Frequency Response Analysis

Introduction

We have examined the use of test signals such as a step, an impulse and a ramp signal. In this chapter we will use a sinusoidal input signal and consider the steady-state response of the system as the frequency of the sinusoid is varied. Thus, we will look at the response of the system to a changing frequency, ω.

We will examine the transfer function $G(j\omega)$, when $s = j\omega$ and develop several forms of plotting the complex number for $G(j\omega)$ when ω is varied, [1]

These plots provide insight regarding the performance of a system. We are able to develop several time-domain performance measures in terms of the frequency response of a system.

Industrial control systems are often designed by use of frequency-response methods. Frequency-response tests are, in general, simple and can be made accurately by use of readily available sinusoidal generators and precise measurement equipment. Often, the transfer function of complicated components can be determined experimentally by frequency-response tests.

The frequency response of a system is defined as the steady-state response of the system to a sinusoidal input signal. The sinusoid is a unique input signal and the resulting output signal, for a linear system, as well as signals throughout the system, is sinusoidal in the steady-state; it differs from the input waveform only in amplitude and phase angle, [1].

Consider the system shown in Figure 1 where the input $r(t)$ is a sinusoid: $r(t) = A sin\omega t$. The Laplace transform of the input is:
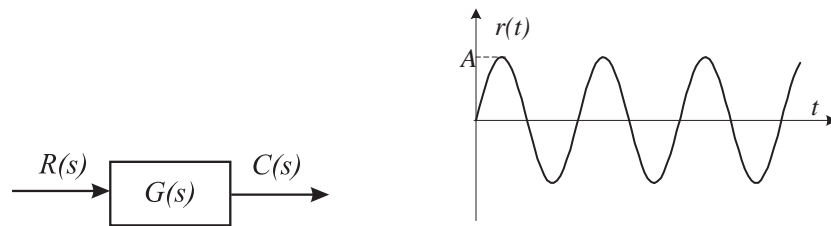

Figure 1: System with sinusoidal input

$$R(s) = \frac{A\omega}{s^2 + \omega^2}$$

and the transfer function:

$$G(s) = \frac{m(s)}{q(s)} = \frac{m(s)}{\prod_{i=1}^{n}(s + p_i)}$$

Then, the output $C(s)$ is given by:

$$C(s) = G(s)R(s)$$

or, in partial fraction form:

$$C(s) = \frac{k_1}{s + p_1} + ... + \frac{k_n}{s + p_n} + \frac{\alpha s + \beta}{s^2 + \omega^2}$$

1

If the system is stable, then all p_i have negative nonzero real parts and the terms corresponding to the poles in $c(t)$ will be zero at steady-state, or:

$$\lim_{t\to\infty} \mathcal{L}^{-1}\left[\frac{k_i}{s+p_i}\right] = 0$$

In the limit, for $c(t)$, we obtain for $t \to \infty$ (the steady -state):

$$c(t) = \mathcal{L}^{-1}\left[\frac{\alpha s + \beta}{s^2 + \omega^2}\right] = \frac{1}{\omega}|A\omega G(j\omega)|sin(\omega t + \varphi) = A|G(j\omega)|sin(\omega t + \varphi)$$

where $\varphi = \angle G(j\omega)$.

Example Consider a first-order system with the transfer function

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

The input is sinusoidal: $r(t) = A\sin\omega t$. We shall calculate the output signal $c(t)$ at steady state $(t \to \infty)$.

$$R(s) = \mathcal{L}[A sin\omega t] = \frac{A\omega}{s^2 + \omega^2}$$

$$C(s) = R(s)G(s) = \frac{A\omega}{(s^2 + \omega^2)(s + 1)} = \frac{A\omega}{\omega^2 + 1}\frac{1}{s+1} + \frac{A\omega}{\omega^2 + 1}\frac{1-s}{s^2 + \omega^2}$$

$$c(t) = \mathcal{L}^{-1}[C(s)] = \frac{A\omega}{\omega^2 + 1}\mathcal{L}^{-1}\left[\frac{1}{s+1} + \frac{1-s}{s^2 + \omega^2}\right]$$

$$c(t) = \frac{A\omega}{\omega^2 + 1}e^{-t} + \frac{A\omega}{\omega^2 + 1}\mathcal{L}^{-1}\left[\frac{1-s}{s^2 + \omega^2}\right]$$

At steady state $(t \to \infty)$, the exponential $e^{-t} \to 0$. The steady state of the output signal is given by:

$$c(t) = \frac{A\omega}{\omega^2 + 1}\mathcal{L}^{-1}\left[\frac{1-s}{s^2 + \omega^2}\right] = \frac{A\omega}{\omega^2 + 1}\mathcal{L}^{-1}\left[\frac{\omega}{s^2 + \omega^2}\frac{1}{\omega} - \frac{s}{s^2 + \omega^2}\right]$$

$$c(t) = \frac{A\omega}{\omega^2 + 1}\left[\frac{1}{\omega}\sin\omega t - \cos\omega t\right]$$

If we replace $s = j\omega$ into $G(s)$ we obtain a complex quantity with the magnitude and phase angle:

$$|G(j\omega| = \left|\frac{1}{j\omega + 1}\right| = \frac{1}{\sqrt{\omega^2 + 1}}$$

$$\angle G(j\omega) = \varphi = \angle\frac{1}{j\omega + 1} = -\arctan\omega$$

$$\tan\varphi = -\omega; \quad \frac{\sin\varphi}{\cos\varphi} = -\omega; \quad \cos\varphi = \frac{1}{\sqrt{\omega^2 + 1}}$$

By replacing these results into $c(t)$:

$$c(t) = \frac{A\omega}{\omega^2 + 1}\frac{\sin\omega t - \omega\cos\omega t}{\omega} = \frac{A}{\omega^2 + 1}\frac{1}{\cos\varphi}[\sin\omega t \cos\varphi + \sin\varphi\cos\omega t]$$

$$c(t) = \frac{A}{\omega^2 + 1}\frac{1}{\cos\varphi}\sin(\omega t + \varphi) = \frac{A}{\sqrt{\omega^2 + 1}}\sin(\omega t + \varphi)$$

2

$$c(t) = A|G(j\omega)| \sin{(\omega t + \angle G(j\omega))}$$

Thus, the steady-state output signal depends only on the magnitude and phase of $G(j\omega)$ at a specific frequency ω. Notice that the steady-state response as described before is true only for stable systems, $G(s)$.

We see that a stable linear time-invariant system subjected to a sinusoidal input will, at steady-state, have a sinusoidal output of the same frequency as the input. But the amplitude and phase of the output will be different from those of the input. In fact, the amplitude of the output is given by the product of that of the input and $|G(j\omega)|$ while the phase angle differs from that of the input by the amount $\varphi = \angle G(j\omega)$. An example of input and output sinusoidal signals is shown in Figure 2.
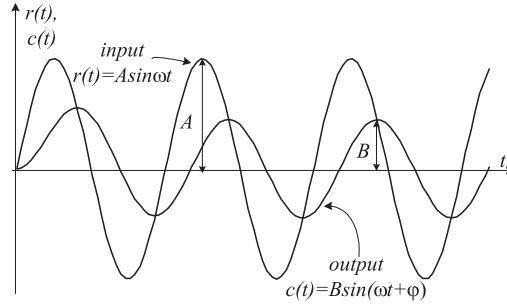

Figure 2: Input and output sinusoidal signals

Thus, for sinusoidal inputs the amplitude ratio of the output sinusoid to the input sinusoid is:
$$|G(j\omega| = \left| \frac{C(j\omega)}{R(j\omega)} \right|$$
and the phase shift of the output sinusoid with respect to the input sinusoid:
$$\angle G(j\omega) = \angle \frac{C(j\omega)}{R(j\omega)}$$

Hence, the response characteristics of a system to a sinusoidal input can be obtained directly from:
$$\frac{C(j\omega)}{R(j\omega)} = G(j\omega)$$

The function $G(j\omega)$ is called **the sinusoidal transfer function**. It is a complex quantity and can be represented by the magnitude and phase angle with frequency as a parameter. A negative phase angle is called **phase lag** and a positive phase angle is called **phase lead.**

The sinusoidal transfer function of a linear system is obtained by substituting $j\omega$ for s in the transfer function of the system. To completely characterize a linear system in the frequency domain, we must specify both the amplitude ratio and the phase angle as functions of the frequency ω.

Example. Consider the system shown in Figure 3.

The transfer function $G(s)$ is
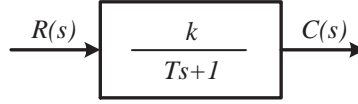$$G(s) = \frac{k}{Ts + 1}$$

Figure 3: First-order system

For the sinusoidal input $r(t) = A \sin \omega t$, the steady-state output $c_{ss}(t)$ can be found as follows:
- Substituting $j\omega$ for s in $G(s)$ yields:

$$G(j\omega) = \frac{k}{jT\omega + 1}$$

- The amplitude ratio of the output is:

$$|G(j\omega)| = \frac{k}{\sqrt{1 + T^2\omega^2}}$$

while the phase angle of the output φ is

$$\varphi = \angle G(j\omega) = -arctan T\omega$$

Thus, for the input $r(t)$, the steady-state output $c_{ss}(t)$ can be obtained as:

$$c_{ss}(t) = \frac{Ak}{\sqrt{1 + T^2\omega^2}} \sin(\omega t - arctan T\omega)$$

Logarithmic plots. Bode diagrams

The sinusoidal transfer function, a complex function of the frequency ω is characterized by its magnitude and phase angle, with frequency as parameter. A sinusoidal transfer function may be represented by two separate plots, one giving the magnitude versus frequency and the other the phase angle versus frequency.

The logarithmic plots are called Bode plots in the honor of H.W.Bode who used them extensively in his studies of feedback amplifiers.

The transfer function in the frequency domain is:

$$G(j\omega) = |G(\omega)| e^{j\Phi(\omega)}$$

A Bode diagram consists of two graphs: one is the plot of the logarithm of the magnitude of a sinusoidal transfer function, and the other one is a plot of the phase angle, both are plotted against the frequency in logarithmic scale, or *decades*:

$$\omega^{dec} = log_{10}\omega$$

The standard representation of the logarithmic magnitude of $G(j\omega)$, denoted M^{dB}, is $20 log_{10} |G(j\omega)|$. The unit used in this representation is the decibel, usually abbreviated dB:

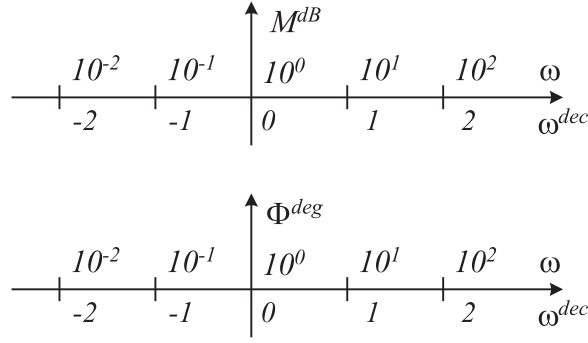$$M^{dB} = |G(j\omega)|^{dB} = 20 log_{10} |G(j\omega)|$$

4

Figure 4: Axes for a Bode diagram

The phase angle of the transfer function, $\phi(\omega)$, is represented (in degrees or radians) versus the logarithmic frequency.

The axes for a Bode diagram are shown in Figure 4.

The main advantage of using the logarithmic plot is that multiplication of magnitudes can be converted into addition. Furthermore, a simple method for sketching an approximate log-magnitude curve is available. It is based on asymptotic approximations.

The general form of a sinusoidal transfer function emphasizing first and second order factors as well as the poles (or zeros) at the origin, is:

$$G(j\omega) = \frac{k \prod_{i=1}^{m_1}(T_i(j\omega) + 1) \prod_{p=1}^{m_2}(\frac{1}{\omega_p^2}(j\omega)^2 + \frac{2\zeta_p}{\omega_p}(j\omega) + 1)}{(j\omega)^n \prod_{l=1}^{n_1}(T_l(j\omega) + 1) \prod_{k=1}^{n_2}(\frac{1}{\omega_k^2}(j\omega)^2 + \frac{2\zeta_k}{\omega_k}(j\omega) + 1)}$$

The logarithmic magnitude of $G(j\omega)$, expressed in decibels, is:

$$M^{dB} = |G(j\omega)|^{dB} = 20logk + 20\sum_{i=1}^{m_1} log|T_i(j\omega) + 1| +$$

$$+20\sum_{p=1}^{m_2} log|(\frac{1}{\omega_p^2}(j\omega)^2 + \frac{2\zeta_p}{\omega_p}(j\omega) + 1| - 20log|j\omega|^n -$$

$$20\sum_{l=1}^{n_1} log|T_l(j\omega) + 1| - 20\sum_{k=1}^{n_2} log|\frac{1}{\omega_k^2}(j\omega)^2 + \frac{2\zeta_k}{\omega_k}(j\omega) + 1|$$

The phase angle of $G(j\omega)$ can be calculated as:

$$\Phi = \angle(G(j\omega)) = \angle\frac{k}{(j\omega)^n} + \sum_{i=1}^{m_1} \angle(T_i(j\omega) + 1) +$$

$$+\sum_{p=1}^{m_2} \angle((\frac{1}{\omega_p^2}(j\omega)^2 + \frac{2\zeta_p}{\omega_p}(j\omega) + 1) - \sum_{l=1}^{n_1} \angle(T_l(j\omega) + 1) -$$

$$-\sum_{k=1}^{n_2} \angle(\frac{1}{\omega_k^2}(j\omega)^2 + \frac{2\zeta_k}{\omega_k}(j\omega) + 1)$$

Therefore three different kinds of factors that may occur in a transfer function are as follows:

5

1. Gain k and the integral or derivative factors $k/(j\omega)^n$, (n can be positive or negative)

2. First-order factors, $(T(j\omega) + 1)^{\pm1}$

3. Quadratic factors $[(\frac{1}{\omega_n^2}(j\omega)^2 + \frac{2\zeta}{\omega_n}(j\omega) + 1)]^{\pm1}$

It is possible to construct composite plots for any general form of $G(j\omega)$ by sketching the plot for each factor and adding individual curves graphically because adding the logarithm of the gains correspond to multiplying them together.

1. Gain k and the integral or derivative factors $k/(j\omega)^n$. The logarithmic gain, in decibels, is written as:

$$M_1^{dB} = 20log\left|\frac{k}{(j\omega)^n}\right| = 20logk - 20nlog\omega = k^{dB} - 20n\omega^{dec}$$

In a coordinate system (M_1^{dB}, ω^{dec}) as shown in figure 4, this is the equation of a straight line which crossed the vertical axis at k^{dB} and has the slope $-20n$ (dB/dec), because

$$\frac{dM_1^{dB}}{d\omega^{dec}} = -20n, \quad (dB/dec)$$

The phase angle is:

$$\Phi_1 = -n \cdot arctan\frac{\omega}{0} = -90^o \cdot n$$

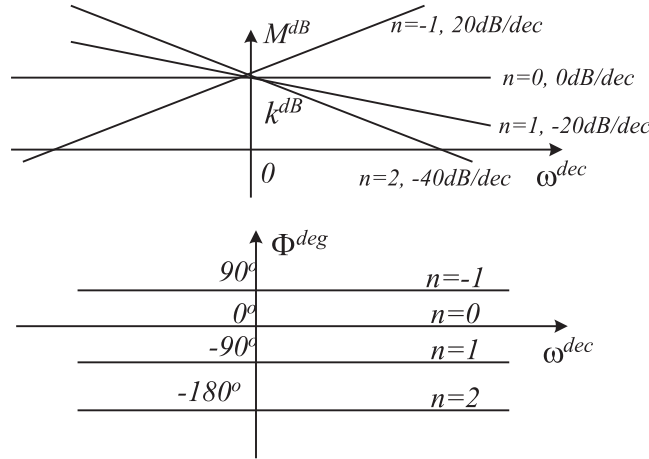The Bode diagram corresponding to integral or derivative factors plus the gain is shown in Figure 5.

Figure 5: Logarithmic plot for integral and derivative factors

2. First-order factors, $(T(j\omega) + 1)^{\pm1}$

   The log-magnitude of the first-order factor $1/(T(j\omega) + 1)$ is:

   $$M_2^{dB} = 20log\left|\frac{1}{Tj\omega + 1}\right| = -20log\sqrt{T^2\omega^2 + 1}$$

6

For low frequencies, such that $\omega \ll 1/T$, the log-magnitude may be approximated by:

$$M_2^{dB}|_{\omega \ll} \cong -20log1 = 0 \ dB$$

Thus, the log-magnitude curve at low frequencies is the constant 0 dB line. For high frequencies, such that $\omega \gg 1/T$, the magnitude in decibels may be approximated by:

$$M_2^{dB}|_{\omega \gg} \cong -20log\omega T \ dB.$$

or

$$M_2^{dB}|_{\omega \gg} \cong -20log\omega - 20logT = -20\omega^{dec} - 20logT$$

Since the representation is made for M^{dB} versus ω^{dec}, this is the equation of a straight line with the slope:

$$\frac{dM_2^{dB}}{d\omega^{dec}} = -20, \ (dB/dec)$$

The frequency ω_c, at which the two asymptotes meet is called the *corner frequency* or *break frequency*, and is calculated from:

$$M_2^{dB}|_{\omega \ll} = M_2^{dB}|_{\omega \gg}$$

or

$$0 = -20\omega_c^{dec} - 20logT$$

and we get:

$$\omega_c^{dec} = -logT = log\frac{1}{T}$$

The error in the magnitude curve caused by the use of asymptotes can be calculated. The maximum error occurs at the corner frequency and is calculated by replacing the expression of the corner frequency in the exact relation for log-magnitude:

$$M_2^{dB}|_{\omega = \omega_c} = 20log\sqrt{1 + T \cdot \frac{1}{T}} = 20log\sqrt{2} \cong 3.03 \ dB$$

Since the asymptotes are quite easy to draw and are sufficiently close to the exact curve, the use of such approximations in drawing Bode diagrams is convenient in establishing the general nature of the frequency-response characteristics.

The exact phase angle Φ_2 of the factor $1/(Tj\omega + 1)$ is:

$$\Phi_2 = -arctan \ \omega T$$

At zero frequency, the phase angle is 0°. At the corner frequency, the phase angle is

$$\Phi_2 = -arctan\frac{T}{T} = -arctan \ 1 = -45^o$$

At infinity, when $\omega \to \infty$ the phase angle becomes:

$$\Phi_2 = -arctan \ \infty = -90^o$$

Since the phase angle in given by an inverse-tangent function, the phase angle is skew symmetric about the inflexion point at -45°.
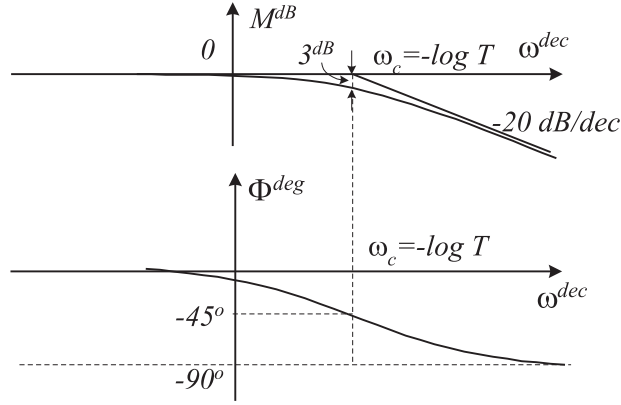
Figure 6: Logarithmic plot for first-order factors

The exact log-magnitude curve, the asymptotes and the phase angle curve are shown in Figure 6.

An advantage of the Bode diagram is that for the reciprocal factors, for example, the factor $Tj\omega + 1$, the log-magnitude and the phase angle curves need only be changed in sign. Since

$$20log\,|T\omega + 1| = 20log\left|\frac{1}{T\omega + 1}\right|$$

and

$$\angle(Tj\omega + 1) = -\angle\left(\frac{1}{T\omega + 1}\right)$$

the corner frequency is the same for both cases. The slope of the high-frequency curve is $20\,dB/dec$ and the phase angle varies from 0 to 90° as the frequency is increased from zero to infinity. The log-magnitude, together with the asymptotes and the phase angle curve for the factor $Tj\omega + 1$ are shown in Figure 7.
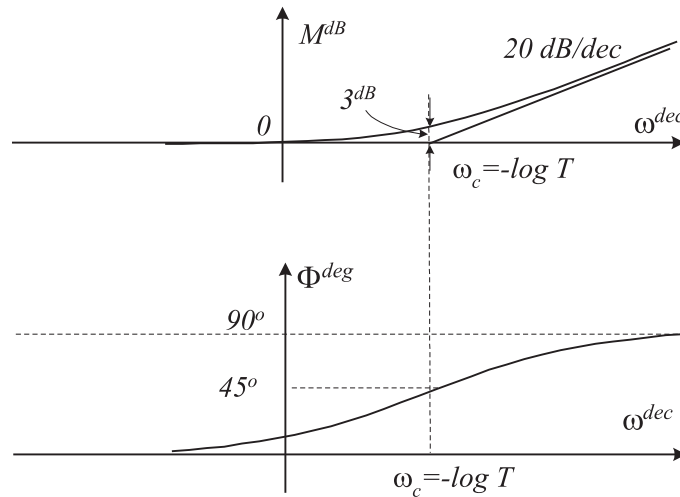

Figure 7: Logarithmic plot for first-order factors

3. Quadratic factors $[(\frac{1}{\omega_n^2}(j\omega)^2 + \frac{2\zeta}{\omega_n}(j\omega)+1)]^{\pm 1}$ Control systems often posses quadratic factors of the form:

$$\frac{1}{\frac{1}{\omega_n^2}(j\omega)^2 + \frac{2\zeta}{\omega_n}(j\omega) + 1}$$

If $\zeta > 1$ this quadratic factor can be expressed as a product of two first-order ones with real poles. If $0 < \zeta < 1$, this quadratic factor has complex-conjugate poles. The asymptotic frequency-response curve may be obtained as follows: Since:

$$M_3^{dB} = 20log \left| \frac{1}{\frac{1}{\omega_n^2}(j\omega)^2 + \frac{2\zeta}{\omega_n}(j\omega) + 1} \right| = -20log\sqrt{(1 - \frac{\omega^2}{\omega_n^2})^2 + (2\zeta\frac{\omega}{\omega_n})^2}$$

for low frequencies such that $\omega \ll \omega_n$, the log-magnitude becomes:

$$M_3^{dB}|_{\omega \ll} = -20log1 = 0dB$$

The low-frequency asymptote is thus a horizontal line at $0dB$. For high frequencies such that $\omega \gg \omega_n$, the log-magnitude becomes:

$$M_3^{dB}|_{\omega \gg} = -20log\frac{\omega^2}{\omega_n^2} = -40log\frac{\omega}{\omega_n} = -40log\omega - 40log\omega_n \ dB$$

The equation for high-frequency asymptote is a straight line having the slope $-40 \ dB/dec$ since:

$$\frac{dM_3^{dB}}{d\omega^{dec}} = -40 \ dB/dec$$

The high-frequency asymptote intersects the low-frequency asymptote when:

$$M_3^{dB}|_{\omega \gg} = M_3^{dB}|_{\omega \ll}$$

or

$$-40log\omega_c - 40log\omega_n = 0$$

The corner (or break) frequency ω_c is:

$$\omega_c = \omega_n$$

The two asymptotes just derived are independent of the value of ζ. Near the corner frequency ω_c, a resonant peak occurs as may be expected. The damping ration ζ determines the magnitude of this peak.

The maximum error caused by the use of asymptotes occurs at the corner frequency and is calculated by replacing the expression of the corner frequency $\omega = \omega_c = \omega_n$ in the exact relation for log-magnitude:

$$M_3^{dB}|_{\omega=\omega_c} = -20log\sqrt{(1 - \frac{\omega_n^2}{\omega_n^2})^2 + (2\zeta\frac{\omega_n}{\omega_n})^2}$$

$$= -20log(2\zeta) = -20log2 - 20log\zeta \cong -6^{dB} - \zeta^{dB}$$

The magnitude of the error depends on the value of ζ. It is large for small values of ζ. Figure 8 shows the exact log-magnitude curves together with the aymptotes for several values of ζ.

The phase angle of the quadratic factor $1/[(\frac{1}{\omega_n^2}(j\omega)^2 + \frac{2\zeta}{\omega_n}(j\omega) + 1)]$ is:

$$\Phi_3 = \angle \left(\frac{1}{(\frac{1}{\omega_n^2}(j\omega)^2 + \frac{2\zeta}{\omega_n}(j\omega) + 1)} \right) = -arctan\frac{2\zeta\frac{\omega}{\omega_n}}{1 - \left(\frac{\omega}{\omega_n}\right)^2}$$

At $\omega = 0$, the phase angle equals 0.

At the corner frequency $\omega = \omega_c = \omega_n$, the phase angle is -90° regardless of ζ since:

$$\Phi_3 = -arctan\frac{2\zeta}{0} = -arctan\infty = -90^o$$

At $\omega \to \infty$, the phase angle becomes -180°. The phase angle is skew symmetric about the inflexion point, where $\Phi_3 = -90^o$. Figure 8 shows the exact phase angle for different values of ζ.
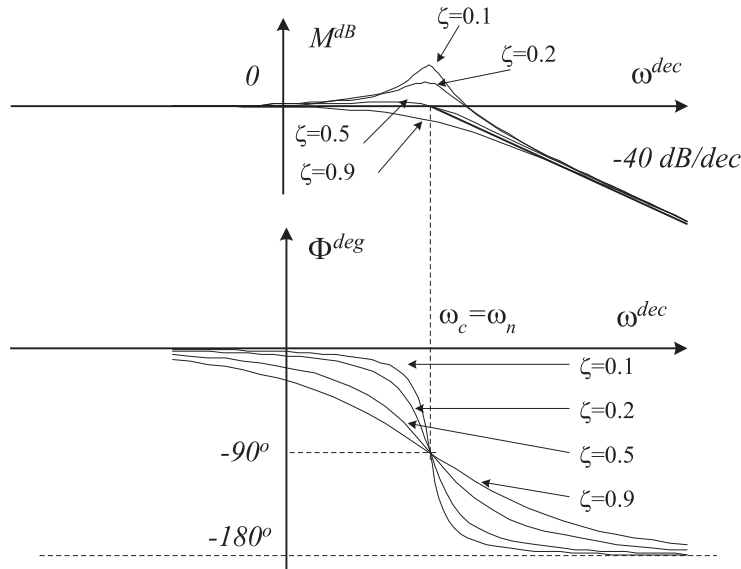
Figure 8: Logarithmic plot for second-order factors

The frequency response curves for the factor

$$\frac{1}{\omega_n^2}(j\omega)^2 + \frac{2\zeta}{\omega_n}(j\omega) + 1$$

can be obtained by reversing the sign of the log-magnitude and the phase angle of the factor $1/[\frac{1}{\omega_n^2}(j\omega)^2 + \frac{2\zeta}{\omega_n}(j\omega) + 1]$. The frequency response curves are presented in Figure 9.

Example. Draw the Bode diagram for the following transfer function:

$$G(s) = \frac{10^3(s + 10)}{s(s + 1)(s^2 + 10s + 100)}$$

10

Figure 9: Logarithmic plot for second-order factors

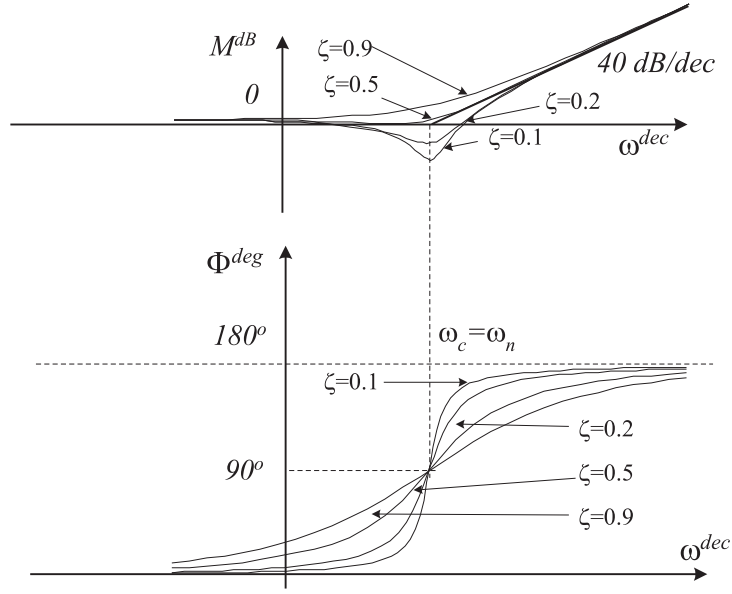First write the transfer function in the general form shown before, emphasizing the time constants, natural frequencies, damping factors and the gain:

$$G(s) = \frac{10^2(\frac{1}{10}s + 1)}{s(s+1)(\frac{1}{100}s^2 + \frac{1}{10}s + 1)}$$

Replacing s with $j\omega$, this function is composed of the following factors:

$$G_1(j\omega) = \frac{10^2}{j\omega} = \frac{k}{(j\omega)^1};$$

$$G_2(j\omega) = \frac{1}{10}(j\omega) + 1 = T_1(j\omega) + 1;$$

$$G_3(j\omega) = \frac{1}{j\omega + 1} = \frac{1}{T_2 j\omega + 1};$$

$$G_4(j\omega) = \frac{1}{\frac{1}{100}(j\omega)^2 + \frac{1}{10}(j\omega) + 1} = \frac{1}{\frac{1}{\omega_n^2}(j\omega)^2 + \frac{2\zeta}{\omega_n}(j\omega) + 1} \tag{1}$$

The Bode diagram is plotted for every one of these factors and than the curves are added graphically. For simplicity reasons, only the asymptotes will be plotted for the log-magnitude curves. The gain k, time constants T_1 and T_2, the damping factor ζ and the natural frequency ω_n, can be identified from equations (1), and we obtain:

$$k = 10^2; \ \ T_1 = \frac{1}{10}; \ \ T_2 = 1; \ \ \omega_n = 10; \ \ \zeta = 0.5.$$

The log-magnitude curve of $G_1(j\omega))$ is a straight line with the slope -20 dB/dec which crosses the vertical axis at $k^{dB} = 20log10^2 = 40dB$. The phase angle corresponding to this factor is a constant line at -90°. The plots are shown in Figure 10.

11

The magnitude curve of the first order factor at the numerator $G_2(j\omega)$ has two asymptotes: one at 0 dB for low frequencies and the other is a line with the slope 20 dB/dec. Both meet at the corner frequency

$$\omega_{c1}^{dec} = -logT_1 = -log\frac{1}{10} = log10 = 1\ dec$$

The phase angle is an arctangent function which approaches the 0 dB line at low frequencies and 90° at very high frequencies with an inflexion point at $(\omega_{c1},\ 45^o)$.

For the first-order factor at the denominator $G_3(j\omega)$ we obtain in a similar way the corner frequency

$$\omega_{c2}^{dec} = -logT_2 = -log1 = 0\ dec$$

and the asymptotes (the one at high frequencies will have a slope of 20 dB/dec).

The phase angle is an arctangent function which approaches the 0 dB line at low frequencies and -90° at very high frequencies with an inflexion point at $(\omega_{c2},\ -45^o)$.

The log-magnitude curve of the second order factor at the denominator $G_4(j\omega)$ has a corner frequency:

$$\omega_{c3} = \omega_n = 10; \quad \omega_{c3}^{dec} = log\ 10 = 1dec;$$

and two asymptotes: at 0 dB for low frequencies and one with a slope of -40 dB/dec at high frequencies.

The phase angle approaches 0 dB at low frequencies and -180° at high frequencies with an inflexion point at $(\omega_{c3},\ -90^o)$.
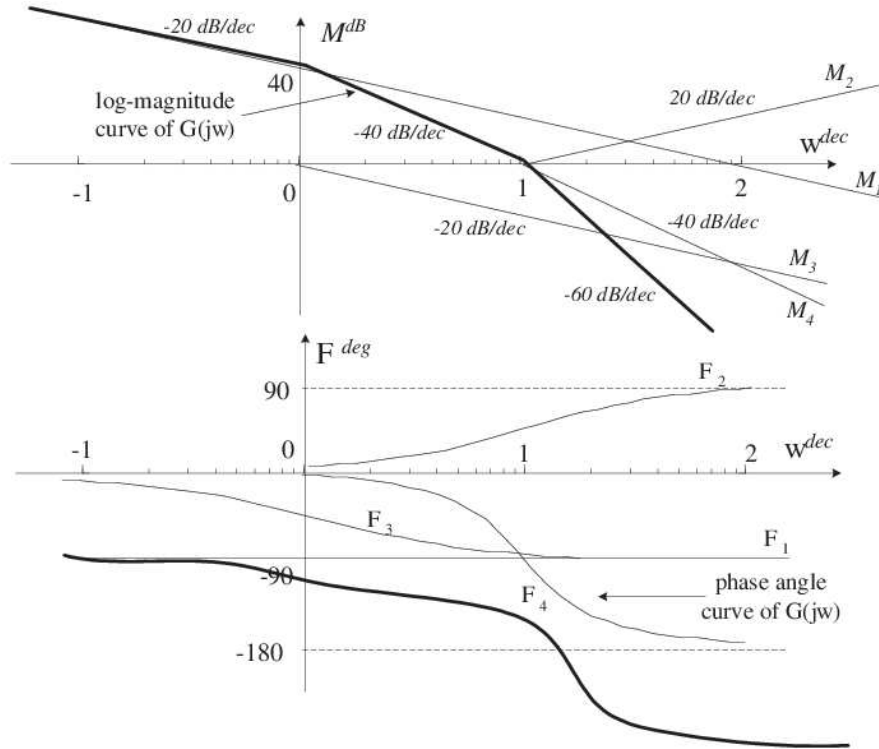

Figure 10: Logarithmic plot for Example 1

Adding the asymptotes. The total asymptotic magnitude can be plotted by adding algebraically the asymptotes due to each factor, as shown by the solid line in Figure 10.

Below $\omega^{dec} = 0$ all the asymptotes approach the 0 dB line except for the plot of $G_1(j\omega)$ where the plot has a slope of -20 dB/dec. At the first corner frequency $\omega_{c2}^{dec} = 0$, the slope changes to -40 dB/dec because of asymptote at high frequency of $G_3(j\omega)$ (which has a slope of -20 dB/dec) was added. The final plot changes again the slope at $\omega^{dec} = 1$, where another two asymptotes at high frequencies add their slope to the resultant (-40dB/dec +20 dB/dec). Thus, for frequencies higher than $\omega^{dec} = 1$ the final plot will have a slope of -60 dB/dec.

For plotting the complete phase-angle curve, the phase angle curves for all factors have to be sketched. The algebraic sum of these curves provides the complete phase-angle curve as shown in Figure 10.

Reading Bode Diagrams. If a sinusoidal signal $X(j\omega)$ with frequency ω is applied as the input to an open-loop system with the sinusoidal transfer function $G(j\omega)$ the output signal, $Y(j\omega)$ will have the same frequency. The amplitude of the output can be obtained by reading from a Bode diagram the magnitude of the transfer function corresponding to frequency ω and we obtain:
$$|Y(j\omega)| = |G(j\omega)| \cdot |X(j\omega)|$$
If $|G(j\omega)| > 1$ (or $M^{dB} = |G(j\omega)|^{dB} > 0$), the output will be amplified. If $|G(j\omega)| < 1$ (or $M^{dB} = |G(j\omega)|^{dB} < 0$), the output is attenuated.

The phase shift of the output signal with respect to in input is the phase angle of the transfer function represented in a Bode diagram. When $\Phi > 0$ the output is shifted with a positive angle comparing to the input and the system has phase lead. If $\Phi < 0$ the output is shifted with a negative angle with respect to the input (phase lag).

The log-magnitude curve shown in Figure 10 is positive for frequencies lower than $\omega^{dec} = 1$ or $\omega = 10 \ rad/sec$ and negative for the high frequencies. This means that for all low frequencies the system will amplify the input signal and will attenuate the signals with high frequencies. The system in this case is a low-pass filter.

A system which amplifies the high frequencies and attenuate the low ones is a high-pass filter.

Frequency Response Specifications

Pages from K. Ogata (Prentice Hall, 2002)

The Resonant Frequency ω_r and the Resonant Peak Value M_r. The magnitude of

$$G(j\omega) = \frac{1}{1 + 2\zeta\left(j\dfrac{\omega}{\omega_n}\right) + \left(j\dfrac{\omega}{\omega_n}\right)^2}$$

is

$$|G(j\omega)| = \frac{1}{\sqrt{\left(1 - \dfrac{\omega^2}{\omega_n^2}\right)^2 + \left(2\zeta\dfrac{\omega}{\omega_n}\right)^2}} \tag{8-9}$$

If $|G(j\omega)|$ has a peak value at some frequency, this frequency is called the *resonant* frequency. Since the numerator of $|G(j\omega)|$ is constant, a peak value of $|G(j\omega)|$ will occur when

$$g(\omega) = \left(1 - \frac{\omega^2}{\omega_n^2}\right)^2 + \left(2\zeta\frac{\omega}{\omega_n}\right)^2 \tag{8-10}$$

is a minimum. Since Equation (8-10) can be written

$$g(\omega) = \left[\frac{\omega^2 - \omega_n^2(1 - 2\zeta^2)}{\omega_n^2}\right]^2 + 4\zeta^2(1 - \zeta^2) \tag{8-11}$$

the minimum value of $g(\omega)$ occurs at $\omega = \omega_n\sqrt{1 - 2\zeta^2}$. Thus the resonant frequency ω_r is

$$\omega_r = \omega_n\sqrt{1 - 2\zeta^2}, \qquad \text{for } 0 \le \zeta \le 0.707 \tag{8-12}$$

As the damping ratio ζ approaches zero, the resonant frequency approaches ω_n. For $0 < \zeta \le 0.707$, the resonant frequency ω_r is less than the damped natural frequency $\omega_d = \omega_n\sqrt{1 - \zeta^2}$, which is exhibited in the transient response. From Equation (8-12), it can be seen that for $\zeta > 0.707$, there is no resonant peak. The magnitude $|G(j\omega)|$ decreases monotonically with increasing frequency ω. (The magnitude is less than 0 dB for all values of $\omega > 0$. Recall that, for $0.7 < \zeta < 1$, the step response is oscillatory, but the oscillations are well damped and are hardly perceptible.)

Figure 8–10
M_r-versus-ζ curve for the second-order system $1/[1 + 2\zeta(j\omega/\omega_n) + (j\omega/\omega_n)^2]$.
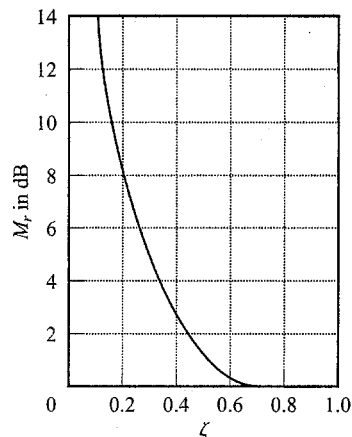
The magnitude of the resonant peak, M_r, can be found by substituting Equation (8–12) into Equation (8–9). For $0 \leq \zeta \leq 0.707$,

$$M_r = |G(j\omega)|_{\max} = |G(j\omega_r)| = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} \qquad (8\text{–}13)$$

For $\zeta > 0.707$,

$$M_r = 1 \qquad (8\text{–}14)$$

As ζ approaches zero, M_r approaches infinity. This means that if the undamped system is excited at its natural frequency, the magnitude of $G(j\omega)$ becomes infinity. The relationship between M_r and ζ is shown in Figure 8–10.

The phase angle of $G(j\omega)$ at the frequency where the resonant peak occurs can be obtained by substituting Equation (8–12) into Equation (8–8). Thus, at the resonant frequency ω_r,

$$\underline{/G(j\omega_r)} = -\tan^{-1}\frac{\sqrt{1 - 2\zeta^2}}{\zeta} = -90° + \sin^{-1}\frac{\zeta}{\sqrt{1 - \zeta^2}}$$

Resonant Peak Magnitude M_r and Resonant Frequency ω_r. Consider the standard second-order system shown in Figure 8–76. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{8-16}$$

where ζ and ω_n are the damping ratio and the undamped natural frequency, respectively. The closed-loop frequency response is

$$\frac{C(j\omega)}{R(j\omega)} = \frac{1}{\left(1 - \dfrac{\omega^2}{\omega_n^2}\right) + j2\zeta\dfrac{\omega}{\omega_n}} = Me^{j\alpha}$$

where

$$M = \frac{1}{\sqrt{\left(1 - \dfrac{\omega^2}{\omega_n^2}\right)^2 + \left(2\zeta\dfrac{\omega}{\omega_n}\right)^2}}, \qquad \alpha = -\tan^{-1}\frac{2\zeta\dfrac{\omega}{\omega_n}}{1 - \dfrac{\omega^2}{\omega_n^2}}$$

As given by Equation (8–12), for $0 \le \zeta \le 0.707$, the maximum value of M occurs at the frequency ω_r, where

$$\omega_r = \omega_n\sqrt{1 - 2\zeta^2} \tag{8-17}$$

The frequency ω_r is the resonant frequency. At the resonant frequency, the value of M is maximum and is given by Equation (8–13), rewritten

$$M_r = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} \tag{8-18}$$

where M_r is defined as the *resonant peak magnitude*. The resonant peak magnitude is related to the damping of the system.

The magnitude of the resonant peak gives an indication of the relative stability of the system. A large resonant peak magnitude indicates the presence of a pair of dominant closed-loop poles with small damping ratio, which will yield an undesirable transient response. A smaller resonant peak magnitude, on the other hand, indicates the absence of a pair of dominant closed-loop poles with small damping ratio, meaning that the system is well damped.

Remember that ω_r is real only if $\zeta < 0.707$. Thus, there is no closed-loop resonance if $\zeta > 0.707$. [The value of M_r is unity only if $\zeta > 0.707$. See Equation (8–14).] Since the values of M_r and ω_r can be easily measured in a physical system, they are quite useful for checking agreement between theoretical and experimental analyses.
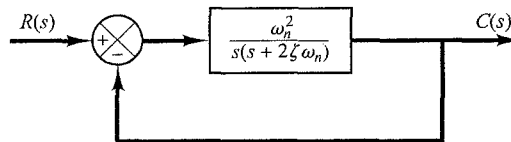

Figure 8–76
Standard second-order system.

It is noted, however, that in practical design problems the phase margin and gain margin are more frequently specified than the resonant peak magnitude to indicate the degree of damping in a system. .

Correlation between Step Transient Response and Frequency Response in the Standard Second-Order System. The maximum overshoot in the unit-step response of the standard second-order system, as shown in Figure 8–76, can be exactly correlated with the resonant peak magnitude in the frequency response. Hence, essentially the same information about the system dynamics is contained in the frequency response as is in the transient response.

For a unit-step input, the output of the system shown in Figure 8–76 is given by Equation (5-12), or

$$c(t) = 1 - e^{-\zeta \omega_n t}\left(\cos \omega_d t + \frac{\zeta}{\sqrt{1 - \zeta^2}} \sin \omega_d t \right), \qquad \text{for } t \geq 0$$

where

$$\omega_d = \omega_n \sqrt{1 - \zeta^2} \tag{8-19}$$

On the other hand, the maximum overshoot M_p for the unit-step response is given by Equation (5–21), or

$$M_p = e^{-(\zeta/\sqrt{1-\zeta^2})\pi} \tag{8-20}$$

This maximum overshoot occurs in the transient response that has the damped natural frequency $\omega_d = \omega_n \sqrt{1 - \zeta^2}$. The maximum overshoot becomes excessive for values of $\zeta < 0.4$.

Since the second-order system shown in Figure 8–76 has the open-loop transfer function

$$G(s) = \frac{\omega_n^2}{s(s + 2\zeta\omega_n)}$$

for sinusoidal operation, the magnitude of $G(j\omega)$ becomes unity when

$$\omega = \omega_n \sqrt{\sqrt{1 + 4\zeta^4} - 2\zeta^2}$$

which can be obtained by equating $|G(j\omega)|$ to unity and solving for ω. At this frequency, the phase angle of $G(j\omega)$ is

$$\underline{/G(j\omega)} = -\underline{/j\omega} - \underline{/j\omega + 2\zeta\omega_n} = -90° - \tan^{-1} \frac{\sqrt{\sqrt{1 + 4\zeta^4} - 2\zeta^2}}{2\zeta}$$

Thus, the phase margin γ is

$$\gamma = 180° + \underline{/G(j\omega)}$$

$$= 90° - \tan^{-1} \frac{\sqrt{\sqrt{1 + 4\zeta^4} - 2\zeta^2}}{2\zeta}$$

$$= \tan^{-1} \frac{2\zeta}{\sqrt{\sqrt{1 + 4\zeta^4} - 2\zeta^2}} \tag{8-21}$$

Equation (8–21) gives the relationship between the damping ratio ζ and the phase margin γ. (Notice that the phase margin γ is a function *only* of the damping ratio ζ.)

Figure 8–77
Curve γ (phase margin) versus ζ for the system shown in Figure 8–76.
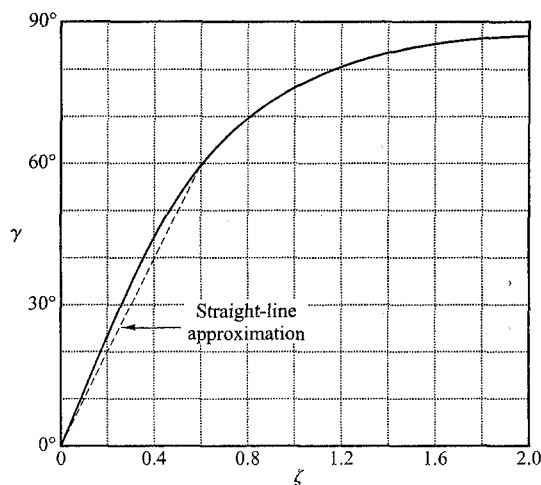
In the following, we shall summarize the correlation between the step transient response and frequency response of the standard second-order system given by Equation (8–16):

1. The phase margin and the damping ratio are directly related. Figure 8–77 shows a plot of the phase margin γ as a function of the damping ratio ζ. It is noted that for the standard second-order system shown in Figure 8–76, the phase margin γ and the damping ratio ζ are related approximately by a straight line for $0 \le \zeta \le 0.6$, as follows:
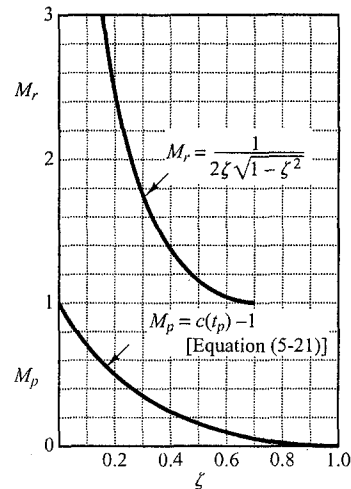
$$\zeta = \frac{\gamma}{100}$$

Thus a phase margin of $60°$ corresponds to a damping ratio of 0.6. For higher-order systems having a dominant pair of closed-loop poles, this relationship may be used as a rule of thumb in estimating the relative stability in the transient response (that is, the damping ratio) from the frequency response.

2. Referring to Equations (8–17) and (8–19), we see that the values of ω_r and ω_d are almost the same for small values of ζ. Thus, for small values of ζ, the value of ω_r is indicative of the speed of the transient response of the system.

3. From Equations (8–18) and (8–20), we note that the smaller the value of ζ is, the larger the values of M_r and M_p are. The correlation between M_r and M_p as a function of ζ is shown in Figure 8–78. A close relationship between M_r and M_p can be seen for $\zeta > 0.4$. For very small values of ζ, M_r becomes very large $\left(M_r \gg 1\right)$, while the value of M_p does not exceed 1.

Correlation between Step Transient Response and Frequency Response in General Systems. The design of control systems is very often carried out on the basis of the frequency response. The main reason for this is the relative simplicity of this approach compared with others. Since in many applications it is the transient response of the system to aperiodic inputs rather than the steady-state response to sinusoidal inputs that is of primary concern, the question of correlation between transient response and frequency response arises.

For the standard second-order system shown in Figure 8–76, mathematical relationships correlating the step transient response and frequency response can be obtained

Figure 8–78
Curves M_r versus ζ and M_p versus ζ for the system shown in Figure 8–76.

In the figure:

$$M_r = \frac{1}{2\zeta\sqrt{1-\zeta^2}}$$

$$M_p = c(t_p) - 1 \quad \text{[Equation (5-21)]}$$

easily. The time response of the standard second-order system can be predicted exactly from a knowledge of the M_r and ω_r of its closed-loop frequency response.

For nonstandard second-order systems and higher-order systems, the correlation is more complex, and the transient response may not be predicted easily from the frequency response because additional zeros and/or poles may change the correlation between the step transient response and the frequency response existing for the standard second-order system. Mathematical techniques for obtaining the exact correlation are available, but they are very laborious and of little practical value.

The applicability of the transient-response–frequency-response correlation existing for the standard second-order system shown in Figure 8–76 to higher-order systems depends on the presence of a dominant pair of complex-conjugate closed-loop poles in the latter systems. Clearly, if the frequency response of a higher-order system is dominated by a pair of complex-conjugate closed-loop poles, the transient-response–frequency-response correlation existing for the standard second-order system can be extended to the higher-order system.

For linear, time-invariant, higher-order systems having a dominant pair of complex-conjugate closed-loop poles, the following relationships generally exist between the step transient response and frequency response:

1. The value of M_r is indicative of the relative stability. Satisfactory transient performance is usually obtained if the value of M_r is in the range $1.0 < M_r < 1.4$ $\left(0 \text{ dB} < M_r < 3 \text{ dB}\right)$, which corresponds to an effective damping ratio of $0.4 < \zeta < 0.7$. For values of M_r greater than 1.5, the step transient response may exhibit several overshoots. (Note that, in general, a large value of M_r corresponds to a large overshoot in the step transient response. If the system is subjected to noise signals whose frequencies are near the resonant frequency ω_r, the noise will be amplified in the output and will present serious problems.)

2. The magnitude of the resonant frequency ω_r is indicative of the speed of the transient response. The larger the value of ω_r, the faster the time response is. In other words, the rise time varies inversely with ω_r. In terms of the open-loop frequency response, the damped natural frequency in the transient response is somewhere between the gain crossover frequency and phase crossover frequency.

3. The resonant peak frequency ω_r and the damped natural frequency ω_d for the step transient response are very close to each other for lightly damped systems.

The three relationships just listed are useful for correlating the step transient response with the frequency response of higher-order systems, provided that they can be approximated by the standard second-order system or a pair of complex-conjugate closed-loop poles. If a higher-order system satisfies this condition, a set of time-domain specifications may be translated into frequency-domain specifications. This simplifies greatly the design work or compensation work of higher-order systems.

In addition to the phase margin, gain margin, resonant peak M_r, and resonant frequency ω_r, there are other frequency-domain quantities commonly used in performance specifications. They are the cutoff frequency, bandwidth, and the cutoff rate. These will be defined in what follows.

Cutoff Frequency and Bandwidth. Referring to Figure 8–79, the frequency ω_b at which the magnitude of the closed-loop frequency response is 3 dB below its zero-frequency value is called the *cutoff frequency*. Thus

$$\left|\frac{C(j\omega)}{R(j\omega)}\right| < \left|\frac{C(j0)}{R(j0)}\right| - 3 \text{ dB}, \qquad \text{for } \omega > \omega_b$$

For systems in which $|C(j0)/R(j0)| = 0$ dB,

$$\left|\frac{C(j\omega)}{R(j\omega)}\right| < -3 \text{ dB}, \qquad \text{for } \omega > \omega_b$$

The closed-loop system filters out the signal components whose frequencies are greater than the cutoff frequency and transmits those signal components with frequencies lower than the cutoff frequency.

The frequency range $0 \leq \omega \leq \omega_b$ in which the magnitude of the closed loop does not drop -3 dB is called the *bandwidth* of the system. The bandwidth indicates the frequency where the gain starts to fall off from its low-frequency value. Thus, the bandwidth indicates how well the system will track an input sinusoid. Note that for a given ω_n, the rise time increases with increasing damping ratio ζ. On the other hand, the bandwidth decreases with the increase in ζ. Therefore, the rise time and the bandwidth are inversely proportional to each other.

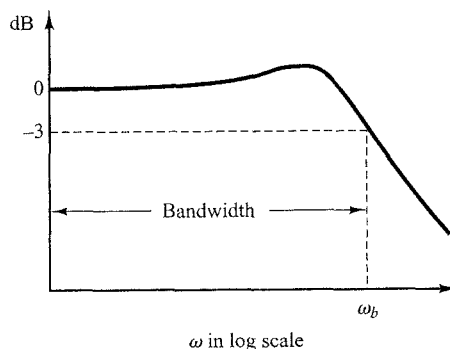The specification of the bandwidth may be determined by the following factors:

Figure 8–79
Plot of a closed-loop frequency response curve showing cutoff frequency ω_b and bandwidth.

1. The ability to reproduce the input signal. A large bandwidth corresponds to a small rise time, or fast response. Roughly speaking, we can say that the bandwidth is proportional to the speed of response. (For example, to decrease the rise time in the step response by a factor of 2, the bandwidth must be increased by approximately a factor of 2.)

2. The necessary filtering characteristics for high-frequency noise.

For the system to follow arbitrary inputs accurately, it must have a large bandwidth. From the viewpoint of noise, however, the bandwidth should not be too large. Thus, there are conflicting requirements on the bandwidth, and a compromise is usually necessary for good design. Note that a system with large bandwidth requires high-performance components, so the cost of components usually increases with the bandwidth.

Cutoff Rate. The cutoff rate is the slope of the log-magnitude curve near the cutoff frequency. The cutoff rate indicates the ability of a system to distinguish the signal from noise.

It is noted that a closed-loop frequency response curve with a steep cutoff characteristic may have a large resonant peak magnitude, which implies that the system has a relatively small stability margin.

EXAMPLE 8–23 Consider the following two systems:

$$\text{System I:} \quad \frac{C(s)}{R(s)} = \frac{1}{s+1}, \qquad \text{System II:} \quad \frac{C(s)}{R(s)} = \frac{1}{3s+1}$$

Compare the bandwidths of these two systems. Show that the system with the larger bandwidth has a faster speed of response and can follow the input much better than the one with the smaller bandwidth.

Figure 8–80(a) shows the closed-loop frequency-response curves for the two systems. (Asymptotic curves are shown by dashed lines.) We find that the bandwidth of system I is $0 \le \omega \le 1$ rad/sec and that of system II is $0 \le \omega \le 0.33$ rad/sec. Figures 8–80(b) and (c) show, respectively, the unit-step response and unit-ramp response curves for the two systems. Clearly, system I, whose bandwidth is three times wider than that of system II, has a faster speed of response and can follow the input much better.
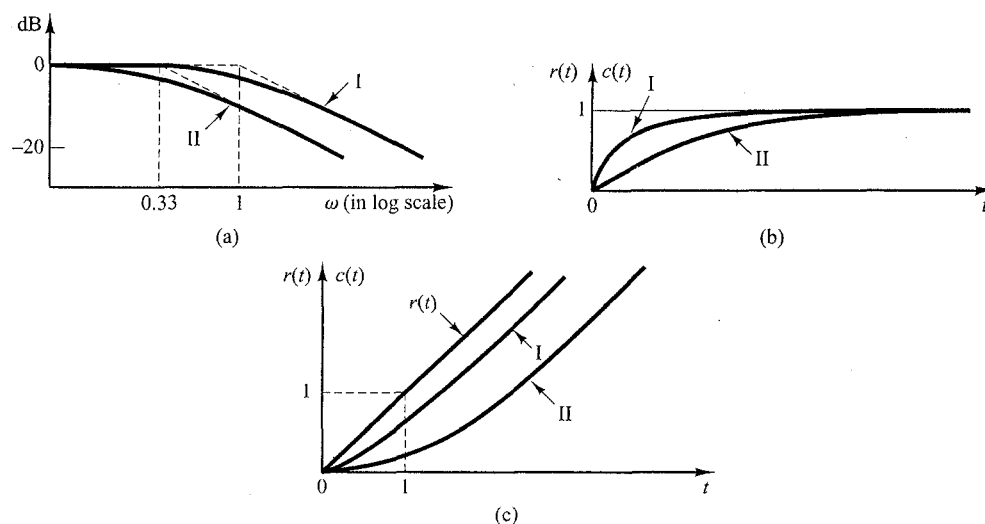
Figure 8–80
Comparison of dynamic characteristics of the two systems considered in Example 8–23. (a) Closed-loop frequency-response curves; (b) unit-step response curves; (c) unit-ramp response curves.

5. STABILITY OF LINEAR SYSTEMS

For every reader, the term (or, in a rather general expression) the concept of system stability (including the stability of a technical system) looks rather intuitive or even expressive. However, for a reader who is not informed in the field, this term is often less specific. A better known and relatively accurate formulation of the concept of stability is that related to the equilibrium state of a thermodynamic system. In this case, the equilibrium state of the system is considered to be stable if the system – subjected to external or internal conditions, with permanent (stationary) or transitory action – evolves to a new equilibrium / steady-state state or – from case to case – it evolves back to the initial equilibrium state.

The technical applications of systems make use of the ***concept of stability*** in order to highlight the capacity of systems to keep – in certain conditions – a steady-state equilibrium state or to carry out the transition from one equilibrium state to another one. Moreover, the concept of stability can also be referred to an operating regime (which can be, for example, a steady-state operating point) that proves "to be or not to be" stable.

The quality of a system to be ***stable*** is influenced as an effect of modifying:
- one / several system inputs,
- the system parameters,
- the system structure.

All these causes have repercussions on the situation and regime in which the system was prior to the modification(s) and also effects on the evolution of system output (state).

The idea of stability is generally familiar to readers. Knowing that an unstable system will exhibit an erratic and destructive response, it is searched to ensure that a system is stable and exhibits a bounded transient response.

The concept of stability can be **illustrated** in terms of considering a right circular cone placed on a plane horizontal surface. If the cone is resting on its base and is tipped slightly, it returns to its original equilibrium position; this position and response is said to be stable, and it is illustrated in Fig. 5.1 (a). However, if the cone rests on its side and is displaced slightly, it rolls with no tendency to leave the position on its side; this position is designated as the neutral stability or marginal stability, and it is illustrated in Fig. 5.1 (b). On the other hand, if the cone is placed on its tip and released, it falls onto its side; this position is said to be unstable, and it is illustrated in Fig. 5.1 (c).
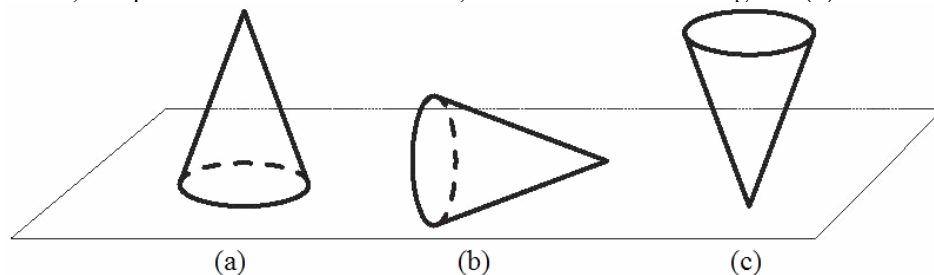

Fig. 5.1. *Illustration of cone stability: stable (a), marginally stable (b), unstable (c).*

This chapter will treat only **engineering aspects** concerning:
- the definition of the concept of stability,
- the main aspects of the theory of linear system stability,
- criteria to check the stability of linear systems,

that can be treated in a deep manner using, for example, the classical books [1]–[5].

5.1. The concept of stability

It is obvious that technical systems are usable only if they are *stable*. **Whether a linear system is stable or unstable is a property of the system itself and does not depend on the input**.

The stability of technical systems can be expressed using both intuitive formulations (definitions) – of the type expressed in the previous section, which are usually subjected to experimental testing (but often associated with rather high risks) – and mainly formulations concerning the stability of the dynamical system (DS, namely mathematical model, MM) associated to the physical system (PS), which can be tested analytically. The intuitive formulations (in the form of definitions) can be given in relation to the stability of both PS and DS. Two such formulations will be given as follows with focus on the SISO system characterized by the (linear) SS-MM

$$\mathbf{x}'(t) = \mathbf{A}\,\mathbf{x}(t) + \mathbf{b}\,u(t),$$
$$y(t) = \mathbf{c}^T\,\mathbf{x}(t),$$

(5.1.1-a)

with

$$\mathbf{x}'(t) = \begin{cases} \dot{\mathbf{x}}(t),\ t \in T \subset \mathbf{R}\ \text{for continuous - time systems,} \\ \mathbf{x}_{k+1},\ k \in \mathbf{Z}\,(\mathbf{N})\ \text{for discrete - time systems,} \end{cases}$$

or the equivalent IO-MM given by the transfer function (t.f.)

$$H(\lambda) = \mathbf{c}^T (\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{b},$$

(5.1.1-b)

with

$$\lambda = \begin{cases} s,\ \text{for continuous - time systems,} \\ z,\ \text{for discrete - time systems.} \end{cases}$$

A) The external stability of a system or the input-output stability. IO-MMs characterize the input-output relationships of systems without explicitly describing its inner structure. In the context of this characterization, the most frequently used definition of external stability is the so-called **BIBO** (**B**ounded **I**nput – **B**ounded **O**utput) stability. This property is considered under the action of an external cause that acts on the system.

Definition 1: A dynamical system (DS) of given in (5.1.1) – and the PS characterized by this DS – is referred to as **BIBO stable** if for any initial time moment $t_0 \in T \subset \Re$ and initial state (vector) $\mathbf{x}_0 = \mathbf{0}$, applying a bounded variation of system input $u(t)$:

$$|u(t)| < L_u, \tag{5.1.2}$$

makes the system respond with a bounded variation of system output $y(t)$:

$$|y(t)| < L_y. \tag{5.1.3}$$

Contrarily, the system is called unstable.

Remark 1: The values of the margins L_u and L_y of the bounds and their correlation are not stated. They are, as far as this presentation is concerned, at the disposal of the person that carried out the stability analysis.

The BIBO stability can be **intuitively tested** by the shape of system response. Fig. 5.2 shows typical step system responses of BIBO stable systems ((1), (2) and (3)) and BIBO unstable systems ((4) and (5)) as well.
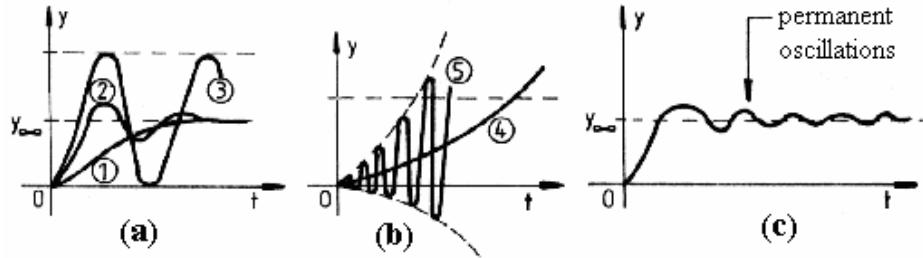

Fig. 5.2. *Graphical interpretation of BIBO stability in terms of step responses.*

Remark 2: In the sense of Definition 1, the systems with responses of type (3), i.e. undamped (or critically damped or sustained) oscillations with constant amplitude (illustrated in Fig. 5.2 (a)), can be considered to belong to the class of stable systems. However, from a practical point of view they are considered as unstable systems in automation and control systems; they are widely used in oscillators designed by electronic engineers.

A particular situation is that illustrated in Fig. 5.2 (c), where a permanent (or stationary) oscillatory regime is installed in the system. Depending on the interpretation of Definition 1, the system can be considered as stable (an example in this regard is the case of control systems with two-position nonlinear controllers that are widely used in home applications as washing machines and clothing irons) or unstable (for example, such unwanted oscillation that occur in control systems with linear controllers with control signal (control input) limitation, $u_m \leq |u(t)| \leq u_M$, will prove its instability).

R.-E. Precup, UPT, 2024

Remark 3: In case of control systems, the technical limitations deliberately imposed at the level of various variables avoid the increase of system output $y(t)$ over some safety values (the dotted line in Fig. 5.2 (b)). Contrarily, the technical process (plant) could be damaged. The control loop does not have anymore the task to get next the system out of the limitation state.

Although simple, due to safety reasons, the experimental testing of system stability in terms of Definition 1 is risky.

B) *The internal stability of a system* or *the stability of system state*. In order to characterize the internal stability of a PS / DS, the formulations are related to a certain system state (vector) $\mathbf{x}_0$, which can be stable or unstable. The concept of internal stability is referred to as ***stability in the sense of Lyapunov*** or ***Lyapunov stability***.

Definition 2 (in an engineering formulation): The equilibrium state (or point) $\mathbf{x}_0 = \mathbf{0} \in \mathfrak{R}^n$ of the DS (5.1.1) and accordingly the state $\mathbf{x}_0 = \mathbf{0}$ of the PS characterized by the DS (5.1.1) is a **stable state** if by getting the system out of this state:

$$| \Delta x_i(0_+) |< L_{x0}, \ i = 1...n, \tag{5.1.4}$$

with $L_{x0} > 0$, the system will evolve back as follows after removing the cause:

♦ in the initial stable state $\mathbf{x}_0$ or
♦ in an acceptable vicinity of this state,
and the resulted state trajectories, $\Delta\mathbf{x}(t)$, $t > 0$, fulfill the condition

$$| \Delta x_i(t) |< L_x, \ i = 1...n, \tag{5.1.5}$$

where $L_x > 0$, $L_x = f(L_{x0})$ such that $L_x > L_{x0}$ [6], [7]. If the requirements imposed in Definition 2 are not fulfilled, the system is called unstable.

The interpretation of stability according to Definition 2 is illustrated in Fig. 5.3 for second-order systems (n=2) making use of possible **phase trajectories** of these systems as follows: the trajectories (1), (2), (3) and (4) highlight stable systems or behaviors, and the trajectories (5) and (6) highlight unstable systems or behaviors.
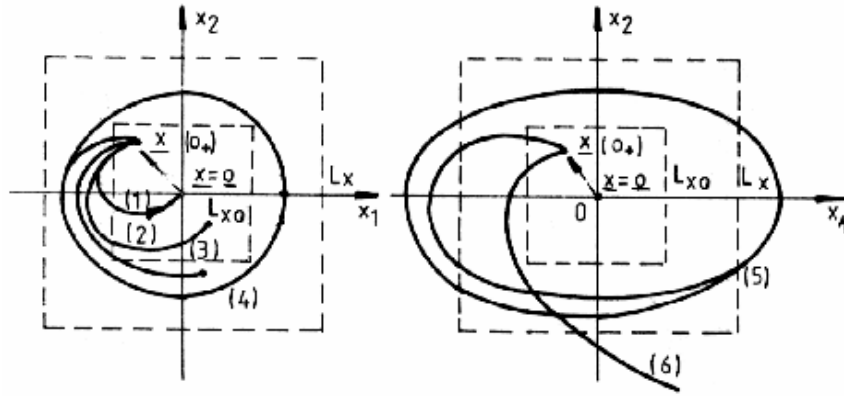
Fig. 5.3. *Graphical interpretation of internal stability using phase trajectories.*

The trajectories (4) and (5) are generally ellipses and indicate that the system is in a stationary oscillatory regime. The difference between (4) and (5) is that (4) characterizes by the fixed $L_x > 0$ a system that is accepted to be stable, and (5) characterizes by the fixed $L_x > 0$ an unstable system.

The trajectory (1) indicates a remarkable aspect: **after removing the cause**, the system **evolves back exactly** in the initial state $\mathbf{x}_0$. Systems with such behavior are referred to as ***asymptotically stabile systems***.

C) Final remarks. The testing of stability of a DS (PS) on the basis of the definitions is not advantageous as:

- if an experimental testing is conducted, that can lead to effects that are often unwished;
- it the testing is conducted by the digital simulation of DS behavior, the time dedicated to checking can be significantly larger than that for the theoretical analysis of system stability.

These are the reasons why the dynamical system theory operates with ***theorems that set the necessary and often sufficient*** for the stability of a DS (and, accordingly, PS) to be stable in the conditions of the above definitions. The qualitative testing of stability conditions of a DS – which can be accompanied in certain cases by quantitative appreciations on the stability degree – can be performed in an operative manner by means of ***stability criteria***, expressed as theorems. However, these stability theorems / criteria impose requirements specific to the type of system, i.e. continuous-time or discrete-time, linear (linearized) or nonlinear, with fix (fixed) structure and constant (fixed) parameter values, etc.

The presentations given in the sequel will concern only the stability of continuous- and discrete-time linear systems, with fixed structure and possible variable parameter values.

The stability is an **(internal) property of systems**, which plays its role if the input and / or disturbances (of any type) act on the DS. The ***structural property*** label should be viewed in two senses, namely:
* modifications of system structure can affect this property;
* considering a fix (fixed) system structure, the modification of parameters that characterize the system can affect the property.

From this point of view, ***the systems with fixed structure are*** as follows:
* **absolutely stable** systems, where the property does not depend on the values of parameters that characterize the structure;
* **absolutely stable** systems, for which given the fixed structure, they are unstable for any parameter values
* **conditionally stable systems**, for which given the fixed structure, they are stable for a set of parameter values, but can become (are) unstable for another set of parameter values.

5.2. Stability of continuous-time linear time invariant systems (C-LTIS). The fundamental stability theorem of C-LTIS

A) The fundamental stability theorem of C-LTIS. Let us consider the SISO C-LTIS characterized by
SS-MM:

$$\dot{\mathbf{x}} = \mathbf{A}\,\mathbf{x} + \mathbf{b}\,u,$$
$$y = \mathbf{c}^T\,\mathbf{x}, \tag{5.2.1-a}$$

or IO-MM:

$$\sum_{\nu=0}^{n} a_\nu\, y^{(\nu)}(t) = \sum_{\mu=0}^{m} b_\mu\, u^{(\mu)}(t),\ m < n. \tag{5.2.1-b}$$

Generally speaking, the two models reflect the same reality, pointed out, for example, by the same t.f. $H(s)$:

$$H(s) = \begin{cases} \mathbf{c}^T (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \mathbf{c}^T \dfrac{\mathrm{adj}(s\mathbf{I} - \mathbf{A})}{\det(s\mathbf{I} - \mathbf{A})}\mathbf{b}, & (5.2.2\text{-a}) \\[2mm] \dfrac{b_m s^m + ... + b_1 s + b_0}{a_n s^n + ... + a_1 s + a_0}. & (5.2.2\text{-b}) \end{cases}$$

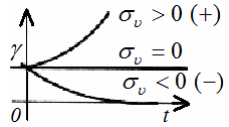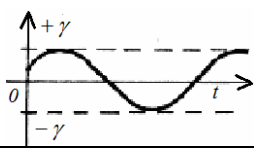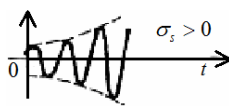The characteristic equation of the system, $\Delta(s) = 0$, is unique, and can be expressed in one of the two forms
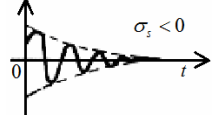
$$\Delta(s) = \begin{cases} \det(s\mathbf{I} - \mathbf{A}) = 0, & (5.2.3\text{-a}) \\ \displaystyle\sum_{v=0}^{n} a_v s^v = 0. & (5.2.3\text{-b}) \end{cases}$$

The theory of ordinary differential equations states that the homogenous solution to the differential equation given in (5.2.3) is given by the roots of the characteristic equation

$$\Delta(s) = a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0 = 0. \tag{5.2.4}$$

Depending on the roots of this equation, the homogenous solution (or the system unforced response or the system free response) will contain components that lead to different shapes of the solutions as function of time. Such shapes are exemplified in Table 5.1.

Table 5.1. Shapes of free (unforced) responses for C-LTIS (5.2.1).

Types of roots and additional conditions		Components of solution(s) corresponding to the root(s)	Graphic(s) of solution(s) corresponding to the root(s)
Real $s_v = \sigma_v$	$\sigma_v > 0$	$\gamma \cdot e^{\sigma_v t}$ $(\gamma > 0)$	
	$\sigma_v = 0$		
	$\sigma_v < 0$		
Pure imaginary (pair) $s_{v,v+1} = \pm j\omega$	$\sigma_v = 0$ $\omega_v > 0$	$\gamma \sin(\omega_v t + \varphi_v)$ $(\gamma > 0)$	
Complex conjugated $s_{v,v+1} = \sigma_v \pm j\omega_v$	$\sigma_v > 0$ $\omega_v > 0$	$\gamma \cdot e^{\sigma_v t} \sin(\omega_v t + \varphi_v)$ $(\gamma > 0)$	
	$\sigma_v < 0$ $\omega_v > 0$		

The analysis of results given in Table 5.1 points out that the free solutions, which are vanishing in time for $t \to \infty$, are characterized by roots with **strictly negative real parts**, i.e.

$$\text{Re}(s_v) < 0. \tag{5.2.5}$$

This means that, in order to have a system response – unforced (free) or next forced – damped in time, it should be characterized by **roots of the characteristic equation that are strictly placed in the left half-plane of the roots plane**.

These aspects are synthesized in *the fundamental stability theorem of continuous-time linear time invariant systems (C-LTIS)*:

Theorem 1: A C-LTIS characterized by the SS-MM / IO-MM expressed in (5.2.1) – and, accordingly, the PS described by this DS – is stable **if and only if** the *n* roots s_v, $v = 1...n$, of the characteristic equation (5.2.4) (i.e., the system poles) have **strictly negative real parts**, i.e.

$$\text{Re}(s_v) < 0, \ v = 1...n. \tag{5.2.6}$$

In other words, **the roots are placed inside the left half-plane of the "s" plane** illustrated in Fig. 5.4.

A simple graphical interpretation of the fundamental stability theorem is given in Fig. 5.4 in terms of the representation of system poles in the "s" roots plane and their relation to system impulse responses.
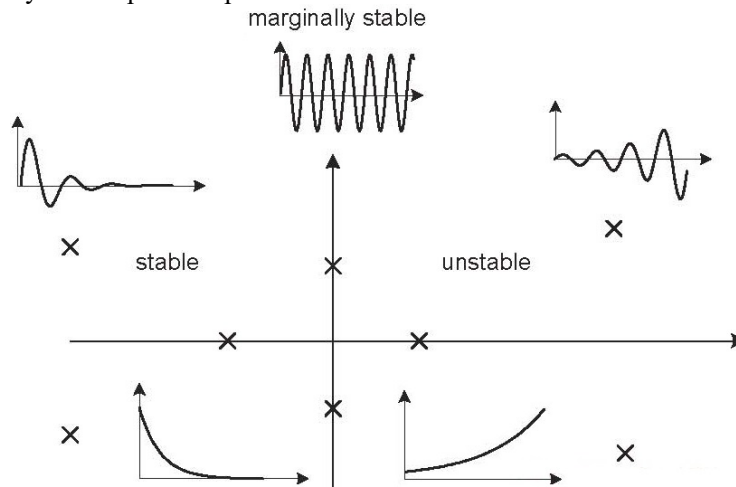

Fig. 5.4. *Graphical interpretation of fundamental stability theorem in terms of poles positions and their relation to system impulse responses.*

A deeper graphical interpretation of the fundamental stability theorem is given in Fig. 5.5 (a) in terms of the representation of system poles in the "s" roots plane. From practical reasons due to:

- the possibility of "migration" of poles because of possible modifications of system parameters, and this migration could be even drastically in certain situations, leading to system instability, Fig. 5.5 (b),
- the oscillatory transient regimes, which should be damped as fast as possible for a PS,

a **stability margin (or reserve) zone** is needed and highlighted in Fig. 5.5 (b). This will prevent shortcomings that can occur because of the above mentioned aspects. The quantitative definition of this zone should account for:

- the magnitude of the time constants that characterize the system (the repartition of system poles),
- the possible tendencies regarding the modification of system parameters resulting in the modification of poles positions.

The second aspect can be studied analytically on the basis of *system sensitivity theory* [8].
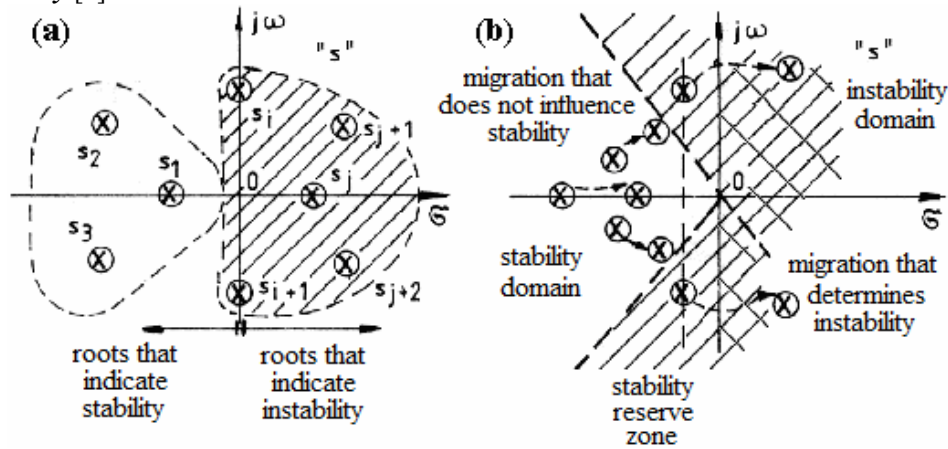
Fig. 5.5. *Graphical interpretation of fundamental stability theorem of C-LTIS [9], [10].*

Remark 1: Since the systems zeros have an important effect on system responses, their reparation cannot be neglected.

Remark 2: The results presented in this section are not based on any assumption on the system structure, i.e. if it is organized in open-loop or closed-loop. The theory is applicable in the same manner irrespective of system structure.

B) Specific aspects of the stability analysis of continuous-time conventional control systems (control loops, C-CCS) characterized by IO-MM. Many control systems have are organized in terms of C-CCS structures for which the following operational input-output relation is expressed:

R.-E. Precup, UPT, 2024

$$z(s) = H_{zr}(s)r(s) + H_{zd}(s)d(s), \tag{5.2.7}$$

where r is the reference input (or the setpoint), d is the disturbance input, z is the system output (the assessed output), $H_{zr}(s)$ is the C-CCS t.f. with respect to the reference input, and $H_{zd}(s)$ is the C-CCS t.f. with respect to the disturbance input. The following relation is used to express the characteristic equation of C-CCS:

$$\Delta(s) = 1 + H_0(s) = 0, \; H_0(s) = H_C(s)H_{CP}(s), \tag{5.2.8}$$

where $H_0(s)$ is the open-loop t.f., $H_C(s)$ is the controller t.f., and $H_{CP}(s)$ is the controlled process t.f.

In certain particular situations, targeted by the designer, the t.f. $H_0(s)$ can contain certain **common factors** in its nominator and denominator, which can be **simplified** from a mathematical point of view.

Remark 3: Such a (series) structure is referred to as ***system with inverse model*** (the controller model is equivalent to the inverse of process model).

However, the above mentioned simplification can result in shortcomings with negative effects as far as the control system stability analysis is concerned. These effects are **totally negative** if the common factors characterize poles and zeros (with the same values) that are placed in **the right half-plane of roots plane**. This aspect is illustrated in the next example.

Example 5.1 [9], [10]: Conduct the stability analysis of the linear control system with the structure given in Fig. 5.6, where:

- case (I):

$$H_C(s) = \frac{1+4s}{1+s}, H_{CP}(s) = \frac{1}{(1+2s)(1+4s)};$$

- case (II):

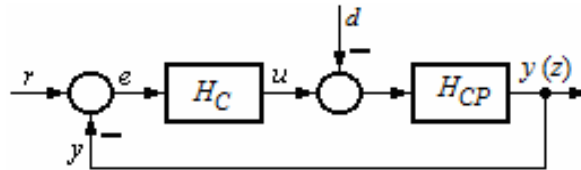$$H_C(s) = \frac{1-4s}{1+s}, H_{CP}(s) = \frac{1}{(1+2s)(1-4s)}.$$

Fig. 5.6. *Conventional control system structure (control loop).*

Solution:

Case (I): The characteristic equation is expressed:

$$\Delta(s) = 1 + H_0(s) = 1 + \frac{1+4s}{1+s} \cdot \frac{1}{(1+2s)(1+4s)} = 0.$$

(1) Without the simplification of the common factor (1+4s) (correctly done), the result is

$$\Delta(s) = (1+4s)(2+3s+2s^2) = 0,$$

with the roots:

$$s_1 = -\frac{1}{4}, \ s_{2,3} = \frac{-3 \pm \sqrt{7}}{4};$$

therefore, the system is stable.

(2) Simplifying the common factor (1+4s), the characteristic equation becomes

$$\Delta'(s) = (1+2s)(1+s) + 1 = 2s^2 + 3s + 2 = 0,$$

with the roots:

$$s'_{1,2} = \frac{-3 \pm \sqrt{7}}{4};$$

therefore, the system is stable.

Case (II): The characteristic equation is expressed:

$$\Delta(s) = 1 + H_0(s) = 1 + \frac{1-4s}{1+s} \cdot \frac{1}{(1+2s)(1-4s)} = 0.$$

(1) Without the simplification of the common factor (1−4s) (correctly done), the result is

$$\Delta(s) = (1-4s)(2+3s+2s^2) = 0,$$

with the roots:

$$s_1 = \frac{1}{4}, \ s_{2,3} = \frac{-3 \pm \sqrt{7}}{4};$$

therefore, the system is unstable.

(2) Simplifying the common factor (1−4s), the characteristic equation becomes

$$\Delta'(s) = (1+2s)(1+s) + 1 = 2s^2 + 3s + 2 = 0,$$

with the roots:

$$s'_{1,2} = \frac{-3 \pm \sqrt{7}}{4};$$

this indicates that the system is stable, which is a wrong conclusion.

If in case (I), the (forbidden) simplification of the common factor did not affect the stability property, in case (II), this simplification affects the correct conclusion which states that the system is unstable.

Concluding, if the controlled process (CP) has only poles placed in the left half-plane, which aiming the objective of compensation (for faster control system dynamics), will be zeros of the controller (C) as well, and by their simplification the stability property seems not to be affected, in case of CP with poles placed in the right half-plane, their compensation with zeros placed in the right-half plane (brought by C) will be forbidden. The conclusion on the stability of the system obtained after the simplification is completely wrong as illustrated in the example 5.1.

Concluding, the following **formulation** will be used in the stability of conventional control systems:

Theorem 2: A C-CCS is stable if and only if **all roots of its characteristic equation**

$$\Delta(s) = 1 + H_0(s) = 0 \tag{5.2.9}$$

are placed strictly in the left half-plane of the roots plane, without allowing any simplification in $H_0(s)$ of factors that would affect the system order.

The simplification of factors is an important issue accounting for the fact that the design methods based on the compensation of large time constants of the CP by the C are based on creating such factors in the C structure (its t.f., in the context of inverse process model).

A remarkable special case is that of **time delay systems** (T_d), where the t.f. $H_0(s)$ is expressed as

$$H_0(s) = \frac{B_0(s)}{A_0(s)} \cdot e^{-sT_d}, \tag{5.2.10}$$

where $A_0(s)$ and $B_0(s)$ are polynomials. Accordingly, the characteristic equation is the next transcendental equation:

$$A_0(s) + B_0(s)e^{-sT_d} = 0, \tag{5.2.11}$$

for which there is no sense to speak about the degree and the system order.

That is the reason why, the correct and complete formulation of the fundamental stability theorem refers to all roots of the characteristic equation and not to the *n* roots of the characteristic equation.

Generally speaking, the stabilization of systems (processes) with large time delays leads to big problems in the development of control systems with very good performance, where guaranteeing the stability is the minimum requirement.

5.3. Stability criteria for continuous-time linear time invariant systems

The efficient quantitative assessment of stability is carried out on the basis of criteria for the assessment of the stability of linear systems. These criteria are briefly called stability criteria. Two types of stability criteria based on IO-MM have been developed:

- algebraic criteria,
- frequency criteria, which are usually expressed in terms of graphic-analytic formulations.

5.3.1. The Hurwitz criterion for the assessment of stability of continuous-time linear systems

The Hurwitz criterion is an algebraic criterion that sets – depending on the values of the coefficients of the characteristic equation – the necessary and sufficient conditions for its roots to be placed in the left half-plane of the complex roots plane.

The following theorem specific to algebra is reminded, without proof, prior to the presentation of the Hurwitz criterion:

Theorem 3: The **necessary** (but not sufficient) condition for the roots of an algebraic equation

$$\Delta(s) = a_n s^n + a_{n-1} s^{n-1} + ... + a_1 s + a_0 = 0 \qquad (5.3.1)$$

to be placed in the left half-plane of roots plane is that **all coefficients** a_ν of the algebraic equation should be **strictly positive**:

$$a_\nu > 0, \ \nu = 0...n. \qquad (5.3.2)$$

Consequently, if the characteristic equation of a C-LTIS expressed in (5.3.1) **has at least one coefficient a_ν that is zero or negative**, the system will be **certainly unstable**; therefore, the further application of any stability criterion is useless.

The presentation of the Hurwitz criterion makes use of a C-LTIS with SS-MM or IO-MM expressed in (5.2.1) and known, with its characteristic equation (5.2.3) also given in (5.3.1).

R.-E. Precup, UPT, 2024

The assessment of the stability of C-LTIS with the coefficients $a_\nu > 0, \ \nu = 0...n$, starts with testing if all coefficients fulfill the conditions (5.3.2).

❑ if not, the system will certainly be unstable;

❑ if yes, the second step consists in building **the Hurwitz matrix** $\mathbf{H} \in \mathfrak{R}^{n \times n}$, with the following structure:

$$\mathbf{H} = \begin{bmatrix} a_{n-1} & a_{n-3} & a_{n-5} & ...\ 0 & 0 \\ a_n & a_{n-2} & a_{n-4} & ...\ 0 & 0 \\ 0 & a_{n-1} & a_{n-3} & ...\ 0 & 0 \\ 0 & a_n & a_{n-2} & ...\ 0 & 0 \\ ... & & & & \\ 0 & 0 & 0 & ...\ a_0 & 0 \\ 0 & 0 & 0 & ...\ a_1 & 0 \\ 0 & 0 & 0 & ...\ a_2 & a_0 \end{bmatrix}. \tag{5.3.3}$$

The following determinants will be next computed on the basis of the Hurwitz matrix $\mathbf{H}$:

• the Hurwitz determinant, i.e. the determinant of the matrix $\mathbf{H}$, and

• the leading principal minors of $\mathbf{H}$ in terms of

$$\det(\mathbf{H}_n) = \det(\mathbf{H}) = a_0 \det(\mathbf{H}_{n-1}),...,$$

$$\det(\mathbf{H}_2) = \det\left(\begin{bmatrix} a_{n-1} & a_{n-3} \\ a_n & a_{n-2} \end{bmatrix} \right), \det(\mathbf{H}_1) = a_{n-1}. \tag{5.3.4}$$

In this context, **the Hurwitz stability criterion is formulated** as follows:

The condition for a C-LTIS with the characteristic equation (5.3.1) to be *stable* is that **the Hurwitz determinant and all its leading principal minors should be strictly positive**.

The following **steps** are proceeded to apply the Hurwitz criterion:

♦ the system whose stability is analyzed is separated and its IO-MM or SS-MM are derived;

♦ the characteristic equation is expressed in terms of (5.3.1); the system order, n, must be kept;

♦ the requirements Theorem 3 are tested, i.e. the condition (5.3.2) is tested for all coefficients; if **the condition (5.3.2) is fulfilled**, then:

♦ the Hurwitz matrix **H** is built and the Hurwitz determinant, $\det(\mathbf{H}_n)$, and the leading principal minors of the Hurwitz matrix, i.e. $\det(\mathbf{H}_1)$, $\det(\mathbf{H}_2)$ … and $\det(\mathbf{H}_{n-1})$ are computed;

♦ the following stability condition is tested for all these minors:

$$\det(\mathbf{H}_i) > 0, \ i = 1...n. \tag{5.3.5}$$

If (5.3.5) is fulfilled, then the system is stable. If not, the system is unstable.

Since the formulation of the criterion does not concern the system type, namely open-loop or closed-loop, it results immediately that the Hurwitz criterion is applied in the same manner irrespective of system type. The computations are based on the characteristic polynomial $\Delta(s)$ of the system whose stability is analyzed.

The Hurwitz criterion has the following **advantages** (A) and **shortcomings** (S):

A/S: The Hurwitz criterion is applied relatively simply to system with a reduced order ($n < 4$ or 5), Once the system order is increased, the "manual" computations are tedious.

A: In some simplified forms, the Hurwitz criterion is subjected to a relatively organization as an algorithm that can be next implemented in a digital version. That is the reason why the Hurwitz criterion is also called ***the Routh-Hurwitz criterion*** (the Routh version of the Hurwitz criterion).

A: The Hurwitz criterion can also be used in the stability analysis of linear systems that depend of one or more parameters of the control system within a certain domain or a domain that can be calculated.

S/A: Accounting for the formulation of the Hurwitz criterion, it cannot be applied to time delay systems. However, if the time delay (T_d) proves to he sufficiently small with respect to the large time constants of the system (for example, it is of the order of magnitude of the small time constants of the system), then, the use of a Pàde approximation (with finite dimension) of the expression $\exp(-s\, T_d)$ makes the criterion applicable.

The simplest and most frequently used Pàde approximation is

$$e^{-sT_d} = \exp(-s\, T_d) \approx \frac{1}{1 + s\, T_d}. \tag{5.3.6}$$

S: The Hurwitz criterion does not give information on the stability degree of the system. It also does not give information on the trends related to the modification of the system stability.

Example 5.2 [9], [10]: Let us consider the simplified control system structure of the angular speed (frequency) ω control system of a hydrogenerator given in Fig. 5.7. The t.f.s of the blocks are

- controller (C):

$$H_C(s) = k_C \frac{1+8s}{1+20s},$$

- pipeline/penstock-turbine (PT):

$$H_{PT}(s) = \frac{1-4s}{1+2s},$$

- synchronous generator (SG):

$$H_{SG}(s) = \frac{1}{\alpha+7s},$$

with $\alpha \in (0.1, 1.2]$, which characterizes the strength of connection of SG to the power system, and $k_C > 0$. It is required:
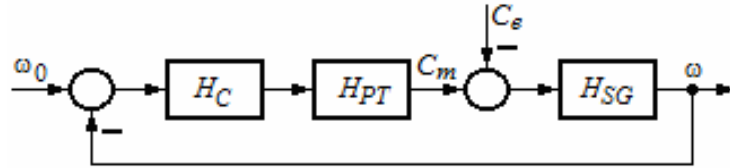

Fig. 5.7. *Control loop of hydrogenertor speed control system.*

1. Considering $\alpha = 1$, the domain of values of $k_C > 0$ for which the system is stable should be determined, where k_{Cmax} is the upper limit of that domain.
2. Setting $k_{C0} = k_{Cmax}/4$, it should be analyzed if for $\alpha = 0.2$ the control system is or is not stable and the lower limit of k_C for which the system remains stable should be computed.

Solution: 1. Considering $\alpha = 1$, the characteristic equation of the control system is expressed in terms of

$$\Delta(s) = 1 + H_0(s),$$

where

$$H_0(s) = H_C(s)H_{PT}(s)H_{SG}(s) = \frac{k_C(1+8s)(1-4s)}{(1+20s)(1+2s)(1+7s)};$$

therefore, the characteristic equation becomes

R.-E. Precup, UPT, 2024

$$\Delta(s) = 280s^3 + (194 - 32k_C)s^2 + (29 + 4k_C)s + 1 + k_C = 0.$$

The necessary stability conditions (5.3.2) specified in Theorem 3 are imposed:

$$194 - 32k_C > 0,$$
$$29 + 4k_C > 0,$$
$$1 + k_C > 0.$$

Since $k_C > 0$, this system of inequations leads to

$$k_C \in (0, 6.0625).$$

The Hurwitz matrix (n=3) is next built, and the Hurwitz determinant is

$$\det(\mathbf{H}) = \det(\mathbf{H}_3) = \begin{vmatrix} 194 - 32k_C & 1 + k_C & 0 \\ 280 & 29 + 4k_C & 0 \\ 0 & 194 - 32k_C & 1 + k_C \end{vmatrix}.$$

The stability conditions are imposed as follows:

$$\det(\mathbf{H}_1) = 194 - 32k_C > 0 \Rightarrow k_C < 194/32 \Leftrightarrow k_C < 6.0625; \qquad (1)$$

$$\det(\mathbf{H}_2) = \begin{vmatrix} 194 - 32k_C & 1 + k_C \\ 280 & 29 + 4k_C \end{vmatrix} > 0$$
$$\Leftrightarrow (194 - 32k_C)(29 + 4k_C) - 280(1 + k_C) > 0$$
$$\Rightarrow k_C \in (-8.3668, 4.9919); \qquad (2)$$

$$\det(\mathbf{H}_3) = \begin{vmatrix} 194 - 32k_C & 1 + k_C & 0 \\ 280 & 29 + 4k_C & 0 \\ 0 & 194 - 32k_C & 1 + k_C \end{vmatrix} > 0 \Leftrightarrow (1 + k_C)\det(\mathbf{H}_2) > 0$$

$$\Leftrightarrow k_C \in (-1, 0) \cap (-8.3668, 4.9919)$$
$$\Rightarrow k_C \in (-1, 4.9919). \qquad (3)$$

Concluding, since $k_C > 0$, the intersection of this condition and the conditions (1), (2) and (3) leads to the domain of values of k_C, namely D_{k_C}, for which the system is stable:

$$D_{k_C} = (0, 4.9919),$$

with

R.-E. Precup, UPT, 2024

$k_{C\max} = 4.9919.$

2. The value of k_{C0} is $k_{C0} = k_{C\max}/4 = 4.9919/4 = 1.248.$
Using $\alpha = 0.2$, the t.f. $H_{SG}(\mathbf{s})$ becomes

$$H'_{SG}(s) = \frac{1}{0.2 + 7s} = \frac{5}{1 + 35s}$$

and accordingly

$$1 + H'_0(s) = H_C(s)H_{PT}(s)H'_{SG}(s) = 1 + \frac{5k_{C0}(1 + 8s)(1 - 4s)}{(1 + 20s)(1 + 2s)(1 + 35s)}. \tag{4}$$

The computation of the nominator in (4) leads to:
- the characteristic polynomial

$$\Delta'(s) = 1400s^3 + 810s^2 + 57s + 1 + 5k_{C0}(-32s^2 + 4s + 1)$$
$$= 1400s^3 + 10(81 - 16k_{C0})s^2 + (57 + 20k_{C0})s + 1 + 5k_{C0};$$

- the necessary stability conditions given in (5.3.2):

$$81 - 16k_{C0} > 0,$$
$$57 + 20k_{C0} > 0,$$
$$1 + 5k_{C0} > 0;$$

since $k_{C0} > 0$, this system of inequations leads to

$$k_{C0} \in (0, 5.0625); \tag{5}$$

- the Hurwitz determinant:

$$\det(\mathbf{H}') = \det(\mathbf{H}'_3) = \begin{vmatrix} 10(81 - 16k_{C0}) & 1 + 5k_{C0} & 0 \\ 1400 & 57 + 20k_{C0} & 0 \\ 0 & 10(81 - 16k_{C0}) & 1 + 5k_{C0} \end{vmatrix};$$

- the stability conditions:

$$\det(\mathbf{H}'_1) = 10(81 - 16k_{C0}) > 0 \Rightarrow k_{C0} < 81/16 \Leftrightarrow k_{C0} < 5.0625; \tag{6}$$

$$\det(\mathbf{H}'_2) = \begin{vmatrix} 10(81 - 16k_{C0}) & 1 + 5k_{C0} \\ 1400 & 57 + 20k_{C0} \end{vmatrix} > 0$$

$$\Leftrightarrow 10(81 - 16k_{C0})(57 + 20k_{C0}) - 1400(1 + 5k_{C0}) > 0$$

$$\Rightarrow k_C \in (-3.7279, 3.7529); \tag{7}$$

$$\det(\mathbf{H}_3^{'}) = (1 + 5k_{C0})\det(\mathbf{H}_2^{'}) > 0 \Leftrightarrow (1 + 5k_{C0})\det(\mathbf{H}_2) > 0$$

$$\Leftrightarrow k_{C0} \in (-0.2,0) \cap (-3.7279,3.7529)$$

$$\Rightarrow k_{C0} \in (-0.2,3.7529). \tag{8}$$

The intersection of the conditions (5), (6), (7) and (8) leads to the domain of values of k_{C0}, namely $D_{k_{C0}}$, for which the system is stable:

$$D_{k_{C0}} = (0,3.7529).$$

Since the value $k_{C0} = 1.248$ belongs to the stability domain $D_{k_{C0}}$, the conclusion is that the control system is stable for $k_{C0}=k_{Cmax}/4$.

A careful analysis of the real operating conditions of the control system will lead to the conclusion that the lower limit of k_C for which the system is stable is 0.

Remark: If the real control system is considered, it is possible that the domain D_{k_C} that guarantees the control system stability is more constrained than that obtained in this example. This is due to at least the following reasons:

- the incomplete mathematical modeling of the real system, which can result in approximate values of the coefficients in the t.f.s;
- possible further modifications of the parameters of the process if the modifications of the operating point occur during the real-time operation of the system.

5.4. Stability of discrete-time linear time invariant systems (D-LTIS). Fundamental stability theorem of D-LTIS

The approach in case of D-LTIS is similar to that in case of C-LTIS.

A) The fundamental stability theorem of D-LTIS. Let us consider the SISO D-LTIS characterized by
SS-MM:

$$\mathbf{x}_{k+1} = \mathbf{A}\,\mathbf{x}_k + \mathbf{b}\,u_k,$$
$$y_k = \mathbf{c}^T\,\mathbf{x}_k, \tag{5.4.1-a}$$

or IO-MM:

$$\sum_{\nu=0}^{n} a_\nu y_{k+\nu} = \sum_{\mu=0}^{m} b_\mu u_{k+\mu}, \ m < n. \tag{5.4.1-b}$$

As in case of C-LTIS, generally speaking, the two models reflect the same reality, pointed out, for example, by the same t.f. $H(z)$:

$$H(s) = \begin{cases} \mathbf{c}^T (z\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \mathbf{c}^T \dfrac{\text{adj}(z\mathbf{I} - \mathbf{A})}{\det(z\mathbf{I} - \mathbf{A})}\mathbf{b}, & (5.4.2\text{-a}) \\[4mm] \dfrac{b_m z^m + ... + b_1 z + b_0}{a_n z^n + ... + a_1 z + a_0}. & (5.4.2\text{-b}) \end{cases}$$

The characteristic polynomial $\Delta(z)$ and the characteristic equation $\Delta(z) = 0$ can be expressed in one of the two forms

$$\Delta(z) = \begin{cases} \det(z\mathbf{I} - \mathbf{A}) = 0, & (5.4.3\text{-a}) \\[4mm] \displaystyle\sum_{v=0}^{n} a_v z^v = 0. & (5.4.3\text{-b}) \end{cases}$$

The theory of recurrent equations states that the homogenous solution to the recurrent equation given in (5.4.3) is given by the roots of the characteristic equation

$$\Delta(z) = a_n z^n + a_{n-1} z^{n-1} + ... + a_1 z + a_0 = 0. \tag{5.4.4}$$

Depending on the roots of this equation, the homogenous solution (or the system unforced response or the system free response) will be [11], [12]:

- stable, if $|z_v| < 1$, with z_v, $v = 1...n$ – the root of the characteristic equation (the pole of D-LTIS);
- unstable, if at least one of the roots of the characteristic equation fulfils the condition $|z_v| > 1$;
- for the roots of the characteristic equations that fulfill the conditions $|z_v| = 1$, particular operating regimes are installed in the system, and they indicate instability; in this regard, for:
  $z_v = 1$, the (discrete-time) unit impulse sequence response is constant and equal to 1, and the unit step sequence response is linearly increasing,
  $z_v = -1$, the (discrete-time) unit impulse sequence response is oscillatory, of constant magnitude, and the unit step sequence response has an increasing magnitude.

These aspects are synthesized in ***the fundamental stability theorem of D -LTIS***:

Theorem 4: A D-LTIS characterized by the SS-MM / IO-MM expressed in (5.4.1) – and, accordingly, the PS described by this DS – is stable **if and only** if the n roots z_v, $v = 1...n$, of the characteristic equation (5.4.4) (i.e., the system poles) have **modulus less than one**, i.e.

$$|z_v| < 1, \ v = 1...n. \tag{5.4.5}$$

In other words, **the roots are placed inside the unit disk of the "z" plane** illustrated in Fig. 5.8.

Considering four distinct values of the pole z_1 of a D-LTIS, the graphical interpretation of the fundamental stability theorem is illustrated in Fig. 5.8 in terms of unit impulse sequence responses.
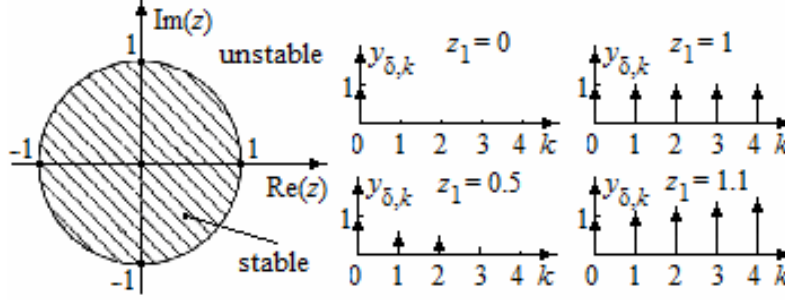

Fig. 5.8. *Graphical interpretation of fundamental stability theorem of D-LTIS using unit impulse sequence responses [9], [10].*

In the discrete-time case, because of practical reasons mentioned for C-LTIS in Section 5.3.1, the pole placement is preferred to be carried out in **favorable zones** that will ensure:

- favorable transient regimes;
- keeping the stability even in the conditions of pole migration due to modifications of system parameters; for example, this is the case of parameters of the continuous controlled process with the t.f. $H_{CP}(s)$ that lead, by discretizing it, to the t.f. $H_{CP}(z)$ used in stability analysis.

Remark: Such pole migrations appear by changing the value of the sampling period T_s. In these conditions, for example, using the same value of k_C (the controller gain), two values of the sampling period, $T_s^{(1)}$ and $T_s^{(2)}$, can lead to different behaviors of the control system.

B) Connections between the "s" plane and the "z" plane. As previously mentioned, the analysis and design of continuous-time control systems, the knowledge on the placement of the "continuous" poles and zeros in the "s" roots plane is an efficient way to get information on the system dynamics. This will be next used to design the control system.

In order to get experience in case of discrete-time control systems, it is useful to investigate the connection between the placement of roots in the "s" plane (corresponding to the continuous time) and the placement of the corresponding roots in the "z" plane (afferent to the discrete time). The following connection is used in this investigation:

$$z = e^{sT_s} \Leftrightarrow z^{-1} = e^{-sT_s}, \tag{5.4.6}$$

which allows for the derivation of connections between the placement of poles in the "s" plane and that in the "z" plane.

Replacing $s = \sigma + j\omega$, in (5.4.6) leads to

$$z = e^{\sigma T_s + j\omega T_s} = e^{\sigma T_s} \cdot e^{j\omega T_s} = \mid z \mid \cdot e^{j\underline{/z}}, \tag{5.4.7}$$

where

$$\mid z \mid = e^{\sigma T_s}, \tag{5.4.8-a}$$

$$\underline{/z} = \omega T_s. \tag{5.4.8-b}$$

The relationship (5.4.8-a) is used to express the stability condition in the continuous-time case, $\sigma < 0$, in the known condition

$$\mid z \mid = e^{\sigma T_s} < 1 \Leftrightarrow \sigma < 1. \tag{5.4.9}$$

In other words, **the two stability conditions stated in the fundamental stability theorems have the same contents**.

Remark: The stability criteria in the discrete-time case are different to those in the continuous-time case.

The relationship (5.4.8-b) points out the fact that **the connection between the "s" plane and the "z" plane is not univocal**; each point in the "z" plane is mapped onto an infinite number of points in the "s" plane. This is explained considering a point z_1 in the "z" plane, which is mapped onto a set of points characterized by the same modulus, $|z_1|$, but with different arguments $\underline{/z_1}$:

$$z_1 = e^{s_1 T_s} = e^{\sigma_1 T_s} \cdot e^{j\omega(T_s + 2m\pi)}, \tag{5.4.10-a}$$

or

$$z_1 = [e^{\sigma_1 T_s} \cdot e^{j\omega T_s}] \cdot e^{j \cdot 2m\pi}, \tag{5.4.10-b}$$

where $m = 0, \pm 1, \pm 2, \dots$, and the factor $[e^{\sigma_1 T_s} \cdot e^{j\omega T_s}]$ corresponds to the "fundamental" frequency (pulsation). This situation is illustrated in Fig. 5.9, which is processed using the initial plots given in [13].

Remarks:

1. Equations (5.4.7) … (5.4.10) regarding the main zone are useful in mapping the specific features in the "s" plane to the "z" one and vice versa. The width of the main zone, i.e. $2 \cdot (\omega_s/2)$, depends on the value of the sampling period T_s:

- if T_s is decreasing, ω_s will increase and the width of the main zone will increase accordingly, allowing for the inclusion of a larger number of the continuous system poles (complex conjugated ones);

- if T_s is increasing, ω_s will decrease and the width of the main zone will decrease accordingly, which prevents the inclusion of possible essential complex conjugated poles; this will falsify the interpretation of the information transmitted by these poles in relation with the system.
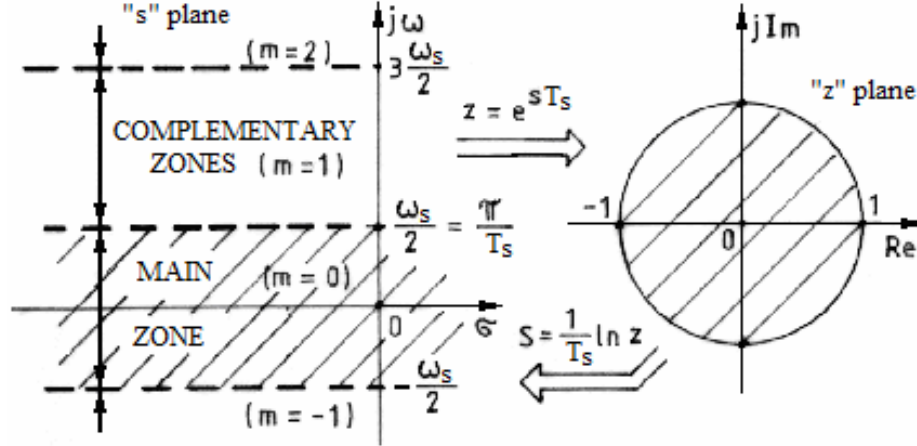

Fig. 5.9. Correspondence between the points in the planes "s" and "z" [9], [10].

2. The relationships presented in this sub-section are correct only in case of D-LTIS by sampling the impulse response function of C-LTIS using the Z transform method. These relationships are more or less correct approximations if the discrete-time mathematical models of D-LTIS are obtained by other discretization methods. This conclusion is normal in the context of D-LTIS analysis in the frequency domain.

A useful information for those who analyze and develop discrete-time control systems on the basis of discretize continuous-time control laws is the way that geometric loci with a constant parameter in the "s" plane are mapped on geometric loci with the same constant parameter in the "z" plane. In this regard, Fig. 5.10 gives two such mappings, which are drawn for the second-order continuous-time system with the transfer function

$$H(s) = \frac{1}{1 + 2\zeta Ts + T^2 s^2} = \frac{\omega_0^2}{\omega_0^2 + 2\zeta\omega_0 s + s^2}, \tag{5.4.11}$$

where $\omega_0 = 1/T$ is the natural frequency, ζ is the damping ratio (the notation $\zeta = D$ is also used for it) and the poles

$$p_{1,2} = -\zeta\omega_0 \pm j\omega_0\sqrt{1 - \zeta^2}, \tag{5.4.12}$$

with the equivalent expression

$$p_{1,2} = -\omega_0 \cos\psi \pm j\omega\sin\psi, \qquad\qquad (5.4.13)$$

where

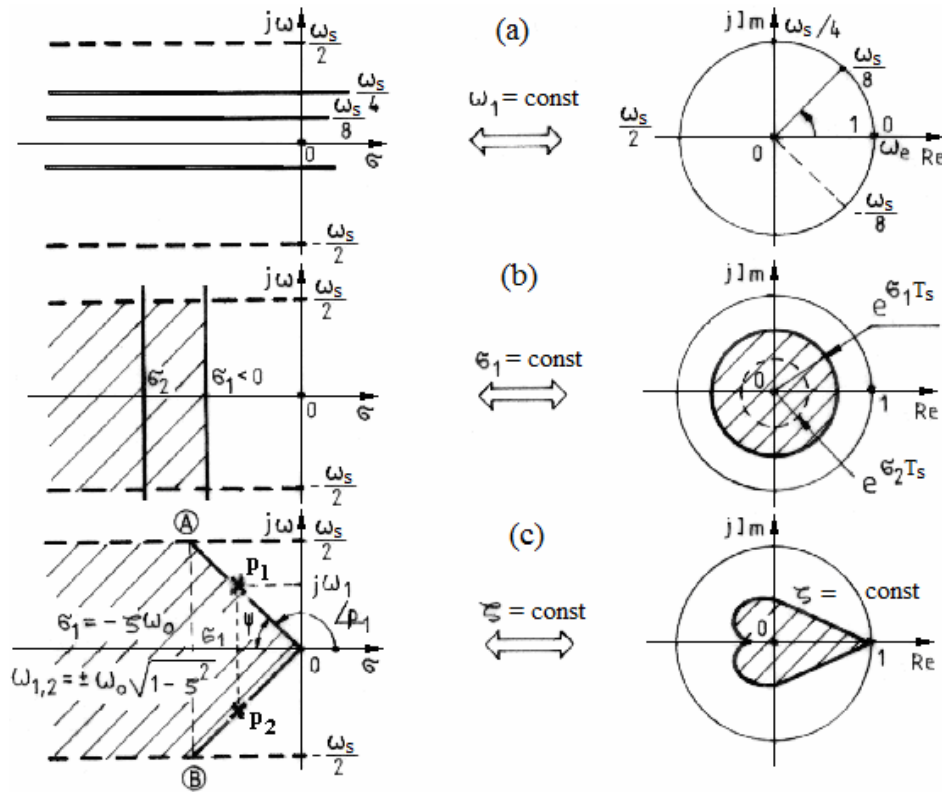$$\text{tg}\psi = \frac{\sqrt{1-\zeta^2}}{\zeta}. \qquad\qquad (5.4.14)$$

Fig. 5.10. *Mappings of geometric loci from the "s" plane to the "z" plane and vice versa [9], [10].*

The system behavior is characterized as follows as function of the value of ζ ($\zeta > 0$) (the condition to place the poles p_1, p_2 in the left half-plane):

$0 < \zeta < 1$: $p_{1,2}$ – complex conjugated $\rightarrow$ oscillatory behavior;

$\zeta > 1$: $p_{1,2}$ – real and distinct $\rightarrow$ aperiodic behavior;

$\zeta = 1$: $p_{1,2}$ – real and equal $\rightarrow$ limit aperiodic behavior.

The discrete-time correspondent to this pair of poles is obtained on the basis of (5.4.10):

$$z_{1,2} = e^{p_{1,2}T_s} = e^{-\zeta\omega_0 T_s} \cdot e^{\pm j\omega_0 T_s \sqrt{1-\zeta^2}}, \tag{5.4.15}$$

where:

$$|z_{1,2}| = e^{-\zeta\omega_0 T_s} = e^{-\omega_0 T \cos\psi}, \quad \underline{/z_{1,2}} = \pm\omega_0 T_s \sqrt{1-\zeta^2} = \pm\omega_0 T_s \sin\psi. \tag{5.4.16}$$

Therefore, the following **mappings of geometric locus of constant parameter** are formulated [13]:

- The geometric locus of constant frequency $\omega_1 =$ const (ω_1 – any value of ω_0 in (5.4.11)) in the "s" plane is mapped onto a ray from the origin in the "z" plane of angle θ with respect to the positive horizontal direction (i.e., the real axis), Fig. 5.10 (a):

$$\theta = \omega_1 T_s. \tag{5.4.17}$$

- The geometric locus of $\sigma_1 = \text{Re}(s_1) =$ const in the "s" plane (a vertical line) is mapped onto a circle centered about at the origin in the "z" plane and with constant radius r, Fig. 5.10 (b):

$$r = |z_1| = e^{\sigma_1 T_s} = \text{const.} \tag{5.4.18}$$

Remark: For the same value of the sampling period T_s, decreasing $|\sigma_1|$ is equivalent to increasing the corresponding time constant T as $T=1/|\sigma_1|$, and the radius of the circle (r) is decreasing.

Consequences:

- the poles in the "s" plane that characterize large time constant and are closer to the $j\omega$ axis of the plane will be placed close to the point $(0, j1)$ in the "z" plane,

- the increase of T_s will lead to the tendency to make those poles closer to the center of the circle (and vice versa).

- The geometric locus of points with constant damping $\zeta =$ const in the "s" plane are mapped on to logarithmic spirals in the "z" plane, Fig. 5.10 (c).

The relations (5.4.11) … (5.4.18) are employed to express the exact mappings of pole placement in the "z" plane as function of the poles of the initial continuous-time transfer functions. These mappings are illustrated in Fig. 5.11 (a).

In order to place the poles of the discrete-time system close to the point $(0, j1)$, which is a favorable placement shown in Fig. 5.11 (b), it is mandatory to connect the natural frequencies of the system (all essential ω_0 frequencies) to the sampling period T_s.

The angles with respect to the positive horizontal direction in the closed zones of the curves $\zeta =$ const ($\zeta > 0.5$) are approximately equal.
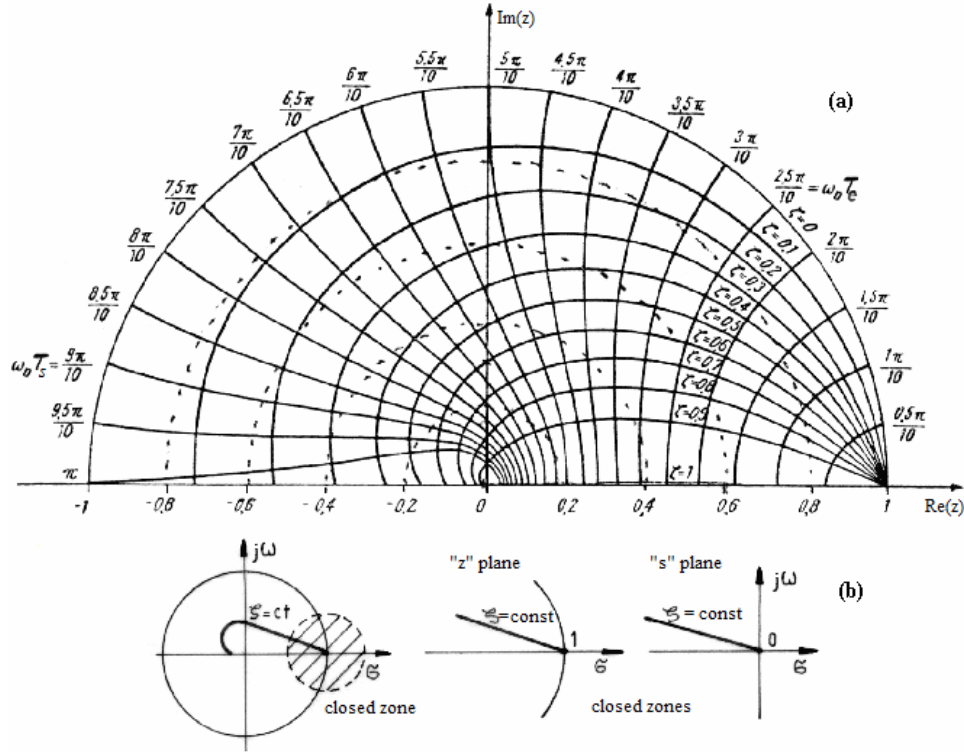
Fig. 5.11. *Mappings of pole placement between C-LTIS and corresponding D-LTIS (a) and characteristics of pole placement in the "z" plane (b) [9], [10].*

5.5. Criteria for discrete-time linear time invariant systems

Similar to the case of C-LTIS, two types of stability criteria for D-LTIS are used:
- algebraic criteria,
- frequency (pulsation) criteria, usually expressed in terms of grapho-analytical formulations.

Accounting for the specific of the placement of D-LTIS poles in the "z" plane, depending on the system stability (or not), it is not possible to carry out the direct transfer of the analysis taken from C-LTIS. That is the reason why two categories of criteria are of interest for D-LTIS:
- criteria based on the extension of the algebraic criteria specific to the continuous-time case (as, for example, the Hurwitz criterion) to discrete-time systems;
- criteria specific to discrete-time systems.

5.5.1. Extension of Hurwitz criterion to discrete-time systems

The extension is based on using the conformal map (mapping)

$$z = \frac{r+1}{r-1} \text{ or } z = \frac{1+w}{1-w}, \qquad (5.5.1)$$

which maps all points inside the unit disk in the "z" plane (that characterize a stable system) to points placed in the left half-plane in the "w" or "r" plane.

Therefore, starting with the system t.f. $H(z)$, assumed to be known, a "pseudo-continuous" t.f. $H(r)$ or $H(w)$ can be computed:

$$H(r) = H(z)\Big|_{z=\frac{r+1}{r-1}} \quad \text{or} \quad H(w) = H(z)\Big|_{z=\frac{1+w}{1-w}}, \qquad (5.5.2)$$

whose poles respect the above mapping. The Hurwitz criterion can next be applied for this t.f., which appears to be extended to D-LTIS as well.

Restricting the presentation to one of the two mappings (the r one to be treated as follows), the expression of the transformed t.f. is

$$H(r) = H(z)\Big|_{z=\frac{r+1}{r-1}} = \frac{B(r)}{A(r)}, \qquad (5.5.3)$$

with the characteristic equation

$$A(r) = \Delta(r) = a_n r^n + a_{n-1} r^{n-1} + ... + a_1 r + a_0 = 0. \qquad (5.5.4)$$

The stability of D-LTIS can next be assessed using the coefficients in (5.5.4) and the extension of Hurwitz criterion.

The following **steps** are proceeded to apply the extended Hurwitz criterion:

♦ the system is analyzed and the t.f. $H(z)$ is derived;
♦ having $H(z)$ (possibly with some coefficients given as parameters) known, the t.f. $H(r)$ or $H(w)$ is computed;
♦ the characteristic polynomial $A(r)$ or $A(w)$ is derived;
♦ all steps specific to the Hurwitz criterion in the continuous-time case are applied.

The conclusion on the **stability** or **instability** obtained after these steps is valid to the discrete-time system as well.

Remark: The mapping (5.5.2) should not be viewed as "mapping back" the poles in the "z" plane as poles of the discretized continuous-time system with the characteristic polynomial $\Delta(s)$ because this mapping is not related to the sampling period T_s. The effect of discretizing on the system stability is embedded in the initial

conversion from continuous-time (s) to discrete-time (z). The extended Hurwitz criterion just checks this effect as far as the stability is concerned.

5.5.2. The Jury criterion for the assessment of stability of discrete-time linear systems

The Jury criterion is specific to discrete-time systems. Accepting that the t.f. $H(z)$ of D-LTIS is known, the characteristic equation

$$\Delta(z) = a_n z^n + a_{n-1} z^{n-1} + ... + a_1 z + a_0 = 0 \text{ with } a_n > 0 \tag{5.5.5}$$

is involved in building **the array for Jury's stability** test (also called **the Jury array**) given in Table 5.2.

Table 5.2. Array for Jury's stability test.

Row	z^0	z^1	z^2	$...z^{n-k}...$	z^{n-2}	z^{n-1}	z^n
1	a_0	a_1	a_2	$...a_{n-k}...$	a_{n-2}	a_{n-1}	a_n
2	a_n	a_{n-1}	a_{n-2}	$...a_k...$	a_2	a_1	a_0
3	b_0	b_1	b_2	$...b_{n-k}...$	b_{n-2}	b_{n-1}	–
4	b_{n-1}	b_{n-2}	b_{n-3}	$...b_k...$	b_1	b_0	–
5	c_0	c_1	c_2	$...c_{n-k}...$	c_{n-2}	–	–
6	c_{n-2}	c_{n-3}	c_{n-4}	$...c_k...$	c_0	–	–
...	...	...	...	...	–	–	–
$2n-5$	p_0	p_1	p_2	p_3	–	–	–
$2n-5$	p_3	p_2	p_1	p_0	–	–	–
$2n-3$	q_0	q_1	q_2	–	–	–	–

As shown in Table 5.3, the elements of the even-numbered rows are the elements of the preceding row in reverse order. The elements of the odd-numbered rows are computed in terms of

$$b_k = \begin{vmatrix} a_0 & a_{n-k} \\ a_n & a_k \end{vmatrix}, \ c_k = \begin{vmatrix} b_0 & b_{n-1-k} \\ b_{n-1} & b_k \end{vmatrix}, \ d_k = \begin{vmatrix} c_0 & c_{n-2-k} \\ c_{n-2} & c_k \end{vmatrix}, ...,$$

$$q_0 = \begin{vmatrix} p_0 & p_3 \\ p_3 & p_0 \end{vmatrix}, \ q_1 = \begin{vmatrix} p_0 & p_2 \\ p_3 & p_1 \end{vmatrix}, \ q_2 = \begin{vmatrix} p_0 & p_1 \\ p_3 & p_2 \end{vmatrix}. \tag{5.5.6}$$

Using the array for Jury's stability test given in Table 5.2, **the Jury stability criterion is expressed as follows**:

The D-LTIS with the characteristic polynomial (5.5.4) is **stable** (i.e., all roots are placed inside the unit disk) if and only if the following $n+1$ **conditions** are fulfilled (with $a_n > 0$):

$\Delta(1) > 0,$ (1)
$\Delta(-1) > 0$ if n is even,   (2)
     < 0 if n is odd,
$|a_0| < a_n,$ (3)
$|b_0| > |b_{n-1}|,$ (4)
$|c_0| > |c_{n-2}|,$ (5)
$|d_0| > |d_{n-3}|,$ (6)
…
$|q_0| > |q_2|.$ (n+1)

Remarks: 1. Although building the array seems to be heavy, gaining experience after few applications makes it rather simple.

2. The coefficients b_k are not related to the coefficients in the nominator of the system t.f.

3. For a second-order system, the array contains only one row.

4. As in the case of the Hurwitz criterion, the Jury criterion has also the shortcoming of not giving information on the stability degree of the system. Since the number of inequalities is rather high, it is difficult to conduct a stability analysis that depends on one or more system parameters.

The following **steps** are proceeded to apply the Jury criterion:
♦ the system whose stability is analyzed is separated, its characteristic polynomial $\Delta(z)$ is expressed and the system order n is identified;
♦ $\Delta(1)$ and $\Delta(-1)$ are computed and the conditions (1), (2) and (3) are tested;
♦ if one of the conditions (1), (2) and (3) is not satisfied, the criterion is stopped and the system is unstable;
♦ otherwise, the Jury array is built and the rest of (n–2) conditions are tested one by one; if one of the conditions is not satisfied, the criterion is stopped and the system is unstable; otherwise, the system is stable.

Example 5.4 [9], [10]: Conduct the stability analysis of the discrete-time linear system with the t.f.

$$H(z) = \frac{11z^2 - 3z + 0.5}{z^3 + 3z^2 + 4z + 0.5}.$$

Solution: The characteristic polynomial of the system is

$$\Delta(z) = z^3 + 3z^2 + 4z + 0.5,$$

with $n = 3$ and $a_3 = 1 > 0$. The first three stability conditions are tested:

$$\Delta(1) = 8.5 > 0, \tag{1}$$

$$\Delta(-1) = -1.5 < 0 \; (n = 3 \text{ is odd}), \tag{2}$$

$$|a_0| = 0.5 < a_3 = 1. \tag{3}$$

Since all these conditions are satisfied, the Jury array is built. It is given in Table 5.3, and its elements are computed as follows:

$$b_0 = \begin{vmatrix} a_0 & a_3 \\ a_3 & a_0 \end{vmatrix} = a_0^2 - a_3^2 = -0.75,$$

$$b_1 = \begin{vmatrix} a_0 & a_2 \\ a_3 & a_1 \end{vmatrix} = a_0 a_1 - a_2 a_3 = -1,$$

$$b_2 = \begin{vmatrix} a_0 & a_1 \\ a_3 & a_2 \end{vmatrix} = a_0 a_2 - a_1 a_3 = -2.5.$$

Table 5.3. *Jury array for the example 5.4.*

Row	z^0	z^1	z^2	z^3
1	0.5	4	3	1
	(a_0)	(a_1)	(a_2)	(a_3)
2	1	3	4	0.5
	(a_3)	(a_2)	(a_1)	(a_0)
3	–0.75	–1	–2.5	–
	(b_0)	(b_1)	(b_2)	
4	–2.5	–1	–0.75	–
	(b_2)	(b_1)	(b_0)	

The last, namely fourth stability condition ($n+1 = 4$) is next tested:

$$|b_0| = 0.75 < |b_2| = 2.5. \tag{4}$$

Since this condition is not satisfied, the conclusion is that the system is unstable.

The same conclusion can be reached if the system poles are computed, for example, using Matlab [14]: $z_1 = -0.139$, $z_{2,3} = -1.43 \pm j1.25$. These poles are placed outside the unit disk, which indicates that the system is unstable.

5.6. Aspects concerning the stability analysis of control loops with continuous-time processes and discrete-time controllers

The situation specific to the majority of practical control applications is characterized by:

- a discrete-time controller (C) characterized by the t.f. $H_C(z)$;
- a continuous-time controlled process (CP) characterized by the t.f. $H_{CP}(s)$.

The interfacing of the two sub-systems, C and CP, is ensured by:

- the analog-to-digital converter (ADC), with the informational equivalent represented by sampler (S) + zero-order hold (ZOH);
- the digital-to-analog converter (DAC), with the informational equivalent represented by sampler (S).

The digital control algorithm specific to the discrete-time controller can be obtained by:

- discretizing a continuous-time control law (by the backward difference rule, the forward difference rule or Tustin's rule) using a fixed sampling period T_s;
- direct design, when the value of the sampling period T_s can be often chosen to be significantly larger.

Therefore, **the stability analysis of discrete-time control system** is carried out in terms of the next **steps**:

- ♦ the discrete-time control algorithm is specified by its t.f. $H_C(z)$;
- ♦ the continuous-time process is specified by its t.f. $H_{CP}(s)$;
- ♦ the discrete (or pulse) t.f. or the process extended with the elements (S+ZOH) and (S) is computed:

$$H_{CP}(z) = \frac{z-1}{z} Z\{L^{-1}\{\frac{1}{s}H_{CP}(s)\Big|_{t=kT_s}\}\};$$ (5.6.1)

- ♦ the discrete t.f. $H_r(z)$ of the control system with respect to the reference input r is next computed

$$H_r(z) = \frac{H_C(z)H_{CP}(z)}{1+H_C(z)H_{CP}(z)} = \frac{H_0(z)}{1+H_0(z)},$$ (5.6.2)

where $H_0(z) = H_C(z)H_{CP}(z)$ is the open-loop t.f., and the characteristic equation that uses the polynomial $\Delta(z)$ is expressed:

$$\Delta(z) = 1 + H_0(z) = a_n z^n + a_{n-1}z^{n-1} + ... + a_1 z + a_0 = 0;$$ (5.6.3)

- ♦ the steps of the stability analysis criterion are applied.

Remark: The increase of the value of the sampling period T_s usually leads to harder stability conditions. Any modification of T_s requires the re-computation of the mathematical models of both the control algorithm and the controlled process, which

result in a different $H_r(z)$ and, therefore, $\Delta(z)$, and carrying out again the stability analysis.

Example 5.5 [9], [10]: Let us consider a control system with discrete-time controller and continuous-time process modeled by the transfer functions

$$H_C(z) = k_c \frac{z\,T_s}{z-1} \tag{1}$$

(a discrete-time integral (I) algorithm) and

$$H_{CP}(s) = \frac{2}{(1+s)^2}.$$

The stability analysis of this system should be carried out for $T_s = 0.2$ s using k_c as a parameter and the domain of values of k_c for which the control system is stable should be determined.

Solution: The t.f. of the extended process, $H_{CP}(z)$, is computed using (5.6.1):

$$H_{CP}(z) = \frac{0.69z - 0.624}{z^2 + 1.367z + 0.67}. \tag{2}$$

Using (1) and (2) in (5.6.2), the t.f. $H_r(z)$ is computed:

$$H_r(z) = \frac{k_c(0.138z^2 - 0.125z)}{z^3 + (0.138k_c - 2.637)z^2 + (2.38 - 0.125k_c)z - 0.67}, \tag{3}$$

and next the characteristic polynomial by the substitution of (3) in (5.6.3):

$$\Delta(z) = z^3 + (0.138k_c - 2.637)z^2 + (2.38 - 0.125k_c)z - 0.67, \tag{4}$$

with $n = 3$.

The extended Hurwitz criterion is applied to conduct the stability analysis in terms of replacing

$$z = \frac{1+w}{1-w}$$

in (4). The expression of the transformed characteristic polynomial $\Delta(w)$ is

$$\Delta(w) = b_3 w^3 + b_2 w^2 + b_1 w + b_0,$$

where:

$$b_3 = 0.616 - 0.296k_c, \; b_1 = 0.269k_c + 0.0657,$$
$$b_2 = 1.319 - 0.013k_c, \; b_0 = 0.013k_c. \tag{5}$$

The Hurwitz determinant is built using the coefficients in (5):

$$\det(\mathbf{H}_3) = \begin{vmatrix} b_2 & b_0 & 0 \\ b_3 & b_1 & 0 \\ 0 & b_2 & b_0 \end{vmatrix}. \tag{5}$$

The stability analysis leads to the condition

$$0 < k_c < 25.17.$$

The reader is advised to conduct again the stability analysis in this example but for $T_s = 0.5$ s and $T_s = 1$ s, followed by the comparison of the results and formulating comments on them.

The aspects presented in this sub-chapter will also be treated in the next chapter. Additional details on the stability analysis of linear systems are given in [9] and [10].

Stability in the Frequency Domain

Bode plots and closed-loop system stability

It is important to determine whether a system is stable and if so, than the degree of stability is important to determine. We can use the frequency response of a transfer function around a feedback loop to provide answers to our questions about system's relative stability.

A frequency domain stability criterion relates the stability of a closed-loop system to the open-loop frequency response and open-loop pole location.

Our problem is the determination of stability from frequency data plots, for the closed-loop system.

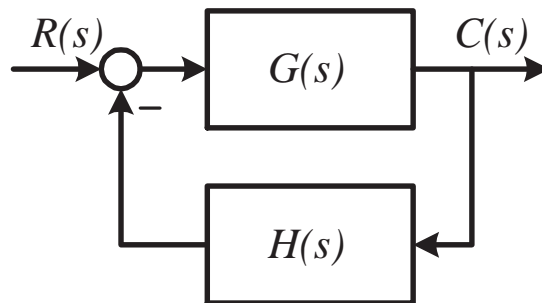Consider the closed-loop control system shown in Figure 11.

Figure 11: Closed-loop system

The zeros of the characteristic equation $1 + G(s)H(s) = 0$ are the poles of the closed-loop transfer function:

$$T(s) = \frac{G(s)}{1 + G(s)H(s)}$$

Thus by examining the open loop transfer function, $G(s)H(s)$ (denoted here $GH(s)$), we can determine the stability (or otherwise) of the closed-loop system.

A point on the imaginary axis $s = j\omega$ will be a solution of the characteristic equation (i.e. the system is critically stable) if $|GH(j\omega)| = 1$ and $\angle GH(j\omega) = \pm 180^o$.

Since we have access to $|GH(j\omega)|$ and $\angle GH(j\omega)$ from a Bode plot, we should be able to determine the imaginary axis crossings by finding the frequencies ω (if any) on the plot that satisfy the conditions:

$$|GH(j\omega)| = 1 \quad \text{and} \quad \angle GH(j\omega) = \pm 180^o$$

We shall use the following terminology:

- **Gain crossover frequency**: This is the frequency ω_g such that $|GH(j\omega_g)| = 1$ (or equivalently, $20 \log_{10} |GH(j\omega_g)| = M^{dB}(\omega_g) = 0$).

- **Phase crossover frequency**: This is the frequency ω_p such that $\angle GH(j\omega_p) = \pm 180^o$.

Example 1. Consider a closed-loop system having the open-loop transfer function:

$$GH(s) = \frac{K}{s(s+1)(\frac{1}{100}s+1)}$$

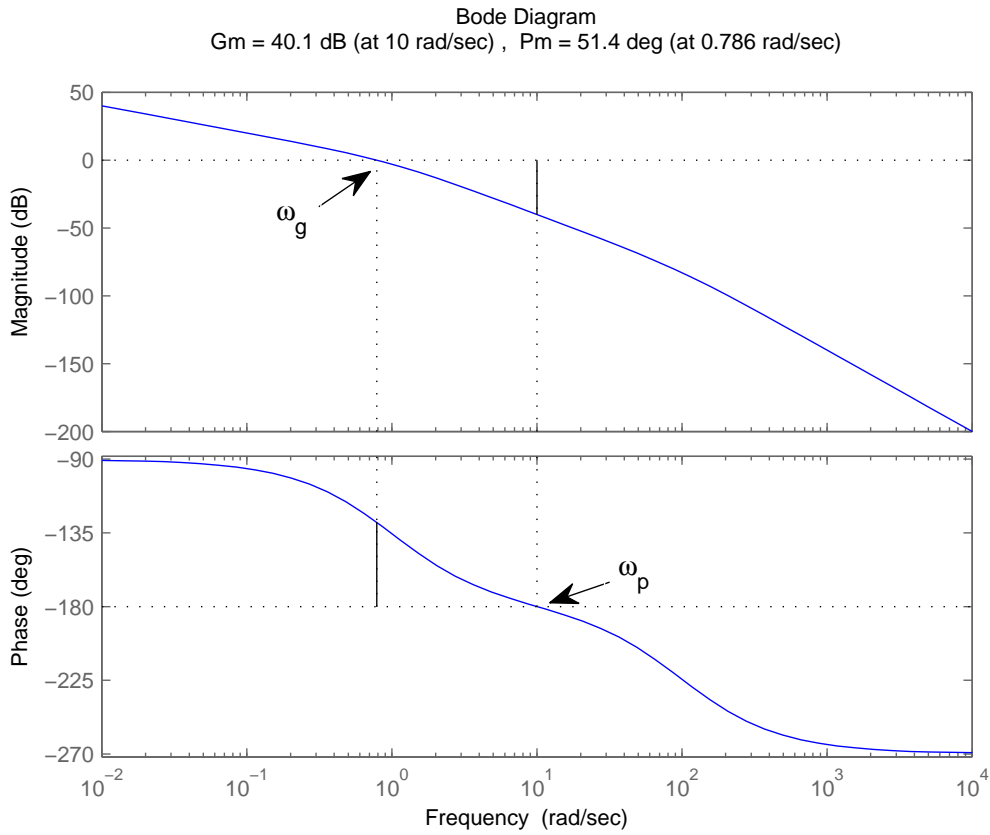The Bode plot is given in Figure 12 for $K = 1$.

Figure 12: Example 1. Bode plot

In this example, we have $\omega_g \approx 1[rad/sec]$ (or $\omega_g^{dec} = 0$ dec). The phase at ω_g is approximately -135°, and so the feedback configuration with K=1 does not have any closed-loop poles on the imaginary axis. We may determine whether there is another value of K for which the closed-loop system will have poles on the imaginary axis (and this cross the boundary between stability to instability). Note that:

$$20 \cdot \log_{10} K \cdot GH(j\omega) = 20 \cdot \log_{10} K + 20 \cdot \log_{10} GH(j\omega)$$

and

$$\angle K \cdot GH(j\omega) = \angle GH(j\omega), \quad \text{for} \quad K > 0$$

16

Thus, K has no effect on the phase and it affects the magnitude plot by shifting it up and down by K^{dB}. For example, when $K = 10$, the entire magnitude gets shifted up by $20 \log_{10} 10 = 20 dB$. Examining the magnitude plot, we see that it has to be shifted up by approximately $40 dB$ in order to set $\omega_g = \omega_p$, which can be accomplished by setting $K = 100$ (or $K^{dB} = 40$). One can easily see from the Bode plot that this is the only positive value of K for which this will happen. Figure 13 shows the changes in the magnitude plot for various values of K.
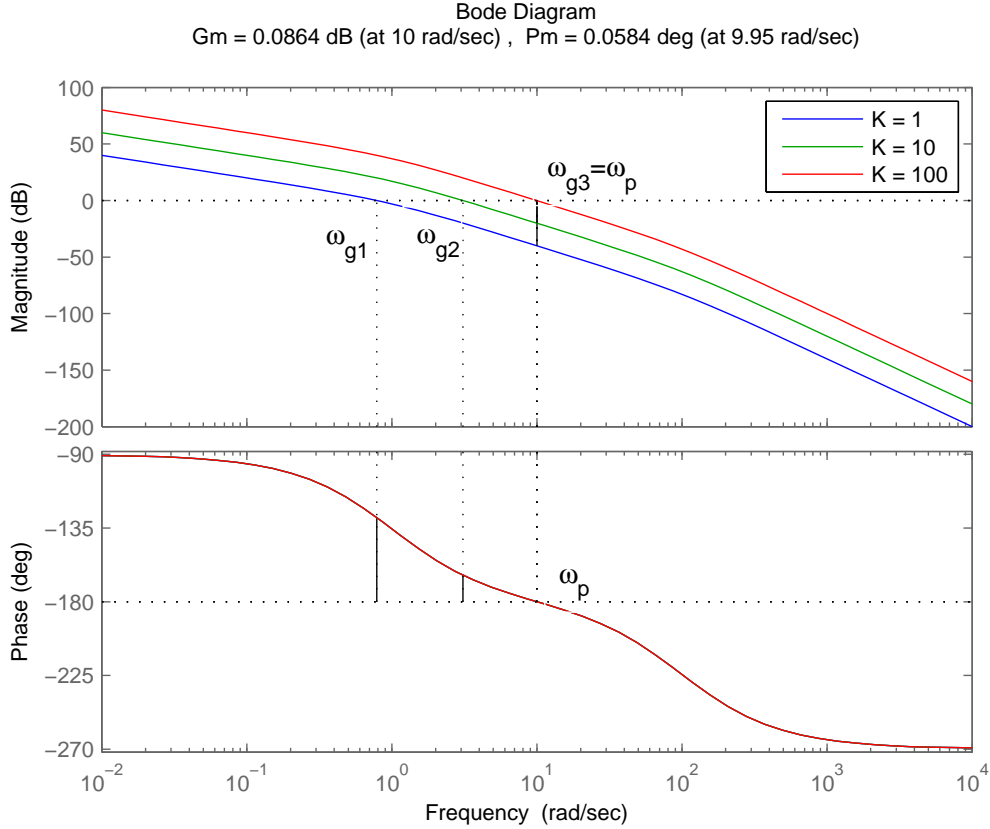

Figure 13: Example 1. Bode plot for various values of K

Relative stability

In designing a control system, we require that the closed-loop system be stable. Furthermore, it is necessary that the system have adequate relative stability. In this section we shall show that the Bode plot indicates not only whether or not a system is stable but also the degree of stability of a stable system.

Phase and Gain Margin

Assuming that the closed-loop system is stable, we can determine how far from instability is the system. There are two metrics to evaluate this:

- **Gain margin**: is the amount by which K can be multiplied before $|KGH(j\omega_p)| = 1$,

or $20log|KGH(j\omega_p)| = M^{dB}(\omega_p) = 0dB$ (i.e. the gain crossover frequency and phase crossover frequencies coincide). In other words, the *gain margin* is the reciprocal of the gain $|GH(j\omega)|$ at the frequency at which the phase angle reaches -180°. The gain margin indicates how much the gain can be increased before the system becomes unstable.

- **Phase margin**: is the amount by which the phase at ω_g exceeds -180°.

More specifically, they are defined by:

- **Gain margin K_g**:

$$K_g = \frac{1}{|GH(j\omega_p)|}, \quad \text{for} \quad \angle GH(j\omega_p) = -180^o$$

or, in log scale:
$$K_g^{dB} = -M^{dB}(\omega_p)$$

- **Phase margin, γ**:

$$\gamma = 180^o + \angle GH(j\omega_g), \quad \text{for} \quad M(\omega_g) = |GH(j\omega_g)| = 1, \quad \text{or} \quad M^{dB}(\omega_g) = 0$$

,

Stability analysis

- A closed-loop system will be stable if the gain margin and phase margin are positive.

- Negative margins indicate instability.

- For satisfactory performance, the phase margin should be between 30° and 60° and gain margin should be greater than $6dB$.

- Either the gain margin or the phase margin alone does not give a sufficient indication of the relative stability. Both should be given in order to determine the relative stability.

- For first order and second order system, gain margin is always infinity.

Example1. Stable closed-loop system

Consider a **closed-loop** control system having the **open-loop** transfer function

$$GH(s) = \frac{1}{s(s+1)^2}$$

The closed-loop system stability and relative stability will be assessed using the Bode plot for the open loop system, shown in Figure 14.

From the plot, the gain margin is evaluated at the frequency ω_p at which the phase crosses -180° line. The magnitude (in dB) is negative, thus the gain margin will be positive: $K^{dB} = -M^{dB}(\omega_p)$.
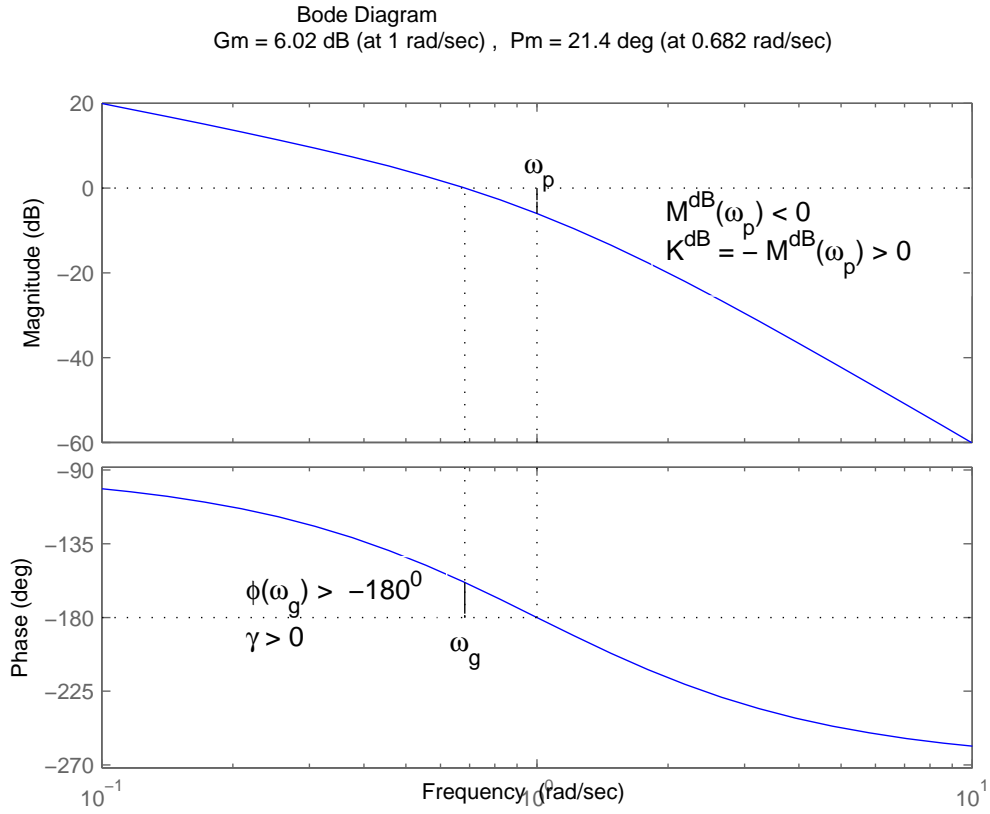
Figure 14: Phase margin and gain margin for a stable closed-loop system

The phase margin is read at the frequency ω_g, when $M^dB(\omega_g) = 0$. It is the difference between -180° and the phase angle of the open-loop system measured at this frequency. Note that in the relation:

$$\gamma = 180^o + \angle GH(j\omega_g),$$

the angle $\angle GH(j\omega_g)$ is negative, thus the phase angle $\gamma > 0$.

For this example, the gain and phase margin are:

$$K^{dB} = 6.02dB, \quad \gamma = 21.4^o$$

and the closed-loop system is stable.

Example2. Unstable closed-loop system

Consider a **closed-loop** control system having the **open-loop** transfer function

$$GH(s) = \frac{100}{s(s+1)^2}$$

The closed-loop system stability and relative stability will be assessed using the Bode plot for the open loop system, shown in Figure 15.
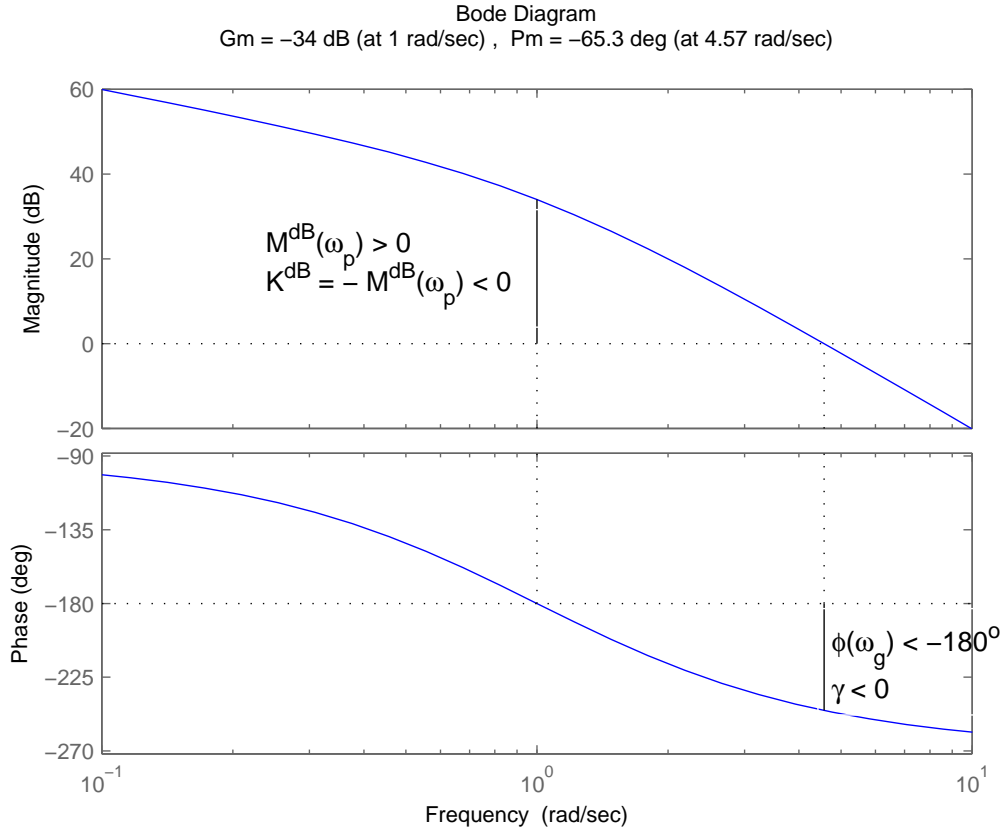
Figure 15: Phase margin and gain margin for an unstable closed-loop system

In this case, the gain margin (expressed in dB) is negative because the magnitude is positive at the phase crossover frequency. The phase margin is negative too, because the phase angle of the open-loop system measured at the gain crossover frequency is less than 180°.

For this example, the gain and phase margin are:

$$K^{dB} = -34 dB dB, \quad \gamma = -65.3^o$$

and the closed-loop system is unstable.

Example3

Consider a **closed-loop** control system having the **open-loop** transfer function

$$GH(s) = \frac{10k(s+1)}{s^2(s+10)}$$

The closed-loop system stability and relative stability will be assessed using the Bode plot for the open loop system, shown in Figure 16, for $k = 1$.

For this example, the gain and phase margin are:

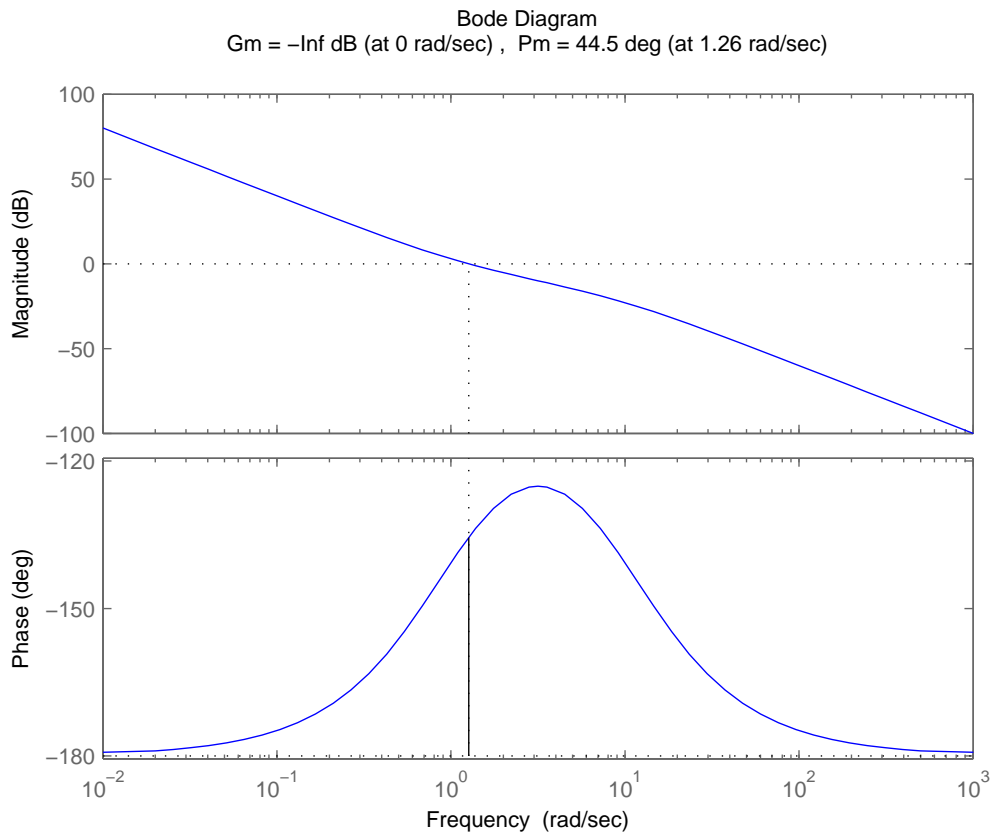$$K^{dB} = Inf dB, \quad \gamma = 44.5^o$$

Figure 16: Infinite gain margin

and the closed-loop system is stable.

The gain margin cannot be evaluated from the Bode plot because the phase angle plot does not cross the -180° line. In this case, the closed-loop system is stable whatever the value of the gain K. This can be verified from a root locus plot or by using the Routh-Hurwitz criterion.

The Nyquist criterion applied to discrete-time systems

The t.f. of the discrete-time (control) system is

$$H(z) = \frac{H_C(z)H_{CP}(z)}{1+H_C(z)H_{CP}(z)} = \frac{H_0(z)}{1+H_0(z)},$$ (5.7.1)

where $H_0(z) = H_C(z)H_{CP}(z)$ is the open-loop t.f.

The stability analysis is carried out, as in the continuous-time case, starting with the discrete open-loop t.f. $H_0(z)$. The Nyquist criterion is applied in a similar way to the continuous-time case, and the theory in the continuous-time case presented in the previous sections is still valid.

A specific problem concerns building the Nyquist plot and the Bode plots. Several versions for computation and building these plots are given in the literature, which lead by their approximations to more or less accurate results [15].

An advantageous version to apply the criterion is based on using the exact frequency response

$$H_0(e^{i\omega T_s}) = H_0(z)\Big|_{z=e^{sT_s},\ s=j\omega},$$ (5.7.2)

for which the plots are drawn considering the values of the frequency ω within the interval $0 \leq \omega \leq \omega_s/2$.

All coefficients of $H_0(z)$ depend on the value of T_s. Changing this value requires to re-consider the system stability.

References

[1] T. L. Dragomir and S. Preitl, *Elemente de teoria sistemelor şi reglaj automat*, lectures (in Romanian), vol. 1 and 2, Litografia I.P.T.V. Timişoara, Romania, 1979.

[2] V. Răsvan, *Teoria stabilităţii* (in Romanian), Editura Ştiinţifică şi Enciclopedică, Bucharest, Romania, 1987.

[3] M. Voicu, *Tehnici de analiză a stabilităţii sistemelor automate* (in Romanian), Editura Tehnică, Bucharest, Romania, 1986.

[4] R. C. Dorf and R. H. Bishop, *Modern Control Systems*, 12th Ed., Prentice Hall, Upper Saddle River, NJ, USA, 2011.

[5] K. Ogata, *Modern Control Engineering*, 5th Ed., Prentice Hall, Upper Saddle River, NJ, USA, 2010.

[6] W. L. Brogan, *Modern Control Theory*, Quantum, New York, USA, 1974.

[7] G. Ionescu, V. Ionescu, T. Ionescu, R. Dobrescu, S. Şerban, C. Soare and R. Vârbănescu, *Automatica de la A la Z* (in Romanian), Editura Ştiinţifică şi Enciclopedică, Bucharest, Romania, 1987.

[8] O. Föllinger, *Regelungstechnik*, Elitera Verlag, Berlin, 1972.

[9] S. Preitl and R.-E. Precup, *Introducere în ingineria reglării automate* (in Romanian), Editura Politehnica, Timişoara, Romania, 2001.

[10] S. Preitl and R.-E. Precup, *Automatizări* (in Romanian), Editura Orizonturi Universitare, Timişoara, Romania, 2001.

[11] R. Isermann, *Digitale Regelungssysteme*, vol. I, II, Springer-Verlag, Berlin, Germany, 1977.

[12] I. D. Landau, *Identificarea şi comanda sistemelor* (in Romaian), Editura Tehnică, Bucharest, Romania, 1997.

[13] S. Preitl, *Teoria sistemelor şi reglaj automat* (in Romanian), vol. 1, part 1, Litografia U.T. Timişoara, Romania, 1992.

[14] *Matlab. User's Guide*, Mathworks Inc., Natick, MA, USA, 1988.

[15] B. C. Kuo, *Digital Control Systems*, 2nd Ed., Oxford University Press, New York, NY, USA, 2007.