european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness
INVESTMENTS IN EDUCATION DEVELOPMENT

TECHNICAL UNIVERSITY OF OSTRAVA

FACULTY OF MECHANICAL ENGINEERING

BASIC PRINCIPLES OF AUTOMATIC CONTROL

Antonín Víteček, Miluše Vítečková, Lenka Landryová

Ostrava 2012

Lektor: Prof. Ing. Zora Jančíková, CSc.

Basic Principles of Automatic Control

ISBN 978-80-248-4062-8

PREFACE

The major mission of this textbook is to highlight the importance of basic principles of an automatic control by covering the most important areas from analog automatic control, digital control, and two- and three-position control.

Hopefully, this textbook will stimulate new ideas by giving the reader basic points of view of control system theory as well as appreciation of its use and adaptability into complex systems.

The contents of this textbook originate in many texts and papers written by the authors on their own, as well as hours of working on their approaches to the basic methodology and experience with teaching it to students of control engineering.

Since the textbook is concerned with the basic concepts of automatic control, therefore the textbook does not have any given references itself. For deepening your knowledge and extending your study materials the authors recommend the references mentioned below for further reading:

DORF, R.C. – BISHOP, R. *Modern Control Systems* (*12th ed.*). Prentice-Hall, Upper Saddle River, New Jersey 2011

FRANKLIN, G.F. – POWELL, J.D. – EMAMI-NAEINI, A. *Feedback Control of Dynamic Systems* (*4th ed.*). Prentice-Hall, Upper Saddle River, New Jersey 2002

THE authors thank Prof. Ing. Zora Jančíková, CSc. for her valuable suggestions.

Many key control techniques in use today have been founded on the very basic principles of the past and we must not forget those ingenious individuals of old who solved control problems with truly original solutions. The textbook would like to point out these ideas which blended into our technologies and are now taken for granted not only by students interested in control engineering. Good technical ideas are precious and need to be respected by properly obeying the basics when developing modern technological systems. If you enjoy reading the book then the authors' efforts were worthwhile.

CONTENT

1 INTRODUCTION

We meet with "control" or "drive" every day and all the time. The word "control" is used in common cases, but the word "drive" is often used to mean manual control. We drive (ride) a bicycle, a motorcycle, a car, etc. In these cases, it is a manual control. An example of the simplified control of a car is shown in Fig. 1.1. A driver tries to keep a desired path that is a desired lateral displacement $w(t)$ on the right side of the road with a steering wheel angle $u(t)$ regardless of disturbances $v(t)$, i.e. current car velocity, the road condition and its behavior (slopes, bends, zigzag bends etc.). The effect of his driving is the true lateral displacement from the middle of the right side of the road $y(t)$, see Fig. 1.2.

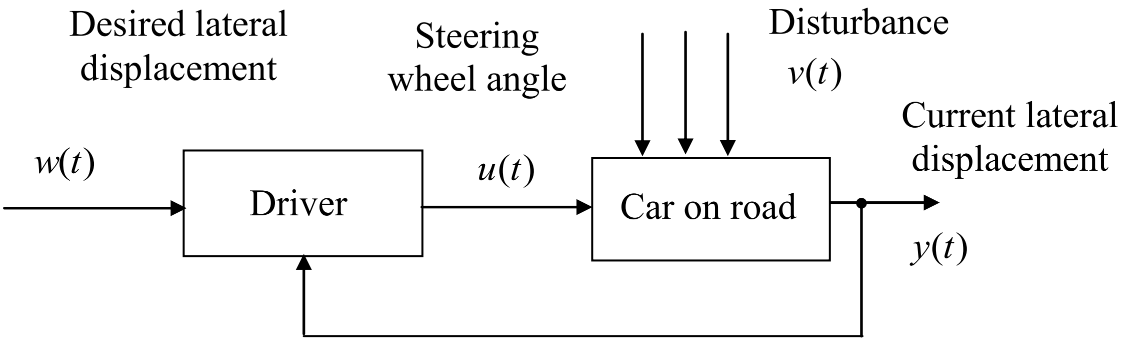


Fig. 1.1 – Control of car on a road

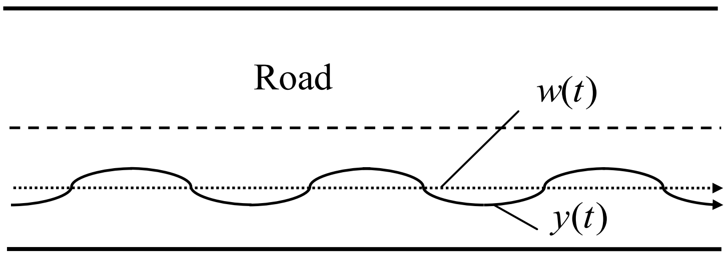


Fig. 1.2 – Courses of a current $y(t)$ and desired $w(t)$ car displacement from the middle of the right side of the road

The driver evaluates the current lateral displacement $y(t)$ and by suitably turning the steering wheel with angle $u(t)$ he tries to minimize the difference

$$e(t) = w(t) - y(t) \rightarrow 0 \tag{1.1}$$

which can be written in the equivalent form

$$y(t) \rightarrow w(t) \tag{1.2}$$

The relations (1.1) or (1.2) equivalently express the **control objective**.

We deal with automatic control so often that we do not perceive it. There are controls for an iron´s temperature, the water temperature and level control in the washing machine, the refrigerator and freezer temperature control, the room temperature control, etc. in our homes.

Iron temperature control is shown in Fig. 1.3. The controlling device is made from a bimetal strip, which bends when heated and the strip's bending measures the current temperature of a heating body $y(t)$. When this temperature is lower than the adjusted desired temperature $w(t)$, then the bimetal strip switches on the heating body and it is supplied by a voltage $u(t)$ (mostly 230 V). When reaching the desired temperature $y(t) \approx w(t)$, the bimetal strip switches off the heating body and it begins to cool down. After decreasing the heating body temperature $y(t)$ below the desired temperature $w(t)$, the bimetal strip switches on again. This process repeats.

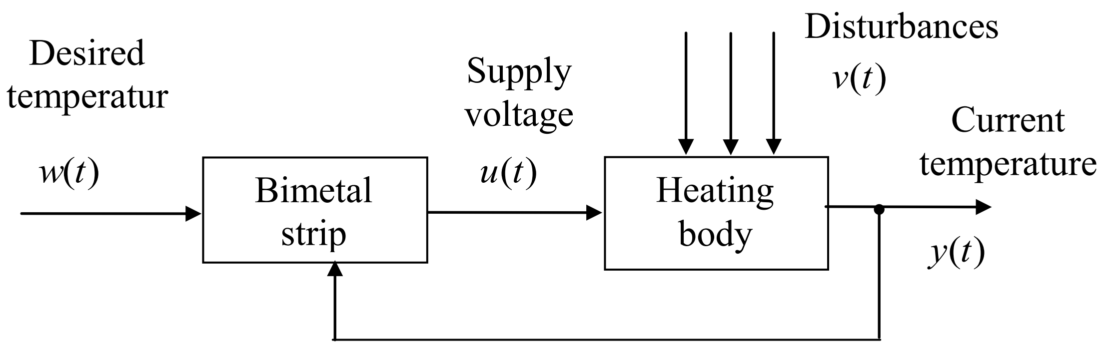


Fig. 1.3 – Iron temperature control

In this case the disturbances $v(t)$ can be e.g. the different moisture and temperature of laundry. If the disturbances $v(t)$ are constant, then the heating and cooling processes are periodic.

It is obvious that in this case the bimetal strip fulfills these conditions (1.1) or equivalently (1.2). The bimetal strip of an iron is one of the simplest controlling devices. Therefore it operates in two states “switch-on” and “switch-off”, it is called an“**ON-OFF” controller** or **two-position controller**.

There are different control systems in the present-day radio and television sets, e.g. the automatic volume control, the automatic frequency control, voltage and current stabilization, automatic brightness control, etc. Nowadays every compact camera contains automatic focusing, automatic image stabilization, the automatic white balancing, an automatic aperture and shutter setting, the automatic tracking of an object, etc.

Very complex automatic control systems are especially used in automobile, aviation, rocket and military technology.

Both control systems in Figs 1.1 and 1.2 can be generally presented by a block diagram in Figs 1.4 and 1.5, where in the first case (Fig. 1.1) the controller is implemented by a driver – a man (human) and in the second case

(Fig. 1.2) the controller is implemented by the bimetal strip – an automatic two-position controller.

It is obvious that the **sensor** (measuring device) must be accurate and fast and that is why its behavior is very often neglected or added to **a plant** or **process** (controlled device). **The control cannot be more accurate than the sensor's accuracy is**. Similarly, the behavior of **an actuator** (actuating device) is added to a plant or to **a controller** (the controlling device) and a comparative element is set apart in a separate **summing node** (a comparison device). The disturbances are often aggregated into one or two selected disturbances. Then the **closed-loop control system** or the **feedback control system** can be obtained, where the desired output $w(t)$ is the **desired** or **reference variable**, the current controlled output $y(t)$ is the **controlled variable**, the controller output $u(t)$ is the **control**, **actuating** or **manipulated variable**, the summing node output $e(t)$ is the **control error**, the aggregated disturbances $v_1(t)$ and $v_2(t)$ are the **disturbance variables**.

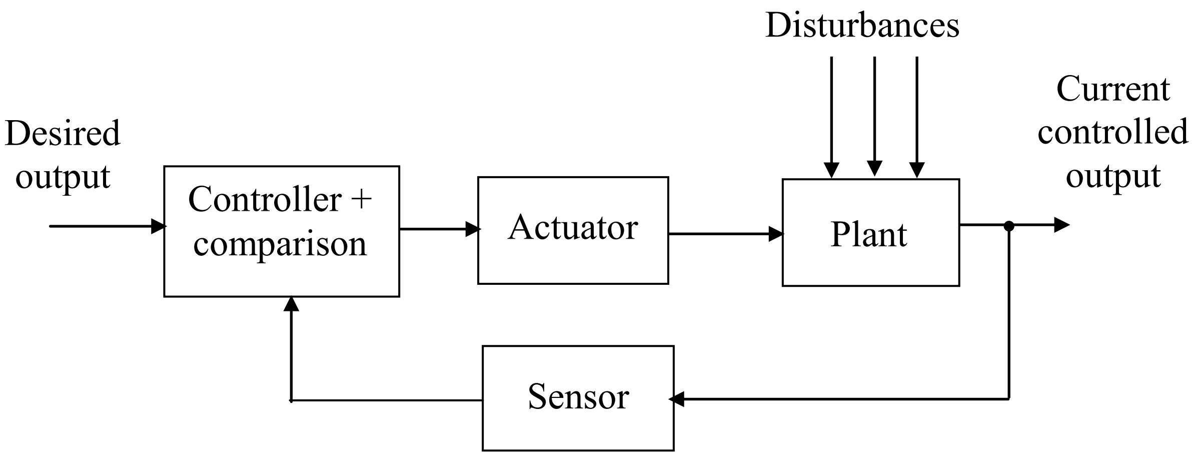


Fig. 1.4 – General control system

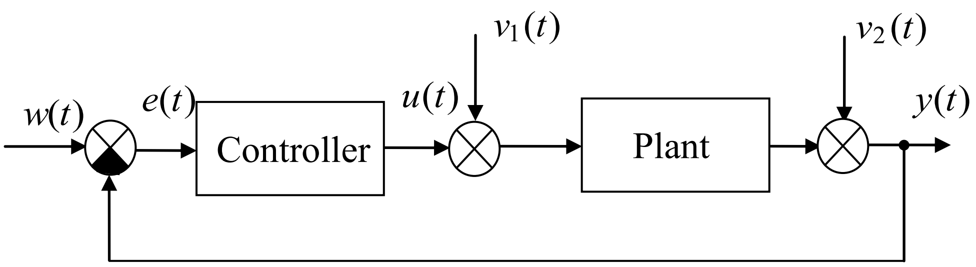


Fig. 1.5 – Closed-loop control system

Negative feedback is very important, because determination of the control error $e(t)$ wasn't enabled and the controller could not hold the demand (1.1) or (1.2).

The demand (1.1) or (1.2) is called a **control objective**. Two controller tasks follow from it. The first task is tracking a desired variable by the controlled variable – the **servo problem** (**set-point tracking**), and the second task is the rejection of the disturbances – the **regulatory problem**. The rejection of a disturbance which is caused in the input of a process/plant is the most frequent problem considered in the second case.

An **open-loop control system** or the **feedforward control system** can be used in some simple cases, when the disturbances are negligible or they have not influenced a control process. They are mostly very simple logical systems, e.g. the traffic control, the washing machine etc. A traffic control is shown in Fig. 1.6). The traffic light sequence and switching (green, amber, red) are preprogrammed in accordance with the expected traffic flow depending on the time of day and the kind of day (working day, holiday etc.). A simplified block diagram of an open-loop system is shown in Fig. 1.7.

The behavior of both open-loop and closed-loop control systems is explained below.

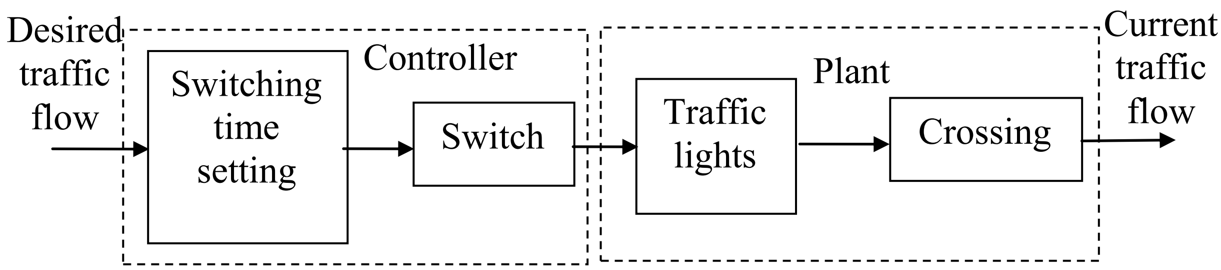


Fig. 1.6 – Traffic Flow Control

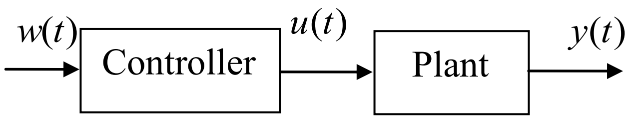


Fig. 1.7 – Open-loop control system

For example, consider the simple control systems in Fig. 1.8, where a controller's behavior is expressed by the gain $K_P > 0$ and a plant by the gain $k_1 > 0$ too.

We can perform an analysis of both open-loop (Fig. 1.8a) and closed-loop (Fig. 1.8b) control systems a)

b)

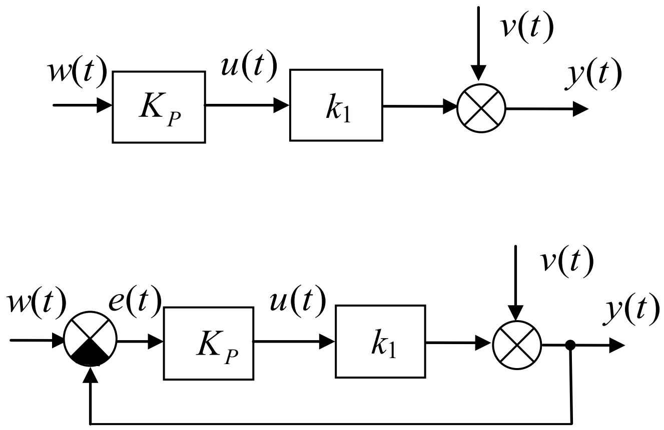


Fig. 1.8 – A control system: a) open-loop structure, b) closed-loop structure

a) Open-loop control system (Fig. 1.8a)

In accordance with Fig. 1.8a we can write

$$y(t) = K_P k_1 w(t) + v(t) \tag{1.3}$$

On condition that the disturbance $v(t)$ doesn't cause a problem to an open-loop control system, i.e. $v(t) = 0$, it is

$$y(t) \to w(t) \;\Rightarrow\; K_P \to \frac{1}{k_1} \tag{1.4}$$

which follows from the control objective (1.2).

If the disturbance $v(t) \neq 0$, it will cause a problem to an open-loop control system (Fig. 1.8a) and at the same time (1.4) will hold, then there can be obtained

$$y(t) \to w(t) + v(t) \tag{1.5}$$

We can see that the open-loop control system is unable to reject the disturbance $v(t)$, i.e. its influence on the controlled variable $y(t)$.

If the behavior of a plant changes or is known with an accuracy $\pm\Delta k_1$ then (1.5) has the form

$$y(t) \to \frac{k_1 \pm \Delta k_1}{k_1} w(t) + v(t) = \left(1 \pm \frac{\Delta k_1}{k_1}\right) w(t) + v(t) \tag{1.6}$$

From (1.6) it is obvious that the changes of a plant (uncertainty) $\pm\Delta k_1$ fully come out on the controlled variable $y(t)$.

For example, for $k_1 = 1$ and $\Delta k_1 / k_1 = 0.5$ (50 %) there is obtained

$$y(t) = (1 \pm 0.5) w(t) + v(t)$$

We can see that the change of the plant behavior and the disturbance fully come out on the controlled variable. It is obvious that the open-loop structure is suitable only for cases when the plant behavior is invariant and disturbances are negligible.

b) Closed-loop control system (Fig. 1.8b)

We can write on the basis of Fig. 1.8b

$$y(t) = K_P k_1 [w(t) - y(t)] + v(t) \Rightarrow$$

$$y(t) = \frac{1}{\frac{1}{K_P k_1} + 1} w(t) + \frac{1}{1 + K_P k_1} v(t) \tag{1.7}$$

From (1.7) for

$$K_P \to \infty \text{ or } K_P k_1 \to \infty \tag{1.8}$$

relation

$$y(t) \to w(t) \tag{1.9}$$

is obtained.

We can see that for the sufficient high controller gain K_P or the product $K_P k_1$ the control objective (1.2) holds for a plant with an arbitrary finite gain k_1 and at the same time the negative influence of a disturbance $v(t)$ on a controlled variable $y(t)$ will be rejected. The same conclusion holds for plant changes or uncertainties expressed by an increment of plant gain $\pm \Delta k_1$:

$$y(t) = \frac{1}{\frac{1}{K_P k_1 \left(1 \pm \frac{\Delta k_1}{k_1}\right)} + 1} w(t) + \frac{1}{1 + K_P k_1 \left(1 \pm \frac{\Delta k_1}{k_1}\right)} v(t) \tag{1.10}$$

If conditions (1.8) are fulfilled then (1.9) is obtained again.

For example, for $K_P = 100$, $k_1 = 1$ and $\Delta k_1 / k_1 = 0.5$ (50 %) there is obtained on the basis of (1.10)

$$y(t) = \left(0.9901 \begin{matrix} +0.0033 \\ -0.0097 \end{matrix}\right) w(t) + \left(0.0099 \begin{matrix} -0.0033 \\ +0.0097 \end{matrix}\right) v(t)$$

We can see that the control objective (1.2)

$$y(t) \to w(t)$$

holds with an accuracy better than 2 % even for a relatively small value of a controller gain $K_P = 100$ and for ± 50 % changes of plant behavior, i.e. its gain k_1. At the same time the negative influence of a disturbance $v(t)$ is reduced to less than 2 % as well.

A closed-loop control system enables superior control considerably more than the open-loop control system. It is caused by the existence of the negative feedback, which is a necessary condition not only for high-quality control but for any meaningful activity of living beings and thus for a man. **Living isn't possible without the existence of negative feedback.**

It is very important that the high controller's gain K_P occurs in the forward path (branch).

A closed-loop control system even works out the non-linear plant. In Fig. 1.9 there is a control system with a non-linear plant, which is described by a non-linear function

$$y(t) = f[u(t)] + v(t) \tag{1.11}$$

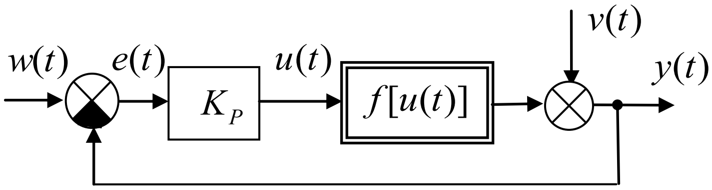


Fig. 1.9 – A closed-loop control system with a non-linear plant

In accordance with Fig. 1.9, we can write

$$e(t) = w(t) - y(t) \;\Rightarrow\; y(t) = w(t) - e(t) \tag{1.12}$$

$$e(t) = \frac{u(t)}{K_P} = \frac{f^{-1}[y(t) - v(t)]}{K_P} \tag{1.13}$$

After substituting (1.13) in (1.12) there is obtained

$$y(t) = w(t) - \frac{f^{-1}[y(t) - v(t)]}{K_P} \tag{1.14}$$

It is obvious that the relation holds

$$K_P \to \infty \Rightarrow y(t) \to w(t)$$

We can see again that for a satisfactory high controller gain K_P the control objective (1.2) is available even for a non-linear plant and for the negative influence of the disturbance (1.11).

At the end of this chapter the general system in Fig. 1.10 is considered. We can symbolically describe the system by a following relation

$$y(t) = Su(t)$$

where S is an operator, which symbolically expresses the system's behavior.

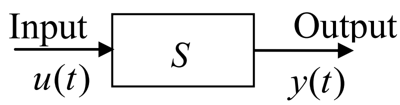


Fig. 1.10 – General system

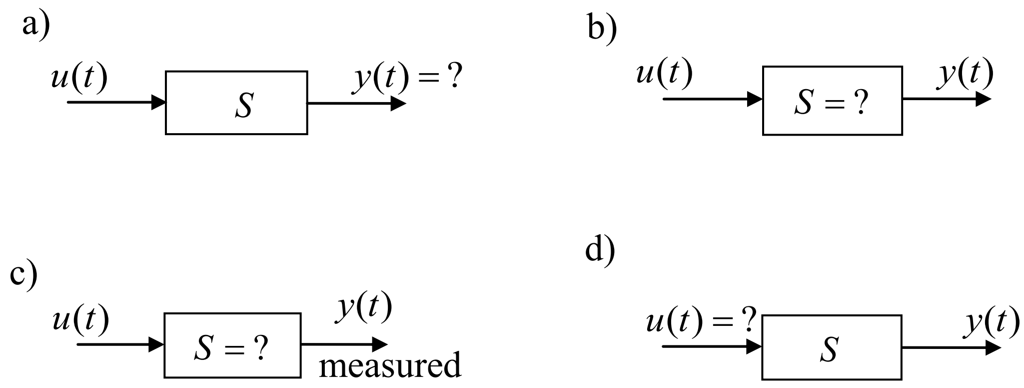


Fig. 1.11 – Basic problems in automatic control: a) analysis, b) synthesis, c) identification, d) control

The basic problems with a system in Fig. 1.10 in automatic control are:

The analysis problem. The system's behavior S and the input $u(t)$ are given and we want to determine output $y(t)$. The solution to this problem is generally unique.

The synthesis problem. The input $u(t)$ and the output $y(t)$ are given and we want to determine (design) a corresponding system's behavior S. The solution to this problem isn't unique and it demands a further criterion for selecting a suitable system's behavior S.

The identification problem. The input $u(t)$ is given and the system is given, but its behavior S isn't known. We can measure the output $y(t)$ and we want to determine the mathematical model of a system's behavior S. This problem relates to the black (color, gray) box problem.

The control problem. The system's behavior S is known and the desired output $y(t)$ is given and we want to determine a corresponding input $u(t)$, which ensures the desired output $y(t)$.

2 MATHEMATICAL MODELS OF SYSTEMS

We will consider the SISO (single-input single-output) system (Fig. 2.1).

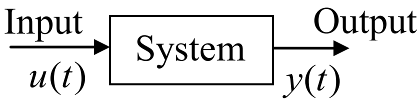


Fig. 2.1 – Block representation of the SISO system

The dependence of a system output y(t) on its input u(t) expresses its static and dynamic behavior. The time changes on a system input are called the **action** or **excitation** and a corresponding system output time changes are called **reaction** or **response**. A real, existing system has to hold the **physical realizability condition** or the **causality condition**, which means that **the reaction – consequence cannot precede the action – cause**.

The control systems are analyzed on their mathematical models. An analogy is employed, which keeps the most important behavior of original systems. If there is no difference between the original system and its mathematical model behaviour and it does not cause any confusion, then a mathematical model is called the original system. The input time functions are called – **inputs**, **input signals** or **input variables** and similarly the output time functions are called – **outputs**, **output signals** or **output variables**.

A mathematical model of the SISO system has often a form of a differential equation in a time domain

$$g[y^{(n)}(t),\dots,\dot{y}(t),y(t),u^{(m)}(t),\dots,\dot{u}(t),u(t)]=0 \tag{2.1}$$

with initial conditions

$$\begin{aligned} &y(0)=y_0,\dot{y}(0)=\dot{y}_0,\dots,y^{(n-1)}(0)=y_0^{(n-1)} \\ &u(0)=u_0,\dot{u}(0)=\dot{u}_0,\dots,u^{(m-1)}(0)=u_0^{(m-1)} \end{aligned} \tag{2.2}$$

$$\begin{aligned} &\dot{y}(t)=y^{(1)}(t),y^{(i)}(t)=\frac{\mathrm{d}^i y(t)}{\mathrm{d}t^i};\ i=1,2,\dots,n \\ &\dot{u}(t)=u^{(1)}(t),u^{(j)}(t)=\frac{\mathrm{d}^j u(t)}{\mathrm{d}t^j};\ j=1,2,\dots,m \end{aligned} \tag{2.3}$$

where $u(t)$ is an input variable, $y(t)$ – an output variable, n – an order of a differential equation and at the same time the order of an original system, g – is generally a non-linear function.

If a mathematical model (2.1) satisfies the inequality

$$n > m \tag{2.4}$$

then the mathematical model is **strongly** physically realizable.

For

$$n = m \tag{2.5}$$

it satisfies only a **weak** physical realizability condition and for

$$n < m \tag{2.6}$$

the mathematical model **isn't** physically realizable, i.e. a mathematical model similar to this doesn't correspond to any real existing system.

If on the basis of a differential equation (2.1) for

$$\begin{aligned} &\lim_{t\to\infty} y^{(i)}(t) = 0; \; i = 1,2,\dots,n \\ &y = \lim_{t\to\infty} y(t) \\ &\lim_{t\to\infty} u^{(j)}(t) = 0; \; j = 1,2,\dots,m \\ &u = \lim_{t\to\infty} u(t) \end{aligned} \tag{2.7}$$

an equation can be obtained

$$y = f(u) \tag{2.8}$$

then this equation describes the **static characteristic** of a given model and, at the same time, the original system (Fig. 2.2).

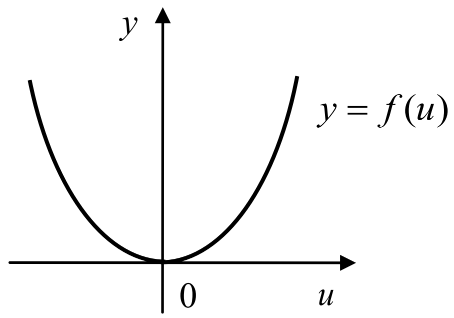


Fig. 2.2 – Non-linear static characteristic

A static characteristic expresses the dependency between an output *y* and input *u* variables in a **steady state**.

The course of an output *y*(*t*) or input *u*(*t*) variables between two steady states is called a **transient process**.

If in the equation (2.1) the derivatives (2.3) don't arise, i.e..

$$g[y(t), u(t)] = 0 \quad \text{or} \quad g(y,u) = 0 \tag{2.9}$$

then a mathematical model (2.9) describes a **static system**. The derivatives (2.3) are basic attributes for dynamic behaviors, and therefore a differential equation (2.1) describes a **dynamic system**.

2.1 Linear Mathematical Models

The **linear models** create a very important class of mathematical models. Their most important behavior is **linearity**. The linearity of a dynamic system in Fig. 2.1 can be expressed by two partial behaviors:

additivity (superposition):

$$\left.\begin{matrix} u_1(t) \Rightarrow y_1(t) \\ u_2(t) \Rightarrow y_2(t) \end{matrix}\right\} \quad u_1(t)+u_2(t) \Rightarrow y_1(t)+y_2(t) \tag{2.10a}$$

homogeneity

$$u(t) \Rightarrow y(t), \quad au(t) \Rightarrow ay(t) \tag{2.10b}$$

Both partial behaviors (2.10a) and (2.10b) can be expressed together

$$\left.\begin{matrix} u_1(t) \Rightarrow y_1(t) \\ u_2(t) \Rightarrow y_2(t) \end{matrix}\right\} \quad a_1u_1(t)+a_2u_2(t) \Rightarrow a_1y_1(t)+a_2y_2(t) \tag{2.11}$$

where a, a_1, a_2 are any constants; $u(t)$, $u_1(t)$, $u_2(t)$ – the input variables; $y(t)$, $y_1(t)$, $y_2(t)$ – the output variables.

The linearity of a dynamic system means that a weighting sum of output variables corresponds to a weighting sum of input variables.

Another very important behavior of linear dynamic models (systems) is: every local behavior of a linear dynamic system is at the same time its global behavior.

A linear SISO system can be described in the time domain by a linear differential equation with constant coefficients (with lumped parameters)

$$a_n y^{(n)}(t)+\ldots+a_1\dot{y}(t)+a_0y(t)=b_m u^{(m)}(t)+\ldots+b_1\dot{u}(t)+b_0u(t) \tag{2.12}$$

with initial conditions

$$y(0)=y_0, \dot{y}(0)=\dot{y}_0, \ldots, y^{(n-1)}(0)=y_0^{(n-1)} \tag{2.13a}$$

$$u(0)=u_0, \dot{u}(0)=\dot{u}_0, \ldots, u^{(m-1)}(0)=u_0^{(m-1)} \tag{2.13b}$$

A static characteristic of a linear dynamic system is a straight line, which goes through a co-ordinate´s origin (Fig. 2.3). It can be obtained simply from a differential equation (2.12) for (2.7)

$$y(t)=\frac{b_0}{a_0}u(t), \; a_0 \neq 0 \tag{2.14}$$

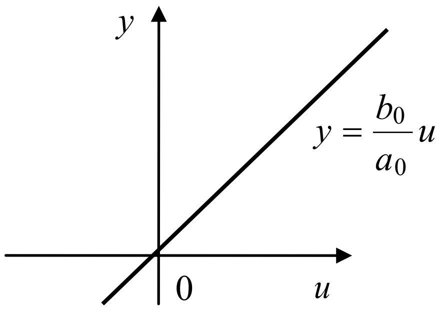


Fig. 2.3 – Linear static characteristic

If a linear dynamic system is described by a linear differential equation (2.12), then for the given initial conditions (2.13) and the given course of an input variable $u(t)$, it is possible to determine the course of output variable $y(t)$. This task is very demanding in a time domain, because it requires very good knowledge of a differential equation theory. The use of the Laplace transform is considerably easier. After an application of Laplace transform on a linear differential equation (2.12) together with initial conditions (2.13) an algebraic equation is obtained

$$(a_n s^n + \ldots + a_1 s + a_0)Y(s) - L(s) = (b_m s^m + \ldots + b_1 s + b_0)U(s) - R(s) \quad (2.15)$$

where $Y(s)$ is an output variable $y(t)$ transform; $U(s)$ – an input variable $u(t)$ transform; s – a complex variable in Laplace transform; $L(s)$ – a polynomial of the highest degree $(n - 1)$, which is determined by initial conditions (2.13a); $R(s)$ – a polynomial of the highest degree $(m - 1)$, which is determined by initial conditions (2.13b).

The dimension of complex variable s is [s^{-1}], generally [time^{-1}].

The transform of the solution can be determined from (2.15)

$$Y(s) = \frac{M(s)}{N(s)}U(s) + \frac{L(s) - R(s)}{N(s)} \quad (2.16)$$

$$N(s) = a_n s^n + \ldots + a_1 s + a_0 = a_n (s - s_1)(s - s_2)\ldots(s - s_n) \quad (2.17)$$

$$M(s) = b_m s^m + \ldots + b_1 s + b_0 = b_m (s - z_1)(s - z_2)\ldots(s - z_m) \quad (2.18)$$

where $N(s)$ is a **characteristic polynomial** of the degree n of a linear differential equation (2.12) (as well as a linear dynamic system), which is determined by its left-hand side coefficients; $M(s)$ – a polynomial of the degree m, which is determined by its right-hand side coefficients; s_i –roots of the characteristic polynomial (2.17), z_j –roots of a polynomial (2.18).

The **original of the solution** $y(t)$ for $t \geq 0$ can be obtained from the transform of the solution (2.16) on the basis of Laplace transform

$$y(t) = \mathrm{L}^{-1}\{Y(s)\} \tag{2.19}$$

The procedure is shown in Fig. 2.4.

The first part of the solution (2.16) is a transform of the response to an input variable $u(t)$, the second part of the solution (2.16) is the response to initial conditions (2.13).

On the assumption that initial conditions are zeros, i.e.

$$L(s) = 0 \text{ and } R(s) = 0$$

the transform of the solution has a form

$$Y(s) = G(s)U(s) \tag{2.20}$$

where the expression

$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_m s^m + \ldots + b_1 s + b_0}{a_n s^n + \ldots + a_1 s + a_0} = \frac{M(s)}{N(s)} \tag{2.21a}$$

is the **transfer function** of a linear dynamic system.

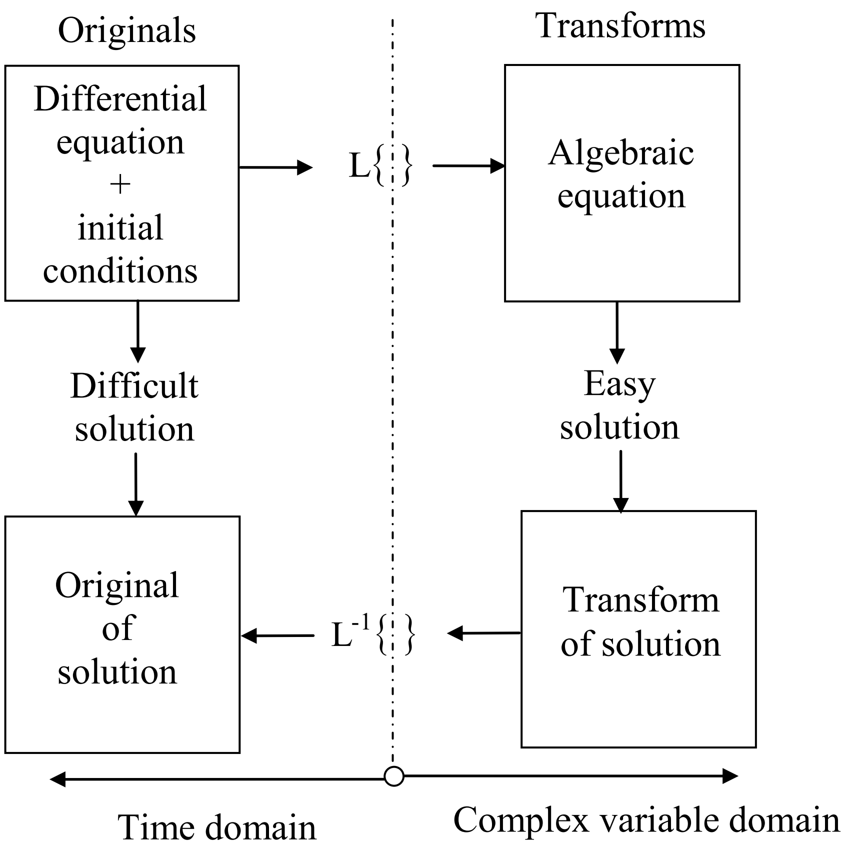


Fig. 2.4 – Solving a differential equation by the Laplace transform

The physical realizability conditions are given by relations (2.4) – (2.6).

A transfer function (2.21a) expresses a mathematical model of a given linear dynamic system for zero initial conditions in a **complex variable domain** and can be presented by the block diagram in Fig. 2.5.

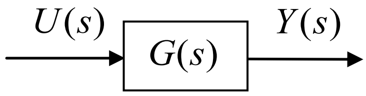


Fig. 2.5 – Block diagram of a system

For the following text zero initial conditions are supposed.

A transfer function (2.21a) can be written by means of linear dynamic system poles s_i ($i = 1, 2,\ldots, n$) and zeros z_j ($j = 1, 2,\ldots, m$)

$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_m(s - z_1)(s - z_2)\ldots(s - z_m)}{a_n(s - s_1)(s - s_2)\ldots(s - s_n)} \tag{2.21b}$$

A static characteristic of a linear dynamic system can be easily obtained from its transfer function ($a_0 \neq 0$)

$$y = \left\lfloor \lim_{s\to 0} G(s) \right\rfloor u \tag{2.22}$$

For a given course of the input variable $u(t)$ a corresponding course of a system response, i.e. the output variable $y(t)$ can be determined in accordance with the scheme

$$\begin{aligned} U(s) &= \mathrm{L}\{u(t)\} \\ Y(s) &= G(s)U(s) \\ y(t) &= \mathrm{L}^{-1}\{Y(s)\} \end{aligned} \tag{2.23}$$

For a linear dynamic system the responses to the **unit (Dirac) impulse** (Fig. 2.6)

$$\delta(t) = \begin{cases} 0 & \text{for } t \neq 0 \\ \infty & \text{for } t = 0 \end{cases}, \quad \int_{-\varepsilon}^{\varepsilon} \delta(t)\mathrm{d}t = 1, \ \varepsilon > 0 \tag{2.24a}$$

$$\mathrm{L}\{\delta(t)\} = 1 \tag{2.24b}$$

and the unit (Heaviside) step (Fig. 2.7)

$$\eta(t) = \begin{cases} 1 & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases} \tag{2.25a}$$

$$\mathrm{L}\{\eta(t)\} = \frac{1}{s} \tag{2.25b}$$

are very important.

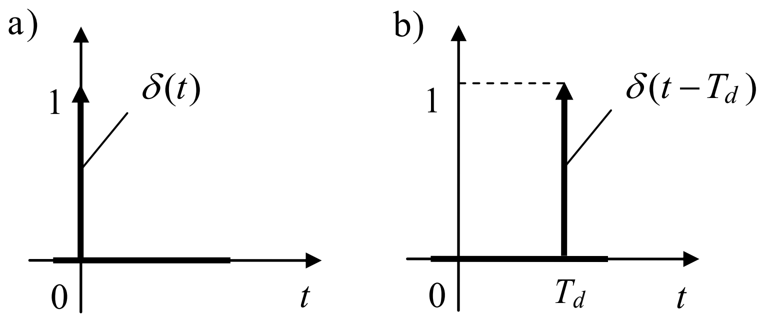


Fig. 2.6 – Unit impulse: a) undelayed, b) delayed

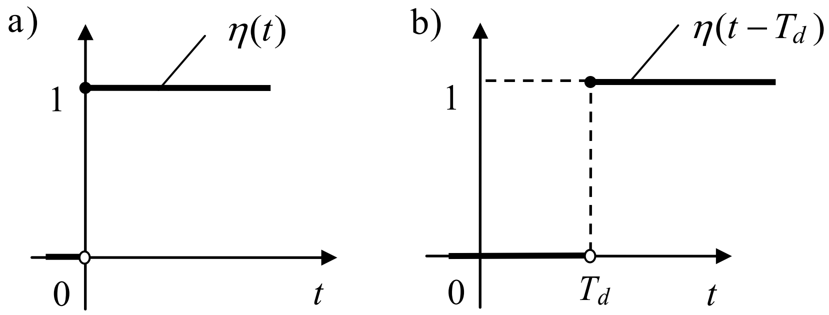


Fig. 2.7 – Unit step: a) undelayed, b) delayed

A linear dynamic system response to the unit impulse can be obtained on the basis of (2.23) and (2.24b)

$$y(t) = \mathrm{L}^{-1}\{Y(s)\} = \mathrm{L}^{-1}\{G(s)\} = g(t) \qquad (2.26)$$

The time function $g(t)$ is the original of a transfer function $G(s)$. It is called the **(unit) impulse response** (Fig. 2.8).

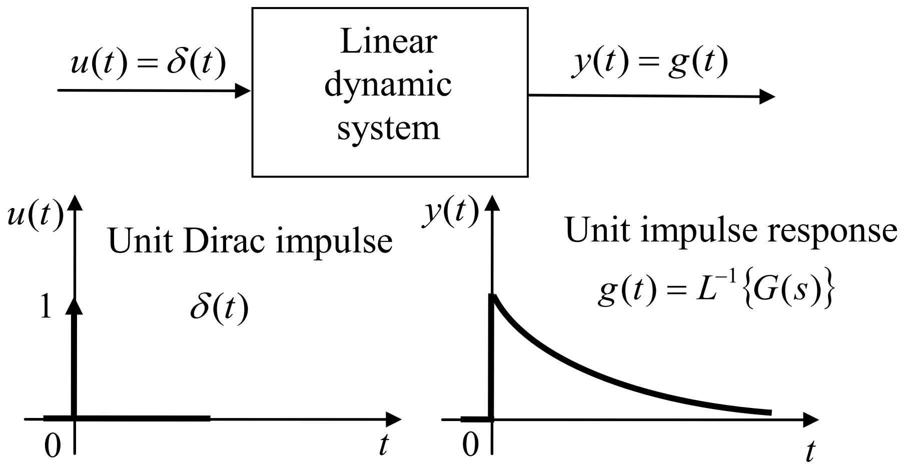


Fig. 2.8 – Unit impulse response of a linear dynamic system

A static characteristic (if it exists) is given by the relation

$$y = \left[\lim_{t\to\infty} \int_0^t g(\tau)\mathrm{d}\tau \right] u \tag{2.27}$$

For $|g(0)| < \infty$ a linear dynamic system is strongly physically realizable and for $g(0)$ containing the Dirac impulse $\delta(t)$ it is only weakly physically realizable.

A linear dynamic system response to the unit Heaviside step can be obtained on the basis of (2.23) and (2.25b)

$$y(t) = \mathrm{L}^{-1}\{Y(s)\} = \mathrm{L}^{-1}\left\{\frac{1}{s}G(s)\right\} = h(t) \tag{2.28}$$

A time function $h(t)$ is called the **(unit) step response** (Fig. 2.9).

A static characteristic (if it exists) is given by the relation

$$y = [\lim_{t\to\infty} h(t)]u \tag{2.29}$$

For $h(0) = 0$ a linear dynamic system is strongly physically realizable and for $0 < |h(0)| < \infty$ it is only weakly physically realizable.

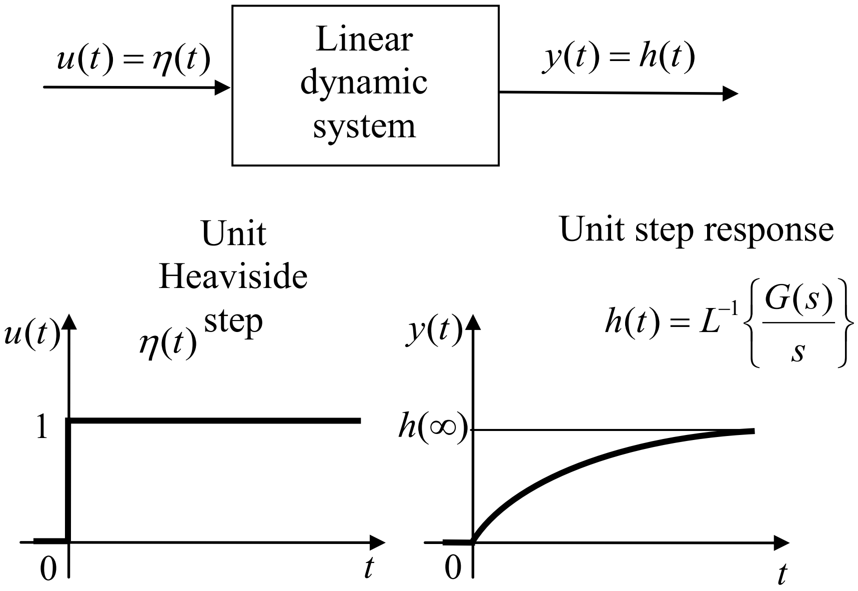


Fig. 2.9 – Unit step response of a linear dynamic system

The use of a generalized derivative is advantageous. It is defined by the relations

$$\begin{aligned} \dot{x}(t) &= \dot{x}_{or}(t) + \sum_i \Delta_i \delta(t - t_i) \\ \Delta_i &= \lim_{t \to t_{i+}} x(t) - \lim_{t \to t_{i-}} x(t) \end{aligned} \tag{2.30}$$

where t_i is the discontinuity points of the first kind with the steps Δ_i, $\dot{x}_{or}(t)$ – the ordinary derivative, which is determined out of the discontinuity points.

On use of the generalized derivative (2.30) it is possible to write

$$\delta(t) = \frac{\mathrm{d}\eta(t)}{\mathrm{d}t} \Leftrightarrow \eta(t) = \int_0^t \delta(\tau)\mathrm{d}\tau \tag{2.31}$$

$$g(t) = \frac{\mathrm{d}h(t)}{\mathrm{d}t} \Leftrightarrow h(t) = \int_0^t g(\tau)\mathrm{d}\tau \tag{2.32}$$

$$G(s) = sH(s) \Leftrightarrow H(s) = \frac{1}{s}G(s) \tag{2.33}$$

A mathematical model of a linear dynamic system in a state space has the form (Fig. 2.10)

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}u(t),\ \boldsymbol{x}(0) = \boldsymbol{x}_0 \text{ – the state equation} \tag{2.34a}$$

$$y(t) = \boldsymbol{c}^T \boldsymbol{x}(t) + du(t) \text{ – the output equation} \tag{2.34b}$$

where $\boldsymbol{A}$ is the square system matrix (n x n), $\boldsymbol{b}$ – the column input vector (n x 1), $\boldsymbol{c}^T$ – the row output vector (1 x n), d – the transfer constant, $\boldsymbol{x}(t)$ – the vector of the **state variables**.

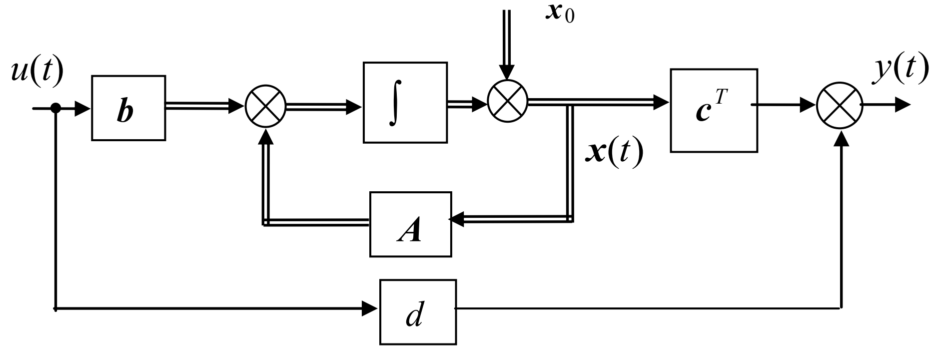


Fig. 2.10 – Block diagram of a SISO state space model

For $d = 0$ a mathematical model (2.34) satisfies the strong physical condition and for $d \neq 0$ only the weak physical realizability condition.

If a mathematical model (2.34) fulfills the **controllability condition**

$$\text{rank}[\boldsymbol{b}, \boldsymbol{A}\boldsymbol{b}, \ldots, \boldsymbol{A}^{n-1}\boldsymbol{b}] = n \;\Leftrightarrow\; \det[\boldsymbol{b}, \boldsymbol{A}\boldsymbol{b}, \ldots, \boldsymbol{A}^{n-1}\boldsymbol{b}] \neq 0 \qquad (2.35)$$

and the observability condition

$$\text{rank}[\boldsymbol{c}, \boldsymbol{A}^T\boldsymbol{c}, \ldots, (\boldsymbol{A}^T)^{n-1}\boldsymbol{c}] = n \;\Leftrightarrow\; \det[\boldsymbol{c}, \boldsymbol{A}^T\boldsymbol{c}, \ldots, (\boldsymbol{A}^T)^{n-1}\boldsymbol{c}] \neq 0 \qquad (2.36)$$

then on the assumption that the initial conditions are zeros, on the basis of the Laplace transform from (2.34) the transfer function can be determined

$$\left.\begin{aligned} s\boldsymbol{X}(s) &= \boldsymbol{A}\boldsymbol{X}(s) + \boldsymbol{b}U(s) \\ Y(s) &= \boldsymbol{c}^T\boldsymbol{X}(s) + dU(s) \end{aligned}\right\} \; G(s) = \frac{Y(s)}{U(s)} = \boldsymbol{c}^T(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{b} + d \qquad (2.37)$$

where rank is a matrix rank, det – a determinant of the square matrix.

The relation (2.37) for practical use is not suitable, because it demands an inversion of the functional matrix. Considerably preferable is the following relation

$$G(s) = \frac{Y(s)}{U(s)} = \frac{\det(s\boldsymbol{I} - \boldsymbol{A} + \boldsymbol{b}\boldsymbol{c}^T) - \det(s\boldsymbol{I} - \boldsymbol{A})}{\det(s\boldsymbol{I} - \boldsymbol{A})} + d \qquad (2.38)$$

A characteristic polynomial of a linear dynamic system with a mathematical model (2.37) is given in accordance with (2.38)

$$\begin{aligned} N(s) = \det(s\boldsymbol{I} - \boldsymbol{A}) &= s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0 = \\ &= (s - s_1)(s - s_2)\ldots(s - s_n) \end{aligned} \qquad (2.39)$$

where s_i are the **eigenvalues** of the matrix $\boldsymbol{A}$.

It is obvious that the poles s_i of a linear dynamic system are given by the eigenvalues of a square system matrix $\boldsymbol{A}$.

A static characteristic (if it exists) can be determined from a transfer function (2.37) or (2.38) on the basis of (2.22).

On the assumption of zero initial conditions and fulfillment of the controllability (2.35) and observability (2.36) conditions a transfer function (2.37) or (2.38) is determined uniquely. A transformation of the transfer function in a state space model is more complicated and non-unique. A state space model of a linear dynamic system can have many different forms. It depends on the choice of the **state variables** $\boldsymbol{x}(t) = [x_1(t), x_2(t),\ldots, x_n(t)]^T$. These variables are "internal" variables, and therefore a state space model is often called the **internal model** in contrast to the previous mathematical models, which are called the **external models**.

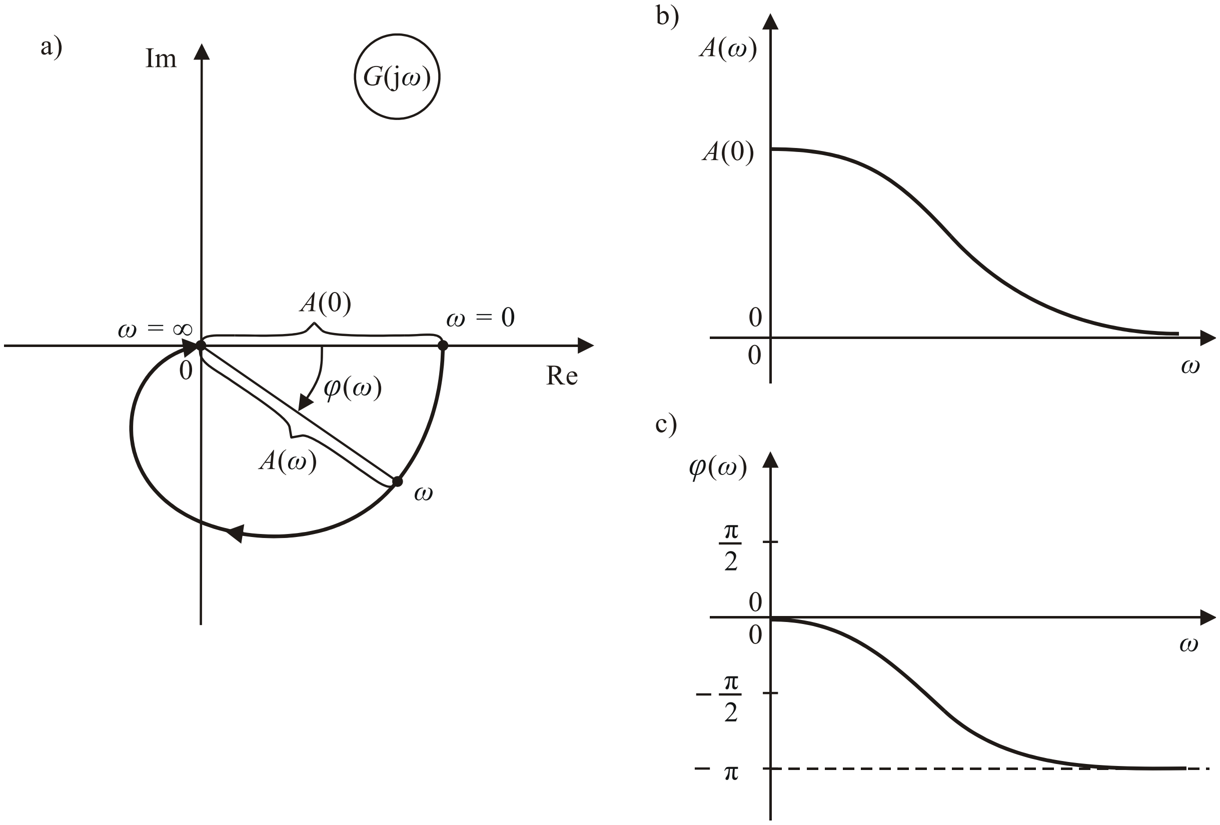


Fig. 2.11 – Frequency responses: a) polar plot, b) magnitude frequency response, c) phase frequency response

A description of the linear dynamic system in the **frequency domain** is very important. This description is based on the **frequency transfer function**, which can be obtained from a transfer function $G(s)$ by replacement of the complex variable s with "complex frequency" jω, i.e.

$$G(\mathrm{j}\omega) = G(s)\big|_{s=\mathrm{j}\omega} = \frac{b_m(\mathrm{j}\omega)^m + \ldots + b_1\mathrm{j}\omega + b_0}{a_n(\mathrm{j}\omega)^n + \ldots + a_1\mathrm{j}\omega + a_0} \tag{2.40}$$

$$A(\omega) = \operatorname{mod} G(\mathrm{j}\omega) = |G(\mathrm{j}\omega)| \tag{2.41a}$$

$$\varphi(\omega) = \arg G(\mathrm{j}\omega) \qquad (2.41b)$$

where ω is the **angular frequency** or **pulsation**, $\mathrm{j} = \sqrt{-1}$ – the imaginary unit, $A(\omega)$ – the **modulus** or **magnitude** of the frequency transfer function, $\varphi(\omega)$ – the **phase** or **phase-angle** of the frequency transfer function.

The dimension of an angular frequency ω is the same as the dimension of a complex variable s, i.e. [s^{-1}] or generally [$time^{-1}$], but for the reason to make a distinction of the "ordinary" frequency with unit [s^{-1}] and the name Hz from an angular frequency, the unit [$rad.s^{-1}$] is very often used.

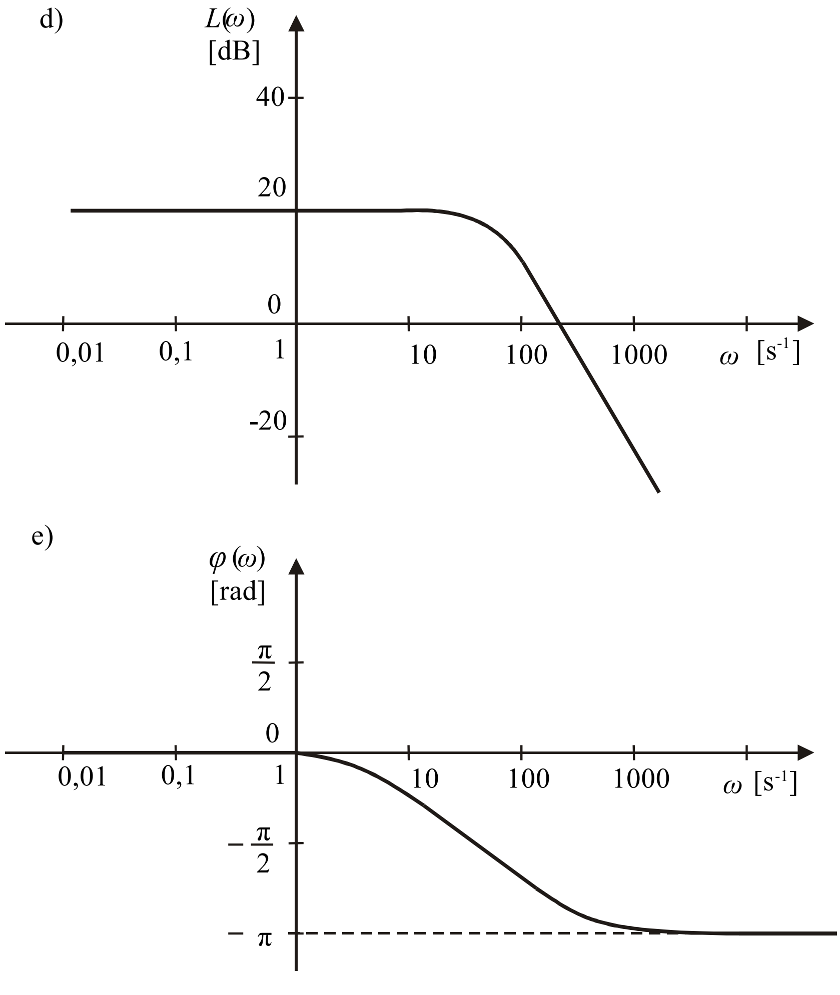


Fig. 2.11 – Frequency responses: d) Bode magnitude plot, e) Bode phase plot

Mapping of a frequency transfer function to the angular frequency in a complex plane from $\omega = 0$ to $\omega = \infty$ is called a **polar plot** or **frequency response** (Fig. 2.11a). A selected mapping of the modulus (magnitude) $A(\omega)$ and the phase $\varphi(\omega)$ from $\omega = 0$ to $\omega = \infty$ is called the **magnitude frequency response** (Fig. 2.11b) and the **phase frequency response** (Fig. 2.11c). For

$$L(\omega) = 20 \log A(\omega) \qquad (2.41c)$$

Bode plots are obtained, i.e. **Bode magnitude plot** (Fig. 2.11d) and **Bode phase plot** (Fig. 2.11e). $L(\omega)$ is the **logarithmic modulus** or **logarithmic magnitude (gain)** [dB] of a frequency transfer function (2.40). For Bode plots the approximation is used on the basis of the line sections and asymptotical lines, i.e. (Fig. 3.5).

The frequency transfer function is very important for practice, because for every angular frequency ω it expresses the magnitude (amplitude) $A(\omega)$ and the phase $\varphi(\omega)$ of the steady-state harmonic response to the harmonic input with a unit amplitude and a zero phase. It means that the frequency response can be obtained experimentally, and therefore it can be used for the **experimental identification** (Fig. 2.12).

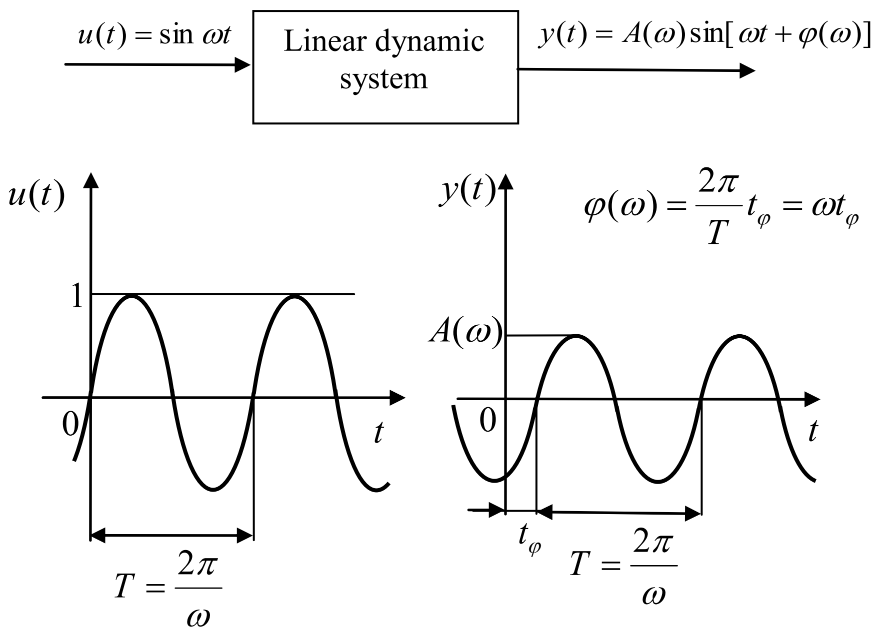


Fig. 2.12 – Interpretation of a frequency response of a linear dynamic system

The physical realizability conditions are given by relations (2.4) – (2.6). In case of a frequency transfer function (2.40) they have a very visual physical interpretation. Since a frequency transfer function $G(\mathrm{j}\omega)$ describes the transmission of a harmonic signal through a linear dynamic system for different angular frequencies ω, it is obvious that the real linear dynamic system cannot transmit a signal with infinity angular frequency, and this is why it must hold for mathematical models of the physically realizable linear dynamic systems

$$\left.\begin{aligned} &\lim_{\omega\to\infty} G(\mathrm{j}\omega) = 0 \\ &\lim_{\omega\to\infty} A(\omega) = 0 \\ &\lim_{\omega\to\infty} L(\omega) = -\infty \end{aligned}\right\} \Rightarrow n > m$$

It is a strong realizability condition. For the steady-state $t \to \infty \Rightarrow \omega \to 0$ holds, and therefore the static characteristic is given

$$y = [\lim_{\omega \to 0} G(\mathrm{j}\omega)]u\,, \quad a_0 \neq 0 \tag{2.42}$$

2.2 Block Diagram Algebra

A great advantage of the description of the linear dynamic systems by the transfer functions is the possibility to use the block diagrams. Every linear dynamic system is presented by a block with its inscribed transfer function (Fig. 2.13a), the addition or subtraction of the variables (signals) are presented by the summing nodes (Fig. 2.13b) and the variable (signal) branching is presented by the information node (Fig. 2.13c).

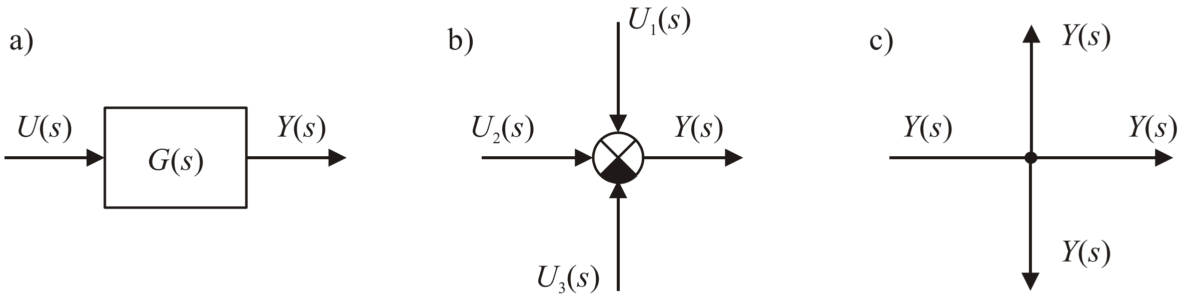


Fig. 2.13 – Representation: a) a linear dynamic system by a block, b) variable addition or subtraction by a summing node, c) variable branching by an information node

For a block in Fig. 2.13a it holds

$$Y(s) = G(s)U(s)$$

and for the summing node in Fig. 2.13b

$$Y(s) = U_1(s) + U_2(s) - U_3(s)\,.$$

Only one output from the summing node can go out.

The filled segment of the summing node expresses the minus sign. Besides the filled segment the sign “-“ is often used too.

The function of an information node is obvious.

On the basis of the blocks and on the summing and information nodes very complicated block diagrams can be created, which can always be reduced into three basic block interconnections: serial (cascade), parallel and feedback.

Serial Interconnection

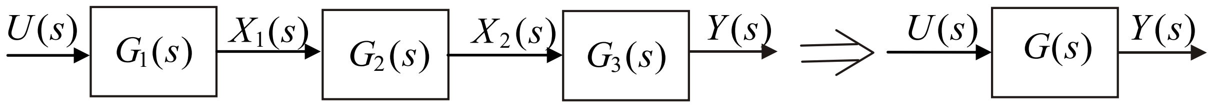


Fig. 2.14 – Serial interconnection of blocks

For the serial (cascade) interconnection of the blocks in Fig. 2.14 it holds

$$\left.\begin{aligned} Y(s) &= G_3(s)X_2(s) \\ X_2(s) &= G_2(s)X_1(s) \\ X_1(s) &= G_1(s)U(s) \end{aligned}\right\} \Rightarrow \frac{Y(s)}{U(s)} = G(s) = G_1(s)G_2(s)G_3(s)\,. \qquad (2.43)$$

For the serial interconnection of the blocks the resultant transfer function is given by the multiplication of the transfer functions of the separate blocks (it does not depend on the succession of the transfer functions).

Parallel Interconnection

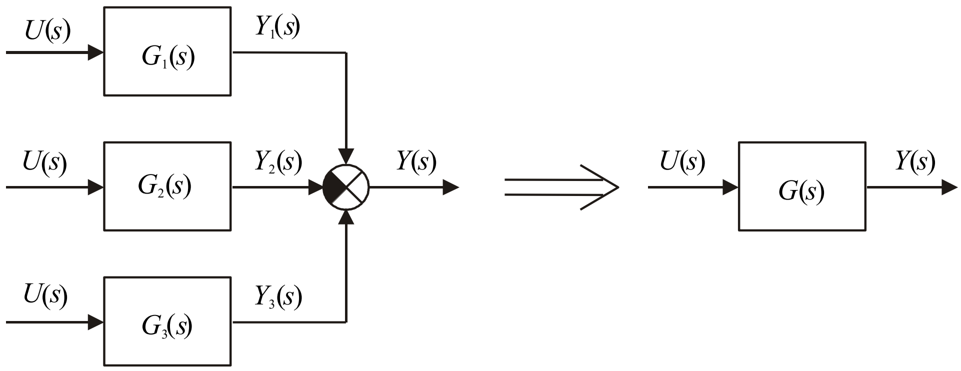


Fig. 2.15 – Parallel interconnection of blocks

For the parallel interconnection of the blocks in Fig. 2.15 it holds

$$\left.\begin{aligned} Y(s) &= Y_1(s) - Y_2(s) + Y_3(s) \\ Y_1(s) &= G_1(s)U(s) \\ Y_2(s) &= G_2(s)U(s) \\ Y_3(s) &= G_3(s)U(s) \end{aligned}\right\} \Rightarrow \frac{Y(s)}{U(s)} = G(s) = G_1(s) - G_2(s) + G_3(s) \qquad (2.44)$$

For the parallel interconnection of the blocks the resultant transfer function is given by the summation of the transfer functions of the separate blocks (the signs of the separate transfer functions must be taken into account, the signs at the summing node).

It is obvious that the number of blocks for the serial (cascade) and parallel interconnections can be arbitrary.

Feedback Interconnection

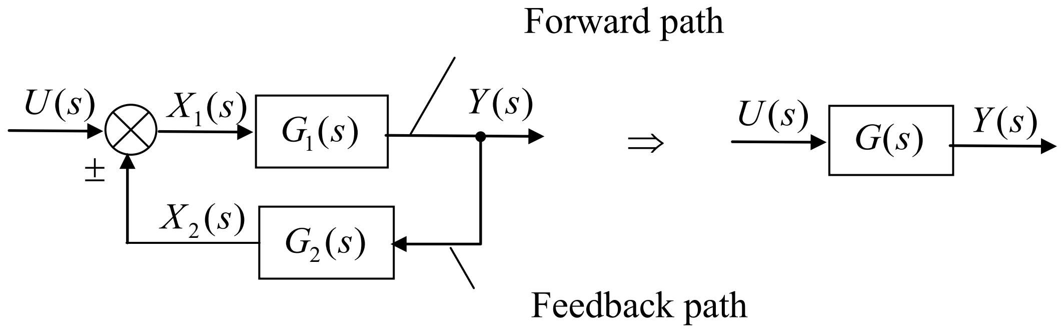


Fig. 2.16 – Feedback interconnection of blocks

The feedback interconnection of the blocks in Fig. 2.16 is very important, because it is the ground for all theory of automatic control. For the feedback interconnection of the blocks in Fig. 2.16 it holds

$$\left.\begin{aligned} Y(s) &= G_1(s)X_1(s) \\ X_1(s) &= U(s) \pm X_2(s) \\ X_2(s) &= G_2(s)Y(s) \end{aligned}\right\} \Rightarrow \frac{Y(s)}{U(s)} = G(s) = \frac{G_1(s)}{1 \mp G_1(s)G_2(s)} \qquad (2.45)$$

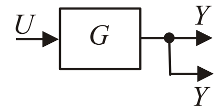
For the feedback interconnection of the blocks the resultant transfer function is given by the transfer function in the forward path (branch) divided by the negative (in case of sign "+") or the positive (in case of sign "-") product of the transfer function in the forward path and the transfer function in the feedback path increased by one. The transfer function of the branch without the block (a transfer function) is a unit.

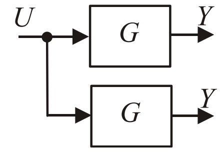
If we know these three basic interconnections of the blocks we can reduce any complicated block diagram. We can use the Tab. 2.1. For the reason of simplicity the independent variable *s* is not often explicitly written in the block diagrams.

If the block diagram contains more input and output variables, for every output variable the input variables are successively considered. The input variables, which are not considered, are supposed to be zero (they aren't drawn). The resultant transfer functions are given on the basis of the linearity principle by the summation of the influence of the separate input variables. For the reason of unity the resultant transfer function uses subscripts. The first subscript indicates the input variable and the second subscript the output variable.

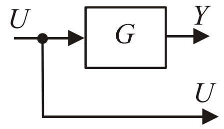
Tab. 2.1 – Basic Block Diagram Transformations

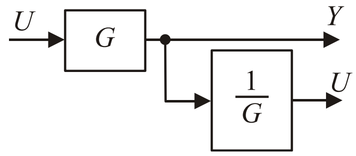
Moving an information node ahead of a block



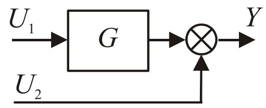


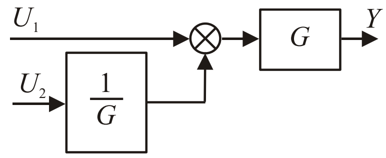
Moving an information node behind a block



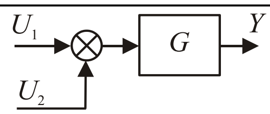


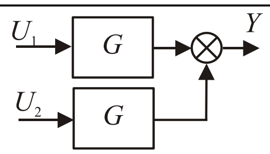
Moving a summing node behind a block



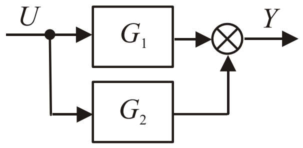


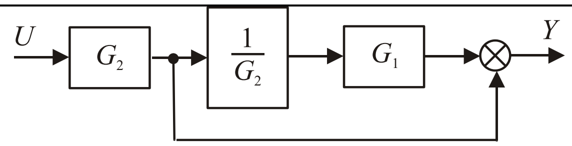
Moving a summing node ahead of a block



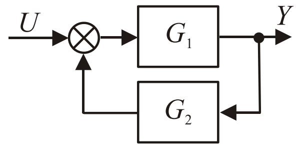


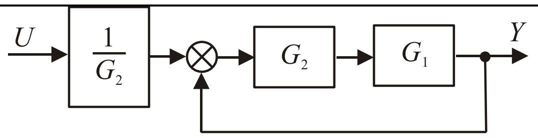
Moving a block from a parallel interconnection





Moving a block from a feedback interconnection





2.3 Linearization

In the previous subchapters we considered that all real systems (elements, plants, processes etc.) are linear. In reality all real systems are non-linear, i.e.

their static and dynamic behaviors can be non-linear. If the non-linear behavior of a given dynamic system is not substantial, then its behavior can be described for small variable changes in the surroundings of the **operating point** by a linear mathematical model. The linear mathematical model for a given or selected operating point can be obtained from a non-linear mathematical model by the **linearization**.

There exist many different linearization methods. The simplest method only linearizes the non-linear static characteristics by analytical or graphical ways. The more complex methods use optimization of some criteria. The least squares method and its different modifications are often used.

If a static mathematical model of a system has only one output variable y and m input variables $u_1, u_2, \ldots, u_m$, i.e.

$$y = f(u_1, u_2, \ldots, u_m) \tag{2.46}$$

then it is suitable to use in the operating point

$$y_0 = f(u_{10}, u_{20}, \ldots, u_{m0}) \tag{2.47}$$

an approximation on the basis of the tangent plane

$$\hat{y} = y_0 + \Delta y \tag{2.48}$$

where

$$\Delta y = k_1 \Delta u_1 + k_2 \Delta u_2 + \ldots + k_m \Delta u_m \tag{2.49}$$

is an increment of the output variable, i.e. $\Delta y = y - y_0$; and $\Delta u_1 = u_1 - u_{10}$, $\Delta u_2 = u_2 - u_{20}, \ldots, \Delta u_m = u_m - u_{m0}$ are the increments of the corresponding input variables, and

$$k_1 = \left.\frac{\partial f}{\partial u_1}\right|_0, k_2 = \left.\frac{\partial f}{\partial u_2}\right|_0, \ldots, k_m = \left.\frac{\partial f}{\partial u_m}\right|_0 \tag{2.50}$$

are the partial derivatives determined in the operating point (2.47), and $\hat{y}$ is an output variable in the absolute form, which was obtained after linearization. From a geometrical interpretation for one input (Fig. 2.17) it follows that the coefficient k_1 is the angular coefficient of a tangent line.

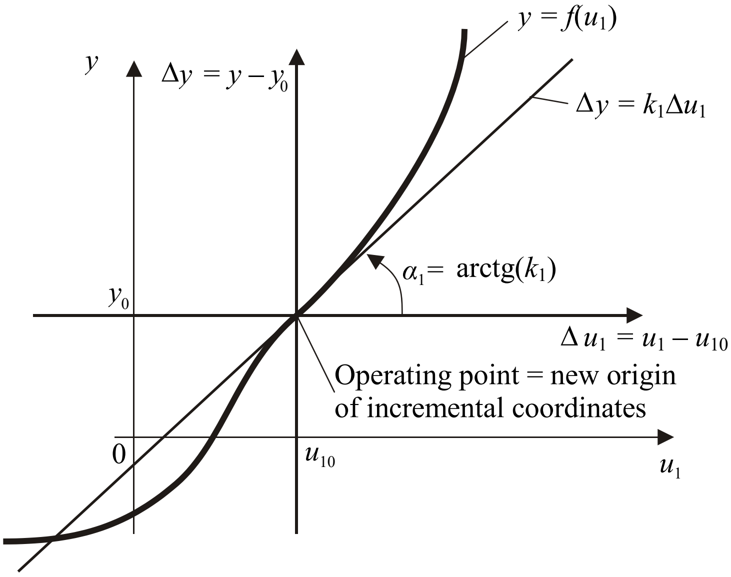


Fig. 2.17 – Geometrical interpretation of linearization by a tangent line for one input

The linearization on the basis of the tangent plane can be only used in a case that the partial derivatives (2.50) exist and they are continuous. After linearization the new origin in incremental coordinates (variables) must be regarded in the operating point (2.47), see Fig. 2.17.

It is obvious that the linearization on the basis of the tangent plane can keep its quality only for the small surrounding of the operating point.

In case of the differential equations, e.g. for the derivative of the i order with respect to time it holds

$$\frac{\mathrm{d}^i y(t)}{\mathrm{d}t^i} = \frac{\mathrm{d}^i[y_0 + \Delta y(t)]}{\mathrm{d}t^i} = \frac{\mathrm{d}^i \Delta y(t)}{\mathrm{d}t^i} \tag{2.51}$$

because y_0 = const.

If the linearized mathematical model is complex, then it is useful to divide it into simpler relations (models), and to linearize these simpler relations and then to determine the resultant linear relation by the substitution. The algebra of a block diagram can be used to great advantage.

3 FEEDBACK CONTROL SYSTEMS

This chapter is devoted to a description and an analysis of a control system. Conventional linear analog controllers and simple identification methods for basic plants are presented. The verification of the stability of the control systems is described.

3.1 Controllers

A control system in Fig. 3.1 is considered, where $G_C(s)$ is the controller transfer function, $G_P(s)$ – the plant transfer function, $G_S(s)$ – the sensor transfer function, $G_V(s)$ – the disturbance allocating transfer function, $W(s)$ – the transform of the desired (reference) variable $w(t)$, $E(s)$ – the transform of the control error $e(t)$, $U(s)$ – the transform of the control (manipulated, actuating) variable $u(t)$, $Y(s)$ – the transform of the controlled variable $y(t)$.

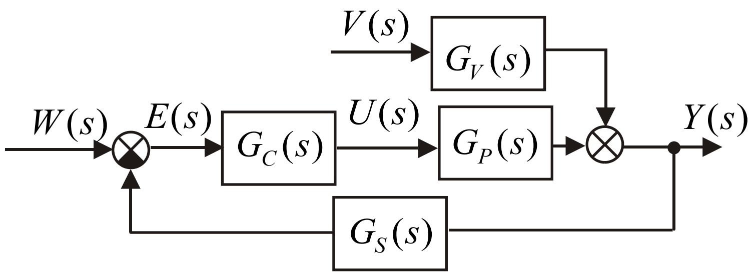


Fig. 3.1 – Block diagram of a common control system

For the reason of simplicity in lieu of the term “transform of variable” we will only use “variable”.

A sensor (measuring device) with a transfer function $G_S(s)$ must measure precisely and fast, therefore we may suppose that in practical cases its transfer function is unit, i.e.

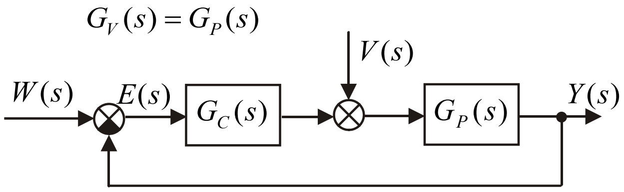
$$G_S(s) = 1 \tag{3.1}$$

The controlled variable $Y(s)$ can be obtained from the sensor, that’s why a sensor is very often assigned to the plant.

The transfer function $G_V(s)$ enables allocating the disturbance $V(s)$ in any place in a control system. Two most important cases are in Fig. 3.2.

If disturbance variables cannot be measured or they are uncertain, then they are aggregated in a one disturbance variable $V(s)$ and a disturbance is then allocated in the least advantageous place of a control system. In this case, it is the plant’s input of an integrating plant (Fig. 3.2a) and the plant’s output in the case of a proportional plant (Fig. 3.2b).

a) b)



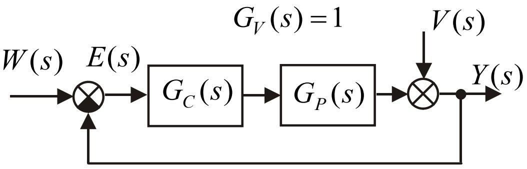


Fig. 3.2 – Control system with disturbance: a) in the input of a plant, b) in the output of a plant

As noted previously, with the assumption that the condition (3.1) holds (the closed-loop control system with a unit feedback), the control objective for the control system in Fig. 3.1 can be expressed in two equivalent forms.

The control objective in the form:

$$y(t) \to w(t) \mathrel{\hat{=}} Y(s) \to W(s) \tag{3.2}$$

In accordance with Fig. 3.1 and (3.1) for the controlled variable holds

$$Y(s) = G_{wy}(s)W(s) + G_{vy}(s)V(s) \tag{3.3}$$

where

$$G_{wy}(s) = \frac{G_C(s)G_P(s)}{1 + G_C(s)G_P(s)} \tag{3.4}$$

is the desired variable to the controlled variable transfer function or the **closed-loop transfer function** (the control system transfer function) and

$$G_{vy}(s) = \frac{G_V(s)}{1 + G_C(s)G_P(s)} = [1 - G_{wy}(s)]G_V(s) \tag{3.5}$$

is the disturbance variable to the controlled variable transfer function or the **disturbance transfer function**.

It is obvious that for fulfillment of the control objective (3.2) for any desired variable $W(s)$ and any disturbance variable $V(s)$ the conditions

$$G_{wy}(s) \to 1 \qquad \text{servo (tracking) problem} \tag{3.6}$$

and

$$G_{vy}(s) \to 0 \qquad \text{regulatory problem} \tag{3.7}$$

must hold.

The first condition for the closed-loop transfer function (3.6) expresses the controller function, which consists in the following desired variable $W(s)$ by the controlled variable $Y(s)$ – it is the **servo** or **tracking problem**. The second condition for the disturbance transfer function (3.7) expresses the controller function, which consists in the disturbance $V(s)$ rejection (attenuation) – it is the **regulatory problem**.

The control objective in the form:

$$e(t) \to 0 \mathrel{\hat{=}} E(s) \to 0 \tag{3.8}$$

In accordance with Fig. 3.1 and (3.1) for the control error holds

$$E(s) = G_{we}(s)W(s) + G_{ve}(s)V(s) \tag{3.9}$$

where

$$G_{we}(s) = \frac{1}{1 + G_C(s)G_P(s)} = 1 - G_{wy}(s) \tag{3.10}$$

is the desired variable to the control error transfer function and

$$G_{ve}(s) = -\frac{G_V(s)}{1 + G_C(s)G_P(s)} = -[1 - G_{wy}(s)]G_V(s) \tag{3.11}$$

is the disturbance variable to the control error transfer function.

It is obvious that for fulfillment of the control objective (3.8) for any desired variable $W(s)$ and any disturbance variable $V(s)$ the conditions

$$G_{we}(s) \to 0 \qquad \text{servo (tracking) problem} \tag{3.12}$$

and

$$G_{ve}(s) \to 0 \qquad \text{regulatory problem} \tag{3.13}$$

must hold.

Similarly like in previous case, the first condition for the desired variable to the control error transfer function (3.12) expresses the servo problem and the second condition for the disturbance variable to the control error transfer function (3.13) expresses the regulatory problem.

It is obvious that both formulations (3.2) and (3.8) of the control objective are equivalent and therefore further we will use the control objective in the form (3.2).

The controller will operate correctly if the conditions (3.6) and (3.7) [or (3.12) and (3.13)] will hold at the same time. If the disturbance variable $V(s)$ is effected in the plant output (Fig. 3.2b) then both conditions are equivalent (it is the most frequent case), i.e. if the condition (3.6) holds then automatically the condition (3.7) holds. Therefore, in automatic control theory attention is devoted to the closed-loop transfer function (3.4). The transfer functions (3.4), (3.5), (3.10) and (3.11) are called the **basic transfer functions** of the control system.

In accordance with (3.4) for the frequency closed-loop transfer function there can be written

$$G_{wy}(\mathrm{j}\omega) = \frac{G_C(\mathrm{j}\omega)G_P(\mathrm{j}\omega)}{1 + G_C(\mathrm{j}\omega)G_P(\mathrm{j}\omega)} = \frac{1}{\dfrac{1}{G_C(\mathrm{j}\omega)G_P(\mathrm{j}\omega)} + 1} \tag{3.14}$$

and it is obvious that relations

$$\left.\begin{array}{l} |G_C(\mathrm{j}\omega)| \to \infty \\ G_P(\mathrm{j}\omega) \neq 0 \end{array}\right\} \Rightarrow G_{wy}(\mathrm{j}\omega) \to 1 \Rightarrow G_{wy}(s) \to 1 \tag{3.15}$$

or

$$|G_C(\mathrm{j}\omega)G_P(\mathrm{j}\omega)| \to \infty \Rightarrow G_{wy}(\mathrm{j}\omega) \to 1 \Rightarrow G_{wy}(s) \to 1 \tag{3.16}$$

hold.

From (3.15) it follows that if the satisfactory high controller modulus will be ensured

$$A_C(\omega) = \operatorname{mod} G_C(\mathrm{j}\omega) = |G_C(\mathrm{j}\omega)|, \tag{3.17}$$

then the condition (3.6) will hold with adequate accuracy and for non-singular $G_V(s)$ the condition (3.7) as well.

If the plant behavior expressed by the transfer function $G_P(s)$ is known then it is easier to ensure the high modulus of the frequency open-loop transfer function

$$A_o(\omega) = \operatorname{mod} G_o(\mathrm{j}\omega) = |G_o(\mathrm{j}\omega)| = |G_C(\mathrm{j}\omega)G_P(\mathrm{j}\omega)| \tag{3.18}$$

see (3.16).

The high moduli $A_C(\omega)$ or $A_o(\omega)$ must be ensured for the band of the operating frequencies and at the same time for the **stability** and **desired performance** of the control system. This is practical by a suitable **controller choice** and its successive **controller tuning**.

The industrial controllers are made in different versions and modifications, and therefore only basic structures and modifications of the commonly used controllers will be presented.

Analog (continuous) conventional controllers are implemented as a combination of three components (terms): **proportional – P**, **integral – I** and **derivative – D**. The controller with all three components is called the **proportional plus integral plus derivative controller** or the **PID controller**. Its behavior can be described by the relation

$$u(t)=\underbrace{K_P e(t)}_{P}+\underbrace{K_I\int_0^t e(\tau)\,\mathrm{d}\tau}_{I}+\underbrace{K_D\frac{\mathrm{d}e(t)}{\mathrm{d}t}}_{D}=K_P\left[e(t)+\frac{1}{T_I}\int_0^t e(\tau)\,\mathrm{d}\tau+T_D\frac{\mathrm{d}e(t)}{\mathrm{d}t}\right] \tag{3.19}$$

where K_P, K_I and K_D are the **proportional**, **integral** and **derivative component weights**, K_P – the **controller gain** (the proportional component weight), T_I – the **integral time**, T_D – the **derivative time**.

In industrial controllers the **proportional band**

$$pp=\frac{100}{K_P}\,[\%] \tag{3.20}$$

is often used.

The dimension of the proportional component weight K_P, i.e. the controller gain is given by the dimension of the control variable $u(t)$ divided by the dimension of the control error $e(t)$. The time constants T_I and T_D have the dimension of time [s]. The dimension of the integral component weight K_I is given by the dimension of K_P divided by time and the dimension of the derivative component weight K_D is given by the product of the dimension of K_P and time.

The parameters K_P, K_I and K_D or K_P, T_I and T_D are **adjustable controller parameters**. The task of controller tuning is to ensure the desired control performance by suitable tuning (setting) of the adjustable controller parameters for a given plant. Among the adjustable controller parameters the conversion relations hold

$$K_I=\frac{K_P}{T_I},\quad K_D=K_P T_D \tag{3.21}$$

or

$$T_I=\frac{K_P}{K_I},\quad T_D=\frac{K_D}{K_P} \tag{3.22}$$

After using the Laplace transform on relation (3.19) the controller transfer function

$$G_C(s)=\frac{U(s)}{E(s)}=K_P+\frac{K_I}{s}+K_D s=K_P\left(1+\frac{1}{T_I s}+T_D s\right) \tag{3.23}$$

is obtained.

In Fig. 3.3 there are drawn the courses of the moduli of the controller components P, I and D. From Fig. 3.3 it follows that the integral component (I) ensures the high value of the frequency transfer function modulus of the PID controller for small angular frequencies and especially for the steady state (ω = 0), the derivative component (D) for high angular frequencies and the proportional component for all angular frequencies (mainly for medium frequencies). In fact by a suitable choice of the particular components P, I and D, i.e. by the suitable setting of the adjustable controller parameter K_P, K_I and K_D or K_P, T_I and T_D it is possible to achieve a high modulus of the frequency controller transfer function (3.17) or a high modulus of the frequency open-loop transfer function (3.18), in order to fulfill the conditions (3.15) or (3.16).

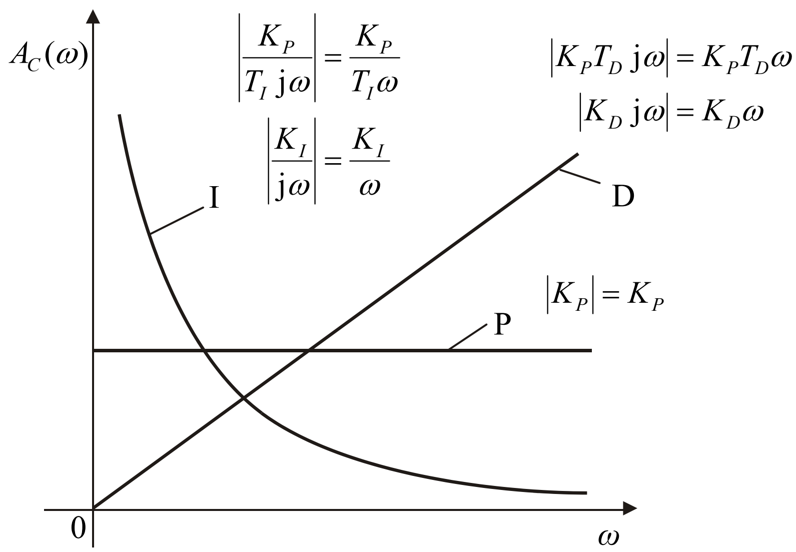


Fig. 3.3 – Dependence of partial controller components P, I and D of PID on angular frequency

Tab. 3.1 – Conventional analog controller transfer functions

	Type	Transfer function $G_C(s)$
1	P	K_P
2	I	$\frac{1}{T_I s}$
3	PI	$K_P\left(1+\frac{1}{T_I s}\right)$
4	PD	$K_P(1+T_D s)$
5	PID	$K_P\left(1+\frac{1}{T_I s}+T_D s\right)$
6	PID$_\mathrm{i}$	$K'_P\left(1+\frac{1}{T'_I s}\right)(1+T'_D s)$

In industrial practice simpler controllers are often used. They are: the P (proportional) controller, the I (integral) controller, the PI (proportional plus integral) controller and PD (proportional plus derivative) controller. Their transfer functions are in Tab. 3.1 (rows 1 – 5). The single D component is unusable because it only reacts to the derivative $\dot{e}(t)$ and therefore in a steady state it causes a disconnection of the control system.

The block diagram of the PID controller with the transfer function (3.23) is in Fig. 3.4a. From the Fig. 3.4a it follows that it has a parallel structure. The adjustable parameters of this controller can be tuned independently. Therefore this PID controller is **without interaction (non-interacting)**.

a)

b)

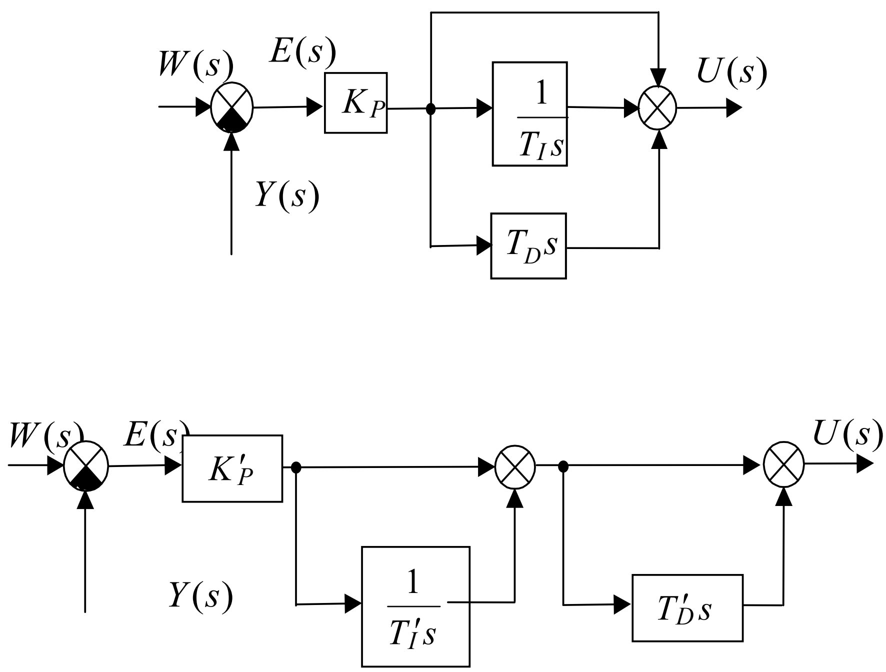


Fig. 3.4 – Block diagram of a PID controller with a structure: a) parallel (without interaction), b) serial (with interaction)

Sometimes the PID controller form with weights (3.23) is only considered as a controller with a parallel structure and the PID controller form with the time constants is considered as a standard form according to ISA (The International Society of Automation formerly Instrument Society of America).

The PID controller can be implemented by the serial (cascade) structure (Fig. 3.4b), which is described by relation

$$G_C(s) = \underbrace{K'_P\left(1+\frac{1}{T'_I s}\right)}_{\text{PI}}\underbrace{\left(1+T'_D s\right)}_{\text{PD}} = K'_P\,\frac{(T'_I s+1)(T'_D s+1)}{T'_I s} \tag{3.24}$$

This relation may be transformed into a parallel structure (3.23)

$$G_C(s) = \underbrace{K'_P \frac{T'_I + T'_D}{T'_I}}_{K_P}(1 + \underbrace{\frac{1}{T'_I + T'_D}}_{\frac{1}{T_I}} \frac{1}{s} + \underbrace{\frac{T'_I T'_D}{T'_I + T'_D}}_{T_D} s) \tag{3.25}$$

From (3.25) it follows that the change of the integral time T'_I or derivative T'_D time comes to change all values of the adjustable controller parameters K_P, T_I and T_D, which corresponds to the parallel structure, i.e. the interaction among the adjustable controller parameters happens. Therefore the PID controller with the serial structure is called the PID controller **with interaction (interacting)** and it is marked like the PID_i controller (Tab. 3.1, row 6). Among the adjustable controller parameters for the parallel and serial structure the following relations hold

$$K_P = K'_P i, \quad T_I = T'_I i, \qquad T_D = \frac{T'_D}{i}, \qquad i = 1 + \frac{T'_D}{T'_I} \tag{3.26}$$

$$K'_P = K_P \beta, \quad T'_I = T_I \beta, \qquad T'_D = \frac{T_D}{\beta}, \qquad \beta = \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{T_D}{T_I}} \tag{3.27}$$

The coefficient i is called the **interaction factor**. Most of controller tuning methods suppose the PID controller (without interaction) and therefore the adjustable controller parameters K'_P, T'_I and T'_D of the PID_i controller (with interaction) must be recounted for parameters K_P, T_I and T_D on the basis of (3.26).

For the PID_i controller in accordance with (3.27) the restriction

$$\frac{T_D}{T_I} \le \frac{1}{4} \tag{3.28}$$

there arises.

The approximate Bode plots of the PID_i controller [with interaction (3.24)] are shown in Fig. 3.5.

If the condition (3.28) holds then the approximate Bode plots of the PID controller [without interaction (3.23)] have the same courses as in Fig. 3.5, but the relations (3.27) must be considered.

From Fig. 3.5 it follows again that the integral component ensures the high value of the controller modulus for low angular frequencies firstly for steady states, the derivative component for high angular frequencies and the proportional component for all angular frequencies in the operating band. The serial structure of the PID_i controller has some advantages. It can be simply implemented by the serial interconnection of the PI and PD controllers [Fig. 3.4b and (3.24)] and therefore it is cheaply manufactured. For $T'_D = T_D = 0$ both structures are equivalent to the PI controller.

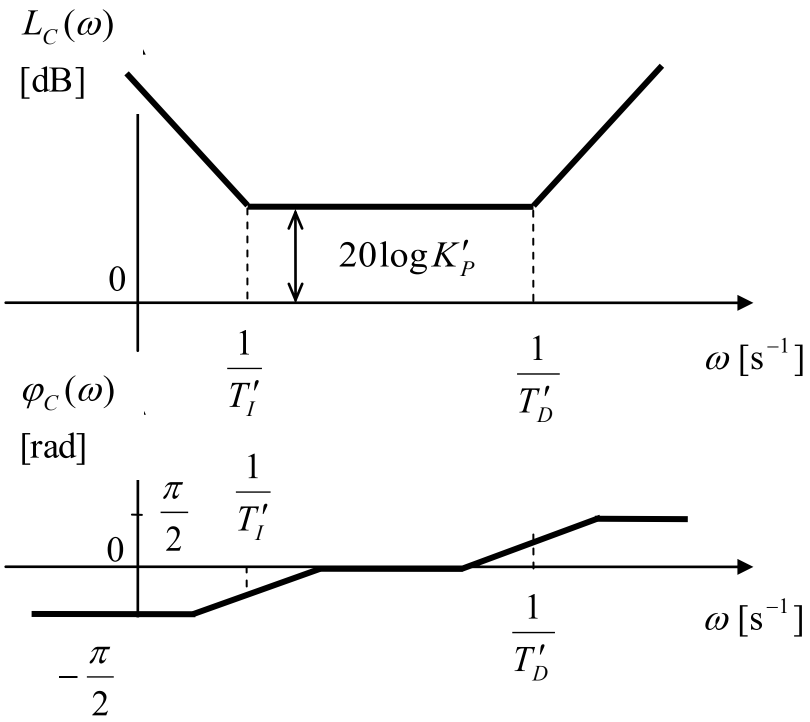


Fig. 3.5 – Bode plots of PID_i controller

From a theoretical point of view the derivative component has a positive stabilizing effect on the control process, but from a practical point of view it has very unpleasant behavior, which consists in the amplification of high frequency noise and fast changes (Fig. 3.3 and 3.5). E.g. if the derivative component of the PD or PID controllers

$$K_D \frac{\mathrm{d}\,e(t)}{\mathrm{d}t} = K_P T_D \frac{\mathrm{d}\,e(t)}{\mathrm{d}t} \tag{3.29}$$

processes the control error $e(t)$, which contains harmonic noise with the amplitude a_n and the angular frequency ω_n

$$e(t) + a_n \sin \omega_n t$$

then the derivative component (3.29) output is

$$K_D[\frac{\mathrm{d}\,e(t)}{\mathrm{d}t} + a_n \omega_n \cos \omega_n t] \tag{3.30}$$

where $K_D \frac{\mathrm{d}\,e(t)}{\mathrm{d}t}$ is the useful part of the derivative component output and $K_D a_n \omega_n \cos \omega_n t$ is the parasitic part of the derivative component output.

It is obvious that for high angular frequencies ω_n the parasitic part will dominate over the useful part and then the output of the derivative component can cause an incorrect controller function, thereby even over all the control system. Hence the ideal derivative operation is practically unusable. For attenuation of the parasitic part a **filter** of the derivative component is used. Its transfer function is given

$$\frac{1}{\frac{T_D}{N}s+1}=\frac{1}{\alpha T_D s+1}, \qquad \alpha=\frac{1}{N} \tag{3.31}$$

where $N = 5 \div 20$ or $\alpha = 0.05 \div 0.2$.

The task of the filter is to attenuate the parasitic noise in the controlled variable $y(t)$. For $\alpha \leq 0.1$ the filter doesn't have a principle effect on the resultant controller behavior, therefore during controller tuning it isn't considered. In industrial controllers the filter (3.31) is often preset at a value $\alpha = 0.1$ ($N = 10$).

The transfer function of the PID controller with the filter has the form

$$G_C(s)=K_P\left(1+\frac{1}{T_I s}+\frac{T_D s}{\alpha T_D s+1}\right) \tag{3.32}$$

A very unpleasant effect, which appears in controllers with the integral component, is the **windup**. The windup is caused by limiting the control variable, when the integration goes on and big overshoots arise. For windup removal a special mechanism must be used – the **antiwindup**.

3.2 Plants

The mathematical models of the plants may have different forms. For the linear plants the transfer functions with time constants are frequently used. The time constants are marked so that inequalities

$$T_i \geq T_{i+1}, \qquad i = 0,1,2,\ldots \tag{3.33}$$

hold, i.e. the time constant with a lower subscript has a higher or the same value than the time constant with a higher subscript.

The obtaining of the mathematical model of the real plant (object) is called the **identification**. The identification can be **analytical** or **experimental**. The practical identification methods lie between these two marginal cases. It is always useful to find the approximate relations describing given plant in the theoretical way and then experimentally to determine model parameters more precisely. For better prepared analytical relations experimental measurements are shorter and cheaper.

Every concrete plant demands a different identification method. Finding the most suitable identification approach supposes some intuition and experience.

Furthermore, some simpler experimental identification methods will be shown, which use step responses. It is supposed that the courses of the step responses are suitably prepared (filtered, smoothed etc.) and that all variables are in incremental forms, i.e. the courses begin in the origin of coordinates.

Proportional non-oscillating plants

If the plant is non-oscillating and has the step response $h_P(t)$ as in Fig 3.6a then the simplest identification method consists in the determination of the time delay $T_u = T_d = T_{d1}$ and the time constant $T_n = T_1$. The first order plus time delay (FOPTD) plant transfer function has the form

$$G_P(s) = \frac{k_1}{T_1 s + 1} \mathrm{e}^{-T_{d1} s} \tag{3.34}$$

a) b)

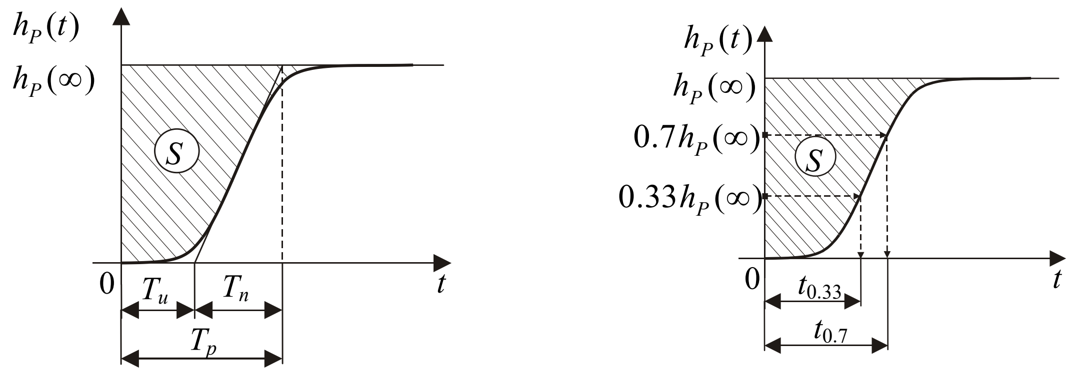


Fig. 3.6 – FOPTD plant identification on the basis of:
a) time delay $T_u = T_{d1}$ and time constant $T_n = T_1$, b) times $t_{0.33}$ and $t_{0.7}$

The plant gain k_1 for proportional plants for the unit step of the input variable, i.e. $\Delta u(t) = \eta(t)$ is given by the steady state in the step response

$$k_1 = h_P(\infty) \tag{3.35}$$

because $h_P(0) = 0$.

For general value of the step $\Delta u(t) = \Delta u$ the plant gain k_1 is given

$$k_1 = \frac{h_P(\infty)}{\Delta u} \tag{3.36}$$

The dimension of the plant gain k_1 is given by the ratio of the dimension of the output variable $y_P(t) = h_P(t)$ to the dimension of the input variable $\Delta u(t)$.

A very good mathematical model can be obtained by the **Strejc method**. It is suitable for proportional non-oscillating plants. The approximate value of the time delay T_d' must be determined at first and then on the basis of the times T_u and T_n the ratio

$$\frac{T_u - T_d'}{T_n}$$

is computed and in Tab. 3.2 the nearest lower value of the ratio

$$\frac{T_u - T_d' - \Delta T_d'}{T_n} = \frac{T_u - T_{di}}{T_n} \tag{3.37}$$

must be found and then the plant order i is determined. The plant transfer function is given by the formula

$$G_S(s) = \frac{k_1}{(T_i s + 1)^i} \mathrm{e}^{-T_{di} s} \tag{3.38}$$

where time delay is

$$T_{di} = T_d' + \Delta T_d' \tag{3.39}$$

and T_i is determined from row 3 or 4 ($\Delta T_d'$ is the correction of the estimation T_d').

Tab. 3.2 – Strejc method of experimental identification

i	**1**	**2**	**3**	**4**	**5**	**6**
$\frac{T_u - T_{di}}{T_n}$	0	0.104	0.218	0.319	0.410	0.493
$\frac{T_u - T_{di}}{T_i}$	0	0.282	0.805	1.425	2.100	2.811
$\frac{T_n}{T_i}$	1	2.718	3.695	4.463	5.119	5.699

If the times $t_{0.33}$ and $t_{0.7}$ (Fig. 3.6b) are used for the experimental identification, then for the FOPTD plant (3.34) the formulas can be used

$$\begin{aligned} T_1 &= 1.245(t_{0.7} - t_{0.33}) \\ T_{d1} &= 1.498 t_{0.33} - 0.498 t_{0.7} \end{aligned} \tag{3.40}$$

For the second order plus time delay (SOPTD) plant with the transfer function

$$G_S(s)=\frac{k_1}{(T_2 s+1)^2}\mathrm{e}^{-T_{d2}s} \tag{3.41}$$

the formulas

$$\begin{aligned} T_2 &= 0.794(t_{0.7}-t_{0.33}) \\ T_{d2} &= 1.937t_{0.33}-0.937t_{0.7} \end{aligned} \tag{3.42}$$

can be used.

The relation

$$iT_i + T_{di} \approx \frac{S}{h_P(\infty)} \tag{3.43}$$

can be used for the approximate verification of the (3.34), (3.38) and (3.41), where S is the complementary area over the step response $h_P(t)$, see Fig. 3.6.

The relations (3.40) were obtained analytically and the relations (3.42) numerically from the correspondences of the original step response and the approximate step response in the values $h_P(0) = 0$, $h_P(t_{0.33}) = 0.33h_P(\infty)$, $h_P(t_{0.7}) = 0.7h_P(\infty)$ and $h_P(\infty)$.

A very good approximation of the SOPTD plant with different time constants T_1 and T_2 is given by the following formulas

$$G_P(s)=\frac{k_1}{(T_1 s+1)(T_2 s+1)}\mathrm{e}^{-T_{d2}s} \tag{3.44}$$

where

$$\begin{aligned} &T_1=\frac{1}{2}\left(D_2+\sqrt{D_2^2-4D_1^2}\right),\ T_2=\frac{1}{2}\left(D_2-\sqrt{D_2^2-4D_1^2}\right) \\ &T_{d2}=1.937t_{0.33}-0.937t_{0.7} \\ &D_1=0.794(t_{0.7}-t_{0.33}), \qquad D_2=\frac{S}{h_S(\infty)}-T_{d2} \end{aligned} \tag{3.45}$$

In order to use the transfer function in the form (3.44), the inequality $D_2 > 2D_1$, must hold otherwise the transfer function (3.41) must be used.

For fast conversion of the transfer function (3.38) on the simpler transfer functions (3.34) and (3.41) in accordance with the scheme

$$\frac{1}{(T_i s+1)^i}\mathrm{e}^{-T_{di}s}$$

$$\swarrow \qquad \searrow \tag{3.46}$$

$$\frac{1}{T_1 s+1}\mathrm{e}^{-T_{d1}s} \longleftrightarrow \frac{1}{(T_2 s+1)^2}\mathrm{e}^{-T_{d2}s}$$

on the basis of Tab. 3.3 can be used.

Tab. 3.3 – Table for fast transfer function conversion in accordance with scheme (3.46)

$\frac{1}{(T_i s+1)^i}\mathrm{e}^{-T_{di}s}$	i	1	2	3	4	5	6
$\frac{1}{T_1 s+1}\mathrm{e}^{-T_{d1}s}$	$\frac{T_1}{T_i}$	1	1.568	1.980	2.320	2.615	2.881
	$\frac{T_{d1}-T_{di}}{T_i}$	0	0.552	1.232	1.969	2.741	3.537
$\frac{1}{(T_2 s+1)^2}\mathrm{e}^{-T_{d2}s}$	$\frac{T_2}{T_i}$	0.638	1	1.263	1.480	1.668	1.838
	$\frac{T_{d2}-T_{di}}{T_i}$	* –0.352	0	0.535	1.153	1.821	2.523

* Applicable for $T_{d1} > 0.352T_1$.

Tab. 3.3 was obtained numerically on condition that the values $h_P(0)$, $h_P(t_{0.33})$, $h_P(t_{0.7})$ a $h_P(\infty)$ of the original and the conversed step responses are the same.

Non-oscillating integrating plants

The identification of the integral plus first order plus time delay (IFOPTD) plants with the transfer function

$$G_P(s)=\frac{k_1}{s(T_1 s+1)}\mathrm{e}^{-T_{d1}s} \tag{3.47}$$

can be made on the basis of their step responses $h_P(t)$ in accordance with Fig. 3.7a. The dimension of the plant gain k_1 is given by the ratio of the dimension of the output variable $y_P(t) = h_P(t)$ and the dimensions of the input variable $\Delta u(t)$

and time.

All previous methods for identification of the proportional plants can be used for identification of the simple integrating plants if we use the impulse response (the derivative of the step response)

$$\frac{\mathrm{d}h_P(t)}{\mathrm{d}t} = g_P(t)$$

in lieu of the step response $h_P(t)$.

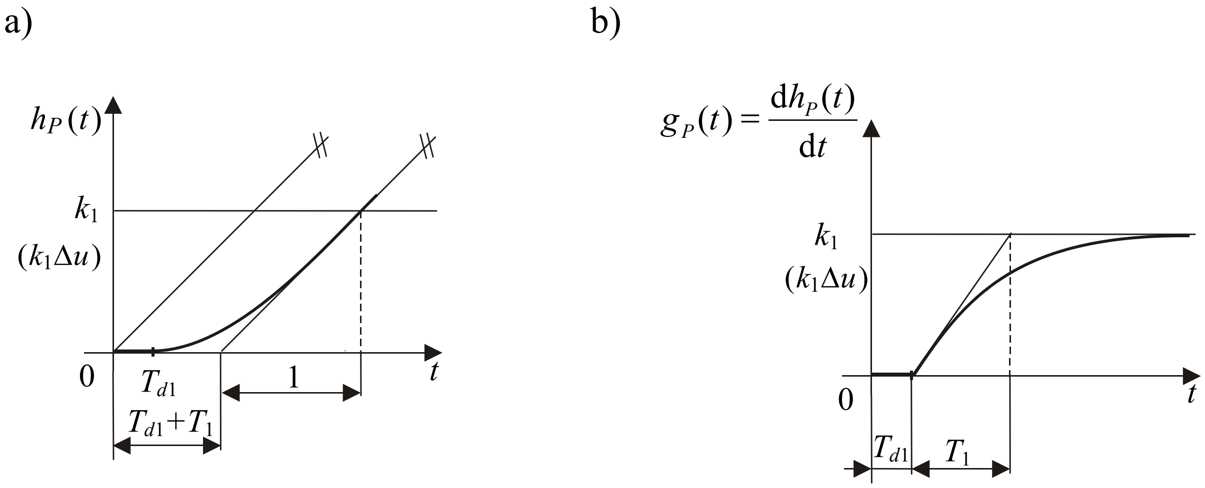


Fig. 3.7 – Identification of integrating plants on the basis of:
a) step response $h_P(t)$, b) impulse response $g_P(t)$

It is shown in Fig. 3.7b for the IFOPTD plant with the transfer function (3.47).

If the step of the input variable isn't a unit, i.e. $\Delta u(t) \neq \eta(t)$ but it is $\Delta u(t) = \Delta u$, then it is necessary to consider the values, which are in parentheses in Fig. 3.7.

Conversion of plant transfer functions

Some of the methods for the analysis and synthesis of control systems demand that the plant transfer functions have specific forms. These forms can be obtained by the simple transfer function conversion.

The conversion of the transfer function in the form (3.38) on the 1st or 2nd order form can be made on the basis of scheme (3.46) and Tab. 3.3.

The simple conversions of the transfer functions without derivations are given below. These conversions come from the equality of supplementary areas over the step responses.

Proportional plants

a)

$$\frac{k_1}{(T_1 s+1)\prod_{i=2}^{n}(T_i s+1)} \approx \frac{k_1}{(T_1 s+1)(T_\Sigma s+1)}$$

$$T_\Sigma = \sum_{i=2}^{n} T_i,\ T_1 >> T_i,\ i = 2,3\ldots,n \tag{3.48}$$

b)

$$\frac{k_1}{(T_1 s+1)\prod_{i=2}^{n}(T_i s+1)} \approx \frac{k_1}{(T_1 s+1)} \mathrm{e}^{-T_d s}$$

$$T_d = \sum_{i=2}^{n} T_i,\ T_1 >> T_i,\ i = 2,3\ldots,n \tag{3.49}$$

c)

$$\frac{k_1}{(T_1 s+1)(T_2 s+1)\prod_{i=3}^{n}(T_i s+1)} \approx \frac{k_1}{(T_1 s+1)(T_2 s+1)} \mathrm{e}^{-T_d s}$$

$$T_d = \sum_{i=3}^{n} T_i,\ T_1 \ge T_2 >> T_i,\ i = 3,4,\ldots,n \tag{3.50}$$

d)

$$\frac{k_1}{(T_0^2 s^2 + 2\xi_0 T_0 s+1)\prod_{i=1}^{n}(T_i s+1)} \approx \frac{k_1}{T_0^2 s^2 + 2\xi_0 T_0 s+1} \mathrm{e}^{-T_d s}$$

$$T_d = \sum_{i=1}^{n} T_i,\ T_0 >> T_i,\ i = 1,2,\ldots,n \tag{3.51}$$

Integrating plants

e)

$$\frac{k_1}{s\prod_{i=1}^{n}(T_i s+1)} \approx \frac{k_1}{s(T_\Sigma s+1)},\quad T_\Sigma = \sum_{i=1}^{n} T_i \tag{3.52}$$

f)

$$\frac{k_1}{s\prod_{i=1}^{n}(T_i s+1)} \approx \frac{k_1}{s}\mathrm{e}^{-T_d s}, \quad T_d=\sum_{i=1}^{n} T_i \tag{3.53}$$

g)

$$\frac{k_1}{s(T_1 s+1)\prod_{i=2}^{n}(T_i s+1)} \approx \frac{k_1}{s(T_1 s+1)}\mathrm{e}^{-T_d s}$$

$$T_d=\sum_{i=2}^{n} T_i, \ T_1 >> T_i, \ i=2,3,\ldots,n \tag{3.54}$$

The use of a combination of the summary time constant T_Σ and the substitute time delay T_d is advantageous.

If in the numerator of the plant transfer function stands up the binomials

$$1 \pm \tau_i s \tag{3.55}$$

then each binomial can be substituted by the term

$$\mathrm{e}^{\pm \tau_i s} \tag{3.56}$$

on condition that the resultant time delay will be non-negative.

The "**half rule**" is very simple and simultaneously effective.

On the assumption that the plant transfer function has a form with unstable zeros

$$G_P(s)=\frac{\prod_j (1-\tau_{j0}s)}{\prod_i (T_{i0}s+1)}\mathrm{e}^{-T_{d0}s} \tag{3.57}$$

$$T_{i0} \ge T_{i+1,0}, \ \tau_{j0} \ge 0, \ T_{d0} \ge 0$$

then on the basis of the "half rule" we can obtain

$$T_1=T_{10}+\frac{T_{20}}{2}, \quad T_{d1}=T_{d0}+\frac{T_{20}}{2}+\sum_{i\ge 3} T_{i0}+\sum_j \tau_{j0} \tag{3.58}$$

for the transfer function (3.34) or

$$T_1=T_{10}, \quad T_2=T_{20}+\frac{T_{30}}{2}, \quad T_{d2}=T_{d0}+\frac{T_{30}}{2}+\sum_{i\ge 4} T_{i0}+\sum_j \tau_{j0} \tag{3.59}$$

for the transfer function (3.44).

The resultant time delay T_{d1} or T_{d2} must be always non-negative.

3.3 Control System Stability

Stability of the linear control system is defined as its ability to fix all variables on finite values if input variables are fixed. The input variables are the desired variable $w(t)$ and all disturbance variables, which are often aggregated into one disturbance variable $v(t)$.

It is obvious that the following stability definition is equivalent. **The linear control system is stable if for any bounded input the output is always bounded.** It is so-called BIBO (bounded-input bounded-output) stability.

From both definitions it follows that stability is the characteristic behavior of the given control system, which doesn't depend on the inputs and outputs (it doesn't hold for non-linear systems).

Therefore the control system is fully described by the equation (3.3)

$$Y(s) = G_{wy}(s)W(s) + G_{vy}(s)V(s)$$

or (3.9)

$$E(s) = G_{we}(s)W(s) + G_{ve}(s)V(s)$$

it is obvious that stability is given by the term, which figures in all the basic transfer functions, i.e. $G_{wy}(s)$ and $G_{vy}(s)$ or $G_{we}(s)$ and $G_{ve}(s)$. From relations (3.4) and (3.5) or (3.10) and (3.11) it follows that this term is their denominator

$$1 + G_C(s)G_P(s) = 1 + G_o(s) = 1 + \frac{M_o(s)}{N_o(s)} = \frac{N_o(s) + M_o(s)}{N_o(s)} = \frac{N(s)}{N_o(s)} \qquad (3.60)$$

where $G_o(s)$ is the **open-loop transfer function** of the control system (it is generally given by the product of all transfer functions in the loop), $N_o(s)$ – the characteristic polynomial of the open-loop of the control system (the denominator of the open-loop transfer function), $M_o(s)$ – the polynomial of the numerator of the open-loop transfer function.

The polynomial

$$N(s) = N_o(s) + M_o(s) \qquad (3.61)$$

is the **characteristic polynomial** of the control system and after its equating to zero the **characteristic equation** of the control system

$$N(s) = 0$$

is obtained.

The characteristic polynomial (3.61) rises after its arrangement in the denominators of all basic transfer functions of the control system, i.e. (3.4), (3.5), (3.10) and (3.11) and therefore it is simultaneously the characteristic polynomial of the relevant linear differential equation, which describes the given control system.

A necessary and sufficient condition for (asymptotic) stability of the linear differential equation and the corresponding linear dynamic system is that the roots $s_1, s_2,..., s_n$ of the characteristic polynomial (or the characteristic equation)

$$N(s) = a_n s^n + \ldots + a_1 s + a_0 = a_n (s - s_1)(s - s_2)\ldots(s - s_n) \tag{3.62}$$

have negative real parts, i.e. (see Fig. 3.8)

$$\operatorname{Re} s_i < 0, \qquad \text{for} \qquad i = 1,2,\ldots,n \tag{3.63}$$

It is obvious that the conditions of the negativeness of the real parts of the roots (i.e. poles) (3.63) of the characteristic polynomial of the control system **(3.61) [(3.62)]** are the necessary and sufficient conditions for (asymptotic) stability of the given linear control system**.**

Because the concept of the stability of the non-linear systems has a rather different meaning, it is necessary in some cases when the necessary and sufficient conditions hold to use a more precise concept of "asymptotic" stability.

The complex roots, i.e. poles of the control system rise always in the conjugate couple (i.e. in the symmetry of the real axis in the s-complex plane). It is very important that the poles $s_1, s_2,..., s_n$ of the control system are at the same time the poles of all of its basic transfer functions. It doesn't hold for the zeros of the basic transfer functions. **The poles of the control system determine its dynamic behavior.**

The necessary and sufficient condition for stability (3.63) of the control system can be obtained in another way.

Consider any basic transfer function of the control system, e.g.

$$G_{wy}(s) = \frac{M(s)}{N(s)} \tag{3.64}$$

and the desired variable transform

$$W(s) = \frac{M_w(s)}{N_w(s)} \tag{3.65}$$

where $M(s)$, $M_w(s)$ and $N_w(s)$ are the polynomials and $N(s)$ is the characteristic polynomial of the control system.

On condition that the characteristic polynomial of the control system $N(s)$ has the simple roots $s_1, s_2,..., s_n$ and the polynomial $N_w(s)$ has the simple roots $s_1^w, s_2^w, \ldots, s_p^w$ [p is the degree of the polynomial $N_w(s)$], the transform of the controlled variable (response)

$$Y(s) = G_{wy}(s)W(s) = \frac{M(s)}{N(s)} \frac{M_w(s)}{N_w(s)} \tag{3.66}$$

can be written in the form of the sum of the partial fractions

$$Y(s) = \underbrace{\sum_{i=1}^{n} \frac{A_i}{s - s_i}}_{Y_T(s)} + \underbrace{\sum_{j=1}^{p} \frac{B_j}{s - s_j^w}}_{Y_S(s)} = Y_T(s) + Y_S(s) \tag{3.67}$$

where $Y_T(s)$ is the transform of the transient response part, $Y_S(s)$ – the transform of the steady response part.

The original of the controlled variable $y(t)$ can be obtained from (3.67) on the basis of the Laplace transform

$$y(t) = y_T(t) + y_S(t) = \sum_{i=1}^{n} A_i \,\mathrm{e}^{s_i t} + \sum_{j=1}^{p} B_j \,\mathrm{e}^{s_j^w t} \tag{3.68}$$

The constants A_i and B_j in the relations (3.67) and (3.68) generally depend on the forms of the transfer function $G_{wy}(s)$ and the desired variable transform $W(s)$, see (3.64) and (3.65).

The course of the transient part of the controlled variable $y_T(t)$ depends on the roots of the characteristic polynomial of the control system, i.e. on its poles and it is given as

$$y_T(t) = \sum_{i=1}^{n} A_i \,\mathrm{e}^{s_i t}$$

The course of the steady part of the controlled variable

$$y_S(t) = \sum_{j=1}^{p} B_j \,\mathrm{e}^{s_j^w t}$$

is given by the course of the desired variable $w(t)$.

Here by its steady course it is necessary to understand the given time function, e.g. $y_S(t) = Bt$, $y_S(t) = B\sin\omega t$ etc. in contrast to the steady (static) state, e.g. $y_S(t) = y_S = \text{const.}$

From (3.68) it follows that for the bounded input variable – the desired variable $w(t)$ ($\mathrm{Re}\, s_j^w < 0$ for $j = 1, 2,..., p$) the output variable – the controlled variable $y(t)$ will be bounded if and only if its the transient part $y_T(t)$ will be bounded, i.e. the condition (3.63) will hold. Therefore for the stable control system the transient part $y_T(t)$ must vanish for $t \to \infty$, i.e.

$$\lim_{t\to\infty} y_T(t) = 0 \tag{3.69}$$

therefore for $t \to \infty$

$$y(t) \to y_S(t) \tag{3.70}$$

holds.

From the last relation it follows that control system stability is its ability to steady the output controlled variable $y(t) \to y_S(t)$ for the steady input desired variable $w(t) \to w_S(t)$.

For the control system from the control objective $y(t) \to w(t)$ it follows the obvious demand $y_S(t) \to w_S(t)$.

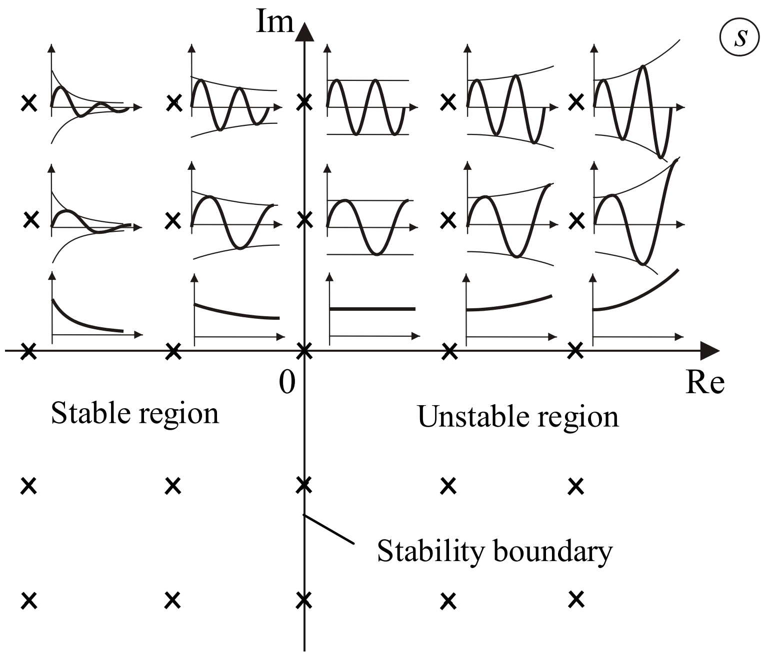


Fig. 3.8 – The influence of the position of control system poles on the transient part of the response

It is obvious that similar conclusions will hold for multiple poles of the polynomial $N(s)$ and $N_w(s)$ in the relation (3.66), because adding negligible small numbers to the multiple poles changes their simple poles and this small change can't have a substantial effect on the behavior of the given control system.

The influence of the position of the control system poles on the transient part of the response is shown in Fig. 3.8. It is necessary to reason out that the oscillating responses are evoked by the conjugate complex couple of the poles.

The transfer function of the open-loop control system with a time delay has the form [compare with (3.60)]

$$G_o(s) = \frac{M_o(s)}{N_o(s)} e^{-T_d s} \tag{3.71}$$

On the basis of the (3.71) we can easily obtain the **characteristic quasipolynomial** of the control system [compare with (3.61)]

$$N(s) = N_o(s) + M_o(s) e^{-T_d s} \tag{3.72}$$

The characteristic quasipolynomial (3.72) has an infinite number of roots, i.e. the control system with the time delay has an infinite number of poles. That is why the stability verification by the necessary and sufficient conditions (3.63) by the direct computation isn't real.

Control system stability is the only necessary condition for its proper operation. For verification of control system stability different stability criteria are used, which enable checking the fulfillment of inequalities (3.63) without labored computation of all roots of the control system characteristic polynomial or quasipolynomial $N(s)$.

Furthermore, the three stability criteria are given without derivation: Hurwitz, Mikhailov and Nyquist criteria.

Hurwitz stability criterion

The Hurwitz stability criterion is an algebraic criterion and therefore it isn't suitable for the control systems with a time delay (the exponential function isn't an algebraic function). It can be used only for approximate stability verification in the case that the time delay will be substituted by its algebraic approximation.

The Hurwitz stability criterion can be formulated in the form:

„The linear control system with the characteristic polynomial

$$N(s) = a_n s^n + \ldots + a_1 s + a_0$$

is (asymptotic) stable [i.e. the conditions (3.63) hold] if and only if, when:

− all coefficients $a_0, a_1, ..., a_n$ exist and are positive (**it is a necessary stability condition** formulated by Slovak technician A. Stodola)

− the main corner minors (subdeterminants) of the Hurwitz matrix

$$\boldsymbol{H} = \begin{bmatrix} a_{n-1} & a_{n-3} & a_{n-5} & \cdots & 0 \\ a_n & a_{n-2} & a_{n-4} & \cdots & 0 \\ 0 & a_{n-1} & a_{n-3} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a_0 \end{bmatrix} \tag{3.73}$$

$$H_1 = a_{n-1},\ H_2 = \begin{vmatrix} a_{n-1} & a_{n-3} \\ a_n & a_{n-2} \end{vmatrix},\ \ldots,\ H_n = |\boldsymbol{H}|$$

are positive."

Because the equality $H_1 = a_{n-1}$, $H_n = a_0 H_{n-1}$ hold, it is enough to check only the positiveness of $H_2, H_3, ..., H_{n-1}$. If some of the Hurwitz minors are zero, then it determines the **stability boundary**. E.g. for $a_0 = 0 \Rightarrow H_n = 0$ one pole is zero (it is the origin of the coordinates in the complex plane s). This case characterizes the **non-oscillating** stability boundary. For $H_{n-1} = 0$ two poles are imaginary and conjugate (they are on the imaginary axes in symmetry by the origin of the coordinates in the complex plane s). This case characterizes the **oscillating** stability boundary, see Fig. 3.8.

If the Stodola necessary condition of the stability holds, then the simplified **Lineard – Chipart stability criterion** can be used, which consists only in checking of the positiveness of all odd or all even Hurwitz minors.

The disadvantage of the Hurwitz criterion is its high demandingness of computation for $n \geq 5$.

Mikhailov stability criterion

The Mikhailov stability criterion is a frequency criterion with a wide range of use. Here only a simple formulation will be given, which is suitable for control systems without a time delay.

The Mikhailov stability criterion uses the control system characteristic polynomial $N(s)$, from it after substituting $s = \mathrm{j}\omega$ the Mikhailov function

$$N(\mathrm{j}\omega) = N(s)\big|_{s=\mathrm{j}\omega} = N_P(\omega) + \mathrm{j}N_Q(\omega) \tag{3.74}$$

is obtained, where

$$N_P(\omega) = \operatorname{Re} N(\mathrm{j}\omega) = a_0 - a_2\omega^2 + a_4\omega^4 - \ldots \tag{3.75a}$$

is the real part and

$$N_Q(\omega) = \operatorname{Im} N(\mathrm{j}\omega) = a_1\omega - a_3\omega^3 + a_5\omega^5 - \ldots \tag{3.75b}$$

is the imaginary part of the Mikhailov function.

The Mikhailov stability criterion can be formulated in the form:

„The linear control system is (asymptotic) stable if and only if its Mikhailov function (plot) $N(\mathrm{j}\omega)$ for $0 \leq \omega \leq \infty$ begins on the positive real axis and successively passes through n quadrants in a positive direction (anticlockwise).“

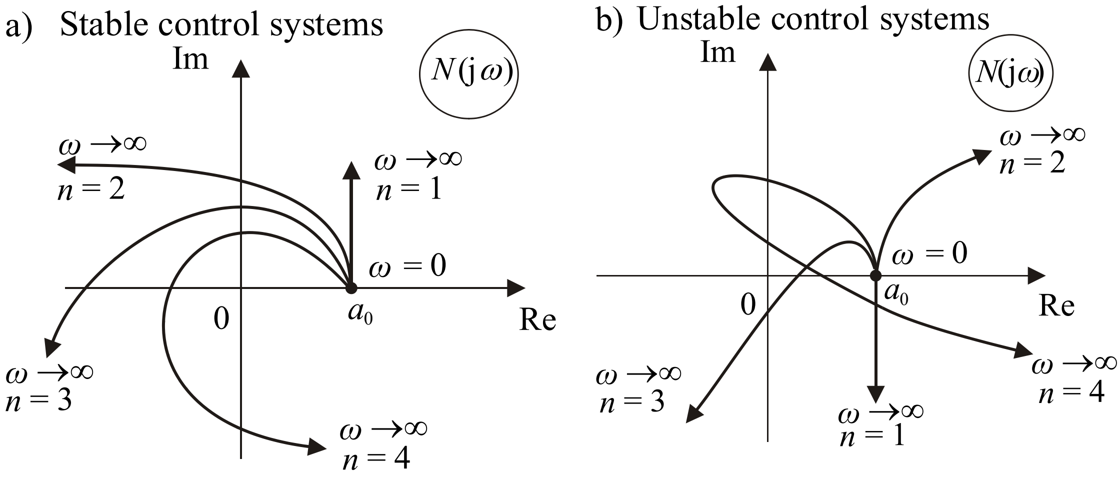


Fig. 3.9 – Mikhailov plots for control systems:
a) stable, b) unstable

This formulation can be written in a form for changing the argument (angle) of the Mikhailov function

$$\Delta \arg_{0 \le \omega \le \infty} N(\mathrm{j}\omega) = n\frac{\pi}{2} \tag{3.76}$$

where n is the characteristic polynomial $N(s)$ degree.

The courses of the Mikhailov functions (plots) for the stable control systems are in Fig. 3.9a and for unstable control systems in Fig. 3.9b.

The Mikhalov function can be employed for an analytical determination of the ultimate (critical) angular frequency ω_c and **the** ultimate (critical) controller gain K_{Pc} or **the** ultimate (critical) controller integral time T_{Ic}.

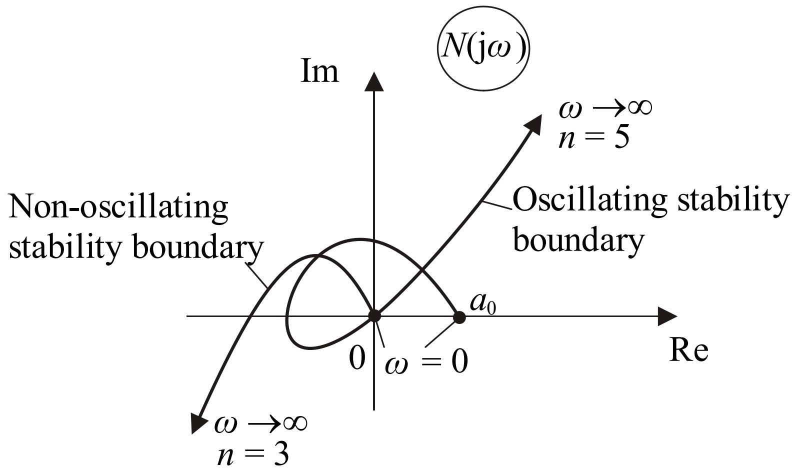


Fig. 3.10 – Mikhailov plots for control systems on the stability boundary

For this the equations

$$N_P = 0 \text{ and } N_Q = 0 \tag{3.77}$$

are used.

The ultimate parameters (ω_c, K_{Pc} or T_{Ic}), more correctly their values, cause that the control system is on a stability boundary, i.e. in the critical state between stability and instability. In this case a slight change of these values causes stability or instability of a given control system.

Nyquist stability criterion

The Nyquist stability criterion is a frequency criterion, which in contrast to the Hurwitz and Mikhailov criteria uses the open-loop frequency transfer function $G_o(j\omega)$. It is very general and it can be extended for unstable open-loop control systems and even for non-linear control systems.

The control system in Fig. 3.11 is considered. It is obvious that when oscillations arise with a constant amplitude and an angular frequency on the stability boundary [for $W(s) = V(s) = 0$] it is necessary that oscillations in the feedback path must be the same as oscillations in the forward path but with a negative sign, see Fig. 3.11. It can be written in the transforms

$$G_o(s) = -1 \Rightarrow G_o(j\omega_c) = -1 \tag{3.78}$$

where $G_o(s) = G_C(s)G_P(s)$ is the open-loop transfer function (it is generally given by the product of all transfer functions in the loop), ω_c – the ultimate angular frequency.

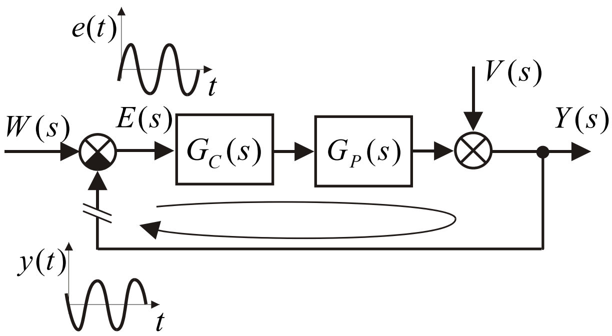


Fig. 3.11 – Control system on stability boundary

It is obvious that this conclusion can be made on condition that the open-loop control system is stable (otherwise the stable oscillations in the control loop wouldn't be possible).

The relation (3.78) expresses the given control system condition for the oscillating stability boundary. It can be obtained from the same denominators of the basis transfer functions of the control systems [see e.g. (3.4), (3.5), (3.10) and (3.11)], where the term $1 + G_o(s)$ stands out. It is obvious that the critical state arises when this term will be equal to zero, which corresponds with (3.78).

The relation (3.78) expresses the fact that if the control system is on the oscillating stability boundary, then the frequency response (polar plot) of the open-loop control system comes through the point –1+j0 on the negative real axis. This point is called the **critical point**. The frequency response of the open-loop control system is called the **Nyquist plot**.

Furthermore, from the relation (3.78) and Fig. 3.14 it follows that if the value e.g. $G_o(j\omega_p) = -0.5$ in lieu of $G_o(j\omega_p) = G_o(j\omega_c) -1$ was in it, the oscillations would decrease (i.e. the control system is stable) and vice versa for value e.g. $G_o(j\omega_p) = -2$ the oscillations would increase (the control system is unstable).

The Nyquist stability criterion can be formulated in the form:

„The linear control system is (asymptotic) stable if and only if when the frequency response of the stable open-loop control system, i.e. the Nyquist plot $G_o(j\omega)$ for $0 \le \omega \le \infty$ doesn't enclose the critical point –1+j0 on the negative real axis."

The main cases of the Nyquist plots $G_o(j\omega)$ are shown in Fig. 3.12. The integrating elements in the forward path or feedback path (i.e. in the loop) from the point of view of the Nyquist stability criterion aren't considered as unstable (they are in fact neutral elements). The number of these integrating elements q is called the **control system type**.

In the case the integrating elements exist the decision about if that the Nyquist plot encloses or doesn't enclose the critical point –1+j0 must be made in accordance with the Fig. 3.13.

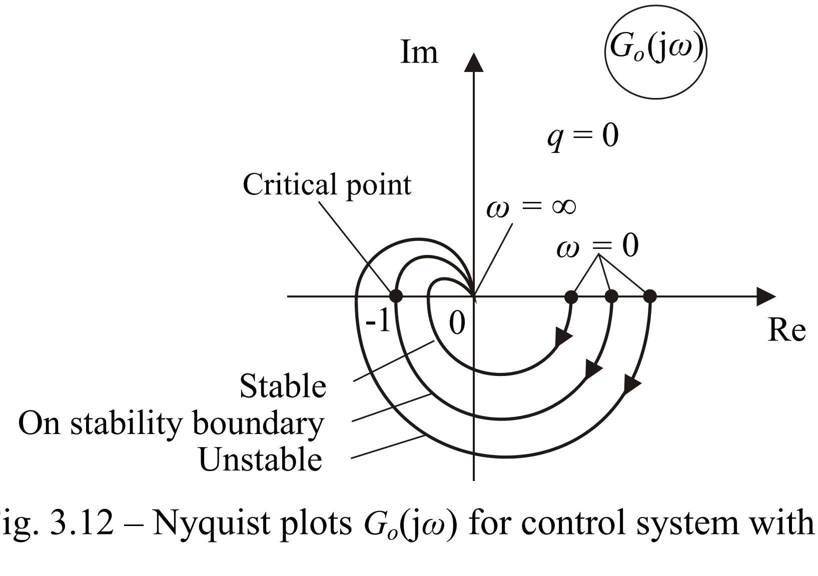


Fig. 3.12 – Nyquist plots $G_o(j\omega)$ for control system with $q = 0$

If the Nyquist plot $G_o(j\omega)$ for $q = 2$ has the course as in Fig. 3.13 then the control system is **conditionally stable**, because the increasing or decreasing of the $A_o(\omega)$ for the phase $-\pi$ can cause the instability of the control system.

Above the geometrical form of the Nyquist stability was formulated. The analytical formulation of the Nyquist stability criterion is also very useful. We can write

$$A_o(\omega_g) = 1 \tag{3.79}$$

$$\varphi_o(\omega_p) = -\pi \tag{3.80}$$

where ω_g is the gain crossover angular frequency, ω_p – the phase crossover angular frequency.

For the oscillating stability boundary holds

$$\omega_c = \omega_g = \omega_p \tag{3.81}$$

Now the Nyquist stability criterion can be written in different analytical forms:

$$G_o(\mathrm{j}\omega_p) = \mathrm{Re}G_o(\mathrm{j}\omega_p) > -1, \quad A_o(\omega_p) < 1 \tag{3.82}$$

$$\varphi_o(\omega_g) > -\pi \tag{3.83}$$

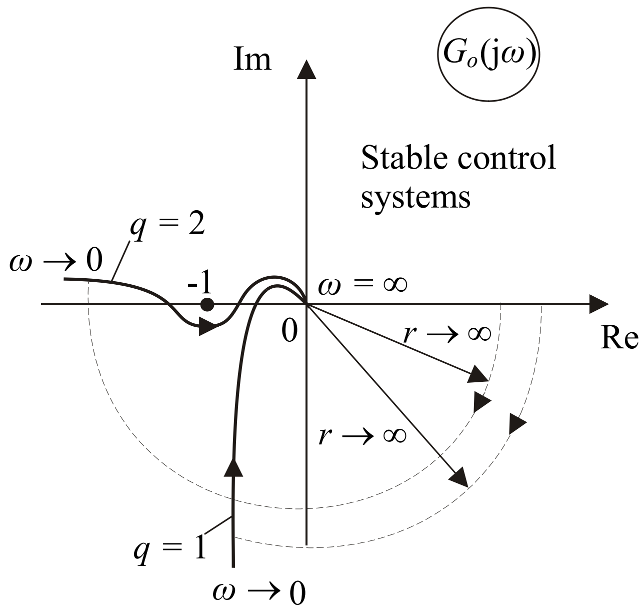


Fig. 3.13 – Nyquist plots $G_o(\mathrm{j}\omega)$ for stable control systems with q = 1 and q = 2

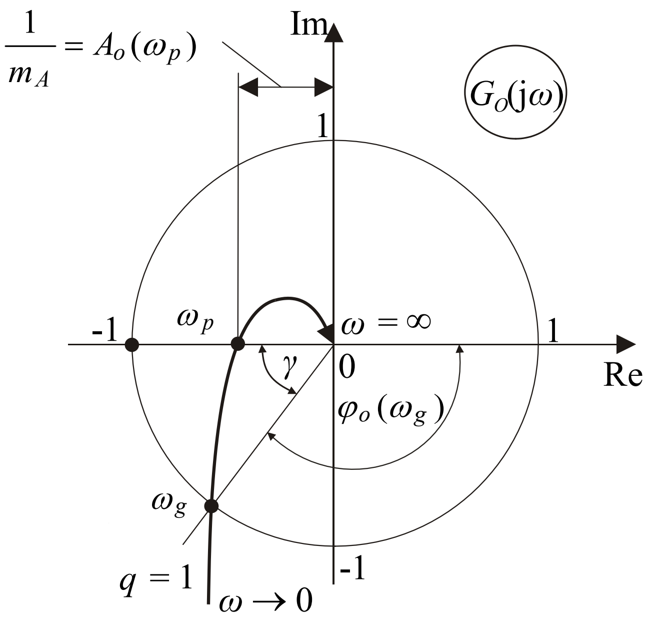


Fig. 3.14 – Gain margin m_A and phase margin γ

It is obvious that these simple analytical formulations hold for nonconditionally stable control systems. For conditionally stable systems these formulations can be easily extended.

On the basis of the angular frequencies ω_g and ω_p further important indices can be defined (Fig. 3.14):

the gain margin

$$m_A = \frac{1}{A_o(\omega_p)} \tag{3.84}$$

and the phase margin

$$\gamma = \pi + \varphi_o(\omega_g) \tag{3.85}$$

The gain margin m_A expresses how many times the magnitude $A_o(\omega_p)$ can be increased (how many times the open-loop gain k_o can be increased) in order for the control system to reach the stability boundary. Similarly the phase margin γ expresses how much the phase $\varphi_o(\omega_g)$ (in the absolute value) can be increased in order for the control system to reach the stability boundary.

Because the controller integral component brings the negative phase in the open-loop of the control system (see Fig. 3.5), i.e. it decreases the phase margin γ, therefore the **controller integral component destabilizes (i.e. it deteriorates the stability) the control system**. On the other hand the controller derivative component brings the positive phase in the open-loop of the control system (see Fig. 3.5), i.e. it increases the phase margin γ, therefore the **controller derivative component stabilizes (i.e. improves the stability) the control system** [of course for suitable filtration, see e.g. (3.31 and 3.32)].

Regarding the controller gain K_P, it is obvious that by its increasing it simultaneously increases the open-loop gain k_o and hence the gain margin is decreased, therefore the **controller proportional component destabilizes the control system** (it doesn't hold for conditionally stable control systems).

Time delay is very dangerous for the control system stability. The frequency transfer function of the time delay has the form

$$G(\mathrm{j}\omega) = \mathrm{e}^{-T_d\,\mathrm{j}\omega} = A(\omega)\,\mathrm{e}^{\mathrm{j}\varphi(\omega)} \tag{3.86}$$

$$A(\omega) = 1 \tag{3.87}$$

$$\varphi(\omega) = -T_d\omega \tag{3.88}$$

From the relations (3.86) – (3.88) it follows that the time delay doesn't change the modulus (magnitude) [see (3.87)] but linearly increases the negative phase [see (3.88)], i.e. it decreases the phase margin γ. Therefore **the time delay always essentially destabilizes the control system**.

4 CONTROL SYSTEM SYNTHESIS

The chapter is devoted to process control performance and the linear control system synthesis, i.e. to controller choices and their tuning. Basic known and new controller tuning methods are brought up. Some of them are also for the digital controller.

4.1 Process Control Performance

The control objective expressed in two equivalent forms (3.2) and (3.8) or by couple relations (3.6), (3.7) and (3.12), (3.13) [see as well (3.15), (3.16)] can be held with a different **process control performance** and only on the condition that a given control system is stable. It is obvious that process control performance can be reviewed in: the time domain, the frequency domain and the complex variable domain. Different criteria and indices can be used for it.

Time Domain

The time domain is very popular among the control system technicians and designers because it enables the fast and intuitive evaluation of process control performance on the basis of the step responses $y(t)$ caused by the step changes of the desired variable $w(t)$ or the disturbance variable $v(t)$. It is useful to inscribe the responses with subscripts in accordance with the input variables.

For simultaneous actuating the desired variable $w(t)$ and the disturbance variable $v(t)$ on the basis of the linearity principle it holds

$$Y(s) = G_{wy}(s)W(s) + G_{vy}(s)V(s) = Y_w(s) + Y_v(s) \Rightarrow$$

$$y(t) = y_w(t) + y_v(t) \tag{4.1}$$

where $y_w(t)$ is the response caused by the desired variable $w(t)$ for $v(t) = 0$, $y_v(t)$ – the response caused by the disturbance variable $v(t)$ for $w(t) = 0$.

The typical control system oscillatory and non-oscillatory responses in incremental variables (i.e. in increments from the operation point) are shown in Figs 4.1 and 4.2. A very important conclusion comes from them. If the disturbance variable $v(t)$ influences the plant output then for the same input steps the servo (setpoint) response and regulatory response are in principle the same as well (the regulatory response is turned up and moved, see Figs 4.1 and 4.2). It is given by relation $G_{vy}(s) = 1 - G_{wy}(s)$. The steady-state errors $e_v(\infty)$ for the control systems in Fig. 3.2 and the steps of the disturbance variable $v(t)$ have negative values, see Figs 4.2b and 4.4b and relation (3.11).

The servo and regulatory responses for the disturbance variable caused in the plant output with the zero steady-state errors in Fig. 4.1 correspond to a case when the open-loop contains at least one integrating element, i.e. the control system type $q \geq 1$. The integrating element (component) can be included in the controller or in the plant.

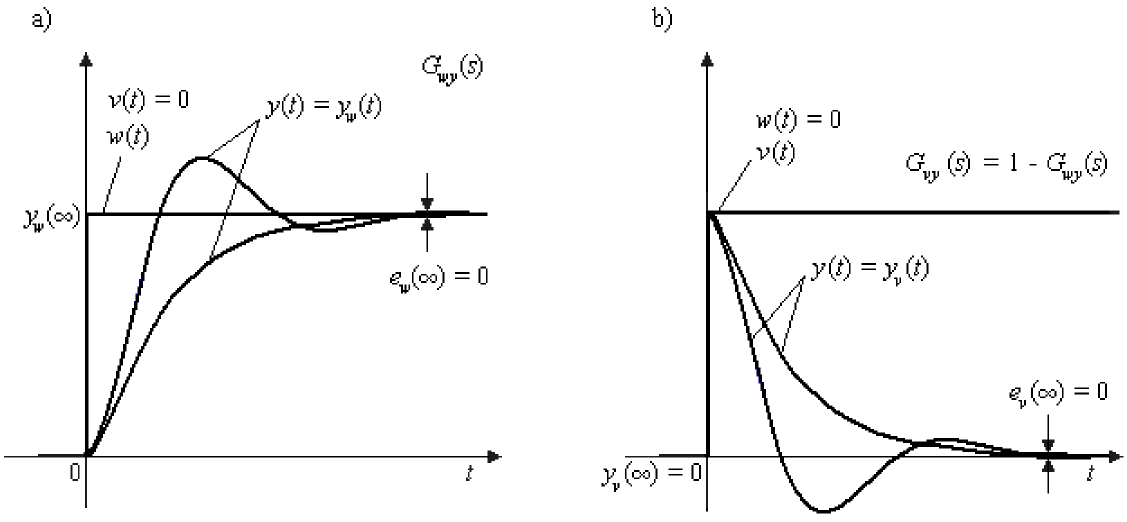


Fig. 4.1 – Control system step responses in the case of zero steady-state errors: a) servo (setpoint) responses, b) regulatory responses for disturbance variable in the plant output

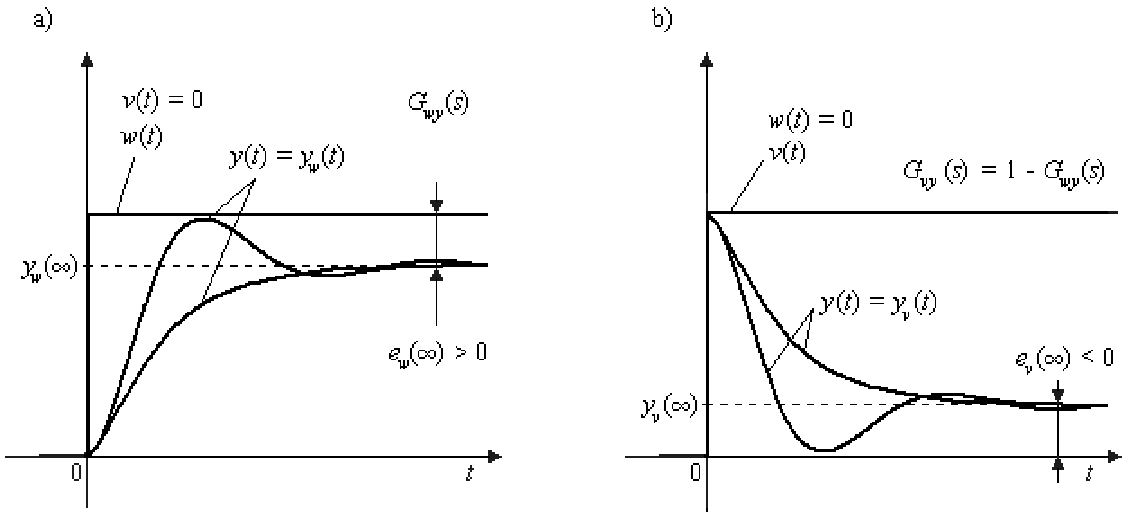


Fig. 4.2 – Control system step responses in case non-zero steady-state errors: a) servo (setpoint) responses, b) regulatory responses for disturbance variable in plant output

The servo and regulatory responses for the disturbance variable caused in the plant output with non-zero steady-state errors in Fig. 4.2 correspond to the case when the open-loop doesn’t contain any integrating element, i.e. the control system type $q = 0$

If the disturbance variable $v(t)$ influences the plant input (in Figs 4.3 and 4.4 the oscillating responses are only shown), then it is necessary to distinguish the causes if the plant contains the integrating elements (it has an integrating character) or doesn't contain the integrating elements (it has a proportional character).

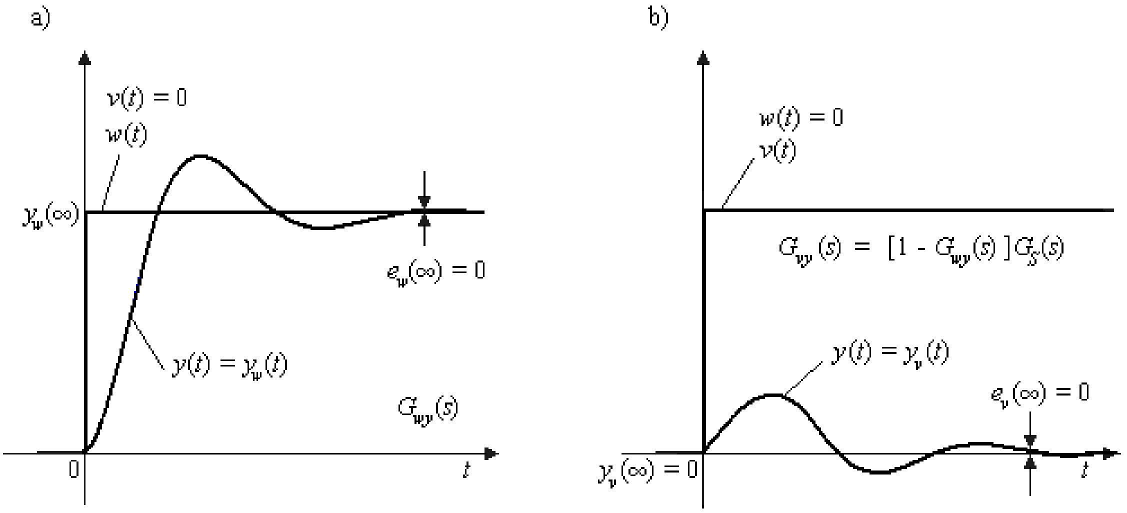


Fig. 4.3 – Control system step responses for a controller with integral component and proportional plant: a) servo (setpoint) response, b) regulatory response for a disturbance variable in plant input

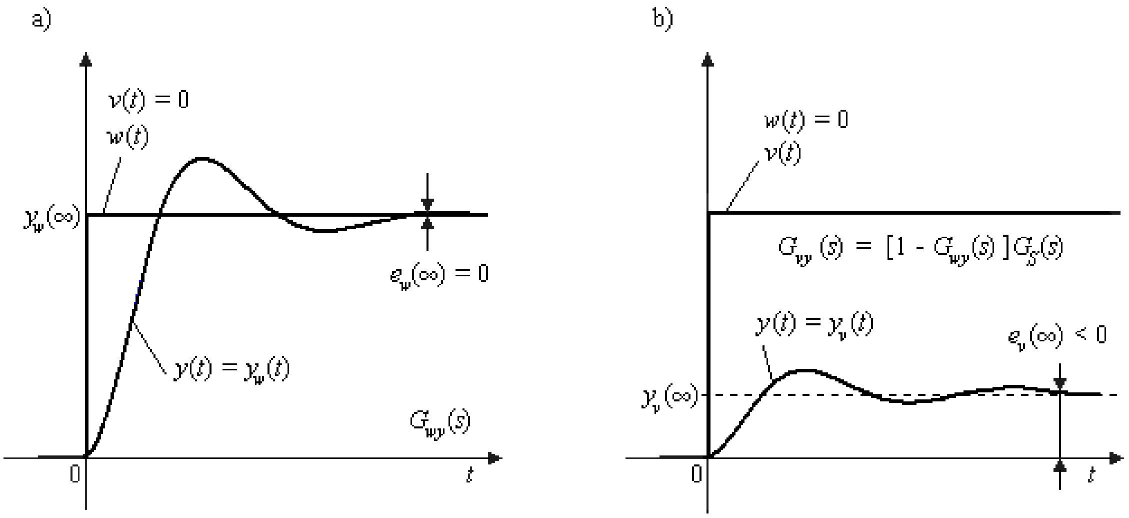


Fig. 4.4 – Control system step responses for a controller without an integral component and integrating plant: a) servo (setpoint) response, b) regulatory response for a disturbance variable in the plant input

If the plant has a proportional character and the controller contains the integral component (e.g. I, PI, PID) then $q = 1$ and the steady-state errors are zero, see Fig. 4.3. From Fig. 4.3 it follows that the regulatory response $y_v(t)$ is often very well attenuated by the plant. It is caused by the filtration (inertia) behavior of the plant. Therefore the controller can be tuned more aggressively, i.e. it is possible to increase the controller gain K_P or to decrease the integral time T_I.

If the plant has an integrating character (only one integrating element is considered) then in the case of the use of the controllers without the integral component (e.g. P, PD) the control system is type q = 1 but still for the disturbance in the plant input the regulatory response will be with a non-zero error, see Fig. 4.4b. For controllers with the integral component the steady-state errors $e_w(\infty)$ and $e_v(\infty)$ will be zero for the input steps. In this case the control system type q is 2.

The steady-state errors can be determined on the basis of the following relations

$$E(s) = G_{we}(s)W(s) + G_{ve}(s)V(s) = E_w(s) + E_v(s) \tag{4.2}$$

$$e_w(\infty) = \lim_{s \to 0} sE_w(s), \qquad e_v(\infty) = \lim_{s \to 0} sE_v(s) \tag{4.3}$$

where $e_w(\infty)$ is the steady-state error caused by the desired variable $w(t)$, $e_v(\infty)$ – the steady-state error caused by the disturbance variable $v(t)$.

The mentioned relations (4.2) and (4.3) generally hold for any changes of the input variables $w(t)$ and $v(t)$, e.g. for the velocity or acceleration steps etc.

The steady-state errors can be decreased by increasing the controller gain K_P (in the case of the I controller by decreasing the integral time T_I).

If the plant has an integrating character and the disturbance variable $v(t)$ causes in the plant input then it is necessary to reason it out in controller tuning.

By ensuring suitable behavior of the control system from the point of view of the desired variable $w(t)$, the corresponding behavior of the control system from the point of view of the disturbance variable $v(t)$ (for a disturbance caused in the plant output it always holds) will be ensured in most cases too. Therefore further the servo (tracking) problem is solved first of all and that is why the subscripts w will not be mostly used.

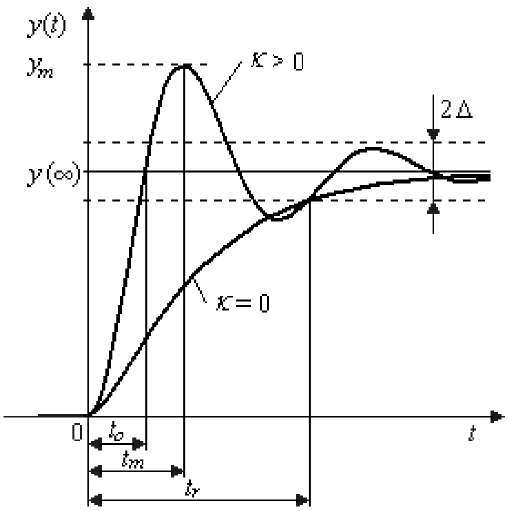


Fig. 4.5 – Servo (setpoint) responses with marked control performance indices

In Fig. 4.5 two typical courses of the servo (setpoint) response are shown. From the practical point of view the most important performance indices are: the **settling time** t_r and the **relative overshoot**

$$\kappa = \frac{y_m - y(\infty)}{y(\infty)}, \quad y_m = y(t_m) \tag{4.4}$$

where y_m is the maximum value of the controlled variable $y(t_m)$ (the first maximum or peak), t_m – the time of reaching the value y_m (the peak time), $y(\infty)$ – the steady state value of the controlled variable. The settling time is determined by the time when the controlled variable $y(t)$ gets in the band with a width 2Δ, i.e. $y(\infty) \pm \Delta$, where the **control tolerance** is given

$$\Delta = \delta\, y(\infty), \delta = 0.01 \div 0.05 \qquad (1 \div 5)\ \% \tag{4.5}$$

The **relative control tolerance** δ mostly has a value 0.05 or 0.02.

For the settling time tr the relative control tolerance δ must be mentioned otherwise it is supposed $\delta = 0.05$ (5 %).

The case $\kappa = 0$ corresponds to a non-oscillating (aperiodic) control process, which is used for processes where the overshoot can cause undesirable effects (e.g. thermal and chemical processes, assembly robots and manipulators etc.).

For the non-oscillating control process, the minimum of the settling time is demanded very often. This control process is called the **marginal** non-oscillating control process.

For $\kappa > 0$ the control process is oscillating and faster then the non-oscillating process. The time for reaching the value $y(\infty)$ is the **rise time** t_o. Very often the rise time is defined like the time required for the response to go from $0.1y(\infty)$ to $0.9y(\infty)$.

The control process with the relative overshoot κ about 0.05 (5 %) is acceptable for most plants. If the minimum of the settling time t_r is simultaneously ensured then this control process is regarded as practically "optimal". It is widely accepted everywhere that the small overshoot doesn't matter or is desirable, e.g. for the indicator measuring and recording devices (in this case the small overshoot enables a faster interpolating of the indicator position).

The integral **criteria** are very useful for the complex evaluating of the control performance. The shade area in Fig. 4.6 expresses the so-called **control area**.

It is obvious that the control area will be smaller and the control performance will be higher. It is suitable to work with the control error $e(t) = w(t) - y(t)$ (see Figs 4.6b, c, d) on condition $e(\infty) = e_w(\infty) = 0$. If $e(\infty) \neq 0$, then in all relations for the integral criteria the term $e(t) - e(\infty)$ must be substituted in lieu of $e(t)$

Integral of error (Fig. 4.6b)

$$I_{IE} = \int_0^\infty e(t)\,\mathrm{d}t \to \min \tag{4.6}$$

The integral of error I_{IE} (IE = **I**ntegral of **E**rror) is the simplest integral criterion. It isn't suitable for oscillating control processes, because $I_{IE} = 0$ for the control process on the oscillating stability boundary (the areas marked with signs + and – are mutually subtracted). Its best advantage is that it can be easily computed (see appendix)

$$I_{IE} = \lim_{s \to 0} E(s) = \lim_{s \to 0} \int_0^\infty e(t)\,\mathrm{e}^{-st}\,\mathrm{d}t = \int_0^\infty e(t)\,\mathrm{d}t \tag{4.7}$$

Integral of absolute error (Fig. 4.6c)

$$I_{IAE} = \int_0^\infty |e(t)|\,\mathrm{d}t \to \min \tag{4.8}$$

The integral of absolute error I_{IAE} (IAE = **I**ntegral of **A**bsolute **E**rror) removes the disadvantage of the previous integral criterion I_{IE} (see Fig. 4.6c), and therefore it is applicable for both non-oscillating and oscillating control processes. It has a very unpleasant behavior and generally cannot be calculated analytically but only numerically or by simulation.

It is obvious that the control area in Fig. 4.6a is (4.8) too.

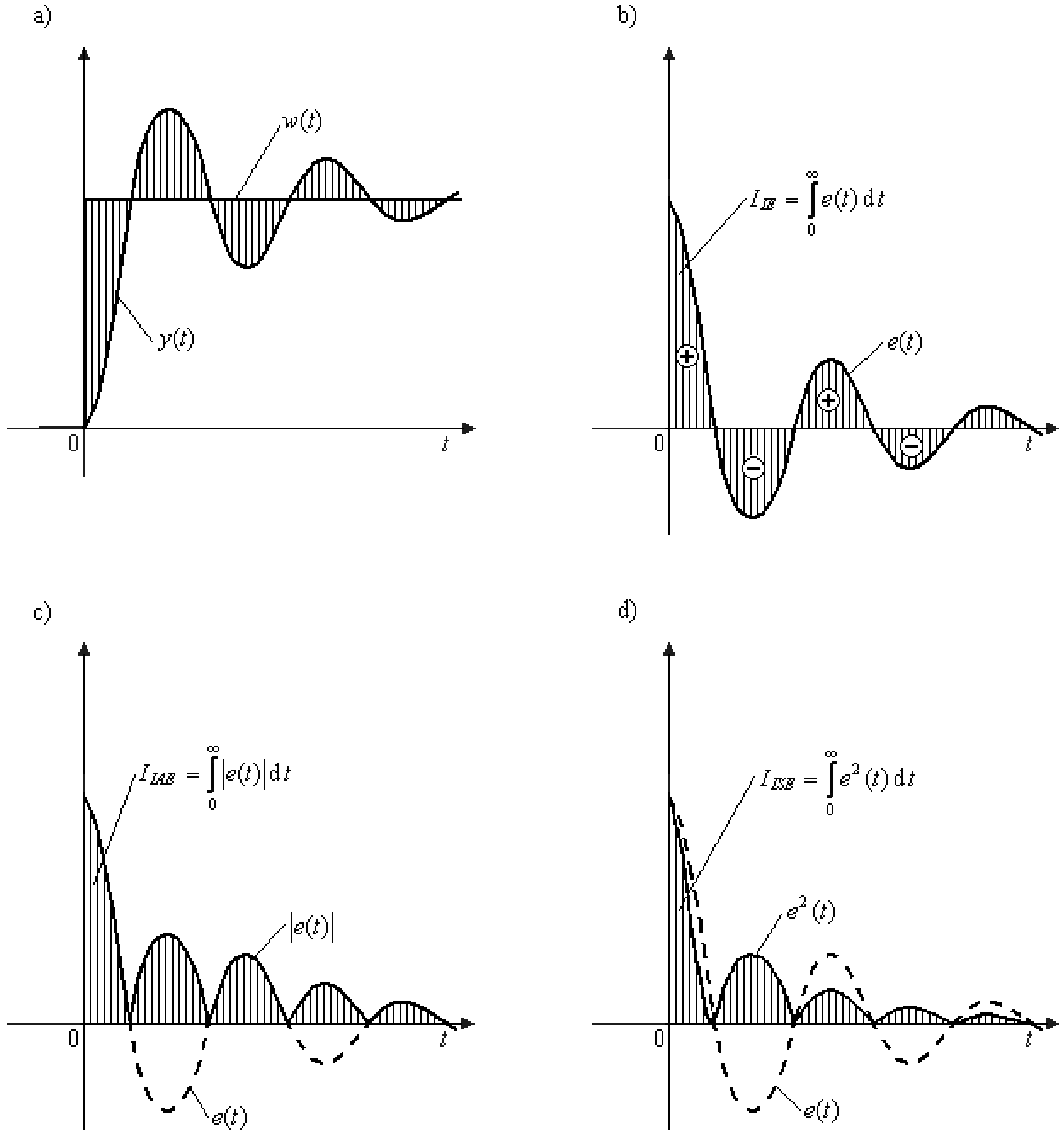


Fig. 4.6 – Geometrical interpretation of integral criteria: a) control area, b) integral of error I_{IE}, c) integral of absolute error I_{IAE}, d) integral of squared error I_{ISE}

Integral of squared error (Fig. 4.6d)

$$I_{ISE} = \int_0^\infty e^2(t)\,\mathrm{d}t \rightarrow \min \tag{4.9}$$

The integral of squared error I_{ISE} (ISE = **I**ntegral of **S**quared **E**rror) removes the disadvantages of both previous integral criteria I_{IE} and I_{IAE}. It can be used for non-oscillating and oscillating control processes and its value can be calculated in an analytical way. It is very suitable in these cases when the desired $w(t)$ and the disturbance $v(t)$ variables have a random character. Some disadvantage of the integral of squared error consists in that the control process is too oscillating.

For the control error transform

$$E(s)=\frac{b_{n-1}s^{n-1}+\ldots+b_1s+b_0}{a_ns^n+\ldots+a_1s+a_0} \tag{4.10}$$

can be computed:

$$n=1 \quad I_{ISE}=\frac{b_0^2}{2a_0a_1} \tag{4.11}$$

$$n=2 \quad I_{ISE}=\frac{a_0b_1^2+a_2b_0^2}{2a_0a_1a_2} \tag{4.12}$$

$$n=3 \quad I_{ISE}=\frac{a_0a_1b_2^2+(b_1^2-2b_0b_2)a_0a_3+a_2a_3b_0^2}{2a_0a_3(a_1a_2-a_0a_3)} \tag{4.13}$$

For higher degree n the formulas are very complex.

ITAE criterion

$$I_{ITAE}=\int_0^\infty t\left|e(t)\right|\mathrm{d}t \rightarrow \min \tag{4.14}$$

The ITAE criterion I_{ITAE} (ITAE = **I**ntegral of **T**ime multiplied by **A**bsolute **E**rror) contains the time and the error and therefore it simultaneously minimalizes both the settling time and the error. This integral criterion is very popular among technicians though its value can be determined generally by simulation.

For the given control system type q and the characteristic polynomial $N(s)$ with degree n so-called **standard forms** of the control system transfer functions were determined by simulation for minimum of the ITAE criterion.

Below are shown the standard forms only for $q = 1$, $n = 2$ and 3:

$$n=2,\ G_{wy}(s)=\frac{a^2}{s^2+1.4as+a^2} \ \Rightarrow\ G_o(s)=\frac{a^2}{s^2+1.4as}=\frac{a^2}{s(s+1.4a)} \tag{4.15}$$

$$n=3,\ G_{wy}(s)=\frac{a^3}{s^3+1.75as^2+2.15a^2s+a^3} \ \Rightarrow$$
$$G_o(s)=\frac{a^3}{s^3+1.75as^2+2.15a^2s}=\frac{a^3}{s(s^2+1.75as+2.15a^2)} \tag{4.16}$$

The parameter a matches the time scales of the original system and its model in a standard form. From both transfer functions of the open-loop control system $G_o(s)$ it follows that they contain one integrating element, i.e. $q = 1$.

Only the most important integral criteria were briefly described. By their minimization the optimal values of the adjustable controller parameters can be obtained. The minimization is generally done by simulation.

The integral criteria I_{IAE} and I_{ITAE} can be used for control performance comparison and assessment of the different control processes.

Frequency Domain

The frequency domain is also suitable for assessing the control performance. It is the most favorite for the control system designers. Most often three frequency transfer functions are used (Fig. 4.7):

the frequency (closed-loop) control system transfer function

$$G_{wy}(\mathrm{j}\omega) = \frac{G_C(\mathrm{j}\omega)G_P(\mathrm{j}\omega)}{1 + G_C(\mathrm{j}\omega)G_P(\mathrm{j}\omega)} = T(\mathrm{j}\omega) \tag{4.17}$$

the frequency open-loop transfer function

$$G_o(\mathrm{j}\omega) = G_C(\mathrm{j}\omega)G_P(\mathrm{j}\omega) \tag{4.18}$$

the frequency disturbance transfer function (for the disturbance in the plant output)

$$G_{vy}(\mathrm{j}\omega) = \frac{1}{1 + G_C(\mathrm{j}\omega)G_P(\mathrm{j}\omega)} = 1 - G_{wy}(\mathrm{j}\omega) = S(\mathrm{j}\omega) \tag{4.19}$$

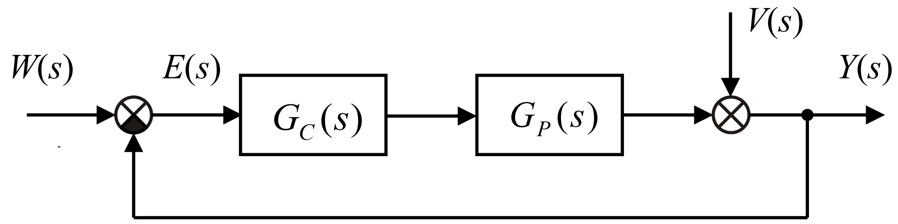


Fig. 4.7 – Control system

From the frequency control system transfer function (4.17) the modulus (magnitude) or logarithmic modulus (magnitude) can be obtained

$$A_{wy}(\omega) = \operatorname{mod} G_{wy}(\mathrm{j}\omega) = |G_{wy}(\mathrm{j}\omega)| \quad \text{or} \quad L_{wy}(\omega) = 20\log A_{wy}(\omega) \tag{4.20}$$

The typical course of the magnitude response of the control system $A_{wy}(\omega)$ is in Fig. 4.8. From Fig. 4.8 the following control performance indices can be get: $A_{wy}(\omega_R)$ – the **peak resonance (resonant magnitude)**, ω_R – the **resonant angular frequency**, ω_m – the **cutoff angular frequency**.

For the well-tuned control system the relations

$$A_{wy}(\omega_R) \leq 1.1 \div \mathbf{1.5} \quad \text{or} \quad L_{wy}(\omega_R) \leq (0.8 \div \mathbf{3.5})\ \text{dB} \qquad (4.21)$$

hold.

A too high value of peak resonance gives high oscillation and a great overshoot.

The cutoff angular frequency ω_m determines the operating bandwidth, i.e. the **region of the operating angular frequencies.** Its higher value enables the control system to better process higher angular frequencies. The cutoff angular frequency ω_m is given by a decrease of the modulus $A_{wy}(\omega)$ [$L_{wy}(\omega)$] on the value $\frac{1}{\sqrt{2}} A_{wy}(0) \doteq 0.707 A_{wy}(0)$ [$L_{wy}(0)$ = – 3 dB] and for the big peak resonance $A_{wy}(\omega_R)$ by increasing the modulus $A_{wy}(\omega)$ [$L_{wy}(\omega)$] to the value $\sqrt{2} A_{wy}(0) \doteq 1.414 A_{wy}(0)$ [$L_{wy}(0) = +3$ dB].

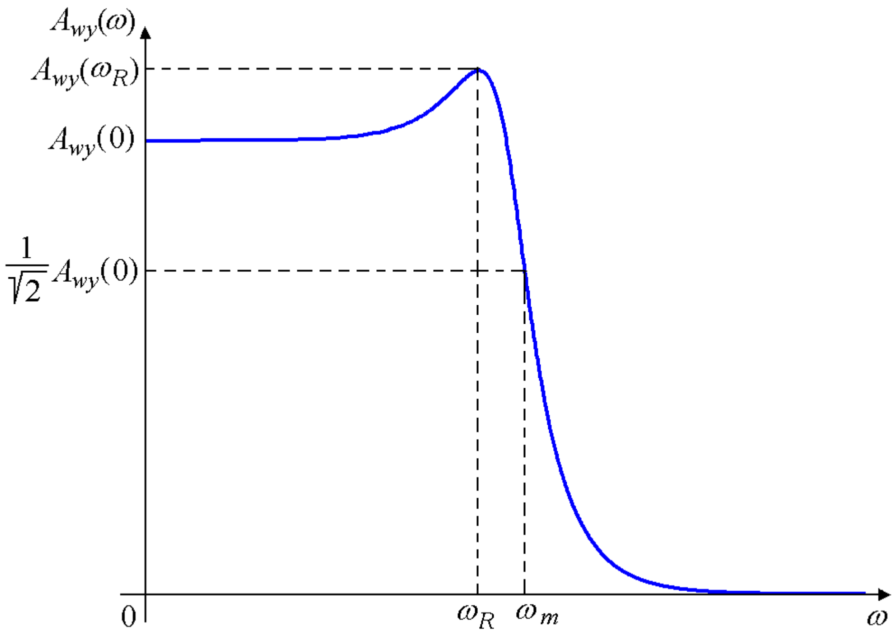


Fig. 4.8 – Magnitude response of a control system

On the basis of the magnitude response of the control system $A_{wy}(\omega)$ the control system type q can be determined because relations

$$A_{wy}(0) = 1 \quad \text{or} \quad L_{wy}(0) = 0 \ \Rightarrow \ q \geq 1 \qquad (4.22a)$$

$$A_{wy}(0) < 1 \quad \text{or} \quad L_{wy}(0) < 0 \ \Rightarrow \ q = 0 \qquad (4.22b)$$

hold.

The control system type q can be determined on the basis of the frequency response of the open-loop control system $G_o(\mathrm{j}\omega)$ for $\omega \to 0$, see Figs 3.12 ÷ 3.14 and also Fig. 4.10.

The frequency response of the open-loop control system $G_o(\mathrm{j}\omega)$ is very useful because it enables pointing out very important control performance indices like the gain margin m_A and the phase margin γ, see Figs 3.14 and 4.10. For common control systems there are recommended following values:

$$m_A = \mathbf{2} \div 5 \quad \text{or} \quad m_L = 20\log m_A = (\mathbf{6} \div 14)\ \text{dB} \tag{4.23a}$$

$$\gamma = \mathbf{30}° \div 60° \quad \left(\frac{\pi}{6} \div \frac{\pi}{3}\right) \tag{4.23b}$$

The bold values should not be exceeded.

The frequency transfer functions $G_{wy}(\mathrm{j}\omega)$ and $G_{vy}(\mathrm{j}\omega)$ [see Fig. 4.8 and relations (4.17), (4.19)] have the fundamental meaning for the theory of automatic control and therefore they are specially inscribed by symbols $G_{wy}(\mathrm{j}\omega) = T(\mathrm{j}\omega)$ and $G_{vy}(\mathrm{j}\omega) = S(\mathrm{j}\omega)$ and they have special names. From the relation (4.19) it follows

$$G_{wy}(\mathrm{j}\omega) + G_{vy}(\mathrm{j}\omega) = 1 \;\Leftrightarrow\; T(\mathrm{j}\omega) + S(\mathrm{j}\omega) = 1 \tag{4.24}$$

The $S(\mathrm{j}\omega)$ is called the sensitivity function and the $T(\mathrm{j}\omega)$ is the complementary sensitivity function.

The name of the $S(\mathrm{j}\omega)$ "sensitivity function" follows from the next considerations.

From

$$Y(\mathrm{j}\omega) = G_{wy}(\mathrm{j}\omega)W(\mathrm{j}\omega) \tag{4.25}$$

for $W(\mathrm{j}\omega)$ = constant the relation

$$\frac{\mathrm{d}Y(\mathrm{j}\omega)}{Y(\mathrm{j}\omega)} = \frac{\mathrm{d}G_{wy}(\mathrm{j}\omega)}{G_{wy}(\mathrm{j}\omega)} \tag{4.26}$$

is obtained, i.e. the relative change of the controlled variable (its transform) is equal to the relative change of the control system behavior (its transfer function). Similarly on the basis of (4.17) the relation

$$\frac{\mathrm{d}G_{wy}(\mathrm{j}\omega)}{G_{wy}(\mathrm{j}\omega)} = \frac{1}{1 + G_C(\mathrm{j}\omega)G_P(\mathrm{j}\omega)}\left[\frac{\mathrm{d}G_C(\mathrm{j}\omega)}{G_C(\mathrm{j}\omega)} + \frac{\mathrm{d}G_P(\mathrm{j}\omega)}{G_P(\mathrm{j}\omega)}\right]$$

or

$$\frac{\mathrm{d}Y(\mathrm{j}\omega)}{Y(\mathrm{j}\omega)} = \frac{\mathrm{d}G_{wy}(\mathrm{j}\omega)}{G_{wy}(\mathrm{j}\omega)} = S(\mathrm{j}\omega)\left[\frac{\mathrm{d}G_C(\mathrm{j}\omega)}{G_C(\mathrm{j}\omega)} + \frac{\mathrm{d}G_P(\mathrm{j}\omega)}{G_P(\mathrm{j}\omega)}\right] \tag{4.27}$$

can be obtained, which expresses the influence of the relative changes of the controller and the plant behaviors (their transfer functions) on the relative change of the control system (its transfer function), and hence on a relative change of the controlled variable (its transform). It is obvious that this influence expresses just the sensitivity function $S(\mathrm{j}\omega)$. For its small value the influence of the relative changes of the controller and plant behaviors on the behavior of the control system and therefore on the controlled variable will be small too.

It has a small value if the relations (3.15) or (3.16) hold.

The sensitivity function $S(\mathrm{j}\omega)$ then expresses the **sensitivity** of the control system to small unspecified changes of the control system elements, first of all the plant.

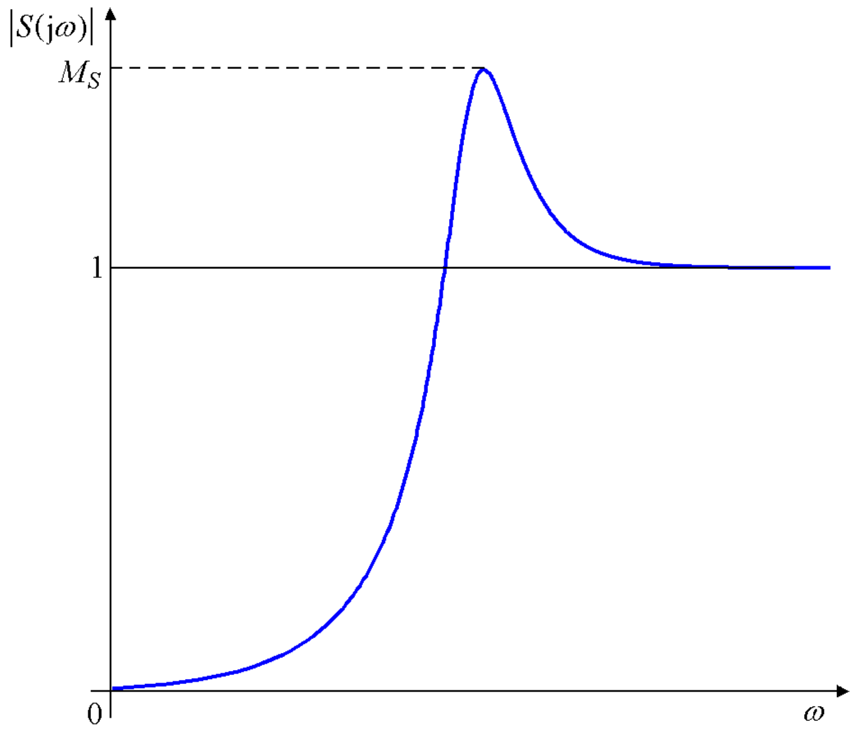


Fig. 4.9 – Course of the modulus of the sensitivity function

In Fig. 4.9 the typical course of the modulus of the sensitivity function $|S(\mathrm{j}\omega)| = \operatorname{mod} S(\mathrm{j}\omega)$ is shown. The scale of the angular frequency ω is often logarithmic.

The maximum value of the sensitivity function modulus

$$M_S = \max_{0 \le \omega < \infty} |S(\mathrm{j}\omega)| = \max_{0 \le \omega < \infty} \left| \frac{1}{1 + G_C(\mathrm{j}\omega) G_P(\mathrm{j}\omega)} \right| \tag{4.28}$$

has a very important interpretation.

The inverted value of the maximum of the sensitivity function modulus $1/M_S$ is the shortest distance of the open-loop frequency response $G_o(j\omega)$ to the critical point -1 + j0, see Fig. 4.10.

This value M_S for a well-tuned control system should not be more than 2 and it ought be in the interval

$$1.3 \le M_S \le \mathbf{2} \tag{4.29}$$

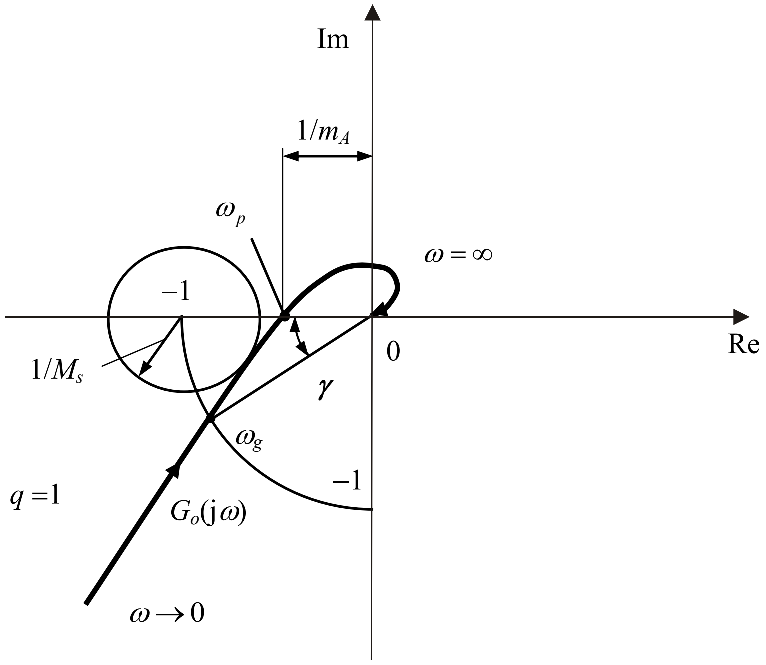


Fig. 4.10 – Geometrical interpretation of the maximum of the sensitivity function modulus

The estimations follow from Fig. 4.10 – the gain margin

$$m_A > \frac{M_S}{M_S - 1} \tag{4.30}$$

and the phase margin

$$\gamma > 2 \arcsin \frac{1}{2M_S} \tag{4.31}$$

The maximum of the sensitivity function modulus M_S is the complex control performance index because from the relation (4.30) and (4.31) it follows that for $M_S \le 2$ it ensures the gain margin $m_A \ge 2$ and the phase margin $\gamma > 29$ °. The reversed statement doesn't hold, i.e. the values m_A and γ don't ensure the corresponding value M_S.

The sensitivity of the control system is related to its **robustness**. The robustness of the control system is its ability to hold the control objective for the given changes mostly of the plant (or other control system elements) behavior.

The control performance can go down in the determined range but the **control system stability must be always ensured**.

S-domain

The control system pole placement, i.e. the control system transfer function $G_{wy}(s)$ pole placement has a principal influence on control performance. The influence the control system transfer function $G_{wy}(s)$ pole placement on control system behavior is shown in Fig. 3.8. It is supposed that the control system is stable, i.e. all its poles lie in the left half of the *s*-complex plane. The influence on dynamic behavior is best seen on the second order oscillating system with the transfer function

$$G(s) = \frac{Y(s)}{U(s)} = \frac{1}{T_0^2 s^2 + 2\xi_0 T_0 s + 1} = \frac{\omega_0^2}{s^2 + 2\xi_0 \omega_0 s + \omega_0^2} \tag{4.32}$$

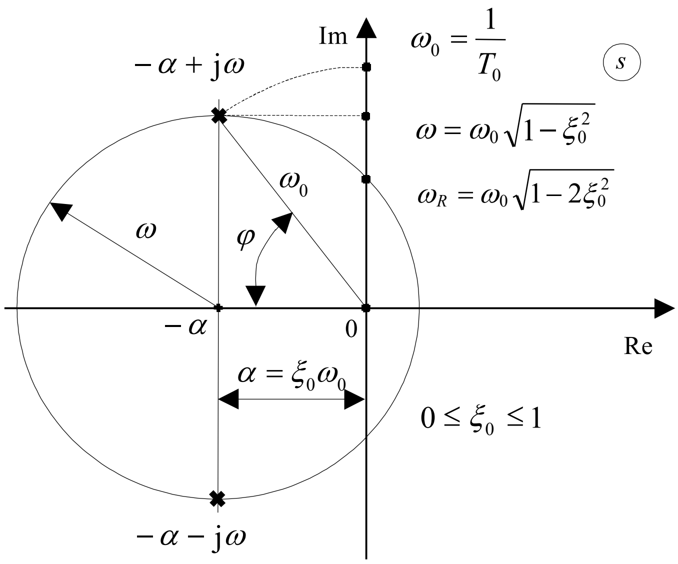


Fig. 4.11 – Geometrical interpretation of the second order oscillating system parameters

and the step response

$$h(t) = \mathrm{L}^{-1}\left\{ \frac{1}{s(T_0^2 s^2 + 2\xi_0 T_0 s + 1)} \right\} = 1 - C\,\mathrm{e}^{-\alpha t} \sin(\omega t + \varphi) \tag{4.33}$$

$$C = \frac{1}{\omega T_0} = \frac{\omega_0}{\omega}, \qquad \alpha = \frac{\xi_0}{T_0} = \xi_0 \omega_0, \qquad \omega_0 = \frac{1}{T_0}$$

$$\omega = \frac{1}{T_0}\sqrt{1-\xi_0^2} = \omega_0\sqrt{1-\xi_0^2}, \quad \omega_R = \omega_0\sqrt{1-2\xi_0^2}, \quad \varphi = \mathrm{arctg}\frac{\omega}{\alpha} = \arccos \xi_0$$

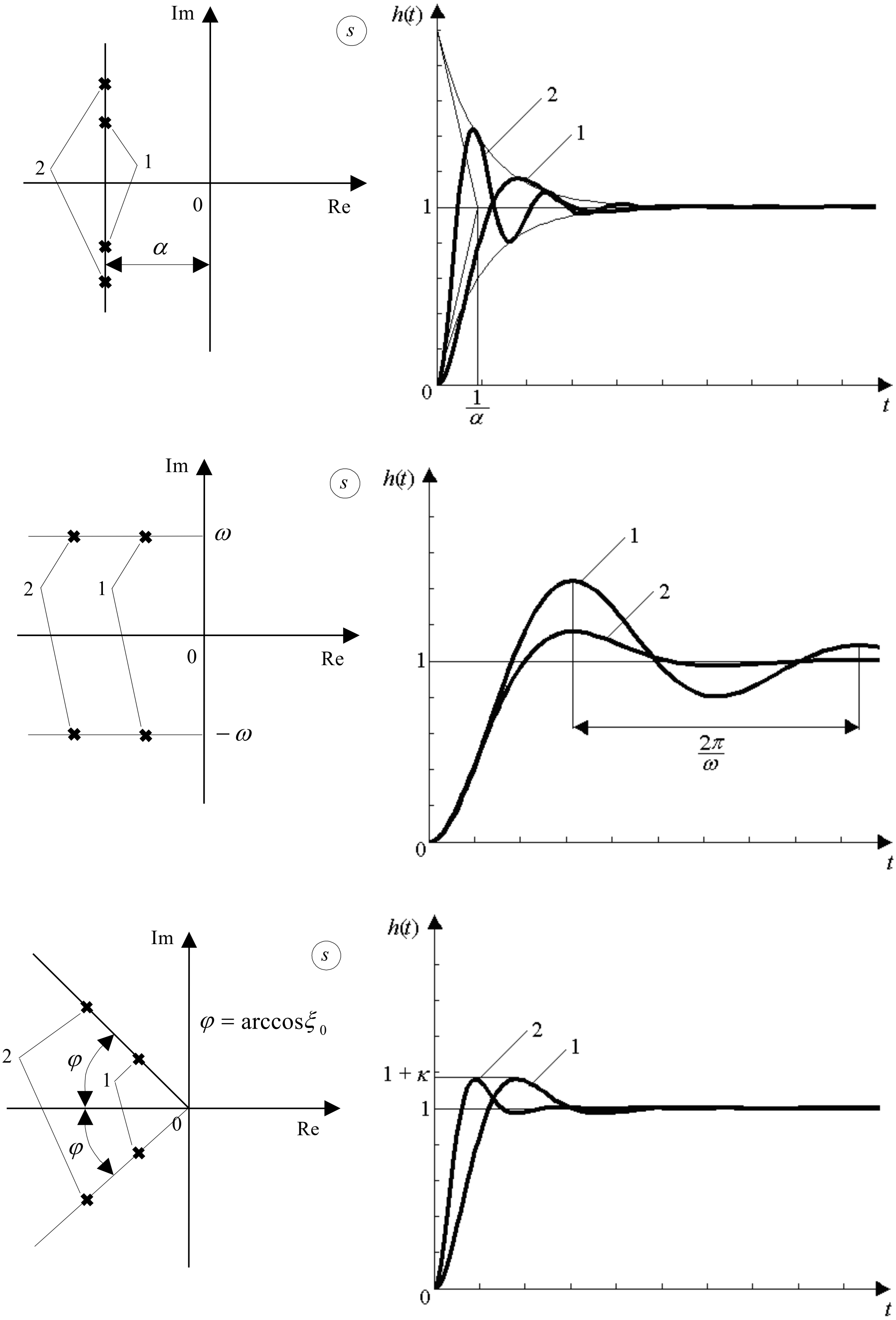


Fig. 4.12 – Influence of complex conjugate poles of the second order oscillating system on its step responses

The geometrical interpretation of the second order oscillating system parameters is shown in Fig. 4.11 and the influence of the second order oscillating system poles on its step responses is in Fig. 4.12. Some of these parameters have special names: ω_0 is the **natural angular frequency**, ω – the **damped angular frequency**, ω_R – the resonant angular frequency, ξ_0 – the **damping ratio**, α – the **stability degree (damping)**. The dimension of the stability degree α ($\alpha > 0$) is [time^{-1}] in contrast to the dimensionless damping ratio ξ_0 and expresses the distance of the couple poles from the imaginary axis. It indicates the exponential fall rate of the step response $h(t)$, i.e. the exponential approaching the steady state $h(\infty)$ [see relation (4.33) and Fig. 4.12].

The meaning of the stability degree α is shown for the first order plant (Fig. 4.13a) and for the second order (Fig. 4.13b). From both figures it is obvious that for the higher stability degree, α the settling time t_r is shorter.

The damping ratio ξ_0 determines the relative overshoot κ (Fig. 4.12). Two half lines correspond to the constant damping ratio ξ_0, which make the negative real axis the angle φ [the complex roots (poles) always rise in the complex conjugate couples].

Then it is obvious that on the basis of the control performance requirements, which are expressed for the given control system by the maximum settling time t_r and the maximum relative overshoot κ it is possible to determine the admissible region in the left half of the s-complex plane in that the all control system poles must lie, see Fig. 4.14. The poles lying the closest to the admissible region boundary are called the **dominant poles** (sometimes as the dominant poles are thought the ones which are the closest to the imaginary axis). Furthermore, it is supposed that the poles lying far from the admissible region boundary have a negligible influence on control system behavior.

The admissible region boundary in Fig. 4.14 is determined by the relations

$$\alpha_w \geq (3 \div 5)\frac{1}{t_r} \tag{4.34}$$

$$\varphi_w \leq \arccos \xi_w \tag{4.35}$$

In the case of the one dominant pole the smaller number in (4.34) is considered and in case of the double dominant pole there is considered the greater number. The first relation is given for the control tolerance at about 5 %. From the second relation for the maximum relative overshoot $\kappa = 0.25$ it is possible to get

$$\kappa \leq 0.25 \;\Rightarrow\; \xi_0 \geq 0.404 \;\Rightarrow\; \varphi_w \leq 66^\circ \;(1.15\,\text{rad})$$

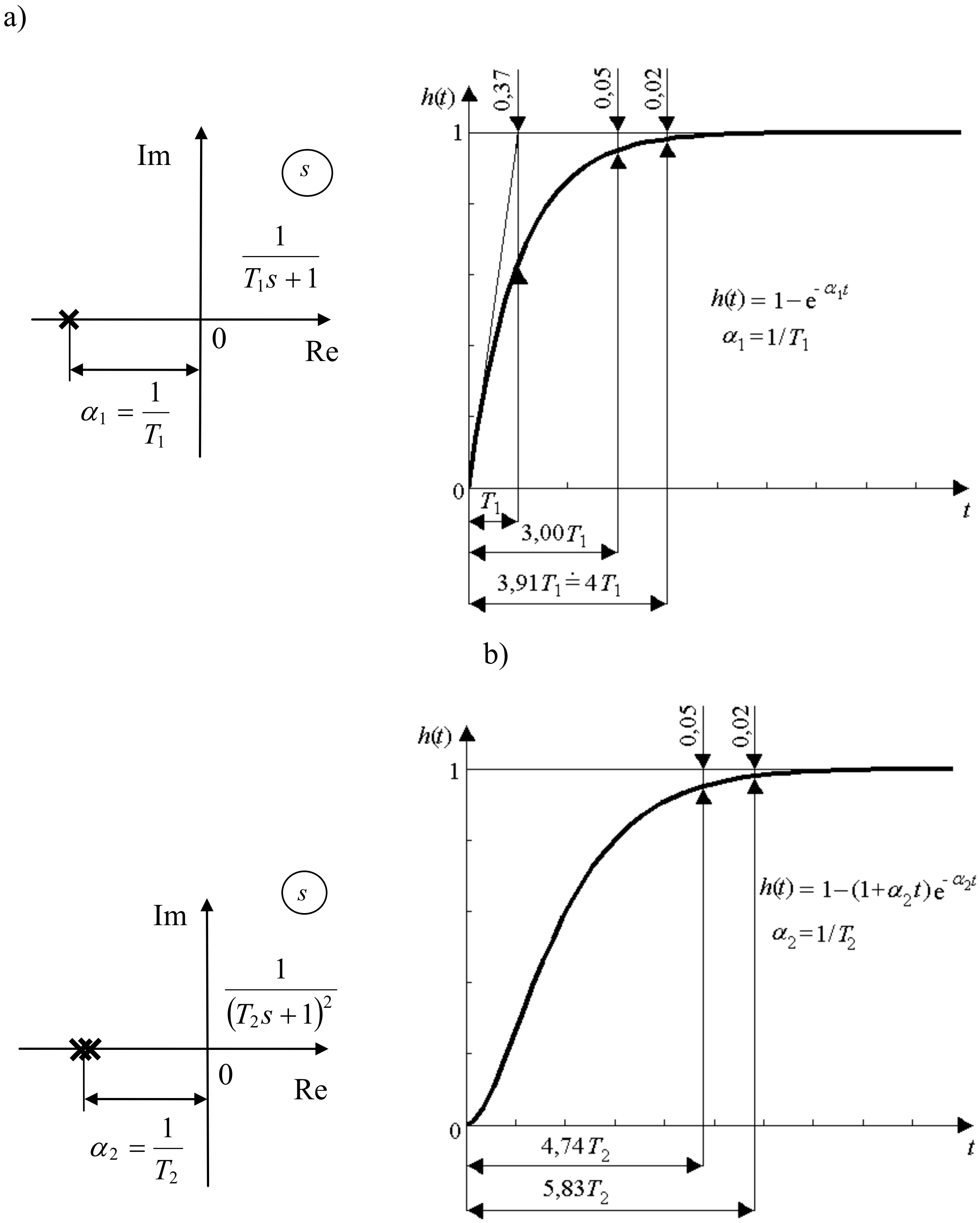


Fig. 4.13 – Influence of stability degree (damping) on the step response and settling time for a non-oscillating system of: a) the first order, b) the second order

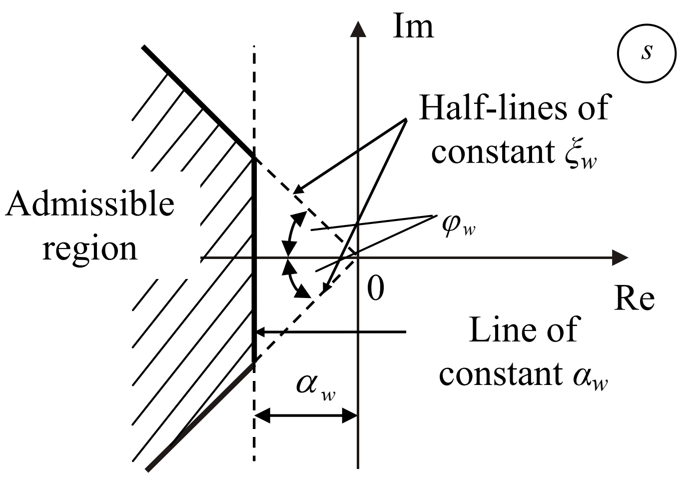


Fig. 4.14 – Determination of admissible region for control system poles

4.2 Controller Tuning

The synthesis belongs to the most important procedures in control system design. It consists of the choice of the suitable controller type and its subsequent tuning from the point of view of given control performance requirements. A rise of the steady state errors is mostly undesirable and therefore the control system type $q = 1$ is mostly chosen. The higher control system type q ensures the zeroness of the steady state errors but it simultaneously increases a disposition for control system instability and makes it difficult for controller tuning. The control system type $q = 0$ can be used only for very simple control systems with a desired low control performance. In the case of control systems with a time delay, the steady-state errors would be inadmissibly great. Generally it holds that the controller with more components (terms) gives the better control performance.

The task of the controller consists in the fulfillment of the control objective (3.2) [or (3.8)] with the desired control performance. It was shown in subchapter 3.1 that it is possible in the case of fulfillment of the conditions (3.15) or (3.16) of course for a sufficient stable control system. All these conditions can hold by choosing the corresponding controller and its suitable tuning.

The conditions (3.15) or (3.16) are very important because their fulfillment ensures the low value of the sensitivity function $S(j\omega)$ [see (4.27)] and therefore the small influence of the relative controller and plant behaviors changes on the relative controlled variable changes.

It is important that for the "smooth" extreme (i.e. minimum or maximum) the small changes of the parameters on which it depends have little influence on its optimal value (the gradient for the smooth extreme is zero), see Fig. 4.15. This figure shows the dependency of the chosen performance index (criterion) I on the controller gain K_P. Therefore it is useful to have the values of the adjustable controller parameters for the given performance index (criterion) determined by optimization.

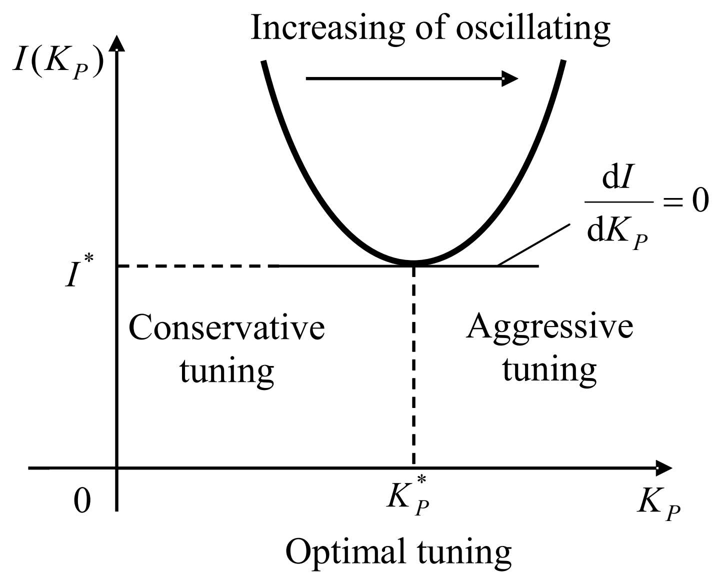


Fig. 4.15 – Dependence of performance index I on controller gain K_P

From all these arguments it follows that appropriate attention must be given to the controller choice and its tuning for the "nominal" (i.e. given or identified) plant.

Conventional controller tuning methods are experimental, analytical and combined.

Experimental methods „trial and error"

The **„trial and error" methods** belong to the basic experimental methods. These methods are often used in practice because they operate with a real (true) closed-loop control system and therefore they don't demand in principle any knowledge about plant behavior. These methods are applied on the existing control systems, which must be fine-tuned or tuned after redesign or repair.

From the many existing "trial and error" methods there will be described only one method which is simple and effective.

Procedure:

1. All connection of the control system and the functionality of its devices must be checked.
2. The desired variable (setpoint) value $w(t)$ is set and in the manual mode $y_w(t) \approx w(t)$ is set too, the integral and the derivative components shut down (i.e. $T_I \to \infty$ and $T_D \to 0$), the controller gain K_P is decreased and the controller is switched to the automatic mode.

3. The controller gain K_P is subsequently increased so as the desired step response $y_w(t)$ is obtained (the steady-state error doesn't matter).
4. The controller gain K_P is decreased on the 3/4 of the previous value and the integral time T_I is slowly decreased so as the possible steady state-error is removed and the desired step response $y_w(t)$ is obtained. It is often suitable that this step response is marginally non-oscillating.
5. The final desired step response $y_w(t)$ is obtained by fine-tuning.
6. In the case of using the derivative component (term) the derivative time T_D is set to value 1/10 T_I. If noises arise or the manipulated variable $u(t)$ is too active then using the derivative component isn't proper and it is shut down. If by using the derivative component the control performance is better than the derivative time T_D rises to the value 1/4 T_I, the controller gain K_P rises about 1/4 of the previous value (i.e. the value obtained in step 5) and the integral time T_I decreases about 1/3 of the previous value (i.e. the value obtained in step 4).

The described tuning procedure is simply and easy to use.

Experimental Ziegler – Nichols methods

The experimental Ziegler – Nichols methods belong among classical experimental controller tuning methods. They are suitable for preliminary tuning of the conventional controllers because they mostly give a big overshoot in the range from 10 % to 60 %, at average for different plants around 25 % (the quarter-decay criterion), see Figs 4.16 and 4.18.

For the PID controller the constant ratio

$$\frac{T_D^*}{T_I^*} = \frac{1}{4} \tag{4.36}$$

is very interesting.

The controller tuning by the experimental Ziegler – Nichols methods is suitable in cases when the disturbance variable $v(t)$ influences the plant input.

Further two original Ziegler – Nichols methods and the one modification which derives from them are described.

Open-loop method

The **open-loop method** (the step response method) comes from the step response of the plant. The time delay T_u, the time constant T_n and the plant gain k_1 are determined in accordance with Fig. 3.6a and on the basis of Tab. 4.1 the values of the adjustable controller parameters are computed.

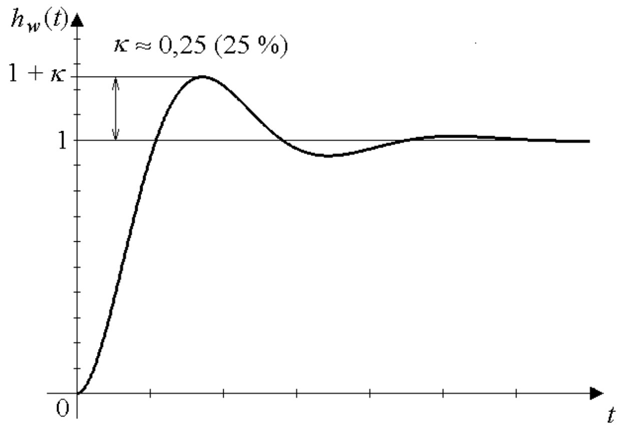


Fig. 4.16 – „Average“ step response of control system tuned by experimental Ziegler – Nichols methods

Tab. 4.1 – Values of adjustable controller parameters for Ziegler – Nichols open-loop method

Controller	K_P^*	T_I^*	T_D^*
P	$\dfrac{T_n}{k_1 T_u}$	–	–
PI	$0.9\dfrac{T_n}{k_1 T_u}$	$3.33T_u$	–
PID	$1.2\dfrac{T_n}{k_1 T_u}$	$2T_u$	$0.5T_u$

The destabilizing influence of the integral component of the PI controller evokes decreasing the controller gain K_P^* in comparison with the P controller and the stabilizing influence of the derivative component of the PID controller evokes increasing the controller gain K_P^* (compare Tab.4.1 with Tab. 4.2).

The PID controller transfer function

$$G_C(s) = K_P^*\left(1+\frac{1}{T_I^* s}+T_D^* s\right) = 1.2\frac{T_n}{k_1 T_u}\left(1+\frac{1}{2T_u s}+\frac{T_u}{2}s\right) = 0.6\frac{T_n}{k_1 T_u^2}\frac{(T_u s+1)^2}{s} \tag{4.37}$$

is interesting. It shows that the PID controller tuned by the Ziegler – Nichols open-loop method has the double zero $z_2 = -1/T_u$.

Procedure:

1. From the plant step response the plant gain k_1 and the times T_u and T_n are determined (see subchapter 3.2, Fig. 3.6).

2. On the basis of Tab. 4.1 for a chosen controller the values of its adjustable parameters are computed.

Closed-loop method

The **closed-loop method** (the ultimate parameters method) comes from the real (true) closed-loop control system. The ultimate (critical) value of the controller gain K_{Pc} and the ultimate period T_c (Fig. 4.17) for the P controller are determined. Then on the basis of Tab. 4.2 the values of the adjustable controller parameters are computed.

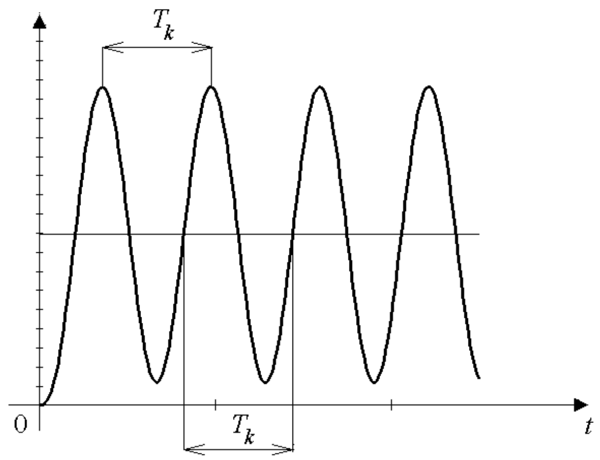


Fig. 4.17 – Determination of ultimate period T_c

Tab. 4.2 – Values of adjustable controller parameters for the Ziegler – Nichols closed-loop method

Controller	K_P^*	T_I^*	T_D^*
P	$0.5K_{Pc}$	–	–
PI	$0.45K_{Pc}$	$0.83T_c$	–
PID	$0.6K_{Pc}$	$0.5T_c$	$0.125T_c$

The PID controller transfer function tuned by the closed-loop method has an interesting form too

$$G_C(s) = K_P^*\left(1+\frac{1}{T_I^* s}+T_D^* s\right) = 0.6K_{Pc}\left(1+\frac{1}{0.5T_c s}+0.125T_c s\right) = 1.2\frac{K_{Pc}}{T_c}\frac{\left(\frac{T_c}{4}s+1\right)^2}{s} \tag{4.38}$$

From comparison of (4.37) and (4.38) it follows

$$K_{Pc} = 2\frac{T_n}{k_1 T_u}, \qquad T_c = 4T_u \tag{4.39}$$

The relations (4.39) for $T_u < T_n$ can be used for approximately determining the ultimate parameters K_{Pc} and T_c.

From the first relation (4.39) and Tab. 4.2 it follows that both Ziegler – Nichols methods in the case of the use the P controller have the same gain margin m_A = 2, i.e. for doubly increasing the controller gain K_P the control system reaches the oscillating stability boundary.

The closed-loop method is applicable even for the I controllers. In this case the closed-loop control system is brought up on the stability boundary by decreasing the integral time T_I. On the stability boundary the ultimate (critical) integral time T_{Ic} is determined and then for tuning the value

$$T_I^* = 2T_{Ic} \tag{4.40}$$

is used. Even in this case the gain margin is the same $m_A = 2$.

If the non-oscillating control process is demanded then there is chosen

$$T_I^* = (4 \div 6)T_{Ic} \tag{4.41}$$

with the gain margin $m_A = 4 \div 6$.

The closed-loop Ziegler – Nichols method is useful above all because it doesn't suppose any a priori knowledge of the plant behavior and that it operates with the real (true) plant and controller. Its basic disadvantage is that it must bring up the control system to stability boundary, i.e. the control system must oscillate which could cause damage to the plant or its non-linear behavior can arise.

In case the plant doesn't contain the time delay and its behavior is known then the ultimate parameters K_{Pc} and T_c or T_{Ic} can be obtained analytically by the use of the Mikhailov stability criterion (see subchapter 3.3).

Procedure:

1. and 2. the same steps like for the „trial and error" method.
3. The controller gain K_P is subsequently increased as for small change of the desired value $w(t)$ the oscillating stability boundary arises.
4. From the periodic course of any variable, the ultimate period T_c and from the P controller setting the ultimate gain K_{Pc} are determined.
5. For the chosen controller on the basis of Tab. 4.2 the values of its adjustable parameters are computed.

Quarter-decay method

The **quarter-decay method** is a specific modification of the closed-loop Ziegler – Nichols method. In contrast to it the quarter-decay method doesn't suppose to bring up the control system to the oscillating stability boundary which enables operation in the linear region and use for more plants.

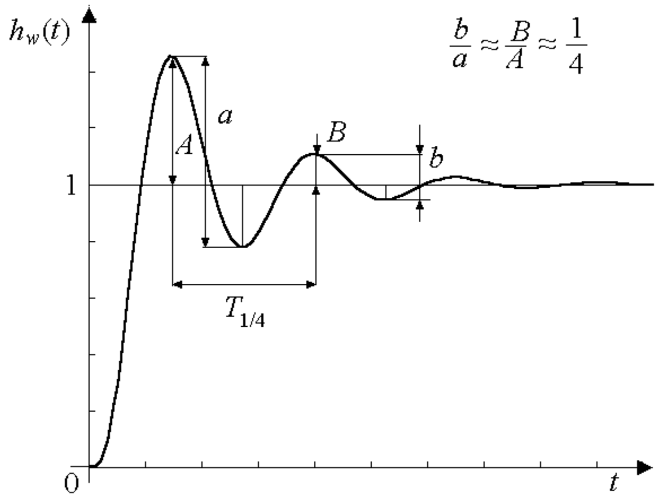


Fig. 4.18 – Control system tuning by the quarter-decay method

Tab. 4.3 – Values of adjustable controller parameters for the quarter-decay method

Controller	K_P^*	T_I^*	T_D^*
P	$K_{P1/4}$	–	–
PI	$0.9K_{P1/4}$	$T_{1/4}$	–
PID	$1.2K_{P1/4}$	$0.6T_{1/4}$	$0.15T_{1/4}$

Procedure:

1. and 2. the same steps like for the „trial and error" method.
3. The controller gain K_P is subsequently increased as the step response $h_w(t)$ holds that the ratio of the two consecutive amplitudes is equal to ¼, see Fig. 4.18.
4. From the step response $h_w(t)$ the time $T_{1/4}$ and from the P controller setting the controller gain $K_{P1/4}$ are determined.
5. For the chosen controller on the basis of Tab. 4.3 the values of its adjustable parameters are computed.

„Universal" experimental method

The **„universal" experimental method** was elaborated in the former Soviet Union. It is supposed the plants with the transfer functions

$$G_P(s) = \frac{k_1}{T_1 s + 1} \mathrm{e}^{-T_d s} \tag{4.42}$$

Tab. 4.4 – Values of adjustable controller parameters for the „universal" experimental method – transfer function (4.42)

$\frac{k_1}{T_1 s+1}\mathrm{e}^{-T_d s}$		Control process				
		Fastest response without overshoot		Fastest response with overshoot 20 %		Minimum of ISE
Controller type		Tuning from point of view				
		Desired variable w	Disturbance variable v	Desired variable w	Disturbance variable v	Disturbance variable v
P	K_P^*	$0.3\frac{T_1}{k_1 T_d}$	$0.3\frac{T_1}{k_1 T_d}$	$0.7\frac{T_1}{k_1 T_d}$	$0.7\frac{T_1}{k_1 T_d}$	–
PI	K_P^*	$0.35\frac{T_1}{k_1 T_d}$	$0.6\frac{T_1}{k_1 T_d}$	$0.6\frac{T_1}{k_1 T_d}$	$0.7\frac{T_1}{k_1 T_d}$	$\frac{T_1}{k_1 T_d}$
	T_I^*	$1.17T_1$	$0.8T_d + 0.5T_1$	T_1	$T_d + 0.3T_1$	$T_d + 0.35T_1$
PID	K_P^*	$0.6\frac{T_1}{k_1 T_d}$	$0.95\frac{T_1}{k_1 T_d}$	$0.95\frac{T_1}{k_1 T_d}$	$1.2\frac{T_1}{k_1 T_d}$	$1.4\frac{T_1}{k_1 T_d}$
	T_I^*	T_1	$2.4T_d$	$1.36T_1$	$2T_d$	$1.3T_d$
	T_D^*	$0.5T_d$	$0.4T_d$	$0.64T_d$	$0.4T_d$	$0.5T_d$

and

$$G_P(s) = \frac{k_1}{s} \mathrm{e}^{-T_d s} \tag{4.43}$$

The "universal" experimental method enables conventional controller tuning both from the point of view of the desired variable $w(t)$ and from the point of view of the disturbance variable $v(t)$ which acts on plant input for three control performance indices (criteria). These control performance indices are: the fastest response without overshoot, the fastest response with the relative overshoot $\kappa = 0.2$ (20 %) and the minimum of the integral of the squared error. This method, as with the control process without the overshoot, considers the control process with a maximum relative overshoot from 0.02 (2 %) to 0.05 (5 %).

Tab. 4.5 – Values of adjustable controller parameters for the "universal" experimental method– transfer function (4.43)

$\frac{k_1}{s}\mathrm{e}^{-T_d s}$		Control process				
		Fastest response without overshoot		Fastest response with overshoot 20 %		Minimum of ISE
Controller type		Tuning from point of view				
		Desired variable w	Disturbance variable v	Desired variable w	Disturbance variable v	Disturbance variable v
P	K_P^*	$0.37\frac{1}{k_1T_d}$	$0.37\frac{1}{k_1T_d}$	$0.7\frac{1}{k_1T_d}$	$0.7\frac{1}{k_1T_d}$	–
PI	K_P^*	$0.37\frac{1}{k_1T_d}$	$0.46\frac{1}{k_1T_d}$	$0.7\frac{1}{k_1T_d}$	$0.7\frac{1}{k_1T_d}$	$\frac{1}{k_1T_d}$
	T_I^*	∞	$5.75T_d$	∞	$3T_d$	$4.3T_d$
PID	K_P^*	$0.65\frac{1}{k_1T_d}$	$0.65\frac{1}{k_1T_d}$	$1.1\frac{1}{k_1T_d}$	$1.1\frac{1}{k_1T_d}$	$1.36\frac{1}{k_1T_d}$
	T_I^*	∞	$5T_d$	∞	$2T_d$	$1{,}6T_d$
	T_D^*	$0.4T_d$	$0.23T_d$	$0.53T_d$	$0.37T_d$	$0.5T_d$

Procedure:

1. The plant transfer function must be converted on one form (4.42) or (4.43) on the basis of the methods described in subchapter 3.2.
2. On the basis of the control performance requirements the suitable controller, the kind of the control process (without an overshoot, with the relative overshoot $\kappa = 0.2$, minimum of ISE) and the purpose (the tuning from point of view of the desired $w(t)$ or disturbance $v(t)$ variables) are chosen based on Tab. 4.4 for the plant transfer function (4.42) or Tab. 4.5 for the plant transfer function (4.43) the values of the adjustable controller parameters are computed.

Modulus optimum method

The **modulus** optimum **method** belongs among the analytical controller tuning methods. It comes from desired condition for the modulus of the frequency control sytem transfer function [see (3.6)]

$$G_{wy}(s) \to 1 \Rightarrow G_{wy}(\mathrm{j}\omega) \to 1 \Rightarrow A_{wy}(\omega) \to 1 \qquad (4.44)$$

It is supposed that the desired course of the modulus $A_{wy}(\omega)$ would be a monotone decreasing function in accordance with Fig. 4.19.

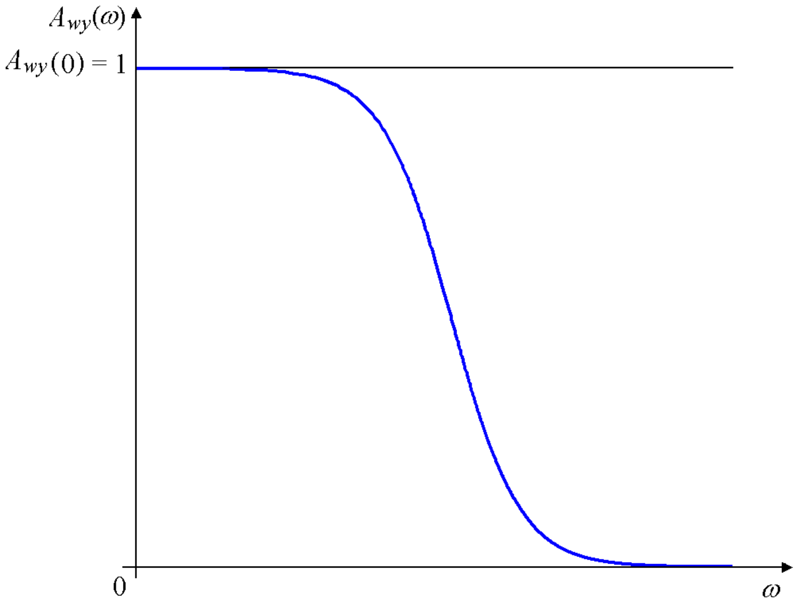


Fig. 4.19 – Desired course of the modulus of a frequency control system transfer function

It is obvious that the relation

$$A_{wy}(\omega) \to 1 \Leftrightarrow A_{wy}^2(\omega) \to 1 \tag{4.45}$$

holds.

It is important because it is easier to operate with the square power and further the equality

$$(\alpha + \mathrm{j}\omega)(\alpha - \mathrm{j}\omega) = \alpha^2 + \omega^2 = |\alpha + \mathrm{j}\omega|^2 \tag{4.46}$$

holds and therefore for the control system transfer function

$$G_{wy}(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \ldots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0}, \qquad n \geq m \tag{4.47}$$

it is possible to write

$$A_{wy}^2(\omega) = G_{wy}(\mathrm{j}\omega) G_{wy}(-\mathrm{j}\omega) = \frac{B_m \omega^{2m} + B_{m-1} \omega^{2(m-1)} + \ldots + B_1 \omega^2 + B_0}{A_n \omega^{2n} + A_{n-1} \omega^{2(n-1)} + \ldots + A_1 \omega^2 + A_0} \tag{4.48}$$

where

$$
\begin{aligned}
&A_0 = a_0^2 && B_0 = b_0^2 \\
&A_1 = a_1^2 - 2a_0a_2 && B_1 = b_1^2 - 2b_0b_2 \\
&A_2 = a_2^2 - 2a_1a_3 + 2a_0a_4 && B_2 = b_2^2 - 2b_1b_3 + 2b_0b_4 \\
&\quad\vdots && \quad\vdots \\
&A_i = a_i^2 + 2\sum_{j=1}^{i}(-1)^j a_{i-j}a_{i+j} && B_i = b_i^2 + 2\sum_{j=1}^{i}(-1)^j b_{i-j}b_{i+j} \\
&\quad\vdots && \quad\vdots \\
&A_{n-1} = a_{n-1}^2 - 2a_{n-2}a_n && B_{m-1} = b_{m-1}^2 - 2b_{m-2}b_m \\
&A_n = a_n^2 && B_m = b_m^2
\end{aligned}
\tag{4.49}
$$

If the equalities

$$\frac{B_0}{A_0} = \frac{B_1}{A_1} = \frac{B_2}{A_2} = \cdots = \frac{B_i}{A_i} = \cdots \tag{4.50}$$

hold and the numerator degree m will be equal to the denominator degree n in the transfer function (4.47) then the square of the modulus $A_{wy}^2(\omega)$ and therefore the modulus $A_{wy}(\omega)$ would be independent from the angular frequency ω. From the point of view of the physical realizability the inequality $n > m$ always holds in technical practice and therefore the independence on the angular frequency ω cannot be reached. The control process will be satisfactory if the square of the modulus $A_{wy}^2(\omega)$ will be a monotone decreasing function with an increasing angular frequency ω, i.e.

$$A_{wy}^2(0) = \frac{B_0}{A_0} \geq \frac{B_i}{A_i} \tag{4.51}$$

When the modulus optimum method is used then the conditions (4.51) are used in the same number as there is the number of adjustable controller parameters p, i.e.

$$A_iB_0 = A_0B_i, \qquad i = 1,2,\ldots,p \tag{4.52}$$

For the control system with $q = 1$ ($b_0 = a_0 \Leftrightarrow B_0 = A_0$) the equalities

$$A_i = B_i, \qquad i = 1,2,\ldots,p \tag{4.53}$$

are used.

Because the conditions (4.52) or (4.53) don't consider all the characteristic polynomial coefficients

$$N(s) = a_ns^n + a_{n-1}s^{n-1} + \ldots + a_1s + a_0 \tag{4.54}$$

arising in the denominator of the control system transfer function (4.47) the modulus optimum method generally doesn't ensure the control system stability and so neither the desired control performance. It means that after using the modulus optimum method the stability must be checked and the control performance would be preferably verified by simulation.

If the plant transfer function $G_P(s)$ has some of the forms given in Tab. 4.6 then for the recommended controllers and given values of the adjustable controller parameters ($T = 0$) the **standard form** of the control system transfer function

$$G_{wy}(s) = \frac{1}{T_w^2 s^2 + 2\xi_w T_w s + 1}, \qquad \xi_w = \frac{1}{\sqrt{2}}, \qquad T_w = \sqrt{2} T_i \tag{4.55}$$

is obtained, where the rows 1 and 2 in Tab. 4.6 $i = 1$, for the rows 3 and 4 $i = 2$ and for the row 5 $i = 3$.

Tab. 4.6 – Values of adjustable controller parameters for the modulus optimum method

Plant		Controller < analog $T = 0$, digital $T > 0$			
		Type	K_P^*	T_I^*	T_D^*
1	$\frac{k_1}{T_1 s + 1}$	I	–	$2k_1(T_1 - 0.5T)$	–
2	$\frac{k_1}{s(T_1 s + 1)}$	P	$\frac{1}{2k_1 T_1}$	–	–
3	$\frac{k_1}{(T_1 s + 1)(T_2 s + 1)}$ $T_1 \geq T_2$	PI	$\frac{T_I^*}{2k_1 T_2}$	$T_1 - 0.5T$	–
4	$\frac{k_1}{s(T_1 s + 1)(T_2 s + 1)}$ $T_1 \geq T_2$	PD	$\frac{1}{2k_1(T_2 + 0.5T)}$	–	$T_1 - 0.5T$
5	$\frac{k_1}{(T_1 s + 1)(T_2 s + 1)(T_3 s + 1)}$ $T_1 \geq T_2 \geq T_3$	PID	$\frac{T_I^*}{2k_1(T_3 + 0.5T)}$	$T_1 + T_2 - T$	$\frac{T_1 T_2}{T_1 + T_2} - \frac{T}{4}$

In this case it isn't necessary to verify control system stability because the form (4.55) is the standard form for the ITAE criterion, see (4.15).

For controller tuning in accordance with Tab. 4.6 the **time constant compensation** was used. It consists in the mutual reduction one of the plant stable binomials by the one binomial of the PI or PD controllers or two of the plant stable binomials by the two binomials of the PID controller. The dynamics of the control system is simplified during the compensation but simultaneously the response slowdown can rise because the stable zeros of the numerator of the control system transfer function $G_{wy}(s)$ can cause the response to accelerate.

Tab. 4.6 can be used as well for the analog controller ($T = 0$) as for the digital controllers ($T > 0$), see Chapter 5.

The modulus optimum method is used for $q \leq 1$ first of all for the control of the electrical drives, where the small time constants (electrical) are substituted by the summary time constant, see subchapter 3.2.

Procedure:

1. The plant transfer function is converted to a suitable form in accordance with Tab. 4.6 and then for the recommended controller the values of its adjustable parameters are computed.
2. If the plant transfer function cannot be converted to some of the forms in Tab. 4.6 or another controller instead of the recommended controller is used then for the determination of the p adjustable parameters of the selected controller are for $q = 0$ computed from the relations (4.52) and for $q = 1$ from the relations (4.53). The time constant compensation can be used as well.
3. In the case of another form than the standard form for the modulus optimum method (4.55) for control system stability it is necessary to verify if the control system is unstable (then the modulus optimum method cannot be used) and the control performance would be preferably verified by simulation.

Desired model method

The **desired model method** is a combined (analytical-experimental) controller tuning method, which comes from the desired model of the closed-loop control system, i.e. from the desired control system transfer function

$$G_{wy}(s) = \frac{Y(s)}{W(s)} = \frac{k_o}{s + k_o \mathrm{e}^{-T_d s}} \mathrm{e}^{-T_d s} \tag{4.56}$$

where k_o is the open-loop gain.

It is very simple tuning method, which makes use of the time constant compensation and it ensures the control system type $q = 1$ (i.e. the zeros of the steady-state errors steps of the desired variable $w(t)$ and the disturbance variable

$v(t)$ in the plant input) and by a suitable choice of the open-loop gain and it makes it possible to ensure the desired relative overshoot κ in the range from 0 to 0.5 (0 to 50 %).

The dependency of the relative overshoot κ for some special values of the open-loop gain k_o is shown in Fig. 4.20.

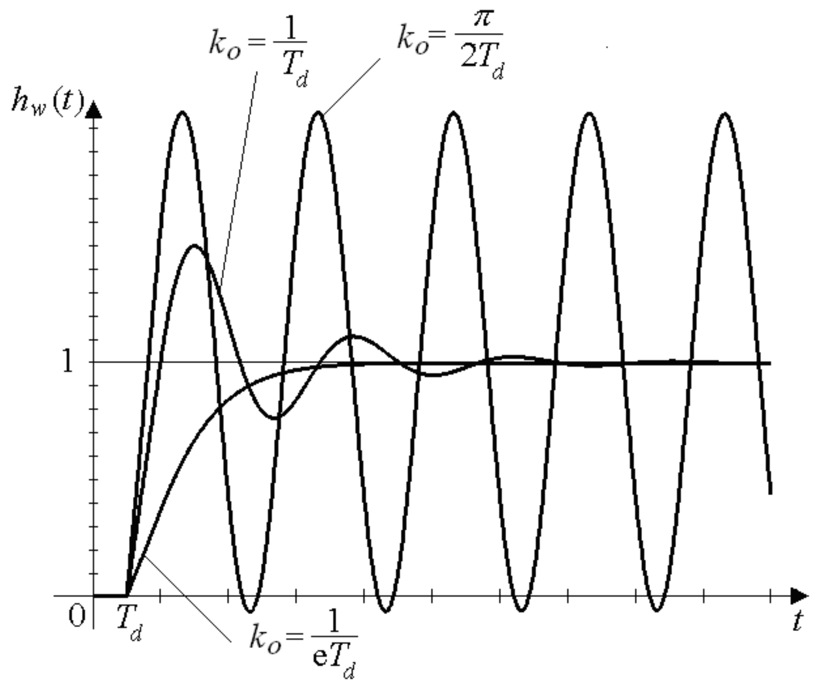


Fig. 4.20 – Influence of open-loop gain k_o on control system step responses

Tab. 4.7 – Dependence of coefficients α and β on relative overshoot κ

κ	0	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
α	1.282	0.984	0.884	0.832	0.763	0.697	0.669	0.640	0.618	0.599	0.577
β	2.718	1.944	1.720	1.561	1.437	1.337	1.248	1.172	1.104	1.045	0.992

The open-loop gain k_o can be obtained analytically for the non-oscillating control process ($k_o = \frac{1}{eT_d}$) and for the oscillating stability boundary ($k_o = \frac{\pi}{2T_d}$). For the other values of the relative overshoot κ the dependency of the open-loop gain k_o on the time delay T_d was determined by the simulation (see Tab. 4.7)

$$k_o = \frac{1}{\beta T_d} \qquad (4.57)$$

The suitable plant transfer functions for the desired model method are given in Tab. 4.8 together with the recommended controllers and values of their adjustable parameters.

The transfer function of the recommended controller $G_C(s)$ for some of the plants with the transfer function $G_P(s)$ for the desired control transfer function (4.56) can be obtained from the formula for direct synthesis

$$G_{wy}(s) = \frac{G_C(s)G_P(s)}{1+G_C(s)G_P(s)} \Rightarrow G_C(s) = \frac{1}{G_P(s)}\frac{G_{wy}(s)}{1-G_{wy}(s)} \tag{4.58}$$

Tab. 4.8 – Values of adjustable controller parameters for the desired model method

Plant		Controller $<$ analog $T=0$, digital $T>0$			
		Type	K_P^*	T_I^*	T_D^*
1	$\frac{k_1}{s}\mathrm{e}^{-T_d s}$	P	$\frac{1}{(\alpha T + \beta T_d)k_1}$	–	–
2	$\frac{k_1}{T_1 s + 1}\mathrm{e}^{-T_d s}$	PI	$\frac{T_I^*}{(\alpha T + \beta T_d)k_1}$	$T_1 - \frac{T}{2}$	–
3	$\frac{k_1}{s(T_1 s + 1)}\mathrm{e}^{-T_d s}$	PD	$\frac{1}{(\alpha T + \beta T_d)k_1}$	–	$T_1 - \frac{T}{2}$
4	$\frac{k_1}{(T_1 s + 1)(T_2 s + 1)}\mathrm{e}^{-T_d s}$, $T_1 \geq T_2$	PID	$\frac{T_I^*}{(\alpha T + \beta T_d)k_1}$	$T_1 + T_2 - T$	$\frac{T_1 T_2}{T_1 + T_2} - \frac{T}{4}$
5	$\frac{k_1}{T_0^2 s^2 + 2\xi_0 T_0 s + 1}\mathrm{e}^{-T_d s}$, $0.5 < \xi_0 \leq 1$	PID	$\frac{T_I^*}{(\alpha T + \beta T_d)k_1}$	$2\xi_0 T_0 - T$	$\frac{T_0}{2\xi_0} - \frac{T}{4}$

E.g. for the plant with the transfer function

$$G_P(s) = \frac{k_1}{T_1 s + 1}\mathrm{e}^{-T_d s}$$

after substitution in (4.58) and considering (4.56) the controller transfer function

$$G_C(s) = \frac{T_1 s + 1}{k_1\,\mathrm{e}^{-T_d s}} \frac{\dfrac{k_o}{s + k_o\,\mathrm{e}^{-T_d s}}\mathrm{e}^{-T_d s}}{1 - \dfrac{k_o}{s + k_o\,\mathrm{e}^{-T_d s}}\mathrm{e}^{-T_d s}} = \frac{k_o(T_1 s + 1)}{k_1 s} = k_P^*\left(1 + \frac{1}{T_I^* s}\right)$$

is obtained (see the row 2 in Tab. 4.8 for $T = 0$), where

$$K_P^* = \frac{k_o T_I^*}{k_1}, \quad T_I^* = T_1$$

or after considering (4.57)

$$K_P^* = \frac{T_I^*}{k_1 \beta T_d}, \; T_I^* = T_1$$

In a similar way for $T = 0$ the remaining rows were obtained in Tab. 4.8. Tabs 4.7 and 4.8 can be used for $T > 0$ also for the digital controllers, see Chapter 5.

For a control system tuned by the desired model method the values of the most important control performance indices were computed, see Tab. 4.9.

Tab. 4.9 – Values of the most important control performance indices

κ	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
$\omega_R T_d$	0	0.3	0.7	0.8	0.95	1.0	1.1	1.2	1.2	1.3	1.3
$A_{wy}(\omega_R)$	1	1.00	1.06	1.14	1.25	1.37	1.51	1.68	1.88	2.10	2.37
$L_{wy}(\omega_R)$ [dB]	0	0.02	0.47	1.15	1.92	2.72	3.59	4.50	5.47	6.46	7.51
M_S	1.4	1.6	1.7	1.9	2.0	2.1	2.3	2.5	2.67	2.9	3.2
γ [°]	69	60	57	53	50	47	44	41	38	35	32
m_A	4.3	3.0	2.7	2.5	2.3	2.1	2.0	1.8	1.7	1.6	1.6
m_L [dB]	12.6	9.7	8.6	7.8	7.1	6.4	5.85	5.3	4.8	4.3	3.9

From Tab. 4.9 it follows that for control systems with the analog controllers tuned by the desired model method for the relative overshoot $\kappa \le 0.2$ (20 %) the values of all the most important control performance indices satisfy the recommendations for well-tuned control systems. Therefore after using the desired model method for $\kappa \le 0.2$ it can be expected that besides the desired control performance the high control system robustness will hold.

From Tab. 4.9 the conclusion follows that because the product of the resonant angular frequency ω_R and the time delay T_d is for the given relative overshoot κ constant, it is obvious that the time delay T_d strongly restricts the range of the operating angular frequencies.

Procedure:

1. The plant transfer function is converted to a suitable form in accordance with Tab. 4.8.
2. For the desired relative overshoot κ from Tab. 4.7 the coefficient β is chosen and on the basis of Tab. 4.8 for the recommended controller and for $T = 0$ the values of its adjustable parameters are computed.

SIMC method

The **SIMC method** comes from the internal model control (IMC). Its author, Skogestad, recommends the abbreviation SIM to be understood as „SIMple Control“ or „Skogestad IMC“.

For the determination of the controller transfer function the formula for direct synthesis [see (4.58)]

$$G_C(s) = \frac{1}{G_P(s)} \frac{G_{wy}(s)}{1 - G_{wy}(s)} \tag{4.59}$$

is used on the assumption that the control system transfer function has the form

$$G_{wy}(s) = \frac{1}{T_w s + 1} \mathrm{e}^{-T_d s} \tag{4.60}$$

where T_w is the time constant of the closed-loop control system.

E.g. for the plant with the transfer function

$$G_P(s) = \frac{k_1}{T_1 s + 1} \mathrm{e}^{-T_d s}$$

it is obtained

$$G_C(s) = \frac{T_1 s + 1}{k_1} \frac{1}{T_w s + 1 - \mathrm{e}^{-T_d s}} \tag{4.61}$$

After use of the approximation

$$\mathrm{e}^{-T_d s} \approx 1 - T_d s$$

from relation (4.61) the transfer function of the PI controller

$$G_C(s) = K_P\left(1 + \frac{1}{T_I s}\right), \quad K_P = \frac{T_1}{k_1(T_w + T_d)}, \quad T_I = T_1$$

is obtained.

By a suitable choice of the time constant T_w the different fast responses can be obtained. The time constant T_w can be considered as the tuning parameter. There is most often recommended $T_w = T_d$ and the integral time T_I is determined on the basis of the relation

$$T_I = \min(T_1, 8T_d)$$

Then the values of the adjustable parameters of the PI controller are given (see rows 2 and 3 in Tab. 4.10)

$$K_P^* = \frac{T_1}{2k_1 T_d}, \quad \begin{array}{lll} T_I^* = T_1 & \text{for} & T_1 \le 8T_d \\ T_I^* = 8T_d & \text{for} & T_1 > 8T_d \end{array}$$

In a similar way the remaining rows in Tab. 4.10 were obtained.

The cases in the rows 2, 4 and 6 in Tab. 4.10 are equivalent to the desired model method for the relative overshoot $\kappa \approx 0.05$ (5 %).

Tab. 4.10 – Values of adjustable controller parameters for the SIMC method

Plant		Controller				Note
		Type	K_P^*	T_I^*	T_D^*	
1	$k_1 \mathrm{e}^{-T_d s}$	I	–	$2k_1 T_d$	–	–
2	$\frac{k_1}{T_1 s + 1} \mathrm{e}^{-T_d s}$	PI	$\frac{T_1}{2k_1 T_d}$	T_1	–	$T_1 \le 8T_d$
3			$\frac{T_1}{2k_1 T_d}$	$8T_d$	–	$T_1 > 8T_d$
4	$\frac{k_1}{(T_1 s + 1)(T_2 s + 1)} \mathrm{e}^{-T_d s}$ $T_1 \ge T_2$	PID$_\mathrm{i}$	$\frac{T_1}{2k_1 T_d}$	T_1	T_2	$T_1 \le 8T_d$
5			$\frac{T_1}{2k_1 T_d}$	$8T_d$	T_2	$T_1 > 8T_d$
6		PID	$\frac{T_1 + T_2}{2k_1 T_d}$	$T_1 + T_2$	$\frac{T_1 T_2}{T_1 + T_2}$	$T_1 \le 8T_d$
7			$\frac{T_1(T_2 + 8T_d)}{16k_1 T_d^2}$	$T_2 + 8T_d$	$\frac{8T_2 T_d}{T_2 + 8T_d}$	$T_1 > 8T_d$
8	$\frac{k_1}{s} \mathrm{e}^{-T_d s}$	PI	$\frac{1}{2k_1 T_d}$	$8T_d$	–	–
9	$\frac{k_1}{s(T_2 s + 1)} \mathrm{e}^{-T_d s}$	PID$_\mathrm{i}$	$\frac{1}{2k_1 T_d}$	$8T_d$	T_2	–
10		PID	$\frac{T_2 + 8T_d}{16k_1 T_d^2}$	$T_2 + 8T_d$	$\frac{8T_2 T_d}{T_2 + 8T_d}$	–
11	$\frac{k_1}{s^2} \mathrm{e}^{-T_d s}$	PID$_\mathrm{i}$	$\frac{1}{16k_1 T_d^2}$	$8T_d$	$8T_d$	–
12		PID	$\frac{1}{8k_1 T_d^2}$	$16T_d$	$4T_d$	–

Procedure:

1. The plant transfer function is converted to a suitable form in accordance with Tab. 4.10.
2. For the recommended controller on the basis of Tab. 4.10 the values of its adjustable parameters are computed.

5 DIGITAL CONTROL

This chapter is devoted to a brief description of the control systems with digital controllers. A simple approximate design method for digital controllers is shown.

Lately **digital controllers** have most frequently been used in control engineering. It is caused by the recent development of digital technologies and simultaneously the decreasing of their prices. Conventional digital controllers mostly implement the same control algorithms, like analog ones but in discrete forms. Further in the text it is supposed that the **quantization error is negligibly small** and therefore the concept "**digital**" (discrete in magnitude and time) and "**discrete**" (discrete in time but continuous in magnitude) are equivalent. For example, the digital PID controller

$$\begin{aligned} u(kT) &= K_P\left\{e(kT)+\frac{T}{T_I}\sum_{i=0}^{k}e(iT)+\frac{T_D}{T}\{e(kT)-e[(k-1)T]\}\right\} = \\ &= \underbrace{K_P e(kT)}_{P}+\underbrace{K_I\sum_{i=0}^{k}e(iT)}_{I}+\underbrace{K_D\{e(kT)-e[(k-1)T]\}}_{D} \end{aligned} \tag{5.1}$$

$k = 0,1,2,\ldots$

corresponds to the analog PID controller (3.19), where K_P, K_I and K_D are the **proportional, summation** and **difference component weights**, T – the **sampling period**, kT – the **discrete time**.

From the adjustable digital PID controller parameters it holds that

$$K_I = K_P\frac{T}{T_I}, \qquad K_D = K_P\frac{T_D}{T} \tag{5.2}$$

or

$$T_I = \frac{K_P}{K_I}T, \quad T_D = \frac{K_D}{K_P}T \tag{5.3}$$

It is obvious that for the digital controllers the further adjustable parameter arises – the sampling period T. Its proper choice is very important from the point of view of control performance. The sampling period T increases the influence of the summation component (the summation component always destabilizes the control process) and decreases the influence of the difference component (the difference component stabilizes the control process), therefore the **sampling period´s influence on the control performance and stability is always negative**. Also, from this follows that between the sampling instants $kT < t < (k + 1)T$ the digital controller hasn't any information about the current value of the control error $e(t)$, see Fig. 5.1 and therefore it cannot perform and control well.

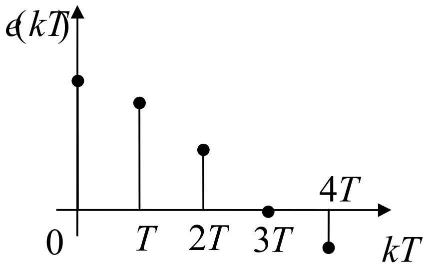


Fig. 5.1 – Control error course in a control system with a digital controller

The **analog-to-digital (A/D) converter** processes the conversion from the analog (continuous) variable to the digital (discrete) variable. It is often plugged in the feedback (Fig. 5.2). The output variable of the digital controller (DC) is the discrete control variable $u(kT)$, which the **digital-to-analog (D/A) converter** converts to the continuous in the time control variable $u_T(t)$ with a staircase course (Fig. 5.3), which is the input variable of the plant (P).

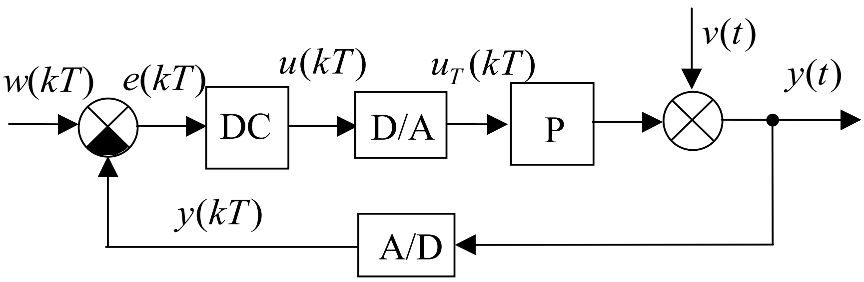


Fig. 5.2 – Control system with a digital controller

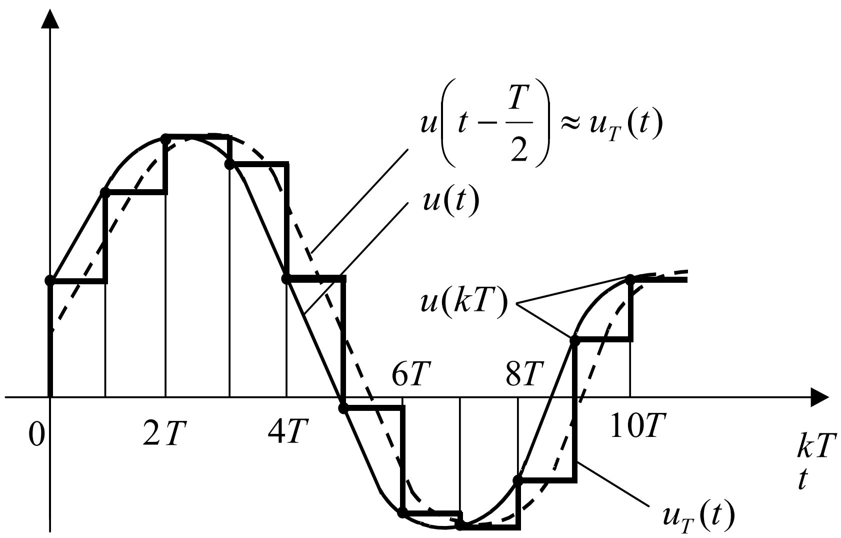


Fig. 5.3 – Control variable courses in a control system with a digital controller

From Fig. 5.3 it follows that the staircase control variable $u_T(t)$ for the small sampling period T value can be substituted by smooth control variable $u(t)$, which is delayed by half the sampling period, i.e. $u(t - T/2)$. It is obvious that this substitution will be better for the smaller sampling period. Therefore for the approximate analysis and synthesis of the control system with the digital controller the substitute control system in Fig. 5.4 can be used. The digital controller is substituted by the analog controller of the corresponding type and the time delay is assigned to the plant. If methods not suitable for the time delay are used for analysis or synthesis then the time delay must be approximated by one of the following relations

$$\mathrm{e}^{-\frac{T}{2}s} \approx \frac{1-\frac{T}{4}s}{1+\frac{T}{4}s} \tag{5.4}$$

or

$$\mathrm{e}^{-\frac{T}{2}s} \approx \frac{1}{1+\frac{T}{2}s} \tag{5.5}$$

The more accurate approximation is not used. The obtained results must be carefully interpreted with a sense of the approximate approach.

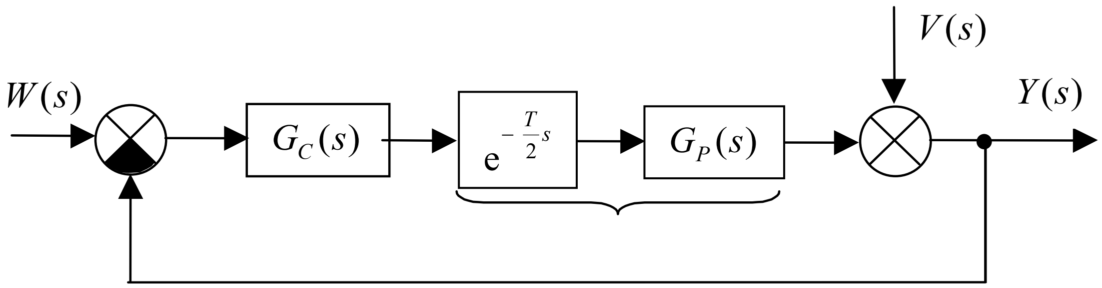


Fig. 5.4 – Substitute control system with the digital controller

The digital PID controller is the most complex conventional controller. In technical practice simpler digital controllers are used:

the digital **PI** controller

$$u(kT) = K_P\left[e(kT) + \frac{T}{T_I}\sum_{i=0}^{k} e(iT)\right] \tag{5.6}$$

the digital **PD** controller

$$u(kT) = K_P\left\{e(kT) + \frac{T_D}{T}\{e(kT) - e[(k-1)T]\}\right\} \tag{5.7}$$

the digital **I** controller

$$u(kT) = \frac{T}{T_I} \sum_{i=0}^{k} e(iT) \tag{5.8}$$

and the digital **P** controller

$$u(kT) = K_P e(kT) \tag{5.9}$$

The summation and difference components (terms) are often implemented using other different methods (the forward rectangular method, trapezoidal method etc.).

For the suitable choice of the sampling period T these distinctions aren't substantial and in addition the manufacturers very often don't give any information about the summation and difference component implementation.

For the digital difference component the input variable must be always suitably filtered.

For choosing the sampling period T definite rules and recommendations don´t exist. For a rough choice the following recommendations can be used.

Sampling period *T*	**Plant (Process)**
(10 ÷ 500) μs	the accurate control, the electrical and power systems, the accurate control robots
(0.5 ÷ 20) ms	the stabilization of the power systems, the flight and drive simulators
(10 ÷ 100) ms	image processing, virtual reality, artificial vision
(0.5 ÷ 1) s	the control and monitoring of the processes, the chemical processes, the power systems
(1 ÷ 3) s	flow control
(1 ÷ 5) s	pressure control
(5 ÷ 10) s	level control
(10 ÷ 20) s	temperature control

The more accurate determination of the sampling period comes from the behavior of the plant or a closed-loop control system. For example, for the proportional non-oscillating plant it is recommended that

$$T = \left(\frac{1}{15} \div \frac{1}{6} \right) t_{0.95} \tag{5.10}$$

where $t_{0.95}$ is the time when the step response reaches 95 % of the steady-state value.

For the plant with the dominant time delay T_d the relation

$$T = \left(\frac{1}{8} \div \frac{1}{3}\right) T_d \tag{5.11}$$

is recommended.

For digital controllers with the difference component the sampling period T must be chosen in accordance with the relation

$$T = (0.1 \div 0.5) T_D \tag{5.12}$$

Some controller tuning methods are processed and derived also for the digital controllers (see Tab. 4.6 ÷ 4.8) and therefore they can be used directly.

6 TWO- AND THREE-POSITION CONTROL

The chapter is devoted to the two- and tree-position control, which belongs among the simplest of control technologies.

The **two- and three-position (relay) control** is widely and commonly used in home equipment and devices. Especially in every house, the two-position (ON-OFF) control is used, e.g. for the electric iron temperature (see Fig. 1.3), water temperature and the level in the washing machine, the room temperature etc.

The main reason of the use of the two- and three-position control is its very low price and relatively high reliability.

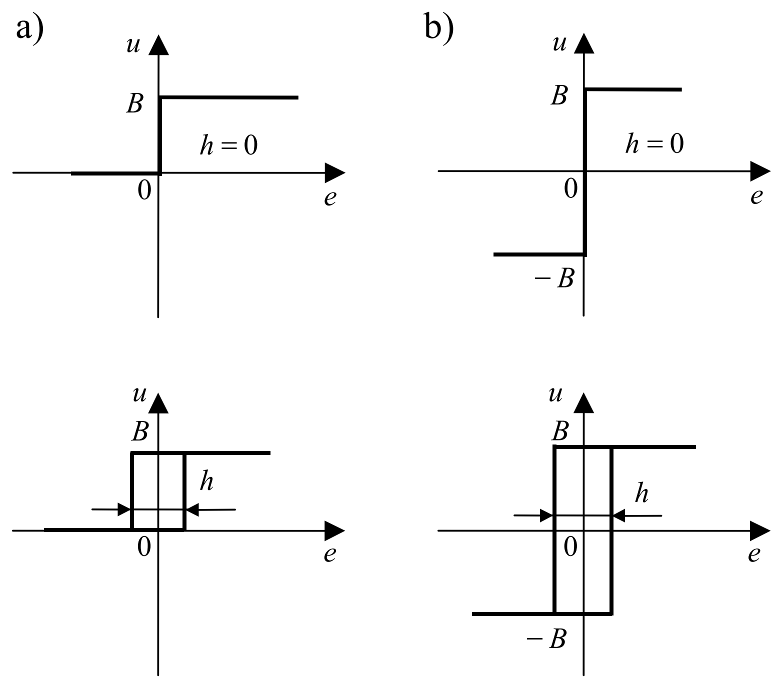


Fig. 6.1 – Different characteristics of a two-position controller: a) asymmetric without hysteresis (h = 0) and with hysteresis (h > 0), b) symmetric without hysteresis (h = 0) and with hysteresis (h > 0)

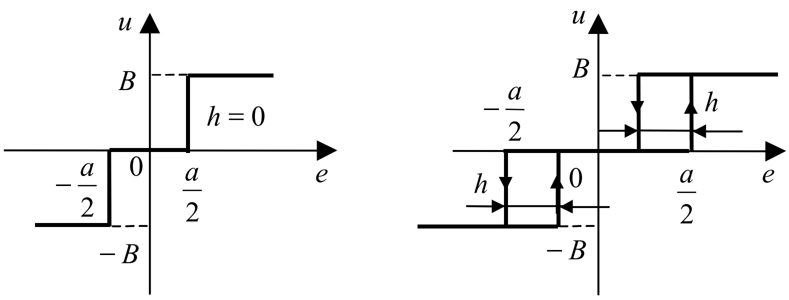


Fig. 6.2 – Characteristic of a symmetric three-position controller without hysteresis (h = 0) and with hysteresis (h > 0)

The two- and three-position controllers are strongly non-linear. Their characteristics are relay characteristics shown in Figs 6.1 and 6.2, where B is the relay amplitude, h – the hysteresis width, a – the dead zone. If the controller characteristic is without hysteresis (i.e. without memory) then it is the controller static characteristic. In the case of the controller characteristic with the hysteresis (i.e. with memory) this characteristic isn't in an exact sense "static" and therefore it is just called the "characteristic".

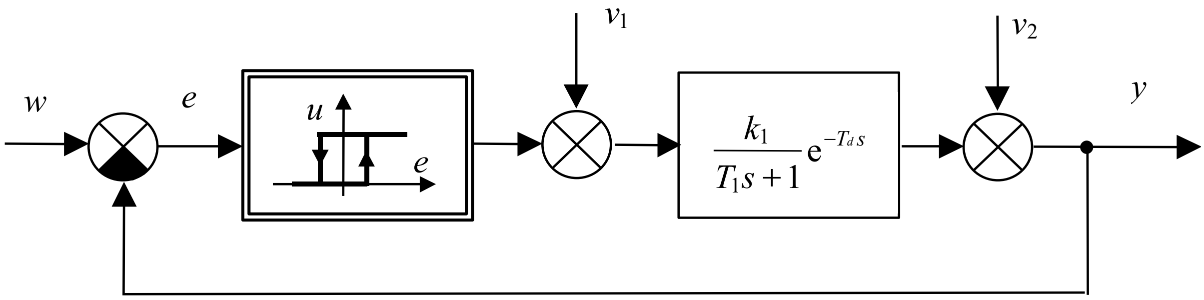


Fig. 6.3 – Control system with ON-OFF controller

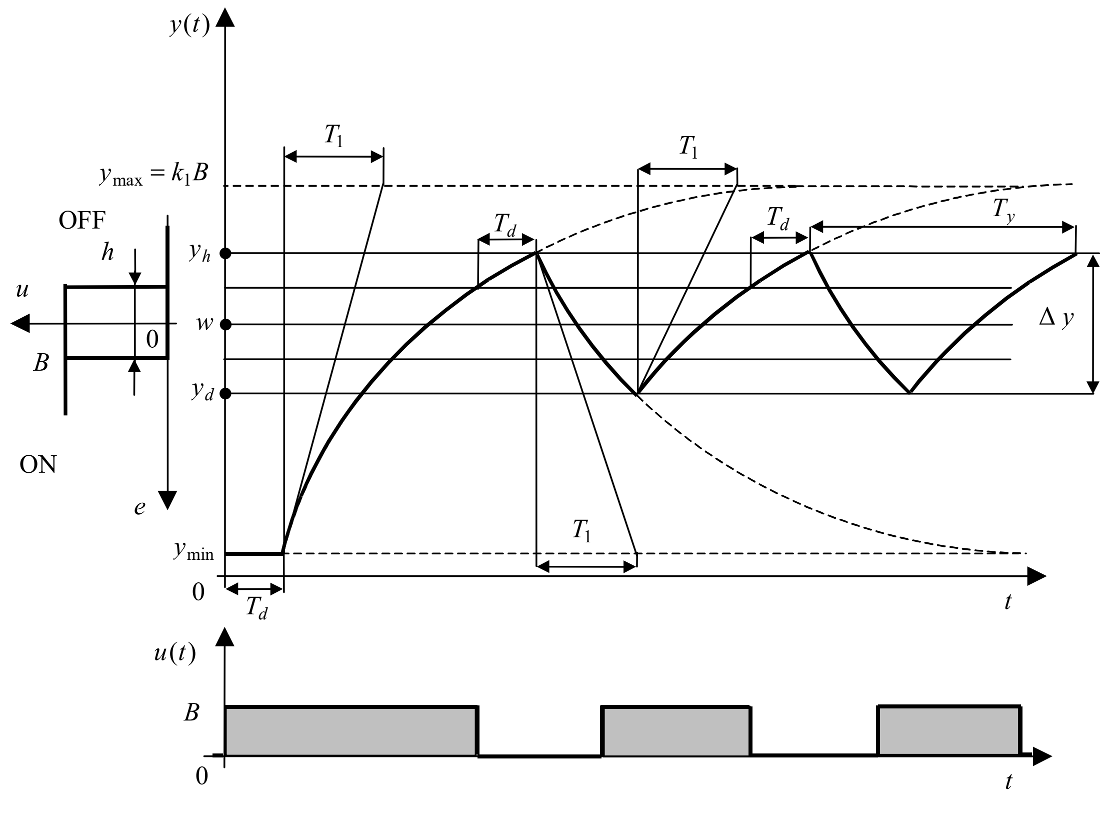


Fig. 6.4 – Courses of controlled $y(t)$ and control $u(t)$ variables in control system with an ON-OFF controller

Two-position controllers with the characteristic as in Fig. 6.1a very often operate in the mode "switch-on" and "switch-off" (e.g. the heating is on and the heating is off) and as Fig. 6.1b they operate in the mode "switch-on plus" and "switch-on minus" (e.g. the heating is on and the cooling is on). The three-position controllers in Fig. 6.2 are the two-position controller (in Fig. 6.1b) with extension of the third position "switch-off". They often operate in mode "switch-on plus", "switch-off" and "switch-on minus" (e.g. the heating is on, the heating and cooling are off and the cooling is on). The typical control system with the ON-OFF controller is in Fig. 6.3. Since both the originals of the variables and their transforms stand out the variables are written without their arguments and in lower case letters. The operation of the control system in Fig. 6.3 is following. It is supposed that at the beginning the controlled variable value is $y(0) = y_{\min}$. Because $e(0) > h/2$ the control variable $u(t) = B$ (the state: switch – ON) and therefore the initial course of the controlled variable $y(t)$ is given by the relation (Fig. 6.4)

$$y(t) = y_{\min} + (y_{\max} - y_{\min})\left(1 - \mathrm{e}^{-\frac{t-T_d}{T_1}}\right)\eta(t - T_d), \qquad t \geq 0 \tag{6.1}$$

After reaching the value $y(t) = w + \frac{h}{2}$ the control variable $u(t) = 0$ (the state: switch – OFF), the controlled variable $y(t)$ at first rises during the time delay T_d and then it falls until it reaches the value $y(t) = w - \frac{h}{2}$, the control variable $u(t) = B$ (the state: switch – ON), it further falls during the time delay T_d and then it rises etc. The whole control process periodically repeats. Because the control system with the ON-OFF controller is strongly non-linear therefore the analytical description of the course of the controlled variable $y(t)$ is relatively complicated. While its graphical construction is very easy and it follows directly from Fig. 6.4.

For the well-designed control system with the ON-OFF controller the desired variable (set-point) value approximately holds

$$w \approx \frac{y_{\max} - y_{\min}}{2} \tag{6.2}$$

If it is equal then the 100 % abundance of the actuator power is given and the average controlled variable value is $y_{av} = w$. For the higher power abundance the inequality $y_{av} > w$ holds and for the smaller one the opposite inequality $y_{av} < w$ holds. In both the last cases the courses of the controlled variable $y(t)$ are asymmetric.

If the disturbance variables $v_1(t)$ and $v_2(t)$ influence the control system in Fig. 6.3 they cause the controlled variable $y(t)$ to fall under value $w - \frac{h}{2}$, the control variable $u(t) = B$ (the state: switch – ON), the controlled variable $y(t)$

after the time delay T_d begins to rise and again the periodical control process arises.

It is obvious that if the control error $e(t)$ arises [it doesn't matter whether it was caused by the desired $w(t)$ or disturbance $v_1(t)$ and $v_2(t)$ variables or by the plant behavior change] then the ON-OFF controller makes efforts to remove it by the maximum value of the control variable, i.e. $u_{max} = B$ or $u_{min} = 0$. Therefore **if the ON-OFF control is applicable then it is highly robust**.

The applicability of the ON-OFF control decides the obtained control performance. It is given by the **oscillation band width** Δy of the controlled variable, which can be determined on the basis of the relations (see Fig. 6.4)

$$y_h = w + \frac{h}{2} + \left(y_{\max} - w - \frac{h}{2} \right)\left(1 - \mathrm{e}^{-\frac{T_d}{T_1}} \right) = y_{\max} - \left(y_{\max} - w - \frac{h}{2} \right)\mathrm{e}^{-\frac{T_d}{T_1}}$$

$$y_d = w - \frac{h}{2} - \left(w - \frac{h}{2} - y_{\min} \right)\left(1 - \mathrm{e}^{-\frac{T_d}{T_1}} \right) = y_{\min} + \left(w - \frac{h}{2} - y_{\min} \right)\mathrm{e}^{-\frac{T_d}{T_1}}$$

$$\Delta y = y_h - y_d = (y_{\max} - y_{\min})\left(1 - \mathrm{e}^{-\frac{T_d}{T_1}} \right) + h\,\mathrm{e}^{-\frac{T_d}{T_1}} \tag{6.3}$$

After using of the approximation

$$\mathrm{e}^{-\frac{T_d}{T_1}} \approx 1 - \frac{T_d}{T_1} \qquad \text{and} \qquad -h\frac{T_d}{T_1} \approx 0$$

the last relation in (6.3) can be simplified

$$\Delta y \approx (y_{\max} - y_{\min})\frac{T_d}{T_1} + h \tag{6.4}$$

From the approximate formula (6.4) it is obvious that both the hysteresis width h and the time delay T_d have a negative influence on the oscillation band width Δy. The time delay T_d can be sometimes decreased by the suitably placed sensor but it is mostly given by the plant behavior and therefore it cannot be decreased.

The time delay T_d is the greatest enemy of the ON-OFF control (anywhere in the control) and therefore it demands

$$\frac{T_u}{T_n} < 0.2 \tag{6.5}$$

would hold (in order to fulfil the condition $T_u = T_d = T_{d1}$, $T_n = T_1$, see Fig. 3.6a).

Therefore if the desired control performance isn't reached for $h = 0$, then the ON-OFF controller cannot be used.

From a practical point of view the **oscillating period** T_y is very important because its inverse value

$$f_y = \frac{1}{T_y} \tag{6.6}$$

expresses the **switching frequency** (i.e. number of switch-on or switch-off) per time unit. The switching frequency f_y has a direct influence on the lifetime of the controller or actuator. From the Fig. 6.4, it follows that the oscillating period T_y will be greater if the time delay T_d and the hysteresis width h will be greater. It is obvious that these requirements on the minimal oscillation band width Δy and the maximal oscillation period T_y are contradictory to each other and therefore it is necessary to choose a compromise solution.

For the electronic two-position controller the oscillation period T_y can be increased by the adjustable dwell time.

It is obvious that all considerations can be applied to a two-position symmetric controller (Fig. 6.1b) for $y_{\min} = -k_1B$.

The two-position symmetric controller (Fig. 6.1b) is sometimes used together with the integrating device (most frequently with the electric drive). Its disadvantage is the continuous switching, therefore the use of the three-position controller (Fig. 6.2) is more suitable in accordance with Fig. 6.5. This connection is often used for the actuator (valve) setting.

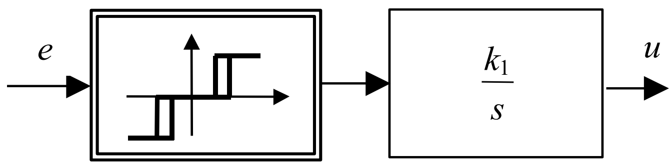


Fig. 6.5 – Three-position controller with integrating device

The great oscillating band width Δy for the two- and three-position controllers can be decreased by the dynamic feedback, see Fig. 6.6. For both interconnections in Fig. 6.6 the two- or three-position controller can be approximately substituted by the gain $k_n \to \infty$ and then holds

$$G_C(s) = \frac{U(s)}{E(s)} \approx \frac{k_n}{1 + k_n G_{FB}(s)} = \frac{1}{\frac{1}{k_n} + G_{FB}(s)}$$

$$k_n \to \infty \Rightarrow G_C(s) \approx \frac{1}{G_{FB}(s)} \tag{6.7}$$

where $G_{FB}(s)$ is the feedback transfer function.

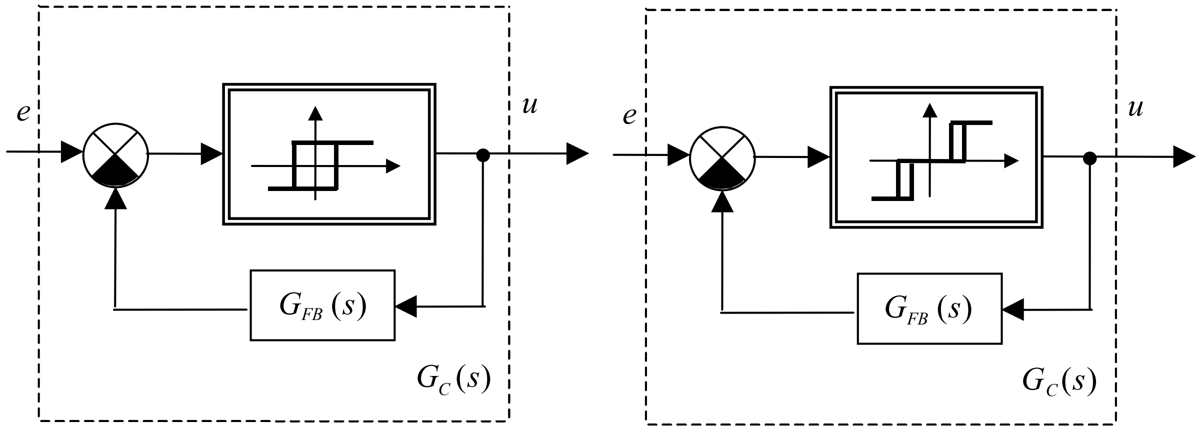


Fig. 6.6 – Two- and three-position controller with dynamic feedback

It is obvious that the two- or three-position controller with a dynamic feedback approximately implements the inversion of the feedback, i.e. (6.7).

For example, for

$$G_{FB}(s) = \frac{k_{FB}}{T_{FB}s+1}$$

the approximate PD controller

$$G_C(s) = \frac{U(s)}{E(s)} \approx \frac{1}{G_{FB}(s)} = k_P(1+T_D s)$$
$$k_P \approx \frac{1}{k_{FB}}, \quad T_D \approx T_{FB} \tag{6.8}$$

can be obtained.

Similarly for

$$G_{FB}(s) = \frac{k_{FB}s}{T_{FB}s+1}$$

the approximate PI controller

$$G_C(s) = \frac{U(s)}{E(s)} \approx k_P(1+\frac{1}{T_I s})$$
$$k_P \approx \frac{T_{FB}}{k_{FB}}, \quad T_I \approx T_{FB} \tag{6.9}$$

is obtained and for

$$G_{FB}(s) = \frac{k_{FB}s}{(T_{FB1}s+1)(T_{FB2}s+1)}, \quad T_{FB1} > T_{FB2}$$

the approximate PID$_i$ (with the interaction) controller is implemented [see (3.24)]

$$G_C(s) = \frac{U(s)}{E(s)} \approx k'_P(1 + \frac{1}{T'_I s})(1 + T'_D s)$$
$$k'_P \approx \frac{T_{FB1}}{k_{FB}}, \qquad T'_I \approx T_{FB1}, \qquad T'_D \approx T_{FB2} \tag{6.10}$$

The **PI step controller** is obtained for interconnection in accordance with Fig. 6.7. Its transfer function is approximately given

$$G_C(s) = \frac{U(s)}{E(s)} \approx k_P(1 + \frac{1}{T_I s})$$
$$k_P \approx \frac{k_1}{k_{FB}} T_{FB}, \qquad T_I \approx T_{FB} \tag{6.11}$$

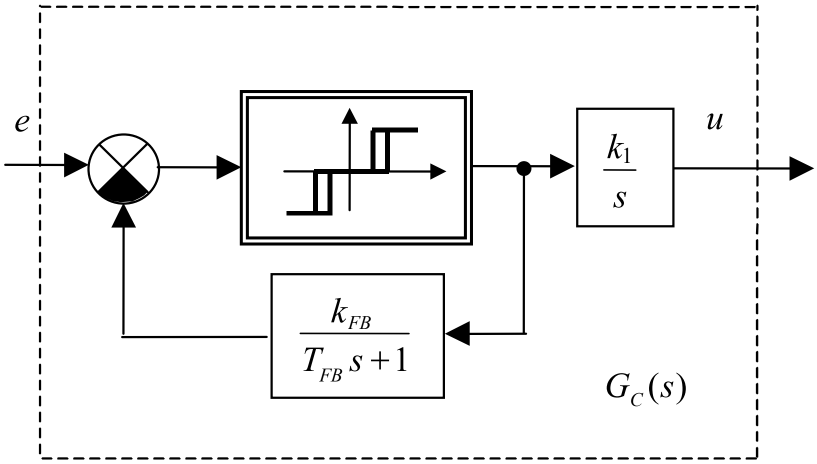


Fig. 6.7 – PI step controller

7 CONCLUSION

After reading this book every control engineering student is now able to understand what the control objective is, why negative feedback is important, when open loop control can be used, and the four main principles for every general system: the analysis, synthesis, identification and control.

In order to be able to see the behavior of systems when they respond to signals on their inputs, the tools for modeling them and methods for visualizing the output results are presented.

The analysis part starts with the role of controllers and their influence on the stability of systems as well as methods on how to check them for that. The synthesis part continues with a detailed look into controller tuning methods and procedures, both for analog and digital control, so our reader can decide what is more suitable, when, and under which conditions. A special chapter is devoted to two- and three-position (relay) control, since it is widely and commonly used in home equipment and devices.

For deeper study and a wider view, it is possible to use the recommended references.

8 REFERENCES

ÅSTRÖM, K. – HÄGGLUND, T. *Advanced PID Control.* Instrument Society of America, Research Triangle Park, 2006, 460 p.

CHEN, CH.T. Analog and Digital Control System Design: Transfer-Function, State-Space, and Algebraic Methods. Oxford University Press, New York – Oxford 1993, 600 p.

DORF, R.C. – BISHOP, R. *Modern Control Systems* (*12th ed.*). Prentice-Hall, Upper Saddle River, New Jersey 2011, 1082 p.

FRANKLIN, G.F. – POWELL, J.D. – EMAMI-NAEINI, A. *Feedback Control of Dynamic Systems* (*4th ed.*). Prentice-Hall, Upper Saddle River, New Jersey 2002, 910 p.

GÓRECKI, H., FUKSA, S., GRABOWSKI, P., KORYTOWSKI, A. *Analysis and Synthesis of Time Delay Systems*. PWN-Polish Scientific Publishers – John Wiley&Sons, Warszawa – Chichester, 1989, 369 p.

KOWAL, J. *Podstawy automatyki* (Tom I). Uczelniane wydawnictwa naukovo-dydaktyczne AGH, Kraków, 2006, 301 str.

LANDAU, I.D. – ZITO, G. Digital Control Systems. Design, Identification and Implementation. Springer-Verlag, London 2006, 484 p.

LEVINE, W.S. (Editor) *The Control Handbook.* CRC Press, Boca Raton, Florida 1996, 1548 p.

MIKLEŠ, J., FIKAR, M. *Process Modelling, Identification, and Control*. Springer-Verlag, Berlin, 2007, 480 p.

NISE, N.S. *Control Systems Engineering* (*2nd ed.*). The Benjamin/Cummings Publishing Company, Redwood City 1995, 851 p.

O'DWYER, A. *Handbook of PI and PID Controllers Tuning Rules* (*3rd ed.*). Imperial College Press, World Scientific, London 2009, 608 p.

OGUNNAIKE, B.A. – RAY, W.H. *Process Dynamics, Modeling and Control.* Oxford University Press, New York – Oxford 1994, 1260 p.

SKOGESTAD, S. *Probably the Best Simple PID Tuning Rules in the World.* Paper No. 276h presented at AIChE Annual Meeting, pp. 1-28, Reno, USA, November 19, 2001

SKOGESTAD, S. Simple Analytic Rules for Model Reduction and PID Controller Tuning. *Modeling, Identification and Control,* Vol. 25, No. 2, 2004, pp. 52-120

VÍTEČKOVÁ, M., VÍTEČEK, A. *Základy automatické regulace*. 2. Přepracované vydání. FS VŠB-TUO, Ostrava, 2008, 244 str.

ZÍTEK, P. Time Delay Control System Design Using Functional State Models. CTU Publishing House, Prague, 1998, 93 p.

1 LAPLACE TRANSFORM - BASIC RELATIONS AND PROPERTIES

	Definition formulas
1	$X(s)=\mathrm{L}\{x(t)\}=\int_0^\infty x(t)\mathrm{e}^{-st}\,\mathrm{d}t$
2	$x(t)=\mathrm{L}^{-1}\{X(s)\}=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} X(s)\mathrm{e}^{st}\,\mathrm{d}s$
	Linearity
3	$\mathrm{L}\{a_1x_1(t)\pm a_2x_2(t)\}=a_1X_1(s)\pm a_2X_2(s)$
	Similarity theorem
4	$\mathrm{L}\{ax(at)\}=X\left(\frac{s}{a}\right),\ a>0$
	Convolution in time domain
5	$\mathrm{L}\left\{\int_0^t x_1(t-\tau)x_2(\tau)d\tau\right\}=\mathrm{L}\left\{\int_0^t x_2(t-\tau)x_1(\tau)d\tau\right\}=X_1(s)X_2(s)=X_2(s)X_1(s)$
	Real shifting in time domain (on the right)
6	$\mathrm{L}\{x(t-a)\}=\mathrm{e}^{-as}\,X(s),\ a\ge 0$
	Real shifting in time domain (on the left)
7	$\mathrm{L}\{x(t+a)\}=\mathrm{e}^{as}\left[X(s)-\int_0^a x(t)\mathrm{e}^{-st}\,\mathrm{d}t\right],\ a\ge 0$
	Complex shifting in complex domain
8	$\mathrm{L}\{x(t)\mathrm{e}^{\mp at}\}=X(s\pm a)$
	Derivative in time domain
9	1 order derivative $\mathrm{L}\left\{\frac{\mathrm{d}x(t)}{\mathrm{d}t}\right\}=sX(s)-x(0)$
10	n order derivative $\mathrm{L}\left\{\frac{\mathrm{d}^n x(t)}{\mathrm{d}t^n}\right\}=s^nX(s)-\sum_{i=1}^{n}s^{n-i}\frac{\mathrm{d}^{i-1}x(0)}{\mathrm{d}t^{i-1}}$
	Derivative in complex variable domain
11	$\mathrm{L}\{tx(t)\}=-\frac{\mathrm{d}X(s)}{\mathrm{d}s}$
	Integral in time domain
12	$\mathrm{L}\left\{\int_0^t x(\tau)\mathrm{d}\tau\right\}=\frac{1}{s}X(s)$

	Integral value
13	$\int_0^\infty x(t)\,\mathrm{d}t = \lim_{s\to 0} X(s)$
14	$\int_0^\infty tx(t)\,\mathrm{d}t = -\lim_{s\to 0} \frac{\mathrm{d}X(s)}{\mathrm{d}s}$
	Periodical function transform
15	$\mathrm{L}\{x(t)+x(t-a)+x(t-2a)+\ldots\} = X(s)\frac{1}{1-\mathrm{e}^{-as}}$ a – period, $a>0$
	Initial value in time domain (if it exists)
16	$x(0) = \lim_{t\to 0_+} x(t) = \lim_{s\to\infty} sX(s)$
	Final value in time domain (if it exists)
17	$x(\infty) = \lim_{t\to\infty} x(t) = \lim_{s\to 0} sX(s)$
	Mathematical operation with respect to independent parameter
18	$\mathrm{L}\{x(t,a)\} = X(s,a)$
19	$\mathrm{L}\{\lim_{a\to a_0} x(t,a)\} = \lim_{a\to a_0} X(s,a)$
20	$\mathrm{L}\left\{\frac{\partial x(t,a)}{\partial a}\right\} = \frac{\partial X(s,a)}{\partial a}$
21	$\mathrm{L}\left\{\int_{a_1}^{a_2} x(t,a)\,\mathrm{d}a\right\} = \int_{a_1}^{a_2} X(s,a)\,\mathrm{d}a$
	Inverse transform by residues
22	$x(t) = \sum_i \operatorname{res}_{s=s_i}\left[X(s)\mathrm{e}^{st}\right] = \sum_i \left\{\frac{1}{(r_i-1)!}\lim_{s\to s_i}\frac{\mathrm{d}^{r_i-1}}{\mathrm{d}s^{r_i-1}}\left[(s-s_i)^{r_i} X(s)\mathrm{e}^{st}\right]\right\}$ r_i – the multiplicity of transform pole s_i $n = \sum_i r_i$ – the polynomial degree in the transform denominator

2 LAPLACE TRANSFORM - CORRESPONDENCES

	Transform $X(s)$	Original $x(t)$
1	s	$\dot{\delta}(t)$
2	1	$\delta(t)$
3	$\frac{1}{s}$	$\eta(t)$
4	$\frac{1}{s^n}, \quad n = 1,2,\dots$	$\frac{t^{n-1}}{(n-1)!}$
5	$\frac{s}{T_1 s + 1}$	$\alpha_1\left[\delta(t) - \alpha_1 \,\mathrm{e}^{-\alpha_1 t}\right], \quad \alpha_1 = \frac{1}{T_1}$
6	$\frac{1}{T_1 s + 1}$	$\alpha_1 \,\mathrm{e}^{-\alpha_1 t}, \quad \alpha_1 = \frac{1}{T_1}$
7	$\frac{1}{s(T_1 s + 1)}$	$1 - \mathrm{e}^{-\alpha_1 t}, \quad \alpha_1 = \frac{1}{T_1}$
8	$\frac{1}{s^2(T_1 s + 1)}$	$\frac{1}{\alpha_1}\left(\mathrm{e}^{-\alpha_1 t} - 1\right) + t, \quad \alpha_1 = \frac{1}{T_1}$
9	$\frac{b_1 s + 1}{s(T_1 s + 1)}$	$1 + (\alpha_1 b_1 - 1)\mathrm{e}^{-\alpha_1 t}, \quad \alpha_1 = \frac{1}{T_1}$
10	$\frac{b_1 s + 1}{s^2(T_1 s + 1)}$	$C_1\left(1 - \mathrm{e}^{-\alpha_1 t}\right) + t, \;\; C_1 = b_1 - \frac{1}{\alpha_1}, \quad \alpha_1 = \frac{1}{T_1}$
11	$\frac{s}{(T_1 s + 1)^2}$	$\alpha_1^2(1 - \alpha_1 t)\mathrm{e}^{-\alpha_1 t}, \quad \alpha_1 = \frac{1}{T_1}$
12	$\frac{1}{(T_1 s + 1)^2}$	$\alpha_1^2 t \,\mathrm{e}^{-\alpha_1 t}, \quad \alpha_1 = \frac{1}{T_1}$
13	$\frac{1}{s(T_1 s + 1)^2}$	$1 - (1 + \alpha_1 t)\mathrm{e}^{-\alpha_1 t}, \quad \alpha_1 = \frac{1}{T_1}$
14	$\frac{1}{s^2(T_1 s + 1)^2}$	$t - \frac{2}{\alpha_1} + \left(\frac{2}{\alpha_1} + t\right)\mathrm{e}^{-\alpha_1 t}, \quad \alpha_1 = \frac{1}{T_1}$
15	$\frac{b_1 s + 1}{(T_1 s + 1)^2}$	$\alpha_1^2\left[b_1 + (1 - \alpha_1 b_1)t\right]\mathrm{e}^{-\alpha_1 t}, \quad \alpha_1 = \frac{1}{T_1}$
16	$\frac{b_1 s + 1}{s(T_1 s + 1)^2}$	$1 - \left[1 + \alpha_1(1 - \alpha_1 b_1)t\right]\mathrm{e}^{-\alpha_1 t}, \quad \alpha_1 = \frac{1}{T_1}$

	Transform $X(s)$	Original $x(t)$
17	$\dfrac{b_1 s+1}{s^2(T_1 s+1)^2}$	$t+C_1-(C_1-C_2 t)\mathrm{e}^{-\alpha_1 t}$ $C_1=b_1-\dfrac{2}{\alpha_1},\ \ C_2=1-\alpha_1 b_1,\ \ \alpha_1=\dfrac{1}{T_1}$
18	$\dfrac{s}{(T_1 s+1)^n},\quad n=2,3,\ldots$	$\alpha_1^n\dfrac{t^{n-2}}{(n-1)!}(n-1-\alpha_1 t)\mathrm{e}^{-\alpha_1 t},\quad \alpha_1=\dfrac{1}{T_1}$
19	$\dfrac{1}{(T_1 s+1)^n},\quad n=1,2,\ldots$	$\alpha_1^n\dfrac{t^{n-1}}{(n-1)!}\mathrm{e}^{-\alpha_1 t},\quad \alpha_1=\dfrac{1}{T_1}$
20	$\dfrac{1}{s(T_1 s+1)^n},\quad n=1,2,\ldots$	$1-\mathrm{e}^{-\alpha_1 t}\sum_{i=0}^{n-1}\alpha_1^i\dfrac{t^i}{i!},\quad \alpha_1=\dfrac{1}{T_1}$
21	$\dfrac{1}{s^2(T_1 s+1)^n},\quad n=1,2,\ldots$	$t-\dfrac{n}{\alpha_1}+\mathrm{e}^{-\alpha_1 t}\sum_{i=0}^{n-1}\alpha_1^{i-1}(n-i)\dfrac{t^i}{i!},\quad \alpha_1=\dfrac{1}{T_1}$
22	$\dfrac{s}{(T_1 s+1)(T_2 s+1)},\quad T_1\neq T_2$	$C_1\mathrm{e}^{-\alpha_1 t}-C_2\mathrm{e}^{-\alpha_2 t},\ \ \alpha_1=\dfrac{1}{T_1},\ \ \alpha_2=\dfrac{1}{T_2}$ $C_1=\dfrac{1}{T_1(T_2-T_1)},\ \ C_2=\dfrac{1}{T_2(T_2-T_1)}$
23	$\dfrac{1}{(T_1 s+1)(T_2 s+1)},\quad T_1\neq T_2$	$C_1(\mathrm{e}^{-\alpha_1 t}-\mathrm{e}^{-\alpha_2 t}),\ \ C_1=\dfrac{1}{T_1-T_2},\ \ \alpha_1=\dfrac{1}{T_1},\ \ \alpha_2=\dfrac{1}{T_2}$
24	$\dfrac{1}{s(T_1 s+1)(T_2 s+1)},\quad T_1\neq T_2$	$1+C_1\mathrm{e}^{-\alpha_1 t}-C_2\mathrm{e}^{-\alpha_2 t},\ \ \alpha_1=\dfrac{1}{T_1},\ \ \alpha_2=\dfrac{1}{T_2}$ $C_1=\dfrac{T_1}{T_2-T_1},\ \ C_2=\dfrac{T_2}{T_2-T_1}$
25	$\dfrac{1}{s^2(T_1 s+1)(T_2 s+1)},\quad T_1\neq T_2$	$t-C_0+C_1\mathrm{e}^{-\alpha_1 t}-C_2\mathrm{e}^{-\alpha_2 t},\ \ C_0=T_1+T_2$ $C_1=\dfrac{T_1^2}{T_1-T_2},\ \ C_2=\dfrac{T_2^2}{T_1-T_2},\ \ \alpha_1=\dfrac{1}{T_1},\ \ \alpha_2=\dfrac{1}{T_2}$
26	$\dfrac{b_1 s+1}{(T_1 s+1)(T_2 s+1)},\quad T_1\neq T_2$	$C_1\mathrm{e}^{-\alpha_1 t}-C_2\mathrm{e}^{-\alpha_2 t},\ \ \alpha_1=\dfrac{1}{T_1},\ \ \alpha_2=\dfrac{1}{T_2}$ $C_1=\dfrac{T_1-b_1}{T_1(T_1-T_2)},\ \ C_2=\dfrac{T_2-b_1}{T_2(T_1-T_2)}$
27	$\dfrac{b_1 s+1}{s(T_1 s+1)(T_2 s+1)},\quad T_1\neq T_2$	$1+C_1\mathrm{e}^{-\alpha_1 t}+C_2\mathrm{e}^{-\alpha_2 t},\ \ \alpha_1=\dfrac{1}{T_1},\ \ \alpha_2=\dfrac{1}{T_2}$ $C_1=\dfrac{b_1-T_1}{T_1-T_2},\ \ C_2=\dfrac{T_2-b_1}{T_1-T_2}$

	Transform $X(s)$	Original $x(t)$
28	$\dfrac{b_1 s+1}{s^2(T_1 s+1)(T_2 s+1)},\quad T_1 \neq T_2$	$t + C_0 + C_1 e^{-\alpha_1 t} + C_2 e^{-\alpha_2 t},\quad C_0 = -T_1 - T_2 + b_1$ $C_1 = \dfrac{(b_1 - T_1)T_1}{T_2 - T_1},\quad C_2 = \dfrac{(T_2 - b_1)T_2}{T_2 - T_1},\quad \alpha_1 = \dfrac{1}{T_1},\quad \alpha_2 = \dfrac{1}{T_2}$
29	$\dfrac{s}{\prod_{i=1}^{n}(T_i s+1)},\quad n = 2,3,\ldots$ T_i – different	$-\sum_{i=1}^{n} C_i e^{-\alpha_i t},\quad C_i = \dfrac{T_i^{n-3}}{\prod_{k=1,k\neq i}^{n}(T_i - T_k)},\quad \alpha_i = \dfrac{1}{T_i}$
30	$\dfrac{1}{\prod_{i=1}^{n}(T_i s+1)},\quad n = 2,3,\ldots$ T_i – different	$\sum_{i=1}^{n} C_i e^{-\alpha_i t},\quad C_i = \dfrac{T_i^{n-2}}{\prod_{k=1,k\neq i}^{n}(T_i - T_k)},\quad \alpha_i = \dfrac{1}{T_i}$
31	$\dfrac{1}{s\prod_{i=1}^{n}(T_i s+1)},\quad n = 2,3,\ldots$ T_i – different	$1-\sum_{i=1}^{n} C_i e^{-\alpha_i t},\quad C_i = \dfrac{T_i^{n-1}}{\prod_{k=1,k\neq i}^{n}(T_i - T_k)},\quad \alpha_i = \dfrac{1}{T_i}$
32	$\dfrac{1}{s^2\prod_{i=1}^{n}(T_i s+1)},\quad n = 2,3,\ldots$ T_i – different	$t - C_0 + \sum_{i=1}^{n} C_i e^{-\alpha_i t},\quad \alpha_i = \dfrac{1}{T_i}$ $C_i = \dfrac{T_i^{n}}{\prod_{k=1,k\neq i}^{n}(T_i - T_k)},\quad C_0 = \sum_{i=1}^{n} T_i$
33	$\dfrac{\omega}{s^2+\omega^2}$	$\sin \omega t$
34	$\dfrac{s}{s^2+\omega^2}$	$\cos \omega t$
35	$\dfrac{s}{T_0^2 s^2 + 2\xi_0 T_0 s + 1},$ $0 \le \xi_0 < 1$	$-C_1 e^{-\gamma t}\sin(\omega t - \varphi),\quad C_1 = \dfrac{1}{\omega T_0^3},\quad \gamma = \dfrac{\xi_0}{T_0}$ $\omega = \dfrac{1}{T_0}\sqrt{1-\xi_0^2},\quad \varphi = \operatorname{arctg}\dfrac{\omega}{\gamma}$
36	$\dfrac{1}{T_0^2 s^2 + 2\xi_0 T_0 s + 1},$ $0 \le \xi_0 < 1$	$C_1 e^{-\gamma t}\sin \omega t,\quad C_1 = \dfrac{1}{\omega T_0^2},\quad \gamma = \dfrac{\xi_0}{T_0},\quad \omega = \dfrac{1}{T_0}\sqrt{1-\xi_0^2}$
37	$\dfrac{1}{s(T_0^2 s^2 + 2\xi_0 T_0 s + 1)},$ $0 \le \xi_0 < 1$	$1 - C_1 e^{-\gamma t}\sin(\omega t + \varphi),\quad C_1 = \dfrac{1}{\omega T_0},\quad \gamma = \dfrac{\xi_0}{T_0}$ $\omega = \dfrac{1}{T_0}\sqrt{1-\xi_0^2},\quad \varphi = \operatorname{arctg}\dfrac{\omega}{\gamma}$
38	$\dfrac{1}{s^2(T_0^2 s^2 + 2\xi_0 T_0 s + 1)},$ $0 \le \xi_0 < 1$	$t - C_0 + C_1 e^{-\gamma t}\sin(\omega t + 2\varphi),\quad C_0 = 2\xi_0 T_0^2$ $C_1 = \dfrac{1}{\omega}, \gamma = \dfrac{\xi_0}{T_0},\quad \omega = \dfrac{1}{T_0}\sqrt{1-\xi_0^2},\quad \varphi = \operatorname{arctg}\dfrac{\omega}{\gamma}$

	Transform $X(s)$	Original $x(t)$
39	$\dfrac{b_1 s+1}{T_0^2 s^2+2\xi_0 T_0 s+1}$, $0 \le \xi_0 < 1$	$C_1\,\mathrm{e}^{-\gamma t}\sin(\omega t+\varphi)$, $C_1=\dfrac{1}{\omega T_0^3}\sqrt{(1-2b_1\gamma)T_0^2+b_1^2}$ $\gamma=\dfrac{\xi_0}{T_0}$, $\omega=\dfrac{1}{T_0}\sqrt{1-\xi_0^2}$, $\varphi=\operatorname{arctg}\dfrac{\omega b_1}{1-\gamma b_1}$
40	$\dfrac{b_1 s+1}{s(T_0^2 s^2+2\xi_0 T_0 s+1)}$, $0 \le \xi_0 < 1$	$1+C_1\,\mathrm{e}^{-\gamma t}\sin(\omega t-\varphi)$, $C_1=\dfrac{1}{\omega T_0^2}\sqrt{(1-2b_1\gamma)T_0^2+b_1^2}$ $\gamma=\dfrac{\xi_0}{T_0}$, $\omega=\dfrac{1}{T_0}\sqrt{1-\xi_0^2}$, $\varphi=\operatorname{arctg}\dfrac{\omega T_0^2}{b_1-\gamma T_0^2}$

b_1, b_2 – the real constants, $T_i > 0$, $i = 0, 1, \ldots$

Authors:	Prof. Ing. Antonín Víteček, CSc., Dr.h.c. Prof. Ing. Miluše Vítečková, CSc. Doc. Ing. Lenka Landryová, CSc.
Department:	Control Systems and Instrumentation
Title:	Basic Principles of Automatic Control
Place, Year, Edition:	Ostrava, 2012, 1st
Pages:	118
Published:	VŠB – Technical University of Ostrava 17. listopadu 15/2172 708 33 Ostrava - Poruba

ISBN 978-80-248-4062-8