AUTOMATIC CONTROL REFRESHER

Stefan Pettersson and Fredrik Bruzelius
Edited 2024: Jonas Fredriksson
Division of Systems and Control
Department of Electrical Engineering

Automatic control is the science of controlled systems and is one of the basic subjects within the engineering sciences. A control system consists of a system to be controlled, called the plant (or process), as well as the system that exercises control over the plant, called the controller. A controller could be either a human or an artificial device, where the latter is the focus of this material and most books in the control literature. This material comprises, in a condensed form, the essential topics treated in a basic course in automatic control at an undergraduate level.

1 Introduction

1.1 Open-loop and closed-loop control systems

In a control system, in order to produce a desired response $r(t)$, the controller supplies the plant with signals $u(t)$, called the control input (or the input to the plant or actuating signal), and the plant responds with a signal $y(t)$, called the output from the plant (or controlled variables). When referring to an isolated system, the terms input and output are used to describe the signals that enter the system and the signals that exit the system, respectively.

A control system in which the control input is applied without knowledge of the plant output is called an open-loop control system. Figure 1 shows a block diagram of an open-loop control system, where the sub-systems (controller and plant) are shown as rectangular blocks, with arrows indicating input and output to each block.

The output from the plant y in an open-loop controlled system will be close to its desired response $r(t)$ only if the controller has a good prior knowledge of the plant's behavior. Because of their simplicity and economy, open-loop control systems may be found in non-critical applications. Since the behavior of many systems cannot exactly be anticipated, and disturbances (or noise) $v(t)$ affect the plant, the output

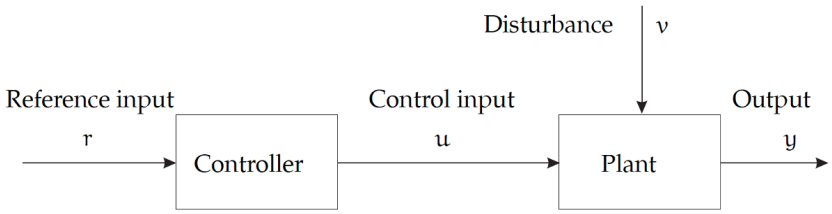

Figure 1: An open-loop control system; the controller applies the control input without knowing the plant output.

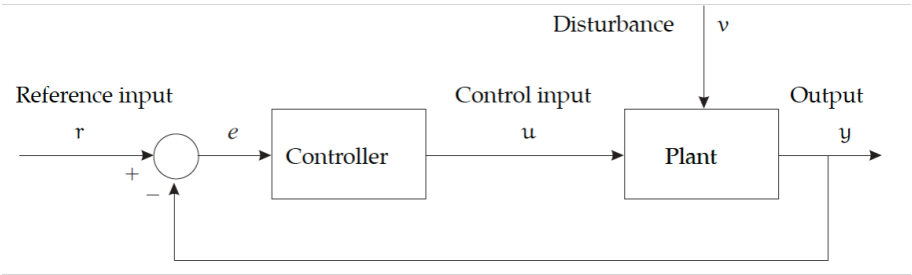

Figure 2: A closed-loop control system with feedback; the controller applies a control input based on the plant output.

from the plant y may be far away from its desired response $r(t)$. However, by measuring the plant output and using this information to control the plant better, a closed-loop control system is obtained, see Figure 2.

The closed-loop control system compensates for disturbances by measuring the output response, feeding that measurement back through a feedback path, and comparing that response to the input at the summing junction (which adds the signals leading into it with the appropriate signs which are indicated adjacent to the respective arrowheads). If there is an error $e(t)$ between these two, the control system drives the plant to make a correction. If there is no error, the control system does not change the control input since the plant's response is already the desired response.

1.2 A closed-loop system with a feedforward path

The closed-loop control system in Figure 2 can be extended with a feedforward controller using a feedforward path from the reference input $r(t)$ to the control input $u(t)$, see Figure 3.

The feedforward controller incorporates some a priori knowledge of the plant's behavior, thereby reducing the burden on the feedback controller in controlling the plant. Note that if the feedback controller is removed from Figure 3, the resulting control system becomes an open-loop control system.

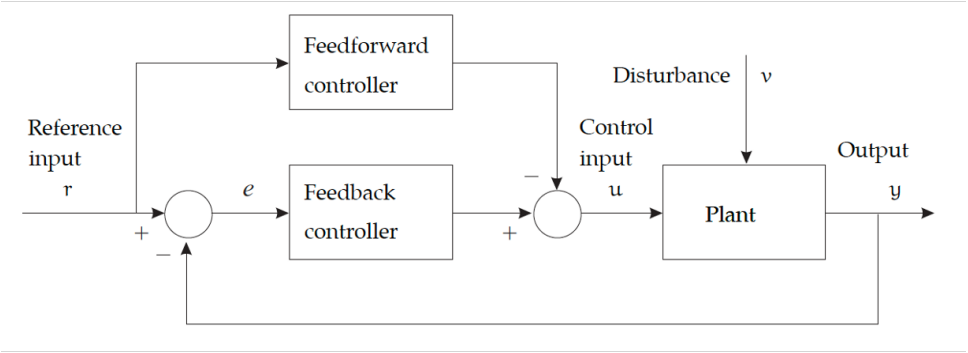

Figure 3: A closed-loop control system with a feedforward path.

1.3 Design of control systems

Control systems are an important part of modern society. Numerous applications are all around us, such as quality control of manufactured products, machine-tool control, space technology, robotics, the process industry, mobile telephones, temperature control, and so on. Specifically, there are many control systems in modern vehicles, like anti-brake and anti-spin control, cruise control, idle control, torque control, etc.

To succeed in the design of a controller, such that the control system behaves in a satisfactory manner, broad knowledge of dynamical systems is required, and this material tries to summarize the essential topics treated in a first basic course in automatic control. The usual steps that are involved in the entire design process of a control system can be summarized as follows:

- **Defining the control problem** - In this first step, the designer has to identify the system that is to be controlled, the control problem, and what behavior is desired of the control system. The behavior is usually denoted by the specification and comprises requirements of the closed-loop control system that have to be fulfilled. The control signals $u(t)$, the output signals $y(t)$, and possibly the disturbances $v(t)$ have to be identified.

- **Modelling** - The better knowledge about the plant under control, the better the chance to control it such that the specifications are fulfilled. Usually, mathematical models are used to describe the dynamics of the plant. Such models can be stated by physical modeling, which means that physical relations between quantities in a dynamical system are used to obtain the model. There are other possibilities to obtain a mathematical description of a plant, e.g., system identification, which means that the mathematical model is derived from experimental data, i.e., input and output signals of the plant.

- **Controller design** - There are numerous choices on how to design controllers satisfying the specification of the closed-loop system. In basic courses in automatic control, the structure of the controller is P-, PI-, and PID-type, possibly in series with low-pass filters.

- **Implementation** - Finally, the designed controller has to be implemented in some hardware. This step reveals if the control system meets the stated specification. If everything works as expected, we have succeeded in the design; otherwise, we have to return to some of the steps above.

The results of the second and third steps above are usually analyzed and validated also by simulations, which means that the tedious calculations to see how the plant and/or control system responds to different input signals are carried out by numerical routines, usually implemented in some programming language on a computer.

The following chapters aim to explore the different steps involved in the analysis and design of control systems. We are only focusing on the last three steps above. The first step is usually application-related and is illustrated in other parts of this chapter.

2 Linear models for dynamical systems

2.1 Differential equations

The dynamical properties of a broad class of physical systems can accurately be described by differential equations. A differential equation defines a relation between system signals using their time derivatives. For a signal $y(t)$ defined over time t, its time-derivatives will be denoted by

$$\frac{dy(t)}{dt} = y^{(1)}(t), \qquad \frac{d^2y(t)}{dt^2} = y^{(2)}(t), \qquad \frac{d^3y(t)}{dt^3} = y^{(3)}(t), ...$$

The first and second derivative is often written using the "dot"-notation

$$\dot{y}(t) = y^{(1)}(t), \qquad \ddot{y}(t) = y^{(2)}(t).$$

Consider the (ordinary time-invariant) linear differential equation describing the relation between the input and output of a physical system

$$a_0 y^{(n)}(t) + ... + a_{n-1}y^{(1)}(t) + a_n y(t) = b_0 u^{(n)}(t) + ... + b_{n-1}u^{(1)}(t) + b_n u(t). \quad (1)$$

If we want to study the output behavior $y(t)$ in the case of different input signals $u(t)$, the calculations will be very tedious (unless we are using some numerical simulation tool) since we have to solve the differential equation. Furthermore, if a plant or control system consists of several subsystems that are connected, finding a solution to the coupled differential equations will be hard. A tool that is very useful in solving differential equations and analyzing a system connected by several subsystems is the Laplace transform. Using the Laplace transform, it is much easier to solve differential equations by algebraic manipulations. Furthermore, much information on the dynamic properties of the system can easily be concluded from the Laplace transform.

2.2 Laplace transform

The Laplace transform is defined as:

$$F(s) = \int_0^\infty f(t)e^{-st}dt \tag{2}$$

where s denotes the Laplace variable (a complex number) and $F(s)$ is called the Laplace transform of $f(t)$. The Laplace transform of a function $f(t)$ is defined only if the infinite integral in (2) exists.

Table 1 gives some important properties of the Laplace transform, and Table 2 shows the Laplace transform of common signals.

Table 1: Laplace transform properties.

Superposition	$\mathcal{L}\{a_1 f_1(t) + a_2 f_2(t)\} = a_1 F_1(s) + a_2 F_2(s)$
Differentiation	$\mathcal{L}\left\{\frac{df(t)}{dt}\right\} = sF(s) - f(0)$
	$\mathcal{L}\left\{f^{(k)}(t)\right\} = s^k F(s) - s^{k-1}f(0) - ... - f^{(k-1)}(0)$
Integration	$\mathcal{L}\left\{\int_0^t f(\tau)d\tau\right\} = \frac{1}{s}F(s)$
Initial value theorem[1]	$\lim_{t\to 0} f(t) = \lim_{s\to\infty} sF(s)$
Final value theorem[2]	$\lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s)$
Time shift theorem	$\mathcal{L}\{f(t-T))\} = e^{-sT}F(s)$
Frequency shift theorem	$\mathcal{L}\{e^{-at}f(t)\} = F(s+a)$
Convolution	$\mathcal{L}\left\{\int_0^t f_1(\tau)f_2(t-\tau)d\tau\right\} = F_1(s)F_2(s)$

[1]) For the initial value theorem to be valid, $f(t)$ must be continuous or have a step discontinuity at $t = 0$ (that is, no impulses or their derivatives at $t = 0$).

[2]) For the final value theorem to yield correct finite results, all roots of the denominator of $sF(s)$ must have negative real parts (that is, not more than one pole at the origin of $F(s)$).

Table 2: Laplace transform table.

$f(t)$ $(f(t) = 0 \quad t < 0)$	$F(s)$
$\delta(t)$	1
$\sigma(t)$	$\frac{1}{s}$
t	$\frac{1}{s^2}$
e^{-at}	$\frac{1}{s+a}$
$\frac{t^{m-1}e^{-at}}{(m-1)!}$	$\frac{1}{(s+a)^m}$
$1 - e^{-at}$	$\frac{a}{s(s+a)}$
$e^{-at} - e^{-bt}$	$\frac{b-a}{(s+a)(s+b)}$
$t - \frac{1}{a}(1 - e^{-at})$	$\frac{a}{s^2(s+a)}$
$1 - (1 + at)e^{-at}$	$\frac{a^2}{s(s+a)^2}$
$e^{-at}\sin\omega t$	$\frac{\omega}{(s+a)^2+\omega^2}$
$e^{-at}\cos\omega t$	$\frac{s+a}{(s+a)^2+\omega^2}$

2.3 Transfer functions

When the system evolves from a state of rest, i.e. zero initial conditions ($u^{(i)}(0^-) = y^{(i)}(0^-) = 0$), the relation between $u(t)$ and $y(t)$ can be readily found using differentiation in Table 1,

$$(a_0 s^n + ... + a_{n-1}s + a_n)Y(s) = (b_0 s^n + ... + b_{n-1}s + b_n)U(s).$$

The transfer function between u and y becomes

$$Y(s) = G(s)U(s) \qquad G(s) = \frac{b_0 s^n + ... + b_{n-1}s + b_n}{a_0 s^n + ... + a_{n-1}s + a_n} = \frac{B(s)}{A(s)}, \tag{3}$$

where A and B are polynomials in the s-variable. The degree of the A-polynomial, n, is referred to as the order of the system.

A transfer function is called strictly proper if the denominator has a higher degree than the numerator and proper if the degree is higher than or equal to the numerator. Technical system models will always be proper (which can be related to physical interpretations of, e.g., causality).

2.4 Solving differential equations using Laplace transforms

Using the Laplace transform, it is easy to study the output behavior $y(t)$ in the case of an input signal $u(t)$ without directly solving the differential equation (1). The steps are as follows:

- Calculate the Laplace transform of $u(t)$, i.e. $U(s) = \mathcal{L}\{u(t)\}$, using Table 2.
- The output signal becomes $Y(s) = G(s)U(s)$, where $G(s)$ is the transfer function (3) of the differential equation (1).
- The output $y(t)$ is transformed back to the time domain $y(t) = \mathcal{L}^{-1}\{Y(s)\}$ using Table 2.

It should be pointed out that there exists a formal definition of the inverse Laplace transform in the last step above, which allows us to find $y(t)$ given $Y(s)$. However, since we have already produced Table 2, the table will be advantageously used in this step as well. This requires, however, that the output response $Y(s)$ is formed in a way such that Table 2, possibly with the help of the properties in Table 1, can be used. This means that complicated functions $Y(s)$ have to be converted to a sum of simpler terms for which we know the (inverse) Laplace transform for each form. This procedure is known as the partial-fraction expansion of a transfer function.

2.5 Output responses

The output $Y(s)$ is determined by a multiplication, $G(s)U(s)$, in the s-domain. This equals a convolution in the time domain, according to Table 1. Hence $y(t)$ is given by the integral

$$y(t) = \int_0^t g(t-\tau)u(\tau)d\tau,$$

where $g(t) = \mathcal{L}^{-1}\{G(s)\}$ is called the weighting function. Since $g(t)$ also is the response to $u(t) = \delta(t)$, it is also denoted the impulse response. The step response becomes, not surprisingly, the integration of the impulse response since $U(s) = 1/s$ gives

$$y(t) = \mathcal{L}^{-1}\left\{\frac{G(s)}{s}\right\} = \int_0^t g(\tau)d\tau.$$

2.6 Poles and zeros

The poles and zeros of a transfer function (3) are defined according to

- $s = p_i : A(p_i) = 0$ is called a *pole*
- $s = z_i : B(z_i) = 0$ is called a *zero*

The transfer function $G(s)$ in (3) has n poles and n zeros. The equation that determines the poles of the transfer function is called the characteristic equation

$$A(s) = 0.$$

2.7 Stability

The poles have a strong influence on the dynamic properties of a linear system (a system described by a linear differential equation (1)). It can be shown that the impulse response of a transfer function (3) becomes

$$g(t) = \mathcal{L}^{-1}\{G(s)\} = g_1(t)e^{p_1 t} + ... + g_m(t)e^{p_m t}$$

where p_i are distinct poles of $G(s)$ with degree l_i and $g_i(t)$ are polynomials in t with degree $l_i - 1$. Since the exponential terms dominate over polynomial terms, we see that one pole in the right half complex plane implies that the impulse response $g(t) \to \pm\infty$ when $t \to \infty$, and all poles in the left half complex plane implies that the impulse response goes to zero when $t \to \infty$. Therefore, a stable linear system is characterized by an impulse response for which

$$g(t) \to 0, \quad when \quad t \to \infty,$$

which implies that the poles must lie in the left half complex plane. Hence, to conclude stability given a transfer function, the location of the poles must (directly or indirectly) be determined. This can be done analytically for low-order systems or by using numerical routines. However, later on, indirect stability conditions will be given with the advantage of being able to conclude stability where the system (or controller) parameters are not exactly known or a part of the stability problem or for feedback control systems.

2.8 Static gain

A fundamental characteristic of a system is the static gain which is defined according to

$$K = \frac{y_\infty}{u_0}$$

where $y_\infty = \lim_{t\to\infty} y(t)$ and $u(t) = u_0$ is constant. This definition is only valid if the system is stable, which means that a constant input signal $u(t) = u_0$ results in a constant output signal $y(t) = y(\infty)$. Since the final value theorem in this case is valid, we have

$$y_\infty = \lim_{t\to\infty} y(t) = \lim_{s\to 0} sY(s) = \lim_{s\to 0} sG(s)U(s) = \lim_{s\to 0} sG(s)\frac{u_0}{s} = G(0)u_0.$$

Hence, the static gain becomes

$$K = \frac{y_\infty}{u_0} = G(0).$$

2.9 Reduction of multiple subsystems

More complicated systems are represented by the interconnection of many subsystems. Since the response of a single transfer function can be calculated, we want to represent multiple subsystems as a single transfer function. We can then apply the analytical methods for one system and obtain transient response information about the entire connected system.

As mentioned in the first section, block diagrams can be used to represent systems with inputs and outputs graphically, and the block in between is the transfer function in the case of linear systems. We will now examine some common structures for interconnecting subsystems and derive the single transfer function representation for each of them. These form the basis for reducing more complicated systems to a single block.

Cascade form Figure 4 shows an example of a cascaded subsystem. The output y_1 of the first subsystem is equal to the input of the second u_2. Each signal is derived from the product of the input times the transfer function. Since

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

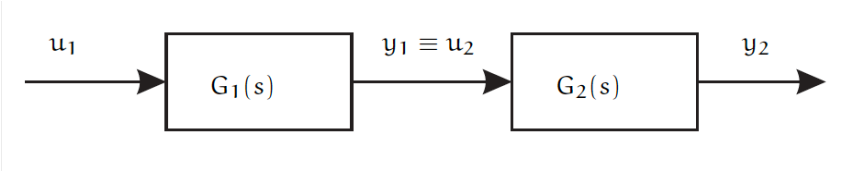

Figure 4: Cascaded subsystems.

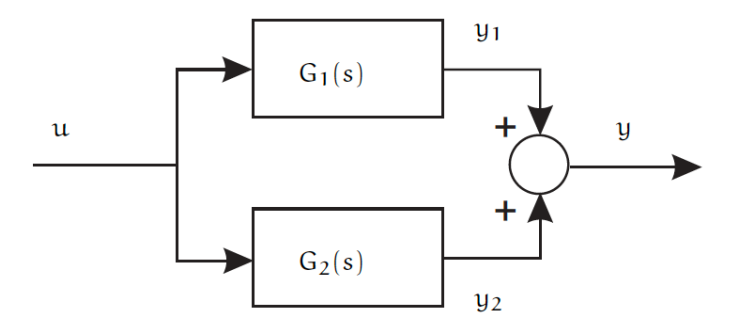

Figure 5: Parallel subsystems.

the equivalent transfer function $G(s)$, relating the input u_1 to the output y_2, becomes

$$G(s) = G_2(s)G_1(s),$$

which is the product of the subsystem's transfer functions.

Parallel form Figure 5 shows an example of a parallel subsystem. Again, by writing the output of each subsystem and summing them up, we have

$$Y(s) = Y_1(s) + Y_2(s) = G_1(s)U(s) + G_2(s)U(s) = (G_1(s) + G_2(s))U(s),$$

implying that the equivalent transfer function becomes

$$G(s) = G_1(s) + G_2(s),$$

which is the sum of the subsystem's transfer functions.

Feedback form The third topology is the feedback form, which is the fundamental basis for our study of control system engineering. The typical feedback configuration is shown in Figure 6, where the feedback summation is done with a negative sign. The transfer function $G(s)$ represents the cascade of the controller and plant subsystems, and $H(s)$ represents the dynamics of the feedback loop (which is equal to 1 in many cases). Since

$$X(s) = R(s) - H(s)Y(s),$$

and

$$Y(s) = G(s)X(s) = G(s)(R(s) - H(s)Y(s)),$$

we obtain, by rearranging the terms,

$$Y(s) = \frac{G(s)}{1 + G(s)H(s)}U(s),$$

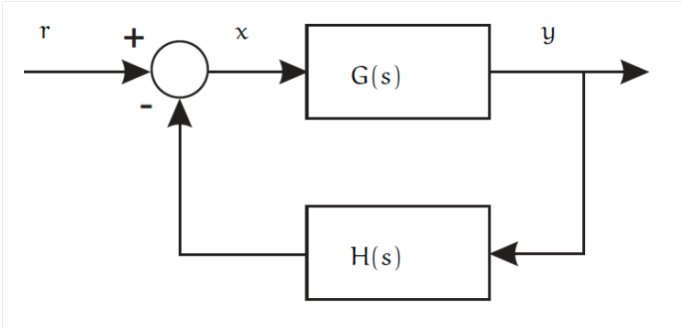

Figure 6: Feedback control system.

and hence, the equivalent closed-loop transfer function becomes

$$G_{ry}(s) = \frac{G(s)}{1 + G(s)H(s)}. \tag{4}$$

The transfer function

$$L(s) = G(s)H(s),$$

is called the open-loop transfer function or loop gain. This is obtained by encircling the feedback path one revolution where the different sub-transfer functions and -1 (which is the minus-sign at the summation junction) are multiplied. The characteristic equation of (4) is

$$1 + G(s)H(s) = 1 + L(s) = 0,$$

which gives the poles of the feedback system. Hence, (4) is stable if all roots to this equation lie strictly in the left half complex plane.

The controller error $e(t)$ in a feedback configuration is defined as

$$e(t) = r(t) - y(t),$$

which is equivalent to $x(t)$ only if $H(s) = 1$. The steady-state error is defined as

$$e_s = \lim_{t\to\infty} e(t) = \lim_{t\to\infty} (r(t) - y(t)),$$

which, in case of a stable feedback configuration and a step input signal $R(s) = r_0/s$, becomes

$$e_s = (1 - G_{ry}(0))r_0 = \frac{r_0}{1 + L(0)}.$$

3 Time domain specifications

Specifications of the closed-loop control system may be expressed in the time domain, often in terms of the step response performance (i.e. the characteristics of the output

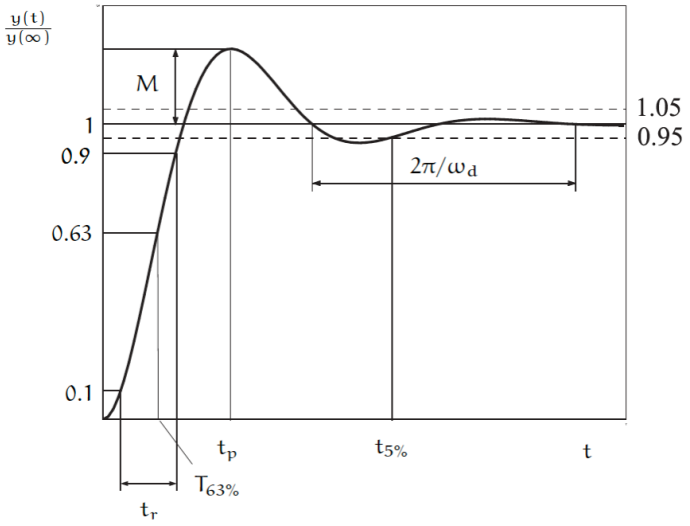

Figure 7: Specifications of a step response.

signal when the input signal is a step). The step response behavior of a physical system can also be used to derive a simple model. The most common performance parameters of a step response (for a stable dynamic system) are given in Figure 7, where:

- *Rise time:* t_r = the time it takes for the output signal to increase from 10% to 90% of its final value.
- *Settling time:* $t_{5\%}$ = the time it takes for the output signal $y(t)$ to remain in the interval $0.95y(\infty) < y(t) < 1.05y(\infty)$.
- *Equivalent time constant:* $T_{63\%}$ = the time it takes for the output signal $y(t)$ to reach 63% of its final value.
- *Relative overshoot:* $M = \frac{\max(y(t))-y(\infty)}{y(\infty)}$.
- t_p = the time when $\max(y(t))$ occurs.
- *Damped natural frequency:* $\omega_d = 2\pi/T_p$, where T_p = time period of the damped oscillation of the step response.

3.1 First-order systems

For first-order systems with delay,

$$G(s) = \frac{K}{1+Ts}e^{-T_d s}$$

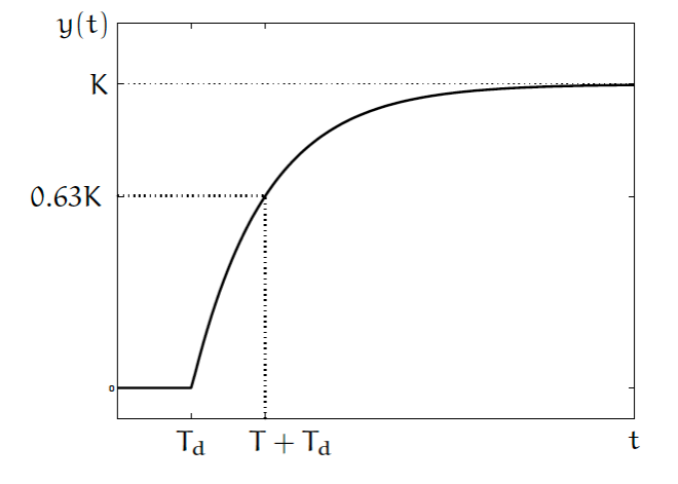

Figure 8: Step response of a first-order system.

the analytic expression for the step response is given by,

$$y(t) = \begin{cases} K(1 - e^{-(t-T_d)/T}) & t \geq T_d \\ 0 & t < T_d \end{cases}$$

which is illustrated in Figure 8. The relevant parameters become

Rise time	$t_r = 2.2T,$
Settling time	$t_{5\%} = 3.0T$
Equivalent time constant	$T_{63\%} = T$

3.2 Second-order systems

For second-order systems with real poles

$$G(s) = \frac{K}{(1 + Ts)(1 + \alpha Ts)},$$

the step response has the form as in Figure 9, and the relevant parameters are given approximately by

Rise time	$t_r \approx (2.2 + \alpha)T,$
Settling time	$t_{5\%} \approx (3.0 + 1.6\alpha)T$
Equivalent time constant	$T_{63\%} \approx (1 + 1.1\alpha)T$

For second-order systems

$$G(s) = \frac{K\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{K}{1 + 2\zeta s/\omega_n + (s/\omega_n)^2}$$

with complex poles in

$$s = -a \pm j\omega_d \quad where \quad \begin{cases} a = \zeta\omega_n \\ \omega_d = \omega_n\sqrt{1 - \zeta^2} \end{cases}$$

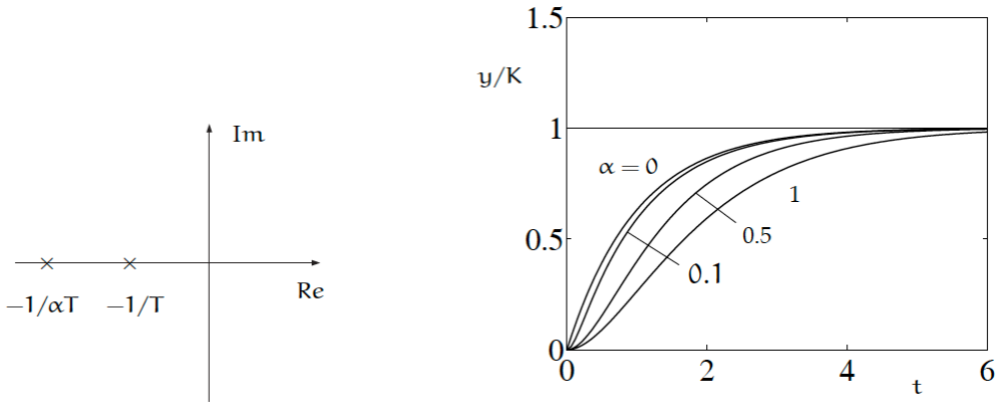

Figure 9: Left: location of the poles on the real-axis in the left complex half plane. Right: corresponding step response.

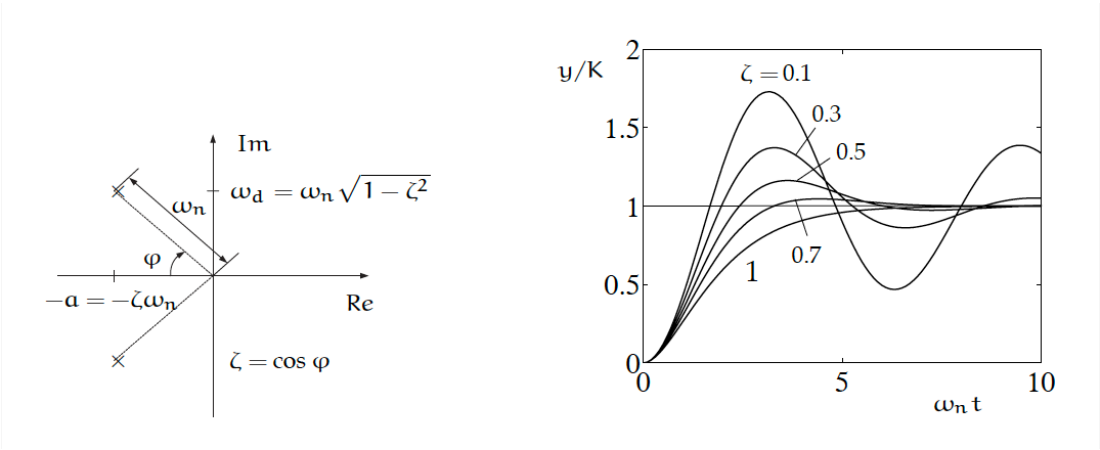

Figure 10: Left: location of the poles in the left complex half plane. Right: corresponding step response.

the step response is given by,

$$y(t) = K(1 - e^{-at}\frac{1}{\sqrt{1-\zeta^2}}\sin(\omega_d t + \varphi)) \quad where \quad \varphi = arccos(\zeta),$$

which is illustrated in Figure 10, and the relevant parameters are given approximately by

Rise time	$t_r \approx (1 + 0.3\zeta + 2\zeta^2)/\omega_n,$
Settling time	$t_{5\%} \approx 3/a \qquad \zeta \leq 0.9$
Relative overshoot	$M = e^{-\pi a/\omega_d} = e^{-\pi\zeta/\sqrt{1-\zeta^2}}$
Damped natural frequency	$\omega_d = \omega_n\sqrt{1-\zeta^2}$

4 Analysis of feedback systems

In this section, the analysis of feedback systems will be discussed. The model is given by a transfer function

$$\frac{Y(s)}{U(s)} = G(s)$$

By introducing a feedback controller, it is possible to design a controller such that the dynamic behavior of the closed-loop control system (the relation between the reference $r(t)$ and the output $y(t)$) fulfills the specification. The most fundamental property of dynamic systems is stability. The stability of a dynamic system represented by a transfer function is determined by the location of the poles of the transfer function. Computing the location of the poles in the complex plane can be done directly or by using the Routh Hurwitz algorithm. The Bode and Nyquist diagrams can be used to determine the stability of feedback systems. In addition, the diagrams can reveal how close the feedback systems are to instability.

4.1 Routh Hurwitz

Stability for a rational transfer function, $G(s) = B(s)/A(s)$, is determined from the location of its poles. The pole locations are given by the roots to the characteristic equation

$$A(s) = a_0 s^n + a_1 s^{n-1} + ... + a_{n-1} s + a_n = 0$$

Note that this applies to systems in feedback configurations

$$\frac{Y(s)}{R(s)} = \frac{G(s)F(s)}{1 + G(s)F(s)},$$

as long as the model $G(s)$ does not contain any delay. In case of a delay in the model, a Padé approximation

$$e^{-sT_d} \approx \frac{2 - sT_d}{2 + sT_d}$$

can be used to transform the model into a rational transfer function of polynomials. The poles of the transfer function are given by the roots of the characteristic equation

$$1 + G(s)F(s) = 0, \quad \Rightarrow \quad P(s) = p_0 s^n + ... + p_{n-1} s + p_n = 0$$

The Routh Hurwitz method gives the number of poles located in the right half of the complex plane, RHP, i.e., the number of unstable poles.

Generate the following table from the coefficients of the characteristic equation

$$\begin{array}{c|cccc} s^n & p_0 & p_2 & p_4 & \dots \\ s^{n-1} & p_1 & p_3 & p_5 & \dots \\ s^{n-2} & c_0 & c_1 & c_2 & \dots \\ s^{n-3} & d_0 & d_1 & d_2 & \dots \\ \vdots & \vdots & & & \\ s^0 & & & & \end{array}$$

where

$$c_0 = (p_1 p_2 - p_3 p_0) p_1^{-1}, \quad c_1 = (p_1 p_4 - p_5 p_0) p_1^{-1}, \quad \cdots$$
$$d_0 = (c_0 p_3 - c_1 p_1) c_0^{-1}, \quad d_1 = (c_0 p_5 - p c_2 p_1) c_0^{-1}, \quad \cdots$$

Stability condition: *all* coefficients in the first column should be positive or negative. The number of sign changes in the first column equals the number of poles in the RHP.

If some element in the first column becomes zero, substitute it with $\epsilon > 0$ in order to be able to complete the table computations. When all calculations in the table are done, let $\epsilon \to 0$. A system with a zero in the first column is either unstable or marginally stable.

4.2 Bode diagrams

Frequency interpretation of linear systems A stable linear system $G(s)$ with input $u(t) = \sin(\omega t)$ has the stationary output

$$y(t) = |G(j\omega)| \sin(\omega t + \arg\{G(j\omega)\}),$$

after the transient behavior.

The Bode diagram is a diagram of the amplitude $|G(j\omega)|$ in a log-log scale and the phase $\arg\{G(j\omega)\}$ versus the frequency ω.

Examples The first-order transfer function,

$$G(s) = \frac{1}{10s + 1},$$

has an amplitude and a phase function

$$|G(j\omega)| = \frac{1}{\sqrt{1 + 10^2\omega^2}}, \qquad \arg\{G(j\omega)\} = \arctan(10\omega).$$

The amplitude in the logarithmic scale becomes,

$$\log|G(j\omega)| = -\frac{1}{2}\log(1 + 10^2\omega^2)$$

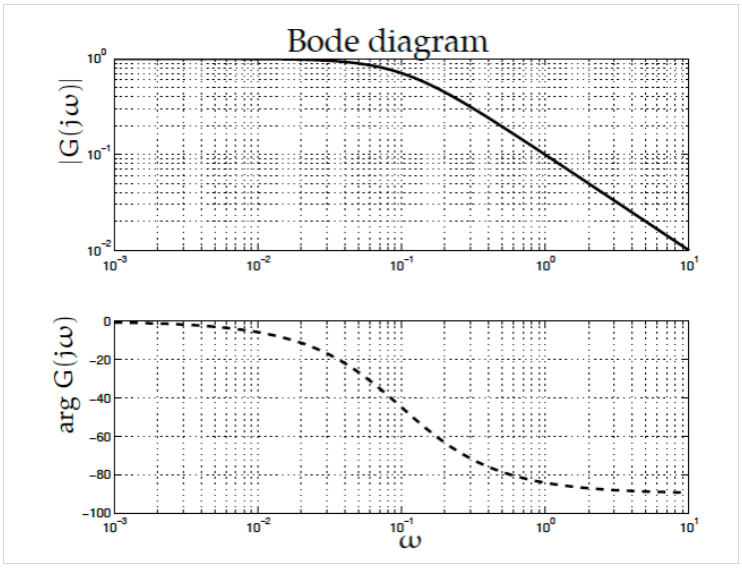

Figure 11: Bode diagram, first-order system.

For large ω, the amplitude becomes "linearly" decaying in $\log(\omega)$, and for small ω, a constant. The breakpoint between constant and "linearly" decaying is $\omega = 1/10$ as one can see in Figure 11.

The bandwidth ω_B of a system is the frequency where the amplitude has decreased a factor $1/\sqrt{2}$ from its low-frequency amplitude. In the first-order example above, the bandwidth is equal to the breakpoint ($\omega_B = 1/10$).

For second-order systems,

$$G(s) = \frac{1}{1 + 2\zeta s/\omega_n + (s/\omega_n)^2}$$

where ω_n is the natural frequency and ζ is the relative damping, the bode diagram is given in Figure 12.

The frequency $\omega = \omega_n\sqrt{1 + 2\zeta^2}$ corresponds to a resonance frequency, i.e., the amplitude has a maximum, for small values of the relative damping ($\zeta < 1/\sqrt{2}$).

Observe that Bode diagrams for higher-order systems can obtained by factorizing the system in first- and second-order systems and by "adding" the factorized systems graphically in the log-log scale.

In feedback systems, the open-loop transfer function $L(s) = G(s)F(s)$, which relates to the closed-loop control system according to

$$\frac{Y(s)}{R(s)} = \frac{L(s)}{1 + L(s)},$$

see Figure 13, which can be used to determine the closed-loop feedback system stability. Intuitively, the closed-loop control system becomes unstable if the frequency

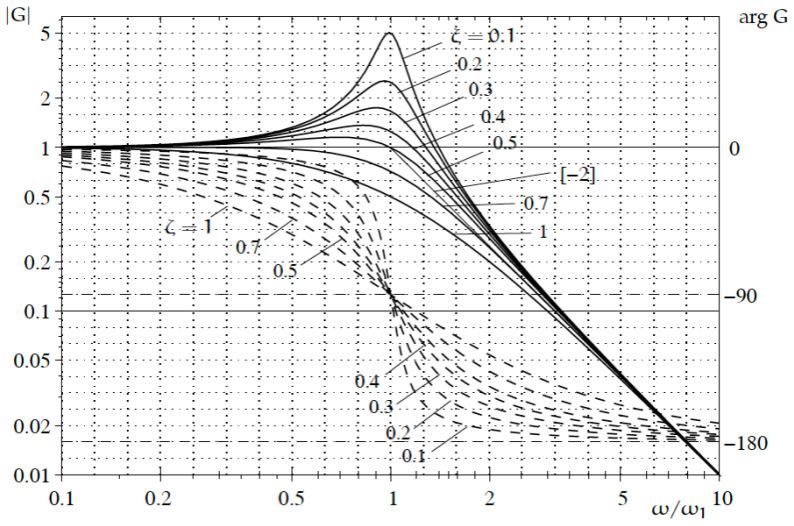

Figure 12: Bode diagram, second-order system.

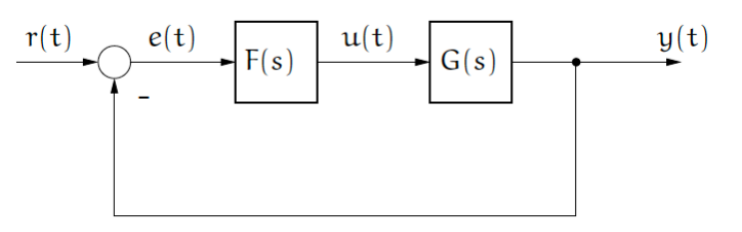

Figure 13: Feedback system.

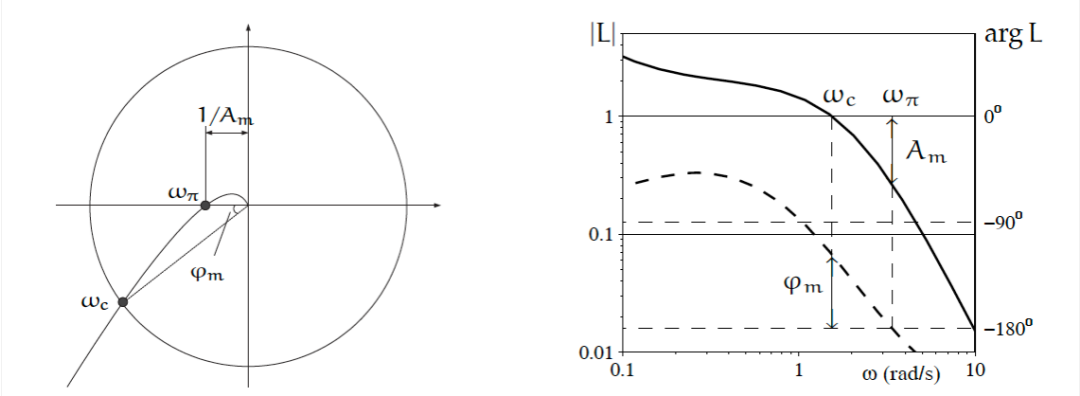

Figure 14: Nyquist (left) and Bode (right) diagrams with the phase and amplitude margins.

ω corresponding to a phase-lag of $-180°$ (i.e., $\arg L(j\omega) = -180°$) is amplified more than one (i.e., $|L(j\omega)| > 1$), due to the negative sign in the feedback loop. The closeness to this critical gain can be expressed in two different measures, the amplitude margin A_m and the phase margin ϕ_m:

$$\phi_m = \arg\{L(j\omega_c)\} + 180°, \quad where \quad |L(j\omega_c)| = 1$$

$$A_m = \frac{1}{|L(j\omega_\pi)|}, \quad where \quad \arg\{L(j\omega_\pi)\} = -180°,$$

which can be determined from the Bode diagram according to Figure 14.

4.3 Nyquist diagrams

A Nyquist diagram is a plot of the complex-valued function $L(j\omega)$ for positive angular frequencies ω. The simplified Nyquist criterion (which applies to stable loop transfers $L(s)$, possibly with time delays) states that the closed loop system is stable if the curve $L(j\omega)$ intersects with the negative real axis to the right of -1 in the complex plane (i.e., -1 is located left to the curve $L(j\omega)$). The amplitude and phase margin can be found in the Nyquist diagram as in Figure 14.

Whenever the loop transfer is unstable, the full Nyquist criterion must be used (the full Nyquist criterion may, of course, always be used). The image of $L(s)$ is plotted when s is described by the contour Γ in Figure 6.5. The contour Γ is expanded (the radius of the left figure's outer half circle tends to infinity) such that the right half complex plane is enclosed. The contour must not pass through singularities of the loop transfer $L(s)$. For example, if $L(s)$ has a pole in the origin, a half circle with a radius that tends to zero must be placed around the origin according to Figure 6.5. The number of poles in the right half plane of the closed loop system Z is given by

$$Z = P + N,$$

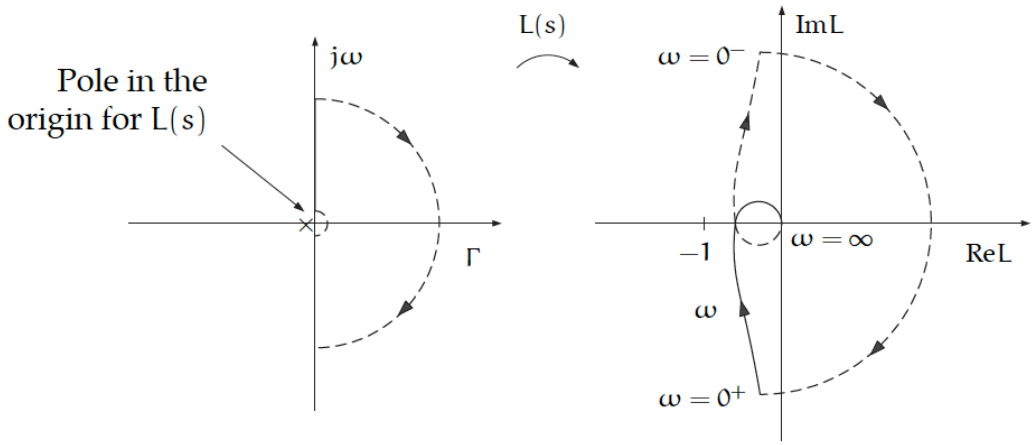

Figure 15: Nyquist contour (left) and its image (right).

where P is the number of poles in the right half plane of the loop transfer $L(s)$ and N is the number of clockwise encirclements around -1 of the image of $L(s)$ in the Nyquist diagram.

5 Controllers and control design

As mentioned earlier, the purpose of feedback control is to keep the controlled variables close to their desired responses. In the basic courses in automatic control, PID and lead- and lag-compensators are usually introduced.

5.1 PID controllers

A proportional–integral–derivative controller (PID controller) is the most widely used controller. A PID controller continuously calculates an error $e(t) = r(t) - y(t)$ as the difference between a desired response $r(t)$ and a measured plant output $y(t)$ and applies a correction based on proportional, integral, and derivative terms of the error

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau)d\tau + K_d \frac{de(t)}{dt},$$

an alternative representation of the PID controller commonly used:

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

or in Laplace:

$$U(s) = (K_p + \frac{K_i}{s} + sK_d)E(s) = F(s)E(s)$$

$$U(s) = K_p(1 + \frac{1}{T_i s} + sT_d)E(s) = F(s)E(s)$$

where K_p, K_i, and K_d denote the coefficients for the proportional, integral, and derivative terms, respectively. By setting some of the coefficients to zero, different controllers can be formed, such as P and PI controllers.

There are many ways to tune a PID controller. It is common to develop some form of plant model and then choose the control parameters based on the dynamic model parameters. Manual tuning methods are also common but can be relatively time-consuming, particularly for systems with long response times. One of the most well-known methods is the Ziegler-Nichols method.

5.1.1 Ziegler–Nichols tuning method

The Ziegler–Nichols tuning method is a heuristic method for tuning a PID controller. It is performed by setting the K_i and K_d to zero. The proportional gain, K_p is then increased (from zero) until it reaches the *ultimate gain* K_u, at which the output of the control loop has stable and consistent oscillations. K_u and the oscillation period T_u are then used to set the K_p, K_i, and K_d gains depending on the type of controller used and behavior desired:

Control type	K_p	K_i	K_d
P	$0.5K_u$	-	-
PI	$0.45K_u$	$0.45K_u/T_u$	-
PID	$0.6K_u$	$1.2K_u/T_u$	$3K_uT_u/40$

5.2 Controller design in frequency domain

The design of a feedback controller $F(s)$ in a configuration as in Figure 13 can be viewed as a shaping of the loop transfer $L(j\omega) = F(j\omega)G(j\omega)$ in the Bode or Nyquist diagram. In order to satisfy specifications on the phase margin ϕ_m of the closed loop system, a compensator (lead compensator) that increases the phase of $L(j\omega)$ for a certain frequency is introduced. In order to meet steady-state error tolerance ($\lim_{t\to\infty} e(t)/\lim_{t\to\infty} r(t)$), a compensator that increases the amplitude of the loop transfer at low frequencies is introduced (lag compensator).

5.2.1 Lead compensator

If a stability margin and performance requirement in the form of a phase margin ϕ_m and crossover frequency ω_c are specified for the closed-loop control system, a lead compensator can be introduced

$$F_{lead}(s) = N\frac{s+b}{s+bN},$$

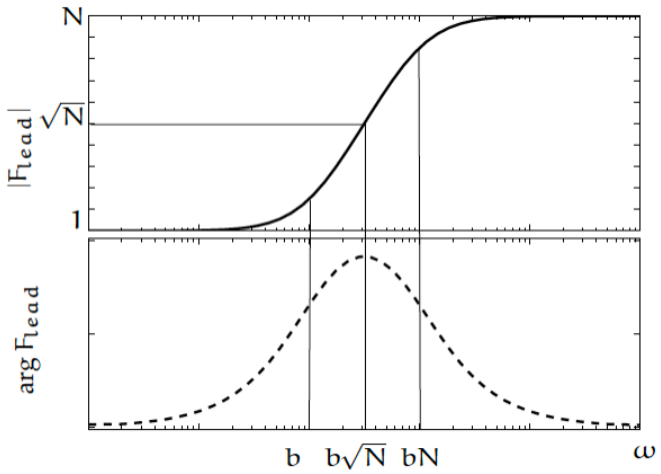

Figure 16: Bode plot of a lead compensator.

such that the phase of $F_{lead}(j\omega)G(j\omega)$ is increased relative to the phase of $G(j\omega)$ for the crossover frequency ω_c. The lead compensator is also known as a PD controller which usually is parameterized as

$$F_{PD}(s) = 1 + \frac{T_d}{T_v s + 1},$$

where T_d is the D-action of the PD controller and T_v is the cut-off time constant.

The lead compensator has a Bode diagram according to Figure 16. The maximum phase of the lead compensator is

$$\max(\arg\{F_{lead}\}) = \arctan\left(\frac{1}{2}(\sqrt{N} - \frac{1}{\sqrt{N}})\right),$$

and occurs at the frequency $\omega = b\sqrt{N}$. The maximum of the phase of the lead compensator can be placed at the crossover frequency ω_c, i.e., $\omega_c = b\sqrt{N}$. Hence, the phase increment that the lead filter must have in order to meet the phase margin specification can be by N. If the needed phase increment is larger than 90◦, additional lead compensators can be introduced. Finally, a gain K_p is determined such that,

$$K_p|F_{lead}(j\omega_c)||G(j\omega_c)| = 1$$

i.e., such that $|L(j\omega_c)| = 1$, and the controller is given by

$$F(s) = K_p F_{lead}(s)$$

.

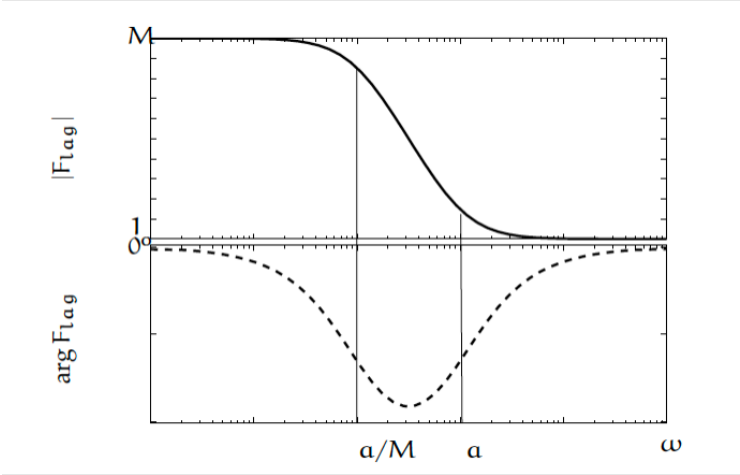

Figure 17: Bode plot of a lag compensator.

5.2.2 Lag compensator

If a requirement of the steady-state error is imposed on the closed-loop control system, a lag compensator can be introduced

$$F_{lag}(s) = \frac{s+a}{s+a/M}.$$

The steady-state error (i.e. when the reference is a unit constant) is given by

$$\lim_{t\to\infty} e(t) = \frac{1}{1+L(0)}$$

In order to decrease this value the amplitude of the open-loop transfer function at low frequency must be increased. The lag compensator has a Bode diagram according to Figure 17. If M is chosen to infinity a conventional PI controller

$$F_{PI}(s) = 1 + \frac{1}{T_i s},$$

is achieved. The controller parameter M is chosen such that the steady-state error requirement e_0 is satisfied, i.e.,

$$e_0 = \frac{1}{1+M|G(0)|}.$$

The other parameter a can be chosen sufficiently small such that high-frequency properties of the loop transfer are unaffected. A rule of thumb is $a = 0.1\omega_c$, which implies that the negative phase contribution due to the lag compensator is of the order of $-6°$ at the crossover frequency ω_c.