



Education and Culture

Leonardo da Vinci

Pilot Project No: CZ/02/B/F/PP/134001

DynLAB

Course on Dynamics of multidisciplinary and
controlled Systems

Part III System Control

Module Version 3.2.2

October 31, 2004

Chr. Schmid

Lehrstuhl für Automatisierungstechnik und Prozessinformatik

Ruhr-Universität Bochum

Preface

This document contains a course both on the basics of control and on extensions to get deeper insight into this area. The text contains many examples to illustrate the subjects. Demonstration examples, problems, interactive questions and demonstration examples with on-line simulations and models visualised in virtual reality are linked to the Internet. A CD version completes the set.

All the ideas, material and the facilities of this course would not have been implemented and completed without the hard work of the following persons, who have directly or indirectly contributed to this course (in alphabetical order):

Abid Ali
Derek Atherton
Carsten Fritsch
Arnd Grosse-Frintrop
Christoph Hackstein
Norman Markgraf
Andrea Marschall
Tom Robert
Heinz Unbehauen

Bochum, October 31, 2004
Christian Schmid

Contents

1	Introduction into System Control	1-1
1.1	Control objectives	1-1
1.2	Open loop vs closed loop	1-2
1.3	The basic structure of closed-loop systems	1-4
2	The Laplace transform	2-1
2.1	Definition	2-1
2.2	Correspondences of the Laplace transform	2-2
2.3	Main theorems of the Laplace transform	2-3
2.4	The inverse Laplace transform	2-5
2.5	Solving linear differential equations using the Laplace transform	2-8
2.6	Laplace transform of the impulse $\delta(t)$ and step $\sigma(t)$	2-12
3	Transfer functions	3-1
3.1	Definition	3-1
3.2	Interpretation of the transfer function	3-2
3.3	Realisability and properness of transfer functions	3-3
3.4	Transfer functions with dead time	3-3
3.5	Poles and zeros of the transfer function	3-3
3.6	Using transfer functions for calculations	3-5
4	Frequency Response	4-1
4.1	Definitions	4-2
4.2	Nyquist plot of a frequency response	4-4
4.3	Bode plot	4-5
4.4	Some important transfer function elements	4-7
4.4.1	The proportional element (P element)	4-7
4.4.2	The integrator (I element)	4-7
4.4.3	The derivative element (D element)	4-8
4.4.4	The 1st-order lag element (PT_1 element)	4-9

4.4.5	The proportional plus derivative element (PD element)	4-12
4.4.6	The derivative lag element (DT_1 element)	4-12
4.4.7	The 2nd-order lag element (PT_2 element and PT_2S element)	4-13
4.4.8	Bandwidth of a system	4-17
4.4.9	Example for the construction of a Bode plot	4-19
4.5	Systems with minimum and non-minimum phase behaviour	4-20
4.6	Systems with dead time	4-22
5	Stability of linear control systems	5-1
5.1	Stable and unstable systems	5-2
5.2	Definition of stability and stability conditions	5-2
5.3	Algebraic stability criteria	5-3
5.3.1	The Hurwitz criterion	5-3
5.3.2	Routh criterion	5-6
5.3.3	Nyquist criterion	5-7
5.3.4	Nyquist criterion using Nyquist plots	5-8
5.3.5	Simplified forms of the Nyquist criterion	5-10
5.3.6	The Nyquist criterion using Bode plots	5-11
6	The root-locus method	6-1
6.1	Introduction and basic ideas	6-1
6.2	General rules for constructing root loci	6-5
6.3	Example of an application of the root-locus method	6-10
7	Behaviour of linear continuous-time control systems	7-1
7.1	Dynamical behaviour of a closed loop system	7-1
7.2	Static properties of the closed loop	7-6
7.2.1	Transfer function $G_0(s)$ with delayed P behaviour	7-7
7.2.2	Transfer function $G_0(s)$ with delayed I behaviour	7-7
7.2.3	Transfer function $G_0(s)$ with delayed I_2 behaviour	7-8
7.3	Performance indices	7-9
7.3.1	Time-response specifications	7-9
7.3.2	Integral performance indices	7-10
7.3.3	Determination of quadratic performance indices	7-12
8	PID control and associated controller types	8-1
8.1	The classical three-term PID controller	8-1
8.2	Optimal tuning of PID controllers	8-3
8.3	Advantages and disadvantages of the different types of controllers	8-4

8.4	Empirical tuning rules according to Ziegler and Nichols	8-6
9	Design of controllers using pole-zero compensators	9-1
9.1	Characteristics in frequency and time domain	9-1
9.2	Controller design using frequency domain characteristics	9-10
9.3	Application of the design using frequency domain characteristics	9-15
9.4	Controller design using the root-locus method	9-18
10	Compensator design methods	10-1
10.1	Basic ideas of compensator design	10-2
10.2	Design by specifying the closed-loop transfer function	10-2
10.3	The method of Truxal and Guillemin	10-3
10.4	Generalised compensator design method	10-7
10.4.1	The basic idea	10-7
10.4.2	Zeros of the closed loop	10-9
10.4.3	The synthesis equations	10-11
10.4.4	Application of the method	10-12
10.5	Compensator design for reference and disturbances	10-16
10.5.1	Structure of the closed loop	10-16
10.5.2	The design procedure	10-17
10.5.3	Design of the pre-filter	10-21
10.5.4	Application of the design method	10-23
11	Improving the control behaviour by more complex loop structures	11-1
11.1	Problem	11-1
11.2	Disturbance feed-forward control	11-2
11.2.1	Disturbance feed-forward on the controller	11-2
11.2.2	Disturbance feed-forward on the manipulated variable	11-3
11.3	Control systems with an auxiliary manipulated variable	11-4
11.4	Cascade control systems	11-5
11.5	Control system with auxiliary manipulated variable	11-7
11.6	Control system with anti-windup measure	11-9
12	State-space representation	12-1
12.1	State-space representation of single-input-single-output systems	12-1
12.2	State-space representation of multi-input-multi-output systems	12-2
12.3	The relationship between transfer functions and the state-space representation	12-3
12.4	State-space vs transfer function approach	12-4
12.5	Uniqueness of the state variables	12-4

12.6	Controllability and observability	12-6
13	Design of state-feedback control systems	13-1
13.1	Structures and properties of state-feedback control systems	13-2
13.1.0.1	State-feedback control in the frequency domain	13-3
13.1.0.2	Steady state of the closed-loop system	13-3
13.2	State-feedback control with integrator	13-4
13.3	Design of state-feedback controllers by pole placement	13-5
13.3.1	Design of a system in controller canonical form	13-6
13.3.1.1	Design of a system not in a canonical form	13-7
13.3.2	Design using Ackermann's formula	13-8
13.4	State reconstruction using observers	13-9
13.4.1	Structure of an observer	13-10
13.4.2	Design of observers	13-10
13.5	Combined observer-controllers	13-11
13.6	Example of a state-feedback control system	13-12
14	Introduction to fuzzy techniques	14-1
14.1	Crisp and fuzzy logic	14-1
14.2	Why use fuzzy logic for control ?	14-3
14.3	Ideas of the fuzzy control methodology	14-4
15	Basics of fuzzy sets	15-1
15.1	Fuzzy sets	15-1
15.2	Membership functions	15-2
15.3	Elementary operators for fuzzy sets	15-5
15.4	Fuzzy relations	15-6
15.5	Fuzzy composition	15-8
16	Fuzzy systems	16-1
16.1	Fuzzification	16-2
16.2	Fuzzy inference machine	16-2
16.3	Defuzzification	16-5
16.3.1	Centre of gravity method (COG)	16-5
16.3.2	Centre of singleton method (COS)	16-5
16.3.3	Maximum methods	16-7
16.3.4	Margin properties of the centroid methods	16-7
16.4	The Takagi-Sugeno fuzzy system	16-7
16.5	The components of a fuzzy system	16-8

17 Fuzzy control	17-1
17.1 Basic structure of a fuzzy controller	17-1
17.2 Transfer behaviour of fuzzy controllers	17-2
17.2.1 Representation using 2D characteristics	17-3
17.2.2 Influence of the membership functions and rule base on the characteristic	17-3
17.2.3 Representation using 3D characteristics	17-7
17.3 Example of a fuzzy control system	17-8
17.3.1 Loading crane plant model	17-9
17.3.2 Fuzzy control system design	17-9
17.4 Contribution of fuzzy control	17-11
A Mathematical and table appendix	A-1
A.1 The Laplace transform	A-2
A.1.1 Convergence	A-2
A.1.2 The inverse Laplace transform	A-2
A.1.3 Main theorems of the Laplace transform	A-3
A.1.3.1 Derivative theorem	A-3
A.1.3.2 Integral theorem	A-3
A.1.3.3 Convolution in the time domain	A-3
A.1.3.4 Convolution in the frequency domain	A-4
A.1.3.5 Initial value theorem	A-5
A.1.3.6 Final value theorem	A-5
A.2 The complex G -plane	A-6
A.3 Detailed analysis of 2nd-order lag elements	A-7
A.3.1 Determining resonances of 2nd-order lag elements	A-7
A.3.2 Poles and step responses of 2nd-order lag elements	A-8
A.4 The law of Bode and the Hilbert transformation	A-10
A.5 Stability considerations using the weighting function	A-11
A.6 Equivalence of the Hurwitz and Routh criteria	A-11
A.7 Determination of J_{ISE} using determinants	A-12
A.8 Tables	A-15
Index	A-22

Module 1

Introduction into System Control

Module units

1.1 Control objectives	1-1
1.2 Open loop vs closed loop	1-2
1.3 The basic structure of closed-loop systems	1-4

Module overview. *This is an introductory module to explain the structures, basic components and terminology of control systems. The difference between open-loop and closed-loop control is explained by a room-heating system. The user can explore the principles of control systems using several virtual experiments with a level control system.*

Module objectives. *When you have completed this module you should be able to:*

1. Understand the operation of a control system.
2. Understand how a control system works.
3. Distinguish between open-loop and closed-loop control.
4. Know the main components of a control system.

Module prerequisites. *Basics in dynamical system models and modelling. Describing dynamical systems by block diagrams.*

1.1 Control objectives

The use of automatic control systems permeates life in all advanced societies today. Such systems act as a catalyst for promoting progress and development. Control systems are an integral component of any industrial society and are necessary for the production of goods. Technological developments have made it possible to travel to the moon and outer space. The successful production of chemical components depends on the proper functioning of a large number of control systems used in lines for their production. As this fact is seldom apparent control engineering is often called a hidden technology.

Control engineering deals with the task of affecting a temporally changing process in such a way that the process behaves in a given way. Such tasks are not only found in technology, but also in daily life in very large number. For example the ambient temperature in a room must be held between given limits, despite temporal changes due to sun exposure and other influences. The grip arm of a robot must move along the edge of a workpiece or be led as fast as possible from one point to another in order to grip a workpiece. The same applies to the grip arm of a crane, which is to carry bricks to a certain place on the building site.

In all of these cases, a manipulated variable must be selected in such a way that the given goal is achieved. As this selection depends on how well the goal is reached, a control loop arises that consists of the given process and a new feature, the controller. In the first example, the room was the process and the thermal valve the automatic controller, which measures the current air temperature and lets more or less heat into the heater depending on the deviation from the target temperature. In the robot example the control equipment has the task of steering the grip arm on a given course and/or to a given point, whereby the control is based on information that is supplied by the sensors installed on the grip arm. In the third example, the automatic controller is the crane operator, who determines the current grip arm position by sight and steers the crane.

1.2 Open loop vs closed loop

The terms open-loop control and closed-loop control are often not clearly distinguished. Therefore, the difference between open-loop control and closed-loop control is demonstrated in the following example of a room heating system. In the case of *open-loop control* of the room temperature ϑ_R according to Figure 1.2.1 the outdoor temperature ϑ_A will be measured by a temperature sensor and fed into a control

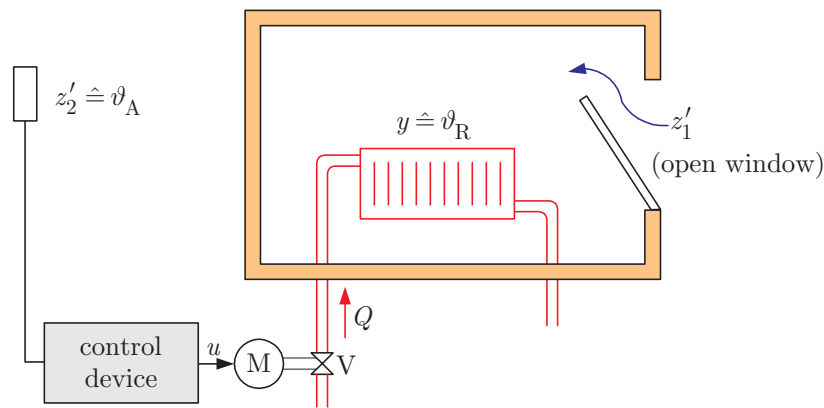


Figure 1.2.1: Open-loop control of a room heating system

device. In the case of changes in the outdoor temperature ϑ_A ($\hat{=}$ disturbance z'_2) the control device adjusts the heating flow Q according to the characteristic $Q = f(\vartheta_A)$ of Figure 1.2.2 using the motor M and the valve V. The slope of this characteristic can be tuned at the control device. If the room temperature ϑ_R is changed by opening a window ($\hat{=}$ disturbance z'_1) this will not influence the position of the valve, because only the outdoor temperature will influence the heating flow. This control principle will not compensate the effects of all disturbances.

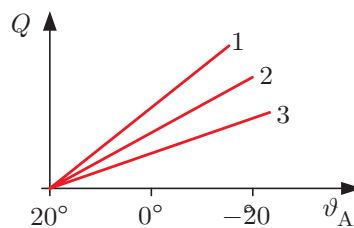


Figure 1.2.2: Characteristic of a heating control device for three different tuning sets (1, 2, 3)

In the case of *closed-loop control* of the room temperature as shown in Figure 1.2.3 the room temperature ϑ_R is measured and compared with the set-point value w , (e.g. $w = 20^\circ\text{C}$). If the room temperature deviates from the given set-point value, a controller (C) alters the heat flow Q . All changes of the room temperature ϑ_R , e.g. caused by opening the window or by solar radiation, are detected by the controller and removed.

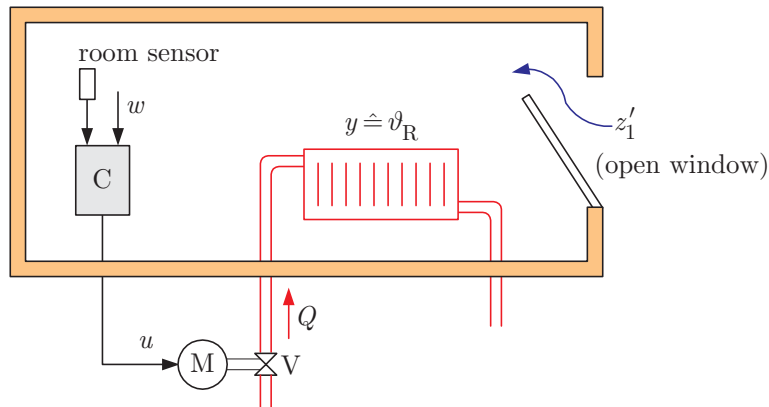


Figure 1.2.3: Closed-loop control of a room heating system

The block diagrams of the open-loop and the closed-loop temperature control systems are shown in Figures 1.2.4 and 1.2.5, and from these the difference between open- and closed-loop control is readily apparent.

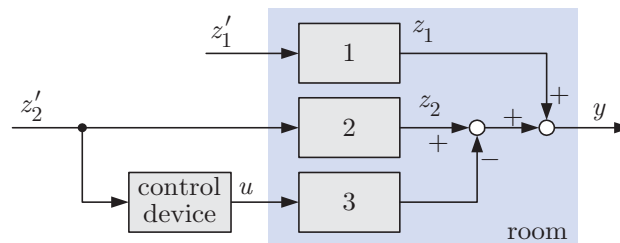


Figure 1.2.4: Block diagram of the open-loop control of the heating system

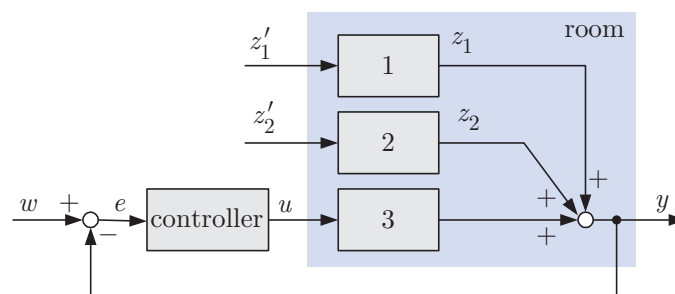


Figure 1.2.5: Block diagram of the closed-loop control of the heating system

The order of events to organise a closed-loop control is characterised by the following steps:

- Measurement of the controlled variable y ,
- Calculation of the control error $e = w - y$ (comparison of the controlled variable y with the set-point value w),
- Processing of the control error such that by changing the manipulated variable u the control error is reduced or removed.

Comparing open-loop control with closed-loop control the following differences are seen:

Closed-loop control

- shows a closed-loop action (closed control loop);
- can counteract against disturbances (negative feedback);
- can become unstable, i.e. the controlled variable does not fade away, but grows (theoretically) to an infinite value.

Open-loop control

- shows an open-loop action (controlled chain);
- can only counteract against disturbances, for which it has been designed; other disturbances cannot be removed;
- cannot become unstable – as long as the controlled object is stable.

Summarising these properties we can define:

Systems in which the output quantity has no effect upon the process input quantity are called open-loop control systems.

Systems in which the output has an effect upon the process input quantity in such a manner as to maintain the desired output value are called closed-loop control systems.

1.3 The basic structure of closed-loop systems

In this section the general structure of control systems having a closed loop will be analysed in more detail. According to Figure 1.3.1 a closed-loop system consists of the following four *main components*:

plant, measurement device, controller and actuator.

The signals in the closed loop will be denoted by symbols. It means:

y	controlled variable (actual value)	u	manipulated variable
w	command variable (set point),	z	disturbance.
e	control error (deviation)		

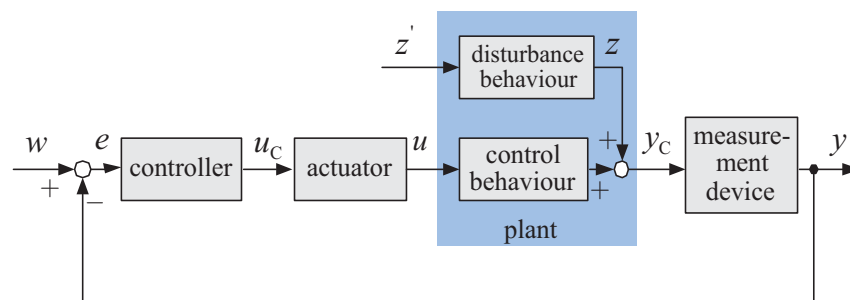


Figure 1.3.1: Basic block diagram of a control system

From this block diagram it can be realised that the task of controlling a process (*plant*) consists of holding the *controlled value* $y(t)$, acquired by the *measurement device*, either on a constant *set point* $w(t) = \text{const}$ (fixed command control) or tracking a time-varying *reference variable* $w(t) \neq \text{const}$ (variable command control), independent of external *disturbances* $z(t)$. This task is performed by a *controller*. The controller processes the control error $e(t) = w(t) - y(t)$, which is the difference between the set point $w(t)$ and the actual value $y(t)$ of the controlled variable. The control signal $u_c(t)$ generated by the controller will act via the *actuator* as the *manipulated variable* $u(t)$ on the plant, such that it counteracts in the case of fixed command control against the disturbance $z(t)$. A closed-loop control system is characterised by

this closed signal path, whereby the controller function consists in cancelling the occurring control error $e(t)$ or at least holding it very small.

Closed-loop control problems can be reduced to this basic structure. In most cases it is not possible to identify all the basic functions clearly. It is therefore proper to aggregate a control loop only into two blocks. Hereby we distinguish between the plant, which may also aggregate the measurement device, and the controlling system that usually contains the actuator, as shown in Figure 1.3.2.

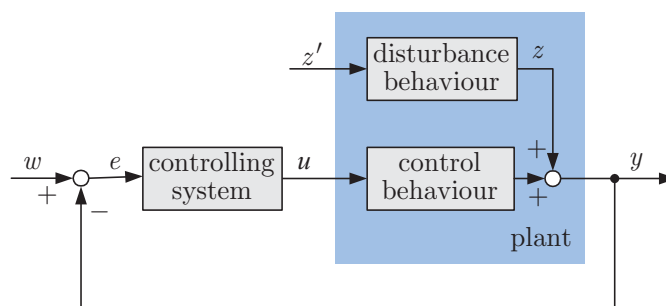


Figure 1.3.2: Simplified block diagram of a closed-loop control system

From Figures 1.3.1 and 1.3.2 it becomes obvious that the comparison of the set-point value w and the actual value y of the controlled variable for generating the control error e will become possible just through the *negative feedback* of the controlled variable y . Only because of this negative sign at the summing point of both signals is the control error generated, which is used by the controller to build the control signal u using special mathematical functions (e.g. proportional, integrating, differentiating). The principle of negative feedback, shortly also called the *feedback principle*, is a characteristic for every control loop.

The principles of closed-loop control are demonstrated by the following examples. Click on the links to start the animation with your Web browser.

Demonstration Example 1.1

[Water tank level without control and with disturbances](#)

Demonstration Example 1.2

[Water tank level without control and with set point](#)

Demonstration Example 1.3

[Water tank level manual control](#)

Demonstration Example 1.4

[Water tank level min/max control](#)

Demonstration Example 1.5

[Water tank level closed-loop control](#)

Demonstration Example 1.6

[Main components of a control system](#)

Module 2

The Laplace transform

Module units

2.1	Definition	2-1
2.2	Correspondences of the Laplace transform	2-2
2.3	Main theorems of the Laplace transform	2-3
2.4	The inverse Laplace transform	2-5
2.5	Solving linear differential equations using the Laplace transform	2-8
2.6	Laplace transform of the impulse $\delta(t)$ and step $\sigma(t)$	2-12

Module overview. *This module is a mathematical section to establish a base for the theory of control systems. This is a tool and it is indispensable as most of linear system dynamics are described in a mapped space that can only be understood when the main theorems of the Laplace transform are known. The module contains only the essential results, which are explained by several examples from the area of differential equations and their solutions. Some additional mathematical details can be found in the mathematical appendix module. The correspondences of the Laplace transform are given in tabular form to be simply used for the forward and back transformation. Special focus is put on the solution of differential equations using the Laplace transform and on special signals, e.g. impulse or step.*

Module objectives. *When you have completed this module you should be able to:*

1. Apply the Laplace transform to differential equations.
2. Solve linear differential equations.
3. Apply the main theorems of the Laplace transform.
4. Know how useful this techniques is to handle dynamical systems.

Module prerequisites. *Mathematics: integrals, differential equations, complex numbers, rational and analytical functions.*

2.1 Definition

The Laplace transform is an important tool for solving systems of linear differential equations with constant coefficients. The differential equations to be solved for control tasks normally fulfil the conditions that must be met for taking the Laplace transform. The Laplace transform is an *integral transformation*, which maps a large class of *original functions* $f(t)$ in the time domain unambiguously reversible into

image functions $F(s)$ in the s domain. This mapping is performed via the *Laplace integral* of $f(t)$, that is

$$F(s) = \int_0^{\infty} f(t) e^{-st} dt, \quad (2.1.1)$$

where in the argument of the *Laplace transform* $F(s)$ the complex variable $s = \sigma + j\omega$ appears. For the application of Eq. (2.1.1) to causal systems considered here the following two conditions for the time function $f(t)$ must be met:

1. $f(t) = 0$ for $t < 0$;
2. the integral in Eq. (2.1.1) must converge.

To show the correspondence between the original and mapped functions it is useful to use the operator notation

$$F(s) = \mathcal{L} [f(t)] .$$

Another possibility of correspondence is to use the sign $\bullet \rightarrow \circ$ in the following way:

$$F(s) \bullet \rightarrow \circ f(t) .$$

During the treatment of control systems usually the original function $f(t)$ is a function of time. As the complex variable s contains the frequency ω , the image function $F(s)$ will often be called a frequency function. Therefore, the Laplace transform allows one to make a transition from the 'time domain' into the 'frequency domain' according to Eq. (2.1.1).

2.2 Correspondences of the Laplace transform

The so called back transformation or inverse Laplace transformation, i.e. the determination of the original function from the mapped function, is described by the *inverse integral* shown in section A.1.2. For this inverse Laplace transform an operator notation in the form

$$f(t) = \mathcal{L}^{-1} [F(s)]$$

can be used.

The Laplace transformation is an *unambiguously reversible* mapping between the original function and the mapped function. $f(t)$ and $F(s)$ are referred to as transform pairs and have a unique correspondence. This is the reason why in most cases one does not need to use the inverse integral. A *Correspondence table*, as shown in Table 2.2.1, will suffice. For the inverse transformation case ones goes from the right column to the left column. In addition some theorems on the Laplace transform given in the next section may be useful.

Table 2.2.1: Corresponding elements of the Laplace transform

Nr.	time response $f(t)$, $f(t) = 0$ for $t < 0$	Laplace transformed $F(s)$
1	δ pulse $\delta(t)$	1
2	unit step $\sigma(t)$	$\frac{1}{s}$
3	t	$\frac{1}{s^2}$
4	t^2	$\frac{2}{s^3}$
5	$\frac{t^n}{n!}$	$\frac{1}{s^{n+1}}$

Table 2.2.1 continued

6	e^{-at}	$\frac{1}{s+a}$
7	te^{-at}	$\frac{1}{(s+a)^2}$
8	t^2e^{-at}	$\frac{2}{(s+a)^3}$
9	$t^n e^{-at}$	$\frac{n!}{(s+a)^{n+1}}$
10	$1 - e^{-at}$	$\frac{a}{s(s+a)}$
11	$\frac{1}{a^2}(e^{-at} - 1 + at)$	$\frac{1}{s^2(s+a)}$
12	$(1 - at)e^{-at}$	$\frac{s}{(s+a)^2}$
13	$\sin \omega_0 t$	$\frac{\omega_0}{s^2 + \omega_0^2}$
14	$\cos \omega_0 t$	$\frac{s}{s^2 + \omega_0^2}$
15	$e^{-at} \sin \omega_0 t$	$\frac{\omega_0}{(s+a)^2 + \omega_0^2}$
16	$e^{-at} \cos \omega_0 t$	$\frac{s+a}{(s+a)^2 + \omega_0^2}$
17	$\frac{1}{a} f\left(\frac{t}{a}\right)$	$F(as) \ (a > 0)$
18	$e^{at} f(t)$	$F(s-a)$
19	$f(t-a)$ for $t > a \geq 0$ 0 for $t < a$	$e^{-as} F(s)$
20	$-t f(t)$	$\frac{dF(s)}{ds}$
21	$(-t)^n f(t)$	$\frac{d^n F(s)}{ds^n}$
22	$f_1(t) f_2(t)$	$\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F_1(p) F_2(s-p) dp$

2.3 Main theorems of the Laplace transform

a) *Superposition theorem:*

For arbitrary constants a_1 and a_2 it follows that

$$\mathcal{L} \{a_1 f_1(t) + a_2 f_2(t)\} = a_1 F_1(s) + a_2 F_2(s) . \quad (2.3.1)$$

The Laplace transformation is a linear integral transformation.

b) *Similarity theorem:*

For an arbitrary constant $a > 0$

$$\mathcal{L} \{f(at)\} = \frac{1}{a} F\left(\frac{s}{a}\right) \quad (2.3.2)$$

is valid. This follows from Eq. (2.1.1) by the substitution of $\tau = at$.

c) *Real Shifting theorem:*

For an arbitrary constant $a > 0$

$$\mathcal{L} \{f(t-a)\} = e^{-as} F(s) \quad (2.3.3)$$

is valid. This follows directly from Eq. (2.1.1) by the substitution of $\tau = t - a$.

d) *Complex Shifting theorem:*

For an arbitrary constant $a > 0$

$$\mathcal{L} \{e^{-at} f(t)\} = F(s+a) \quad (2.3.4)$$

is valid. This follows directly from Eq. (2.1.1).

e) *Derivative theorem:*

For a causal function of time, $f(t)$, for which the derivative for $t > 0$ exists, then as shown in section A.1.3.1, one obtains

$$\mathcal{L} \left\{ \frac{df(t)}{dt} \right\} = sF(s) - f(0+) , \quad (2.3.5)$$

and in the case of multiple differentiation

$$\mathcal{L} \left\{ \frac{d^n f(t)}{dt^n} \right\} = s^n F(s) - \sum_{i=1}^n s^{n-i} \frac{d^{(i-1)} f(t)}{dt^{(i-1)}} \Big|_{t=0+} . \quad (2.3.6)$$

f) *Complex differentiation theorem:*

This theorem shows that a differentiation of the mapped function $F(s)$ corresponds to a multiplication with the time t in the time domain:

$$\mathcal{L} \{t^k f(t)\} = (-1)^k \frac{d^k F(s)}{ds^k} . \quad (2.3.7)$$

g) *Integral theorem:*

The integral of a function is mapped by

$$\mathcal{L} \left\{ \int_0^t f(\tau) d\tau \right\} = \frac{1}{s} F(s) . \quad (2.3.8)$$

as shown in section A.1.3.2.

h) *Convolution in the time domain:*

The convolution of two functions of time $f_1(t)$ and $f_2(t)$, presented by the symbolic notation $f_1(t) * f_2(t)$, is defined as

$$f_1(t) * f_2(t) = \int_0^t f_1(\tau) f_2(t-\tau) d\tau . \quad (2.3.9)$$

In section A.1.3.3 it is shown that the convolution of the two original functions corresponds to the multiplication of the related mapped functions, that is

$$\mathcal{L} \{f_1(t) * f_2(t)\} = F_1(s) F_2(s) . \quad (2.3.10)$$

i) *Convolution in the frequency domain:*

Whereas in h) the convolution of two functions of time was given, a similar result for the convolution of two functions in the frequency domain exists and is given by

$$\mathcal{L} \{f_1(t) f_2(t)\} = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F_1(p) F_2(s-p) dp . \quad (2.3.11)$$

Here $F_1(s) \bullet \bullet f_1(t)$ and $F_2(s) \bullet \bullet f_2(t)$ is valid. Furthermore, p is the complex variable of integration. According to this theorem the Laplace transform of the product of two functions of time is equal to the convolution of $F_1(s)$ and $F_2(s)$ in the mapped domain. This is shown in detail in section A.1.3.4.

j) *Initial and final value theorems:*

The *theorem of the initial condition* allows the direct calculation of the function value $f(0+)$ of a causal function of time $f(t)$ from the Laplace transform $F(s)$. If the Laplace transform of $f(t)$ and $\dot{f}(t)$ exist, then

$$f(0+) = \lim_{t \rightarrow 0+} f(t) = \lim_{s \rightarrow \infty} s F(s) \tag{2.3.12}$$

is valid if the $\lim_{t \rightarrow 0} f(t)$ exists, see section A.1.3.5.

Using the *theorem of the final value* the value of $f(t)$ for $t \rightarrow \infty$ can be determined from $F(s)$, if the Laplace transform of $f(t)$ and $\dot{f}(t)$ exist and the limit $\lim_{t \rightarrow \infty} f(t)$ also exists. Then it follows from section A.1.3.6 that

$$f(\infty) = \lim_{t \rightarrow \infty} f(t) = \lim_{s \rightarrow 0} s F(s) . \tag{2.3.13}$$

One has to observe that

$$\lim_{t \rightarrow \infty} f(t) \quad \text{or} \quad \lim_{t \rightarrow 0} f(t)$$

can be calculated only from the corresponding Laplace transform $\mathcal{L}\{f(t)\}$ by application of the theorems of the initial or final value, if the existence of the related limit in the time domain is a priori assured. The following two examples should explain this:

Example 2.3.1

$$f(t) = e^{\alpha t} (\alpha > 0) \circ \bullet F(s) = \frac{1}{s - \alpha}$$

The limit $\lim_{t \rightarrow \infty} e^{\alpha t}$ does not exist so that the final value theorem may not be applied. ■

Example 2.3.2

$$f(t) = \cos \omega_0 t \circ \bullet F(s) = \frac{s}{s^2 + \omega_0^2}$$

The limit $\lim_{t \rightarrow \infty} \cos \omega_0 t$ does not exist and therefore the final value theorem may not be applied. ■

It can be concluded from the last two examples that the following general statement is valid: If the Laplace transform $F(s)$ has, apart from a single pole at the origin $s = 0$, poles on the imaginary axis or in the right-half s plane, then the initial or final value theorems cannot be applied.

2.4 The inverse Laplace transform

The inverse Laplace transform is described by Eq. (A.1.2). As already mentioned in section 2.2 in many cases a direct evaluation of the complex inverse integral is not necessary, as the most important elementary functions are given in Table 2.2.1. A complicated function, $F(s)$, not given in Table 2.2.1 must be decomposed into a sum of simple functions of s that is

$$F(s) = F_1(s) + F_2(s) + \dots + F_n(s) , \tag{2.4.1}$$

which have a known inverse Laplace transform:

$$\begin{aligned} \mathcal{L}^{-1}\{F(s)\} &= \mathcal{L}^{-1}\{F_1(s)\} + \mathcal{L}^{-1}\{F_2(s)\} + \dots + \mathcal{L}^{-1}\{F_n(s)\} \\ &= f_1(t) + f_2(t) + \dots + f_n(t) = f(t) . \end{aligned} \tag{2.4.2}$$

For many problems in control the function $F(s)$ is a ratio of polynomials in s , known as *rational fraction*, that is

$$F(s) = \frac{n_0 + n_1 s + \dots + n_m s^m}{d_0 + d_1 s + \dots + s^n} = \frac{N(s)}{D(s)} , \tag{2.4.3}$$

where $N(s)$ and $D(s)$ are the numerator and the denominator, respectively.

If $m > n$, then $N(s)$ is divided by $D(s)$, where a polynomial in s and a ratio of polynomials are obtained. The numerator of the fraction $N_1(s)$ has a lower order than n . E.g, if $m = n + 2$, then

$$\frac{N(s)}{D(s)} = k_2 s^2 + k_1 s + k_0 + \frac{N_1(s)}{D(s)}, \quad (2.4.4)$$

whereby degree $N_1(s) < n$ and k_0 , k_1 and k_2 are constants.

A rational fraction $F(s)$ given in Eq. (2.4.3) can be decomposed into more simple functions by application of *partial fraction decomposition*, as shown in Eq. (2.4.1). In order to perform this decomposition the denominator polynomial $D(s)$ must be factorised into the form

$$F(s) = \frac{N(s)}{(s - s_1)(s - s_2) \dots (s - s_n)}. \quad (2.4.5)$$

For a denominator polynomial of n -th order one obtains n roots or zeros $s = s_1, s_2, \dots, s_n$. The zeros of $D(s)$ are also known as the *poles* of $F(s)$, since they define where $F(s)$ is infinite. The partial fraction decomposition for different types of poles is shown in the following.

Case 1: $F(s)$ has only *single poles*.

Here $F(s)$ can be expanded into the form

$$F(s) = \sum_{k=1}^n \frac{c_k}{s - s_k}, \quad (2.4.6)$$

where the *residuals* c_k are real or complex constants. Using the table of correspondences one immediately can obtain the corresponding function of time

$$f(t) = \sum_{k=1}^n c_k e^{s_k t} \quad \text{for } t > 0. \quad (2.4.7)$$

The values c_k can be determined either by comparing the coefficients or by using the theorem of residuals from the theory of functions according to

$$c_k = \frac{N(s_k)}{D'(s_k)} = (s - s_k) \frac{N(s)}{D(s)} \Big|_{s=s_k} \quad (2.4.8)$$

for $k = 1, 2, \dots, n$ with $D'(s_k) = dD/ds|_{s=s_k}$.

Case 2: $F(s)$ has *multiple poles*.

For multiple poles of $F(s)$ each with multiplicity r_k ($k = 1, 2, \dots, l$) the corresponding partial fraction decomposition is

$$F(s) = \sum_{k=1}^l \sum_{\nu=1}^{r_k} \frac{c_{k\nu}}{(s - s_k)^\nu} \quad \text{with } n = \sum_{k=1}^l r_k. \quad (2.4.9)$$

The back transformation of Eq. (2.4.9) into the time domain is

$$f(t) = \sum_{k=1}^l e^{s_k t} \sum_{\nu=1}^{r_k} \frac{c_{k\nu} t^{\nu-1}}{(\nu-1)!} \quad \text{for } t > 0. \quad (2.4.10)$$

The real or complex coefficients $c_{k\nu}$ for $\nu = 1, 2, \dots, r_k$ determined by the theorem of residuals are

$$c_{k\nu} = \frac{1}{(r_k - \nu)!} \left\{ \frac{d^{(r_k - \nu)}}{ds^{(r_k - \nu)!}} [F(s) (s - s_k)^{r_k}] \right\}_{s=s_k}. \quad (2.4.11)$$

This general relation also contains the case of single poles of $F(s)$. The poles may be real or complex.

Case 3: $F(s)$ has also *conjugate complex poles*.

As both, the numerator $N(s)$ and the denominator $D(s)$ of the function $F(s)$ are rational algebraic functions, complex factors always arise as conjugate complex pairs. If $F(s)$ has a conjugate complex pair of poles $s_{1,2} = \sigma_1 \pm j\omega_1$, then for the function $F_{1,2}(s)$ in the partial fraction decomposition of

$$F(s) = \frac{N(s)}{D(s)} = F_{1,2}(s) + F_3(s) + \dots + F_n(s)$$

Eq. (2.4.6) can be applied to give

$$F_{1,2}(s) = \frac{c_1}{s - (\sigma_1 + j\omega_1)} + \frac{c_2}{s - (\sigma_1 - j\omega_1)}, \quad (2.4.12)$$

where the residuals

$$c_{1,2} = \delta_1 \pm j\varepsilon_1$$

are also a conjugate complex pair. Therefore, both fractions of $F_{1,2}(s)$ can be combined, and one obtains

$$F_{1,2}(s) = \frac{\beta_0 + \beta_1 s}{\alpha_0 + \alpha_1 s + s^2} \quad (2.4.13)$$

with the real coefficients

$$\left. \begin{aligned} \alpha_0 &= \sigma_1^2 + \omega_1^2 & ; & \alpha_1 = -2\sigma_1 \\ \beta_0 &= -2(\sigma_2\beta_1 + \omega_1\varepsilon_1) & ; & \beta_1 = 2\delta_1 \end{aligned} \right\} . \quad (2.4.14)$$

The determination of the coefficients β_0 and β_1 is performed again using the theorem of residuals by

$$(\beta_0 + \beta_1 s) \Big|_{s=s_1} = (s - s_1)(s - s_2) \frac{N(s)}{D(s)} \Big|_{s=s_1} . \quad (2.4.15)$$

As s_1 is complex, both sides of this equation are complex. Comparing the real and imaginary parts of both sides one gets two equations for the calculation of β_0 and β_1 . This procedure is demonstrated now using the following example.

Example 2.4.1

Find the inverse Laplace transform $f(t)$ of

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

The partial fraction decomposition of $F(s)$ is

$$F(s) = F_{1,2}(s) + F_3(s) = \frac{\beta_0 + \beta_1 s}{s^2 + 2s + 2} + \frac{c_3}{s + 2} ,$$

where the function $F_{1,2}(s)$ contains the conjugate pair of poles

$$s_{1,2} = -1 \pm j .$$

In addition the third pole of $F(s)$ is

$$s_3 = -2 .$$

For the coefficients β_0 and β_1 it follows from Eq. (2.4.15)

$$\begin{aligned} (\beta_0 + \beta_1 s) \Big|_{s=s_1} &= \frac{1}{s + 2} \Big|_{s=s_1} \\ (\beta_0 - \beta_1) + j\beta_1 &= \frac{1}{-1 + j + 2} = \frac{1}{2} - j\frac{1}{2} . \end{aligned}$$

Comparing the real and imaginary parts on both sides one obtains

$$\beta_0 - \beta_1 = \frac{1}{2} \quad \text{and} \quad \beta_1 = -\frac{1}{2}$$

and from this finally $\beta_0 = 0$.

Using Eq. (2.4.8) the residual is

$$c_3 = (s + 2) \frac{1}{(s^2 + 2s + 2)(s + 2)} \Big|_{s=s_3} = \frac{1}{2}.$$

The partial fraction decomposition of $F(s)$ is thus

$$F(s) = -\frac{1}{2} \left[\frac{s}{s^2 + 2s + 2} \right] + \frac{1}{2} \frac{1}{s + 2},$$

which can be rearranged in the form

$$F(s) = -\frac{1}{2} \left[\frac{s + 1}{(s + 1)^2 + 1} - \frac{1}{(s + 1)^2 + 1} - \frac{1}{s + 2} \right]$$

such that correspondences given in Table 2.2.1 can be directly applied to find the inverse transformation. Using the correspondences 16, 15 and 6 of this table, it follows that

$$f(t) = -\frac{1}{2} [e^{-t} \cos t - e^{-t} \sin t - e^{-2t}] \quad \text{for } t > 0,$$

which can be rearranged as

$$f(t) = \frac{1}{2} e^{-t} [e^{-t} + \sin t - \cos t] \quad \text{for } t > 0.$$

The graphical representation of $f(t)$ is shown in Figure 2.4.1a. Figure 2.4.1b shows the corresponding poles, marked by a \times , for this $F(s)$ in the complex s plane. ■

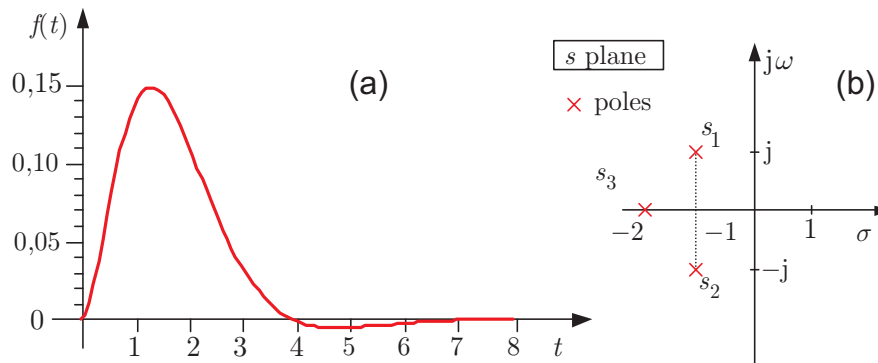


Figure 2.4.1: (a) Graph of the original function $f(t)$ (function in the time domain) and (b) position of the poles of $F(s)$ in the s plane

It can be seen from this example that the position of the poles s_1, s_2 and s_3 affects the shape of the graph of $f(t)$. In this case all poles of $F(s)$ have negative real parts, therefore the graph of $f(t)$ shows a damped behaviour, i.e. it decreases to zero for $t \rightarrow \infty$. If the real part of one pole be positive, then the graph of $f(t)$ would be infinitely large for $t \rightarrow \infty$.

Since in control problems the original function $f(t)$ always represents the time behaviour of a system variable, the behaviour of this system variable $f(t)$ can be judged to a large extent by investigation of the positions of the poles of the corresponding mapped function $F(s)$. This will be further commented on in later sections.

2.5 Solving linear differential equations using the Laplace transform

The Laplace transform, the basics of which have been introduced in the sections above, is an elegant way for fast and schematic solving of linear differential equations with constant coefficients. In the following

the importance of this approach is demonstrated. Instead of solving the differential equation with the initial conditions directly in the original domain, the detour via a mapping into the frequency domain is taken, where only an algebraic equation has to be solved. Thus solving differential equations is performed according to Figure 2.5.1 in the following three steps:

1. Transformation of the differential equation into the mapped space ,
2. Solving the algebraic equation in the mapped space,
3. Back transformation of the solution into the original space.

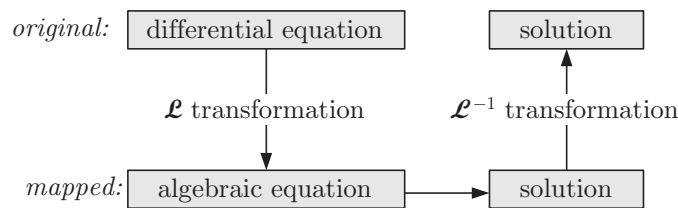


Figure 2.5.1: Schema for solving differential equations using the Laplace transformation

Demonstration Example 2.1
[Here the same in animated form](#)

Whereas the first two steps are trivial, the third step usually demands more effort. The procedure will be demonstrated by the following two examples.

Example 2.5.1

Consider the differential equation

$$\ddot{f}(t) + 3\dot{f}(t) + 2f(t) = e^{-t}$$

with the initial conditions $f(0+) = \dot{f}(0+) = 0$.

Proceeding using the steps given above one has

Step 1:

$$s^2 F(s) + 3sF(s) + 2F(s) = \frac{1}{s + 1}$$

Step 2:

$$F(s) = \frac{1}{s + 1} \frac{1}{s^2 + 3s + 2}$$

Step 3:

The complex function $F(s)$ must be decomposed into partial fractions in order to use the tables of correspondences. This gives

$$F(s) = \frac{1}{s + 2} - \frac{1}{s + 1} + \frac{1}{(s + 1)^2} .$$

By means of the correspondences 6 and 7 of Table 2.2.1 it follows from the inverse Laplace transformation that the solution of the given differential equation is

$$f(t) = e^{-2t} - e^{-t} + t e^{-t} .$$



Example 2.5.2

Given the differential equation

$$\ddot{x} + a_1\dot{x} + a_0x = 0, \quad (2.5.1)$$

where a_0 and a_1 are constants and the initial conditions $\dot{x}(0+)$ and $x(0+)$ are known. Then

Step 1:

$$s^2X(s) - sx(0+) - \dot{x}(0+) + a_1[sX(s) - x(0+)] + a_0X(s) = 0$$

Step 2:

$$\begin{aligned} X(s) &= \frac{s + a_1}{s^2 + a_1s + a_0} x(0+) + \frac{1}{s^2 + a_1s + a_0} \dot{x}(0+), \\ X(s) &= L_0(s)x(0+) + L(s)\dot{x}(0+) \end{aligned} \quad (2.5.2)$$

with the abbreviation

$$L_0(s) = \frac{N_0(s)}{D(s)} = \frac{s + a_1}{s^2 + a_1s + a_0} \quad \text{and} \quad L(s) = \frac{N(s)}{D(s)} = \frac{1}{s^2 + a_1s + a_0}.$$

Step 3:

Case a): two *single real zeros of the denominator*:

This means

$$D(s) = s^2 + a_1s + a_0 = (s - \alpha_1)(s - \alpha_2).$$

For both rational expressions $L_0(s)$ and $L(s)$ it follows by partial fraction decomposition that

$$L_0(s) = \frac{A_1}{s - \alpha_1} + \frac{A_2}{s - \alpha_2} \quad \text{and} \quad L(s) = \frac{B_1}{s - \alpha_1} + \frac{B_2}{s - \alpha_2}.$$

The coefficients A_i and B_i can now be determined by comparing coefficients or by applying Eq. (2.4.8):

$$A_i = \frac{N_0(\alpha_i)}{D'(\alpha_i)}; \quad B_i = \frac{N(\alpha_i)}{D'(\alpha_i)} \quad \text{for} \quad i = 1, 2.$$

Thus for Eq. (2.5.2) follows

$$X(s) = \left[\frac{A_1}{s - \alpha_1} + \frac{A_2}{s - \alpha_2} \right] x(0+) + \left[\frac{B_1}{s - \alpha_1} + \frac{B_2}{s - \alpha_2} \right] \dot{x}(0+),$$

and by applying the correspondence 6 from Table 2.2.1 the solution of the differential equation is

$$\begin{aligned} x(t) &= [A_1e^{\alpha_1 t} + A_2e^{\alpha_2 t}] x(0+) + [B_1e^{\alpha_1 t} + B_2e^{\alpha_2 t}] \dot{x}(0+) \\ &= [A_1x(0+) + B_1\dot{x}(0+)] e^{\alpha_1 t} + [A_2x(0+) + B_2\dot{x}(0+)] e^{\alpha_2 t}. \end{aligned} \quad (2.5.3)$$

Case b): One *double real zero of the denominator*:

Here is

$$D(s) = (s - \alpha)^2.$$

For the two rational expressions $L_0(s)$ and $L(s)$ of Eq. (2.5.2) the partial fraction decomposition is

$$L_0(s) = \frac{A_1}{s - \alpha} + \frac{A_2}{(s - \alpha)^2} \quad \text{and} \quad L(s) = \frac{B_1}{s - \alpha} + \frac{B_2}{(s - \alpha)^2}.$$

The coefficients A_i and B_i are determined by comparing both sides or by evaluation of Eq. (2.4.11):

$$A_1 = \left\{ \frac{d}{ds} \left[\frac{N_0(s)}{D(s)}(s - \alpha)^2 \right] \right\}_{s=\alpha} = 1, \quad A_2 = \left[\frac{N_0(s)}{D(s)}(s - \alpha)^2 \right]_{s=\alpha} = \alpha + a_1$$

and

$$B_1 = \left\{ \frac{d}{ds} \left[\frac{N(s)}{D(s)}(s - \alpha)^2 \right] \right\}_{s=\alpha} = 0, \quad B_2 = \left[\frac{N(s)}{D(s)}(s - \alpha)^2 \right]_{s=\alpha} = 1.$$

From these results one obtains the solution

$$X(s) = \frac{x(0+)}{s - \alpha} + \frac{(\alpha + a_1)x(0+) + \dot{x}(0+)}{(s - \alpha)^2}$$

in the mapped space. By applying the inverse Laplace transformation the required solution of the differential equation is

$$x(t) = x(0+) e^{\alpha t} + [(\alpha + a_1)x(0+) + \dot{x}(0+)] t e^{\alpha t} . \quad (2.5.4)$$

Case c): Two *conjugate complex zeroes of the denominator*:

Here

$$D(s) = (s - \alpha_1)(s - \alpha_2) \quad \text{with} \quad \alpha_{1,2} = \sigma_1 \pm j\omega_1 .$$

Introducing the values of α_1 and α_2 and after multiplication of this expression one obtains

$$D(s) = (s - \sigma_1)^2 + \omega_1^2 .$$

Comparison with the denominator of the original relation, Eq. (2.5.2), gives according to Eq. (2.4.14)

$$a_0 = \sigma_1^2 + \omega_1^2 \quad \text{and} \quad a_1 = -2\sigma_1 .$$

With these coefficients Eq. (2.5.2) is in the form

$$\begin{aligned} X(s) &= \frac{s - 2\sigma_1}{(s - \sigma_1)^2 + \omega_1^2} x(0+) + \frac{1}{(s - \sigma_1)^2 + \omega_1^2} \dot{x}(0+) \\ &= \left[\frac{s - \sigma_1}{(s - \sigma_1)^2 + \omega_1^2} - \frac{\sigma_1}{\omega_1} \frac{\omega_1}{(s - \sigma_1)^2 + \omega_1^2} \right] x(0+) \\ &\quad + \frac{1}{\omega_1} \frac{\omega_1}{(s - \sigma_1)^2 + \omega_1^2} \dot{x}(0+) , \end{aligned}$$

and from this one gets by applying the correspondences 15 and 16 of Table 2.2.1 to $X(s)$ the corresponding time function

$$x(t) = e^{\sigma_1 t} \left[\cos \omega_1 t - \frac{\sigma_1}{\omega_1} \sin \omega_1 t \right] x(0+) + \frac{1}{\omega_1} e^{\sigma_1 t} \dot{x}(0+) \sin \omega_1 t$$

or rearranged

$$x(t) = e^{\sigma_1 t} \left\{ x(0+) \cos \omega_1 t + \left[\frac{1}{\omega_1} \dot{x}(0+) - \frac{\sigma_1}{\omega_1} x(0+) \right] \sin \omega_1 t \right\} . \quad (2.5.5)$$

■

Also from Eq. (2.5.2) of this example the importance of the position of the zeros of $D(s)$, the poles of $X(s)$, on the solution is clear. For all three cases the solution of the differential equation according to Eqs. (2.5.3), (2.5.4) and (2.5.5) is mainly influenced by the position of the poles of $X(s)$. These poles of $X(s)$ are – as one can see from the two examples – only depend on the left side of the corresponding differential equation, i.e. the homogeneous part of it. As is generally known the solution of the homogeneous differential equation describes the *modes* of the system, that is the behaviour, which depends only on the initial conditions. Therefore, consider for the general case only the homogeneous part of an n th-order ordinary homogeneous linear differential equation with constant coefficients that is

$$\sum_{i=0}^n a_i \frac{d^i x_a(t)}{dt^i} = 0 \quad (2.5.6)$$

with all n initial conditions

$$\left(\frac{d^i x_a(t)}{dt^i} \right) \Big|_{t=0+} \quad \text{for} \quad i = 0, 1, \dots, n - 1 .$$

One obtains by Laplace transformation

$$X_a(s) \left[\sum_{i=0}^n a_i s^i \right] - \left[\sum_{i=1}^n a_i \sum_{\nu=1}^i s^{i-\nu} \frac{d^{\nu-1}}{dt^{\nu-1}} x_a(t) \Big|_{t=0+} \right] = 0$$

and a form according to Eq. (2.5.2)

$$X_a(s) = \frac{\sum_{i=1}^n a_i \sum_{\nu=1}^i s^{i-\nu} \frac{d^{\nu-1}}{dt^{\nu-1}} x_a(t) \Big|_{t=0+}}{\sum_{i=0}^n a_i s^i} = \frac{N(s)}{D(s)}, \quad (2.5.7)$$

where $N(s)$ and $D(s)$ are polynomials in s and the initial conditions are only in the numerator polynomial $N(s)$. The poles s_k ($k = 1, 2, \dots, n$) of $X_a(s)$ can be determined directly from the solution of the equation

$$\sum_{i=0}^n a_i s^i = 0. \quad (2.5.8)$$

After factorisation of this equation one obtains

$$a_n (s - s_1) (s - s_2) \dots (s - s_n) = 0. \quad (2.5.9)$$

The poles s_k of $X_a(s)$ make it possible to perform a partial fraction decomposition of $X_a(s)$, e.g. for the case of single poles according to Eq. (2.4.6). For this case one obtains following Eq. (2.4.7) the solution of the homogeneous differential equation, Eq. (2.5.6), in the form

$$x_a(t) = \sum_{k=1}^n c_k e^{s_k t} \quad \text{for } t > 0.$$

From this one can realise that the position of the poles s_k of $X_a(s)$ in the s plane completely characterises the modes or inherent behaviour of the system described by Eq. (2.5.6). Thus one obtains for $\text{Re } s_k < 0$ (left-half s plane) a decreasing and for $\text{Re } s_k > 0$ (right-half s plane) an increasing behaviour of $x_a(t)$, while for pairs of poles with $\text{Re } s_k = 0$ permanent oscillations occur. Therefore, Eq. (2.5.8) or equivalently Eq. (2.5.9), is called the *characteristic equation* and the poles s_k of $X_a(s)$ are often called *eigenvalues* of the equation. Therefore investigation of the characteristic equation provides the most important information about the oscillating behaviour of a system.

2.6 Laplace transform of the impulse $\delta(t)$ and step $\sigma(t)$

The impulse function $\delta(t)$ is *not a function* in the sense of classical analysis, but a distribution (pseudo-function). Therefore without entering the theory of distributions the integral

$$\mathcal{L} \{ \delta(t) \} = \int_0^{\infty} \delta(t) e^{-st} dt \quad (2.6.1)$$

is not defined. The singularity exactly matches with the lower integration limit. The impulse function can be approximately described by the limit

$$\delta(t) = \lim_{\varepsilon \rightarrow 0} r_\varepsilon(t)$$

with the rectangular impulse function

$$r_\varepsilon = \begin{cases} 1/\varepsilon & \text{for } 0 \leq t \leq \varepsilon \\ 0 & \text{otherwise.} \end{cases} \quad (2.6.2)$$

Strictly speaking this representation of $\delta(t)$ is not a distribution, as $r_\varepsilon(t)$ for $0 \leq t \leq \infty$ is not arbitrarily often differentiable. Because of the simple description compared with other functions (e.g. Gaussian functions) this approach is preferred here. From Eq. (2.6.1) it follows that

$$\mathcal{L}\{\delta(t)\} = \int_0^\infty \left[\lim_{\varepsilon \rightarrow 0} r_\varepsilon(t) \right] e^{-st} dt . \quad (2.6.3)$$

As Eq. (2.6.2) can also be represented in the form

$$r_\varepsilon(t) = \frac{1}{\varepsilon} [\sigma(t) - \sigma(t - \varepsilon)] , \quad (2.6.4)$$

where $\sigma(t)$ is the unit step. Since the integration is independent of ε , the limit and integration can be permuted so that

$$\begin{aligned} \mathcal{L}\{\delta(t)\} &= \lim_{\varepsilon \rightarrow 0} \left\{ \frac{1}{\varepsilon} \int_0^\infty [\sigma(t) - \sigma(t - \varepsilon)] e^{-st} dt \right\} \\ \mathcal{L}\{\delta(t)\} &= \lim_{\varepsilon \rightarrow 0} \left\{ \frac{1}{\varepsilon} \frac{1}{s} (1 - e^{-\varepsilon s}) \right\} . \end{aligned}$$

By applying l'Hospital's rule one obtains

$$\mathcal{L}\{\delta(t)\} = \lim_{\varepsilon \rightarrow 0} \frac{s e^{-\varepsilon s}}{s} = 1 . \quad (2.6.5)$$

As the impulse $\delta(t)$ has an area of unity it is also called unit impulse.

Example 2.6.1

Given the differential equation

$$\frac{dy}{dt} = \delta(t) .$$

Find solution $y(t)$.

Remark: The derivative theorem according to Eq. (2.3.5) is – as mentioned in section A.1.3.1 – valid only for classical functions. If, however, a signal consists of a δ function at $t = 0$, then the lower integration limit of Eq. (2.1.1) must be chosen equal to $t = 0-$ and also in Eq. (A.1.11) the left-hand initial condition to $y(0-)$. According to the definition of Eq. (2.1.1) all left-hand initial conditions are always zero.

The solution can be determined in the following three steps:

Step 1:

The Laplace transform of the given differential equation is:

$$sY(s) - y(0-) = 1 \quad \text{with } y(0-) = 0 .$$

Step 2:

The solution of the algebraic equation is:

$$Y(s) = \frac{1}{s} .$$

Step 3:

From the back transformation the solution follows as

$$y(t) = \sigma(t) ,$$

where $\sigma(t)$ is the unit step function. ■

Module 3

Transfer functions

Module units

3.1	Definition	3-1
3.2	Interpretation of the transfer function	3-2
3.3	Realisability and properness of transfer functions	3-3
3.4	Transfer functions with dead time	3-3
3.5	Poles and zeros of the transfer function	3-3
3.6	Using transfer functions for calculations	3-5

Module overview. *Transfer functions simplify working with linear differential equations, which describe the transition behaviour of linear dynamical systems. This module starts with a definition and interpretation of transfer functions. As transfer functions contain at least a rational function of a complex variable, the physical interpretation of poles and zeros is discussed. Finally, the most popular operations of combining systems described by transfer functions are presented. The properties of transfer functions are shown by examples and interactive questions allow the readers to test their knowledge of transfer functions.*

Module objectives. *When you have completed this module you should be able to:*

1. Determine transfer functions from differential equations.
2. Apply transfer functions to describe control systems.
3. Interpret the roles of poles and zeros of transfer functions.

Module prerequisites. *Laplace transform, complex numbers, rational functions.*

3.1 Definition

Linear, continuous-time time-invariant systems with lumped parameters – as initially a dead time is not being taken into account – will be described by the ordinary differential equation

$$\sum_{i=0}^n a_i \frac{d^i x_a(t)}{dt^i} = \sum_{j=0}^m b_j \frac{d^j x_e(t)}{dt^j} . \quad (3.1.1)$$

If all *initial conditions* are set to *zero* and the Laplace transformation is applied to both sides of this equation, one obtains

$$X_a(s) \sum_{i=0}^n a_i s^i = X_e(s) \sum_{j=0}^m b_j s^j ,$$

or reordering

$$\frac{X_a(s)}{X_e(s)} = \frac{b_0 + b_1s + \dots + b_ms^m}{a_0 + a_1s + \dots + a_ns^n} = G(s) = \frac{N(s)}{D(s)}, \quad (3.1.2)$$

where $N(s)$ and $D(s)$ describe the numerator and denominator polynomials, respectively. The quotient of the Laplace-transformed output and input of such a type of system is a rational fraction. The coefficients of this fraction depend only on the structure and parameters of the system. Such a type of function $G(s)$, which describes completely the transfer behaviour of a system, is called the *transfer function* of the system. With such a transfer function the output

$$X_a(s) = G(s) X_e(s) \quad (3.1.3)$$

can be immediately calculated for a known input signal $x_e(t)$, and therefore $X_e(s)$.

3.2 Interpretation of the transfer function

Comparing Eq. (3.1.3) with the convolution theorem from section 2.3, Eqs. (2.3.9) and (2.3.10), it then follows for the representation of Eq. (3.1.3) in the time domain that

$$x_a(t) = \int_0^t g(t - \tau) x_e(\tau) d\tau, \quad (3.2.1)$$

where obviously the inverse Laplace transform of $G(s)$ is the function $g(t)$. This function is generally known as the *weighting function* of the system. In other words, the transfer function is the Laplace-transformed weighting function according to

$$G(s) = \mathcal{L}\{g(t)\}. \quad (3.2.2)$$

If the unit impulse $\delta(t)$ is taken as the input signal $x_a(t)$ for a system described by the transfer function $G(s)$, one obtains according to Eq. (3.1.3)

$$X_a(s) = G(s) \mathcal{L}\{\delta(t)\}$$

and after using the Laplace transform of the unit impulse $\delta(t)$ from Eq. (2.6.5)

$$X_a(s) = G(s)$$

or from Eq. (3.2.2)

$$x_a(t) = g(t).$$

This shows that the response to an unit impulse $\delta(t)$ is the weighting function. Therefore the weighting function is also called the *impulse response*.

Another interpretation is when the system is excited by the input signal

$$x_e(t) = \bar{x}_e e^{\sigma t} \sin(\omega t).$$

For the steady-state case one obtains for the output signal

$$x_{as}(t) = |G(\sigma + j\omega)| \bar{x}_e e^{\sigma t} \sin(\omega t + \phi(\sigma + j\omega)). \quad (3.2.3)$$

This shows that the modulus $|G(\sigma + j\omega)|$ of the transfer function describes the gain, and that $\phi(\sigma + j\omega) = \arg G(\sigma + j\omega)$ describes the phase shift of a sinusoidal function with the frequency ω and with increasing or decreasing amplitude according to $e^{\sigma t}$.

Interactive Questions 3.1

Test your knowledge about impulse and step response

Interactive Questions 3.2

Test your knowledge about convolution

3.3 Realisability and properness of transfer functions

It must be mentioned that a transfer function with $m > n$ is physically not realisable. The transfer function of an ideal differentiator is described by $G(s) = s$ according to Eq. (2.3.5). Any transfer function with $m > n$ can be decomposed into

$$G(s) = \frac{N(s)}{D(s)} = \frac{N_1(s)}{D(s)} + k_0 + k_1s + \dots + k_{m-n}s^{m-n} ,$$

where $\text{degree } N_1(s) = n - 1$ and terms in s with positive powers also occur. Such derivative elements would deliver for input signals of arbitrary high frequency corresponding output signals of arbitrary high amplitude, which are physically not realisable. The *condition of realisability* of the transfer function according to Eq. (3.1.2) is

$$\text{degree } N(s) \leq \text{degree } D(s) \quad \text{or} \quad m \leq n . \quad (3.3.1)$$

This condition is also called as the *condition of properness*. If a transfer function does not follow this condition, it is called an *improper* transfer function. It has the property that $G(s) \rightarrow \infty$ as $s \rightarrow \infty$. A realisable transfer function is called as *proper* and it always follows $G(s) \rightarrow k_0$ as $s \rightarrow \infty$. A transfer function with $k_0 = 0$ is called as *strictly proper*.

3.4 Transfer functions with dead time

If a time delay or *dead time* T_t is introduced in the input signal $x_e(t)$, one obtains instead of Eq. (3.1.1) the differential equation

$$\sum_{i=0}^n a_i \frac{d^i x_a(t)}{dt^i} = \sum_{j=0}^m b_j \frac{d^j x_e(t - T_t)}{dt^j} . \quad (3.4.1)$$

In this case taking the Laplace transformation gives the *transcendental* transfer function

$$G(s) = \frac{N(s)}{D(s)} e^{-sT_t} . \quad (3.4.2)$$

3.5 Poles and zeros of the transfer function

In some cases (e.g. stability analysis) it is expedient to represent the rational transfer function $G(s)$ according to Eq. (3.1.2) in the factorised form

$$G(s) = \frac{N(s)}{D(s)} = k_0 \frac{(s - s_{Z_1})(s - s_{Z_2}) \dots (s - s_{Z_m})}{(s - s_{P_1})(s - s_{P_2}) \dots (s - s_{P_n})} . \quad (3.5.1)$$

For physical reasons only real coefficients a_i, b_j occur. Therefore the *poles* s_{P_i} and the *zeros* s_{Z_j} of $G(s)$, respectively, can be *real* or *complex conjugate* pairs. The terms zeros and poles are chosen, because the transfer function is zero at s_{Z_j} and infinite at s_{P_i} . Zeros and poles can be graphically represented in the complex s plane as shown in Figure 3.5.1. A linear time-invariant system *without* dead time is described completely by the distribution of its poles and zeros and the gain factor k_0 .

Moreover, the poles and zeros of a transfer function have a further significance. Observing a system without input ($x_e(t) \equiv 0$) according to Eq. (3.1.1) and determining the time response $x_a(t)$ for the given n initial conditions, one has to solve the associated homogeneous differential equation

$$\sum_{i=0}^n a_i \frac{d^i x_a(t)}{dt^i} = 0 , \quad (3.5.2)$$

which corresponds exactly to Eq. (2.5.6). For the approach $x_a(t) = e^{st}$ of Eq. (3.5.2) one obtains for the solution in s the *characteristic equation*

$$P(s) = \sum_{i=0}^n a_i s^i = 0 , \quad (3.5.3)$$

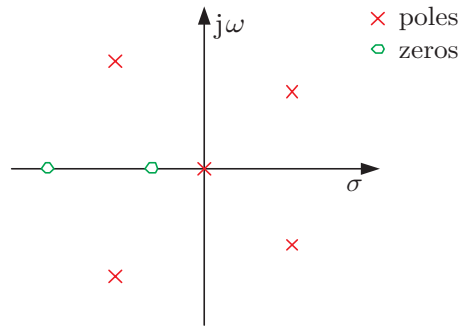


Figure 3.5.1: Example of the pole and zero distribution of a rational transfer function in the complex s plane

which was already mentioned in Eq. (2.5.8). This relation can be directly determined by setting the denominator of $G(s)$ to zero ($D(s) = 0$), as long as $D(s)$ and $N(s)$ have no common factor. The zeros s_k of the characteristic equation are the poles s_{P_i} of the transfer function. As already shown in section 2.5 the modes (i.e. $x_e(t) \equiv 0$) are described by the characteristic equation, so that the poles s_{P_i} of a transfer function contain all of this information.

The zeros of a transfer function are those values $s = s_{Z_j}$ for which $|G(s_{Z_j})| = 0$. This means that the output signal $X_a(s)$ does not contain any components which depend on s_{Z_j} . In order to explain this in more detail a stable system with a transfer function according to Eq. (3.5.1) is excited by the input signal

$$x_e(t) = e^{s_{Z_j} t}.$$

First for simplification the zero $s_{Z_j} = \sigma_{Z_j}$ is assumed to be real. For this case the input signal is $x_e(t) = e^{\sigma_{Z_j} t}$. Because $|G(s_{Z_j})| = 0$ one obtains from Eq. (3.2.3) the steady-state output signal as

$$x_{as}(t) = 0.$$

In the case of complex conjugate pairs of zeros s_{Z_j} , $s_{Z_{j+1}} = s_{Z_j}^*$ both zeros have to be taken into consideration in the input signal

$$\begin{aligned} x_e(t) &= e^{s_{Z_j} t} + e^{s_{Z_j}^* t} \\ &= 2e^{\sigma_{Z_j} t} \cos \omega t \\ &= 2e^{\sigma_{Z_j} t} \sin\left(\omega t + \frac{\pi}{2}\right). \end{aligned}$$

Eq. (3.2.3) leads also to the result $x_{as}(t) = 0$. This shows that a zero s_{Z_j} of a system blocks the transmission of the input signal $e^{\sigma_{Z_j} t}$.

Example 3.5.1

The mass-spring-damper mechanical system in Figure 3.5.2 with the mechanical constants $c_1=1$, $c_2=2$, $d = 1.5$, $m_1=1$ and $m_2=4$ is excited by the force x_e . The transfer function between the force x_e and the position x_a can be shown to be

$$G(s) = \frac{s^2 + 1}{s^4 + 0.5s^3 + 1.75s^2 + 0.5s + 0.5}$$

which has the zeros $s_{Z_{1,2}} = \pm j$. If this system is excited by the sinusoidal input signal

$$\begin{aligned} X_e(s) &= \frac{1}{(s - s_{Z_1})(s - s_{Z_2})} = \frac{1}{s^2 + 1} \\ \mathcal{L}^{-1}\{X_e(s)\} &= \sin t, \end{aligned}$$

which is derived from this pair of zeros, the output signal $x_a(t)$ decays to zero as shown in Figure 3.5.3b even though the input signal is a sinusoidal signal and the mass m_1 shows an undamped oscillation (see Figure 3.5.3c). The system does not pass this oscillation to the mass m_2 when the frequency matches the zeros $s_{Z_{1,2}}$. ■

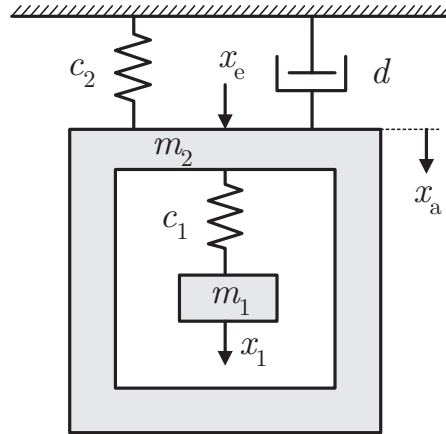


Figure 3.5.2: Mass-spring-damper mechanical system used for the interpretation of zeros

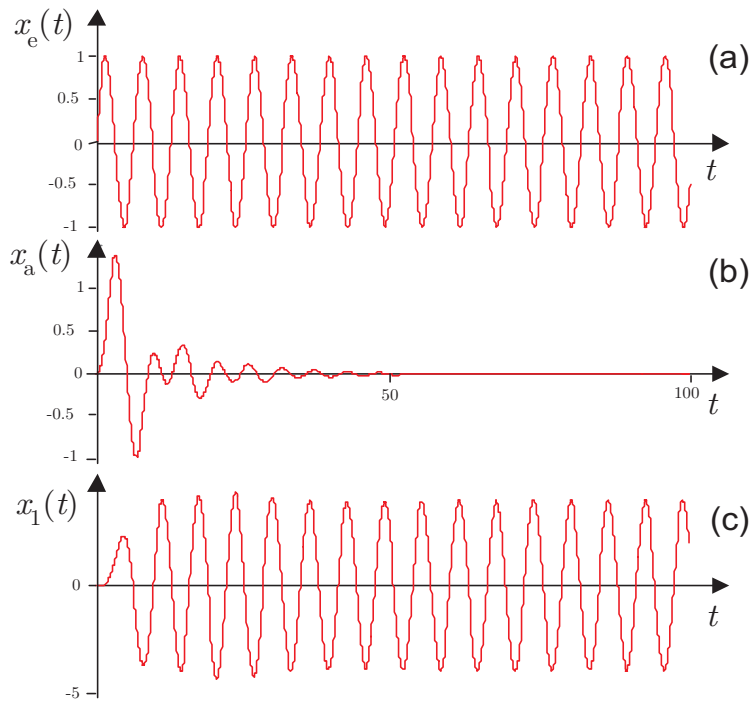


Figure 3.5.3: Response to the sinusoidal input signal $x_a(t) = \sin t$ (a), (b) position of the mass m_2 and (c) position of the mass m_1

3.6 Using transfer functions for calculations

For combinations of transfer functions simple rules for determining the resulting transfer function can be derived. The combinations are of the type that transfer function blocks are connected. In making any transfer function block connection it is assumed that the connection does not load the block to which the connection is being made.

a) *Series connection*

From the diagram in Figure 3.6.1 it follows that

$$\begin{aligned} Y(s) &= G_2(s)X_{e_2}(s) \\ X_{e_2}(s) &= X_{a_1}(s) = G_1(s)U(s) \\ Y(s) &= G_2(s)G_1(s)U(s) . \end{aligned}$$

The total transfer function of this series connection is

$$G(s) = \frac{Y(s)}{U(s)} = G_1(s)G_2(s) . \quad (3.6.1)$$

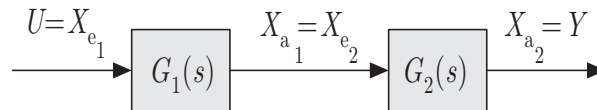


Figure 3.6.1: Series connection of two transfer functions

b) *Parallel connection*

For the output of both functions it follows from Figure 3.6.2 that

$$\begin{aligned} X_{a_1}(s) &= G_1(s)U(s) \\ X_{a_2}(s) &= G_2(s)U(s) . \end{aligned}$$

The output of the total system is

$$Y(s) = X_a(s) = X_{a_1}(s) + X_{a_2}(s) = [G_1(s) + G_2(s)]U(s),$$

and from this the transfer function of a parallel connection is

$$G(s) = \frac{Y(s)}{U(s)} = G_1(s) + G_2(s) . \quad (3.6.2)$$

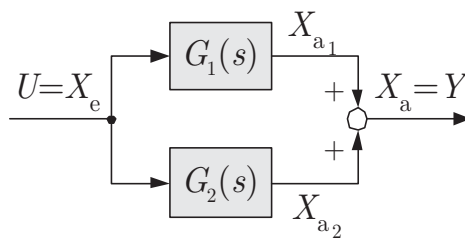


Figure 3.6.2: Parallel connection of two transfer functions

c) *Feedback loop*

From Figure 3.6.3 the output is

$$Y(s) = X_a(s) = [U(s) \underset{(+)}{-} X_{a_2}(s)] G_1(s) .$$

Since

$$X_{a_2}(s) = G_2(s)Y(s)$$

one obtains

$$Y(s) = [U(s) \underset{(+)}{-} G_2(s)Y(s)] G_1(s) ,$$

and from this

$$Y(s) = \frac{G_1(s)}{1 \pm G_1(s)G_2(s)} U(s) .$$

The total transfer function of the feedback loop is

$$G(s) = \frac{Y(s)}{U(s)} = \frac{G_1(s)}{1 \pm G_1(s)G_2(s)} . \quad (3.6.3)$$

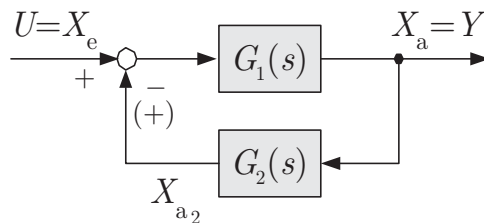


Figure 3.6.3: Feedback using two elements

As the output of $G_1(s)$ is fed back via $G_2(s)$ to the input, this is called *feedback*. One has to distinguish between a positive feedback for positive adding of $X_{a_2}(s)$ and a negative feedback for negative adding of $X_{a_2}(s)$.

Example 3.6.1

For the special case of $G_1(s)$ being a pure amplifier with a high gain $K \rightarrow \infty$, one obtains for negative feedback

$$G(s) = \frac{K}{1 + K G_2(s)} = \frac{1}{\frac{1}{K} + G_2(s)} \approx \frac{1}{G_2(s)} .$$

The entire technique of operational amplifiers is based on this principle. In a feedback loop for which $G_1(s)$ is an amplifier with $K \rightarrow \infty$ then an element $G_2(s)$ can be used to realise any transfer function within certain limits. ■

Interactive Questions 3.3

Test your basic knowledge about transfer functions

Interactive Questions 3.4

Test your knowledge about transfer functions and stability

Interactive Questions 3.5

Test your knowledge about transfer functions and electrical circuits

Interactive Questions 3.6

Test your knowledge about transfer functions and electrical elements

Interactive Questions 3.7

Test your knowledge about simple combinations of transfer functions

Interactive Questions 3.8

Test your knowledge about complicated combinations of transfer functions

Interactive Questions 3.9

Test your knowledge about transfer functions with dead time

Module 4

Frequency Response

Module units

4.1	Definitions	4-2
4.2	Nyquist plot of a frequency response	4-4
4.3	Bode plot	4-5
4.4	Some important transfer function elements	4-7
4.4.1	The proportional element (P element)	4-7
4.4.2	The integrator (I element)	4-7
4.4.3	The derivative element (D element)	4-8
4.4.4	The 1st-order lag element (PT ₁ element)	4-9
4.4.5	The proportional plus derivative element (PD element)	4-12
4.4.6	The derivative lag element (DT ₁ element)	4-12
4.4.7	The 2nd-order lag element (PT ₂ element and PT ₂ S element)	4-13
4.4.8	Bandwidth of a system	4-17
4.4.9	Example for the construction of a Bode plot	4-19
4.5	Systems with minimum and non-minimum phase behaviour	4-20
4.6	Systems with dead time	4-22

Module overview. *This comprehensive module deals with classical frequency-response techniques. Definitions are given starting from the transfer function and its graphical representation in the frequency domain and from the input-output experiment using sinusoidal signals. As this module is the basis for other modules on control system analysis and design using frequency-response techniques, most emphasis is put on the graphical representation techniques of the Nyquist and Bode plots and their application to standard elements to get a feel about the frequency-response characteristics of these elements. Rules for constructing and analysing frequency-response characteristics are given. Finally, the behaviour of non-minimum phase systems is shown. The reader can test his knowledge about frequency response method by some interactive questions, examples and problems.*

Module objectives. *When you have completed this module you should be able to:*

1. Understand the behaviour of linear dynamical systems in the frequency domain.
2. Apply graphical methods (Bode and Nyquist diagram) to describe dynamical systems.
3. Describe and design systems using frequency-response characteristics of standard elements.
4. Understand the behaviour of non-minimum phase systems.

Module prerequisites. *Transfer function.*

4.1 Definitions

In order to represent a transfer function $G(s)$ in graphical form, there are different possibilities. Representing $|G(s)|$ and $\arg G(s)$ over the complex plane $s = \sigma + j\omega$ two three-dimensional diagrams are needed. Figure 4.1.1 shows an example for the magnitude of the transfer function

$$G(s) = \frac{s - 1}{s^2 + s + 1.25} = \frac{s - 1}{(s + 0.5 - j)(s + 0.5 + j)} .$$

The two yellow peaks are at the pole positions where $|G(-0.5 \pm j)| \rightarrow \infty$ and the blue negative peak is at

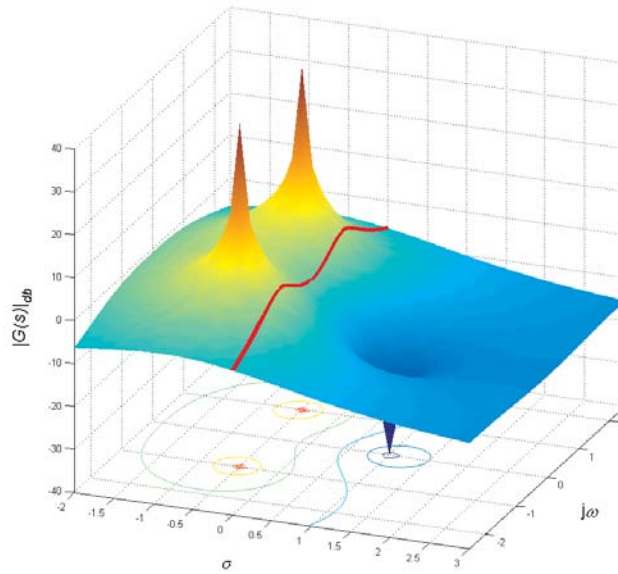


Figure 4.1.1: Magnitude in dB of a transfer function over the complex plane $s = \sigma + j\omega$

the position of the zero $s = 1$ where $|G(1)| = 0$ or $|G(1)|_{\text{dB}} \rightarrow -\infty$, respectively. This three-dimensional representation clearly shows the influence of the poles and zeros on the magnitude of the transfer function. Such diagrams are difficult to handle and are fortunately not necessary for the analysis and design of control systems, because many properties can be investigated by using the *frequency response* $G(j\omega)$. The frequency response is just the cut through the three-dimensional diagrams of $G(s)$ at the imaginary axis ($\sigma = 0$). For the magnitude diagram in Figure 4.1.1 the intersection of this cut and the surface is at the red curve, which represents the magnitude of the frequency response. This curve is symmetric with respect to $\pm\omega$ and therefore the frequency response $G(j\omega)$ is only considered for positive frequencies ω .

Another interesting possibility to represent a transfer function $G(s)$ in graphical form, is to map the s plane into the complex G plane as shown in section A.2. The result is a cartesian representation of a band of curves $G(\sigma + j\omega)$ as shown in Figure A.2.1. For $\sigma = 0$, i.e. the special case $s = j\omega$, the transfer function $G(s)$ migrates into the frequency response $G(j\omega)$ and we obtain the Nyquist plot of the frequency response, which will be discussed later in section 4.2.

The frequency response $G(j\omega)$ is an immediately physically interpretable and measurable quantity. To show this, the frequency response will be represented by the complex entity

$$G(j\omega) = R(\omega) + jI(\omega) \quad (4.1.1)$$

with the real part $R(\omega)$ and the imaginary part $I(\omega)$, and by the *amplitude response* $A(\omega)$ and *phase response* $\varphi(\omega)$ in polar notation as

$$G(j\omega) = A(\omega) e^{j\varphi(\omega)} . \quad (4.1.2)$$

If the system is excited by a sinusoidal input $x_e(t)$ with amplitude \hat{x}_e and frequency ω , i.e.

$$x_e(t) = \hat{x}_e \sin \omega t, \quad (4.1.3)$$

then in the case of a linear continuous-time system the output signal will oscillate in the steady state with the same frequency ω , but with another amplitude \hat{x}_a and with a certain phase shift $\varphi = \omega t_\varphi$ (Figure 4.1.2a). Both oscillations $x_e(t)$ and $x_a(t)$ can be represented by two vectors of length proportional

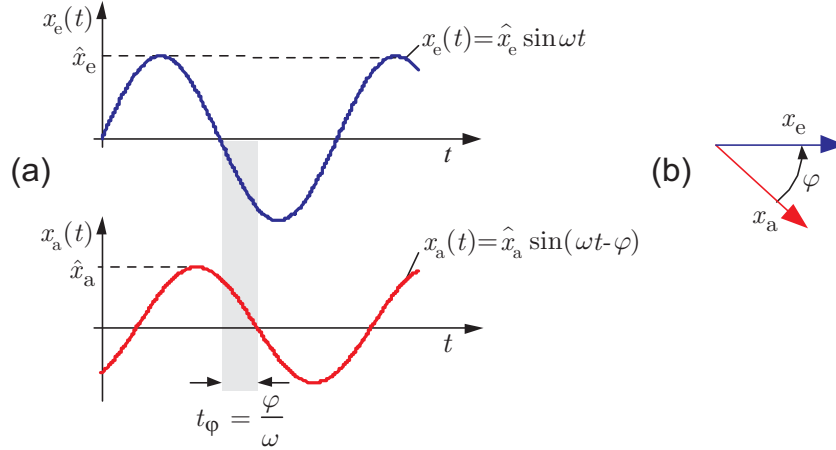


Figure 4.1.2: (a) Sinusoidal input signal $x_e(t)$ and corresponding output signal $x_a(t)$ of a linear element in steady state (b) vector representation of both signals

to their amplitudes \hat{x}_e and \hat{x}_a according to Figure 4.1.2b. They are rotating with the phase shift φ and the same speed ω . For the system output one gets

$$x_a(t) = \hat{x}_a \sin(\omega t + \varphi). \quad (4.1.4)$$

If this experiment is conducted for different frequencies $\omega_\nu (\nu = 0, 1, 2, \dots)$ with $\hat{x}_e = \text{const}$, then one notices that the amplitude \hat{x}_a and the phase φ of the output signal depend on the frequency ω_ν . Therefore for each frequency ω_ν one gets

$$\hat{x}_{a,\nu} = \hat{x}_a(\omega_\nu) \quad \text{and} \quad \varphi_\nu = \varphi(\omega_\nu).$$

Now from the ratio of the values \hat{x}_e and $\hat{x}_a(\omega)$ the *amplitude* of the frequency response

$$A(\omega) = \frac{\hat{x}_a(\omega)}{\hat{x}_e} = |G(j\omega)| = \sqrt{R^2(\omega) + I^2(\omega)} \quad (4.1.5)$$

can be represented as a function of frequency. Furthermore, the phase shift $\varphi(\omega)$ will be represented as the *phase* of the frequency response. For the phase

$$\varphi(\omega) = \arg G(j\omega) = \tan^{-1} \frac{I(\omega)}{R(\omega)} \quad (4.1.6)$$

is valid, where the value of the \tan^{-1} function in the range 0° to 360° must be found through the signs of $R(\omega)$ and $I(\omega)$.

From this experiment it is obvious that the amplitude response $A(\omega)$ and phase response $\varphi(\omega)$ of the frequency response $G(j\omega)$ can be directly measured by applying sinusoidal input signals $x_e(t)$ of different frequencies. The total frequency response $G(j\omega)$ for all frequencies $\omega = 0$ to $\omega \rightarrow \infty$ describes completely the behaviour of a linear continuous-time system, like the the transfer function $G(s)$ or the step response $h(t)$. In some cases only some partial information is sufficient in order to reconstruct the total frequency response. As will be shown later only a knowledge of the real part $R(\omega)$ or of the amplitude response $A(\omega)$ may be necessary.

Between the representations of a linear system in the time domain and frequency domain there are some general and simple relationships. E.g. on the basis of the initial and final value theorems two important relations between the transfer function $G(s)$, the Frequency response $G(j\omega)$ and the step response $h(t)$ or $H(s)$ are valid:

$$\lim_{t \rightarrow 0^+} h(t) = \lim_{s \rightarrow \infty} s H(s) = \lim_{s \rightarrow \infty} G(s) = \lim_{j\omega \rightarrow \infty} G(j\omega) , \tag{4.1.7a}$$

$$\lim_{t \rightarrow \infty} h(t) = \lim_{s \rightarrow 0} s H(s) = \lim_{s \rightarrow 0} G(s) = \lim_{j\omega \rightarrow 0} G(j\omega) . \tag{4.1.7b}$$

The suppositions for the application of the initial and final value theorems are the existence of the related limits in the time domain according to section 2.3.

4.2 Nyquist plot of a frequency response

If in the experiment described above for each value of ω_ν the value of

$$G(j\omega_\nu) = A(\omega_\nu) e^{j\varphi(\omega_\nu)}$$

is plotted in the complex G plane, one obtains the Nyquist plot of the frequency response, which is parameterised by the frequency ω . Figure 4.2.1 shows such a curve drawn from measurements of eight

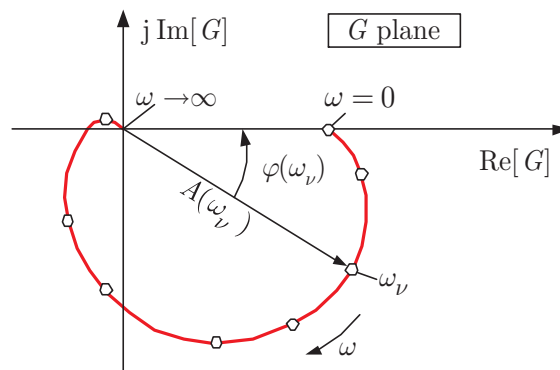


Figure 4.2.1: Example of a Nyquist plot drawn from measurements

frequencies. Using Eqs. (4.1.7a) and (4.1.7b) the initial and final values of the step response $h(t)$ can be estimated from a Nyquist plot based on such measurements (Figure 4.2.2). If $G(j\omega)$ is known analytically

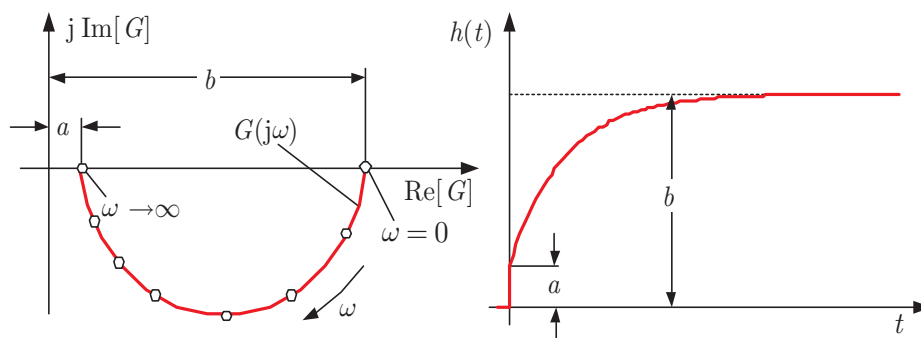


Figure 4.2.2: Relations between initial and final value of the frequency response $G(j\omega)$ and the step response $h(t)$

the step response can be calculated from $G(s)$ by using

$$h(t) = \mathcal{L}^{-1} \left\{ \frac{1}{s} G(s) \right\} . \tag{4.2.1}$$

The graphical representation of a frequency response using a Nyquist plot has among other things the advantage that the frequency responses, both of a series connection and a parallel connection of two systems, can be simply constructed graphically. Here the vectors of the Nyquist plots for the same values of ω must be taken. For the case of a parallel connection the vectors will be added, and for the case of a series connection the vectors will be multiplied by multiplying the lengths of the vectors and adding their angles according to Figure 4.2.3.

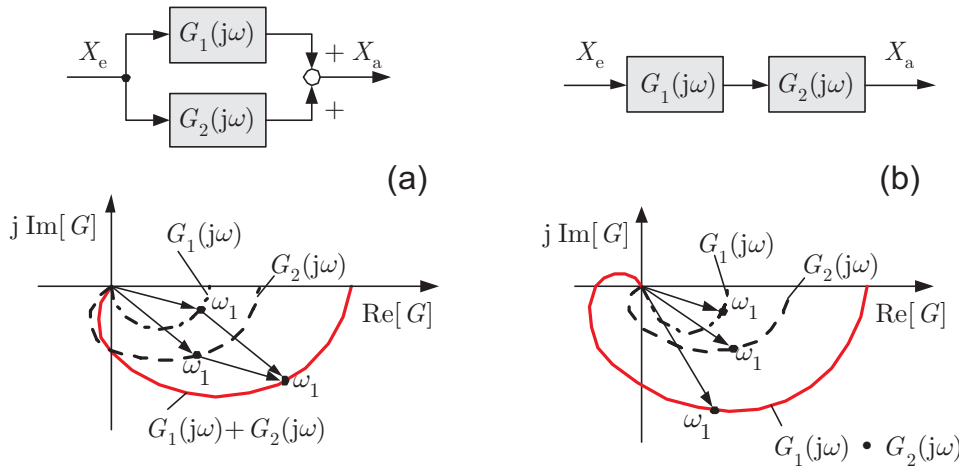


Figure 4.2.3: Addition (a) and multiplication (b) of frequency responses in Nyquist plots

Interactive Questions 4.1
 Test your knowledge about frequency response and Nyquist plots

Interactive Questions 4.2
 Test your knowledge about frequency response and Nyquist plots with other questions

Interactive Questions 4.3
 Test your knowledge about frequency response and Nyquist plots with more demanding questions

4.3 Bode plot

If the absolute value $A(\omega)$ and the phase $\varphi(\omega)$ of the frequency response $G(j\omega) = A(\omega) e^{j\varphi(\omega)}$ are separately plotted over the frequency ω according to Figure 4.3.1, one obtains the *amplitude response* and the *phase response*. Both together are the *frequency response characteristics*. $A(\omega)$ and ω are normally drawn with a logarithm and $\varphi(\omega)$ with a linear scale. This representation is called a *Bode diagram* or *Bode plot*. Usually $A(\omega)$ will be specified in decibels [dB]. By definition this is

$$A(\omega)_{\text{dB}} = 20 \log_{10} A(\omega) \quad [\text{dB}] . \tag{4.3.1}$$

The logarithmic representation of the amplitude response $A(\omega)_{\text{dB}}$ has consequently a linear scale in this diagram and is called the *magnitude*.

The logarithmic representation has some advantages for *series connections* of transfer functions. For complicated frequency responses, e.g. with

$$G(s) = K \frac{(s - s_{Z_1}) \dots (s - s_{Z_m})}{(s - s_{P_1}) \dots (s - s_{P_n})} \tag{4.3.2}$$

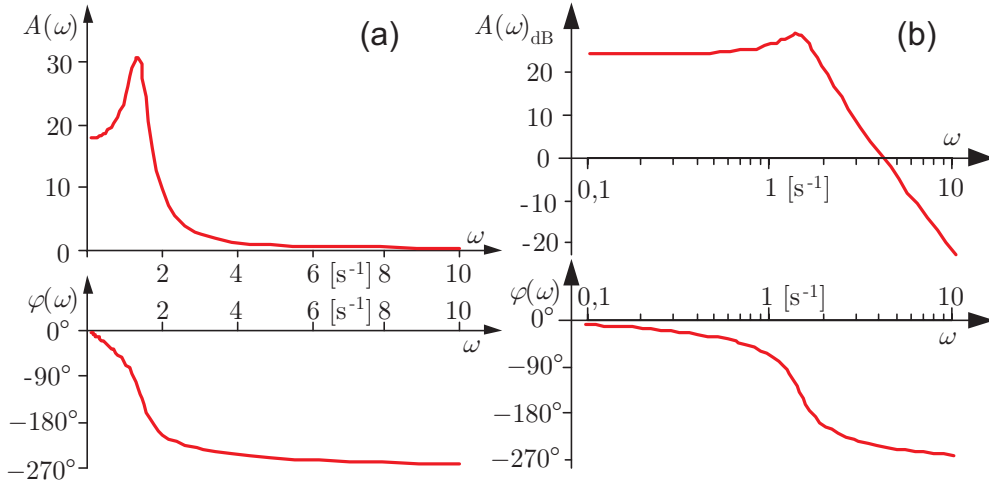


Figure 4.3.1: Plot of a frequency response: (a) linear, (b) logarithmic presentation (ω on a logarithmic scale) (Bode plot)

for $s = j\omega$, it can be represented as series connections of the frequency responses of simple elements of the form

$$G_i(j\omega) = \begin{cases} K = \text{const} & \text{for } i = 0 \\ (j\omega - s_{z_i}) & \text{for } i = 1, 2, \dots, m \end{cases} \quad (4.3.3)$$

and

$$G_i(j\omega) = \frac{1}{j\omega - s_{p_\nu}} \quad \text{for } \begin{matrix} i = m + 1, m + 2, \dots, m + n, \\ \nu = 1, 2, \dots, n. \end{matrix} \quad (4.3.4)$$

From this it follows that

$$G(j\omega) = K G_1(j\omega) \dots G_{m+n}(j\omega), \quad (4.3.5)$$

with

$$G_i(j\omega) = A_i(\omega) e^{j\varphi_i(\omega)} \quad \text{for } i = 1, \dots, m + n.$$

From the representation

$$G(j\omega) = K A_1(\omega) A_2(\omega) \dots A_{m+n}(\omega) e^{j[\varphi_0(\omega) + \varphi_1(\omega) + \varphi_2(\omega) + \dots + \varphi_{m+n}(\omega)]} \quad (4.3.6)$$

and

$$A(\omega) = |K| |G_1(j\omega)| |G_2(j\omega)| \dots |G_{m+n}(j\omega)| = |K| A_1(\omega) A_2(\omega) \dots A_{m+n}(\omega),$$

respectively, one obtains the *logarithmic characteristic of the magnitude*

$$\begin{aligned} A(\omega)_{\text{dB}} &= 20 \log_{10} [|K| A_1(\omega) A_2(\omega) \dots A_{m+n}(\omega)] \\ &= |K|_{\text{dB}} + A_1(\omega)_{\text{dB}} + A_2(\omega)_{\text{dB}} + \dots + A_{m+n}(\omega)_{\text{dB}} \end{aligned} \quad (4.3.7)$$

and the *phase characteristic*

$$\varphi(\omega) = \varphi_0(\omega) + \varphi_1(\omega) + \varphi_2(\omega) + \dots + \varphi_{m+n}(\omega) \quad (4.3.8)$$

with $\varphi_0(\omega) = 0^\circ$ for $K > 0$ and $\varphi_0(\omega) = -180^\circ$ for $K < 0$. Thus, the frequency response of a series connection is obtained by addition of the individual frequency response characteristics.

A further advantage of this logarithmic representation is for the determination of the *inverse* of a frequency response, that is for $1/G(j\omega) = G^{-1}(j\omega)$. Here

$$20 \log_{10} [|G(j\omega)|^{-1}] = -20 \log_{10} |G(j\omega)| = -20 \log_{10} A(\omega)$$

and

$$\arg[G^{-1}(j\omega)] = -\arg[G(j\omega)] \quad (4.3.9)$$

are valid, the curves of $A(\omega)$ and $\varphi(\omega)$ need only to be mirrored at the axes $20 \log_{10} A = 0$ (0-dB line) and $\varphi = 0^\circ$.

Because of the double logarithmic and of the single logarithmic scale of $A(\omega)$ and $\varphi(\omega)$, respectively, the curve of $A(\omega)$ and that of $\varphi(\omega)$ can be approximated by line segments. This approximation by lines allows the analysis and synthesis of control systems using simple geometric constructions. They are important concepts for the control engineer.

4.4 Some important transfer function elements

In the following the transfer function $G(s)$, frequency response $G(j\omega)$, Nyquist plot and the Bode diagram for some important elements will be derived and shown.

4.4.1 The proportional element (P element)

The P element describes a pure proportional relationship between input and output:

$$x_a(t) = K x_e(t) , \quad (4.4.1)$$

with the arbitrary positive or negative constant K . K is called the *gain* of the P element. The transfer function of this system is

$$G(s) = K . \quad (4.4.2)$$

The locus of the frequency response

$$G(j\omega) = K \quad (4.4.3)$$

is for all frequencies a point on the real axis with distance K from the origin. This means that the phase response $\varphi(\omega)$ is zero for $K > 0$ or -180° for $K < 0$. The characteristic of the magnitude is

$$A(\omega)_{\text{dB}} = 20 \log_{10} K = K_{\text{dB}} = \text{const} .$$

4.4.2 The integrator (I element)

The dynamical behaviour of this element is described in the time domain by

$$x_a(t) = \frac{1}{T_I} \int_0^t x_e(\tau) d\tau + x_a(0) \quad (4.4.4)$$

with input and output signals $x_e(t)$ and $x_a(t)$, respectively, and with the *time constant* T_I , which has the dimension 'time'. This element integrates the input signal. Setting $x_a(0) = 0$ one obtains by taking the Laplace transform of Eq. (4.4.4) the transfer function of the I element as

$$G(s) = \frac{1}{s T_I} , \quad (4.4.5)$$

and with $s = j\omega$ the frequency response

$$G(j\omega) = \frac{1}{j\omega T_I} = \frac{1}{\omega T_I} e^{-j\frac{\pi}{2}} . \quad (4.4.6)$$

From this the amplitude and phase responses

$$A(\omega) = \frac{1}{\omega T_I} \quad \text{and} \quad \varphi(\omega) = -\frac{\pi}{2}$$

follow. For the magnitude characteristic one obtains

$$A(\omega)_{\text{dB}} = +20 \log_{10} \frac{1}{\omega T_I} = -20 \log_{10} \omega T_I . \quad (4.4.7)$$

The graphical representation of Eq. (4.4.7) gives a line with a slope of -20dB/decade, or equivalently -6dB/octave, in the Bode diagram, Figure 4.4.1a. The phase response is independent of the frequency. The Nyquist plot of the frequency response

$$G(j\omega) = -j \frac{1}{\omega T_I}$$

coincides with the negative imaginary axis, as shown in Figure 4.4.1b.

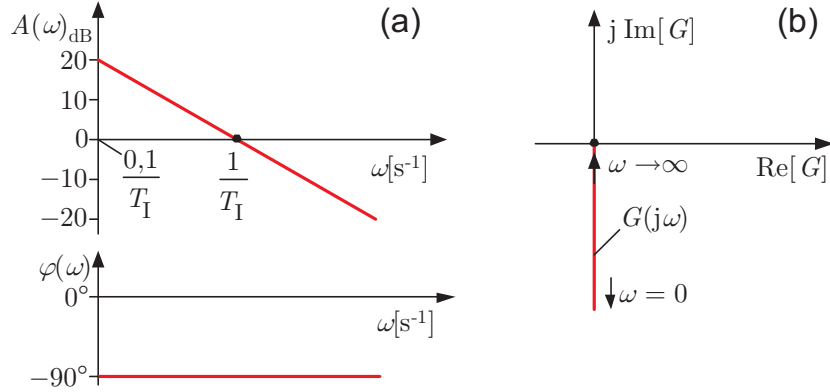


Figure 4.4.1: (a) Magnitude and phase response (b) Nyquist plot of the frequency response of an integrator

4.4.3 The derivative element (D element)

The relationship between the input $x_e(t)$ and output $x_a(t)$ of a derivative element is described by

$$x_a(t) = T_D \frac{d}{dt} x_e(t) . \quad (4.4.8)$$

This element differentiates the input signal $x_e(t)$ and therefore is called a derivative element, or in short D element. The associated transfer function is

$$G(s) = s T_D , \quad (4.4.9)$$

and with $s = j\omega$ it follows that the frequency response

$$G(j\omega) = j\omega T_D = \omega T_D e^{j\frac{\pi}{2}} , \quad (4.4.10)$$

from which the magnitude

$$A(\omega)_{dB} = 20 \log_{10} \omega T \quad (4.4.11)$$

and the phase response

$$\varphi(\omega) = \frac{\pi}{2} \quad (4.4.12)$$

follow. It can easily be seen, that the I and D elements are related by an inversion. Therefore the curves of the magnitude and phase response of the D element can be found – as shown above – by mirroring those of the I element at the 0-dB line and $\varphi = 0$ line, respectively. This is obvious from Eqs. (4.4.11) and (4.4.12). Figure 4.4.2 shows the Bode diagram and the Nyquist plot of the frequency response of the D element. The slope of the line $A(\omega)$ is +20dB/decade and the phase response is independent of the frequency.

The D element discussed here is – as already mentioned in section 3.3 – an idealisation and therefore not a physically realisable element. For practical applications the D element will be approximated by the DT_1 element (see section 4.4.6).

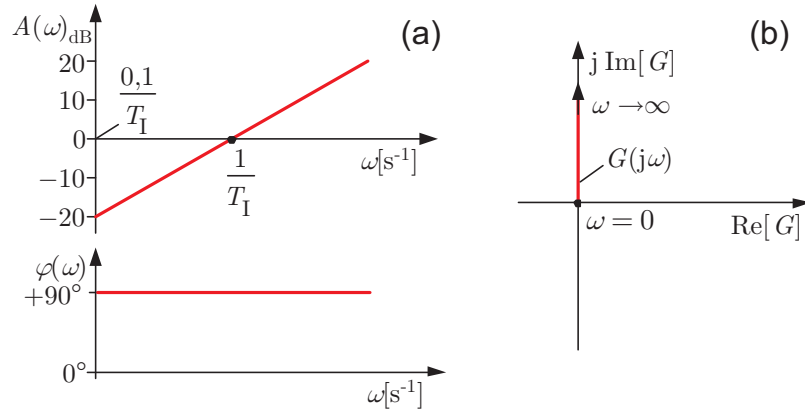


Figure 4.4.2: (a) Magnitude and phase response (b) Nyquist plot of the frequency response of a D element

4.4.4 The 1st-order lag element (PT₁ element)

The 1st-order lag element or in short PT₁ element is an element with an output signal $x_a(t)$ that for a step input $x_e(t)$ has a certain initial slope and approaches asymptotically the final value. An example of such an element is the simple RC lag circuit shown in Figure 4.4.3. When at time $t = 0$ a voltage $u_e = 2V$ is applied, the voltage u_a at the output will approach exponentially with the time constant $T = RC$ the final value $u_a = 2V$.

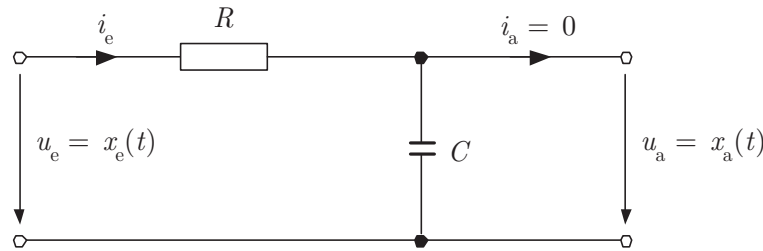


Figure 4.4.3: Simple RC lag as an example of a 1st-order lag element

The circuit involves a single energy storage element, the capacitor C . The differential equation of this RC lag is

$$x_a(t) + RC\dot{x}_a(t) = x_e(t) . \tag{4.4.13}$$

For the general notation of a PT₁ element one obtains the differential equation

$$x_a(t) + T\dot{x}_a(t) = K x_e(t) . \tag{4.4.14}$$

If the initial condition x_a is set to zero, on taking the Laplace transform the transfer function is

$$G(s) = \frac{K}{1 + sT} , \tag{4.4.15}$$

and it follows that with $s = j\omega$ the frequency response is

$$G(j\omega) = K \frac{1}{1 + j\omega T} . \tag{4.4.16}$$

With the *breakpoint frequency* $\omega_B = \frac{1}{T}$ one obtains

$$G(j\omega) = K \frac{1}{1 + j\frac{\omega}{\omega_B}} = K \frac{1 - j\frac{\omega}{\omega_B}}{1 + \left(\frac{\omega}{\omega_B}\right)^2} . \tag{4.4.17}$$

The amplitude response is

$$A(\omega) = |G(j\omega)| = K \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_B}\right)^2}} \tag{4.4.18}$$

and the phase response is

$$\varphi(\omega) = \tan^{-1} \frac{I(\omega)}{R(\omega)} = -\tan^{-1} \frac{\omega}{\omega_B}. \tag{4.4.19}$$

The magnitude characteristic derived from Eq. (4.4.18) is

$$A(\omega)_{dB} = 20 \log_{10} K - 20 \log_{10} \sqrt{1 + \left(\frac{\omega}{\omega_B}\right)^2}. \tag{4.4.20}$$

Eq. (4.4.20) can be asymptotically approximated by lines for:

a) $\frac{\omega}{\omega_B} \ll 1$ by

$$A(\omega)_{dB} \approx 20 \log_{10} K = K_{dB} \quad (\text{initial asymptote}),$$

with

$$\varphi(\omega) \approx 0;$$

b) $\frac{\omega}{\omega_B} \gg 1$ by

$$A(\omega)_{dB} \approx 20 \log_{10} K - 20 \log_{10} \frac{\omega}{\omega_B} \quad (\text{final asymptote}),$$

with

$$\varphi(\omega) \approx -\frac{\pi}{2}.$$

In the Bode diagram $A(\omega)_{dB}$ can be consequently approximated by two lines. The progression of the initial asymptote is horizontal, whereas the final asymptote shows a slope of -20dB/decade. The intersection of both lines (breakpoint) can be determined from

$$20 \log_{10} K = 20 \log_{10} K - 20 \log_{10} \frac{\omega}{\omega_B}$$

and provides the frequency

$$\omega = \omega_B.$$

Therefore $\omega = \omega_B$ is called the breakpoint frequency. As can be easily seen from Figure 4.4.4a, the

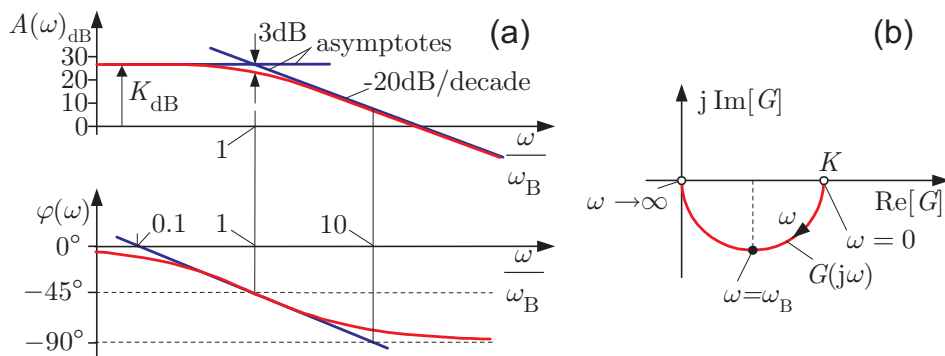


Figure 4.4.4: (a) Magnitude and phase response (b) Nyquist plot of the frequency response of a PT₁ element

deviation between $A(\omega)_{dB}$ and the asymptotes has a maximum at the breakpoint for ω_B . The exact values are

$$A(\omega_B) = K \frac{1}{\sqrt{2}} \quad \text{and} \quad \varphi(\omega_B) = -\frac{\pi}{4}.$$

The deviation of the magnitude characteristic from the asymptotes for $\omega = \omega_B$ is

$$\Delta A(\omega_B)_{dB} = -20 \log_{10} \sqrt{2} \text{ dB} \approx -3 \text{ dB} .$$

The deviations at the other frequencies are symmetrical from the breakpoint on the logarithmic scale, as can be seen directly from Table 4.4.1. This is the reason why the magnitude and phase characteristics can be easily constructed on a Bode diagram. The phase curve is approximately zero up to one tenth of the breakpoint frequency, and -90° beyond 10 time the breakpoint frequency. In between, it is approximately a straight line with slope 45° per decade through -45° at the breakpoint frequency.

Table 4.4.1: Magnitude, phase response and deviation $\Delta A(\omega)$ of the exact magnitude from the asymptotes for a PT_1 element with $K = 1$

$\frac{\omega}{\omega_B}$	$A(\omega)_{dB}$	$\varphi(\omega)$	$\Delta A(\omega)_{dB}$
0.03	0.0	- 2°	0.00
0.1	-0.04	- 6°	-0.04
0.25	-0.26	- 14°	-0.26
0.5	-0.97	- 27°	-0.97
0.76	-2.00	- 37°	-2.00
1.0	-3.00	- 45°	-3.00
1.31	-4.35	- 53°	-2.00
2.0	-6.99	- 63°	-0.97
4.0	-12	- 76°	-0.26
10.0	-20	- 84°	-0.04
30.0	-30	- 88°	0.00

As already shown in section A.2 the locus of the frequency response of a PT_1 element is a semicircle, which starts for $\omega = 0$ at K on the real axis and stops for $\omega \rightarrow \infty$ at the origin, as shown in Figure 4.4.4b.

The constant $T = 1/\omega_B$ in the transfer function and on the frequency response, respectively, is usually called the *time constant* of the PT_1 element. It can be determined also by the point of intersection of the line with the initial slope and the horizontal line of the final asymptote, $h(\infty)$, of the step response $h(t)$ as shown in Figure 4.4.5. This time constant can also be physically interpreted. It is the time when

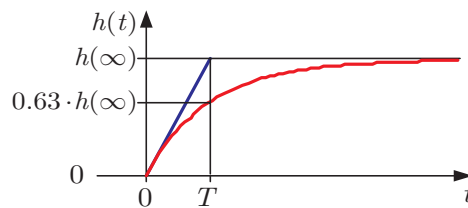


Figure 4.4.5: Graphical representation of the step response, $h(t)$, of a PT_1 element

the step response has reached approx. 63% of the final value, $h(\infty)$. K is – similar to the P element – called the *gain* of the PT_1 element. It is defined as the value of the frequency response at $\omega = 0$.

DYNAST study example 4.1
[Simple automobile model](#)

DYNAST study example 4.2
[D.C. motor - open loop](#)

4.4.5 The proportional plus derivative element (PD element)

The PD element shows both proportional and derivative behaviour and is described by the transfer function

$$G(s) = K(1 + sT) . \tag{4.4.21}$$

Apart from the gain factor K this element is the inverse of the PT_1 element. Hence for $K = 1$ one obtains the magnitude and phase response by mirroring at the 0-dB axis and the $\varphi(\omega) = 0$ line, respectively (compare Figure 4.4.6 with 4.4.4). The locus of the frequency response

$$G(j\omega) = K(1 + j\omega T) \tag{4.4.22}$$

is a semi-line, which starts for $\omega = 0$ on the real axis at K and progresses parallel to the imaginary axis for increasing values of ω .

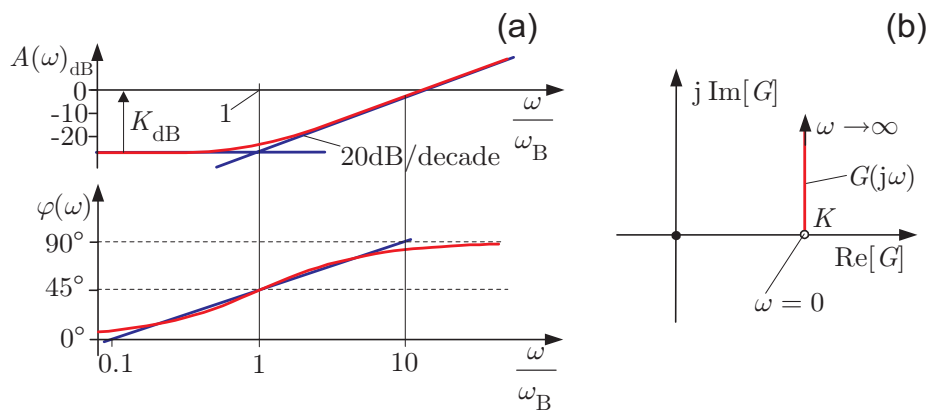


Figure 4.4.6: (a) Magnitude and phase response (b) Nyquist plot of the frequency response of a PD element

4.4.6 The derivative lag element (DT_1 element)

This element has a step response which initially contains a step and then decreases exponential to zero with a characteristic time constant as shown in Figure 4.4.7. Figure 4.4.8 shows an example of such a

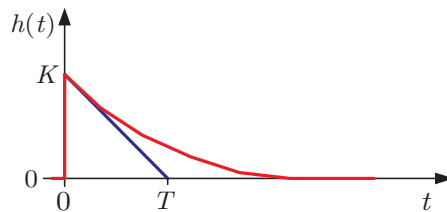


Figure 4.4.7: Graphical representation of the step response, $h(t)$, of a DT_1 element

system; a simple RC high pass filter. The differential equation of this circuit is

$$C \frac{d(u_e - u_a)}{dt} = \frac{u_a}{R} ,$$

which can be written as

$$u_a + RC \frac{du_a}{dt} = RC \frac{du_e}{dt} .$$

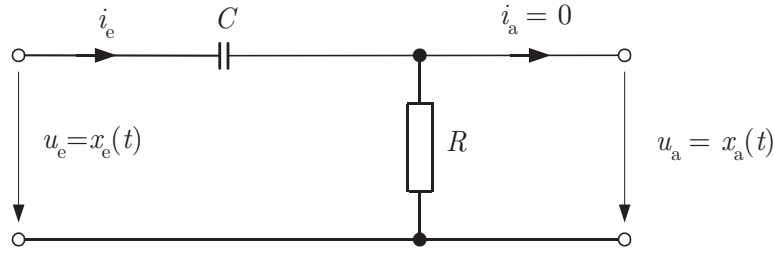


Figure 4.4.8: Simple RC high-pass circuit as an example of a DT_1 element

Applying the Laplace transform one obtains the transfer function

$$G(s) = \frac{U_a(s)}{U_e(s)} = \frac{RCs}{1 + RCs}. \quad (4.4.23)$$

The generalised notation of the DT_1 element is

$$G(s) = K \frac{Ts}{1 + Ts}. \quad (4.4.24)$$

For constructing the Bode plot one starts with the frequency response

$$G(j\omega) = K \frac{j\omega T}{1 + j\omega T}, \quad (4.4.25)$$

which on substituting $\omega_B = \frac{1}{T}$ gives

$$G(j\omega) = Kj \frac{\omega}{\omega_B} \frac{1}{1 + j \frac{\omega}{\omega_B}} = \frac{\omega}{\omega_B} K \frac{\frac{\omega}{\omega_B} + j}{1 + \left(\frac{\omega}{\omega_B}\right)^2}. \quad (4.4.26)$$

From Eq. (4.4.26) it follows that

$$A(\omega) = |G(j\omega)| = \frac{\omega}{\omega_B} K \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_B}\right)^2}},$$

and

$$A(\omega)_{dB} = 20 \log_{10} \frac{\omega}{\omega_B} + 20 \log_{10} K - 20 \log_{10} \sqrt{1 + \left(\frac{\omega}{\omega_B}\right)^2}. \quad (4.4.27)$$

After some calculations the phase response can be shown to be

$$\varphi(\omega) = \frac{\pi}{2} - \tan^{-1} \left(\frac{\omega}{\omega_B} \right). \quad (4.4.28)$$

Comparing Eq. (4.4.27) with Eqs. (4.4.20) and (4.4.23) shows that the magnitude characteristic of the DT_1 element can be obtained by adding the corresponding curves of a PT_1 element and a D element. The same also holds for the phase response $\varphi(\omega)$. With this information the curves of the frequency response characteristics and of the Nyquist diagram can be simply constructed according to Figure 4.4.9.

4.4.7 The 2nd-order lag element (PT_2 element and PT_2S element)

A 2nd-order lag element is characterised by two independent energy storages. Depending on the damping properties and the position of the poles of $G(s)$, respectively, one distinguishes between oscillating and

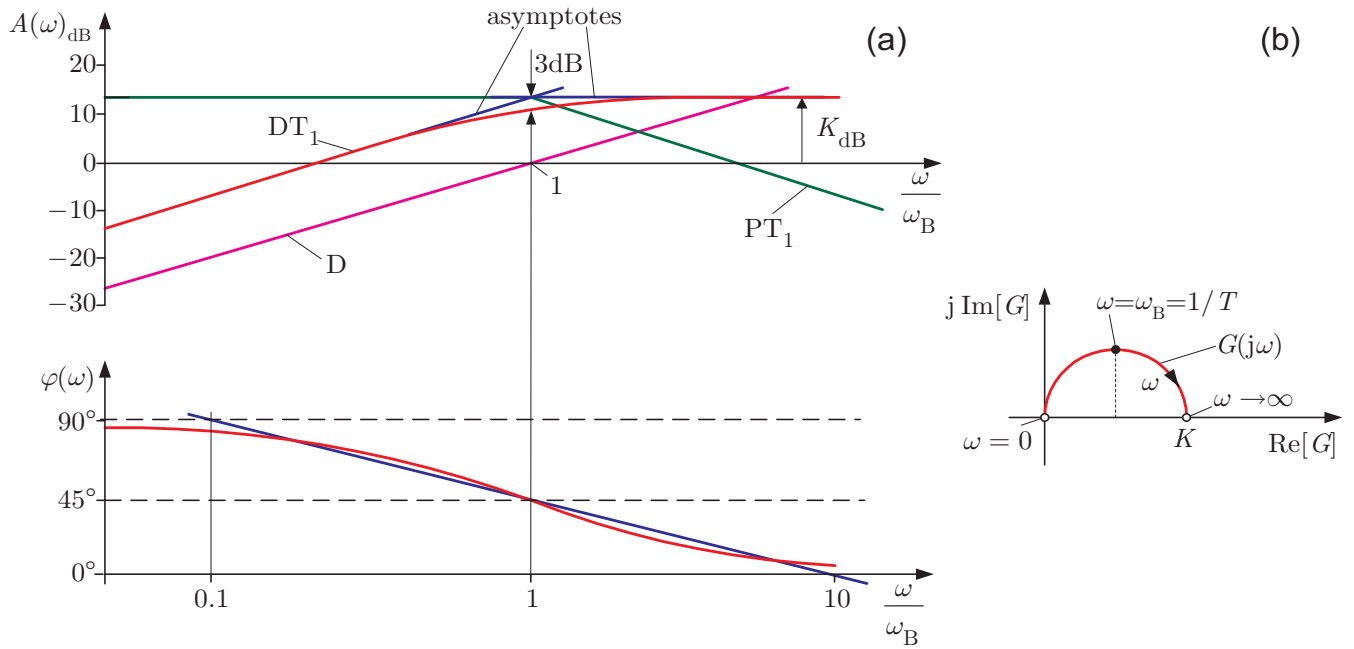


Figure 4.4.9: (a) Magnitude and phase responses (b) Nyquist plot of the frequency response of a DT_1 element

aperiodic behaviour. If a 2nd-order lag element has a conjugate complex pair of poles, then it shows an oscillating behaviour (PT_2S behaviour). If both poles are on the negative real axis, then the element has a lag behaviour (PT_2 behaviour).

The RLC network of Figure 4.4.10 is an example of such an element. From the equation of the mesh

$$i_e R + L \frac{di_e}{dt} + u_a = u_e \quad (4.4.29)$$

with

$$i_e = C \frac{du_a}{dt} \quad (4.4.30)$$

the differential equation becomes

$$LC \frac{d^2 u_a}{dt^2} + RC \frac{du_a}{dt} + u_a(t) = u_e(t) . \quad (4.4.31)$$

Interactive Questions 4.4
[Test using other example](#)

The transfer function is

$$G(s) = \frac{U_a(s)}{U_e(s)} = \frac{1}{1 + RCs + LCs^2} . \quad (4.4.32)$$

For the 2nd-order lag element the general notation of the transfer function

$$G(s) = \frac{K}{1 + T_1 s + T_2^2 s^2} \quad (4.4.33)$$

is chosen. Introducing terms which characterise the time behaviour, that is the *damping ratio*

$$\zeta = \frac{1}{2} \frac{T_1}{T_2} \quad (4.4.34)$$

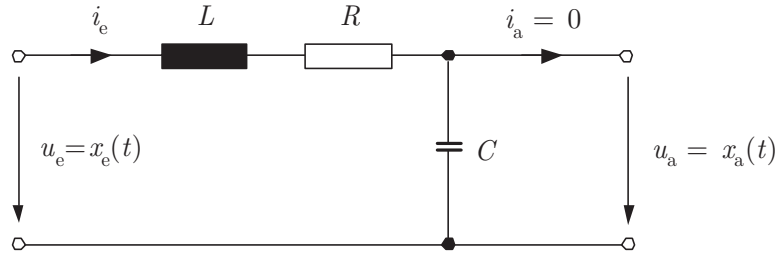


Figure 4.4.10: Simple RLC network as an example of a 2nd-order lag element

and the *natural frequency* (frequency of the undamped oscillation)

$$\omega_0 = \frac{1}{T_2}, \quad (4.4.35)$$

one obtains from Eq. (4.4.33)

$$G(s) = \frac{K}{1 + \frac{2\zeta}{\omega_0}s + \frac{1}{\omega_0^2}s^2} = \frac{K}{D(s)}. \quad (4.4.36)$$

For $s = j\omega$ the frequency response

$$G(j\omega) = \frac{K}{1 + j2\zeta\frac{\omega}{\omega_0} - \frac{\omega^2}{\omega_0^2}} = K \frac{\left[1 - \left(\frac{\omega}{\omega_0}\right)^2\right] - j2\zeta\frac{\omega}{\omega_0}}{\left[1 - \left(\frac{\omega}{\omega_0}\right)^2\right]^2 + \left[2\zeta\frac{\omega}{\omega_0}\right]^2}. \quad (4.4.37)$$

results. The amplitude response is

$$A(\omega) = \frac{K}{\sqrt{\left[1 - \left(\frac{\omega}{\omega_0}\right)^2\right]^2 + \left(2\zeta\frac{\omega}{\omega_0}\right)^2}} \quad (4.4.38)$$

and the phase response

$$\varphi(\omega) = -\tan^{-1} \frac{2\zeta\frac{\omega}{\omega_0}}{1 - \left(\frac{\omega}{\omega_0}\right)^2}. \quad (4.4.39)$$

Here the ambivalence of the \tan^{-1} function has to be observed. For the magnitude characteristic one has from Eq. (4.4.38)

$$A(\omega)_{\text{dB}} = 20 \log_{10} K - 20 \log_{10} \sqrt{\left[1 - \left(\frac{\omega}{\omega_0}\right)^2\right]^2 + \left(2\zeta\frac{\omega}{\omega_0}\right)^2}. \quad (4.4.40)$$

The progression of $A(\omega)_{\text{dB}}$ can be approximated by the following asymptotes:

a) For $\frac{\omega}{\omega_0} \ll 1$ by

$$A(\omega)_{\text{dB}} \approx 20 \log_{10} K \quad (\text{initial asymptote}),$$

with

$$\varphi(\omega) \approx 0.$$

b) For $\frac{\omega}{\omega_0} \gg 1$ by

$$A(\omega)_{\text{dB}} \approx 20 \log_{10} K - 20 \log_{10} \left(\frac{\omega}{\omega_0}\right)^2 \approx 20 \log_{10} K - 40 \log_{10} \left(\frac{\omega}{\omega_0}\right) \quad (\text{final asymptote}),$$

with

$$\varphi(\omega) \approx -\pi .$$

In the Bode diagram the final asymptote is a line with a slope of -40dB/decade. The point of intersection of both asymptotes follows from

$$20 \log_{10} K = 20 \log_{10} K - 40 \log_{10} \left(\frac{\omega}{\omega_0} \right)$$

to be the normalised frequency $\frac{\omega}{\omega_0}$. The exact value of $A(\omega)_{\text{dB}}$ may deviate considerably from the point of intersection at $\omega = \omega_0$, because it is

$$A(\omega_0)_{\text{dB}} = 20 \log_{10} K - 20 \log_{10} 2\zeta$$

according to Eq. (4.4.38). For $\zeta < 0.5$ this value is above, for $\zeta > 0.5$ below the asymptotes.

Figure 4.4.11 shows for $0 < \zeta \leq 2.5$ and $K = 1$ the magnitude and phase responses in a Bode diagram. This graphical representation contains the cases of PT₂S and PT₂ behaviour, which will be discussed in section A.3.2. From Figure 4.4.11 it can be seen that a maximum magnitude exists for some values of the damping ratio ζ . This maximum occurs at the so called *resonant peak frequency* ω_p . A detailed analysis of this resonance can be found in section A.3.1.

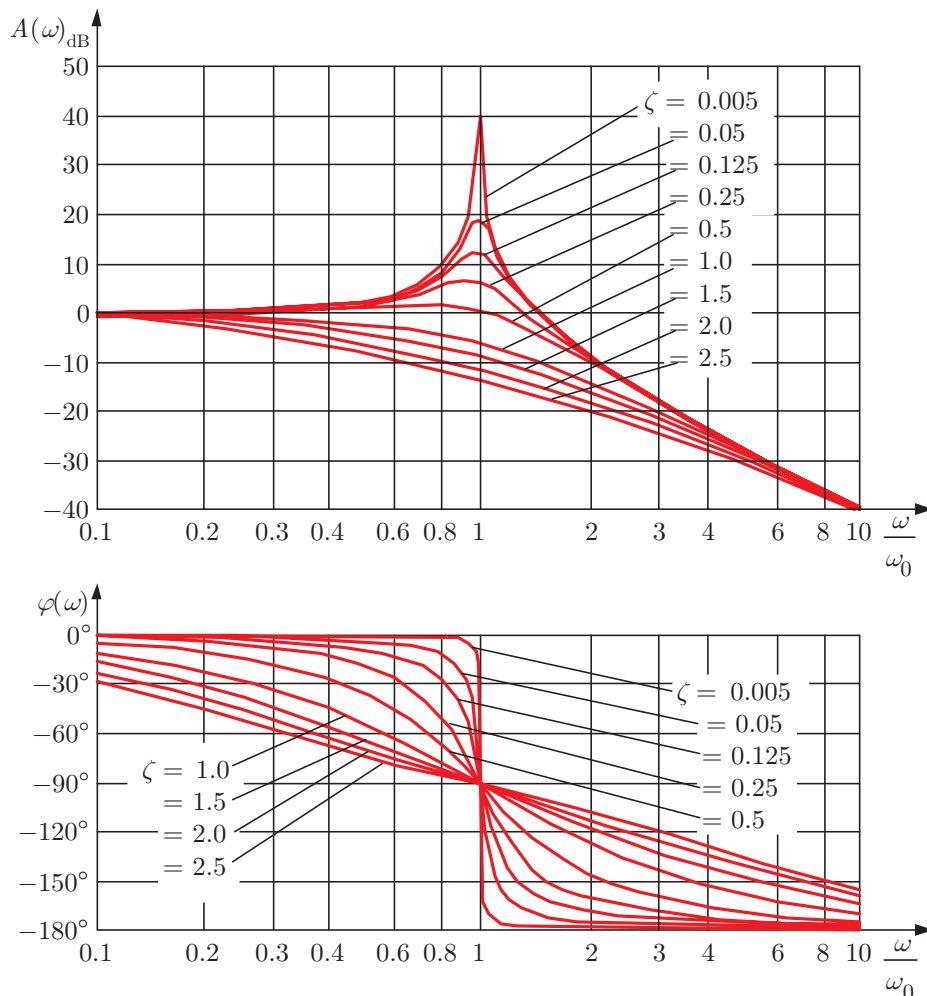


Figure 4.4.11: Bode diagram of a 2nd-order lag element with the transfer function $G(s) = 1/[1 + s2\zeta/\omega_0 + (s/\omega_0)^2]$

Figure 4.4.12 shows the Nyquist plots of 2nd-order lag elements with high and low damping. From Eq. (4.4.37) it follows that for $\omega = \omega_0$ the real part of $G(j\omega)$ is zero. Therefore the locus of the frequency

response intersects with the imaginary axis at $\omega = \omega_0$ independent of the values of ζ , as can be seen in Figure 4.4.12.

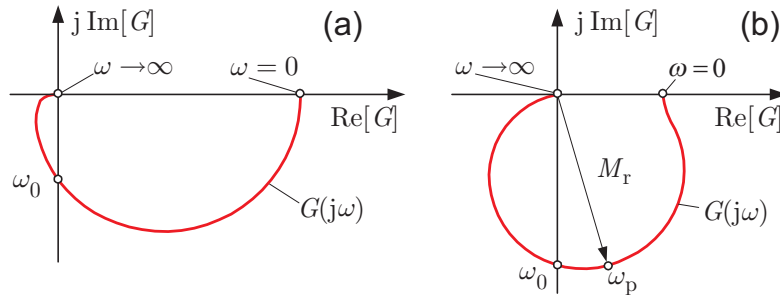


Figure 4.4.12: Nyquist plots of 2nd-order lag elements, (a) PT_2 element, (b) PT_2S element

The modes of a dynamical system are determined according to Eq. (3.5.3) by the roots of the characteristic equation or by the poles of the transfer function, respectively. From the characteristic equation of the 2nd-order lag element

$$P(s) \equiv D(s) = 1 + \frac{2\zeta}{\omega_0}s + \frac{1}{\omega_0^2}s^2 = 0 \tag{4.4.41}$$

one obtains the poles of the transfer function as

$$s_{1,2} = -\omega_0\zeta \pm \omega_0\sqrt{\zeta^2 - 1} . \tag{4.4.42}$$

The oscillating behaviour of a 2nd-order lag element is dependent on the position of the poles in the s plane. For a graphical analysis the step response $h(t)$ is a useful tool. Table 4.4.2 shows step responses and the corresponding poles for different values of ζ . These five cases will be analysed and discussed in more detail in section A.3.2.

4.4.8 Bandwidth of a system

An important term that has not been defined so far is the *bandwidth* of a system. Lag elements with a proportional behaviour, e.g. PT_1 , PT_2 and PT_2S elements as well as PT_n elements (n PT_1 elements in series connection), show a so-called low-pass property. This means that they pass low frequencies whereas high frequencies in signals are attenuated by the strongly decreasing amplitude response of the frequency response. In order to describe this behaviour the concept of bandwidth is introduced. This is the frequency ω_b at which the magnitude of the frequency response is decreased by 3dB from the value of the initial horizontal asymptote, see Figure 4.4.13.

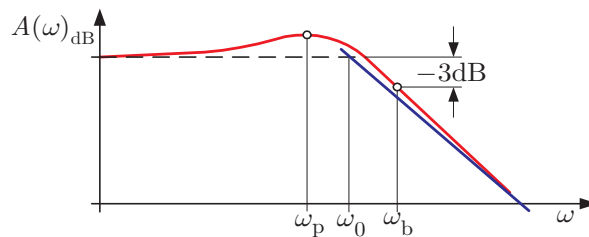


Figure 4.4.13: Definition of the bandwidth ω_b of systems with lags (ω_p resonant peak frequency, ω_0 natural frequency of the undamped oscillation; ω on logarithmic scale)

Table 4.4.2: Pole positions in the s plane and step responses for elements with the transfer function $G(s) = 1/[1 + s2\zeta/\omega_0 + (s/\omega_0)^2]$

damping	pole positions	step response $h(t)$
$0 < \zeta < 1$		
$\zeta = 1$		
$\zeta > 1$		
$\zeta = 0$		
$-1 < \zeta < 0$		

4.4.9 Example for the construction of a Bode plot

In section 4.3 it has been shown how Bode plots of a system with rational transfer functions can be generated by decomposing the system into smaller elements, such as those shown in Table A.8.3. In this section a detailed example will be given.

A Bode plot is to be drawn for a system, which has only real poles and zeros. The given transfer function

$$G(s) = K_1 \frac{(s + 0.1)(s + 2)}{s(s + 5)(s + 20)}, \quad \text{with } K_1 = 890$$

is rearranged into the form

$$G(s) = K_2 \frac{\left(\frac{s}{0.1} + 1\right) \left(\frac{s}{2} + 1\right)}{s \left(\frac{s}{5} + 1\right) \left(\frac{s}{20} + 1\right)}; \quad K_2 = K_1/500 = 1.78 .$$

This system can now be decomposed into an integrator, two PD and two PT₁ elements, that is

$$\begin{aligned} G(s) &= \underbrace{\frac{K_2}{s}}_{G_1(s)} \underbrace{\left(\frac{s}{0.1} + 1\right)}_{G_2(s)} \underbrace{\left(\frac{s}{2} + 1\right)}_{G_3(s)} \underbrace{\frac{1}{\frac{s}{5} + 1}}_{G_4(s)} \underbrace{\frac{1}{\frac{s}{20} + 1}}_{G_5(s)} . \\ &= G_1(s) \quad G_2(s) \quad G_3(s) \quad G_4(s) \quad G_5(s) \end{aligned}$$

From this simple analysis the Bode plot can be determined by adding the Bode plots of the elements G_1 to G_5 according to Figure 4.4.14a. In this figure the variable quantities ω_{B_i} in the terms $(s/\omega_{B_i} + 1)^{\pm 1}$

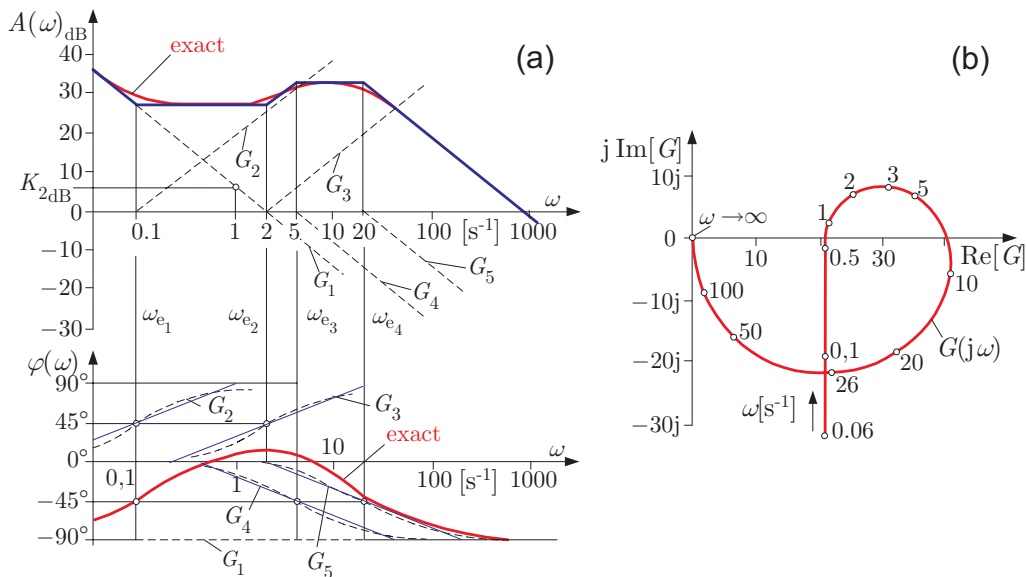


Figure 4.4.14: Representation of a dynamic system by two frequency response diagrams: (a) Bode plot, (b) Nyquist plot

for $i = 1, 2, 3$ and 4 are the breakpoint frequencies in the Bode plot. Also shown in Figure 4.4.14 is the Nyquist plot of the frequency response. Both representations of Figure 4.4.14 basically contain the same information about the system.

Based on the example given above the procedure for constructing a Bode plot of a given system can be recapitulated:

a) The given transfer function must be put into the form

$$G(s) = K \frac{(s - s_{Z_1}) \dots (s - s_{Z_m})}{(s - s_{P_1}) \dots (s - s_{P_n})} = K \frac{\prod_{\mu=1}^m (-s_{Z_\mu})}{\prod_{\substack{\nu=1 \\ s_{P_\nu} \neq 0}}^n (-s_{P_\nu})} \frac{1}{s^k} \frac{\prod_{\mu=1}^m \left(1 + \frac{s}{-s_{Z_\mu}}\right)}{\prod_{\substack{\nu=1 \\ s_{P_\nu} \neq 0}}^n \left(1 + \frac{s}{-s_{P_\nu}}\right)}$$

with $k = 0, 1, 2, \dots$,

where possible poles of $G(s)$ at $s_{P_\nu} = 0$ will be specially considered according to their multiplicity k .

b) Then for $s = j\omega$ the asymptotes of the elements will be used to approximate $A(\omega)_{\text{dB}}$ and $\varphi(\omega)$.

c) If necessary corrections of the approximations can be performed.

4.5 Systems with minimum and non-minimum phase behaviour

Stable systems without dead time, which are described by the transfer function

$$G(s) = \frac{N(s)}{D(s)}$$

and which do not have zeros in the right half s plane, are called *minimum phase* systems. They are characterised by the fact that for a known amplitude response $A(\omega) = |G(j\omega)|$ in the range of $0 \leq \omega < \infty$ the corresponding phase response $\varphi(\omega)$ can be calculated from $A(\omega)$ and that the value of $\varphi(\omega)$ determined has its minimum modulus for the given $A(\omega)$.

If a transfer function has poles and/or zeros in the right half s plane then this system shows *non-minimum phase behaviour*. The modulus of the phase response is then always larger than that for a system with minimum phase behaviour, which has the same amplitude response.

In order to illustrate the non-minimum phase behaviour, two systems will be considered, which have in fact the same amplitude response $A(\omega)$ but differ considerably in the phase response. The transfer functions of the two systems are

$$G_a(s) = \frac{1 + sT}{1 + sT_1} \quad \text{and} \quad G_b(s) = \frac{1 - sT}{1 + sT_1}$$

with $0 < T < T_1$. The distributions of the poles and zeros of $G_a(s)$ and $G_b(s)$ in the s plane is shown in Figure 4.5.1. The amplitude response of the corresponding frequency responses $G_a(j\omega)$ (minimum phase

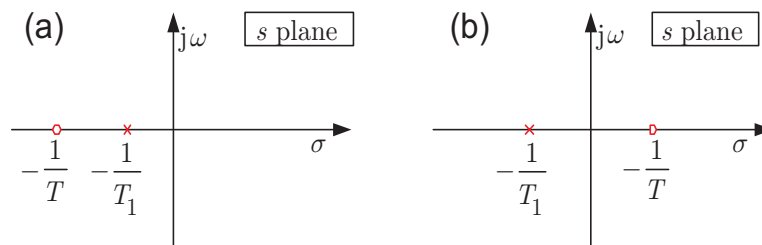


Figure 4.5.1: Distribution of poles and zeros in the s plane (a) $G_a(s)$ and (b) $G_b(s)$

system) and $G_b(j\omega)$ (non-minimum phase system) is in both cases the same, as

$$A_a(\omega) = A_b(\omega) = \sqrt{\frac{1 + (\omega T)^2}{1 + (\omega T_1)^2}}.$$

For the phase responses one obtains

$$\varphi_a(\omega) = -\tan^{-1} \frac{\omega(T_1 - T)}{1 + \omega^2 T_1 T}$$

and

$$\varphi_b(\omega) = -\tan^{-1} \frac{\omega(T_1 + T)}{1 - \omega^2 T_1 T}$$

a different result, which is shown in Figure 4.5.2. For $\varphi_b(\omega)$ the ambivalence of the \tan^{-1} function has to be observed. Here the minimum phase response of $\varphi_a(\omega)$ can be clearly seen.

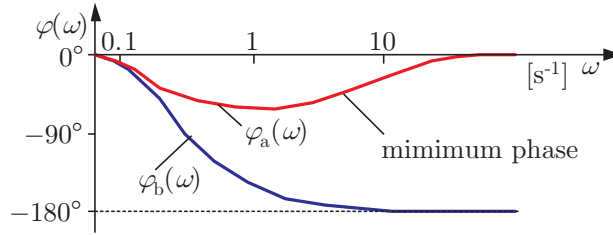


Figure 4.5.2: Phase response of two transfer functions with identical amplitude, but with minimum and non-minimum phase behaviour: $|\varphi_a| < |\varphi_b|$

The transfer functions of non-minimum phase systems, like $G_b(s)$, can always be composed by a series connection of a minimum phase system and a pure phase shift element, which are described by the transfer functions $G_a(s)$ and $G_A(s)$:

$$G_b(s) = G_A(s) G_a(s) . \quad (4.5.1)$$

A pure phase shift element, also called *all-pass element*, is characterised by the fact, that the modulus of its frequency response $G_A(j\omega)$ has a value of unity at all frequencies. For this example one obtains

$$G_b(s) = G_A(s) G_a(s)$$

or

$$\frac{1 - sT}{1 + sT_1} = \frac{1 - sT}{1 + sT} \frac{1 + sT}{1 + sT_1} .$$

From this, the transfer function of the all-pass element (1st order) is

$$G_A(s) = \frac{1 - sT}{1 + sT} ,$$

which has the amplitude response

$$A_A(\omega) = 1 \quad (4.5.2)$$

and the phase response

$$\varphi_A(\omega) = -\tan^{-1} \frac{2\omega T}{1 - (\omega T)^2} = -2 \tan^{-1} \omega T .$$

This all-pass element has a phase response $\varphi_A(\omega)$ from 0° to -180° . The condition of Eq. (4.5.2) is only fulfilled by systems for which the distribution of zeros of the transfer function $G_A(s)$ in the s plane is symmetric to the $j\omega$ axis. This is shown in Figure 4.5.3 for a stable 4th-order all-pass system.

The transfer function of an n th-order all-pass system is

$$G_A(s) = \pm \frac{(s - \tilde{s}_1)(s - \tilde{s}_2) \dots (s - \tilde{s}_n)}{(s - s_1)(s - s_2) \dots (s - s_n)} , \quad (4.5.3)$$

where the corresponding poles s_i and zeros \tilde{s}_i for $i = 1, 2, \dots, n$ only differ in the signs of their real parts ($\text{Re } s_i = -\text{Re } \tilde{s}_i$).

For minimum phase systems, as already mentioned and shown in section A.4, the phase response $\varphi(\omega)$ can be unambiguously determined from the amplitude response $A(\omega)$. This is not valid for non-minimum

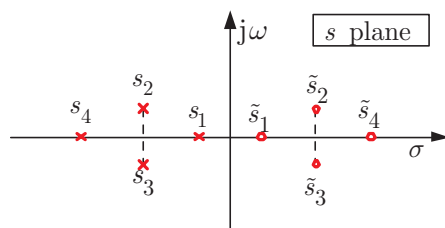


Figure 4.5.3: Poles (x) and zeros (o) of a 4th-order all-pass system

phase systems. The verification of whether a system is non-minimum or minimum phase can be easily estimated from the progression of $\varphi(\omega)$ and $A(\omega)_{\text{dB}}$ for high frequencies. For a minimum phase system, which is described by the transfer function

$$G(s) = \frac{N(s)}{D(s)}$$

with $m = \text{degree } N(s)$ and $n = \text{degree } D(s)$ one obtains for $\omega \rightarrow \infty$ the phase

$$\varphi(\infty) = -90^\circ(n - m) . \quad (4.5.4)$$

For a low-pass system with non-minimum phase behaviour the modulus of this phase value is always larger than that given by Eq. (4.5.4). In both cases the magnitude response has a slope of

$$-20(n - m) \text{ dB/decade}$$

for $\omega \rightarrow \infty$.

4.6 Systems with dead time

A typical system with non-minimum phase behaviour is the *dead time element* (PT_t element), which is described by the transfer function

$$G(s) = e^{-sT_t} . \quad (4.6.1)$$

The frequency response is

$$G(j\omega) = e^{-j\omega T_t}$$

with the amplitude response

$$A(\omega) = |G(j\omega)| = 1$$

and the phase response (in radians)

$$\varphi(\omega) = -\omega T_t .$$

The locus of $G(j\omega)$ is a circle around the origin, starting with $\omega = 0$ on the real axis at $R(\omega) = 1$ and perpetually traversing the circle with increasing values of ω as shown in Figure 4.6.1.

DYNAST study example 4.3

Explanation of plotting frequency responses in DYNAST by way of an example

Problem 4.1

Frequency response - seven problems

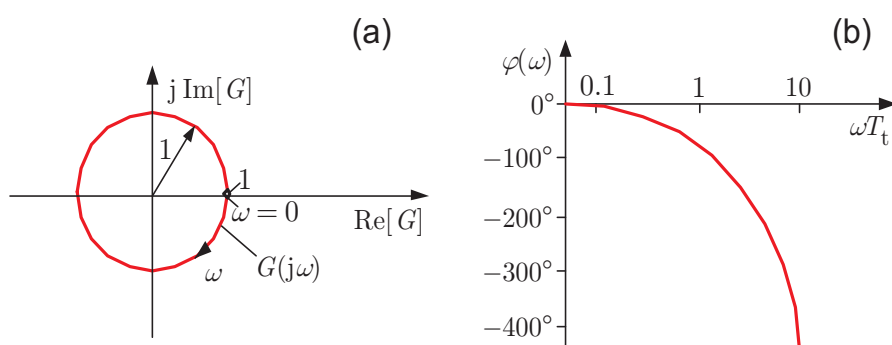


Figure 4.6.1: (a) Nyquist plot and (b) phase of a dead time element

Module 5

Stability of linear control systems

Module units

5.1	Stable and unstable systems	5-2
5.2	Definition of stability and stability conditions	5-2
5.3	Algebraic stability criteria	5-3
5.3.1	The Hurwitz criterion	5-3
5.3.2	Routh criterion	5-6
5.3.3	Nyquist criterion	5-7
5.3.4	Nyquist criterion using Nyquist plots	5-8
5.3.5	Simplified forms of the Nyquist criterion	5-10
5.3.6	The Nyquist criterion using Bode plots	5-11

Module overview. *Stability is the base requirement for the design of a control system. This module gives a first insight into the design of closed-loop systems and the problem of their stability. It is shown by definitions what stability means. Then the most important stability criteria using the characteristic polynomial (Hurwitz, Routh) are introduced. Most emphasis is put on the Nyquist criterion, which can be used with Nyquist and Bode diagrams to design stable closed-loop systems with given stability margins. To simplify the stability test in the frequency domain with Nyquist or Bode diagrams several rules are given. All terms, techniques and rules are illustrated by examples.*

Module objectives. *When you have completed this module you should be able to:*

1. Understand the stability of linear dynamical systems.
2. Understand the algebraic stability criteria for linear systems.
3. Know how to test the stability of linear systems described by transfer functions.
4. Know how to test the stability of linear systems described by frequency-response characteristics.
5. Know how to test the stability of a closed loop from open-loop data.

Module prerequisites. *Transfer function, characteristic polynomial, determinants, frequency response, Bode diagram, Nyquist diagram.*

5.1 Stable and unstable systems

The stability of control systems is an important property. Considering any bounded input signal $x_e(t)$ of a system, and if the output signal $x_a(t)$ of the system to such a signal is also bounded, then the system is called *bounded-input-bounded-output stable*. If the output signal does not show this property, the system is unstable. For illustration see Figure 5.1.1.

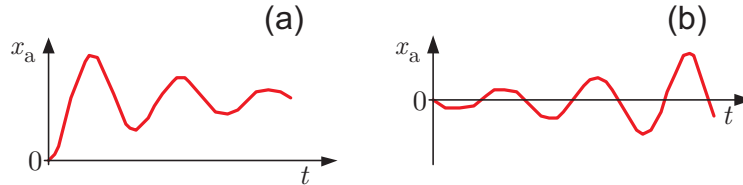


Figure 5.1.1: (a) Stable and (b) unstable system response $x_a(t)$ to a bounded input signal $x_e(t)$

5.2 Definition of stability and stability conditions

Because of its feedback structure a control system can become unstable, e.g. oscillations with increasing amplitudes in the signals can occur. In section 5.1 a signal-based definition of stability is established, which relies on the boundedness of the input-output signals. In this section we focus on a definition of stability for linear systems that is independent of the input-output signals. First the following definition is introduced:

A linear time-invariant system according to Eq. (3.1.3) is called (*asymptotically*) *stable*, if its weighting function decays to zero, i.e. if

$$\lim_{t \rightarrow \infty} g(t) = 0 \quad (5.2.1)$$

is valid. If the modulus of the weighting function increases with increasing t to infinity, the system is called *unstable*.

A special case is a system where the modulus of the weighting function does not exceed a finite value as $t \rightarrow \infty$ or for which it approaches a finite value. Such systems are called *critically stable*. Examples are undamped PT₂S and I elements, see sections 4.4.2 and 4.4.7.

This definition shows that stability is a *system property* for linear systems. If Eq. (5.2.1) is valid, then there exists no initial condition and no bounded input signal which drives the output to infinity. This definition can be directly applied to the stability analysis of linear systems by determining the value of the weighting function for $t \rightarrow \infty$. If this value exists, and if it is zero, the system is stable. However, in most cases the weighting function is not given in an explicit analytic form and therefore it is costly to determine the final value. The transfer function $G(s)$ of a system is often known and as it is the Laplace transform of the weighting function $g(t)$, there is an equivalent stability condition for $G(s)$ according to Eq. (5.2.1). The analysis of this condition – see section A.5 – shows that for the stability analysis it is sufficient to check the poles of the transfer function $G(s)$ of the system, that is the roots s_i of its characteristic equation

$$P(s) \equiv D(s) = a_0 + a_1s + s_2s^2 + \dots + a_ns^n = 0 . \quad (5.2.2)$$

Now the following necessary and sufficient *stability conditions* can be formulated:

a) *Asymptotic stability*

A linear system is only asymptotically stable, if for the roots s_i of its characteristic equation

$$\operatorname{Re} s_i < 0 \quad \text{for all} \quad s_i (i = 1, 2, \dots, n)$$

is valid, or in other words, if *all* poles of its transfer function lie in the left-half s plane.

b) *Instability*

A linear system is only unstable, if at least one pole of its transfer function lies in the right-half s plane, or, if at least one multiple pole (multiplicity $r \geq 2$) is on the imaginary axis of the s plane.

c) *Critical stability*

A linear system is critically stable, if at least one *single* pole exists on the imaginary axis, no pole of the transfer function lies in the right-half s plane, and in addition no multiple poles lie on the imaginary axis.

It has been shown above that the stability of linear systems can be assessed by the distribution of the roots of the characteristic equation in the s plane (Figure 5.2.1). For control problems there is often no need to know these roots with high precision. For a stability analysis it is interesting to know whether all roots of the characteristic equation lie in the left-half s plane or not. Therefore simple criteria are available for easily checking stability, called *stability criteria*. These are partly in algebraic, partly in graphical form.

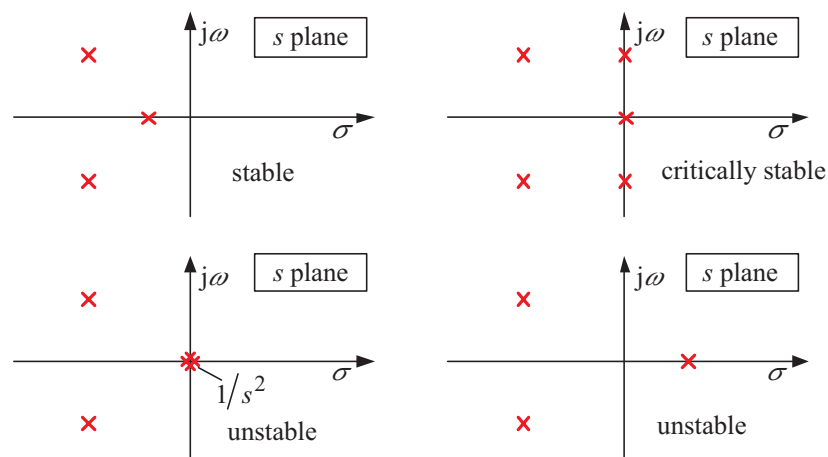


Figure 5.2.1: Stability of a linear system discussed by the distribution of the roots of the characteristic equation in the s plane

Interactive Questions 5.1
[Test your knowledge about stability](#)

5.3 Algebraic stability criteria

The algebraic stability criteria are based on the characteristic equation, Eq. (5.2.2), of the system to be analysed. They contain algebraic conditions as inequalities between coefficients a_i , which are only valid if all roots of the polynomial lie in the left-half s plane.

5.3.1 The Hurwitz criterion

A polynomial

$$P(s) = a_0 + a_1s + \dots + a_n s^n \tag{5.3.1}$$

with k complex conjugate pairs of roots and $(n - 2k)$ real roots can always be represented as

$$P(s) = a_n(s^2 - 2d_1 + d_1^2 + \omega_1^2) \dots (s^2 - 2d_k + d_k^2 + \omega_k^2)(s - s_{2k+1}) \dots (s - s_n) . \tag{5.3.2}$$

If all roots of the polynomial of $P(s)$ are in the left-half s plane then for $a_n > 0$ all constants $-d_j$ and $-s_j$ in Eq. (5.3.2) are positive. From this follows that all coefficients a_j of the polynomial $P(s)$, which are products and sums of positive numbers, are also positive. This result is formulated in the so-called *Stodola criterion*:

For the polynomial to have all roots with negative real parts it is necessary that

$$\text{sgn } a_0 = \text{sgn } a_1 = \dots = \text{sgn } a_n . \tag{5.3.3}$$

These conditions are also sufficient for $n = 1$ and $n = 2$ as can be easily verified by calculating the roots. However, for $n \geq 3$ this is no longer the case.

Example 5.3.1

The polynomial with positive coefficients

$$P(s) = s^3 + s^2 + 4s + 30 = (s + 3)(s^2 - 2s + 10)$$

fulfills the Stodola criterion, but not all the roots $s_1 = -1, s_{2,3} = 1 \pm j3$ have negative real parts. ■

A polynomial for which all roots $s_i (i = 1, 2, \dots, n)$ have negative real parts is called Hurwitzian. Therefore, according to the stability conditions introduced in section 5.2 a linear system is only asymptotically stable, if its characteristic polynomial is Hurwitzian. The Hurwitz criterion for the coefficients of a Hurwitz polynomial is as follows:

A polynomial $P(s)$ is Hurwitzian, if and only if for $a_n > 0$ all determinants

$$\begin{aligned} D_1 &= a_{n-1} > 0 \\ D_2 &= \begin{vmatrix} a_{n-1} & a_n \\ a_{n-3} & a_{n-2} \end{vmatrix} > 0 \\ D_3 &= \begin{vmatrix} a_{n-1} & a_n & 0 \\ a_{n-3} & a_{n-2} & a_{n-1} \\ a_{n-5} & a_{n-4} & a_{n-3} \end{vmatrix} > 0 \end{aligned}$$

until

$$\begin{aligned} D_{n-1} &= \begin{vmatrix} a_{n-1} & a_n & \dots & 0 \\ a_{n-3} & a_{n-2} & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & a_1 \end{vmatrix} > 0 \\ D_n &= a_0 D_{n-1} > 0 . \end{aligned} \tag{5.3.4}$$

are positive.

The following schema of the coefficients can be used to build the Hurwitz determinants:

$$\begin{array}{cccccc} D_1 & \begin{array}{|c|} \hline a_{n-1} \\ \hline \end{array} & \begin{array}{|c|} \hline a_n \\ \hline \end{array} & \begin{array}{|c|} \hline 0 \\ \hline \end{array} & \begin{array}{|c|} \hline 0 \\ \hline \end{array} & \begin{array}{|c|} \hline 0 \\ \hline \end{array} \\ D_2 & \begin{array}{|c|c|} \hline a_{n-3} & a_{n-2} \\ \hline \end{array} & \begin{array}{|c|} \hline a_{n-1} \\ \hline \end{array} & \begin{array}{|c|} \hline a_n \\ \hline \end{array} & \begin{array}{|c|} \hline 0 \\ \hline \end{array} & \\ D_3 & \begin{array}{|c|c|c|} \hline a_{n-5} & a_{n-4} & a_{n-3} \\ \hline \end{array} & \begin{array}{|c|} \hline a_{n-2} \\ \hline \end{array} & \begin{array}{|c|} \hline a_{n-1} \\ \hline \end{array} & \begin{array}{|c|} \hline \dots \\ \hline \end{array} \\ D_4 & \begin{array}{|c|c|c|c|} \hline a_{n-7} & a_{n-6} & a_{n-5} & a_{n-4} \\ \hline \end{array} & \begin{array}{|c|} \hline a_{n-3} \\ \hline \end{array} & \begin{array}{|c|} \hline \dots \\ \hline \end{array} & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \end{array}$$

The Hurwitz determinants D_ν are characterised by the diagonal coefficients $a_{n-1}, a_{n-2}, \dots, a_{n-\nu}$ ($\nu = 1, 2, \dots, n$) and by the increasing indices from left to right. Coefficients with indices larger than n are set to zero. For applying this criterion all determinants until D_{n-1} have to be calculated. Calculation of the last determinant D_n is trivial.

While for a 2nd-order system the conditions of the determinants are automatically fulfilled as soon as the coefficients a_0, a_1, a_2 are positive, for a 3rd-order system one obtains the Hurwitz conditions

$$\begin{aligned} D_1 &= a_2 > 0 \\ D_2 &= \begin{vmatrix} a_2 & a_3 \\ a_0 & a_1 \end{vmatrix} = a_1 a_2 - a_0 a_3 > 0 \\ D_3 &= \begin{vmatrix} a_2 & a_3 & 0 \\ a_0 & a_1 & a_2 \\ 0 & 0 & a_0 \end{vmatrix} = a_0 D_2 > 0 . \end{aligned}$$

It goes without saying that the determinant conditions will be only applied if the easily checkable conditions of Eq. (5.3.3) are fulfilled. The Hurwitz criterion is not only practical for the stability analysis of a system with given coefficients a_i , but also of a system with free parameters. This is the task when the range of parameters must be determined for which the system is asymptotically stable. Therefore the following example is given.

Example 5.3.2

Figure 5.3.1 shows a control loop, for which the range of K_0 must be determined such that the closed loop is asymptotically stable. The time constants T_1 and T_2 of both lag elements are known and positive.

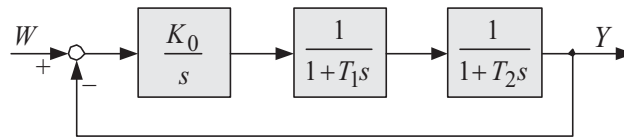


Figure 5.3.1: Stability analysis of a simple control loop

With the transfer function of the open loop

$$\begin{aligned} G_0(s) &= \frac{K_0}{s(1+T_1s)(1+T_2s)} \\ &= \frac{K_0}{s + (T_1+T_2)s^2 + T_1T_2s^3} \end{aligned}$$

one obtains for the closed-loop transfer function

$$G_W(s) = \frac{Y(s)}{W(s)} = \frac{G_0(s)}{1+G_0(s)}$$

and by substituting $G_0(s)$

$$G_W(s) = \frac{K_0}{K_0 + s + (T_1+T_2)s^2 + T_1T_2s^3} .$$

The characteristic equation of the closed loop is

$$P(s) = K_0 + s + (T_1+T_2)s^2 + T_1T_2s^3 = 0 .$$

According to the Stodola and Hurwitz criteria the following conditions must be met for asymptotic stability:

- a) Coefficients $a_0 = K_0$, $a_1 = 1$, $a_2 = (T_1+T_2)$ and $a_3 = T_1T_2$ must be positive. From this the lower limit $K_0 > 0$ follows.

- b) Furthermore

$$(a_1 a_2 - a_3 a_0) > 0$$

must be valid.

With the coefficients given above it follows that

$$T_1 + T_2 - T_1 T_2 K_0 > 0$$

and for the upper limit of K_0

$$K_0 < \frac{T_1 + T_2}{T_1 T_2} .$$

The closed loop is asymptotically stable for the range

$$0 < K_0 < \frac{T_1 + T_2}{T_1 T_2} .$$

■

5.3.2 Routh criterion

For given coefficients a_i of the *characteristic equation* the method of Routh, which is an alternative to the method of Hurwitz, can be applied, see section A.6. Here the coefficients $a_i (i = 0, 1, \dots, n)$ will be arranged in the first two rows of the Routh schema, which contains $n + 1$ rows:

$$\begin{array}{r|cccccc}
 n & a_n & a_{n-2} & a_{n-4} & a_{n-6} & \dots & 0 \\
 n-1 & a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \dots & 0 \\
 \hline
 n-2 & b_{n-1} & b_{n-2} & b_{n-3} & b_{n-4} & \dots & 0 \\
 n-3 & c_{n-1} & c_{n-2} & c_{n-3} & c_{n-4} & \dots & 0 \\
 \vdots & \vdots & & & & & \\
 3 & d_{n-1} & d_{n-2} & 0 & & & \\
 2 & e_{n-1} & e_{n-2} & 0 & & & \\
 1 & f_{n-1} & & & & & \\
 0 & g_{n-1} & & & & &
 \end{array}$$

The coefficients $b_{n-1}, b_{n-2}, b_{n-3}, \dots$ in the third row are the results from cross multiplication the first two rows according to

$$\begin{aligned}
 b_{n-1} &= \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}} \\
 b_{n-2} &= \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}} \\
 b_{n-3} &= \frac{a_{n-1}a_{n-6} - a_n a_{n-7}}{a_{n-1}} \\
 &\vdots
 \end{aligned}$$

Building the cross products one starts with the elements of the first row. The calculation of these b values will be continued until all remaining elements become zero. The calculation of the c values are performed accordingly from the two rows above as follows:

$$\begin{aligned}
 c_{n-1} &= \frac{b_{n-1}a_{n-3} - a_{n-1}b_{n-2}}{b_{n-1}} \\
 c_{n-2} &= \frac{b_{n-1}a_{n-5} - a_{n-1}b_{n-3}}{b_{n-1}} \\
 c_{n-3} &= \frac{b_{n-1}a_{n-7} - a_{n-1}b_{n-4}}{b_{n-1}} \\
 &\vdots
 \end{aligned}$$

From these new rows further rows will be built in the same way, where for the last two rows finally

$$f_{n-1} = \frac{e_{n-1}d_{n-2} - d_{n-1}e_{n-2}}{e_{n-1}}$$

and

$$g_{n-1} = e_{n-2}$$

follows. Now the *Routh criterion* is:

A polynomial $P(s)$ is Hurwitzian, if and only if the following three conditions are valid:

- a) all coefficients $a_i (i = 0, 1, \dots, n)$ are positive,
- b) all coefficients b_{n-1}, c_{n-1} in the first column of the Routh schema are positive.

Example 5.3.3

$$P(s) = 240 + 110s + 50s^2 + 30s^3 + 2s^4 + s^5 .$$

The Routh schema is:

5	1	30	110	0
4	2	50	240	0
3	5	-10	0	
2	54	240		
1	-32.22	0		
0	240			

As in the first row of the Routh schema a coefficient is negative the system is unstable. ■

For proving instability it is sufficient to build the Routh schema only until negative or zero value occurs in the first column. In the example given above the schema could have been stopped at the 5th row.

Another interesting property of the Routh scheme says, that the number of roots with positive real parts is equal to the number of changes of sign of the values in the first column.

5.3.3 Nyquist criterion

This graphical method, which was originally developed for the stability analysis of feedback amplifiers, is especially suitable for different control applications. With this method the closed-loop stability analysis is based on the locus of the open-loop frequency response $G_0(j\omega)$. Since only knowledge of the frequency response $G_0(j\omega)$ is necessary, it is a versatile practical approach for the following cases:

- a) For many cases $G_0(j\omega)$ can be determined by series connection of elements whose parameters are known.
- b) Frequency responses of the loop elements determined by experiments or $G_0(j\omega)$ can be considered directly.
- c) Systems with dead time can be investigated.
- d) Using the frequency response characteristic of $G_0(j\omega)$ not only the stability analysis, but also the design of stable control systems can be easily performed.

5.3.4 Nyquist criterion using Nyquist plots

To derive this criterion one starts with the rational transfer function of the *open loop*

$$G_0(s) = \frac{N_0(s)}{D_0(s)} \tag{5.3.5}$$

and makes the following assumptions:

1. The polynomials $N_0(s)$ and $D_0(s)$ are relatively prime.
- 2.

$$\text{degree } N_0(s) = m \leq n = \text{degree } D_0(s) , \tag{5.3.6}$$

which is always valid for realisable systems, see section 3.3.

The poles β_i of the open loop are the roots of the characteristic equation

$$D_0(s) = 0. \tag{5.3.7}$$

For stability analysis just the poles α_i of the *closed loop* are of interest, i.e. the roots of the characteristic equation, which are determined by setting the denominator of the closed-loop transfer function to zero. From this condition

$$1 + G_0(s) = \frac{N_0(s) + D_0(s)}{D_0(s)} = \frac{N_g(s)}{D_0(s)} = 0 \tag{5.3.8a}$$

and

$$P(s) \equiv N_g(s) = N_0(s) + D_0(s) = 0 \tag{5.3.8b}$$

follows. Because of Eq. (5.3.6) $\text{degree } N_g(s) = n$ is valid. Thus the function $G'(s) = 1 + G_0(s)$ must be investigated in more detail. The zeros of this function match the poles of the closed loop and its poles match the poles of the open loop. Therefore this function can be represented by

$$G'(s) = 1 + G_0(s) = k'_0 \frac{\prod_{i=1}^n (s - \alpha_i)}{\prod_{i=1}^n (s - \beta_i)} , \tag{5.3.9}$$

where α_i are the poles of the closed loop and β_i the poles of the open loop. With respect to the position of the poles it is assumed according to Figure 5.3.2 that

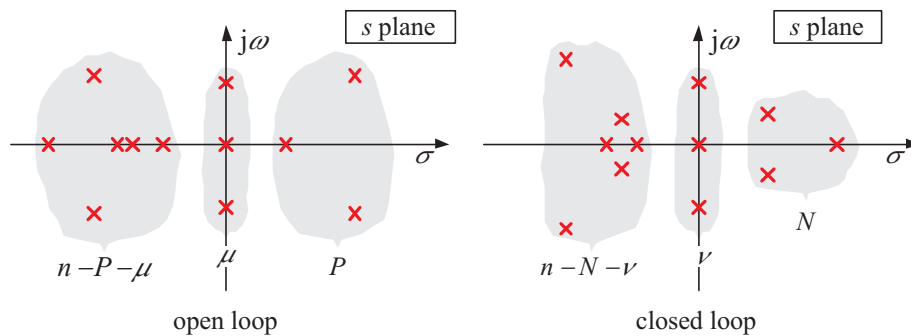


Figure 5.3.2: Poles of the open and closed loop in the s plane (multiple poles are counted according to their multiplicity)

- a) from the n poles α_i of the closed loop
 - N are lying in the right-half s plane,
 - ν are on the imaginary axis, and

$(n - N - \nu)$ in the left-half s plane.

Accordingly,

b) from the n open-loop poles β_i

P are lying in the right-half s plane,

μ on the imaginary axis, and

$(n - P - \mu)$ in the left-half s plane.

P and μ are assumed to be known. Then N and ν will be determined from the knowledge about the frequency response locus of $G_0(j\omega)$. Therefore with $s = j\omega$ the frequency response

$$G'(j\omega) = 1 + G_0(j\omega) = \frac{N_g(j\omega)}{D_0(j\omega)} \quad (5.3.10)$$

is calculated, for which the phase response is given by

$$\varphi(\omega) = \arg[G'(j\omega)] = \arg[N_g(j\omega)] - \arg[D_0(j\omega)] .$$

When ω traverses the range of $0 \leq \omega \leq \infty$, the change in the phase $\Delta\varphi = \varphi(\infty) - \varphi(0)$ consists of the parts from the polynomials $N_g(j\omega)$ and $D_0(j\omega)$ and is given by

$$\Delta\varphi = \Delta\varphi_g - \Delta\varphi_0 .$$

Each root of the polynomials $N_g(s)$ and $D_0(s)$, respectively, provides for $\Delta\varphi_g$ and $\Delta\varphi_0$, respectively, an amount of $+\pi/2$, if they lie in the left-half s plane, and each root on the right-hand side of the imaginary axis provides an amount of $-\pi/2$. These phase changes are continuous with respect to ω .

Each root $j\delta$ on the imaginary axis for $\delta > 0$ causes during the traverse of $j\omega$ at $j\delta$ a stepwise change of π in the phase. This discontinuous part of the phase will not be considered in what follows.

Using the terms given above, for the continuous part $\Delta\varphi_S$ of the phase change $\Delta\varphi$ one obtains

$$\begin{aligned} \Delta\varphi_S &= [(n - N - \nu) - N] \pi/2 - [(n - P - \mu) - P] \pi/2 \\ &= (n - 2N - \nu) \pi/2 - (n - 2P - \mu) \pi/2 , \end{aligned}$$

or

$$\Delta\varphi_S = [2(P - N) + \mu - \nu] \pi/2 . \quad (5.3.11)$$

If besides P and μ , $\Delta\varphi_S$ is also known, then from Eq. (5.3.11) it can be determined, whether $N > 0$ or/and $\nu > 0$ is valid, i.e. whether and how many closed-loop poles are in the right-half s plane and on the imaginary axis.

To determine $\Delta\varphi_S$, the locus $G'(j\omega) = 1 + G_0(j\omega)$ can be drawn on the Nyquist diagram and the phase angle checked. Expediently one moves this curve by 1 to the left in the $G_0(j\omega)$ plane. Thus for stability analysis of the closed loop the locus $G_0(j\omega)$ of the open loop according to Figure 5.3.3 has to be drawn. Here $\Delta\varphi_S$ is the continuous change in the angle of the vector from the so called *critical point* $(-1, j0)$ to the moving point on the locus of $G_0(j\omega)$ for $0 \leq \omega \leq \infty$. Points where the locus passes through the point $(-1, j0)$ or where it has points at infinity correspond to the zeros and poles of $G'(s)$ on the imaginary axis, respectively. These discontinuities are not taken into account for the derivation of Eq. (5.3.11). Figure 5.3.4 shows an example of a $G_0(j\omega)$ where two discontinuous changes of the angle occur. Thereby the continuous change of the angle consists of three parts

$$\begin{aligned} \Delta\varphi_S &= \Delta\varphi_{AB} + \Delta\varphi_{CD} + \Delta\varphi_{DO} \\ &= -\varphi_1 - (2\pi - \varphi_1 - \varphi_2) - \varphi_2 = -2\pi . \end{aligned}$$

The rotation is counter clockwise positive.

As the closed loop is only asymptotically stable for $N = \nu = 0$, then from Eq. (5.3.11) the *general case of the Nyquist criterion* follows:

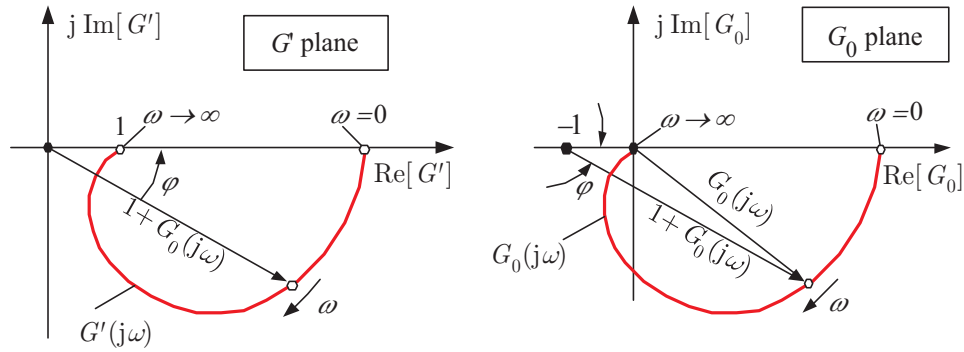


Figure 5.3.3: Nyquist diagrams of $G'(j\omega)$ and $G_0(j\omega)$

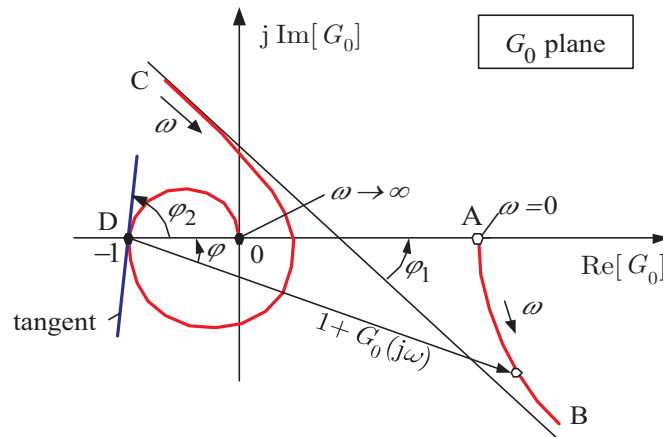


Figure 5.3.4: Determination of continuous changes in the angle $\Delta\varphi_S$

The closed loop is asymptotically stable, if and only if the continuous change in the angle of the vector from the critical point $(-1, j0)$ to the moving point of the locus $G_0(j\omega)$ of the open loop is

$$\Delta\varphi_S = (P + \mu/2)\pi . \tag{5.3.12}$$

For the case with a *negative* gain K_0 of the open loop the locus is rotated by 180° relative to the case with a positive K_0 . The Nyquist criterion remains valid also in the case of a *dead time* in the open loop.

5.3.5 Simplified forms of the Nyquist criterion

It follows from Eq. (5.3.12) that for an open-loop stable system, that is $P = 0$ and $\mu = 0$, then $\Delta\varphi_S = 0$. Therefore the Nyquist criterion can be reformulated as follows:

If the open loop is asymptotically stable, then the closed loop is only asymptotically stable, if the frequency response locus of the open loop does neither revolve around or pass through the critical point $(-1, j0)$.

Another form of the simplified Nyquist criterion for $G_0(s)$ with poles at $s = 0$ is the so called 'left-hand rule':

The open loop has only poles in the left-half s plane with the exception of a single or double pole at $s = 0$ (P, I or I_2 behaviour). In this case the closed loop is only stable, if the critical point $(-1, j0)$ is on the *left* hand-side of the locus $G_0(j\omega)$ in the direction of increasing values of ω .

This form of the Nyquist criterion is sufficient for most cases. The part of the locus that is significant is that closest to the critical point. For very complicated curves one should go back to the general case. The left-hand rule can be graphically derived from the generalised locus according to section A.2. The orthogonal (σ, ω) -net is observed and asymptotic stability of the closed loop is given, if a curve with $\sigma < 0$ passes through the critical point $(-1, j0)$. Such a curve is always on the left-hand side of $G_0(j\omega)$.

5.3.6 The Nyquist criterion using Bode plots

Because of the simplicity of the graphical construction of the frequency response characteristics of a given transfer function the application of the Nyquist criterion is often more simple using Bode plots. The continuous change of the angle $\Delta\varphi_S$ of the vector from the critical point $(-1, j0)$ to the locus of $G_0(j\omega)$ must be expressed by the amplitude and phase response of $G_0(j\omega)$. From Figure 5.3.5 it can be seen that

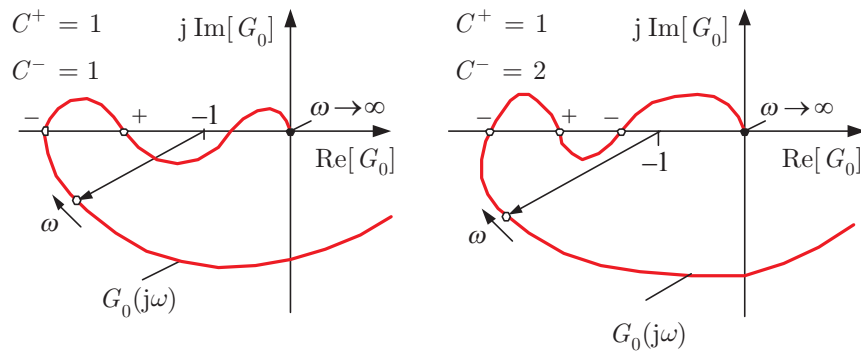


Figure 5.3.5: Positive (+) and negative (-) intersections of the locus $G_0(j\omega)$ with the real axis on the left-hand side of the critical point

this change of the angle is directly related to the count of intersections of the locus with the real axis on the left-hand side of the critical point between $(-\infty, -1)$. The Nyquist criterion can therefore also be represented by the count of these intersections if the gain of the open loop is positive.

Regarding the intersections of the locus of $G_0(j\omega)$ with the real axis in the range $(-\infty, -1)$, the transfer from the upper to the lower half plane in the direction of increasing ω values are treated as *positive intersections* while the reverse transfer are *negative intersections* (Figure 5.3.5). The change of the angle is zero if the count of positive intersections S^+ is equal to the count of negative intersections S^- . The change of the angle $\Delta\varphi_S$ depends also on the number of positive and negative intersections and if the open loop does not have poles on the imaginary axis, the change of the angle is

$$\Delta\varphi_S = 2\pi(C^+ - C^-) .$$

In the case of an open loop containing an integrator, i.e. a single pole in the origin of the complex plane ($\mu = 1$), the locus starts for $\omega = 0$ at $\delta - j\infty$, where an additional $+\pi/2$ is added to the change of the angle. For proportional and integral behaviour of the open loop

$$\Delta\varphi_S = 2\pi(C^+ - C^-) + \mu\pi/2 \quad \mu = 0, 1 \quad (5.3.13)$$

is valid. In principle this relation is also valid for $\mu = 2$, but the locus starts for $\omega = 0$ at $-\infty + j\delta$ (Figure 5.3.6), and this intersection would be counted as a negative one if $\delta > 0$, i.e. if the locus for small ω is in the upper half plane of the real axis. But de facto there is for $\delta > 0$ (and accordingly $\delta < 0$) no intersection. This follows from the detailed investigation of the discontinuous change of the angle, which occurs at $\omega = 0$. As only a continuous change of the angle is taken into account and because of reason of symmetry the start of the locus at $\omega = 0$ is counted as a half intersection, positive for $\delta < 0$ and negative for $\delta > 0$, which is analogous to the definition given above (Figure 5.3.6). For continuous changes of the angle

$$\Delta\varphi_S = 2\pi(C^+ - C^-) \quad (\mu = 2) \quad (5.3.14)$$

is valid. Comparing Eqs. (5.3.13) and (5.3.14), respectively, with Eq. (5.3.12) then the Nyquist criterion can be formulated as:

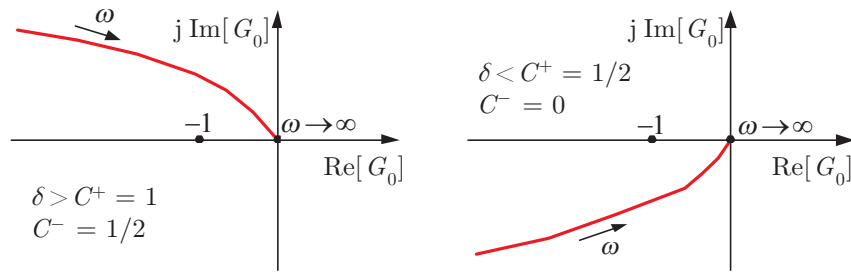


Figure 5.3.6: Count of the intersections on the left-hand side of the critical point for I_2 behaviour of the open loop

The open loop with the transfer function $G_0(s)$ has P poles in the left-half s plane and possibly a single ($\mu = 1$) or double pole ($\mu = 2$) at $s = 0$. If the locus of $G_0(j\omega)$ has C^+ positive and C^- negative intersections with the real axis to the left of the critical point, then the closed loop is only asymptotically stable, if

$$D^* = C^+ - C^- = \begin{cases} \frac{P}{2} & \text{for } \mu = 0, 1 \\ \frac{P+1}{2} & \text{for } \mu = 2 \end{cases} \quad (5.3.15)$$

is valid. For the special case, that the open loop is stable ($P = 0, \mu = 0$), the number of positive and negative intersections must be equal.

From this it follows that the difference of the number of positive and negative intersections in the case of $\mu = 0, 1$ is an integer and for $\mu = 2$ not an integer. From this follows immediately, that for $\mu = 0, 1$ the number P is even, for $\mu = 2$ the number $P + 1$ is uneven and therefore in all cases P is an even number, such that the closed loop is asymptotically stable. This is only valid if $D^* \geq 1$.

The Nyquist criterion can now be transferred directly into the representation using frequency response characteristics. The magnitude response $A_0(\omega)_{dB}$, which corresponds to the locus $G_0(j\omega)$, is always positive at the intersections of the locus with the real axis in the range of $(-\infty, -1)$. These points of intersection correspond to the crossings of the phase response $\varphi_0(\omega)$ with lines $\pm 180^\circ, \pm 540^\circ$ etc., i.e. a uneven multiple of 180° . In the case of a positive intersection of the locus, the phase response at the $\pm(2k + 1) 180^\circ$ lines crosses from below to top and reverse from top to below on a negative intersection as shown in Figure 5.3.7. In the following these crossings will be defined as positive (+) and negative (-)

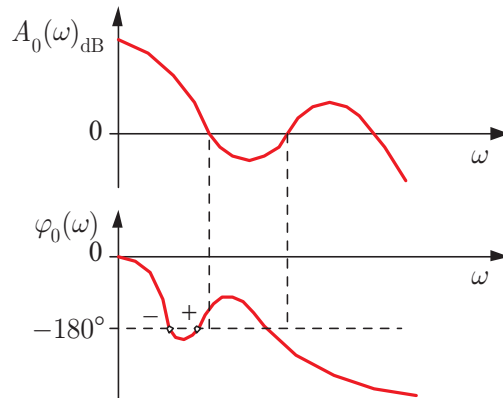


Figure 5.3.7: Frequency response characteristics of $G_0(j\omega) = A_0(\omega) e^{j\varphi_0(\omega)}$ and definition of positive (+) and negative (-) crossings of the phase response $\varphi_0(\omega)$ with the -180° line

(-) crossings of the phase response $\varphi_0(\omega)$ over the particular $\pm(2k + 1) 180^\circ$ lines, where $k = 0, 1, 2, \dots$ may be valid. If the phase response starts at -180° this point is counted as a half crossing with the corresponding sign. Based on the discussions above the Nyquist criterion can be formulated in a form suitable for frequency response characteristics:

The open loop with the transfer function $G_0(s)$ has P poles in the right-half s plane, and possibly a single or double pole at $s = 0$. C^+ are the number of positive and C^- of negative crossings of the phase response $\varphi_0(\omega)$ over the $\pm(2k + 1) 180^\circ$ lines in the frequency range where $A_0(\omega)_{dB} > 0$ is valid. The closed loop is only asymptotically stable, if

$$D^* = C^+ - C^- = \begin{cases} \frac{P}{2} & \text{for } \mu = 0, 1 \\ \frac{P+1}{2} & \text{for } \mu = 2 \end{cases}$$

is valid. For the special case of an open-loop stable system ($P = 0, \mu = 0$)

$$D^* = C^+ - C^- = 0$$

must be valid.

Table 5.3.1 shows some examples of the Nyquist criterion in the representation using frequency response characteristics.

Finally the 'left-hand rule' will be given using Bode diagrams, because this version is for the most cases sufficient and simple to apply.

The open loop has only poles in the left-half s plane with the exception of possibly one single or one multiple pole at $s = 0$ (P, I or I_2 behaviour). In this case the closed loop is only asymptotically stable, if $G_0(j\omega)$ has a phase of $\varphi_0 > -180^\circ$ for the *crossover frequency* ω_C at $A_0(\omega_C)_{dB} = 0$.

This stability criterion offers the possibility of a practical assessment of the 'quality of stability' of a control loop. The larger the distance of the locus from the critical point the farther is the closed loop from the stability margin. As a measure of this distance the terms gain margin and phase margin are introduced according to Figure 5.3.8. The *phase margin*

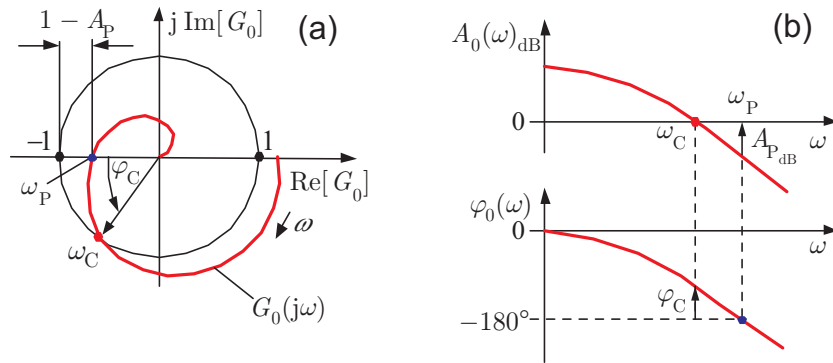


Figure 5.3.8: Phase and gain margin φ_C and A_P or $A_{P_{dB}}$, respectively, in the (a) Nyquist diagram and (b) Bode diagram

$$\varphi_C = 180^\circ + \varphi_0(\omega_C) \tag{5.3.16}$$

is the distance of the phase response from the -180° line at the crossover frequency ω_C , i.e. at the crossing of the magnitude response with the 0dB line ($|G_0| = 1$). The *gain margin*

$$A_{P_{dB}} = A_0(\omega_P)_{dB} \tag{5.3.17}$$

is the distance of the magnitude response from the 0dB line at the phase of $\varphi_0 = -180^\circ$.

A well damped control system should yield the following characteristics:

$$A_{P_{dB}} = \begin{cases} -12 \text{ dB} & \text{to } -20 \text{ dB} & \text{for command response} \\ -3.5 \text{ dB} & \text{to } -9.5 \text{ dB} & \text{for disturbance response} \end{cases}$$

$$\varphi_C = \begin{cases} 40^\circ & \text{to } 60^\circ & \text{for command response} \\ 20^\circ & \text{to } 50^\circ & \text{for disturbance response} . \end{cases}$$

Table 5.3.1: Examples of stability analysis using the Nyquist criterion with frequency response characteristics

No.	Bode Diagram	Stability Analysis
1		$\Rightarrow \left. \begin{array}{l} S^+ = 1 \\ S^- = 2 \\ D^* = -1 \\ P = 2 \end{array} \right\} \Rightarrow D^* \neq P/2: \text{unstable}$
2		$\Rightarrow \left. \begin{array}{l} S^+ = 3/2 \\ S^- = 1 \\ D^* = 1/2 \\ P = 0 \end{array} \right\} \Rightarrow D^* = \frac{P+1}{2}: \text{stable if } 2 \text{ poles in the origin}$
3		$\Rightarrow \left. \begin{array}{l} S^+ = 0 \\ S^- = 1 \\ D^* = -1 \\ P = 0 \end{array} \right\} \Rightarrow D^* \neq P/2: \text{unstable}$
4		$\Rightarrow \left. \begin{array}{l} S^+ = 0 \\ S^- = 0 \\ D^* = 0 \\ P = 0 \end{array} \right\} \Rightarrow D^* = P/2: \text{stable}$

The crossover frequency ω_C is a measure of the dynamical quality of the control loop. The higher ω_C the higher the bandwidth of the closed loop, and the faster the reaction on command inputs or disturbances. As the bandwidth that frequency is understood, at which the magnitude $A(\omega)$ of the closed-loop frequency response has fallen off approximately to zero.

Interactive Questions 5.2

Test your knowledge about gain and phase margin

Problem 5.1

Stability - three questions

Module 6

The root-locus method

Module units

6.1 Introduction and basic ideas	6-1
6.2 General rules for constructing root loci	6-5
6.3 Example of an application of the root-locus method	6-10

Module overview. *Based on the pole and zero distributions of an open-loop system the stability of the closed-loop system can be discussed as a function of one scalar parameter. The root-locus method shown in this module is a technique that can be used as a tool to design control systems. The basic ideas and its relevancy to control system design are introduced and illustrated. Ten general rules for constructing root loci for positive and negative gain are shortly presented such that they can be easily applied. This is demonstrated by some discussed examples, by a table with sixteen examples and by a comprehensive design of a closed-loop system of higher order.*

Module objectives. *When you have completed this module you should be able to:*

1. Understand root-locus diagrams.
2. Draw root-locus diagrams.
3. Apply the root-locus method to design control systems.
4. Discuss the stability of closed-loop systems with respect to variations in one parameter.

Module prerequisites. *Stability, closed loop, transfer function, poles and zeros.*

6.1 Introduction and basic ideas

A designer can determine whether his design for a control system meets the specifications if he knows the desired time response of the controlled variable. By deriving the differential equations for the control system and solving them, an accurate solution of the system's performance can be obtained, but this approach is not feasible for other than simple systems. It is not easy to determine from this solution just what parameters in the system should be changed to improve the response. A designer wishes to be able to predict the performance by an analysis that does not require the actual solution of the differential equations.

The first thing that a designer wants to know about a given system is whether or not it is stable. This can be determined by examining the roots obtained from the characteristic equation

$$1 + G_0(s) = 0 \tag{6.1.1}$$

of the closed loop. The work involved in determining the roots of this equation can be avoided by applying the Hurwitz or Routh criterion as shown in section 5.3. Determining in this way whether the system is stable or unstable does not satisfy the designer, because it does not indicate the degree of stability of the system, i.e., the amount of overshoot and the settling time of the controlled variable for a step input. Not only must the system be stable, but the overshoot must be maintained within prescribed bounds and transients must die out in a sufficiently short time.

The root-locus method described in this section not only indicates whether a system is stable or unstable but, for a stable system, also shows the degree of stability. The root locus is a plot of the roots of the characteristic equation of the closed loop as a function of the gain. This graphical approach yields a clear indication of the effect of gain adjustment with relatively small effort.

With this method one determines the closed-loop poles in the s plane – these are the roots of Eq.(6.1.1) – by using the known distribution of the poles and zeros of the open-loop transfer function $G_0(s)$. If for instance a parameter is varied, the roots of the characteristic equation will move on certain curves in the s plane as shown by the example in Figure 6.1.1. On these curves lie all possible roots of the

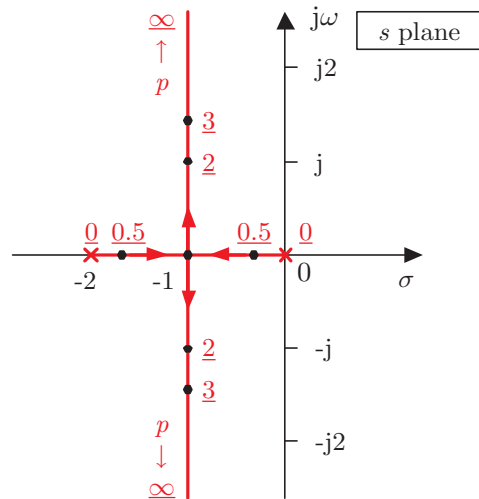


Figure 6.1.1: Plot of all roots of the characteristic equation $s^2 + 2s + p = 0$ for $0 \leq p < \infty$. Values of p are red and underlined.

characteristic equation for all values of the varied parameter from zero to infinity. These curves are defined as the *root-locus plot* of the closed loop. Once this plot is obtained, the roots that best fit the system performance specifications can be selected. Corresponding to the selected roots there is a required value of the parameter which can be determined from the plot. When the roots have been selected, the time response can be obtained. Since the process of finding the root locus by calculating the roots for various values of a parameter becomes tedious, a simpler method of obtaining the root locus is desired. The graphical method for determining the root-locus plot is shown in the following.

An open-loop transfer function with k poles at the origin of the s plane is often described by

$$G_0(s) = \frac{K_0}{s^k} \frac{1 + \beta_1 s + \dots + \beta_m s^m}{1 + \alpha_1 s + \dots + \alpha_{n-k} s^{n-k}} \quad m \leq n, \quad (6.1.2)$$

where K_0 is the gain of the open loop. In order to represent this transfer function in terms of the open-loop poles and zeros it is rewritten as

$$G_0(s) = k_0 \frac{\prod_{\mu=1}^m (s - s_{Z_\mu})}{\prod_{\nu=1}^n (s - s_{P_\nu})} = k_0 G(s) \quad (6.1.3a)$$

or

$$G_0(s) = k_0 \frac{\prod_{\mu=1}^m (-sZ_\mu)}{\prod_{\substack{\nu=1 \\ s_{P_\nu} \neq 0}}^{n-k} (-sP_\nu)} \frac{1}{s^k} \frac{\prod_{\mu=1}^m \left(1 + \frac{s}{-sZ_\mu}\right)}{\prod_{\substack{\nu=1 \\ s_{P_\nu} \neq 0}}^{n-k} \left(1 + \frac{s}{-sP_\nu}\right)} \quad (6.1.3b)$$

with $k_0 > 0$ and $s_{Z_\nu} \neq s_{P_\nu}$. The relationship between the factor k_0 and the open-loop gain K_0 is

$$K_0 = k_0 \frac{\prod_{\mu=1}^m (-sZ_\mu)}{\prod_{\substack{\nu=1 \\ s_{P_\nu} \neq 0}}^{n-k} (-sP_\nu)} \frac{1}{s^k}. \quad (6.1.4)$$

The characteristic equation of the closed loop using Eq. (6.1.3a) is

$$1 + k_0 G(s) = 0 \quad (6.1.5a)$$

or

$$G(s) = -\frac{1}{k_0}. \quad (6.1.5b)$$

All complex numbers $s_i = s_i(k_0)$, which fulfil this condition for $0 \leq k_0 \leq \infty$, represent the root locus.

From the above it can be concluded that the magnitude of $k_0 G(s)$ must always be unity and its phase angle must be an odd multiple of π . Consequently, the following two conditions are formalised for the root locus for all positive values of k_0 from zero to infinity:

a) *Magnitude condition:*

$$|G(s)| = \frac{1}{k_0} \quad (6.1.6)$$

b) *Angle condition*

$$\varphi(s) = \arg G(s) = \pm 180^\circ (2k + 1) \quad \text{for } \begin{array}{l} k = 0, 1, 2, \dots \\ k_0 \geq 0 \end{array} \quad (6.1.7)$$

In a similar manner, the conditions for negative values of k_0 ($-\infty \leq k_0 < 0$) can be determined. The magnitude conditions is the same, but the angle must satisfy the

c) *Angle condition*

$$\varphi(s) = \arg G(s) = \pm k 360^\circ \quad \text{for } \begin{array}{l} k = 0, 1, 2, \dots \\ k_0 < 0 \end{array}. \quad (6.1.8)$$

Apparently the angle condition is independent of k_0 . All points of the s plane that fulfil the angle condition are the loci of the poles of the closed loop by varying k_0 . The calibration of the curves by the values of k_0 is obtained by the magnitude condition according to Eq. (6.1.6). Based upon this interpretation of the conditions the root locus can be constructed in a graphical/numerical way.

Once the open-loop transfer function $G_0(s)$ has been determined and put into the proper form, the poles and zeros of this function are plotted in the s plane.

Example 6.1.1

Consider the example

$$G_0(s) = \frac{K_0}{s(s+2)} = \frac{k_0}{(s-s_{P_1})(s-s_{P_2})}$$

with $s_{P_1} = 0$, $s_{P_2} = -2$ and $k_0 = K_0$. The poles of the closed-loop transfer function

$$G_W(s) = \frac{K_0}{s^2 + 2s + K_0}$$

are the roots s_1 and s_2 of the characteristic equation

$$P(s) = s^2 + 2s + K_0 = 0$$

and are given by

$$s_{1,2} = -1 \pm \sqrt{1 - K_0}.$$

As $s_1 = s_{P_1} = 0$ and $s_2 = s_{P_2} = -2$ it can be seen that for $K_0 = 0$ the poles of the closed loop transfer function are identical with those of the open-loop transfer function $G_0(s)$. For other values K_0 the following two cases are considered:

- $K_0 \leq 1$: Both roots s_1 and s_2 are real and lie on the real axis in the range of $-2 \leq \sigma \leq -1$ and $-1 \leq \sigma \leq 0$;
- $K_0 > 1$: The roots s_1 and s_2 are conjugate complex with the real part $\text{Re } s_{1,2} = -1$, which does not depend on K_0 , and the imaginary part $\text{Im } s_{1,2} = \pm\sqrt{K_0 - 1}$.

The curve has two branches as shown in Figure 6.1.2. At $(s_{P_1} + s_{P_2})/2 = -1$ is the *breakaway point* of

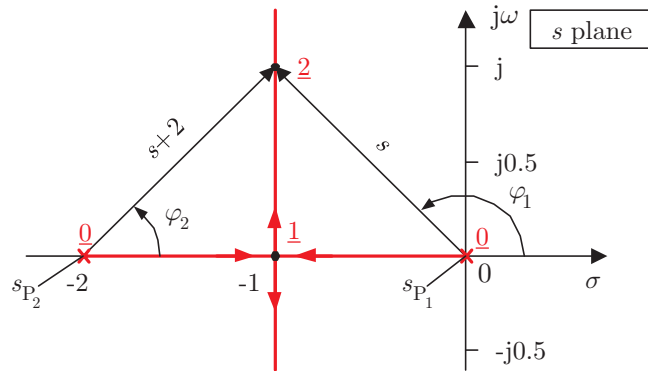


Figure 6.1.2: Root locus of a simple second-order system

the two branches. Checking the angle condition the condition

$$\varphi(s) = \arg\{G(s)\} = \arg\left\{\frac{1}{s(s+2)}\right\} = -\arg s - \arg(s+2) \stackrel{!}{=} \pm 180^\circ(2k+1)$$

must be valid. The complex numbers s and $(s+2)$ have the angles φ_1 and φ_2 and the magnitudes $|s|$ and $|s+2|$. The triangle $(-2, 0, -1+j)$ in Figure 6.1.2 yields the angle condition. Evaluating the magnitude condition according to Eq. (6.1.6)

$$|G(s)| = \left|\frac{1}{s(s+2)}\right| = \frac{1}{K_0}$$

one obtains the value K_0 on the root locus. E.g. for $s = -1 + j$ the gain of the open loop is

$$K_0 = |s(s+2)|_{s=-1+j} = 2.$$

The value of K_0 at the breakaway point $s_B = -1$ is

$$K_0 = |-1(-1+2)| = 1.$$

■

Table 6.1.1 shows further examples of some 1st- and 2nd-order systems.

Table 6.1.1: Root loci of 1st- and 2nd-order systems

$G_0(s)$	root locus	$G_0(s)$	root locus
$\frac{k_0}{s}$		$\frac{k_0}{(s + \sigma_1)^2 + \omega_1^2}$	
$\frac{k_0}{s^2}$		$\frac{k_0}{(s - s_{P_1})(s - s_{P_2})}$	
$\frac{k_0}{s - s_{P_1}}$		$\frac{k_0(s - s_{Z_1})}{(s - s_{P_1})}$ $ s_{Z_1} > s_{P_1} $	
$\frac{k_0}{s^2 + \omega_1^2}$		$\frac{k_0(s - s_{Z_1})}{(s - s_{P_1})}$ $ s_{Z_1} < s_{P_1} $	

6.2 General rules for constructing root loci

To facilitate the application of the root-locus method for systems of higher order than 2nd, rules can be established. These rules are based upon the interpretation of the angle condition and the analysis of the characteristic equation. The rules presented aid in obtaining the root locus by expediting the manual plotting of the locus. But for automatic plotting using a computer these rules provide checkpoints to ensure that the solution is correct.

Though the angle and magnitude conditions can also be applied to systems having dead time, in the following we restrict to the case of the open-loop rational transfer functions according to Eq. (6.1.3a)

$$G_0(s) = k_0 \frac{(s - s_{Z_1})(s - s_{Z_2}) \dots (s - s_{Z_m})}{(s - s_{P_1})(s - s_{P_2}) \dots (s - s_{P_n})}, \quad k_0 \geq 0 \tag{6.2.1a}$$

or

$$G_0(s) = k_0 \frac{b_0 + b_1s + \dots + b_{m-1}s^{m-1} + s^m}{a_0 + a_1s + \dots + a_{n-1}s^{n-1} + s^n} = k_0 \frac{N_0(s)}{D_0(s)}. \tag{6.2.1b}$$

As this transfer function can be written in terms of poles and zeros s_{P_ν} and s_{Z_μ} ($\nu = 1, 2, \dots, n$; $\mu = 1, 2, \dots, m$) $G_0(s)$ can be represented by their magnitudes and angles

$$G_0(s) = k_0 \frac{|s - s_{Z_1}| e^{j\varphi_{Z_1}} |s - s_{Z_2}| e^{j\varphi_{Z_2}} \dots |s - s_{Z_m}| e^{j\varphi_{Z_m}}}{|s - s_{P_1}| e^{j\varphi_{P_1}} |s - s_{P_2}| e^{j\varphi_{P_2}} \dots |s - s_{P_n}| e^{j\varphi_{P_n}}}$$

or

$$G_0(s) = k_0 \frac{\prod_{\mu=1}^m |s - s_{Z_\mu}|}{\prod_{\nu=1}^n |s - s_{P_\nu}|} e^{j \left(\sum_{\mu=1}^m \varphi_{Z_\mu} - \sum_{\nu=1}^n \varphi_{P_\nu} \right)}. \tag{6.2.2}$$

From Eq. (6.1.6) the *magnitude condition*

$$\frac{\prod_{\mu=1}^m |s - s_{Z_\mu}|}{\prod_{\nu=1}^n |s - s_{P_\nu}|} = \frac{1}{k_0} \tag{6.2.3}$$

and from Eq. (6.1.7) the *angle condition*

$$\varphi(s) = \sum_{\mu=1}^m \varphi_{Z_{\mu}} - \sum_{\nu=1}^n \varphi_{P_{\nu}} = \pm 180^{\circ}(2k+1) \quad \text{for } k = 0, 1, 2, \dots \quad (6.2.4)$$

follows. Here $\varphi_{Z_{\mu}}$ and $\varphi_{P_{\nu}}$ denote the angles of the complex values $(s - s_{Z_{\mu}})$ and $(s - s_{P_{\nu}})$, respectively. All angles are considered positive, measured in the counterclockwise sense. If for each point the sum of these angles in the s plane is calculated, just those particular points that fulfil the condition in Eq. (6.2.4) are points on the root locus. This principle of constructing a root-locus curve – as shown in Figure 6.2.1 – is mostly used for automatic root-locus plotting.

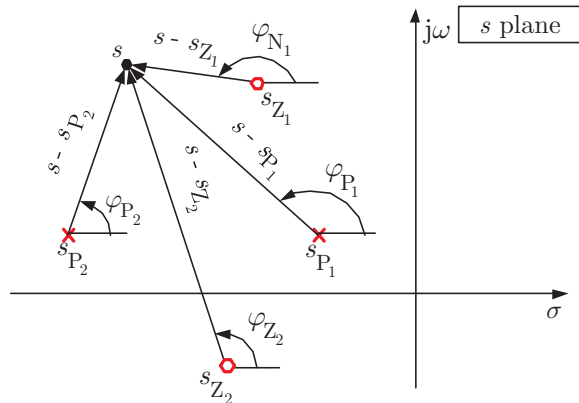


Figure 6.2.1: Pole-zero diagram for construction of the root locus

In the following the most important *rules for the construction of root loci* for $k_0 > 0$ are listed:

Rule 1 Symmetry As all roots are either real or complex conjugate pairs so that the root locus is symmetrical to the real axis.

Rule 2 Number of branches The number of branches of the root locus is equal to the number of poles n of the open-loop transfer function.

Rule 3 Locus start and end points The locus starting points ($k_0 = 0$) are at the open-loop poles and the locus ending points ($k_0 = \infty$) are at the open-loop zeros. $(n - m)$ branches end at infinity. The number of starting branches from a pole and ending branches at a zero is equal to the multiplicity of the poles and zeros, respectively. A point at infinity is considered as an equivalent zero of multiplicity equal to $n - m$.

Rule 4 Real axis locus If the total number of poles and zeros to the right of a point on the real axis is odd, this point lies on the locus.

Rule 5 Asymptotes There are $n - m$ asymptotes of the root locus with a slope of

$$\alpha_k = \arg s = \frac{\pm 180^{\circ}(2k+1)}{n-m} . \quad (6.2.5)$$

For $(n - m) = 1, 2, 3$ and 4 one obtains the asymptote configurations as shown in Figure 6.2.2.

Rule 6 Real axis intercept of the asymptotes The real axis crossing $(\sigma_a, j0)$ of the asymptotes is at

$$\sigma_a = \frac{1}{n-m} \left\{ \sum_{\nu=1}^n \operatorname{Re} s_{P_{\nu}} - \sum_{\mu=1}^m \operatorname{Re} s_{Z_{\mu}} \right\} . \quad (6.2.6)$$

Rule 7 Breakaway and break-in points on the real axis At least one breakaway or break-in point $(\sigma_B, j0)$ exists if a branch of the root locus is on the real axis between two poles or zeros, respectively. Conditions to find such real points are based on the fact that they represent multiple real

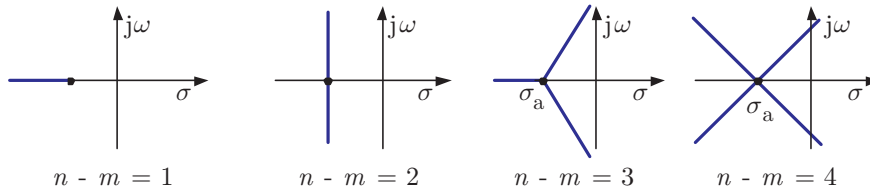


Figure 6.2.2: Asymptote configurations of the root locus

roots. In addition to the characteristic equation (6.1.1) for multiple roots the condition

$$\frac{d}{ds}[1 + G_0(s)] = \frac{d}{ds}G_0(s) = 0. \quad (6.2.7)$$

must be fulfilled, which is equivalent to

$$\sum_{\nu=1}^n \frac{1}{s - s_{P_\nu}} = \sum_{\mu=1}^m \frac{1}{s - s_{Z_\mu}} \quad (6.2.8)$$

for $s = \sigma_B$. If there are no poles or zeros, the corresponding sum is zero.

Rule 8 Complex pole/zero angle of departure/entry The angle of departure of pairs of poles with multiplicity r_{P_ϱ} is

$$\varphi_{P_\varrho, D} = \frac{1}{r_{P_\varrho}} \left\{ - \sum_{\substack{\nu=1 \\ \nu \neq \varrho}}^n \varphi_{P_\nu} + \sum_{\mu=1}^m \varphi_{Z_\mu} \pm 180^\circ(2k+1) \right\} \quad (6.2.9)$$

and the angle of entry of the pairs of zeros with multiplicity r_{Z_ϱ}

$$\varphi_{Z_\varrho, E} = \frac{1}{r_{Z_\varrho}} \left\{ - \sum_{\substack{\mu=1 \\ \mu \neq \varrho}}^m \varphi_{Z_\mu} + \sum_{\nu=1}^n \varphi_{P_\nu} \pm 180^\circ(2k+1) \right\}. \quad (6.2.10)$$

Rule 9 Root-locus calibration The labels of the values of k_0 can be determined by using

$$k_0 = \frac{\prod_{\nu=1}^n |s - s_{P_\nu}|}{\prod_{\mu=1}^m |s - s_{Z_\mu}|}. \quad (6.2.11)$$

For $m = 0$ the denominator is equal to one.

Rule 10 Asymptotic stability The closed loop system is asymptotically stable for all values of k_0 for which the locus lies in the left-half s plane. From the imaginary-axis crossing points the critical values $k_{0, \text{crit}}$ can be determined.

The rules shown above are for positive values of k_0 . According to the angle condition of Eq. (6.1.8) for negative values of k_0 some rules have to be modified. In the following these rules are numbered as above but labelled by a *.

Rule 3* Locus start and end points The locus starting points ($k_0 = 0$) are at the open-loop poles and the locus ending points ($k_0 = -\infty$) are at the open-loop zeros. $(n - m)$ branches end at infinity. The number of starting branches from a pole and ending branches at a zero is equal to the multiplicity of the poles and zeros, respectively. A point at infinity is considered as an equivalent zero of multiplicity equal to $n - m$.

Rule 4* Real axis locus If the total number of poles and zeros to the right of a point on the real axis is even including zero, this point lies on the locus.

Rule 5* Asymptotes There are $n - m$ asymptotes of the root locus with a slope of

$$\alpha_k = \arg s = \frac{\pm 360^\circ k}{n - m} . \quad (6.2.12)$$

Rule 8* Complex pole/zero angle of departure/entry The angle of departure of pairs of poles with multiplicity r_{P_Q} is

$$\varphi_{P_Q,D} = \frac{1}{r_{P_Q}} \left\{ - \sum_{\substack{\nu=1 \\ \nu \neq Q}}^n \varphi_{P_\nu} + \sum_{\mu=1}^m \varphi_{Z_\mu} \pm 360^\circ k \right\} \quad (6.2.13)$$

and the angle of entry of the pairs of zeros with multiplicity r_{Z_Q}

$$\varphi_{Z_Q,E} = \frac{1}{r_{Z_Q}} \left\{ - \sum_{\substack{\mu=1 \\ \mu \neq Q}}^m \varphi_{Z_\mu} + \sum_{\nu=1}^n \varphi_{P_\nu} \pm 360^\circ k \right\} . \quad (6.2.14)$$

The root-locus method can also be applied for other cases than varying k_0 . This is possible as long as $G_0(s)$ can be rewritten such that the angle condition according to Eq. (6.2.4) and the rules given above can be applied. This will be demonstrated in the following two examples.

Example 6.2.1

Given the closed-loop characteristic equation

$$a_0 + a_1 s + \dots + a_{n-1} s^{n-1} + s^n = 0 ,$$

the root locus for varying the parameter a_1 is required. The characteristic equation is therefore rewritten as

$$1 + a_1 \frac{s}{a_0 + a_2 s^2 + \dots + s^n} = 0 .$$

This form then corresponds to the standard form

$$1 + G_0(s) = 1 + a_1 \frac{N_0(s)}{D_0(s)} = 0$$

to which the rules can be applied. ■

Example 6.2.2

Given the closed-loop characteristic equation

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

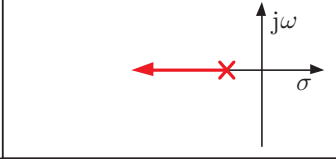
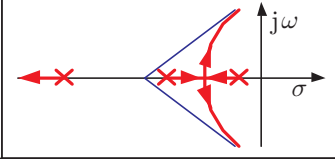
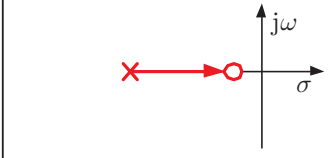
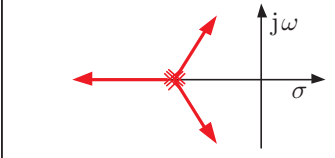
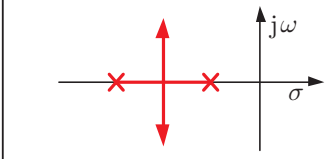
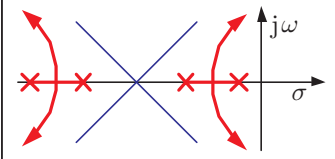
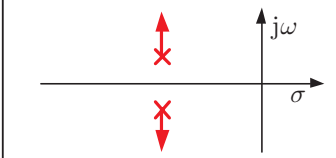
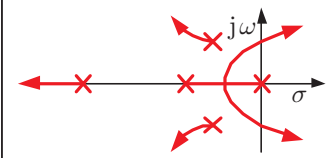
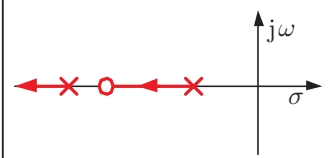
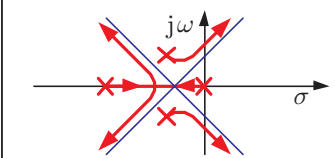
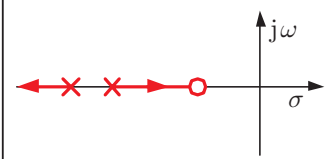
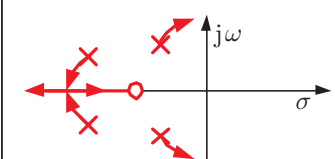
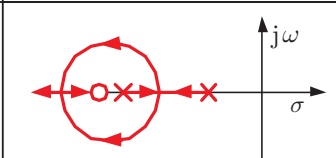
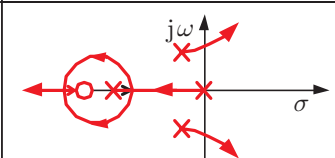
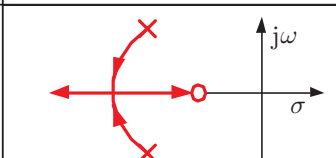
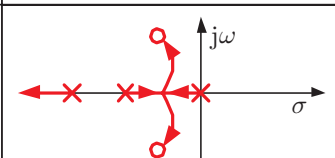
it is required to find the effect of the parameter α on the position of the closed-loop poles. The equation is rewritten into the desired form

$$1 + \alpha \frac{s^2}{s^3 + 3s^2 + 2s + 4} = 0 . \quad \blacksquare$$

Using the rules 1 to 10 one can easily predict the geometrical form of the root locus based on the distribution of the open-loop poles and zeros. Table 6.2.1 shows some typical distributions of open-loop poles and zeros and their root loci.

For the qualitative assessment of the root locus one can use a physical analogy. If all open-loop poles are substituted by a negative electrical charge and all zeros by a commensurate positive one and if a massless negative charged particle is put onto a point of the root locus, a movement is observed. The path that the particle takes because of the interplay between the repulsion of the poles and the attraction of the zeros lies just on the root locus. Comparing the root locus examples 3 and 9 of Table 6.2.1 the 'repulsive' effect of the additional pole can be clearly seen.

Table 6.2.1: Typical distributions of open-loop poles and zeros and the root loci

No.	root locus	No.	root locus
1		9	
2		10	
3		11	
4		12	
5		13	
6		14	
7		15	
8		16	

6.3 Example of an application of the root-locus method

The systematic application of the rules from section 6.2 for the construction of a root locus is shown in the following non-trivial example for the open-loop transfer function

$$G_0(s) = \frac{k_0(s+1)}{s(s+2)(s^2+12s+40)}. \quad (6.3.1)$$

The degree of the numerator polynomial is $m = 1$. This means that the transfer function has one zero ($s_{Z_1} = -1$). The degree of the denominator polynomial is $n = 4$ and we have the four poles ($s_{P_1} = 0$, $s_{P_2} = -2$, $s_{P_3} = -6 + j2$, $s_{P_4} = -6 - j2$). First the poles (x) and the zeros (o) of the open loop are drawn on the s plane as shown in Figure 6.3.1. According to rule 3 these poles are just those points of the root

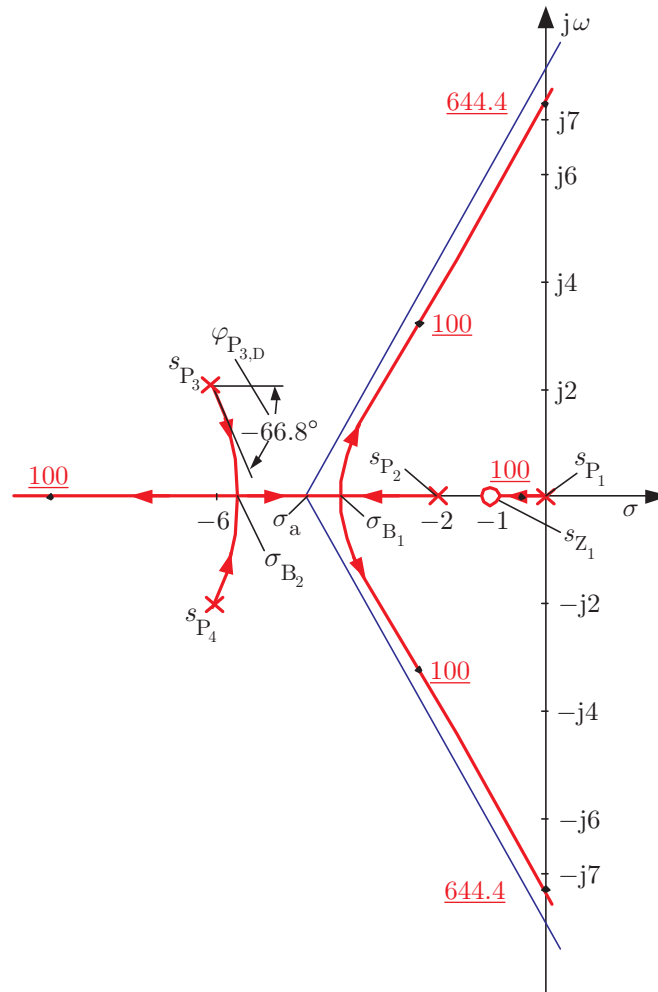


Figure 6.3.1: Root locus of $G_0(s) = \frac{k_0(s+1)}{s(s+2)(s+6+j2)(s+6-j2)}$. Values of k_0 are in red and underlined.

locus where $k_0 = 0$ and the zeros where $k_0 \rightarrow \infty$. We have $(n - m) = 3$ branches that go to infinity and the asymptotes of these three branches are lines which intercept the real axis according to rule 6. From Eq. (6.2.6) the crossing is at

$$\sigma_a = \frac{(0 - 2 - 6 - 6) - (-1)}{3} = -\frac{13}{3} = -4.33 \quad (6.3.2)$$

and the slopes of the asymptotes are according to Eq. (6.2.5)

$$\alpha_k = \frac{\pm 180^\circ(2k+1)}{3} = \pm 60^\circ(2k+1) \quad k = 0, 1, 2, \dots \quad (6.3.3)$$

i.e. $\alpha_0 = 60^\circ, \quad \alpha_1 = +180^\circ, \quad \alpha_2 = -60^\circ .$

The asymptotes are shown in Figure 6.3.1 as blue lines. Using Rule 4 it can be checked which points on the real axis are points on the root locus. The points σ with $-1 < \sigma < 0$ and $\sigma < -2$ belong to the root locus, because to the right of them the number of poles and zeros is odd. According to rule 7 breakaway and break-in points can only occur pairwise on the real axis to the left of -2. These points are real solutions of the Eq. (6.2.8). Here we have

$$\frac{1}{s} + \frac{1}{s+2} + \frac{1}{s+6-j2} + \frac{1}{s+6+j2} = \frac{1}{s+1} \tag{6.3.4}$$

or

$$3s^4 + 32s^3 + 106s^2 + 128s + 80 = 0 .$$

This equation has the solutions $s_{B_1} = -3.68, s_{B_2} = -5.47$ and $s_{B_{3,4}} = -0.76 \pm j0.866$. The real roots $s_{B_1} = -3.68$ and $s_{B_2} = -5.47$ are the positions of the breakaway and the break-in point. The angle of departure $\varphi_{P_{3,D}}$ of the root locus from the complex pole at $s_{P_3} = -6 + j2$ can be determined from Figure 6.3.2 according to Eq. (6.2.9):

$$\begin{aligned} \varphi_{P_{3,D}} &= -90^\circ - 153.4^\circ - 161.6^\circ + 158.2^\circ \pm 180^\circ(2k + 1) \\ &= -246.8^\circ + 180^\circ = -66.8^\circ . \end{aligned} \tag{6.3.5}$$

With this specifications the root locus can be sketched. Using rule 9 the value of k_0 can be determined

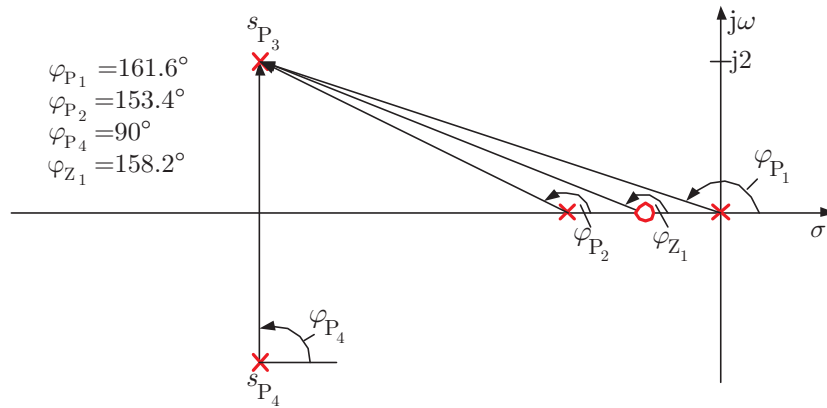


Figure 6.3.2: Calculating the angle of departure $\varphi_{P_{3,D}}$ of the complex pole $s_{P_3} = -6 + j2$

for some selected points. The value at the intersection with the imaginary axis is

$$k_{0,crit} = \frac{7.2 \cdot 7.4 \cdot 7.9 \cdot 11.1}{7.25} = 644.4 .$$

Module 7

Behaviour of linear continuous-time control systems

Module units

7.1 Dynamical behaviour of a closed loop system	7-1
7.2 Static properties of the closed loop	7-6
7.2.1 Transfer function $G_0(s)$ with delayed P behaviour	7-7
7.2.2 Transfer function $G_0(s)$ with delayed I behaviour	7-7
7.2.3 Transfer function $G_0(s)$ with delayed I_2 behaviour	7-8
7.3 Performance indices	7-9
7.3.1 Time-response specifications	7-9
7.3.2 Integral performance indices	7-10
7.3.3 Determination of quadratic performance indices	7-12

Module overview. Before a control system is designed it is necessary to specify what is expected for the closed-loop control behaviour. The dynamical behaviour and its static properties must be described. The basic requirements for control system design are given. These are precisely defined for the case of disturbance and command behaviour and illustrated by examples. Performance indices to measure the quality of the control system in the time domain using standard quantities are introduced, e.g. maximum overshoot, rise time and settling time. Integral performance indices and their determination in the s domain for the case of quadratic forms are also given.

Module objectives. When you have completed this module you should be able to:

1. Know the basic requirements of control system design.
2. Determine static and dynamic properties.
3. Judge and measure the performance of control systems in the time domain.

Module prerequisites. Transfer function, Laplace transform.

7.1 Dynamical behaviour of a closed loop system

Prepared with an understanding of models, transfer functions and basics about control loops, we now consider control systems and the types of feedback design principles available. Figure 7.1.1 shows a block diagram of a closed loop system with the four classical components: controller, actuator, plant and

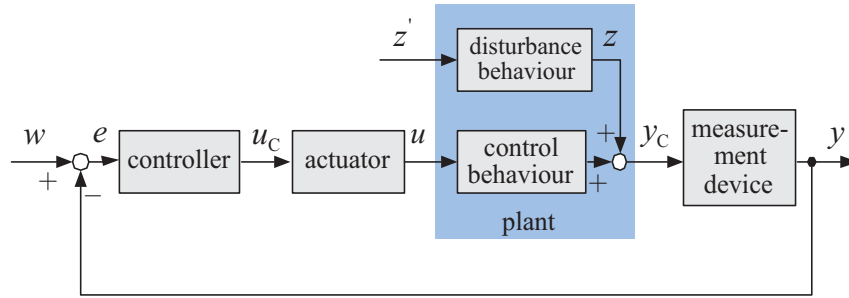


Figure 7.1.1: The basic components of a control loop

measurement device. It is often convenient to combine the controller and actuator into one controller component, while the measurement device is often assigned to the plant. Usually a set of disturbances $z'_i (i = 1, 2, \dots)$ may occur, each of them can enter the plant at different locations. The transition behaviour of the plant and of the parts of the plant between disturbance input and plant output, respectively, is denoted by $G_{Pz_i}(s)$. From this a block diagram of the closed loop system is obtained according to Figure 7.1.2.

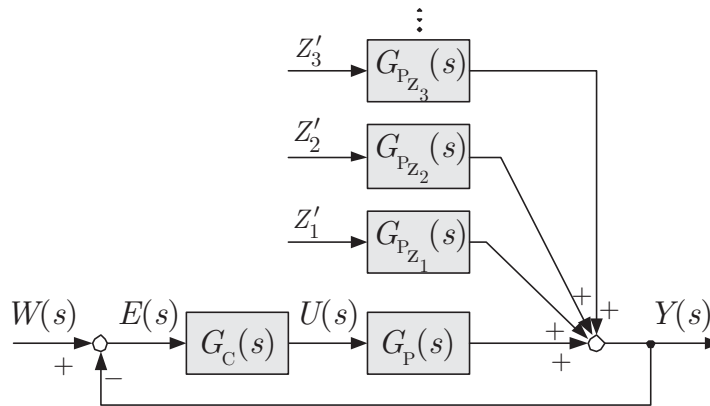


Figure 7.1.2: Block diagram of the closed-loop system

For linear plants all the disturbances Z'_i can be combined into one single cumulative disturbance

$$Z(s) = \sum_{i=1}^n Z'_i(s)G_{Pz_i}(s)$$

according to Figure 7.1.2. This cumulative disturbance will act at the plant output (see Figure 7.1.3). Furthermore, by a suitable choice of $G_{Pz_i}(s)$ it can be shown that the structure from Figure 7.1.2 is also valid for disturbances $z'_i(t)$ entering at other locations in the closed loop.

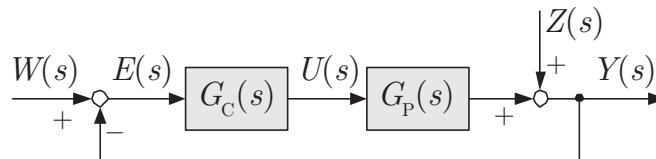


Figure 7.1.3: Block diagram of the closed-loop system with cumulative disturbance $Z(s)$

The transition behaviour of this control loop is specified according to the two inputs (command and disturbance) either command behaviour or disturbance behaviour. The transfer function of the controller

elements – briefly called in the following only controller – is $G_C(s)$ and those of the plant $G_P(s)$. From Figure 7.1.3 the controlled variable of the closed loop is

$$Y(s) = Z(s) + [W(s) - Y(s)] G_C(s) G_P(s) .$$

Rearranging, then it follows

$$Y(s) = \frac{1}{1 + G_C(s) G_P(s)} Z(s) + \frac{G_C(s) G_P(s)}{1 + G_C(s) G_P(s)} W(s) . \quad (7.1.1)$$

Using this equation, the control system tasks already mentioned in section 1.3 can be formulated more precisely as follows:

- a) For $W(s) = 0$ the transfer function of the closed loop for *disturbance behaviour* the *disturbance transfer function*

$$G_Z(s) = \frac{Y(s)}{Z(s)} = \frac{1}{1 + G_C(s) G_P(s)} \quad (7.1.2)$$

is obtained.

- b) Similarly for $Z(s) = 0$ the transfer function of the closed loop for *command behaviour* is the *command transfer function*

$$G_W(s) = \frac{Y(s)}{W(s)} = \frac{G_C(s) G_P(s)}{1 + G_C(s) G_P(s)} . \quad (7.1.3)$$

Both transfer functions $G_Z(s)$ and $G_W(s)$ contain the *dynamical control factor*

$$R(s) = \frac{1}{1 + G_0(s)} \quad (7.1.4)$$

with

$$G_0(s) = G_C(s) G_P(s) . \quad (7.1.5)$$

Opening the closed loop for $W(s) = 0$ and $Z(s) = 0$ according to Figure 7.1.4 at an arbitrary location and defining with respect to the route of the transfer elements the input as $x_e(t)$ and the output as $x_a(t)$, the transfer function of the *open loop*

$$G_{\text{open}}(s) = \frac{X_a(s)}{X_e(s)} = -G_C(s) G_P(s) = -G_0(s) \quad (7.1.6)$$

is obtained.

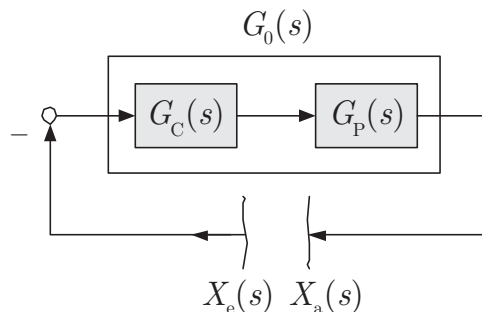


Figure 7.1.4: Open control loop

If $G_0(s)$ can be described by a rational fraction, by setting the denominator of Eq. (7.1.2) or Eq. (7.1.3) to zero one obtains for the closed loop the condition

$$1 + G_0(s) = 0 \quad (7.1.7)$$

analogous to Eq. (3.5.3) for the *characteristic equation* in the form

$$P(s) = a_0 + a_1s + a_2s^2 + \dots + a_ns^n = 0 . \quad (7.1.8)$$

The overall goal in designing a control system is to use the principle of feedback to cause the controlled variable to follow a desired command variable accurately regardless of the command variable's path and to minimise the effect of any external disturbances or changes in the dynamics of the plant. Reaching this goal economically the standard structure of Figure 7.1.3 is a relatively complex task if one must meet the basic requirements listed below:

- a) The minimum requirement is that the closed loop is stable.
- b) Disturbances $z(t)$ should be rejected or they must have a small influence on the controlled variable $y(t)$.
- c) The controlled variable $y(t)$ should track the command input $w(t)$ as precisely and as fast as possible.
- d) The closed loop should be as insensitive as possible with respect to changes in the plant parameters.

In order to fulfil the requirements in the ideal case, the command transfer function must be according to requirement c)

$$G_W(s) = \frac{Y(s)}{W(s)} = \frac{G_0(s)}{1 + G_0(s)} = 1 , \quad (7.1.9)$$

and the disturbance transfer function according to requirement b)

$$G_Z(s) = \frac{Y(s)}{Z(s)} = \frac{1}{1 + G_0(s)} = 0 . \quad (7.1.10)$$

A rigorous realisation of these requirements is not possible for physical and technical reasons. The problem will be illustrated using the following simple example.

Example 7.1.1

A common actuator in control systems is the DC motor. It provides rotary motion for a current input. The dynamical behaviour between current $u(t)$ and speed $y(t)$ is described by the simplified transfer function

$$G_P(s) = \frac{Y(s)}{U(s)} = \frac{K_P}{1 + T_s s} . \quad (7.1.11)$$

In order to compensate the plant dynamics, a candidate controller may be

$$G_C(s) = K_C(1 + T_s s) . \quad (7.1.12)$$

The open-loop transfer function is

$$\begin{aligned} G_0(s) &= G_C(s) G_P(s) \\ &= K_C(1 + T_s s) \frac{K_P}{1 + T_s s} = K_C K_P , \end{aligned} \quad (7.1.13)$$

which shows a proportional behaviour. On step inputs to the controller the speed will jump, which is physically not possible due to the inertia of the motor. According to section 3.3 the controller in Eq.(7.1.12) is not realisable. Adding a pole in the controller transfer function to the left in the s plane at $s_C = -10/T$ will cure this problem, but with a delayed speed response. Figure 7.1.5a shows the controlled speed for a unit step in the command input. The time constant of the closed loop system changes as the feedback gain increases. Increasing the controller gain K_C will speed-up the behaviour and reduce the steady-state error, but will also increase the control effort as shown in Figure 7.1.5b. As the current of the motor is limited for physical reasons the manipulated variable $u(t)$ is also limited. Increasing the controller gain K_C to an arbitrary high value is not suitable. During the design of a controller such limitations have to be taken into account.

It is often true that closed-loop systems have a faster response as the feedback gain is increased, and if there are no other effects, this is generally desirable. However, systems typically also become less well

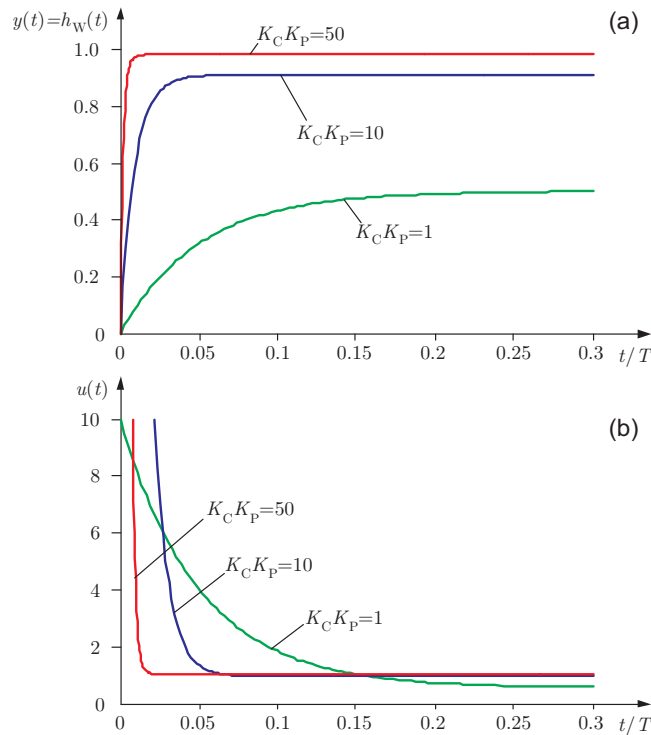


Figure 7.1.5: Step response of the closed loop, (a) speed and (b) current for different open-loop gains

damped and even unstable as the gain increases. This is shown when we mount the same DC motor on a robotic manipulator and control the speed of the manipulator arm using the same type of controller. In this case the speed of the arm movement is the controlled variable $y(t)$. The transfer function between the current of the DC motor and the speed of the arm is

$$G_P(s) = \frac{Y(s)}{U(s)} = \frac{K_P}{(1 + Ts)(1 + \frac{2\zeta}{\omega_0}s + (\frac{s}{\omega_0})^2)} \quad (7.1.14)$$

Figure 7.1.6 shows the step response of this control system. ■

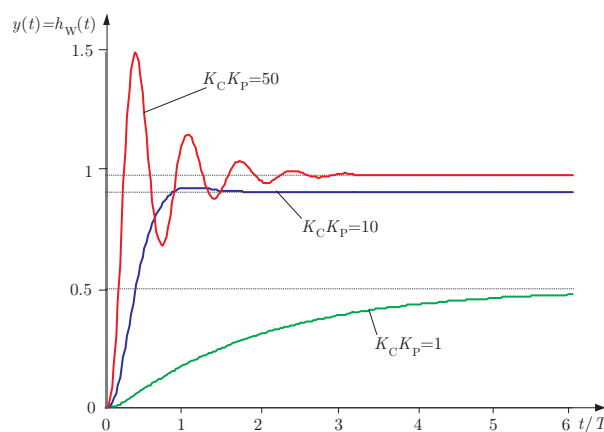


Figure 7.1.6: Step response of the closed loop for different open-loop gains for $\zeta = 4$ and $\omega_0 = 2/T$

From the example given above it can be seen that a definite limit exists on how high we can make the gain. But there is a design tradeoff between gain and steady-state error. Attempts to resolve the conflict between small steady-state errors and good transient or dynamic responses must be undertaken. These two essential aspects of performance are considered when a control system is designed: the transient

performance and the steady-state performance. The following sections deal with these aspects in more detail.

7.2 Static properties of the closed loop

Frequently the behaviour of an open loop (according to Figure 7.1.4 and Eq. (7.1.5)) can be described by a generalised transfer function of the form

$$G_0(s) = \frac{K_0}{s^k} \frac{1 + \beta_1 s + \dots + \beta_m s^m}{1 + \alpha_1 s + \dots + \alpha_{n-k} s^{n-k}} e^{-T_v s} \quad m \leq n, \quad (7.2.1)$$

where the constant $k = 0, 1, 2, \dots$ denotes the type of transfer function $G_0(s)$. K_0 is the gain of the open loop. Therefore $G_0(s)$ shows for

$$\begin{aligned} k = 0; & \quad \textit{delayed proportional} \textit{ behaviour} && \text{(delayed P behaviour)} \\ k = 1; & \quad \textit{delayed integral} \textit{ behaviour} && \text{(delayed I behaviour)} \\ k = 2; & \quad \textit{delayed double integral} \textit{ behaviour} && \text{(delayed I}_2 \textit{ behaviour)}. \end{aligned}$$

We assume now that the term of the rational fraction in Eq. (7.2.1) contains only poles in the left half s plane. For the different types of transfer functions $G_0(s)$ with different forms of the command signal $w(t)$ or of the disturbance $z(t)$ the steady state of the closed loop for $t \rightarrow \infty$ can be analysed.

With

$$E(s) = W(s) - Y(s) \quad (7.2.2)$$

from Eqs. (7.1.1) and (7.1.6) it follows for the control error

$$E(s) = \frac{1}{1 + G_0(s)} [W(s) - Z(s)]. \quad (7.2.3)$$

Under the assumption, that the limit of the control error $e(t)$ for $t \rightarrow \infty$ exists, one obtains by using the final value theorem of the Laplace transform (see section 2.3) the steady-state value of the control error

$$\lim_{t \rightarrow \infty} e(t) = \lim_{s \rightarrow 0} s E(s). \quad (7.2.4)$$

For the case of all disturbances being related to the plant output from Eq. (7.2.3) it follows that – sign apart – both types of inputs, command or disturbance, can be treated equally. Hence in the following to represent both types of input signals the term $X_e(s)$ is chosen as the input signal. Using both Eqs. (7.2.3) and (7.2.4) the steady-state values of the control error for the different signal types of $x_e(t)$ and for different types of transfer functions $G_0(s)$ of the open loop can be obtained. These values characterise the behaviour of the control loop. They are obtained consecutively for the most important cases.

For further treatment the following test signals according to Figure 7.2.1 are used:

a) *Step input signal:*

$$X_e(s) = \frac{x_{e0}}{s}, \quad (7.2.5)$$

where x_{e0} is the height of the step.

b) *Ramp input signal:*

$$X_e(s) = \frac{x_{e1}}{s^2}, \quad (7.2.6)$$

where x_{e1} describes the slope of the ramp signal $x_e(t)$.

c) *Parabolic input signal:*

$$X_e(s) = \frac{x_{e2}}{s^3}, \quad (7.2.7)$$

where x_{e2} is a measure of the acceleration of the parabolic signal $x_e(t)$.

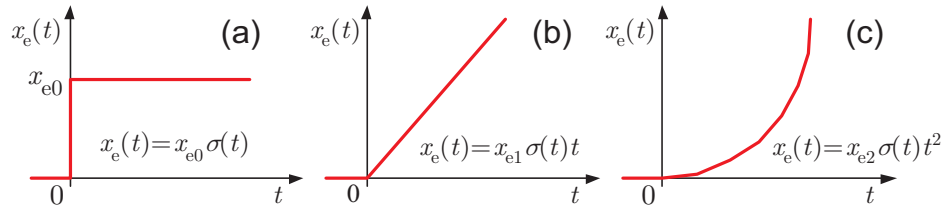


Figure 7.2.1: Different input signals $x_e(t)$, which are frequently used for the disturbance $z(t)$ and command input $w(t)$: (a) step, (b) ramp and (c) parabolic input signal

Following Eq. (7.2.3) the control error is obtained by

$$E(s) = \frac{1}{1 + G_0(s)} X_e(s) , \quad (7.2.8)$$

where the difference between command and disturbance behaviour is only in the sign of $X_e(s)$ (disturbance: $X_e(s) = -Z(s)$; command: $X_e(s) = W(s)$). Inserting this relation into Eqs. (7.2.5) to (7.2.7) the corresponding control error can be obtained for different types of transfer functions $G_0(s)$. This will be demonstrated in the following.

7.2.1 Transfer function $G_0(s)$ with delayed P behaviour

For this case from Eq. (7.2.1) the transfer function is

$$G_0(s) = K_0 \frac{1 + \beta_1 s + \dots + \beta_m s^m}{1 + \alpha_1 s + \dots + \alpha_n s^n} e^{-T_t s} . \quad (7.2.9)$$

This transfer function describes an open control loop with delayed *P behaviour*. The variable K_0 is the gain of this open control loop. In this case it is composed of the gain of the controller K_C and of the plant K_P in multiplicative form

$$K_0 = K_C K_P . \quad (7.2.10)$$

With Eq. (7.2.4) one obtains for the steady-state error of the closed control loop

$$\lim_{t \rightarrow \infty} e(t) = \lim_{s \rightarrow 0} s \frac{1}{1 + G_0(s)} X_e(s) \quad (7.2.11)$$

and for the *step signal* with Eq. (7.2.5)

$$e_\infty = \lim_{t \rightarrow \infty} e(t) = \frac{1}{1 + K_0} x_{e0} . \quad (7.2.12)$$

It can be shown for a *ramp input signal* according to Eq. (7.2.6) that the double pole in Eq. (7.2.8) in the time domain corresponds to $e(t) = \text{const} \cdot t \cdot \sigma(t)$ such that

$$e_\infty = \lim_{t \rightarrow \infty} e(t) \rightarrow \infty , \quad (7.2.13)$$

which is not finite. A similar situation results for a *parabolic input signal* using Eq. (7.2.7).

7.2.2 Transfer function $G_0(s)$ with delayed I behaviour

From Eq. (7.2.1) it follows for this case that

$$G_0(s) = \frac{K_0}{s} \frac{1 + \beta_1 s + \dots + \beta_m s^m}{1 + \alpha_1 s + \dots + \alpha_{n-1} s^{n-1}} e^{-T_t s} . \quad (7.2.14)$$

The inherent open control loop shows delayed I behaviour. With Eq. (7.2.11) one obtains for the steady-state error of the closed loop for the case of a *step input signal*

$$e_{\infty} = \lim_{t \rightarrow \infty} e(t) = 0, \quad (7.2.15)$$

and in the case of a *ramp input signal*

$$e_{\infty} = \lim_{t \rightarrow \infty} e(t) = \frac{1}{K_0} x_{e1}. \quad (7.2.16)$$

Furthermore in the case of *parabolic input signal* we have the same result as Eq. (7.2.13).

DYNAST study example 7.1
Cruise control of a car

DYNAST study example 7.2
D.C. motor position control

DYNAST study example 7.3
Double integrator plant

DYNAST study example 7.4
2nd-order plant

DYNAST study example 7.5
Load disturbance - 2nd-order plant

7.2.3 Transfer function $G_0(s)$ with delayed I_2 behaviour

The transfer function for this particular case is

$$G_0(s) = \frac{K_0}{s^2} \frac{1 + \beta_1 s + \dots + \beta_m s^m}{1 + \alpha_1 s + \dots + \alpha_{n-2} s^{n-2}} e^{-T_t s} \quad (7.2.17)$$

and it describes a system with delayed I_2 behaviour. For the steady-state error of the closed loop system it follows in the case of a stable loop for a *step* and *ramp input signal*

$$e_{\infty} = \lim_{t \rightarrow \infty} e(t) = 0, \quad (7.2.18)$$

and for a *parabolic input signal*

$$e_{\infty} = \lim_{t \rightarrow \infty} e(t) = \frac{1}{K_0} x_{e2}. \quad (7.2.19)$$

From these results, especially from Eqs. (7.2.12), (7.2.16) and (7.2.19) and from Table 7.2.1 it follows, that the steady-state error e_{∞} , which characterises the static behaviour of the control loop is in all cases smaller the larger the *loop gain* K_0 given by Eq. (7.2.10). In the case of delayed P behaviour of the open loop the steady-state error e_{∞} is much smaller the smaller the value of the *static control factor*

$$R = \frac{1}{1 + K_0}. \quad (7.2.20)$$

Often a large loop gain K_0 rapidly leads to instability of the closed loop. For setting K_0 usually a compromise is made unless selecting an appropriate type of controller such that the steady-state error vanishes. Both, the dynamical and especially the static behaviour depend strongly on the choice of the controller. Therefore, in the following the most important types of standard controllers will be introduced.

Table 7.2.1: Steady-state error for different type of systems of $G_0(s)$ and different input signals $x_e(t)$ (command and disturbance signals if all disturbances will act at the plant output)

Type of system of $G_0(s)$ according to Eq. (7.2.1)	Input signal $X_e(s)$	Steady-state error e_∞
$k = 0$ (delayed P behaviour)	$\frac{x_{e_0}}{s}$	$\frac{1}{1 + K_0} x_{e_0}$
	$\frac{x_{e_1}}{s^2}$	∞
	$\frac{x_{e_2}}{s^3}$	∞
$k = 1$ (delayed I behaviour)	$\frac{x_{e_0}}{s}$	0
	$\frac{x_{e_1}}{s^2}$	$\frac{1}{K_0} x_{e_1}$
	$\frac{x_{e_2}}{s^3}$	∞
$k = 2$ (delayed I_2 behaviour)	$\frac{x_{e_0}}{s}$	0
	$\frac{x_{e_1}}{s^2}$	0
	$\frac{x_{e_2}}{s^3}$	$\frac{1}{K_0} x_{e_2}$

7.3 Performance indices

7.3.1 Time-response specifications

Specifications for a control system design often involve certain requirements associated with the time response of the closed-loop system. The requirements are specified by the behaviour of the controlled variable $y(t)$ or by the control error $e(t)$ on well defined test signals. The most important test signal is a unit step on the input of the control system and requirements are placed on the behaviour of the controlled variable $y(t) = h_W(t)$, as shown in Figure 7.3.1. The requirements for a unit step response are expressed in terms of the following standard quantities:

- The *maximum overshoot* M_p is the magnitude of the overshoot after the first crossing of the steady-state value (100%). This value is normally expressed as a percentage of the steady-state value of the controlled variable.
- The *peak time* t_p is the time required to reach the maximum overshoot.
- The *settling time* t_ε is the time for the controlled variable first to reach and thereafter remain within a prescribed percentage $\pm\varepsilon$ of the steady-state value. Common values of ε are 2%, 3% or 5%.
- The *rise time* t_r is the time required to reach first the steady-state value (100%). It may also be defined as the time to reach the vicinity of the steady-state value particularly for a response with no overshoot, e.g. the time between 10% and 90%. The 50% rise time $t_{r,50}$ is defined as the time to first reach 50% of the final value.

Similarly, the behaviour on step disturbances can be characterised as shown in Figure 7.3.2. Here likewise the terms 'maximum overshoot' and 'settling time' are defined.

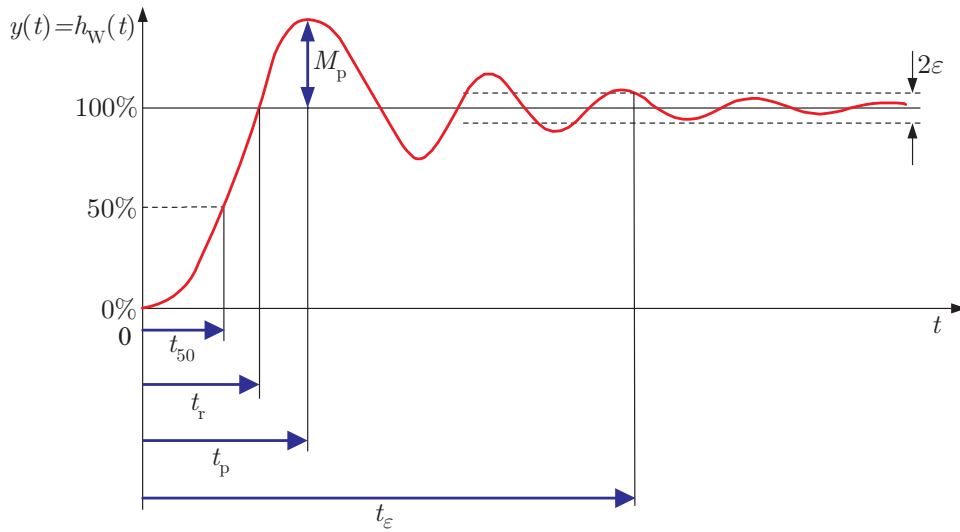


Figure 7.3.1: Typical under-damped response of a control system to step command inputs

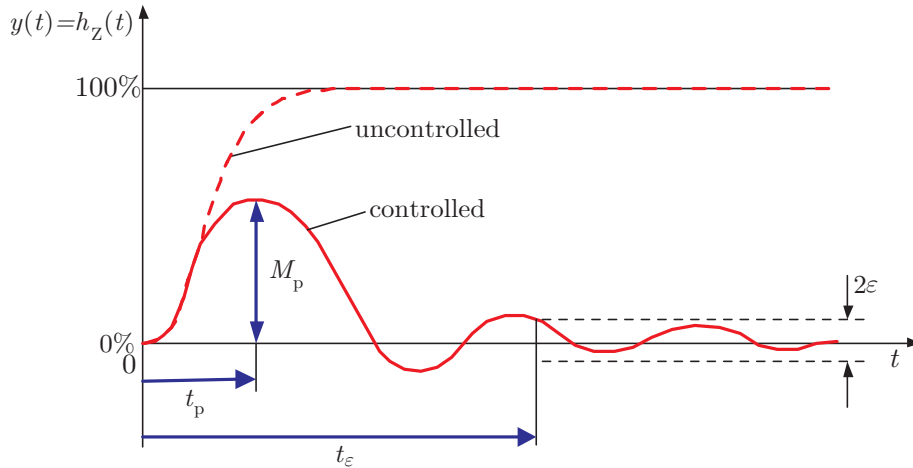


Figure 7.3.2: Typical under-damped response of a control system to step disturbances

These standard quantities are measures of some properties of the control system. M_p and t_ϵ essentially characterise the damping and t_r and t_p the speed, i.e. the dynamics of the control behaviour. The steady-state error e_∞ as described in section 7.2 is a typical characteristic of the static behaviour.

These quantities describe the deviation of the step response from the ideal case described in section 7.1 and the goal of the design of a control system is to hold them as small as possible. In most cases one can restrict the values of the three quantities t_r , t_ϵ and M_p .

7.3.2 Integral performance indices

It generally happens that a control system design problem reaches the point where one or more parameters are to be selected to give the best performance. If a measure or index of performance can be expressed mathematically, the problem can be solved for the best choice of the adjustable parameters. The resulting system is termed optimal with respect to the selected criterion.

Introducing a performance index

$$J_a = w_1 t_r + w_2 t_\epsilon + w_3 M_p, \quad (7.3.1)$$

composed of the quantities as discussed in section 7.3.1 and minimising J_a is unsuitable. This is because one has to subjectively specify the weights w_1 , w_2 and w_3 , and a straight mathematical approach to solve

the minimisation problem is not available.

The selection of an appropriate performance index is as much a part of the design process as calculating the final system. An optimal value of an inappropriate performance measure may result in poor performance. Commonly used performance indices are based on integral performance measures. From Figure 7.3.1 one can see that the area between the 100% line and the step response $h_W(t)$ is surely a measure for the deviation of the control loop from the ideal response for a step command input. Likewise for the deviation of the disturbance rejection from the ideal case Figure 7.3.2 shows the area between the 0% line and $h_Z(t)$ to be a good measure. In both cases the total area below the control error $e(t) = w(t) - y(t)$ is involved. It is obvious to introduce the integral

$$J_k = \int_0^{\infty} f_k[e(t)] dt \quad (7.3.2)$$

as a performance measure, where $f_k[e(t)]$ is one of the functions given in Table 7.3.1, e.g. $e(t)$, $|e(t)|t$ and $e^2(t)$. These types of integral performance indices can also include derivatives of the control error or terms of the manipulated variable $u(t) - u_{\infty}$ that adds a penalty for control effort. Using such performance

Table 7.3.1: The most common integral performance indices

performance index	properties
$J_{IE} = \int_0^{\infty} e(t) dt$	<i>Integral of error:</i> suitable for highly damped or monotonic responses; simple mathematical treatment.
$J_{IAE} = \int_0^{\infty} e(t) dt$	<i>Integral of absolute value of error:</i> suitable for non-monotonic responses; awkward analysis.
$J_{ISE} = \int_0^{\infty} e^2(t) dt$	<i>Integral of squared error:</i> highly penalising large control errors; settling time is larger than for J_{IAE} ; good analytical treatment.
$J_{ITAE} = \int_0^{\infty} e(t) t dt$	<i>Integral of time multiplied by the absolute value of error:</i> effect like J_{IAE} ; regards the permanence of the control error.
$J_{IST^qE} = \int_0^{\infty} [e(t) t^q]^2 dt$	<i>Integral of the squared time to the q times error:</i> effect like J_{ISE} , but adds a heavy penalty for errors that do not die out rapidly; does not try to eliminate too quickly the inevitable initial error for a step input; less overshoot M_P and shorter settling time t_{ε} ; good analytical treatment.
$J_{GISE} = \int_0^{\infty} [e^2(t) + \alpha \dot{e}^2(t)] dt$	<i>Generalised integral of squared error:</i> responses more favourable than J_{ISE} ; choice of the weight α subjective.
$J_{ISESC} = \int_0^{\infty} [e^2(t) + \beta(u(t) - u_{\infty})^2] dt$	<i>Integral of squared error and squared control effort:</i> Larger overshoot M_P , but essentially shorter settling time t_{ε} ; choice of the weight β subjective.

indices one can now define the integral criterion as follows:

A control system is better the smaller the value of J_k of the selected performance index.

Therefore the integral criterion

$$J_k = \int_0^{\infty} f_k[e(t)] dt = J_k(c_1, c_2, \dots) \stackrel{!}{=} \text{Min} \quad (7.3.3)$$

always requires minimising J_k by adjusting the controller parameters c_1, c_2, \dots

7.3.3 Determination of quadratic performance indices

The integral performance indices J_{IAE} and J_{ITAE} from Table 7.3.1 have the disadvantage that they have to be evaluated in the time domain, either by laborious computation or by simulation. All the other squared error type of indices are more pleasant because of calculating its value in the s-domain rather than in the time domain. Therefore the analysis is simpler. This will be shown in the following for J_{ISE} .

The performance index is given by

$$J_{ISE} = \int_0^{\infty} e^2(t) dt \quad (7.3.4)$$

Applying the convolution theorem in the frequency domain from Eq. (2.3.11) for $s = c = 0$ and $f_1(t) = f_2(t) = e(t)$, one obtains *Parseval's theorem*

$$J_{ISE} = \int_0^{\infty} e^2(t) dt = \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} E(s) E(-s) ds . \quad (7.3.5)$$

When $E(s)$ is expressed as a ratio of polynomials

$$E(s) = \frac{b_0 + b_1 s + \dots + b_{n-1} s^{n-1}}{a_0 + a_1 s + \dots + a_n s^n} , \quad (7.3.6)$$

different methods are available to evaluate the integral. For calculation of the integral a recursion formula is available. Its solution has also been tabulated up to quite high values of n in terms of the coefficients of the polynomials. Table 7.3.2 below gives a short list. For a detailed analysis, e.g. when the integral depends on some parameters, a general algebraic approach using determinants is more suitable as shown in section A.7.

Table 7.3.2: Values for the integral

n	J_{ISE}
1	$\frac{b_0^2}{2a_0 a_1}$
2	$\frac{b_1^2 a_0 + b_0^2 a_2}{2a_0 a_1 a_2}$
3	$\frac{b_2^2 a_0 a_1 + (b_1^2 - 2b_0 b_2) a_0 a_3 + b_0^2 a_2 a_3}{2a_0 a_3 (-a_0 a_3 + a_1 a_2)}$

The more general form

$$J_{IST^qE} = \int_0^{\infty} [e(t) t^q]^2 dt \quad (7.3.7)$$

of a squared performance index can be easily evaluated. Since the Laplace transform of $te(t)$ according to the complex differentiation theorem Eq. (2.3.7) is equal to $-\frac{dE(s)}{ds}$, one obtains

$$J_{IST^qE} = \int_0^{\infty} [e(t) t^q]^2 dt = \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \frac{d^q}{ds^q} E(s) \frac{d^q}{ds^q} E(-s) ds . \quad (7.3.8)$$

Since it is easy to compute the derivatives of a polynomial the above integral can be computed using simple algebra and the aforementioned recursive formula.

Example 7.3.1

Determining the best damping ratio for a second-order system is a simple example of the use of performance indices. Let us assume that the command transfer function of a control system is described by Eq. (4.4.36) with $K = 1$ as

$$G_W(s) = \frac{\omega_0^2}{\omega_0^2 + 2\zeta\omega_0 s + s^2} .$$

The control error for a unit step input is

$$E(s) = \frac{1}{s} \left[1 - \frac{\omega_0^2}{\omega_0^2 + 2\zeta\omega_0 s + s^2} \right] = \frac{2\zeta\omega_0 + s}{\omega_0^2 + 2\zeta\omega_0 s + s^2} . \quad (7.3.9)$$

From Table 7.3.2 we have

$$J_{\text{ISE}} = \frac{1 + 4\zeta^2}{4\zeta\omega_0} .$$

As this function is a parabola in ζ , with minimum given by

$$\frac{dJ_{\text{ISE}}}{d\zeta} = \frac{4\zeta^2 - 1}{4\zeta^2\omega_0} = 0$$

the minimum square error to a step input occurs for $\zeta = 0.5$. ■

Demonstration Example 7.1

A virtual experiment using manual control

Module 8

PID control and associated controller types

Module units

8.1	The classical three-term PID controller	8-1
8.2	Optimal tuning of PID controllers	8-3
8.3	Advantages and disadvantages of the different types of controllers	8-4
8.4	Empirical tuning rules according to Ziegler and Nichols	8-6

Module overview. *Proportional-integral-derivative (PID) control constitutes the heuristic approach to controller design that has found wide acceptance in industrial applications. This type of controller family is introduced and its behaviour discussed in detail. Both, empirical tuning using the rules according to Ziegler and Nichols and the optimal tuning using quadratic performance indices are shown. A comparison of the controller members of this family using performance indices is given to demonstrate the differences in their control behaviour.*

Module objectives. *When you have completed this module you should be able to:*

1. Understand PID-type controllers.
2. Tune PID-type controllers using the rules of Ziegler and Nichols.

Module prerequisites. *Transfer function, performance indices.*

8.1 The classical three-term PID controller

We have seen in section 7.1 that proportional feedback control can reduce error responses but that it still allows a non-zero steady-state error for a proportional system. In addition, proportional feedback increases the speed of response but has a much larger transient overshoot. When the controller includes a term proportional to the integral of the error, then the steady-state error can be eliminated, as shown in section 7.2. But this comes at the expense of further deterioration in the dynamic response. Addition of a term proportional to the derivative of the error can damp the dynamic response. Combined, these three kinds of actions form the classical PID controller, which is widely used in industry.

This principle mode of action of the PID controller can be explained by the parallel connection of the P, I and D elements shown in Figure 8.1.1. From this diagram the *transfer function* of the PID controller is

$$G_C(s) = \frac{U_C(s)}{E(s)} = K_P + \frac{K_I}{s} + K_D s . \quad (8.1.1)$$

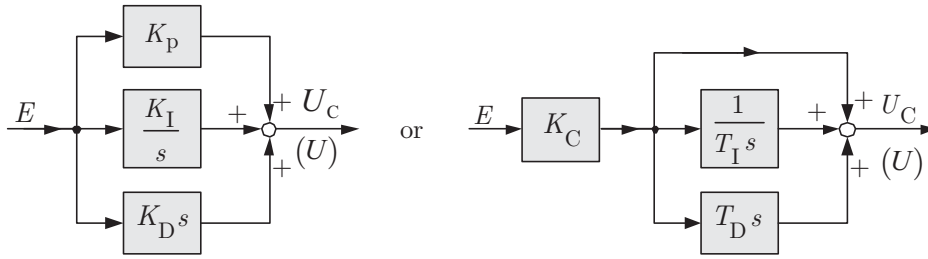


Figure 8.1.1: Block diagram of the PID controller

The controller variables are

$$\begin{aligned}
 K_C &= K_p && \text{gain} \\
 T_I &= \frac{K_p}{K_I} && \text{integral action time} \\
 T_D &= \frac{K_D}{K_p} && \text{derivative action time}
 \end{aligned}$$

Eq. (8.1.1) can be rearranged to give

$$G_C(s) = K_C \left(1 + \frac{1}{T_I s} + T_D s \right) . \tag{8.1.2}$$

These three variables K_C , T_I and T_D are usually tuned within given ranges. Therefore, they are often called the *tuning parameters* of the controller. By proper choice of these tuning parameters a controller can be adapted for a specific plant to obtain a good behaviour of the controlled system.

It follows from Eq. (8.1.2) that the time response of the controller output is

$$u_C(t) = K_C e(t) + \frac{K_C}{T_I} \int_0^t e(\tau) d\tau + K_C T_D \frac{de(t)}{dt} . \tag{8.1.3}$$

Using this relationship for a step input of $e(t)$, i.e. $e(t) = \sigma(t)$, the *step response* $h(t)$ of the PID controller can be easily determined. The result is shown in Figure 8.1.2a. One has to observe that the length of the arrow $K_C T_D$ of the D action is only a measure of the weight of the δ impulse.

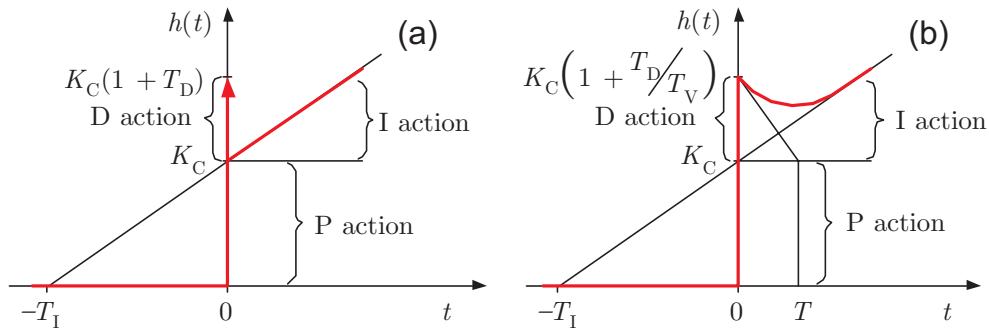


Figure 8.1.2: Step responses (a) of the ideal and (b) of the real PID controller

In the previous considerations it has been assumed that a D behaviour can be realised by the PID controller. This is an ideal assumption and in reality the ideal D element cannot be realised (see section 3.3). In real PID controllers a lag is included in the D behaviour. Instead of a D element in the block diagram of Figure 8.1.1 a DT_1 element with the transfer function

$$G_D(s) = K_D \frac{T_V s}{1 + T_V s} \tag{8.1.4}$$

is introduced. From this the transfer function of the *real* PID controller or more precisely of the PIDT₁ controller follows as

$$G_C(s) = K_p + \frac{K_I}{s} + K_D \frac{T_V s}{1 + T_V s} . \quad (8.1.5)$$

Introducing the controller tuning parameters

$$K_C = K_p, \quad T_I = \frac{K_C}{K_I} \quad \text{and} \quad T_D = \frac{K_D T_V}{K_C}$$

it follows

$$G_C(s) = K_C \left(1 + \frac{1}{T_I s} + T_D \frac{s}{1 + T_V s} \right) . \quad (8.1.6)$$

The step response $h(t)$ of the PIDT₁ controller is shown in Figure 8.1.2b. This response from $t = 0$ gives a large rise, which declines fast to a value close to the P action, and then migrates into the slower I action. The P, I and D behaviour can be tuned independently. In commercial controllers the 'D step' at $t = 0$ can often be tuned 5 to 25 times larger than the 'P step'. A strongly weighted D action may cause the actuator to reach its maximum value, i.e. it reaches its 'limits'.

As *special cases of PID controllers* one obtains for:

- a) $T_D = 0$ the *PI controller* with transfer function

$$G_C(s) = K_C \left(1 + \frac{1}{T_I s} \right) ; \quad (8.1.7)$$

- b) $T_I \rightarrow \infty$ the ideal *PD controller* with the transfer function

$$G_C(s) = K_C (1 + T_D s) \quad (8.1.8)$$

and the PDT₁ controller with the transfer function

$$G_C(s) = K_C \left(1 + T_D \frac{s}{1 + T_V s} \right) ; \quad (8.1.9)$$

- c) $T_D = 0$ and $T_I \rightarrow \infty$ the *P controller* with the transfer function

$$G_C(s) = K_C . \quad (8.1.10)$$

The step responses of these types of controllers are compiled in Figure 8.1.3. A pure *I controller* may also be applied and this has the transfer function

$$G_C(s) = K_I \frac{1}{s} = \frac{K_C}{T_I s} . \quad (8.1.11)$$

8.2 Optimal tuning of PID controllers

The measure of the quality of the transient response of a PID controlled system can be performed by calculating an integral performance index as shown in section 7.3.2. The best controller is one that has the minimum performance index. When this performance index is a minimum for a specified input, the system performance is said to be optimal. When the input signal is specified the quadratic performance index J_{ISE} can be calculated for a given plant transfer function as a function of the tuning parameters, e.g. K_C , T_I , T_D and T_V .

The mathematical calculation of this performance index for given values of the tuning parameters is simple as shown in section 7.3.3. But getting the optimal parameters is a non-trivial task. Though computerised optimisation algorithms are available to calculate the optimal parameter setting, for the case of quadratic

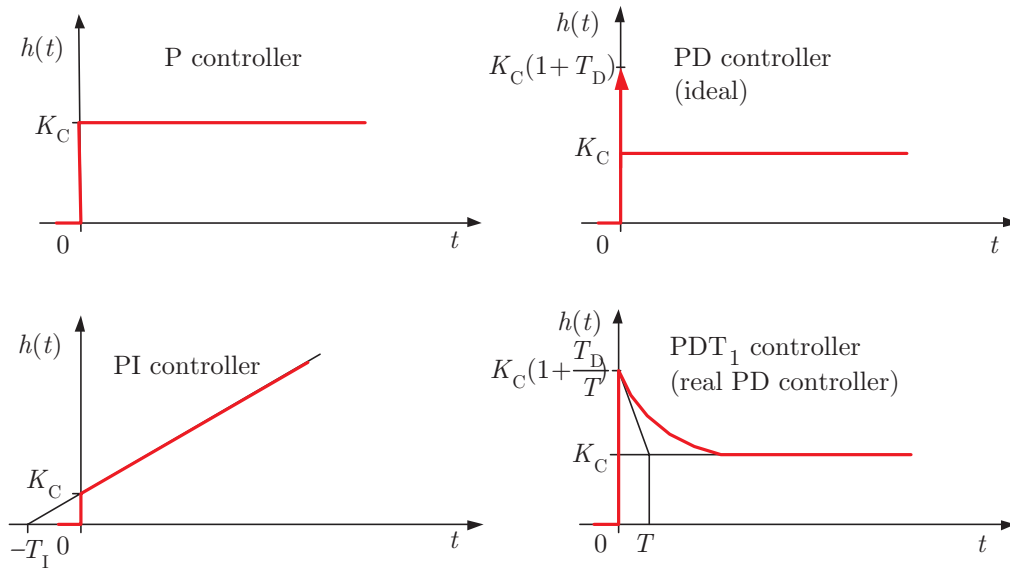


Figure 8.1.3: Step responses of the PID controller family

performance indices a mathematical analysis is possible. The approach shown in section A.7 gives more insight into the controller settings and can be applied to all types of plants and PID controllers.

In the following the command and disturbance behaviour of a control system with a real PID controller and a plant with the transfer function

$$G_P(s) = \frac{K_P}{(1 + Ts)^4}. \quad (8.2.1)$$

will be investigated. The response of the control error to step changes $w(t) = w_0\sigma(t)$ in the command input and $z'(t) = z_0\sigma(t)$ in the plant input is

$$E(s) = \frac{w_0 - z_0 G_C}{1 + G_C G_P} \frac{1}{s}.$$

For the plant (Eq. (8.2.1)) and the real PID controller (Eq. (8.1.6)) one obtains

$$E(s) = \frac{w_0 T_I (1 + Ts)^4 (1 + T_V s) - z_0 K_P T_I (1 + T_V s)}{T_I (1 + Ts)^4 (1 + T_V s)s + K_C K_P [1 + (T_I + T_V)s + (T_D + T_V)s^2]}, \quad (8.2.2)$$

which is in the form of Eq. (7.3.6) or Eq. (A.7.5) for $k = K_C K_P$.

Applying the analysis shown in section A.7 to the J_{ISE} performance index one gets the diagrams in Figure 8.2.1, separately for the command and disturbance inputs. The integral action time constant is normalised by $T_{IN} = \frac{T_I}{4T}$. These diagrams are shown for the optimal value $T_D = T_{Dopt} = \frac{8}{3}T$ of the derivative action time constant. The filter time constant is $T_V = 0.1T_D$. The diagrams show a rather rectangular stability area that makes tuning of K_C and T_I for a fixed T_D easy from the stability point of view. But the performance characteristics are quite different. The optimal parameters for the two cases differ by about a factor of two. Therefore, an optimal tuned controller is in general never optimally tuned for command and disturbance inputs.

8.3 Advantages and disadvantages of the different types of controllers

In the following the disturbance behaviour is investigated using the controllers introduced in section 8.1. Their parameters are tuned optimally according to the performance index J_{ISE} from section 7.3.2. The

plant is given by Eq. (8.2.1). Figure 8.3.1 shows for the different types of controller the responses to a step disturbance $z_0\sigma(t)$ of the controlled variable y , which is normalised by $K_P z_0$. These curves indicate that because $w(t) \equiv 0$ the relation $e(t) = -y(t)$ is valid.

For discussing these curves the term *settling time* $t_{3\%}$ according to section 7.3.1 is used, which is related to the steady state of the uncontrolled case

$$y_{\infty, \text{without}} = K_P z_0 . \tag{8.3.1}$$

In addition, the different cases should be compared with respect to the normalised maximum overshoot $M_p/(K_P z_0)$.

The different cases are discussed below:

- a) The *P controller* shows a relatively high maximum overshoot $M_p/(K_P z_0)$, a long settling time $t_{3\%}$ as well as a steady-state error e_{∞} .

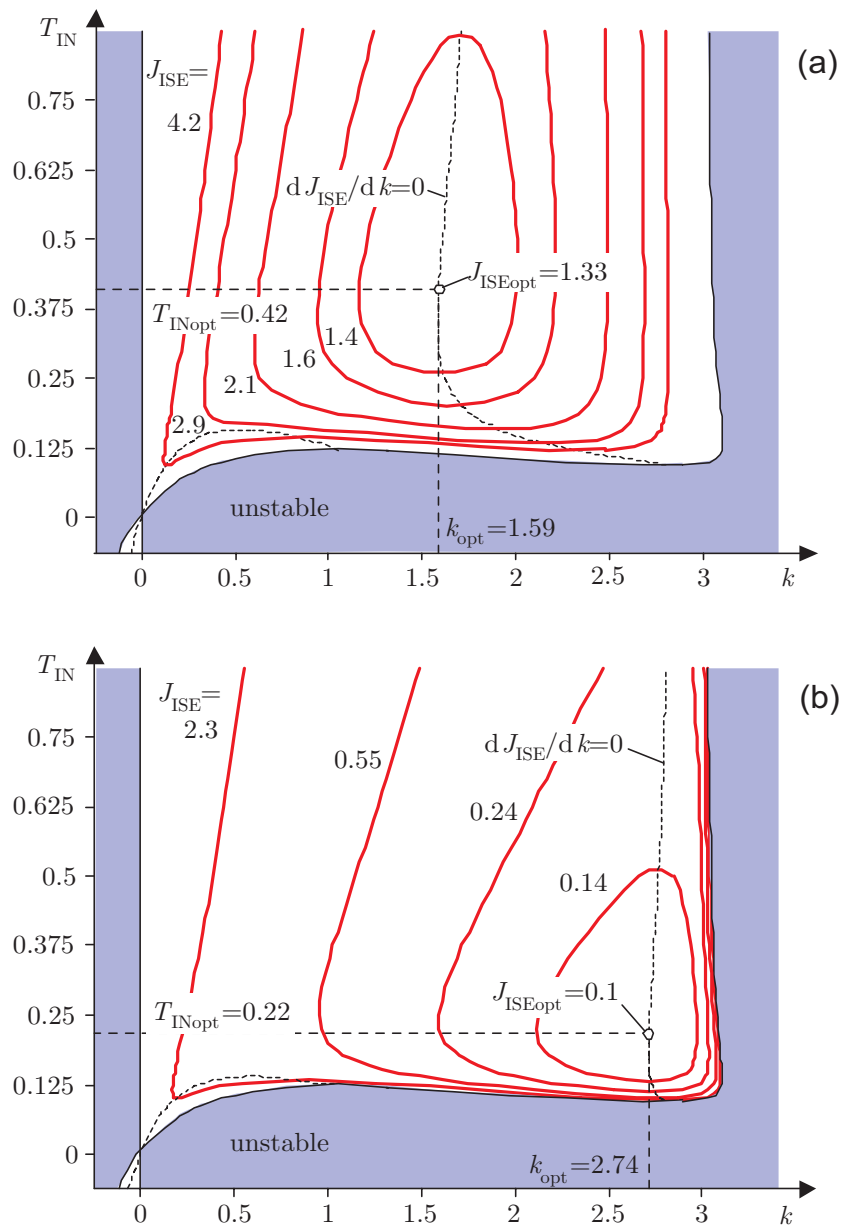


Figure 8.2.1: Stability and performance diagram for step changes (a) in the command input ($w_0 = 1, z_0 = 0$) and (b) in the plant input ($z_0 = 1, w_0 = 0$)

- b) The *I controller* has a higher maximum overshoot than the P controller due to the slowly starting I behaviour, but no steady-state error.
- c) The *PI controller* fuses the properties of the P and I controllers. It shows a maximum overshoot and settling time similar to the P controller but no steady-state error.
- d) The real *PD controller* according to Eq. (8.1.9) with $T_V = T_D/10$ has a smaller maximum overshoot due to the 'faster' D action compared with the controller types mentioned under a) to c). Also in this case a steady-state error is visible, which is smaller than in the case of the P controller. This is because the PD controller generally is tuned to have a larger gain K_C due to the positive phase shift of the D action. For the results shown in Figure 8.3.1 the gain for the P controller is $K_C = 2.68$ and for the PD controller $K_C = 4.74$. The plant has a gain of $K_P = 1$.
- e) The PID controller according to Eq. (8.1.6) with $T_V = T_D/10$ fuses the properties of a PI and PD controller. It shows a smaller maximum overshoot than the PD controller and has no steady state error due to the I action.

The qualitative concepts of this example are also relevant to other type of plants with delayed proportional behaviour. This discussion has given some first insights into the static and dynamic behaviour of control loops.

8.4 Empirical tuning rules according to Ziegler and Nichols

Many industrial processes show step responses with pure aperiodic behaviour according to Figure 8.4.1. This S-shape curve is characteristic of many high-order systems and such plant transfer functions may be approximated by the mathematical model

$$G_P(s) = \frac{K_P}{1 + T_s} e^{-T_t s}, \quad (8.4.1)$$

which contains a 1st-order delay element and a dead time. Figure 8.4.1 shows the approximation by a PT_1T_t element.

Here the step response is characterised by constructing the tangent at the turning point T with the following three values: K_P (gain of the plant), T_r (rise time) and T_u (delay time). Then a rough approximation according to Eq. (8.4.1) is to set $T_t = T_u$ and $T = T_r$.

For a plant of the type described above a lot of tuning rules for standard controllers have been developed. These have been mostly developed empirically from simulation studies. The most famous empirical tuning rules are those of *Ziegler* and *Nichols*. These tuning rules have been derived to provide step responses for the closed loop, where the response shows a decrease of the amplitude of approx. 25% per period. For the application of these rules according to Ziegler and Nichols two different approaches can be used:

- a) *Method of the stability margin(I)*: Here, the following steps are used:
 1. The controller is switched to pure P action.
 2. The gain K_C of the P controller is continuously increased until the closed loop shows permanent oscillations. The value of the gain K_C at this state is denoted as the critical controller gain $K_{C \text{ crit}}$.
 3. The length of period T_{crit} (critical period) of the oscillations is measured.
 4. From $K_{C \text{ crit}}$ and T_{crit} one determines the controller tuning values K_C , T_I and T_D using the formulas given in Table 8.4.1.
- b) *Method of the step response (II)*: In the case of an industrial plant it is often not possible, suitable or allowed to drive the plant into permanent oscillations for determining $K_{C \text{ crit}}$ and T_{crit} . Measuring the step response of the plant does not generally cause difficulties. Therefore, in many cases the second form of the Ziegler-Nichols approach is more expedient. The rules are based directly on the slope K_P/T_r of the tangent at the turning point and on the delay T_u of the step response. One

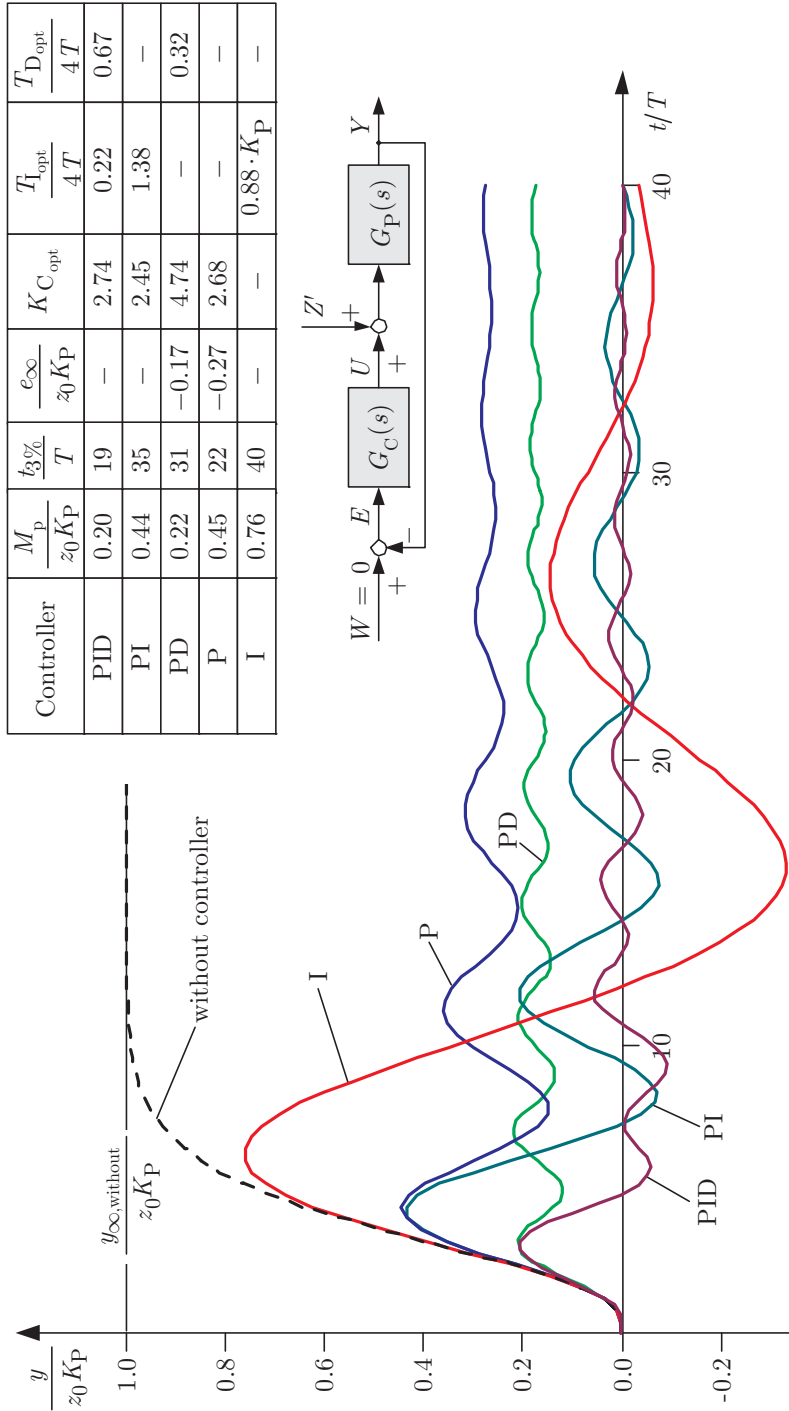


Figure 8.3.1: Behaviour of the normalised controlled variable $y/(z_0 K_P)$ for step disturbance $z' = z_0 \sigma(t)$ at the input to the plant $[G_P(s) = K_P/(1 + Ts)^4; K_P = 1]$ for different types of controllers

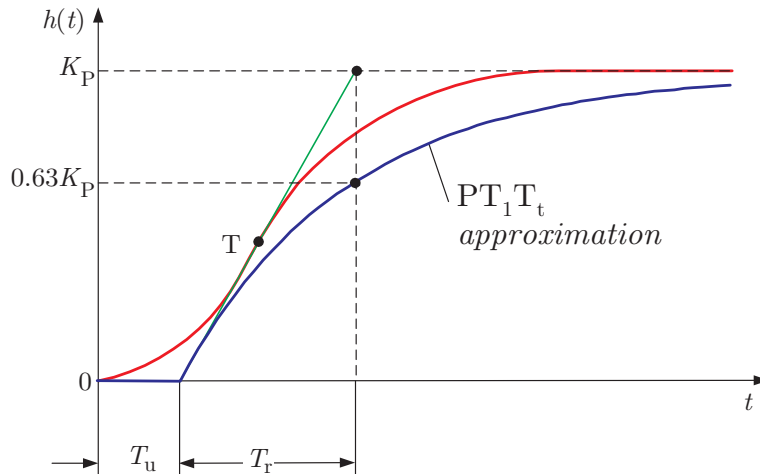


Figure 8.4.1: Describing the step response of a process by the three characteristic values K_P (gain of the plant), T_r (rise time) and T_u (delay time)

has to observe that the measurement of the step response needs only to be taken at the turning point T, as the slope of the tangent already describes the ratio K_P/T_r . Using the measured data T_u and K_P/T_r as well as the formula given in Table 8.4.1 the controller tuning parameters can be determined by simple calculations.

Table 8.4.1: Controller tuning parameters according to Ziegler and Nichols

	Type of controller	Controller parameters		
		K_C	T_i	T_D
Method I	P	$0.5 K_{C\text{crit}}$	-	-
	PI	$0.45 K_{C\text{crit}}$	$0.85 T_{\text{crit}}$	-
	PID	$0.6 K_{C\text{crit}}$	$0.5 T_{\text{crit}}$	$0.12 T_{\text{crit}}$
Method II	P	$\frac{1}{K_P} \frac{T_r}{T_u}$	-	-
	PI	$\frac{0.9}{K_P} \frac{T_r}{T_u}$	$3.33 T_u$	-
	PID	$\frac{1.2}{K_P} \frac{T_r}{T_u}$	$2 T_u$	$0.5 T_u$

Demonstration Example 8.1
[A virtual experiment using PID control for tracking](#)

Demonstration Example 8.2
[A virtual experiment using PID control for high-precision positioning](#)

DYNAST study example 8.1
[PI control of a \$PT_1 T_t\$ plant](#)

DYNAST study example 8.2
[PID control of a \$PT_1 T_t\$ plant](#)

DYNAST study example 8.3
Disturbance response for PI control of a PT_1T_t plant

Module 9

Design of controllers using pole-zero compensators

Module units

9.1	Characteristics in frequency and time domain	9-1
9.2	Controller design using frequency domain characteristics	9-10
9.3	Application of the design using frequency domain characteristics	9-15
9.4	Controller design using the root-locus method	9-18

Module overview. *Frequency-response design is popular primarily because it provides good designs in the face of uncertainty in the plant model. In order to design control systems with good performance the relationships between performance indices in the time and frequency domains are discussed in detail. Based on these, the rules and prerequisites of using a dominant pair of poles are introduced. The systematic design of controllers using lead and lag elements is described in detail and a comprehensive example shows the effectiveness of these methods. The methods are completely conducted using only frequency-response characteristics shown by Bode diagrams. In the sequel, design using the root-locus method counterpart of this design is outlined.*

Module objectives. *When you have completed this module you should be able to:*

1. Judge and measure the performance of control systems in the frequency domain.
2. Design systematically controllers using lead lag elements by Bode diagrams.
3. Apply the root-locus method to design pole zero compensators.

Module prerequisites. *Performance indices, frequency response, Bode diagram, Nyquist diagram, root-locus method.*

9.1 Characteristics in frequency and time domain

In the following the most important characteristics in the frequency domain of the open and closed loop for command inputs for a closed loop having a transfer function with two complex poles will be given. This section is based on the relationship between the frequency characteristics and the performance indices in the time domain for the closed loop introduced in section 7.3.1.

A closed loop showing a step response $h_W(t)$ according to Figure 7.3.1 has a frequency response $G_W(j\omega)$ with a peak as qualitatively shown in Figure 9.1.1. For describing this behaviour the following characteristics mentioned earlier can be used:

- resonant peak frequency ω_p ,
- resonant peak M_r ,
- bandwidth ω_b ,
- phase angle $\varphi_b = \varphi(\omega_b)$.

These characteristics are shown in Figure 9.1.1.

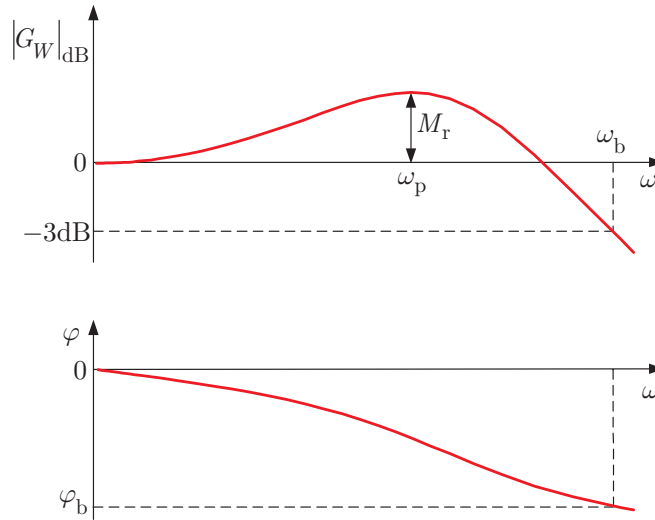


Figure 9.1.1: Bode plot of the closed-loop frequency response

The closed loop for a PT₂S element has the transfer function

$$G_W(s) = \frac{G_0(s)}{1 + G_0(s)} = \frac{\omega_0^2}{s^2 + 2\zeta\omega_0s + \omega_0^2} \tag{9.1.1}$$

according to Eq. (4.4.36) with $K = 1$.

The *natural frequency* ω_0 and the *damping ratio* ζ characterise the control behaviour completely. This can be used as a good approximation for other transfer functions if they contain a *dominant pair of poles* according to Figure 9.1.2.

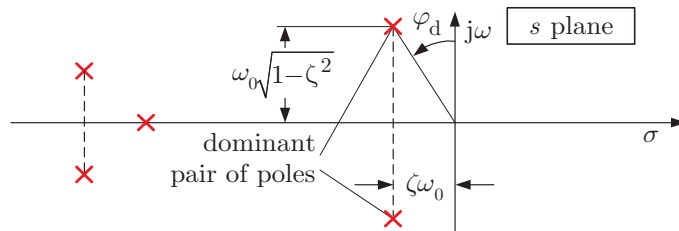


Figure 9.1.2: Distribution of poles of an element with a dominant pair of poles

This pair of poles is assumed to be the closest pair to the $j\omega$ axis in the s domain and therefore it describes the slowest mode and influences the dynamical behaviour of the system very strongly provided the other poles are sufficiently far away on the left-hand side of the s plane.

The step response for the transfer function of Eq. (9.1.1) is

$$h_W(t) = \left\{ 1 - e^{-\zeta\omega_0 t} \left[\cos\left(\sqrt{1-\zeta^2}\omega_0 t\right) + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin\left(\sqrt{1-\zeta^2}\omega_0 t\right) \right] \right\} \sigma(t) \tag{9.1.2a}$$

and according to Eq. (A.3.8) it can be put into the more suitable form

$$h_W(t) = \left\{ 1 - \frac{e^{-\zeta\omega_0 t}}{\sqrt{1-\zeta^2}} \cos \left[\left(\sqrt{1-\zeta^2}\omega_0 t \right) - \varphi_d \right] \right\} \sigma(t), \quad (9.1.2b)$$

where for $\varphi_d = \sin^{-1} \zeta$ or $\zeta = \sin \varphi_d$ is valid. From Eq. (9.1.2b) the weighting function, which follows by differentiation, is

$$g_W(t) = \dot{h}_W(t) = \frac{\omega_0}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_0 t} \sin \left(\sqrt{1-\zeta^2}\omega_0 t \right) \sigma(t). \quad (9.1.3)$$

Therewith the conditions are accomplished in order to determine the maximum overshoot, rise time and settling time that depends on the characteristics in the frequency domain, e.g. natural frequency ω_0 and damping ratio ζ . With ω_0 and ζ the interesting items $A_{W \max \text{dB}}$ and ω_r can be calculated directly by the Eqs. (A.3.1) and (A.3.2).

a) *Determination of the maximum overshoot M_p :*

For calculation of M_p the time $t = t_p > 0$ will be determined at which $\dot{h}_W(t)$ will be first zero according to Eq. (9.1.3). This is when the the sin function in Eq. (9.1.3) has

$$\sqrt{1-\zeta^2}\omega_0 t = \pi.$$

This gives

$$t_p = \frac{\pi}{\omega_0 \sqrt{1-\zeta^2}}. \quad (9.1.4)$$

From Eq. (9.1.2) it follows that the maximum overshoot is

$$M_p = h_W(t_p) - 1 = e^{-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}} = f_1(\zeta). \quad (9.1.5)$$

The maximum overshoot is therefore only a function of the damping ratio ζ as shown in Figure 9.1.3.

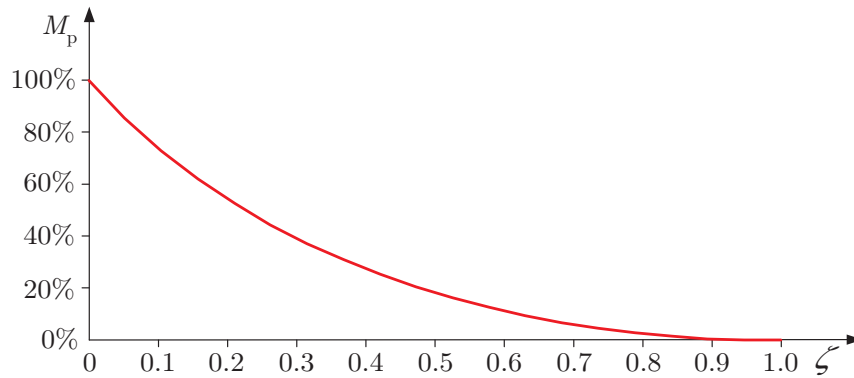


Figure 9.1.3: Maximum overshoot $M_p = f_1(\zeta)$ (in %) relative to $h_{W\infty} = 100\%$ as function of the damping ratio ζ

b) *Determination of the rise time $t_{r,50}$:*

In the following the rise time will not be calculated by the tangent at the turning point, but by the tangent at time $t = t_{50}$ (see Figure 7.3.1), where $h_W(t)$ reaches 50% of the stationary value $h_{W\infty} = 1$. So the time t_{50} must be determined, for which according to Eq. (9.1.2) $h_W(t_{50}) = 0.5$ is valid. From Eq. (9.1.2a) it follows that

$$0.5 = 1 - e^{-\zeta\omega_0 t_{50}} \left[\cos \left(\sqrt{1-\zeta^2}\omega_0 t_{50} \right) + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin \left(\sqrt{1-\zeta^2}\omega_0 t_{50} \right) \right].$$

This equation for the product $\omega_0 t_{50}$ must be evaluated numerically. One gets a function of the form

$$\omega_0 t_{50} = f_2^*(\zeta) . \quad (9.1.6)$$

From Eq. (9.1.3) it follows that

$$t_{r,50} = \frac{1}{\dot{h}_W(t_{50})} = \frac{\sqrt{1-\zeta^2}}{\omega_0 e^{-\zeta\omega_0 t_{50}} \sin\left(\sqrt{1-\zeta^2}\omega_0 t_{50}\right)} ,$$

and from this together with Eq. (9.1.6) the normalised rise time is

$$\omega_0 t_{r,50} = \frac{\sqrt{1-\zeta^2}}{e^{-\zeta f_2^*(\zeta)} \sin\left(\sqrt{1-\zeta^2} f_2^*(\zeta)\right)} = f_2(\zeta) , \quad (9.1.7)$$

which also only depends on the damping ratio ζ . This relationship is shown in Figure 9.1.4.

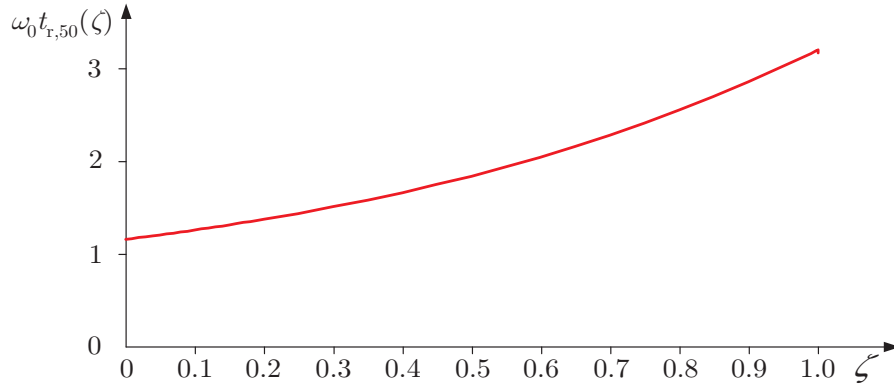


Figure 9.1.4: The product $\omega_0 t_{r,50} = f_2(\zeta)$ (normalised rise time) as a function of the damping ratio ζ

c) *Determination of the settling time t_ε :*

Using Eq. (9.1.2b) the decay of the amplitude to a value less than ε for $t \geq t_\varepsilon$ can be estimated from the envelope of the response

$$\frac{e^{-\zeta\omega_0 t_\varepsilon}}{\sqrt{1-\zeta^2}} \approx \varepsilon .$$

From this the normalised settling time

$$\omega_0 t_\varepsilon \approx \frac{1}{\zeta} \ln \frac{1}{\varepsilon} \frac{1}{\sqrt{1-\zeta^2}} \quad (9.1.8)$$

follows. If $\varepsilon = 3\%$ ($\hat{=} 0.03$) is chosen, one gets

$$\omega_0 t_{3\%} \approx \frac{1}{\zeta} [3.5 - 0.5 \ln(1-\zeta^2)] = f_3(\zeta) . \quad (9.1.9)$$

This relationship is shown in Figure 9.1.5 together with the normalised rise time $\omega_0 t_r$ that will be shown later in Figure 9.1.9.

Comparing the results from Figures 9.1.3 to 9.1.5 one can summarise as follows:

- The maximum overshoot M_p depends only on the damping ratio ζ .
- A change of the damping ratio ζ in the range of approximately $\zeta < 0.9$ behaves contrary to the settling time t_ε compared to the rise time $t_{r,50}$, i.e. an increase of the damping ratio ζ in order to obtain a smaller settling time t_ε increases the rise time $t_{r,50}$.

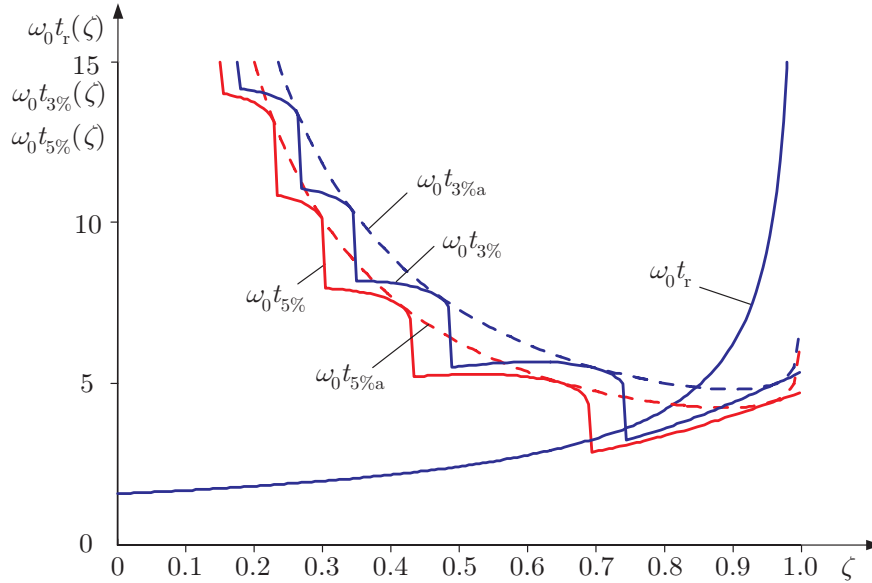


Figure 9.1.5: Normalised settling time $\omega_0 t_{3\%} \approx f_3(\zeta)$ and normalised rise time $\omega_0 t_r$ as functions of the damping ratio ζ

- For a fixed damping ratio ζ the parameter ω_0 determines the speed of the control loop. A large value of ω_0 shows a small settling and rise time.

For the practical application of the diagrams in Figures 9.1.3 to 9.1.5 the following example is given.

Example 9.1.1

The response on step changes in the set-point value $h_W(t)$ of a closed loop with a dominant pair of poles should show a maximum overshoot of $M_p \leq 10\%$, a rise time of $t_{r,50} \leq 1$ s and a settling time of $t_{3\%} \leq 4$ s. How must the damping ratio ζ and the natural frequency ω_0 be chosen?

With the given value of M_p one obtains from Figure 9.1.3 the damping ratio

$$\zeta = 0.58.$$

For this value of ζ with $t_{r,50} = 1$ s the natural frequency

$$\omega_0 = \frac{f_2(0.58)}{1 \text{ s}} = 2.05 \text{ s}^{-1}$$

follows from Figure 9.1.4. But from Figure 9.1.5 for $t_{3\%} = 4$ s the required natural frequency is

$$\omega_0 = \frac{f_3(0.58)}{4 \text{ s}} = 1.6 \text{ s}^{-1}.$$

The rise time of $t_{r,50} = 1$ s is the sharper requirement. Therefore, $\omega_0 = 2.05 \text{ s}^{-1}$ must be chosen. For the pair (ω_0, ζ) from Eq. (A.3.1) the resonant peak frequency

$$\omega_p = \omega_0 \sqrt{1 - 2\zeta^2} = 1.17 \text{ s}^{-1}$$

and from Eq. (A.3.2) the resonant peak

$$M_r = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} = 0.06$$

giving

$$M_r = 0.49 \text{ dB},$$

respectively, can be determined. ■

In order to estimate the bandwidth ω_b for a given damping ratio ζ , the relationship between these two parameters is often needed. Based on the bandwidth ω_b as defined in Figure 4.4.13, that is

$$|G_W(j\omega_b)| = \frac{1}{\sqrt{2}} |G_W(0)| ,$$

it follows after a short calculation using Eq. (9.1.1) for $s = j\omega$ and $\omega = \omega_b$ that

$$\frac{\omega_b}{\omega_0} = \sqrt{1 - 2\zeta^2 + \sqrt{(1 - 2\zeta^2)^2 + 1}} = f_4(\zeta) \quad (9.1.10)$$

and

$$\varphi_b = \tan^{-1} \frac{2\zeta \sqrt{(1 - 2\zeta^2) + \sqrt{(1 - 2\zeta^2)^2 + 1}}}{2\zeta^2 - \sqrt{(1 - 2\zeta^2)^2 + 1}} = f_5(\zeta) . \quad (9.1.11)$$

Furthermore, one obtains using Eqs. (9.1.7) and Eq. (9.1.10)

$$\omega_b t_{r,50} = f_2(\zeta) f_4(\zeta) = f_6(\zeta) . \quad (9.1.12)$$

The graphs of the functions $f_4(\zeta)$, $f_5(\zeta)$ and $f_6(\zeta)$ are shown in Figure 9.1.6.

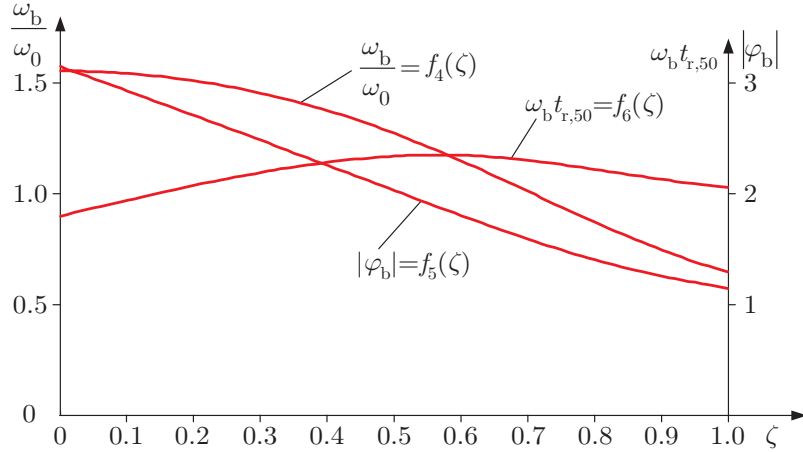


Figure 9.1.6: Characteristics $\omega_b/\omega_0 = f_4(\zeta)$, $\varphi_b = f_5(\zeta)$ and $\omega_b t_{r,50} = f_6(\zeta)$ depending on the damping ratio ζ of the closed loop with PT₂S behaviour

By approximation of $f_4(\zeta)$, $f_5(\zeta)$ and $f_6(\zeta)$ the following 'rules of thumb' can be determined:

$$1. \quad \frac{\omega_b}{\omega_0} \approx 1.84 - 1.21 \zeta \quad \text{for } 0.3 < \zeta < 1.0 , \quad (9.1.13)$$

$$2. \quad |\varphi_b| \approx \pi - 2.23 \zeta \quad \text{for } 0 \leq \zeta \leq 0.8 , \quad (9.1.14)$$

$$3. \quad \omega_b t_{r,50} \approx 2.3 \quad \text{for } 0.3 < \zeta < 0.8 . \quad (9.1.15)$$

Applying these rules to Example 9.1.1 with $\omega_0 = 2.05 \text{ s}^{-1}$ and $\zeta = 0.58$, the bandwidth ω_b can be determined either from Eq. (9.1.13) as

$$\omega_b \approx 2.05 (1.84 - 1.21 \cdot 0.58) = 2.33 \text{ s}^{-1}$$

or with $t_{t,50} = 1 \text{ s}$ from Eq. (9.1.15) as

$$\omega_b \approx 2.3 \text{ s}^{-1} .$$

The Bode plot of a typical corresponding open-loop frequency response $G_0(j\omega)$ is shown in Figure 9.1.7. From this and from Eqs. (5.3.16) and (5.3.17) one can use the characteristics:

- crossover frequency ω_C ,

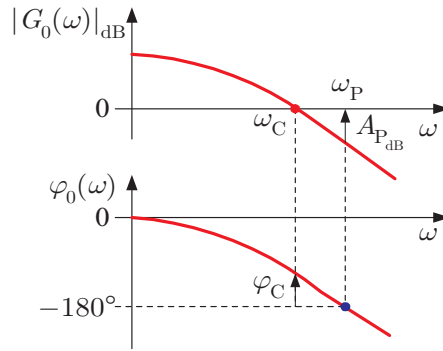


Figure 9.1.7: Bode plot of the open loop

- phase margin $\varphi_C = 180^\circ + \varphi(\omega_C)$,
- gain margin $A_{P\text{dB}} = |G_0(\omega_P)|_{\text{dB}}$.

Since the closed-loop transfer function has been assumed to be approximated by Eq. (9.1.1), the corresponding open-loop transfer function is

$$G_0(s) = \frac{G_W(s)}{1 - G_W(s)} = \frac{\omega_0^2}{s(s + 2\zeta\omega_0)} \tag{9.1.16a}$$

or

$$G_0(s) = \frac{K_0}{s} \frac{1}{1 + Ts} \tag{9.1.16b}$$

with $K_0 = \omega_0/(2\zeta)$ and $T = 1/(2\zeta\omega_0)$. The frequency response of Eq. (9.1.16) is shown in Figure 9.1.8. This Bode plot is considerably different from that of Figure 9.1.7. The system in Figure 9.1.7 does not

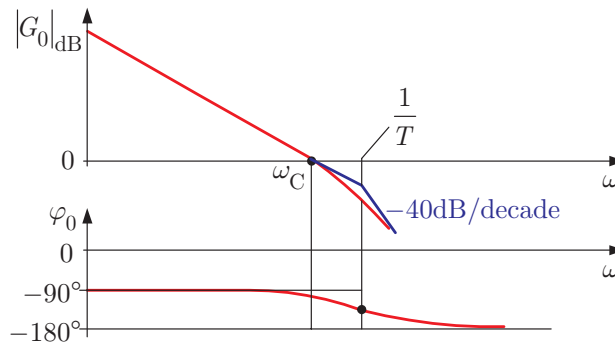


Figure 9.1.8: Bode plot of the open loop with $G_0(s)$ according to Eq. (9.1.16b)

have an integrator. Furthermore, it is of order higher than two, as the phase characteristic exceeds the value of -180° . But close to the crossover frequency ω_C , both Bode plots show a similar behaviour. If for the magnitude response of a given system $|G_0(j\omega)| \gg 1$ is valid for $\omega \ll \omega_D$ and $|G_0(j\omega)| \approx 0$ for $\omega \gg \omega_D$, then $G_0(s)$ can often be approximated in the vicinity of the crossover frequency ω_C by Eqs. (9.1.16a) and (9.1.16b). The associated transfer function $G_W(s)$ contains a dominant conjugate complex pair of poles. In order to transfer the known performance indices of a second-order system to control systems of higher order, the design must be performed such that the magnitude response $|G_0(j\omega)|$ decreases by 20dB/decade in the vicinity of ω_C . For Eq. (9.1.16b) this is only possible if $\omega_C < 1/T$ is valid (compare with Figure 9.1.8). From Eq. (9.1.16a) one obtains under the condition

$$|G_0(j\omega_C)| = 1$$

after a short calculation

$$\frac{\omega_C}{\omega_0} = \sqrt{\sqrt{4\zeta^4 + 1} - 2\zeta^2} = f_7(\zeta) . \tag{9.1.17}$$

With $T = 1/(2\zeta\omega_0)$ for $\omega_C < 1/T$ from

$$\sqrt{\sqrt{4\zeta^4 + 1} - 2\zeta^2} < 2\zeta$$

the condition $\zeta > 0.42$ follows. When for the damping ratio a value of $\zeta > 0.42$ is chosen, then it is guaranteed that the magnitude response $|G_0|_{\text{dB}}$ of the open loop falls off in the vicinity of the crossover frequency ω_C by 20dB/decade. Figure 9.1.9 shows that only the interval $0.5 < \zeta < 0.7$ is a range of suitable damping ratios, since both, the rise time and the maximum overshoot, show acceptable values from the performance index point of view. This also means that the phase and gain margin φ_C and $A_{\text{P dB}}$ show proper values.

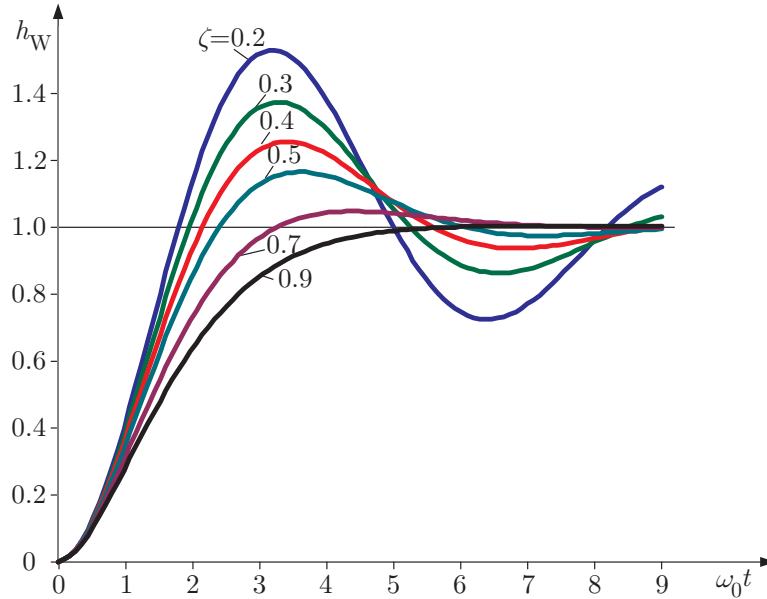


Figure 9.1.9: Step response $h_W(t)$ of the closed loop with PT_2S behaviour according to the transfer function $G_W(s)$ of Eq. (9.1.1)

From these considerations one can conclude that for control systems with minimum-phase behaviour, which can be approximately described by a PT_2S element, the magnitude response $|G_0(j\omega)|_{\text{dB}}$ of the open loop must decrease by 20dB/decade in the vicinity of the crossover frequency ω_C if a good performance is to be achieved, i.e. a sufficient large phase margin φ_C .

As already mentioned in section 5.3.6, the crossover frequency ω_C is an important performance index of the dynamical behaviour of the closed loop. The larger ω_C , the larger is the bandwidth ω_b of $G_W(j\omega)$ in general, and the faster is the reaction to set-point changes. For the frequency response for set-point changes one gets approximately

$$G_W(j\omega) = \frac{G_0(j\omega)}{1 + G_0(j\omega)} \approx \begin{cases} 1 & \text{for } |G_0(j\omega)| \gg 1 \\ G_0(j\omega) & \text{for } |G_0(j\omega)| \ll 1 \end{cases} \quad (9.1.18)$$

From this, the asymptote of the magnitude response of $G_W(j\omega)$ can be determined (Figure 9.1.10). If $|G_0(j\omega)|_{\text{dB}}$ decreases in the vicinity of ω_C by 20dB/decade, then for this range

$$G_0(j\omega) \approx \frac{\omega_C}{j\omega}$$

is valid, and thus it follows that

$$G_W(j\omega) \approx \frac{1}{1 + j\frac{\omega}{\omega_b}} .$$

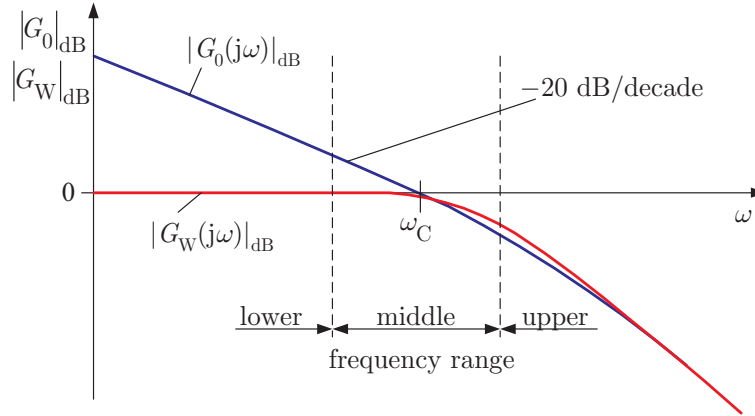


Figure 9.1.10: Piecewise determination of $|G_W(j\omega)|_{dB}$ from $|G_0(j\omega)|_{dB}$ in the Bode diagram

$G_W(j\omega)$ behaves in this range as a PT_1 element. As generally known, the magnitude response of a PT_1 element decreases by 3dB at the breakpoint frequency (here $\omega_B = \omega_C$). Therefore, the crossover frequency ω_C of the open loop is just the bandwidth ω_b of the closed loop, i.e. $\omega_C = \omega_b$. From this it follows that for minimum phase systems the frequency response of $G_W(j\omega)$ can be determined piecewise from $G_0(j\omega)$ according to Figure 9.1.10. Thereby for fulfilling Eq. (9.1.18) in the *lower frequency range* the value of $|G_0(j\omega)|$ and therefore also the loop gain K_0 must be large to hold the steady-state error as small as possible. This lower frequency range of $|G_0(j\omega)|$ is responsible for the steady-state behaviour of the closed loop, whereas the *middle frequency range* is essential for the transient behaviour and is characteristic for the damping. In order to avoid non-suppressable high-frequency disturbances of the set point $w(t)$ in the closed loop, $|G_0(j\omega)|$ and therefore also $|G_W(j\omega)|$ must decrease quickly in the *upper frequency range*.

From these ideas it is now possible to specify besides Eq. (9.1.17) additional important relationships between the characteristics of the time response of the closed loop and the characteristics of the frequency response of the open and partly of the closed loop. Using Eq. (9.1.7) and Eq. (9.1.17) it follows immediately that

$$\omega_{CTr,50} = f_2(\zeta) f_7(\zeta) = f_8(\zeta) . \quad (9.1.19)$$

Figure 9.1.11 shows the graphical representation of $f_8(\zeta)$. It is easy to check that this curve can be described in the range of $0 < \zeta < 1$ by the approximation

$$\omega_{CTr,50} \approx 1.5 - \frac{M_p[\%]}{250} \quad (9.1.20a)$$

or

$$\omega_{CTr,50} \approx 1.5 \quad \text{for} \quad M_p \leq 20\% \quad \text{or} \quad \zeta > 0.5 . \quad (9.1.20b)$$

A further relationship may be determined from the crossover frequency ω_C for the phase margin as

$$\begin{aligned} \varphi_C &= 180^\circ + \arg G_0(j\omega_C) \\ &= 180^\circ - 90^\circ - \tan^{-1} \left(\frac{1}{2\zeta} \frac{\omega_C}{\omega_0} \right) , \end{aligned}$$

which yields

$$\varphi_C = \tan^{-1} \left(2\zeta \frac{\omega_0}{\omega_C} \right) = \tan^{-1} \left[2\zeta \frac{1}{f_7(\zeta)} \right] = f_9(\zeta) . \quad (9.1.21)$$

Figure 9.1.11 also shows this function. By superposition of $f_9(\zeta)$ with $f_1(\zeta)$ one can show that in the range of the mainly interesting values of the damping $0.3 \leq \zeta \leq 0.8$ the approximation

$$\varphi_C[^\circ] + M_p[\%] \approx 70 \quad (9.1.22)$$

is valid. This 'rule of thumb' can only be applied for values of the variables with the given dimensions in squared brackets.

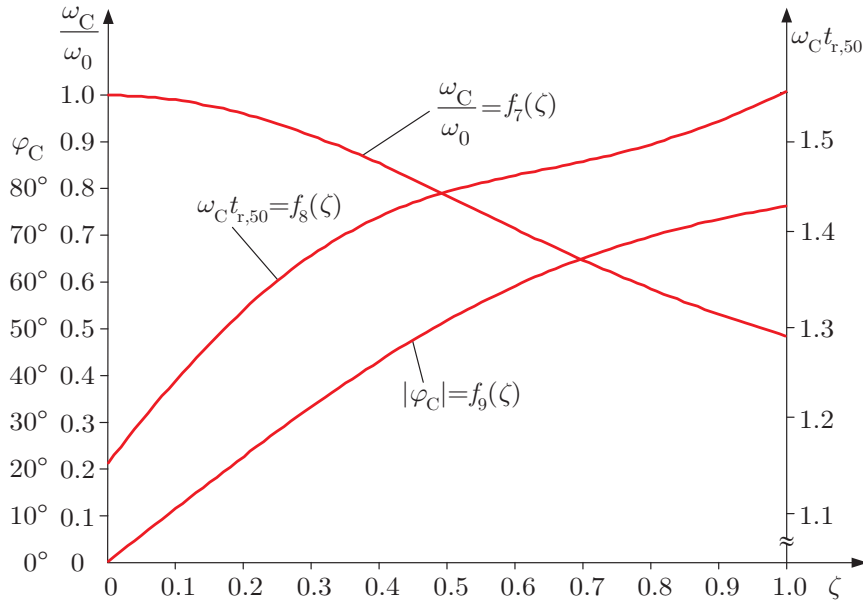


Figure 9.1.11: Frequency domain characteristics of the open loop, ω_C and φ_C , depending on the damping ratio ζ of the closed loop with PT₂S behaviour

9.2 Controller design using frequency domain characteristics

The relations $f_i(\zeta)$ ($i = 1, 2, \dots, 9$) derived in the previous section for the closed loop behaviour of a PT₂S element, can be applied also to higher-order systems as long as these systems have a dominant pair of poles. For this class of systems an efficient synthesis method exists as shown in the following. The starting point of this method is the representation of the frequency response $G_0(j\omega)$ of the open loop on a Bode diagram. The specifications of the closed loop that must be met are first given as characteristics of the open loop according to the above section. The synthesis requires in the choice of a controller transfer function $G_C(s)$, which modifies the open-loop transfer function such that the required characteristics are met. The method consists of the following steps:

Step 1: In general, the characteristics of the time response of the closed loop, M_p , $t_{r,50}$ and e_∞ , are given. On the basis of these values from Table 7.2.1 the gain K_0 , from the rule of thumb for $\omega_C t_{r,50} \approx 1.5$ according to Eq. (9.1.20b), the crossover frequency ω_C and from $\varphi_C [^\circ] \approx 70 - M_p [\%]$ the phase margin φ_C will be determined, and from $f_1(\zeta)$ the damping ratio ζ .

Step 2: First a P element will be chosen as controller such that the gain K_0 determined during step 1 will be met. By inserting additional elements in series (often called *compensator* or *correction elements*) G_0 will be changed such that the other values from step 1, ω_C and φ_C , can be achieved while the amplitude response $|G_0(j\omega)|_{dB}$ decreases by 20dB/decade in the vicinity of the crossover frequency ω_C .

Step 3: It must be checked whether the response meets the required specifications. This can be performed directly by determining M_p , $t_{r,50}$ and e_∞ by simulation, or indirectly by using the formula in section 9.1 for the resonant peak M_r according to Eq. (A.3.2) and the bandwidth ω_b according to Eq. (9.1.15). These values must be verified by calculation of the closed-loop frequency domain characteristic

$$G_W(j\omega) = \frac{G_0(j\omega)}{1 + G_0(j\omega)}$$

from the open-loop characteristic. In the case of too large deviations from the approximations of M_r and ω_b , step 2 must be repeated in a modified form.

This method does not inevitably deliver a proper controller during the first run and it is a trial-and-error method that leads generally to satisfactory results after some recursions.

For the design of this controller the methods given in section 8 for a standard controller are usually not sufficient. The controller must be composed of different elements – as shown above in step 2. In this procedure two special elements are of important interest, which have to perform a phase shift as shown below:

a) The lead element

The increasing phase shift element is used to increase the phase in a certain frequency range. The transfer function of this element is

$$G_C(s) = \frac{1 + Ts}{1 + \alpha Ts} = \frac{1 + \frac{s}{1/T}}{1 + \frac{s}{1/(\alpha T)}}, \quad 0 < \alpha < 1. \quad (9.2.1)$$

For $s = j\omega$ the frequency response

$$G_C(j\omega) = \frac{1 + j\frac{\omega}{\omega_Z}}{1 + j\frac{\omega}{\omega_N}} \quad (9.2.2)$$

follows with the two breakpoint frequencies

$$\omega_Z = \frac{1}{T} \quad (9.2.3a)$$

and

$$\omega_N = \frac{1}{\alpha T}. \quad (9.2.3b)$$

A further characteristic is the frequency ratio

$$m_h = \frac{\omega_N}{\omega_Z} = \frac{1}{\alpha} > 1. \quad (9.2.4)$$

From Eq. (9.2.1) the frequency response

$$\begin{aligned} G_C(j\omega) &= A(\omega) e^{j\varphi(\omega)} \\ &= \sqrt{\frac{1 + T^2\omega^2}{1 + \alpha^2 T^2\omega^2}} e^{j(\tan^{-1} T\omega - \tan^{-1} \alpha T\omega)} \end{aligned} \quad (9.2.5)$$

follows. The Nyquist plot is shown in Figure 9.2.1 and it is a semicircle. The maximum phase shift

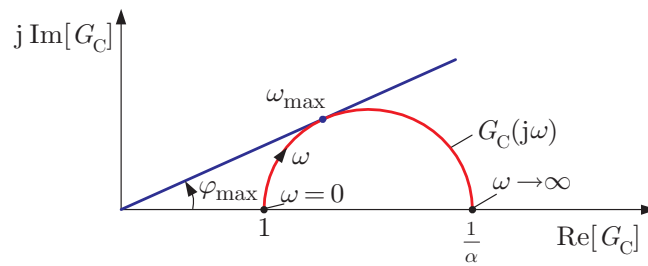


Figure 9.2.1: Nyquist plot of a lead element

$$\varphi(\omega) = \tan^{-1} T\omega - \tan^{-1} \alpha T\omega$$

can be determined from the condition $d\varphi(\omega)/d\omega = 0$ for

$$\omega_{\max} = \sqrt{\omega_Z \omega_N} = \omega_N \frac{1}{\sqrt{m_h}} = \omega_Z \sqrt{m_h} = \frac{1}{T} \sqrt{m_h}. \quad (9.2.6)$$

As shown by the Bode plot in Figure 9.2.2 the lead element has at high frequencies an undesirable increase in the magnitude response of

$$|\Delta G_C|_{\text{dB}} = 20 \log_{10}(1/\alpha) = 20 \log_{10} m_h.$$

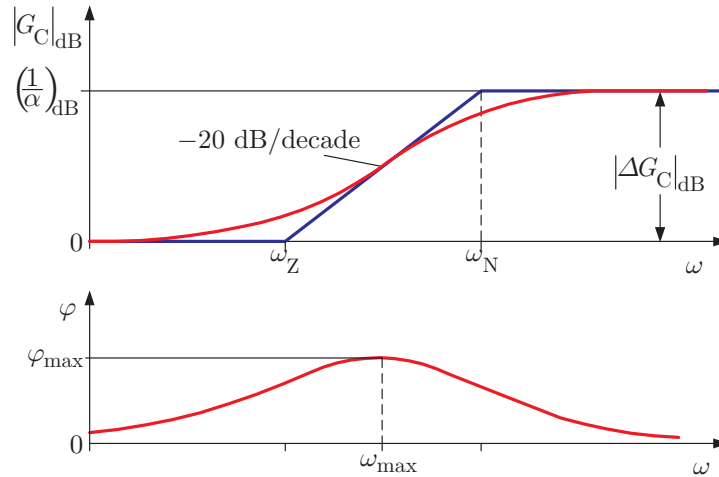


Figure 9.2.2: Bode diagram of the lead element

If Eq. (9.2.1) is broken down as

$$\begin{aligned} G_C(s) &= \frac{1 + \alpha Ts}{1 + \alpha Ts} + \frac{(1 - \alpha)Ts}{1 + \alpha Ts} \\ &= 1 + \left(\frac{1}{\alpha} - 1\right) \frac{\alpha Ts}{1 + \alpha Ts}, \end{aligned} \quad (9.2.7)$$

the lead element consists of a parallel connection of a P element with gain 1 and a DT₁ element, which is a special PDT₁ controller (compare Eq. (8.1.9)). For the step response one obtains

$$h_C(t) = \sigma(t) \left[1 + \left(\frac{1}{\alpha} - 1\right) e^{-\frac{t}{\alpha T}} \right], \quad (9.2.8)$$

which is shown in Figure 9.2.3. For the practical design of lead elements the normalised phase diagram

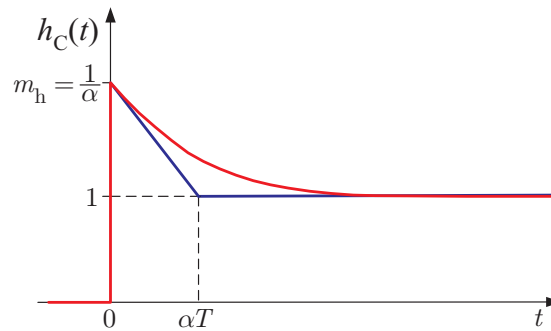


Figure 9.2.3: Step response of the lead element

in Figure 9.2.4 is helpful. If the frequency ω_{\max} is known, from this diagram the frequency ratio m_h can be determined. The lower breakpoint frequency ω_Z can be either read from the diagram directly or calculated from Eq. (9.2.6).

Example 9.2.1

The phase response of a transfer function must be shifted by $\Delta\varphi = 30^\circ$ at $\omega = \omega_{\max} = 4 \text{ s}^{-1}$. The maximum of the phase shift of $\Delta\varphi = 30^\circ$ is from Figure 9.2.4 for $\bar{\omega} = \omega/\omega_Z \approx 1.7$ and $m_h = 3$. With $\omega = \omega_{\max} = 4 \text{ s}^{-1}$ follows for the lower breakpoint frequency $\omega_Z \approx \omega_{\max}/1.7$ or from Eq. (9.2.6) $\omega_Z = \omega_{\max}/\sqrt{m_h} = 4/\sqrt{3} = 2.31 \text{ s}^{-1}$ and with Eq. (9.2.4) for the upper breakpoint frequency $\omega_N = m_h\omega_Z = 6.93 \text{ s}^{-1}$. ■

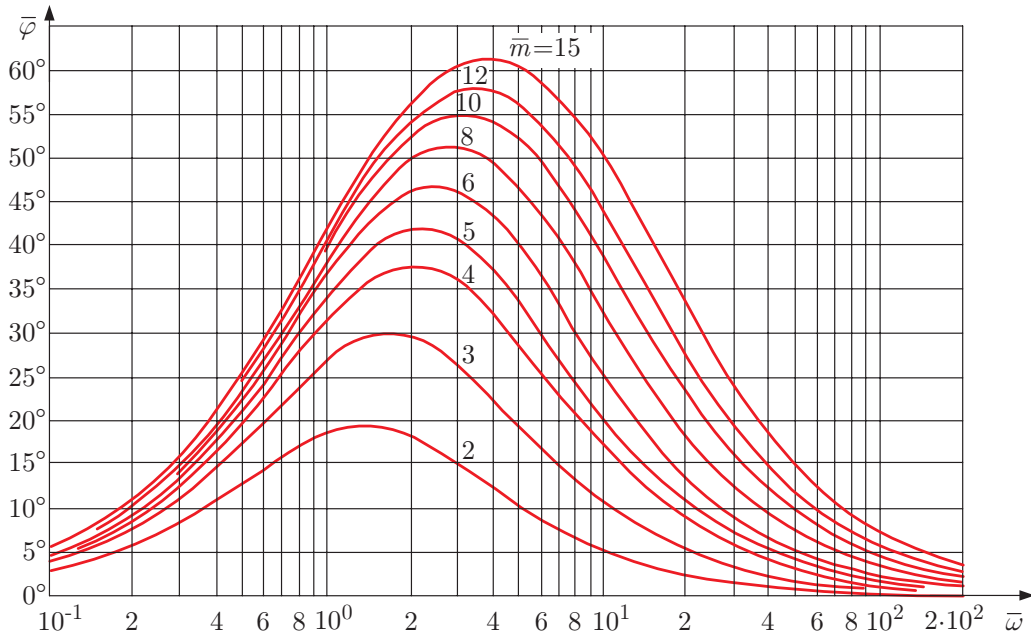


Figure 9.2.4: Normalised phase responses of the lead element:

$\bar{\varphi} = \varphi$; $\bar{\omega} = \omega/\omega_Z$; $\bar{m} = m_h$;

ω_Z = lower breakpoint frequency; $m_h\omega_Z$ = upper breakpoint frequency;
normalised phase responses of the lag element:

$\bar{\varphi} = -\varphi$; $\bar{\omega} = \omega/\omega_N$; $\bar{m} = m_s$;

ω_N = lower breakpoint frequency; $m_s\omega_N$ = upper breakpoint frequency

b) The lag element

The lag element is used to decrease the magnitude response above a certain frequency. Hereby a undesirable decrease of the phase response occurs in a certain frequency range. The transfer function of this lag element is

$$G_C(s) = \frac{1 + Ts}{1 + \alpha Ts} = \frac{1 + \frac{s}{(1/T)}}{1 + \frac{s}{(1/\alpha T)}} \quad \text{with } \alpha > 1 . \tag{9.2.9}$$

For $s = j\omega$ and the breakpoint frequencies $\omega_Z = \frac{1}{T}$ and $\omega_N = \frac{1}{\alpha T}$ the frequency response is

$$G_C(j\omega) = \frac{1 + j\frac{\omega}{\omega_Z}}{1 + j\frac{\omega}{\omega_N}} . \tag{9.2.10}$$

Also in this case a frequency ratio can be defined as

$$m_s = \frac{\omega_Z}{\omega_N} = \alpha > 1 . \tag{9.2.11}$$

The decrease of the amplitude response at high frequencies is

$$|\Delta G_C|_{dB} = -20 \log_{10} \frac{1}{\alpha} = 20 \log_{10} m_s . \tag{9.2.12}$$

Figure 9.2.5 shows the Nyquist plot and Figure 9.2.6 the Bode diagram of the lag element. The rearrangement of Eq. (9.2.9) into

$$G_C(s) = \frac{1}{\alpha} + \frac{1 - \frac{1}{\alpha}}{1 + \alpha Ts} \tag{9.2.13}$$

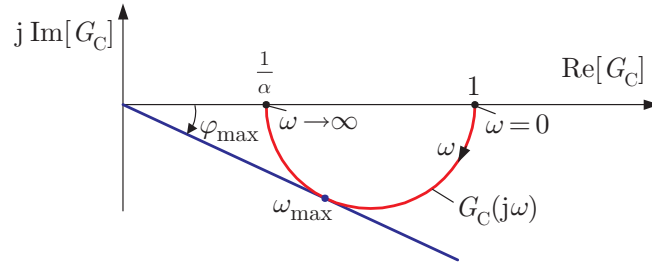


Figure 9.2.5: Nyquist plot of the lag element

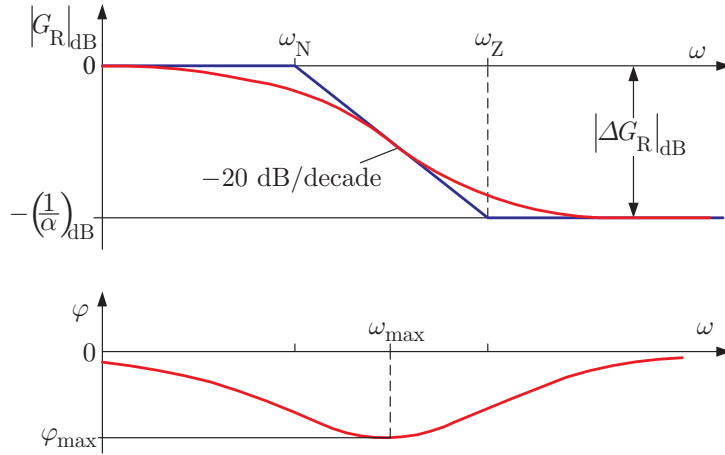


Figure 9.2.6: Bode diagram of the lag element

shows that the lag element consists of a parallel connection of a P element with gain $1/\alpha$ and a PT₁ element with gain $(1 - 1/\alpha)$ and time constant αT . The step response of this lag element follows from Eq. (9.2.13) as

$$h_C(t) = \sigma(t) \left[\frac{1}{\alpha} + \left(1 - \frac{1}{\alpha}\right) \left(1 - e^{-\frac{t}{\alpha T}}\right) \right] \tag{9.2.14}$$

and is shown in Figure 9.2.7. It is easy to see that this relation is equal to Eq. (9.2.8) but with $\alpha > 1$.

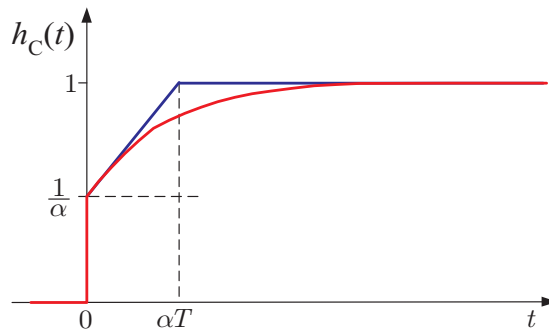


Figure 9.2.7: Step response of the lag element

For practical working the phase diagram of Figure 9.2.4 can be used, which is in this case in principle the same as that for the lead element but with different parameters and flipped over.

Example 9.2.2

The magnitude response $|G_0|_{dB}$ of an open-loop system should be decreased at $\omega = 10 \text{ s}^{-1}$ by 20dB, whereby the maximum phase shift must be 10° . From Eq. (9.2.12) it follows that $|\Delta G_C|_{dB} = 20 \text{ dB} = 20 \log_{10} m_s$ and from this $m_s = 10$. With $\varphi = -10^\circ$ and $m_s = 10$ one obtains from the phase response

$\bar{\omega} = \omega/\omega_N \approx 50$, and with $\omega = 10 \text{ s}^{-1}$ for the breakpoint frequencies $\omega_N = 10/50 \text{ s}^{-1} = 0.2 \text{ s}^{-1}$ and $\omega_Z = m_s \omega_N = 2 \text{ s}^{-1}$. ■

9.3 Application of the design using frequency domain characteristics

In the following the design using the frequency domain characteristics will be demonstrated by an example. The plant is given with the transfer function

$$G_P(s) = \frac{1}{(1+s) \left(1 + \frac{s}{3}\right)}. \quad (9.3.1)$$

For the step response of set-point changes $h_W(t)$ of the closed loop the rise time

$$t_{r,50} = 0.7 \text{ s} \quad (9.3.2a)$$

and the maximum overshoot

$$M_p = 25\% \quad (9.3.2b)$$

is specified. Furthermore, for a ramp command input

$$w(t) = w_1 \sigma(t) t$$

of the set point, the closed loop should show a steady-state error of

$$e_\infty = \frac{1}{20}. \quad (9.3.2c)$$

The design method will be performed according to section 9.2, assuming that the resultant system will have a dominant pair of complex poles, using the following steps:

Step 1:

One obtains from Eq. (9.1.20) with the specification according to Eq. (9.3.2a) the approximation of the crossover frequency as

$$\omega_C \approx \frac{1}{t_{r,50}} \left(1.5 - \frac{M_p[\%]}{250}\right) = \frac{1}{0.7} (1.5 - 0.1) \approx 2 \text{ s}^{-1} \quad (9.3.3a)$$

and from Eq. (9.1.22) with Eq. (9.3.2b) the phase margin as

$$\varphi_C[^\circ] \approx 70 - M_p[\%] = 45. \quad (9.3.3b)$$

From the specification Eq. (9.3.2c) one obtains for a ramp command input according to Table 7.2.1 (case $k = 1$) for $x_{e1} \equiv w_1 = 1$ the open-loop gain as

$$K_0 = K_C K_P = 20. \quad (9.3.3c)$$

Step 2:

- A)** First, an I controller is chosen, $G_{C_1}(s) = K_C/s$, with the gain K_C such that Eq. (9.3.3c) is fulfilled, here $K_C = 20$. The Bode diagram of the open loop transfer function

$$G_{0_1}(s) = G_{C_1}(s) G_P(s) = \frac{20}{s(1+s) \left(1 + \frac{s}{3}\right)} \quad (9.3.4)$$

is plotted in Figure 9.3.1. From this it can be seen that in order to achieve the requirements of Eqs. (9.3.3a) and (9.3.3b)

1. the phase of $G_{0_1}(j\omega)$ must be increased at $\omega = \omega_C$ by 53° , and
2. the magnitude of $G_{0_1}(j\omega)$ decreased at $\omega = \omega_C$ by 11dB .

B) In order to fulfill the first requirement, the I controller will be extended by a lead element, whose phase response at $\omega = \omega_C = 2 \text{ s}^{-1}$ has a maximum of $(53^\circ + 6^\circ)$. Here in this case a 6° larger value is used, as by the usage of a lag element in step 3 a small unavoidable decrease of the phase occurs. From the phase diagram (Figure 9.2.4) one reads for $\varphi_{\max} = 59^\circ$ the frequency ratio of

$$m_h \approx 12 .$$

From this, one obtains for $\omega = \omega_{\max} = \omega_C$ with Eq. (9.2.6) or likewise from Figure 9.2.4 the breakpoint frequencies

$$\omega_Z = \frac{\omega_C}{\sqrt{m_h}} \approx 0.6 \text{ s}^{-1}$$

and

$$\omega_N = \omega_Z m_h \approx 7.2 \text{ s}^{-1} .$$

The transfer function of the extended controller is

$$G_{C_2}(s) = 20 \frac{1 + \frac{s}{0.6}}{s(1 + \frac{s}{7.2})} . \quad (9.3.5)$$

The open loop has now the transfer function

$$G_{0_2}(s) = G_{C_2}(s) G_P(s) = 20 \frac{1 + \frac{s}{0.6}}{s(1+s) \left(1 + \frac{s}{3}\right) \left(1 + \frac{s}{7.2}\right)} . \quad (9.3.6)$$

The associated frequency-domain characteristics are shown in Figure 9.3.1. Because of the additional lead element also the magnitude response of $G_{0_2}(s)$ is slightly changed. Hence, in the following step the magnitude response at $\omega = \omega_C$ must be decreased by 22dB instead of 11dB.

C) In order to reach this decrease of the magnitude at $\omega = \omega_C = 2 \text{ s}^{-1}$ the controller will be further extended by a lag element. From Eq. (9.2.12) it follows that

$$20 \log_{10} m_s = 22 \text{ dB}$$

and from this

$$m_s = 12.6 .$$

The phase at $\omega_C = 2 \text{ s}^{-1}$ must not be influenced too much such that the upper breakpoint frequency ω_Z , and therefore also the lower breakpoint frequency ω_N must be sufficiently far left from ω_C . By the special choice of the lead element in step 1, the lag element introduced during this step may have a maximum phase decrease of 6° . From this, one obtains with $m_s = 12.6$ from Figure 9.2.4

$$\frac{\omega}{\omega_N} = 125$$

and especially for $\omega = \omega_C = 2 \text{ s}^{-1}$

$$\omega_N = \frac{\omega_C}{125} = 0.016 \text{ s}^{-1} .$$

For the upper breakpoint frequency it follows that

$$\omega_Z = \omega_N m_s = 0.2 \text{ s}^{-1} .$$

The transfer function of the final controller is given by

$$G_C(s) = 20 \frac{1 + \frac{s}{0.6}}{s(1 + \frac{s}{7.2})} \frac{1 + \frac{s}{0.2}}{1 + \frac{s}{0.016}} . \quad (9.3.7)$$

The transfer function of the open loop is

$$G_0(s) = 20 \frac{\left(1 + \frac{s}{0.2}\right) \left(1 + \frac{s}{0.6}\right)}{s \left(1 + \frac{s}{0.016}\right) (1 + s) \left(1 + \frac{s}{3}\right) \left(1 + \frac{s}{7.2}\right)}$$

The associated frequency response characteristics are shown in Figure 9.3.1.

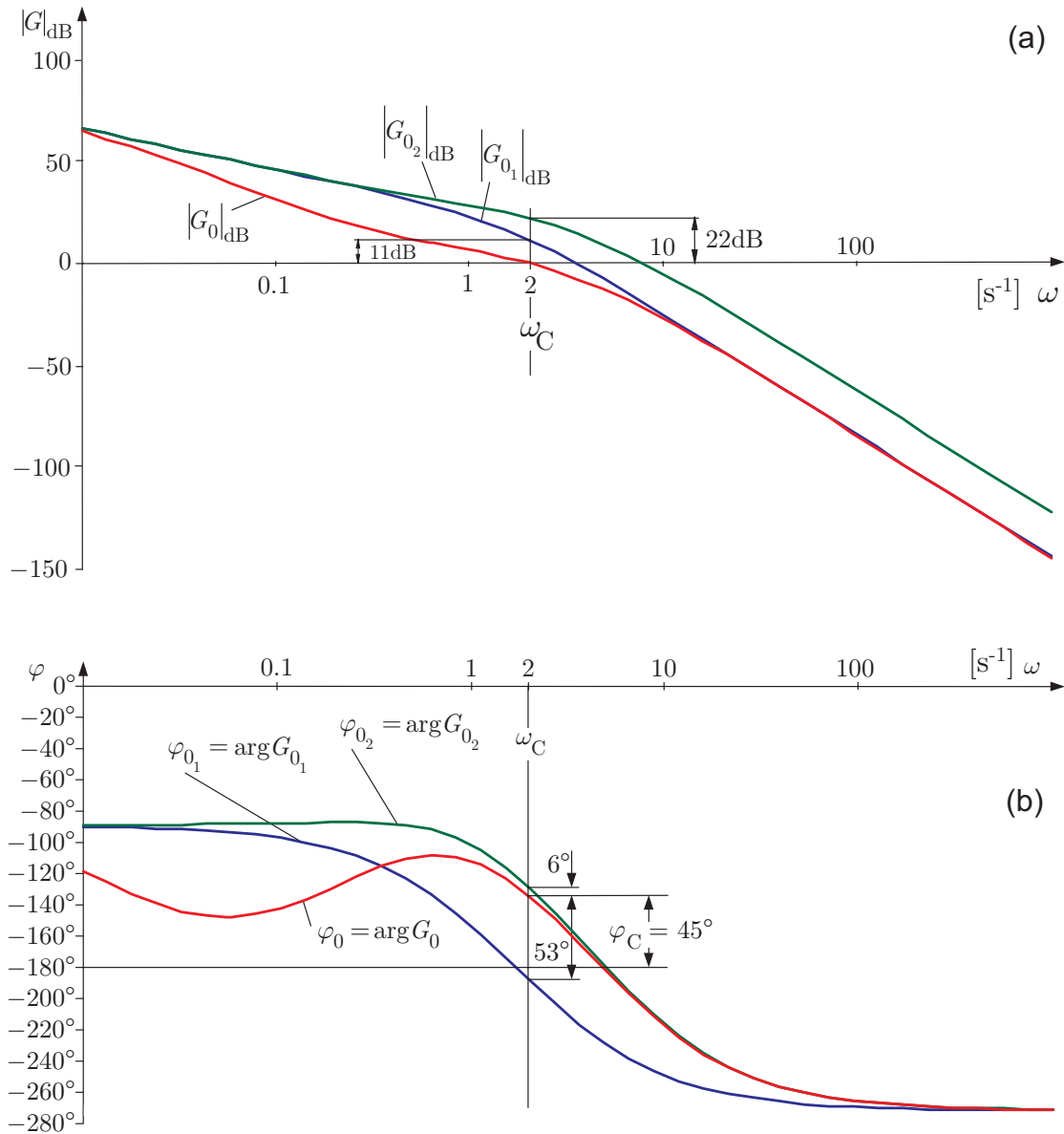


Figure 9.3.1: The magnitude (a) and phase (b) response of $G_{01}(j\omega)$, $G_{02}(j\omega)$ and $G_0(j\omega)$

Step 3:

As the design based on frequency-domain characteristics is an approximate method, one should verify the results by simulation studies and make sure of the match of the the results with the initial specifications. The simulation results for the time responses are shown in Figure 9.3.2. The required specifications for $t_{r,50}$, M_p and e_∞ are fulfilled. Furthermore, the manipulated variable $u(t)$ must be checked for realisability. If this in not the case, the initial specifications for the control system must be modified.

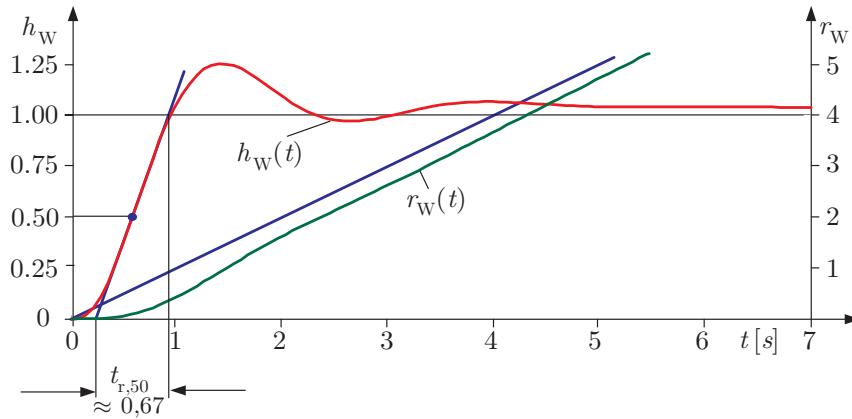


Figure 9.3.2: Step response $h_W(t)$ and ramp response $r_W(t)$ of the designed control system

9.4 Controller design using the root-locus method

The design using the root-locus method is directly connected to the considerations in section 9. There the specifications of maximum overshoot, rise and settling time for a closed loop having a dominant pair of poles have been converted into the conditions for the damping ratio ζ and for the natural frequency ω_0 of the related transfer function $G_W(s)$. With ζ and ω_0 the poles of the transfer function G_W are directly tightened as shown in Figure 9.1.2. An open-loop transfer function $G_0(s)$ must now be determined such that the closed loop has a dominant pair of poles at the desired position, which is given by the values of ω_0 and ζ . Such an approach is also called *pole assignment*.

The root-locus method is – as generally known – a graphical method, which is used to analyse the position of the closed-loop poles. This method offers the possibility to combine in the complex s plane the desired dominant pair of poles with the root locus of the fixed part of the loop and to deform the root locus by adding poles and zeros such that two of the branches traverse through the desired dominant pair of poles at a certain gain K_0 . Figure 9.4.1 shows, how the root locus can be deformed to the right by adding a pole and to the left by adding a zero.

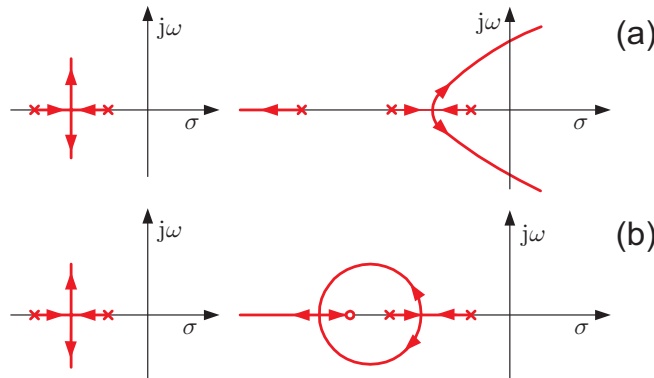


Figure 9.4.1: Deforming of the root locus (a) to the right by an additional pole, (b) to the left by an additional zero in the open loop

The principal strategy during the controller design by the root-locus method will be shown in the following using two examples.

Example 9.4.1

Given is a plant described by the transfer function

$$G_P(s) = \frac{1}{s(s + 3)(s + 5)} . \tag{9.4.1}$$

For this plant a controller must be designed, such that the step response $h_W(t)$ of the closed loop shows the following properties:

$$M_p = 16\% \quad \text{and} \quad t_{r,50} = 0.6 \text{ s} .$$

First, these specifications will be transformed. The conditions for ζ and ω_0 are from Figures 9.1.3 and 9.1.4

$$\zeta \geq 0.5 \quad \text{and} \quad \omega_0 \geq \frac{1.85}{0.6 \text{ s}} = 3.1 \text{ s}^{-1} .$$

In order to have a geometrical interpretation one should consider Figure 9.4.2, where a pair of complex poles

$$s_{a,b} = -\zeta\omega_0 \pm j\omega_0\sqrt{1-\zeta^2}$$

is shown. The distance d^* of both poles $s_{a,b}$ from the origin is

$$d^* = \sqrt{\omega_0^2\zeta^2 + (1-\zeta^2)\omega_0^2} = \omega_0 . \tag{9.4.2}$$

The angle α is

$$\cos \alpha = \frac{\omega_0\zeta}{\omega_0} = \zeta \tag{9.4.3a}$$

or

$$\alpha = \cos^{-1} \zeta , \tag{9.4.3b}$$

where for the current case of $\zeta \geq 0.5$ the condition $\alpha \geq 60^\circ$ is met. The damping ratio ζ describes the angle α , the frequency ω_0 the distance d^* of the dominant pair of poles from the origin.

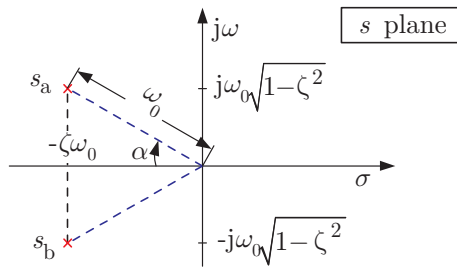


Figure 9.4.2: Pair of conjugate complex poles in the s plane

During the design one will try to increase the damping ratio not unnecessarily high as this step causes an increase in the rise time $t_{r,50}$ for a given natural frequency ω_0 (Figure 9.1.4). Increasing the natural frequency ω_0 implies an increase of the speed of the control loop. But this parameter should not be unnecessarily increased, otherwise the dominance of the pair of poles may be lost.

Figure 9.4.3 shows the root locus of the closed loop using a P controller. Potential positions for the dominant pair of poles are drawn by the two thick blue lines H_1 and H_2 . It is obvious that the design using a pure P controller (changes in the gain K_0) does not lead to the goal, as the root locus does not traverse the two lines H_1 and H_2 . Equally it is clear which steps have to be taken, such that the two branches of the root locus under consideration traverse the two lines H_1 and H_2 . If the two poles $s_2 = -3$ and $s_3 = -5$ are shifted further left, the centre of gravity of the poles will move left and with it the total curve without changing the structure of the system. One possibility is to perform this shift by a simple lead element which compensates the pole $s_2 = -3$ by a zero and a pole $s_4 = -10$. The resulting controller transfer function is

$$G_C(s) = K_C \frac{1 + s/3}{1 + s/10} = 3.33 K_C \frac{s + 3}{s + 10} \tag{9.4.4}$$

and the transfer function of the modified open loop

$$G_0(s) = 3.33 K_C \frac{1}{s(s + 10)(s + 5)} . \tag{9.4.5}$$

The root locus of the closed loop is shown in Figure 9.4.4. It traverses the line H_1 at s_k . The associated

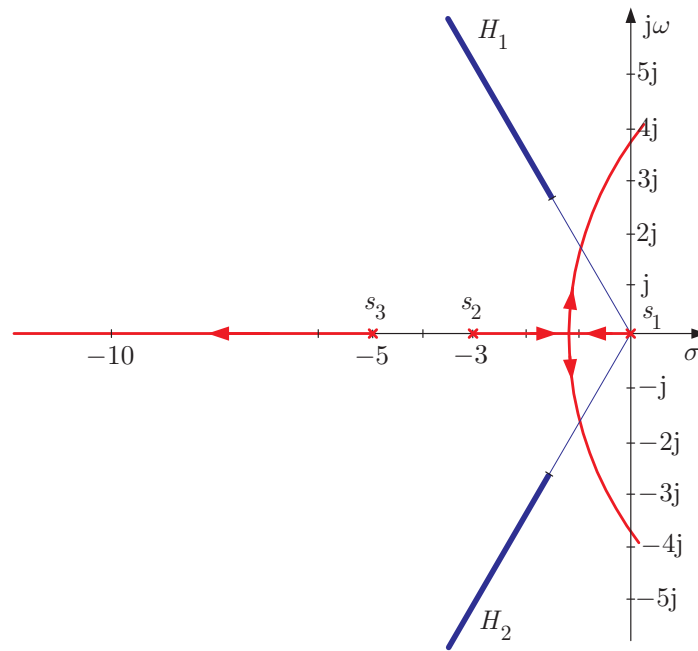


Figure 9.4.3: Root locus of $G_W(s)$ (plant with P controller) and potential positions of the dominant pairs of poles (blue thick lines)

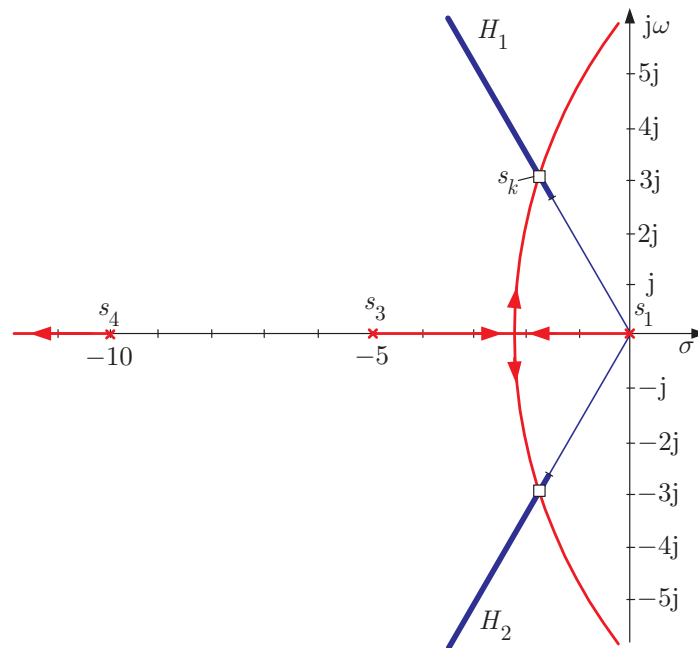


Figure 9.4.4: Root locus of $G_W(s)$ with modified controller according to Eq. (9.4.4)

gain at this point can be determined via the distances to the three poles from Eq. (6.2.11)

$$k_0 = 3.33 K_C = |s_k - 0| |s_k + 5| |s_k + 10| = 3.3 \cdot 4.4 \cdot 8.8$$

as

$$K_C = 38.4 .$$

The step response of the closed loop in Figure 9.4.5 shows that the specifications are achieved.

It must be mentioned that a complete compensation (cancelling) of the pole $s_2 = -3$ cannot be realised exactly, as the plant parameters are not exactly known or may change within some bounds. Thus the

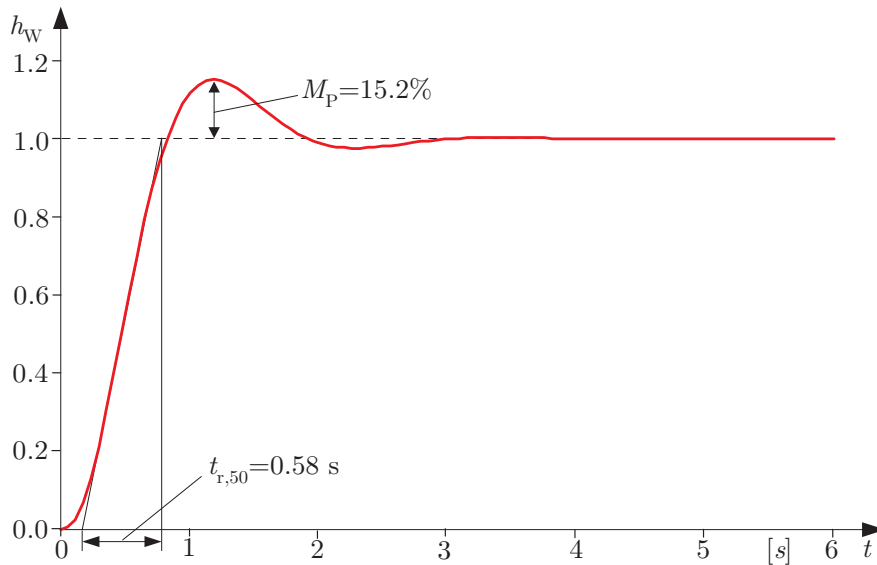


Figure 9.4.5: Closed-loop step response with modified controller according to Eq. (9.4.4)

root locus will differ in the vicinity of the compensated pole from the ideal case of Figure 9.4.4, but intersection with the two lines H_1 and H_2 is almost invariant. ■

One realises by means of this example that the root-locus method is well suited to provide a fast and first overview about the principal possibilities of corrections. Often the knowledge about the asymptotes is already sufficient. The root-locus method is especially suited for stabilising unstable plants. This will be explained by the second example.

Example 9.4.2

Given is an unstable plant with the transfer function

$$G_P(s) = \frac{1}{(s + 1)(s + 5)(s - 1)} . \tag{9.4.6}$$

At first, it is close to compensate the pole $s_1 = 1$ by a corresponding zero. This would be possible, e.g. using a first-order all-pass element with the transfer function

$$G_C(s) = \frac{s - 1}{s + 1} .$$

But because of the reason given above, the compensation of the unstable pole will be practically never complete. So it must be performed without this procedure for stability reasons.

Another possibility is to provide feed back around the unstable plant so that the original unstable pole in the closed loop is shifted into the left-half s plane. A simple P controller would yield the root locus shown in Figure 9.4.6. This configuration is not stabilisable, as the two branches on the right-hand side remain in the right-half s plane for all gain values. But if a controller is used that has a double zero at $s = -1$ and a pole at $s = 0$, the pole at $s = s_2 = -1$ will be substituted by a zero. The root locus will be deformed to the left as shown in Figure 9.4.7.

This distribution of poles and zeros can be simply realised by a PID controller with the transfer function

$$\begin{aligned} G_C(s) &= K_C \frac{(s + 1)^2}{s} = K_C \left(2 + \frac{1}{s} + s \right) \\ &= 2K_C \left(1 + \frac{1}{2s} + 0.5s \right) . \end{aligned}$$

From Figure 9.4.7 it can be seen for higher gain values than the critical gain $K_{0\text{crit}}$ the closed loop is stable, as then all poles remain in the left-half s plane.

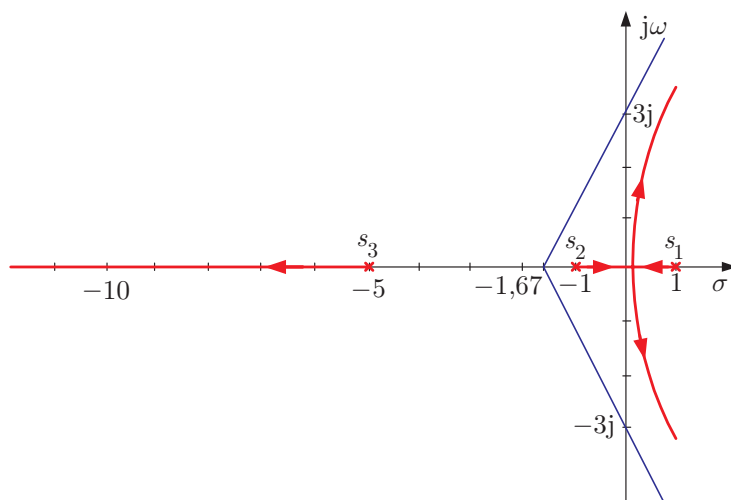


Figure 9.4.6: Root locus of the closed loop consisting of an unstable plant and a P controller

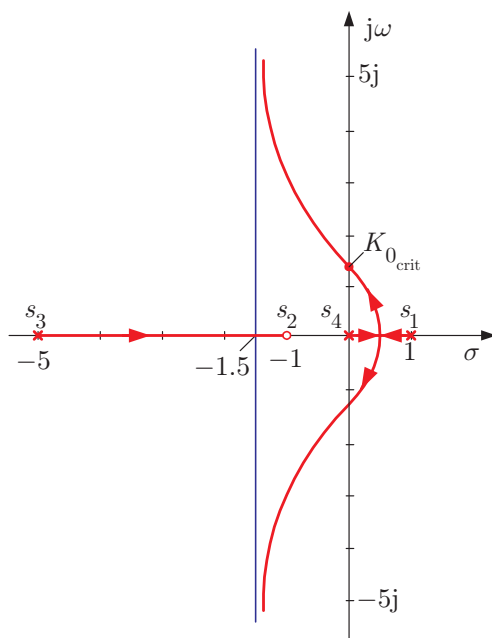


Figure 9.4.7: Root locus of the closed loop consisting of an unstable plant and a PID controller

As the stability of the closed-loop system is not influenced by the left-half-plane poles, a compensation of these poles is possible. Even if this compensation is not completely possible, the system will be stable. A compensation of the right-half-plane poles – as discussed above – should not be done. ■

Demonstration Example 9.1
 A virtual experiment stabilising a pendulum

Module 10

Compensator design methods

Module units

10.1 Basic ideas of compensator design	10-2
10.2 Design by specifying the closed-loop transfer function	10-2
10.3 The method of Truxal and Guillemin	10-3
10.4 Generalised compensator design method	10-7
10.4.1 The basic idea	10-7
10.4.2 Zeros of the closed loop	10-9
10.4.3 The synthesis equations	10-11
10.4.4 Application of the method	10-12
10.5 Compensator design for reference and disturbances	10-16
10.5.1 Structure of the closed loop	10-16
10.5.2 The design procedure	10-17
10.5.3 Design of the pre-filter	10-21
10.5.4 Application of the design method	10-23

Module overview. *Analytical design methods are of interest when the plant model is given as a transfer function and the plant parameters are known. Whereas the pole zero compensator systematic design procedures are indirect methods based on performance indices and their graphical interpretation occur in Bode diagrams, the compensator designs given here are direct methods based on a given transfer function for the closed loop. First, an introduction into the specifications of closed-loop transfer functions is given, then the method of Truxal and Guillemin including different examples of stable, unstable and non-minimum phase plants is shown. This method is the basis of a generalised compensator design procedure that can cope simultaneously both with disturbances and command changes, where the disturbance can occur at different locations. A comprehensive example demonstrates the application.*

Module objectives. *When you have completed this module you should be able to:*

1. Use standard forms to specify the closed-loop behaviour.
2. Design compensator controllers for stable, unstable and non-minimum phase plants.
3. Cope with disturbances for compensator-based control systems.

Module prerequisites. *Transfer function, performance indices.*

10.1 Basic ideas of compensator design

In the following some analytical design methods will be discussed, which will lead directly to the design solution in a strongly systematic way. In contrary to these *direct* design methods, the methods hitherto discussed, e.g. the design method using frequency-domain characteristics or the root-locus method, are *indirect* methods based more on systematic trial and error techniques iterating through some design steps. The success depends strongly on the experience and skill of the designer. The starting point was always the open loop, which was modified iteratively by adding lead and lag elements until the closed loop shows the desired behaviour.

Whereas in the direct design methods one will always start from the behaviour of the closed loop. Mostly a desired transfer function $G_W(s) \equiv K_W(s)$ is given. In general this follows from the specification of some performance indices, which, for example are required for the step response $h_W(t)$. For a series of appropriate transfer functions a table of numerator and denominator polynomials of the associated transfer function $K_W(s)$ and its distribution of zeros and poles to yield a specific response are given. Then for a known plant behaviour the required controller can be directly calculated.

The controllers designed in this way are not always optimal. They guarantee the compliance of the desired specification, e.g. maximum overshoot and settling time. A drawback of these methods is that they cannot be applied directly to systems with dead time.

10.2 Design by specifying the closed-loop transfer function

The desired transfer function of the closed loop is given by

$$K_W(s) = \frac{\alpha(s)}{\beta(s)} = \frac{\alpha_0 + \alpha_1 s + \dots + \alpha_v s^v}{\beta_0 + \beta_1 s + \dots + \beta_u s^u}, \quad u > v, \quad (10.2.1)$$

where $\alpha(s)$ and $\beta(s)$ are polynomials in s . In the following design methods, the distribution of poles and zeros of $K_W(s)$ will be chosen, such that the performance indices for the step response $h_W(t)$ are fulfilled. Often a detailed investigation of the distribution of poles and zeros of the desired transfer function is not necessary, especially when the transfer function does not contain zeros and when – because of the requirement $K_W(0) \equiv 1$ – just yields $\alpha_0 = \beta_0$ and in the simplest case

$$K_W(s) = \frac{\beta_0}{\beta_0 + \beta_1 s + \dots + \beta_u s^u}. \quad (10.2.2)$$

For a closed loop with the transfer function according to Eq. (10.2.2) different possibilities exist, the so called *standard forms*, which can be used by a table lookup for the step response $h_W(t)$, distribution of poles of $K_W(s)$ and the coefficients of the denominator polynomial $\beta(s)$.

A first possibility is a distribution of poles with a real multiple pole at $s = -\omega_0$. Here and in the following sections, the term ω_0 is a relative frequency, not the natural frequency. Thus one obtains for the step response of the desired behaviour

$$K_W(s) = \frac{\omega_0^u}{(s + \omega_0)^u}. \quad (10.2.3)$$

This is a series connection of u PT₁ elements with the same time constant $T = 1/\omega_0$. This representation is also called a *binomial form*. The standard polynomials $\beta(s)$ of different order u are given in Table A.8.2. As this table further shows, the normalised step response $h_W(\omega_0 t)$ will become slower with increasing order u . A design using this binomial form is only considered when the step response h_W is required to have no overshoot.

A further possibility of a standard form for $K_W(s)$ of Eq. (10.2.2) is the *Butterworth form*. In this form, all u poles of $K_W(s)$ are equally distributed on a semicircle with radius ω_0 in the left-half s plane and centred at the origin. Table A.8.2 contains the standard polynomials $\beta(s)$ and the associated normalised step responses $h_W(\omega_0 t)$.

Numerous further possibilities for the development of standard forms of Eq. (10.2.2) can be derived from the integral criteria given in Table 7.3.1. For example, the minimum performance index J_{ITAE} is the

basis of a standard form that is also shown in Table A.8.2. Further, often the minimum settling time t_ε is used as the criterion. This table contains for $\varepsilon = 5\%$ the corresponding standard form.

Furthermore, for pole assignment the Weber method can be used. This specifies the desired closed-loop transfer function

$$K_W(s) = \frac{5^k(1 + \kappa^2)\omega_0^{k+2}}{(s + \omega_0 + j\kappa\omega_0)(s + \omega_0 - j\kappa\omega_0)(s + 5\omega_0)^k} \quad (10.2.4)$$

by a real pole with multiplicity $k = u - 2$ and a pair of complex poles. Table A.8.1 contains for different values of k and κ the normalised step responses $h_W(\omega_0 t)$. By a proper choice of k , κ and ω_0 a closed-loop transfer function can be found that fulfils in many instances the desired performance.

10.3 The method of Truxal and Guillemin

For the closed loop shown in Figure 10.3.1 the behaviour is described by the transfer function

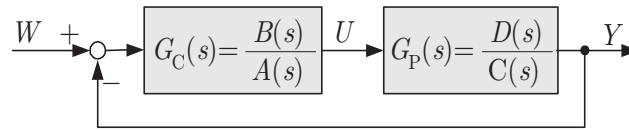


Figure 10.3.1: Block diagram of the closed loop to be designed

$$G_P(s) = \frac{d_0 + d_1s + d_2s^2 + \dots + d_ms^m}{c_0 + c_1s + c_2s^2 + \dots + c_ns^n} = \frac{D(s)}{C(s)}, \quad (10.3.1)$$

where the numerator and denominator polynomials $D(s)$ and $C(s)$ must have no common roots. Furthermore, $G_P(s)$ is normalised to $c_n = 1$ and $m < n$ must be valid.

It is assumed that $G_P(s)$ is stable and minimum phase. For the controller to be designed, the transfer function

$$G_C(s) = \frac{b_0 + b_1s + b_2s^2 + \dots + b_ws^w}{a_0 + a_1s + a_2s^2 + \dots + a_zs^z} = \frac{B(s)}{A(s)} \quad (10.3.2)$$

is chosen and normalised to $a_z = 1$. Because of the realisability of the controller the relation $w = \text{degree } B(s) \leq \text{degree } A(s) = z$ must be valid. Now, the controller must be designed such that the closed loop behaves like a given transfer function for Eq. (10.2.1), whereby $K_W(s)$ should be freely chosen under the condition of the realisability of the controller. From the closed-loop transfer function

$$G_W(s) = \frac{G_C(s)G_P(s)}{1 + G_C(s)G_P(s)} \stackrel{!}{=} K_W(s) \quad (10.3.3)$$

one obtains the controller transfer function

$$G_C(s) = \frac{1}{G_P(s)} \frac{K_W(s)}{1 - K_W(s)} \quad (10.3.4)$$

or with the numerator and denominator polynomials given above

$$G_C(s) = \frac{B(s)}{A(s)} = \frac{C(s)\alpha(s)}{D(s)[\beta(s) - \alpha(s)]}. \quad (10.3.5)$$

The *condition of realisability* for the controller is

$$\text{degree } B(s) = w = n + v \leq \text{degree } A(s) = z = u + m$$

or

$$u - v \geq n - m. \quad (10.3.6)$$

The pole excess $(u - v)$ of the desired closed-loop transfer function $K_W(s)$ must be larger than or equal to the pole excess $(n - m)$ of the plant. Within these constraints the order of $K_W(s)$ is free.

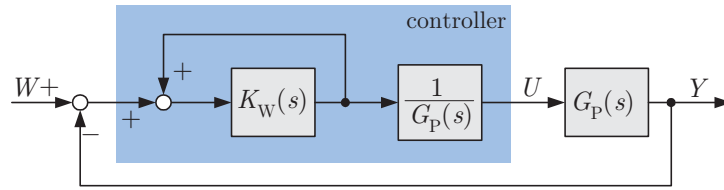


Figure 10.3.2: Compensation of the plant

According to Eq. (10.3.4) the controller contains the inverse plant transfer function $1/G_P(s)$. This is a total compensation of the plant as shown in the block diagram of Figure 10.3.2. For the realisation of the controller Eq. (10.3.5) is used, not the controller structure as shown in this figure with the plant inverse $1/G_P(s)$. As the controller implicitly contains the plant inverse, i.e. the plant zeros are in the set of the controller poles and the plant poles are in the set of the controller zeros, the plant must be stable and minimum phase as mentioned at the beginning. Otherwise, the manipulated variable and/or the controlled variable will show unstable behaviour.

Example 10.3.1

The plant transfer function is given as

$$G_P(s) = \frac{5}{s(1 + 1.4s + s^2)} . \quad (10.3.7)$$

The pole excess of the plant is $n - m = 3$. According to (10.3.6) the pole excess of the desired closed-loop transfer function $K_W(s)$ must be

$$u - v \geq 3 .$$

■

The coefficients of the transfer function $K_W(s)$ that obeys the realisability condition (10.3.6) are subjected to practical constraints, like the maximum range of the manipulated variable, plant parameter errors and measurement noise in the controlled variable, which is disturbing the controller output. The procedure for the design of $G_C(s)$ will be demonstrated by the following example.

Example 10.3.2

For a plant with the transfer function

$$G_P(s) = \frac{1}{(1 + s)^2(1 + 5s)} = \frac{1}{1 + 7s + 11s^2 + 5s^3} = \frac{D(s)}{C(s)} \quad (10.3.8)$$

a controller should be designed such that the closed loop shows optimal behaviour in the sense of the performance index J_{ITAE} and has a rise time of $t_{r,50} = 2.4$ s.

First, it follows from the realisability condition Eq. (10.3.6) and from $n - m = 3 - 0 = 3$ that the pole excess of the desired transfer function $K_W(s)$ is

$$u - v \geq 3 .$$

Inspecting Table A.8.2 one obtains from the J_{ITAE} form for $u = 3$ and $v = 0$ the standard polynomial

$$\beta(s) = s^3 + 1.75\omega_0 s^2 + 2.15\omega_0^2 s + \omega_0^3 . \quad (10.3.9)$$

From the associated step response $h_W(\omega_0 t)$ it follows from Table A.8.2 that the normalised rise time

$$\omega_0 t_{r,50} = 2.4 ,$$

and with this value from the specified rise time $t_{r,50} = 2.4$ s the relative frequency is $\omega_0 = 1$ s⁻¹. Eq. (10.3.9) is now

$$\beta(s) = s^3 + 1.75s^2 + 2.15s + 1 .$$

As for the chosen standard form for $K_W(s)$ the numerator polynomial is $\alpha(s) = 1$, so it follows from Eq. (10.3.5) that the compensator transfer function is

$$G_C(s) = \frac{C(s) \alpha(s)}{D(s) [\beta(s) - \alpha(s)]} = \frac{1 + 7s + 11s^2 + 5s^3}{1 + 2.15s + 1.75s^2 + s^3 - 1}$$

or

$$G_C(s) = \frac{1 + 7s + 11s^2 + 5s^3}{s(2.15 + 1.75s + s^2)}.$$

This controller contains an integrator. The time responses are shown in Figure 10.3.3. ■

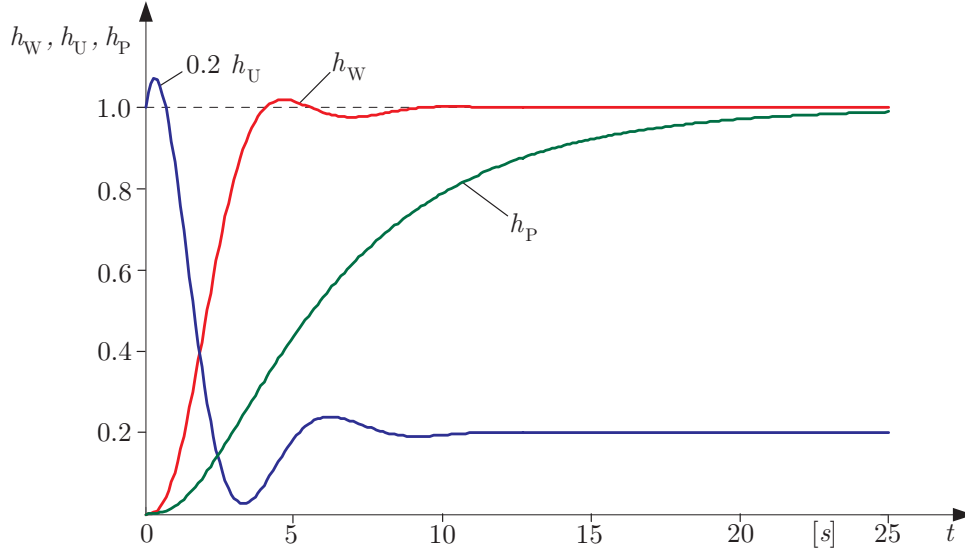


Figure 10.3.3: Closed-loop behaviour for the example 10.3.2: $h_W(t)$ step response of the controlled variable on step in the set point, $h_U(t)$ step response of the associated controlled variable, $h_P(t)$ step response of the uncontrolled plant

If as a further example the plant given by Eq. (10.3.7) instead of Eq. (10.3.8) is taken, then for the same $K_W(s)$ the controller is

$$G_C(s) = \frac{1 + 1.4s + s^2}{(2.15 + 1.75s + s^2)5} = \frac{1 + 1.4s + s^2}{10.75 + 8.75s + 5s^2}.$$

For these two very different plants the same closed-loop behaviour for the controlled variable can be achieved.

In the considerations of this section it has been hitherto assumed that $G_P(s)$ is stable and minimum phase. For plants that do not have this properties this design method cannot be applied in this form. The method must be extended to the following:

A direct compensation of the plant poles and zeros by the controller must be avoided, otherwise stability problems would arise. In these cases, the closed-loop transfer function $K_W(s)$ cannot be arbitrary. For a stable *non-minimum phase plant* the transfer function $K_W(s)$ must be given such that the zeros of $K_W(s)$ contain the right-half-plane zeros of $G_P(s)$. Whereas for an *unstable plant* the zeros of the transfer function $1 - K_W(s)$ must contain the right-half-plane poles of $G_P(s)$. Of course, this restricts the choice of $K_W(s)$ as the following examples demonstrate.

Example 10.3.3

For an all-pass plant with the transfer function

$$G_P(s) = \frac{1 - Ts}{1 + Ts}$$

a controller is to be designed such that the closed loop has the desired transfer function

$$G_W(s) \equiv K_W(s) = \frac{1}{1 + T_1 s} .$$

Using Eq. (10.3.5) one gets for the controller transfer function

$$G_C(s) = \frac{1 + Ts}{1 - Ts} \frac{1}{T_1 s} .$$

This controller gives a direct compensation (cancellation) of the plant zero. This is undesirable as already discussed above, and $K_W(s)$ must be selected as

$$K_W(s) = \frac{1 - Ts}{(1 + T_1 s)^2} .$$

With Eq. (10.3.5) one obtains the controller transfer function as

$$G_C(s) = \frac{1 + Ts}{s[(2T_1 + T) + sT_1^2]} .$$

Because of this choice of $K_W(s)$, the closed loop shows also all-pass behaviour. This effect is more intense the smaller the time constant T_1 . Figure 10.3.4 shows the time responses of this control system. ■

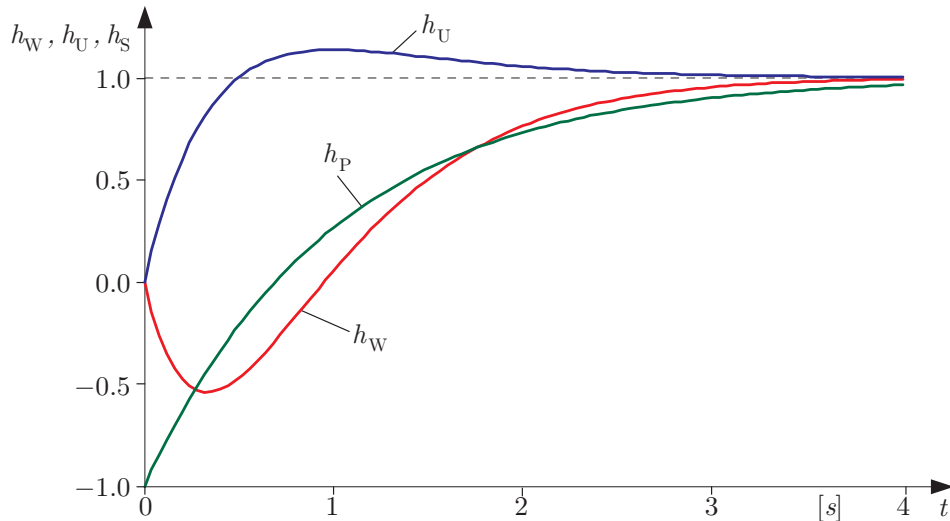


Figure 10.3.4: Closed-loop behaviour of the example 10.3.3: $h_W(t)$ step response of the controlled variable on step in the set point, $h_U(t)$ step response of the associated controlled variable, $h_P(t)$ step response of the uncontrolled plant ($T = 1$ s; $T_1 = 0.5$ s)

Example 10.3.4

The transfer function of the unstable plant

$$G_P(s) = \frac{1}{1 - sT}$$

is given and a controller $G_C(s)$ is required for which $K_W(s)$ fulfills the realisability condition $u - v \geq 1$ and for which the zeros of $1 - K_W(s)$ must contain the plant pole $s = +1/T$. This is expressed by the approach

$$1 - K_W(s) = \frac{\beta(s) - \alpha(s)}{\beta(s)} = \frac{(1 - sT) K(s)}{\beta(s)} ,$$

whereby $K(s)$ is chosen such that

$$\text{degree}[(1 - sT) K(s)] = \text{degree} \beta(s)$$

is valid. In the present case $K_W(s)$ should be chosen such that degree $\beta(s) = u = 2$. From this follows degree $K(s) = 1$. In order to have a stable $K_W(s)$ one can take

$$K(s) = -T_1 s .$$

and obtain

$$\beta(s) - \alpha(s) = (1 - Ts) (-T_1 s) .$$

Observing the realisability condition, it follows that

$$(\beta_0 - \alpha_0) + (\beta_1 - \alpha_1) s + \beta_2 s^2 = -T_1 s + T_1 T s^2 .$$

Comparing the coefficients and taking $\beta_2 = 1$ one obtains

$$\beta_0 - \alpha_0 = 0, \quad \beta_1 - \alpha_1 = -T_1 \quad \text{and} \quad T_1 T = 1$$

and finally

$$T_1 = \frac{1}{T} \quad \text{and} \quad \beta_0 = \alpha_0 .$$

The parameters β_0 and β_1 are still free and may now be chosen such that taking an acceptable behaviour of the manipulated variable into account, a given damping ratio and natural frequency for K_W can be reached. Without going into details,

$$\beta_0 = 1 \quad \text{and} \quad \beta_1 = 2T_1 = \frac{2}{T}$$

will be chosen for the present case and from this it follows that

$$\alpha_0 = 1 \quad \text{and} \quad \alpha_1 = 3T_1 = \frac{3}{T} .$$

The desired closed-loop transfer function will be

$$K_W(s) = \frac{1 + (3/T) s}{1 + (2/T) s + s^2}$$

and

$$1 - K_W(s) = \frac{(1 - Ts)(-s/T)}{1 + (2/T) s + s^2} .$$

The conditions for the design are fulfilled and for the controller transfer function one obtains from Eq. (10.3.4) or Eq. (10.3.5)

$$G_C(s) = \frac{1 + (3/T) s}{(1/T) s} = 3 \left(1 + \frac{T}{3} \frac{1}{s} \right) .$$

This design obviously produces a PI controller. The time responses of this control system are shown in Figure 10.3.5 for $T = 1$ s. The relatively large maximum overshoot cannot be avoided with an acceptable behaviour of the manipulated variable. ■

10.4 Generalised compensator design method

10.4.1 The basic idea

With the following method a control system according to Figure 10.3.1 using the controller given by Eq. (10.3.2) will be designed for a plant described by Eq. (10.3.1) such that the closed loop behaves like the desired transfer function Eq. (10.2.1). Hereby the orders of the controller numerator and denominator polynomials are equal, i.e. $w = \text{degree } B(s) = \text{degree } A(s) = z$. The closed-loop poles are the roots of the characteristic equation, which one obtains from

$$1 + G_C(s) G_P(s) = 0 .$$

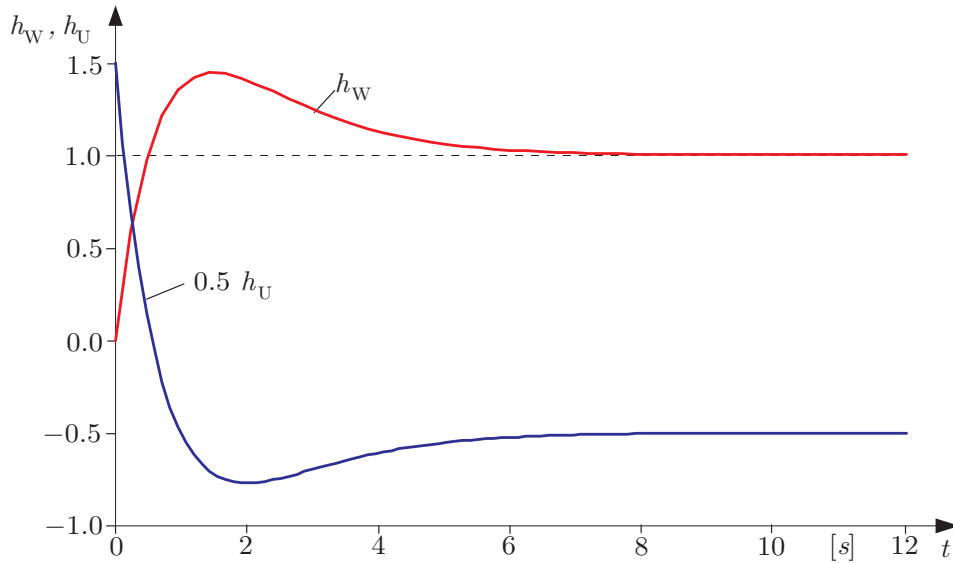


Figure 10.3.5: Closed-loop behaviour of the example 10.3.4: $h_W(t)$ step response of the controlled variable on step in the set point, $h_U(t)$ step response of the associated controlled variable

With respect to the polynomials defined in Eqs. (10.3.1) and (10.3.2) this gives

$$P(s) \equiv \beta(s) = A(s)C(s) + B(s)D(s) = 0. \quad (10.4.1a)$$

On the other hand it follows from Eq. (10.2.1) that

$$P(s) \equiv \beta(s) = \beta_0 + \beta_1 s + \dots + \beta_u s^u = \beta_u \prod_{i=1}^u (s - s_i) = 0. \quad (10.4.1b)$$

This polynomial has order $u = 2 + n$, the coefficients depend linearly on the plant and controller parameters. Comparing both equations, the first coefficient is

$$\beta_0 = a_0 c_0 + b_0 d_0, \quad (10.4.2a)$$

and the last because of $m < n$ and $a_z = c_n = 1$

$$\beta_u = a_z c_n = 1. \quad (10.4.2b)$$

A general representation is given by

$$\begin{aligned} \beta_i = & b_0 d_i + b_1 d_{i-1} + \dots + b_w d_{i-w} \\ & + a_0 c_i + a_1 c_{i-1} + \dots + a_z c_{i-z}, \end{aligned} \quad (10.4.2c)$$

whereby

$$\begin{aligned} d_k = 0 & \quad \text{for } k < 0 \text{ and } k > m \\ c_k = 0 & \quad \text{for } k < 0 \text{ and } k > n, \end{aligned}$$

and $w = z$. The coefficients β_i are obtained from the poles. For the first, second last and last one gets

$$\beta_0 = \prod_{i=1}^u (-s_i) \quad (10.4.3a)$$

$$\beta_{u-1} = \sum_{i=1}^u (-s_i) \quad (10.4.3b)$$

$$\beta_u = 1. \quad (10.4.3c)$$

While the coefficients β_i according to Eqs. (10.4.3) are directly given by the closed-loop poles, the coefficients β_i of Eq. (10.4.2) contain the required controller parameters. Comparing both sides of the latter equation one obtains the *synthesis equation*, which is a system of linear equations for $2z + 1$ unknown controller coefficients $a_0, \dots, a_{z-1}, b_0, b_1, \dots, b_z$. The number of equations is $u = z + n$. A unique solution exists if $z = n - 1$.

A detailed analysis shows, however, that a controller obtained in this way does not usually achieve the desired goals. Because of its small gain a finite steady-state error may occur. This must be taken into consideration during the design. For plants with integral behaviour an order $z = n - 1$ for the controller is sufficient; for proportional behaviour or when disturbances at the input of an integral plant are taken into consideration, the gain must be influenced so that an integral behaviour of the controller can be obtained. This happens if the order of the controller is increased by one, i.e. $z = n$, such that the system of equations is of lower rank. This gives an additional degree of freedom and allows one to choose the controller gain K_C , which is usually introduced as a reciprocal gain factor:

$$\frac{1}{K_C} = c_C = \frac{a_0}{b_0} . \tag{10.4.4}$$

Indeed, the order of the closed loop will be increased; it is now double that of the plant order.

10.4.2 Zeros of the closed loop

In the method presented above, the zeros of the closed loop transfer function for command changes

$$K_W(s) \stackrel{!}{=} G_W(s) = \frac{B(s) D(s)}{A(s) C(s) + B(s) D(s)} \tag{10.4.5}$$

are obtained automatically. In fact, the zeros of the plant, i.e. the roots of $D(s)$, can be considered during the choice of the pole distribution and may be compensated, but the polynomial $B(s)$ arises not in the design and must possibly be compensated after this step. This can be done by introducing a pre-filter in the feed-forward path according to 10.4.1a with a transfer function

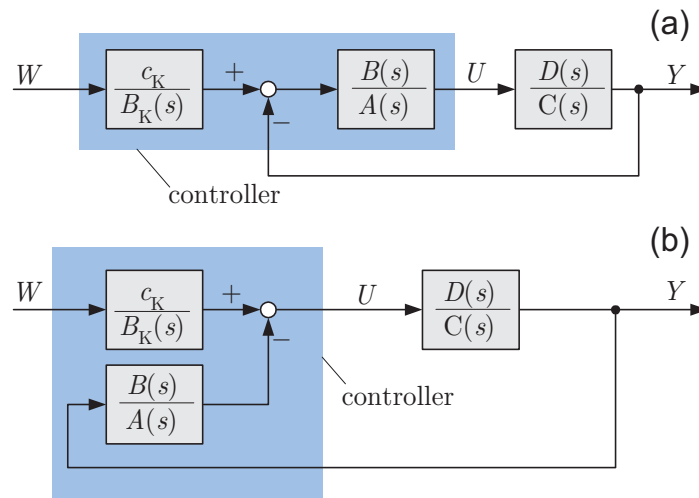


Figure 10.4.1: Compensation of the plant zeros (a) with a controller in the feed-forward path and (b) in the feedback path

$$G_K(s) = \frac{c_K}{B_K(s)} .$$

The zeros of the controller and plant are compensated in this way. For stability reasons, this is only possible for left-half-plane zeros. If $B^+(s)$ and $D^+(s)$ are polynomials with only left-half-plane zeros and $B^-(s)$ and $D^-(s)$ the corresponding polynomials with only right-half-plane zeros including the imaginary

axis, the polynomials of $B(s)$ and $D(s)$ can be factorised as

$$B(s) = B^-(s) B^+(s) \quad (10.4.6)$$

$$D(s) = D^-(s) D^+(s) \quad (10.4.7)$$

with

$$B^-(s) = \sum_{i=0}^{w^-} b_i^- s^i \quad (10.4.8a)$$

$$w = w^+ + w^-$$

$$B^+(s) = \sum_{i=0}^{w^+} b_i^+ s^i \quad (10.4.8b)$$

and

$$D^-(s) = \sum_{i=0}^{m^-} d_i^- s^i \quad (10.4.9a)$$

$$m = m^+ + m^-$$

$$D^+(s) = \sum_{i=0}^{m^+} d_i^+ s^i . \quad (10.4.9b)$$

For the case that $B(s)$ and $C(s)$, and $A(s)$ and $D(s)$ do not have common divisors, i.e. the controller does not compensate plant poles and zeros, the denominator polynomial of the pre-filter can be determined as

$$B_K(s) = B^+(s) D^+(s) . \quad (10.4.10)$$

The transfer function for a command input is then

$$\begin{aligned} G_W(s) &= \frac{c_K}{B_K(s)} \frac{B(s) D(s)}{A(s) C(s) + B(s) D(s)} \\ &= \frac{c_K B^-(s) D^-(s)}{A(s) C(s) + B(s) D(s)} . \end{aligned} \quad (10.4.11)$$

If both, the controller and the plant, show minimum phase behaviour and their transfer functions do not have zeros on the imaginary axis, all zeros of the closed loop can be compensated, such that one obtains instead of Eq. (10.4.11)

$$G_W(s) = \frac{c_K}{A(s) C(s) + B(s) D(s)} . \quad (10.4.12)$$

If the closed-loop transfer function also contains given zeros, the transfer function $G_K(s)$ should have a corresponding numerator polynomial. The coefficient c_K in the numerator is used to make the gain K_W of the closed-loop transfer function $G_W(s)$ equal to 1. From Eq. (10.4.11) it therefore follows that

$$K_W = G_W(0) = c_K \frac{b_0^- d_0^-}{a_0 c_0 + b_0 d_0} = 1 . \quad (10.4.13)$$

The expression in the denominator is the first coefficient β_0 of the characteristic polynomial $\beta(s)$, and therefore with Eq. (10.4.13)

$$c_K = \frac{\beta_0}{b_0^- d_0^-} . \quad (10.4.14)$$

For a controller with integral action the coefficient a_0 is zero and according to Eq. (10.4.4) $c_K = 0$. From Eqs. (10.4.13) and (10.4.6) to (10.4.9) it follows directly that

$$c_K = b_0^+ d_0^+ . \quad (10.4.15)$$

When the controller is inserted into the feedback path according to Figure 10.4.1b the inherent closed-loop dynamics will not be changed compared with the configuration according to Figure 10.4.1a, because

the denominator polynomial of the transfer function, and therefore the characteristic equation of the closed loop, are preserved. Indeed, the zeros of the controller transfer function do no longer arise, but their poles as zeros in the closed-loop transfer function. Analogous considerations for $A_K(s)$ lead to

$$A_K(s) = A^+(s)D^+(s) \text{ ,} \quad (10.4.16)$$

whereby the polynomial $A^+(s)$ contains the poles of the controller and $D^+(s)$ the plant zeros in the left-half plane. The transfer function

$$G_W(s) = \frac{c_K A^-(s) D^-(s)}{A(s)C(s) + B(s)D(s)} \quad (10.4.17)$$

is the same as for the case of a stable controller and a minimum-phase plant according to Eq. (10.4.12).

The constant c_K for a proportional controller is

$$c_K = \frac{\beta_0}{a_0^- d_0^-} \text{ .} \quad (10.4.18)$$

For an integral controller in the feedback loop a feed-forward path is not realisable.

10.4.3 The synthesis equations

The system of equations described by Eq. (10.4.2c) can be rewritten in matrix notation. Thus the required controller parameters are combined into one parameter vector. The matrix of the plant parameters applies for both the cases, controller order $z = n - 1$ and $z = n$.

For *integral plants* ($c_0 = 0$) with controller order $z = n - 1$ and normalised $c_n = 1$ the system is:

$$\left[\begin{array}{cccc|cccc} d_0 & & & & 0 & & & \\ d_1 & d_0 & \mathbf{0} & & c_1 & 0 & \mathbf{0} & \\ d_2 & d_1 & d_0 & & c_2 & c_1 & 0 & \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \\ & & & & c_{n-2} & c_{n-3} & \cdots & c_1 & 0 \\ \hline d_{n-1} & d_{n-2} & & d_1 & d_0 & c_{n-1} & c_{n-2} & \cdots & c_2 & c_1 \\ \hline 0 & d_{n-1} & d_{n-2} & \cdots & d_1 & 1 & c_{n-1} & c_{n-2} & \cdots & c_2 \\ & & & & & 1 & c_{n-1} & \cdots & c_3 & \\ \vdots & & & \ddots & \vdots & & \ddots & \ddots & \vdots & \\ & \mathbf{0} & & & & \mathbf{0} & & \ddots & c_{n-1} & \\ \hline 0 & & & & d_{n-1} & & & & 1 & \end{array} \right] \left[\begin{array}{c} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_{n-2} \\ b_{n-1} \\ a_0 \\ a_1 \\ \vdots \\ a_{n-3} \\ a_{n-2} \end{array} \right] = \left[\begin{array}{c} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{n-2} \\ \beta_{n-1} \\ \beta_n \\ \beta_{n+1} \\ \vdots \\ \beta_{2n-3} \\ \beta_{2n-2} \end{array} \right] - \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-2} \\ c_{n-1} \end{array} \right] \quad (10.4.19a)$$

and

$$a_{n-1} = \beta_u \text{ .} \quad (10.4.19b)$$

For *proportional plants* or in the case of disturbances at the input of integral plants, where the order of the controller is increased by one to $z = n$, with Eqs. (10.4.4) and (10.4.2a) it follows that

$$a_0 = c_R b_0 \quad (10.4.20)$$

$$b_0 = \frac{\beta_0}{d_0 + c_R c_0} \text{ .} \quad (10.4.21)$$

Here the system is:

$$\begin{bmatrix} d_0 & & & & & c_0 \\ d_1 & d_0 & & & & c_1 & c_0 & & & \mathbf{0} \\ d_2 & d_1 & d_0 & & & c_2 & c_1 & c_0 & & & \\ \vdots & \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \\ & & & & & & & & & & c_0 \\ \hline d_{n-1}d_{n-2} \cdots d_1 & d_0 & & & & c_{n-1}c_{n-2} \cdots c_2 & c_1 & & & & \\ 0 & d_{n-1}d_{n-2} \cdots d_1 & & & & & 1 & c_{n-1}c_{n-2} \cdots c_2 & & & \\ & & & & & & & 1 & c_{n-1} \cdots c_3 & & \\ \vdots & & & \ddots & \vdots & & & & \ddots & \vdots & \\ & & \mathbf{0} & & & & \mathbf{0} & & & & c_{n-1} \\ 0 & & & & d_{n-1} & & & & & & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_n \\ \beta_{n+1} \\ \vdots \\ \beta_{2n-1} \end{bmatrix} - b_0 \begin{bmatrix} d_1 + c_R c_1 \\ d_2 + c_R c_2 \\ \vdots \\ d_{n-1} + c_R c_{n-1} \\ c_R \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix} \quad (10.4.22a)$$

and

$$a_n = \beta_u . \quad (10.4.22b)$$

The $(2n - 1) \times (2n - 1)$ matrices on the left side of Eqs. (10.4.19a) and (10.4.22a) are equal for $c = 0$. This matrix is always regular and therefore the solution is always unique.

10.4.4 Application of the method

Example 10.4.1

An integral plant has the transfer function

$$G_P(s) = 0.25 \frac{1 + 5s}{s(1 + 0.25s)} = \frac{1 + 5s}{4s + s^2} .$$

As we do not consider a disturbance at the input of the plant, the controller coefficients can be calculated by Eqs. (10.4.19a) and (10.4.19b).

According to section 10.4.1 one obtains for this second-order plant the order $z = n - 1 = 1$ of the controller. The closed loop has therefore order $u = z + n = 3$. The step response of this loop should comply with the binomial form given in Table A.8.2, and $t_{r,50} \approx 2.5 s$ should be met. This corresponds to a value of $\omega_0 \approx 1s^{-1}$. The associated characteristic polynomial is

$$\beta(s) = (1 + s)^3 = 1 + 3s + 3s^2 + s^3 ,$$

and Eqs. (10.4.19a) and (10.4.19b) provide the synthesis equation

$$\begin{bmatrix} 1 & 0 & | & 0 \\ 5 & 1 & | & 4 \\ \hline 0 & 5 & | & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$$

with

$$a_1 = 1,$$

from which the controller coefficients follow as

$$\begin{aligned} a_0 &= -\frac{9}{19} , & a_1 &= 1 , \\ b_0 &= 1 , & b_1 &= -\frac{2}{19} . \end{aligned}$$

The transfer function of the controller is

$$G_C(s) = \frac{1 - \frac{2}{19}s}{-\frac{9}{19} + s} .$$

It can be seen that this design leads to an unstable controller, which is in addition non-minimum phase. The closed-loop transfer function is

$$G_{W_1}(s) = \frac{\left(1 - \frac{2}{19}s\right)(1 + 5s)}{(1 + s)^3},$$

and it contains in the numerator polynomial, besides the zero of the plant, the right-half-plane zero of the controller. According to the considerations in section 10.4.2 this zero cannot be compensated by a pre-filter for stability reasons. Also this would not be necessary, as the plant zero is dominant and has a stronger influence on the closed-loop step response (see Figure 10.4.2). The denominator polynomial $B_K(s)$ for the transfer function $G_K(s)$ of the pre-filter is determined as

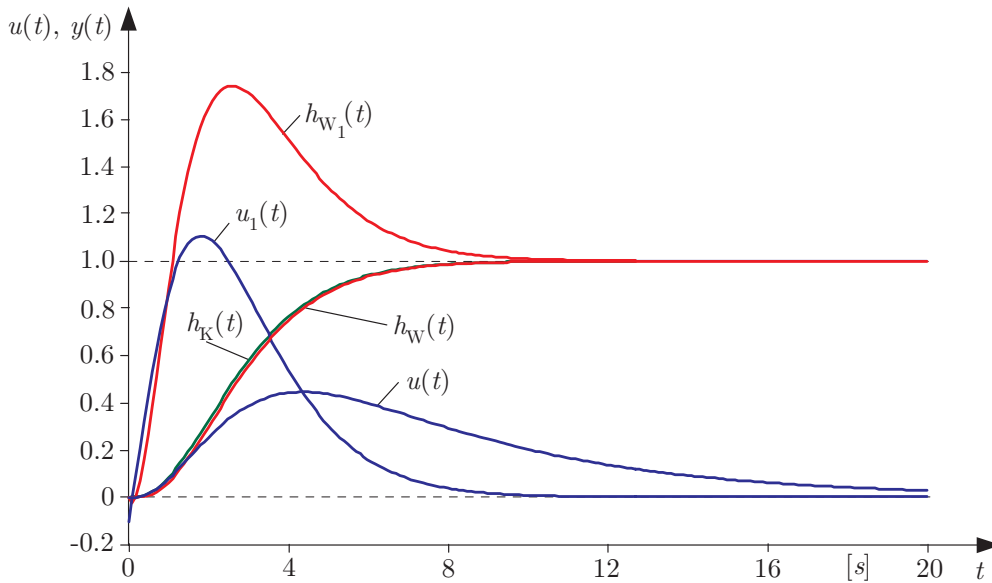


Figure 10.4.2: Step responses of the controlled variable $y(t)$ for the case: (a) without pre-filter $h_{W_1}(t)$, (b) with pre-filter $h_W(t)$, and the associated manipulated variable $u_1(t)$ and $u(t)$, and the step response $h_K(t)$ of the given closed-loop transfer function $K_W(s)$.

$$B_K(s) = D^+(s)B^+(s) = 1 + 5s,$$

and the numerator is

$$c_K = 1.$$

The closed-loop transfer function including the pre-filter is thus

$$G_W(s) = \frac{1 - \frac{2}{19}s}{(1 + s)^3}.$$

This in fact still contains in the numerator polynomial the controller zero, but as can be seen from Figure 10.4.2, the corresponding step response $h_W(t)$ does not show a large deviation from the step response $h_K(t)$ of the given transfer function

$$K_W(s) = \frac{1}{(1 + s)^3}.$$

Conspicuous is the fact that the step response $h_{W_1}(t)$ of the closed loop without pre-filter has a large overshoot and does not show any similarity with the other step responses, though all three other systems have the same inherent behaviour. Here, the dominant behaviour of the plant zero $s_N = -0.2$ has a noticeable effect. ■

Example 10.4.2

Given the third-order plant transfer function

$$G_P(s) = \frac{1}{(1+s)^2(1+5s)} = \frac{0.2}{0.2 + 1.4s + 2.2s^2 + s^3} = \frac{d_0}{C(s)}.$$

The step response $h_P(t)$ of this plant is shown in Figure 10.4.3. A controller should be designed such that the step response $h_W(t)$ of the closed loop has a desired standard form chosen from Table A.8.1. In the current case of a PT_n plant the order m of the controller must be chosen equal to the order n of the plant, that is

$$z = n.$$

Thus one obtains a sixth-order closed-loop transfer function $G_W(s)$. The desired transfer function according to Eq. (10.2.4), which is the basis for Table A.8.1, has then exactly the total order $z + n = 6$, if

$$k = 4$$

is chosen. For this case the step response with

$$\kappa = 4$$

(see Table A.8.1) will be desired. The last unspecified parameter of the transfer function $K_W(s)$ is the relative frequency ω_0 . All step responses in Table A.8.1 are normalised by this value such that the time scale can still be chosen. Consequently, by a proper choice of ω_0 the normalised step response from Table A.8.1 can be scaled to the desired time scale. E.g., for a value of $\omega_0 = 0.4 \text{ s}^{-1}$ a rise time $t_{r,50} \approx 1.6 \text{ s}$ would follow, for $\omega_0 = 2 \text{ s}^{-1}$ this would be $t_{r,50} \approx 0.32 \text{ s}$. If a large value is taken for ω_0 to obtain a small rise time, this would result in controller coefficients of very different order such that this controller would not be realisable from a numerical point of view. Therefore here

$$\omega_0 = 0.4 \text{ s}^{-1}$$

is a good choice. For the determined values of k , κ and ω_0 one obtains for the desired closed-loop transfer function from Eq. (10.2.4)

$$K_W(s) = \frac{43.52}{43.52 + 99.84s + 106.88s^2 + 72.96s^3 + 33.12s^4 + 8.9s^5 + s^6} = \frac{\beta_0}{\beta(s)}.$$

The Eqs. (10.4.20) and (10.4.21) deliver at first the absolute coefficients of the controller transfer function $G_C(s)$

$$b_0 = \frac{\beta_0}{d_0 \left(1 + c_C \frac{c_0}{d_0}\right)} = 217.6 \frac{1}{1 + c_C}$$

and

$$a_0 = c_C b_0 = 217.6 \frac{c_C}{1 + c_C}.$$

The remaining coefficients are obtained from Eq. (10.4.22a) by solving

$$\begin{bmatrix} 0.2 & 0 & 0 & | & 0.2 & 0 \\ 0 & 0.2 & 0 & | & 1.4 & 0.2 \\ 0 & 0 & 0.2 & | & 2.2 & 1.4 \\ \hline 0 & 0 & 0 & | & 1 & 2.2 \\ 0 & 0 & 0 & | & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \hline a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 99.84 \\ 106.88 \\ 72.96 \\ \hline 33.12 \\ 8.8 \end{bmatrix} - 217.6 \cdot \frac{1}{1 + c_C} \begin{bmatrix} 1.4c_C \\ 2.2c_C \\ c_C \\ \hline 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0.2 \\ \hline 1.4 \\ 2.2 \end{bmatrix}.$$

The solution is

$$b_1 = 482 - 1523.2 \frac{c_C}{1 + c_C}$$

$$b_2 = 407.4 - 2393.6 \frac{c_C}{1 + c_C}$$

$$b_3 = 128.4 - 1088 \frac{c_C}{1 + c_C}$$

$$a_1 = 17.2$$

$$a_2 = 6.6.$$

If the reciprocal gain is chosen according to Eq. (10.4.4) as

$$c_C = 0.2 \quad \text{for} \quad K_C = 5 ,$$

then the controller coefficients have the same order of magnitude and the controller transfer function is

$$\begin{aligned} G_C(s) &= \frac{181.33 + 228.13s + 8.467s^2 - 52.93s^3}{36.26 + 17.2s + 6.6s^2 + s^3} = \\ &= \frac{-52.93(s - 2.468)(s + 1.154 + j0.236)(s + 1.154 - j0.236)}{(s + 4.573)(s + 1.013 + j2.627)(s + 1.013 - j2.627)} . \end{aligned}$$

From Eq. (10.4.14) one obtains for the coefficient c_K of the pre-filter

$$c_K = \frac{\beta_0}{b_0^- d_0^-} = \frac{43.52}{2.468 \cdot 52.93 \cdot 1} = 0.33 .$$

The transfer function of the pre-filter is

$$G_K = \frac{0.33}{0.2(s + 1.154 + j0.236)(s + 1.154 - j0.236)} = \frac{1.67}{1.39 + 2.31s + s^2} .$$

The step response of the controlled variable $h_{W_1}(t)$ of the closed loop is shown in Figure 10.4.3. For

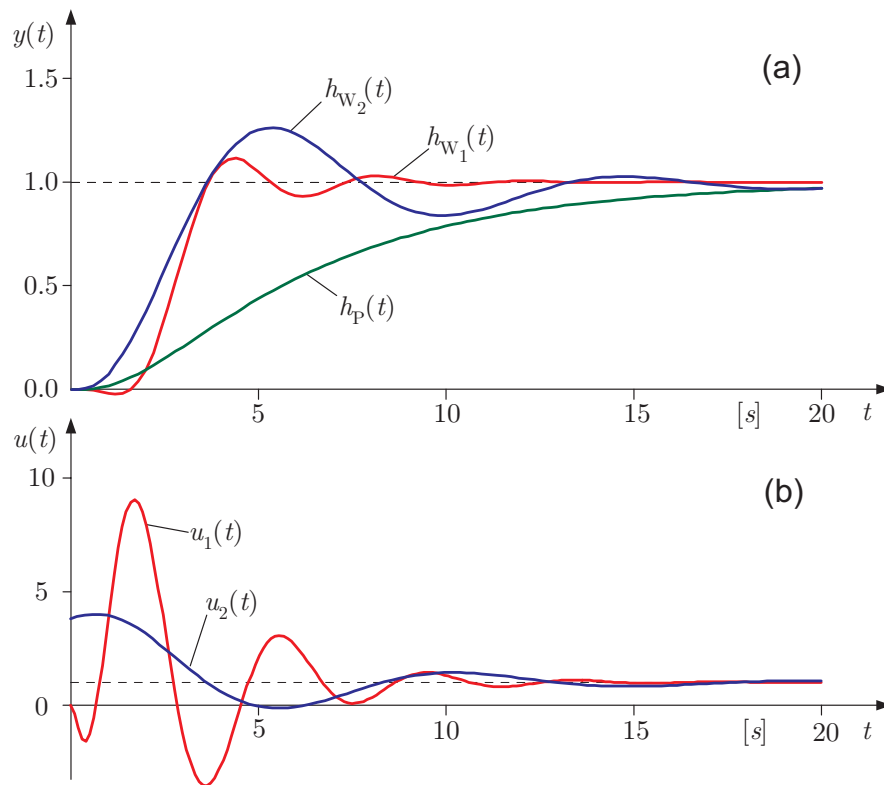


Figure 10.4.3: (a) Step responses of the controlled variable $y(t)$ for the plant in the uncontrolled case $h_P(t)$ and in the controlled case for the designed compensator $h_{W_1}(t)$ and for the optimal PI controller $h_{W_2}(t)$; (b) Step responses of the associated manipulated variables $u_1(t)$ and $u_2(t)$

comparison this figure also shows the corresponding step response $h_{W_2}(t)$ of the closed loop using a PI controller, which is optimal in the sense of the performance index J_{ISE} according to section 7.3.3. Both, the maximum overshoot and the rise time of this control system with a PI controller are clearly worse than for the controller designed here. From the behaviour of the manipulated variables $u_1(t)$ and $u_2(t)$, respectively, it can be seen that in general a smaller rise time must be bought by a larger amplitude of the controller output. Because of the always existing limitation on the value of the manipulated variable, too demanding specifications for the transfer function $K_W(s)$ cannot be realised. ■

10.5 Compensator design for reference and disturbances

10.5.1 Structure of the closed loop

In the previous sections different design methods have been described for the design of controllers for a given plant transfer function $G_P(s)$ and for a desired closed-loop transfer function $G_W(s) = K_W(s)$. If in addition a disturbance is considered, which always exists in real control systems, and if one wants to exert an influence on the disturbance behaviour, the pre-filter $G_K(s)$ is often necessary. Figure 10.5.1 shows this structure. Furthermore, different entry points for the disturbances must be taken into account, because disturbances at the plant input or output act differently on the controlled variable.

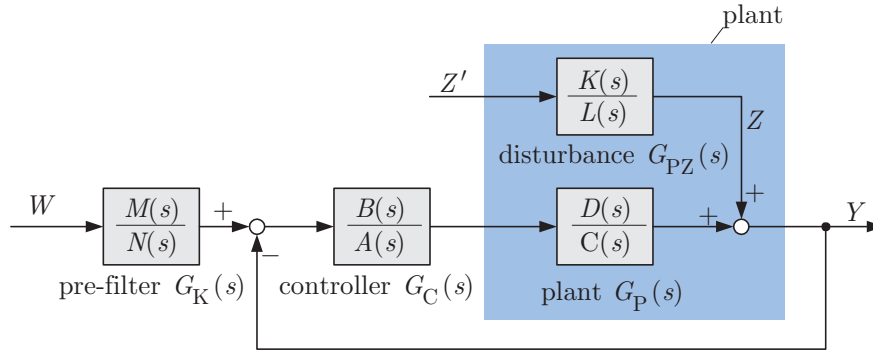


Figure 10.5.1: Control system designed for reference and disturbance behaviour

For the design of the controller and pre-filter the following transfer functions are introduced:

$$G_K(s) = \frac{M(s)}{N(s)} = \frac{m_0 + m_1s + \dots + m_x s^x}{n_0 + n_1s + \dots + n_y s^y}, \quad y \geq x; \quad (10.5.1)$$

$$G_C(s) = \frac{B(s)}{A(s)} = \frac{b_0 + b_1s + \dots + b_w s^w}{a_0 + a_1s + \dots + a_z s^z}, \quad z \geq w; \quad (10.5.2)$$

$$G_P(s) = \frac{D(s)}{C(s)} = \frac{d_0 + d_1s + \dots + d_m s^m}{c_0 + c_1s + \dots + c_n s^n}, \quad n \geq m. \quad (10.5.3)$$

If for the control system the desired closed-loop transfer function for the command input

$$K_W(s) = \frac{\alpha(s)}{\beta(s)} = \frac{\alpha_0 + \alpha_1s + \dots + \alpha_v s^v}{1 + \beta_1s + \dots + \beta_u s^u}, \quad u \geq v \quad (10.5.4)$$

and the desired disturbance transfer function

$$K_Z(s) = \frac{\gamma(s)}{\sigma(s)} = \frac{\gamma_0 + \gamma_1s + \dots + \gamma_q s^q}{1 + \sigma_1s + \dots + \sigma_p s^p}, \quad p \geq q \quad (10.5.5)$$

are given, then on step changes in the command variable and in the disturbance a steady-state error must not occur. Therefore, the following must be valid:

$$K_W(0) = \lim_{s \rightarrow 0} \frac{\alpha(s)}{\beta(s)} = 1, \quad \text{i.e. } \alpha_0 = 1 \quad (10.5.6)$$

and

$$K_Z(0) = \lim_{s \rightarrow 0} \frac{\gamma(s)}{\sigma(s)} = 0, \quad \text{i.e. } \gamma_0 = 0. \quad (10.5.7)$$

For the synthesis of the closed loop it is additionally required to choose the numerator and denominator polynomials of $K_W(s)$ and $K_Z(s)$ of as minimum an order as possible and according to given criteria.

10.5.2 The design procedure

The closed-loop transfer function is given by

$$G_W(s) = \frac{Y(s)}{W(s)} = \frac{G_V(s) G_C(s) G_P(s)}{1 + G_C(s) G_P(s)} \stackrel{!}{=} K_W(s) . \quad (10.5.8)$$

and the disturbance transfer function can be obtained from

$$G_Z(s) = \frac{Y(s)}{Z(s)} = \frac{G_{PZ}(s)}{1 + G_C(s) G_P(s)} \stackrel{!}{=} K_Z(s) . \quad (10.5.9)$$

Then from Eq. (10.5.9) it follows that the controller transfer function is

$$G_C(s) = \frac{G_{PZ}(s) - K_Z(s)}{G_P(s) K_Z(s)} . \quad (10.5.10)$$

For disturbances at the *input* of the plant, $G_{PZ}(s) = G_P(s)$ is valid and therefore one obtains with Eq. (10.5.10) the controller transfer function as

$$G_C(s) = \frac{D(s) \sigma(s) - \gamma(s) C(s)}{D(s) \gamma(s)} . \quad (10.5.11)$$

Disturbances at the plant *output* are taken into account by $G_{PZ} = 1$ in Eq. (10.5.10). In this case the controller transfer function is

$$G_C(s) = \frac{C(s) [\sigma(s) - \gamma(s)]}{D(s) \gamma(s)} = \frac{1}{G_P(s)} \frac{\sigma(s) - \gamma(s)}{\gamma(s)} . \quad (10.5.12)$$

Design for disturbances at the plant input

Contemplating Eq. (10.5.11), the degree w of the numerator polynomial is given either by $m + p$ or by $q + n$. The degree z of the denominator polynomial is fixed by $m + q$. Because of the realisability condition $z \geq w$ of the controller

$$q \geq p \quad \text{and} \quad m \geq n$$

should be valid. But, as this realisability condition cannot hold because of Eqs. (10.5.3) and (10.5.5), the higher-order terms in s of the numerator polynomial of Eq. (10.5.11)

$$D(s) \sigma(s) - \gamma(s) C(s)$$

must disappear. This is only possible with a choice of

$$m + p = q + n \quad \text{or} \quad n - m = p - q . \quad (10.5.13)$$

Eq. (10.5.13) tells us that the pole excess both of the plant and the disturbance transfer function must be the same. Furthermore, it can be seen that in the numerator polynomial of Eq. (10.5.11) exactly

$$(m + p) - (m + q) = p - q = n - m$$

higher-order terms in s must disappear, i.e. they must be compensated (cancelled). From this it follows that the degree of the plant numerator and denominator polynomials is

$$w = z = (m + p) - (p - q) = m + q .$$

As $n - m$ higher-order terms in the numerator polynomial of Eq. (10.5.11) must be compensated, and the degree of the polynomial $\gamma(s)$ must be minimal, the degree of this polynomial must be chosen as

$$q = n - m . \quad (10.5.14)$$

Therefore it follows with Eq. (10.5.13) that

$$p = q + n - m = 2(n - m) . \quad (10.5.15)$$

The previous determined degree of the numerator and denominator polynomials of $G_C(s)$ can be finally obtained using Eq. (10.5.14) as

$$w = z = m + q = n . \quad (10.5.16)$$

In the following the design will be demonstrated by some simple examples.

Example 10.5.1

The plant transfer function

$$G_P(s) = \frac{D(s)}{C(s)} = \frac{d_0 + d_1s}{c_0 + c_1s + c_2s^2}$$

with $n = 2$ and $m = 1$ is given. According to the Eqs. (10.5.14) and (10.5.15) both polynomials $\sigma(s)$ and $\gamma(s)$ have the degrees

$$p = 2(2 - 1) = 2$$

and

$$q = 2 - 1 = 1 .$$

For the disturbance transfer function one gets

$$K_Z(s) = \frac{\gamma_1(s)}{1 + \sigma_1s + \sigma_2s^2} = \frac{\gamma(s)}{\sigma(s)} .$$

If the polynomials $\gamma(s)$ and $\sigma(s)$ are now inserted into Eq. (10.5.11), the controller transfer function

$$G_C(s) = \frac{d_0 + (d_0\sigma_1 + d_1 - \gamma_1c_0)s + (d_0\sigma_2 + d_1\sigma_1 - \gamma_1c_1)s^2 + (d_1\sigma_2 - \gamma_1c_2)s^3}{d_0\gamma_1s + d_1\gamma_1s^2}$$

follows. With $\gamma_1 = \sigma_2d_1/c_2$ the highest-order term in the numerator polynomial of $G_C(s)$ disappears, and one obtains the realisable controller transfer function of

$$G_C(s) = \frac{b_0 + b_1s + b_2s^2}{a_1s + a_2s^2}$$

and the disturbance transfer function of

$$K_Z(s) = \frac{\frac{\sigma_2d_1}{c_2}s}{1 + \sigma_1s + \sigma_2s^2} .$$

This example clearly shows that the disturbance behaviour cannot be freely designed, but the inherent dynamics influences the disturbance response, i.e. the poles of the disturbance transfer function. In this case the coefficients of the numerator polynomial $\gamma(s)$ depend on those of the denominator polynomial $\sigma(s)$. ■

Example 10.5.2

Similarly to the previous example a second-order plant is again considered, but with the pole excess $n - m = 2$. The transfer function is

$$G_P(s) = \frac{D(s)}{C(s)} = \frac{d_0}{c_0 + c_1s + c_2s^2} .$$

The degree of the numerator and denominator polynomials of the disturbance transfer function can be obtained from Eqs. (10.5.14) and (10.5.15) as

$$q = n - m = 2 \quad \text{and} \quad p = 2(n - m) = 4 .$$

Therefore, the disturbance transfer function has the form

$$K_Z(s) = \frac{\gamma_1s + \gamma_2s^2}{1 + \sigma_1s + \sigma_2s^2 + \sigma_3s^3 + \sigma_4s^4} .$$

According to Eq. (10.5.16) the degree of the numerator and denominator polynomials of the controller transfer function will be

$$w = z = m + q = n = 2 .$$

On the other hand one obtains formally from Eq. (10.5.11) the controller transfer function as

$$G_C(s) = \frac{b_0 + b_1s + b_2s^2 + (d_0\sigma_3 - \gamma_1c_2 - \gamma_2c_1)s^3 + (d_0\sigma_4 - \gamma_2c_2)s^4}{d_0\gamma_1s + d_0\gamma_2s^2} .$$

However, for realisability reasons $n - m = 2$ higher-order terms in s in the numerator polynomial must disappear. From this it follows for the disturbance transfer function that

$$\gamma_2 = \frac{d_0 \sigma_4}{c_2}$$

and

$$\gamma_1 = \frac{d_0 \sigma_3 c_2 - d_0 \sigma_4 c_1}{c_2^2} , .$$

■

Design for disturbances at the plant output

Starting from Eq. (10.5.12) similar results as in the previous section will be obtained. The degree of the controller numerator polynomial in this case is given either by $n + p$ or by $q + n$. For realisability reasons a compensation of the supernumerary terms in the numerator of Eq. (10.5.12) is necessary. This is possible if

$$p = q \tag{10.5.17}$$

is valid. That p must equal q is also understandable from the fact that the disturbance acts directly at the output of the plant and therefore the disturbance transfer function has always a feed-through term in the transfer function. The $(n + p) - (m + q) = n - m$ higher-order terms in the numerator polynomial of Eq. (10.5.12) must be compensated. Therefore, the degree of the numerator and denominator polynomials of $G_C(s)$ is given by

$$w = z = (q + n) - (n - m) = m + q .$$

As in the numerator polynomial of Eq. (10.5.12) $n - m$ terms must be compensated and the degree of the the polynomial $\gamma(s)$ must be minimal, its degree is chosen as

$$q = n - m . \tag{10.5.18}$$

From this it follows with Eq. (10.5.17) that

$$p = n - m . \tag{10.5.19}$$

The previous given degree of the polynomials of the controller transfer function $G_C(s)$ will now be with Eq. (10.5.18)

$$w = z = m + q = n . \tag{10.5.20}$$

It can be seen from Eq. (10.5.12) that the poles of the plant transfer function are cancelled by the zeros of the controller transfer function. If the plant parameters are varying slightly, in the case of an unstable plant the closed loop will also be unstable. For unstable plants, $\gamma(s)$ will therefore chosen such that it contains the unstable terms $C^-(s)$ of degree n^- , that is

$$\gamma(s) = C^-(s) \psi(s) .$$

Consequently the controller transfer function $G_C(s)$ has the form

$$G_C(s) = \frac{C^+(s) [\sigma(s) - C^-(s) \psi(s)]}{D(s) \psi(s)} ,$$

whereby $C^+(s)$ is that part of the polynomial $C(s)$ which has zeros in the left-half s plane.

The polynomial $\psi(s)$ of degree λ must be chosen such that the polynomial $\sigma(s) - C^-(s) \psi(s)$ itself has only zeros in the left-half s plane. The order of the disturbance transfer function is given by $p = n^- + \lambda$, as for realisability reasons in the numerator polynomial those $n - n^- + p - m - \lambda$ higher-order terms in s must be compensated. The controller transfer function is then of order $m + \lambda$.

A detailed example of the design for an unstable plant can be found in section 10.5.4. The procedure of the design will be demonstrated now in the following using some simple examples with stable plants.

Example 10.5.3

For the plant with the transfer function

$$G_P(s) = \frac{d_0 + d_1 s}{c_0 + c_1 s + c_2 s^2}$$

and the pole excess $n - m = 1$, the degree will be using Eqs. (10.5.18) and (10.5.19) $q = p = 1$. For the disturbance transfer function $K_Z(s)$ follows according to Eqs. (10.5.5) and (10.5.7)

$$K_Z(s) = \frac{\gamma_1 s}{1 + \sigma_1 s}.$$

Inserted in Eq. (10.5.12) the controller transfer function arises formally as

$$G_C(s) = \frac{c_0 + [c_1 + c_0(\sigma_1 - \gamma_1)]s + [c_2 + c_1(\sigma_1 - \gamma_1)]s^2 + c_2(\sigma_1 - \gamma_1)s^3}{d_0\gamma_1 s + d_1\gamma_1 s^2}.$$

With the realisability condition $\sigma_1 = \gamma_1$ one obtains the realisable controller transfer function as

$$G_C(s) = \frac{c_0 + c_1 s + c_2 s^2}{d_0\sigma_1 s + d_1\sigma_1 s^2}$$

and for the disturbance transfer function

$$K_Z(s) = \frac{\sigma_1 s}{1 + \sigma_1 s}.$$

■

Example 10.5.4

For the plant transfer function

$$G_P(s) = \frac{d_0}{c_0 + c_1 s + c_2 s^2}$$

is $n - m = 2$ and therefore $p = q = 2$. From this it follows that the disturbance transfer function is

$$K_Z(s) = \frac{\gamma_1 s + \gamma_2 s^2}{1 + \sigma_1 s + \sigma_2 s^2}.$$

The realisability condition of the controller transfer function

$$G_C(s) = \frac{b_0 + b_1 s + b_2 s^2 + [c_1(\sigma_2 - \gamma_2) + c_2(\sigma_1 - \gamma_1)]s^3 + (\sigma_2 - \gamma_2)s^4}{d_0\gamma_1 s + d_0\gamma_2 s^2}$$

is

$$\gamma_2 = \sigma_2$$

and

$$\gamma_1 = \sigma_1.$$

The realisable controller transfer function is

$$G_C(s) = \frac{b_0 + b_1 s + b_2 s^2}{d_0\gamma_1 s + d_0\gamma_2 s^2}.$$

■

Recapitulating, we can ascertain that the order for the disturbance transfer function depends on the entry point of the disturbance and on the pole excess ($n - m$) of the plant transfer function. If the inherent dynamics of $K_Z(s)$ is chosen, the zeros of $K_Z(s)$ are given by the realisability conditions for the controller.

10.5.3 Design of the pre-filter

The starting point for the synthesis of the pre-filter transfer function $G_K(s)$ are the Eqs. (10.5.8) and (10.5.9). From these equations it follows directly that

$$G_K(s) = \frac{G_{PZ}(s) K_W(s)}{G_C(s) G_P(s) K_Z(s)} . \quad (10.5.21)$$

If the disturbance acts at the *input* of the plant, $G_{PZ}(s) = G_P(s)$ is valid and therefore the pre-filter transfer function is

$$G_K(s) = \frac{K_W(s)}{G_C(s) K_Z(s)} . \quad (10.5.22)$$

In the case of disturbances at the plant *output* the pre-filter will be, because $G_{PZ}(s) = 1$, given by

$$G_K(s) = \frac{K_W(s)}{G_C(s) G_P(s) K_Z(s)} . \quad (10.5.23)$$

Design of the pre-filter for disturbances at the plant input

With Eq. (10.5.22) one obtains the pre-filter transfer function as

$$G_K(s) = K_W(s) \frac{1}{G_C(s)} \frac{1}{K_Z(s)} = \frac{\alpha(s) A(s) \sigma(s)}{\beta(s) B(s) \gamma(s)} .$$

Taking with Eqs. (10.5.2) and (10.5.11)

$$A(s) = D(s) \gamma(s)$$

into consideration, the pre-filter transfer function can also be written as

$$G_K(s) = \frac{\alpha(s) D(s) \sigma(s)}{\beta(s) B(s)} = \frac{M(s)}{N(s)} . \quad (10.5.24)$$

For the determination of the degrees v and u of the polynomials $\alpha(s)$ and $\beta(s)$, respectively, the realizability condition for the pre-filter transfer function $G_K(s)$ will be investigated. Accordingly, one obtains with the results from section 10.5.2 with $p = 2(n - m)$ and $w = n$ the condition $u + n \geq v + m + 2(n - m)$ or

$$u \geq n - m + v . \quad (10.5.25)$$

According to Eq. (10.5.24) the degree of the numerator polynomial $M(s)$ and of the denominator polynomial $N(s)$ is given by

$$x = m + v + 2(n - m) = 2n - m + v \quad (10.5.26)$$

and

$$y = n + u , \quad (10.5.27)$$

as far as $M(s)$ and $N(s)$ have no common divisor. For example, if the degree of the numerator of the closed-loop transfer function for command input is given by $v = 0$, the following is valid:

$$u \geq n - m .$$

For $v > 0$ the polynomial $\alpha(s)$ can be used to cancel the unavoidable zeros of the controller transfer function in the right-half s plane. Otherwise this would lead to an unstable pre-filter. But the non-minimum-phase behaviour caused by the controller still remains.

Example 10.5.5

The transfer functions $G_P(s)$, $K_Z(s)$ and $G_C(s)$ are taken from example 10.5.1. According to Eq. (10.5.24) the pre-filter can be obtained as

$$G_K(s) = \frac{\alpha(s) D(s)}{\beta(s) B(s)} \sigma(s) .$$

With $v = 0$ and $u = 1$ the closed-loop transfer function is chosen as

$$K_W(s) = \frac{\alpha(s)}{\beta(s)} = \frac{1}{1 + \beta_1 s}.$$

Herewith, the degree of the pre-filter numerator and denominator polynomial is given by $x = y = 3$. After substitution of the results from example 10.5.1 in Eq. (10.5.24) the pre-filter transfer function in general form is

$$G_K(s) = \frac{m_0 + m_1 s + m_2 s^2 + m_3 s^3}{n_0 + n_1 s + n_2 s^2 + n_3 s^3}.$$

■

Example 10.5.6

If we have as in example 10.5.2 a pole excess of the plant transfer function of $n - m = 2$, and if the results for $K_Z(s)$ and $G_C(s)$ are considered, then for $v = 0$ the closed-loop transfer function can be chosen as

$$K_W(s) = \frac{1}{1 + \beta_1 s + \beta_2 s^2}.$$

From Eqs. (10.5.24), (10.5.26) and (10.5.27) the pre-filter transfer function

$$G_K(s) = \frac{m_0 + m_1 s + m_2 s^2 + m_3 s^3 + m_4 s^4}{n_0 + n_1 s + n_2 s^2 + n_3 s^3 + n_4 s^4}$$

follows.

■

Design of the pre-filter for disturbances at the plant output

Similarly as in section 10.5.1 one obtains here from Eq. (10.5.23)

$$G_K(s) = \frac{A(s) C(s) \sigma(s) \alpha(s)}{B(s) D(s) \gamma(s) \beta(s)} = \frac{M(s)}{N(s)}.$$

If the Eqs. (10.5.2) and (10.5.11) are considered, one obtains with

$$A(s) = D(s) \gamma(s)$$

from the above equation

$$G_K(s) = \frac{C(s) \sigma(s) \alpha(s)}{B(s) \beta(s)}. \quad (10.5.28)$$

The first task is again to specify the degrees u and v of the polynomials $\alpha(s)$ and $\beta(s)$ such that the pre-filter becomes realisable. For the polynomials $M(s)$ and $N(s)$ previous results from section 10.5.2 can be considered and from this follows

$$n + u \geq n + n - m + v$$

and

$$u \geq n - m + v. \quad (10.5.29)$$

From Eq. (10.5.28) one obtains finally for the degree of the numerator and denominator polynomials $M(s)$ and $N(s)$ of $G_K(s)$ that

$$x = 2n - m + v \quad (10.5.30a)$$

and

$$y = n + u \quad (10.5.30b)$$

as far as $M(s)$ and $N(s)$ have no common divisor. If during the design of $N(s)$ unstable poles arise, then one can proceed as in the case of disturbances at the plant input. The degree v of the numerator polynomial of $K_W(s)$ will be increased by the number of compensated terms.

For the case of unstable plants with $C(s) = C^+(s)C^-(s)$ and the unstable part $C^-(s)$ and with the denominator polynomial $\gamma(s) = C^-(s)\psi(s)$ of the disturbance transfer function $K_Z(s)$, the pre-filter transfer function can be obtained from the previous section by using the controller transfer function

$$G_C(s) = \frac{C^+(s) [\sigma(s) - C^-(s)\psi(s)]}{D(s)\psi(s)} = \frac{B(s)}{A(s)}$$

as

$$G_K(s) = \frac{C^+(s)\sigma(s)a(s)}{B(s)\beta(s)}. \quad (10.5.31)$$

Example 10.5.7

Again as in example 10.5.3 the plant transfer function

$$G_P(s) = \frac{D(s)}{C(s)} = \frac{d_0 + d_1s}{c_0 + c_1s + c_2s^2}$$

is given. Considering earlier results

$$K_Z(s) = \frac{\gamma(s)}{\sigma(s)} = \frac{\gamma_1s}{1 + \sigma_1s}$$

and

$$G_C(s) = \frac{B(s)}{A(s)} = \frac{b_0 + b_1s + b_2s^2}{a_1s + a_2s^2},$$

one obtains for $v = 0$ the closed-loop transfer function for command input using Eq. (10.5.29) as

$$K_W(s) = \frac{\alpha(s)}{\beta(s)} = \frac{1}{1 + \beta_1s}.$$

Substituted in Eq. (10.5.28) the pre-filter transfer function follows as

$$G_K(s) = \frac{m_0 + m_1s + m_2s^2 + m_3s^3}{n_0 + n_1s + n_2s^2 + n_3s^3}.$$

■

Example 10.5.8

Similarly as in example 10.5.4 the pole excess of the plant transfer function is given as $n - m = 2$. With

$$K_Z(s) = \frac{\gamma_1s + \gamma_2s^2}{1 + \sigma_1s + \sigma_2s^2}$$

and

$$G_C(s) = \frac{b_0 + b_1s + b_2s^2}{a_1s + a_2s^2}$$

follows with the choice of $v = 0$ and applying Eq. (10.5.29) the pre-filter transfer function is

$$G_V(s) = \frac{m_0 + m_1s + m_2s^2 + m_3s^3 + m_4s^4}{n_0 + n_1s + n_2s^2 + n_3s^3 + n_4s^4}.$$

■

10.5.4 Application of the design method

The design procedure for compensators described in the previous sections will be applied in this concluding section to an example already introduced in section 10.3. It is the unstable plant with the transfer function

$$G_P(s) = \frac{D(s)}{C(s)} = \frac{1}{1 - sT}, \quad T = 1 \text{ s}$$

(see example 10.3.4).

Disturbance at the plant input

If a disturbance at the plant input is assumed, the degree of the numerator and denominator polynomial of the free selectable disturbance transfer function according to Eqs. (10.5.14) and (10.5.15) is given by $q = 1$ and $p = 2$. The disturbance transfer function will be obtained as

$$K_Z(s) = \frac{\gamma(s)}{\sigma(s)} = \frac{\gamma_1 s}{1 + \sigma_1 s + \sigma_2 s^2}$$

and the the controller transfer function formally according to Eq. (10.5.11) as

$$G_C(s) = \frac{B(s)}{A(s)} = \frac{1 + (\sigma_1 - \gamma_1) s + (\sigma_2 + \gamma_1 T) s^2}{\gamma_1 s} .$$

For the realisability condition of the controller it follows that

$$\gamma_1 = -\frac{\sigma_2}{T} .$$

The final realisable controller transfer function is then

$$G_C(s) = \frac{1 + \left(\sigma_1 + \frac{\sigma_2}{T}\right) s}{-\frac{\sigma_2}{T} s} .$$

Furthermore, the above equation for γ_1 provides for the disturbance transfer function

$$K_Z(s) = \frac{-\frac{\sigma_2}{T} s}{1 + \sigma_1 s + \sigma_2 s^2} .$$

The procedure for the design of the pre-filter for disturbances at the plant input starts with the given order of the transfer function for command input according to Eq. (10.5.25). For the command behaviour $u = 2$ and $v = 0$ is chosen. According to that the transfer function for command input is

$$K_W(s) = \frac{\alpha(s)}{\beta(s)} = \frac{1}{1 + \beta_1 s + \beta_2 s^2} .$$

The synthesis of the controller is required to be performed such that the behaviour for a command input will be characterised by approximately 10% maximum overshoot and $t_{3\%} \approx 3$ s settling time. In this case of a second-order system these demands are met by a natural frequency $\omega_0 = 2$ s⁻¹ and by a damping ratio $\zeta = 0.6$ according to Figures 9.1.3 and 9.1.5. From this the transfer function for command input follows as

$$K_W(s) = \frac{1}{1 + 0.6s + 0.25s^2} .$$

If for the inherent behaviour of the closed loop the same coefficients are taken for the disturbance case, then

$$\sigma(s) = \beta(s) ,$$

and therefore the disturbance transfer function is

$$K_Z(s) = \frac{-0.25s}{1 + 0.6s + 0.25s^2}$$

and the controller transfer function is

$$G_C(s) = \frac{1 + 0.85s}{-0.25s} .$$

The pre-filter design, which is given by

$$G_K(s) = \frac{D(s) \alpha(s) \sigma(s)}{B(s) \beta(s)} ,$$

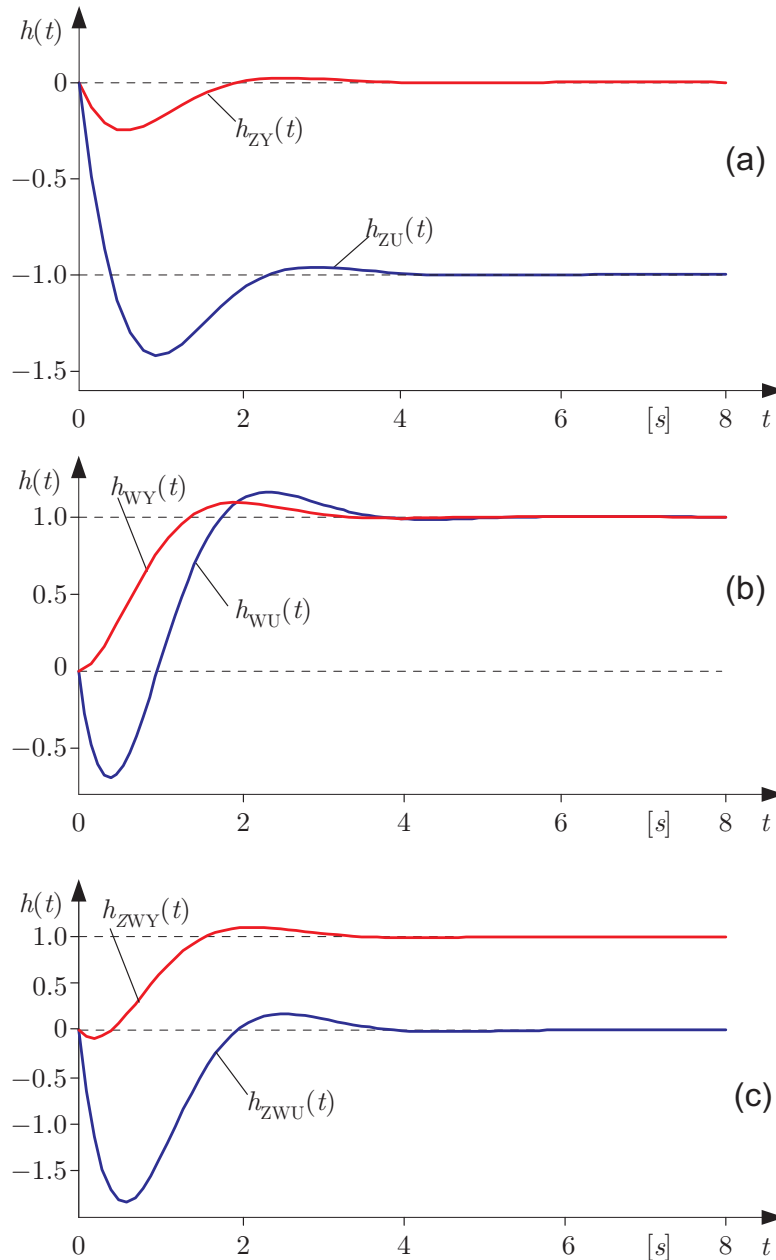


Figure 10.5.2: Step responses for the design for a step disturbance $z = \sigma(t)$ at the plant input:
 (a) controlled variable $h_{ZY}(t)$ and manipulated variable $h_{ZU}(t)$ for $z = \sigma(t)$
 (b) controlled variable $h_{WY}(t)$ and manipulated variable $h_{WU}(t)$ for $w = \sigma(t)$
 (c) controlled variable $h_{ZWY}(t)$ and manipulated variable $h_{ZWU}(t)$ for the simultaneous application of $z = \sigma(t)$ and $w = \sigma(t)$

is simplified by $\beta(s) = \sigma(s)$ to

$$G_K(s) = \frac{D(s)\alpha(s)}{B(s)} = \frac{1}{1 + 0.85s} .$$

Step responses for the disturbance and command inputs are shown in Figure 10.5.2.

Disturbance at the plant output

If the closed loop is designed for a disturbance at the plant output, for the unstable plant the controller transfer function

$$G_C(s) = \frac{C^+(s) [\sigma(s) - C^-(s)\psi(s)]}{D(s)\psi(s)}$$

from section 10.5.2 is significant. With

$$C^+(s) = 1, \quad C^-(s) = 1 - sT$$

and with the approach

$$\psi(s) = a + bs$$

it follows – as in the previous case – for

$$\sigma(s) = \beta(s)$$

that

$$G_C(s) = \frac{1 - a + (\beta_1 - b + aT)s + (\beta_2 + bT)s^2}{a + bs}.$$

The realisability condition for the controller is thus

$$b = -\frac{\beta_2}{T},$$

and if for simplicity a is set equal to zero, then the controller transfer function is

$$G_C(s) = \frac{1 + \left(\beta_1 + \frac{\beta_2}{T}\right)s}{-\frac{\beta_2}{T}s} = \frac{1 + 0.85s}{-0.25s},$$

as before. The corresponding disturbance transfer function is

$$K_Z(s) = \frac{\psi(s)C^-(s)}{\sigma(s)} = \frac{-\frac{\beta_2}{T}s + \beta_2s^2}{1 + \beta_1s + \beta_2s^2}.$$

The transfer function of the pre-filter is calculated according to Eq. (10.5.23), that is

$$G_K(s) = \frac{A(s)C(s)\sigma(s)\alpha(s)}{B(s)D(s)\gamma(s)\beta(s)}.$$

With the assumptions

$$\begin{aligned} K_W(s) &= \frac{\alpha(s)}{\beta(s)} = \frac{1}{1 + \beta_1s + \beta_2s^2}, \\ \sigma(s) &= \beta(s), \\ \gamma(s) &= C^-(s)\psi(s), \\ A(s) &= D(s)\psi(s) \end{aligned}$$

the pre-filter is simplified to

$$G_K(s) = \frac{1}{B(s)} = \frac{1}{1 + \left(\beta_1 + \frac{\beta_2}{T}\right)s} = \frac{1}{1 + 0.85s},$$

again as before. The step responses for this case are shown in Figure 10.5.3.

Problem 10.1
Compensator design - two questions

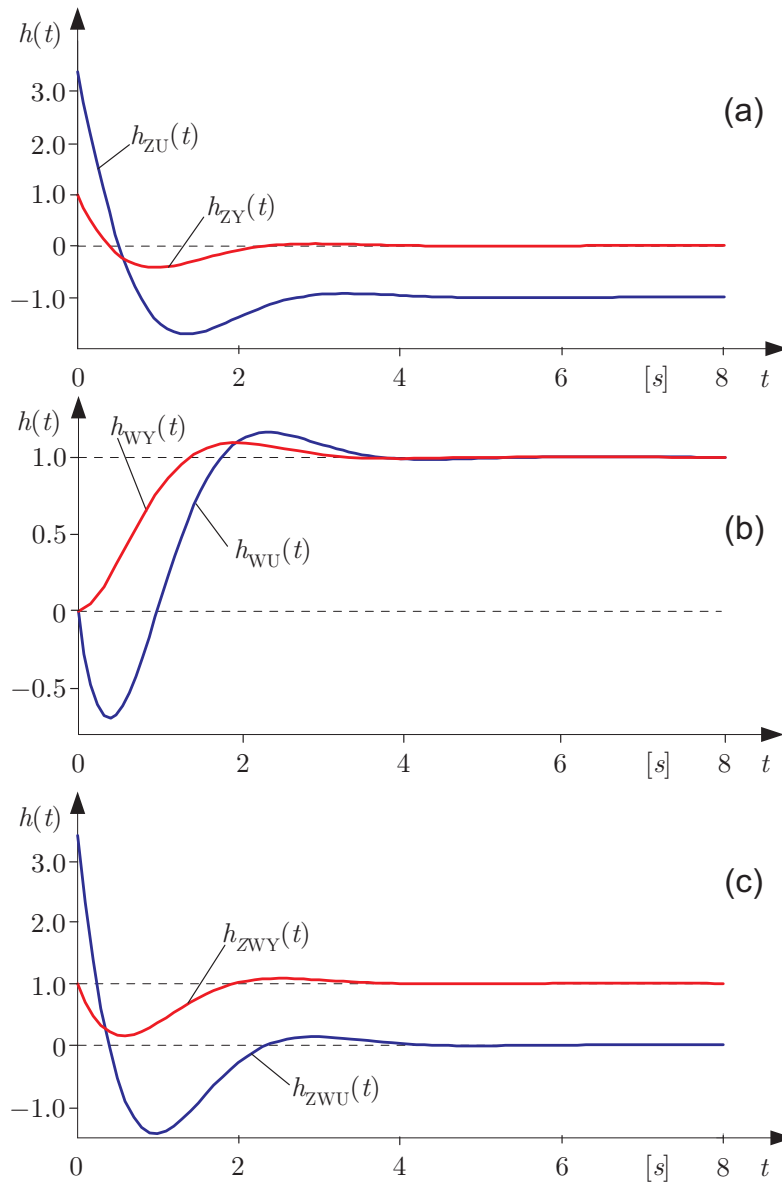


Figure 10.5.3: Step responses for the design for a step disturbance $z = \sigma(t)$ at the plant output:
 (a) controlled variable $h_{ZY}(t)$ and manipulated variable $h_{ZU}(t)$ for $z = \sigma(t)$
 (b) controlled variable $h_{WY}(t)$ and manipulated variable $h_{WU}(t)$ for $w = \sigma(t)$
 (c) controlled variable $h_{ZWY}(t)$ and manipulated variable $h_{ZWU}(t)$ for the simultaneous application of $z = \sigma(t)$ and $w = \sigma(t)$

Module 11

Improving the control behaviour by more complex loop structures

Module units

11.1 Problem	11-1
11.2 Disturbance feed-forward control	11-2
11.2.1 Disturbance feed-forward on the controller	11-2
11.2.2 Disturbance feed-forward on the manipulated variable	11-3
11.3 Control systems with an auxiliary manipulated variable	11-4
11.4 Cascade control systems	11-5
11.5 Control system with auxiliary manipulated variable	11-7
11.6 Control system with anti-windup measure	11-9

Module overview. *The simple closed-loop structure is not suitable for all applications. In those cases where additional information, for example, about disturbances or additional system measurements are available, the basic structure can be extended to improve the dynamical behaviour. Such cases are collected in this module. The different feed-forward structures of disturbances and auxiliary variables with and without a cascade control structure are introduced and demonstrated by examples. Finally, the windup problem of controllers and anti-windup measures are discussed.*

Module objectives. *When you have completed this module you should be able to:*

1. Know how to improve the control behaviour by disturbance feed-forward.
2. Understand the possibilities which are available for use of additional output signals from the plant for performance improvement.
3. Know how to cope with the controller windup problem.

Module prerequisites. *Transfer function, properties of closed-loop systems.*

11.1 Problem

The control systems discussed hitherto are single-loop control systems. These control systems may not meet extra high requirements even in an optimal design case for higher-order plants and plants with dead time concerning the maximum overshoot M_p , rise time t_r and settling time t_ϵ . This is apparent especially in the case of large disturbances and when large delays occur between the actuator and measurement device. An improvement of the control behaviour can be obtained if the signal paths between actuator

and disturbance are shortened, or if disturbances are already compensated by a separate *pre-controller* before they enter the plant. In this case the disturbance must be measurable and controllable via an actuator. Shortening the signal paths within a control system leads to a structural expansion of the basic control loop and therefore to more complex loop structures. In the following, the most important basic structures of these complex control systems will be discussed.

11.2 Disturbance feed-forward control

The disturbance feed-forward control corresponds to the basic control loop, which is superimposed by an open-loop control scheme with the goal to compensate the disturbance by a control element $G_{FF_i}(s)$ ($i = 1, 2, \dots$) as far as possible before it fully acts on the controlled variable y . This configuration is of course only realisable, if the disturbance is measurable at the plant input. With regard to a feed-forward configuration, the following two different cases are distinguished, whereby the following transfer functions, see Figure 11.2.1, will be used:

$$G_C(s) = \frac{B(s)}{A(s)}; \quad G_{FF_i}(s) = \frac{B_{FF_i}(s)}{A_{FF_i}(s)} \quad (i = 1, 2)$$

$$G_P(s) = \frac{D(s)}{C(s)}; \quad G_{PZ}(s) = \frac{D_Z(s)}{C_Z(s)} .$$

11.2.1 Disturbance feed-forward on the controller

According to Figure 11.2.1 the disturbance z' will feed via the transfer function $G_{FF_1}(s)$ to the controller, which will compensate the influence of the disturbance. From this diagram the controlled variable directly

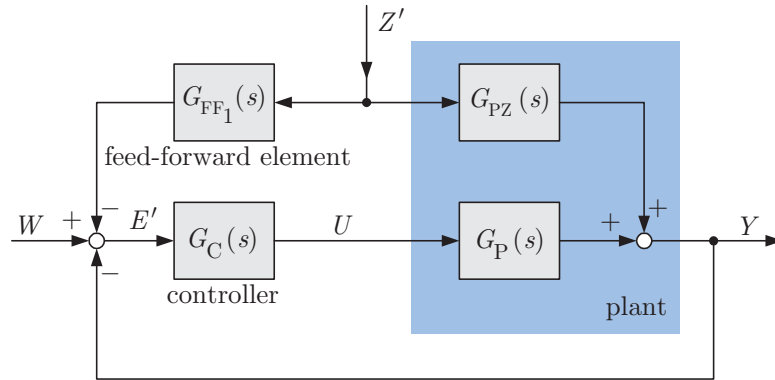


Figure 11.2.1: Block diagram of the feed-forward on the controller

follows as

$$Y(s) = [W(s) - Y(s) - Z'(s) G_{FF_1}(s)] G_C(s) G_P(s) + Z'(s) G_{PZ}(s) . \quad (11.2.1)$$

With some manipulations one obtains from this

$$Y = \frac{G_{PZ} - G_{FF_1} G_C G_P}{1 + G_C G_P} Z' + \frac{G_C G_P}{1 + G_C G_P} W , \quad (11.2.2a)$$

which gives

$$Y = \frac{A_{FF_1} A C D_Z - B_{FF_1} B D C_Z}{A_{FF_1} C_Z (A C + B D)} Z' + \frac{B D}{A C + B D} W , \quad (11.2.2b)$$

where for brevity the argument s is omitted. From the transfer functions of Eq. (11.2.2b) one can see that the characteristic equation is

$$A_{FF_1} C_Z (A C + B D) = 0 \quad (11.2.3a)$$

with regard to disturbance behaviour and

$$A C + B D = 0 \quad (11.2.3b)$$

with regard to the reference behaviour. The disturbance will be fully compensated if

$$G_{PZ} = G_{FF_1} G_C G_P , \quad (11.2.4)$$

from which the required transfer function for the feed-forward element is

$$G_{FF_1} = \frac{G_{PZ}}{G_C G_P} = \frac{A C D_Z}{B D C_Z} . \quad (11.2.5)$$

This approach can only be realised by a controller if the pole excess of G_P is not larger than that of G_{PZ} . Otherwise a total compensation is not possible. Moreover, the polynomial $B D C_Z$ must be Hurwitzian.

For the frequent case that the disturbance and control behaviour are equal, i.e. the case of $G_{PZ} = G_P$, the transfer function of the feed-forward elements is

$$G_{FF_1} = \frac{1}{G_C} = \frac{A}{B} . \quad (11.2.6)$$

As the total compensation of a disturbance in a plant with P behaviour is only possible by a controller with I behaviour, the transfer function of the feed-forward element, according to Eq. (11.2.6), should thus show ideal D behaviour. If there is a PI controller in the loop, the feed-forward element must be designed as a DT_1 element.

Often the feed-forward element cannot be realised as ideally designed according to Eqs. (11.2.5) or (11.2.6), because G_C , besides pure I behaviour, normally contains delay elements. Also in these cases a DT_1 element is recommended.

11.2.2 Disturbance feed-forward on the manipulated variable

The configuration with feed-forward on the manipulated variable or on the actuator, respectively, is shown in Figure 11.2.2. From this for the controlled variable it follows that

$$Y = [(W - Y) G_C - Z' G_{FF_2}] G_P + Z' G_{PZ}$$

and after rearranging

$$Y = \frac{G_{PZ} - G_{FF_2} G_P}{1 + G_C G_P} Z' + \frac{G_C G_P}{1 + G_C G_P} W \quad (11.2.7a)$$

or

$$Y = \frac{A(A_{FF_2} C D_Z - B_{FF_2} D C_Z)}{A_{FF_2} C_Z (A C + B D)} Z' + \frac{B D}{A C + B D} W . \quad (11.2.7b)$$

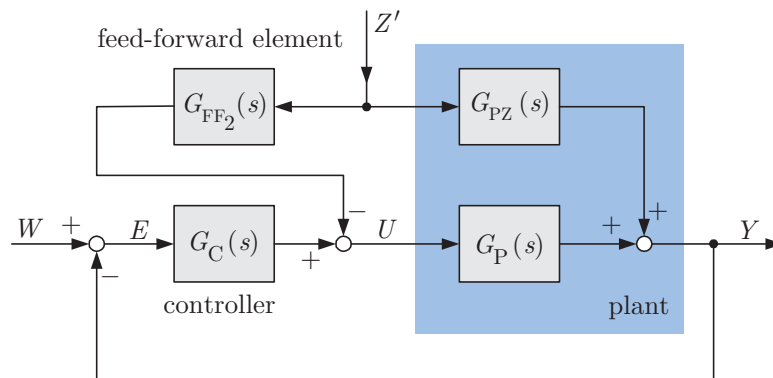


Figure 11.2.2: Block diagram of the feed-forward on the manipulated variable

The characteristic equations are the same as in the previous case with feed-forward to the controller. For the ideal compensation of the disturbances it follows from Eq. (11.2.7) that

$$G_{PZ} = G_{FF_2} G_P , \quad (11.2.8)$$

from which the transfer function of the feed-forward element follows is

$$G_{FF_2} = \frac{G_{PZ}}{G_P} = \frac{C D_Z}{D C_Z}. \quad (11.2.9)$$

For the special case of $G_{PZ} = G_P$, where the disturbance z acts directly at the plant input, the compensation is performed by $G_{FF_2} = 1$ directly at the plant input..

Similarly as in the case of Eq. (11.2.5) the realisation of the feed-forward element according to Eq. (11.2.9) is not possible if

$$\text{degree } D_Z + \text{degree } C > \text{degree } C_Z + \text{degree } D \quad (11.2.10)$$

with $G_{PZ} = D_Z/C_Z$ and $G_P = D/C$ is valid, as G_{FF_2} must be realised by PD elements. Also in the case of a non-minimum phase behaviour of G_P or of instability of G_{PZ} the Eq. (11.2.9) cannot be realised, as the required feed-forward element is unstable. In those cases in which a dynamical compensation according to Eq. (11.2.9) is not possible one must be content with a static compensation using a P element

$$G_{FF_2} = \frac{K_{PZ}}{K_P}, \quad (11.2.11)$$

where K_{PZ} and K_P are the gains of the transfer functions G_{PZ} and G_P .

Figure 11.2.3 shows disturbance feed-forward configurations on (a) the controller and (b) the manipulated

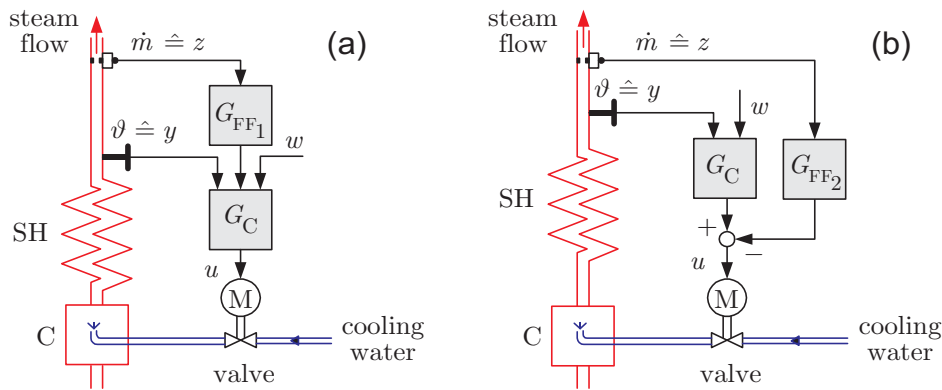


Figure 11.2.3: Examples of disturbance feed-forward configurations (a) on the controller and (b) on the actuator of a steam superheater temperature control system

variable, for the case of a temperature control system of a steam superheater (SH) in a power station. The steam temperature ϑ at the superheater outlet is the controlled variable. The manipulated variable is the cooling water flow in the spray-water cooler (C). Fluctuations of the steam flow \dot{m} have an influence on the steam temperature and are treated as disturbances. The steam flow (disturbance z) is measured and fed via G_{FF_1} or G_{FF_2} to the controller or to the manipulated variable, respectively.

11.3 Control systems with an auxiliary manipulated variable

For plants with a distinctive delayed behaviour, besides the actual controlled variable y , a secondary variable can often be measured and used as an auxiliary variable y_A . The auxiliary control loop consists, as shown in Figure 11.3.1, of the first part of the plant with the transfer function $G_{P_1}(s)$ and the auxiliary controller with the transfer function $G_{CA}(s)$. The controlled variable then follows directly as shown in Figure 11.3.1. Now

$$Y = \left\{ \left[(W - Y) G_C - \frac{Y}{G_{P_2}} G_{CA} \right] G_{P_1} + Z' G_{PZ} \right\} G_{P_2} \quad (11.3.1)$$

or on rearranging

$$Y = \frac{G_{PZ} G_{P_2}}{1 + (G_C G_{P_2} + G_{CA}) G_{P_1}} Z' + \frac{G_C G_{P_1} G_{P_2}}{1 + (G_C G_{P_2} + G_{CA}) G_{P_1}} W \quad (11.3.2a)$$

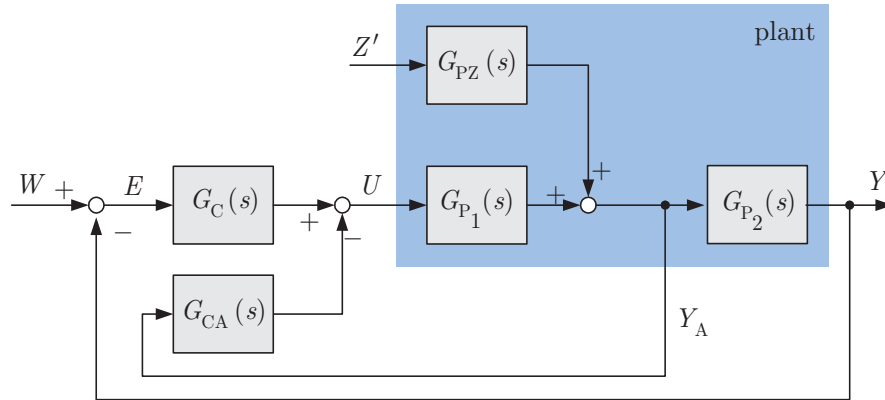


Figure 11.3.1: Block diagram of a control system with auxiliary manipulated variable y_A

giving

$$Y = \frac{A C_1 D_2 D_Z}{C_Z [A C_2 (C_1 A_A + D_1 B_A) + D_1 D_2 B A_A]} Z' + \frac{B D_1 D_2 A_A}{A C_2 (C_1 A_A + D_1 B_A) + D_1 D_2 B A_A} W \quad (11.3.2b)$$

with

$$G_{CA} = \frac{B_A}{A_A} ; \quad G_{P1} = \frac{D_1}{C_1} ; \quad G_{P2} = \frac{D_2}{C_2} .$$

The characteristic equation with regard to the disturbance behaviour is

$$C_Z [A C_2 (C_1 A_A + D_1 B_A) + D_1 D_2 B A_A] = 0 \quad (11.3.3a)$$

and with regard to the reference behaviour is

$$A C_2 (C_1 A_A + D_1 B_A) + D_1 D_2 B A_A = 0 . \quad (11.3.3b)$$

From this it is obvious that the introduction of the auxiliary controlled variable has an influence on the stability of the control system.

By a proper choice of G_{CA} , on the one hand a reduction of the effects of the disturbance on the second section of the plant (G_{P2}) can be achieved and on the other an improvement of the behaviour of the main control loop. The location where the auxiliary measurement y_A is taken should be after the entry of the disturbance but as close as possible to the plant input. If the first section of the plant contains only short delays then a P controller is sufficient for G_{CA} . Often the need for an auxiliary controller can be avoided if the auxiliary variable is connected directly to the input of the main controller via a PT_1 element.

Figure 11.3.2 again shows the example of the temperature control of a steam superheater system, where here it is configured with an auxiliary controller that uses the steam temperature measurement ϑ_1 at the superheater inlet as an auxiliary controlled variable.

DYNAST study example 11.1
[PID and PI-D control of a \$PT_1T_t\$ plant](#)

DYNAST study example 11.2
[PID and PI-D control of a 3rd-order plant](#)

11.4 Cascade control systems

Cascade control systems are special cases of control systems with auxiliary controlled variables. Here, as shown in Figure 11.4.1, the main controller with the transfer function G_{C2} does not directly effect

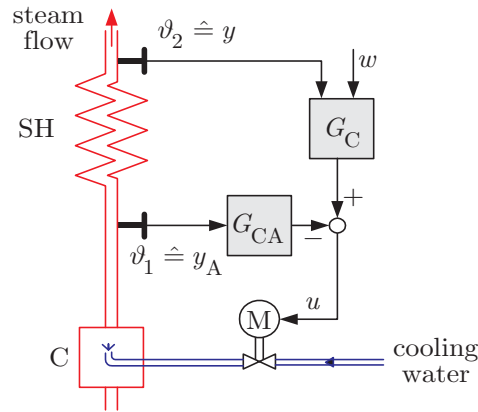


Figure 11.3.2: Example of a steam superheater temperature control system with auxiliary controlled variable y_A

the actuator, but provides the reference value for the underlying auxiliary controller with the transfer function G_{C_1} . This auxiliary controller forms together with the first plant section G_{P_1} the auxiliary control loop, which is inside the main control loop. Disturbances in the first plant section will be already controlled by the auxiliary controller such that they have less influence on the second section. The main controller has then only to act slightly.

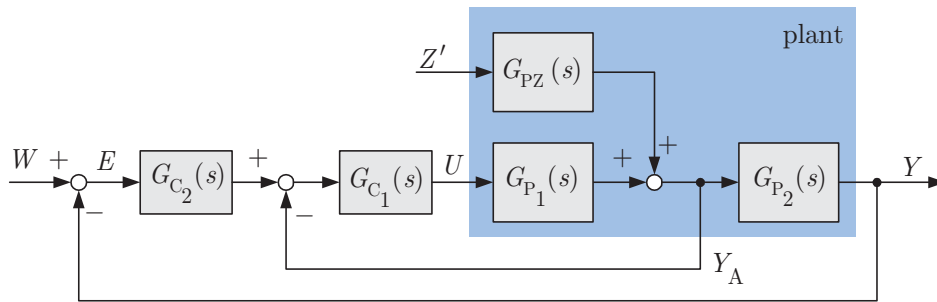


Figure 11.4.1: Block diagram of a cascade control system

When multiple auxiliary variables are measured, multiple cascade control systems can be built. For the cascade control system of Figure 11.4.1 the controlled variable y is given by

$$Y = \left\{ \left[(W - Y) G_{C_2} - \frac{Y}{G_{C_2}} \right] G_{C_1} G_{P_1} + Z' G_{PZ} \right\} G_{P_2} , \tag{11.4.1}$$

which gives

$$Y = \frac{G_{PZ} G_{P_2}}{1 + G_{C_1} G_{P_1} (1 + G_{C_2} G_{P_2})} Z' + \frac{G_{C_1} G_{C_2} G_{P_1} G_{P_2}}{1 + G_{C_1} G_{P_1} (1 + G_{C_2} G_{P_2})} W \tag{11.4.2a}$$

and can be written as

$$Y = \frac{A_1 A_2 C_1 D_2 D_Z}{C_Z [A_1 A_2 C_1 C_2 + B_1 D_1 (A_2 C_2 + B_2 D_2)]} Z' + \frac{B_1 B_2 D_1 D_2}{A_1 A_2 C_1 C_2 + B_1 D_1 (A_2 C_2 + B_2 D_2)} W \tag{11.4.2b}$$

with

$$G_{C_1} = \frac{B_1}{A_1} ; \quad G_{C_2} = \frac{B_2}{A_2} .$$

The characteristic equation with regard to the disturbance behaviour is

$$C_Z [A_1 A_2 C_1 C_2 + B_1 D_1 (A_2 C_2 + B_2 D_2)] = 0 \quad (11.4.3a)$$

and with regard to the reference behaviour

$$A_1 A_2 C_1 C_2 + B_1 D_1 (A_2 C_2 + B_2 D_2) = 0 . \quad (11.4.3b)$$

It is obvious that the stability is influenced by the underlying auxiliary control loop. If the reference behaviour of the auxiliary control loop in Figure 11.4.1 is summarised as

$$G_A = \frac{G_{C_1} G_{P_1}}{1 + G_{C_1} G_{P_1}} , \quad (11.4.4)$$

then Eq. (11.4.2a) can be represented by

$$Y = \frac{G_{P_2}}{1 + G_{C_2} G_A G_{P_2}} \frac{G_{P_Z}}{1 + G_{C_1} G_{P_1}} Z' + \frac{G_{C_2} G_A G_{P_2}}{1 + G_{C_2} G_A G_{P_2}} W . \quad (11.4.5)$$

From this equation the block diagram of a single-loop control system can be drawn as shown in Figure 11.4.2, which describes the same system as Figure 11.4.1. The auxiliary control system with the

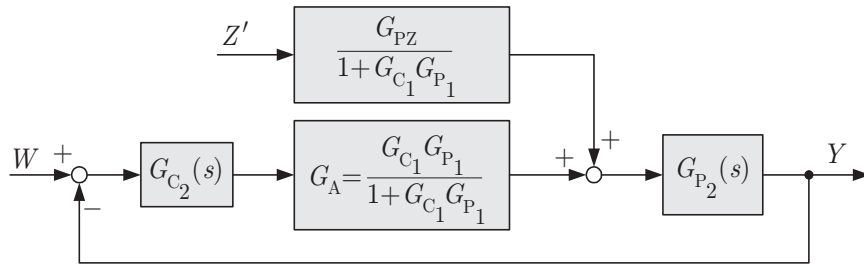


Figure 11.4.2: Rearranged block diagram of a cascade control system

transfer function G_A is an element of the basic control loop. Therefore, it is possible to design a cascade control system in the following steps:

1. Design of the auxiliary control system, i.e. parametrisation of the controller transfer function G_{C_1} for the given section G_{P_1} of the plant for disturbances. The auxiliary control system must be fast (high bandwidth) and therefore mostly a P or PD controller is chosen for G_{C_1} .
2. Design of the main controller transfer function G_{C_2} for the plant transfer function $G_A G_{P_2}$. G_{C_2} has the task to remove steady-state errors in the controlled variable. Therefore, it is expedient to use a PI controller as long as the plant has PT_n behaviour.

Figure 11.4.3 shows two examples of cascade control systems.

Demonstration Example 11.1
[A virtual experiment using cascade control](#)

11.5 Control system with auxiliary manipulated variable

Disturbances can also be counteracted by a configuration where an additional auxiliary control signal u_A is inserted between the actuator of the main control loop and the controlled variable. This auxiliary manipulated variable will be generated by an additional auxiliary controller G_{CA} . Figure 11.5.1 shows the corresponding block diagram.

From this diagram it directly follows that the controlled variable is

$$Y = [(W - Y) G_C G_{P_1} + Z' G_{P_Z} + (W - Y) G_{CA}] G_{P_2} \quad (11.5.1)$$

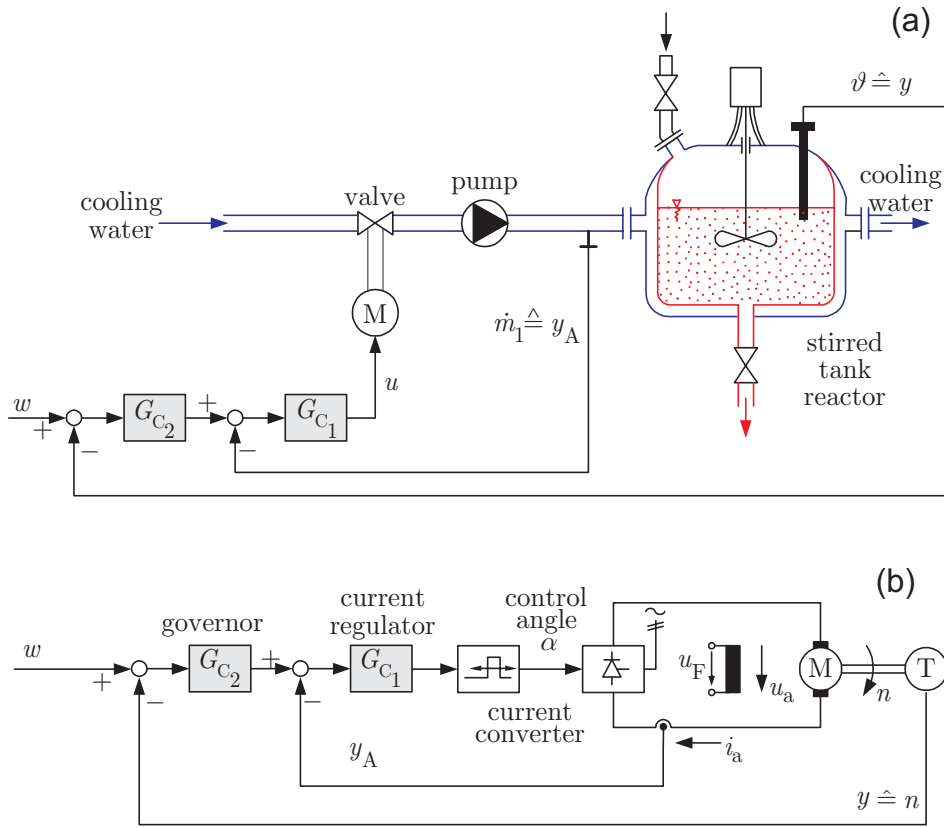


Figure 11.4.3: Examples of cascade control systems: (a) temperature control of a stirred tank reactor, (b) governor of a DC motor with motor current control

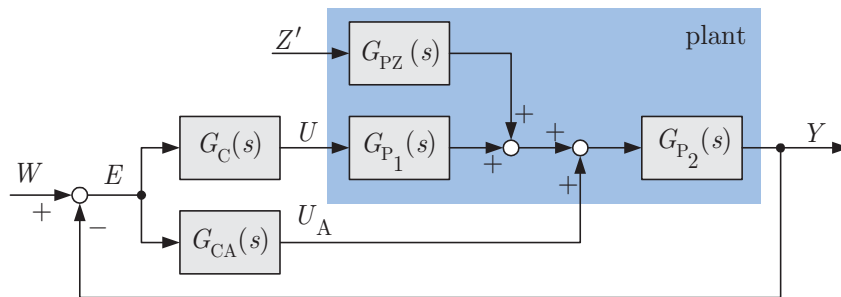


Figure 11.5.1: Block diagram of a control system with an auxiliary manipulated variable u_A

or rearranged

$$Y = \frac{G_{PZ}G_{P_2}}{1 + (G_C G_{P_1} + G_{CA}) G_{P_2}} Z' + \frac{G_{P_2}(G_{CA} + G_C G_{P_1})}{1 + (G_C G_{P_1} + G_{CA}) G_{P_2}} W, \quad (11.5.2a)$$

which becomes

$$Y = \frac{A C_1 D_2 A_A D_Z}{C_Z [A C_1 (C_2 A_A + D_2 B_A) + B D_1 D_2 A_A]} Z' + \frac{D_2 (B D_1 A_A + A C_1 B_A)}{A C_1 (C_2 A_A + D_2 B_A) + B D_1 D_2 A_A} W. \quad (11.5.2b)$$

The characteristic equation with regard to the disturbance behaviour is

$$C_Z [A C_1 (C_2 A_A + D_2 B_A) + B D_1 D_2 A_A] = 0 \quad (11.5.3a)$$

and with regard to the reference behaviour is

$$A C_1 (C_2 A_A + D_2 B_A) + B D_1 D_2 A_A = 0. \quad (11.5.3b)$$

The stability of the main control loop is influenced by adding the auxiliary manipulated variable u_A . During the choice of the auxiliary manipulated variable it must be observed that the second plant section G_{P_2} should have an as small as possible delay, because the auxiliary controller can cancel disturbances faster. In the steady state u_A must be zero, if the steady state must be influenced only by u . This is possible when for G_{CA} a DT_1 element is used.

11.6 Control system with anti-windup measure

In practical applications the manipulated variable u must not exceed given extreme values. This is the case due to either the bounded power of the actuator or to the physical constraints of the plant. In most cases these hard limitations of the manipulated variable must be respected. This means that the modulus of the manipulated variable must not exceed given bounds

$$|u(t)| \leq u_{\max} . \quad (11.6.1)$$

For control system design using linear methods it is difficult to cope with this problem and to abide by the bounds on the manipulated variable. When the design is performed such that the amplitudes of the manipulated variable are small and do not reach the bounds, the actuator is not fully exploited and, thus, the control response is slow. On the other hand, when the bounds are exceeded for a reasonable period of time undesirable control behaviour may be obtained.

In order to discuss the problem, the bounds are described by a saturation element, as shown in Figure 11.6.1. The variable u_C is the manipulated variable obtained from the controller and u the manipu-

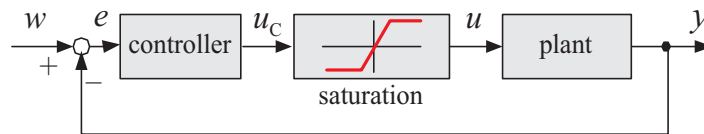


Figure 11.6.1: Block diagram of a control system with a bounded manipulated variable

lated variable acting on the plant, which is determined from

$$u(t) = \begin{cases} -u_{\max} & \text{for } u_C(t) < -u_{\max} \\ u_C(t) & \text{for } |u_C(t)| \leq u_{\max} \\ u_{\max} & \text{for } u_C(t) > u_{\max} \end{cases} . \quad (11.6.2)$$

When the bounds are exceeded the nonlinear saturation characteristic will take effect and influence the dynamical behaviour. In some cases the closed-loop system may also become unstable or show an oscillating behaviour.

This undesired phenomenon, called the *windup effect*, occurs in all control systems where an integrator is used in the controller. This integrator is necessary to have a zero steady-state control error. In order to demonstrate this effect, the example from section 9.3 is taken. The plant is given by Eq. (9.3.1) and the controller by Eq. (9.3.7). The step response of the closed loop without saturation is shown in Figure 11.6.2, where the response of the controlled variable is the same as in Figure 9.3.2. If the manipulated variable is bounded by $u_{\max} = 1.5$ (with saturation) the rise time increases due to the smaller values of the manipulated variable in the period from 0.1 s to 2 s. The increased maximum overshoot and settling time reflect a worse control behaviour. The reason for this is the following: From the beginning, the control error decreases and changes sign at 1.6 s. As u_C is very large at this time (≈ 2.1), the manipulated variable u cannot be reduced despite the negative control error. This will only occur when u_C falls below u_{\max} at 2 s. The problem is obviously that the controller continues to integrate though the manipulated variable has already reached its bound. As the controller output u_C further grows unnecessarily, this is called the windup effect.

The goal of an anti-windup measure is to counteract the integration of the controller. This can be performed by feeding back the difference $u - u_C$ to the controller. Figure 11.6.3 shows a simple approach

for an anti-windup measure, where the difference is weighted by the factor γ and fed into the controller. Figure 11.6.2 shows the improvement of the behaviour for $\gamma = 1$. The settling time is close to the case without saturation, but the maximum overshoot is half of that without saturation.

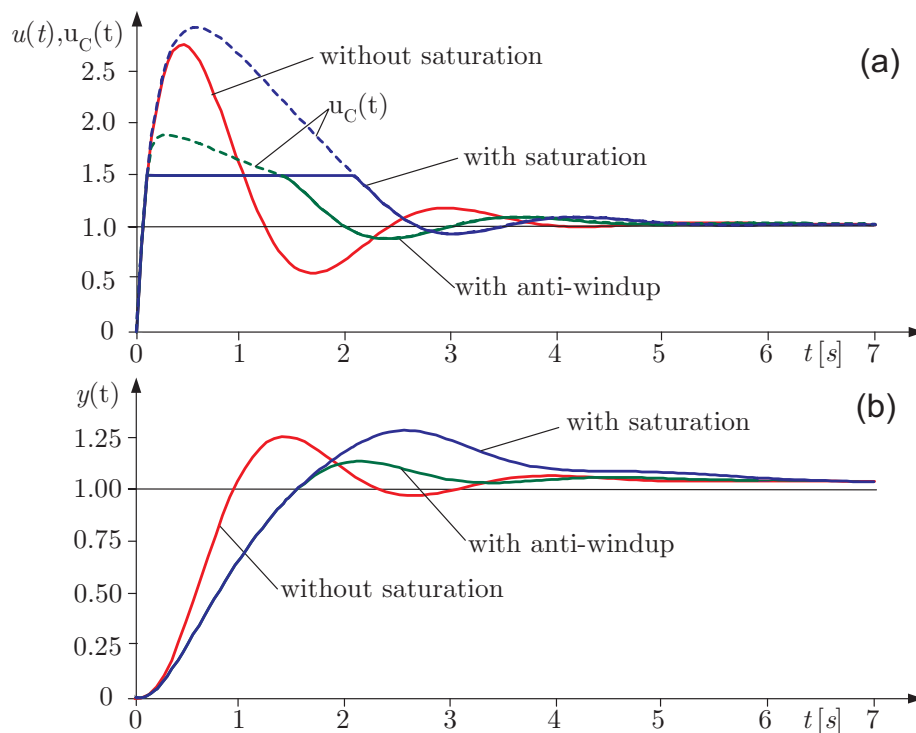


Figure 11.6.2: Step response of the closed loop system with and without anti-windup measure, (a) manipulated variable u and controller output u_C , (b) controlled variable y

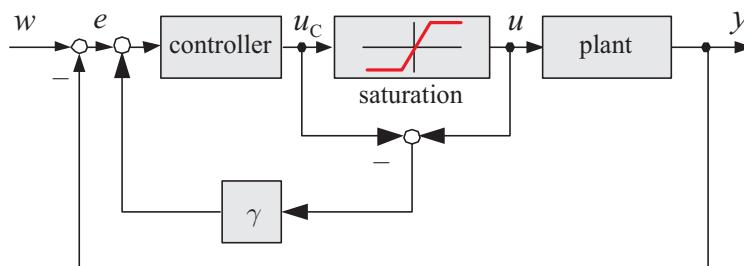


Figure 11.6.3: Block diagram of a control system with an anti-windup measure

Module 12

State-space representation

Module units

12.1 State-space representation of single-input-single-output systems	12-1
12.2 State-space representation of multi-input-multi-output systems	12-2
12.3 The relationship between transfer functions and the state-space representation	12-3
12.4 State-space vs transfer function approach	12-4
12.5 Uniqueness of the state variables	12-4
12.6 Controllability and observability	12-6

Module overview. *This introductory module begins by considering the reasons and advantages for use of a state-space description. The description is introduced using a specific example. The relationship to the transfer-function form is outlined and discussed. Fundamentals, like transformations, controllability and observability are briefly introduced.*

Module objectives. *When you have completed this module you should be able to:*

1. Represent a system in state-space form.
2. Determine the transfer function from a description in state-space form.
3. Transform a system into other state-space forms.

Module prerequisites. *Differential equations, transfer function.*

12.1 State-space representation of single-input-single-output systems

In the following a short introduction into the representation of systems using state-variable techniques is given. For this purpose, the example of the *RLC* network from Figure 4.4.10 is used. The dynamical behaviour of this network is completely defined for $t \geq t_0$, if the

- initial conditions $u_a(t_0), i_e(t_0)$

and the

- input variable $u_e(t)$ for $t \geq t_0$

are known. For these specifications the variables $i_e(t)$ and $u_a(t)$ can be determined for all $t \geq t_0$. The variables $i_e(t)$ and $u_a(t)$ characterise the 'state' of the network and are therefore called *state variables* of the network.

The differential equations of Eqs. (4.4.29) and (4.4.30) describe the dynamical behaviour of this network. Inserting Eq. (4.4.30) into Eq. (4.4.29) one obtains the 2nd-order differential equation according to Eq. (4.4.31), which completely describes the system with respect to the input-output behaviour. But one can also use the two original differential equations and can write them in vector notation so that the 1st-order vector differential equation

$$\begin{bmatrix} \frac{di_e(t)}{dt} \\ \frac{du_a(t)}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{R}{L} & -\frac{1}{L} \\ \frac{1}{C} & 0 \end{bmatrix} \begin{bmatrix} i_e(t) \\ u_a(t) \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} u_e(t) \quad (12.1.1)$$

with the initial condition

$$\begin{bmatrix} i_e(t_0) \\ u_a(t_0) \end{bmatrix}$$

is obtained. This linear 1st-order vector differential equation describes the connection between the input variable and the state variables. To complete a state-space system, one needs an additional equation that describes the dependence of the output variable on the state variables. In this example, it is the direct relationship

$$y(t) = u_a(t) .$$

Introducing the state vector

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} i_e(t) \\ u_a(t) \end{bmatrix}$$

into Eq. (12.1.1), with the vectors

$$\mathbf{x}_0 = \mathbf{x}(t_0) = \begin{bmatrix} i_e(t_0) \\ u_a(t_0) \end{bmatrix} ,$$

$$\mathbf{b} = \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{c}^T = [0 \quad 1] ,$$

with the matrix

$$\mathbf{A} = \begin{bmatrix} -\frac{R}{L} & -\frac{1}{L} \\ \frac{1}{C} & 0 \end{bmatrix}$$

and with the scalar variables

$$u(t) = u_e(t) \quad \text{and} \quad d = 0$$

one obtains the general state-space representation of a linear time-invariant single-input-single-output system:

$$\dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{b} u(t) \quad \mathbf{x}(t_0) \text{ initial condition} \quad (12.1.2)$$

$$y(t) = \mathbf{c}^T \mathbf{x}(t) + d u(t) . \quad (12.1.3)$$

The Eq. (12.1.2) is the *state equation*, and in the general case it is a linear system of 1st-order differential equations of n state variables x_1, x_2, \dots, x_n , which are combined in the state vector $\mathbf{x} = [x_1 x_2 \dots x_n]^T$. Eq. (12.1.3) is the *output equation*, which maps the states and inputs linearly to the output. This is an algebraic equation, whereas the state equation is a differential equation.

12.2 State-space representation of multi-input-multi-output systems

The Eqs. (12.1.2) and (12.1.3) describe an n th-order linear time-invariant single-input-single-output system. For linear multi-input-multi-output systems of order n with r inputs and m outputs these equations become

$$\dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{B} \mathbf{u}(t) \quad \text{with the initial condition} \quad \mathbf{x}(t_0) \quad (12.2.1)$$

$$\mathbf{y}(t) = \mathbf{C} \mathbf{x}(t) + \mathbf{D} \mathbf{u}(t) , \quad (12.2.2)$$

where the following notation is used:

state vector	$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}$	$(n \times 1)$ vector
input vector	$\mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_r(t) \end{bmatrix}$	$(r \times 1)$ vector
output vector	$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_m(t) \end{bmatrix}$	$(m \times 1)$ vector
system matrix	\mathbf{A}	$(n \times n)$ matrix
input matrix	\mathbf{B}	$(n \times r)$ matrix
output matrix	\mathbf{C}	$(m \times n)$ matrix
feedthrough matrix	\mathbf{D}	$(m \times r)$ matrix

It goes without saying that the general representation of Eqs. (12.2.1) and (12.2.2) also includes the single-input-single-output case. The matrices \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{D} have constant elements. If these elements are time-varying, the matrices of the corresponding time-varying system are substituted by matrix functions of time, e.g. $\mathbf{A}(t)$.

12.3 The relationship between transfer functions and the state-space representation

In the following, the Eqs. (12.2.1) and (12.2.2) will be transformed into the s domain using the Laplace transform, which will be done analogously to the scalar case in section 2.5. For this, the operator notation

$$\mathbf{F}(s) = \mathcal{L}[\mathbf{f}(t)]$$

from section 2.1 is adopted and when applying it to Eq. (12.2.1), one obtains

$$s\mathbf{X}(s) - \mathbf{x}(t_0) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s) ,$$

or rearranged

$$(s\mathbf{I} - \mathbf{A})\mathbf{X}(s) = \mathbf{x}(t_0) + \mathbf{B}\mathbf{U}(s) .$$

The solution of the state equation in the s domain is then given by

$$\mathbf{X}(s) = \mathbf{\Phi}(s) \mathbf{x}(t_0) + \mathbf{\Phi}(s) \mathbf{B}\mathbf{U}(s) \quad (12.3.1)$$

with

$$\mathbf{\Phi}(s) = (s\mathbf{I} - \mathbf{A})^{-1} . \quad (12.3.2)$$

Similarly, for Eq. (12.2.2) yields

$$\mathbf{Y}(s) = \mathbf{C}\mathbf{X}(s) + \mathbf{D}\mathbf{U}(s) .$$

Substituting $\mathbf{X}(s)$ from Eq. (12.3.1), the system output in the s domain is

$$\mathbf{Y}(s) = \mathbf{C}\mathbf{\Phi}(s) \mathbf{x}(t_0) + [\mathbf{C}\mathbf{\Phi}(s) \mathbf{B} + \mathbf{D}] \mathbf{U}(s) .$$

To obtain the relationship with transfer functions, the initial condition $\mathbf{x}(t_0)$ has to be set to zero. For a single-input-single-output system according to Eqs. (12.1.2) and (12.1.3) the system output is

$$Y(s) = [\mathbf{c}^T \mathbf{\Phi}(s) \mathbf{b} + d] U(s) .$$

Comparing this equation with Eq. (3.1.3) the transfer function is given by

$$G(s) = \mathbf{c}^T \Phi(s) \mathbf{b} + d . \quad (12.3.3)$$

The matrix $\Phi(s)$ from Eq. (12.3.2) is a matrix of rational functions of s , which can always be represented by

$$\Phi(s) = \frac{1}{\det(s\mathbf{I} - \mathbf{A})} \Psi(s) , \quad (12.3.4)$$

where $\Psi(s)$ is a matrix with polynomial elements in s . From Eq. (3.1.2) it is obvious that

$$N(s) = \mathbf{c}^T \Psi(s) \mathbf{b} + d D(s) \quad (12.3.5)$$

and

$$D(s) = \det(s\mathbf{I} - \mathbf{A}) , \quad (12.3.6)$$

which is the characteristic polynomial of the system. The zeros of this polynomial are the poles of the transfer function and at the same time eigenvalues of the system matrix \mathbf{A} . If the system in the state-space representation is fully controllable and observable (see section 12.6), then the number of poles are equal to the number of eigenvalues.

12.4 State-space vs transfer function approach

The key advantage of transfer functions is in their compactness, which makes them suitable for frequency-domain analysis and stability studies. However, the transfer function approach suffers from neglecting the initial conditions. Not only does state-space representation serve as an alternative to transfer functions, but also it is not limited to linear and time-invariant systems and it has the following advantages:

1. Single-input-single-output and multi-input-multi-output systems can be formally treated equal.
2. The state-space representation is best suited both for the theoretical treatment of control systems (analytical solutions, optimisation) and for numerical calculations.
3. The determination of the system response in the homogeneous case with the initial condition $\mathbf{x}(t_0)$ is very simple.
4. This representation gives a better insight into the inner system behaviour. General system properties, for example, the system controllability or observability can be defined and determined, see section 12.6.

12.5 Uniqueness of the state variables

Initially it sounds paradoxical that the choice of the state variables is *not unique*. This means that for one and the same system with the input \mathbf{u} , the output \mathbf{y} and n state variables, there exist an infinite number of state-space representations. For each value of time t one gets the state in the n -dimensional *state space*. The n values are the cartesian coordinates of the state

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \dots + x_n \mathbf{e}_n ,$$

where the unit vectors \mathbf{e}_i are n -dimensional linear independent vectors. Their elements are – besides the i th element, which has the value of 1 – all zero. For describing the state also other basis vectors can be used. Candidates are all n linear independent and n -dimensional vectors \mathbf{t}_i . Therefore, it is always possible, to write the state as

$$\mathbf{x} = z_1 \mathbf{t}_1 + z_2 \mathbf{t}_2 + \dots + z_n \mathbf{t}_n . \quad (12.5.1)$$

After introducing the new state

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

and the quadratic matrix

$$\mathbf{T} = [t_1, t_2, \dots, t_n]$$

one can rewrite Eq. (12.5.1) as

$$\mathbf{x} = \mathbf{T} \mathbf{z} . \quad (12.5.2)$$

The constant and regular matrix \mathbf{T} is a so-called *transformation matrix*. Now, instead of the state \mathbf{x} the new state \mathbf{z} can be used and its behaviour analysed. Describing the system in the new coordinates the state \mathbf{x} from Eq. (12.5.2) is inserted into Eqs. (12.2.1) and (12.2.2) and one obtains

$$\dot{\mathbf{z}}(t) = (\mathbf{T}^{-1} \mathbf{A} \mathbf{T}) \mathbf{z}(t) + (\mathbf{T}^{-1} \mathbf{B}) \mathbf{u}(t) \quad (12.5.3)$$

$$\mathbf{y}(t) = (\mathbf{C} \mathbf{T}) \mathbf{z}(t) + \mathbf{D} \mathbf{u}(t) \quad (12.5.4)$$

with the initial condition

$$\mathbf{z}(t_0) = \mathbf{T}^{-1} \mathbf{x}(t_0) .$$

The benefits of the transformation of systems into different state-space representations are:

- Most system properties do not depend on the choice of the state variables. They remain unchanged after a regular transformation and may be analysed in an appropriate representation form.
- The computational determination and analysis of system properties can be tremendously simplified if the representation form is specifically selected. In particular certain canonical forms are of interest.

Example 12.5.1

In order to demonstrate a transformation, the example from Eq. (12.1.1) is used with the system parameters $R = 3$, $L = 1$ and $C = 0.5$. The initial condition is assumed to be zero and therefore omitted. With these values one obtains the state equation as

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -3 & -1 \\ 2 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t)$$

and the output equation as

$$y(t) = [0 \quad 1] \mathbf{x}(t) .$$

For the regular transformation matrix

$$\mathbf{T} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

one obtains the matrices for Eqs. (12.5.3) and (12.5.4) as

$$\mathbf{T}^{-1} \mathbf{A} \mathbf{T} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} -3 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\mathbf{T}^{-1} \mathbf{B} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\mathbf{C} \mathbf{T} = [0 \quad 1] \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = [-1 \quad 2] .$$

The new representation

$$\begin{aligned} \dot{\mathbf{z}}(t) &= \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{z}(t) + \begin{bmatrix} 2 \\ 1 \end{bmatrix} u(t) \\ y(t) &= [-1 \quad 2] \mathbf{z}(t) \end{aligned}$$

consists of two decoupled differential equations with respect to the state variables z_1 and z_2 . ■

The analysis and treatment of a system in such a structured representation form, as shown in the example above, is doubtless more simple. As the representation form must be specifically selected depending on the type of analysis or synthesis problem, the different representation forms, for example, the canonical forms, are not discussed separately and are introduced in the following sections when they are needed.

12.6 Controllability and observability

Controllability and observability are important structural properties of a dynamic system. Controllability can be defined as follows:

The system of Eqn. (12.1.2) is controllable if there exists a control signal $u(t)$ that will take the state of the system from any initial state $\mathbf{x}(t_0)$ to any desired final state \mathbf{x}_f in a finite time interval.

This condition is equivalent to the following condition:

The system of Eqn. (12.1.2) is controllable if the controllability matrix

$$\mathbf{S} = [\mathbf{b} \quad \mathbf{A}\mathbf{b} \quad \dots \quad \mathbf{A}^{n-2}\mathbf{b} \quad \mathbf{A}^{n-1}\mathbf{b}] \quad (12.6.1)$$

has full rank n .

The concept of observability is parallel to that of controllability and all can be transformed to statements about observability by invoking the property of duality, as discussed later in section 13.4.2. The observability definition analogous to those for controllability are as follows:

The system of Eqs. (12.1.2) and (12.1.3) is observable if, for any $\mathbf{x}(t_0)$, there is a finite time τ such that $\mathbf{x}(t_0)$ can be determined from $u(t)$ and $y(t)$ for $0 \leq t \leq \tau$.

This condition is equivalent to the following:

The system of Eqs. (12.1.2) and (12.1.3) is observable if the observability matrix

$$\mathbf{O} = \begin{bmatrix} \mathbf{c}^T \\ \mathbf{c}^T \mathbf{A} \\ \vdots \\ \mathbf{c}^T \mathbf{A}^{n-1} \end{bmatrix} \quad (12.6.2)$$

has full rank n .

Module 13

Design of state-feedback control systems

Module units

13.1 Structures and properties of state-feedback control systems	13-2
13.1.0.1 State-feedback control in the frequency domain	13-3
13.1.0.2 Steady state of the closed-loop system	13-3
13.2 State-feedback control with integrator	13-4
13.3 Design of state-feedback controllers by pole placement	13-5
13.3.1 Design of a system in controller canonical form	13-6
13.3.1.1 Design of a system not in a canonical form	13-7
13.3.2 Design using Ackermann's formula	13-8
13.4 State reconstruction using observers	13-9
13.4.1 Structure of an observer	13-10
13.4.2 Design of observers	13-10
13.5 Combined observer-controllers	13-11
13.6 Example of a state-feedback control system	13-12

Module overview. *The main design approach for systems described in state-space form is the use of state feedback. One selects pole locations to achieve a satisfactory dynamic response and develops the control law for the closed-loop system that corresponds to satisfactory dynamic response. One has to design an estimator for the states, because these are generally not measurable. This estimator is an observer that delivers the information about the states so that they can be used for control. The combined observer-controller problem is discussed. Several pole-placement designs for controllers with proportional and integral state feedback and observers based on controller canonical form are given. A comprehensive example shows the overall design procedure.*

Module objectives. *When you have completed this module you should be able to:*

1. Design control systems using state feedback for pole placement.
2. Design observers by pole placement.
3. Transform a state-space representation into controller canonical form.

Module prerequisites. *State space representation, transfer function, poles and zeros.*

13.1 Structures and properties of state-feedback control systems

In the following, the design of state-feedback controllers for single-input-single-output systems described by

$$\dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{b} u(t) \quad \mathbf{x}(t_0) \text{ initial condition} \quad (13.1.1)$$

$$y(t) = \mathbf{c}^T \mathbf{x}(t) \quad (13.1.2)$$

is discussed in detail.

The dynamical characteristics, for example, stability, decay of oscillations or sensitivity to disturbances, are determined by the distribution of the eigenvalues of the system matrix \mathbf{A} in the s plane. The goal is to influence the system specifically so that it shows a desired behaviour. In the sense of command input regulation the control system is configured as shown in Figure 13.1.1. It is assumed, that all state

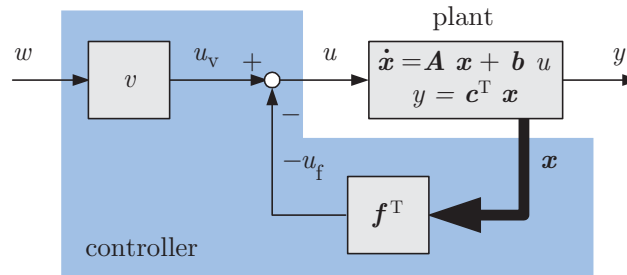


Figure 13.1.1: Basic structure of a state-feedback control system

variables are measured. A linear combination of the state variables is fed back by

$$u_f(t) = -\mathbf{f}^T \mathbf{x}(t) \quad (13.1.3)$$

with

$$\mathbf{f} = [f_1, f_2, \dots, f_n] . \quad (13.1.4)$$

While this feedback determines the dynamical behaviour, the feedforward of the command variable w

$$u_v(t) = v w(t) \quad (13.1.5)$$

using the scalar gain v influences the static behaviour. The manipulated variable is obtained from

$$u(t) = u_f(t) + u_v(t) = -\mathbf{f}^T \mathbf{x}(t) + v w(t) . \quad (13.1.6)$$

After the substitution of the manipulated variable into Eq.(13.1.1) using Eq. (13.1.6), one obtains the closed-loop system as

$$\dot{\mathbf{x}}(t) = [\mathbf{A} - \mathbf{b} \mathbf{f}^T] \mathbf{x}(t) + \mathbf{b} v w(t) \quad \mathbf{x}(t_0) \text{ initial condition} \quad (13.1.7)$$

$$y(t) = \mathbf{c}^T \mathbf{x}(t) . \quad (13.1.8)$$

For the dynamical and static behaviour the following specifications must be fulfilled:

- The dynamical behaviour of the closed loop should be specified by given poles. As these closed-loop poles are the eigenvalues of the closed-loop system matrix $(\mathbf{A} - \mathbf{b} \mathbf{f}^T)$, the desired distribution of the eigenvalues in the s plane for this matrix is specified.
- In the steady state the control error must vanish so that the plant output follows

$$y_\infty = \lim_{t \rightarrow \infty} y(t) = \mathbf{c}^T \lim_{t \rightarrow \infty} \mathbf{x}(t) = \lim_{t \rightarrow \infty} w(t) = w_\infty . \quad (13.1.9)$$

13.1.0.1 State-feedback control in the frequency domain

As the dynamical behaviour of the closed loop is specified in the s domain, it is appropriate to discuss some properties in the frequency domain. The initial point of this discussion is the closed-loop transfer function

$$\frac{Y(s)}{W(s)} = G_W(s) \quad (13.1.10)$$

Applying the Laplace transform as shown in section 12.3 to Eqs. (13.1.7) and (13.1.8) one obtains the closed-loop transfer function as

$$G_W(s) = \mathbf{c}^T (s\mathbf{I} - \mathbf{A} + \mathbf{b}\mathbf{f}^T)^{-1} \mathbf{b} v . \quad (13.1.11)$$

The same procedure for the plant from Eqs. (13.1.1) and (13.1.2) gives

$$G_P(s) = \mathbf{c}^T (s\mathbf{I} - \mathbf{A})^{-1} \mathbf{b} . \quad (13.1.12)$$

With Eq. (12.3.2) the closed-loop and plant transfer functions are rewritten as

$$G_W(s) = \mathbf{c}^T (\Phi^{-1}(s) + \mathbf{b}\mathbf{f}^T)^{-1} \mathbf{b} v \quad (13.1.13)$$

and

$$G_P(s) = \mathbf{c}^T \Phi \mathbf{b} = \frac{N(s)}{D(s)} . \quad (13.1.14)$$

Applying the matrix inversion lemma to the inner part $(\Phi^{-1}(s) + \mathbf{b}\mathbf{f}^T)^{-1}$ of Eq. (13.1.13), one obtains

$$\begin{aligned} G_W(s) &= \mathbf{c}^T \left\{ \Phi(s) - \Phi(s) \mathbf{b} \left[1 + \mathbf{f}^T \Phi(s) \mathbf{b} \right]^{-1} \mathbf{f}^T \Phi(s) \right\} \mathbf{b} v \\ &= \left\{ \mathbf{c}^T \Phi(s) \mathbf{b} - \frac{\mathbf{c}^T \Phi(s) \mathbf{b} \mathbf{f}^T \Phi(s) \mathbf{b}}{1 + \mathbf{f}^T \Phi(s) \mathbf{b}} \right\} v \\ &= \frac{\mathbf{c}^T \Phi(s) \mathbf{b} v}{1 + \mathbf{f}^T \Phi(s) \mathbf{b}} . \end{aligned} \quad (13.1.15)$$

After inserting Eqs. (12.3.4), (12.3.6) and (13.1.14) into Eq. (13.1.15) the closed-loop transfer function is given by

$$\begin{aligned} G_W(s) &= \frac{\mathbf{c}^T \Psi(s) \mathbf{b} v}{D(s) + \mathbf{f}^T \Psi(s) \mathbf{b}} \\ &= \frac{N(s) v}{D(s) + \mathbf{f}^T \Psi(s) \mathbf{b}} , \end{aligned} \quad (13.1.16)$$

where the denominator $[D(s) + \mathbf{f}^T \Psi(s) \mathbf{b}]$ is the characteristic polynomial of the closed-loop system and $D(s)$ is that of the open-loop system. As only the plant numerator polynomial $N(s)$ appears in the closed-loop transfer function, the closed-loop zeros are the same as the open-loop zeros. This means that the zeros cannot be influenced by a state-feedback controller; it only moves the poles.

13.1.0.2 Steady state of the closed-loop system

The steady state from Eq. (13.1.9) can only be reached if the state of the closed-loop system for $t \rightarrow \infty$ approaches the final value

$$\mathbf{x}_\infty = \lim_{t \rightarrow \infty} \mathbf{x}(t) .$$

The condition

$$\dot{\mathbf{x}}(t) = 0 = (\mathbf{A} - \mathbf{b}\mathbf{f}^T) \mathbf{x}_\infty + \mathbf{b} v w_\infty$$

is obviously valid, from which the steady state can be obtained as

$$\mathbf{x}_\infty = -(\mathbf{A} - \mathbf{b}\mathbf{f}^T)^{-1} \mathbf{b} v w_\infty ,$$

which is always possible, because for an asymptotically stable closed-loop the matrix $(\mathbf{A} - \mathbf{b} \mathbf{f}^T)$ has always full rank. For the output, one obtains

$$y_\infty = \mathbf{c}^T \mathbf{x}_\infty = -\mathbf{c}^T (\mathbf{A} - \mathbf{b} \mathbf{f}^T)^{-1} \mathbf{b} v w_\infty, \quad (13.1.17)$$

where it must be observed that this value must not vanish, i.e.

$$\mathbf{c}^T (\mathbf{A} - \mathbf{b} \mathbf{f}^T)^{-1} \mathbf{b} \neq 0.$$

This condition can be fulfilled, if the closed-loop transfer function $G_W(s)$ from Eq. (13.1.11) has no zero at $s = 0$. From the discussion about the zeros in section 13.1.0.1 it is clear that the plant transfer function from Eq. (13.1.12) must not have a zero at $s = 0$. This means that

$$\mathbf{c}^T \mathbf{A}^{-1} \mathbf{b} \neq 0. \quad (13.1.18)$$

With the condition given in Eq. (13.1.18) all given values y_∞ are reachable with command input signals that have a constant steady-state value of w_∞ .

From the conditions in Eqs. (13.1.17) and (13.1.9) the feed-forward gain of the controller is obtained as

$$v = \frac{-1}{\mathbf{c}^T (\mathbf{A} - \mathbf{b} \mathbf{f}^T)^{-1} \mathbf{b}}. \quad (13.1.19)$$

13.2 State-feedback control with integrator

For a given feedback vector \mathbf{f} the feedforward gain v can be calculated according to Eq. (13.1.19) so that the closed-loop system in Figure 13.1.1 shows the desired static behaviour with a vanishing control error. The control structure used is not very robust with respect to the control error, because this error is not fed back. Uncertainties in the plant model parameters or disturbances acting on the plant may cause steady-state control errors. In order to reject such effects, one can use a similar approach to that shown in section 11.4 for cascade control. The cascade control system of Figure 13.2.1 has the state-feedback

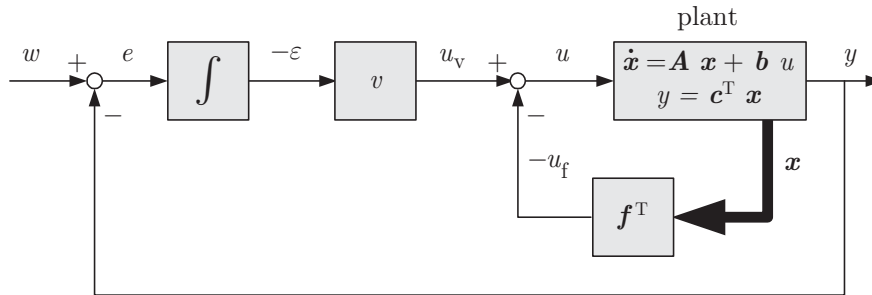


Figure 13.2.1: Basic structure of a cascade state-feedback control system with integrator

configuration from Figure 13.1.1 as its inner loop. The controlled variable y is compared with the set-point value w and the control error e is fed back to an integrator. The former feedforward gain v is now the gain of the integrator. From the closed-loop point of view with this configuration the gain like \mathbf{f} is a feedback parameter. This will be shown in the following.

From Figure 13.2.1 the equations of the closed-loop system can be directly determined as follows:

$$\dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{b} u(t) \quad \mathbf{x}(t_0) \text{ initial condition} \quad (13.2.1)$$

$$y(t) = \mathbf{c}^T \mathbf{x}(t) \quad (13.2.2)$$

$$\dot{\epsilon}(t) = y(t) - w(t) \quad (13.2.3)$$

$$u(t) = u_f(t) + u_v(t) = -\mathbf{f}^T \mathbf{x}(t) - v \epsilon(t). \quad (13.2.4)$$

Combining Eqs. (13.2.1) to (13.2.3) of the open-loop system in matrix notation, one obtains

$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\epsilon}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{c}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \epsilon(t) \end{bmatrix} + \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} \mathbf{0} \\ -1 \end{bmatrix} w(t). \quad (13.2.5)$$

Eq. (13.2.4) for the manipulated variable can be represented by

$$u(t) = - \begin{bmatrix} \mathbf{f} \\ v \end{bmatrix}^T \begin{bmatrix} \mathbf{x}(t) \\ \epsilon(t) \end{bmatrix}. \quad (13.2.6)$$

With the abbreviations

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{c}^T & 0 \end{bmatrix}, \quad \mathbf{b}^* = \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix}, \quad \mathbf{v}^* = \begin{bmatrix} \mathbf{0} \\ -1 \end{bmatrix}, \quad \mathbf{x}^*(t) = \begin{bmatrix} \mathbf{x}(t) \\ \epsilon(t) \end{bmatrix}$$

Eqs. (13.2.5) and (13.2.6) are

$$\dot{\mathbf{x}}^*(t) = \mathbf{A}^* \mathbf{x}^*(t) + \mathbf{b}^* u(t) + \mathbf{v}^* w(t) \quad \mathbf{x}^*(t_0) \text{ initial condition} \quad (13.2.7)$$

$$u(t) = -\mathbf{f}^{*T} \mathbf{x}^*(t) \quad (13.2.8)$$

and for the closed-loop system after inserting Eq. (13.2.8) into Eq. (13.2.7)

$$\dot{\mathbf{x}}^*(t) = \left[\mathbf{A}^* - \mathbf{b}^* \mathbf{f}^{*T} \right] \mathbf{x}^*(t) + \mathbf{v}^* w(t) \quad \mathbf{x}^*(t_0) \text{ initial condition}, \quad (13.2.9)$$

which is formally the same system as that with the simple state-feedback controller in Eq. (13.1.7). Here, the controller parameters in \mathbf{f}^* instead of those in \mathbf{f} must be determined for the desired specifications. This extended problem is now reduced to the original problem, and for both cases the same design procedure can be applied. It means, that for the state-feedback controller with integrator the same design procedures for the controller parameters can be applied as for the original system. This simplifies the design of the extended problem significantly.

13.3 Design of state-feedback controllers by pole placement

The difficulty of this design consists essentially in the determination of the feedback vector \mathbf{f} so that the n eigenvalues of the system matrix $(\mathbf{A} - \mathbf{b} \mathbf{f}^T)$ have the desired distribution. After that, the determination of the feedforward gain v in the control structure without integrator is very simple, see Eq. (13.1.19).

The characteristic polynomial $Q(s) = \det [s \mathbf{I} - (\mathbf{A} - \mathbf{b} \mathbf{f}^T)]$ of the closed loop system from Eq. (13.1.7) is a monic polynomial of order n . The coefficients q_i of it are functions of the controller parameters h_i :

$$\det [s \mathbf{I} - (\mathbf{A} - \mathbf{b} \mathbf{f}^T)] = q_0(\mathbf{f}) + q_1(\mathbf{f})s + q_2(\mathbf{f})s^2 + \dots + q_{n-1}(\mathbf{f})s^{n-1} + s^n.$$

By a proper choice of the vector \mathbf{f} this polynomial $Q(s)$ should be made equal to the desired polynomial $P(s)$ with n zeros, which are the desired poles or eigenvalues s_i , respectively, of the closed-loop system:

$$P(s) = \prod_{i=1}^n (s - s_i). \quad (13.3.1)$$

Multiplying all factors on the right-hand side of Eq. (13.3.1) one obtains this polynomial as

$$P(s) = p_0 + p_1 s + p_2 s^2 + \dots + p_{n-1} s^{n-1} + s^n. \quad (13.3.2)$$

For all values of s , the condition $Q(s) = P(s)$ must be fulfilled. A comparison of the corresponding terms of both sides gives the coefficients as

$$q_0(\mathbf{f}) = p_0, \quad q_1(\mathbf{f}) = p_1, \quad q_2(\mathbf{f}) = p_2, \quad \dots \quad q_{n-1}(\mathbf{f}) = p_{n-1}, \quad (13.3.3)$$

from which the controller parameters f_i can be obtained. This approach, however, has the following drawbacks:

- The Eqs. (13.3.3) for determining the controller parameters are complicated, in general nonlinear.
- For higher-order systems the computational effort is large.
- There is no systematic way to solve the equations.

The determination of the feedback parameters can be significantly simplified when the invariance of the eigenvalues of a system under a regular transformation is observed and used. The idea is to transform the system into a form which is suitable for the determination of the controller parameters. In the sections below this idea is used.

13.3.1 Design of a system in controller canonical form

A state-space system in controller canonical form has the following structure:

$$\dot{\mathbf{z}}(t) = \mathbf{A}_c \mathbf{z}(t) + \mathbf{b}_c u(t) \quad \mathbf{z}(t_0) \text{ initial condition} \quad (13.3.4)$$

$$y(t) = \mathbf{c}_c^T \mathbf{z}(t) \quad (13.3.5)$$

with

$$\mathbf{A}_c = \begin{bmatrix} 0 & 1 & 0 & \cdot & 0 \\ 0 & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & 1 \\ -a_0 & -a_1 & -a_2 & \cdot & -a_{n-1} \end{bmatrix}, \quad \mathbf{b}_c = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{c}_c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}. \quad (13.3.6)$$

This canonical form has the following properties:

- The characteristic polynomial can be directly determined from the last line of \mathbf{A}_c , which is

$$P_c(s) = \det(s\mathbf{I} - \mathbf{A}_c) = a_0 + a_1 s + a_2 s^2 + \dots + a_{n-1} s^{n-1} + s^n. \quad (13.3.7)$$

- A system of this structure is always controllable as its controllability matrix according to Eq. (12.6.1) has always full rank.
- The transfer function of the system is immediately given by

$$G(s) = \frac{c_1 + c_2 s + c_3 s^2 + \dots + c_n s^{n-1}}{a_0 + a_1 s + a_2 s^2 + \dots + a_{n-1} s^{n-1} + s^n}.$$

The feedback is now defined as

$$u(t) = -\mathbf{f}_c^T \mathbf{z}(t) \quad (13.3.8)$$

with

$$\mathbf{f}_c = \begin{bmatrix} f_{c1} \\ f_{c2} \\ \vdots \\ f_{cn} \end{bmatrix}.$$

For the closed-loop system the system matrix is

$$\mathbf{A}_c - \mathbf{b}_c \mathbf{f}_c^T = \begin{bmatrix} 0 & 1 & 0 & \cdot & 0 \\ 0 & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & 1 \\ -a_0 - f_{c1} & -a_1 - f_{c2} & -a_2 - f_{c3} & \cdot & -a_{n-1} - f_{cn} \end{bmatrix} \quad (13.3.9)$$

and its characteristic polynomial is

$$Q(s) = (a_0 + f_{c1}) + (a_1 + f_{c2})s + (a_2 + f_{c3})s^2 + \dots + (a_{n-1} + f_{cn})s^{n-1} + s^n.$$

Equating this polynomial with the polynomial with the desired poles from Eqs. (13.3.2) one obtains by comparison of the corresponding terms of both polynomials the controller parameters as

$$f_{ci} = p_{i-1} - a_{i-1} \quad \text{for } i = 1, 2, \dots, n. \quad (13.3.10)$$

In the controller canonical form the calculation of the controller feedback parameters is reduced to the calculation of a simple difference between the coefficients of two polynomials.

13.3.1.1 Design of a system not in a canonical form

In general, when a system is not given in the controller canonical form, one has to transform it by a regular transformation

$$\mathbf{z}(t) = \mathbf{T} \mathbf{x}(t), \quad (13.3.11)$$

which brings the system into the desired canonical form according to Eqs. (13.3.4) to (13.3.6). The determination of the controller parameters in \mathbf{f}_c is performed according to Eq. 13.3.10. The feedback law in the original state is, using Eq. (13.3.11) given by

$$u(t) = -(\mathbf{f}_c^T \mathbf{T}) \mathbf{x}(t) = -\mathbf{f}^T \mathbf{x}(t). \quad (13.3.12)$$

Finally, the feedback vector is transformed back to

$$\mathbf{f}^T = \mathbf{f}_c^T \mathbf{T}. \quad (13.3.13)$$

The main task in the pole-placement design for systems that are not in controller canonical form, is the determination of the transformation matrix \mathbf{T} . When the original state equation from Eq. (13.1.1) is transformed by Eq. (13.3.11), one obtains the transformed entities

$$\mathbf{A}_c = \mathbf{T} \mathbf{A} \mathbf{T}^{-1} \quad (13.3.14)$$

and

$$\mathbf{b}_c = \mathbf{T} \mathbf{b}. \quad (13.3.15)$$

Eq. (13.3.14) will be further analysed in the form by right-multiplying with \mathbf{T}

$$\mathbf{A}_c \mathbf{T} = \mathbf{T} \mathbf{A}.$$

With the row vectors \mathbf{t}_i^T of the matrix \mathbf{T} this equation is

$$\begin{bmatrix} 0 & 1 & 0 & \cdot & 0 \\ 0 & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & 1 \\ -a_0 & -a_1 & -a_2 & \cdot & -a_{n-1} \end{bmatrix} \begin{bmatrix} \mathbf{t}_1^T \\ \mathbf{t}_2^T \\ \cdot \\ \mathbf{t}_{n-1}^T \\ \mathbf{t}_n^T \end{bmatrix} = \begin{bmatrix} \mathbf{t}_1^T \\ \mathbf{t}_2^T \\ \cdot \\ \mathbf{t}_{n-1}^T \\ \mathbf{t}_n^T \end{bmatrix} \mathbf{A} \quad (13.3.16)$$

and after multiplication:

$$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \mathbf{t}_2^T & \cdot & \cdot & \cdot \\ \cdot & \mathbf{t}_3^T & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \mathbf{t}_n^T \\ -a_0 \mathbf{t}_1^T - a_1 \mathbf{t}_2^T - \dots - a_{n-2} \mathbf{t}_{n-1}^T - a_{n-1} \mathbf{t}_n^T & \cdot & \cdot & \cdot & \cdot \end{bmatrix} = \begin{bmatrix} \mathbf{t}_1^T \mathbf{A} \\ \mathbf{t}_2^T \mathbf{A} \\ \cdot \\ \mathbf{t}_{n-1}^T \mathbf{A} \\ \mathbf{t}_n^T \mathbf{A} \end{bmatrix}. \quad (13.3.17)$$

From the first $n - 1$ rows of both sides one obtains the recursive relationship

$$\mathbf{t}_{i+1}^T = \mathbf{t}_i^T \mathbf{A} \quad \text{for } i = 1, 2, \dots, n - 1,$$

and when \mathbf{t}_1^T is known the remaining rows of the matrix \mathbf{T} are

$$\mathbf{t}_{i+1}^T = \mathbf{t}_1^T \mathbf{A}^i \quad \text{for } i = 1, 2, \dots, n - 1. \quad (13.3.18)$$

The first row \mathbf{t}_1^T is obtained from Eq. (13.3.15), which is also valid. Using the results from Eq. (13.3.18), the right-hand side of Eq. (13.3.15) has the form:

$$\mathbf{b}_c = \mathbf{T} \mathbf{b} = \begin{bmatrix} \mathbf{t}_1^T \\ \mathbf{t}_2^T \\ \vdots \\ \mathbf{t}_{n-1}^T \\ \mathbf{t}_n^T \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{t}_1^T \mathbf{b} \\ \mathbf{t}_1^T \mathbf{A} \mathbf{b} \\ \vdots \\ \mathbf{t}_1^T \mathbf{A}^{n-2} \mathbf{b} \\ \mathbf{t}_1^T \mathbf{A}^{n-1} \mathbf{b} \end{bmatrix} \quad (13.3.19)$$

or in transposed form:

$$\mathbf{b}_c^T = \mathbf{t}_1^T [\mathbf{b} \quad \mathbf{A} \mathbf{b} \quad \dots \quad \mathbf{A}^{n-2} \mathbf{b} \quad \mathbf{A}^{n-1} \mathbf{b}] = \mathbf{t}_1^T \mathbf{S}. \quad (13.3.20)$$

The matrix \mathbf{S} is the controllability matrix from Eq. (12.6.1), which has full rank if the system is completely controllable. Under this condition one can obtain the first row by

$$\mathbf{t}_1^T = \mathbf{b}_c^T \mathbf{S}^{-1}, \quad (13.3.21)$$

which is the last row of the inverse controllability matrix, because \mathbf{b}_c is the n -th unit vector.

Summarising the procedure of the pole-placement design for a system that is not in a canonical form, the following steps are necessary:

1. Calculation of the controllability matrix \mathbf{S} from Eq. (12.6.1), its inverse and extracting the last row according to Eq. (13.3.21). This is the first row of the transformation matrix \mathbf{T} .
2. Row-wise calculation of the remaining rows of the transformation matrix \mathbf{T} using Eq. (13.3.18).
3. Choice of the n eigenvalues s_i .
4. Calculation of the coefficients p_i of the desired polynomial according to Eqs. (13.3.1) and (13.3.2).
5. Transformation of the system matrix \mathbf{A} according to Eq. (13.3.14) and extracting the coefficients a_i of the characteristic polynomial of the open-loop system from the last row of the transformed matrix.
6. Calculation of the coefficients of the feedback vector \mathbf{f}_c^T according to Eq. (13.3.10).
7. Back transformation of the feedback vector according to Eq. (13.3.13).

13.3.2 Design using Ackermann's formula

The method described in the previous section 13.3.1.1 needs the computation of the full transformation matrix \mathbf{T} . This is not necessary as shown by the simplified procedure of Ackermann.

Steps 5 and 6 from the end of the previous section 13.3.1.1 will be combined by using the relationship for the transformation matrix \mathbf{T} from Eqs. (13.3.13) and (13.3.18):

$$\mathbf{f}^T = [p_0 - a_0, p_1 - a_1, \dots, p_{n-1} - a_{n-1}] \begin{bmatrix} \mathbf{t}_1^T \\ \mathbf{t}_1^T \mathbf{A} \\ \vdots \\ \mathbf{t}_1^T \mathbf{A}^{n-1} \end{bmatrix}.$$

After multiplying, the feedback vector can be represented by two sums as

$$\mathbf{f}^T = \mathbf{t}_1^T \sum_{i=0}^{n-1} p_i \mathbf{A}^i - \mathbf{t}_1^T \sum_{i=0}^{n-1} a_i \mathbf{A}^i.$$

Adding the term $\mathbf{t}_1^T \mathbf{A}^n$ to both sums, one obtains

$$\mathbf{f}^T = \mathbf{t}_1^T P(\mathbf{A}) - \mathbf{t}_1^T P_c(\mathbf{A}). \quad (13.3.22)$$

$P(\mathbf{A})$ is the desired polynomial from Eq. (13.3.2), where the scalar variable s is substituted by the matrix \mathbf{A} . The same holds for the characteristic polynomial $P_c(s)$ of the system matrix \mathbf{A} from Eq. (13.3.7). According to the Cayley-Hamilton theorem a quadratic matrix always satisfies its own characteristic polynomial, i.e. $P_c(\mathbf{A}) = \mathbf{0}$, Eq. (13.3.22) is simplified to

$$\mathbf{f}^T = \mathbf{t}_1^T P(\mathbf{A}) . \quad (13.3.23)$$

As \mathbf{t}_1^T is the last row of the inverse controllability matrix \mathbf{S}^{-1} the calculation of the feedback vector is simpler than in the previous case of section 13.3.1.1. The following steps are necessary:

1. Calculation of the controllability matrix \mathbf{S} from Eq. (12.6.1), its inverse and extracting the last row.
2. Choice of the n eigenvalues s_i .
3. Calculation of the coefficients p_i of the desired polynomial according to Eqs. (13.3.1) and (13.3.2).
4. Calculation of the term $P(\mathbf{A})$.
5. Calculation of the coefficients of the feedback vector \mathbf{f}^T according to Eq. (13.3.23).

13.4 State reconstruction using observers

The state-feedback controllers designed in the previous sections assume that all state variables can be measured. This may not be the case. That is, certain components of the state vector may correspond to inaccessible internal variables, which may not be available for measurement. The theory of observers has been proposed to reconstruct an approximation of the state vector based only on available measurements. Being a dynamical system in itself, the observer is designed to track the n state variables of the original system. The block diagram in Figure 13.4.1 shows the system structure. The observer has two inputs,

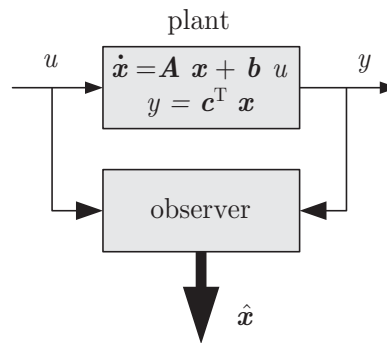


Figure 13.4.1: Principle of an observer

$u(t)$ and $y(t)$, and it provides an estimate $\hat{\mathbf{x}}(t)$ of the plant state $\mathbf{x}(t)$. The estimate should satisfy the condition

$$\lim_{t \rightarrow \infty} \hat{\mathbf{x}}(t) = \lim_{t \rightarrow \infty} \mathbf{x}(t) .$$

Thus, the error between the plant state and the estimate

$$\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t)$$

should converge asymptotically to zero, i.e.

$$\lim_{t \rightarrow \infty} \tilde{\mathbf{x}}(t) = \mathbf{0} . \quad (13.4.1)$$

13.4.1 Structure of an observer

A candidate for a dynamical system in state space form which can fulfil the condition from Eq. (13.4.1), may have the following form:

$$\dot{\hat{\mathbf{x}}}(t) = \hat{\mathbf{A}} \hat{\mathbf{x}}(t) + \hat{\mathbf{b}} u(t) + \hat{\mathbf{g}} y(t) \quad \hat{\mathbf{x}}(t_0) \text{ initial condition} . \quad (13.4.2)$$

In order to show this, Eq. (13.4.2) is subtracted from Eq. (13.1.1) and using Eq. (13.1.2) the state equation

$$\dot{\tilde{\mathbf{x}}}(t) = [\mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T] \tilde{\mathbf{x}}(t) - \hat{\mathbf{A}} \hat{\mathbf{x}}(t) + [\mathbf{b} - \hat{\mathbf{b}}] u(t) \quad \tilde{\mathbf{x}}(t_0) \text{ initial condition} \quad (13.4.3)$$

is obtained. If the observing error $\tilde{\mathbf{x}}$ is to be independent of the plant input $u(t)$, one should make the choice

$$\hat{\mathbf{b}} = \mathbf{b}, \quad (13.4.4)$$

and to have a homogeneous error state equation, one has to set the observer system matrix to

$$\hat{\mathbf{A}} = \mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T . \quad (13.4.5)$$

Then the error state equation is

$$\dot{\tilde{\mathbf{x}}}(t) = [\mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T] \tilde{\mathbf{x}}(t) \quad \tilde{\mathbf{x}}(t_0) \text{ initial condition} . \quad (13.4.6)$$

The parameter $\hat{\mathbf{g}}$ is the only design parameter of the observer, which has to be specified so that the error converges asymptotically to zero, independent of the unknown initial error $\tilde{\mathbf{x}}(t_0)$. This is only the case if all eigenvalues of the matrix $(\mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T)$ have negative real parts.

In order to show how an observer works and to gain the interpretation of the observer equations, $\hat{\mathbf{A}}$ in Eq. (13.4.2) is substituted by Eq. (13.4.5) and introducing the output error term

$$\tilde{y}(t) = y(t) - \mathbf{c}^T \hat{\mathbf{x}}(t) ,$$

the observer state equation is

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{A} \hat{\mathbf{x}}(t) + \mathbf{b} u(t) + \hat{\mathbf{g}} \tilde{y}(t) \quad \hat{\mathbf{x}}(t_0) \text{ initial condition} .$$

From this equation one can see, that the observer consists of a copy of the plant model and of a feedback of the error between the plant model output and the estimated plant output using the state estimates from the observer. The feedback gains in $\hat{\mathbf{g}}$ are the design parameters of the observer.

13.4.2 Design of observers

The design parameter $\hat{\mathbf{g}}$ has to be specified so that the matrix

$$\hat{\mathbf{A}} = \mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T$$

has n given eigenvalues \hat{s}_i in the left-half s plane. This means that the characteristic polynomial is given by

$$\det [s \mathbf{I} - (\mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T)] = \prod_{i=1}^n (s - \hat{s}_i) .$$

The determination of the design parameter is solved in an elegant way by observing that the characteristic polynomial of a matrix is invariant to transposition. Therefore, the above equations can be rewritten with transposed matrices as:

$$\hat{\mathbf{A}}^T = \mathbf{A}^T - \mathbf{c} \hat{\mathbf{g}}^T$$

$$\det [s \mathbf{I} - (\mathbf{A}^T - \mathbf{c} \hat{\mathbf{g}}^T)] = \prod_{i=1}^n (s - \hat{s}_i) .$$

A comparison of these equations with those of the state-feedback controller do not show structural differences. Instead of \mathbf{A} now \mathbf{A}^T , instead of \mathbf{b} now \mathbf{c} and instead of \mathbf{f} now $\hat{\mathbf{g}}$ is used. Therefore, an observer design is the same task as designing the feedback

$$u_f(t) = -\mathbf{g}^T \boldsymbol{\xi}(t)$$

for the system

$$\dot{\boldsymbol{\xi}}(t) = \mathbf{A}^T \boldsymbol{\xi}(t) + \mathbf{c} y(t) .$$

Such a system is called a *dual system*. For this dual system the same design procedure as for the state-feedback controller can be applied, i.e. the design of a state-feedback controller of a dual system is the same as the design of an observer for the original system. Therefore, designing an observer, one takes the dual system and designs a state-feedback controller according to section 13.3. The resulting feedback vector is then the required parameter vector $\hat{\mathbf{g}}$ of the observer.

13.5 Combined observer-controllers

One is led to the observer problem because of the need to obtain the states for use in the controller. Now, only the asymptotically correct estimates $\hat{\mathbf{x}}(t)$ of the states $\mathbf{x}(t)$ rather than the states themselves are available. A natural question is whether the previous result on pole placement via state feedback will continue to hold when these estimates of the actual states are available. There is no other option available than to see what happens with the observer and controller.

In the steady state there should clearly be no loss in using the asymptotic observer, since the error in the estimates will be zero. The question for the present scheme is whether the incorporation of the observer dynamical system into the feedback loop will affect the stability of the overall system. That is, interconnections of stable subsystems may lead to unstable overall systems.

The following analysis on stability is performed by setting up the joint observer-controller system of Figure 13.5.1. For this scheme the equations for the overall system can be written down by inspection as

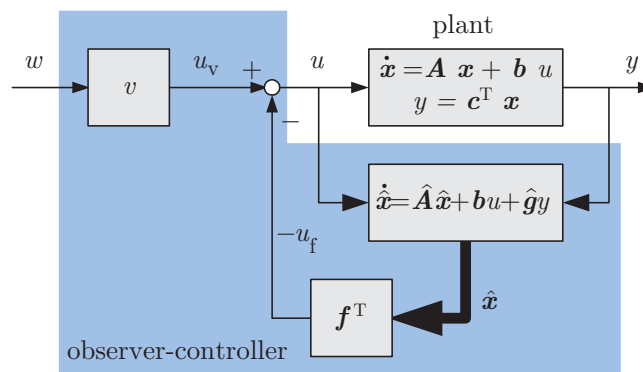


Figure 13.5.1: Basic structure of a combined observer-controller

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A} \mathbf{x}(t) - \mathbf{b} \mathbf{f}^T \hat{\mathbf{x}}(t) + \mathbf{b} v u(t) & \mathbf{x}(t_0) \text{ initial condition} \\ \dot{\hat{\mathbf{x}}}(t) &= \hat{\mathbf{g}} \mathbf{c}^T \mathbf{x}(t) + [\mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T - \mathbf{b} \mathbf{f}^T] \hat{\mathbf{x}}(t) + \mathbf{b} v u(t) & \hat{\mathbf{x}}(t_0) \text{ initial condition} . \end{aligned}$$

To study the overall system in terms of the state $\mathbf{x}(t)$ and its error $\tilde{\mathbf{x}}(t)$ the second equation is subtracted from the first, and one obtains

$$\begin{aligned} \dot{\mathbf{x}}(t) &= [\mathbf{A} - \mathbf{b} \mathbf{f}^T] \mathbf{x}(t) + \mathbf{b} \mathbf{f}^T \tilde{\mathbf{x}} + \mathbf{b} v u(t) & \mathbf{x}(t_0) \text{ initial condition} \\ \dot{\tilde{\mathbf{x}}}(t) &= [\mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T] \tilde{\mathbf{x}}(t) & \tilde{\mathbf{x}}(t_0) \text{ initial condition} \end{aligned}$$

which in matrix form can be written

$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\tilde{\mathbf{x}}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A} - \mathbf{b} \mathbf{f}^T & \mathbf{b} \mathbf{f}^T \\ \mathbf{0} & \mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \tilde{\mathbf{x}}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{b} v \\ \mathbf{0} \end{bmatrix} w(t) \quad \begin{bmatrix} \mathbf{x}(t_0) \\ \tilde{\mathbf{x}}(t_0) \end{bmatrix} \text{ initial condition .}$$

As the system matrix of the above overall system has a triangular block form, the characteristic polynomial of the overall system

$$\det \begin{bmatrix} s \mathbf{I} - (\mathbf{A} - \mathbf{b} \mathbf{f}^T) & -\mathbf{b} \mathbf{f}^T \\ \mathbf{0} & s \mathbf{I} - (\mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T) \end{bmatrix}$$

is just the product of the characteristic polynomial of the observer and the characteristic polynomial of the controlled system

$$\det [s \mathbf{I} - (\mathbf{A} - \mathbf{b} \mathbf{f}^T)] \det [s \mathbf{I} - (\mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T)]$$

assuming perfect knowledge of the states.

This is nice, because it means that the natural frequencies or modes of the overall system can always be arranged to be stable. In fact, they can be chosen completely arbitrarily.

Another useful consequence is that the controller and observer can be designed independently of each other. Whether the true states are available, or only asymptotically correct estimates of the states, is immaterial to the calculation of the feedback vector \mathbf{f} . Similarly, the dynamics of the observer can be calculated from knowledge of \mathbf{A} and \mathbf{c} without consideration of whether the observer is to be combined with a feedback controller or not. This is the so-called *separation property* of the observer-controller design procedure.

13.6 Example of a state-feedback control system

For a 2nd-order plant, described by

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 8 \\ 8 \end{bmatrix} u(t) + \begin{bmatrix} 8 \\ 8 \end{bmatrix} z(t)$$

and

$$y(t) = \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

with the disturbance

$$z(t) = \sigma(t - 6)$$

and the initial condition

$$\mathbf{x}(0) = \begin{bmatrix} 5 \\ -5 \end{bmatrix}$$

a state-feedback controller with and without observer is to be designed so that the poles of the closed-loop system are $s_1 = s_2 = -3$. The eigenvalues of the observer are given as $\hat{s}_1 = \hat{s}_2 = -8$.

The plant is unstable and has the eigenvalues at 1.0 and -0.5. The desired characteristic polynomial with the zeros $s_1 = s_2 = -3$ as the given poles is from Eq. (13.3.1)

$$P(s) = (s + 3)(s + 3) = 9 + 6s + s^2 .$$

The controllability matrix is

$$\mathbf{S} = [\mathbf{b} \quad \mathbf{A} \mathbf{b}] = \begin{bmatrix} 8 & 8 \\ 8 & -4 \end{bmatrix}$$

and its inverse

$$\mathbf{S}^{-1} = \begin{bmatrix} 0.0417 & 0.0833 \\ 0.0833 & -0.0833 \end{bmatrix} .$$

The last row of the inverse is

$$\mathbf{t}_1^T = [0.0833 \quad -0.0833] .$$

With this data, the feedback vector using Ackermann's formula Eq. (13.3.23) is

$$\begin{aligned} \mathbf{f}^T &= [0.0833 \quad -0.0833] \left\{ 9 \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} + 6 \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} \right\} \\ &= [0.0833 \quad -0.0833] \begin{bmatrix} 16 & 0 \\ 0 & 6.25 \end{bmatrix} = [1.3333 \quad -0.5208] . \end{aligned}$$

The feedforward gain according to Eq. (13.1.19) is

$$\begin{aligned} v &= \frac{-1}{[2 \quad -1] \left(\begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} - \begin{bmatrix} 8 \\ 8 \end{bmatrix} [1.3333 \quad -0.5208] \right)^{-1} \begin{bmatrix} 8 \\ 8 \end{bmatrix}} \\ &= \frac{-1}{[2 \quad -1] \left(\begin{bmatrix} -9.6667 & 4.1667 \\ -10.6667 & 3.6667 \end{bmatrix} \right)^{-1} \begin{bmatrix} 8 \\ 8 \end{bmatrix}} \\ &= \frac{-1}{[2 \quad -1] \begin{bmatrix} 0.4074 & -0.4630 \\ 1.1852 & -1.0741 \end{bmatrix} \begin{bmatrix} 8 \\ 8 \end{bmatrix}} \\ &= 0.5625 . \end{aligned}$$

For the observer design the dual system

$$\begin{bmatrix} \dot{\xi}_1(t) \\ \dot{\xi}_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \end{bmatrix} y(t)$$

is used. The desired characteristic polynomial with the zeros $\hat{s}_1 = \hat{s}_2 = -8$ as the given eigenvalues of the observer is from Eq. (13.3.1)

$$P(s) = (s + 8)(s + 8) = 64 + 16s + s^2 .$$

The controllability matrix of the dual system is

$$\mathbf{S} = [\mathbf{c} \quad \mathbf{A}^T \mathbf{c}] = \begin{bmatrix} 2 & 2 \\ -1 & 0.5 \end{bmatrix}$$

and its inverse

$$\mathbf{S}^{-1} = \begin{bmatrix} 0.1667 & -0.6667 \\ 0.3333 & 0.6667 \end{bmatrix} .$$

The last row of the inverse is

$$\mathbf{t}_1^T = [0.3333 \quad 0.6667] .$$

With this data, the feedback vector using Ackermann's formula Eq. (13.3.23) is

$$\begin{aligned} \hat{\mathbf{g}}^T &= [0.3333 \quad 0.6667] \left\{ 64 \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} + 16 \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} \right\} \\ &= [0.3333 \quad 0.6667] \begin{bmatrix} 81 & 0 \\ 0 & 56.25 \end{bmatrix} = [27 \quad 37.5] . \end{aligned}$$

The observer system matrix is

$$\hat{\mathbf{A}} = \mathbf{A} - \hat{\mathbf{g}} \mathbf{c}^T = \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} - \begin{bmatrix} 27 \\ 37.5 \end{bmatrix} [2 \quad -1] = \begin{bmatrix} -53 & 27 \\ -75 & 37 \end{bmatrix}$$

and, finally, the following observer equation is obtained:

$$\begin{bmatrix} \dot{\hat{x}}_1(t) \\ \dot{\hat{x}}_2(t) \end{bmatrix} = \begin{bmatrix} -53 & 27 \\ -75 & 37 \end{bmatrix} \begin{bmatrix} \hat{x}_1(t) \\ \hat{x}_2(t) \end{bmatrix} + \begin{bmatrix} 8 \\ 8 \end{bmatrix} u(t) + \begin{bmatrix} 27 \\ 37.5 \end{bmatrix} y(t) .$$

Figure 13.6.1 shows the time responses of the state-feedback control system with and without observer. The responses are divided into three periods. The control system is started at $t = 0$ s and until $t = 3$ s

the decay of the observing error is demonstrated. At $t = 3$ s one can see the behaviour of the system following the change in the set point $w(t)$. Then from $t = 6$ s the disturbance $z(t)$ is active and one can see the disturbance behaviour. The observer is started at $t = 0$ s with zero initial conditions, whereas the plant state is not zero. The estimated states in Figure 13.6.1c and 13.6.1d converge asymptotically to the real states and the value of $w(t) = 1$ is reached by the controlled variable $y(t)$ as shown in Figure 13.6.1b. The control system follows the set-point change from 1 to 4 applied at $t = 3$ s. However, due to the proportional behaviour of the open-loop system the controlled variable $y(t)$ shows a large steady-state error after the disturbance is applied at $t = 6$ s.

The control structure used does not show the desired static behaviour with a vanishing control error under disturbances. Therefore, a state-feedback controller with an integrator is required to cope with the disturbance problem. For the modified control structure according to Figure 13.2.1 one has to design a state-feedback controller for the extended system according to Eqs. (13.2.5) and (13.2.6):

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{\epsilon}(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.5 & 0 \\ 2 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \epsilon(t) \end{bmatrix} + \begin{bmatrix} 8 \\ 8 \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} w(t) .$$

The initial condition is now

$$\mathbf{x}(0) = \begin{bmatrix} 5 \\ -5 \\ 0 \end{bmatrix} .$$

For the eigenvalues of the closed-loop system a third value must be given, as the additional state $\epsilon(t)$ of the integrator is introduced. For simplicity the choice $s_1 = s_2 = s_3 = -3$ is taken and the desired characteristic polynomial is

$$P(s) = (s + 3)(s + 3)(s + 3) = 27 + 27s + 9s^2 + s^3 .$$

The controllability matrix is

$$\mathbf{S} = [\mathbf{b}^* \quad \mathbf{A}^* \mathbf{b}^* \quad \mathbf{A}^{*2} \mathbf{b}^*] = \begin{bmatrix} 8 & 8 & 8 \\ 8 & -4 & 2 \\ 0 & 8 & 20 \end{bmatrix}$$

and its inverse

$$\mathbf{S}^{-1} = \begin{bmatrix} 0.0625 & 0.0625 & -0.0313 \\ 0.1042 & -0.1042 & -0.0313 \\ -0.0417 & 0.0417 & 0.0625 \end{bmatrix} .$$

The last row of the inverse is

$$\mathbf{t}_1^T = [-0.0417 \quad 0.0417 \quad 0.0625] .$$

With this data, the feedback vector using Ackermann's formula Eq. (13.3.23) is

$$\begin{aligned} \mathbf{f}^{*T} &= [-0.0417 \quad 0.0417 \quad 0.0625] \begin{bmatrix} 64 & 0 & 0 \\ 0 & 15.625 & 0 \\ 74 & -22.75 & 27 \end{bmatrix} \\ &= [1.9583 \quad -0.7708 \quad 1.6875] . \end{aligned}$$

From this the feedback vector of the state-feedback is

$$\mathbf{f}^T = [1.9583 \quad -0.7708]$$

and the feedforward gain

$$v = 1.6875 .$$

As the state of the integrator is known, the same observer can be used as in the case without integrator.

Figure 13.6.2 shows the time responses of the same experiment as in Figure 13.6.1. The responses in the first and second period are similar, the amplitudes are somewhat larger and due to the additional integrator the response on set-point changes is slower, but there is no steady-state error on disturbances. This is for the case with and without observer, though the observer shows a steady-state error when the disturbance $z(t)$ is active.

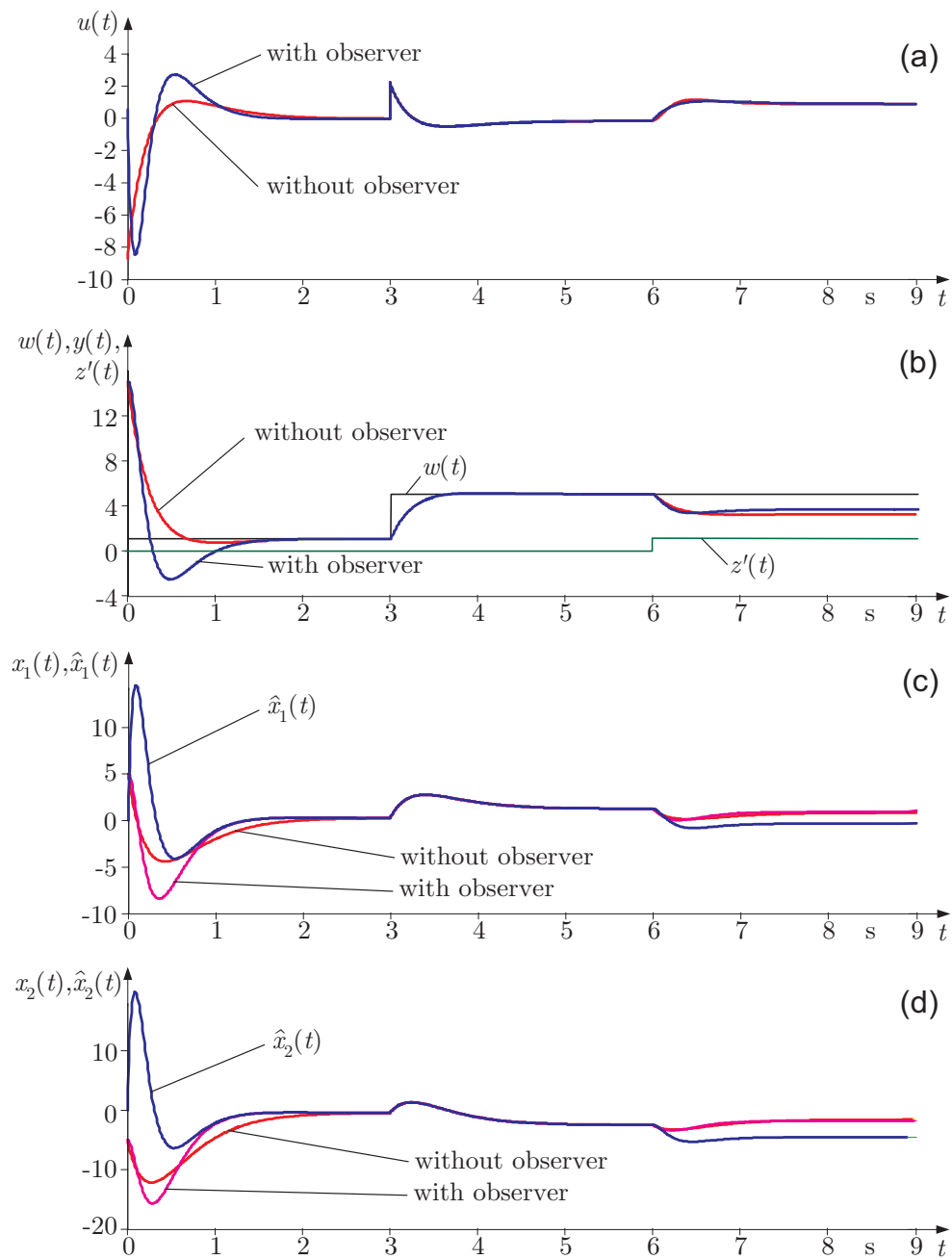


Figure 13.6.1: Time responses of the state-feedback control system without integrator, (a) manipulated variable $u(t)$, (b) controlled value $y(t)$, set point $w(t)$ and disturbance $z'(t)$, (c) state $x_1(t)$ and its estimate $\hat{x}_1(t)$, (d) state $x_2(t)$ and its estimate $\hat{x}_2(t)$

Demonstration Example 13.1

A virtual experiment using state-feedback control of a tank system

Demonstration Example 13.2

A virtual experiment using state-feedback control of a VTOL

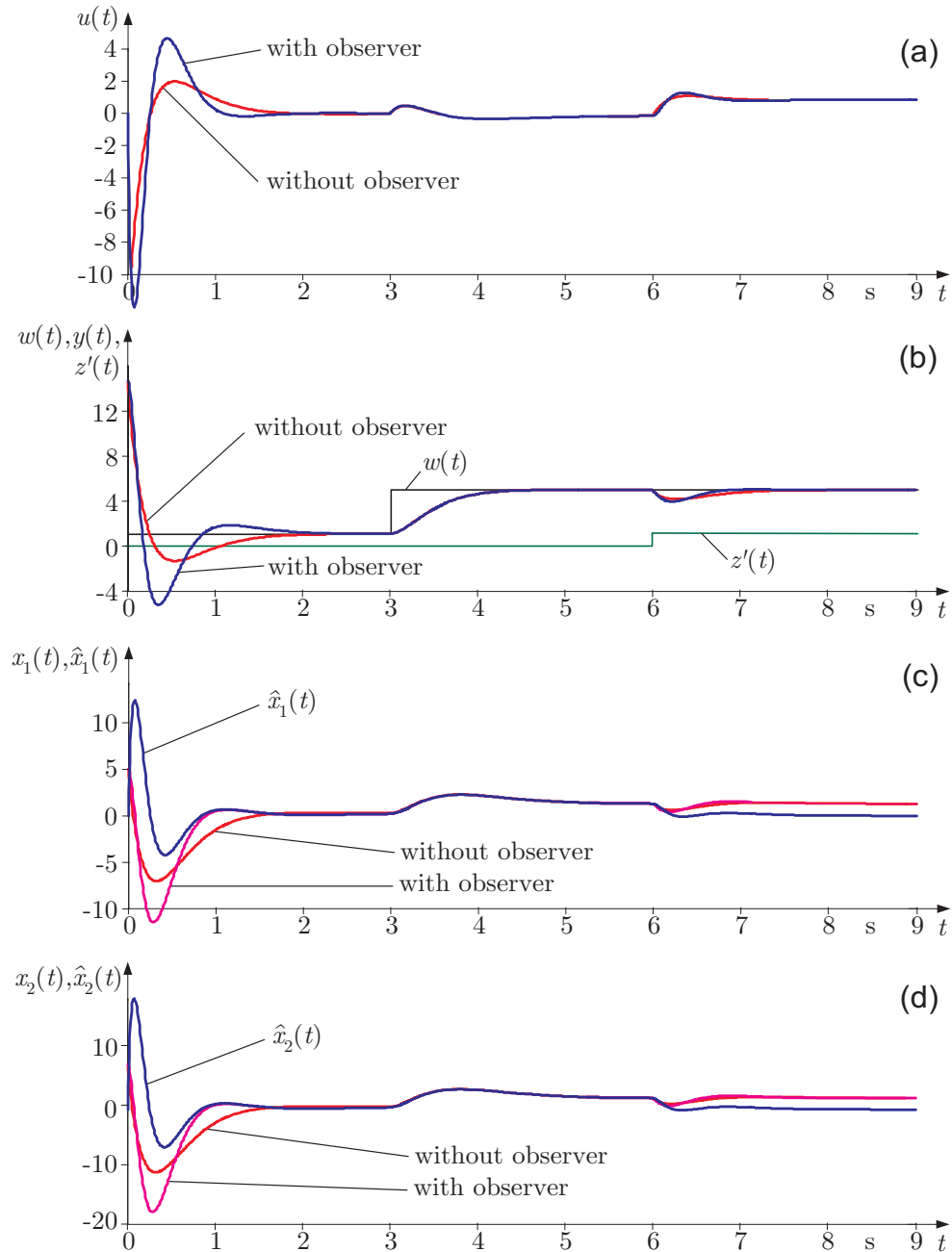


Figure 13.6.2: Time responses of the state-feedback control system with integrator, (a) manipulated variable $u(t)$, (b) controlled value $y(t)$, set point $w(t)$ and disturbance $z'(t)$, (c) state $x_1(t)$ and its estimate $\hat{x}_1(t)$, (d) state $x_2(t)$ and its estimate $\hat{x}_2(t)$

Module 14

Introduction to fuzzy techniques

Module units

14.1 Crisp and fuzzy logic	14-1
14.2 Why use fuzzy logic for control ?	14-3
14.3 Ideas of the fuzzy control methodology	14-4

Module overview. *This is an introductory module to explain what the term fuzzy is and why fuzzy logic is used to solve control problems. First, the difference between the crisp binary valued logic and the fuzzy logic is explained by introducing the grade of truth. This is deepened by considering control system design with unknown information and with descriptions using rules and linguistic terms. The main ideas of the fuzzy control methodology is sketched and summarised in five steps, which are followed in the other modules.*

Module objectives. *When you have completed this module you should be able to:*

1. Understand the basic ideas of fuzzy logic.
2. Understand the main objectives of control system design using fuzzy logic.
3. Understand the basic ideas of fuzzy control.

Module prerequisites. *Boolean algebra.*

14.1 Crisp and fuzzy logic

Fuzzy logic forms a bridge between the two areas of qualitative and quantitative modelling. Although the input-output mapping of such a model is integrated into a system as a quantitative map, internally it can be considered as a set of qualitative linguistic rules. Since the pioneering work of Zadeh in 1965 and Mamdani in 1975, the models formed by fuzzy logic have been applied to many varied types of information processing including control systems.

The term *Fuzzy Logic* is a misnomer. It implies that in some way the methodology is vague or ill-defined. This is in fact far from the case. Fuzzy logic just evolved from the need to model the type of of vague or ill-defined systems that are difficult to handle using conventional binary valued logic, but the methodology itself is based on mathematical theory.

We are all familiar with binary valued logic and set theory. An element belongs to a set of all possible elements and given any specific subset, it can be said accurately, whether that element is or is not a member of it. For example, a person belongs to the set of all human beings, and given a specific subset,

such as all males, one can say whether or not each particular person (element) belongs to this set. This is appealing since it seems to describe the way human reason. Collecting many elements into sets allows to describe many occurrences with few rules. For example, the simple statement

IF person is male AND a parent THEN person is a father

applies to many people across the world with complete precision. The rules are formed using operators. Here, the intersection operator AND is used, which manipulates the sets.

Unfortunately, not everything can be described using binary valued sets. The classifications of persons into males and females is easy, but it is problematic to classify them as being tall or not tall. The set of tall people is far more difficult to define, because there is no distinct cut-off point at which tall begins. This is not a measurement problem, and measuring the height of all elements more precisely is not helpful. Such a problem is often distorted so that it can be described using the well-known existing methodology. Here, one could simply select a height, e.g. 1.80 m, at which the set tall begins, see Figure 14.1.1a. The output of a reasoning system using this definition would not be smooth with respect to the height of a person. A person of height 1.79 m would produce a different output than a person of 1.81 m. In human reasoning this property is not observed and it is also undesirable for reasoning systems that are part of a control system.

Fuzzy logic was suggested by Zadeh as a method for mimicking the ability of human reasoning using a small number of rules and still producing a smooth output via a process of interpolation. It forms rules that are based upon multi-valued logic and so introduced the concept of set membership. With fuzzy logic an element could partially belong to a set and this is represented by the set membership. For example, a person of height 1.79 m would belong to both tall and not tall sets with a particular degree of membership. As the height of a person increases the membership grade within the tall set would increase whilst the membership grade within the not tall set would decrease, see Figure 14.1.1b. The output of a fuzzy reasoning system would produce similar results for similar inputs. Fuzzy logic is simply the extension of conventional logic to the case where partial set membership can exist, rule conditions can be satisfied partially and system outputs are calculated by interpolation and, therefore, have output smoothness over the equivalent binary-valued rule base. This property is particularly relevant to control systems.

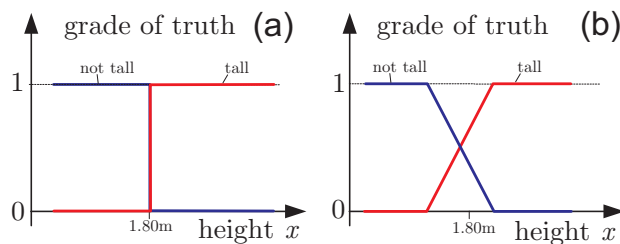


Figure 14.1.1: The difference between the grade of truth in (a) binary valued logic $\{0, 1\}$ and (b) fuzzy logic $[0, 1]$

A fuzzy logic control system is one that has at least one system component that uses fuzzy logic for its internal knowledge representation. Although it is possible for fuzzy systems to communicate information using fuzzy sets, most applications have a single fuzzy system component communicating with conventional system components via deterministic values. In this case, and also in this chapter, fuzzy logic is used purely for internal knowledge representation and, externally, can be considered as any other system component.

Demonstration Example 14.1

Tumbler full or empty ?

Demonstration Example 14.2

The fuzzy set of full and empty tumblers

14.2 Why use fuzzy logic for control ?

Controlling a system means that some characteristics of this system are monitored, and, depending on the values of these characteristics, different controls are applied. An algorithm that transforms sensor inputs into corresponding control values is called a control strategy. The previous chapters deal with the traditional approach of control systems design that consists of the following:

- First, one tries to describe the behaviour of the system in precise mathematical terms, i.e., one comes up with the exact model of the system.
- Second, one tries to describe in precise terms what one wants to achieve. One wants the control that is the best in the sense of some criterion.
- Now that the controlled system is described in precise mathematical terms, and the objective function is described in the same manner, it can be determined for each control strategy and for each initial state how exactly the system will change and what the resulting value of the control will be. The main goal is then to find the control strategy for which the resulting value of the objective function is the largest possible one. This is a well-defined mathematical optimisation problem, and traditional control theory has developed many methods for solving this problem and designing the corresponding control strategies.

Traditional control theory has many important applications. There are, however, practical cases when this theory is not applicable. Indeed, to apply the traditional control theory, one must

- know the model of the controlled system,
- know the objective function formulated in precise terms, and
- be able to solve the corresponding mathematical design problem.

If one of these conditions is not satisfied, then traditional control methodology is not applicable, as in the following cases:

- Sometimes, the model and the objective function is known, but the design problem cannot be solved. This is when the design problem is very complicated, time consuming or when the problem is new and algorithms for solving it have not yet been developed. For example, parking a car is an example of a problem that traditional control theory has not considered until recently.
- Sometimes, the model is known, but the objective function is unknown. For example, if a control system for a vehicle is designed, the intended goal is to make the ride most comfortable, but there is no well-accepted formalism of what comfortable means.
- Sometimes, one does not even know the model of the controlled system. In many practical applications one can in principle measure all the possible variables and determine the model exactly, but this will increase the cost drastically. In other practical situations, the main goal of the controlled system is to explore the unknown, e.g., to control a rover over a terrain of unknown type, or to control surgery instruments. In such situations, the entire objective of the control is to learn as much about the system, and one cannot have a precise model of this system before the control is over.

If traditional control methodology cannot be applied, how can one control? Often, there is an additional expert knowledge available, for example, expert operators who successfully control the desired system. Expert operators know how to operate a plant. Therefore it is desirable to extract the control rules from the expert and use this knowledge in an automatic control system. At first glance, the problem seems very simple. Since the person is a real expert, one simply ask her multiple questions like “suppose that x_1 is equal to 1.2, x_2 is equal to -2.7, ..., what is u ?” After asking all these questions, one will get many pattern, from which one will be able to extrapolate the function $f(x_1, \dots, x_2)$ using one of the known methods. Alas, there are two problems with this idea:

- There is a computational problem. Since one needs to ask a question for each combination of sensor readings, one may end up having to ask too many questions that takes years.
- There is a more serious problem that makes it in most cases impossible to implement. If one asks a car driver a question like “you are driving at 80 km/h when a car which is 20 m in front of you slows down to 50 km/h, for how many seconds do you hit the brakes?”, nobody will give a precise number.

An expert cannot usually express his knowledge in precise numerical terms, like “hit the brakes for 1.27 s”, but he can formulate his knowledge by using words from natural language. The knowledge, which one can extract from an expert consists of statements like “if the velocity is a little bit smaller than maximum, hit the breaks for a while”.

For the fuzzy control methodology one has to

- know the expert’s control rules formulated by words from natural language and
- one wants to produce a precise control strategy.

The methodology that transform the informal expert control rules into a precise control strategy is called *fuzzy control*. The idea was first proposed by Zadeh, and the methodology itself was first proposed and applied by Mamdani. In this chapter it is described exactly how this transformation is done.

14.3 Ideas of the fuzzy control methodology

Before coming to the details, the main ideas of fuzzy control methodology on the example of a situation will be illustrated in which everyone feels himself an expert. One of the most widely used control systems is the simplest rule-based system imaginable, a thermostatic temperature controller. This rule-based system operates with two rules:

- (1) *IF temperature is below set point THEN heat is on*
- (2) *IF temperature is above set point THEN heat is off*

The success of this controller is due to the combination of the properties, that it is simple, robust and does not require a complex process model. The model is: when the heat is on, the temperature rises slowly, and when the heat is off, the temperature falls slowly.

The two IF-THEN clauses above can also be formally rewritten as

$$\text{IF } x \text{ is } A_r \text{ THEN } u \text{ is } B_r \quad \text{for } r = 1, 2$$

In the thermostat example there is only one input variable (linguistic variable), the temperature x . In the general case, there are several input variables x_1, \dots, x_n , so, in addition to the logical connective IF-THEN, another logical connective is needed, AND. Then the IF-THEN clauses are

$$\text{IF } x_1 \text{ is } A_{r1} \text{ AND } x_2 \text{ is } A_{r2} \dots \text{ AND } x_n \text{ is } A_{rn} \text{ THEN } u \text{ is } B_r .$$

Here, $r = 1, \dots, R$ is the rule number, and A_{ri} and B_r are words from natural language (linguistic terms), like “below set point”, “on”, “small”, “large”, “approximately 1.5”, etc. If the standard mathematical notation for IF-THEN and AND is used, the above rules can be re-formulated as follows:

$$A_{r1}(x_1) \wedge A_{r2}(x_2) \dots \wedge A_{rn}(x_n) \rightarrow B_r(u) , \quad (14.3.1)$$

where $r = 1, \dots, R$. The set of rules is usually called a *rule base*. The left-hand side of a rule is called *premise*, the right-hand side *conclusion* and the rule itself *implication*.

The general idea is to represent the rule base in a computer. It has a clear structure. A rule base consists of rules and each rule, in its turn, is obtained from properties expressed by linguistic variables and terms

and using logical connectives. In view of this structure, it is reasonable to represent the rule base by first representing the basic elements of the rule base, premises and conclusions, and then by extending this representation to the rule base as a whole. It makes sense to use the following steps in the methodology:

1. Representation of the basic properties $A_{r_i}(x_i)$ and $B_r(u)$.
2. Representation of the logical connectives.
3. The representations of the basic properties and of the logical connectives is used to get the representations of all the rules.
4. Combination of the representations of different rules into a representation of a rule base.

As a result of these four steps, one obtains an expert system that can give advice for a specialist, who has to make a decision. In control, a system to automatically make a decision based on its own conclusions is wanted. Therefore, for control situations, a fifth follow-up step is needed:

5. Based on facts for x_i and on the rule base a reasoning procedure makes a decision.

In the following section from the very beginning it is described how these five steps are implemented.

Module 15

Basics of fuzzy sets

Module units

15.1 Fuzzy sets	15-1
15.2 Membership functions	15-2
15.3 Elementary operators for fuzzy sets	15-5
15.4 Fuzzy relations	15-6
15.5 Fuzzy composition	15-8

Module overview. *Fuzzy set theory is an extension of classical set theory. Under this aspect the term of the membership function is introduced. The operations with fuzzy sets are introduced and explained and illustrated using membership functions. Starting with elementary operators for fuzzy sets, more complicated operators for modelling of rule bases and reasoning processes are introduced. Interactive demonstration examples round off the module.*

Module objectives. *When you have completed this module you should be able to:*

1. Describe fuzzy sets using membership functions.
2. Operate with fuzzy sets.
3. Model rules using fuzzy sets.
4. Reason using a rule and a fact.

Module prerequisites. *Basics about fuzzy logic.*

15.1 Fuzzy sets

In the classical set theory a set can be represented by enumerating all its elements using

$$\mathcal{A} = \{a_1, a_2, a_3, \dots, a_n\} .$$

If these elements $a_i (i = 1, \dots, n)$ of \mathcal{A} are together a subset of the universal base set \mathcal{X} , the set \mathcal{A} can be represented for all elements $x \in \mathcal{X}$ by its characteristic function

$$\mu_{\mathcal{A}}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{A} \\ 0 & \text{otherwise} . \end{cases} \quad (15.1.1)$$

In classical set theory $\mu_{\mathcal{A}}(x)$ has only the values 0 (“false”) and 1 (“true”), so two values of truth. Such sets are also called *crisp sets*.

Non-crisp sets are called *fuzzy sets*, for which also a characteristic function can be defined. This function is a generalisation of that in Eq.(15.1.1) and called a *membership function*. The membership of a fuzzy set is described by this membership function $\mu_A(x)$ of \mathcal{A} , which associates to each element $x_0 \in \mathcal{X}$ a grade of membership $\mu_A(x_0)$. In contrast to classical set theory a membership function $\mu_A(x)$ of a fuzzy set can have in the normalised closed interval $[0, 1]$ an arbitrary grade of truth, see introductory section 14 and Figure 14.1.1. Therefore, each membership function maps elements of a given universal base set \mathcal{X} , which is itself a crisp set, into real numbers in $[0, 1]$. The notation for the membership function $\mu_A(x)$ of a fuzzy set \mathcal{A}

$$\mathcal{A} : \mathcal{X} \rightarrow [0, 1] \quad (15.1.2)$$

is used. Each fuzzy set is completely and uniquely defined by one particular membership function. Consequently symbols of membership functions are also used as labels of the associated fuzzy sets. That is, each fuzzy set and the associated membership function are denoted by the same capital letter. Since crisp sets and the associated characteristic functions may be viewed, respectively, as special cases of fuzzy sets and membership functions, the same notation is used for crisp sets as well, see Figure 15.1.1

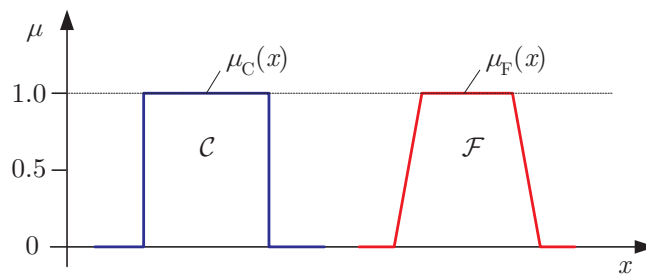


Figure 15.1.1: Membership functions of a crisp set \mathcal{C} and a fuzzy set \mathcal{F}

The base set \mathcal{X} is introduced first above as a universal set. In practical applications, physical or similar quantities are considered that are defined in some interval. When such quantities are described by sets, a base sets can be generalised seamlessly to a crisp base set \mathcal{X} that exists in a defined interval. This is a generalisation of fuzzy sets.

Base sets are not always crisp sets. Another generalisation is that the base set is itself a fuzzy set. This is necessary for multi-dimensional fuzzy sets, which are discussed later in this chapter.

15.2 Membership functions

The membership function $\mu_A(x)$ describes the membership of the elements x of the base set \mathcal{X} in the fuzzy set \mathcal{A} , whereby for $\mu_A(x)$ a large class of functions can be taken. Reasonable functions are often piecewise linear functions, such as triangular or trapezoidal functions.

The grade of membership $\mu_A(x_0)$ of a membership function $\mu_A(x)$ describes for the special element $x = x_0$, to which grade it belongs to the fuzzy set \mathcal{A} . This value is in the unit interval $[0, 1]$. Of course, x_0 can simultaneously belong to another fuzzy set \mathcal{B} , such that $\mu_B(x_0)$ characterises the grade of membership of x_0 to \mathcal{B} . This case is shown in Figure 15.2.1.

Demonstration Example 15.1
Colour as a fuzzy set

Interactive Questions 15.1
[Test yourself here](#)

In the following, a set of important properties and characteristics of fuzzy sets will be described.

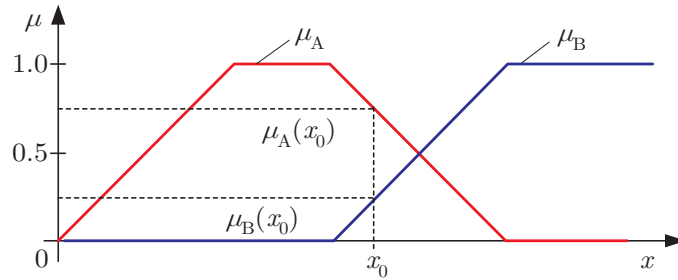


Figure 15.2.1: Membership grades of x_0 in the sets \mathcal{A} and \mathcal{B} : $\mu_{\mathcal{A}}(x_0) = 0.75$ and $\mu_{\mathcal{B}}(x_0) = 0.25$

- Having two fuzzy sets \mathcal{A} and \mathcal{B} based on \mathcal{X} , then both are *equal* if their membership functions are equal, i.e.

$$\mathcal{A} = \mathcal{B} \Leftrightarrow \mu_{\mathcal{A}}(x) = \mu_{\mathcal{B}}(x), \quad x \in \mathcal{X} . \tag{15.2.1}$$

- The *universal set* \mathcal{U} is defined as

$$\mu_{\mathcal{U}}(x) = 1, \quad x \in \mathcal{X} . \tag{15.2.2}$$

- The *height* of a fuzzy set \mathcal{A} is the largest membership grade obtained by any element in that set, i.e.

$$\text{hgt}(\mathcal{A}) = \sup_{x \in \mathcal{X}} \mu_{\mathcal{A}}(x) . \tag{15.2.3}$$

- A fuzzy set \mathcal{A} is called *normal* when $\text{hgt}(\mathcal{A}) = 1$, and it is *subnormal* when $\text{hgt}(\mathcal{A}) < 1$.
- The *support* of a fuzzy set \mathcal{A} is the crisp set that contains all the elements of \mathcal{X} that have nonzero membership grades in \mathcal{A} , i.e.

$$\text{supp}(\mathcal{A}) = \{x \in \mathcal{X} \mid \mu_{\mathcal{A}}(x) > 0\} . \tag{15.2.4}$$

An illustration is shown in Figure 15.2.2.

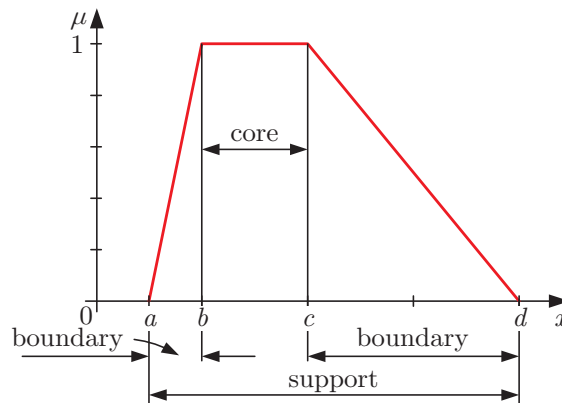


Figure 15.2.2: Some characteristics of a membership function

- The *core* of a normal fuzzy set \mathcal{A} is the crisp set that contains all the elements of \mathcal{X} that have the membership grades of one in \mathcal{A} , i.e.

$$\text{core}(\mathcal{A}) = \{x \in \mathcal{X} \mid \mu_{\mathcal{A}}(x) = 1\} . \tag{15.2.5}$$

- The *boundary* is the crisp set that contains all the elements of \mathcal{X} that have the membership grades of $0 < \mu_{\mathcal{A}}(x) < 1$ in \mathcal{A} , i.e.

$$\text{bnd}(\mathcal{A}) = \{x \in \mathcal{X} \mid 0 < \mu_{\mathcal{A}}(x) < 1\} . \tag{15.2.6}$$

Having two fuzzy sets \mathcal{A} and \mathcal{B} based on \mathcal{X} , then both are *similar* if

$$\text{core}(\mathcal{A}) = \text{core}(\mathcal{B}) \quad \text{and} \quad \text{supp}(\mathcal{A}) = \text{supp}(\mathcal{B}). \tag{15.2.7}$$

- If the support of a normal fuzzy set consists of a single element x_0 of \mathcal{X} , which has the property

$$\text{supp}(\mathcal{A}) = \text{core}(\mathcal{A}) = \{x_0\}, \tag{15.2.8}$$

this set is called a *singleton*.

The type of representation of the membership function depends on the base set. If this set consists of many values, or is the base set a continuum, then a *parametric representation* is appropriate. For that functions are used that can be adapted by changing the parameters. Piecewise linear membership functions are preferred, because of their simplicity and efficiency with respect to computability. Mostly these are trapezoidal or triangular functions, which are defined by four and three parameters, respectively. Figure 15.2.2 shows a trapezoidal function formally described by

$$\mu(x, a, b, c, d) = \begin{cases} 0, & x < a, x > d \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b < x < c \\ \frac{d-x}{d-c}, & c \leq x \leq d \end{cases}, \tag{15.2.9}$$

which migrates for the case $b = c$ into a triangular membership function. For some applications the modelling requires continuously differentiable curves and therefore smooth transitions, which the trapezoids do not have. Here, for example, three of these functions are mentioned, which are shown in Figure 15.2.3. These are

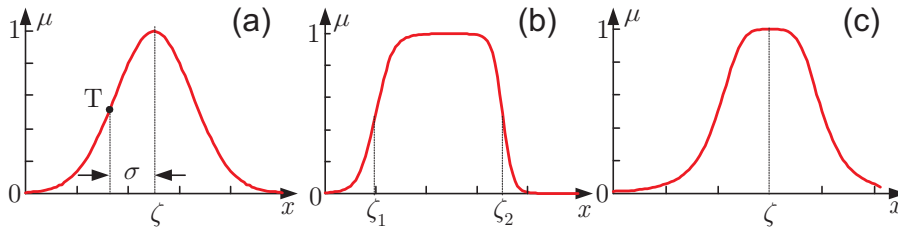


Figure 15.2.3: Membership functions with smooth transitions (Eqs.(15.2.10) to (15.2.12))

- the normalised Gaussian function (Figure 15.2.3a)

$$\mu(x, \zeta, \sigma) = e^{-\frac{(x-\zeta)^2}{2\sigma^2}}, \tag{15.2.10}$$

- the difference of two sigmoidal functions (Figure 15.2.3b)

$$\mu(x, \alpha_1, \zeta_1, \alpha_2, \zeta_2) = \left[1 + e^{-\alpha_1(x-\zeta_1)}\right]^{-1} - \left[1 + e^{-\alpha_2(x-\zeta_2)}\right]^{-1} \tag{15.2.11}$$

and

- the generalised bell function (Figure 15.2.3c)

$$\mu(x, \alpha, \beta, \zeta) = \left[1 + \left|\frac{x-\zeta}{\alpha}\right|^{2\beta}\right]^{-1}. \tag{15.2.12}$$

Interactive Questions 15.2
 What are linguistic terms?

15.3 Elementary operators for fuzzy sets

The basic connective operations in classical set theory are those of intersection, union and complement. These operations on characteristic functions can be generalised to fuzzy sets in more than one way. However, one particular generalisation, which results in operations that are usually referred to us as standard fuzzy set operations, has a special significance in fuzzy set theory. In the following, only the standard operations are introduced. The following operations can be defined:

- The *fuzzy intersection* operator \cap (fuzzy AND connective) applied to two fuzzy sets \mathcal{A} and \mathcal{B} with the membership functions $\mu_{\mathcal{A}}(x)$ and $\mu_{\mathcal{B}}(x)$ is

$$\mu_{\mathcal{A} \cap \mathcal{B}}(x) = \min\{\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)\}, \quad x \in \mathcal{X}. \quad (15.3.1)$$

- The *fuzzy union* operator \cup (fuzzy OR connective) applied to two fuzzy sets \mathcal{A} and \mathcal{B} with the membership functions $\mu_{\mathcal{A}}(x)$ and $\mu_{\mathcal{B}}(x)$ is

$$\mu_{\mathcal{A} \cup \mathcal{B}}(x) = \max\{\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)\}, \quad x \in \mathcal{X}. \quad (15.3.2)$$

- The *fuzzy complement* (fuzzy NOT operation) applied to the fuzzy set \mathcal{A} with the membership function $\mu_{\mathcal{A}}(x)$ is

$$\mu_{\overline{\mathcal{A}}}(x) = 1 - \mu_{\mathcal{A}}(x), \quad x \in \mathcal{X}. \quad (15.3.3)$$

Whilst the operations according to Eqs. (15.3.1) and (15.3.2) are based on min/max operations, the complement is an algebraic one. Union and intersection can also be defined in an algebraic manner but giving different results as:

- The *fuzzy intersection* operator \cap (fuzzy AND connective) can be represented as the *algebraic product* of two fuzzy sets \mathcal{A} and \mathcal{B} , which is defined as the multiplication of their membership functions:

$$\mu_{\mathcal{A} \cap \mathcal{B}}(x) = \mu_{\mathcal{A}}(x) \mu_{\mathcal{B}}(x), \quad x \in \mathcal{X}. \quad (15.3.4)$$

- The *fuzzy union* operator \cup (fuzzy OR connective) can be represented as the *algebraic sum* of two fuzzy sets \mathcal{A} and \mathcal{B} , which is defined as:

$$\mu_{\mathcal{A} \cup \mathcal{B}}(x) = \mu_{\mathcal{A}}(x) + \mu_{\mathcal{B}}(x) - \mu_{\mathcal{A}}(x) \mu_{\mathcal{B}}(x), \quad x \in \mathcal{X}. \quad (15.3.5)$$

Demonstration Example 15.2

What is Fuzzy AND ?

Demonstration Example 15.3

What is Fuzzy OR ?

Demonstration Example 15.4

What is Fuzzy NOT ?

The standard connective operations for fuzzy sets are now defined. As one can easily see, these operations perform precisely as the corresponding operations for crisp sets when the range of membership grades is restricted to the set $\{0, 1\}$. That is, the standard fuzzy operations are generalisations of the corresponding classical set operations. However, they are not the only possible generalisation. As shown above, the fuzzy intersection, union and complement are not unique operations, contrary to their crisp counterparts. Different functions may be appropriate to represent these operations in different contexts. The capability to determine appropriate membership functions and meaningful fuzzy operations in the context of each particular application is crucial for making fuzzy set theory practically useful.

When fuzzy operators are later applied within more complex fuzzy logic operations for rules and fuzzy reasoning, one has to take care of the right combinations of fuzzy operations. For example, in classical

set theory, the operations of intersection and union are dual with respect to the complement in the sense that they satisfy the De Morgan laws

$$\overline{\mathcal{A} \cap \mathcal{B}} = \overline{\mathcal{A}} \cup \overline{\mathcal{B}} \quad \text{and} \quad \overline{\mathcal{A} \cup \mathcal{B}} = \overline{\mathcal{A}} \cap \overline{\mathcal{B}} .$$

It is desirable that this duality be satisfied for fuzzy sets as well. Other combinations need equivalences for commutativity, associativity and distributivity. From Table 15.3 the type of operations can be determined for which operations are valid. Only distributivity is not given in the arithmetic case. Therefore, one has to be careful in applications where arithmetic operations are performed.

Table 15.3.1: Validity of fuzzy equivalences

		max/min	arithmetic
commutativity	$\mathcal{A} \cup \mathcal{B} = \mathcal{B} \cup \mathcal{A}$ $\mathcal{A} \cap \mathcal{B} = \mathcal{B} \cap \mathcal{A}$	*	*
associativity	$\mathcal{A} \cup (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cap \mathcal{C}$ $\mathcal{A} \cap (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cup \mathcal{C}$	*	*
distributivity	$\mathcal{A} \cup (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cap (\mathcal{A} \cup \mathcal{C})$ $\mathcal{A} \cap (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{C})$	*	
De Morgan	$\overline{(\mathcal{A} \cup \mathcal{B})} = \overline{\mathcal{A}} \cap \overline{\mathcal{B}}$ $\overline{(\mathcal{A} \cap \mathcal{B})} = \overline{\mathcal{A}} \cup \overline{\mathcal{B}}$	*	*

Interactive Questions 15.3
What do you know from fuzzy operators ?

Interactive Questions 15.4
The fuzzy NOT properties

15.4 Fuzzy relations

In the introduction to the fuzzy control methodology, section 14.3, rules have been introduced, which in mathematical notation are connective operations over fuzzy sets. For example, the operations on premises in Eq.(14.3.1) can be handled for each rule already by the elementary standard operators introduced in section 15.3. The means are now available to handle steps 1 and 2 of the methodology. But to cope with step 3 something more is needed to complete the modelling of rules. That is now added in this section.

First, relations are explained by a simple example from daily life using discrete fuzzy sets. Let us describe the relationship between the colour of a fruit x and the grade of maturity y and characterise the linguistic variable colour by a crisp set \mathcal{X} with three linguistic terms as

$$\mathcal{X} = \{ \textit{green}, \textit{yellow}, \textit{red} \} ,$$

and similarly the grade of maturity as

$$\mathcal{Y} = \{ \textit{verdant}, \textit{half-mature}, \textit{mature} \} .$$

One knows that a crisp formulation of a relation $\mathcal{X} \rightarrow \mathcal{Y}$ between the two crisp sets would look like this in tabular form:

	<i>verdant</i>	<i>half-mature</i>	<i>mature</i>
<i>green</i>	1	0	0
<i>yellow</i>	0	1	0
<i>red</i>	0	0	1

The zeros and ones describe the grade of membership to this relation. This relation is now a new kind of crisp set that is built from the two crisp base sets \mathcal{X} and \mathcal{Y} . This new set is now called \mathcal{R} and can be expressed also by the rules:

- (1) *IF the colour is green THEN the fruit is verdant*
- (2) *IF the colour is yellow THEN the fruit is half-mature*
- (3) *IF the colour is red THEN the fruit is mature*

As can be seen from this example, a relation, which is called a rule or rule base, can be used to provide a model.

Demonstration Example 15.5

Test this with your own tomatoes

This crisp relation \mathcal{R} represents the presence or absence of association, interaction or interconnection between the elements of these two sets. This can be generalised to allow for various degrees of strength of association or interaction between elements. Degrees of association can be represented by membership grades in a fuzzy relation in the same way as degrees of the set membership are represented in a fuzzy set. Applying this to the fruit example, the table can be modified to

	<i>verdant</i>	<i>half-mature</i>	<i>mature</i>
<i>green</i>	1	0.5	0
<i>yellow</i>	0.3	1	0.4
<i>red</i>	0	0.2	1

where there are now real numbers in $[0, 1]$. This table represents a fuzzy relation and models the connectives in a fuzzy rule base. It is a two-dimensional fuzzy set and the question now is, how can this set be determined from its elements.

Demonstration Example 15.6

Test this with your own strawberries

In order to do this, the elements are generalised. In the above example, the linguistic terms were treated as crisp terms. For example, when one represents the colours on a colour spectrum scale, the colours would be described by their spectral distribution curves that can be interpreted as membership functions and then a particular colour is a fuzzy term. Treating also the grades of maturity as fuzzy terms, the above relation is a two-dimensional fuzzy set over two fuzzy sets. For example, taking from the fruit example the relation between the linguistic terms *red* and *mature*, and represent them by the membership functions as shown in Figure 15.4.1a, a fruit can be characterised by the property *red* AND *mature*. This expression can be re-written in mathematical form using elementary connective operators (see Eqs. (15.3.1) or (15.3.4)) for the membership functions by

$$\mu_{\mathcal{R}}(x, y) = \min\{\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(y)\} \quad (15.4.1)$$

or

$$\mu_{\mathcal{R}}(x, y) = \mu_{\mathcal{A}}(x) \mu_{\mathcal{B}}(y) . \quad (15.4.2)$$

Figure 15.4.1b shows a 3-dimensional view of these two fuzzy terms and Figure 15.4.1c the result of the connective operation according to Eq.(15.4.1). This result combines the two fuzzy sets by an operation that is a Cartesian product

$$\mathcal{R} : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1] . \quad (15.4.3)$$

From this example it is obvious that the connective operation in a rule for the \rightarrow operation is simply performed by a fuzzy intersection in two dimensions. For this, both intersection operators from Eqs. (15.3.1) or (15.3.4) can be used.

Combining rules into a rule base the example from above may help when it is rewritten as

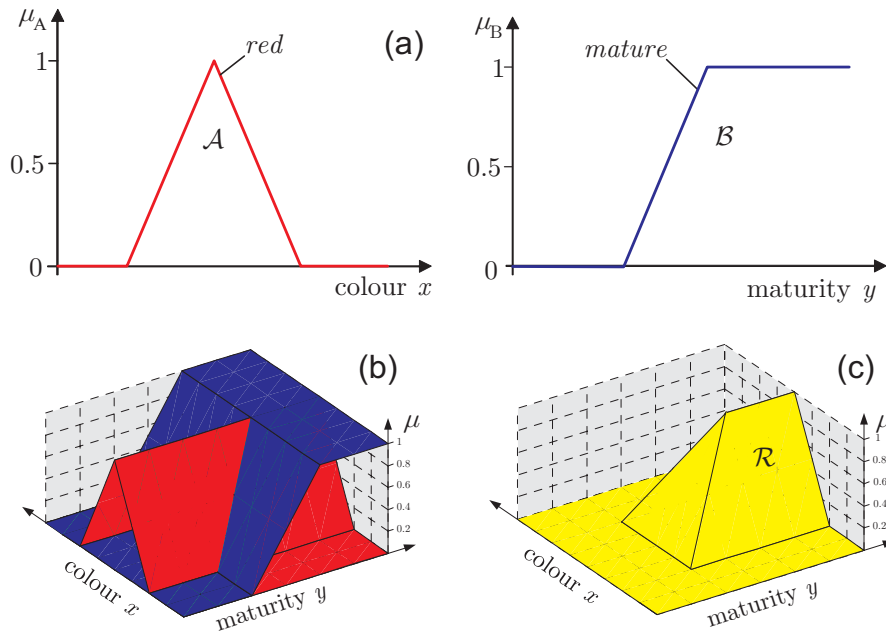


Figure 15.4.1: Relation between two fuzzy sets: (a) membership functions, (b) 3-D view of the membership functions, (c) membership function of the relation after applying the min operation to (b)

- (1) *IF the colour is green THEN the fruit is verdant*
 OR
 (2) *IF the colour is yellow THEN the fruit is half-mature*
 OR
 (3) *IF the colour is red THEN the fruit is mature*

which describes in a linguistic way a union of three rules. For the complete rule base \mathcal{R} one can combine the relations formed for each individual rule with a fuzzy union operator, which is the fuzzy OR according to Eqs. (15.3.2) or (15.4.1).

Now, step 4 of the methodology introduced in section 14.3 can be specified by taking the rule base from Eq. (14.3.1) and applying the union operator by writing the rule base with max/min operators as follows:

$$\mu_{\mathcal{R}}(x_1, x_2, \dots, x_n, u) = \max_r \{ \min \{ \mu_{P_r}(x_1, x_2, \dots, x_n), \mu_{B_r}(u) \} \}, \quad (15.4.4)$$

where $\mu_{P_r}(x_1, x_2, \dots, x_n)$ is the premise of the r th rule. This representation is the standard max/min representation of a rule base that will be later used for fuzzy controllers. Instead of the max/min representation a so called max-prod representation is also usual, where the algebraic product

$$\mu_{\mathcal{R}}(x_1, x_2, \dots, x_n, u) = \max_r \{ \mu_{P_r}(x_1, x_2, \dots, x_n) \mu_{B_r}(u) \} \quad (15.4.5)$$

is used to build the relation between the premise and the conclusion.

15.5 Fuzzy composition

In order to make step 5 of the methodology introduced in section 14.3, a reasoning method is needed. Explaining reasoning by the fruit example, it is assumed that one has a crisp fact: a green fruit. The decision from the rule base is obvious: the fruit is verdant, and this is similar for the other facts: yellow and red. But if one has a fact like: the fruit is orange, one does not know how to determine which rule fires the decision and what the decision is. In the following, a new fuzzy operation type is introduced, that allows to operate with a given fact and a fuzzy relation to produce an output that represents the

decision in a fuzzy way. This operation is called *fuzzy reasoning*, which is a special case of the more general operation called *fuzzy composition*.

Two relations of the form given in Eq. (15.4.3)

$$\begin{aligned} \mathcal{R} &: \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1] \\ \mathcal{S} &: \mathcal{Y} \times \mathcal{Z} \rightarrow [0, 1] , \end{aligned} \tag{15.5.1}$$

can be composed to one relation

$$\mathcal{T} : \mathcal{X} \times \mathcal{Z} \rightarrow [0, 1] . \tag{15.5.2}$$

This process is known as composition and, using the max and min operators for union and intersection, one can express the composition operation $\mathcal{T} = \mathcal{R} \circ \mathcal{S}$ by the corresponding membership functions as follows:

$$\mu_{\mathcal{T}}(x, z) = \max_{y \in \mathcal{Y}} \{ \min \{ \mu_{\mathcal{R}}(x, y), \mu_{\mathcal{S}}(y, z) \} \} . \tag{15.5.3}$$

When one takes the above fruit example again with the colour-maturity relation \mathcal{R}

\mathcal{R}	<i>verdant</i>	<i>half-mature</i>	<i>mature</i>
<i>green</i>	1	0.5	0
<i>yellow</i>	0.3	1	0.4
<i>red</i>	0	0.2	1

and define for \mathcal{S} a maturity-taste relation

\mathcal{S}	<i>sour</i>	<i>tasteless</i>	<i>sweet</i>
<i>verdant</i>	1	0.2	0
<i>half-mature</i>	0.7	1	0.3
<i>mature</i>	0	0.7	1

then by applying Eq. (15.5.3) to the elements of these two tables, the following is obtained:

$\mathcal{T} = \mathcal{R} \circ \mathcal{S}$	<i>sour</i>	<i>tasteless</i>	<i>sweet</i>
<i>green</i>	1	0.5	0.3
<i>yellow</i>	0.7	1	0.4
<i>red</i>	0.2	0.7	1

When the fuzzy set \mathcal{S} is now interpreted as a rule base and the fuzzy set \mathcal{R} as a fact obtained from some measurement data, then the fuzzy set \mathcal{T} is the result from the reasoning process, which is in this case a relation.

Demonstration Example 15.7
 Before you continue taste your fruits

In the following, first, a one-dimensional crisp fact is taken. Define the fruit colour green as a fact by the singleton

$$\mathcal{C}' = \{ 1 \ 0 \ 0 \} ,$$

where the numbers are the intensity grades of the colours green, yellow and red. When one calculates the composition $\mathcal{T}' = \mathcal{C}' \circ \mathcal{R}$ by applying Eq. (15.5.3), where in this case the first operand has only one dimension, the fuzzy set for the maturity

$$\mathcal{T}' = \{ 1 \ 0.5 \ 0 \}$$

is obtained. The result is obvious from the first rule of the rule base \mathcal{R} . When a different colour is taken than included in the rule base entries, say orange as

$$\mathcal{C}' = \{ 0 \ 0.5 \ 0.5 \} ,$$

then there is no problem to obtain the value for the maturity

$$\mathcal{T}' = \{ 0.3 \quad 0.5 \quad 0.5 \}$$

by applying the composition formula. The reasoning process is now solved.

In the same manner as relations can be composed, the one-dimensional facts can be composed with the rule base to realise the reasoning operation. This can now be precisely re-formulated for the general case of a rule base according to Eq. (14.3.1).

If for the rule base

$$\mathcal{R} : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$$

its membership function is described by Eq. (15.4.4) or (15.4.5) and if there is a fact described by the fuzzy set

$$\mathcal{A}' : \mathcal{X} \rightarrow [0, 1]$$

and its membership function $\mu_{\mathcal{A}'}(x)$, the result

$$\mathcal{B}' = \mathcal{A}' \circ \mathcal{R} : \mathcal{Y} \rightarrow [0, 1]$$

of the fuzzy reasoning is represented by the membership function

$$\mu_{\mathcal{B}'}(y) = \max_{x \in \mathcal{X}} \{ \min \{ \mu_{\mathcal{A}'}(x), \mu_{\mathcal{R}}(x, y) \} \} . \quad (15.5.4)$$

Module 16

Fuzzy systems

Module units

16.1 Fuzzification	16-2
16.2 Fuzzy inference machine	16-2
16.3 Defuzzification	16-5
16.3.1 Centre of gravity method (COG)	16-5
16.3.2 Centre of singleton method (COS)	16-5
16.3.3 Maximum methods	16-7
16.3.4 Margin properties of the centroid methods	16-7
16.4 The Takagi-Sugeno fuzzy system	16-7
16.5 The components of a fuzzy system	16-8

Module overview. *The idea of a fuzzy system is introduced in this module. Its main components are described and discussed in detail using examples. The results from the fuzzification module drives the rule base. The fuzzy inference machine is developed which solves the reasoning. The methods to transform the fuzzy results of the reasoning process to crisp data is shown in detail. The mainstream in the fuzzy system of this module is the Mamdani approach which needs the defuzzification step, but the Takagi-Sugeno-type of fuzzy system is also introduced which avoids defuzzification.*

Module objectives. *When you have completed this module you should be able to:*

1. Design a fuzzy system.
2. Fuzzify input information.
3. Describe a fuzzy inference machine using fuzzy sets.
4. Defuzzify results from the reasoning process.

Module prerequisites. *Fuzzy sets.*

In the previous section, elementary fuzzy terms and fuzzy logic operations have been introduced. In this section, the application to the treatment of rule-based knowledge follows. For this a rule-based fuzzy system is needed, containing a rule base and a reasoning algorithm, which is used to process crisp or fuzzy input values x_i , $i = 1, \dots, n$ to a crisp or fuzzy output value y , see Figure 16.0.1. Using multiple inputs and one output implies no restriction as a multi-input-multi-output fuzzy system can always be decomposed into multiple systems according to Figure 16.0.1. Such systems are the basis for the realisation of fuzzy controllers. As there are mostly crisp input values x_i from measurements and for controllers only a crisp output y , a fuzzy system must contain additional components, fuzzification and defuzzification.

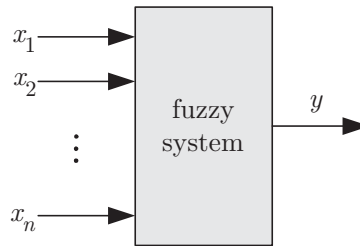


Figure 16.0.1: Rule-based fuzzy system with n inputs and one output

16.1 Fuzzification

The fuzzification comprises the process of transforming crisp values into grades of membership for linguistic terms of fuzzy sets. The membership function is used to associate a grade to each linguistic term.

Example 16.1.1

For the fuzzification of the car speed value $x_0 = 70$ km/h the two membership functions μ_A and μ_B from Figure 16.1.1 can be used, which characterise a low and a medium speed fuzzy set, respectively. The given speed value of $x_0 = 70$ km/h belongs with a grade of $\mu_A(x_0) = 0.75$ to the fuzzy set “low” and with a grade of $\mu_B(x_0) = 0.25$ to the fuzzy set “medium”. ■

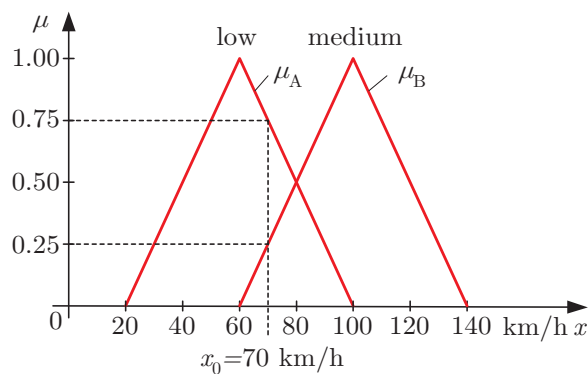


Figure 16.1.1: Fuzzification of a car speed

Demonstration Example 16.1 Fuzzification example

16.2 Fuzzy inference machine

The core section of a fuzzy system is that part, which combines the facts obtained from the fuzzification with the rule base and conducts the fuzzy reasoning process. This is called a *fuzzy inference machine*. Here rule and composition operations are applied from sections 15.4 and 15.5.

In the following, for simplicity it is assumed that there is only one input $x_1 = x$ and the rule base is described with max/min operators from Eq. (15.4.4) by

$$\mu_R(x, u) = \max_r \{ \min \{ \mu_{P_r}(x), \mu_{B_r}(u) \} \}, \quad (16.2.1)$$

then inserting Eq. (16.2.1) into Eq. (15.5.4) yields

$$\mu_{B'}(u) = \max_{x \in \mathcal{X}} \{ \min \{ \mu_{A'}(x), \max_r \{ \min \{ \mu_{P_r}(x), \mu_{B_r}(u) \} \} \} \}.$$

The max operations can be reordered such that only the relevant operands are on the right-hand side. Then

$$\begin{aligned} \mu_{B'}(u) &= \max_r \{ \min \{ \underbrace{\max_{x \in \mathcal{X}} \{ \min \{ \mu_{A'}(x), \mu_{P_r}(x) \} \}}_{H_r}, \mu_{B_r}(u) \} \} \\ &= \max_r \{ \min \{ H_r, \mu_{B_r}(u) \} \} \end{aligned} \tag{16.2.2}$$

is obtained for the reasoning process. The inner term H_r , which combines the fact with the premise, is a constant and is called *degree of relevance* of the rule r . It characterises the relevance of the fired rule r and can be treated as a de-normalised universal fuzzy set. The following example will help to interpret and provide understanding of this reasoning operation by applying graphical means.

Example 16.2.1

A simple fuzzy system is given, which models the brake behaviour of a car driver depending on the car speed. The inference machine should determine the brake force for a given car speed. The speed is specified by the two linguistic terms “low” and “medium”, and the brake force by “moderate” and “strong”. The rule base includes the two rules

- (1) IF the car speed is low THEN the brake force is moderate
- (2) IF the car speed is medium THEN the brake force is strong

For the linguistic terms in the premises “speed is low” and “speed is medium” those from Figure 16.1.1 are used. The terms “force is moderate” and “force is strong” are defined in Figure 16.2.1. For these two rules

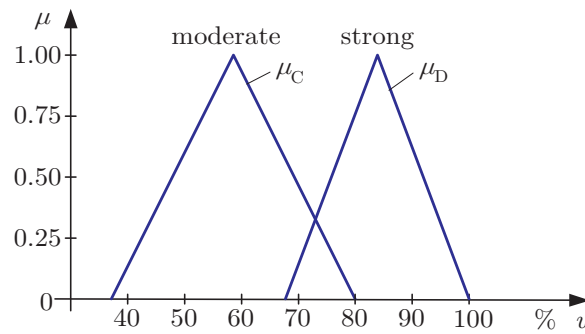


Figure 16.2.1: Fuzzy set of a car brake force

the membership functions for the premises are $\mu_{P_1} = \mu_A$, $\mu_{P_2} = \mu_B$, and for the conclusions $\mu_{B_1} = \mu_C$, $\mu_{B_2} = \mu_D$. The car speed is $x_0 = 70$ km/h, which is a crisp value and therefore can be represented in fuzzy notation by the singleton $\mu_{A'}(x) = \{70 \text{ km/h}\}$. Inserting these values into Eq. (16.2.2), one obtains

$$\mu_{B'}(u) = \max \{ \min \{ H_1, \mu_C(u) \}, \min \{ H_2, \mu_D(u) \} \}$$

with

$$H_1 = \max_{x \in \mathcal{X}} \{ \min \{ \mu_{A'}(x), \mu_A(x) \} \}$$

and

$$H_2 = \max_{x \in \mathcal{X}} \{ \min \{ \mu_{A'}(x), \mu_B(x) \} \} .$$

In rule 1, there is only a relation between μ_A and μ_C . The degree of relevance H_1 for rule 1 is a horizontal line of height 0.75, which is the value from the fuzzification as shown in Figure 16.2.2a. The reasoning operation $\min \{ H_1, \mu_C(u) \}$ with rule 1 cuts the membership function $\mu_C(u)$ by this line (yellow area). The rule 2 is evaluated analogously, as shown in Figure 16.2.2b. Figure 16.2.2c shows the final result from the union operation over all yellow areas from the inferences obtained with the individual rules. ■

Demonstration Example 16.2
An inference example with one rule

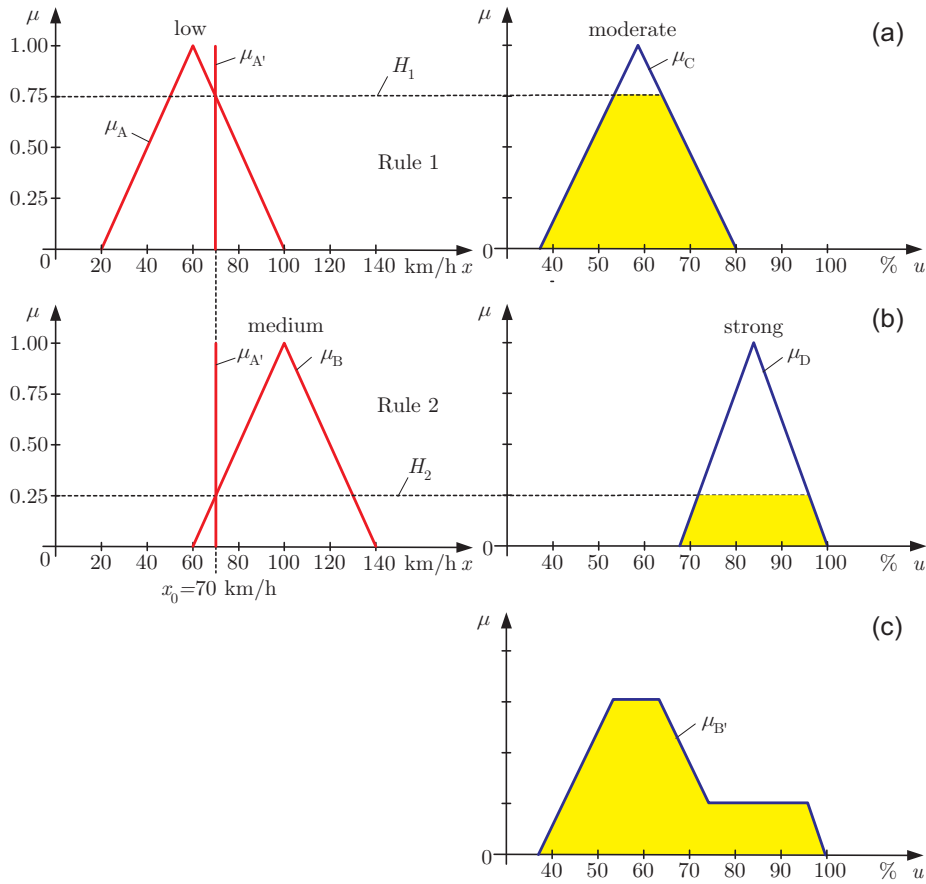


Figure 16.2.2: Fuzzy inference example: (a) inference with rule 1, (b) with rule 2, and (c) final fuzzy set

Demonstration Example 16.3
An inference example with more rules

The inference shown above is based on the max/min representation of a rule base from Eq. (15.4.4). Taking alternatively the max-prod representation from Eq. (15.4.5)

$$\begin{aligned} \mu_{B'}(u) &= \max_r \{ \underbrace{\max_{x \in \mathcal{X}} \{ \min \{ \mu_{A'}(x), \mu_{P_r}(x) \} \}}_{H_r} \mu_{B_r}(u) \} \\ &= \max_r \{ H_r \mu_{B_r}(u) \} \end{aligned} \tag{16.2.3}$$

is obtained for the max-prod inference. The difference between the max/min and max-prod inference is that in the first case the membership function of the conclusion is cut and in the second case scaled. Figure 16.2.3 illustrates this.

Example 16.2.2

For a fuzzy system with the two inputs x_1 and x_2 and the output u the inference of the two fuzzy rules

- (1) IF ($x_1 = P$) AND ($x_2 = M$) THEN ($u = M$)
- (2) IF ($x_1 = N$) OR ($x_2 = S$) THEN ($u = S$),

should be evaluated. The linguistic term S stands for “small”, M for “medium”, L for “large”, N for “negative” and P for “positive”.

Each rule contains two premises, which are differently connected. In rule 1 the connective operation is the intersection, which can be performed by the min operation according to Eq. (15.3.1) for $\mu_{A_{11}}(x_1)$ and

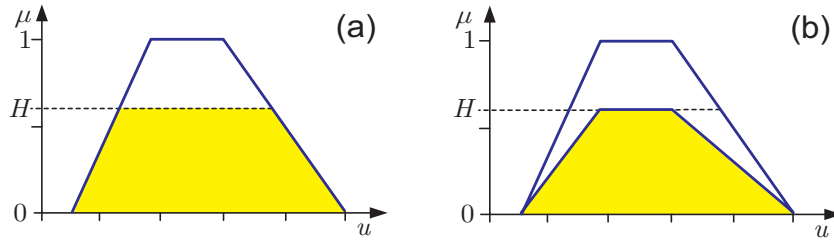


Figure 16.2.3: Membership function of a conclusion using (a) max/min and (b) max-prod inference

$\mu_{A_{12}}(x_2)$, and in rule 2 the premise is a union of the two premises ($x_1 = N$) and ($x_2 = S$), which can be performed by the max operation according to Eq. (15.3.2) for $\mu_{A_{21}}(x_1)$ and $\mu_{A_{22}}(x_2)$. In the first case the degree of relevance is $H_1 = \mu_{A_{12}}(x_2)$ and in the second case $H_2 = \mu_{A_{22}}(x_2)$. The membership functions of the conclusion of each rule will be determined using the degree of relevance of the corresponding rule by applying either the max/min inference method according to Eq. (16.2.2) or the max-prod inference method according to Eq. (16.2.3). The reasoning process using both inference methods is visualised in Figure 16.2.4. ■

16.3 Defuzzification

As a result of applying the previous steps, one obtains a fuzzy set $\mu_{B'}(u)$ from the reasoning process that describes, for each possible value u , how reasonable it is to use this particular value. In other words, for every possible value u , one gets a grade of membership that describes to what extent this value u is reasonable to use. Using a fuzzy system as a controller, one wants to transform this fuzzy information into a single value u' that will actually be applied. This transformation from a fuzzy set to a crisp number is called a *defuzzification*. It is not a unique operation as different approaches are possible. The most important ones for control are described in the following.

16.3.1 Centre of gravity method (COG)

This approach has its origin in the idea to select a value u' that, on average, would lead to the smallest error in the sense of a criterion. If u' is chosen, and the best value is u , then the error is $u' - u$. Thus, to determine u' the least squares method can be used. As weights for each square $(u' - u)^2$, one can take the grade of membership $\mu_{B'}(u)$ with which u is a reasonable value. As a result one has to find

$$u' = \arg \min_{u'} \int_{\mathcal{U}} \mu_{B'}(u) (u' - u)^2 du . \tag{16.3.1}$$

Differentiating with respect to the unknown u' and equating the derivative to zero, the formula

$$u' = \frac{\int_{\mathcal{U}} u \mu_{B'}(u) du}{\int_{\mathcal{U}} \mu_{B'}(u) du} \tag{16.3.2}$$

is obtained, which determines the value of the abscissa of the centre of gravity of the area below the membership function $\mu_{B'}(u)$.

16.3.2 Centre of singleton method (COS)

The defuzzification can be strongly simplified if the membership functions $\mu_{B_r}(u)$ of the conclusions are singly defuzzified for each rule such that each function is reduced to a singleton that has the position u'_r .

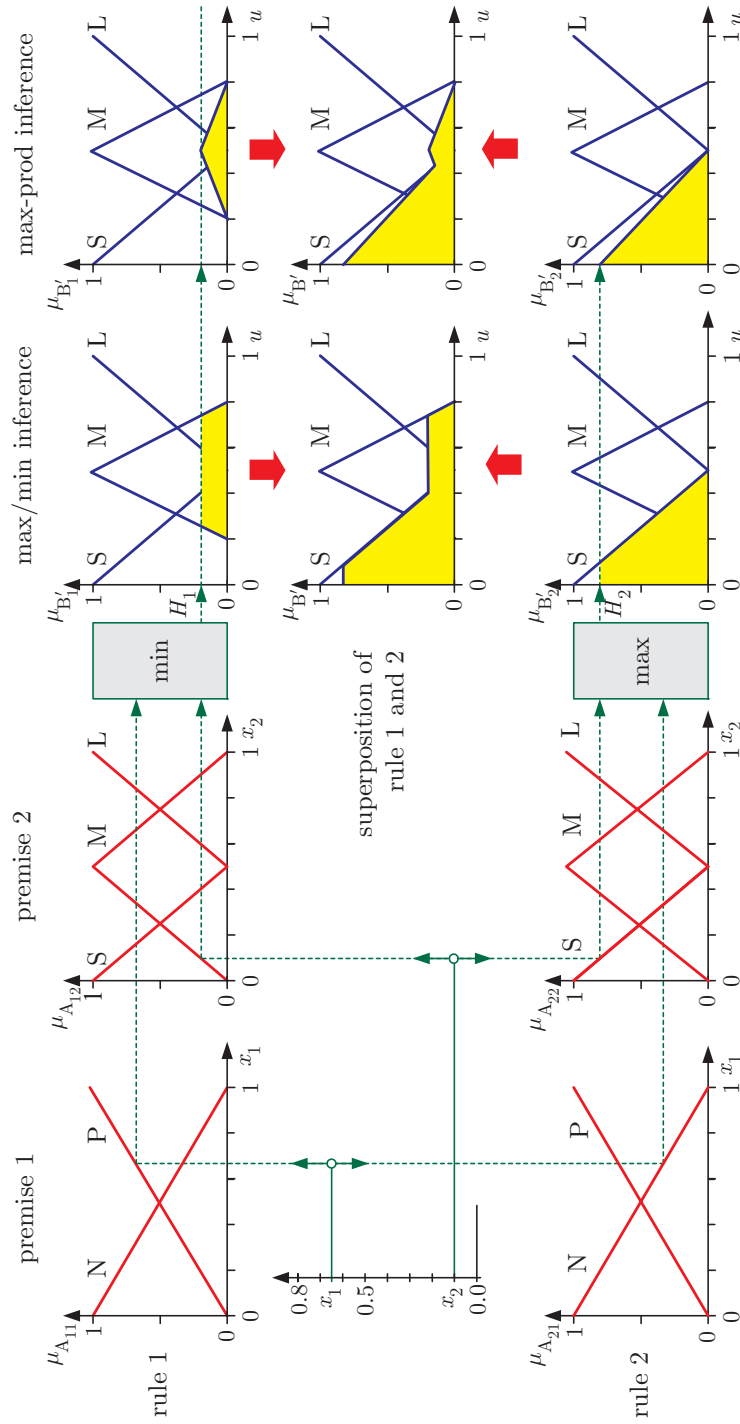


Figure 16.2.4: Example of the application with two premises with (a) max/min inference and (b) max-prod inference

of the individual membership function's centre of gravity. The centre of singletons is calculated by using the degree of relevance as follows:

$$u' = \frac{\sum_R u'_r H_r}{\sum_R H_r} . \quad (16.3.3)$$

The simplification consists in that the singletons can be determined already during the design of the fuzzy system and that the membership function $\mu_{B'}(u)$ with its complicated geometry is no longer needed. The defuzzification using this formula is an approximation of the defuzzification by Eq. (16.3.2). Experiences from control show that there are slight differences between both approaches, which can be in most cases neglected.

16.3.3 Maximum methods

This class of methods determines u' by selecting the membership function with the maximum value. If the maximum is a range, either the lower, upper or the middle value is taken for u' depending on the method. Using these methods, the rule with the maximum activity always determines the value, and therefore they show discontinuous and step output on continuous input. This is the reason why these types of method are not attractive for use in controllers.

16.3.4 Margin properties of the centroid methods

As the centre of gravity of the area below the membership functions cannot reach the margins of \mathcal{U} , the membership functions, which are at the margins, must be symmetrically expanded when obtaining the centre of gravity. This is necessary in order to have the full range of \mathcal{U} available. This is shown in Figure 16.3.1. The same expansion is also necessary for the COS method.

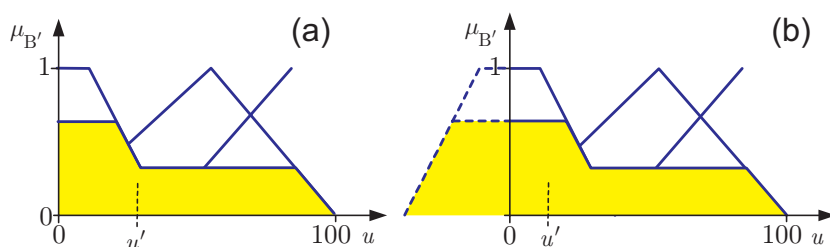


Figure 16.3.1: Margin of $\mu_{B'}(u)$ (a) original and (b) expanded

Demonstration Example 16.4

Properties of the defuzzification methods

16.4 The Takagi-Sugeno fuzzy system

In the rule bases described hitherto with the IF-THEN rules of this chapter fuzzy sets both in the premises and in the conclusions are used. This kind of inference is called Mamdani inference. A modified inference scheme, developed by Takagi and Sugeno, represents the conclusions by functions. A rule of this form will be

$$\text{IF } x_1 \text{ is } A_{r1} \text{ AND } x_2 \text{ is } A_{r2} \dots \text{ AND } x_n \text{ is } A_{rn} \text{ THEN } u = f_r(x_1, x_2, \dots, x_n) .$$

The structure of the premises are the same as for the Mamdani inference. However, in the conclusion all linguistic terms B_r are substituted by the functions f_r , and therefore it is not necessary to define a priori linguistic terms $B_r(u)$ for the conclusions as in Eq. (14.3.1). The function f_r represents a direct mapping from the input space $\mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_n$ with the input values x_1, x_2, \dots, x_n to the output space \mathcal{U} .

The connective operation in a rule is in this case performed via the degree of relevance H_r of the premise of the rule R_r and the function f_r in the conclusion. The final output is determined as a weighted mean value over all R rules according to

$$u' = \frac{\sum_R H_r f_r(x_1, x_2, \dots, x_n)}{\sum_R H_r}. \quad (16.4.1)$$

The effort of performing a defuzzification is saved, as the crisp value u' is directly determined by the inference operation and this makes this method attractive.

The Takagi-Sugeno fuzzy system builds an overall combination of functions f_r , which are valid in some range. If the membership functions of the fuzzy sets in the premises are overlapping, the transition between the functions is always continuous. For the special case of linear functions

$$f_r(x_1, x_2, \dots, x_n) = \sum_{\nu=1}^n c_{r\nu} x_\nu \quad (16.4.2)$$

the coefficients $c_{r\nu}$ can be determined by some identification procedure.

16.5 The components of a fuzzy system

In the introduction to this fuzzy system section, in Figure 16.0.1 a fuzzy system is drawn as a black box with some inputs and an output. Figure 16.5.1 shows the contents of a fuzzy system. Now it is clear,

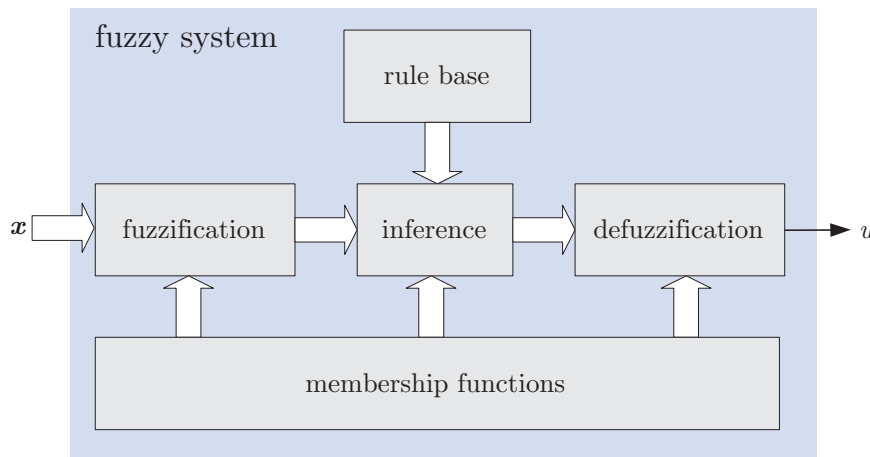


Figure 16.5.1: Components of a fuzzy system

what is in this black box. The input signals combined to the vector $\mathbf{x} = [x_1, x_2, \dots, x_q]^T$ are crisp values, which are transformed into fuzzy sets in the fuzzification block as discussed in section 16.1. The output u comes out directly from the defuzzification block, which transforms an output fuzzy set back to a crisp value using methods from section 16.3. The set of membership functions responsible for the transforming part and the rule base as the relational part contain as a whole the modelling information about the system, which is processed by the inference machine from section 16.2. This rule-based fuzzy system is the basis of a fuzzy controller, which is described in the following section.

Module 17

Fuzzy control

Module units

17.1 Basic structure of a fuzzy controller	17-1
17.2 Transfer behaviour of fuzzy controllers	17-2
17.2.1 Representation using 2D characteristics	17-3
17.2.2 Influence of the membership functions and rule base on the characteristic	17-3
17.2.3 Representation using 3D characteristics	17-7
17.3 Example of a fuzzy control system	17-8
17.3.1 Loading crane plant model	17-9
17.3.2 Fuzzy control system design	17-9
17.4 Contribution of fuzzy control	17-11

Module overview. *A fuzzy controller is a special fuzzy system that can be used as a controller component in a closed-loop system. The integration of a fuzzy system into a closed loop is shown. Special emphasis is put onto the transfer behaviour of fuzzy controllers, which is analysed using different configurations of standard membership functions. An example for the design of a fuzzy controller for a loading crane is given. Finally, the module series is closed by a general discussion about the contribution of fuzzy control.*

Module objectives. *When you have completed this module you should be able to:*

1. Design a fuzzy control system.
2. Know the transfer behaviour of a fuzzy control system.
3. Change the membership functions to influence the transfer behaviour of a fuzzy control system.

Module prerequisites. *Fuzzy systems, fuzzy sets, PID control.*

17.1 Basic structure of a fuzzy controller

A fuzzy controller can be handled as a system that transmits information like a conventional controller with inputs containing information about the plant to be controlled and an output that is the manipulated variable. From outside, there is no vague information visible, both, the input and output values are crisp values. The input values of a fuzzy controller consist of measured values from the plant that are either plant output values or plant states, or control errors derived from the set-point values and the controlled variables.

A control law represented in the form of a fuzzy system is a static control law. This means that the fuzzy rule-based representation of a fuzzy controller does not include any dynamics, which makes a fuzzy controller a static transfer element, like the standard state-feedback controller. In addition to this, a fuzzy controller is in general a fixed nonlinear static transfer element, which is due to those computational steps of its computational structure that have nonlinear properties. In what follows the computational structure of a fuzzy controller will be described by presenting the computational steps involved. The computational structure of a fuzzy controller consists of three main steps as illustrated by the three blocks in Figure 17.1.1:

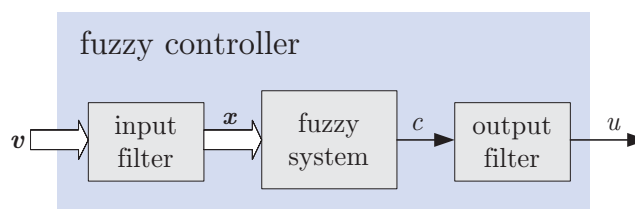


Figure 17.1.1: Basic structure of a fuzzy controller

1. signal conditioning and filtering at the input (input filter),
2. fuzzy system according to Figure 16.5.1, and
3. signal conditioning and filtering at the output (output filter).

The input and output filters are for signal conditioning. The external input signals \mathbf{v} must be scaled such that they can be fed as signals \mathbf{x} into the fuzzification part of the fuzzy system. In many cases, the signals in \mathbf{v} are the control error e and its derivative \dot{e} . In this case the input filter contains a differentiating element. Also other dynamical elements can be in the input filter, e.g. integrators for the control error. Additionally auxiliary signals from the plant measurements may be used that represent plant states or disturbances acting on the plant. The design of this input filter depends on the application, which will be illustrated later by an example.

The fuzzy system contains the control strategy and consists of those components already discussed in section 16.5. For example, a linguistic formulation of a proportional control strategy would be expressed by the following rules of the fuzzy system:

- (1) IF (control error positive) THEN (manipulated variable positive),
- (2) IF (control error zero) THEN (manipulated variable zero),
- (3) IF (control error negative) THEN (manipulated variable negative).

A proper rule base can be found either by asking experts or by evaluation of measurement data using data mining methods.

The output filter is for the adaptation of the crisp output c from the fuzzy system to the manipulated variable u of the plant. In principle, there are many dynamical and static operations possible. Often, the output c of the fuzzy system describes an increment of the manipulated variable, and thus an integration of this increment must occur.

17.2 Transfer behaviour of fuzzy controllers

As already stated, fuzzy systems describe the static behaviour of a fuzzy control strategy. Therefore it is obvious that the transfer behaviour of such a system can be represented by nonlinear characteristics. This provides a fine opportunity to compare fuzzy control strategies amongst themselves and with conventional nonlinear ones.

17.2.1 Representation using 2D characteristics

To provide a first approach for the design of membership functions for a fuzzy controller component, some prototype membership functions are introduced. For such prototypes, linguistic terms are introduced. For example, the following seven terms

NL	negative large	PL	positive large
NM	negative medium	PM	positive medium
NS	negative small	PS	positive small
AZ	approximately zero		

can be used to characterise the triangular shaped fuzzy sets according to Figure 17.2.1. It is important to recognise that the fuzzy sets defined in this figure and the seven linguistic terms are only a reasonable

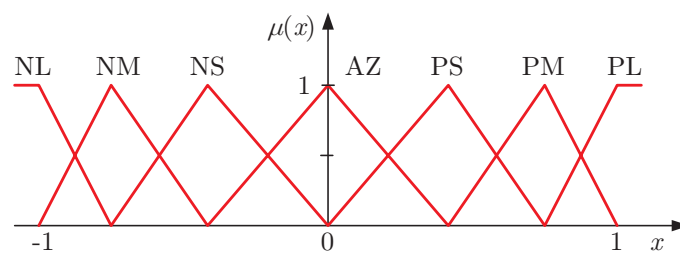


Figure 17.2.1: Prototype membership functions for a fuzzy set with seven linguistic terms

example. For various reasons, emerging from specific applications, other shapes of membership functions might be used over the given ranges. Moreover, different fuzzy sets may be defined for different variables. The prototype membership functions are usually chosen only as a preliminary candidate. They may later be modified by the designer.

For a fuzzy system using this kind of prototype with membership functions which overlaps, both for the premise and the conclusion the transfer characteristic is nonlinear, as shown in the following example.

Example 17.2.1

For a proportional fuzzy controller with the control error $e = w - y$ as input, with the manipulated variable u as output, with the rule base

- (1) IF $e = \text{NL}$ THEN $u = \text{NL}$
- (2) IF $e = \text{NM}$ THEN $u = \text{NM}$
- (3) IF $e = \text{NS}$ THEN $u = \text{NS}$
- (4) IF $e = \text{AZ}$ THEN $u = \text{AZ}$
- (5) IF $e = \text{PS}$ THEN $u = \text{PS}$
- (6) IF $e = \text{PM}$ THEN $u = \text{PM}$
- (7) IF $e = \text{PL}$ THEN $u = \text{PL}$

and with both membership functions of the form shown in Figure 17.2.1, the static nonlinear characteristic $u = u(e)$ is as shown in Figure 17.2.2 using the assumptions given below. ■

17.2.2 Influence of the membership functions and rule base on the characteristic

The above example shows that a fuzzy controller is a nonlinear controller. The form of the characteristic depends only on the rule base and the membership functions of e and u . In the following discussions about the influence of membership functions the following assumptions for the fuzzy controller with the input signal e and the output signal u will be used:

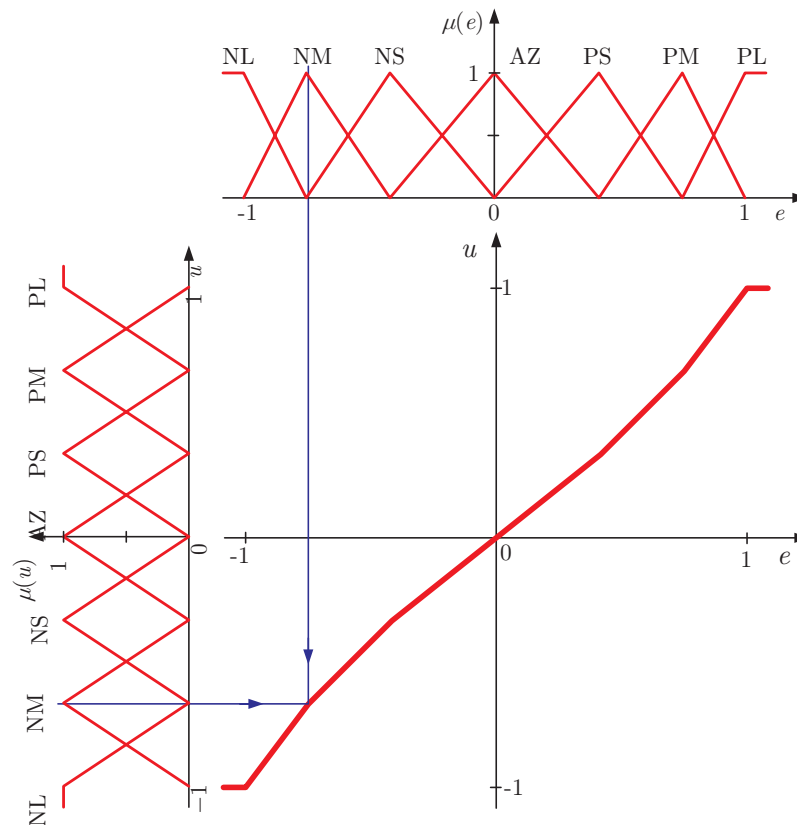


Figure 17.2.2: Nonlinear characteristic of a fuzzy controller

- For the AND connectives the min and for the OR connectives the max operator will be used.
- The max/min inference will be used.
- The defuzzification will be performed by the COG method with symmetrical membership functions at the margins.

Input and output values are normalised to the interval $[-1, 1]$, and at first, only the three linguistic terms NS (negative small), AZ (approximate zero) and PS (positive small) are considered. The rule base is that of a proportional fuzzy controller

- (1) IF $e = \text{NS}$ THEN $u = \text{NS}$,
- (2) IF $e = \text{AZ}$ THEN $u = \text{AZ}$,
- (3) IF $e = \text{PS}$ THEN $u = \text{PS}$,

with the membership functions shown in Figure 17.2.3a and b. The static characteristic in Figure 17.2.3c is odd symmetrical about the origin due to the symmetry of the membership functions. Because of the different supports of the membership function for the fuzzy sets AZ of both functions, the characteristic is approximately piecewise linear and has three distinct levels. The membership functions of the input e have two overlaps in the intervals $[-0.6, -0.4]$ and $[0.4, 0.6]$ that correspond precisely with the ranges with the positive slope of the curve. The reason for this is just that two rules in these ranges are simultaneously active. On the other hand, in the non-overlapping ranges only one rule is active. The membership function of the output depends in this case only on the degree of relevance and thus the centre of gravity of the membership function remains constant.

If the number of linguistic terms for the input and output is increased, the characteristic is similar, but with more sections. The number of sections depends only on the number of linguistic terms and the width of the sections depends on the degree of overlapping. In the special case without overlapping in

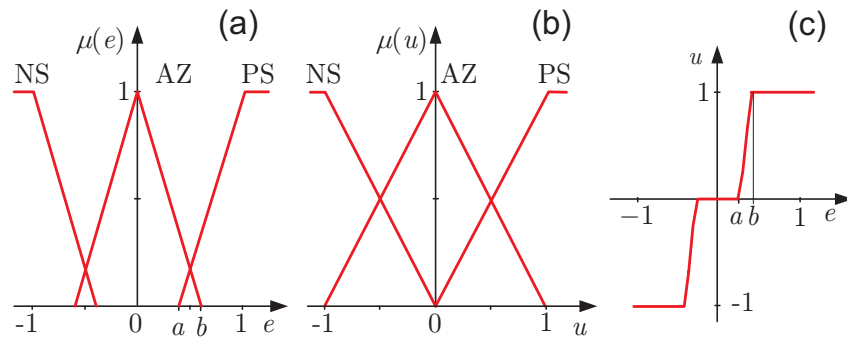


Figure 17.2.3: Membership functions and static characteristic of the fuzzy controller

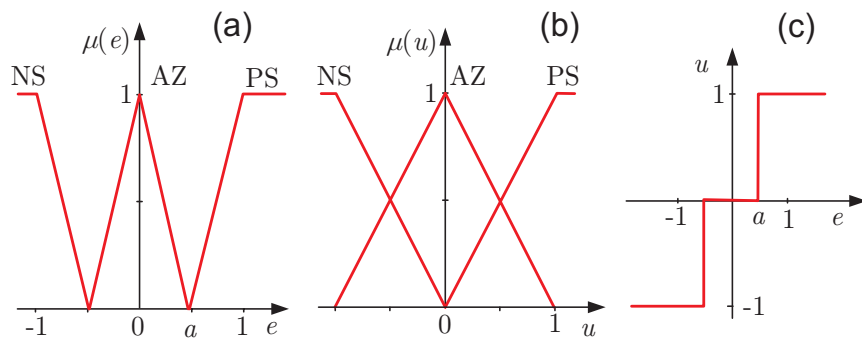


Figure 17.2.4: Influence of the (c) characteristic of a proportional fuzzy controller (a) without overlapping in the input membership functions and (b) with full overlapping in the output membership functions

the input one obtains the characteristic of a three-level controller, as shown in Figure 17.2.4. In this case only one rule is active such that only the three crisp values -1, 0 and 1 are generated. Now, consider varying the degree of overlap of the output membership functions. Figure 17.2.5 shows the case with full overlap on input and output, where the result is approximately a linear behaviour.

A modification of the output membership functions so that they do not overlap will cause the characteristic to become close to that of Figure 17.2.5, compare Figure 17.2.6. Therefore one can establish the fact that the degree of overlap in the input membership functions has a strong influence on the static characteristic of a fuzzy controller. While small overlaps in the input membership functions generate step characteristics, with a higher degree of overlap the curves become smoother. The influence of overlap in the output membership functions has less effect on the characteristic. For a reduction of the support of the output membership functions the characteristic of Figure 17.2.7 is obtained which does not differ significantly from that of Figure 17.2.5.

The size of the individual output membership function has a strong influence on the characteristic. Figure 17.2.8 shows the case for a very small support of the output membership function AZ, which generates an S-type characteristic with a high gain at the origin. Widening the support of the membership function AZ inverts the S-curve with a small gain at the origin, as shown in Figure 17.2.9. Thus the form of the characteristic depends strongly on the support of the individual output membership function.

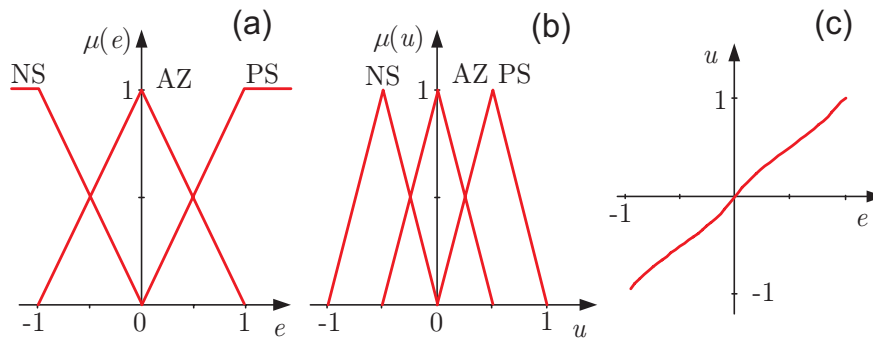


Figure 17.2.5: Influence on the (c) characteristic of a proportional fuzzy controller with (a) full overlap in the input membership functions and (b) full overlap in the output membership functions

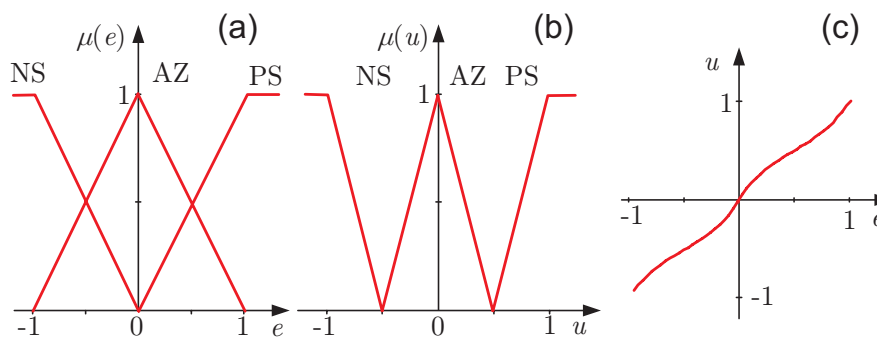


Figure 17.2.6: Influence on the (c) characteristic of a proportional fuzzy controller with (a) full overlap in the input membership functions and (b) without overlap in the output membership functions

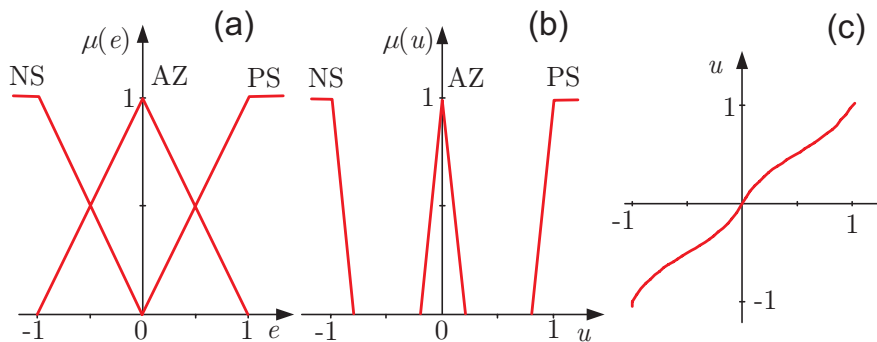


Figure 17.2.7: Influence on the (c) characteristic of a proportional fuzzy controller with (a) full overlap in the input membership functions and with (b) reduced support in the output membership functions

The effects of a modified rule base will be demonstrated by an example. The same full overlapping membership functions are used as in Figure 17.2.5a and b. A modified rule base of the form

- (1) IF $e = \text{NS}$ THEN $u = \text{PS}$,
- (2) IF $e = \text{AZ}$ THEN $u = \text{AZ}$,
- (3) IF $e = \text{PS}$ THEN $u = \text{NS}$,

will give a modulus-type of characteristic, as shown in Figure 17.2.10.

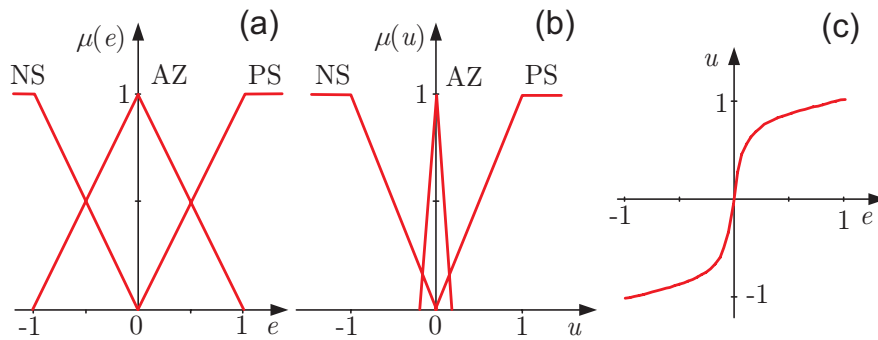


Figure 17.2.8: Influence on the (c) characteristic of a proportional fuzzy controller with (a) full overlap in the input membership functions and with (b) a small support in the output membership function AZ

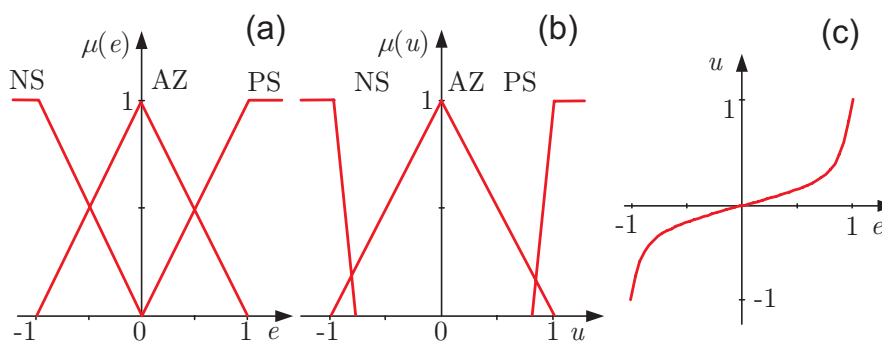


Figure 17.2.9: Influence on the (c) characteristic of a proportional fuzzy controller with (a) full overlap in the input membership functions and with (b) a large support in the output membership function AZ

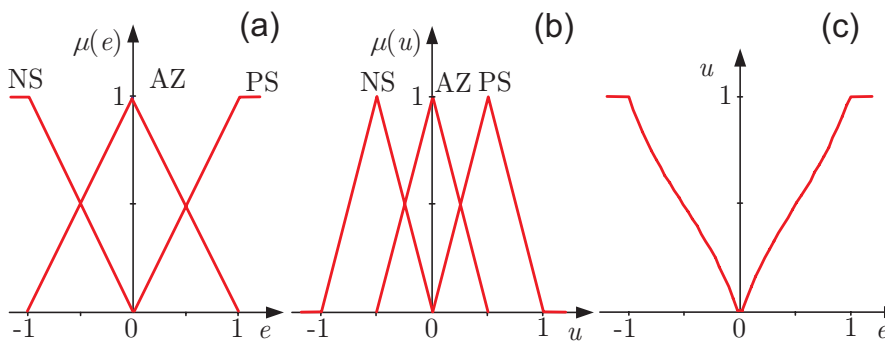


Figure 17.2.10: Influence on the rule base on the (c) characteristic of a proportional fuzzy controller with (a) full overlap in the input membership functions and (b) full overlap in the output membership functions

17.2.3 Representation using 3D characteristics

Up to now, fuzzy controllers with only one input and one output of the fuzzy system have been considered. The same arguments with respect to the degrees of freedom of a fuzzy system are also valid in the case of multiple inputs. A graphical representation of the characteristic is not as easy as in the 2D cases of section 17.2.2. Moreover, for the case of two inputs, a 3D representation is possible. For a fuzzy controller with a fuzzy system having two inputs e_1 and e_2 and one output u one gets a band of characteristics in a 2D discrete representation, or a 3D representation with the output over the two inputs, as shown in Figure 17.2.11 for a PD-type of fuzzy controller. For the case with more than two inputs, projections on

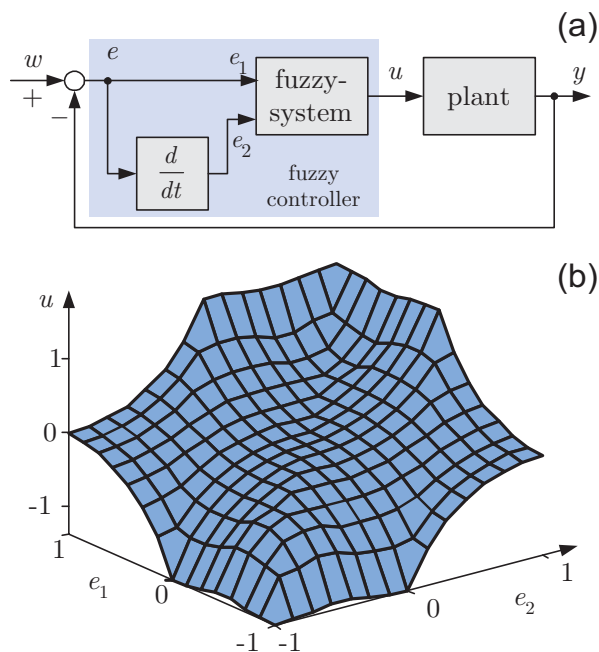


Figure 17.2.11: Control system with PD-type fuzzy controller: (a) block diagram and (b) 3D representation of the characteristics of the fuzzy system with $e_1 = e$ and $e_2 = \dot{e}$

the 3D space can be used to generate multiple 3D diagrams, but in general these representations have only a limited usefulness.

A fuzzy controller is typically a classical controller using nonlinear characteristics. But the design and parametrisation is entirely different.

17.3 Example of a fuzzy control system

In the following the design and functioning of a fuzzy control system will be presented using the example of the portal-type loading crane shown in Figure 17.3.1



Figure 17.3.1: View of a portal-type loading crane

17.3.1 Loading crane plant model

The schematic diagram in Figure 17.3.2 shows the principle of a loading crane, which consists of a crab moving on rails. The load (freight) hangs on a rope from the crab such that the rope and load together

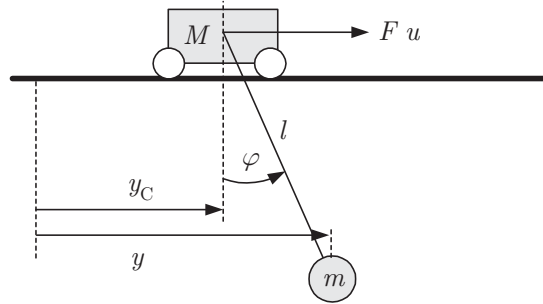


Figure 17.3.2: Schematic representation of the loading crane

can be treated as a pendulum. The load moves only in the plane that contains the direction of the rails. The force of the electrical drive, which moves the crab, is proportional to the control signal, fed into the drive system. The position of the crab and the rope (pendulum) angle are measured. The real manipulated variable to control the drive is limited to ± 10 V.

The plant dynamical behaviour can be described by the two coupled nonlinear differential equations

$$\ddot{y}_C = \frac{1}{M + m \sin^2 \varphi} [m \sin \varphi (g \cos \varphi + l \dot{\varphi}^2) + F u - S \operatorname{sgn}(\dot{y}_C) - D \dot{y}_C], \quad (17.3.1)$$

$$\ddot{\varphi} = \frac{-1}{l (M + m \sin^2 \varphi)} \quad (17.3.2)$$

$$\cdot [m \sin \varphi (g + l \dot{\varphi}^2 \cos \varphi) + M g \sin \varphi + \cos \varphi (F u - S \operatorname{sgn}(\dot{y}_C) - D \dot{y}_C)]$$

$$y = y_C + l \sin \varphi \quad (17.3.3)$$

with the following parameters and variables of the crane:

position of the load:	y ,
position of the crab:	y_C ,
pendulum angle:	φ ,
mass of the crab:	$M = 1000$ kg,
mass of the load:	$m = 20 \dots 1250$ kg,
pendulum length:	$l = 10 \dots 20$ m,
acceleration due to gravity:	$g = 9.80665$ m/s ² ,
drive coefficient:	$F = 1000$ N/V,
static friction force:	$S = 500$ N,
dynamical friction coefficient:	$D = 777$ kg/s,
experimental initial position:	$y_0 = -9$ m,
experimental target position:	$y_T = +9$ m.

The control task is to move the load from the initial position y_0 to the target position y_T such that the load does not swing at the target position and the transition goes smoothly with a minimum of oscillations and no overshoot.

17.3.2 Fuzzy control system design

In general, the design of a fuzzy controller is characterised by the fuzzy methodology as described above. For designing a fuzzy control system from scratch more or less heuristic methods are available. If no expert or operator is available, one cannot tackle the design problem without some information, typical a mathematical model, of the plant. The controller of the crane has been designed and tested using simulation studies by the following generic procedure:

1. Identification of the relevant input and output variables of the controller, i.e. choice of the linguistic variables,
2. setting of the possible ranges of the input and output values, i.e scaling of the linguistic variables,
3. definition of meaningful linguistic terms and their membership functions for each linguistic variable,
4. setting up the rule base, and
5. simulation of the closed loop if possible or testing at the plant site.

As the maximum number of rules R , considering all possible connective operations, increases strongly with the number of inputs n and membership functions m according to

$$R = m^n, \quad (17.3.4)$$

one has to try to keep it low. Hence, as a first approach a control structure according to Figure 17.2.11a is taken, and for the linguistic variables the control error $e = w - y$ and its derivative \dot{e} are chosen, each with the three membership functions N (negative), Z (zero) and P (positive) including the output. Thus from Eq. (17.3.4) 9 rules for a full rule base have to be specified. The control error is directly calculated as the difference between the set-point value w and the position of the load y , not of the crab. For determining the derivative \dot{e} a differentiating filter may be used. As the set point is changed only in steps when the pendulum has settled, the derivative \dot{e} is equal to the derivative $-\dot{y}$. Most crane systems measure the speed of the crab for the internal drive control. Therefore it is reasonable to avoid the differentiating filter and use the speed $-\dot{y}_C$ of the crab instead. The input filter is now reduced to gains for scaling signals such that they fit to the range of the linguistic variables. In order to compensate step disturbances, e.g. caused by static friction S , the output filter is provided with an integrator in parallel to the proportional channel.

This relatively simple fuzzy control system shows a reasonable behaviour, as shown in Figure 17.3.7a. The maximum amplitude of the oscillations of the load is about 7° , which is still relatively large. The control performance can be improved when additional information about the angle φ and its speed $\dot{\varphi}$ are exploited in the new structure shown in Figure 17.3.3. As $\dot{\varphi}$ is not measured, an input filter (DT_1 element) is necessary to generate this signal. Now due to the doubling of the number of inputs, the

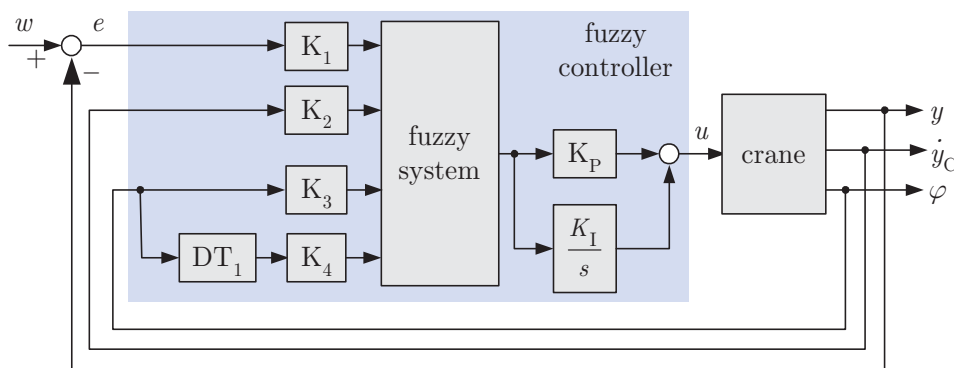


Figure 17.3.3: Structure of the loading crane fuzzy control system

problem of an efficient rule base arises. According to Eq. (17.3.4), 81 rules are now necessary if all inputs enter into all premises. It is thus difficult to formulate such a large number of useful rules. Therefore one can use a simpler approach of using two separate rule bases, the one for position control that has already been used for the linguistic variables of e and $-\dot{y}_C$ and the other for angle control for the linguistic variables of φ and $\dot{\varphi}$. Both rule bases are linked by the fuzzy union operation. The maximum number of rules is now reduced to 18 and can be further reduced to 14 by removing all unfired rules. Figure 17.3.4 shows the membership functions for the five linguistic variables. The rules are illustrated in tabular form in Figure 17.3.5. In Figure 17.3.6 two 3D characteristics of the fuzzy system are shown. The nonlinear behaviour can be clearly recognised.

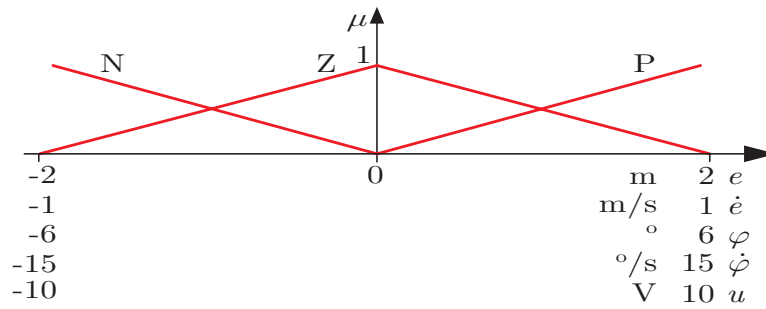


Figure 17.3.4: Triangular membership functions of the five linguistic variables

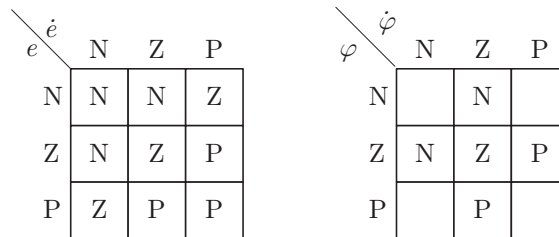


Figure 17.3.5: Rule bases for position and angle control

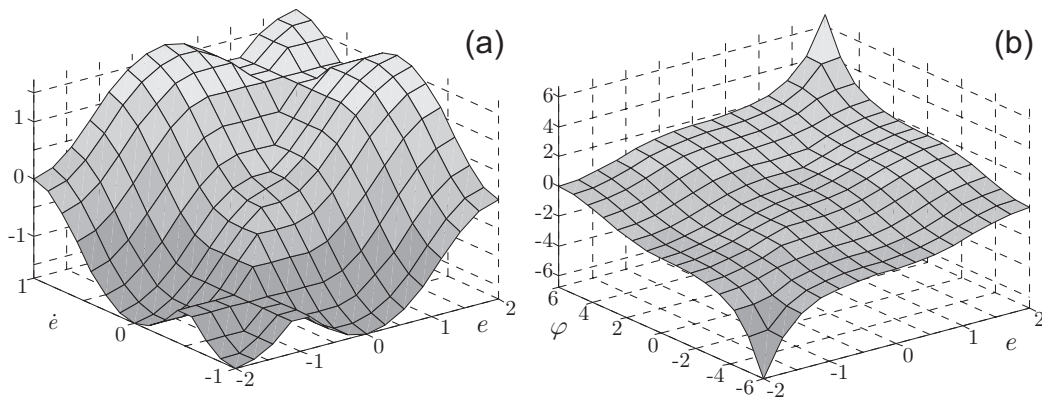


Figure 17.3.6: 3D characteristics of the fuzzy system (a) $u = u(e, \dot{e})$, (b) $u = u(e, \varphi)$

Figure 17.3.7 shows the results of the fuzzy control system when moving the load from y_0 to y_T . The input signals are scaled such that they use almost the full range of the linguistic variables. The scaling factors from Figure 17.3.3 are $K_1 = 0.1$, $K_2 = -0.5$, $K_3 = 1$ and $K_4 = 6$. The parameters of the output filter are $K_P = 2$ and $K_I = 1.4$. The additional angle feedback damps the oscillations to about a half the number with pure position control, as shown in Figure 17.3.7b. The control system shows robustness when the length of the rope is doubled as shown in Figure 17.3.7c.

17.4 Contribution of fuzzy control

In many real-life situations, there is a need to automatically control a system such as a car, a chemical reactor or a crane. In some situations, a reasonable model of the controlled object and how it will react to different controls are known. Also one can often describe precisely the objective of the control – usually, to maximise or minimise a certain characteristic such as the plant’s output, say in response to a system

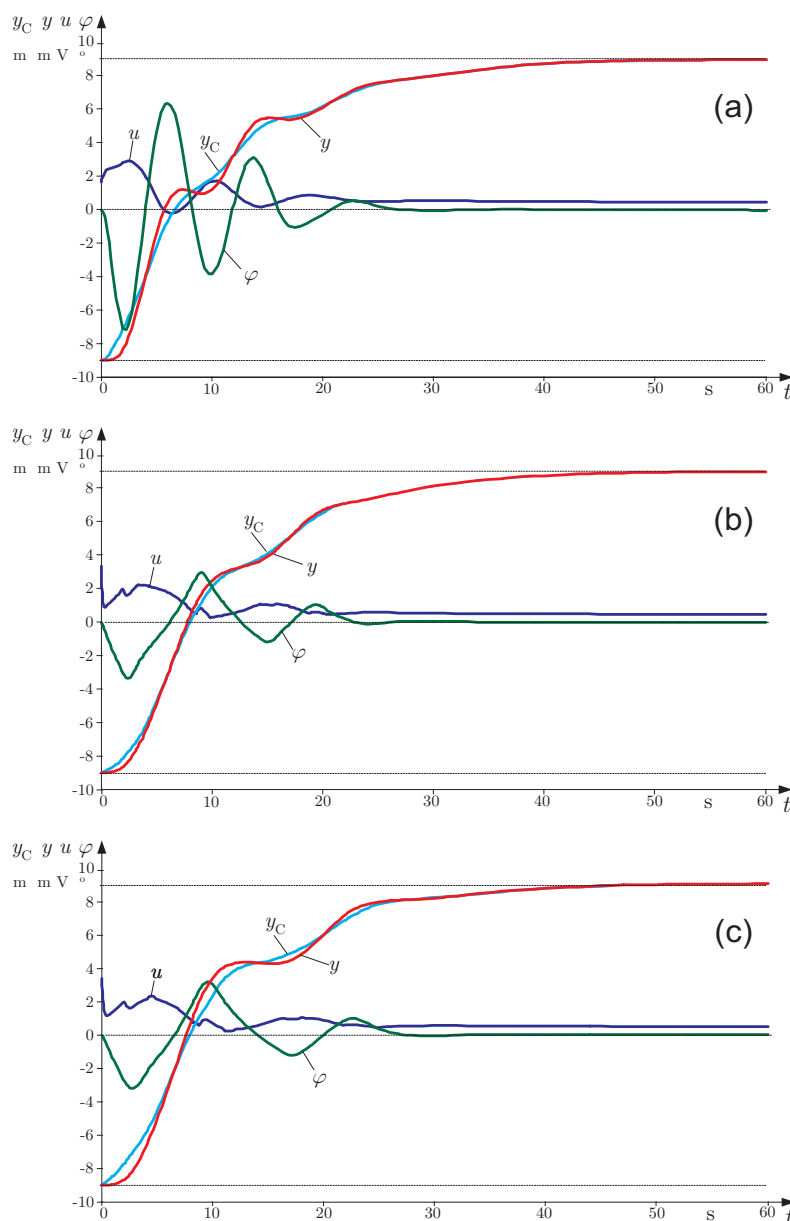


Figure 17.3.7: Responses of the portal-type loading crane with fuzzy control from initial to target position for load $m = 1000$ kg: (a) without angle feedback, $l = 10$ m, (b) with angle feedback, $l = 10$ m, (c) with angle feedback, $l = 20$ m

disturbance can be defined. In such cases, the search for the optimal control strategy can be reformulated as a precise mathematical problem. Thus, in many real-life cases one can explicitly solve the optimisation problem and thus find the desired control.

In many other situations, however, one does not have a good description of the controlled system, or a good model is available, but the corresponding optimisation problem is too difficult to solve. In such situations, one may have the expertise of skilled operators who have the experience of controlling the system. For example, the experience of drivers who control cars, the experience of chemical engineers who successfully control chemical plants and the experience of crane operators who successfully operate their cranes. It may be appropriate to transform this expert experience into an automatic control strategy.

It is often difficult to come up with such a transformation, because expert operators are often unable to describe their experience in precise terms. Instead they describe their control by using words of natural

language, which do not have a precise meaning and are, in this sense, *fuzzy*. There is thus a need for a methodology that translates such fuzzy rules into a precise control strategy. Fuzzy control is such a methodology.

Fuzzy control is successful in many real-life problems in which traditional control methods fails or at least is not so successful. Does this mean that traditional control as shown in the other chapters is the thing of the past and only fuzzy methods should be used? Of course not, fuzzy control has its limitations too.

The main limitation of fuzzy control is that it is applicable only in the situation of uncertainty, when there is the complete knowledge about the controlled system not available. Fuzzy control is therefore good but not optimal. Often, as one gains more and more experience of controlling the system, one gets a better and better understanding of how the system works. Eventually, this understanding leads to a precise description of the system, which allows to find an optimal control, which is better than any other control and in particular, which is better than a fuzzy control strategy.

From this viewpoint, fuzzy control is a temporary phenomenon. This does not mean when our knowledge grows that fuzzy control will be used less and less. As one obtains more and more knowledge about the system that is controlled for a long time, new systems and objects attract our attention, and one needs to be able to control them. For example, the car manufacturer finds a precise description of a certain type of motor and the engineers learn how to optimally control the motor of this type, but when new improved motors appear, for these new devices, the engineers do not have the exact model, and thus, they have to use fuzzy control or similar techniques. As the progress intensifies, more and more new objects and systems appear that have to be controlled, and therefore, the relative use of fuzzy control increases. Fuzzy control methodology has its limitations, but it does not have limits.

Appendix A

Mathematical and table appendix

Module units

A.1 The Laplace transform	A-2
A.1.1 Convergence	A-2
A.1.2 The inverse Laplace transform	A-2
A.1.3 Main theorems of the Laplace transform	A-3
A.1.3.1 Derivative theorem	A-3
A.1.3.2 Integral theorem	A-3
A.1.3.3 Convolution in the time domain	A-3
A.1.3.4 Convolution in the frequency domain	A-4
A.1.3.5 Initial value theorem	A-5
A.1.3.6 Final value theorem	A-5
A.2 The complex G-plane	A-6
A.3 Detailed analysis of 2nd-order lag elements	A-7
A.3.1 Determining resonances of 2nd-order lag elements	A-7
A.3.2 Poles and step responses of 2nd-order lag elements	A-8
A.4 The law of Bode and the Hilbert transformation	A-10
A.5 Stability considerations using the weighting function	A-11
A.6 Equivalence of the Hurwitz and Routh criteria	A-11
A.7 Determination of J_{ISE} using determinants	A-12
A.8 Tables	A-15

Appendix overview. *In the text, it is assumed that one does not need to read all details about theorems, mathematical derivations and background information. Most comprehensive tables would disturb the flow of reading the text. Therefore, all this information is moved into this appendix. The beginning is more a mathematical appendix, and in the last part all large tables can be found.*

A.1 The Laplace transform

A.1.1 Convergence

With respect to the *range of convergence* of the Laplace integral

$$F(s) = \int_0^{\infty} f(t) e^{-st} dt, \quad (\text{A.1.1})$$

now the following considerations are taken:

If the function to be transformed $f(t)$ is stepwise continuous and if there are real numbers α and σ' such that for all $t \geq 0$

$$|f(t)| < \alpha e^{\sigma' t}$$

is valid, then the Laplace integral will converge for all s with $\text{Re } s > \sigma'$. Particularly, if for σ' the smallest possible value σ_0 is taken, the condition $\text{Re } s > \sigma_0$ is the smallest possible convergence range. Consequently the Laplace integral exists only within some part of the complex s plane. This part is called convergence area, as shown in Figure A.1.1. The variable σ_0 is called the abscissa of convergence. For

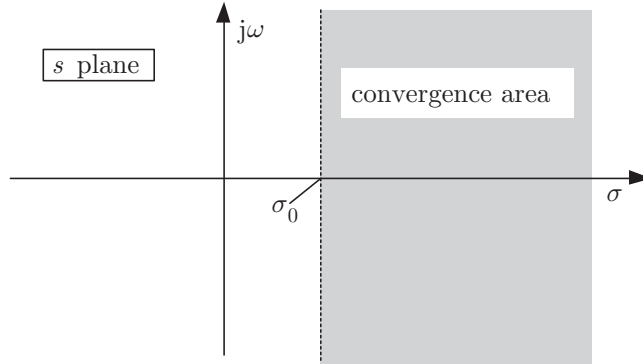


Figure A.1.1: Convergence area of the Laplace integral

values of s with $\text{Re } s > \sigma_0$ Eq. (A.1.1) makes no sense. Thus for $\sigma > \sigma_0$ the limit value of $f(t) e^{-\sigma t}$ for $t \rightarrow \infty$ must go to zero, but not for $\sigma < \sigma_0$.

Example A.1.1

$$f(t) = t^n$$

The Laplace integral converges for all s with $\sigma > 0$, as $e^{-\sigma t}$ for $t \rightarrow \infty$ decreases faster than any power of t increases. ■

A.1.2 The inverse Laplace transform

The so called back transformation or inverse Laplace transformation, i.e. the determination of the original function from the mapped function, is given by the *inverse integral*

$$f(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(s) e^{st} ds \quad t > 0, \quad (\text{A.1.2})$$

where $f(t) = 0$ for $t < 0$ is valid. The variable c must be chosen such that the path of integration is in the convergence area along a line parallel to the imaginary axis at distance c from it, where c must be larger than the real parts of all singular values of $F(s)$.

It must be observed that Eq. (A.1.2) at a step location $t = t_s$ delivers the arithmetic mean value of the left and right limits $[f(t_s+) + f(t_s-)]/2$, particularly for $t = 0$ at the origin, as $[f(0+) + f(0-)]/2 = f(0+)/2$.

A.1.3 Main theorems of the Laplace transform

A.1.3.1 Derivative theorem

For a causal function of time $f(t)$, for which the derivative for $t > 0$ exists, and take into account any steps at $t = 0$, the value of $0+$ will be chosen for the lower limit of integration for Eq. (A.1.1). This is necessary to eliminate the case $t = 0$ from the integration interval. This has no influence on the value of the integral as far as we restrict on classical functions (no distributions). So one obtains by partial integration

$$\mathcal{L} \left\{ \frac{df(t)}{dt} \right\} = \int_{0+}^{\infty} e^{-st} \frac{df(t)}{dt} dt = [e^{-st} f(t)] \Big|_{0+}^{\infty} + s \int_{0+}^{\infty} e^{-st} f(t) dt$$

or

$$\mathcal{L} \left\{ \frac{df(t)}{dt} \right\} = sF(s) - f(0+). \quad (\text{A.1.3})$$

In the case of multiple differentiation it follows that

$$\mathcal{L} \left\{ \frac{d^n f(t)}{dt^n} \right\} = s^n F(s) - \sum_{i=1}^n s^{n-i} \frac{d^{(i-1)} f(t)}{dt^{(i-1)}} \Big|_{t=0+}. \quad (\text{A.1.4})$$

A.1.3.2 Integral theorem

From

$$\mathcal{L} \left\{ \int_0^t f(\tau) d\tau \right\} = \int_0^{\infty} \int_0^t f(\tau) d\tau e^{-st} dt$$

one obtains by partial integration

$$\begin{aligned} \mathcal{L} \left\{ \int_0^t f(\tau) d\tau \right\} &= -\frac{1}{s} \left[\int_0^t f(\tau) d\tau e^{-st} \right]_0^{\infty} + \frac{1}{s} \int_0^{\infty} f(t) e^{-st} dt \\ &= \frac{1}{s} \int_0^{\infty} f(t) e^{-st} dt \\ \mathcal{L} \left\{ \int_0^t f(\tau) d\tau \right\} &= \frac{1}{s} F(s). \end{aligned} \quad (\text{A.1.5})$$

A.1.3.3 Convolution in the time domain

For the convolution of two functions of time $f_1(t)$ and $f_2(t)$

$$f_1(t) * f_2(t) = \int_0^t f_1(\tau) f_2(t - \tau) d\tau. \quad (\text{A.1.6})$$

It can easily be shown by permutation of the variables that the convolution is a symmetrical operation, so that

$$f_1(t) * f_2(t) = f_2(t) * f_1(t)$$

or

$$\int_0^t f_1(\tau) f_2(t - \tau) d\tau = \int_0^t f_2(\tau) f_1(t - \tau) d\tau.$$

In the following it will be shown that the convolution of two original functions corresponds to multiplication of the related mapped functions, that is

$$\mathcal{L} \{f_1(t) * f_2(t)\} = F_1(s) F_2(s). \quad (\text{A.1.7})$$

The Laplace transform of Eq. (A.1.6) is given by

$$\begin{aligned} \mathcal{L} \{f_1(t) * f_2(t)\} &= \mathcal{L} \left\{ \int_0^t f_1(\tau) f_2(t - \tau) d\tau \right\} \\ &= \int_{t=0}^{\infty} \int_{\tau=0}^t e^{-st} f_1(\tau) f_2(t - \tau) d\tau dt. \end{aligned}$$

Substituting $\sigma = t - \tau$ and $d\sigma = dt$, respectively and using the valid extension of the upper bounds of integration to $\tau \rightarrow \infty$ yields

$$\mathcal{L} \{f_1(t) * f_2(t)\} = \int_{\sigma=-\tau}^{\infty} \int_{\tau=0}^{\infty} e^{-s(\tau+\sigma)} f_1(\tau) f_2(\sigma) d\tau d\sigma.$$

As both functions $f_1(t)$ and $f_2(t)$ have zero values for $t < 0$, it follows with respect to the lower limit of integration that

$$\mathcal{L} \{f_1(t) * f_2(t)\} = \int_0^{\infty} e^{-s\tau} f_1(\tau) d\tau \int_0^{\infty} e^{-s\sigma} f_2(\sigma) d\sigma.$$

The right-hand side of this equation is just the product $F_1(s)F_2(s)$.

A.1.3.4 Convolution in the frequency domain

Whereas in section A.1.3.3 the convolution of two functions of time was the focus of interest, here the convolution of two functions in the frequency domain is of concern and it can be shown that

$$\mathcal{L} \{f_1(t) f_2(t)\} = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F_1(p) F_2(s - p) dp. \quad (\text{A.1.8})$$

Here $F_1(s) \bullet \circ f_1(t)$ and $F_2(s) \bullet \circ f_2(t)$ is valid. Furthermore, p is the complex variable of integration. According to this theorem the Laplace transform of the product of two functions of time is equal to the convolution of $F_1(s)$ and $F_2(s)$ in the mapped domain.

For the product of two causal functions of time

$$f(t) = f_1(t) f_2(t) \quad (\text{A.1.9})$$

with Laplace transforms $F_1(s)$ and $F_2(s)$ and areas of convergence $\text{Re } s > \sigma_1$ and $\text{Re } s > \sigma_2$, respectively, the expression

$$\mathcal{L} \{f(t)\} = F(s) = \int_0^{\infty} f_1(t) f_2(t) e^{-st} dt \quad (\text{A.1.10})$$

follows after taking the Laplace transform of $f(t)$. Using the inverse integral according to Eq. (A.1.2)

$$f_1(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F_1(p) e^{pt} dp \quad c > \sigma_1 \quad (\text{A.1.11})$$

and by substituting this relationship into Eq. (A.1.10) it follows that

$$F(s) = \int_0^{\infty} f_2(t) e^{-st} \left[\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F_1(p) e^{pt} dp \right] dt. \quad (\text{A.1.12})$$

Permuting the sequence of integration (as far as the integrals fulfil the conditions of convergence) one obtains

$$F(s) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F_1(p) dp \int_0^{\infty} f_2(t) e^{-(s-p)t} dt, \quad (\text{A.1.13})$$

where for the second integral one can make the substitution

$$F_2(s-p) = \int_0^{\infty} f_2(t) e^{-(s-p)t} dt. \quad (\text{A.1.14})$$

This integral converges for $\text{Re}(s-p) > \sigma_2$. By substituting Eq. (A.1.14) into Eq. (A.1.13) the validity of Eq. (A.1.8) is shown.

A.1.3.5 Initial value theorem

It is required to show that

$$f(0+) = \lim_{t \rightarrow 0+} f(t) = \lim_{s \rightarrow \infty} s F(s). \quad (\text{A.1.15})$$

One has that

$$\mathcal{L} \{ \dot{f}(t) \} = \int_{0+}^{\infty} \dot{f}(t) e^{-st} dt = s F(s) - f(0+)$$

which as $s \rightarrow \infty$ can be written

$$\lim_{s \rightarrow \infty} \int_{0+}^{\infty} \dot{f}(t) e^{-st} dt = \lim_{s \rightarrow \infty} [s F(s) - f(0+)].$$

As the integration is independent of s , the calculation of the limit and the integration can be permuted provided that the integral converges uniformly. If $\mathcal{L} [f(t)]$ exists, then

$$\lim_{s \rightarrow \infty} \dot{f}(t) e^{-st} = 0$$

is valid. Therefore one gets

$$\lim_{s \rightarrow \infty} s F(s) = f(0+).$$

A.1.3.6 Final value theorem

The final value theorem states that

$$f(\infty) = \lim_{t \rightarrow \infty} f(t) = \lim_{s \rightarrow 0} s F(s). \quad (\text{A.1.16})$$

To prove this one evaluates the limit

$$\lim_{s \rightarrow 0} \int_{0+}^{\infty} \dot{f}(t) e^{-st} dt = \lim_{s \rightarrow 0} [s F(s) - f(0+)].$$

Again one can permute the sequence of determining the limit and the integration provided the integral converges. The result is

$$\int_{0+}^{\infty} \dot{f}(t) dt = \lim_{s \rightarrow 0} [s F(s) - f(0+)],$$

and after integration it follows that

$$\begin{aligned} f(\infty) - f(0+) &= \lim_{s \rightarrow 0} [s F(s) - f(0+)] \\ f(\infty) &= \lim_{s \rightarrow 0} s F(s). \end{aligned}$$

A.2 The complex G -plane

The complex transfer function $G(s)$ describes a local conformal mapping of the s plane to the G plane. Because of the preservation of the angles in this transformation the orthogonal grid of lines parallel to the axes $\sigma = \text{const}$ and $\omega = \text{const}$ of the s plane will be mapped into an orthogonal, but warped net of curves in the G plane, as shown in Figure A.2.1.

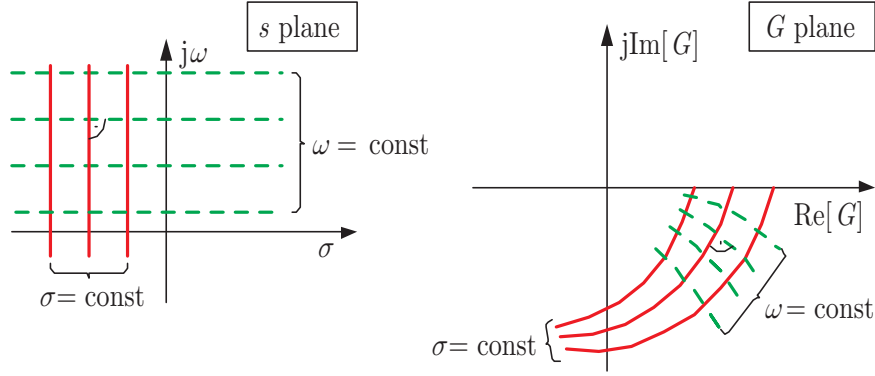


Figure A.2.1: Local conformal mapping of the lines $\sigma = \text{const}$ and $\omega = \text{const}$ of the s plane into the G plane (generalised Nyquist plot)

This mapping property will be discussed in the following using the simple *example* of a 1st-order transfer function

$$G(s) = \frac{K}{1 + sT}. \quad (\text{A.2.1})$$

For $s = \sigma + j\omega$ one obtains from Eq. (A.2.1)

$$G(\sigma + j\omega) = \frac{K}{1 + \sigma T + j\omega T} = K \frac{1 + \sigma T - j\omega T}{(1 + \sigma T)^2 + \omega^2 T^2}.$$

From this it follows that for the real and imaginary parts of $G(s)$

$$\text{Re}\{G(s)\} = K \frac{1 + \sigma T}{(1 + \sigma T)^2 + \omega^2 T^2}, \quad (\text{A.2.2a})$$

$$\text{Im}\{G(s)\} = K \frac{\omega T}{(1 + \sigma T)^2 + \omega^2 T^2}. \quad (\text{A.2.2b})$$

For mapping the following two cases are treated:

a) *Mapping of the lines $\sigma = \text{const}$*

From ω from Eqs. (A.2.2a) and (A.2.2b) one obtains

$$\omega T = -(1 + \sigma T) \frac{\text{Im}\{G(s)\}}{\text{Re}\{G(s)\}} \quad (\text{A.2.3})$$

and substituting Eq. (A.2.3) in Eq. (A.2.2a) shows, after a simple rearrangement, that

$$\left[\text{Re}\{G(s)\} - \frac{K}{2(1 + \sigma T)} \right]^2 + \text{Im}^2\{G(s)\} = \left[\frac{K}{2(1 + \sigma T)} \right]^2. \quad (\text{A.2.4})$$

This relationship represents for the variables $\text{Re}\{G(s)\}$ and $\text{Im}\{G(s)\}$ the equation of a band of circles with the parameter σ . The centres of these circles are on the real axis of the G plane at $K/[2(1 + \sigma T)]$. The radii are $K/[2(1 + \sigma T)]$. For $\omega \geq 0$, $K > 0$ and $T > 0$ the lines $\sigma = \text{const}$ map into semicircles in the lower G plane, as Figure A.2.2 shows. The semicircles for $\sigma = \text{const}$ are parameterised by the values of ω . The circles start with $\omega = 0$ on the real G axis and end for $\omega \rightarrow \infty$ at the origin of the G plane.

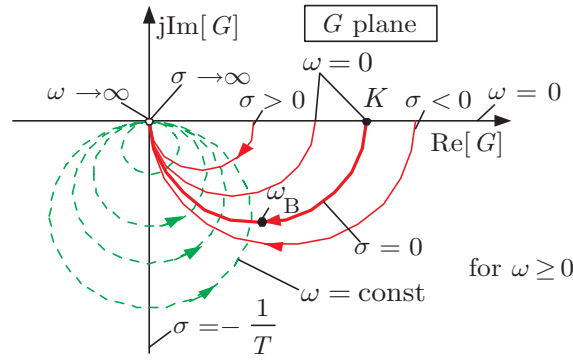


Figure A.2.2: Conform mapping of the upper s plane ($\omega > 0$) into the G plane for the example $G(s) = K/(1 + sT)$

A very important case is the semicircle with the parameter $\sigma = 0$. It represents the conformal mapping of the positive imaginary axis of the s plane and is called the *frequency response locus*, $G(j\omega)$, for the system. This semicircle starts for $\omega = 0$ with a value of K on the positive real axis of the G plane and it has for $|\text{Re}\{G(j\omega)\}| = |\text{Im}\{G(j\omega)\}|$ the frequency $\omega = \omega_B = 1/T$, which is also called the *breakpoint frequency*.

From Eq. (A.2.4) it is obvious, that for $\sigma > 0$ the radius of the semicircle will decrease until it becomes zero for $\sigma \rightarrow \infty$ and the final semicircle will coincide with the origin of the G plane. For $\sigma < 0$, however, the radius will increase to infinity for $\sigma = -1/T$, and the semicircle will degenerate to the negative imaginary axis of the G plane.

- *Mapping of the lines $\omega = \text{const}$*

If Eq. (A.2.3) is solved for σT then

$$\sigma T = - \left(1 + \omega T \frac{\text{Re}\{G(s)\}}{\text{Im}\{G(s)\}} \right) \quad (\text{A.2.5})$$

and using this in Eq. (A.2.2b), one obtains by elementary manipulations

$$\left[\text{Im}\{G(s)\} + \frac{K}{2\omega T} \right]^2 + \text{Re}^2\{G(s)\} = \left[\frac{K}{2\omega T} \right]^2. \quad (\text{A.2.6})$$

This relationship also represents a band of circles, but for the parameter ω . The centres of the circles for $\omega \geq 0$ are on the negative imaginary axis at $-K/(2\omega T)$ and because of the radii of size $K/(2\omega T)$ they pass to the origin of the G plane. For $\omega = 0$ the radius will be infinite and the circle degenerates to a line that is the real axis of the G plane. For $\omega \rightarrow \infty$ the radius shrinks to zero and the circle degenerates to the origin of the G plane. It can be easily shown that both bands of circles according to Eqs. (A.2.4) and (A.2.6) are orthogonal.

The transfer function $G(s) = K/(1 + sT)$ belongs to a special class of local conformal mappings, the so called linear mappings. A mapping, described by the equation $G(s) = (As + B)/(Cs + D)$ always maps the circles in the s plane to circles in the G plane. Here lines are treated as a special case of circles. Introducing the complex G plane one gets for the special case $\sigma = 0$ of $G(s)$ the frequency response locus $G(j\omega)$. The description of systems using frequency responses $G(j\omega)$ in the G plane is of high importance for practical applications as the frequency response is a directly measurable description of a dynamical system.

A.3 Detailed analysis of 2nd-order lag elements

A.3.1 Determining resonances of 2nd-order lag elements

From Eq. (4.4.38) the maximum of the magnitude $A(\omega)_{\max} = A(\omega_p) = M_r$ and the resonant peak frequency ω_p can be simply determined. $A(\omega)$ is at a maximum, when the denominator of Eq. (4.4.38),

i.e.

$$a(\omega) = \left[1 - \left(\frac{\omega}{\omega_0} \right)^2 \right]^2 + \left(2\zeta \frac{\omega}{\omega_0} \right)^2,$$

is at minimum. Setting the derivative of 1st order to zero one obtains for $\omega = \omega_p$

$$\left. \frac{da(\omega)}{d\omega} \right|_{\omega_p} = 0 = -1 + \left(\frac{\omega_p}{\omega_0} \right)^2 + 2\zeta^2.$$

From this the resonant peak frequency is

$$\omega_p = \omega_0 \sqrt{1 - 2\zeta^2} \quad \left(\text{for } \zeta < \frac{1}{\sqrt{2}} \right), \quad (\text{A.3.1})$$

and the maximum of the amplitude for $K = 1$ is

$$M_r = \frac{1}{2\zeta \sqrt{1 - \zeta^2}}. \quad (\text{A.3.2})$$

From Eq. (A.3.1) it follows that a maximum only exists for $(1 - 2\zeta^2) > 0$, which gives $\zeta < 1/\sqrt{2}$. For $\zeta = 1/\sqrt{2} = 0.707$, $\omega_p = 0$ is valid and $M_r = 1$ or $M_{r\text{dB}} = 0$. For $\zeta = 0$, $\omega_p = \omega_0$ and $M_r = \infty$ is valid.

A.3.2 Poles and step responses of 2nd-order lag elements

From the characteristic equation of the 2nd-order lag element

$$P(s) \equiv D(s) = 1 + \frac{2\zeta}{\omega_0}s + \frac{1}{\omega_0^2}s^2 = 0 \quad (\text{A.3.3})$$

one obtains the poles of the transfer function

$$s_{1,2} = -\omega_0\zeta \pm \omega_0\sqrt{\zeta^2 - 1}. \quad (\text{A.3.4})$$

The oscillating behaviour of a 2nd-order lag element depends on the position of the poles in the s plane and this is discussed in the following sections:

a) **Case I:** $0 < \zeta < 1$ (*oscillating behaviour: PT₂S element*)

For this case Eq. (A.3.4) gives a conjugate complex pair of poles

$$s_{1,2} = -\omega_0\zeta \pm j\omega_0\sqrt{1 - \zeta^2}. \quad (\text{A.3.5})$$

The corresponding step response is obtained from

$$H(s) = \frac{K}{1 + 2\frac{\zeta}{\omega_0}s + \frac{1}{\omega_0^2}s^2} \frac{1}{s} \quad (\text{A.3.6})$$

which can be written

$$H(s) = \frac{K\omega_0^2}{(s - s_1)(s - s_2)s}. \quad (\text{A.3.7})$$

Writing in partial fractions and finding the inverse Laplace transform gives

$$h(t) = K \left\{ 1 - e^{-t\omega_0\zeta} \left[\cos(\omega_0\sqrt{1 - \zeta^2}t) + \frac{\zeta}{\sqrt{1 - \zeta^2}} \sin(\omega_0\sqrt{1 - \zeta^2}t) \right] \right\} \sigma(t). \quad (\text{A.3.8})$$

The decay of the oscillations is influenced by the quantity $T_A = 1/(\omega_0\zeta)$, which is therefore called the *decay time constant*. From the position of the poles s_1 and s_2 of $G(s)$ the time constant T_A can be found. The ratio

$$\rho = \left| \frac{\text{Re}(s_i)}{\text{Im}(s_i)} \right| \quad \text{for } i = 1 \text{ or } 2,$$

or

$$\rho = \frac{\zeta}{\sqrt{1-\zeta^2}} = \tan \varphi_d. \quad (\text{A.3.9})$$

is defined as the *relative damping ratio* of the oscillation.

If a certain value of ζ is given then the angle φ_d between a line from the origin to a complex pole and the imaginary axis in the s plane is fixed. Also, from the damped oscillation of the step response $h(t)$, which has the damped natural frequency of

$$\omega_d = \omega_0 \sqrt{1-\zeta^2} < \omega_0, \quad (\text{A.3.10})$$

then from Eq. (A.3.8) the ratio of a consecutive overshoot to overshoot or vice versa (see Table 4.4.2)

$$\frac{h_{n+1/2}}{h_n} = e^{-\pi \frac{\zeta}{\sqrt{1-\zeta^2}}}$$

can be determined, and from this one obtains the damping ratio as

$$\zeta = \frac{\ln \frac{h_n}{h_{n+1/2}}}{\sqrt{\pi^2 + \left[\ln \frac{h_n}{h_{n+1/2}} \right]^2}}. \quad (\text{A.3.11})$$

b) **Case 2:** $\zeta = 1$ (*critical damping: PT₂ element*)

For Eq. (A.3.4) the two poles of $G(s)$ are

$$s_{1,2} = -\omega_0.$$

This is a double pole on the negative real axis. If a time constant

$$T = \frac{1}{\omega_0} \geq 0,$$

is defined, one obtains the transfer function as

$$G(s) = \frac{K}{(1+Ts)(1+Ts)}, \quad (\text{A.3.12})$$

which is a series connection of two 1st-order lag elements with identical time constants. The step response follows from Eq. (A.3.8) with $\zeta = 1$ as

$$h(t) = K [1 - e^{-t\omega_0}(1 + \omega_0 t)] \sigma(t). \quad (\text{A.3.13})$$

c) **Case 3:** $\zeta > 1$ (*aperiodic behaviour: PT₂ element*)

In this case from Eq. (A.3.4) $G(s)$ has two negative real poles

$$s_{1,2} = -\omega_0 \zeta \pm \omega_0 \sqrt{\zeta^2 - 1}.$$

With the definition of the time constants

$$T_1 = -\frac{1}{s_1} \quad \text{and} \quad T_2 = -\frac{1}{s_2}$$

one obtains the transfer function as

$$G(s) = \frac{K}{(1+T_1s)(1+T_2s)}. \quad (\text{A.3.14})$$

Here the system is composed of two PT₁ elements with different time constants in series connection. This element also shows typical PT₂ behaviour. During the calculation of the step response $h(t)$

according to Eq. (A.3.8) the arguments of sin and cos will be complex. By applying the hyperbolic functions

$$\cos jx = \cosh x \quad \text{and} \quad \sin jx = j \sinh x$$

one obtains directly using Eq. (A.3.8)

$$h(t) = K \left\{ 1 - e^{-t\omega_0\zeta} \left[\cosh(\omega_0\sqrt{\zeta^2 - 1}t) + \frac{\zeta}{\sqrt{\zeta^2 - 1}} \sinh(\omega_0\sqrt{\zeta^2 - 1}t) \right] \right\} \sigma(t). \quad (\text{A.3.15})$$

d) **Case 4:** $\zeta = 0$ (*undamped behaviour: oscillating element*)

Eq. (A.3.4) gives for this case a pair of imaginary poles of $G(s)$ at

$$s_{1,2} = \pm j\omega_0.$$

With Eq. (4.4.36) and $\zeta = 0$ one obtains the transfer function

$$G(s) = \frac{K}{1 + \frac{1}{\omega_0^2}s^2} = K \frac{\omega_0^2}{\omega_0^2 + s^2}. \quad (\text{A.3.16})$$

The step response $h(t)$ is an undamped oscillation

$$h(t) = K(1 - \cos \omega_0 t) \sigma(t) \quad (\text{A.3.17})$$

of frequency ω_0 , which is usually called the natural frequency.

e) **Case 5:** $\zeta < 0$ (*unstable element*)

In this case both poles of $G(s)$ are in the right half s plane and may be real or conjugate complex:

$$s_{1,2} = \omega_0 |\zeta| \pm \omega_0 \sqrt{\zeta^2 - 1}. \quad (\text{A.3.18})$$

Eq. (A.3.8) is also valid for this case, but the exponential term has now a positive sign. Therefore the oscillations will not decay. The amplitudes of the oscillations in $h(t)$ will increase exponentially with time. This behaviour, for which $h(t)$ for $t \rightarrow \infty$ increases indefinitely, is defined as *unstable*.

A.4 The law of Bode and the Hilbert transformation

For a given amplitude response $A(\omega)$ there exists only one minimum phase system that realises this amplitude response. The phase $\varphi(\omega)$ of it can be determined for discrete frequencies ω_ν according to the *law of Bode*

$$\varphi(\omega_\nu) = \frac{2\omega_\nu}{\pi} \int_0^\infty \frac{\ln A(\omega) - \ln A(\omega_\nu)}{\omega^2 - \omega_\nu^2} d\omega. \quad (\text{A.4.1})$$

The connection between the real part $R(\omega) = \text{Re}\{G(j\omega)\}$ and the imaginary part $I(\omega) = \text{Im}\{G(j\omega)\}$ of the frequency response, which is valid both for minimum phase and non-minimum phase systems, is given by the *Hilbert transformation*

$$R(\omega_\nu) = R(\infty) - \frac{2}{\pi} \int_0^\infty \frac{\omega I(\omega)}{\omega^2 - \omega_\nu^2} d\omega \quad (\text{A.4.2a})$$

and

$$I(\omega_\nu) = \frac{2\omega_\nu}{\pi} \int_0^\infty \frac{R\omega(\omega)}{\omega^2 - \omega_\nu^2} d\omega. \quad (\text{A.4.2b})$$

From this it is obvious that with knowledge of only the real or the imaginary part, the imaginary part or the real part and therefore the frequency response $G(j\omega)$ can always be completely reconstructed.

A.5 Stability considerations using the weighting function

As a transfer function $G(s)$ is the Laplace transformed weighting function $g(t)$, the stability condition according to Eq. (5.2.1) can also be reformulated in terms of $G(s)$. In order to show this, the inverse Laplace transform from section 2.4 is used. If $G(s)$ is given as a rational fraction

$$G(s) = \frac{N(s)}{D(s)} = \frac{N(s)}{a_0 + a_1s + \dots + a_ns^n}, \quad (\text{A.5.1})$$

and $s_k = \sigma_k + j\omega_k$ are the poles of the transfer function $G(s)$, i.e. the roots of the denominator polynomial

$$D(s) = a_n(s - s_1)(s - s_2) \dots (s - s_n) = \sum_{i=0}^n a_i s^i, \quad (\text{A.5.2})$$

then according to Eq. (2.4.7) and Eq. (2.4.10) the weighting function

$$g(t) = \sum_{j=1}^{\nu} g_j(t) \quad (\text{A.5.3})$$

consists of $\nu \geq n$ terms

$$g_j(t) = c_j t^\mu e^{s_i t}, \quad \mu = 0, 1, 2, \dots, \quad j = 1, 2, \dots, \nu, \quad i = 1, 2, \dots, n.$$

In general c_j is a complex constant, and for multiple poles s_i of multiplicity r the exponent is $\mu = r - 1 > 0$. For the modulus of this function one obtains

$$|g_j(t)| = |c_j t^\mu e^{s_i t}| = |c_j| t^\mu e^{\sigma_i t}.$$

For $\sigma_i < 0$ the exponential function decays to zero for $t \rightarrow \infty$, and therefore so also does $|g_j(t)|$, even if $\mu > 0$, because the exponential function decays faster to zero than any other finite power of t increases.

By this consideration it is obvious, that Eq. (5.2.1) is only valid, if all poles of $G(s)$ have a negative real part. If the real part of only one pole is positive or of a multiple pole is zero, the weighting function grows with t beyond all limits.

A.6 Equivalence of the Hurwitz and Routh criteria

The validity of the Routh criterion can be verified easily by the equivalence with the Hurwitz criterion. From the coefficients of the first row of the Routh schema the connection with the Hurwitz determinants can be seen directly:

$$\begin{aligned} D_1 &= a_{n-1} \\ D_2 &= a_{n-1}b_{n-1} = D_1b_{n-1} \\ D_3 &= a_{n-1}b_{n-1}c_{n-1} = D_2c_{n-1} \\ &\vdots \\ D_n &= a_{n-1}b_{n-1}c_{n-1} \dots d_{n-1}e_{n-1}f_{n-1}g_{n-1} = D_{n-1}g_{n-1} \end{aligned}$$

The coefficients $b_{n-1}, c_{n-1} \dots$ in the first column of the Routh schema are just the quotients of consecutive Hurwitz determinants. When all Hurwitz determinants are positive, then their quotients and therefore also the coefficients in the first column of the Routh schema are positive. When the coefficients of the Routh schema are positive, then also, as $a_{n-1} = D_1$, all Hurwitz determinants are positive. Thus, the Routh criterion is equivalent to the Hurwitz criterion.

A.7 Determination of J_{ISE} using determinants

Writing the quadratic performance indices in terms of determinants is a general algebraic approach that is more suitable for a detailed analysis. Eq. (7.3.5) can be rewritten as

$$J_{\text{ISE}} = \frac{(-1)^{n+1} \det \mathbf{C}_n^b}{2a_n \det \mathbf{C}_n}. \quad (\text{A.7.1})$$

For odd n the matrix \mathbf{C}_n is

$$\mathbf{C}_n^{\text{odd}} = \begin{bmatrix} a_n & a_{n-2} & \cdots & a_3 & a_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & a_n & a_{n-2} & \cdots & a_3 & a_1 & 0 \\ \vdots & \cdots & 0 & a_{n-1} & a_{n-3} & \cdots & a_2 & a_0 \\ \vdots & \cdots & a_{n-1} & a_{n-3} & \cdots & a_2 & a_0 & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{n-1} & a_{n-3} & \cdots & a_2 & a_0 & 0 & \cdots & 0 \end{bmatrix} \quad (\text{A.7.2})$$

and for even n

$$\mathbf{C}_n^{\text{even}} = \begin{bmatrix} a_{n-1} & a_{n-3} & \cdots & a_3 & a_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & 0 & a_{n-1} & a_{n-3} & \cdots & a_3 & a_1 & 0 \\ \vdots & 0 & a_n & a_{n-2} & a_{n-4} & \cdots & a_2 & a_0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_n & a_{n-2} & a_{n-4} & \cdots & a_2 & a_0 & \cdots & 0 \end{bmatrix}. \quad (\text{A.7.3})$$

Both matrices consist in the upper part of a Toeplitz matrix with the odd coefficients and in the lower part of a Hankel matrix with the even coefficients of the denominator polynomial $A(s)$ of the control error in Eq. (7.3.4). The determinant of this matrix is nothing other than the Hurwitz determinant from Eq. (5.3.4). The other determinant in the numerator of Eq. (A.7.1) is from the same matrix as \mathbf{C}_n , but the coefficients in the last row in Eq. (A.7.2) and in the first row in Eq. (A.7.3) are exchanged for those of the polynomial

$$C(s) = B(s)B(-s) = c_0 + c_1s^2 + \cdots + c_{n-1}s^{2(n-1)}. \quad (\text{A.7.4})$$

Example A.7.1

Example 7.3.1 of determining the best damping ratio ζ for a second-order system is rewritten using determinants. For the control error from Eq. (7.3.9) we have the polynomial $C(s)$ from Eq. (A.7.4) as

$$C(s) = 4\zeta^2\omega_0^2 - s^2$$

and from Eqs. (7.3.9) and (A.7.3) follows

$$J_{\text{ISE}} = \frac{-1}{2} \frac{\det \begin{bmatrix} -1 & 4\zeta^2\omega_0^2 \\ 1 & \omega_0^2 \end{bmatrix}}{\det \begin{bmatrix} 2\zeta\omega_0 & 0 \\ 1 & \omega_0^2 \end{bmatrix}} = \frac{1 + 4\zeta^2}{4\zeta\omega_0}.$$

As this function is a parabola in ζ , with minimum given by

$$\frac{dJ_{\text{ISE}}}{d\zeta} = \frac{4\zeta^2 - 1}{4\zeta^2\omega_0} = 0$$

the minimum square error to a step input occurs for $\zeta = 0.5$. ■

The procedure shown in this example can be generalised for all quadratic performance indices. If the control error can be written as

$$E(s) = \frac{N(s)}{A(s) + kB(s)}, \quad (\text{A.7.5})$$

where k is a feedback gain to be chosen or for which the optimal value in the sense of a criterion according to Eq. (7.3.3) must be found, the performance index can be represented as

$$J_{\text{ISE}} = \frac{(-1)^{n+1}}{2(a_n + kb_n)} \frac{\det(\mathbf{C} + k\mathbf{D})}{\det(\mathbf{A} + k\mathbf{B})}. \quad (\text{A.7.6})$$

The matrices \mathbf{A} and \mathbf{B} correspond to the matrix given by Eq. (A.7.2) or (A.7.3), respectively, and \mathbf{C} to \mathbf{A} where the coefficients in the upper or lower row are exchanged for those of the polynomial

$$C(s) = N(s)N(-s) = c_0 + c_1s^2 + \dots + c_{n-1}s^{2(n-1)} \quad (\text{A.7.7})$$

and \mathbf{D} to \mathbf{B} where the coefficients in the upper or lower row are zeros. The determinants in Eq. (A.7.6) are polynomials in k . Therefore the performance index can be rewritten as

$$J_{\text{ISE}} = -\frac{Q(k)}{P(k)} \quad (\text{A.7.8})$$

or

$$1 + \frac{1}{J_{\text{ISE}}} \frac{Q(k)}{P(k)} = 0. \quad (\text{A.7.9})$$

As Eq. (A.7.9) corresponds directly to the characteristic equation (6.1.5a), the root-locus method can be applied for the analysis of the performance index in the k plane.

As we are interested in real values of k , only the behaviour on the real axis of the complex k plane is needed. The branches are calibrated by the inverse of J_{ISE} . The denominator in Eq. (A.7.6) or (A.7.8) is the highest Hurwitz determinant. Therefore the real poles (real zeros of $P(k)$) represent critical values of k , where the closed loop has poles on the imaginary axis and where $J_{\text{ISE}} \rightarrow \infty$. One can take the stability margins directly from the branches on the real axis in the k plane.

Now inspecting real breakaway or break-in points. If we apply rule 7 from section 6.2 on Eq. (A.7.9), from Eq. (6.2.7) we have the condition

$$\frac{d}{dk} \frac{Q(k)}{P(k)} = 0, \quad (\text{A.7.10})$$

which is the same as the condition $\frac{d}{dk} J_{\text{ISE}} = 0$ for an extremum of the performance index. Therefore real breakaway or break-in points in the k plane represent a minimum or maximum in the performance index and can be used to find optimal adjustments of parameters.

Example A.7.2

Consider the following plant transfer function:

$$G_P(s) = \frac{4 - 3s - 17s^3 + s^4}{4 + 17s + 28s^2 + 21s^3 + 7s^4 + s^5}$$

and the controller transfer function

$$G_C(s) = K_C \frac{1 + (1 + T_I)s + 2T_I s^2}{T_I s(1 + s)},$$

which contains two free parameters, the gain K_C and the time constant T_I . The control system should be optimised on step changes in the command or in the plant input using the performance index J_{ISE} .

For both inputs the control error can be described by Eq. (A.7.5) for $k = K_C$, where it differs only in the numerator polynomial $N(s)$. While determining the optimal gain by evaluation of the condition in Eq. (A.7.10) the other parameter T_I is held constant. In the same manner the stability margins can be determined by using $P(k) = 0$. For constant values of J_{ISE} one gets the gains according to Eq. (A.7.9).

Scanning over a given range of T_I and plotting the results will generate a combined stability and performance diagram as shown in Figure A.7.1. From these diagrams one can see that the optimal controller structure does not differ significantly for the two system inputs. ■

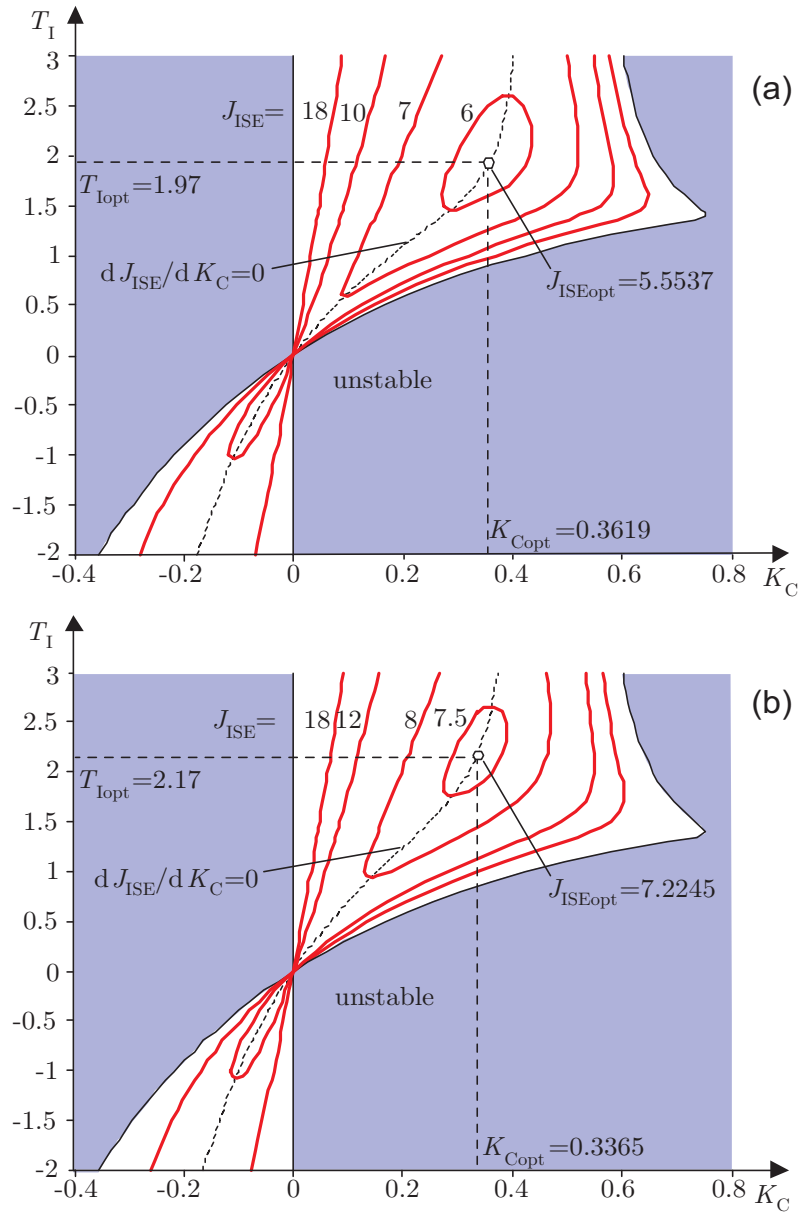


Figure A.7.1: Stability and performance diagram for step changes (a) in the command and (b) plant input

A.8 Tables

Table A.8.1: Step responses for a given pair of complex poles and a real pole with multiplicity k according to Eq. (10.2.4)

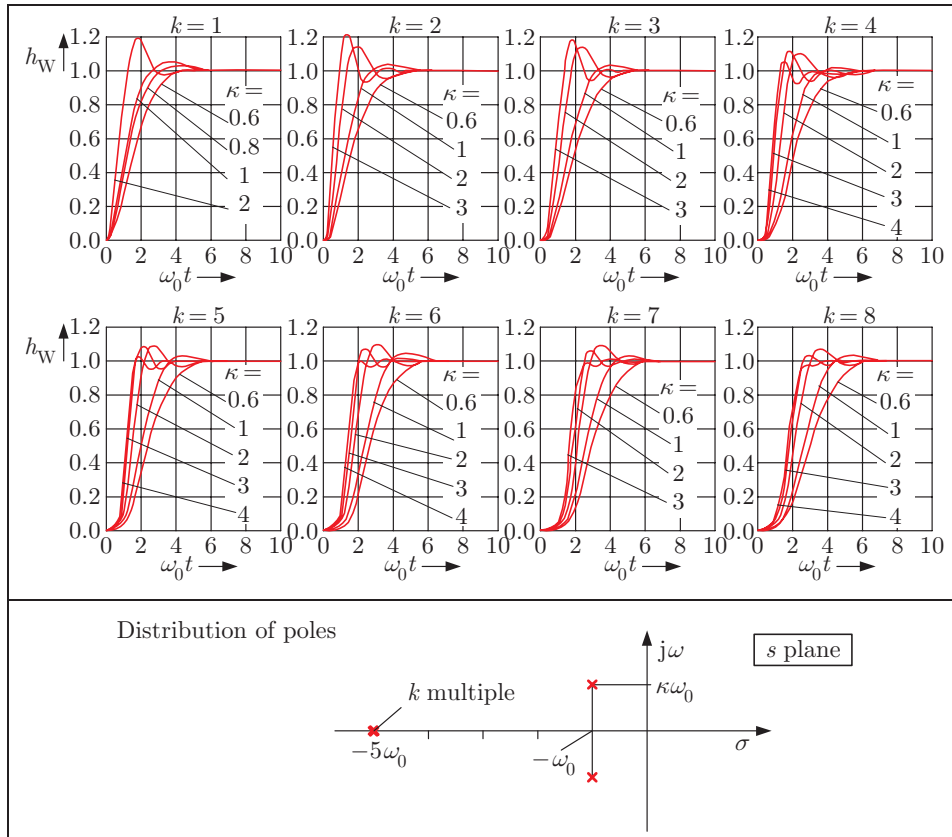


Table A.8.2: Standard polynomials $\beta(s)$ of Eq. (10.2.2) and corresponding step responses of different order u

	Standard polynomials $\beta(s)$ for $u = 1, 2, \dots$	Step response $h_W(\omega_0 t)$
Binomial form	$s + \omega_0$ $s^2 + 2\omega_0 s + \omega_0^2$ $s^3 + 3\omega_0 s^2 + 3\omega_0^2 s + \omega_0^3$ $s^4 + 4\omega_0 s^3 + 6\omega_0^2 s^2 + 4\omega_0^3 s + \omega_0^4$ $s^5 + 5\omega_0 s^4 + 10\omega_0^2 s^3 + 10\omega_0^3 s^2 + 5\omega_0^4 s + \omega_0^5$ $s^6 + 6\omega_0 s^5 + 15\omega_0^2 s^4 + 20\omega_0^3 s^3 + 15\omega_0^4 s^2 + 6\omega_0^5 s + \omega_0^6$	
Butterworth form	$s + \omega_0$ $s^2 + 1.4\omega_0 s + \omega_0^2$ $s^3 + 2\omega_0 s^2 + 2\omega_0^2 s + \omega_0^3$ $s^4 + 2.6\omega_0 s^3 + 3.4\omega_0^2 s^2 + 2.6\omega_0^3 s + \omega_0^4$ $s^5 + 3.24\omega_0 s^4 + 5.24\omega_0^2 s^3 + 5.24\omega_0^3 s^2 + 3.24\omega_0^4 s + \omega_0^5$ $s^6 + 3.86\omega_0 s^5 + 7.46\omega_0^2 s^4 + 9.14\omega_0^3 s^3 + 7.46\omega_0^4 s^2 + 3.86\omega_0^5 s + \omega_0^6$	

Table A.8.2 continued

	Standard polynomials $\beta(s)$ for $u = 1, 2, \dots$	Step response $h_W(\omega_0 t)$
JITAE form	$ \begin{aligned} & s + \omega_0 \\ & s^2 + 1.4\omega_0 s + \omega_0^2 \\ & s^3 + 1.75\omega_0 s^2 + 2.15\omega_0^2 s + \omega_0^3 \\ & s^4 + 2.1\omega_0 s^3 + 3.4\omega_0^2 s^2 + 2.7\omega_0^3 s + \omega_0^4 \\ & s^5 + 2.8\omega_0 s^4 + 5\omega_0^2 s^3 + 5.5\omega_0^3 s^2 + 3.4\omega_0^4 s + \omega_0^5 \\ & s^6 + 3.25\omega_0 s^5 + 6.6\omega_0^2 s^4 + 8.6\omega_0^3 s^3 + 7.45\omega_0^4 s^2 + 3.95\omega_0^5 s + \omega_0^6 \end{aligned} $	
$t_{5\%}$ settling time form	$ \begin{aligned} & s + \omega_0 \\ & s^2 + 1.4\omega_0 s + \omega_0^2 \\ & s^3 + 1.55\omega_0 s^2 + 2.1\omega_0^2 s + \omega_0^3 \\ & s^4 + 1.6\omega_0 s^3 + 3.15\omega_0^2 s^2 + 2.45\omega_0^3 s + \omega_0^4 \\ & s^5 + 1.575\omega_0 s^4 + 4.05\omega_0^2 s^3 + 4.1\omega_0^3 s^2 + 3.025\omega_0^4 s + \omega_0^5 \\ & s^6 + 1.45\omega_0 s^5 + 5.1\omega_0^2 s^4 + 5.3\omega_0^3 s^3 + 6.25\omega_0^4 s^2 + 3.425\omega_0^5 s + \omega_0^6 \end{aligned} $	

Table A.8.3: Standard elements in control engineering

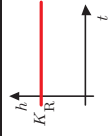
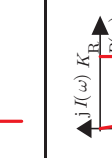
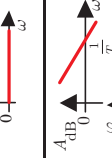
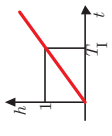
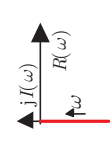
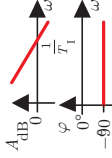
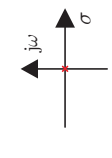
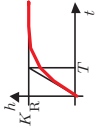

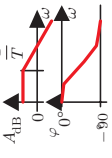
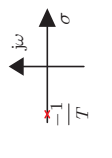
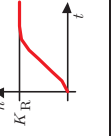

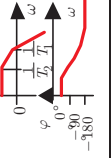
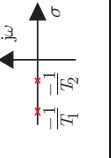
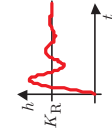
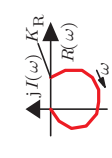
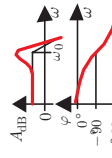
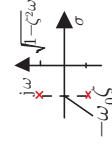
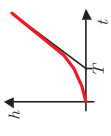
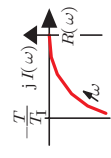
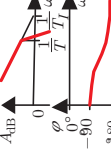
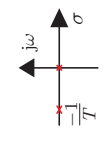
No.	Name	Step response	$h(t) =$	$G(s) =$	Nyquist plot	Bode plot	Poles x and zeros o in s domain
1	P		$K_R \sigma(t)$	K_R			none
2	I		$\frac{t}{T_I} \sigma(t)$	$\frac{1}{sT_I}$			
3	PT ₁		$K_R \left(1 - e^{-\frac{t}{T}}\right) \sigma(t)$	$\frac{K_R}{1+sT}$			
4	PT ₂		$K_R \left(1 - \frac{T_1 - T_2}{T_1 - T_2} e^{-\frac{t}{T_1}} + \frac{T_2}{T_1 - T_2} e^{-\frac{t}{T_2}}\right) \sigma(t)$	$\frac{K_R}{(1+sT_1)(1+sT_2)}$			
5	PT ₂ S		$K_R \left\{ 1 - e^{-\zeta \omega_0 t} \left[\cos(\sqrt{1-\zeta^2} \omega_0 t) + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin(\sqrt{1-\zeta^2} \omega_0 t) \right] \right\} \sigma(t)$	$\frac{K_R}{1+2\frac{\zeta}{\omega_0}s + \frac{1}{\omega_0^2}s^2}$ $\zeta < 1$			
6	IT ₁ S		$\left[\frac{t}{T_I} + \frac{T_I}{T} \left(e^{-\frac{t}{T}} - 1 \right) \right] \sigma(t)$	$\frac{1}{T_I s(1+sT)}$			

Table A.8.3 continued

7	PI		$K_R \left(1 + \frac{t}{T_1} \right) \sigma(t)$	$K_R \frac{1+sT_1}{sT_1}$			
8	D		$T_D \delta(t)$	sT_D			
9	DT1		$e^{-\frac{t}{T}} \sigma(t)$	$\frac{sT}{1+sT}$			
10	PD		$K_R [\sigma(t) + T_D \delta(t)]$	$K_R (1 + sT_D)$			
11	PID		$K_R \left(1 + \frac{t}{T_1} \right) \sigma(t) + T_D \delta(t)$	$K_R \frac{1+sT_1+s^2T_1T_D}{sT_1}$			
12	PIDT1		$K_R \left[1 - \frac{t}{T_1} + \left(\frac{T_D}{T_1} + \frac{t}{T_1} - 1 \right) e^{-\frac{t}{T_1}} + \frac{t}{T_1} \right] \sigma(t)$	$K_R \frac{1+sT_1+s^2T_1T_D}{sT_1(1+sT)}$			
13	PTt		$K_R \sigma(t - T_t)$	$K_R e^{-sT_t}$			

$4 T_D < T_1$
for ω_1 and ω_2 see No. 12

$4 T_D < T_1$
 $\omega_1 = -A+B$
 $\omega_2 = -A-B$
 $A = \frac{1}{2T_D}$
 $B = \sqrt{A^2 - \frac{1}{T_1 T_D}}$

Table A.8.3 continued

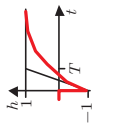
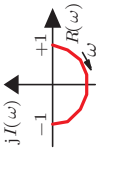
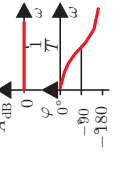
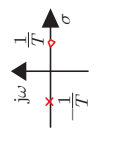
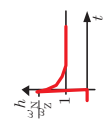
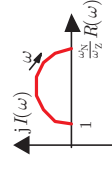
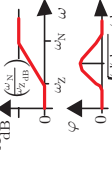
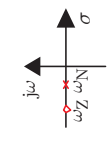
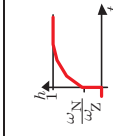
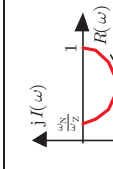
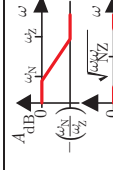
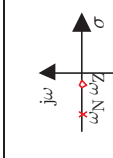
14	1st-order all pass		$\left[1 - 2e^{-\frac{t}{T}} \right] \sigma(t)$	$\frac{1-sT}{1+sT}$			
15	Lead		$\left[1 + \left(\frac{\omega_N}{\omega_Z} - 1 \right) e^{-\omega_N t} \right] \sigma(t)$ $\omega_N > \omega_Z$	$1 + \frac{s}{\omega_Z} \frac{\omega_N}{1 + \frac{s}{\omega_N}}$			
16	Lag		$\left[1 + \left(\frac{\omega_N}{\omega_Z} - 1 \right) e^{-\omega_N t} \right] \sigma(t)$ $\omega_N < \omega_Z$	$1 + \frac{s}{\omega_Z} \frac{\omega_N}{1 + \frac{s}{\omega_N}}$			

Table A.8.4: Padé table for e^{-x}

		numerator			
	denominator	0	1	2	3
0	$\frac{1}{1+x}$	$\frac{1}{1}$	$\frac{1-x}{1}$	$\frac{1-x+\frac{x^2}{2!}}{1}$	$\frac{1-x+\frac{x^2}{2!}-\frac{x^3}{3!}}{1}$
1	$\frac{1}{1+x}$	$\frac{1}{1+x}$	$\frac{1-\frac{1}{2}x}{1+\frac{1}{2}x}$	$\frac{1-\frac{2}{3}x+\frac{1}{3}\frac{x^2}{2!}}{1+\frac{1}{3}x}$	$\frac{1-\frac{3}{4}x+\frac{2}{4}\frac{x^2}{2!}-\frac{1}{4}\frac{x^3}{3!}}{1+\frac{1}{4}x}$
2	$\frac{1}{1+x+\frac{x^2}{2!}}$	$\frac{1}{1+x+\frac{x^2}{2!}}$	$\frac{1-\frac{1}{3}x}{1+\frac{2}{3}x+\frac{1}{3}\frac{x^2}{2!}}$	$\frac{1-\frac{1}{2}x+\frac{1}{6}\frac{x^2}{2!}}{1+\frac{1}{2}x+\frac{1}{6}\frac{x^2}{2!}}$	$\frac{1-\frac{3}{5}x+\frac{3}{10}\frac{x^2}{2!}-\frac{1}{10}\frac{x^3}{3!}}{1+\frac{2}{5}x+\frac{1}{10}\frac{x^2}{2!}}$
3	$\frac{1}{1+x+\frac{x^2}{2!}+\frac{x^3}{3!}}$	$\frac{1}{1+x+\frac{x^2}{2!}+\frac{x^3}{3!}}$	$\frac{1-\frac{1}{4}x}{1+\frac{3}{4}x+\frac{2}{4}\frac{x^2}{2!}+\frac{1}{4}\frac{x^3}{3!}}$	$\frac{1-\frac{2}{5}x+\frac{1}{10}\frac{x^2}{2!}}{1+\frac{3}{5}x+\frac{3}{10}\frac{x^2}{2!}+\frac{1}{10}\frac{x^3}{3!}}$	$\frac{1-\frac{1}{2}x+\frac{1}{5}\frac{x^2}{2!}-\frac{1}{20}\frac{x^3}{3!}}{1+\frac{1}{2}x+\frac{1}{5}\frac{x^2}{2!}+\frac{1}{20}\frac{x^3}{3!}}$
4	$\frac{1}{1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}}$	$\frac{1}{1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}}$	$\frac{1-\frac{1}{5}x}{1+\frac{4}{5}x+\frac{3}{5}\frac{x^2}{2!}+\frac{2}{5}\frac{x^3}{3!}+\frac{1}{5}\frac{x^4}{4!}}$	$\frac{1-\frac{1}{3}x+\frac{1}{15}\frac{x^2}{2!}}{1+\frac{2}{3}x+\frac{2}{5}\frac{x^2}{2!}+\frac{1}{5}\frac{x^3}{3!}+\frac{1}{15}\frac{x^4}{4!}}$	$\frac{1-\frac{3}{7}x+\frac{1}{7}\frac{x^2}{2!}-\frac{1}{35}\frac{x^3}{3!}}{1+\frac{4}{7}x+\frac{2}{7}\frac{x^2}{2!}+\frac{4}{35}\frac{x^3}{3!}+\frac{1}{35}\frac{x^4}{4!}}$

Index

- absolute value of the frequency response, 4-5
- Ackermann, 13-8
- actual value, 1-4, 1-5
- actuator, 1-4, 7-1, 7-2, 8-3, 11-1, 11-2, 11-6, 11-7
- addition
 - frequency response characteristic, 4-6
 - using Nyquist plots, 4-5
- algebraic
 - equation, 2-9
- algebraic product, 15-5
- Algebraic stability criteria, 5-3
- algebraic sum, 15-5
- all-pass element, 4-21, 9-21
- all-pass plant, 10-5
- amplitude response, 4-5, 4-22
 - definition, 4-2
 - logarithmic, 4-22, 5-12
 - of a minimum/non-minimum phase system, 4-20
- angle condition
 - root-locus method, 6-3, 6-6
- angle of departure
 - root-locus method, 6-7, 6-8, 6-11
- angle of entry
 - root-locus method, 6-7, 6-8
- anti-windup, 11-9
- aperiodic behaviour, A-9
- approximation by a PT_1T_t element, 8-6
- approximation by lines, 4-7
- asymptote
 - deviation of the magnitude, 4-10
 - final, 4-10, 4-15, 4-16
 - initial, 4-10, 4-15
 - point of intersection, 4-16
 - slope, 4-10, 4-16
- asymptotic stability, 5-2
- auxiliary control loop, 11-4, 11-6
 - underlying, 11-6
- auxiliary controlled variable, 11-5
- auxiliary controller, 11-4, 11-5
- auxiliary manipulated variable, 11-7, 11-9
- auxiliary variable, 11-4

- bandwidth, 5-15
 - closed loop, 9-2, 9-6, 9-10
 - of a system, 4-17
- basic control loop, 11-2
- behaviour
 - aperiodic, 4-14, 8-6, A-9
 - dynamical, 7-8, 9-2
 - inherent, 10-24
 - minimum phase, 4-20
 - non-minimum phase, 4-20, 10-5, 10-21, 11-4
 - of a continuous-time system, 4-3
 - of a minimum/non-minimum phase system, 4-20
 - of an open loop, 7-6
 - oscillating, 4-14
 - oscillating PT_2S element), A-8
 - permanent oscillation, 8-6
 - PT_2 , 4-14
 - PT_2S element, A-8
 - set-point, 9-8
 - static, 7-6, 7-8
 - steady state, 9-9
 - undamped, A-10
 - unstable, 10-19
- bell function, 15-4
- binary valued logic, 14-1, 14-2
- binomial form, 10-2
- block diagram
 - closed loop, 10-3
 - closed-loop control, 1-3
 - closed-loop control system, 1-5
 - closed-loop system, 7-2
 - combined observer-controller, 13-11
 - compensator, 10-4
 - control system, 1-4
 - feedback loop, 3-7
 - observer principle, 13-9
 - of a series connection of transfer function elements, 3-6
 - open-loop control, 1-3
 - open-loop system, 7-3
 - PID controller, 8-2
 - state-feedback control system, 13-2
 - state-feedback control system with integrator, 13-4
- Bode diagram, 4-5, 4-19
 - lag element, 9-14
 - lead element, 9-12
 - of a PT_2S element, 4-16
- Bode plot
 - of the open loop, 9-7
- breakaway point
 - root-locus method, 6-4
- breakpoint frequency, 4-10, 4-19
 - PT_1 element, 9-9
 - lag element, 9-13, 9-15

- lead element, 9-11
- PT_1 element, 4-9, A-7
- Butterworth form, 10-2

- calibration of the root-locus, 6-3
- car brake, 16-3
- cascade control, 13-4
- cascade control systems, 11-5
- causal function of time, 2-4, 2-5, A-3, A-4
- causal system, 2-2
- centre of gravity method, 16-5
- centre of singleton method, 16-5
- centroid methods
 - margins, 16-7
- characteristic equation, 3-4, 5-5, 5-6, 6-3
 - coefficients, 5-6
 - definition, 2-12
 - of a cascade control system, 11-7
 - of a system with auxiliary manipulated variable, 11-8
 - of a system with auxiliary variable, 11-5
 - of the closed loop, 5-5, 6-3, 7-4
 - position of roots, 5-2
 - stability analysis, 5-2
 - system with feed-forward on the controller, 11-2
 - system with feed-forward on the manipulated variable, 11-3
- characteristic function, 15-1, 15-2
- characteristics of the closed loop in the frequency domain, 9-1
 - bandwidth, 9-2, 9-6, 9-10
 - damping ratio, 9-2
 - natural frequency, 9-2
 - phase angle, 9-2
 - resonant peak, 9-2, 9-10
 - resonant peak frequency, 9-2
- characteristics of the open loop in frequency domain
 - crossover frequency, 5-13
 - gain margin, 5-13
 - phase margin, 5-13
- characteristics of the open loop in the frequency domain
 - crossover frequency, 9-6
 - gain margin, 9-7, 9-8
 - phase margin, 9-7, 9-8, 9-10
- closed loop
 - step response, 10-3
 - transfer function, 10-2
 - bandwidth, 9-2, 9-10
 - basic structure, 1-4
 - behaviour, 7-3, 10-2
 - block diagram, 10-3
 - characteristic equation, 6-3, 7-4
 - characteristics in the frequency domain, 9-1
 - command behaviour, 7-3
 - compensator, 10-4
 - disturbance behaviour, 7-3
 - instability, 7-8, 10-19
 - order, 10-9
 - phase angle, 9-2
 - pole-zero distribution, 10-2
 - poles, 5-8, 10-7, 10-9
 - resonant peak, 9-2
 - resonant peak frequency, 9-2
 - response to step disturbance, 7-10
 - signals, 1-4
 - single loop, 11-1
 - stability, 5-7
 - standard quantities in time domain, 7-9
 - static behaviour, 7-6
 - static property, 7-6
 - structure, 10-16
 - time response, 9-9, 9-10
 - transfer function, 5-5, 7-3, 13-3, A-16, A-17
 - with cumulative disturbance, 7-2
 - zeros, 10-9, 13-3
- closed-loop transfer function, 10-3, 10-16
- closed-loop transfer function for command input, 10-17
 - desired, 10-16
- closed-loop control, 1-2
 - definition, 1-3
- closed-loop system, 13-3
- COG, 16-5
- command behaviour, 7-2, 7-7, 7-11
- command input, 7-2, 7-6
 - step, 7-10
- command signal, 7-6
- command variable
 - step, 10-16
- common divisor, 10-21, 10-22
- comparison of the different types of controllers, 8-4
- compensation, 10-19
 - disturbance, 11-3
 - of disturbances, 11-2
 - of poles, 9-20, 10-5
 - of zeros, 10-5
 - plant, 10-4
 - static, 11-4
- compensator, 9-10, 10-5
 - generalised design method, 10-7
- compensator design methods, 10-2
- complement, 15-5
- complex differentiation theorem of the Laplace transform, 2-4
- complex shifting theorem of the Laplace transform, 2-4
- condition of properness
 - of a transfer function, 3-3
- condition of realisability
 - controller, 10-3
- conformal mapping, A-6, A-7

- continuous-time system, 3-1, 4-3
- control
 - cascade, 11-5
- control behaviour, 11-1
- control device, 1-2
- control error, 1-3–1-5
 - dynamical, 7-9
 - steady-state, 7-10
 - steady-state value, 7-6
- control factor
 - dynamical, 7-3
 - static, 7-8
- control loop
 - auxiliary, 11-4
 - bandwidth of the closed loop, 9-6
 - block diagram, 1-4, 1-5
 - Bode plot of an open, 9-7
 - Bode plot of the open loop, 9-7
 - closed, 1-4
 - frequency response of the open, 9-6
 - resonant peak frequency of the closed loop, 9-5
 - step response of a closed, 9-8
- control system, 10-16
 - complex loop structures, 11-2
 - single-loop, 11-7
 - state-feedback, 13-2
 - with auxiliary manipulated variable, 11-7
 - with auxiliary variable, 11-5
- controllability, 12-6
- controllability matrix, 12-6, 13-8
- controlled chain, 1-4
- controlled value, 1-4
- controlled variable, 1-3, 1-4, 7-3, 10-16, 11-4, 11-6, 11-7
- controller, 1-4, 7-1–7-3, 7-8
 - integral, 10-10
 - state-feedback, 13-2
 - tuning parameters, 8-2, 8-3
- controller transfer function, 9-10, 10-3, 10-7, 10-17, 10-21, 10-23–10-26
- controller canonical form, 13-6, 13-7
 - transformation, 13-7
- controller coefficients, 10-9
- controller gain, 10-9
 - critical, 8-6
- controller order, 10-11
- controller output, 8-2
- controller parameter, 10-8, 10-9
- controller transfer function, 8-1
- convergence
 - Laplace transform, 2-1, A-2
- convolution, 3-2
- convolution in the frequency domain, 2-4, A-4
- convolution in the time domain, 2-4, A-3
- convolution integral, 3-2
- convolution of two functions in the frequency domain, 2-4
 - correction elements, 9-10
- correspondences of Laplace transform, 2-2
- COS, 16-5
- crisp set, 15-1
- critical controller gain, 8-6
- critical damping, A-9
- critical period, 8-6
- critical point, 5-9–5-11
- critical stability, 5-2, 5-3
- crossover frequency, 5-13, 9-6
- cumulative disturbance, 7-2

- D action, 8-3, 8-6
- D behaviour, 8-2, 8-3, 11-3
- D element, 4-8
- D step, 8-3
- damped natural frequency, A-9
- damping ratio, 4-14, 9-2, 9-10, 10-24
 - relative, A-9
- dead time, 3-1, 8-6, 11-1
- decay of the amplitude, 9-4
- decay time constant, A-8
- deforming the root locus, 9-18
- defuzzification, 16-1, 16-5, 16-7, 16-8, 17-4
- delay time, 8-6
- delayed behaviour, 11-4
- delta function, 2-12
- demonstration example, 1-5, 2-9, 7-13, 8-8, 9-22, 11-7, 13-15, 14-2, 15-2, 15-5, 15-7, 15-9, 16-2–16-4, 16-7
- derivative action time, 8-2
- derivative element, 4-8
- derivative theorem of the Laplace transform, 2-4, A-3
- design
 - root-locus method, 10-2
 - direct, 10-2
 - for an unstable plant, 9-21, 10-23
 - for reference and disturbances, 10-16
 - frequency-domain characteristics, 10-2
 - indirect, 10-2
 - state-feedback, 13-5–13-8
- design in controller canonical form, 13-6–13-8
- design method
 - root locus, 9-18
 - compensator, 10-2
 - complex loop structures, 11-1
 - empirical, 8-6
 - frequency domain characteristics method, 9-1
 - pre-filter, 10-21
- design of observer, 13-10
- desired transfer function, 10-3
- differential equation
 - initial condition, 3-1
 - of a RC lag, 4-9
 - solving using the Laplace transform, 2-8

- distribution, 2-12, 2-13
- distribution of poles, 10-2
- disturbance, 1-2, 1-4, 7-2, 7-6, 11-1, 11-5, 11-7
 - at the plant input, 10-16, 10-17, 10-21, 10-24
 - at the plant output, 10-16, 10-17, 10-19, 10-21, 10-25
 - compensation, 11-2, 11-3
 - entry point, 10-16, 10-20
 - feed-forward on the controller, 11-2
 - feed-forward on the manipulated variable, 11-3
 - high-frequency, 9-9
 - step, 10-16
 - transfer function, 7-3
- disturbance behaviour, 7-2, 7-7, 7-9, 10-16
 - closed loop, 7-3
- disturbance control, 1-4, 7-3
- disturbance feed-forward, 11-2
- disturbance reduction, 11-5
- disturbance rejection, 7-11
- disturbance transfer function, 10-17, 10-20, 10-23, 10-24, 10-26
 - desired, 10-16
- dominant pair of poles, 9-2, 9-5, 9-7, 9-10, 9-15, 9-18
- DT₁ element, 4-12, 8-2, 11-3, 11-9
- dual system, 13-11
- dynamical behaviour, 13-2
 - of the closed loop, 7-8, 9-2
- dynamical control factor, 7-3
- dynamical quality, 5-15
- dynamics of the control behaviour, 7-10
- DYNAST study example, 4-11, 4-22, 7-8, 8-8, 8-9, 11-5

- eigenvalue, 2-12, 12-4, 13-2
- empirical tuning rules, 8-6
- equation
 - algebraic, 2-9
 - characteristic, 2-12, 3-4, 5-2, 5-5, 5-6, 6-3, 7-4, 11-2, 11-3
- equivalence
 - Hurwitz and Routh criteria, A-11
- error
 - steady state, 9-9, 10-16

- feed-forward
 - auxiliary manipulated variable, 11-4
- feed-forward gain
 - state-feedback control, 13-4
- feedback, 3-7
 - negative, 1-5, 3-7
- feedback loop
 - of transfer functions, 3-6, 3-7
- feedback principle, 1-5
- feedforward
 - state-feedback controller, 13-2
- feedforward gain, 13-4
- feedthrough matrix, 12-3

- final asymptote, 4-10, 4-15, 4-16
- final value
 - frequency response, 4-4
- final value theorem of the Laplace transform, 2-5, 4-4, 7-6, A-5

- fraction
 - rational, A-11
- frequency
 - breakpoint, 9-9, 9-11, 9-13, 9-15
 - crossover, 9-6
 - damped natural, A-9
 - resonant, 9-5, A-7
- frequency domain, 2-2, 2-9, 4-4
 - state-feedback controller, 13-3
- frequency domain characteristics method
 - application example, 9-15
 - design method, 9-1
 - synthesis of controllers, 9-10
- frequency function, 2-2
- frequency range, 9-9, 9-11, 9-13
- frequency ratio
 - lag element, 9-13
 - lead element, 9-11, 9-12
- frequency response, 4-22, 9-9
 - absolute value, 4-5
 - addition, 4-5
 - all-pass element, 4-21
 - amplitude response, 4-2, 4-20, 4-21, 5-11
 - closed loop, piecewise determination from $G_W(j\omega)$, 9-9
 - dead time, 4-22
 - final value, 4-4
 - imaginary part, 4-2, A-10
 - initial value, 4-4
 - locus, A-7
 - logarithmic amplitude response, 4-5, 4-22, 5-12
 - magnitude, 4-3
 - multiplication, 4-5
 - Nyquist plot, 4-4, 4-5, 4-19, 4-22
 - of the open loop, 9-6
 - phase response, 4-2, 4-3, 4-5, 4-20, 4-21, 5-11, 5-12
 - real part, 4-2, 4-3, A-10
- frequency response characteristic
 - addition, 4-6
 - Bode diagram, 4-5
 - calculation, 4-5
 - definition, 4-5
 - Nyquist criterion, 5-7, 5-10
 - simple transfer function elements, 4-7
- frequency response characteristics, 4-5
- function
 - bell, 15-4
 - characteristic, 15-1, 15-2
 - Gauss-, 15-4
 - hyperbolic, A-10
 - sigmoidal, 15-4

- function of time, 2-2
 - causal, 2-4, 2-5, A-3, A-4
- fuzzification, 16-1–16-3, 16-8, 17-2
- fuzzy AND operator, 15-5
- fuzzy composition, 15-8
- fuzzy control, 14-1, 17-1, 17-8
- fuzzy controller, 17-1
- fuzzy controller design, 17-9
- fuzzy inference machine, 16-2
- fuzzy logic, 14-1, 14-2, 15-1
- fuzzy NOT operator, 15-5
- fuzzy OR operator, 15-5
- fuzzy relations, 15-6
- fuzzy set, 15-2
- fuzzy system, 16-1
 - components, 16-8

- gain, 7-7, 8-2
 - controller, 8-2, 8-6, 10-9
 - of a PT_1 element, 4-11
 - of a transfer function element, 4-11
 - of the controller, 7-7
 - of the open loop, 6-2, 7-6, 7-7
 - P element, 4-7
 - PID controller, 8-2
 - plant, 7-7, 8-6, 8-8, 11-4
- gain margin, 5-13, 9-7, 9-8
- Gaussian function, 2-13, 15-4
- generalised design method
 - compensator, 10-7
- generalised integral of squared error, 7-11
- GISE, 7-11
- grade of membership, 15-2
- grid
 - orthogonal, A-6

- Hilbert transformation, A-10
- Hurwitz conditions, 5-5
- Hurwitz criterion, 5-3, 5-5, A-11
 - example, 5-5
- Hurwitz determinant, 5-4
- Hurwitz polynomial, 5-4, 5-7
- hyperbolic function, A-10

- I action, 8-3, 8-6
- I behaviour, 7-6–7-9, 8-3, 8-6, 11-3
- I controller, 8-3, 10-10
- I element, 4-7, 4-8, 4-19
- I_2 behaviour, 7-6, 7-8, 7-9
- IAE, 7-11
- IE, 7-11
- image function, 2-2
- impulse
 - Laplace transform, 2-12
- inherent behaviour, 2-12, 3-4, 9-2, 10-24
- inherent dynamics, 10-20
- initial asymptote, 4-10, 4-15
- initial condition, 12-1, 12-2, 12-5, 13-2, 13-4–13-6, 13-10–13-12
 - of a differential equation, 3-1
- initial value
 - frequency response, 4-4
- initial value theorem of the Laplace transform, 2-5, 4-4, A-5
- input
 - Laplace transform, 3-2
- input matrix, 12-3
- input signal
 - bounded, 5-2
 - sinusoidal, 4-3
- input variable, 12-1
- input vector, 12-3
- input-output behaviour, 12-2
- inputs
 - multi-input-multi-output system, 12-2
- instability
 - definition, 5-2, 5-3
 - of the closed loop, 7-8
 - proving, 5-7
- integral action time, 8-2
- integral of error, 7-11
- integral of absolute value of error, 7-11
- integral of squared error, 7-11
- integral of squared error and squared control effort, 7-11
- integral of time multiplied by the absolute value of error, 7-11
- integral performance index, 7-10
- integral plant, 10-11
- integral theorem of the Laplace transform, 2-4, A-3
- integral transformation, 2-1, 2-3
- integrator
 - state-feedback, 13-4
- intersection operator, 15-5
- intersections
 - of the locus, 5-11
- inverse frequency response, 4-6
- inverse integral
 - Laplace transform, 2-2, 2-5
- inverse Laplace transform, 2-2, 2-5, A-2, A-11
- inverse plant transfer function, 10-4
- ISE, 7-11
- ISESC, 7-11
- IST^qE, 7-11
- ITAE, 7-11

- lag, 4-9
- lag element, 9-13
- Laplace integral, 2-2
- Laplace transform, 2-1, 2-2, A-2
 - abscissa of convergence, A-2
 - complex differentiation theorem, 2-4
 - complex shifting theorem, 2-4

- convergence, 2-1, A-2
- convergency, A-2
- convergency area, A-2
- convolution in the frequency domain, 2-4, A-4
- convolution in the time domain, 2-4, A-3
- correspondences, 2-2
- derivative theorem, 2-4, A-3
- examples, A-2
- final value theorem, 2-5, 4-4, 7-6, A-5
- impulse, 2-12
- initial value theorem, 2-5, 4-4, A-5
- input, 3-2
- integral theorem, 2-4, A-3
- inverse, 2-2, 2-5, A-2, A-11
- inverse integral, 2-2, 2-5, A-2
- main theorems, 2-3, A-3
- operator notation, 2-2
- output, 3-2
- real shifting theorem, 2-4
- similarity theorem, 2-3
- solving linear differential equation, 2-8
- state-space representation, 12-3
- superposition theorem, 2-3
- weighting function, 3-2, 5-2, A-11
- law of Bode, A-10
- lead element, 9-11
- left-hand rule of the Nyquist criterion, 5-10
 - for Bode diagram, 5-13
- linear mapping, A-7
- linear plant, 7-2
- linear system, 3-1, 3-3, 4-3, 4-4
- linguistic term, 14-4, 15-4, 15-7, 16-2–16-4, 16-7, 17-3, 17-4, 17-10
- linguistic variable, 14-4, 15-6, 17-10
- loading crane, 17-9, 17-10
- logarithmic amplitude response, 4-5
 - logarithmic, 4-5
- loop
 - poles of the open loop, 9-18
 - zeros of the open loop, 9-18
- loop gain, 7-8, 9-9
- low-pass property, 4-17
- lumped parameters, 3-1

- magnitude condition
 - root-locus method, 6-3–6-5
- magnitude of the frequency response, 4-3
- magnitude response
 - increase, 9-11
- main control loop, 11-5–11-7
- main controller, 11-5, 11-6
- main theorems of the Laplace transform, 2-3, A-3
- Mamdani fuzzy system, 16-7
- manipulated variable, 1-3, 1-4
- mapped function, 2-2, 2-4, A-2
- mapped space, 2-9
- mapping
 - conformal, A-6, A-7
 - linear, A-7
- mathematical model, 8-6
- maximum methods, 16-7
- maximum overshoot, 8-5, 8-6, 10-24
 - as function of the damping ratio, 9-3
 - calculation, 9-3
 - definition, 7-9
- measurement device, 1-4, 7-2, 11-1
- membership, 14-2
- membership function, 15-2
 - boundary, 15-3
 - core, 15-3
 - height, 15-3
 - normal, 15-3
 - singleton, 15-4, 15-9, 16-3
 - subnormal, 15-3
 - support, 15-3
- method
 - frequency domain characteristics, 9-10
 - Truxal and Guillemin, 10-3
 - Weber, 10-3, A-15
- minimum phase behaviour, 4-20
- minimum phase system, 4-20–4-22, A-10
- mode, 9-2
- modes, 2-11, 2-12, 3-4
- multi-input-multi-output system
 - inputs, 12-2
 - outputs, 12-2
 - state-space representation, 12-2
- multiple pole
 - real, 10-2
- multiplication
 - using Nyquist plots, 4-5

- natural frequency, 4-15, 4-17, 9-2, 10-24, A-10
 - damped, A-9
- negative feedback, 1-5, 3-7
- non-minimum phase behaviour, 4-20, 10-5, 10-21, 11-4
- non-minimum phase system, 4-20, 4-22, A-10
- nonlinear characteristics, 17-2, 17-3, 17-7, 17-11
- numerator degree
 - transfer function, 3-3
- Nyquist criterion, 5-7
 - frequency response characteristic, 5-7, 5-11
 - general case, 5-9
 - left-hand rule, 5-10
 - left-hand rule for Bode diagram, 5-13
 - simplified case, 5-10
 - theory, 5-7
- Nyquist criterion using Nyquist plots, 5-8
 - of the open loop, 5-9
- Nyquist plot
 - lag element, 9-14
 - lead element, 9-11

- observability, 12-6

- observability matrix, 12-6
- observer, 13-9, 13-10
 - design, 13-10
 - state reconstruction, 13-9
 - state-feedback controller, 13-11
 - structure, 13-10
- observer-controller, 13-11
- open control loop, 7-3
- open loop, 7-3
 - frequency response, 9-10
 - characteristics, 9-10
 - gain, 6-2, 7-6, 7-7
 - pole and zero distribution, 6-8, 6-9
 - poles, 5-8, 5-9
 - transfer function, 5-5, 5-8, 6-2, 7-6
 - zeros, 13-3
- open-loop control, 1-2
 - definition, 1-3, 1-4
- operator
 - complement, 15-5
 - intersection, 15-5
 - union, 15-5
- operator notation
 - Laplace transform, 2-2
- operators
 - for fuzzy sets, 15-5
- optimal tuning
 - PID controller, 8-3
- order of controller, 10-9
- original function, 2-1, 2-2, 2-4, 2-8
- original space, 2-9
- orthogonal grid, A-6
- output
 - Laplace transform, 3-2
- output equation, 12-2
- output matrix, 12-3
- output signal
 - bounded, 5-2
 - sinusoidal, 4-3
- output vector, 12-3
- outputs
 - multi-input-multi-output system, 12-2
- overshoot, maximum, 11-1

- P behaviour, 7-6–7-9, 8-3, 11-3
- P controller, 8-4, 9-22, 10-11
- P element, 4-7, 11-4
- P step, 8-3
- Padé table, A-21
- pair of poles
 - dominant, 9-2, 9-5, 9-7, 9-10, 9-15, 9-18
 - imaginary, A-10
- parabolic input signal, 7-6–7-8
- parallel connection, 4-5
 - of transfer function elements, 3-6, 4-5
- parameter
 - lumped, 3-1
- Parseval's theorem, 7-12
- partial fraction decomposition, 2-6, 2-12, A-8
 - conjugate complex poles, 2-7
 - example, 2-7
 - multiple poles, 2-6
- PD controller, 8-4
- PD element, 4-12, 4-19, 11-4
- PDT₁ controller, 8-4, 8-5, 9-12
- peak time
 - definition, 7-9
- performance diagram, 8-5
- performance index, 7-10, 7-11
 - quadratic, 7-11
- performance index in time domain
 - delay time, 8-6
 - generalised integral of squared error, 7-11
 - integral of absolute value of error, 7-11
 - integral of error, 7-11
 - integral of squared error, 7-11
 - integral of squared error and squared control effort, 7-11
 - integral of time multiplied by the absolute value of error, 7-11
 - maximum overshoot, 7-9, 8-5, 9-3
 - peak time, 7-9
 - rise time, 7-9, 8-6, 9-3, 11-1
 - settling time, 7-9, 9-3, 9-4, 11-1
- period
 - critical, 8-6
- permanent oscillations, 2-12
- phase
 - non-minimum, 10-5
- phase angle
 - maximum, of a lead element, 9-11
 - of the closed loop, 9-2
- phase behaviour
 - minimum, 4-20, 4-21
 - non-minimum, 4-20
 - of a minimum/non-minimum phase system, 4-20
- phase margin, 5-13, 9-7–9-10
- phase response, 4-5, 4-22, 5-12
 - definition, 4-2
 - lead element, 9-13
 - logarithmic representation, 4-6
 - Nyquist criterion, 5-12
 - of a minimum/non-minimum phase system, 4-20
 - of a non-minimum phases system, 4-20
- phase shift, 4-3
 - continuous, 5-9
 - lag element, 9-13
 - lead element, 9-11
 - step, 5-9
- PI controller, 8-4, 11-3
- PID controller, 8-1, 9-22
 - block diagram, 8-2
 - ideal, 8-2
 - optimal tuning, 8-3

- real, 8-2
 - plane, complex
 - s plane, 9-18, 9-21, 10-19, 10-21
 - G plane, 4-4, 5-9, A-6, A-7
 - left-half s plane, 2-12, 5-3, 5-9, 7-6
 - right-half s plane, 5-3, 5-9
 - s plane, 2-8, 2-12, 4-20, 4-21, 5-3, 5-8, 5-9, 6-3, A-2, A-6, A-7
 - plant, 1-4, 7-1, 7-2
 - all-pass, 10-5
 - gain, 8-8, 11-4
 - integral, 10-9, 10-11
 - linear, 7-2
 - parameter, 9-20, 10-8, 10-19
 - proportional, 10-9, 10-11
 - step response, 8-6
 - unstable, 9-21, 9-22, 10-5, 10-19, 10-23, 10-25
 - zeros, 10-9
 - plant transfer function, 10-16, 10-23
 - inverse, 10-4
 - plant output, 7-6
 - point of intersection
 - asymptote, 4-16
 - pole
 - conjugate complex, 2-6
 - double, 7-7
 - multiple, 2-6, 5-3
 - of the closed loop, 9-18
 - plant, 10-19
 - single, 2-6, 5-3
 - unstable, 10-22
 - pole and zero distribution
 - of a minimum/non-minimum phase system, 4-20
 - of a rational transfer function, 3-4, 6-2
 - of an all-pass element, 4-20, 4-22
 - of an open loop, 6-8, 6-9
 - pole assignment, 9-18
 - pole distribution, 5-2, 5-3
 - of a PT_2 element, 4-18
 - pole excess, 10-3, 10-17, 10-20, 11-3
 - pole placement, 13-5
 - state-feedback, 13-5
 - pole-zero distribution
 - of a closed loop, 10-2
 - poles, 2-6, 12-4
 - of the closed loop, 5-8
 - of the open loop, 5-8, 5-9, 9-18
 - of the transfer function, 3-3, 3-4
 - positive feedback, 3-7
 - pre-controller, 11-2
 - pre-filter, 10-9, 10-16
 - design, 10-21, 10-24
 - unstable, 10-21
 - preservation of angles, A-6
 - problem, 4-22, 5-15, 10-26
 - product, algebraic, 15-5
 - properness
 - of a transfer function, 3-3
 - PT_1 element, 4-9–4-11, 4-17, 4-19, 9-9, 11-5
 - PT_1T_t element, 8-6
 - PT_2 element, 4-13, 4-17
 - PT_2S element, 4-17, A-8
 - PT_n element, 4-17
 - PT_t element, 4-22
-
- quadratic performance index, 7-11, 7-12
 - calculation, 8-3, A-12
 - quality
 - dynamical, 5-15
 - questionnaire, 3-2, 3-7, 4-5, 4-14, 5-3, 5-15, 15-2, 15-4, 15-6
-
- ramp input signal, 7-6–7-8
 - rational fraction, 2-5, 2-6, A-11
 - rational transfer function, 3-3
 - RC high pass element, 4-12, 4-13
 - RC lag, 4-9
 - real part
 - negative, A-11
 - real shifting theorem of the Laplace transform, 2-4
 - realisability, 10-19
 - realisability condition, 10-20, 10-24
 - controller, 10-17, 10-19, 10-26
 - of a transfer function, 3-3, 5-8
 - pre-filter, 10-21, 10-22
 - reference behaviour, 7-6
 - reference value, 11-6
 - relative damping ratio, A-9
 - residual, 2-6–2-8
 - theorem, 2-6
 - resonant peak, 9-2, 9-5, 9-10
 - of the closed loop, 9-2
 - resonant peak frequency, A-8
 - of a PT_2S element, 4-16
 - of the closed loop, 9-2, 9-5
 - rise time, 8-6, 11-1
 - calculation, 9-3
 - definition, 7-9
 - RLC lag, 4-14–4-17, 12-1
 - root-locus, 6-4
 - definition, 6-2
 - root-locus method
 - angle condition, 6-3, 6-6
 - angle of departure, 6-7, 6-8, 6-11
 - angle of entry, 6-7, 6-8
 - application for controller design, 9-18
 - breakaway point, 6-4
 - examples, 6-10
 - magnitude condition, 6-3–6-5
 - rules for constructing, 6-5
 - Routh criterion, 5-6, 5-7, A-11
 - example, 5-7
 - Routh schema, 5-6

- rule base, 14-4, 14-5, 15-7–15-10, 16-1–16-4, 17-2, 17-4, 17-6, 17-7, 17-10, 17-11
- s domain, 9-2
 - state equation, 12-3
- scales
 - double logarithmic, 4-7
 - single logarithmic, 4-7
- separation property, 13-12
- series connection, 4-5
 - of transfer function elements, 3-5, 3-6, 4-5, 4-6, 4-21
- set
 - crisp, 15-1
 - fuzzy, 15-2
- set point, 1-4, 1-5, 13-4
 - constant/not constant, 1-4
- set theory, 15-1
- set-point value, 1-2
- settling time, 10-24
 - as function of the damping ratio, 9-5
 - calculation, 9-4
 - definition, 7-9
- sigmoidal function, 15-4
- signal path, 1-5
- signals in the closed loop, 1-4
- similarity theorem of the Laplace transform, 2-3
- single-input-single-output system
 - state-space representation, 12-1–12-3
 - time-invariant, 12-2
- singleton, 15-4, 16-5
- singularity, 2-12
- specification
 - time response, 7-9
- spray-water cooler, 11-4
- stabilising of an unstable plant, 9-21
- stability, 5-2
 - asymptotic, 5-2
 - closed loop, 5-7
 - conditions, 5-2
 - critical, 5-2, 5-3
 - definition, 5-2
- stability analysis, 5-2, 5-7
- stability criteria
 - algebraic, 5-3
- stability criterion, 5-3, 5-6
 - Nyquist, 5-7
- stability diagram, 8-5
- stable system, 5-2
- standard form of the transfer function, 10-2
- standard forms
 - Butterworth form, 10-2
 - ITAE form, 10-2
 - settling time form, 10-3
- state, 12-2
- state equation, 12-2
 - s domain, 12-3
- state reconstruction using observers, 13-9
- state variables, 12-2
 - uniqueness, 12-4
- state vector, 12-3
- state-feedback, 13-2
 - with integrator, 13-4
 - design, 13-5
 - pole placement, 13-5
- state-feedback control
 - feed-forward gain, 13-4
 - steady state, 13-3
 - zeros, 13-3
- state-feedback control system, 13-2
- state-feedback control with integrator, 13-4
- state-feedback controller, 13-2, 13-5
 - frequency domain, 13-3
 - observer, 13-11
 - structure, 13-2
- state-space transformation, 12-5
 - example, 12-5
- state-space representation, 12-1
 - controllability, 12-6
 - Laplace transform, 12-3
 - multi-input-multi-output system, 12-2
 - observability, 12-6
 - single-input-single-output system, 12-1–12-3
 - transfer function, 12-3
 - transformation, 12-5
- static behaviour, 7-6
 - of the closed loop, 7-6, 7-8
- static control factor, 7-8
- statical behaviour, 13-2
- steady state, 8-5, 11-9, 13-2
 - state-feedback control, 13-3
- steady state behaviour, 9-9
- steady-state error, 10-9
- steam flow, 11-4
- steam superheater, 11-4, 11-5
- steam temperature, 11-4, 11-5
- step command input, 7-10
- step disturbance, 8-5
- step input signal, 7-6–7-9
- step response, 10-25
 - lag element, 9-14
 - lead element, 9-12
 - normalised, 10-2
 - of a closed loop with PT₂S behaviour, 9-8
 - of a PT₂ element, 4-18, A-9
 - of a PT₂S element, A-8, A-10
 - of standard forms, A-16, A-17
 - PD controller, 8-3
 - PDT₁ controller, 8-3
 - PI controller, 8-3
 - PID controller, 8-2, 8-3
 - PIDT₁ controller, 8-3
 - PT₁T_t element, 8-8
- Stodola criterion, 5-5

- structure
 - state-feedback controller, 13-2
- structure of an observer, 13-10
- sum, algebraic, 15-5
- summing point, 1-5
- superposition theorem of the Laplace transform, 2-3
- synthesis
 - for an unstable plant, 10-19
- synthesis equation, 10-9
- synthesis method
 - for an unstable plant, 10-25
 - frequency domain characteristics, 9-10
 - pre-filter design, 10-24
- system
 - causal, 2-2
 - continuous-time, 3-1, 4-3
 - invariant, 3-3
 - linear, 3-1, 3-3, 4-3, 4-4
 - minimum phase, 4-20
 - multi-input-multi-output, 12-2
 - non-minimum phase, 10-5, 11-4
 - output signal, 4-3
 - realisable, 5-8
 - single-input-single-output, 12-1, 12-2
 - stable, 4-20, 5-2
 - time-invariant, 3-1, 12-2
 - unstable, 5-2
 - with dead time, 5-7
 - with lumped parameters, 3-1
 - without dead time, 3-3, 4-20
- system matrix, 12-3
- system of differential equations, 12-2
- system property, 5-2

- Takagi-Sugeno fuzzy system, 16-7
- tangent at the turning point, 8-6
- temperature control, 11-4, 11-5
- test signals, 7-6
- theorem
 - Parseval, 7-12
- theorem of residuals, 2-6, 2-7
- time constant
 - of a PT_1 element, 4-9, 4-11
 - of a PT_2 element, 5-5, A-9
 - of an I element, 4-7
- time domain, 2-2, 2-5, 4-4
- time response
 - of the closed loop, 9-9, 9-10
- time-invariant system, 3-1, 3-3
- time-response specifications, 7-9
- time-varying system, 12-3
- transcendental transfer function, 3-3
- transfer function
 - all-pass element, 4-21
 - combinations, 3-5
 - dead time, 3-3, 4-22
 - definition, 3-2
 - desired, 10-2
 - disturbance, 10-16, 10-17, 10-23, 10-24
 - feedback loop, 3-6, 3-7
 - for calculations, 3-5
 - I controller, 8-3
 - I element, 4-7
 - improper, 3-3
 - inverse, 10-4
 - lag element, 9-13
 - lead element, 9-11
 - numerator degree, 3-3
 - of a cascade control system, 11-5
 - of the closed loop, 5-5
 - of the open loop, 5-5, 5-8, 6-2, 7-3, 7-6
 - P controller, 8-3
 - P element, 4-7
 - parallel connection, 3-6, 4-5
 - PD controller, 8-3
 - PD element, 4-12
 - PDT_1 controller, 8-3
 - PI controller, 8-3
 - PID controller, 8-1
 - $PIDT_1$ controller, 8-3
 - poles and zeros, 3-3, 3-4
 - pre-filter, 10-16, 10-21, 10-26
 - proper, 3-3
 - properness, 3-3
 - PT_1 element, 4-9
 - PT_2 element, 4-14, A-9
 - PT_2S element, 4-15, A-10
 - PT_t element, 4-22
 - rational, 3-3, 3-4, 4-19, 4-22
 - realisability, 3-3
 - realisable, 10-24
 - series connection, 3-5, 3-6, 4-5
 - standard forms, 10-2
 - state-space representation, 12-3
 - strictly proper, 3-3
 - transcendental, 3-3
- transfer function elements, 4-7
 - series connection, 4-6
- transfer function for command input, 10-24
- transformation
 - controller canonical form, 13-7
 - state-space representation, 12-5
- transformation matrix, 12-5, 13-7
- tuning parameters of the controller, 8-2, 8-3
- tuning rules
 - empirical, 8-6
- turning point
 - tangent, 8-6

- unambiguously reversible mapping, 2-2
- undamped behaviour, A-10
- union operator, 15-5
- universal set, 15-3

unstable

– PT₂S element, [A-10](#)

unstable behaviour, [10-19](#)

unstable plant, [9-21](#), [9-22](#), [10-5](#), [10-19](#), [10-23](#),
[10-25](#)

unstable system, [5-2](#)

value of truth, [15-1](#), [15-2](#)

variation of the plant parameters, [10-19](#)

weighting function, [3-2](#), [A-11](#)

– Laplace transform, [5-2](#), [A-11](#)

windup, [11-9](#)

windup effect, [11-9](#)

zeros

– controller transfer function, [10-19](#), [10-21](#)

– of the closed loop, [10-9](#)

– of the open loop, [9-18](#)

– of the plant, [10-9](#)

– of the transfer function, [3-3](#), [3-4](#)

– plant, [10-19](#)

– state-feedback control, [13-3](#)

Ziegler-Nichols tuning rules, [8-6](#)

– method of the stability margin, [8-6](#)

– method of the step response, [8-6](#)