability the control system "runs away" even if it is excited only temporarily by some noise, e.g. an impulse acts at its input. Figure 5.5 shows the signals in the case of $K = 2$ when the reference signal is a short time impulse of amplitude unity.

5.3 Mathematical Formulation of the Stability of Continuous Time Linear Control Systems

If an un-excited closed-loop control system is asymptotically stable then the time function describing its transients contains components that are decreasing functions of time. The transient time function is a combination of exponential components whose exponents are the roots of the characteristic equation of the system.

In a controllable and observable control system (when the zeros of the controller do not cancel the poles of the plant) the roots of the characteristic equation are identical to the poles of the overall transfer function of the closed-loop. Formally, the characteristic equation of the differential equation describing the system is equivalent to the denominator of the overall transfer function of the closed-loop system.

That is, the overall transfer function of the closed-loop between the output signal y and the reference signal r is

$$T(s) = \frac{Y(s)}{R(s)} = \frac{C(s)P(s)}{1 + C(s)P(s)} = \frac{C(s)P(s)}{1 + L(s)}. \tag{5.4}$$

The differential equation of the system is the inverse Laplace transform of

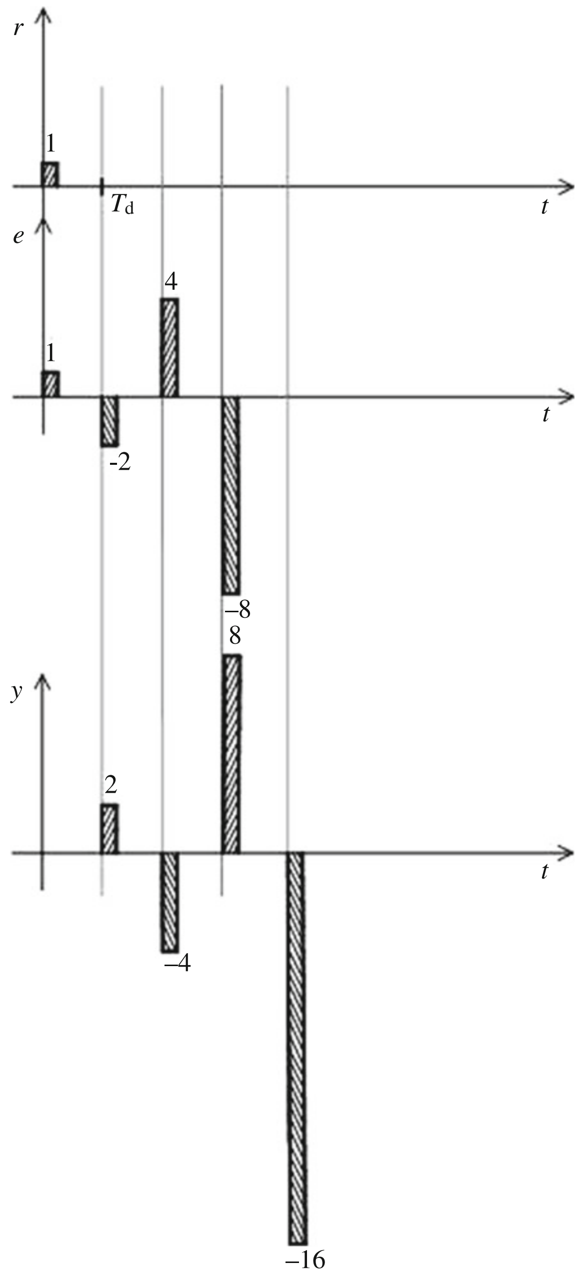
$$[1 + L(s)]Y(s) = C(s)P(s)R(s) \tag{5.5}$$

and the characteristic equation is, formally,

$$1 + L(s) = 0. \tag{5.6}$$

Thus the roots of the characteristic equation are the same as the poles of the overall transfer function of the closed-loop system.

Fig. 5.5 The signals at the output of an unstable system "run away" even if a short time signal acts at its input



If the loop transfer function is a rational fraction, i.e. $L(s) = \mathcal{N}(s)/\mathcal{D}(s)$ where $\mathcal{N}(s)$ and $\mathcal{D}(s)$ are polynomials, then the characteristic equation can also be given in the following form:

$$\mathcal{A}(s) = \mathcal{D}(s) + \mathcal{N}(s) = 0 \tag{5.7}$$

or

$$a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_o = a_n (s - p_1)(s - p_2) \ldots (s - p_n) = 0. \tag{5.8}$$

If the system is described by its state equation with state matrix $\boldsymbol{A}$, then the characteristic equation can be given by the relationship

$$\det(s\boldsymbol{I} - \boldsymbol{A}) = 0. \tag{5.9}$$

(see also Chap. 3).

If the coefficients of the characteristic equation are real numbers, then the roots of the equation are real numbers or pairs of complex conjugate numbers.

The condition for asymptotic stability is that the real part of the poles p_i of the closed-loop have negative real parts, as this condition ensures that the transients are decreasing function of time. This condition can also be formulated as follows: a closed-loop control system is asymptotically stable if all of its poles lie in the left half-plane of the complex plane.

If any of the poles lies in the right half-plane, the system is unstable. If besides the poles in the left half-plane there are poles in the origin, then there is an integrating effect in the system, for step input its output signal goes to infinity. If there are pairs of complex conjugate simple poles on the imaginary axis, then steady oscillations do appear in the transients. In the case of multiple poles, the amplitudes of the oscillations are increasing. In practice, only asymptotic stability is acceptable.

5.4 Analytical Stability Criteria

Stability can be decided from the location of the roots of the characteristic equation which are the poles of the closed-loop system.

If there is no dead-time, the characteristic equation is an algebraic equation, whose roots can be given analytically provided the degree is less than 5 (Galois theorem). For higher degrees, numerical root searching methods can be applied which determine the roots with a given accuracy.

If the system contains dead-time, then the characteristic equation is a transcendental equation $\mathcal{D}(s) + \mathcal{N}(s)e^{-sT_\mathrm{d}} = 0$, whose solution is not simple, and in the case of instability it is difficult to decide how to stabilize the system. In this case, the characteristic equation can be approximated by a rational functional approximation of the dead-time, or the investigation has to be done in the frequency domain (see Sect. 5.6).

Several procedures have been elaborated to determine the stability without solving the characteristic equation. These procedures are referred to as stability criteria.

If there is no dead-time, then based on the relationships between the roots and the coefficients of the algebraic equation it can be checked with analytical stability criteria whether all the roots lie in the left half of the complex plane, i.e., whether the system is stable or not.

A necessary condition for stability is that all the coefficients of the characteristic equation must be of the same sign and none of the coefficients can be zero. This can be seen easily based on Eq. (5.8). That is, if all the poles have negative real parts, then by multiplying the root factors, all the coefficients will be positive. If there are also pairs of complex conjugate roots with negative real parts, then multiplying the root factors the obtained coefficients are also positive. Suppose $p_{1,2} = -\alpha \pm j\beta$, where $\alpha > 0$ and $\beta > 0$. Let us multiply together the two corresponding factors $[s-(-\alpha+j\beta)][s-(-\alpha-j\beta)] = s^2+2\alpha s+\alpha^2+\beta^2$. The coefficients are evidently positive. In the first- and second-degree cases the sameness of the signs of the coefficients is not only a necessary, but also a sufficient condition for stability. In the sequel, two analytical methods will be given, without proof for checking stability.

5.4.1 Stability Analysis Using the Routh Scheme

Let us build the following scheme from the coefficients of the characteristic polynomial given in (5.8):

$$\begin{array}{ccccc} a_n & a_{n-2} & a_{n-4} & a_{n-6} & \cdots \\ a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \cdots \\ b_{n-2} & b_{n-4} & b_{n-6} & b_{n-8} & \cdots \\ c_{n-3} & c_{n-5} & c_{n-7} & c_{n-9} & \cdots \\ \vdots & & & & \end{array} \tag{5.10}$$

where

$$\begin{aligned} b_{n-2} &= \frac{a_{n-1}a_{n-2}-a_na_{n-3}}{a_{n-1}},\ b_{n-4}=\frac{a_{n-1}a_{n-4}-a_na_{n-5}}{a_{n-1}},\ b_{n-6}=\frac{a_{n-1}a_{n-6}-a_na_{n-7}}{a_{n-1}},\ldots \\ c_{n-3} &= \frac{b_{n-2}a_{n-3}-a_{n-1}b_{n-4}}{b_{n-2}},\ c_{n-5}=\frac{b_{n-2}a_{n-5}-a_{n-1}b_{n-6}}{b_{n-2}},\ldots \end{aligned} \tag{5.11}$$

The length of the rows is decreasing. If the degree of the characteristic polynomial is n, the scheme consists of $n+1$ rows. The arrangement given by (5.10) and, (5.11) is called the Routh scheme.

A system is stable if all the coefficients of its characteristic equation are positive and all the elements of the first column of its Routh scheme are positive. If not all the elements in the first column are positive, the system is unstable, and the number of the changes in the signs gives the number of poles of the closed-loop system that lie in the right half-plane. A zero in the first column indicates that the characteristic equation has a root on the imaginary axis. In this case, the scheme can be continued by taking an arbitrarily small ε value instead of zero.

Example 5.1 Suppose the loop transfer function of a control circuit is $L(s) = K/s(1+s)(1+5s)$. In a closed-loop circuit, a unit negative feedback is applied. Let us determine the value of the critical gain K that brings the control system to the stability limit. The characteristic equation is

$$1 + L(s) = 1 + \frac{K}{s(1+s)(1+5s)} = 0,$$

or

$$5s^3 + 6s^2 + s + K = 0.$$

As all the coefficients have to be positive, the necessary condition for stability is $K > 0$.

$$\text{The Routh scheme is}: \quad \begin{array}{cc} 5 & 1 \\ 6 & K \\ \frac{6-5K}{6} & 0 \\ K & \end{array}.$$

To ensure stability, all the elements of the first column have to be positive. Thus the condition for stability is

$$0 < K < 1.2.$$ ■

5.4.2 Stability Analysis Using the Hurwitz Determinant

Let us build the following Hurwitz determinant of dimension $n \times n$ from the coefficients of the characteristic polynomial (5.8)

$$\begin{vmatrix} a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \cdots \\ a_n & a_{n-2} & a_{n-4} & a_{n-6} & \cdots \\ 0 & a_{n-1} & a_{n-3} & a_{n-5} & \cdots \\ 0 & a_n & a_{n-2} & a_{n-4} & \cdots \\ 0 & 0 & a_{n-1} & a_{n-3} & \cdots \\ \vdots & & & & \end{vmatrix} \tag{5.12}$$

Elements with negative indices are taken to be zeros. The system is stable if all the coefficients of the characteristic equation are positive and all the subdeterminants along the main diagonal are also positive: $\Delta_i > 0$. The subdeterminants are:

$$\Delta_1 = |a_{n-1}|,\ \Delta_2 = \begin{vmatrix} a_{n-1} & a_{n-3} \\ a_n & a_{n-2} \end{vmatrix},\ \Delta_3 = \begin{vmatrix} a_{n-1} & a_{n-3} & a_{n-5} \\ a_n & a_{n-2} & a_{n-4} \\ 0 & a_{n-1} & a_{n-3} \end{vmatrix}, \ldots, \Delta_n. \tag{5.13}$$

Example 5.2 Let us investigate the stability of the system analyzed in Example 5.1 on the basis of the Hurwitz determinant. The characteristic equation is

$$5s^3 + 6s^2 + s + K = 0.$$

As all the coefficients have to be positive, $K > 0$.
The Hurwitz determinant is

$$\begin{vmatrix} 6 & K & 0 \\ 5 & 1 & 0 \\ 0 & 6 & K \end{vmatrix}. \tag{5.14}$$

The subdeterminants along the main diagonal are

$$\Delta_1 = 6 > 0;\ \Delta_2 = 6 - 5K > 0 \text{ and } \Delta_3 = K\Delta_2 > 0.$$

Thus the condition of stability is $0 < K < 1.2$. ■

5.5 Stability Analysis Using the Root Locus Method

The root locus gives the location of the roots of the characteristic equation of the closed-loop system in the complex plane as a parameter (generally the loop gain) changes between zero and infinity.

If the roots are in the left half-plane, the system is stable. At the critical gain the root locus crosses the imaginary axis. At gains where the root locus has moved to the right half-plane, the system becomes unstable.

From the root locus, not only the stability of the closed-loop system can be checked, but from the location of the roots, also the dynamic properties can be determined approximately.

For drawing the root locus, the characteristic equation has to be solved for different parameter values. Today's computer techniques and *CAD* programs provide considerable help in drawing the root locus branches. But often there is the need for a rapid qualitative analysis to assist the designer in design considerations. Therefore, several rules have been elaborated to support the quick sketching of the root locus. (It is also called the Evans method, after the name of the developer of the method.)

5.5.1 Basic Relationships of the Root Locus Method

The characteristic equation of the closed-loop control circuit $1 + L(s) = 0$ can be written also in the following form:

$$L(s) = -1 = k \frac{\prod_{j=1}^{Z} (s - z_j)}{\prod_{i=1}^{P} (s - p_i)}, \tag{5.15}$$

where Z denotes the number of zeros, P is the number of poles and k, is the loop gain factor of the pole-zero form.

For all the points of the root locus the *absolute value condition*

$$|L(s)| = 1 \tag{5.16}$$

and the phase condition

$$\varphi = \pm N180°; \quad N = 1, 3, 5, \ldots \tag{5.17}$$

have to be fulfilled.

This means that for the construction of the root locus, those points in the complex plane are to be looked for that fulfill both the phase condition and the absolute value condition.

Let us denote the absolute value of the vector connecting the zero z_j with an arbitrary point s of the complex plane by C_j, and its phase angle with the positive real axis by γ_j. The absolute value of the vector connecting the pole p_i with the same point s is denoted by D_i, while its phase angle is denoted by δ_i (Fig. 5.6). That is

$$s - z_j = C_j e^{j\gamma_j} \tag{5.18}$$

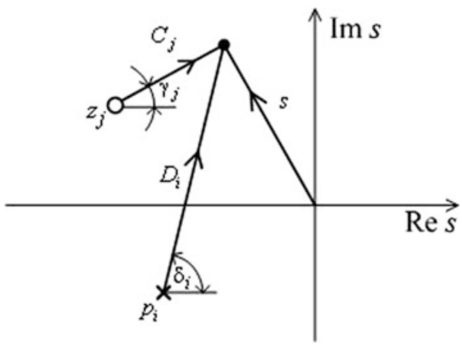


Fig. 5.6 Notation for the vectors connecting the points of the root locus with the poles and zeros of the open-loop

and

$$s - p_i = D_i e^{j\delta_i}. \tag{5.19}$$

The phase condition can be given in the following form:

$$\sum_{j=1}^{Z} \gamma_j - \sum_{i=1}^{P} \delta_i = \pm N180^\circ; \quad N = 1, 3, 5, \ldots \tag{5.20}$$

or

$$\sum_{i=1}^{P} \delta_i - \sum_{j=1}^{Z} \gamma_j = \mp N180^\circ; \quad N = 1, 3, 5, \ldots \tag{5.21}$$

For the absolute value condition, the following relationship holds:

$$\frac{\prod_{i=1}^{P} D_i}{\prod_{j=1}^{Z} C_j} = k. \tag{5.22}$$

A point in the complex plane is a point of the root locus if for that point both the phase condition and the absolute value condition are fulfilled.

The phase condition can also be formulated as follows: a point s on the complex plane is the point of the root locus if from the sum of the angles of the vectors connecting the zeros of the open-loop with that point s one subtracts the sum of the angles of the vectors connecting the poles of the open-loop with s and gets an odd multiple of $\pm 180^\circ$.

The absolute value condition states that a point s is the point of the root locus if dividing the product of the absolute values of the vectors connecting the poles with point s by the product of the absolute values of the vectors connecting the zeros with point s yields the loop gain factor.

Generally, the points of the root locus are determined from the phase condition, and the value of the loop gain factor corresponding to the considered point is obtained from the absolute value condition. Then from the loop gain factor, the loop gain K belonging to the time constant form of the transfer function of the open-loop is calculated by

$$K = L(s)\big|_{s=0} = k \frac{\prod_{j=1}^{Z} (-z_j)}{\prod_{i=1}^{P} (-p_i)}. \tag{5.23}$$

5.5.2 Rules for Drawing Root Locus

There are some simple rules which facilitate drawing the root locus:

1. The root locus is symmetrical with respect to the real axis.
2. The number of its branches is equal to the number of poles of the open-loop transfer function.
3. The root locus starts from the poles of the open-loop when $K = 0$ and runs to the zeros or to infinity when $K \to \infty$. If the number of poles is P and the number of zeros is Z, then Z branches of the root locus run to the zeros and $P - Z$ branches run to infinity. If $P = Z$, the whole root locus is located in a finite range of the complex plane.
4. Sections of the root locus will be on the real axis if to the right of the considered point the sum of the poles and zeros is odd. (It is sufficient to count the real poles and zeros, as the complex poles or complex zeros appear in pairs.)
5. The direction of the asymptotes of the root locus is given by the angles
$$\alpha = \mp \frac{N180^\circ}{P - Z}; \quad N = 1, 3, 5, \ldots \tag{5.24}$$
6. The asymptotes of the root locus cross the real axis at the point calculated by the following relationship:
$$x_o = \frac{\sum_{i=1}^{P} p_i - \sum_{j=1}^{Z} z_j}{P - Z} = \frac{\sum_{i=1}^{P} \mathrm{Re} p_i - \sum_{j=1}^{Z} \mathrm{Re} z_j}{P - Z}. \tag{5.25}$$
7. The location of leaving or entering the real axis can be determined by the equation
$$\sum_{i=1}^{P} \frac{1}{x - p_i} - \sum_{j=1}^{Z} \frac{1}{x - z_j} = 0 \tag{5.26}$$
8. The critical gain factor can be determined from the characteristic equation by the Routh scheme or the Hurwitz determinant. The crossing points with the imaginary axis can be calculated from the characteristic equation assuming that in this case two of its roots are pure imaginary complex conjugate roots.

Explanation of the drawing rules

1. As the coefficients of the characteristic equation are real numbers, its roots are real or complex conjugate pairs. Therefore the root locus is symmetrical to the real axis.
2. The degree of the characteristic equation is equal to the number of the poles of the open-loop. Namely, if the transfer function of the open-loop is a rational fraction, $L(s) = \mathcal{N}(s)/\mathcal{D}(s)$, the characteristic equation is $1 + L(s) = 1 + \mathcal{N}(s)/\mathcal{D}(s) = 0$, or $\mathcal{D}(s) + \mathcal{N}(s) = 0$. As the degree of $\mathcal{D}(s)$ is

greater than or equal to the degree of $\mathcal{N}(s)$, the degree of the characteristic equation will be equal to the degree of $\mathcal{D}(s)$, and the number of its roots will be equal to the number of poles of the open-loop. Thus with a change of the loop gain the root locus will have as many branches as the number of poles of the open-loop.

3. From relationship (5.15)

$$-k = \frac{\prod_{i=1}^{P}(s - p_i)}{\prod_{j=1}^{Z}(s - z_j)}; \quad P \geq Z. \tag{5.27}$$

$k = 0$ holds if $s = p_i$. Thus the root locus starts from the poles of the open-loop when $k = 0$. $k = \infty$ holds if $s = z_j$ or $s \to \infty$. Thus if $k \to \infty$ the roots of the characteristic equation run into the zeros of the open-loop, or if $P > Z$ then the number $P - Z$ of the roots goes to infinity.

4. If a point s of the root locus is on the real axis, the vectors connecting it with the complex conjugate poles (or zeros) make an angle of 0° or 360° considering the pairs, therefore they can be disregarded. The real poles or zeros if they are to the left of the considered point s, make an angle of 0°, while if they are to the right of the point, they make an angle of 180°. To fulfill the phase condition (5.17), the sum of the number of the poles and the zeros to the right of s has to be odd.

5. The asymptotes approach the very distant points of the root locus, from where the p_i poles and the z_j zeros of the open-loop are all seen under the same angle α:

$$P\alpha - Z\alpha = \mp N180^\circ \tag{5.28}$$

hence the angles of the asymptotes are

$$\alpha = \mp\frac{N180^\circ}{P - Z}; \quad N = 1, 3, 5, \ldots \tag{5.29}$$

6. Taking into consideration the poles with weight +1 and the zeros with weight −1, the crossing point of the asymptotes with the real axis is just at the center of gravity, as looking at the system from a longer distance it can be replaced by its center of gravity. The rule can be derived analytically as well.

7. The phase condition is fulfilled also for a point x where the root locus steps out or arrives at the real axis. According to Fig. 5.7 leaving the real axis with a small

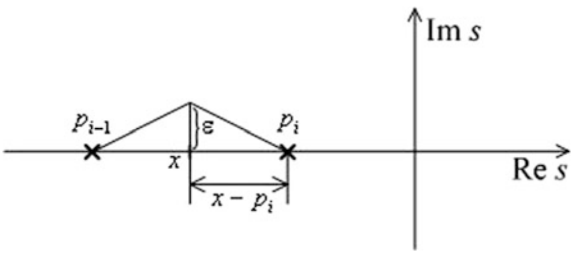


Fig. 5.7 Determination of the place where the root locus leaves the real axis

ε distance perpendicularly and replacing the small angles with their tangents the following relationship can be written:

$$\sum_{i=1}^{P}\frac{\varepsilon}{x-p_i}-\sum_{j=1}^{Z}\frac{\varepsilon}{x-z_j}=0 \tag{5.30}$$

Now (5.26) follows from (5.30).

8. On the borderline between stability and unstability, the characteristic equation has roots on the imaginary axis.

5.5.3 Examples of the Root Locus Method

Example 5.3 Let us consider the system given in Examples 5.1 and 5.2. The loop transfer function is $L(s)=K/s(1+s)(1+5s)$. A negative feedback of unity is applied. The loop transfer function in zero-pole form is

$$L(s)=\frac{k}{s(s+1)(s+0.2)},$$

where $k=0.2K$ is the loop gain. Determine the root locus. The varying parameter is the loop gain k.

On the basis of the construction rules it can be seen that the root locus has three branches. The branches start from $s_1=0$, $s_2=-0.2$, and $s_3=-1$, the poles of the loop transfer function, and go to infinity. On the real axis the root locus has a section between the points 0 and -0.2, and in the range between -1 and $-\infty$. Between the points 0 and -0.2 the root locus steps off of the real axis. The angle of the asymptotes going to infinity is $\alpha=\pm N180°/(3-0)$: at $N=1$ the angle is $\pm 60°$, and at $N=3$ it is $180°$. The asymptotes cross the real axis at $-1.2/3=-0.4$. The point where the root locus steps out of the real axis is calculated by solving equation $\frac{1}{x}+\frac{1}{x+1}+\frac{1}{x+0.2}=0$. The solutions are: $x_1=-0.7055$ and $x_2=-0.0945$. Only x_2 can be a solution, since the root locus may not have a point at x_1. Figure 5.8 shows the root locus. The critical loop gain k_{cr} can be determined from the characteristic equation by either the Routh or the Hurwitz criterion. The root locus crosses the imaginary axis at this gain. In Examples 5.1 and 5.2 its value was calculated by both methods. The stability range of the system is $0<K<1.2$ or $0<k<0.24$, respectively. The characteristic equation at the critical value $k_{cr}=k=0.24$ is:

$$s(s+1)(s+0.2)+0.24=s^3+1.2s^2+0.2s+0.24=0$$

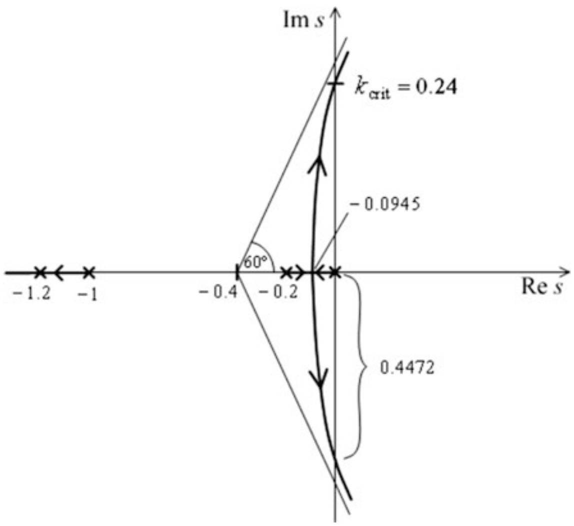


Fig. 5.8 Root locus of an integrating two lag element with negative unity feedback

Two of the roots are on the imaginary axis. Thus

$$\begin{aligned} s^3 + 1.2s^2 + 0.2s + 0.24 &= (s+\gamma)(s+j\eta)(s-j\eta) = (s+\gamma)\left(s^2+\eta^2\right) \\ &= s^3 + \gamma s^2 + \eta^2 s + \gamma\eta^2. \end{aligned}$$

Comparing the coefficients, we obtain

$$\gamma = 1.2 \quad \text{and} \quad \eta = \sqrt{0.2} = 0.4472.$$

The oscillation frequency is determined by the η interception with the imaginary axis. ■

Further examples for root loci

The root loci of some systems (without proper scaling) are shown in Table 5.2. Comparing the figures, it can be seen that a new pole pushes away the branches of the root locus, while a new zero attracts them.

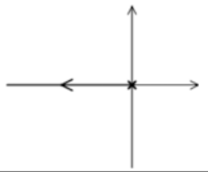
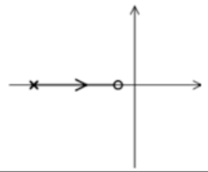
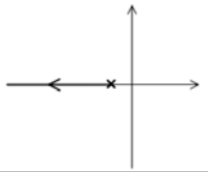
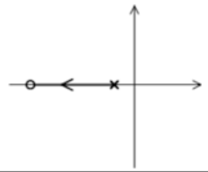
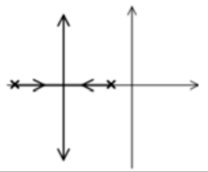
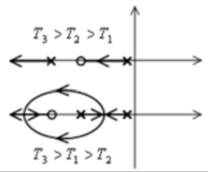
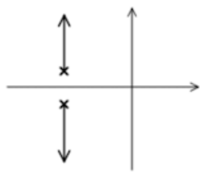
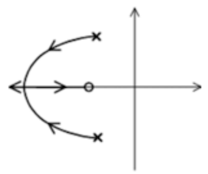
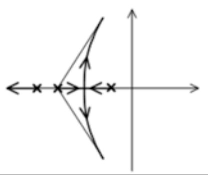
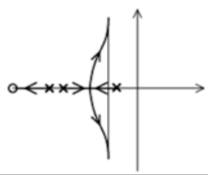
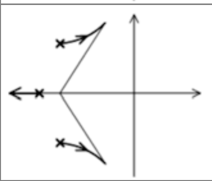
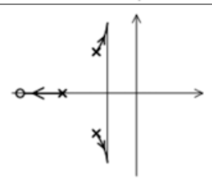
Figure 5.9 shows that in the case of three poles, the introduction of a zero, modifies the shape of the root locus. By appropriate location of the zero the closed-loop system can be stabilized over the whole range of the gain factor.

Figure 5.10 gives the root locus of an unstable open-loop system. The transfer function of the open-loop is

$$L(s) = \frac{k(s+1)}{s(s-1)(s+6)}.$$

This open-loop system has an unstable pole. Inserting an additional zero can ensure that the closed-loop system becomes stable with for appropriate choice of the gain $(k > k_{cr} = 7.5)$.

Table 5.2 Root loci of typical systems

P	Z			
	0		1	
	Transfer function	Root locus	Transfer functions	Root locus
1	$\frac{K}{s}$		$K\frac{1+sT_1}{1+sT_2}$ $T_1 > T_2$	
	$\frac{K}{1+sT}$		$K\frac{1+sT_1}{1+sT_2}$ $T_2 > T_1$	
2	$\frac{K}{(1+sT_1)(1+sT_2)}$		$\frac{K(1+sT_2)}{(1+sT_1)(1+sT_3)}$	
	$\frac{K}{1+s2\xi T+s^2T^2}$		$\frac{K(1+sT_1)}{1+s2\xi T+s^2T^2}$	
3	$\frac{K}{(1+sT_1)(1+sT_2)(1+sT_3)}$		$\frac{K(1+sT_4)}{(1+sT_1)(1+sT_2)(1+sT_3)}$	
	$\frac{K}{(1+s2\xi T_1+s^2T_1^2)(1+sT_2)}$		$\frac{K(1+sT_2)}{(1+s2\xi T_1+s^2T_1^2)(1+sT_3)}$	

The shape of the root locus shows an analogy to the electrostatic field. If positive and negative charges are located in a plane, the asymptotes of the electrostatic field take the shape of the root locus, if the positive charges are replaced by the poles and the negative charges by the zeros. (Generally the analogy with the potential field of sources and sinks can be considered.)

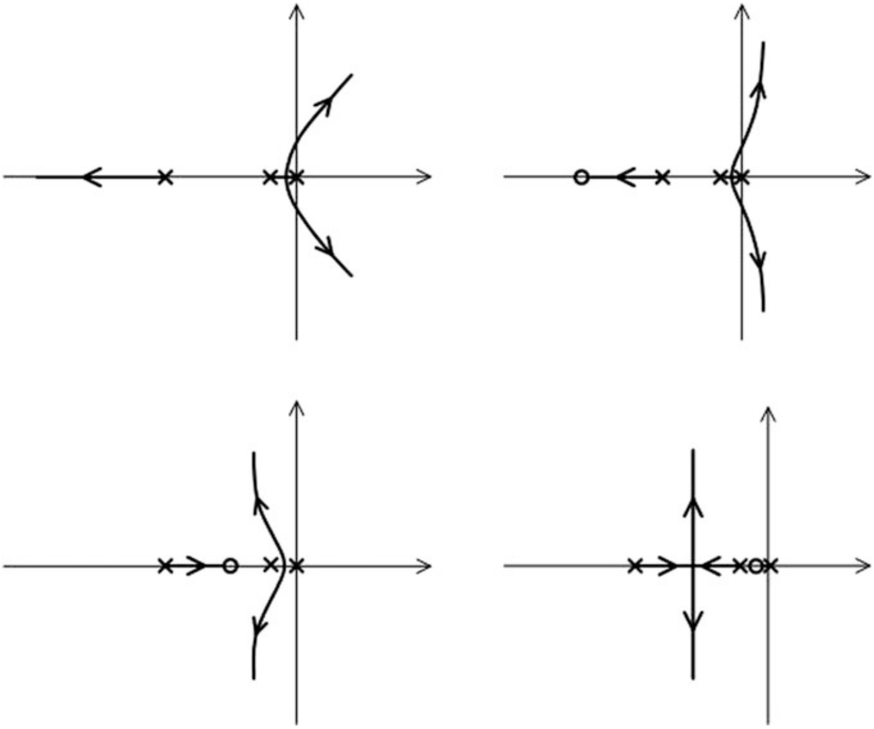
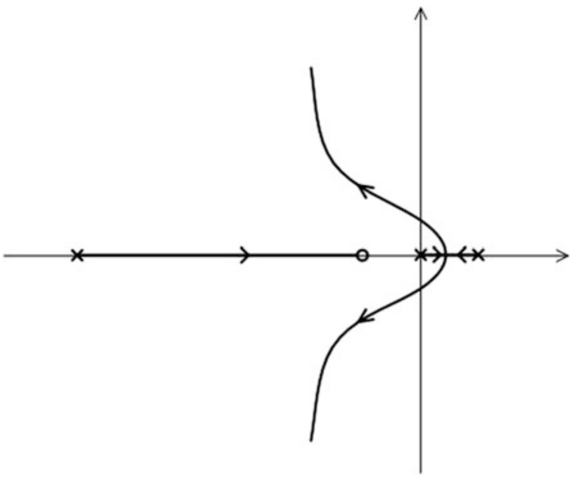


Fig. 5.9 The effect of a zero to the root locus

Fig. 5.10 Stabilization of an unstable open-loop with negative feedback by inserting a zero



5.5.4 Root Locus in the Case of Varying a Parameter Different from the Gain

If the root locus is to be determined as a function of a parameter different from the gain factor, then the characteristic equation has to be transformed to the form

$$\alpha H(s) = -1$$

where α is the varying parameter and $H(s)$ is the transfer function obtained as a result of the transformation. Drawing the root locus, α takes the role of the gain and $H(s)$ is a constructed loop transfer function.

Example 5.4 The procedure will be presented when the open-loop is a proportional element with two time lags where, instead of the gain factor, a pole (the time constant) of the system varies from zero to infinity. The transfer function of the open-loop is

$$L(s) = \frac{10}{(s+\alpha)(s+2)}.$$

The varying parameter is now alpha (the pole is $-\alpha$). The characteristic equation is

$$(s+\alpha)(s+2) + 10 = 0$$

or

$$s(s+2) + \alpha(s+2) + 10 = 0.$$

Rearranging yields

$$1 + \alpha\frac{s+2}{s^2+2s+10} = 0.$$

The root locus is determined for the transfer function

$$H(s) = \alpha\frac{s+2}{s^2+2s+10}$$

(see Fig. 5.11). It can be seen that for small values of α $(0<\alpha<8.3246)$ there are decaying oscillations in the closed-loop system. If α increases further, the transients is aperiodic. ■

Today's modern computer techniques make possible—beside the effect of the change of the loop gain—to observe the effect of an additional parameter as well. In this case the usual root locus is calculated for the discrete values of the other parameter (e.g., α), and an array (in layers) of curves is drawn in three dimensions

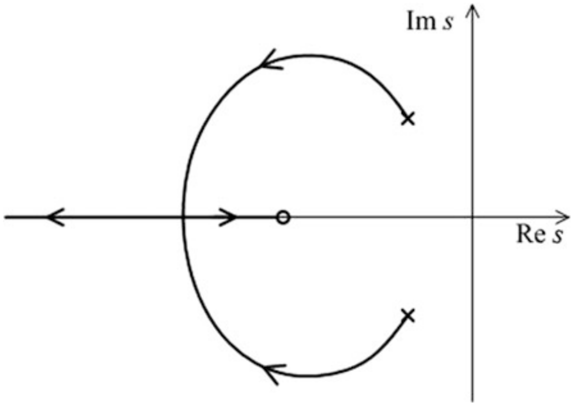


Fig. 5.11 Root locus of a proportional system with two time lags when one of its poles is varied

(3D). The fundamental two dimensions are represented by the complex plane itself, above it the further root-loci are plotted "in layers". Thus the third axis is for the variable α. For 3D graphical representation, a variety of powerful software tools are known, which makes it possible to depict very useful surfaces.

5.6 The Nyquist Stability Criteria

With the analytical Routh-Hurwitz stability criteria, the stability of a closed-loop control system can be determined based on the coefficients of the characteristic equation, but in the case of instability it is difficult to tell how to change the parameters of the system to ensure the appropriate dynamical performance.

The root locus gives an expressive picture of the change of the location of the roots of the closed-loop characteristic equation in the complex plane versus a parameter, thus a comprehensive view can be obtained of the stability and dynamical properties of the system.

With the Nyquist stability criterion, the stability of the *closed-loop* control system can be determined based on the frequency diagram of the *open-loop*. The method is expressive, and in the case of instability it can be easily determined how to modify expediently the structure and the parameters of the system. By appropriately forming the frequency function—i.e., introducing new zeros and poles—the prescribed properties of the closed-loop system, in addition to its stability, as well as its required static and dynamical properties can be ensured.

5.6.1 Illustration of the Evolution of Undamped Oscillations in the Frequency Domain

The characteristic equation of a closed-loop control system is $1+L(s)=0$, where $L(s)$ is the open-loop transfer function. Substituting $s=j\omega$ it can be checked

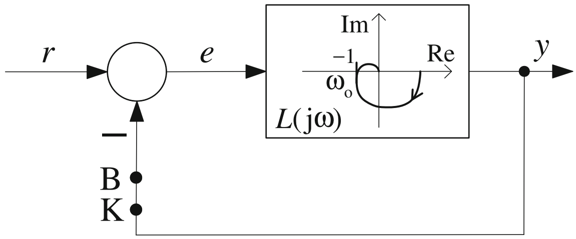


Fig. 5.12 Nyquist diagram of an open-loop control system, where the closed-loop is working at the stability limit

whether the equation has a solution on the imaginary axis. If there exists a frequency ω_o fulfilling the condition $1 + L(j\omega_o) = 0$, that is $L(j\omega_o) = -1$, then in the closed-loop system, an un-damped oscillation arises with this frequency, thus the system gets to the borderline of stability. In this case the Nyquist diagram of the open-loop goes through the point $-1 + 0j$ of the complex plane.

The evolution of un-damped oscillations can be illustrated as follows. Let us consider the control loop in Fig. 5.12. The Nyquist diagram of the open-loop goes through the $-1 + 0j$ point at frequency ω_o. Imagine that the system is opened at points B-K. Let the reference signal r be a sinusoidal signal with frequency ω_o. The system transfers this signal with the same amplitude but with opposite sign. If now the points B-K are connected again, because of the negative feedback the error signal e coincides with the sinusoidal input signal. This un-damped sinusoidal signal will be maintained in the system even if the reference signal is removed. Oscillations with this frequency do appear in the system even in the case when the reference signal is not the considered sinusoidal signal, but a different deterministic signal, e.g., a unit step. That is, since in the frequency spectrum of the reference signal all the frequencies do appear, the reference signal can be built from these sinusoidal components. A component of frequency ω_o is maintained in the system.

5.6.2 The Simple Nyquist Stability Criterion

Let us suppose that the transfer function of the open-loop has no poles on the right half of the complex plane, thus the open-loop is stable.

Let us draw the frequency function in the complex plane for the domain $-\infty < \omega < \infty$ (the complete Nyquist diagram). Go through the Nyquist diagram in the direction of increasing frequencies.

If the Nyquist diagram does not encircle the point $-1 + 0j$, the closed-loop control system is stable.
If the Nyquist diagram crosses the point $-1 + 0j$, the system is at the stability limit.
If the Nyquist diagram encircles the point $-1 + 0j$, the system is unstable.

In a simpler formulation, it is sufficient to draw the Nyquist diagram only for positive ω. If we go through the diagram from $\omega = 0$ to ∞, and the point $-1 + 0j$ is to the left of the curve, the closed-loop control system is stable. If the curve crosses

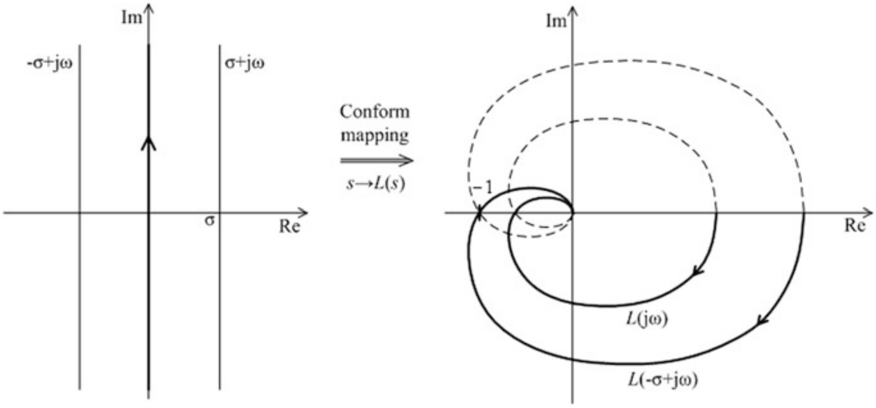


Fig. 5.13 The simple Nyquist stability criterion can be proved by conformal mapping

the point $-1+0j$, the system is at the stability limit. If the point $-1+0j$ is to the right of the curve, the system is unstable.

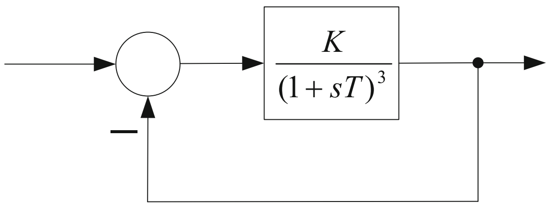
The simple Nyquist stability criterion can be proved based on conformal mapping. The Nyquist diagram of $L(j\omega)$ is the conformal mapping of the imaginary axis by the function $L(s)$ as ω changes between $-\infty$ and $+\infty$ (Fig. 5.13). Let us consider the straight lines $-\sigma+j\omega$ and $\sigma+j\omega$, which are parallel to the imaginary axis. Here, σ is a given positive number.

Conformal mapping preserves the angles and ratios. Therefore a conformal mapping of the straight line $-\sigma+j\omega$ according to $L(-\sigma+j\omega)$ lies to the left of the curve $L(j\omega)$, while conformal mapping of the straight line $\sigma+j\omega$ according to $L(\sigma+j\omega)$ lies to its right. So if the curve $L(j\omega)$ crosses the real axis to the right of the point $-1+0j$, and thus does not encircle it, then the equation $L(s_i)=-1$ can be fulfilled only for roots with negative real part, i.e. the transients are decreasing. In this case the closed-loop control system is stable. Similarly, if the curve $L(j\omega)$ crosses the real axis to the left of the point $-1+0j$, and thus encircles it, then the equation $L(s_i)=-1$ can be fulfilled only for roots with positive real part, therefore the amplitude of the transients is increasing and the system is unstable.

Example 5.5 Let us consider the closed-loop control circuit in Fig. 5.14. Let us determine the critical loop gain based on the Nyquist stability criterion.

Figure 5.15 shows the Nyquist diagram of the open-loop for the case of the stability limit. The Nyquist diagram goes through the point $-1+0j$ of the complex

Fig. 5.14 Stability analysis of a proportional system with three time lags



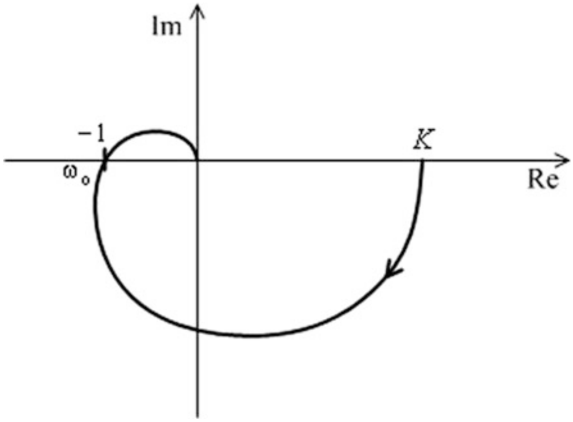


Fig. 5.15 NYQUIST diagram of a proportional system with three time lags at the stability limit

plane at frequency ω_o. At this frequency the phase angle of the frequency function is $-180°$ and its absolute value is 1. So we have

$$\varphi(\omega_o) = -3\mathrm{arctg}(\omega_o T) = -180°,$$

whence $\omega_o T = \sqrt{3}$. Thus $K_{krit} = \left(\sqrt{1+\omega_o^2 T^2}\right)^3 = 8$, which does not depend on the value of the time constant T. ■

Example 5.6 The NYQUIST stability criterion can also be applied to systems with dead-time. Let us consider a control system containing dead-time (see Fig. 5.3). The NYQUIST diagram of the open-loop is a circle with radius K which keeps on circling itself infinitely many times as the frequency increases (Fig. 5.16). At the

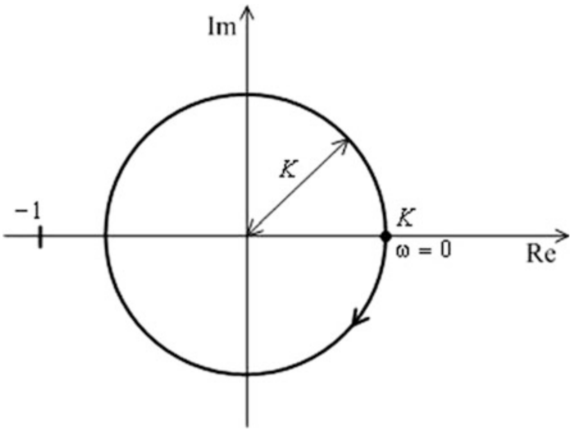


Fig. 5.16 NYQUIST diagram of a pure dead-time system

stability limit, it crosses the point $-1+0j$, thus $K_{\mathrm{crit}}=1$, in agreement with Fig. 5.4., and the convergence condition given in Table 5.1. ■

5.6.3 The Generalized NYQUIST Stability Criterion

The generalized NYQUIST stability criterion gives a condition for stability even for the case when the open-loop has poles in the right half-plane, i.e., the open-loop is unstable. The question is whether the closed-loop can be stabilized with negative feedback.

The generalized NYQUIST stability criterion can be formulated as follows: If the open-loop is unstable and the number of its poles lying in the right half-plane is P, then the closed-loop control system is asymptotically stable if the complete NYQUIST diagram $(-\infty<\omega<\infty)$ of the open-loop encircles the point $-1+0j$ counter-clockwise (considered as the positive direction) as the number of the poles of the open-loop is in the right half-plane (i.e., P times).

The complete NYQUIST diagram is given now more precisely than in the formulation in the previous subsection. In the s plane the straight line $s=j\omega$ $(-\infty<\omega<\infty)$ is closed with a half-circle on the right side with infinite radius, as in Fig. 5.17. The conformal mapping of this closed curve by the function $L(s)$ gives the complete NYQUIST diagram of the open-loop. (If the degree of the denominator of the rational fraction $L(s)$ is higher than the degree of its numerator, then the half-circle of infinite radius is mapped into the zero point.) If $L(s)$ has a pole on the

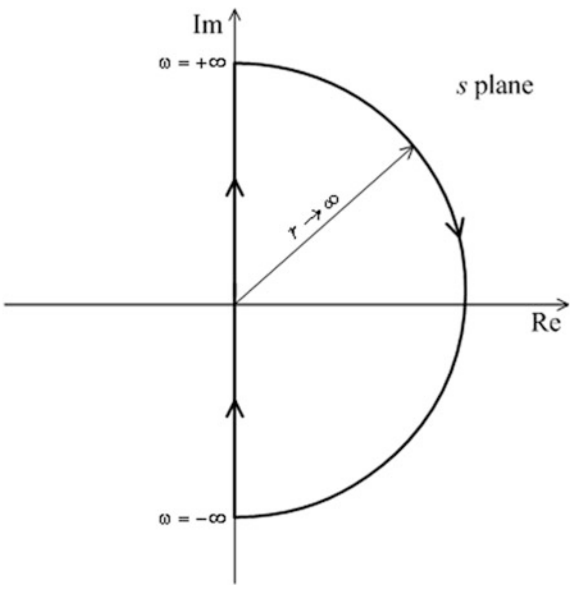


Fig. 5.17 Creating the complete NYQUIST diagram

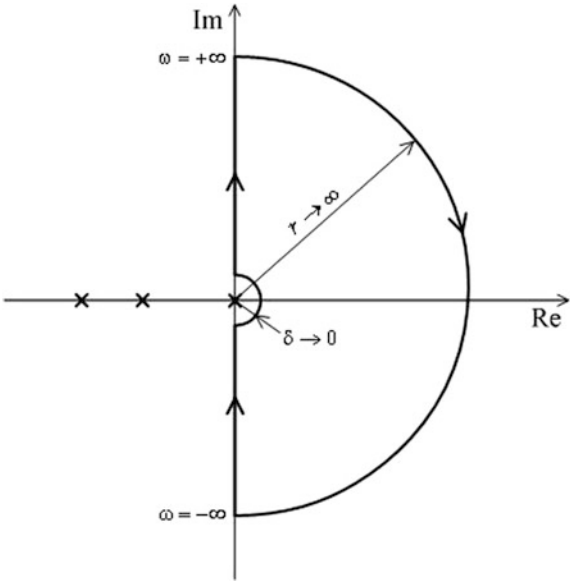


Fig. 5.18 The closed curve to be mapped when $L(s)$ has a pole on the imaginary axis

imaginary axis, then the closed curve is modified to get around the given point from the right or from the left with a half-circle of infinitesimal radius δ. If the curve gets around the pole on the imaginary axis from the right as pictured in Fig. 5.18., then the pole can be considered as a pole in the left half-plane. If the roundabout is executed from the left, then the pole is considered as being on the right side.

The generalized Nyquist stability criterion can be demonstrated through considerations related to complex functions. Let $f(s)$ be the following function of the complex variable s: $f(s) = (s - s_o)^m$, where s_o is a given point. Let us investigate how the vector $f(s)$ changes if the final point of the vector $s - s_o$ goes through a closed curve on the s-plane clockwise, where on this curve the function $f(s)$ is regular (differentiable).

If s_o is inside the closed curve (Fig. 5.19a), then the vector $s - s_o$ starting from an initial point and passing through the curve clockwise gets into its original position, and its phase angle changes by -2π. In the meantime the mapping by the function $f(s)$ rotates from the starting point by an angle of $-m2\pi$ on the curve determined by $f(s)$. This curve encircles the origin a total of m turns clockwise (m is positive) or counter-clockwise (m negative) (Fig. 5.19b). But if the point s_o is outside the closed curve (Fig. 5.19c), passing through the closed curve the angle of vector $s - s_o$ first is increasing in one direction, then it is decreasing with the same value in the other direction, and finally the curve described by $f(s)$ does not encircle the origin (Fig. 5.19d).

Let us apply the above considerations to the characteristic function of a closed-loop control system. Let the transfer function of the open-loop be a rational

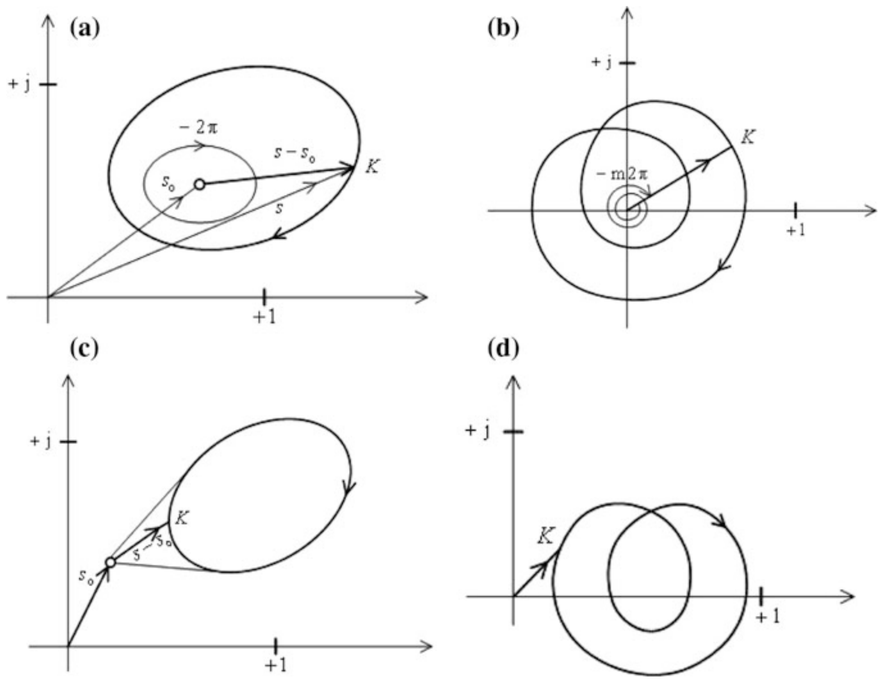


Fig. 5.19 Considerations in the complex plane

fraction whose numerator and denominator are the polynomials $\mathcal{N}(s)$ and $\mathcal{D}(s)$, so that, $L(s) = \mathcal{N}(s)/\mathcal{D}(s)$. The characteristic function is then

$$1 + L(s) = 1 + \frac{\mathcal{N}(s)}{\mathcal{D}(s)} = \frac{\mathcal{D}(s) + \mathcal{N}(s)}{\mathcal{D}(s)} = k\frac{(s - z_1)(s - z_2)\ldots(s - z_n)}{(s - p_1)(s - p_2)\ldots(s - p_n)}. \quad (5.31)$$

The roots of the numerator are denoted by z_i, which are the zeros of $1 + L(s)$. The roots of the denominator are denoted by p_i, which are the poles of $1 + L(s)$. Here, k is a constant. The poles of $1 + L(s)$ coincide with the poles of the transfer function of the open-loop. (Multiple poles appear in the expression when there are multiple, i.e. repeated factors.)

Let us consider the closed curve on the complex plane shown in Fig. 5.17. Go through the curve on the imaginary axis from $-\infty$ to $+\infty$, then close the curve with a half-circle on the right half plane whose radius tends to infinity. Map this curve according to the characteristic function given by (5.31). For all the factors in Eq. (5.31), the above considerations related to complex functions are valid. (For the zeros $s_o = z_1, z_2, \ldots, z_n$ and $m = 1$, while for the poles $s_o = p_1, p_2, \ldots, p_n$ and $m = -1$.) The phase angle of $1 + L(s)$ is the sum of the phase angles of the individual factors taken with the appropriate signs. If the function $1 + L(s)$ has Z zeros and P poles in the right half-plane, inside of the curve in Fig. 5.17, then the

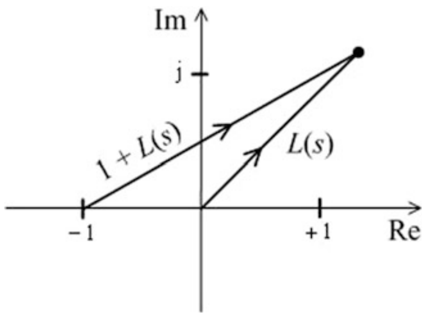


Fig. 5.20 Relationship of vectors $L(s)$ and $1+L(s)$

number of times the conformal mapping by the function $1+L(s)$ of the considered closed curve encircles the origin clockwise is the difference of the number of zeros and the number of poles inside this curve. The difference between the phase angles of the initial and the final states is $-2\pi(Z-P)$, and the number R of windings around the origin is

$$R = P - Z \tag{5.32}$$

where a counterclockwise encirclement is defined as positive (see the detailed derivation in A.5.1. of Appendix A.5).

Simply consider the function $1+L(s)$ as if looking at the curve produced by $L(s)$ from $-1+0j$ (Fig. 5.20). The mapping of $L(s)$ along the closed curve in Fig. 5.17 (the so-called complete Nyquist diagram) encircles the point $-1+0j$ point $R = P - Z$ times.

Now, P is the number of the poles of the characteristic function in the right half-plane. But these poles, according to (5.31), coincide with the right side unstable poles of the open-loop. Also, Z is the number of the zeros of the characteristic equation in the right half-plane. In the case of stable behavior, the characteristic equation has no zeros in the right half-plane. Thus the condition for stability is

$$Z = 0 \quad \text{i.e.} \quad R = P. \tag{5.33}$$

The simple Nyquist stability criterion can be derived from the generalized Nyquist stability criterion. If the open-loop system has no poles in the right half-plane, i.e. if, $P = 0$, the closed-loop is stable if $R = 0$, so the Nyquist diagram does not encircle the point $-1+0j$. In most practical cases the open-loop is stable, and it is in the closed-loop system that the feedback may cause unstable behavior. But sometimes unstable processes have to be dealt with, that is they are to be

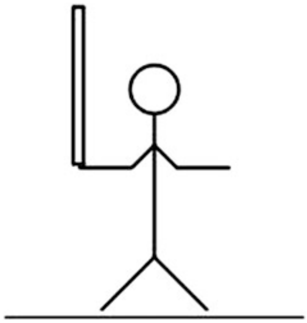


Fig. 5.21 A juggler can balance the rod underpinned at its bottom edge

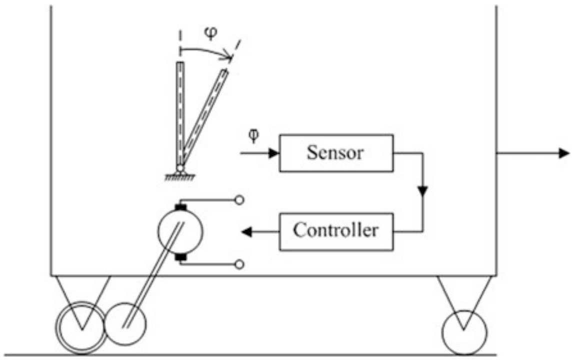


Fig. 5.22 Stabilizing the motion of an inverted pendulum

stabilized by control systems with negative feedback. For example the inverted pendulum is an unstable process. A juggler in the circus is able to balance the leaning rod using the appropriate motions, which are faster than the dynamics of the rod (Fig. 5.21). Thus his body realizes a controller in a closed-loop control system. The automatic solution for stabilizing the motion of the inverted pendulum is shown in Fig. 5.22.

5.6.4 Examples of the Application of the Nyquist Stability Criteria

Example 5.7 Consider the open-loop transfer function

$$L(s) = \frac{5}{1-s} = -\frac{5}{s-1}.$$

Let us analyze the stability of the closed-loop control system.

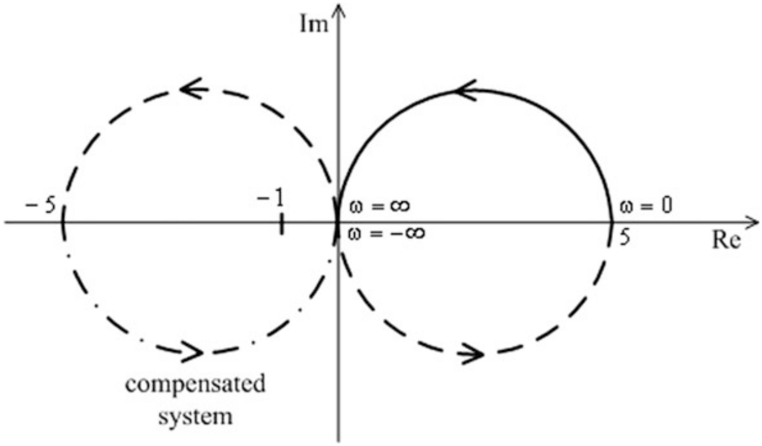


Fig. 5.23 Stability analysis of an unstable system with negative feedback

The system has a pole in the right half-plane, thus $P = 1$. The Nyquist diagram is shown in Fig. 5.23. As the Nyquist diagram does not encircle the point $-1 + 0j$, $R = 0$, thus the closed-loop is unstable.

The system can be stabilized if a so called compensation element of a constant gain by $A = -1$ is connected into the forward path. This element changes the sign of the points of the Nyquist diagram reflecting it about the origin (dashed-dotted curve). Thus the number of windings around $-1 + j0$ will be $R = P = 1$. ■

Example 5.8 Let us consider for example the case when the open-loop is an $L(s) = K_I/s$ integrator, whose pole is at the origin. The closed curve is created by getting around the pole from the right. By mapping this curve according to $L(s)$ the complete Nyquist diagram shown in Fig. 5.24a is obtained. The case involving getting around the pole from the left is demonstrated in Fig. 5.24b. In the s-plane, the points denoted by 1, 2 and 3 on the small circle surrounding the pole are mapped into the points $1'$, $2'$ and $3'$ in the $L(s)$-plane. In case (a) $P = 0$ and $R = 0$, in case (b) $P = 1$ and $R = 1$, thus in both cases the stable behavior of the system can be established. ■

Example 5.9 Let the transfer function of an open-loop be a proportional element with three time lags, $L(s) = K/[(1 + sT_1)(1 + sT_2)(1 + sT_3)]$. The poles $p_1 = -1/T_1, p_2 = -1/T_2, p_3 = -1/T_3$ are all in the left half-plane, thus $P = 0$. Let us apply the generalized Nyquist stability criterion. The complete Nyquist diagram obtained by mapping of the curve given in Fig. 5.17 is shown in Fig. 5.25. If the Nyquist diagram goes through $-1 + j0$, the system is at the stability limit. If the Nyquist diagram does not include the point $-1 + j0$ (K_1 loop gain), $R = P = 0$, thus the control system is stable. If the Nyquist diagram includes the point $-1 + 0j$ (K_2 loop gain), $R \neq P$, thus the control system is unstable. To determine the number of windings R, let us put the spike of an imaginary compass on the point $-1 + 0j$, and with the other end of the compass pass through the Nyquist diagram from $\omega = -\infty$ to $+\infty$. The number of windings is $R = -2$ (clockwise). The

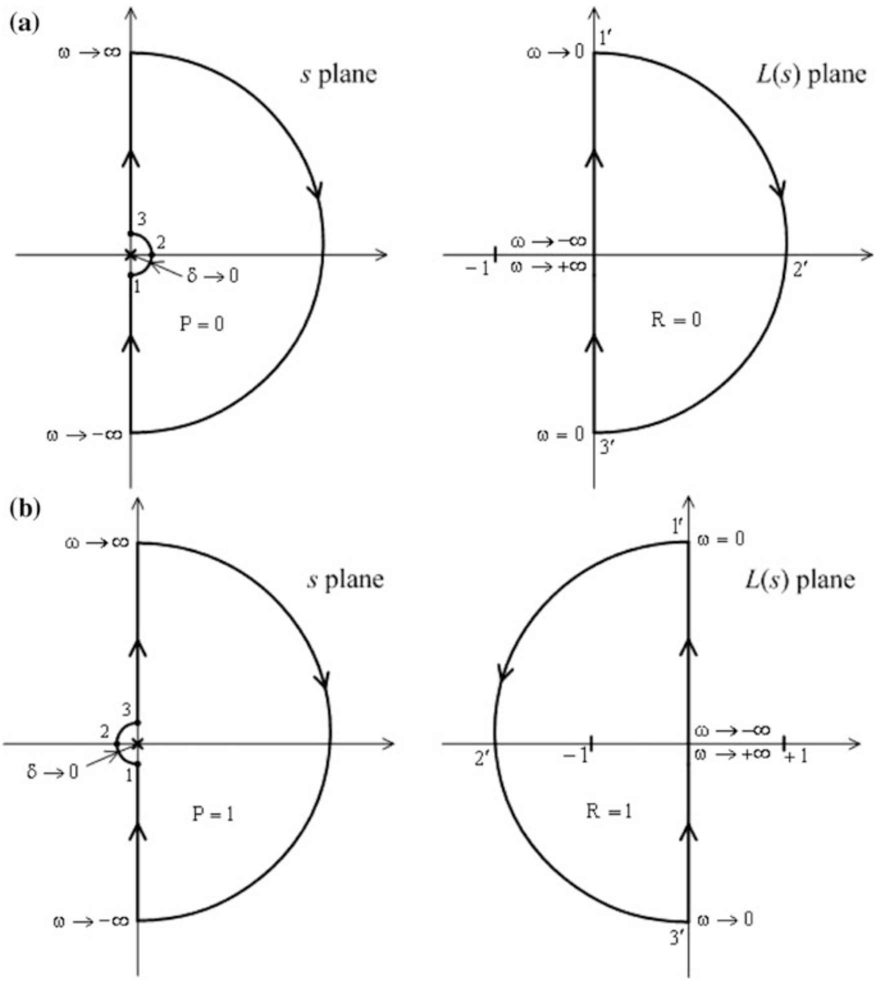


Fig. 5.24 Stability analysis of a control circuit (an integrator is fed back by a unity constant gain)

characteristic equation has two roots in the right half-plane, so, $Z = 2$, and as $R = -2 = P - Z = 0 - Z$, in this case the system is unstable. ■

In the case of a stable open-loop, it is sufficient to use the simple NYQUIST stability criterion. In the stable case $-1 + 0j$ lies to the left of the NYQUIST diagram drawn for positive frequencies, whereas in the unstable case it is to the right of that curve. The simplified stability investigation can be applied also to the cases when the open-loop contains integrators, and thus there are poles at the origin.

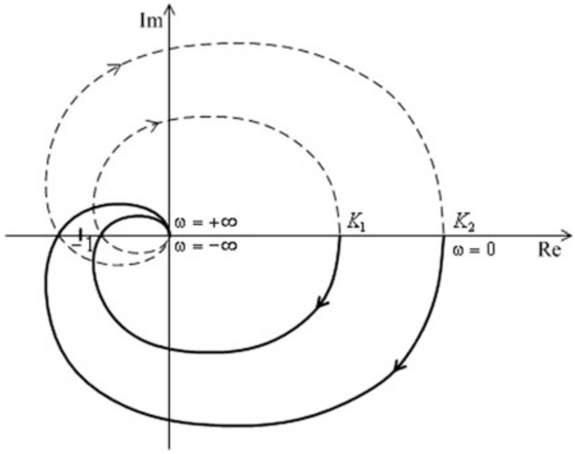


Fig. 5.25 Stability analysis of a proportional system with three time lags

5.6.5 *Practical Stability Measures*

In case of a stable open-loop, the closed-loop is stable if the NYQUIST diagram of the open-loop does not encircle the point $-1+0j$. It can be said that the system has a certain amount of stability reserve, if the NYQUIST diagram is kept sufficiently far from the point $-1+0j$.

Some measures can be defined indicating how far is the NYQUIST diagram of the open-loop from the point $-1+0j$. Such measures include the *phase margin*, the *gain margin*, the *modulus margin* and the *delay margin*.

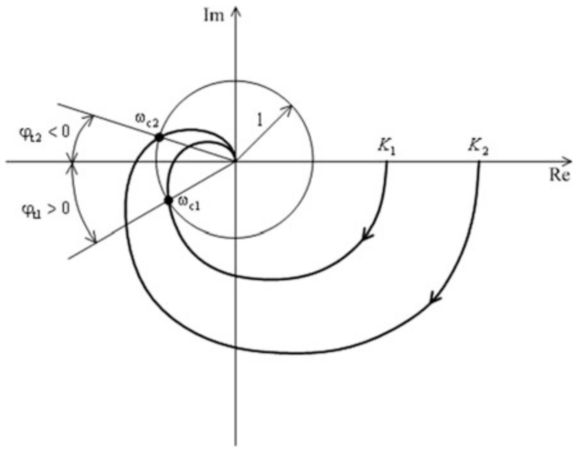


Fig. 5.26 Interpretation of the phase margin

Phase Margin

Let us draw the Nyquist diagram of the open-loop for positive frequencies. Let us determine the intersection point of the Nyquist diagram with the circle of unit radius. The frequency belonging to this point is called the *cut-off frequency* and is denoted by ω_c. Let us connect the origin and the intersection point with a straight line. The angle formed by this straight line with the negative real axis is called the *phase margin* (Fig. 5.26):

$$\varphi_t = \varphi(\omega_c) + 180^\circ = \arg L(j\omega_c) + 180^\circ. \tag{5.34}$$

If the phase margin is positive, the system is stable. If the phase margin is zero, the system is at the stability limit. If the phase margin is negative, the system is unstable.

Thus for the stability of the control system the following statements can be made:

$$\begin{array}{ll} \varphi_t > 0 & \text{Stable system} \\ \varphi_t = 0 & \text{Boundary of stability} \\ \varphi_t < 0 & \text{Unstable system} \end{array} \tag{5.35}$$

The stability of the system can be evaluated based on the phase margin as a single measure only if the Nyquist diagram of the open-loop crosses the unit circle only once.

Gain Margin

Let us determine the intersection point of the Nyquist diagram with the negative real axis and also the distance $\kappa = |1 + L(j\omega_{180})|$ of this point from the point $-1 + 0j$ (Fig. 5.27). the distance κ is called the *gain margin*. It is apparent that for $\kappa > 0$ the stability domain of the simple Nyquist criterion is obtained. The stability

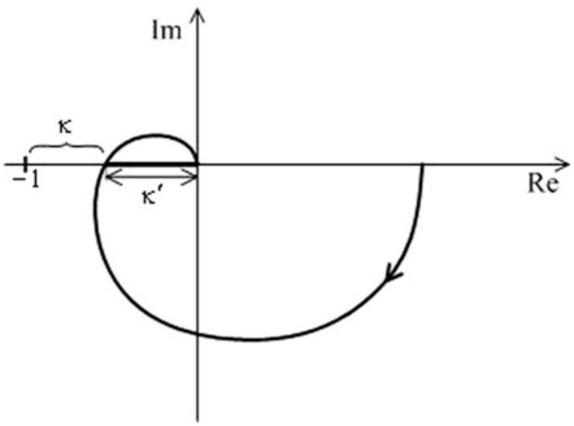


Fig. 5.27 Interpretation of the gain margins

of the system can be evaluated based on the gain margin as a single measure only if the NYQUIST diagram of the open-loop crosses the negative real axis only once.

The *modified gain margin* κ', is defined by the intercept $\kappa' = L(j\omega_{180}) = 1 - \kappa$ seen in Fig. 5.27. If $\kappa' < 1$, the system is stable. If $\kappa' = 1$, the system is at the stability limit. If $\kappa' > 1$, the system is unstable. Thus for the stability of the control system the following statements can be made:

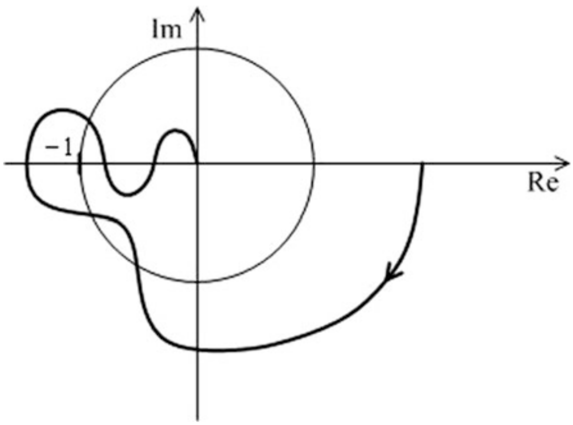
$$\begin{array}{ll} \kappa' < 1 & \text{Stable system} \\ \kappa' = 1 & \text{Boundary of stability} \\ \kappa' > 1 & \text{Unstable system} \end{array} \tag{5.36}$$

The meaning of κ is more expressive than that of κ', however the reciprocal of κ' specifies the factor by which multiplying the actual loop gain the system reaches the stability limit. Therefore it is straightforward to also use the measure $g_t = 1/\kappa' = 1/|L(j\omega_{180})|$ as the relative gain margin. Multiplying the loop gain by g_t the value of the critical gain is obtained. With simple considerations, the inequalities $g_t \geq M_m/(M_m - 1)$ and $\varphi_t \geq 2\arcsin(1/M_m)$ can be derived. (See (4.25) for the interpretation of M_m.)

Figure 5.28 shows the NYQUIST diagram of a system where neither the phase margin nor the gain margin can be interpreted. (Such a NYQUIST diagram is formed if oscillating elements and zeros are included in the transfer function of the system.) In this case the whole NYQUIST diagram has to be considered. Based on the simple NYQUIST stability criterion the stability can be evaluated: as going through the curve the point $-1 + 0j$ is to the right side of the curve, the system is unstable.

Besides stability, the relevant transient performance is also required. To ensure an overshoot less than 10% in the step response of a closed-loop system, the desired phase margin is about 60°, and the desired relative gain margin g_t is about 2 ($\kappa \approx \kappa' \approx 0.5$). These values can be considered characteristic if there are no resonant frequencies in the loop frequency function.

Fig. 5.28 When the phase margin and the gain margin can not be interpreted



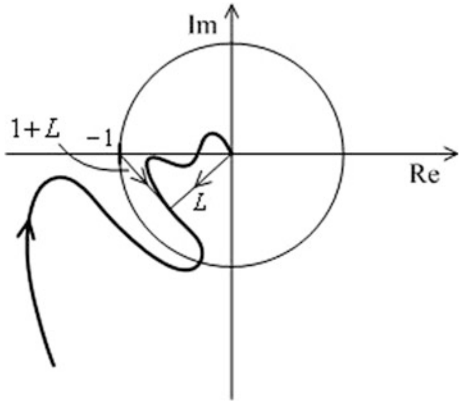


Fig. 5.29 When the dynamic behavior of the system is not satisfactory even if the phase margin and the gain margin are of appropriate values

The appropriate phase margin and gain margin do not give reliable information about the stability margin of the system in every case. Let us consider, e.g. the NYQUIST diagram in Fig. 5.29. In spite of the fact that both the phase margin and the gain margin are appropriate, there may occur high amplifications in the frequency function $L/(1+L)$ of the closed-loop in the vicinity of the cut-off frequency, as with increasing frequency the amplitude of L is only slightly changed, while the amplitude of $1+L$ decreases significantly. High amplification in the amplitude-frequency curve of the closed-loop in the vicinity of the cut-off frequency indicates oscillations in the unit step response. Furthermore, if the parameters of the plant change a little, the closed-loop system may even become unstable. The phase margin and the gain margin characterize the stability properties of the system only if the NYQUIST diagram does not go too close to the unit circle before and after the cut-off frequency.

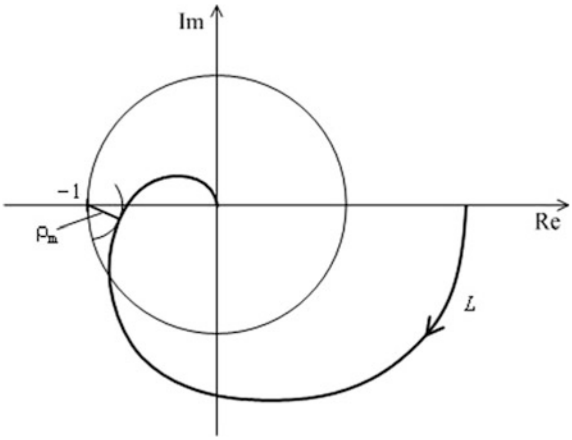


Fig. 5.30 Interpretation of the modulus margin

Modulus Margin

The ρ_m is the minimum of the distance of point $-1+0j$ from the Nyquist diagram, i.e., it is the radius of the smallest circle tangential to the diagram and centered at $-1+0j$ (Fig. 5.30). The modulus margin shows how far the most sensitive point of the system is from the stability limit. As a reasonable prescription, let the modulus margin be $\rho_m > 0.5$.

The modulus margin is also called Nyquist stability margin. An important formula is that ρ_m can be expressed as the reciprocal of the maximum of the absolute value of the sensitivity function (see Chap. 6):

$$\rho_m = \frac{1}{\max\limits_{\omega}|S(j\omega)|} = \min\limits_{\omega}\left|S^{-1}(j\omega)\right| = \min\limits_{\omega}|1+L(j\omega)| \tag{5.37}$$

The three margins $\varphi_t, \kappa, \rho_m$ are analogous concepts, as each of them tries to guarantee somehow the distance from the point $-1+0j$.

Delay Margin

The *delay margin* gives the smallest value of the dead-time T_{min} by which—inserting it serially as an extra dead-time into the loop—the closed-loop control system would reach the stability limit. The delay margin can be calculated from the phase margin measured in radians by the following formula:

$$T_{min} = \frac{\varphi_t}{\omega_c}, \tag{5.38}$$

where ω_c denotes the cut-off frequency.

With these stability margins, not only can the stability be evaluated, but it can also be established, "how far" the system is from the stability limit.

5.6.6 Structural and Conditional Stability

Let us suppose that the open-loop is stable, thus the stability of the closed-loop can be evaluated according to the simplified Nyquist stability criterion.

Most systems generally are stable for small loop gains: they reach the stability limit at a given critical gain, then increasing the gain further they show unstable behavior. (Such a control system is obtained by negative feedback of a proportional plant with three time lags, see Examples 5.5 and 5.9.)

But there are also systems which—because of their structure—remain stable at any value of the loop gain. Such systems are called structurally stable systems. For example a first order or a second order lag element or a pure integrator or an integrator serially connected to a first order lag with negative constant feedback have this property, as their Nyquist diagram does not encircle the point $-1+0j$ even if the loop gain is arbitrarily increased. By increasing the loop gain the system

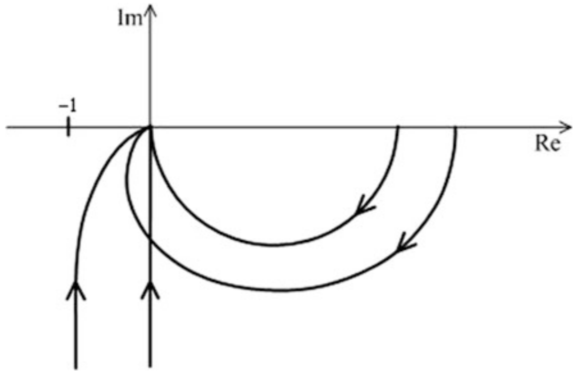


Fig. 5.31 Nyquist diagrams of structurally stable systems

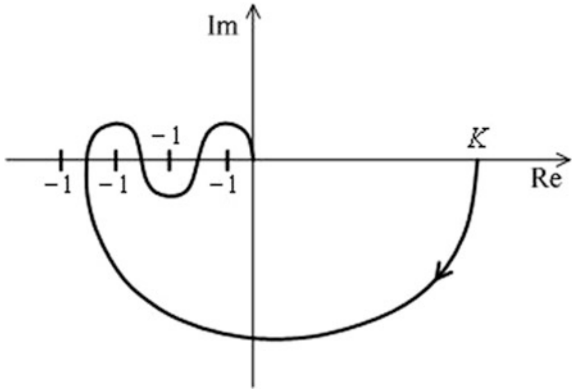


Fig. 5.32 Nyquist diagram of a conditionally stable system

will not become unstable, but its stability margins do decrease. The Nyquist diagrams of such systems are shown in Fig. 5.31.

There are systems which are stable in given regions of the loop gain, while in other regions they show unstable behavior. In the case of such systems the loop gain has to be set carefully. These systems are called conditionally stable systems. Figure 5.32 shows an example of the Nyquist diagram of a conditionally stable system. Besides being influenced by the poles, the course of the Nyquist diagram is influenced by the zeros as well. If the gain is small, the point $-1+0j$ is to the left of the Nyquist diagram, so for small gains the control system is stable. By increasing the gain, the point $-1+0j$ will be to the right of the diagram, so the control system becomes unstable. By increasing the gain further, the point $-1+0j$ will get to the left of the diagram, so the control system will become stable again. Increasing the gain even more the diagram will encircle again the point $-1+0j$, causing again unstable performance.

5.6.7 *Stability Criteria Based on the* Bode *Diagrams*

The phase margin and the gain margin can also be read from the Bode diagram. The absolute value of the frequency function at the cut-off frequency ω_c is 1. The Bode amplitude-frequency diagram crosses the horizontal 0 dB axis at this frequency. The deviation of the phase angle from $-180°$ at this frequency gives the phase margin. The absolute value at the frequency where the phase angle is $\varphi = -180°$ gives the value of the parameter κ in dB-s, and from this the gain margin can be determined (Fig. 5.33).

If the open-loop is of minimum phase (i.e. its transfer function does not contain zeros or poles in the right half-plane), and furthermore the control system does not contain dead-time, the stability can be determined very simply from the approximate Bode amplitude-frequency curve of the open-loop.

Then from the Bode amplitude diagram the course of the phase angle follows unambiguously, as the phase angle belonging to the poles is negative and the phase angle belonging to the zeros is positive, changing according to arctangent curves.

A minimum phase system which does not contain dead-time is stable if the asymptotic Bode amplitude diagram of the open-loop crosses the frequency axis at a straight line section of slope −20 dB/decade. The system is surely unstable if the slope of the crossing is equal to or greater than −60 dB/decade. If the slope of the intersection is −40 dB/decade, then the system can be stable or unstable depending on the phase margin, which in this case is surely very small (Fig. 5.34).

The above statement can be shown based on the following deliberations. Let us consider the asymptotic Bode diagram in Fig. 5.35. The cut-off frequency lies on a straight line of slope $-20\text{dB}/\text{decade}$. The phase angle in the vicinity of the cut-off

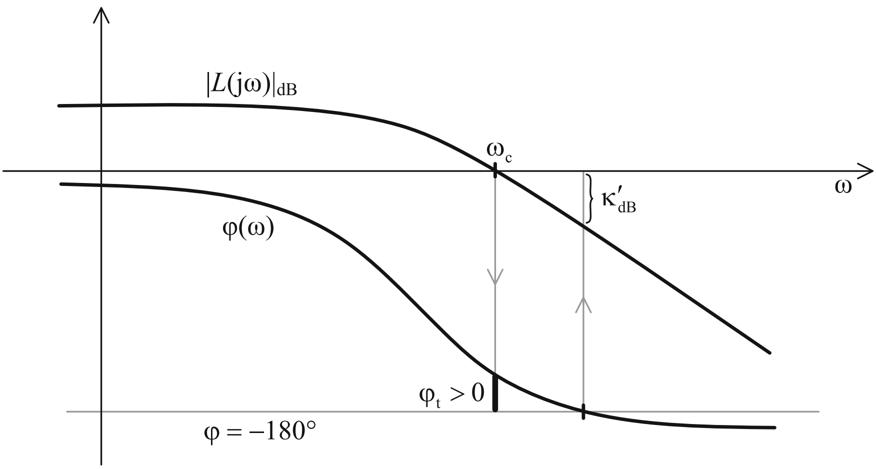


Fig. 5.33 Reading the phase margin and the gain margin from the Bode diagram

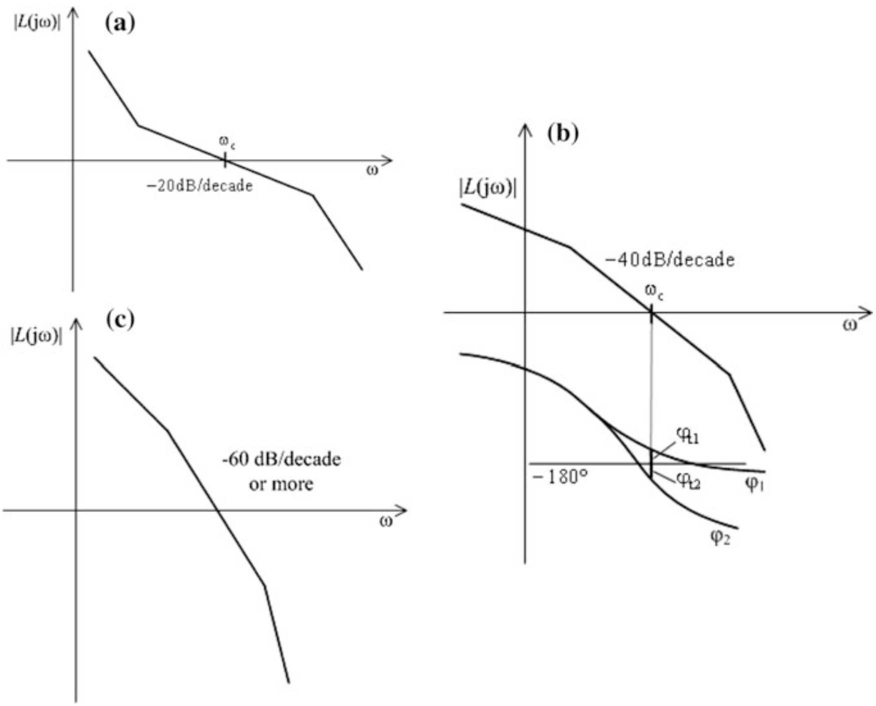


Fig. 5.34 Stability of a minimum phase system without dead-time can be determined from the approximate Bode amplitude-frequency diagram

frequency ω_c approximates $-90°$. The phase angle resulting from a breakpoint which is to the right of the cut-off frequency (especially if it is far away from ω_c, at least located at $5\omega_c$ or at a still higher frequency) will only slightly affect the phase angle at ω_c. Before the straight line of slope $-20\text{dB}/\text{decade}$ the approximate Bode diagram might have a horizontal section or a section with slope −20 or −40 dB/decade. The courses of the phase angle resulting from these parts of the Bode diagram are indicated in the figure by dashed, dotted and dashed-dotted lines, respectively. The effect of these sections on the phase angle at frequency ω_c is small (especially if the breakpoint before ω_c is far away, located at less than $\omega_c/5$). Thus the system surely has a positive phase margin, which is expected to be satisfactory not only for ensuring stability, but also for guaranteeing the appropriate transient behavior.

If the cut-off frequency is located at a straight line of slope −40 dB/decade, the phase angle at frequency ω_c may approach or even exceed $-180°$ with the phase angle resulting from the previous breakpoints. Thus the system will get close to the stability limit (Fig. 5.36). In this case, evaluating the stability requires calculating the value of the phase margin.

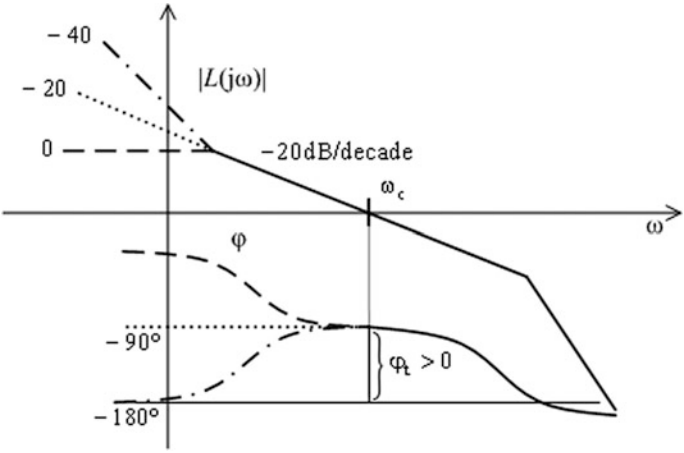


Fig. 5.35 The system surely has positive phase margin if the cut-off frequency is located on a straight line of slope −20 dB/decade

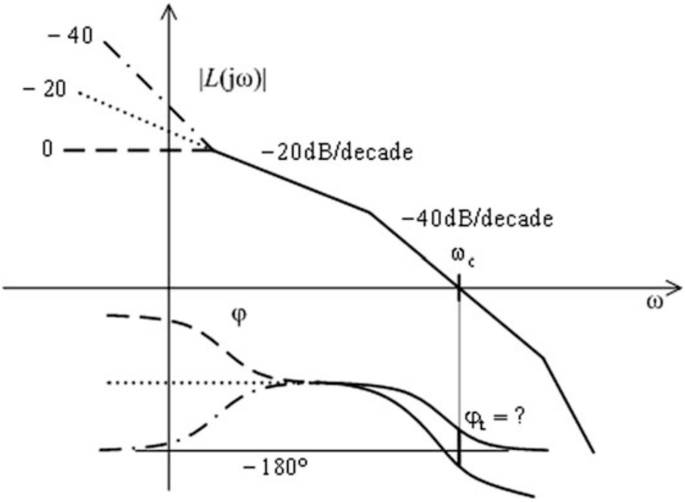


Fig. 5.36 The system is close to the stability limit if the cut-off frequency is located on a straight line of slope −40 dB/decade

If the cut-off frequency is located on a straight line of slope −60 dB/decade or more, then the phase margin surely will become negative.

Thus, to ensure stability, the cut-off frequency has to be located on a straight line of slope −20 dB/decade. (This section has to be long enough to ensure a satisfactory phase margin of about 60°.)

If the open-loop is of non-minimum phase type, or contains also dead-time, then the stability can not be evaluated considering only the BODE amplitude diagram. In this case the BODE amplitude and phase diagram have to be jointly taken into account.

5.7 Robust Stability

Generally, the parameters of the plant are determined from measurement data. The parameters may change around their nominal values in a given range. The closed-loop control system has to be stable throughout the given uncertainty ranges of the parameters.

Suppose that the open-loop is stable. The controller designed for the nominal plant ensures the stability of the nominal closed-loop control system. Let us analyze whether the system remains stable with the parameter uncertainties of the open-loop. Stability is maintained if the NYQUIST diagram of the modified open-loop does not encircle the point $-1+0j$.

The uncertainty of the plant is expressed by the absolute model error

$$\Delta P = P - \hat{P} \tag{5.39}$$

and the relative model error

$$\ell = \frac{\Delta P}{\hat{P}} = \frac{P - \hat{P}}{\hat{P}}, \tag{5.40}$$

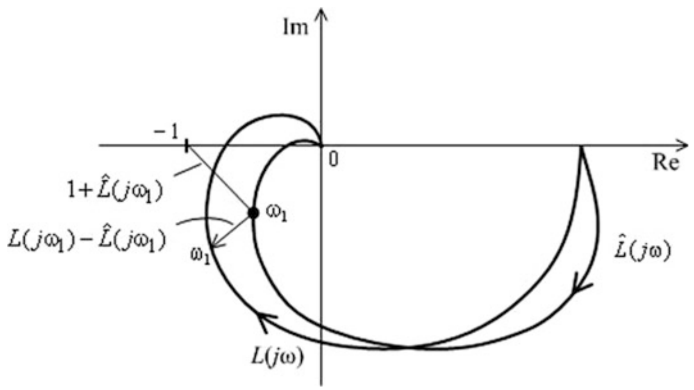


Fig. 5.37 Change of the NYQUIST diagram of an uncertain system

where $\hat{P}$ is the available nominal model used for the design, and P is the real plant. If there is an uncertainty of ΔP (or parameter change) in the transfer function of the plant, then applying the same controller this uncertainty appears in the absolute error $\Delta L = C\Delta P$ of the loop transfer function, while its relative error

$$\ell_{\mathrm{L}} = \frac{\Delta L}{\hat{L}} = \frac{L - \hat{L}}{\hat{L}} = \frac{CP - C\hat{P}}{C\hat{P}} = \frac{P - \hat{P}}{\hat{P}} = \ell \tag{5.41}$$

is equal to the relative model error. Here $\hat{L}$ denotes the nominal, while L denotes the real loop transfer function.

Robust stability means, that the closed-loop control system should not reach an unstable behavior even in the worst case of the parameter changes. The bound for ΔL can be formulated based on Fig. 5.37 taking into account simple geometrical considerations: the NYQUIST diagram will not encircle the point $-1 + 0j$, if the following relationship is satisfied for all the frequencies:

$$|\Delta L(j\omega)| = |\ell(j\omega)|\left|\hat{L}(j\omega)\right| < \left|1 + \hat{L}(j\omega)\right| \quad \forall\omega. \tag{5.42}$$

With further straightforward manipulations on (5.42) the necessary and sufficient condition for robust stability can be obtained as

$$|\ell(j\omega)| < \left|\frac{1 + \hat{L}(j\omega)}{\hat{L}(j\omega)}\right| = \frac{1}{\left|\hat{T}(j\omega)\right|} \quad \forall\omega, \tag{5.43}$$

where $\hat{T} = \hat{L}/\left(1 + \hat{L}\right)$ is the nominal supplementary sensitivity function. Condition (5.43) can also be expressed as

$$\left|\hat{T}(j\omega)\right| < \frac{1}{|\ell|} \quad \forall\omega. \tag{5.44}$$

It is a common practice to express the above inequalities for robust stability also in the following form:

$$\left|\hat{T}(j\omega)\right||\ell| < 1 \quad \forall\omega. \tag{5.45}$$

This form is also called the dialectic relationship of robust stability. In the design process, the first factor $\left|\hat{T}(j\omega)\right|$ is calculated for the supposed (known) nominal parameters of the plant, thus it depends on the designer. The second factor $|\ell|$ does not (or only partly) depend on the designer, as it contains the uncertainties in the knowledge of the plant or its unexpected parameter changes. In those frequency ranges where the uncertainty is large, unfortunately only a small transfer gain can be designed for the closed-loop. Where $\left|\hat{T}(j\omega)\right|$ is high, very accurate information has to be available to guarantee a small error. The higher the absolute value of the complementary sensitivity function, the smaller the permissible parameter uncertainty.

Condition (5.43), which considers the whole frequency range, is fairly strict, therefore generally it is replaced by a more practical condition if the maximum value of $\left|\hat{T}(j\omega)\right|$ is known. Suppose

$$\hat{T}_{\mathrm{m}} = \max_{\omega}\left|\hat{T}(j\omega)\right| \tag{5.46}$$

With this value, (5.43) can be simplified to the following satisfactory condition:

$$|\ell(j\omega)| < \frac{1}{\hat{T}_{\mathrm{m}}} \quad \forall\omega \tag{5.47}$$

(Let us refer to Chap. 4, where $M(\omega)$ is defined as $M(\omega) = |T(j\omega)|$.)

If the open-loop is unstable, and the feedback stabilizes the nominal system, then the closed-loop system remains stable with the parameter uncertainties if the number of the poles of the open-loop in the right half-plane does not change, and the number of windings of the Nyquist diagram around the point $-1 + 0j$ does not change either.