Chapter 4
Negative Feedback

The aim of a control system is to ensure reference signal tracking as well as disturbance rejection. The control system must not be not very sensitive to measurement noise or to plant/model mismatch.

The designed control system has to ensure various quality specifications. Also, it has to be technically realizable and eligible in terms of economic and other (e.g., environmental protection or safety) viewpoints.

4.1 Control in Open- and Closed-Loop

If, when deciding whether intervention in a process is necessary, the information is taken not from the output of the process but from another source, or a priori knowledge about the process or its environment is used, then the realized structure is called open-loop control (Fig. 4.1). Here P denotes the transfer function of the process (plant), C is the transfer function of the controller (regulator), r denotes the reference signal, y is the output signal, while y_{ni} and y_{no} denote the input and the output disturbances, respectively.

The reference signal tracking would be ideal if the control device realized the inverse of the transfer function of the plant. With open-loop control, the reference signal tracking can be realizable, but open-loop control is not able to reject the effect of the disturbances.

The effect of the measurable output disturbance could be eliminated by feed forward of the disturbance according to Fig. 4.2.

Generally the perfect inverse of the transfer function of the plant is not realizable (If e.g., the process contains dead time, its inverse would mean the prediction of a future output value. The signal transfer is also non-realizable if the degree of the numerator of the inverse transfer function is higher than that of its denominator.). Approximating the inverse of the process can be realized by feeding back an amplifier of high gain K through the transfer function of the plant (Fig. 4.3).

L. Keviczky et al., *Control Engineering*, Advanced Textbooks in Control and Signal Processing, https://doi.org/10.1007/978-981-10-8297-9_4

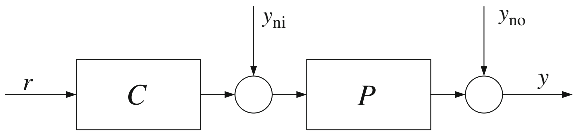


Fig. 4.1 Open-loop control

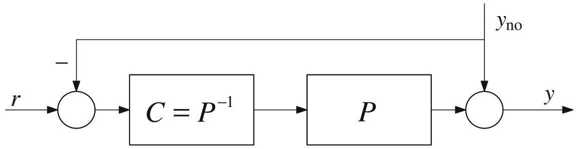


Fig. 4.2 Open-loop control with feed forward

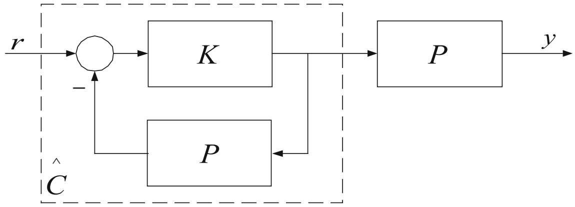


Fig. 4.3 Open-loop control with an element approximating the inverse of the process

Using the above structure the transfer function of the controller is:

$$\hat{C} = \frac{K}{1 + KP(s)} = \frac{1}{\frac{1}{K} + P(s)} \approx \frac{1}{P(s)}. \tag{4.1}$$

The control is realized through negative feedback if the input signal (the manipulated variable) of the process is affected by the difference between the measured output signal and its desired prescribed value. The measured output value is generally noisy because of the noise component y_z released by the measurement equipment. Based on the error signal e the controller C generates the manipulated variable u, which modifies the output signal of the process P. The output signal of the process changes according to the dynamics of the process until it reaches its desired value. Control via negative feedback is called closed-loop control. The block diagram of a closed-loop control system is given in Fig. 4.4. Often the reference signal is filtered by a filter element given by the transfer function F (denoted by the dotted line in the figure).

Comparing Figs. 4.3 and 4.4 shows that the two systems are the same if the disturbances and the measurement noise are not considered, the filter is supposed to be unity ($F = 1$) and in the closed-loop system the controller is proportional, chosen as $C = K$. But as in the closed-loop system the output signal is fed back, not the control signal, in addition to reference signal tracking, the closed-loop system is also able to reject the effects of the disturbances and the measurement noise. Whatever effect is causing a deviation of the output signal from its desired value,

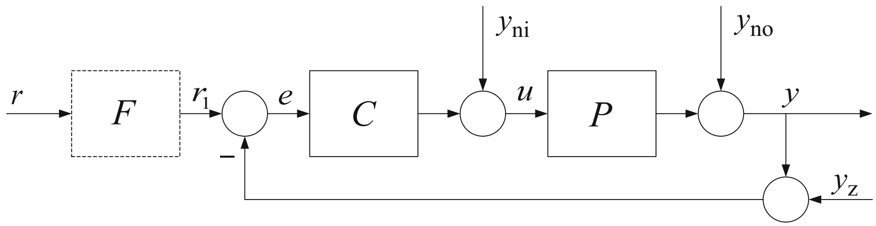


Fig. 4.4 Closed-loop control system

the error signal will be different from zero, creating a control signal to eliminate the deviation.

In a closed-loop control structure the best reference signal tracking is achieved by adjusting the controller C to ensure that the relationship between the control signal u and the reference signal r according to $U(s)/R(s) = C/(1 + CP)$ would provide the inverse of the process model. If the exact inverse is non-realizable, then its best realizable approximation can be employed. (It has to be mentioned that generally the inverse of the process could be well approximated only within a given frequency range.)

A comparison of open-loop and closed-loop control was given in Chap. 1. In the sequel, the main properties of closed-loop control will be discussed.

4.2 The Basic Properties of the Closed Control Loop

The main properties of closed-loop control systems will be illustrated through some simple examples.

Stability. A basic requirement for a closed-loop control system is that for a finite change in the input signal it should respond with a finite change in the output signal, i.e. a steady state should be reached. In a control system realized by negative feedback oscillations with steady or increasing amplitudes may occur. The reason for this is that the execution of the decision to change the process output is delayed by the process dynamics. High gains in the control system may increase the unfavorable inertial change of the signals to such an extent that the control system will not be able to reach a steady state. Stability will be discussed in detail in Chap. 5.

Reference signal tracking. With a closed-loop control realized by negative feedback, the output signal should follow the reference signal as accurately as possible. In the control system in Fig. 4.5 the plant is described by a first-order lag and the controller is a proportional element. If a step reference signal is to be tracked, there will be a steady state error in the system, as only a steady constant input signal is able to produce a constant signal value at the output of the first-order lag. The value of the steady state error will be $e_{steady} = 1/(1 + A_P A)$. This error will

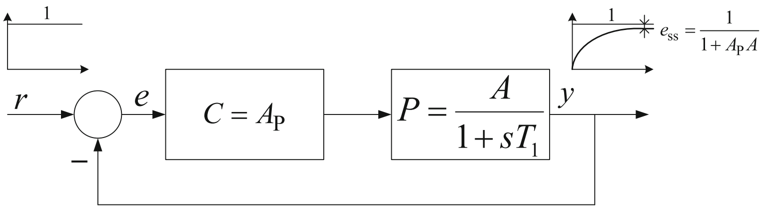


Fig. 4.5 Closed-loop control system with proportional controller

be small if the loop-gain $K = A_P A$ is high. Static accuracy in the steady state could be ensured by applying an integrating controller instead of a proportional controller. The property of an integrator is that its output can be constant only if its input (the error signal) finally has reached the value zero.

Stabilization of an unstable process. An unstable process can be stabilized by negative feedback.

Example 4.1 Let us consider the system given in Fig. 4.6. For unit step input, the output of the closed-loop system for $t \geq 0$ is $y(t) = \mathcal{L}^{-1}\{K/s(s-2)\}$ $= K(e^{2t} - 1)/2$, which tends to infinity if $t \to \infty$. With a proportional negative feedback β, the resulting transfer function is

$$T(s) = \frac{Y(s)}{R(s)} = \frac{\frac{K}{s-2}}{1 + \frac{K\beta}{s-2}} = \frac{K}{s + K\beta - 2}.$$

The feedback system is stable, i.e. its transients decay for any β satisfying $K\beta > 2$. ■

Decreasing the effect of the disturbance in the output signal. In the open-loop control in Fig. 4.7 the disturbance appears entirely in the output. In the feedback system the effect of the disturbance in the output signal is decreased by $1/(1 + A\beta)$, i.e., the higher the value of the loop gain $A\beta$, the better the feedback reduces the effect of the disturbance.

Feedback can improve the transient response. Let us consider the first-order lag element in Fig. 4.8. With a constant feedback β the resulting transfer function is

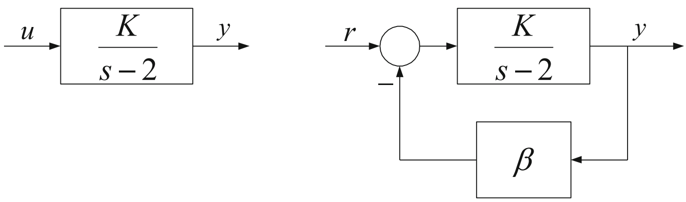


Fig. 4.6 An unstable system can be stabilized by negative feedback

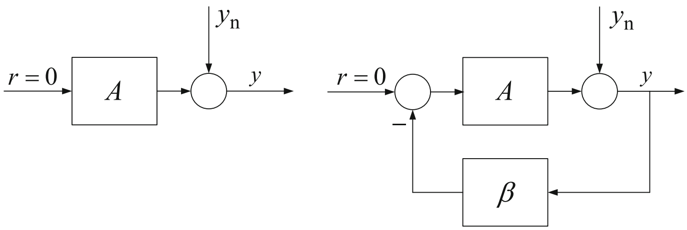


Fig. 4.7 Negative feedback decreases the effect of a disturbance

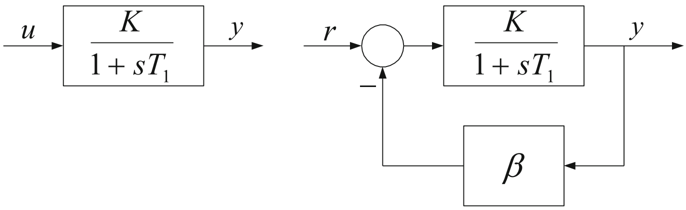


Fig. 4.8 Feedback modifies the transient response

$$T(s) = \frac{Y(s)}{R(s)} = \frac{K}{1+\beta K}\frac{1}{1+s\frac{T_1}{1+\beta K}} = K'\frac{1}{1+sT_1'}.$$

The time constant has been decreased, so the system is faster. At the same time the gain also has been decreased, which generally has to be compensated using a filter of the appropriate gain (as shown in Fig. 4.4).

Feedback decreases the sensitivity of the process to parameter changes.

In Fig. 4.9 the gain of the proportional element without feedback is 10. Suppose the gain of the feedback system is the same: $A_1/(1 + A_1\beta) = 10$. Choose the value of A_1 to be 1000. With this value, $\beta = 0.099$ is obtained. If the value of the input signal is 10, then in both systems the value of the output signal is 100. Reduce both values A and A_1 by 2%. Then $A = 9.8$ and $A_1 = 980$. In the original system then the output value decreases to 98, while in the feedback system it remains quite close to

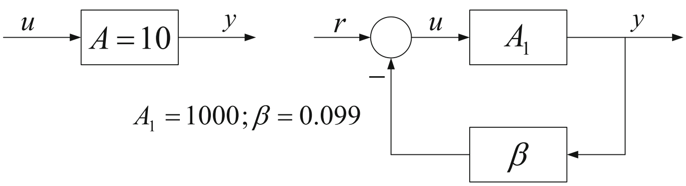


Fig. 4.9 With feedback the system becomes less sensitive to parameter changes

100 (99.98). In the feedback system the error appearing because of the parameter change is $(1+A_1\beta)$ times less than in the original system.

In the range of high gains the feedback system creates the approximate inverse of the feedback element.

Let us consider the circuit in Fig. 4.10. The element given by the transfer function $H_1(s)$ is fed back with negative feedback through an element given by $H_2(s)$. The resulting transfer function is $H(s) = H_1(s)/[1+H_1(s)H_2(s)]$. In the frequency range where $|H_1(j\omega)H_2(j\omega)|$ is much higher than 1, the resulting transfer function approximates the inverse of the transfer function H_2. In that portion of the frequency domain, where the absolute value of the loop frequency function is much less than 1, the resulting frequency function approximates the frequency function H_1 of the forward path.

Feedback has a linearizing effect.

Let us consider the static non-linear characteristics in Fig. 4.11a. The characteristics can be divided into three linearized ranges, where the linear transfer gain of the individual ranges is determined by the slope A of the straight line fitted to the curve at the given operating point. Suppose the value of the proportional feedback gain is β. In the feedback system the slope of the individual linearized ranges of the static characteristics is $A/(1+A\beta)$. The bigger is $A\beta$, the better the transfer gain approximates the value of $1/\beta$, becoming independent of the slopes A of the individual ranges of the static characteristics. Figure 4.11b shows the gains of the linearized individual parts with feedback gain $\beta = 10$. It can be seen that the slopes in the different ranges are almost the same, the characteristic is approximately linear in the whole domain. For $\beta = 100$ the linearization is still better (with slopes 0.00998, 0.00995 and 0.0099).

It has to be emphasized that while the non-linear characteristics have been linearized to a great extent, considering the input u_1 the ranges of the linearized sections have been changed compared to the original sections. For example, if $\beta = 10$:

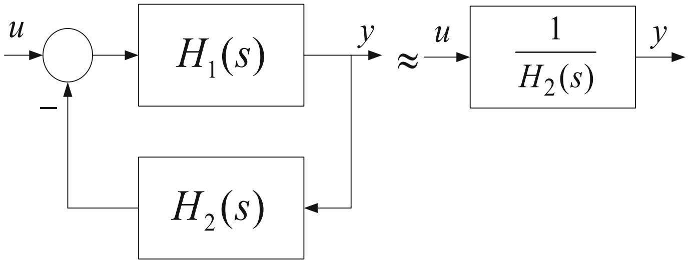
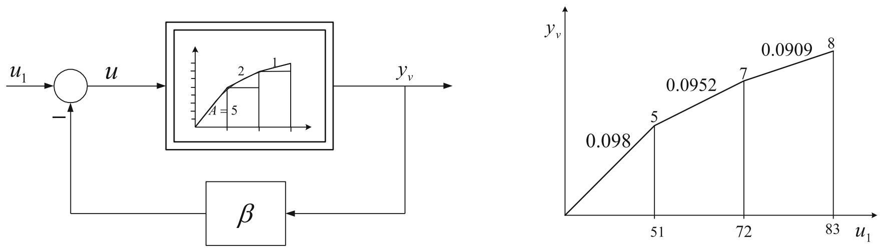


Fig. 4.10 In the range of high gains the feedback creates the approximate inverse of the feedback element



(a) Static characteristics with constant feedback

(b) Feedback has a linearization effect

(Feedback has a linearizing effect, but the ranges of the linearized sections associated to the new input u_1 have been changed)

Fig. 4.11 Linearization by feedback.

If $0 \le u \le 1$, then $y_v = 5u$ and $u = u_1 - 50u$, hence $u_1 = 51u$ and $y_v = \frac{5}{51} u_1$.

If $1 < u \le 2$, then $y_v = 3 + 2u$ and $u = u_1 - 10(3 + 2u)$, hence $u = \frac{1}{21} u_1 - \frac{30}{21}$ and $y_v = \frac{3}{21} + \frac{2}{21} u_1$.

If $2 < u \le 3$, then $y_v = 5 + u$ and $u = u_1 - 10(5 + u)$, hence $u = \frac{1}{11} u_1 - \frac{50}{11}$ and $y_v = \frac{5}{11} + \frac{1}{11} u_1$.

Feeding back an integrator by a static non-linear element results in the inverse of the non-linear characteristics.

Let us consider the circuit given in Fig. 4.12. Negative feedback is applied to an integrator through a static quadratic non-linear element. As the output of the integrator can be constant only if its input, i.e. the error signal, becomes zero, $r = y^2$, and $y = \sqrt{r}$, i.e. the circuit realizes the inverse of the non-linearity in the feedback path.

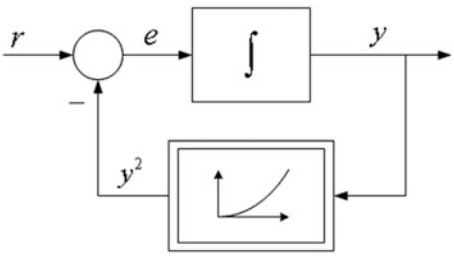


Fig. 4.12 Feeding back an integrator by a non-linear static element realizes the inverse characteristics

4.3 The Feedback Operational Amplifier

With the invention of the telephone and the development of telecommunication, high gain amplifiers were used to compensate the damping of signals over long transmission lines. Invented by Black, the amplifier with negative feedback (1927) ensured a stable solution to decrease the sensitivity of vacuum tube amplifiers to the change of their characteristics, and at the same time it linearized, to a great extent, the nonlinear characteristics of the amplifier.

Operational amplifiers built of integrated circuits are also used in control circuits for amplification and compensation.

Let us analyze the signal transfer properties of the feedback operational amplifier. Its circuit is shown in Fig. 4.13. In the input and feedback path, resistors, capacitors or an interconnection of resistors and capacitors can be employed. For the sake of simplicity let us consider resistors both in the forward and the feedback path. The gain G of the amplifier is of a very high value (in the range of $10^4 - 10^8$).

Let us determine the transfer function and the corresponding block diagram of the operational amplifier.

The output voltage can be expressed as

$$U_2 = -GU. \tag{4.2}$$

If the input current I can be neglected (e.g. the input resistance of the amplifier is high) then the following Kirchhoff voltage law equation can be written for the input point of the amplifier:

$$\frac{U_1 - U}{R_1} = \frac{U - U_2}{R_2}. \tag{4.3}$$

Let us express the variable U using this equation.

$$U = \frac{R_2}{R_1 + R_2}\left(U_1 + \frac{R_1}{R_2}U_2\right). \tag{4.4}$$

Substituting this expression into (4.2), the following equation is obtained after some manipulations:

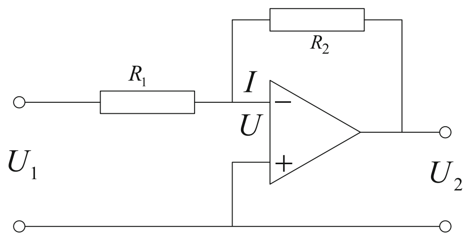


Fig. 4.13 Feedback operational amplifier

$$\frac{U_2}{U_1} = -\frac{R_2}{R_1}\frac{1}{1+\frac{1}{G}\left(1+\frac{R_2}{R_1}\right)}. \tag{4.5}$$

It can be seen that if $G \to \infty$, the resulting transfer gain is determined by the ratio of the two resistances. For high values of G, the transfer gain keeps its value quite close to its nominal value even in the case of possible changes in G. (Note that if impedances Z_1 and Z_2 are used in the input and the feedback path instead of resistors, the transfer function of the operational amplifier will be approximately $-Z_2/Z_1$, and depending on the representation of the impedances, different mathematical operations can be realized.)

Based on the above relationships, a block diagram of the feedback operational amplifier can be found. Figure 4.14 shows three equivalent schemes.

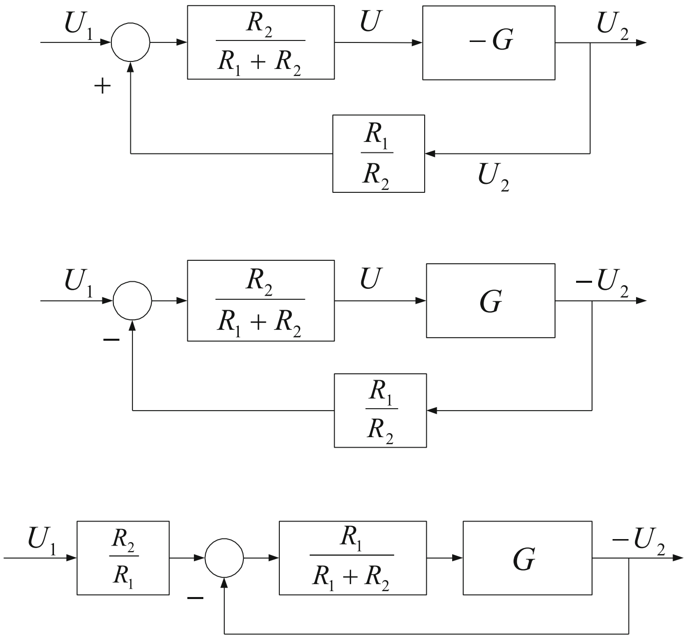


Fig. 4.14 Equivalent block diagrams of the feedback operational amplifier

4.4 The Transfer Characteristics of the Closed Control Loop

The behavior of a closed-loop control system can be investigated by the overall transfer functions exhibiting the relationships between the output and the input signals.

As the systems are assumed to be linear, the superposition theorem can be applied. The effect of the various external signals can simply be summed to obtain the output signal.

Let us determine the overall transfer functions between the controlled signal y, the error signal e, and the control signal u as the output signals, and the reference signal r, the output disturbance y_{no}, the input disturbance y_{ni}, and the measurement noise y_{z} as input signals.

According to Fig. 4.4, the relationships between these input and output signals are

$$Y(s) = \frac{F(s)C(s)P(s)}{1+C(s)P(s)}R(s) + \frac{1}{1+C(s)P(s)}Y_{\mathrm{no}}(s) + \frac{P(s)}{1+C(s)P(s)}Y_{\mathrm{ni}}(s) - \frac{C(s)P(s)}{1+C(s)P(s)}Y_{\mathrm{z}}(s), \tag{4.6}$$

$$E(s) = \frac{F(s)}{1+C(s)P(s)}R(s) - \frac{1}{1+C(s)P(s)}Y_{\mathrm{no}}(s) - \frac{P(s)}{1+C(s)P(s)}Y_{\mathrm{ni}}(s) - \frac{1}{1+C(s)P(s)}Y_{\mathrm{z}}(s), \tag{4.7}$$

$$U(s) = \frac{F(s)C(s)}{1+C(s)P(s)}R(s) - \frac{C(s)}{1+C(s)P(s)}Y_{\mathrm{no}}(s) - \frac{C(s)P(s)}{1+C(s)P(s)}Y_{\mathrm{ni}}(s) - \frac{C(s)}{1+C(s)P(s)}Y_{\mathrm{z}}(s), \tag{4.8}$$

On the basis of these relationships, the output signals can be determined with the knowledge of the input signals. From the time evolution of the output signals it can be verified whether the control system satisfies the quality specifications or not.

It has to be emphasized that the frequency ranges of the different input signals are generally different. The reference signal and the disturbances generally contain low frequency components, whereas the measurement noise generally is a zero mean signal containing high frequency components. If the absolute value of the frequency function obtained from an overall transfer function by substituting $s = j\omega$—considering a given input signal—is approximately unity over a significant frequency range, then the system tracks the signal, but if the transfer function approximates zero, the system attenuates the considered input signal.

It can be seen that all the overall transfer functions have the same denominator, namely $1 + C(s)P(s)$, which is the characteristic polynomial of the closed-loop control system. The roots of the characteristic polynomial determine the stability and the dynamic properties of the transients of the control system. For stable performance it is required that the transients of the output signal should decrease, i.e. the roots of the characteristic equation should be at the left hand side of the complex plane. Chapter 5 deals in detail with methods of stability investigation.

From Eq. (4.7) it can be seen, that if the filter $F(s)$ is a proportional element with gain unity, then the error of reference signal tracking and the error of output disturbance rejection are the same, i.e., the control system follows the reference signal with the same dynamics and the same static error as it rejects the effect of the output disturbance in the output signal. With the appropriate choice of filter $F(s)$ it can be ensured that the properties of reference signal tracking and of disturbance rejection would be different.

If $F(s) = 1$ the control system is called a *One-Degree of Freedom* (*ODOF*) system, while if $F(s)$ is given by a non-unity transfer function, it is called *Two-Degree of Freedom* (*TDF*) system. In the case of an *ODOF*, 4 overall transfer functions determine the overall signal transfer properties between the output signals (the controlled signal y and the manipulated variable u) and the input signals (the reference signal, the disturbances, and the measurement noise), but in the *TDOF* case, 6 overall transfer functions are needed for this determination.

As the disturbance y_{ni} can always be transformed to an equivalent output disturbance, and the signs do not have to be considered, it is sufficient to investigate the following 6 overall transfer functions.

$$\begin{aligned} \frac{Y}{R} &= \frac{FCP}{1+CP}; \quad \frac{Y}{Y_{\mathrm{z}}} = \frac{-CP}{1+CP}; \quad \frac{Y}{Y_{\mathrm{ni}}} = \frac{P}{1+CP} \\ \frac{U}{R} &= \frac{FC}{1+CP}; \quad \frac{U}{Y_{\mathrm{z}}} = \frac{-C}{1+CP}; \quad \frac{E}{Y_{\mathrm{no}}} = \frac{1}{1+CP} \end{aligned} \tag{4.9}$$

The first column characterizes reference signal tracking, the second column characterizes the properties of the disturbance rejection and the third column characterizes the rejection of measurement noise. If $F(s) = 1$, the second and third columns give the 4 characterizing transfer functions. Arranging these functions into matrix form, a transfer function matrix of the closed-loop control system is obtained. To ensure the stability of the closed-loop control system, all the overall transfer functions have to be stable. Also, all the overall transfer functions have to ensure the prescribed dynamic behavior between the given input and output signals.

One of the usual controller design procedures is the cancellation of the unfavorable poles of the plant P with the zeros of the controller C. But it can be seen that the dynamics of the plant P remains in the expression of the overall transfer function between the output signal and the input disturbance. It is not allowed to cancel the unstable poles of the plant, as though they become invisible in the relationship between the output signal and the reference signal, they do appear in the transfer relationship between the output signal and the input disturbance. (It has to be mentioned that even regarding the relationship between the output and the reference signal the unstable pole can not be cancelled quite accurately, as its value generally is obtained by measurements or modeling which certainly involves errors, or its value may change over time; therefore the pole cancellation is never perfectly accurate and instability persists in the system.)

It is reasonable to design a controller in two steps. First the controller C is to be designed to ensure the appropriate rejection of the disturbances and the measurement noise, then the filter F is to be designed for appropriate reference signal tracking.

For good reference signal tracking if $F(s) = 1$, the so called *complementary sensitivity function* $T = CP/(1 + CP)$ has to approximate 1 on those frequencies which characterize the input signal. This means that at these frequencies, the condition $CP \gg 1$ has to be fulfilled. Consider the overall transfer function $S(s) = 1/[1 + C(s)P(s)]$ giving the relationship between the error signal and the reference signal. $S(s)$ is also called the *sensitivity function*. Time domain analysis shows that the error signal contains signal components originating from the poles of the closed-loop and also from the poles of the reference signal. Once the transients decay in the error signal, the quasistationary components originating from the poles of the reference signal are maintained. If $C(s)$ contains the poles of the reference signal, in the error signal the poles of the controller cancel the poles of the reference signal. In this case tracking the reference signal $R(s)$ the steady-state error becomes zero. Thus $C(s) = K_c R(s)$, where $K_c \gg 1$. Considering the disturbance rejection, if the condition $CP \gg 1$ is fulfilled, then the Laplace transform of the error signal as a response for the input and the output disturbances is approximately: $E(s) \approx -[1/C(s)P(s)]Y_{no}(s) - [1/C(s)]Y_{ni}(s)$. For good rejection of the input disturbance it is suggested to choose $C(s) = K_c Y_{ni}(s)$ for the controller dynamics, where $K_c \gg 1$. Appropriate output disturbance rejection can be reached by choosing the controller dynamics according to $C(s) = K_c Y_{no}(s)$, again with $K_c \gg 1$ (supposing that the amplitudes of the frequency function of the plant are not too high in the characteristic frequency range of the disturbance). To ensure good reference signal tracking and good disturbance rejection the controller has to contain the dynamics of both the reference and the disturbance signals. The following example demonstrates the effects of the designed controller to the behaviour of the control system.

Example 4.2 The transfer function of the plant is $P(s) = 1/(1 + 0.5s)^3$. Suppose the transfer function of the controller is $C(s) = 0.5(1 + 0.5s)/s$. Let us accelerate the reference signal tracking of the system with an appropriate prefilter F. Its gain is 1, and let it compensate the complex conjugate poles of the closed-loop control system, replacing them with two identical (real and faster) poles. Apply an $F(s) = (s^2 + 1.161s + 0.7044)/\left(0.7044(1 + 0.4s)^2\right)$ transfer function as the prefilter. Figure 4.15 shows the unit step responses of the different output signals in the closed-loop control circuit. It can be seen that the dynamic behavior of the control system is different for the reference signal and for the input disturbance. It can be

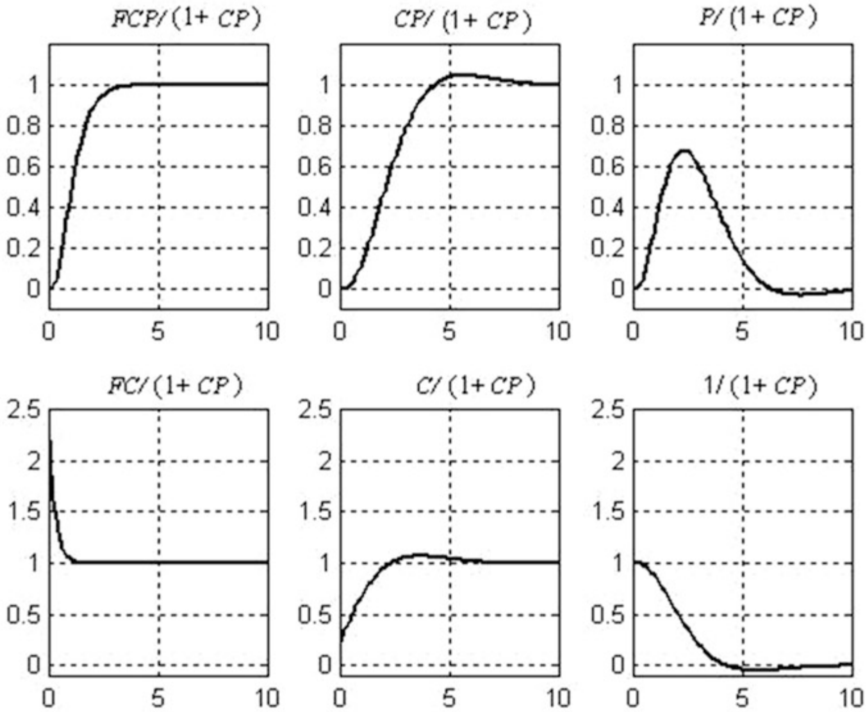


Fig. 4.15 Typical unit step responses of the closed-loop control system

observed that applying the filter accelerates the settling of the controlled output signal. The price paid for this is an overexcitation in the control signal. Figure 4.16 shows the frequency functions of the closed-loop control system. From the course of the frequency functions some evaluation of the time responses can also be derived. The frequency range where the disturbance rejection is efficient is also observable. For example, the third curve in the figure shows that the output will have its highest amplitude around frequency $\omega = 1$ for a sinusoidal input disturbance. From the sixth curve it can be concluded that the system attenuates the effect of the sinusoidal output disturbances up to the frequency $\omega = 1$, but beyond this frequency it tracks the disturbances.

From the second curve of Fig. 4.15 or from the equivalent left upper curve of Fig. 4.17 it can be seen that the control system tracks the unit step reference signal without steady state error. The controller contains an integrating element, whose pole is at the origin in the complex plane. Thus the controller contains the pole of the unit step signal (whose LAPLACE transform is $1/s$). Let us investigate the time evolution of the output signal with the given controller with prefilter $F = 1$ provided an exponential reference signal by $r(t) = \exp(-0.1t)$. The LAPLACE transform of the reference signal is $R(s) = 1/(s + 0.1)$. The reference signal and the output signal are shown on the right upper curve of Fig. 4.17. It can be seen that the

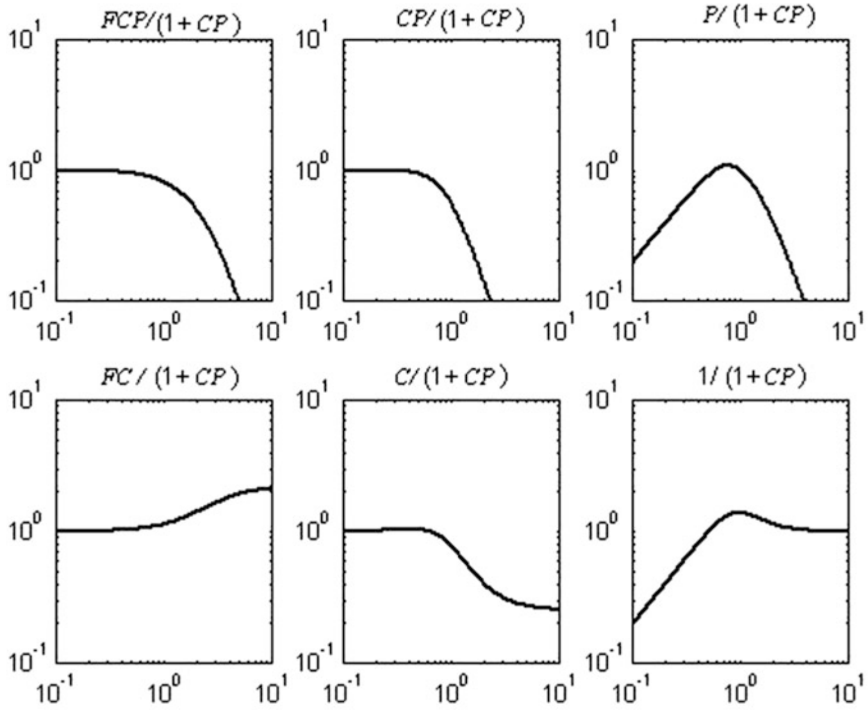


Fig. 4.16 Amplitude-frequency functions of the closed-loop control system

tracking is not accurate: after the transients decay, the output does not fit exactly the input signal. Let us change now the controller according to $C_1(s) = 4(1+0.5s)/(1+10s) = 0.4(1+0.5s)/(s+0.1)$. Now the pole of the controller is the same as the pole of the input signal. The right lower curve shows that after the transient period the output signal exactly tracks the input signal. But now the controller transfer function does not contain the pole of the unit step reference signal, therefore in the unit step response there will be a static deviation (left lower figure). ■

4.5 The Static Transfer Characteristics

If the closed-loop control system is stable, its steady state (or static in other words) properties can be determined on the basis of Eqs. (4.6)–(4.8) using the final value theorem of the LAPLACE transformation.

The signal transfer properties of closed-loop control circuits in steady state, i.e., the accuracy of reference signal tracking and of disturbance rejection in steady state depends on the so called type number and the loop gain of the system. The static

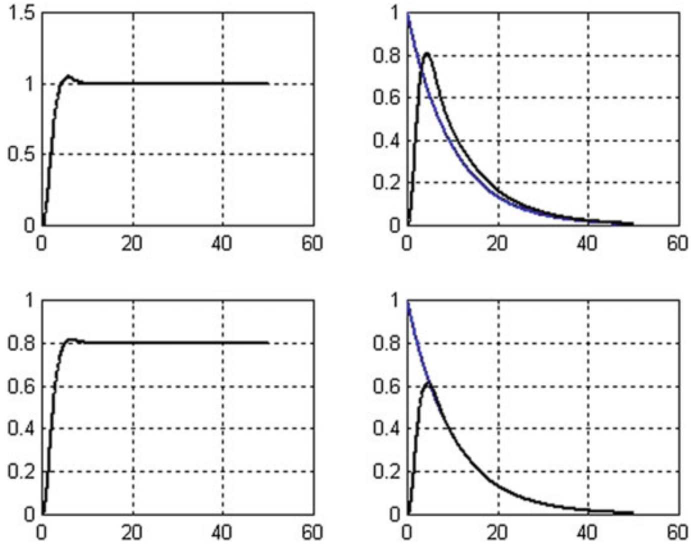


Fig. 4.17 Reference signal tracking is realized without steady-state error, if the controller contains the pole of the reference signal

accuracy also depends on the time evolution of the reference or the disturbance signal.

Let us suppose that $L(s) = C(s)P(s)$, the transfer function of the open-loop (the so-called loop transfer function) is given in its time constant form:

$$L(s) = C(s)P(s) = \frac{K}{s^i} \frac{\prod_{j=1}^{c}\left(1+s\tau_j\right)\prod_{j=1}^{d}\left(1+2\zeta_j\tau_{oj}s+s^2\tau_{oj}^2\right)}{\prod_{j=1}^{e}\left(1+sT_j\right)\prod\limits_{j=1}^{f}\left(1+2\xi_j T_{oj}s+s^2T_{oj}^2\right)} e^{-sT_\mathrm{d}} = \frac{K}{s^i}L_\mathrm{t}(s). \tag{4.10}$$

Here the variable i is the type number, indicating the number of the integrators in the loop (in practice its value can be 0, 1 or 2), K denotes the loop gain. $L_\mathrm{t}(s)$ represents the transfer function determining the transient response of the control circuit. Its important property is that it does not influence the steady state behavior, i.e., $L_\mathrm{t}(s=0) = 1$.

The overall transfer function between the error signal and the reference signal in the case where $F(s) = 1$ is

$$E(s) = \frac{1}{1+L(s)}R(s) = \frac{s^i}{s^i + KL_\mathrm{t}(s)}R(s). \tag{4.11}$$

The steady state value of the error signal is

$$\lim_{t \to \infty} e(t) = \lim_{s \to 0} sE(s). \tag{4.12}$$

Let us analyze the reference signal tracking properties of the closed-loop control system for unit step, unit ramp and parabolic input signals. The LAPLACE transforms of these reference signals are $R(s) = 1/s^j$, where $j = 1$ for the unit step, $j = 2$ for the unit ramp, and $j = 3$ for the parabolic reference input signal.

In case of a 0-type system, the steady state error is:

$$\begin{aligned}
&\text{for a unit step reference signal,} && \lim_{t \to \infty} e(t) = \lim_{s \to 0} s \frac{1}{s} \frac{1}{1 + KL_t(s)} = \frac{1}{1+K}; \\
&\text{for a unit ramp reference signal,} && \lim_{t \to \infty} e(t) = \lim_{s \to 0} s \frac{1}{s^2} \frac{1}{1 + KL_t(s)} = \infty; \\
&\text{for a parabolic reference signal,} && \lim_{t \to \infty} e(t) = \lim_{s \to 0} s \frac{1}{s^3} \frac{1}{1 + KL_t(s)} = \infty.
\end{aligned} \tag{4.13}$$

For a 1-type system, the steady state error is:

$$\begin{aligned}
&\text{for a unit step reference signal,} && \lim_{t \to \infty} e(t) = \lim_{s \to 0} s \frac{1}{s} \frac{s}{s + KL_t(s)} = 0; \\
&\text{for a unit ramp reference signal,} && \lim_{t \to \infty} e(t) = \lim_{s \to 0} s \frac{1}{s^2} \frac{s}{s + KL_t(s)} = \frac{1}{K}; \\
&\text{for a parabolic reference signal,} && \lim_{t \to \infty} e(t) = \lim_{s \to 0} s \frac{1}{s^3} \frac{s}{s + KL_t(s)} = \infty.
\end{aligned} \tag{4.14}$$

For a 2-type system, the steady state error is:

$$\begin{aligned}
&\text{for a unit step reference signal,} && \lim_{t \to \infty} e(t) = \lim_{s \to 0} s \frac{1}{s} \frac{s^2}{s^2 + KL_t(s)} = 0 \\
&\text{for a unit ramp reference signal,} && \lim_{t \to \infty} e(t) = \lim_{s \to 0} s \frac{1}{s^2} \frac{s^2}{s^2 + KL_t(s)} = 0 \\
&\text{for a parabolic reference signal,} && \lim_{t \to \infty} e(t) = \lim_{s \to 0} s \frac{1}{s^3} \frac{s^2}{s^2 + KL_t(s)} = \frac{1}{K}
\end{aligned} \tag{4.15}$$

In the following table, the values of the steady state errors are summarized.

Type number	$i = 0$	$i = 1$	$i = 2$
unit step reference signal, $j = 1$	$\frac{1}{1+K}$	0	0
unit ramp reference signal, $j = 2$	∞	$\frac{1}{K}$	0
parabolic reference signal, $j = 3$	∞	∞	$\frac{1}{K}$

A 0-type system tracks the step reference signal with steady state (static) error, whose value is less if the loop gain of the control circuit is higher (Fig. 4.18). But a high loop gain may cause an unstable behavior of the control system. A 0-type system is not able to track the ramp or the parabolic reference signals.

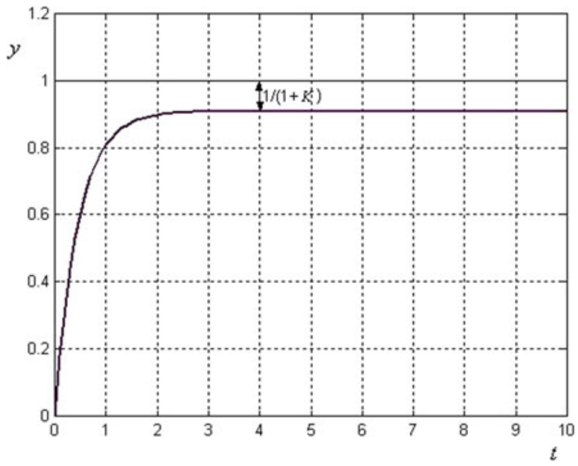


Fig. 4.18 A 0-type system tracks the unit step reference signal with steady state error

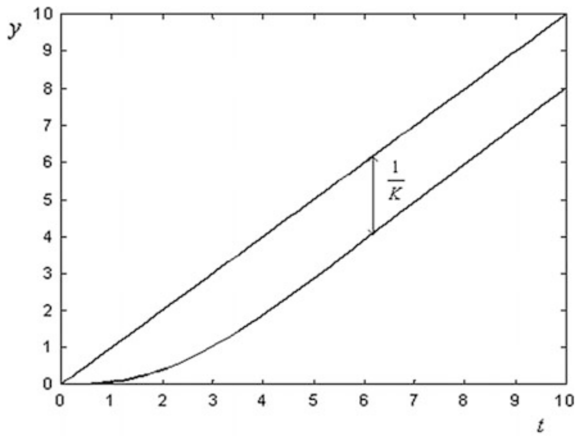


Fig. 4.19 A 1-type system tracks the ramp reference signal with steady state error

A 1-type control system containing one integrator tracks the step reference signal without steady state error. It can follow the ramp reference signal with a steady state error (Fig. 4.19). But it can not track the parabolic reference signal.

A 2-type system containing two integrators tracks the step and the ramp signals without steady state error (Fig. 4.20), and is able to follow the parabolic reference signal with a static error.

It can be seen, that coinciding with the previous statement related to the conditions of accurate reference signal tracking, the closed-loop control system is capable of tracking a reference signal whose Laplace transform contains poles at the origin of the complex plane without steady state error only if the loop transfer function contains as many poles at zero (integrators) as there are poles at zero of the Laplace transform of the reference signal. If the plant does not contain the required

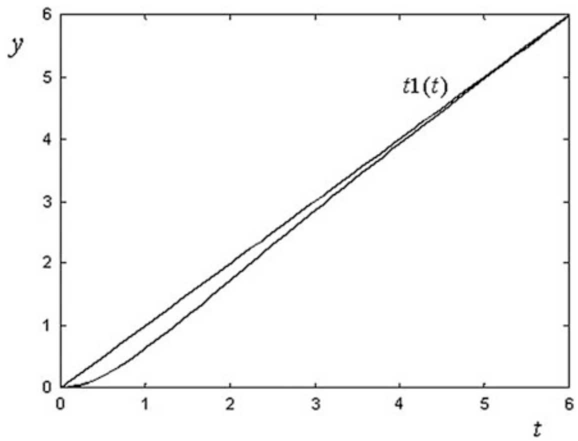


Fig. 4.20 A 2-type system tracks the ramp reference signal without static error

number of integrators ensuring the desired static accuracy, the integrators have to be put in the controller.

The effect of improving the static accuracy by inserting integrators in the control loop can be demonstrated by the following considerations. If the control system is of type 0, i.e. it is proportional, a constant signal value at its output can be maintained only by a constant input signal. Therefore it is necessary that also the error signal take a constant value. The property of the integrator is that its output reaches a constant value when its input finally becomes zero. If there is an integrator in the forward path of the closed-loop control circuit, then for a unit step reference signal the output signal will increase until the error signal—the input signal of the integrator—reaches zero. If the reference signal is a unit ramp, then at the output of the integrator a signal change with constant slope can be reached by a constant input signal, which means a constant error signal, i.e. a constant static deviation.

It can be seen that increasing the number of the integrators in the loop improves the static properties of the closed-loop control system. More specifically, increasing the loop gain reduces the static tracking error. But the number of integrators can not be increased to more than two, as this would lead to stability problems which could not be handled easily. Increasing the gain may also cause stability problems.

Static accuracy and stability are contradictory requirements. With controller design a satisfactory compromise has to be created to satisfy both requirements.

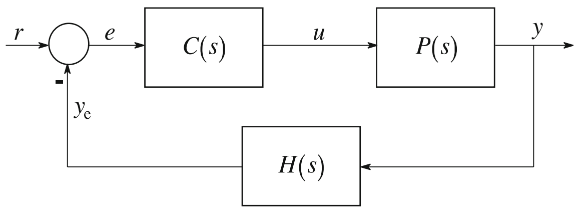


Fig. 4.21 Block diagram of a closed-loop control circuit

The evolution of the steady-state signal values in a closed-loop control circuit could also be demonstrated by a four-quarter-plane figure. On the four axes the error signal (e), the measured signal (y_e), the controlled signal (y) and the control signal (u) are indicated, respectively. Generally positive signal values are supposed. The block diagram is shown in Fig. 4.21. Generally the static characteristics of the plant and of the sensor are non-linear (but usually they are linearized in a vicinity of a given operating point). Stable behavior is supposed.

For a 0-type system, the four-quarter-plane curves are shown in Fig. 4.22. The right upper quarter represents the element creating the difference signal, where the location of the straight lines depends on the signal r. The static characteristic of the controller (left upper quarter) is generally linear, possibly saturating. The static characteristics of the plant (generally non-linear) is in the left lower quarter, here the effects of a disturbance and of parameter changes on the characteristics can be demonstrated. The right lower quarter shows the characteristic of the sensor, i.e. the measurement equipment, which sometimes is also non-linear. In the case of a 0-type system there is a steady state error ($e(\infty) \neq 0$).

In Fig. 4.22, which shows the four-quarter-plane static relations, it can be seen how the static states change if, e.g., the reference signal change. It can be investigated how the steady states change if $y(u)$, i.e., the static characteristic of the plant, changes as a consequence of parameter changes. (A similar four-quarter-plane representation was first introduced by Szilágyi.)

Figure 4.23 shows the static curves for a 1-type closed-loop control system. As now the static error for a step reference signal is zero, only the two lower quadrants of the plane are of interest. Now it would be sufficient to draw only these two, but for the sake of comparability the same coordinate system is given as before.

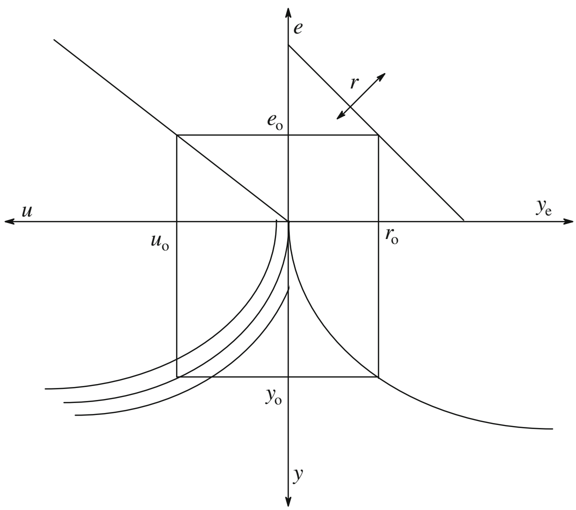


Fig. 4.22 Static characteristic curves of a 0-type closed-loop control system

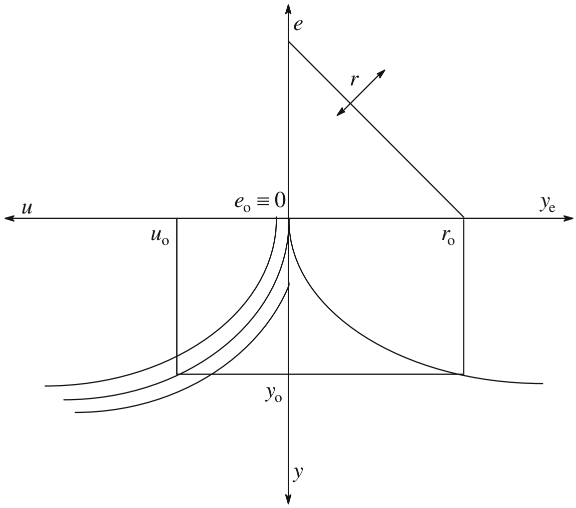


Fig. 4.23 Static characteristics curves of a 1-type closed-loop control system

4.6 Relationships Between Open- and Closed-Loop Frequency Characteristics

In a closed-loop control system $L = CP$ is called the loop transfer function (Fig. 4.24). The overall transfer function of a closed-loop realized by negative feedback (Fig. 4.25) calculated between the output signal and the reference signal is $T = CP/(1 + CP) = L/(1 + L)$, which is also called the complementary sensitivity function. Observe that $T = 1 - S = 1 - 1/(1 + L)$, where $S = 1/(1 + CP) = 1/(1 + L)$ is the sensitivity function. Regarding the frequency course of this function, approximate considerations can be given.

In the frequency range where

$$|L(j\omega)| \gg 1, \quad |T(j\omega)| \approx 1; \tag{4.16}$$

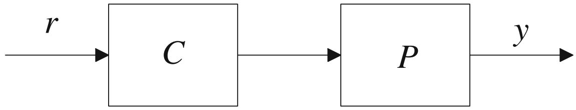


Fig. 4.24 Open-loop control

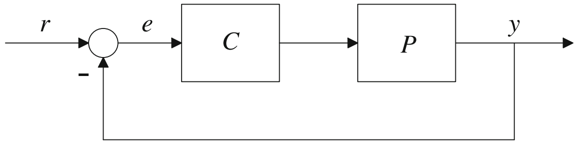


Fig. 4.25 Closed-loop control

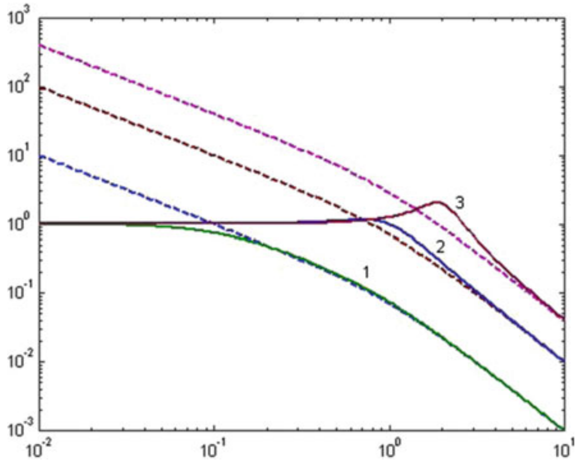


Fig. 4.26 Typical course of the amplitude-frequency functions of the open- and the closed-loop

while in the frequency range where

$$|L(j\omega)| \ll 1, \quad |T(j\omega)| \approx |L(j\omega)|. \tag{4.17}$$

The approximations are not valid in the vicinity of the cut-off frequency. Figure 4.26 shows typical amplitude-frequency curves of the open- and the closed-loop. The open-loop is a first order lag element serially connected to an integrating element. Curves 1, 2, 3 give the BODE amplitude-frequency diagrams of the open- and the closed-loop for three different loop gains. The highest gain is in the case of curve 3. It can be seen that the closed-loop diagrams approximate the value 1 up to the cut-off frequency of the open-loop, and then follow the course of the open-loop diagrams. For higher loop gains, the closed-loop curve shows an amplification in the vicinity of the cut-off frequency, which indicates the appearance of complex conjugate poles in the closed-loop transfer function and transients with decreasing oscillations in the unit step response. Figure 4.27 gives the unit step responses of the closed-loop system with the three different loop gains.

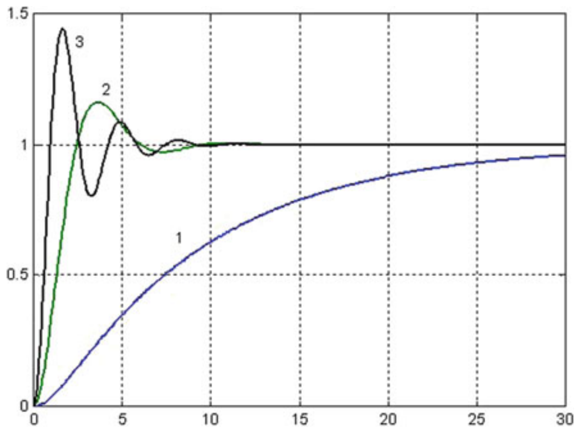


Fig. 4.27 Unit step responses of the closed-loop system

If there is no amplification in the amplitude diagram of the closed-loop, the closed-loop can be approximated by a first order lag element with unit gain and time constant reciprocal to the cut-off frequency ω_c: $T(s) \approx 1/(1 + s/\omega_c)$. The next time constant which changes the slope of the approximating amplitude curve to −40 dB/decade can be neglected in this case. The unit step response approximates the steady state exponentially, and approximately within 3 time constants reaches its steady state within 5% accuracy. Increasing the loop gain the slope of the curve around the cut-off frequency will be −40 dB/decade, and the time of decaying of the oscillations can be approximated by 10 times the time constant ($1/\omega_c$) of the second order oscillating element. Thus the settling time can be given by the following approximate relationship:

$$\frac{3}{\omega_c} < t_s < \frac{10}{\omega_c}. \tag{4.18}$$

To avoid oscillations a long section of slope −20 dB/decade has to be created around the cut-off frequency (before and after it) in the Bode amplitude-frequency diagram of the open-loop. To accelerate the system the cut-off frequency has to be set to higher values.

4.6.1 The $M - \alpha$ and $E - \beta$ Curves

For a deeper analysis of the relationship between the frequency functions of the open and the closed-loop systems, let us analyze the following considerations. The complementary sensitivity function of the closed-loop system is given by

$$T(s) = \frac{C(s)P(s)}{1 + C(s)P(s)} = \frac{L(s)}{1 + L(s)}. \tag{4.19}$$

In controller design, the relationship between the transfer function $T(s)$ of the closed-loop and the transfer function $L(s) = C(s)P(s)$ of the open-loop is taken into account. This relationship seems to be simple, but actually it means a conformal non-linear mapping from the $L(s)$ complex plane to the $T(s)$ complex plane. The complexity of this non-linear relationship is the reason why the controller can not always be designed unambiguously using simple methods.

For each point of the complex plane the mapping point (complex vector) according to relationship (4.19) can be determined. The absolute value of this vector is depicted on the vertical axis in Fig. 4.28. (The phase angle can also be visualized similarly.) Let us plot on the complex plane the Nyquist diagram of the open-loop (shown as a thick line in the figure). If the points of this Nyquist curve are projected to the three dimensional curve, the absolute values of the frequency function of the closed-loop are obtained. The Bode amplitude-frequency diagram of the closed-loop is visualized drawing these values versus the frequency.

The frequency function of the closed-loop can be given by its amplitude and phase angle:

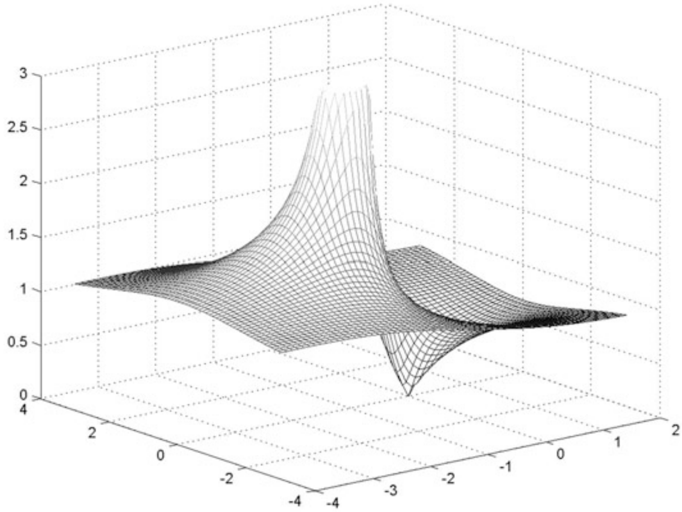


Fig. 4.28 The relationship between the amplitude diagrams of the open- and the closed-loop

$$T(j\,\omega) = \frac{L(j\,\omega)}{1 + L(j\,\omega)} = M(\omega)e^{j\,\alpha(\omega)}. \tag{4.20}$$

From Fig. 4.28 it can be seen that for high amplifications in the open-loop (in the horizontal plane, points which are far from the origin) the amplification of the closed-loop approximates the constant value 1. This relationship is seen also from the approximation

$$|T(j\,\omega)| = \left|\frac{L(j\,\omega)}{L(j\,\omega)+1}\right|_{|L|\gg 1} \approx 1. \tag{4.21}$$

As in control systems the amplification of the open-loop in the low frequency domain is generally high, the amplification of the closed-loop here is approximately 1. Similarly it can also be seen that for points with low amplification in the open-loop (in the horizontal plane points close to the origin) the corresponding closed-loop points are of low amplification values, too.

$$|T(j\,\omega)| = \left|\frac{L(j\,\omega)}{L(j\,\omega)+1}\right|_{|L|\ll 1} \approx |L(j\,\omega)|.$$

As the amplification of physical systems decreases at high frequencies, this relationship shows that at high frequencies the amplifications of the open- and the closed-loop are approximately the same, i.e. negative feedback at these frequencies does not change the open-loop.

It can also be seen that the curve has a singularity at the point $(-1+0j)$ of the complex plane, therefore for controller design the investigation of the neighborhood of this point will have a great importance. The closer we are to $(-1+0j)$, the higher the amplification of the closed-loop will be. When designing a control system, among the given quality specifications the prescribed value of the allowed overshoot is an important requirement. The overshoot in the step response of the closed-loop system is a time domain property, which is determined by the amplification of the amplitude in the frequency domain. Therefore it is important to investigate the location of the points in the complex plane where the closed-loop amplitudes $|T| = M$ are identical. The points of the frequency function of the closed-loop system where the amplitudes are identical are located on circles in the complex plane. This can be seen easily, solving the equation

$$M = \left|\frac{L(j\omega)}{1+L(j\omega)}\right| = \left|\frac{u+jv}{1+u+jv}\right| = \sqrt{\frac{u^2+v^2}{1+2u+u^2+v^2}}. \tag{4.22}$$

The equation of the curves belonging to a constant amplitude M are obtained by rearranging the above equation as

$$\left(u - \frac{M^2}{1-M^2}\right)^2 + v^2 = \left(\frac{M}{1-M^2}\right)^2. \tag{4.23}$$

This is the equation of a circle with radius r and center point (u_o, v_o), where

$$r = \left|\frac{M}{1-M^2}\right| \quad , \quad u_o = \frac{M^2}{1-M^2} \quad \text{and} \quad v_o = 0. \tag{4.24}$$

The circles belonging to different constant M amplitude values of the closed-loop are shown in Fig. 4.29.

The $M = 1$ constant curve is a vertical line at $u = -0.5$. For $M > 1$ the curves are to the left, and for $M < 1$ they are to the right of this line. If M tends to infinity, the curves shrink to point $(-1+0j)$, and if M tends to zero, the circle will be of infinitesimal radius around the origin.

Similarly to the circles belonging to constant M values, curves belonging to constant α values can also be given (4.20), which are also circles. These circles (both for constant M and constant α values) are called Archimedes circles. The two curve systems together are called $M - \alpha$ curves.

If the Nyquist diagram of the open-loop is plotted in the complex plane where the constant M curves are also drawn, the amplitude-frequency diagram of the closed-loop can be obtained by reading the appropriate M values corresponding to the individual points of the Nyquist diagram. The highest amplitude of the closed-loop is determined by how close the Nyquist diagram of the open-loop approaches $(-1+0j)$. The highest value of M will be determined by the circle-tangential to the Nyquist diagram.

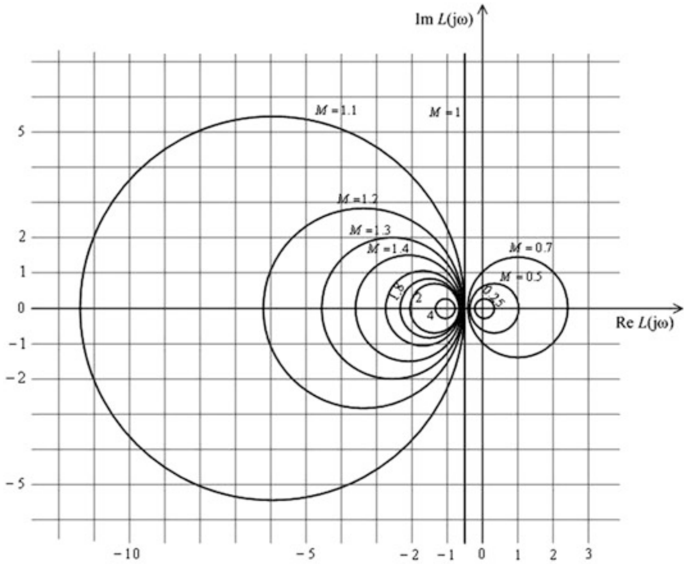


Fig. 4.29 Constant M curves

Some characteristic features of the M curves are shown on the complex plane in Fig. 4.30. The frequency ω_b, where the frequency function $L(j\omega)$ intercepts the circle of $M = 1/\sqrt{2}$, gives the so called bandwidth of the closed-loop system. In the figure the cut-off frequency ω_c and the frequency ω_a where $M = \sqrt{2}$ are also

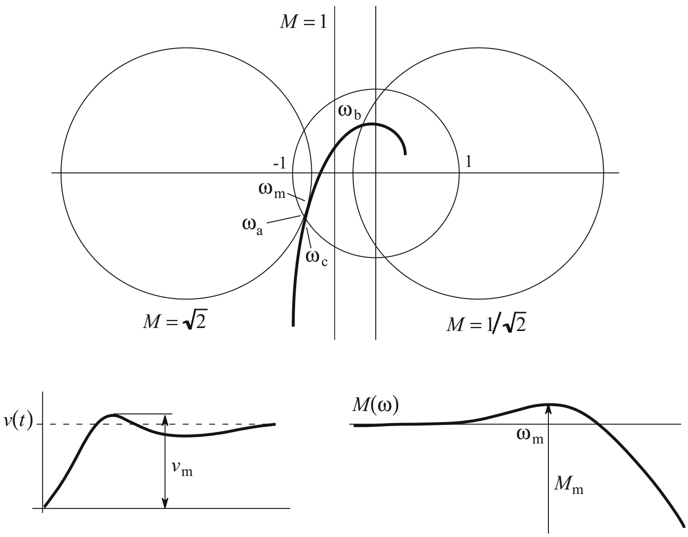


Fig. 4.30 Some characteristic features of the M curves in the complex plane: shape of the unit step response and the amplitude-frequency characteristics

indicated. At the frequency ω_m, the closed-loop system has its maximal amplitude value. The maximal value is the highest value of M belonging to the circle-tangential to the Nyquist curve of the open-loop. The amplitude-frequency diagram of the closed-loop has amplification only if the Nyquist diagram of the open-loop intersects the vertical line corresponding to $M = 1$, thus there is a frequency range where the Nyquist curve is to the left of this line. At the intersection frequency $|T(j\omega)| = 1$.

Approximate relationships can be given between the maximal amplification $M_m = M_{max}$ of the closed-loop amplitude-frequency curve and the maximum value v_m of the step response (Fig. 4.30).

$$\begin{array}{ll} M_m \geq 1.5 & v_m \leq M_m - 0.1 \\ 1.25 \leq M_m \leq 1.5 & v_m \approx M_m \\ M_m \leq 1.25 & v_m < M_m \end{array} \tag{4.25}$$

To avoid oscillations and a big overshoot in the time response, high amplification is not allowed in the amplitude-frequency diagram of the closed-loop. The ideal and the real frequency curves of the closed-loop are shown in Fig. 4.31. Here ω_{cc} is the cut-off frequency of the closed-loop.

Similarly to the $M - \alpha$ curves of the $T(j\omega)$ frequency function, the so called $E - \beta$ curves can be constructed based on the overall error transfer function $S(j\omega)$ (sensitivity function).

$$S(j\omega) = \frac{1}{1 + L(j\omega)} = E(\omega)e^{j\beta(\omega)}. \tag{4.26}$$

Drawing the curves belonging to constant values of E is very simple, as

$$E = |S(j\omega)| = \frac{1}{|1 + L(j\omega)|} \tag{4.27}$$

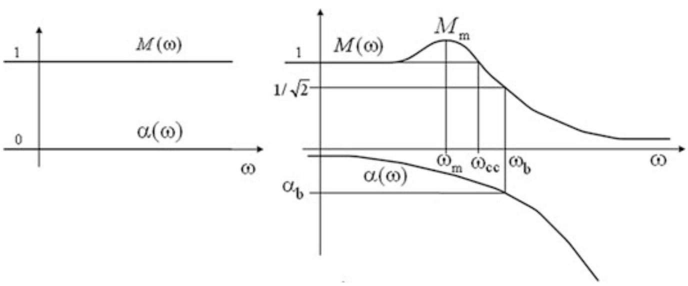


Fig. 4.31 Ideal and real frequency function of the closed-loop system

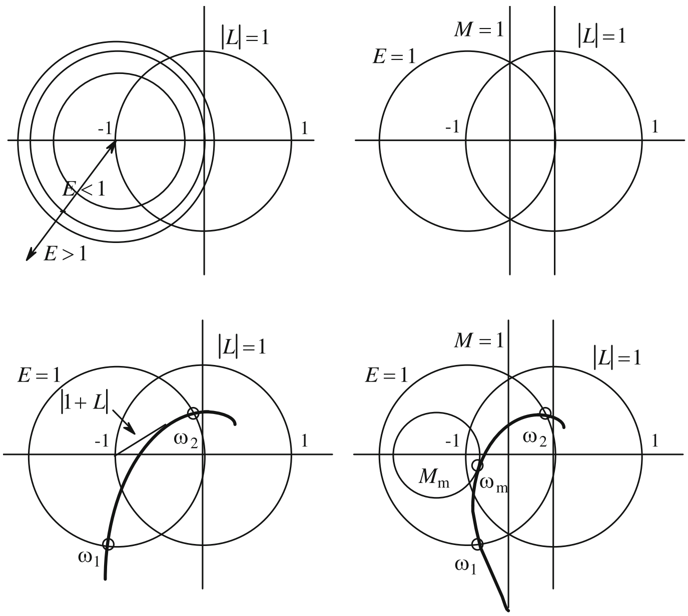


Fig. 4.32 Curves corresponding to $M = 1$, $E = 1$ and $|L| = 1$

and the $|1 + L(j\omega)|$ in the denominator is equal to the distance from the point $(-1 + j0)$. These curves are concentric circles around $(-1 + 0j)$ with radius $1/E$. The curve belonging to $E = 1$ has special significance.

In Fig. 4.32 the curves belonging to $M = 1$, $E = 1$, $|L| = 1$ also the distance $|1 + L|$ are indicated. It is shown how to determine the maximal value M_m with the open-loop Nyquist diagram. The interception points ω_1 and ω_2 of the $L(j\omega)$ characteristics and the $E = 1$ circle indicate the range where $|S(j\omega)| < 1$.

4.7 The Sensitivity of a Closed Control Loop to Parameter Uncertainties

The parameters of a process are never known quite accurately. Also, the process may change over time. The environment of the process may change and as a consequence the parameters of the process may also change within a given range. Negative feedback decreases the sensitivity of the system to parameter changes. In controller design it is advisable to take the possible parameter changes into consideration. The behavior of the control system has to be acceptable not only for the nominal parameter values, but throughout the whole possible range of the parameter changes.

Let us analyze the behavior of the system if the transfer function of the process changes from its nominal value $P_o(s)$ to $P(s) = P_o(s) + \Delta P(s)$. The overall transfer function of the open-loop is $L = CP$, whose small change is

$$\Delta L = \frac{\partial L}{\partial P}\Delta P = C\Delta P. \tag{4.28}$$

The relative change is expressed by

$$\frac{\Delta L}{L} = \frac{C\Delta P}{CP} = \frac{\Delta P}{P}. \tag{4.29}$$

The overall transfer function of the closed-loop realized by negative feedback (Fig. 4.25) is

$$T = \frac{CP}{1+CP}, \tag{4.30}$$

whose small change is

$$\Delta T = \frac{\partial T}{\partial P}\Delta P = \frac{C}{(1+CP)^2}\Delta P. \tag{4.31}$$

The relative value of the change is

$$\frac{\Delta T}{T} = \frac{1}{1+CP}\frac{\Delta P}{P} = S\frac{\Delta P}{P}, \tag{4.32}$$

where S is the sensitivity function of the closed-loop:

$$S = \frac{\Delta T/T}{\Delta P/P} = \frac{1}{1+CP}. \tag{4.33}$$

The sensitivity function shows how much a relative change of the process $(\Delta P/P)$ influences the relative change of the resulting transfer function $(\Delta T/T)$. In the frequency range where $|L(j\omega)| \to \infty$, the sensitivity function takes small values, thus even big parameter changes in the process have a small effect on the resulting closed-loop transfer function, and also on the output signal of the closed-loop.

For an infinitesimally small change ($\Delta P \to 0$):

$$\frac{\partial T}{T} = S\frac{\partial P}{P}, \tag{4.34}$$

whence

$$S = \frac{\partial T/T}{\partial P/P} = \frac{\partial \ln T}{\partial \ln P}. \tag{4.35}$$

The resulting transfer function T of the closed-loop is also called the complementary sensitivity function, as the following relationship holds:

$$S + T = 1. \tag{4.36}$$

Typical amplitude-frequency curves of the loop transfer function L, the sensitivity function S, and the complementary sensitivity function T are shown in Fig. 4.33.

Let us consider now the sensitivity of the control system with respect to parameter changes in the feedback element (Fig. 4.34). This sensitivity function can be defined by the following relationship:

$$S_{\mathrm{H}} = \frac{\Delta T/T}{\Delta H/H}. \tag{4.37}$$

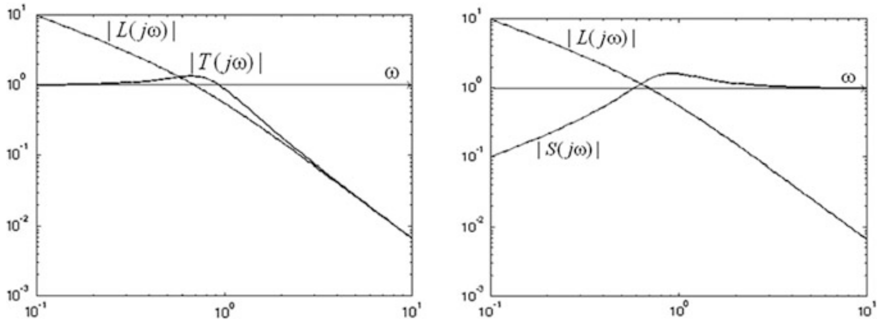
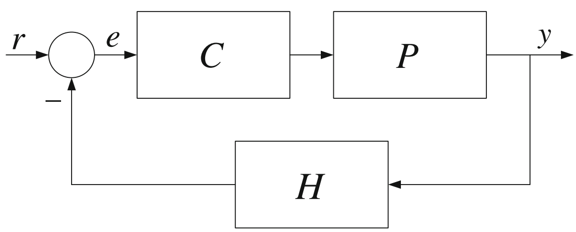


Fig. 4.33 Typical amplitude-frequency curves of the loop, the sensitivity function and the complementary sensitivity function

Fig. 4.34 Feedback control circuit, the sensor has dynamic characteristics



Now

$$T = \frac{CP}{1+CPH}, \quad \text{thus} \quad \Delta T = \frac{\partial T}{\partial H}\Delta H = -\frac{(CP)^2}{(1+CPH)^2}\Delta H, \tag{4.38}$$

and

$$\frac{\Delta T}{T} = S_{\mathrm{H}}\frac{\Delta H}{H} = -\frac{CPH}{1+CPH}\frac{\Delta H}{H} = -\frac{L}{1+L}\frac{\Delta H}{H}. \tag{4.39}$$

As $S_{\mathrm{H}} = -L/(1+L) = -T$ has to take approximately the value of 1 in a wide frequency range to ensure good reference signal tracking, the parameter changes in the feedback element may significantly influence the output signal. Therefore it is required to measure the output signal very accurately, or to realize unity feedback.

Formulas (4.6)–(4.8) giving the relationships between the input and the output signals can also be given by the sensitivity functions.

$$Y(s) = F(s)T(s)R(s) + S(s)Y_{\mathrm{no}}(s) + P(s)S(s)Y_{\mathrm{ni}}(s) - T(s)Y_{\mathrm{z}}(s) \tag{4.40}$$

$$E(s) = F(s)S(s)R(s) - S(s)Y_{\mathrm{no}}(s) - P(s)S(s)Y_{\mathrm{ni}}(s) - S(s)Y_{\mathrm{z}}(s) \tag{4.41}$$

$$U(s) = F(s)C(s)S(s)R(s) - C(s)S(s)Y_{\mathrm{no}}(s) - T(s)Y_{\mathrm{ni}}(s) - C(s)S(s)Y_{\mathrm{z}}(s) \tag{4.42}$$

So with the sensitivity functions, not only the effects of parameter changes can be investigated, but also the signal transfer properties of the control system can be analyzed.

4.8 Requirements for Closed-Loop Control Design

A closed-loop control system has to meet prescribed quality specifications. These specifications depend on the control aims, on the technology of the considered process and also on the process itself.

In a rolling-mill, e.g., the uniform thickness of the steel sheet has to be ensured with high accuracy. The aim of the utilization of the steel sheet will also influence the desired accuracy. In a heat treatment process the temperature has to be set according to a given program. In the treated material undesirable alterations should not happen. This requirement influences the prescribed accuracy of the reference signal tracking. The accuracy of directing an airplane into a path and then tracking the path is important to reach the destination station while ensuring the avoidance of other airplanes. The prescription of the required settling time is also important. This requirement has to consider the dynamics of the process. In case of a very slow process, a big acceleration can not be expected, as this would require too high, practically unrealizable manipulating input signals. The prescriptions should be tailored to the opportunities. The prescriptions consider both the static and the dynamic properties of the closed-loop control system.

The requirements set for a closed-loop control system are:

– stability
– appropriate static accuracy for reference signal tracking and disturbance rejection
– attenuation of the effect of measurement noise
– insensitivity to parameter changes
– prescribed dynamic (transient) behavior
– consideration of the restrictions stemming from the practicality of the realization.

A linear closed-loop control system is stable, its steady state is achieved, if the roots of the characteristic equation are on the left side of the complex plane (see Sects. 4.2 and 4.5).

Static accuracy of the control system for typical input signals (step, ramp, parabolic input) is determined by the number of the integrators in the open-loop (Sect. 4.5).

Disturbance rejection, attenuation of measurement noise, and the effects of parameter changes can be investigated by the sensitivity functions (Sects. 4.6 and 4.7).

The prescribed dynamic behavior is generally given by the characteristic parameters of the unit step response $v(t)$ of the closed-loop system (Fig. 4.35).

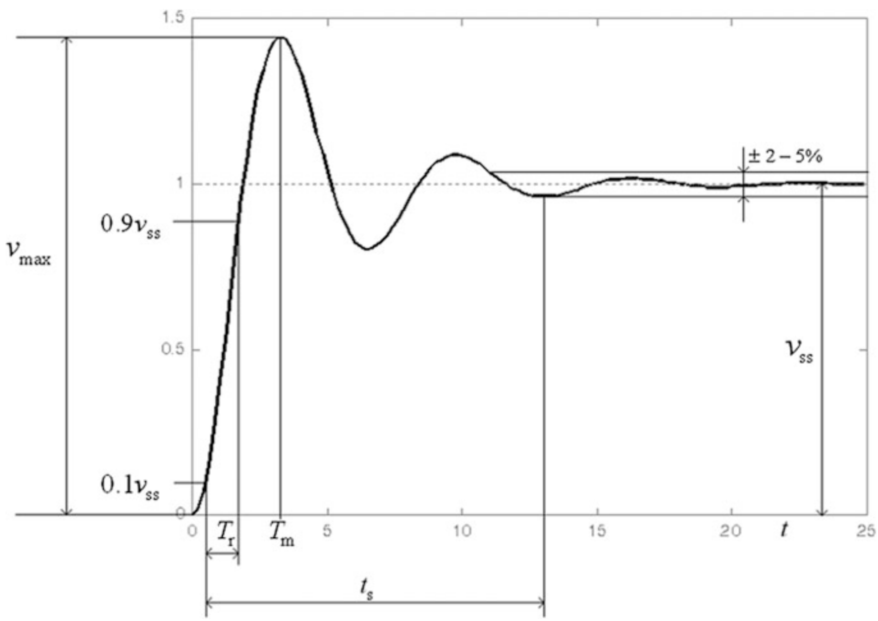


Fig. 4.35 Dynamic specifications of a control system

The static error for a unit step input is

$$1 - v_{ss}. \tag{4.43}$$

The overshoot of the unit step response is

$$\sigma = \frac{v_{max} - v_{ss}}{v_{ss}} 100\%. \tag{4.44}$$

The settling time t_s is the time when the unit step response of the closed-loop system reaches its steady state within $\pm(2-5)\%$ accuracy.

During the rise time T_r the step response starting from 10% reaches 90% of its steady state value. The time of reaching the maximum value is denoted by T_m.

In control techniques the main point of controller design is to find an acceptable compromise between a large overshoot and a long settling time. This compromise can be formulated on the one hand by prescribing the distance of the loop frequency function from the point $(-1+0j)$, which characterizes the stability limit (see Sect. 5.6). On the other hand, a quality index can be formulated, which can be the minimum (optimum) value of an integral criterion. This optimum value indicates a balance between the two extreme transients. In this case the quality of the control performance is evaluated on the basis of an integral of the error signal $e(t) = v(\infty) - v(t)$. The controller parameters are chosen to reach the minimum of this error integral.

The formulas for the different criteria involving integrals are as follows:

$$I_1 = \int_0^\infty e(t)\mathrm{d}t \quad \text{linear control error area} \tag{4.45}$$

(it can be applied only to aperiodic systems, it can be evaluated analytically)

$$I_2 = \int_0^\infty e^2(t)\mathrm{d}t \quad \text{quadratic control error area} \tag{4.46}$$

(it can be calculated analytically)

$$I_3 = \int_0^\infty |e(t)|\mathrm{d}t = IAE \quad \text{Integral of Absolute value Error} \tag{4.47}$$

$$I_4 = \int_0^\infty t|e(t)|\mathrm{d}t = ITAE \quad \text{Integral of Time multiplied by Absolute value Error} \tag{4.48}$$

I_3 and I_4 can be evaluated only by simulation.

Linear control error area.

With simple considerations the following relationship can be obtained:

$$I_1 = \lim_{t\to\infty} \int_0^t e(\tau)\mathrm{d}\tau = \lim_{s\to 0} s \frac{E(s)}{s} = E(0), \tag{4.49}$$

where $E(s) = \mathcal{L}\{e(t)\}$ is the Laplace transform of the error signal. For an aperiodic control system given by the transfer function $T(s)$ where $T(0) = A$ (see Fig. 4.36)

$$T(s) = A \frac{\prod_{k=1}^{m} (1 + s\tau_k)}{\prod_{j=1}^{n} (1 + sT_j)} = AT'(s), \tag{4.50}$$

let us calculate the linear control error area. According to (4.49),

$$\begin{aligned} I_1 &= \int_0^\infty (A - v(t))\ \mathrm{d}t = \left[(A - T(s))\frac{1}{s}\right]_{s=0} = A\left[\frac{1 - T'(s)}{s}\right]_{s=0} \\ &= A\left[\frac{\prod_{j=1}^{n}(1 + sT_j) - \prod_{k=1}^{m}(1 + s\tau_k)}{s\prod_{j=1}^{n}(1 + sT_j)}\right]_{s=0} = A\left(\sum_{j=1}^{n} T_j - \sum_{k=1}^{m}\tau_k\right) \end{aligned} \tag{4.51}$$

The time constants in the denominator of the transfer function increase the linear control error area, but the time constants in its numerator decrease it. Thus by introducing zeros, the system can be accelerated.

For aperiodic transients an equivalent dead time T_e can be defined, which is the dead time of the step function of amplitude A measured from time point $t = 0$, whose linear control error area is equal to the control error area of the step response of the considered transfer function.

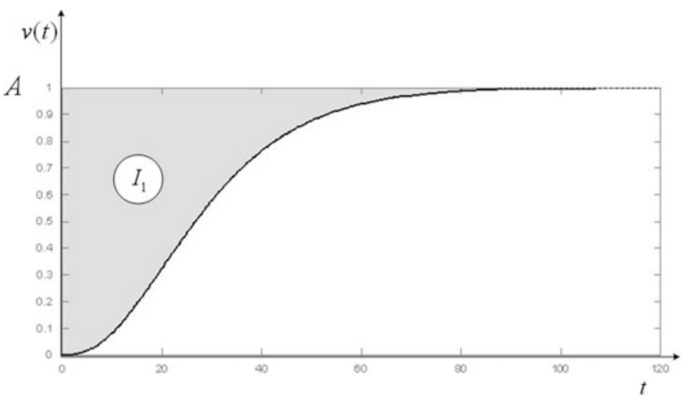


Fig. 4.36 Linear control area of an aperiodic process

$$T_e = \frac{I_1}{A} = \sum_{j=1}^{n} T_j - \sum_{k=1}^{m} \tau_k. \tag{4.52}$$

Quadratic control error area

The quadratic control error area can be evaluated also in the frequency domain using the PARSEVAL theorem.

$$I_2 = \int_0^{\infty} e^2(t)\mathrm{d}t = \frac{1}{2\pi j} \int_{-\infty}^{\infty} E(-s)E(s)\mathrm{d}s = \frac{1}{\pi} \int_0^{\infty} |E(j\omega)|^2 \mathrm{d}\omega. \tag{4.53}$$

The quadratic control error area can be calculated analytically. For lower degree cases for strictly proper LAPLACE transforms of the error signal ($m < n$) of the form

$$E(s) = \frac{\sum_{i=0}^{m} c_i s^i}{\sum_{i=0}^{n} d_i s^i} \tag{4.54}$$

calculation formulae have been derived for evaluation of the I_2 integral for a given degree and for given c_i and d_i parameters. A general formula in algorithmic form can also be given, which provides a special, not too complex recursive algorithm. It should be mentioned that minimizing the quadratic error area as a function of a controller parameter generally results in a flat minimum. Unfortunately the optimal transient generally gives a quite high overshoot (20–25%), so this optimal controller can not be used in high quality control systems.

Absolute value criteria

It is difficult to evaluate a criterion using the absolute value of the error. Instead of analytical calculation the minimum can rather be determined by simulation or with searching optimization methods. The minimum of the cost function is generally sharp. The *Integral of Time multiplied Absolute value Error* (*ITAE*) criterion punishes the error values at the beginning of the time scale less than those occurring at later time points. The optimum (minimum) of this criterion provides beautiful transients with $\sim 5\%$ overshoot.

4.9 Improving the Disturbance Elimination Properties of the Closed-Loop

An adequately designed closed-loop control system ensures good reference signal tracking and also the rejection of the effects of input and output disturbances. If along the path from the disturbance to the output there are signal components with large time constants, then the disturbance rejection will be slow. Of course, in controller design considerations related to disturbance rejection have to be also taken into account.

Disturbance rejection can be improved if not only the effects caused by the disturbance in the output signal are utilized for disturbance rejection, but possibly some internal measurable signals are also used in which the effect of the disturbance appears already earlier than in the output signal. Utilizing the available information in the control circuit, better, deliberate decisions can be made, and thus the quality of the control system can be improved.

4.9.1 Disturbance Elimination Scheme (Feedforward)

If the disturbance is measurable, the quality of the control system, especially its disturbance rejection properties, can be significantly improved by letting it drive a feedforward. Based on the measured value of the disturbance it is possible to execute actions to reject it before its effect would appear in the controlled variable. The block diagram of feedforward control is shown in Fig. 4.37. With appropriate design of the feedforward controller $C_{\mathrm{n}}(s)$ the effect of the disturbance can be significantly decreased or even totally compensated. The disturbance acts on the output through two paths. The resulting transfer function between the output and the disturbance is

$$\frac{Y(s)}{Y_{\mathrm{n}}(s)} = \frac{P_{\mathrm{n}}(s) + C_{\mathrm{n}}(s)P(s)}{1 + C(s)P(s)}. \tag{4.55}$$

The effect of the disturbance will not appear in the output signal if the numerator of the above expression is zero, i.e. if

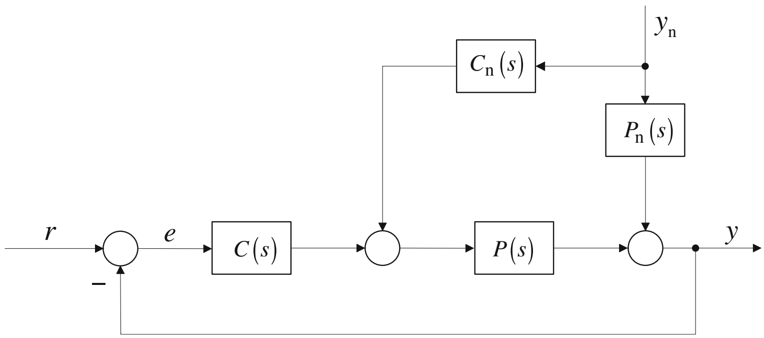


Fig. 4.37 Block diagram of feedforward control

$$P_{\mathrm{n}}(s) + C_{\mathrm{n}}(s)P(s) = 0. \tag{4.56}$$

Hence the transfer function of the feedforward compensator is:

$$C_{\mathrm{n}}(s) = -\frac{P_{\mathrm{n}}(s)}{P(s)}. \tag{4.57}$$

If this transfer function is realizable (i.e. if the degree of its numerator is not higher than the degree of its denominator and furthermore $P(s)$ does not contain dead time), the effect of the feedforward is perfect: the effect of the disturbance does not appear at all in the output signal. If $C_{\mathrm{n}}(s)$ is non-realizable, its transfer function has to be approximated by the best realizable controller.

Feedforward supplements the closed-loop control circuit with an open-loop path. The efficiency of the feedforward compensation depends on how accurately the effect of the disturbance on the output signal is known, and how much it is possible to compensate it with the available manipulations.

As an example let us consider the control scheme of a belt dryer furnace shown in Fig. 4.38. In the electrically heated furnace the material to be dried goes through the conveyor G driven by the motor M. The controlled signal is the moisture content of the material leaving the furnace. At a given conveyor speed the material abides in the furnace for a given time. The manipulated variable is the heating power, which can be changed by a voltage u across the resistance R. The humidity of the material leaving the furnace is measured. It is compared to the reference signal. In case of deviation, the heating power is modified through a *PI* controller. (A *PI* controller consists of a proportional (*P*) and an integrating (*I*) element connected in parallel, see Chap. 8). The main disturbance source is the change of the humidity of the incoming material. Time is needed to eliminate the effect of the disturbance. The control system comes into operation only after the effect of the disturbance has been detected at the output. Thus for a certain time the humidity of the outcoming material will differ from its desired value. If the humidity of the incoming material is measurable, then based on this measured value the heating power could be immediately set to a value which on the basis of a priori knowledge expectably would be needed to ensure the prescribed humidity value through the P part of the controller. (The integrating part of the controller can not be included in the feedforward path, as its output can not reach a finite steady state because of the constant input signal.) The feedforward part of the controller is denoted by the dashed line in Fig. 4.38. Then the closed-loop control circuit has to eliminate only the error component resulting from the inaccuracy of the a priori knowledge.

4.9.2 Cascade Control Schemes

Several times the processes can be separated into serially connected parts, and besides the output signal the intermediate signals can also be measured. Figure 4.39 shows the block diagram of a process which consists of two serially connected

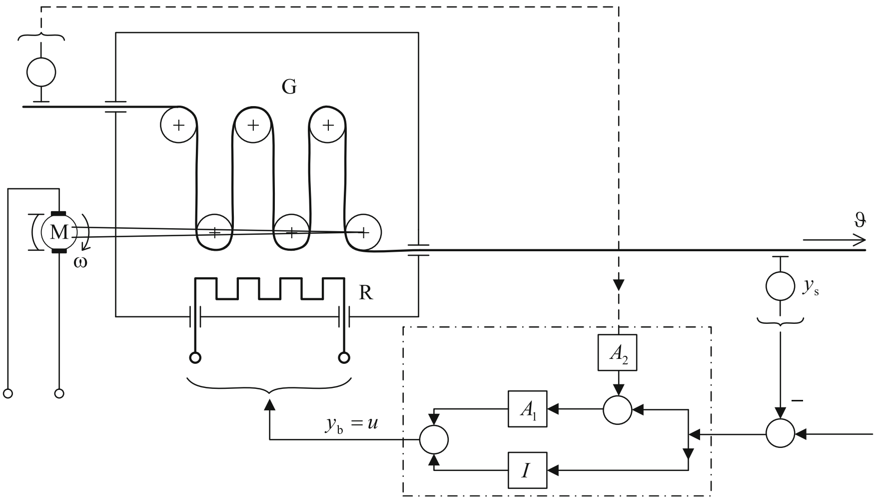


Fig. 4.38 Control scheme of a belt dryer furnace with feedforward

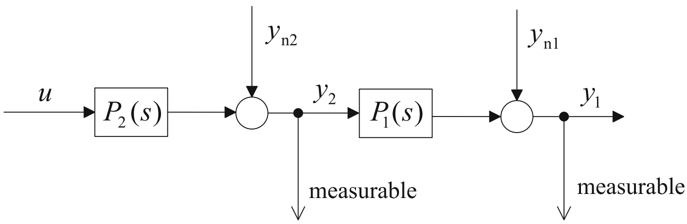


Fig. 4.39 A process which can be separated into two serially connected parts

parts. The disturbances may act on the output or between the two parts of the process. It is supposed that the disturbances themselves are not measurable.

The block diagram of the conventional feedback control is shown in Fig. 4.40. The closed-loop control system is able to track the reference signal and also to reject the effect of the disturbances. To activate the disturbance rejection it is necessary that the effect of the disturbance should appear at the output. Then an

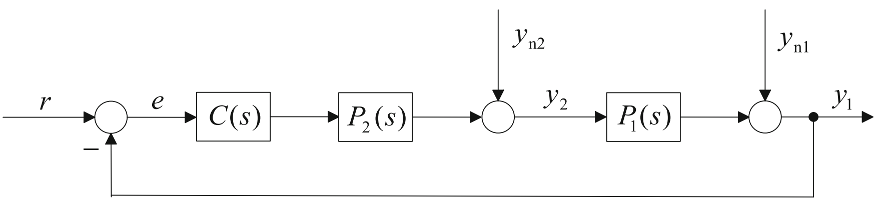


Fig. 4.40 Block diagram of feedback control

error signal appears in the closed-loop which activates the control circuit to eliminate the effect of the disturbance. If the $P_1(s)$ part of the process contains the larger time constants, the rejection of the disturbance y_{n2} acting between the two parts of the process will be slow.

It is worthwhile to create an inner loop using the measurable y_2 signal, which is capable of supporting a fast rejection of the inner disturbance. As the effect of the inner disturbance appears sooner in signal y_2 than in the output y_1, the inner loop can rather quickly decrease the effect of the inner disturbance. The outer loop ensures good reference signal tracking, the rejection of the output disturbance and further attenuation of the effect of the inner disturbance which has been already decreased by the inner loop. The block diagram of the control circuit with two loops —called a cascade control—is shown in Fig. 4.41. The advantage of cascade control compared to a single-loop feedback control is manifested if part $P_1(s)$ of the plant contains the large time constants and/or dead time, while part $P_2(s)$ contains the smaller time constants. The controller of the inner loop $C_2(s)$, is designed for fast performance of the internal loop, thus the inner loop will quickly reject the internal disturbance. With the controller $C_1(s)$ of the outer loop, good reference signal tracking and rejection of the external disturbance is to be ensured. The inner controller could be of structure *P* or *PD*. In the inner loop the feedback provides acceleration, thus because of the smaller time constants the compensation of the outer loop will be easier. The controller in the outer loop which ensures the quality specifications could be of structure *PI* or *PID*. (A *PID* controller consists of parallel connected proportional (*P*), integrating (*I*) and differentiating (*D*) elements, see Chap. 8.)

In some applications it is expedient to put a saturation after the external controller. As the output of the external controller provides the reference signal of the internal loop, by restricting its value, the internal signal y_2 can also be kept within prescribed limits.

Of course if the process can be separated into more than two components, where the internal signals are measurable, a cascade control can be realized with several nested control circuits.

Cascade control is applied generally in the speed or position control of electrical drives, where the output variable is the speed or the position, and the internal variable is the current. In this case the aim of cascade control is mainly the restriction of the armature current. Namely, the current may reach very high values when starting, breaking or loading the motor, while the speed is developed more slowly because of the mechanical inertia of the system. Thus it is not enough to feed back only the speed, the current also

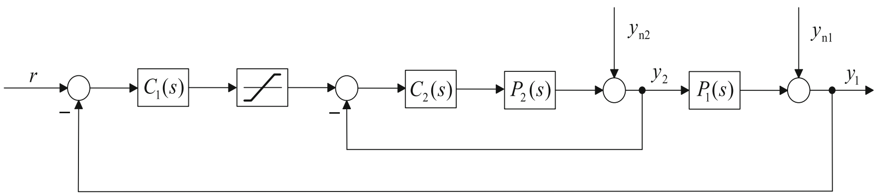


Fig. 4.41 Block diagram of cascade control

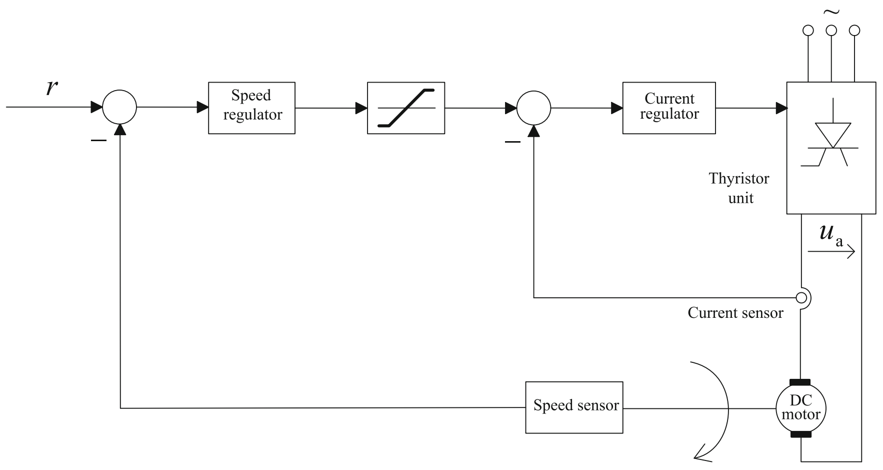


Fig. 4.42 Cascade control of a DC motor

has to be observed, and its value has to be kept within the allowed range. The cascade control of a DC motor is shown schematically in Fig. 4.42.

Figure 4.43 shows a cascade control solution for room temperature control. The controlled variable is the ϑ temperature of room T, which is set to the required value by the air blown across the steam heated heat exchanger H. The manipulated variable is the steam blowing through the heat exchanger, which is set by valve B. The main disturbance is the pressure of the steam, as the amount of the steam, i.e. the heating power entering the heat exchanger H depends on the pressure in a given valve position. For the cascade stage the internal controlled variable could be the temperature ϑ_k of the steam coming out of the heat exchanger, as the effect of the change of the heating power is observed sooner in ϑ_k than in the room temperature ϑ.

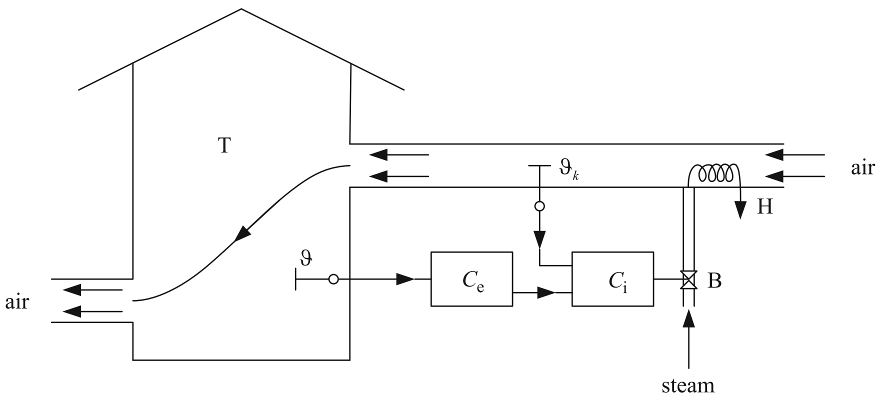


Fig. 4.43 Cascade solution for room temperature control

In a single loop closed-loop control circuit only the output signal is fed back. In cascade control, besides the output signal, one or more measurable internal signals are also fed back, thus improving the quality of the control. The control system will be faster, and could reject the internal disturbances more effectively. These internal variables generally are the state variables of the system.

In a system, the internal variables, the so called state variables, determine the dynamic behavior of the system. Their instantaneous values depend on the previous moves of the input signal. With the knowledge of the actual values of the state variables and the input signal, the states of the system and the output signal at the next time point can be determined.

When building a closed-loop control system not only the measurement of a single output signal or of some additional inner signals considered in cascade loops is important, but it is essential to measure and feed back all the state variables (in a system described by a differential equation of order n, their number is n). This control concept is called state feedback, which can be considered as a generalization of the cascade control concept. Chapter 10 deals with state feedback control in detail.

4.10 Compensation by Feedback Blocks

If some internal signals of a process are measurable, applying feedback on them the performance of the control system can favorably be influenced. The block diagram of feedback compensation is shown in Fig. 4.44. The equivalent series compensation can be determined in a straightforward way. The advantage of feedback compensation is that besides modifying the performance of the closed-loop system it linearizes the relationship between the output and the input signals of the internal loop and considerably decreases the effect of parameter changes. The internal loop is also effective in rejecting internal disturbances. Compensation by a feedback block may also show an advantageous behavior when the control signal is saturated.

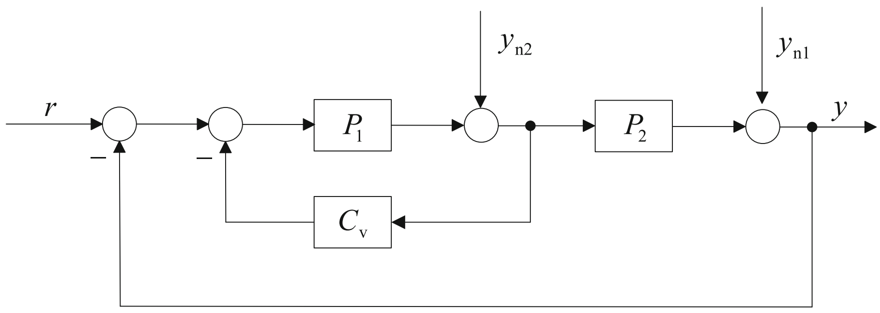


Fig. 4.44 Block diagram of compensation by a feedback block

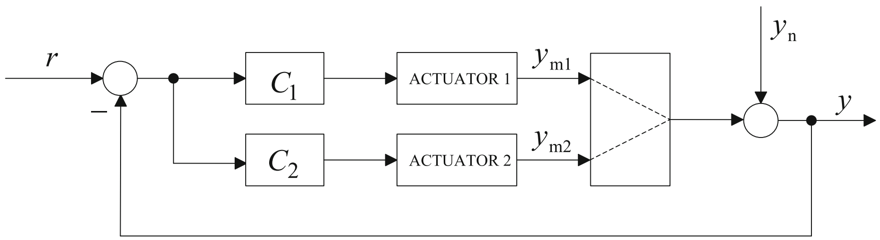


Fig. 4.45 Block diagram of a control system with two loops applying an auxiliary manipulated variable

With appropriate internal feedback, the inverse of the process can be generated, providing a favorable control solution.

Generally it is sufficient to have an internal loop working only in the transient state, while in the steady state only the external loop is efficient. This can be accomplished by feedback through a differentiating element combined with a first order lag: $C_{\mathrm{v}}(s) = A_{\mathrm{v}} s\tau/(1 + sT_1)$.

4.11 Control with Auxiliary Manipulated Variables

Typically there are several possibilities for manipulation at the process input. Depending on the properties of the process one of the manipulated variables is fundamental, while the others can be used as auxiliary possibilities applied, in general, only temporarily.

The block diagram of a control system with an auxiliary loop is shown in Fig. 4.45. As an application example, let us consider the control of the belt dryer furnace shown earlier in Fig. 4.38. If the humidity of the material coming into the furnace changes abruptly, without feedforward its effect on the output is recognized only when the furnace is already full of the material of changed quality. In this case, if only the heating of the furnace is modified, a relatively large amount of the material comes out of the furnace with humidity that differs from the required value, as the thermal inertia of the furnace is big and the temperature can only be changed slowly. The manipulation becomes more effective if the speed of the conveyor is also changed temporarily by changing the speed of the motor M. Thus the residence time of the material in the furnace is shortened. The auxiliary manipulation can only be temporary, which can be achieved by e.g. using a proportional controller in the auxiliary loop, that influences the armature voltage of the motor. In the main control loop, a *PI* controller is applied, thus in steady state the error signal is zero, and then the auxiliary circuit becomes inactive.