7

Identification and Model Reduction Techniques

7.1 Introduction

In this chapter, the issue of the system identification, in the context of PID tuning and control, is addressed. Rather than present an exhaustive review of the existing methodologies for the estimation of a (parametric or non parametric) model of the process, which would be a very huge task, the aim of the following sections is to point out possible issues that might arise when selecting the identification procedure. For this purpose, some techniques are presented and their main features are highlighted. In particular, techniques based on the evaluation either of an open-loop step response or of a relay feedback test are considered, in order to estimate the parameters of a FOPDT or a SOPDT transfer function. This choice is motivated by the fact that methods of this kind are the most adopted in practical cases because of their simplicity. The analysis focuses on self-regulating processes which do not exhibit an oscillatory dynamics.
In addition, the issue of designing a PID controller when a high-order model of the process is available is addressed. In particular, two approaches in the Internal Model Control (IMC) framework (Morari and Zafiriou, 1989) are analysed and discussed. In the first, the (high-order) controller that results from considering the high-order process model is reduced through a Maclaurin series expansion in order to obtain a PID controller. In the second, the process model is first reduced (different techniques are considered for this purpose) in order to naturally obtain a PID controller.

7.2 FOPDT Systems

The great majority of PID tuning rules actually assume that a FOPDT model of the process is available, namely the process is described by the following transfer function:

$$P(s) = \frac{K}{Ts+1}e^{-Ls} \qquad T > 0, \quad L > 0 \tag{7.1}$$

where K is the estimated gain, T is the estimated time constant and L is the estimated (apparent) dead-time. This is motivated by the fact that many processes can be described effectively by this dynamics and, most of all, that this suits well with the simple structure of a PID controller.
Different methods have been therefore proposed in the literature to estimate the three parameters by performing a simple experiment on the plant. They are typically based either on an open-loop step response or on a closed-loop relay feedback experiment.

7.2.1 Open-loop Identification Techniques

The identification techniques based on an open-loop experiment generally derive the FOPDT transfer function parameters based on the evaluation of the process step response (often denoted as the process reaction curve). This can be done in many ways. Some techniques proposed in the literature are explained hereafter with the aim of highlighting their main features.

The Tangent Method

The tangent method consists of drawing the tangent of the process response at the inflection point. Then, the process gain can be determined simply by dividing the steady-state change in the process output y by the amplitude of the input step A. Then, the apparent dead time L is determined as the time interval between the application of the step input and the intersection of the tangent line with the time axis. Finally, the value of $T+L$ is determined as the time interval between the application of the step input and the intersection of the tangent line with the straight line $y = y_\infty$ where y_∞ is the final steady-state value of the process output. Alternatively, the value of $T + L$ can be determined as the time interval between the application of the step input and the time when the process output attains the 63.2% of its final value y_∞. From this value the time constant T can be trivially calculated by subtracting the previously estimated value of the time delay L. The method is sketched in Figure 7.1.
It is worth stressing that the method is based on the fact that it gives exact results for a true FOPDT process. The main drawback of this technique is that it relies on a single point of the reaction curve (*i.e.*, the inflection point) and that it is very sensible to the measurement noise. In fact, the measurement noise might cause large errors in the estimation of the point of inflection and of the first time derivative of the process output.

The Area Method

A technique that is more robust to the measurement noise is the so-called area method. By taking into account that the process gain K can be determined

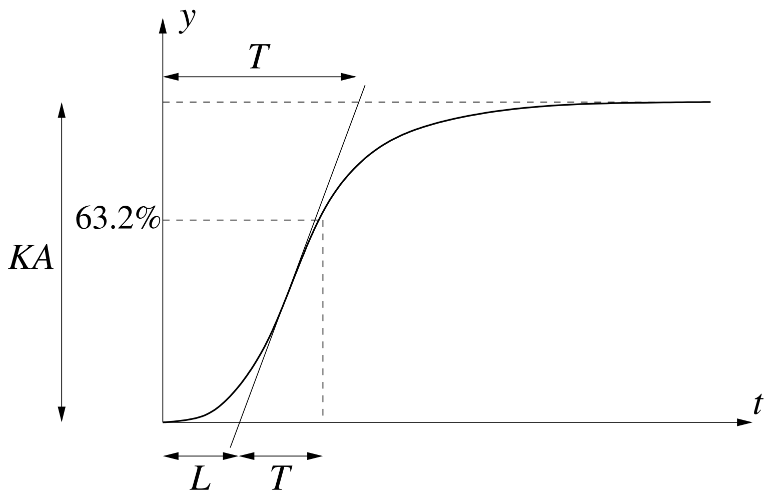

Fig. 7.1. Application of the tangent method for the estimation of a FOPDT transfer function

as for the tangent method, it consists of first calculating the area between the process output and the straight line $y = y_\infty$, namely:

$$A_1 := \int_{t_0}^{\infty} (y_\infty - y(t))dt \tag{7.2}$$

where t_0 is the time instant of the input step change. Then, the value of $T + L$ can be determined by the following expression:

$$L + T = \frac{A_1}{K}. \tag{7.3}$$

Subsequently, the area A_2 between the process output and the time axis in the time interval from t_0 to $T + L$ is evaluated, namely,

$$A_2 := \int_{t_0}^{T+L} (y(t) - y_0)dt \tag{7.4}$$

where y_0 is the initial process output steady-state value. Finally, the values of T and L are determined by means of the following expressions:

$$T = \frac{eA_2}{K} \qquad L = \frac{A_1 - KT}{K} \tag{7.5}$$

The procedure is depicted in Figure 7.2. It is worth noting that the previous expressions are derived by considering the response of a FOPDT system. In other words, as for the tangent method, a perfect parameter estimation occurs

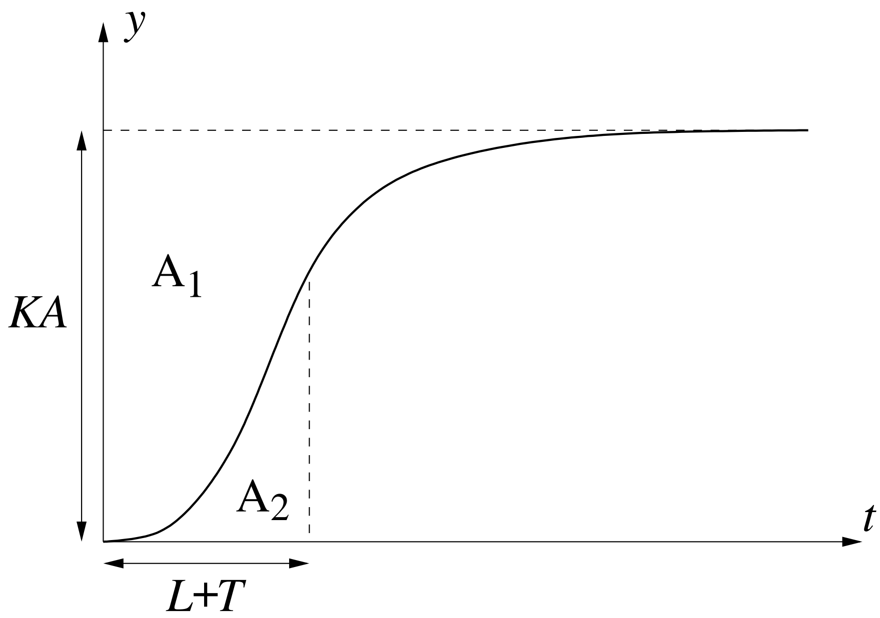

Fig. 7.2. Application of the area method for the estimation of a FOPDT transfer function

if the process has exactly a FOPDT dynamics. In any case it is also possible to apply it for (moderately) undershooting, overshooting or oscillatory responses, provided that the part of $y(t)$ that is less than y_0 be truncated to y_0 and the part of $y(t)$ that is greater than y_∞ be "mirrored" with respect to y_∞ (Leva *et al.*, 2001).

Being based on the calculus of integrals, this approach is more relevant from the computational point of view (the final result is difficult to derive by hand) but has the remarkable feature of being much more robust to the measurement noise than the tangent method. However, it has a drawback in the possible determination of a negative value of the time delay L when the process exhibits a nonlinear lag-dominant dynamics.

Consider for example the nonlinear process described by the following differential equation (note that this can be a model of a tank system where the process variable y is the fluid level, the manipulated variable u is the inflow and $Q_o = 1.2\sqrt{y}$ is the outflow):

$$\dot{y}(t) = \frac{1}{16}(u(t-1) - 1.2\sqrt{y}). \tag{7.6}$$

The unitary step response is plotted in Figure 7.3 (note that there is no measurement noise). The straightforward application of the area method gives the following results: $K = 0.69$, $T = 18.8$ and $L = -0.15$. Obviously, if

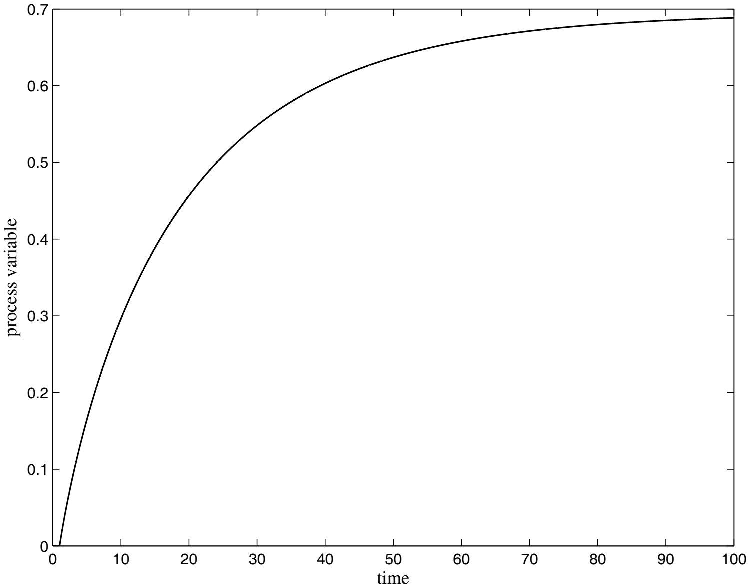

Fig. 7.3. Example of a step response for which the application of the area method gives a negative dead time

the process is not lag-dominant, a significant underestimation of the dead time results in any case. This might be a problem from the point of view of the tuning of the PID parameters because a more aggressive controller than expected might result.

Two-points-based Method

A method that is based on the estimation of two time instants of the reaction curve has been proposed in (Sundaresan and Krishnaswamy, 1978) (it is also reported in (Seborg *et al.*, 2004)). It consists in determining the time instants t_1 and t_2 when the process output attains 35.3% and 85.3% of its final steady-state respectively. Then, the dead time and the time constant are calculated by means of the following formulae:

$$T = 0.67(t_2 - t_1) \qquad L = 1.3t_1 - 0.29t_2. \tag{7.7}$$

The gain of the process is determined as in the previous methods. The previous formulae have been found, by considering many data sets, in order to minimise the difference between the experimental process response and the model response. It is worth noting that the method is very simple (indeed, it can be applied by hand easily).

This technique, in addition to the problem of being sensible to the measurement noise in the estimation of the two times t_1 and t_2, suffers from the same problem as the area method. Indeed, if it is applied to the same transient response obtained by Process (7.6) (see Figure 7.3), it results in $K = 0.69$, $T = 19.69$ and $L = -0.64$. Thus, the same considerations for the area method apply also in this case. As a consequence, care should be devoted in choosing the appropriate context for applying these techniques.

Least-squares-based (with Model Reduction) Methods

A possible way to obtain a FOPDT model is to obtain first a high-order model and then to reduce it to FOPDT form. A method that can be exploited in this context has been proposed in (Sung *et al.*, 1998). The first step is to estimate an arbitrarily high-order transfer function (denoted by $G(s)$) by means of a least-squares approach. A remarkable robustness with respect to measurement noise is achieved by considering the integrals of the input and output signals instead of their derivatives. Then, a low-order model can be derived by applying a model-reduction algorithm. A salient feature of the methodology is that it does not require any special input to the process, but it can be applied in different operating conditions.
Here the case where a step input is applied to the process and a FOPDT model is determined starting from the obtained high-order model is considered. While the value of the gain K can be found as usual by dividing the steady-state change in the process output y by the amplitude of the input step, again a least-squares-based approach is employed in order to find the time constant of the process model (7.1) that minimise the difference between the magnitude of the frequency response of the high-order model and of the FOPDT one. Formally, the value of T that satisfies the following equation is determined as:

$$|G(j\omega_i)| = \frac{K}{\sqrt{T^2\omega_i^2 + 1}}, \qquad 0 < \ldots < \omega_i < \ldots < \omega_u \tag{7.8}$$

where ω_u is the ultimate frequency of $G(s)$. Finally, the apparent dead time of the process is determined as the value that gives the same phase angle (*i.e.*, $-\pi$) of the high-order model at the ultimate frequency ω_u:

$$L = \frac{\pi - \arctan(T\omega_u)}{\omega_u}. \tag{7.9}$$

It appears that this method requires much more computational effort than the previous ones (note that ω_u has to be calculated since it is not available). Critical choices in this context are the selection of the order (and of the relative order) of the rational transfer function $G(s)$ and the portion of the step response to be considered (it is obviously meaningless to use data after the steady-state has been attained). For the first issue, by taking into account

that eventually a FOPDT model is determined, a sensible choice is to select a fourth-order transfer function with a relative order equal to one (this means that the dead time term is approximated by three zeros and three poles). For the second issue, it is sufficient to consider the settling time at 2% of the steady-state value.
It is worth noting at this point that other model-reduction methodologies (that result in a rational transfer function) will be presented in Section 7.5.

Method Based on Laguerre Functions

Laguerre functions are a set of complete orthonormal functions defined as:

$$\begin{aligned}
l_1(t) &= \sqrt{2p}e^{-pt} \\
l_2(t) &= \sqrt{2p}(-2pt+1)e^{-pt} \\
&\vdots \\
l_i(t) &= \sqrt{2p}\left[(-1)^{i-1}\frac{(2p)^{i-1}}{(i-1)!}t^{i-1} + (-1)^i\frac{(i-1)(2p)^{i-2}}{(i-2)!}t^{i-2} + \right. \\
&\quad \left. (-1)^{i-1}\frac{(i-1)(i-2)(2p)^{i-3}}{2!(i-3)!}t^{i-3} + \cdots + 1\right] e^{-pt}
\end{aligned} \tag{7.10}$$

where $p > 0$ is called the time scaling factor. In the context of system identification, the property that an arbitrary function $g(t)$ can be expanded with respect to a set of functions that is orthonormal and complete over the interval $(0, \infty)$ can be exploited. In particular, if $g(t)$ is the unit impulse response of a process, it can be written as

$$g(t) = c_1 l_1(t) + c_2 l_2(t) + \cdots + c_i l_i(t) + \cdots \tag{7.11}$$

where the c_i's are the coefficients of the expansion. By applying the Laplace transform we obtain

$$G(s) = c_1 L_1(s) + c_2 L_2(s) + \cdots + c_i L_i(s) + \cdots \tag{7.12}$$

where

$$L_i(s) = \frac{\sqrt{2p}(s-p)^{i-1}}{(s+p)^i} \qquad i = 0, 1, \ldots \tag{7.13}$$

are often referred as the Laguerre filters. In theory, the expansion expresses in Equation (7.11) requires an infinite number of terms to converge to the true impulse response. However, an arbitrarily good approximation can be obtained by truncating the series after N terms. In any case, from Expression (7.13) it can be easily deduced that the estimated transfer function has coincident poles.

Starting from the measured step response $y(t)$, the coefficients of the expansion can be calculated as:

$$c_1(t) = p\int_0^{T_s} y(t)l_1(t)dt + \bar{y}l_1(T_s)$$

$$c_2(t) = 2p\int_0^{T_s} y(t)l_1(t)dt + p\int_0^{T_s} y(t)l_2(t)dt + \bar{y}l_2(T_s)$$

$$\vdots$$

$$c_i(t) = 2p\int_0^{T_s} y(t)l_1(t)dt + 2p\int_0^{T_s} y(t)l_2(t)dt + \ldots + p\int_0^{T_s} y(t)l_i(t)dt + \bar{y}l_i(T_s) \qquad (7.14)$$

where T_s is the time at which the process attains the steady-state and $\bar{y}$ is the steady-state value. The choice of the time scaling factor p (*i.e.*, of the location of the approximating system poles) affects the accuracy of the approximation, in the sense that a poor choice of p requires more terms in order to provide a desired model accuracy. For this reason, methodologies for a sound selection of p have been investigated (Wang and Cluett, 1994). In particular, it is proposed to search for the optimal value of p (in the sense that it gives the best approximation for a given value of N) in the interval $[p_{min}, p_{max}]$, where $p_{min} = 4/T_s$ and $p_{max} = 5p_{min}$ if $N \leq 4$ and $p_{max} = 10p_{min}$ if $N > 4$. This interval is then discretised and for any value of p the Laguerre coefficients c_N and c_{N+1} are determined by means of Formulae (7.14). The values of p for which $c_N c_{N+1} = 0$ are selected as possible candidates. Among them, the one that produces the maximum value of $\sum_{i=1}^{N} c_i^2$ is selected as the best one.
A detailed analysis of the use of Laguerre functions in this context can be found in (Wang and Cluett, 2000). It is shown that this modelling technique based on the step response has nice statistical properties: it is very robust to the measurement noise and a simple strategy for the pretreating of the data can be implemented in order to cope with disturbances.
In any case, it has to be stressed that the technique requires a somewhat computational effort and the high-order model that results has to be subsequently reduced to a FOPDT model. The method presented in the previous subsection (or others presented in Section 7.5) can be used for this purpose. By applying a similar reasoning, the unique user-chosen parameter N can be chosen as equal to four.
Finally, it is worth noting that a closed-loop approach based on the use of Laguerre functions and a least-squares technique has been proposed in (Park *et al.*, 1997).

Optimisation-based Method

Another technique that is worth to being considered is to estimate the three transfer function parameters K, T and L by solving the following optimisation problem:

$$\min_{K,T,L} \int_0^{\infty} |y(t) - y_m(t)|dt \qquad (7.15)$$

where $y(t)$ denotes the experimental step response and $y_m(t)$ denotes the model step response. In other words, the model parameters are searched in order to minimise the difference between the experimental step response and the model step response.
In order to solve the posed optimisation problem, genetic algorithms (Mitchell, 1998) can be employed. In any case, obviously, the computational effort is significant and this is the major drawback of this method.

7.2.2 Closed-loop Identification Techniques

The closed-loop identification techniques employed in industrial settings typically rely on a relay-feedback experiment. The initial idea of the use of the relay-feedback controller (Åström and Hägglund, 1984) is to evaluate the obtained process output oscillation (see Section 1.3) in order to obtain a non-parametric model of the process, namely its ultimate gain K_u and the ultimate frequency ω_u, in analogy with the original idea of the ultimate sensitivity experiment of Ziegler–Nichols (Ziegler and Nichols, 1942), where the control system is led to the stability limit.

However, recently, different techniques for the determination of a FOPDT parametric model based on a relay-feedback experiment have been also devised. A few of them are presented hereafter, again with the aim of highlight possible issues that might arise when they are applied in a practical context.

Standard Relay-feedback Method

The original relay-feedback experiment proposed in (Åström and Hägglund, 1984) involves the use of a standard symmetrical relay in order to generate a persistent oscillatory response of the process output. Denoting by h the amplitude of the relay and by A the amplitude of the output oscillations, the value of the ultimate gain can be derived, by applying the describing function theory, as:

$$K_u = \frac{4h}{\pi A}. \tag{7.16}$$

The ultimate period T_u is simply the period of the obtained output oscillation. Based on these two values, many PID tuning rules can be applied (O'Dwyer, 2006). Only the amplitude h of the relay has to be selected by the user. This should be done in order to provide an output oscillation of sufficient amplitude to be well distinguished from the measurement noise, but at the same time it has not to be too high so that the process is perturbed as less as possible (and the normal production is not interrupted). Indeed, it is worth stressing that the estimation of the output oscillation is sensible to the measurement noise and therefore some filtering technique has to be applied (Wang *et al.*, 1999*c*) (this is a drawback with respect to the open-loop least-squares-based methods considered in the previous section). In addition to having just one parameter to

be selected by the user and to be performed in closed-loop, so that the process is kept close to the set-point value, the main advantage of this identification technique is that a short time is necessary to run the test (with respect to the use of a pseudo-random binary sequence (PBRS) (Ljung, 1996)). Further, possible load disturbances that might occur during the experiment can be easily detected by the change to asymmetric pulses in the control variable.
In any case, the obtained values of the ultimate gain and ultimate period are approximated, because of the adoption of the describing function theory and the estimation may not be accurate enough for some applications, for example when the process exhibits a long dead time (Li *et al.*, 1991). In order to improve the estimation of the actual values of K_u and T_u, different methods have been proposed in the literature (see, for example, (Majhi and Atherton, 2000; Atherton, 2000)). In any case, if it is desired to implement a model-based controller, the knowledge of a transfer function is required. A FOPDT transfer function can be derived by employing the following two relations, which can be derived by calculating the ultimate gain and period for Process (7.1) (Luyben, 1987):

$$T = \frac{\tan(\pi - L\omega_u)}{\omega_u}, \tag{7.17}$$

$$T = \frac{\sqrt{(KK_u)^2 - 1}}{\omega_u}. \tag{7.18}$$

It can be noted that there are two equations for three parameters. Thus, the gain of the process has to be estimated in an other way. Then, Equation (7.18) can be employed to estimate the value of T and subsequently the value of L can be determined by means of Equation (7.17). Alternatively, the dead time of the process can be estimated in an other way (for example at the beginning of the experiment, with considerations analogous to those made for the open-loop experiments) and then the time constant T and the process gain K are subsequently calculated. However, in this case the resulting time constant and process gain might incorrectly result to be negative (Vivek and Chidambaram, 2005*a*) and therefore this approach should be avoided.
In order to cope with the inaccuracies due to the presence of the describing function approximation, in (Yu, 1999) it is proposed to substitute Equation (7.17) with the following one:

$$T = \frac{\pi}{\omega_u \ln(2e^{\frac{L}{T}} - 1)}. \tag{7.19}$$

Alternative Calculation of the FOPDT Parameters

An alternative way of identifying the FOPDT transfer function by means of a symmetrical relay-feedback experiment has been proposed in (Vivek and Chidambaram, 2005*a*). It consists of first evaluating the integral

$$y(s_1) = \int_0^\infty y(t)e^{-st}dt \tag{7.20}$$

for $s_1 = 8/t_s$, where t_s is the time at which three repeated cycles of oscillations appear in the process output after the initial transient has ended. Analogously, the integral

$$u(s_1) = \int_0^\infty u(t)e^{-st}dt \tag{7.21}$$

is also evaluated for $s_1 = 8/t_s$. With the resulting values, the following equation can be posed:

$$\frac{K}{Ts_1 + 1}e^{-Ls_1} = \frac{y(s_1)}{u(s_1)}. \tag{7.22}$$

Then, the frequency response of the process transfer function can be written as

$$P(j\omega_u) = \frac{y(j\omega_u)}{u(j\omega_u)} = \frac{c_1 - jd_1}{c_2 - jd_2} \tag{7.23}$$

where

$$\begin{aligned} c_1 &= \textstyle\int_0^{T_u} y(t)\cos(\omega_u t)dt \\ d_1 &= \textstyle\int_0^{T_u} y(t)\sin(\omega_u t)dt \\ c_2 &= \textstyle\int_0^{T_u} u(t)\cos(\omega_u t)dt \\ d_2 &= \textstyle\int_0^{T_u} u(t)\sin(\omega_u t)dt \end{aligned} \tag{7.24}$$

where ω_u is the frequency of the oscillation obtained in the process output and $T_u = 2\pi/\omega_u$. The values of c_1, d_1, c_2 and d_2 can be evaluated numerically based on the process input and output data $u(t)$ and $y(t)$ obtained from the relay test. Thus, Equation (7.23) can be rewritten as

$$P(j\omega_u) = p + jq \tag{7.25}$$

where

$$p = \frac{c_1c_2 + d_1d_2}{c_2^2 + d_2^2} \qquad q = \frac{d_2c_1 - d_1c_2}{c_2^2 + d_2^2}. \tag{7.26}$$

By taking into account that

$$P(j\omega_u) = \frac{K}{Tj\omega_u + 1}e^{-Lj\omega_u} \tag{7.27}$$

it can be easily deduced that

$$p + jq = \frac{K(\cos(L\omega_u) - j\sin(L\omega_u))}{Tj\omega_u + 1}. \tag{7.28}$$

Finally, by equating the real and imaginary parts, it can be written

$$p - q\omega_u T - K\cos(L\omega_u) = 0, \tag{7.29}$$

$$q + p\omega_u T + K\sin(L\omega_u) = 0. \tag{7.30}$$

The three process parameters can be obtained by means of Equations (7.22), (7.29) and (7.30). It is worth stressing that a numerical solution has to be derived. In any case the merit of the methodology is that all the three parameters can be found with a single relay test.

Use of an Asymmetrical Relay

If a biased relay is adopted for the experiment, the process gain K can be determined by using the process input and output data $u(t)$ and $y(t)$ according to the expression (Shen *et al.*, 1996):

$$K = \frac{\int_0^{2\pi} e(t)d(\omega_u t)}{\int_0^{2\pi} u(t)d(\omega_u t)}. \tag{7.31}$$

Then, the other process parameters T and L can be calculated by means of Equations (7.29) and (7.30), for which an analytical solution exists (Srinivasan and Chidambaram, 2003). Obviously, in this case both the up-amplitude and the down-amplitude of the relay have to be selected. Further, the use of an asymmetrical relay represents a sort of disturbance to the process since it cause the operating point to drift.

Use of a Relay with Hysteresis

As already mentioned, the relay-feedback test is sensitive to the measurement noise. The easiest way to reduce the influence of the noise is to employ a relay with a hysteresis, whose width is usually chosen as twice the noise band. Denoting again by A and T_u the amplitude and the period of the resulting oscillation, and assuming that the process gain K is known, the process time constant can be determined as (Wang *et al.*, 1997):

$$T = \frac{1}{2}T_u \left(\ln \frac{hK + A}{hK - A} \right)^{-1} \tag{7.32}$$

where h is the amplitude of the (symmetrical) relay. Then, the dead time is estimated as

$$L = \frac{1}{2}T_u \left(\ln \frac{hK - \varepsilon}{hK - A} \right) \left(\ln \frac{hK + A}{hK - A} \right)^{-1} \tag{7.33}$$

where ε is the width of the hysteresis. It is worth stressing that the method requires a previous estimation of the process gain.

Use of a Biased Relay with Hysteresis

The identification of a FOPDT process can be performed also by means of a biased relay (Wang *et al.*, 1997) (see Figure 7.4). By denoting the periods and the amplitudes of oscillations as shown in Figure 7.5, the process parameters can be determined as follows. First, the process gain is calculated again as:

$$K = \frac{\int_0^{T_{u1}+T_{u2}} y(t)dt}{\int_0^{T_{u1}+T_{u2}} u(t)dt}. \tag{7.34}$$

Then, the normalised dead time $\Theta = L/T$ is obtained as:

$$\Theta = \ln \frac{(h+h_0)K - \varepsilon}{(h+h_0)K - A_u} \tag{7.35}$$

or

$$\Theta = \ln \frac{(h-h_0)K - \varepsilon}{(h+h_0)K + A_d}. \tag{7.36}$$

Then, the process time constant can be calculated as

$$T = T_{u1}\left(\ln \frac{2hKe^{\Theta} + h_0K - hK + \varepsilon}{h_0K + hK - \varepsilon}\right)^{-1} \tag{7.37}$$

or

$$T = T_{u1}\left(\ln \frac{2hKe^{\Theta} - h_0K - hK + \varepsilon}{h_0K - hK - \varepsilon}\right)^{-1}. \tag{7.38}$$

The dead time is finally determined by simply calculating $L = \Theta T$. As already mentioned for the simple asymmetrical relay, the technique suffers from the drawback of drifting the process away from the operating point.

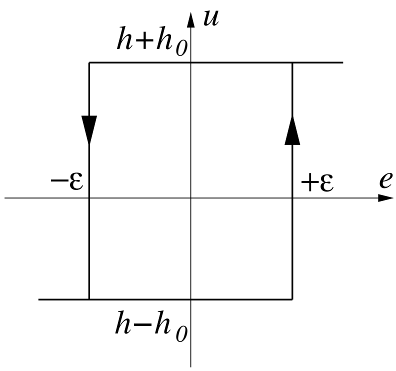

Fig. 7.4. The biased relay

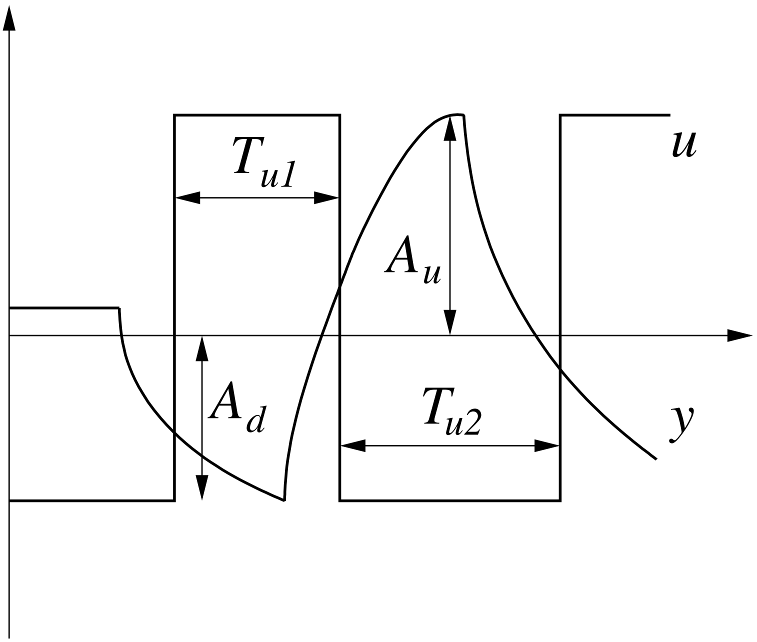

Fig. 7.5. Illustrative example of the use of a biased relay

Method Based on the Output Curve Shape

In the previous sections, it has been shown how to estimate a FOPDT transfer function starting from the data collected after a unique relay-feedback test. However, it has been recognised by many author that for some processes, in particular those with a large dead time, the knowledge of just the ultimate gain and of the ultimate frequency is actually insufficient for the effective design of the (PID) controller.

Indeed, it has been shown that the shape of the output oscillation depends on the process dynamics and analytical expressions are derived in (Panda and Yu, 2003). This fact has been exploited in (Luyben, 2001*b*), where the curve shape obtained by a standard (not biased and without hysteresis) relay is analysed in order to derive a FOPDT model. In particular, the following algorithm is proposed (see Figure 7.6), where A denotes, as usual, the amplitude of the oscillations.

1. Determine $K_u = 4h/(\pi A)$, where h is the relay amplitude, and evaluate the ultimate period T_u (equivalently, the ultimate frequency ω_u).
2. Draw a vertical line passing through the peak in the curve and denote the corresponding time as t_2.
3. Draw a horizontal line at $A/2$.

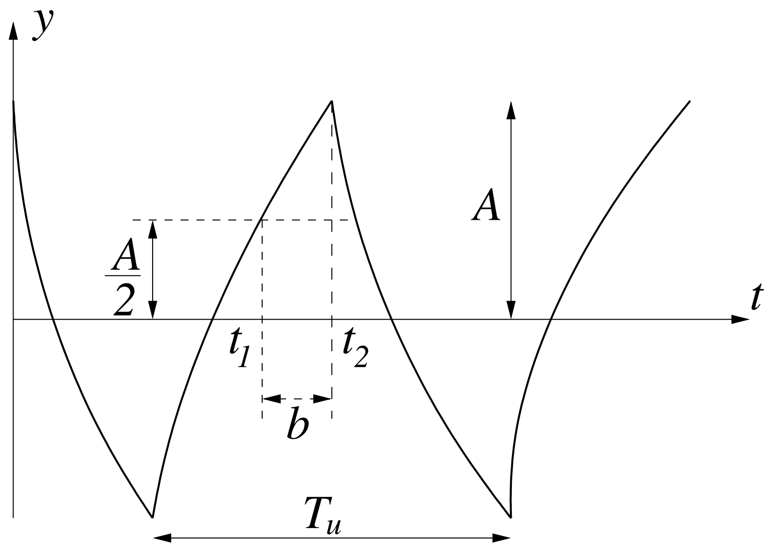

Fig. 7.6. Illustration of the method based on the output curve shape

4. Draw a vertical line passing through the intersection of the curve with the line drawn at step 3 and denote the corresponding time as t_1.
5. Set $b = t_2 - t_1$.
6. Calculate a curvature factor F as $F = 4b/T_u$. This actually indicates if the curve has a shape more similar to a triangle (this results when the dead time is small with respect to the time constant) or more similar to a rectangle (when the dead time is big).
7. Calculate $R := L/T$ by means of the following expression (determined by interpolating results for different processes):

$$\ln\left(\frac{L}{T}\right) = -5.2783 + 12.7147F - 9.8974F^2 + 2.6788F^3. \tag{7.39}$$

8. Substitute $L = RT$ in the following equation

$$-\omega_u L - \arctan(\omega_u T) = -\pi. \tag{7.40}$$

and solve iteratively for T. Then, determine $L = RT$.
9. Determine K by means of the equation

$$\frac{K}{\sqrt{1 + (\omega_u T)^2}} = \frac{1}{K_u}. \tag{7.41}$$

The technique has the great merit of exploiting, in a simple way, the shape of process variable oscillation. Possible drawbacks of the method are its sensitivity to the noise, the somewhat significant computational effort and a possibly inaccurate estimation of the process gain (for example, for the noise-free

step response of the process $P(s) = 1/(s+1)^2 e^{-s}$ the results are $K = 2.97$, $T = 5.71$ and $L = 1.29$). Thus, an appropriate tuning procedure should be applied in this context (Scali *et al.*, 1999).
It is worth stressing that the idea of exploiting the shape factor has been developed in (Thyagarajan and Yu, 2003; Panda and Yu, 2005) where the model structure (FOPDT or SOPDT) is also conveniently selected based on the shape of the obtained oscillation.

7.3 SOPDT Systems

Even if the majority of the existing tuning rules are based on FOPDT transfer functions of the process, there are also many rules that relies on the estimation of SOPDT transfer functions (Panda *et al.*, 2004), as they include overdamped, critically damped and underdamped systems and the presence of the two poles can be handled by the two zeros of the controller.
Usually, such a transfer function can be expressed in two ways, namely:

$$P(s) = \frac{K}{(T_1 s + 1)(T_2 s + 1)} e^{-Ls} \qquad T_1 > T_2 > 0, \quad L > 0 \tag{7.42}$$

or, alternatively,

$$P(s) = \frac{K}{T^2 s^2 + 2\xi T s + 1} e^{-Ls} \qquad T > 0, \quad \xi > 0, \quad L > 0. \tag{7.43}$$

It has to be noted that Expression (7.43) is more general than Expression (7.42) since it includes the cases of both real and complex conjugate poles (when $\xi \geq 1$ and $\xi < 1$ respectively), while in (7.42) the poles are assumed to be real. However, this latter case is highlighted since it is significant in the context of PID control, as many tuning rules assume that the PID controller is in a series form and the derivative time constant is selected in order to cancel the pole associated with the smallest time constant (see for example (Skogestad, 2003)). Further, processes that present an oscillatory dynamics are rarely found in industrial settings, although the case is of concern when a closed-loop dynamics is considered (for example, the secondary loop of a cascade control system, see Section 9.2.1).
Note also that, for the sake of simplicity, it has been assumed that the process has no zeros. This fact is in any case briefly addressed hereafter. As for the estimation of a FOPDT model, both (step-based) open-loop identification techniques and (relay-feedback-based) closed-loop identification techniques are addressed.

7.3.1 Open-loop Identification Techniques

As for FOPDT models, different techniques have been devised for the estimation of a SOPDT transfer function by evaluating open-loop process step responses (not necessarily for the purpose of tuning a PID controller). Some of them are reviewed hereafter, always with the aim of highlighting their practical issues in the context of PID control.

Two-points-based Method

The method described in (Åström and Hägglund, 1995) is based on the numerical solution of two equations that, in case of an overdamped (monotonic) step response, imposes that the experimental step responses and that provided by Model (7.42) matches exactly when the process output attains 33% and 67% of its final value. The (apparent) dead-time L is previously determined by applying the tangent method (*i.e.*, by considering the intersection between the baseline and the tangent line of the response in its inflection point), while the process gain K is previously calculated as usual by dividing the steady-state change in the process output by the amplitude of the input step.
Being based on the selection of single points in the step response, the method is sensitive to the measurement noise and some filtering technique might be required. Possible problems with this technique arise when a process with two coincident poles and a time delay is considered. For example, if the (noise-free) step response of the process

$$P(s) = \frac{1}{(2s+1)^2}e^{-s} \tag{7.44}$$

is considered, the estimated parameters of Model (7.42) are $K = 1$, $T_1 = 3.14$, $T_2 = 0.52$ and $L = 1.56$, which are quite different from the actual ones (indeed, it seems that the dominant dynamics is of first order). Some problems occur also when the dominant dynamics of the process is of first order. For example, consider the step response of the process

$$P(s) = \frac{1}{(0.1s+1)(0.1^2s+1)(0.1^3s+1)(0.1^4s+1)}. \tag{7.45}$$

In this case the result of the application of the method is $K = 1$, $T_1 = 0.05$, $T_2 = 0.05$, $L = 0.01$. It appears that the estimated process has a dominant dynamics of second order. Obviously, the effectiveness of the identification methodology has to be evaluated in conjunction with the employed tuning procedure. However, from the above considerations it might be useful to employ this technique with another one devoted to the estimation of FOPDT processes and to evaluate which of the two estimated transfer functions fits better the experimental data. In case an oscillatory response is detected, in order to estimate T and ξ, two solutions can be adopted. In the first one, the parameters of Model (7.43) are determined by imposing that the step response of the estimated model attains the same peak amplitude y_M at the same time t_M of the experimental response. The time constant T and the damping ratio ξ are therefore determined as:

$$\xi = \frac{\eta}{\sqrt{1+\eta^2}} \tag{7.46}$$

and

$$T = \frac{1}{\pi}\sqrt{\frac{t_M}{1+\xi^2}} \tag{7.47}$$

where

$$\eta = \left|\frac{\ln(y_M - y_\infty)}{\pi}\right| \tag{7.48}$$

where y_∞ is the final steady-state value of the step response.
In the second case, the values of the first minimum y_m (attained at time t_m) and of the second maximum y_{M2} (attained at time t_{M2}) are also considered. Once the decay ratio is determined as

$$d = \frac{y_{M2} - y_M}{y_m - y_M}, \tag{7.49}$$

the two model parameters are calculated by means of the following equations:

$$\varphi = \left|\frac{\log(1-d)}{\pi}\right|, \tag{7.50}$$

$$\xi = \frac{\varphi}{\sqrt{1+\varphi^2}}, \tag{7.51}$$

$$T = \frac{(t_{M2} - t_M)\sqrt{1-\xi^2}}{2\pi}. \tag{7.52}$$

Note that in both cases the dead time is determined as for overdamped responses.

Harriot's Method

The method proposed in (Harriot, 1964), and described also in (Johnson and Moradi (eds.), 2005), is based on the fact that almost all the step responses of processes described by transfer function (7.42) reach 73% of their steady-state values approximately at a time of $1.3(T_1 + T_2)$ and separate from each other most widely at time $0.5(T_1 + T_2)$. Oscillatory responses are not addressed in this case. Thus, the technique consists of first determining the value of A_1 according to Expression (7.2). Then, the value of the dead time can be estimated from the following equation

$$L = A_1 - \frac{t_{73}}{1.3} \tag{7.53}$$

where t_{73} is the time at which the process output attains the 73% of its final value. Then, the sum of the two time constants $T_1 + T_2$ can be derived as

$$T_1 + T_2 = A_1 - L. \tag{7.54}$$

At this point it is possible to evaluate the value y^* of the step response at time $t = 0.5(T_1 + T_2)$. From the plot of Figure 7.7, the value of the ratio $r = T_1/(T_1+T_2)$ can be derived (note that the plot can be easily reconstructed by considering different systems with different values of r). Finally, the values of T_1 and T_2 are determined as

$$T_1 = r\frac{t_{73} - t_0}{1.3} \tag{7.55}$$

and

$$T_2 = (1 - r)\frac{t_{73} - t_0}{1.3}. \tag{7.56}$$

Although Harriot's method is somewhat robust to the measurement noise, its main drawback is that it might result in an estimation of a small dead time value, with respect to other methods. This might imply that the resulting controller is more aggressive than expected and this fact is actually detrimental in practical cases.
Indeed, in some cases it might occur that a negative value of the time delay

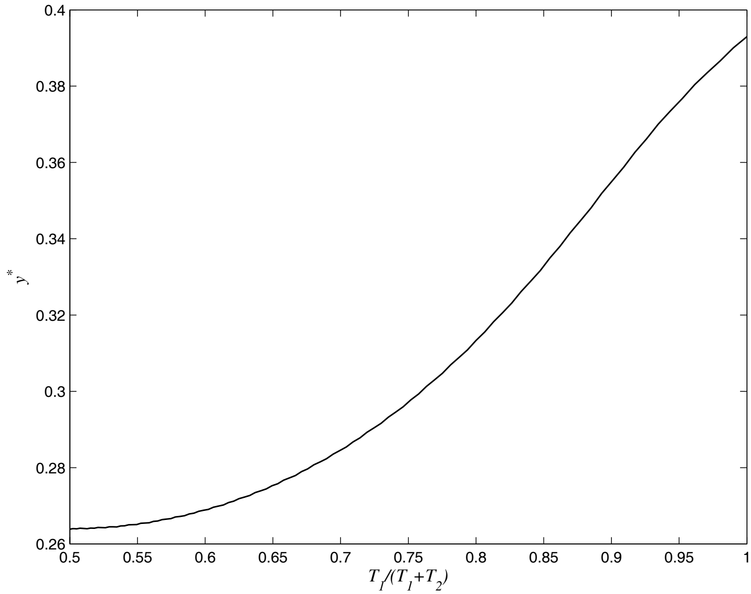

Fig. 7.7. Relation between y^* and $T_1/(T_1 + T_2)$

results (as in the case of the area method for FOPDT processes). For example, if the (noise-free) step response of the process

$$P(s) = \frac{1}{(10s+1)^2} e^{-s} \tag{7.57}$$

is considered, the resulting estimated parameters of the Model (7.42) are $K = 1$, $T_1 = 10.17$, $T_2 = 10.17$ and $L = -0.23$.

Area–Tangent Method

In the technique described in (Sundaresan *et al.*, 1978) both monotonic and oscillatory step responses are considered and one of the two models (7.42)–(7.43) is automatically selected (as already mentioned, this implies that only Model (7.43) can be adopted and if $\xi \geq 1$ then Model (7.42) can be easily derived). After having calculated the process gain K as usual by looking at the input and output steady-state values, the estimation procedure consists of first determining the area between the process output and the straight line $y = y_\infty$, namely:

$$A_1 := \int_{t_0}^{\infty} (y_\infty - y(t))dt \tag{7.58}$$

where t_0 is the time instant of the input step change. Then, the tangent of the process response is drawn at the inflection point and its slope is denoted by M_i and its intersection with the straight line $y = y_\infty$ is denoted as t_m. An auxiliary variable $\lambda = (t_m - A_1)M_i$ can be easily calculated for the purpose of selecting Model (7.42) or Model (7.43). In particular, if $\lambda < e^{-1}$, then Model (7.42) is considered and its parameters are determined by means of the following formulae (which are derived starting from the analytical expression of the step response):

$$T_1 = \frac{\eta^{\frac{\eta}{1-\eta}}}{M_i} \tag{7.59}$$

$$T_2 = \frac{\eta^{\frac{1}{1-\eta}}}{M_i} \tag{7.60}$$

$$L = A_1 - \frac{\eta^{\frac{1}{1-\eta}}}{M_i}\frac{\eta+1}{\eta} \tag{7.61}$$

where the auxiliary variable η is determined as the solution of the equation

$$\lambda = \ln\left(\frac{\eta}{\eta-1}\right) \exp\left(-\frac{\eta}{\eta-1}\right). \tag{7.62}$$

The case $\lambda = e^{-1}$ corresponds to a critically damped system, for which the previous expressions reduce to

$$T_1 = T_2 = \frac{1}{M_i e} \tag{7.63}$$

and

$$L = A_1 - \frac{2}{M_i e}. \tag{7.64}$$

Finally, if $\lambda > e^{-1}$, an oscillatory dynamics results and Model (7.43) is selected. Its parameters are selected by solving the following equations:

$$\lambda = \frac{\cos^{-1}\xi}{\sqrt{1-\xi^2}} \exp\left(\frac{-\xi}{\sqrt{1-\xi^2}} \cos^{-1}\xi\right), \tag{7.65}$$

$$T = \frac{\sqrt{1-\xi^2}}{\cos^{-1}\xi}(t_m - A_1), \tag{7.66}$$

$$L = A_1 - 2\xi T. \tag{7.67}$$

It appears that the technique, being based also on the drawing of the tangent line in the inflection point, has a somewhat high noise sensitivity. It requires also a somewhat significant computational effort (a few equations have to be solved numerically). Further, it has to be stressed that Model (7.43) may result even if a monotonic step response occurs. For example, if the (noise-free) step response of the process

$$P(s) = \frac{1}{(s+1)^3} \tag{7.68}$$

is considered, the resulting estimated parameters are $K = 1$, $T = 0.69$, $\xi = 0.77$ and $L = 1.94$. As already mentioned, this fact is relevant especially if it is intended to employ a PID controller in series (interacting) form, since in this case the design is often based on pole-zero cancellation.

Four-points-based Method

The methodology proposed in (Huang and Huang, 1993) provides a SOPDT model expressed in the form (7.43) by evaluating four points of the process step response. The algorithm can be summarised as follows (equations are derived by applying a least-squares method).

1. Determine the process gain K by dividing the steady-state change in the process output by the amplitude of the step input.
2. Calculate

$$\alpha = \frac{t_9 - t_6}{t_3 - t_1} \tag{7.69}$$

   where t_1, t_3, t_6, t_9 are the time at which the step response attains 10%, 30%, 60%, 90% of its final value.

3. Calculate ξ as

$$\xi = 7.40898 \cdot 10^{-40} e^{16.3329\alpha} + \frac{100\alpha}{4.55048\alpha + 1.57083} + 1.79015 \cdot 10^{-2}\alpha^3$$
$$+2.25401 \cdot 10^{-2}\alpha^2 - 1.14789\alpha - 16.007 \tag{7.70}$$

which has a usable range $2.005 \leq \alpha \leq 5.508$ ($0.707 \leq \xi \leq 3.0$).

4. Calculate T as

$$T = \frac{4\sum t_i f_i(\xi) - \sum f_i(\xi) \sum t_i}{4\sum f_i^2(\xi) - \left(\sum f_i(\xi)\right)^2} \tag{7.71}$$

and L as

$$L = \frac{\sum t_i \sum f_i^2(\xi) - \sum f_i(\xi) \sum t_i f_i(\xi)}{4\sum f_i^2(\xi) - \left(\sum f_i(\xi)\right)^2} \tag{7.72}$$

where

$$\sum t_i = t_1 + t_3 + t_6 + t_9, \tag{7.73}$$

$$\sum f_i(\xi) = f_1(\xi) + f_3(\xi) + f_6(\xi) + f_9(\xi), \tag{7.74}$$

$$\sum f_i^2(\xi) = f_1^2(\xi) + f_3^2(\xi) + f_6^2(\xi) + f_9^2(\xi), \tag{7.75}$$

$$\sum t_i f_i(\xi) = t_1 f_1(\xi) + t_3 f_3(\xi) + t_6 f_6(\xi) + t_9 f_9(\xi), \tag{7.76}$$

and

$$f_1(\xi) = 0.45465 + 0.06033\xi + 0.01674\xi^2, \tag{7.77}$$

$$f_3(\xi) = 0.848967 + 0.071809\xi + 0.19753\xi^2 - 0.021823\xi^3, \tag{7.78}$$

$$f_6(\xi) = 1.08111 + 0.40977\xi + 0.634313\xi^2 - 0.093324\xi^3, \tag{7.79}$$

$$f_9(\xi) = 0.581618 + 0.875726\xi + 3.64626\xi^2 - 1.35143\xi^3 + 0.173916\xi^4. \tag{7.80}$$

It is worth noting that, as for the area–tangent method, Model (7.43) (with $\xi \geq 0.707$) may result even if the method deals only with monotonic step responses. Thus, it is not possible to apply a tuning rule for a series PID controller where the derivative action is employed to cancel a pole of the process. Indeed, this is in accordance to the fact that the four-points-based method proposed in (Huang and Huang, 1993) aims at estimating a SOPDT transfer function without any relationship with the tuning of a PID controller and it is recognised that the range $0.707 \leq \xi < 1$ can be applied also to nonoscillatory processes.

Three-points-based Method

Similar to the four-points-based method, a three-points-based method has been developed (on a more theoretical basis) in (Rangaiah and Krishnaswamy, 1994). It consists of finding the parameters of the Model (7.43) by applying the following algorithm:

1. Determine the process gain K by dividing the steady-state change in the process output by the amplitude of the step input.
2. Calculate
$$\alpha = \frac{t_3 - t_2}{t_2 - t_1} \tag{7.81}$$
where t_1, t_2 and t_3 are the time at which the step response attains 14%, 55%, and 91% of its final value.
3. Calculate β and ξ as
$$\beta = \ln\left(\frac{\alpha}{2.485 - \alpha}\right) \tag{7.82}$$
$$\begin{aligned}\xi = &\ 0.50906 + 0.51743\beta - 0.076284\beta^2 + 0.041363\beta^3 \\ &-0.0049224\beta^4 + 0.00021234\beta^5\end{aligned} \tag{7.83}$$
which has a usable range $1.2323 < \alpha < 2.4850$ that corresponds to $0.707 < \xi < 3.0$.
4. Calculate T and L from the following equations
$$\frac{t_2 - t_1}{T} = 0.85818 - 0.62907\xi + 1.2897\xi^2 - 0.36859\xi^3 + 0.038891\xi^4, \tag{7.84}$$
$$\frac{t_2 - L}{T} = 1.3920 - 0.52536\xi + 1.2991\xi^2 - 0.36859\xi^3 + 0.037605\xi^4. \tag{7.85}$$

The method appears to be simpler than the four-points-based methods but similar considerations can be applied, since also in this case it is assumed that that a model with two complex-conjugate poles (with a damping factor greater than 0.707) can accurately model a process with an overdamped step response.

Method with Model Structure Identification

In the method proposed in (Huang *et al.*, 2001), the model structure is selected according to the shape of the step response. In particular, two model structures are considered, namely,

$$P(s) = \frac{K(as+1)}{(Ts+1)(\eta Ts+1)}e^{-Ls} \qquad 0 < \eta \leq 1, \tag{7.86}$$

and

$$P(s) = \frac{K(as+1)}{T^2s^2 + 2\xi Ts + 1} e^{-Ls} \qquad 0 < \xi < 1. \tag{7.87}$$

It can be remarked that processes with a positive zero (*i.e.*, with inverse response) and with a negative zero can be addressed by this method. Here, for the sake of simplicity, the analysis is restricted to the case of processes with a monotonic step response, for which it is set $a = 0$. Note that in this case it is trivial to derive Model (7.42) from Model (7.86) by simply setting $T_2 = \eta T_1$. Then, the following algorithm is applied.

1. Determine the process gain K by dividing the steady-state change in the process output by the amplitude of the step input.
2. Calculate
$$R_{0.5} = \frac{A_1 - t_{0.3}}{t_{0.5} - t_{0.3}} \tag{7.88}$$
$$R_{0.9} = \frac{A_1 - t_{0.7}}{t_{0.9} - t_{0.7}} \tag{7.89}$$
where A_1 is determined as in Equation (7.58) and t_x is the time when $y(t_x)/y_\infty = x$ (*i.e.*, the time when the process output attains the x% of its steady-state value).
3. If $1.5573 < R_{0.5} < 1.9108$ and $-0.303 < R_{0.9} < -0.0736$ then select Model (7.86) and determine $\eta_{0.5}$ and $\eta_{0.9}$ by solving the following equations:
$$\begin{aligned} R_{0.5} = {} & 1.9108 + 0.2275\eta_{0.5} - 5.5504\eta_{0.5}^2 + 12.8123\eta_{0.5}^3 \\ & -11.8164\eta_{0.5}^4 + 3.9735\eta_{0.5}^5 \end{aligned} \tag{7.90}$$
$$\begin{aligned} R_{0.9} = {} & -0.1871 + 0.0736\eta_{0.9} - 1.2329\eta_{0.9}^2 + 2.1814\eta_{0.9}^3 \\ & -1.5317\eta_{0.9}^4 + 0.3937\eta_{0.9}^5 \end{aligned} \tag{7.91}$$
else, select Model (7.87) and determine $\xi_{0.5}$ and $\xi_{0.9}$ by solving the following equations:
$$R_{0.5} = -3.1623 + 9.3343\xi_{0.5} - 5.7804\xi_{0.5}^2 + 1.1588\xi_{0.5}^3 \tag{7.92}$$
$$R_{0.9} = -6.1991 + 14.6087\xi_{0.5} - 12.1250\xi_{0.5}^2 + 3.4080\xi_{0.5}^3 \tag{7.93}$$
4. Calculate
$$\eta = \frac{\eta_{0.5} + \eta_{0.9}}{2} \tag{7.94}$$
or
$$\xi = \frac{\xi_{0.5} + \xi_{0.9}}{2}. \tag{7.95}$$
5. If Model (7.86) is selected, then determine
$$\bar{t}_{0.3} = 0.3548 + 1.1211\eta - 0.5914\eta^2 + 0.2145\eta^3 \tag{7.96}$$

$$\bar{t}_{0.5} = 0.6862 + 1.1682\eta - 0.1704\eta^2 + 0.0079\eta^3 \tag{7.97}$$

$$\bar{t}_{0.7} = 1.1988 + 1.0818\eta - 0.4043\eta^2 - 0.2501\eta^3 \tag{7.98}$$

$$\bar{t}_{0.9} = 2.3063 + 0.9017\eta + 1.0214\eta^2 + 0.3401\eta^3 \tag{7.99}$$

else, (if Model (7.87) is selected) determine

$$\bar{t}_{0.3} = 0.7954 + 0.2204\xi + 0.0631\xi^2 + 0.0184\xi^3 \tag{7.100}$$

$$\bar{t}_{0.5} = 1.0472 + 0.3952\xi + 0.1577\xi^2 + 0.0784\xi^3 \tag{7.101}$$

$$\bar{t}_{0.7} = 1.2662 + 0.6045\xi + 0.2834\xi^2 + 0.2868\xi^3 \tag{7.102}$$

$$\bar{t}_{0.9} = 1.4655 + 0.9862\xi - 0.1236\xi^2 + 1.5732\xi^3 \tag{7.103}$$

6. Calculate

$$T = \frac{1}{3}\left(\frac{t_{0.9} - t_{0.7}}{\bar{t}_{0.9} - \bar{t}_{0.7}} + \frac{t_{0.7} - t_{0.5}}{\bar{t}_{0.7} - \bar{t}_{0.5}} + \frac{t_{0.5} - t_{0.3}}{\bar{t}_{0.5} - \bar{t}_{0.3}}\right). \tag{7.104}$$

7. Calculate

$$L = \frac{t_{0.9} + t_{0.7} + t_{0.5} + t_{0.3}}{4} - \frac{\bar{t}_{0.9} + \bar{t}_{0.7} + \bar{t}_{0.5} + \bar{t}_{0.3}}{4} \tag{7.105}$$

In the application of this method it has to be taken into account that, although the model structure is automatically selected, a model with two complex conjugate poles might result even for a monotonic step response. Further, the resulting time delay can be negative. For example, if the process described by the transfer function (7.57) is considered again, the resulting values of the estimated SOPDT transfer function are $K = 1$, $T = 12.24$, $\xi = 0.85$ and $L = -0.97$ (note that a noise-free step response has been evaluated). Similarly to Harriot's method, it can be deduced that, in general, the estimated dead time is quite small. This might imply that the resulting controller is more aggressive than expected (especially if a tuning rule based on the normalised dead time is selected).

Least-squares-based Method

A least-squares method that provides directly a SOPDT transfer function from a step response without any iteration has been presented in (Wang *et al.*, 2001; Wang and Zhang, 2001). Indeed, with respect to the method presented in (Sung *et al.*, 1998) (see Section 7.2.1), it does not require a model reduction phase, since the process parameters are determined directly from the least-squares equations. It can be remarked that the approach is very robust to the measurement noise, being based on the use of process output integrals in the regression equations.
The method assumes that the process is described by the following SOPDT transfer function

$$P(s) = \frac{b_1 s + b_2}{s^2 + a_1 s + a_2} e^{-Ls} \tag{7.106}$$

and therefore processes with a stable and an unstable zero are considered. The following definitions are required:

$$\gamma(t) = y(t), \tag{7.107}$$

$$\phi^T(t) = \left[-\int_0^t y(\tau)d\tau, -\int_0^t \int_0^\tau y(\tau_1)d\tau_1 d\tau, A, tA, \frac{1}{2}t^2 A \right], \tag{7.108}$$

$$\theta^T = \left[a_1, a_2, -b_1 L + \frac{1}{2} b_2 L^2, b_1 - b_2 L, b_2 \right]. \tag{7.109}$$

After choosing t_i, $i = 1, 2, \dots, N$ such that $L \leq t_1 < t_2 < \cdots < t_N$, let

$$\Gamma = [\gamma(t_1), \gamma(t_2), \dots, \gamma(t_N)], \tag{7.110}$$

and

$$\Phi = [\phi(t_1), \phi(t_2), \dots, \phi(t_N)]^T. \tag{7.111}$$

Then, the equation

$$\Gamma = \Phi\theta \tag{7.112}$$

can be solved by applying the ordinary least-squares approach. From the resulting vector θ, the process parameters can be found from Equation (7.109):

$$\begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \\ L \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \beta \\ \theta_5 \\ \dfrac{-\theta_4 + \beta}{\theta_5} \end{bmatrix} \tag{7.113}$$

where

$$\beta = \begin{cases} -\sqrt{\theta_4^2 - 2\theta_5\theta_3} & \text{if an inverse response is detected} \\ \sqrt{\theta_4^2 - 2\theta_5\theta_3} & \text{otherwise} \end{cases} \tag{7.114}$$

Practical issues, such as how to cope effectively with the measurement noise and how to choose t_1, t_N and N are given in (Wang *et al.*, 2001; Wang and Zhang, 2001). Further, the knowledge of the process gain can be easily exploited by slightly modifying the technique. In any case, the methodology can be easily applied also by assuming a process model of second order without any zero (see (7.42)). However, it is worth stressing that, as for other techniques, it is not guaranteed that a process with two real poles results when a monotonic step response is considered.

Optimisation-based Method

Analogously to what has been explained for FOPDT transfer functions (see Section 7.2.1), the transfer function parameters K, T_1, T_2 and L (or, alternatively, K, T, ξ and L) can be obtained by solving the following optimisation problem:

$$\min_{K,T_1,T_2,L} \int_0^{\infty} |y(t) - y_m(t)|dt \tag{7.115}$$

where $y(t)$ denotes the experimental step response and $y_m(t)$ denotes the model step response. Actually, the model parameters are searched in order to minimise the difference between the experimental step response and the model step response.
A practical way to solve the posed optimisation problem is to use genetic algorithms (Mitchell, 1998). In this way, it can be easily imposed that the resulting process poles be real, but, in any case, as already mentioned the computational effort is significant and this is the major drawback of this method.

7.3.2 Closed-loop Identification Techniques

Identification techniques based on closed-loop experiments can be employed also for the estimation of SOPDT transfer functions. For example, the use of an asymmetrical relay-feedback experiment is suggested in (Ramakrishnan and Chidambaram, 2003). The approach is similar to that described in (Srinivasan and Chidambaram, 2003) for FOPDT system (see Section 7.2.2). It consists of assuming that the process is described by Model (7.42) (the extension to Model (7.43) is trivial). Then, the obtained (ultimate) period of oscillations is denoted by T_u. The process gain can be determined by calculating

$$K = \frac{\int_{t_0}^{t_0+T_u} y(t)dt}{\int_{t_0}^{t_0+T_u} u(t)dt}. \tag{7.116}$$

At this point, the following equation can be written:

$$\frac{K}{(T_1 s_1 + 1)(T_2 s_1 + 1)} e^{-Ls_1} = \frac{y(s_1)}{u(s_1)} \tag{7.117}$$

where $u(s_1)$ and $y(s_1)$ are determined as in (7.20) and (7.21). Then, other two equations can be considered, namely,

$$p(1 - T_1 T_2 \omega_u^2) - q\omega_u(T_1 + T_2) - K\cos(L\omega_u) = 0 \tag{7.118}$$

and

$$q(1 - T_1 T_2 \omega_u^2) + p\omega_u(T_1 + T_2) + K\sin(L\omega_u) = 0 \tag{7.119}$$

where p and q are determined as in (7.26). The solution of the (nonlinear) system given by Equations (7.116)–(7.119) provides the four process parameters.

It is worth stressing that a numerical procedure is necessary to find the parameters and therefore the computational complexity is somewhat relevant. Further, two complex conjugate poles might result even when the actual process dynamics is not oscillatory.

7.4 Discussion

A review of some (step-based) open-loop and (relay-feedback) closed-loop identification techniques for FOPDT and SOPDT transfer functions has been presented in the previous sections with the aim of highlighting the importance of the choice of the identification method in the context of PID design. A short description of each of the considered methods has been given in order to understand the rationale of the approach and its complexity. Details have been omitted (they can be found in the references).

In general, it has been shown that different options are available both for the open-loop and the closed-loop approach, each of them showing interesting features and possible drawbacks in a given application. In particular, it is worth highlighting that when an integral is employed in the context of the techniques based on step responses, the robustness to measurement noise increases significantly, but it might be possible that the final result is incorrect, because a negative (or, in any case, a too small) value of the time delay might result. This fact has been actually often overlooked in the literature.

The techniques based on a relay-feedback experiment are in general less robust to the measurement noise with respect to the step-based ones (but performing a closed-loop experiment might be significantly advantageous since the operating point of the process does not change). In general, it has been shown that there is a variety of strategies in the context of relay-feedback-based methodologies and it is actually difficult to choose the most appropriate one in a given application. At this point it is worth noting that the technique presented in the previous sections generally provide the process transfer function from the estimated values of the (approximate) ultimate gain and frequency of the process. An alternative procedure has been presented in (Friman and Waller, 1997) where a two-channel relay is employed. In particular, the adoption of two relays operating in parallel on the process output and on the integral of the process output allows to estimate a user-chosen point in the third quadrant of the complex plane. This feature can be exploited for an effective tuning of the PID controller. In any case, the model of the process is still basically determined starting from the knowledge of a single point of the frequency response of the process.

Obviously, the identification of multiple points would improve the accuracy of the obtained model. For this purpose, different techniques have been proposed

in the literature. For example, the use of a delay term in addition to the relay element provides the identification of a point in the frequency response that is different from the ultimate one (Li *et al.*, 1991; Leva, 1993). This fact has been exploited in (Scali *et al.*, 1999) for the identification of a completely unknown process (a suitable model order is selected automatically in the devised procedure). The obvious drawback of this method is that a multiple experiment has to be run. This can be avoided if a technique based on the Discrete Fourier Transform is applied (Wang *et al.*, 1999*a*; Wang *et al.*, 1999*b*), but a significant additional computational effort is required.
It has also to be noted again that in the previous sections the case of processes that can be well described by FOPDT and SOPDT models (7.1) and (7.42) have been particularly addressed. However, it has to be taken into account that there exists also a variety of methodologies capable to deal with processes with an oscillatory dynamics (see, for example, (Huang and Chou, 1994; Rangaiah and Krishnaswamy, 1996; Panda, 2006) in addition to the works referenced in Section 7.3), with processes with a stable and an unstable zero, with integral processes (see, for example, (Kwak *et al.*, 1997)) and with unstable processes (see, for example, (Vivek and Chidambaram, 2005*b*)).
Summarising, from the above (very simple) analysis it turns out that the choice of the identification strategy is indeed a crucial issue in the context of PID controllers if the tuning of the parameters is based on an estimated model of the plant. Different aspects have been pointed out and they have to be considered in order to provide the most satisfactory performance from a cost/benefit point of view.

7.5 PID Control of High-order Systems

In the previous sections it has been mentioned that the great majority of PID tuning rules assumes that the process model is described as a FOPDT or a SOPDT transfer function. In this context different techniques that aim at obtaining such a models directly from simple experiments on the process have been analysed.
From another point of view, it is recognised that many identification techniques are available nowadays in order to obtain accurate high-order models when the process exhibits a somewhat complex dynamics. Actually, it has to be taken into account that, in many cases, an apparent time delay is indeed due to the presence of a high-order dynamics (Leva, 2005). This has motivated a significant research interest in the last years for the design of PID controllers for high-order processes. It is realised that, because of the relative low-order of the controller, a model reduction has necessarily to be performed. In this context, two approaches can be actually followed:

1. design a model-based high-order controller by considering the (full) high-order dynamics of the process and then reduce the controller to a PID form;

2. reduce first the process model to an appropriate low-order form so that a model-based controller results directly to be in PID form.

In this section this two approaches are analysed and compared in the Internal Model Control (IMC) framework (Morari and Zafiriou, 1989), which has been extensively adopted for the purpose of PID controller tuning, in order to assess their advantages and disadvantages from the point of view of the achievable performance and of the ease of use.

7.5.1 Internal Model Control Design

The IMC methodology has been widely adopted for the purpose of PID controller tuning (though, being based on a pole-zero cancellation approach it is not suitable for lag-dominant processes subject to load disturbances (Scali and Semino, 1991; Shinskey, 1996)). Indeed, it provides the user with a desirable feature as a tuning parameter that handles the trade-off between robustness and aggressiveness of the controller.
In a general form, the IMC design can be applied to a standard unity-feedback control system (see Figure 7.8). The (stable) process to be controlled can be described by the model:

$$P(s) = p_m(s)p_a(s) \tag{7.120}$$

where $p_a(s)$ is the all-pass portion of the transfer function containing all the nonminimmum phase dynamics (note that it has to be $p_a(0) = 1$ in order to add the integral action to the resulting controller). The controller transfer function is then chosen as

$$C(s) = \frac{f(s)p_m^{-1}(s)}{1 - f(s)p_a(s)} \tag{7.121}$$

in which

$$f(s) = \frac{1}{(\lambda s + 1)^n} \tag{7.122}$$

is the IMC filter where λ is the adjustable time constant and n is an appropriate order so that the controller is realisable. It has to be noted that

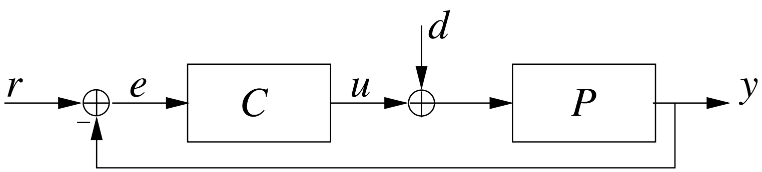

Fig. 7.8. Standard unity-feedback control scheme

the nominal closed-loop transfer function, *i.e.*, the transfer function from the set-point signal r and the process output y, results to be

$$T(s) = \frac{p_a(s)}{(\lambda s + 1)^n} \tag{7.123}$$

and this makes the role of the free design parameter λ clear in selecting the desired closed-loop dynamics (and therefore in handling the trade-off between robustness and aggressiveness, as unavoidable mismatches between the true process dynamics and its model have to be taken into account).
Obviously, in general, the resulting controller is not in PID form, *i.e.*:

$$C(s) = K_p \left(1 + \frac{1}{T_i s} + T_d s\right) \frac{1}{T_f s + 1}, \tag{7.124}$$

if the output-filtered ideal form is implemented. This occurs if the process model has one positive zero and two poles (note that this results if a FOPDT transfer function is considered and a first-order Padè approximation is adopted for the delay term (Rivera *et al.*, 1986)), while a PI controller results if the plant has a simple first-order dynamics. Thus, if a high-order process model is considered, this must be reduced to this suitable form before applying the IMC design or, alternatively, the resulting high-order controller has to be subsequently reduced to a PID form.

7.5.2 Process Model Reduction

Skogestad's Half Rule

The method proposed by Skogestad in (Skogestad, 2003) considers a process model reduction based on the so-called "half rule", which states that the largest neglected (denominator) time constant is distributed evenly to the effective dead time and the smallest retained time constant. In practice, given a high-order transfer function, each numerator term $(\tau_0 s + 1)$ with $\tau_0 > 0$ is first simplified with a denominator term $(T_0 s + 1)$, $T_0 > 0$ using the following rules:

$$\frac{\tau_0 s + 1}{T_0 s + 1} \approx \begin{cases} \tau_0/T_0 & \text{for } \tau_0 \geq T_0 \geq L \\ \tau_0/L & \text{for } \tau_0 \geq L \geq T_0 \\ 1 & \text{for } L \geq \tau_0 \geq T_0 \\ \tau_0/T_0 & \text{for } T_0 \geq \tau_0 \geq 5L \\ \frac{(\tilde{T}_0/T_0)}{(\tilde{T}_0 - \tau_0)s + 1} & \text{for } \tilde{T}_0 \stackrel{\text{def}}{=} \min(T_0, 5L) \geq \tau_0 \end{cases} \tag{7.125}$$

where L is the final effective delay (to be determined subsequently). It has to be noted that T_0 is normally chosen as the closest larger denominator time constant ($T_0 > \tau_0$), except when a larger denominator time constant does not exist or there is a smaller denominator time constant closer to τ_0; this

is true if the ratio between τ_0 and the smaller denominator time constant is less than the ratio between the larger denominator time constant and τ_0 and (both conditions must be satisfied) less than 1.6.
Once this procedure has been terminated for all the positive numerator time constants, the process transfer function is in the following form:

$$\tilde{P}(s) = \frac{\prod_j(-\tau'_{j0}s+1)}{\prod_i(T_{i0}+1)}e^{-L_0 s} \tag{7.126}$$

where $\tau'_{j0} > 0$ and the time constants are ordered according to their magnitude. Then, a SOPDT transfer function

$$P(s) = \frac{k}{(T_1 s+1)(T_2 s+1)}e^{-Ls} \tag{7.127}$$

is obtained by applying the half rule, *i.e.*, by setting

$$T_1 = T_{10}, \quad T_2 = T_{20} + \frac{T_{30}}{2}, \tag{7.128}$$

$$L = L_0 + \frac{T_{30}}{2} + \sum_{i\geq 4} T_{i0} + \sum_j \tau'_{j0}. \tag{7.129}$$

It appears that, being the rules (7.125) based on the final apparent time delay L, in the first part of the algorithm there is the need to guess this final value and to iterate in case at the end the result is incorrect.
Once the SOPDT process model is obtained, the parameters of a series PID controller expressed by the transfer function

$$C(s) = K_p\left(\frac{T_i s+1}{T_i s}\right)\frac{T_d s+1}{T_f s+1} \tag{7.130}$$

are determined by applying the IMC design procedure (and by approximating the delay term as $e^{-Ls} = 1 - Ls$) and by possibly modifying the value of T_i in order to address the case of lag-dominant processes (Skogestad, 2003) (note that this fact in not of concern in the examples presented in Section 7.5.4). It results that the PID parameters in (7.130) are selected as

$$K_p = \frac{T_1}{k(\lambda+L)}, \quad T_i = T_1, \quad T_d = T_2, \quad T_f = 0.01T_d. \tag{7.131}$$

Note that the conversion of the tuning rule (7.131) for the PID controller in the ideal form (7.124) is straightforward and a recommended choice for the desired closed-loop time constant is $\lambda = L$ (Skogestad, 2003).
Summarising, the method is based on simple, easy to remember, tuning rules (indeed, this is one of the main features of the method). However, the possible iterations in the model reduction algorithm makes the overall procedure somewhat difficult to automate.

Isaksson and Graebe's Method

The technique proposed by Isaksson and Graebe in (Isaksson and Graebe, 1999) is also based on a suitable process model reduction before applying the IMC design. The model reduction is performed as follows. Let the initial (high-order) process model be described by the transfer function

$$\tilde{P}(s) = \frac{B(s)}{A(s)} \tag{7.132}$$

Then, the numerator and denominator polynomials are considered separately and the polynomials $B_1(s)$ and $A_1(s)$ that retain only the slowest roots are determined. Subsequently, the polynomials $B_2(s)$ and $A_2(s)$ that retain the low-order coefficients are calculated. Finally, the reduced-order model is obtained as

$$P(s) = \frac{\frac{1}{2}(B_1(s) + B_2(s))}{\frac{1}{2}(A_1(s) + A_2(s))} \tag{7.133}$$

By choosing $B_1(s)$ and $B_2(s)$ of first order and $A_1(s)$ and $A_2(s)$ of second order and by subsequently applying the IMC design (with a first-order filter (7.122)), a PID controller (7.124) naturally arises. If there are no zeros, two solutions can be applied:

1. a second-order denominator is calculated in the reduction procedure and a second-order filter (7.122) is applied in the IMC design, resulting in a PID controller;
2. a first-order denominator is calculated in the reduction procedure and a first-order filter (7.122) is applied in the IMC design, resulting in a PI controller.

It has to be noted that, differently from the Skogestad's half rule, the case of complex conjugate roots is also addressed in (Isaksson and Graebe, 1999), but it will not be considered hereafter (see Section 7.5.4).
Summarising, the Isaksson and Graebe's method can be easily automated, although it is not explicitly based on tuning formulae.

Model Approximation with Step Response Data

The least-squares method presented in (Wang *et al.*, 2001; Wang and Zhang, 2001) and explained in Section 7.3.1 can be employed also for the purpose of model reduction. In particular, given a high-order model of the process, a SOPDT transfer function (with one zero) can be obtained directly from the open-loop step response. Note that there is no need of time consuming and costly experimental results, since a simulation can be performed (Huang, 2003). In this way the overall procedure can be easily automated, although a

relatively significant computational effort is actually necessary.
Starting from the identified model, the tuning rule (7.131) has been adopted. However, for this purpose, the obtained SOPDT model must have real poles and no zeros. Thus, if a zero is determined the half rule is then adopted, while if complex conjugate poles occur, a FOPDT model (obtained with the same identification method) is actually employed. In this latter case a PI controller results.

7.5.3 Controller Reduction

The methods described in Section 7.5.2 are based on the reduction of the process model before applying the IMC design. Conversely, it is possible to apply the IMC procedure described in Section 7.5.1 by considering the full process dynamics and then reduce the obtained high-order controller to a PID controller form. For this purpose, a Maclaurin series expansion can be employed. The expression of the resulting controller can be always written as (Lee *et al.*, 1998*b*; Lee *et al.*, 1998*a*):

$$C(s) = \frac{k(s)}{s} \tag{7.134}$$

and expanding $C(s)$ in a Maclaurin series in s it results:

$$C(s) = \frac{1}{s}\left[k(0) + k'(0)s + \frac{k''(0)}{2}s^2 + \cdots\right] \tag{7.135}$$

It turns out that the first part of the series expansion contains a proportional term, an integral term and a derivative term and therefore, if the high-order terms are neglected, a PID controller (7.124) results (a first-order filter can be easily added in order to make the controller proper and its time constant can be selected sufficiently small so that its dynamics is not significant). Indeed, the following relations hold:

$$\begin{aligned} K_p &= k'(0) \\ T_i &= \frac{k'(0)}{k(0)} \\ T_d &= \frac{k''(0)}{2k'(0)} \end{aligned} \tag{7.136}$$

Hence, the overall procedure can be easily automated, although it is not based on tuning formulae and its computational burden is somewhat considerable. However, it has to be stressed that a wrong choice of the design parameter λ can result in the overall control system being unstable (see Section 7.5.5). Although this can be easily checked before applying the controller, it can be considered as a major drawback of the method.

7.5.4 Simulation Results

In order to analyse and compare the different methodologies, the following processes with high-order dynamics have been considered:

$$P_1(s) = \frac{(15s+1)^2(4s+1)(2s+1)}{(20s+1)^3(10s+1)^3(5s+1)^3(0.5s+1)^3}, \tag{7.137}$$

$$P_2(s) = \frac{(-0.3s+1)(0.08s+1)}{(2s+1)(s+1)(0.4s+1)(0.2s+1)(0.05s+1)^3}, \tag{7.138}$$

$$P_3(s) = \frac{(-45s+1)(4s+1)(2s+1)}{(20s+1)^3(18s+1)^3(5s+1)^3(10s+1)^2(16s+1)(14s+1)(12s+1)}, \tag{7.139}$$

$$P_4(s) = \frac{1}{(s+1)^4}, \tag{7.140}$$

$$P_5(s) = \frac{1}{(s+1)^8}, \tag{7.141}$$

$$P_6(s) = \frac{1}{(s+1)^{20}}. \tag{7.142}$$

The main characteristics of the processes are summarised in Table 7.1. It has to be noted that transfer functions $P_1(s)$ and $P_3(s)$ have been taken from (Wang and Cluett, 2000) (actually, $P_3(s)$ has been modified in order to obtain an inverse response), $P_2(s)$ from (Skogestad, 2003) and $P_4(s) - P_6(s)$ are representative of typical industrial processes (Åström and Hägglund, 2000*a*; Shinskey, 2000).

The reduced-order model that have been adopted for the PI(D) tuning are shown in Table 7.2. Note that two transfer functions might occur for the Isaksson and Graebe's technique, whereas the process dynamics has no zeros, as explained in Section 7.5.2. Indeed, the first one results in a PID controller (with a second-order IMC filter), while the second one results in a PI controller (with a first-order IMC filter). Besides, whereas a FOPDT transfer function is reported for the step response based method, this means that the resulting SOPDT model has complex conjugate poles and has not therefore been employed (see Section 7.5.2).

In order to make a fair comparison, for each method and for each process the value of λ that minimises the integrated absolute error (4.23) for both the set-point and the load disturbance step responses have been selected (*i.e.*, a unit step has been applied on signals r and d separately, see Figure 7.8). For those methods that do not provide the value of the filter time constant T_f explicitly, this has been selected in such a way its dynamics is negligible. The resulting values of the integrated absolute error and the corresponding optimal values of λ are shown in Tables 7.3–7.8. Note again that for the Isaksson and Graebe's method two cases (PID and PI control) might emerge, depending on the fact that a second-order or first-order IMC filter respectively has been adopted (since there are no zeros in the process to be controlled). Analogously, a PID or a PI controller results from the technique based on the step response, depending on the use of a SOPDT model or a FOPDT model (the latter in case the identified SOPDT model has complex conjugate poles). Finally, the resulting (set-point and load) unit step responses are plotted in Figures 7.9–7.14. The process responses obtained with a PI controller resulting from the Isaksson and Graebe's method are not reported.

Table 7.1. Main characteristics of the considered processes

$P_1(s)$	Minimum phase dynamics
$P_2(s)$	Presence of a nondominant positive zero
$P_3(s)$	Presence of a dominant positive zero
$P_4(s)$	Minimum phase dynamics with a small number of coincident poles
$P_5(s)$	Minimum phase dynamics with a medium number of coincident poles
$P_6(s)$	Minimum phase dynamics with a high number of coincident poles

Table 7.2. Resulting reduced models for the different methods

Process	Skogestad	Isaksson and Graebe		step response
$P_1(s)$	$\dfrac{e^{-35.5s}}{(20s+1)(15s+1)}$	$\dfrac{25.5s+1}{2642s^2+73.25s+1}$		$\dfrac{e^{-27.38s}}{44.46s+1}$
$P_2(s)$	$\dfrac{e^{-0.77s}}{(2s+1)(1.2s+1)}$	$\dfrac{-0.26s+1}{3.21s^2+3.38s+1}$		$\dfrac{e^{-0.79s}}{2.48s^2+3.17s+1}$
$P_3(s)$	$\dfrac{e^{-180s}}{(30s+1)(20s+1)}$	$\dfrac{-42s+1}{8564s^2+115.7s+1}$		$\dfrac{e^{-127.7s}}{106.6s+1}$
$P_4(s)$	$\dfrac{e^{-1.5s}}{(1.5s+1)(s+1)}$	$\dfrac{1}{2.25s^2+3s+1}$	$\dfrac{1}{2.5s+1}$	$\dfrac{e^{-1.38s}}{2.71s+1}$
$P_5(s)$	$\dfrac{e^{-5.5s}}{(1.5s+1)(s+1)}$	$\dfrac{1}{14.52s^2+5.02s+1}$	$\dfrac{1}{4.51s+1}$	$\dfrac{e^{-3.88s}}{4.24s+1}$
$P_6(s)$	$\dfrac{e^{-17.5s}}{(1.5s+1)(s+1)}$	$\dfrac{1}{95.98s^2+11.40s+1}$	$\dfrac{1}{10.70s+1}$	$\dfrac{e^{-12.72s}}{7.76s+1}$

Table 7.3. Optimal IAE's (and corresponding values of λ) for process $P_1(s)$

Method		Task set-point	load
Skogestad		71.38 (9.6)	54.45 (0.01)
Isaksson and Graebe	PID	63.08 (54.4)	48.13 (39.1)
	PI		
step response	PID		
	PI	73.49 (37.2)	60.16 (20.8)
Maclaurin		42.07 (4.4)	25.35 (2.8)

Table 7.4. Optimal IAE's (and corresponding values of λ) for process $P_2(s)$

Method		Task set-point	load
Skogestad		1.852 (0.83)	0.869 (0.03)
Isaksson and Graebe	PID	1.963 (1.22)	0.985 (0.37)
	PI		
step response	PID	1.819 (0.78)	0.863 (0.01)
	PI		
Maclaurin		1.783 (0.18)	0.986 (0.07)

Table 7.5. Optimal IAE's (and corresponding values of λ) for process $P_3(s)$

Method		Task set-point	load
Skogestad		376.5 (120.1)	384.3 (115.2)
Isaksson and Graebe	PID	251.7 (122.3)	253.6 (113.0)
	PI		
step response	PID		
	PI	301.1 (159.0)	307.7 (153.6)
Maclaurin		231.6 (11.5)	233.5 (10.9)

Table 7.6. Optimal *IAE*'s (and corresponding values of λ) for process $P_4(s)$

Method		Task set-point	load
Skogestad		3.133 (0.49)	1.952 (0.001)
Isaksson and Graebe	PID	2.722 (0.87)	1.073 (0.39)
	PI	4.099 (3.44)	3.183 (2.33)
step response	PID		
	PI	3.998 (2.06)	3.063 (0.96)
Maclaurin		2.041 (0.41)	0.876 (0.19)

Table 7.7. Optimal *IAE*'s (and corresponding values of λ) for process $P_5(s)$

Method		Task set-point	load
Skogestad		11.58 (3.01)	10.95 (2.06)
Isaksson and Graebe	PID	7.776 (2.78)	6.401 (2.34)
	PI	9.550 (8.83)	8.880 (7.88)
step response	PID		
	PI	9.662 (4.96)	9.030 (4.00)
Maclaurin		6.561 (0.74)	5.394 (0.61)

Table 7.8. Optimal *IAE*'s (and corresponding values of λ) for process $P_6(s)$

Method		Task set-point	load
Skogestad		36.86 (11.2)	36.49 (10.5)
Isaksson and Graebe	PID	23.01 (8.83)	21.88 (8.42)
	PI	25.84 (24.6)	25.22 (23.8)
step response	PID		
	PI	27.36 (12.2)	26.95 (11.4)
Maclaurin		19.85 (0.90)	18.81 (0.86)

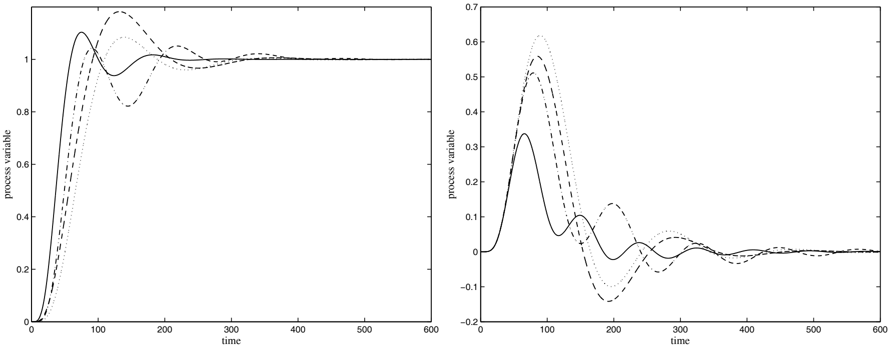

Fig. 7.9. Optimal set-point (left) and load disturbance (right) step responses for $P_1(s)$. Dashed line: Skogestad; dash-dot line: Isaksson and Graebe; dotted line: step response; solid line: Maclaurin.

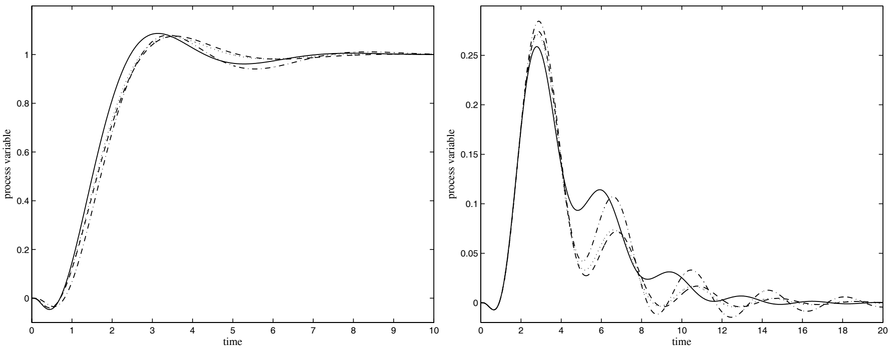

Fig. 7.10. Optimal set-point (left) and load disturbance (right) step responses for $P_2(s)$. Dashed line: Skogestad; dash-dot line: Isaksson and Graebe; dotted line: step response; solid line: Maclaurin.

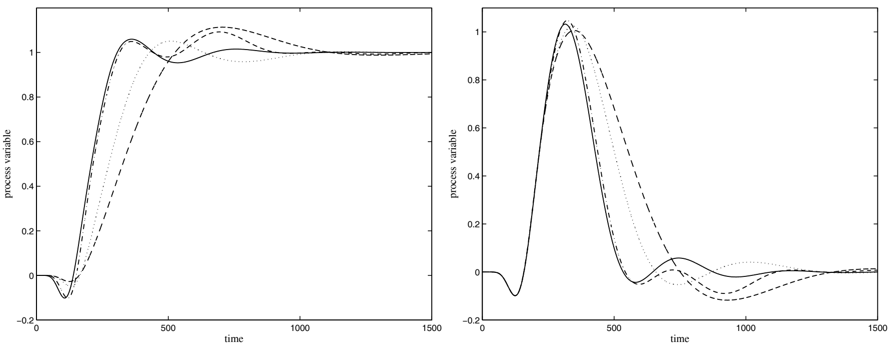

Fig. 7.11. Optimal set-point (left) and load disturbance (right) step responses for $P_3(s)$. Dashed line: Skogestad; dash-dot line: Isaksson and Graebe; dotted line: step response; solid line: Maclaurin.

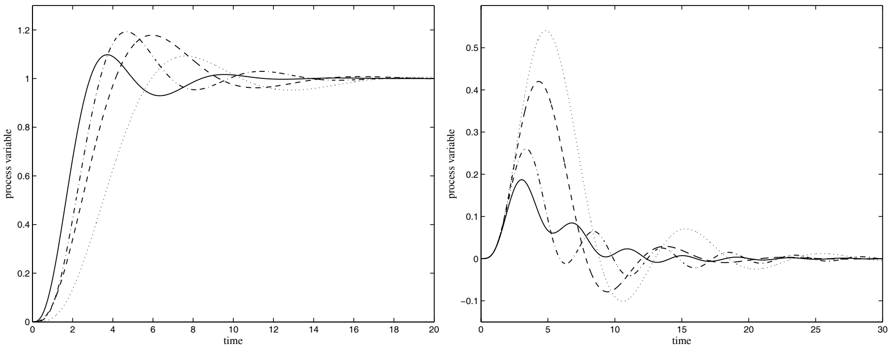

Fig. 7.12. Optimal set-point (left) and load disturbance (right) step responses for $P_4(s)$. Dashed line: Skogestad; dash-dot line: Isaksson and Graebe; dotted line: step response; solid line: Maclaurin.

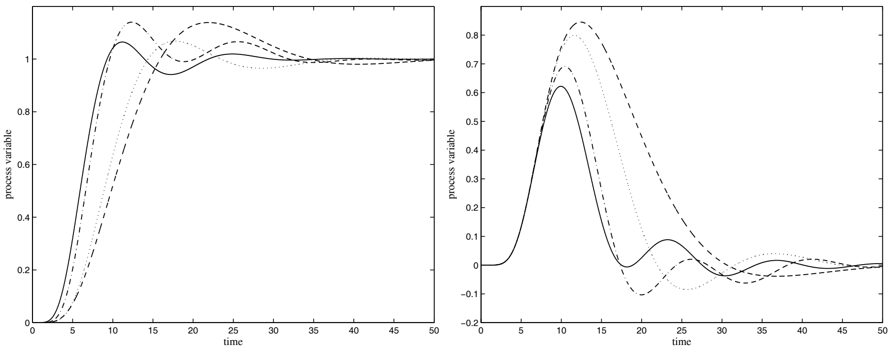

Fig. 7.13. Optimal set-point (left) and load disturbance (right) step responses for $P_5(s)$. Dashed line: Skogestad; dash-dot line: Isaksson and Graebe; dotted line: step response; solid line: Maclaurin.

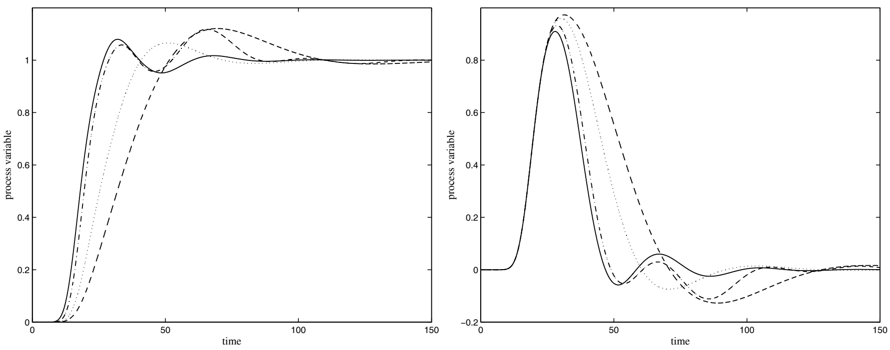

Fig. 7.14. Optimal set-point (left) and load disturbance (right) step responses for $P_6(s)$. Dashed line: Skogestad; dash-dot line: Isaksson and Graebe; dotted line: step response; solid line: Maclaurin.

7.5.5 Discussion

From the results obtained it appears that the approach based on the Maclaurin series expansion provides in general the best performance, both for the set-point following and the load disturbance rejection task. This is due to its capability of providing a higher open-loop crossover frequency without decreasing the phase margin with respect to the other methods (Visioli, 2005*c*). From another point of view, this means that in the set-point step responses a small rise time is achieved without impairing the overshoot and in the load disturbance step responses a small peak error results without the occurrence of significant oscillations.

It turns out that it is better to reduce the model of the controller than that of the plant, because the approximation introduced by adopting only the first three terms of the series expansion is not detrimental in the range of frequencies that is significant for the considered control system. However this is true only if an appropriate value of λ is selected. Indeed, a wrong choice of λ might imply negative parameters of the PID controller or, more remarkably, it might lead the system to instability (even if the PID parameters are positive). For example, for system $P_3(s)$, if $\lambda \leq 6$ or $\lambda \geq 162$ the resulting closed-loop system is unstable and, in any case, if $\lambda \geq 20$ at least one of the PID parameters results to be less than zero. In order to better clarify this fact, consider the value $\lambda = 5.0$. The Maclaurin series approximation of the IMC controller gives a PID controller with $K_p = 1.09$, $T_i = 158.4$ and $T_d = 68.32$ (all the parameters are positive) and the resulting closed-loop system is actually unstable. This fact can be understood by looking at Figures 7.15 and 7.16 where the Bode plots of the two controllers and of the open-loop system with the original IMC controller and with the PID one are reported respectively. It appears that the approximation of the IMC controller around the critical frequency is not sufficiently accurate (indeed, the series expansion is centered at the zero frequency) and therefore the crossover frequency increases too much to provide a positive phase margin.

In general, it might happen that a quite narrow range of values for λ is suitable for a given process. Despite the fact that an unappropriate value of λ can be easily recognised during the design phase, this can be considered as a major drawback of the method, which has been overlooked in the literature. Indeed, this makes the overall design more complicated and, most of all from a practical point of view, the physical meaning of the filter time constant, which should handle the trade-off between aggressiveness and robustness and control activity of the control system, is actually lost (increasing the value of λ does not necessarily correspond anymore to a more sluggish and stable control system).

It has also to be noted that the optimal values of λ differ significantly between the considered methodology, although it appears, as expected, that, in general, a higher order filter (*i.e.*, for the Maclaurin series based technique or when a PID controller is adopted instead of a PI controller in the Isaksson

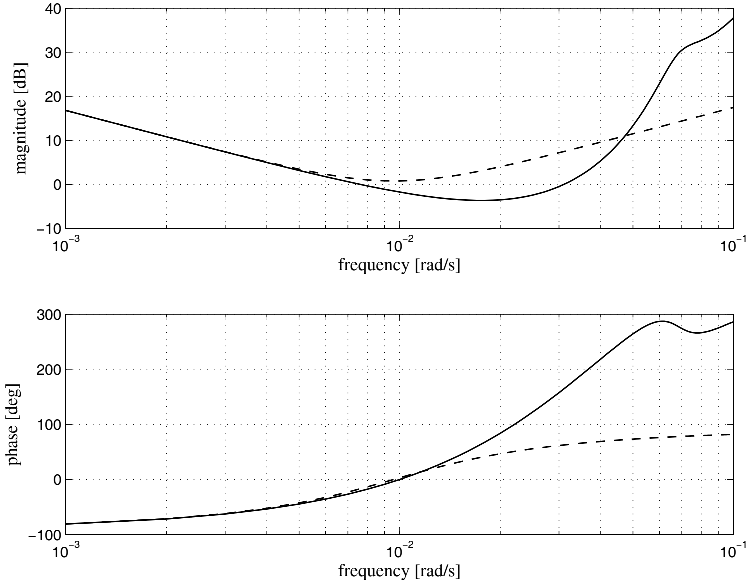

Fig. 7.15. Bode plot of the IMC controller (solid line) and of the approximating PID controller (dashed line)

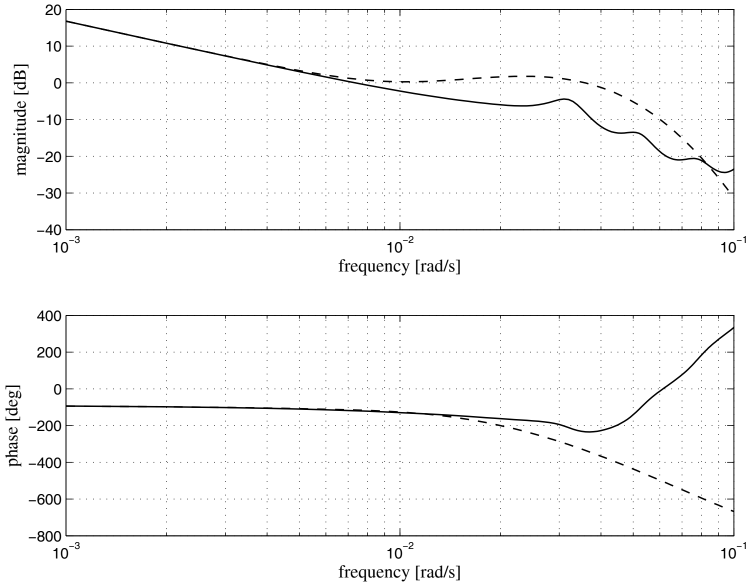

Fig. 7.16. Bode plot of the open-loop transfer function $C(s)P_3(s)$ with IMC controller (solid line) and with the approximating PID controller (dashed line)

and Graebe's method) implies a smaller value of λ (and the load disturbance rejection task requires a smaller value of λ than the set-point following task). From the results obtained, it also appears that the Isaksson and Graebe's method provides in general a better performance that the Skogestad's one and, as expected, the PID controller is better than the PI controller when the analytical PID design technique of Isaksson and Graebe is considered. Note, however, that the tuning rules (7.131) have been conceived with the aim of being applicable to a wide range of processes and of being easy to memorise. Regarding the method based on the step response data, it can be deduced that in general it provides worse performance than the Isaksson and Graebe's one, while no general conclusions can be drawn with respect to the Skogestad's method.

7.6 Conclusions and References

In this chapter, the issue of the identification of the process model and of the design of a PID controller when a high-order process model is available has been addressed. It has been shown that many methods for the estimation of a FOPDT and SOPDT transfer function have been devised, with different characteristics that have to be evaluated in a given application in order to provide a cost-effective PID design. Possible problems, often overlooked in the literature, have been indicated, mainly with the aim of highlighting that the identification method is indeed an integral part of the overall controller design and its choice is critical. It has also been shown that, in case a high-order process model is available, either the approach of reducing first the process model or of reducing the controller at the end can be applied. In the latter case a better performance can be achieved in general but at the expense of a more complicated and less intuitive design.

The review provided for the identification methods is surely not exhaustive. In the context of PID control, an up-to-date review of basic and advanced identification methods can be found in (Johnson and Moradi (eds.), 2005). The subject of relay-feedback is thoroughly addressed in (Wang *et al.*, 2003), while its use for system identification is analysed in (Yu, 1999), where advanced methodologies are also presented. A tutorial review of relay-feedback automatic tuning techniques for process controllers can be found in (Hang *et al.*, 2002).