Chapter 7
Control of Stable Processes

In the initial period, heuristic, trial-and-error methods based on rule of thumb were widely used for controller design. At the same time a great effort was made to elaborate a general mathematical methodology for a theoretical approach to design methods. The transfer function of an mth-order controller has $(2m+1)$ unknown parameters. While tuning the higher order controllers it was noticed that a certain design goal can be reached by many parameter sets, moreover, as a consequence, in many cases the parameters were not independent, so the parameterization of the controller was redundant. The main question is how to parameterize a general, stable controller to solve the basic design tasks of a closed-loop control system with a minimum number of non-redundant parameters. The most important solution is provided by the so-called YOULA-parameterization. The YOULA-parameter, as a matter of fact, is a stable (following from its definition), regular transfer function.

$$Q(s) = \frac{C(s)}{1 + C(s)P(s)} \quad \text{or, for simplicity,} \quad Q = \frac{C}{1 + CP}, \tag{7.1}$$

where $C(s)$ is a stabilizing controller and $P(s)$ is the transfer function of the stable process. The inner stability of a system is defined by the fact, that introducing a bounded input signal at any point of the system provides a bounded output signal at any other point of the loop (see Sect. 5.2). For the investigation of inner stability, the so-called transfer matrix of the closed-loop has to be constructed

$$\mathbf{T}_{\mathrm{t}}(P, C) = \begin{bmatrix} \frac{CP}{1+CP} & \frac{P}{1+CP} \\ \frac{C}{1+CP} & \frac{1}{1+CP} \end{bmatrix} = \frac{1}{1+CP} \begin{bmatrix} CP & P \\ C & 1 \end{bmatrix} \tag{7.2}$$

The transfer matrix represents the connection between two independent outer and two inner signals. The closed-loop is inner stable, if and only if, all elements of $\mathbf{T}_{\mathrm{t}}(P, C)$ are stable.

L. Keviczky et al., *Control Engineering*, Advanced Textbooks in Control and Signal Processing, https://doi.org/10.1007/978-981-10-8297-9_7

7.1 The Youla-Parameterization

The transfer matrix can also be expressed by the Youla parameter $Q(s)$ instead of the controller $C(s)$:

$$\mathbf{T}_{\mathrm{t}}(P,Q) = \begin{bmatrix} QP & P(1-QP) \\ Q & 1-QP \end{bmatrix}. \tag{7.3}$$

Here it can be easily seen that the inner stability is ensured by any stable $Q(s)$ for a stable process.

It follows from the definition of the Youla parameter that the structure of the realizable and stabilizable controller is fixed in the control loop parameterized in this way, i.e.,

$$C(s) = \frac{Q(s)}{1-Q(s)P(s)} \quad \text{or, for simplicity} \quad C = \frac{Q}{1-QP}. \tag{7.4}$$

The Youla-parameterized (*YP*) control loop is shown in Fig. 7.1, where r is the reference signal, e is the error signal, u is the output of the controller (the actuating signal), y_{n} is the disturbance signal affecting the output, and y is the output signal of the process, i.e., the controlled variable.

The overall transfer function of the closed-loop (the complementary sensitivity function) is

$$T(s) = \frac{C(s)P(s)}{1+C(s)P(s)} = Q(s)P(s) \quad \text{or, more simply,} \quad T = \frac{CP}{1+CP} = QP, \tag{7.5}$$

which is linear in $Q(s)$. (This linearity, as will be seen later, will facilitate to a great extent the design of the required dynamics of the one degree of freedom (*ODOF*) closed-loop.) The sensitivity function has the form

$$\begin{aligned} S(s) &= \frac{1}{1+C(s)P(s)} = 1 - Q(s)P(s) \quad \text{or, for simplicity,} \\ S &= \frac{1}{1+CP} = 1 - QP. \end{aligned} \tag{7.6}$$

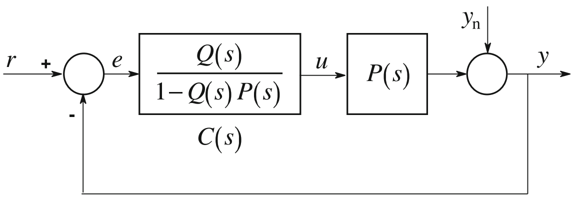

Fig. 7.1 The *YP* control loop

The relationship between the most important signals of the closed-loop can be obtained by simple calculations:

$$\begin{aligned} u &= Qr - Qy_{\mathrm{n}} \\ e &= (1 - QP)r - (1 - QP)y_{\mathrm{n}} = Sr - Sy_{\mathrm{n}} \\ y &= QPr + (1 - QP)y_{\mathrm{n}} = Tr + Sy_{\mathrm{n}} \end{aligned} \qquad (7.7)$$

The effect of r and y_{n} on u and e is completely symmetrical (not considering the sign). Thus in this system the input of the process depends only on the outer signals and on $Q(s)$.

It is interesting to see that the *YP* controller of (7.4) can be realized by the simple control loop of positive feedback shown in Fig. 7.2. Using this scheme the control loop of Fig. 7.1 can be transferred to the equivalent block scheme of Fig. 7.3 by identical conversions. This latter scheme is called an *internal model control* (*IMC*). The basic principle of this control is that only the deviation (ε) (i.e. the error signal) of the process output and the model output is fed back. This error signal is zero in the ideal case when the inner model is completely equivalent to the process. This case is presented in Fig. 7.3. In reality, however, the transfer function $\hat{P}(s)$ of the inner model is only a good approximation of the true process $P(s)$, since the original system is not known. For simplicity, only the ideal case is investigated here.

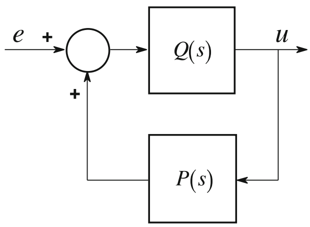

Fig. 7.2 The realization of the *YP* controller

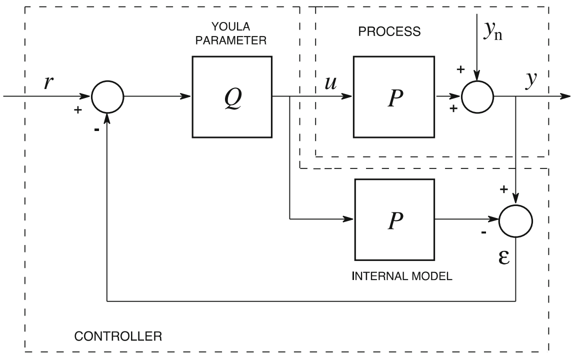

Fig. 7.3 The equivalent *IMC* loop

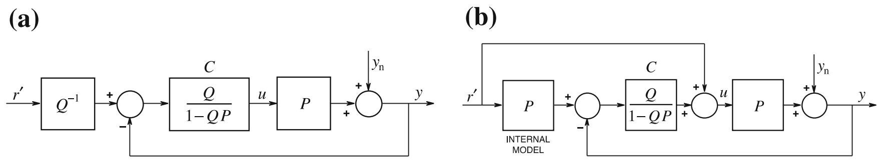

Fig. 7.4 Block schemes for "opening" the closed-loop

Based on the last equation of (7.7) it can be seen that the *IMC* has the transfer function QPr for the tracking of the reference signal. If the inverse of Q is connected in series to the control loop of Fig. 7.1 according to Fig. 7.4a, then the tracking performance becomes independent of Q, i.e., it becomes that given by Pr', which means that practically the closed-loop is "opened". It can be easily checked that this block scheme is equivalent to that of Fig. 7.4b. It has to be noted that here the reference signal has a direct effect on the input of the process: it does not go through the controller and the whole closed-loop. The effect of the controller (concerning the reference signal) is in operation only when the inner model is not equal to the real process.

Following the above train of thought, the extension of the YOULA-parameterization can also be introduced for two-degree of freedom (*TDOF*) control loops. To do this let us simply introduce the parameter Q_r for designing the tracking behavior and connect it in series to the control loop of Fig. 7.4. Then we get the block scheme of Fig. 7.5. The resulting transfer characteristics of this system are

$$\begin{aligned} u &= Q_r y_r - Q y_n \\ e &= (1 - Q_r P) y_r - (1 - QP) y_n = (1 - T_r) y_r - S y_n \\ y &= Q_r P y_r + (1 - QP) y_n = T_r y_r + (1 - T) y_n = T_r y_r + S y_n \end{aligned} \qquad (7.8)$$

where the tracking performance can be designed by choosing the parameter Q_r in $T_r = Q_r P$, while the performance of the disturbance rejection can be designed by choosing Q in $T = QP$. Thus these two performances can be handled separately. The reference signal of the whole system is noted by y_r. The same preconditions are valid for Q_r and Q. The transfer function of the *TDOF* closed-loop referred to the reference signal T_r is analogous to the complementary sensitivity function $T$of the *ODOF* system for the tracking.

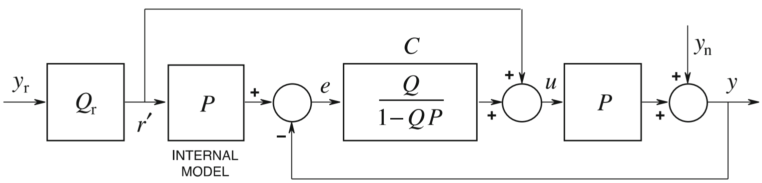

Fig. 7.5 Two degree of freedom version of the *YP-controller*

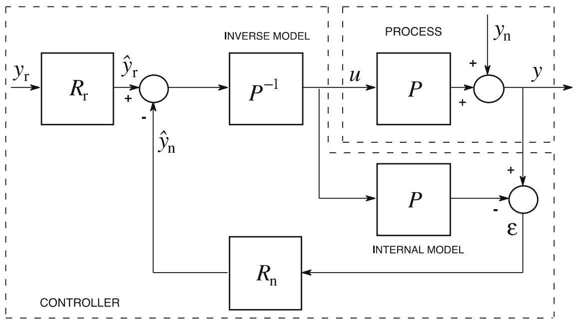

Fig. 7.6 The extension of the control loop based on the ideal *IMC*

The *IMC* of Fig. 7.3 can be further developed according to Fig. 7.6. Here the predicted value $\hat{y}_n$ of the output noise y_n is constructed from the difference ε between the output of the process and the model by the predictor R_n. Similarly the predictor R_r provides the prediction $\hat{y}_r$ of the reference signal y_r. The disturbance compensation of the loop works by giving the predicted value $-\hat{y}_n$ through the inverse of the process to the input of the process, thus in the case of an exact estimation, the disturbance is eliminated. The tracking works in a similar way. Here the operation of R_r can be referred to as a reference model (the desired system dynamics), therefore the introduced predictor is also referred to as a reference model. It is generally required that these predictors have to be strictly proper with unit static gain, i.e., $R_n(\omega = 0) = 1$ and $R_r(\omega = 0) = 1$.

The best operation of a *TDOF* control loop can be attained by the special conditions $R_r = R_n = 1$ or $1 - R_n = 0$, but,—as will be shown later,—it can not be realized in most practical systems.

The block scheme of Fig. 7.6 can be redrawn to the equivalent form of Fig. 7.7. Note that the transfer function—in the ideal case, i.e., when the inverse of the process is realizable and stable—is

$$C_{id} = \frac{(R_n P^{-1})}{1 - (R_n P^{-1})P} = \frac{Q}{1 - QP} = \frac{R_n}{1 - R_n} P^{-1}, \tag{7.9}$$

which is the *YP*-controller with the YOULA parameter

$$Q = R_n P^{-1}. \tag{7.10}$$

For the tracking, however, the parameter is

$$Q_r = R_r P^{-1}. \tag{7.11}$$

It can be seen that the controller C_{id} is realizable if the pole excess of R_n is greater than or equal to that of the process, i.e., a pole excess of j can be easily ensured by the reference model $R_n = 1/(1 + sT_n)^j$.

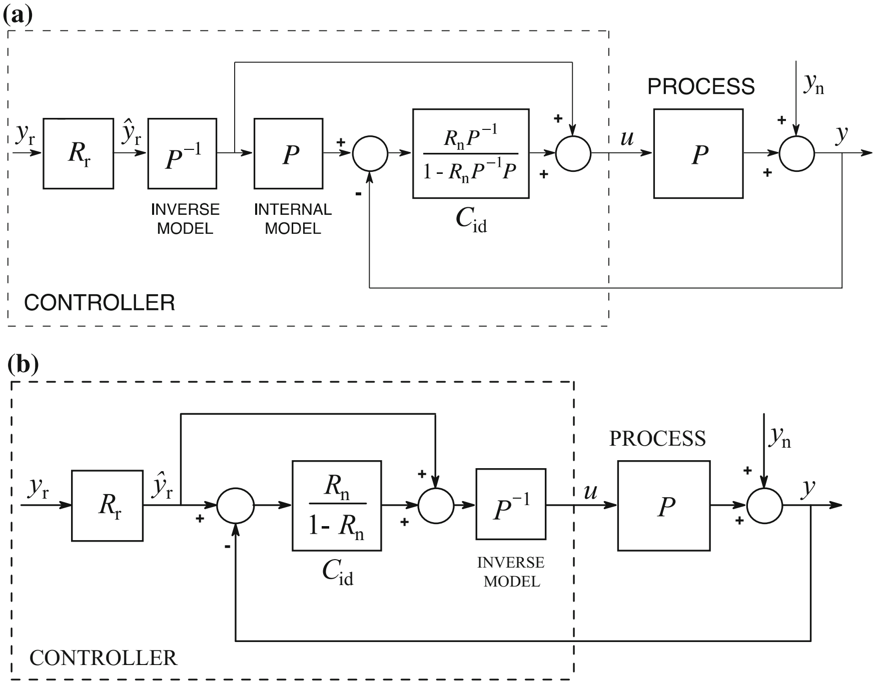

Fig. 7.7 Equivalent ideal control loops using the extended *IMC* principle

The most important signals of the closed-loop for the ideal case are

$$\begin{aligned} u_{id} &= R_r P^{-1} y_r - R_n P^{-1} y_n \\ e_{id} &= (1 - R_r) y_r - (1 - R_n) y_n = \left(1 - T_r^{id}\right) y_r - S_{id} y_n \\ y_{id} &= R_r y_r + (1 - R_n) y_n = T_r^{id} y_r + (1 - T_{id}) y_n = T_r^{id} y_r + S_{id} y_n \end{aligned} \quad (7.12)$$

thus in the ideal case, $T_r^{id} = R_r$ and $T_{id} = R_n$, which are our design goals.

Note that the previously followed thoughts presented only the main idea of the YOULA-parameterization and its equivalency with the control based on *IMC*. There is, however, a very critical point of the realizability of the resulting schemes, namely the realizability of the inverse of the process P. Unfortunately, this, in general—disregarding some rare exceptions—is not true for *continuous-time* (CT) systems. For practical applications, versions of the above approach have to be found where all elements of the *TDOF* system are realizable.

To introduce a generally applicable controller, let us assume the transfer function of the process has the following factored form

$$P(s) = P_+(s)\bar{P}(s)_- = P_+(s)P_-(s)e^{-sT_d}, \quad \text{or, for short,} \\ P = P_+\bar{P}_- = P_+P_-e^{-sT_d}, \tag{7.13}$$

where P_+ is stable, and its inverse is also stable and realizable (*ISR*). The inverse of $\bar{P}_-$ is unstable (*Inverse Unstable*: *IU*) and not realizable (*IUNR*). P_- is inverse unstable (*IU*). Here, in general, the inverse of the dead-time part e^{-sT_d} is not realizable, because it would be an ideal predictor. The generalized *IMC* principle can also be applied to the general process structure that is shown in Fig. 7.8.

The block scheme of Fig. 7.8 can be redrawn into the equivalent form of Fig. 7.9, where the realizable *YP*-controller of optimal structure obtained for the general case has the form

$$C_{opt} = \frac{Q_{opt}}{1 - Q_{opt}P} = \frac{R_n G_n P_+^{-1}}{1 - R_n G_n P_- e^{-sT_d}} = \frac{R_n K_n}{1 - R_n K_n P} = R_n G_n C'_{opt}, \tag{7.14}$$

where the optimal YOULA parameter is

$$Q_{opt} = R_n G_n P_+^{-1} = R_n K_n \quad \text{where} \quad K_n = G_n P_+^{-1} \tag{7.15}$$

and

$$Q_r = R_r G_r P_+^{-1} = R_r K_r \quad \text{where} \quad K_r = G_r P_+^{-1}. \tag{7.16}$$

The obtained general control loop—due to the *YP* – gives structurally the best controller for stable processes. Further optimality of the controller can be set by the embedded transfer functions G_r and G_n. To understand this, let us consider again the most important signals of the *TDOF* closed-loop in the optimal case:

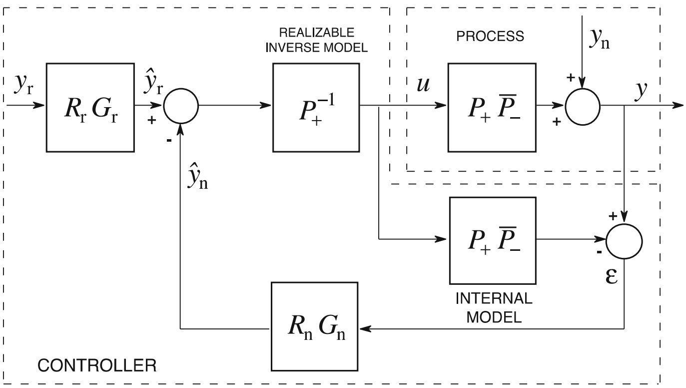

Fig. 7.8 The optimal control loop based on the generalized *IMC* principle

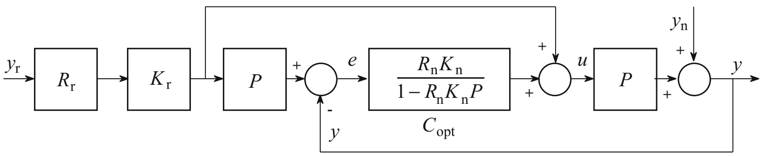

Fig. 7.9 The equivalent optimal control loop corresponding to the generalized *IMC* principle

$$
\begin{aligned}
u_{opt} &= R_r G_r P_+^{-1} y_r - R_n G_n P_+^{-1} y_n \\
e_{opt} &= \left(1 - R_r G_r P_- e^{-sT_d}\right) y_r - \left(1 - R_n G_n P_- e^{-sT_d}\right) y_n = \left(1 - T_r^{opt}\right) y_r - S_n^{opt} y_n \\
y_{opt} &= R_r G_r P_- e^{-sT_d} y_r + \left(1 - R_n G_n P_- e^{-sT_d}\right) y_n = T_r^{opt} y_r + \left(1 - T_n^{opt}\right) y_n = T_r^{opt} y_r + S_n^{opt} y_n
\end{aligned}
\tag{7.17}
$$

where the equalities $T_r^{opt} = R_r G_r P_- e^{-sT_d}$ and $T_n^{opt} = R_n G_n P_- e^{-sT_d}$ occur.

Compare the ideal output y_{id} of (7.12) with the optimal output y_{opt} of (7.17). It can be easily seen that the ideal and the designed transfer functions determined by the reference models R_r and R_n can not be reached, only approximated. The element $P_- e^{-sT_d}$ appearing in the approximate transfer functions can not be eliminated, therefore it is called an invariant factor. Thus the dead-time e^{-sT_d} and the *inverse unstable* (*IU*) term P_- of the process can not be eliminated by any controller. In case of CT processes, this latter term contains the unstable zeros of the non-minimum phase processes and those poles of the stable poles which could not get into the invertible P_+ of (7.13). In practice, only the necessary number of the slowest poles (whose number corresponds to the number of the stable zeros in P) of P are usually included in P_+, the rest should be added to P_-. The effect of the invariant P_- can only be attenuated by the transfer functions G_r and G_n.

The formulation of the deviation of the outputs of the ideal and best reachable (optimal) control loops is

$$
\Delta y = y_{id} - y_{opt} = R_r \left(1 - G_r P_- e^{-sT_d}\right) y_r + R_n \left(1 - G_n P_- e^{-sT_d}\right) y_n, \tag{7.18}
$$

where the error comes from the transfer function $R_x\left(1 - G_x P_- e^{-sT_d}\right)\big|_{x=r,n}$ both for the tracking and noise rejection. The minimization of this error in terms of different criteria (in theoretical investigations, the so-called $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimality) can be accomplished by the optimal choice of $G_x\big|_{x=r,n}$. (The discussion of optimality is not the subject of this book.)

Further equivalent forms of the best reachable optimal control loop are shown in Fig. 7.10. From these the simplest one that is realizable has to be chosen. Figure 7.10b gives advice for the realization of a system having dead-time. The control loop shown in Fig. 7.9 is called the most general (generic) form of a *TDOF* systems. (The Youla-parameterization has been extended to *TDOF* systems by Keviczky and Bányász by introducing two further parameters, R_r and R_n instead of

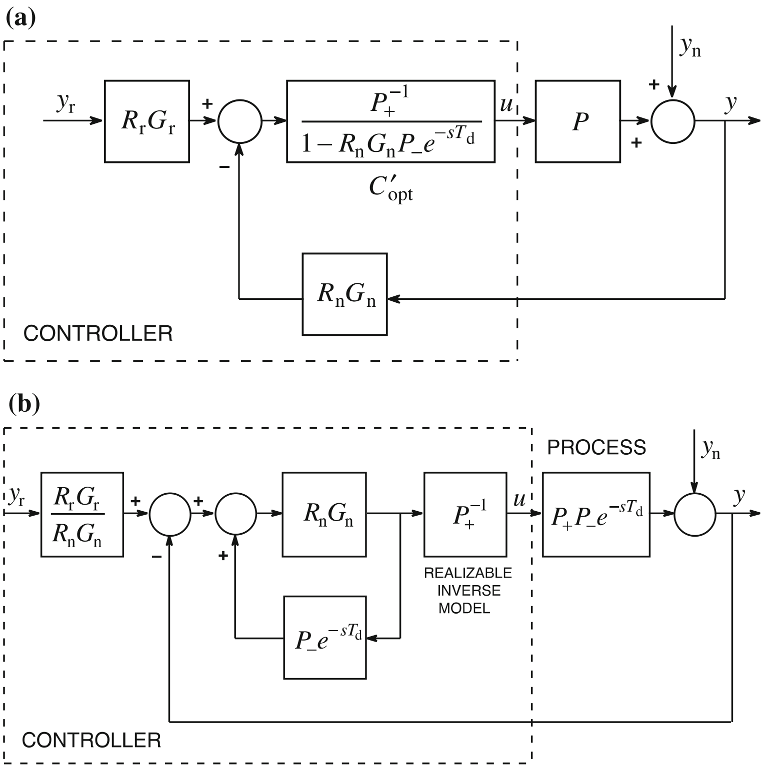

Fig. 7.10 Equivalent forms of the best reachable (optimal) control loops

Q (called the *K-B*-parameterization). The derivation of the generic schemes and their optimization possibilities are also connected to their names).

The questions of realizability can be dealt with in long discussions. If the design of the optimal controller includes also the optimization of G_r and G_n, then the procedure itself must also ensure the realizability of the transfer functions $G_r P_-$ and $G_n P_-$, and the realizability of the other factors, (like $R_r G_r P_+^{-1}$, $R_n G_n P_+^{-1}$ and

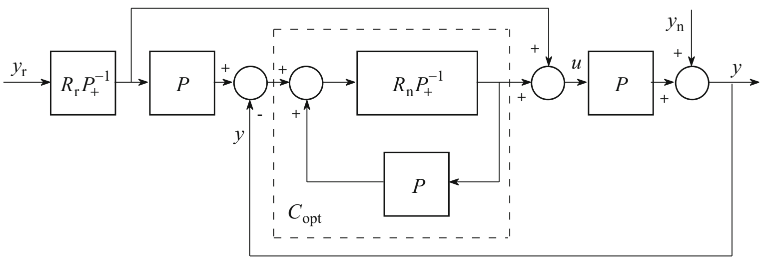

Fig. 7.11 A realizable YOULA-parameterized control loop with the choice $G_r = G_n = 1$

$R_n G_n P_-$), has to also be considered. As was mentioned earlier, the theory concerning the optimality of G_r and G_n is not discussed here. In this case the choice $G_r = G_n = 1$ does not change the invariant P_-, i.e., it appears, as a consequence, unchanged in the signals of the system

$$\begin{aligned} u &= R_r P_+^{-1} y_r - R_n P_+^{-1} y_n \\ e &= \left(1 - R_r P_- e^{-sT_d}\right) y_r - \left(1 - R_n P_- e^{-sT_d}\right) y_n = (1 - T_r) y_r - S_n y_n \\ y &= R_r P_- e^{-sT_d} y_r + \left(1 - R_n P_- e^{-sT_d}\right) y_n = T_r y_r + (1 - T_n) y_n = T_r y_r + S_n y_n \end{aligned} \tag{7.19}$$

furthermore the realizability of the transfer functions $R_r P_+^{-1}$, $R_n P_+^{-1}$ and $R_n P_-$ is required. It is evident that in this case the realizability can be simply handled by the appropriate choice of the order of the reference models R_r and R_n, and of the pole excess, e.g., by prescribing $R_r = 1/(1 + sT_r)^j$ (and the same for R_n). A realizable, but not optimal control loop can be seen in Fig. 7.11.

Although the controller is theoretically realizable, it can not be expected in practice that for CT systems an ideal dead-time element modeling the time-delay of the process can be realized in the inner positive feedback loop of the controller and in the serial compensator. Therefore in the case of time-delay CT systems, the above discussed optimal control scheme has only theoretical importance. In some cases, the time-delay term can be approximated by higher order PADE-series. In computer controlled cases (for sampled DT controls), however, the method can be fully applied (see Chap. 12).

Example 7.1 Let the controlled system be a first order time-delay lag

$$P = \frac{1}{1 + 10s} e^{-5s} \quad \text{i.e.} \quad P_+ = \frac{1}{1 + 10s}; \quad \bar{P}_- = e^{-5s} \quad \text{and} \quad P_- = 1, \tag{7.20}$$

which should be sped up by the control. Let the tracking and disturbance cancellation reference models be

$$R_r = \frac{1}{1 + 4s} \quad \text{and} \quad R_n = \frac{1}{1 + 2s}. \tag{7.21}$$

Since $P_- = 1$, there is nothing to be optimized, i.e., $G_r = 1$ and $G_n = 1$ can be chosen. The optimal controller is

$$\begin{aligned} C_{opt} &= \frac{R_n G_n P_+^{-1}}{1 - R_n G_n P_- e^{-sT_d}} = \frac{1}{1 - R_n e^{-sT_d}} R_n P_+^{-1} = \frac{1}{1 - \frac{1}{1+2s} e^{-5s}} \frac{1 + 10s}{1 + 2s} \\ &= \frac{1 + 10s}{1 + 2s - e^{-5s}} \end{aligned} \tag{7.22}$$

and the serial compensator has the form

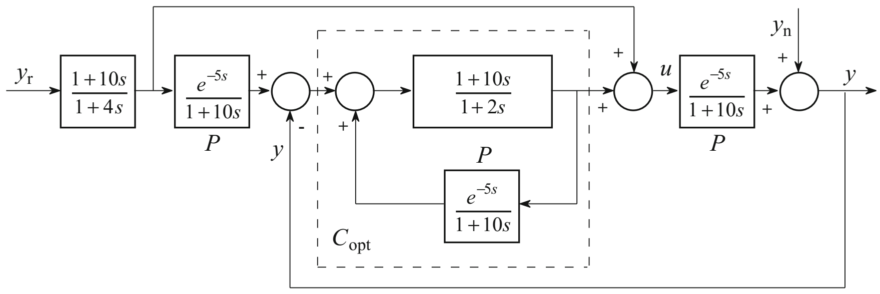

Fig. 7.12 The optimal control loop of Example 7.1

$$R_r K_r = R_r G_r P_+^{-1} = R_r P_+^{-1} = (1+10s)/(1+4s), \tag{7.23}$$

thus the optimal *TDOF* control loop has the structure shown in Fig. 7.12. Observe that $C_{opt}(s=0)=\infty$, i.e., the controller has integrating behavior, which results from the condition $R_n(s=0)=1$.

It can be easily checked, that the output of the closed system is

$$y_{opt} = R_r e^{-sT_d} y_r + \left(1 - R_n e^{-sT_d}\right) y_n = \frac{1}{1+4s} e^{-5s} y_r + \left(1 - \frac{1}{1+2s} e^{-5s}\right) y_n, \tag{7.24}$$

which completely corresponds to the designed *TDOF* control loop. ∎

Example 7.2 Let the controlled process be a second order lag

$$P = \frac{(1+5s)(1+6s)}{(1+10s)(1+8s)} = P_+ \quad \text{i.e.,} \quad P_- = 1. \tag{7.25}$$

Suppose that the tracking and disturbance cancellation models are again of the form (7.21). Since $P_- = 1$, there is nothing to be optimally compensated, i.e., $G_r = 1$ and $G_n = 1$ can be chosen. Now the optimal controller is

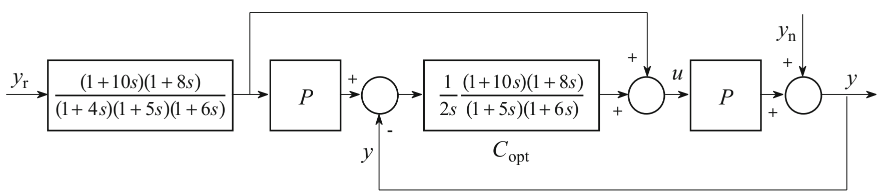

Fig. 7.13 The optimal control loop of Example 7.2

$$C_{\mathrm{opt}} = \frac{R_{\mathrm{n}} G_{\mathrm{n}} P_{+}^{-1}}{1 - R_{\mathrm{n}} G_{\mathrm{n}} P_{-} e^{-sT_{\mathrm{d}}}} = \frac{R_{\mathrm{n}}}{1 - R_{\mathrm{n}}} P^{-1} = \frac{1}{2s} \frac{(1+10s)(1+8s)}{(1+5s)(1+6s)} = C_{\mathrm{id}} \qquad (7.26)$$

thus it corresponds to the ideal controller. The serial compensator, however, has the form

$$R_{\mathrm{r}} K_{\mathrm{r}} = R_{\mathrm{r}} G_{\mathrm{r}} P_{+}^{-1} = R_{\mathrm{r}} P^{-1} = \frac{(1+10s)(1+8s)}{(1+4s)(1+5s)(1+6s)}. \qquad (7.27)$$

It is evident, that the controller is of integrating type, as is also shown in Fig. 7.13.

Note that in ideal case the term $R_{\mathrm{n}}/(1-R_{\mathrm{n}})$ in the controller corresponds to an integrator, whose integrating time is equal to the time constant of the first order reference model R_{n}. ∎

7.2 The Smith Controller

The handling of the time-delay of the processes has required the special attention of the designers of the control loops from the beginning. First Otto Smith suggested a technique by means of which it was thought for a long time that the controller can be designed without the consideration of the dead-time. To understand his method let us consider a simple dead-time process of (7.13)

$$P(s) = P_{+}(s)\bar{P}_{-}(s) = P_{+}(s)e^{-sT_{\mathrm{d}}} \quad \text{or, more simply,} \quad P = P_{+}\bar{P}_{-} = P_{+}e^{-sT_{\mathrm{d}}}, \qquad (7.28)$$

where P_{+} is stable. Figure 7.14a shows the original idea of Smith. Since this figure is equivalent to Fig. 7.14b, his main goal can be clearly seen, namely to separate the original dead-time loop into a closed-loop which does not contain the time-delay

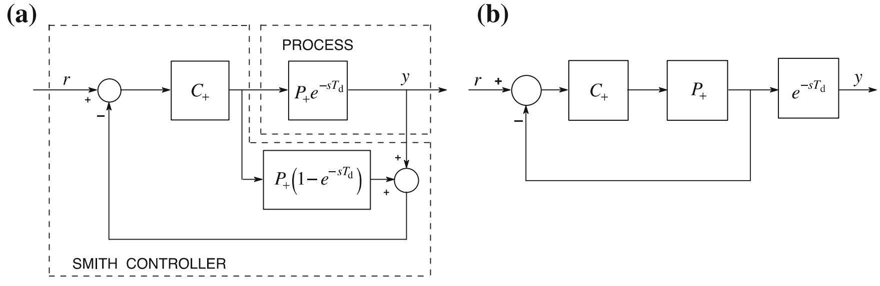

Fig. 7.14 The block scheme of the Smith controller

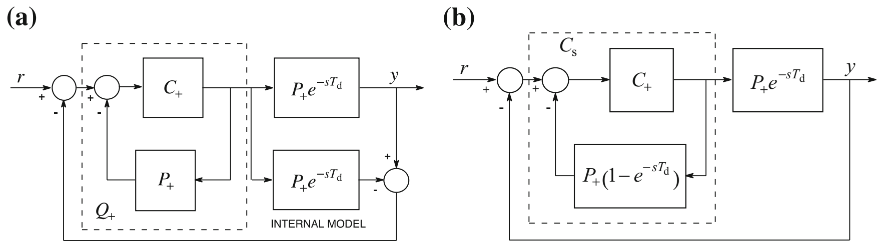

Fig. 7.15 Equivalent SMITH controller block schemes

and a serially connected dead-time. So the controller C_+ regulating the process P_+ can be designed by a conventional method.

By simple block manipulations, Fig. 7.14a can be redrawn to the equivalent forms of Fig. 7.15a, b.

The *IMC* structure of the Fig. 7.15a clearly shows that the SMITH controller is a *YP*-controller with a special YOULA parameter

$$Q_+ = \frac{C_+}{1 + C_+ P_+} = \frac{C_+ P_+}{1 + C_+ P_+} P_+^{-1} = \frac{L_+}{1 + L_+} P_+^{-1} = R_+ P_+^{-1} \tag{7.29}$$

if the controller C_+ stabilizes the delay free part of the process P_+. Here $L_+ = C_+ P_+$ is the loop transfer function of the closed system of Fig. 7.14b, furthermore the complementary sensitivity function

$$T_+ = R_+ = \frac{L_+}{1 + L_+} \tag{7.30}$$

will be the reference model R_+. Since in the *IMC* structure the inner model predicts the output of the process, the name SMITH-predictor derives from this phenomenon. At the time of its introduction the *IMC* principle and the YOULA-parameterization were not yet known.

Figure 7.15b shows the equivalent complete closed control loop, where the serial (YOULA-parameterized) controller is

$$C_s = \frac{Q_+}{1 - Q_+ P_+ e^{-sT_d}} = \frac{C_+}{1 + C_+ P_+ (1 - e^{-sT_d})} = C_+ K_S, \tag{7.31}$$

which, at the same time, also shows the inner closed-loop referring to the realization. Here K_S means the serial transfer function by which the SMITH controller modifies the effect of the original controller C_+.

Thus

$$K_S = \frac{1}{1 + C_+ P_+ (1 - e^{-sT_d})} = \frac{1}{1 + L_+ (1 - e^{-sT_d})}. \tag{7.32}$$

At the stability limit $L_+ = -1$, we get

$$K_S = \frac{1}{1 + (-1)(1 - e^{-sT_d})} = \frac{1}{1 - 1 + e^{-sT_d}} = e^{sT_d}\Big|_{\omega_c} = e^{j\omega_c T_d}, \tag{7.33}$$

which causes the Smith controller to add a significant positive phase advance to the original closed-loop, which is why it can be applied very successfully for stabilization in many cases. At the same time it is very sensitive to a change of the parameters.

Unfortunately it should be repeated that in the practice for CT systems one cannot expect to realize an ideal dead-time element only its higher order lag approximation can be implemented for the application of a Smith controller (see what was stated about Example 7.1 of the previous chapter).

To complete the evaluation of the Smith controller it has to be also mentioned that it can be used only for the design of a one-degree of freedom (*ODOF*) system, i.e., for tracking. The controller designs the tracking of the reference signal only in an indirect way, as the expression $T_+ = R_+$ of (7.30) shows. From the elaboration of the concept of Youla-parameterization it has been known that simple design method is also available for *TDOF* systems both for tracking and disturbance rejections via the design of the reference models.

7.3 The Truxal-Guillemin Controller

Prior to the Youla-parameterization, Truxal and Guillemin recommended a simple algebraic method for the control design of *ODOF* systems. According to the method the required design goal has to be formulated for the transfer function of the closed-loop

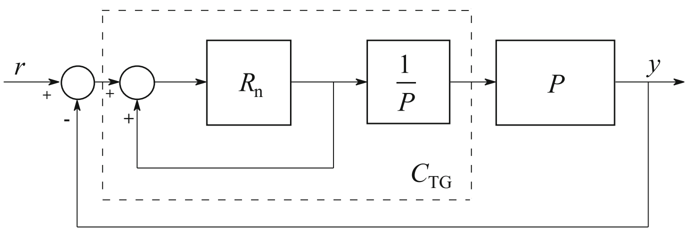

Fig. 7.16 The realization of the Truxal-Guillemin controller

$$R_n = T = \frac{CP}{1 + CP} \tag{7.34}$$

from which a simple algebraic equation results for C:

$$CP = R_n + CPR_n. \tag{7.35}$$

Solving for the controller we get

$$C = \frac{R_n}{1 - R_n} \frac{1}{P} = C_{TG}. \tag{7.36}$$

Note that this form is the same as the simple case of the YOULA controller C_{id} in (7.9). The realization of the controller can be made according to the Fig. 7.16.

Thus R_n corresponds to one of the reference models of the YOULA method. For the *ODOF* case, however, $R_n = R_r$. Let the reference model be $R_n = \mathcal{B}_n / \mathcal{A}_n$, and the process be $P = \mathcal{B}/\mathcal{A}$. So the polynomial form of the controller is

$$C_{TG} = \frac{\mathcal{B}_n}{\mathcal{A}_n - \mathcal{B}_n} \frac{\mathcal{A}}{\mathcal{B}}. \tag{7.37}$$

The controller is realizable if the pole excess of R_n is greater than or equal to that of the process. If R_n has unity gain $(R_n(0) = 1)$, then the type of the controller is one. TRUXAL observed that the loop transfer function $L = \mathcal{F}/s^k\mathcal{D} = \mathcal{CP}$ of type k can be established by the reference model

$$\begin{aligned} R_n = T = \frac{\mathcal{N}}{\mathcal{N} + s^k\mathcal{D}} &= \frac{f_o + f_1 s + \ldots + f_{k-1} s^{k-1}}{f_o + f_1 s + \ldots + f_{k-1} s^{k-1} + s^k + \ldots + d_{n_R + k} s^{n_R + k}} \\ &= \frac{f_o + f_1 s + \ldots + f_{k-1} s^{k-1}}{f_o + f_1 s + \ldots + f_{k-1} s^{k-1} + s^k (1 + \ldots + d_{n_R} s^{n_R})}; n_R - k - 1 \geq n - m \end{aligned} \tag{7.38}$$

where the first k terms of the denominator are equal to the numerator.

7.4 The Effect of a Constrained Actuator Output

The control signals applied in the closed control systems, or the output of the actuator whose task is to increase that signal to the proper level, are always amplitude constrained.

$$|u(t)| \leq U_{max} \tag{7.39}$$

This means that a jump of any size in u, or a significant change in the starting value of the signal related to its final value, i.e., arbitrary overexcitation is

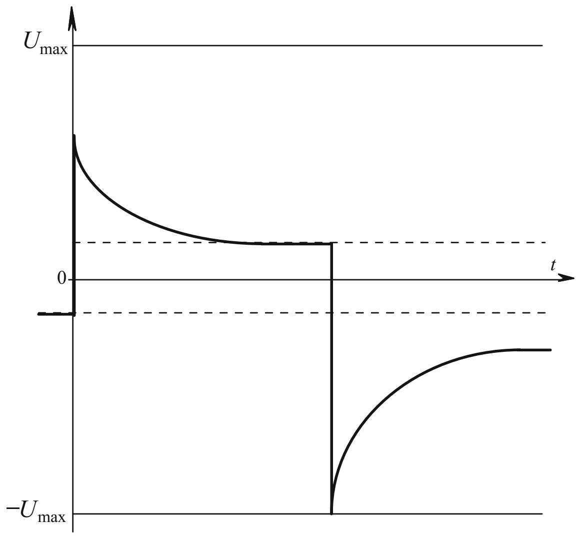

Fig. 7.17 Typical control output (actuator signal) in the case of overexcitation

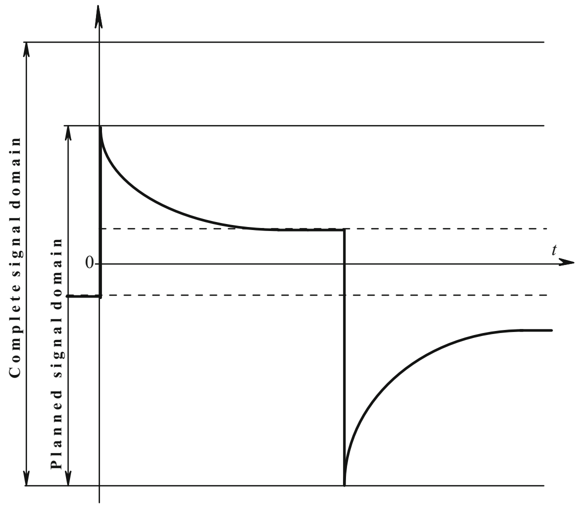

Fig. 7.18 Design of the signal domain for the control output

impossible. It was shown in Sect. 2.4 that a pole cancellation can be made by adding extra zeros which results in speeding up the system. This speeding up always requires overexcitation (energy surplus). The optimal control methods discussed in this chapter almost always applied a certain kind of pole cancellation, i.e., overexcitation (see Fig. 7.17). The above mentioned amplitude constraints in practice mean, that in spite of computing the optimal control parameters, the provided output cannot be realized because of the constraints. The reachable speed-up really depends on the applicable overexcitation.

The design of the signal domain for the control output needs special care and the knowledge of the equipment employed. In general, the working point is set to the center of the signal domain, and the possible changes compared to this point have to be designed to perform without saturation (Fig. 7.18).

In spite of the most careful design it may happen that the control output violates the signal domain. In this case the original design goal has to be reduced. The advantage of the *KB*-parameterization of the generic *TDOF* control loops is that in this case it is enough to redesign only the problematic (very demanding) reference models R_r or R_n by less demanding design conditions. This process can usually be made in small steps by iteration. The iteration steps may include both the model simulation and an experiment on the real system. (In case of lower order reference models it is possible to elaborate explicit design formulas for determining the time constant of the model (bandwidth) with the knowledge of the process model and the amplitude constraints U_{max}.)

In many cases not only the amplitude of the control output has constraints but its changing velocity is also limited in practice. Let us think of the control valves of big pipes, where the motor needs time to transfer the valve from one position to another. Handling these so-called velocity constraints by analytic methods is more difficult, so only simulation and practical experiment remain as a solution. The applied method is the same as earlier: the demand required by the design goal has to be reduced.

Summarizing, it can be stated that the fastest reachable control depends, primarily and to a great extent, on the limitations of the control output. This limitation, however, does not depend on the control design method, but on the type of the equipment used in the given technology. So any improvement can be made only by changing the equipment itself or by redesigning the whole technology.

The Concept and Computation of Dynamic Overexcitation

In control systems the actuator signal $u(t)$ has an important role because this is the input of the process and the physical constraints appear here. Due to the change of the reference signal $r(t)$ transient processes take place according to the dynamics of the system. During a transient, in general, the actuator signal might have a higher value than its static one. The extent or magnitude of this can be described by the dynamic overexcitation. The definition of dynamic overexcitation is

$$u_t = \frac{U_{max}}{u(t \rightarrow \infty)} \cong \frac{u(0)}{u_\infty}.$$

In general the maximum value is the initial value, therefore, for the sake of the simplicity of the computations, this is used as an approximation. The dynamic overexcitation can be interpreted this way only when u_∞ is not equal to zero. This happens when the process contains an integrator, since in this case the system can get to a steady state only if the input of the integrator becomes zero. Here the dynamic overexcitation is replaced by U_{max}.

7.5 The Concept of the Best Reachable Control

7.5.1 General Theory

In the previous section, the basic importance of the constraints relating to the output of the actuator for ensuring the reachable/attainable/obtainable best control has been emphasized. Based on the Sects. 7.1–7.3, however, further constraints not depending on us have to be mentioned. Of these, the dead-time of the process is the most important, which is invariant for any kind of regulation method, i.e., its effect can not be eliminated.

Other such factors are the unstable zeros of the process, which also can not be eliminated by any method. The effect of the invariant zeros on the transients can be compensated (decreased or attenuated) to a certain extent. (This compensation can be made by the optimal choice of the embedded filters G_r and G_n, which is not the topic of this book.)

Thus both the dead-times and the unstable zeros are considered independent features of the process that can not be influenced by the control design methods, only by the redesign of the whole process or technology.

So far, in the long discussion of the different control design methods, it was supposed that the transfer function P of the process is known. In reality the exact transfer function of the process is not known: only its model $\hat{P}$ is available. This distinction was used, until now, only in the discussion of the concept of robust stability in Sect. 5.7. For the closed-loop control design it should be noted that the complementary sensitivity function

$$\hat{T} = \frac{C\hat{P}}{1 + C\hat{P}} \tag{7.40}$$

resulting from the *ODOF* model based design is not equal to the real one

$$T = \frac{CP}{1 + CP} = \hat{T}\frac{1 + \ell}{1 + \hat{T}\ell}. \tag{7.41}$$

Here ℓ means the relative uncertainty of the process model of (5.40). The sensitivity function of the real closed-loop can be written in the following decomposed form:

$$
\begin{aligned}
S &= \underbrace{(1-R_n)}_{S_{des}} + \overbrace{\underbrace{\left(R_n-\hat{T}\right)}_{S_{real}} - \underbrace{\left(T-\hat{T}\right)}_{S_{mod}}}^{S_{perf}} = S_{des} + S_{real} + S_{mod} \\
&= \underbrace{(1-R_n)}_{S_{des}} + \underbrace{(R_n-T)}_{S_{perf}} = S_{des} + S_{perf} = \underbrace{\left(1-\hat{T}\right)}_{S_{contr}} + S_{mod} \\
&= \underbrace{\left(1-\hat{T}\right)}_{S_{des}+S_{real}} + S_{mod} = S_{contr} + S_{mod}
\end{aligned}
\tag{7.42}
$$

Here $S_{des} = (1-R_n)$ is the design loss, $S_{real} = \left(R_n - \hat{T}\right)$ is the realizability loss, and $S_{mod} = -\left(T-\hat{T}\right) = \hat{T} - T$ represents the modeling loss part of the sensitivity function. In the other form $S_{contr} = \left(1-\hat{T}\right)$ is the term referring to the control loss, and $S_{perf} = (R_n - T)$ is the performance loss. Each term can be simply interpreted and explained very easily. The meaning of the reference model R_n has been discussed in the previous chapters. It is obvious that there are trivial equalities $S = 1 - T$ and $\hat{S} = 1 - \hat{T}$, where

$$
\hat{S} = \frac{1}{1+C\hat{P}} \quad \text{and} \quad S = \frac{1}{1+CP} = \hat{S}\frac{1}{1+\hat{T}\ell} = \hat{S} + S_{mod}. \tag{7.43}
$$

The term S_{mod} can be further simplified:

$$
S_{mod} = S - \hat{S} = \hat{T} - T = -\frac{\hat{T}\hat{S}\ell}{1+\hat{T}\ell} = -\hat{T}S\ell\Big|_{\ell\to 0} \approx -\hat{T}\hat{S}\ell. \tag{7.44}
$$

It can be seen easily that $\left|\hat{T}\hat{S}\right|$ has its maximum at the cut-off frequency ω_c, so the model must be the most accurate in the vicinity of this frequency.

For *TDOF* control loops the complete transfer function corresponding to the concept of the complementary sensitivity function is obtained by adding an extension as $T_r = FT$, in general. For the model based control

$$
T_r = \hat{T}_r \frac{1+\ell}{1+\hat{T}\ell} \tag{7.45}
$$

again, as it was in (7.41).

Obviously the triviality $S_r = 1 - T_r$ and the triple decomposition introduced in (7.42)

$$
S_r = (1-R_r) + \left(R_r - \hat{T}_r\right) - \left(T_r - \hat{T}_r\right) = S^r_{des} + S^r_{real} + S^r_{mod} \tag{7.46}
$$

also exist.

The term $S^{\mathrm{r}}_{\mathrm{mod}}$ can be further simplified

$$S^{\mathrm{r}}_{\mathrm{mod}} = \hat{T}_{\mathrm{r}} - T_{\mathrm{r}} = -\frac{\hat{T}_{\mathrm{r}}\hat{S}\ell}{1+\hat{T}\ell} = -\hat{T}_{\mathrm{r}}S\ell\big|_{\ell\to 0} \approx -\hat{T}_{\mathrm{r}}\hat{S}\ell. \tag{7.47}$$

The ideal control loop has to follow the signals prescribed by R_{r} and R_{n} (more exactly $1 - R_{\mathrm{n}}$), thus the ideal output of the closed-loop is

$$y_{\mathrm{id}} = y^{\mathrm{o}} = R_{\mathrm{r}}y_{\mathrm{r}} - (1 - R_{\mathrm{n}})y_{\mathrm{n}} = y^{\mathrm{o}}_{\mathrm{r}} + y^{\mathrm{o}}_{\mathrm{n}} \tag{7.48}$$

according to (7.12).

Theoretically, instead of (7.48) only

$$y = T_{\mathrm{r}}y_{\mathrm{r}} - Sy_{\mathrm{n}} = T_{\mathrm{r}}y_{\mathrm{r}} - (1 - T)y_{\mathrm{n}} \tag{7.49}$$

can be obtained, and even this has to be modified according to the model based design

$$\hat{y} = \hat{T}_{\mathrm{r}}y_{\mathrm{r}} - \hat{S}y_{\mathrm{n}} = \hat{T}_{\mathrm{r}}y_{\mathrm{r}} - \left(1 - \hat{T}\right)y_{\mathrm{n}}. \tag{7.50}$$

The deviation between the ideal and the theoretically reachable output is

$$\Delta y = y^{\mathrm{o}} - y = (R_{\mathrm{r}} - T_{\mathrm{r}})y_{\mathrm{r}} - (R_{\mathrm{w}} - T)y_{\mathrm{n}} = S^{\mathrm{r}}_{\mathrm{perf}}y_{\mathrm{r}} - S^{\mathrm{n}}_{\mathrm{perf}}y_{\mathrm{n}}, \tag{7.51}$$

where $S^{\mathrm{r}}_{\mathrm{perf}}$ refers to the performance loss concerning the tracking and $S^{\mathrm{n}}_{\mathrm{perf}} = S_{\mathrm{perf}}$ means the performance loss concerning the disturbance rejection. A similar expression can be obtained for the deviation between the ideal output (y^{o}) and $\hat{y}$ obtained by the model based design

$$\Delta\hat{y} = y^{\mathrm{o}} - \hat{y} = \left(R_{\mathrm{r}} - \hat{T}_{\mathrm{r}}\right)y_{\mathrm{r}} - \left(R_{\mathrm{n}} - \hat{T}\right)y_{\mathrm{n}} = S^{\mathrm{r}}_{\mathrm{real}}y_{\mathrm{r}} - S^{\mathrm{n}}_{\mathrm{real}}y_{\mathrm{n}}. \tag{7.52}$$

Note that in the above expressions the terms S_{real} and $S^{\mathrm{r}}_{\mathrm{real}}$ can be made zero only in the case of inverse stable systems, while these terms can never be made zero for inverse unstable systems.

The above triple decomposition of the sensitivity functions gives a good insight into the limit-optimality of the closed-loop control systems, i.e., to the characterization of the best reachable control. To this end, distinct optimality criteria have to be created for each term, i.e.,

$$\begin{aligned} J_{\mathrm{tracking}} &\le J^{\mathrm{r}}_{\mathrm{des}} + J^{\mathrm{r}}_{\mathrm{real}} + J^{\mathrm{r}}_{\mathrm{mod}} = \left\|S^{\mathrm{r}}_{\mathrm{des}}\right\| + \left\|S^{\mathrm{r}}_{\mathrm{real}}\right\| + \left\|S^{\mathrm{r}}_{\mathrm{mod}}\right\| \\ J_{\mathrm{control}} &\le J^{\mathrm{n}}_{\mathrm{des}} + J^{\mathrm{n}}_{\mathrm{real}} + J^{\mathrm{n}}_{\mathrm{mod}} = \|S_{\mathrm{des}}\| + \|S_{\mathrm{real}}\| + \|S_{\mathrm{mod}}\| \end{aligned} \tag{7.53}$$

both for the tracking and disturbance rejection behaviors. Here the notation $\|\ldots\|$ is used for expressing the optimality criterion. (Strictly speaking, in mathematical analysis this notation is used to refer to the chosen norm of the transfer function.)

Optimization of the Design Loss

The optimization of the first term primarily means the determination of the best (fastest) reachable reference models $R_{\mathrm{r}} = R_{\mathrm{r}}^{\mathrm{opt}}$ and $R_{\mathrm{n}} = R_{\mathrm{n}}^{\mathrm{opt}}$, i.e. the solution of the optimization task under the following limitations

$$\begin{aligned} R_{\mathrm{r}}^{\mathrm{opt}} &= \arg\left\{ \min_{R_{\mathrm{r}}} \left(J_{\mathrm{des}}^{\mathrm{r}} \right) \Big|_{u \in U} \right\} = \arg\left\{ \min_{R_{\mathrm{r}}} \|1 - R_{\mathrm{r}}\| \Big|_{u \in U} \right\} \\ R_{\mathrm{n}}^{\mathrm{opt}} &= \arg\left\{ \min_{R_{\mathrm{n}}} \left(J_{\mathrm{des}}^{\mathrm{n}} \right) \Big|_{u \in U} \right\} = \arg\left\{ \min_{R_{\mathrm{n}}} \|1 - R_{\mathrm{n}}\| \Big|_{u \in U} \right\} \end{aligned} \tag{7.54}$$

where the chosen criteria $J_{\mathrm{des}}^{\mathrm{r}} = \|1 - R_{\mathrm{r}}\|$ and $J_{\mathrm{des}}^{\mathrm{n}} = \|1 - R_{\mathrm{n}}\|$ express that each reference model has to approach as closely as possible the ideal term. This task has to be solved under the limitation $u \in \mathbb{U}$ concerning the output of the controller. Here $\mathbb{U}$ usually means the allowed domain for u, e.g. the amplitude constraints $\mathbb{U} : |u| \le 1$ [see Sect. 7.4].

Equation (7.54) constitutes a very difficult optimization task because the solution is always on the border of the constrained domain. An analytical solution, except for some low order simple cases, can not be found. The optimal reference models are usually determined by simulation *CAD* tools. Note that for the solution of the task (7.54), under the given constraints, faster reference models can not be applied. Quite the opposite: if no solution is obtained for a reference model providing the required goals under a given constraint, then there is no other possibility than to choose a less demanding (usually slower) reference model. Thus the best (fastest) reachable output of the closed-loop basically depends on the limitations of the controller output. In (7.54), of course, both the controller and the process, i.e., the whole real closed-loop, appear in a very complicated way, therefore its optimality depends on the process, on the model and also on the invariant factors.

Since the reference model is an important parameter of the general Youla design, the condition of robust stability shown in (5.45) can also be guaranteed with it. Based on (7.5) it can be seen easily that the condition (5.45) for *YP* control loops is

$$\left| Q\hat{P}\ell \right| < 1 \quad \forall \omega. \tag{7.55}$$

This condition can be further simplified to the condition

$$|R_{\mathrm{n}}| < \frac{1}{|\ell|} \quad \text{or} \quad |\ell| < \frac{1}{|R_{\mathrm{n}}|} \quad \forall \omega. \tag{7.56}$$

Based on this condition it can be stated that by choosing a first order reference model R_n, robust stability can be ensured even for the case of 100% relative model error.

Optimization of the Realizability Loss

The purpose of this task is to optimize the realizability losses J_{real}^{r} and J_{real}^{n} in terms of

$$\begin{aligned} G_r^{opt} &= \arg\left\{\min_{G_r}\left(J_{real}^{r}\right)\right\} = \arg\left\{\min_{G_r}\left\|R_r - \hat{T}_r\right\|\right\} \\ G_n^{opt} &= \arg\left\{\min_{G_n}\left(J_{real}^{n}\right)\right\} = \arg\left\{\min_{G_n}\left\|R_n - \hat{T}\right\|\right\} \end{aligned} \tag{7.57}$$

which can be ensured by the optimal choice of $G_r = G_r^{opt}$ and $G_n = G_n^{opt}$ (see Sect. 7.1). As was mentioned earlier the conditions $R_r = \hat{T}_r$ and $R_n = \hat{T}_n$ can be theoretically reached in the *ISR* case, which means the trivial solution $G_r = G_n = 1$. For the more general *IU* case, the optimal transfer functions have to be determined.

The Optimization of the Modeling Loss

The optimization of the modeling loss J_{mod}^{r} means the determination of a special optimal excitation $y_r = y_r^{opt}$ applied as a reference signal, and the optimal process model $\hat{P} = \hat{P}^{opt}$ obtained as the solution of the so-called minimax problem below

$$\hat{P}^{opt} = \arg\left\{\min_{\hat{P}}\left[\max_{y_r}\left(J_{mod}^{r}\right)\right]\right\} = \arg\left\{\min_{\hat{P}}\left[\max_{y_r}\left\|S_{mod}^{r}\right\|\right]\right\}. \tag{7.58}$$

This task has two steps: The optimal reference signal (depending on the criterion) usually provides the maximum variance in the output in the case of an amplitude constrained y_r. Measuring the output of the closed-loop the process model ensuring the minimum of the optimality criterion $\left\|S_{mod}^{r}\right\|$ of the modeling loss has to be determined by a proper modeling (identification) method. The task (7.58) is called *worst case* identification task.

It is not an easy task to optimize all the three terms simultaneously. In practice, an iteration technique is used where in a particular step the solution of only one optimality problem is solved.

(As a criterion $\|\ldots\|$ mentioned above, usually the $\mathcal{H}_2$ or $\mathcal{H}_\infty$ norms, frequently applied in the higher level control theory, are chosen. As it was mentioned earlier, these tasks are not discussed in this book, although a short description can be found in Chap. 16.)

7.5.2 Empirical rules

It has already been seen in the investigation of the best reachable control systems that the basic constraint derives from the saturation of the actuator signal and the process dynamics itself. One of the most important dynamical limits is the dead-time of the process, which can not be eliminated, since the system can not respond in a shorter time than the dead-time. The first-order PADE approximation of a dead-time lag has already been discussed, according to which

$$e^{-sT_\mathrm{d}} \approx \frac{1 - sT_\mathrm{d}/2}{1 + sT_\mathrm{d}/2} = \frac{s - 2/T_\mathrm{d}}{s + 2/T_\mathrm{d}} = \frac{s - z_\mathrm{j}}{s + z_\mathrm{j}}, \tag{7.59}$$

i.e., by the right-hand zero z_j approximation it corresponds to a time-delay lag $T_\mathrm{d} = 2/z_\mathrm{j}$. This also means, at the same time, that the right-hand zeros of the process correspond, in any way, to certain limits. An unstable zero with a small value corresponds to a high dead-time.

It can be assumed that the unstable poles of the process can also result in constraints. It can be expected that in order to stabilize an unstable process, a sufficiently fast controller is required.

Summarizing, the constraints can derive from the dead-time and the unstable zeros (z_j) and poles (p_j) (located in the right half-plane) of the process dynamics. Based on fundamental theoretical considerations and practice, the constraints are the following:

– a right-half-plane unstable (RU) zero z_j has the following limit for the cross-frequency

$$\omega_\mathrm{c} < 0.5 z_\mathrm{j} \tag{7.60}$$

A slow RU zero has especially a very bad effect.

– the dead-time also limits the cross-frequency, based on the practice

$$\omega_\mathrm{c} < 0.5 \frac{1}{T_\mathrm{d}} \tag{7.61}$$

– an *RU* pole requires a high cross-frequency

$$\omega_\mathrm{c} > 2 p_\mathrm{j} \tag{7.62}$$

- systems, having both RU zeros and poles simultaneously can not be controlled, in general, only if the poles and zeros are far enough from each other.

$$p_j > 6z_j \tag{7.63}$$

- unstable dead-time systems can not be controlled, unless the separation (distance) condition

$$p_j < 0.16\frac{1}{T_d} \tag{7.64}$$

  is fulfilled.