3

Relay Feedback

Åström and Hägglund [1] suggest the relay feedback test to generate sustained oscillation as an alternative to the conventional continuous cycling technique. It is very effective in determining the ultimate gain and ultimate frequency. Luyben [2] popularizes the relay feedback method and calls this method "ATV" (autotune variation). The acronym also stands for *all-terrain vehicle*, since ATV provides a useful tool for the rough and rocky road of system identification.

As pointed out by Luyben, the motivation for using the relay feedback (ATV) has grown out of a study of an industrial distillation column. The distillation column is an important unit in chemical process industries. It is rather difficult to obtain a linear transfer function model for highly nonlinear columns. Attempts have been made using step or pulse tests. Unfortunately, the system results in an extremely long time constant, *e.g.* $\tau \approx 870 \, \text{h}$ [2]. Moreover, very large deviations occur in the linear model as the size or direction of the input is changed. Simulation studies also reveal that, sometimes, very small changes of magnitude (less than 0.01%) have to be made to get an accurate linear model. This immediately rules out the use of this kind of input design in real plants because plant data are never known to anywhere near this order of accuracy. Luyben shows that the simple relay feedback tests provide an effective way to determine linear models for such processes. It has become a standard practice in chemical process control, as can be seen in recent textbooks in process control [3,4]. Wang et al. [5] discuss various aspects of the relay feedback.

The distinct advantages of the relay feedback are:

1. It identifies process information around the important frequency, the ultimate frequency (the frequency where the phase angle is $-\pi$).

2. It is a closed-loop test; therefore, the process will not drift away from the nominal operating point.

3. For processes with a long time constant, it is a more time-efficient method than conventional step or pulse testing. The experimental time is roughly equal to two to four times the ultimate period.

3.1 Experimental Design

Consider a relay feedback system where $G(s)$ is the process transfer function, y is the controlled output, y^{set} is the SP, e is the error and u is the manipulated input (Figure 3.1A).

An on–off (ideal) relay is placed in the feedback loop. The Åström–Hägglund relay feedback system is based on the observation: when the output lags behind the input by $-\pi$ radians, the closed-loop system may oscillate with a period P_u. Figure 3.1(B) illustrates how the relay feedback system works. A relay of magnitude h is inserted in the feedback loop. Initially, the input u is increased by h. As the output y starts to increase (after a dead time D), the relay switches to the opposite position, $u = -h$. Since the phase lag is $-\pi$, a limit cycle with a period P_u results (Figure 3.1). The period of the limit cycle is the ultimate period. Therefore, the ultimate frequency from this relay feedback experiment is

$$\omega_u = \frac{2\pi}{P_u} \tag{3.1}$$

From the Fourier series expansion, the amplitude a can be considered to be the result of the primary harmonic of the relay output. Therefore, the ultimate gain can be approximated as [1,6]

$$K_u = \frac{4h}{\pi a} \tag{3.2}$$

where h is the height of the relay and a is the amplitude of oscillation. These two values can be used directly to find controller settings. Notice that Equations 3.1
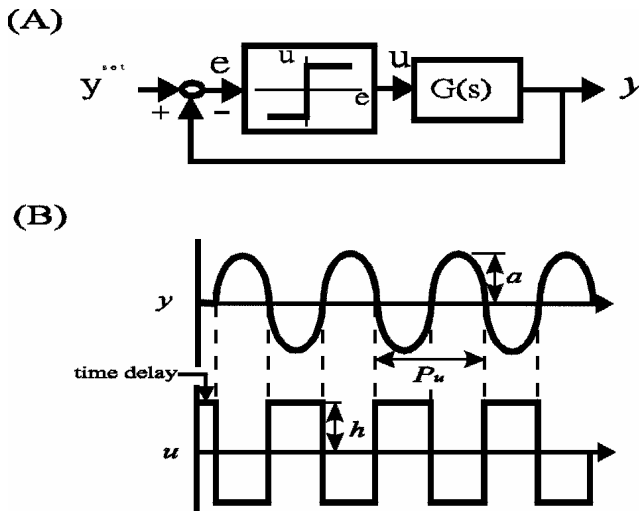
Figure 3.1. (A) Block diagram for a relay feedback system and (B) relay feedback test for a system with positive steady state gain

and 3.2 give approximate values of ω_u and K_u. A more accurate expression will be derived shortly.

The relay feedback test can be carried out manually (without any autotuner). The procedure requires the following steps.

1. Bring the system to steady state.

2. Make a small (*e.g.* 5%) increase in the manipulated input. The magnitude of change depends on the process sensitivities and allowable deviations in the controlled output. Typical values are between 3 and 10%.

3. As soon as the output crosses the SP, the manipulated input is switched to the opposite position (*e.g.* –5% change from the original value).

4. Repeat step 2 until sustained oscillation is observed (Figure 3.1).

5. Read off ultimate period P_u from the cycling and compute K_u from Equation 3.2.

This procedure is relatively simple and efficient. Physically, it implies moving the manipulated input *against* the process. Consider a system with a positive steady state gain (Figure 3.1). When you increase the input (as in step 1), the output y tends to increase also. As a change in the output is observed, you switch the input to the opposite direction. This is meant to bring the output back down to the SP. However, as soon as the output comes down to the SP, you switch the input to the upper position. Consequently, a continuous cycling results, but the amplitude of oscillation is under your control (by adjusting h). More importantly, in most cases, you obtain the information you need for tuning of the controller.

Several characteristics can be seen from the relay feedback test. Consider the most common FOPDT systems.

$$G(s) = \frac{K_p e^{-Ds}}{\tau s + 1} \tag{3.3}$$

where K_p is the steady state gain, D is the dead time and τ is the time constant. Figure 3.2 indicates that, if the normalized dead time D/τ is less than 0.28, the ultimate period is smaller than the process time constant. In terms of plant test, that implies the relay feedback test is more time efficient than the step test. The reason is that it takes almost 3τ to reach 95% of the steady state value in a step test and the time required for the relay feedback is also roughly equal to $3P_u$ (to establish a stable oscillation). Therefore, the relay feedback system is more time efficient than the step test for systems with

$$D/\tau < 0.28 \tag{3.4}$$

Since the dead time cannot be too large (it often comes from the measurement delay), the temperature and composition loops in process industries seem to fall into this category. In other words, Equation 3.4 is fairly typical for many slow chemical processes, especially for units involved with composition changes.
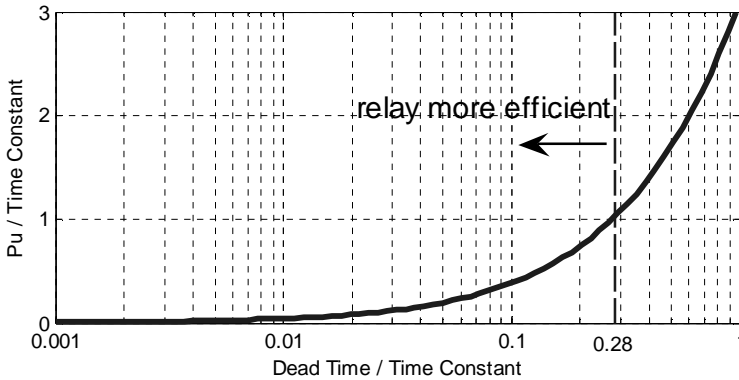
Figure 3.2. P_u/τ as function of the normalized dead time D/τ

3.2 Approximate Transfer Functions: Frequency-domain Modeling

After the relay feedback experiment, the estimated ultimate gain $\hat{K}_u$ and ultimate frequency $\hat{\omega}_u$ can be used directly to calculate controller parameters. Alternatively, it is possible to back-calculate the approximated process transfer functions. The other data useful in finding the transfer function are the dead time D and/or the steady state gain K_p.

In theory, the steady state gain can be obtained from plant data. One simple way to find K_p is to compare the input and output values at two different steady states. That is:

$$K_p = \Delta y / \Delta u \qquad (3.5)$$

where Δy denotes the change in the controlled variable and Δu stands for the deviation in the manipulated input. However, precautions must be taken to make sure that the sizes of the changes in u are made small enough such that the gain in Equation 3.5 truly represents the linearized gain. For highly nonlinear processes, these changes are typically as small as 10^{-3} to 10^{-6} % of the full range [2]. Such small changes would only be feasible using a mathematical model. Trying to obtain reliable steady state gains from plant data is usually impractical.

The dead time D in the transfer function can be easily read off from the initial part of the relay feedback test. It is simply the time it takes for y to start responding to the change in u (Figure 3.1). For the FOPDT system, it is simply the time to reach the peak amplitude in a half period, as will be shown in Chapter 4. Therefore, it is more likely that we will have information on the dead time rather than the steady state gain.

Now we are ready to find an approximate model. Typical transfer functions in process control are assumed and parameters can be calculated. The transfer functions have the following forms:

Model I (integrator plus dead time)

$$G(s) = \frac{K_p e^{-Ds}}{s} \tag{3.6}$$

Model P (pure dead time)

$$G(s) = K_p e^{-Ds} \tag{3.7}$$

Model 1 (FOPDT)

$$G(s) = \frac{K_p e^{-Ds}}{\tau s + 1} \tag{3.8}$$

Model 2a (second-order plus dead time)

$$G(s) = \frac{K_p e^{-Ds}}{(\tau s + 1)^2} \tag{3.9}$$

Model 2b (second-order plus dead time with two unequal lags)

$$G(s) = \frac{K_p e^{-Ds}}{(\tau_1 s + 1)(\tau_2 s + 1)} \tag{3.10}$$

In these five models, model I and model P have two unknown parameters, models 1 and 2a have three unknown parameters and model 2b has four unknown parameters. Therefore, additional information, such as D or K_p, is needed if the last three models are employed. As pointed out by Tyreus and Luyben [7], the simplest integrator-plus-time-delay model (model I) provides good approximation for slow chemical processes, *e.g.* systems showing a small D/τ value. It is the model we recommend for slow processes.

The relay feedback experiment has the following steps:

1. If necessary, the dead time D can be read off from the initial response, or the time to the peak amplitude, and the steady state gain can be obtained from steady state simulation.

2. The ultimate gain $\hat{K}_u$ and ultimate frequency $\hat{\omega}_u$ are computed (Equations 3.1 and 3.2) after the relay feedback experiment.

3. Different model structures (Equations 3.6–3.10) are fitted to the data.

3.2.1 Simple Approach

Once the model is selected, we can back-calculate the model parameters from two equations describing the ultimate gain and the ultimate frequency.

Model I (Friman and Waller [8])

$$K_p = \frac{\omega_u}{K_u} = \frac{2\pi}{K_u P_u} \tag{3.11}$$

$$D = \frac{\pi}{2\omega_u} = \frac{P_u}{4} \tag{3.12}$$

Notice that no *a priori* process knowledge is needed for this model. Moreover, computation of K_p and D is quite straightforward.

Model P

$$K_P = \frac{1}{K_u} \tag{3.13}$$

$$D = \frac{P_u}{2} \tag{3.14}$$

Similar to model I, no *a priori* process knowledge is necessary.

Model 1

$$\tau = \frac{\tan(\pi - D\omega_u)}{\omega_u} \tag{3.15}$$

$$\tau = \frac{\sqrt{(K_p K_u)^2 - 1}}{\omega_u} \tag{3.16}$$

For model 1, either D or K_p is needed to solve for the time constant. For example, if the dead time is read off from the relay test, then we can compute τ from Equation 3.15. Then, K_p can be found by solving Equation 3.16.

Model 2a

$$\tau = \frac{\tan(\pi - D\omega_u)/2}{\omega_u} \tag{3.17}$$

$$\tau = \frac{\sqrt{(K_p K_u) - 1}}{\omega_u} \tag{3.18}$$

The equations describing model 2a are quite similar to those for model 1. Again, we need to know D or K_p before finding model parameters.

Model 2b

$$-\pi = -\omega_u D - \tan^{-1}\left(\omega_u \tau_1\right) - \tan^{-1}\left(\omega_u \tau_2\right) \tag{3.19}$$

$$\frac{1}{\hat{K}_u} = \frac{K_p}{\sqrt{\left[1 + \left(\omega_u \tau_1\right)^2\right]\left[1 + \left(\omega_u \tau_2\right)^2\right]}} \tag{3.20}$$

Since we have four parameters in model 2b, both K_p and D have to be known in order to solve for the two time constants τ_1 and τ_2. This is the most complex model structure in our models, and it is often sufficient for process control applications.

Let us use an FOPDT system to illustrate the parameter estimation procedure.

Example 3.1 WB column [9]

$$G(s) = \frac{12.8\,e^{-s}}{16.8s + 1}$$

This is the transfer function between the top composition x_D and the reflux flow R. From a relay feedback test, we obtain the following ultimate gain and ultimate frequency: $\hat{K}_u = 1.71$ and $\hat{\omega}_u = 1.615$. Note that these two values are only an approximation to the true values: $K_u = 2.1$ and $\omega_u = 1.608$.

Parameters can be calculated for different model structures:

Model I (no prior knowledge on K_p and D)

$$G(s) = \frac{0.94\,e^{-0.97s}}{s}$$

Model P (no prior knowledge on K_p and D)

$$G(s) = 0.58e^{-1.94s}$$

Model 1 (assume D is known, *i.e.* $D = 1$)

$$G(s) = \frac{13.2\,e^{-s}}{\left(14.0s + 1\right)}$$

Model 2a (assume D is known)

$$G(s) = \frac{1.12\,e^{-s}}{\left(0.59s + 1\right)^2}$$

Model 2b (assume K_p and D are known)

$$G(s) = \frac{12.8e^{-s}}{\left(13.5s + 1\right)\left(0.0009s + 1\right)}$$

Despite varying in model parameters, all these four models have the *same* ultimate gain and ultimate frequency. That is, the models are correct around the ultimate

frequency, which is important for the controller design. However, if we extrapolate the model to different frequencies, *e.g.* $\omega = 0$, then the results can be completely misleading. For example, the steady state gain of model 2a is only 1.12, which is less than 10% of the true value. We have to be very cautious when using these models.   ∎

3.2.2 Improved Algorithm

In theory, if the model structure is correct and the ultimate gain and ultimate frequency are correctly identified, then we could have a very good approximation of the transfer function. For example, if the K_u and ω_u in the previous example are close to the true values, then we will not have errors in the steady state gains and time constant for model 1. Unfortunately, since Equations 3.1 and 3.2 only give approximations to the ultimate gain and ultimate frequency, the parameters derived from Equations 3.15 and 3.16 can deviate significantly from the true system parameters. This implies the observed ultimate period $\hat{P}_u$ and the computed ultimate gain are not the true values.

   In order to have a better approximation of the transfer function, fundamental analysis of the relay feedback system is necessary. First, one would like to know what the period of oscillation from the relay feedback experiment really represents. In other words, given a transfer function with known parameters, what is the expression for the period of oscillation observed from the relay feedback experiment, $\hat{P}_u$? The following theorem [1] provides the answer.

Theorem 3.1 Consider the relay feedback system with a transfer function $G(s)$ and an ideal relay (Figure 3.1). Let $HG(T_s, z)$ be the pulse transfer function of $G(s)$ with a sampling time of T_s. If there is a periodic oscillation, then the period of oscillation $\hat{P}_u$ is given by

$$HG\left(\hat{P}_u / 2, -1\right) = 0$$

   Åström and Hägglund [1] prove the theorem starting form the discrete-time state-space equations. The result, $HG\left(\hat{P}_u / 2, -1\right) = 0$, is obtained by finding the z-domain equivalent. The continuous-time response of an ideal relay (Figure 3.1) can be discretized at the point when the relay switches. The z-transforms of the input and output are $h/(z+1)$ and 0 respectively. Since this is a self-oscillation system, the propagation of the input is described by the gain $HG\left(\hat{P}_u / 2, -1\right) = 0$. This equation can be used to find the period of oscillation for a known system. In identification, $\hat{P}_u$ is observed from the response and one is able to use this to back-calculate system parameters. Unlike the continuous-time analysis based on the primary harmonic, the discrete-time expression gives a sound basis for finding the system parameters, since no assumption is made in the derivation.

   Based on the theorem, a better relationship between $\hat{\omega}_u$ (or $\hat{P}_u$) and the system parameters can be derived. For the transfer functions of interest (models 1, 2a and 2b), the following results can be derived from the modified z-transform [10]:

Model 1

$$\tau = \frac{\pi}{\hat{\omega}_u \ln\left(2\exp\left(D/\tau\right)-1\right)}$$

(3.21)

Model 2a

$$\tau = \frac{2\pi\left[m+(m-1)\exp\left(-\dfrac{\pi}{\tau\hat{\omega}_u}\right)\right]}{\hat{\omega}_u\left[1+\exp\left(-\dfrac{\pi}{\tau\hat{\omega}_u}\right)\right]\left[\exp\left(\dfrac{m\pi}{\tau\hat{\omega}_u}\right)\div\left(1+\exp\left(-\dfrac{\pi}{\tau\hat{\omega}_u}\right)\right)-2\right]}$$

(3.22)

where $m = 1 - \dfrac{D\hat{\omega}_u}{\pi}$

Model 2b

$$\tau_1\left[\frac{2\exp\left(-\dfrac{m\pi}{\tau_1\hat{\omega}_u}\div\right)}{1+\exp\left(-\dfrac{\pi}{\tau_1\hat{\omega}_u}\div\right)}\right]-\tau_1 = \tau_2\left[\frac{2\exp\left(-\dfrac{m\pi}{\tau_2\hat{\omega}_u}\right)}{1+\exp\left(-\dfrac{\pi}{\tau_2\hat{\omega}_u}\right)}\right]-\tau_2$$

(3.23)

Equations 3.21–3.23 provide alternative expressions between the observed ultimate period , *e.g.* $\hat{\omega}_u$, and system parameters. For example, Equation 3.21 relates $\hat{\omega}_u$ to D and τ in a way that differs substantially from the standard phase angle equation (*i.e.* Equation 3.15).

$$-\pi = -\hat{\omega}_u D - \tan^{-1}\left(\hat{\omega}_u\tau\right)$$

Again, we can derive a better expression for the amplitude ratio part at the ultimate frequency, since the expression in Equation 3.2 is based on the first harmonic of the Fourier series expansion. The square-wave response of u (Figure 3.1) consists of many frequency components:

$$u(t) = \frac{4h}{\pi}\sum_{n=0}^{\infty}\frac{\sin\left((2n+1)\omega t\right)}{2n+1}$$

(3.24)

Therefore, it becomes obvious that the amplitude observed in the relay feedback response is contributed from multiple frequencies, $\omega = \hat{\omega}, 3\hat{\omega}, 5\hat{\omega}$, *etc.* In theory, one can have a better estimate of the amplitude ratio by employing more terms. An iterative procedure is necessary if more than one term is employed (*e.g.* finding $G(s)$ from the single-term solution and including the higher frequency information, $\omega = 3\hat{\omega}_u$, to find a new $G(s)$ and the procedure is repeated until $G(s)$ converges). However, experimental results show that the estimation of system parameters can be improved substantially by improving the expression for period of oscillation alone, as shown in the next section. Furthermore, for higher order systems, there is little incentive to improve the expression for the amplitude by including more terms, since higher order harmonics (*e.g.* $\omega = 3\hat{\omega}_u$ or $\omega = 5\hat{\omega}_u$) are attenuated by

the process. If only one term is employed, then the equations describing the amplitude ratio are exactly the same as Equations 3.16, 3.18 and 3.20.

3.2.3 Parameter Estimation

From the ongoing analysis, the procedure for the evaluation of the transfer function has the following steps:

1. Select model structure.
2. Compute model parameters according to Table 3.1.

Table 3.1 summarizes the information required and the corresponding equations to find the approximate transfer function. Most of these equation sets can be solved sequentially. Notice that if the improved algorithm is used, then better estimates of the ultimate gain and ultimate frequency can be *calculated* from the *model*. For model 2b, if some information is not known, then a different procedure should be employed. For example, if K_p is not available, we can perform a second relay feedback test [11] or use a biased relay (Chapters 7 and 12) to find additional information. Nonetheless, the equations noted in Table 3.1 are generally applicable regardless of the procedure.

3.2.4 Examples

Several examples are used to illustrate the advantages of the improved algorithm. Consider a first-order plus dead time system.

Example 3.2 FOPDT process

$$G(s) = \frac{16.5e^{-10s}}{20s+1}$$

From a relay feedback experiment with $h = 0.04$ we have $\hat{P}_u = 33.26$ and $a = 0.26$. If D and/or K_p are available, we can back-calculate τ. The τ values calculated from Equations 3.15 and 3.16 are $\tau = 16.3$ and 16.09 respectively. The improved algorithm (Equation 3.21) gives a better estimate in τ, $\tau = 19.97$, by improving the expression in the period of oscillation alone. The result from Equation 3.21 is almost exact (the difference may have resulted from reading off a and P_u from the response curve). Figure 3.3 shows the multiplicative modeling errors, $e_m = \left\| \left(G(i\omega) - \hat{G}(i\omega) \right) / \hat{G}(i\omega) \right\|$, for the transfer function $\hat{G}$ estimated from Equations 3.15, 3.16 and 3.21. The results show that the error e_m is significantly less when τ is calculated from Equation 3.21 alone.   ■

In the following examples, we assume K_p and D are known and the time constant τ for models 1 and 2a is obtained by taking the average of the values calculated from the corresponding equations for the case of the simple algorithm. Next, the effects of dead time on the estimation of the ultimate gain and ultimate frequency are also investigated. In the original ATV method, $\hat{K}_u$ is calculated from

Table 3.1. Equations for different model structures

Model	Simple algorithm	Improved algorithm	Prior information
Model I	Equations 3.11 and 3.12	–	None
Model P	Equations 3.13 and 3.14	–	None
Model 1	Equations 3.15 and 3.16	Equations 3.21 and 3.16	D or K_p
Model 2a	Equations 3.17 and 3.18	Equations 3.22 and 3.18	D or K_p
Model 2b	Equations 3.19 and 3.20	Equations 3.23 and 3.20	D and K_p
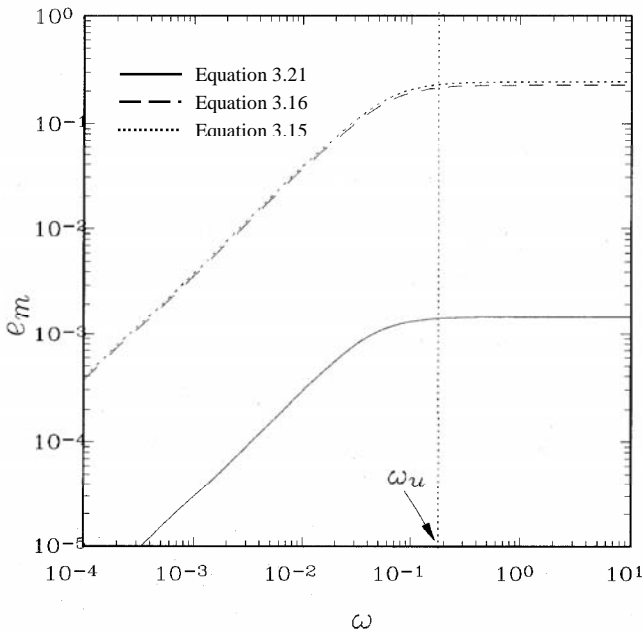
Figure 3.3. Multiplicative errors of an FOPDT system obtained from Equations 3.15, 3.16 and 3.21

Equation 3.2 and $\hat{\omega}_u$ is derived from Equation 3.1. In the proposed method, K_u and ω_u are back-calculated from the estimated transfer function $\hat{G}(s)$. Again, this is shown in the following transfer function:

Example 3.3   Variable dead time

$$G(s) = \frac{16.5e^{-Ds}}{20s + 1}$$

The percentage errors in K_u and ω_u are compared for these two methods over a range of dead time (D = 0.1–60). The results (Figure 3.4) show that the errors in K_u for the simple method are quite significant (5–20%). Furthermore, the error in ω_u is almost nil for the improved method.   ■

Similar behavior can also be observed for a second-order lag with time-delay system.

Example 3.4 Second-order system with two unequal lags

$$G(s) = \frac{37.3e^{-Ds}}{(7200s+1)(2s+1)}$$

Figure 3.5 shows that a better estimation of $\hat{G}(s)$ can be achieved over a range of D $(D < 60)$. Again, improvements can be made in finding the correct K_u and ω_u by using a more accurate expression in the period of oscillation.   ■

Since the estimated transfer function is typically employed in the analysis and design of a feedback control system, the *impact* of the modeling errors in closed-loop performance is evaluated. A model-based controller, IMC, is employed to analyze the performance. One of the advantages of the IMC is that we can specify the desired trajectory in the design. Figure 3.6 compares the SP responses of IMC when different models $\hat{G}'$ are employed in the design of the controllers. Consider
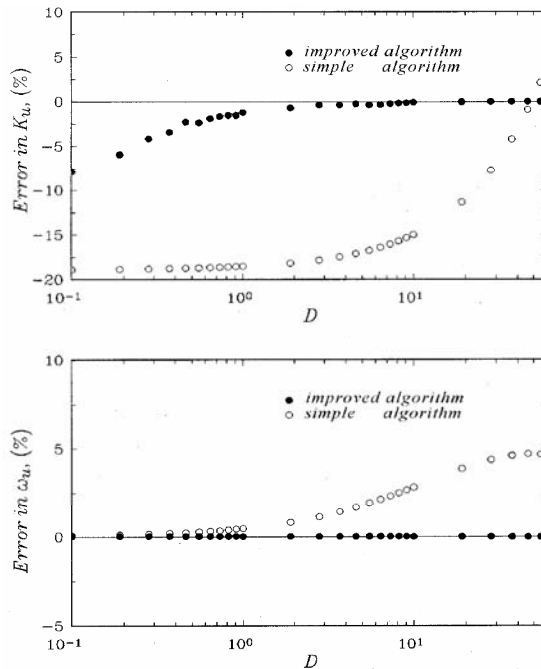
Figure 3.4. Percentage errors in K_u and ω_u for the FOPDT system over a range of dead time D

the FOPDT system

$$G(s) = \frac{16.5e^{-10s}}{20s+1}$$

The SP response of the control system, designed according to $\hat{G}(s)$ from the simple algorithm, tends to be more sluggish than the desired trajectory (Figure 3.6). The proposed method improves this situation, as shown in Figure 3.6. Despite the fact that a tighter response can be achieved by shortening the closed-loop time constant under modeling errors, one has to realize that the value of a model-based controller is that one can foresee the closed-loop response. In other words, a good model always helps.

Generally, the proposed method improves the estimation in $G(s)$ at the nominal condition (with perfect knowledge of K_p and D). The robustness with respect to errors in the dead time is investigated. Since the improved method calculates K_u and ω_u by finding the transfer function $\hat{G}(s)$ first, followed by solving the corresponding equations for them, it is more sensitive to the errors in the dead time than the original method. Let us take another FOPDT system as an example.

Example 3.5 Error in the observed dead time

$$G(s) = \frac{16.5e^{-s}}{20s+1}$$
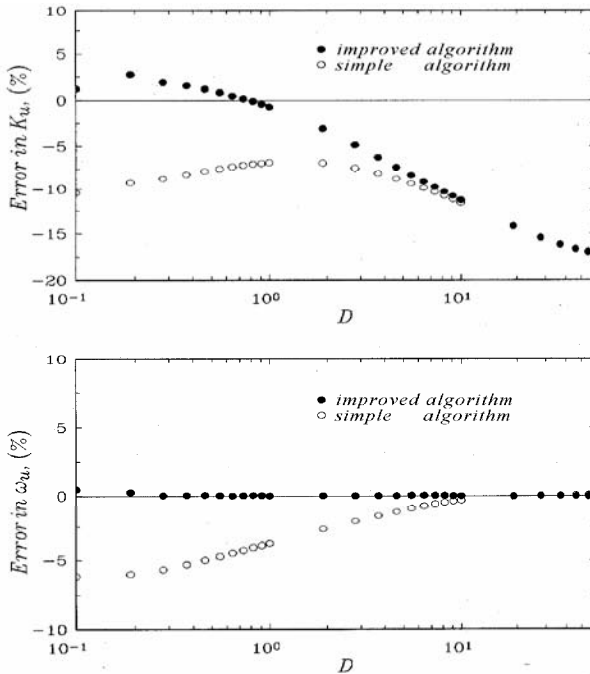
Figure 3.5. Percentage errors in K_u and ω_u for a second-order plus dead time system over a range of delay time D
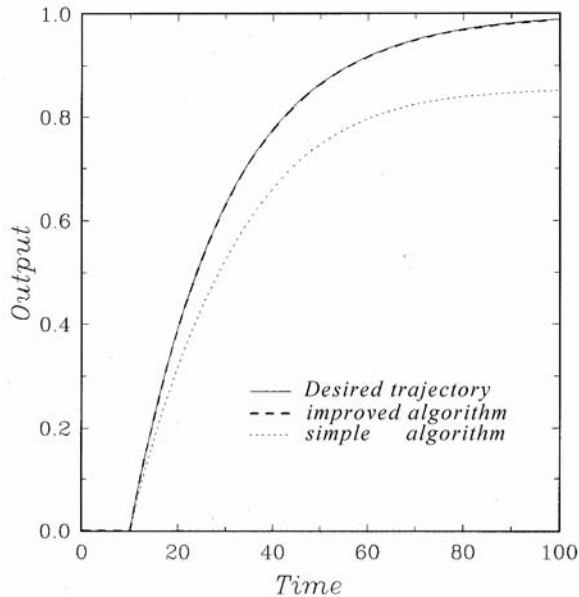
Figure 3.6. SP responses of IMC designed according to the estimated transfer functions $\hat{G}(s)$ (the closed-loop constant is 20 for the desired trajectory)

Figure 3.7 shows the estimate of K_u and ω_u for both methods when the percentage errors in dead time range from -50% to 50%. Despite the fact that the errors in K_u and ω_u are less for the improved method over a reasonable range of errors in dead time, it is more sensitive to the error in D. Therefore, care should be taken in reading off the dead time from the initial responses or the time to the peak amplitude.   ∎

## 3.3 Approximate Transfer Functions: Time-domain Modeling

Up to this point, the model identification is based on the frequency domain approach, which is based on the describing functions. A method to derive FOPDT-type systems was proposed by Wang et al. [12] using a single relay test. In a separate attempt, Majhi and Atherton [3] proposed a technique to identify plant parameters, but the method needs a correct initial guess and convergence is not guaranteed. Kaya and Atherton [14] describe another method (A-locus) to identify low-order process parameters from relay autotuning. Panda and Yu [15] develop analytical models to represent relay responses produced by different systems. The relay output consists of a series of step changes in manipulated variables (with opposite sign). Hence, the stabilized output is a sum of infinite terms of step responses due to those step changes. For systems with dead time D, the actual relay output
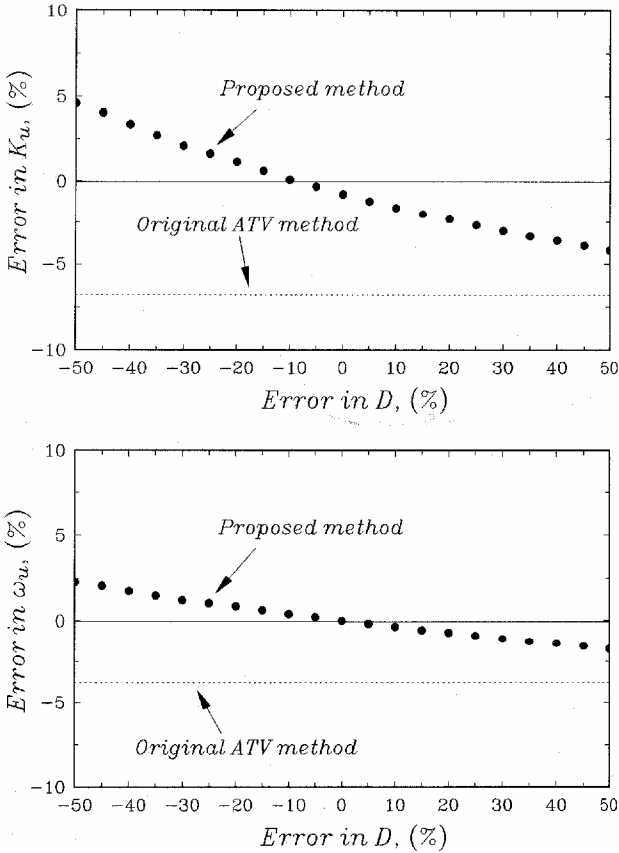
Figure 3.7. Percentage errors in K_u and ω_u for a first-order system over a range of variation in the dead time

lags behind the input by a time unit D. The inputs and outputs can be synchronized by shifting the output forward in time by an amount D, as shown in Figure 3.8B, and, in doing this, the dead time D can be eliminated from the expression for relay responses, as will be shown later. The shifted version of a typical relay feedback response provides the basis for the derivation.

It is assumed that the relay response is formed by n-number of step changes, of opposite directions ($\pm u$), in input. The switching period for each step change is $P_u/2$, except for the initial step change. In Figure 3.9, in the first interval, as time changes from $t=0$ to $t=D$, the response y_1 is produced due to the first step change u_1. Again, in the second interval, time progressing from D to $D+P_u/2$, response y_2 results due to the combined effects of step changes u_1 and u_2. Similarly, the effect of u_1, u_2 and u_3 produces y_3 during the third time interval ($D+P_u/2$ to $D+P_u$). Two half periods ($P_u/2$) are of special interest in Figure 3.9. The even values of n result in descending half period y_{2n}, and the odd values of n formulate the ascending half periods y_{2n+1}. It is interesting to note that the
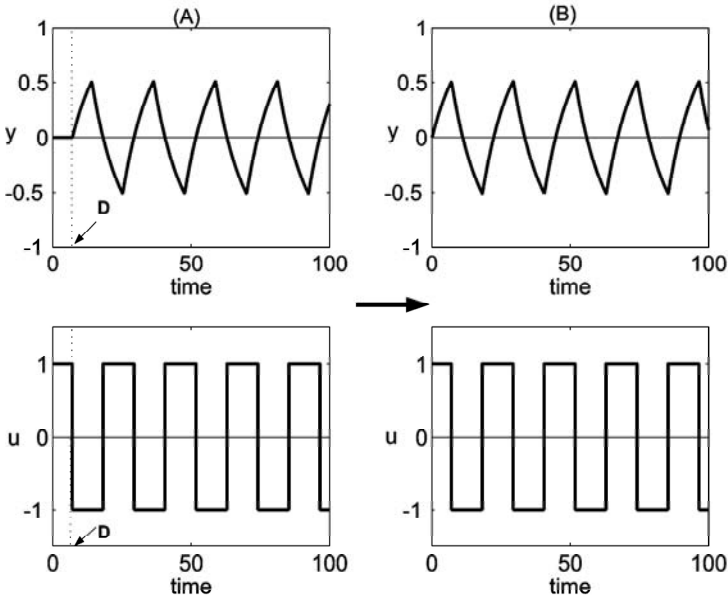
Figure 3.8. Schematic representation of the shifted version of relay feedback response for the development of their analytical expressions: (A) original relay feedback responses and (B) output y shifted by D
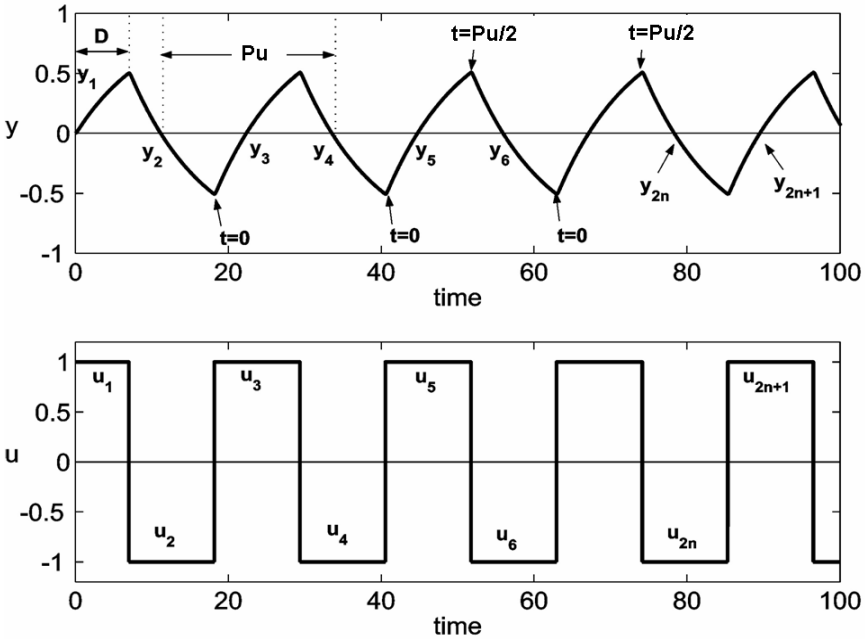
Figure 3.9. Shifted version of relay input u and output y response of a typical SOPDT system

generalized response term y_n slowly forms a convergent series. Let us use a second-order system to illustrate the derivation as they are rich in system dynamics.[1]

3.3.1 Derivation for a Second-order Overdamped System

The transfer function of an SOPDT system with a damping coefficient greater than one can be expressed as $G(s) = K_p e^{-Ds}/[(\tau_1 s+1)(\tau_2 s+1)]$, where K_p is the steady state gain, τ_1 and τ_2 are process time constants with $\tau_1 > \tau_2$, and D is the dead time. The original step response of an overdamped SOPDT can be given by

$$y = K_p[1 - a_1 e^{-(t-D)/\tau_1} + b_1 e^{-(t-D)/\tau_2}]$$

where a_1 and b_1 are given by

$$a_1 = \frac{\tau_1}{\tau_1 - \tau_2} \quad \text{and} \quad b_1 = \frac{\tau_2}{\tau_1 - \tau_2}$$

Under the shifted version (Figure 3.8B), the first segment of the relay response y_1 is simply the step response without dead time in the time index:

$$y_1 = K_p\left[1 - a_1 e^{-t/\tau_1} + b_1 e^{-t/\tau_2}\right] \tag{3.25}$$

At the second instant, the time is reset to zero at the initial point. The step response (relay output) is given by (*i.e.* introducing a time shift by D amount in Equation 3.22)

$$y_2 = K_p\left[1 - a_1 e^{-\frac{t+D}{\tau_1}} + b_1 e^{-\frac{t+D}{\tau_2}}\right] - 2K_p\left[1 - a_1 e^{-\frac{t}{\tau_1}} + b_1 e^{-\frac{t}{\tau_2}}\right]$$

Here, the first term represents the effect of the first step change (occurred at D time earlier) and the second term shows the effect of the second step input, switching to the opposite direction. The above equation can be simplified to

$$y_2 = K_p\left\{[1-2] - a_1 e^{-\frac{t}{\tau_1}}\left(e^{-\frac{D}{\tau_1}} - 2\right) + b_1 e^{-\frac{t}{\tau_2}}\left(e^{-\frac{D}{\tau_2}} - 2\right)\right\} \tag{3.26}$$

The relay response at the third interval is the result of three step changes, lags by an amount $D + P_u/2$ from input. After introducing a time shift of $D + P_u/2$ in Equation 3.22, the net effect becomes

$$y_3 = K_p\left\{\left[1 - a_1 e^{-\frac{t+D+P_u/2}{\tau_1}} + b_1 e^{-\frac{t+D+P_u/2}{\tau_2}}\right]\right.$$
$$\left. - 2\left[1 - a_1 e^{-\frac{t+P_u/2}{\tau_1}} + b_1 e^{-\frac{t+P_u/2}{\tau_2}}\right] + 2\left[1 - a_1 e^{-\frac{t}{\tau_1}} + b_1 e^{-\frac{t}{\tau_2}}\right]\right\}$$

which can be simplified further as

---

[1] One may skip the derivation in Section 3.3.1 and refer directly to Tables 3.2 and 3.3 for the results.

$$y_3 = K_p \left\{ [1-2+2] - a_1 e^{-\frac{t}{\tau_1}} \times \left[e^{-\frac{D+P_u/2}{\tau_1}} - 2e^{-\frac{P_u}{2\tau_1}} + 2 \right] + b_1 e^{-\frac{t}{\tau_2}} \right.$$

$$\left. \times \left[e^{-\frac{D+P_u/2}{\tau_2}} - 2e^{-\frac{P_u}{2\tau_2}} + 2 \right] \right\}$$

(3.27)

It can be seen that the terms in the right-hand side (RHS) of the above equation are slowly forming a series.

With the progress of time, the response becomes stabilized and the general expression for the nth term can be described as

$$y_n = K_p \left\{ [1 - 2 + 2 - \times\!\times] - a_1 e^{-\frac{t}{\tau_1}} \left[e^{-\frac{D+(n-2)P_u/2}{\tau_1}} \right. \right.$$

$$\left. -2e^{-\frac{(n-2)P_u}{2\tau_1}} + 2e^{-\frac{(n-1)P_u}{2\tau_1}} - \times\!\times\!\times + 2e^{-\frac{P_u}{2\tau_1}} - 2 \right]$$

$$\left. + b_1 e^{-\frac{t}{\tau_1}} \left[e^{-\frac{D+(n-2)P_u/2}{\tau_2}} - 2e^{-\frac{(n-2)P_u}{2\tau_2}} + 2e^{-\frac{(n-1)P_u}{2\tau_2}} \right. \right.$$

$$\left. \left. - \times\!\times\!\times + 2e^{-\frac{P_u}{2\tau_2}} - 2 \right] \right\}$$

(3.28)

The RHS of Equation 3.28 has three parts, and each part consists of an infinite series, F_1, F_2 and F_3.

$$y_n = K_p \left\{ F_1 - a_1 e^{-\frac{t}{\tau_1}} F_2 + b_1 e^{-\frac{t}{\tau_2}} F_3 \right\}$$

If n is odd, the first series F_1 is simply

$$F_1 = [1-2+2-2+\times\!\times] = 1$$

The second series becomes:

$$F_2 = \left[e^{-\frac{D}{\tau_1}} r^{n-2} - 2r^{n-2} + 2r^{n-3} - 2r^{n-4} + \times\!\times - 2r + 2 \right]$$

where $r = e^{-v_1}$ and $v_1 = P_u/2\tau_1$. This above series is convergent and can be put into the following form (note that terms are rearranged from the back side of the above expression):

$$F_2 = \lim_{n \to \infty} \left(e^{-D/\tau} r^{n-2} \right) + 2 \left(1 - r + r^2 - r^3 + \times\!\times \right)$$

$$= 2 \left[1 - r + r^2 - r^3 + \times\!\times \right] = \frac{2}{1+r} = \frac{2}{1 + e^{-P_u/2\tau_1}}$$

In a similar way, the F_3 of the RHS of Equation 3.28 can also be simplified. Ultimately, the response can be given by

$$y_n = K_p \left\{ 1 - a_1 e^{-\frac{t}{\tau_1}} \left(\frac{2}{1+e^{-\frac{P_u/2}{\tau_1}}} \div + b_1 e^{-\frac{t}{\tau_2}} \left(\frac{2}{1+e^{-\frac{P_u/2}{\tau_2}}} \right) \right) \right\} \tag{3.29}$$

This represents the ascending response (n is odd). Since this response is dissymmetric, the general form can be employed as

$$y_n = K_p \left\{ -1 + a_1 e^{-\frac{t}{\tau_1}} \left(\frac{2}{1+e^{-\frac{P_u/2}{\tau_1}}} \div - b_1 e^{-\frac{t}{\tau_2}} \left(\frac{2}{1+e^{-\frac{P_u/2}{\tau_2}}} \right) \right) \right\} (-1)^n \tag{3.30}$$

One can refer to Panda and Yu [15] for the derivations for critically damped and underdamped SOPDT systems, as well as for high-order systems.

3.3.2 Results

Different types of transfer function are considered, and the analytical expressions for their relay feedback output response are developed following the above procedure. Table 3.2 gives a list of first-, second-, and third-order plus dead time processes and their corresponding mathematical expressions for the stabilized relay feedback output responses. These equations y_n denote the upward or ascending trend (or sometimes, curves in the lower part of midline for higher order systems) of relay feedback output (while time t changes from 0 to $P_u/2$). The downward or descending trend can be obtained by reversing the sign of the output ($-y_n$).

In Table 3.2, the individual expressions, for relay feedback responses of first-, second- and third-order systems contain terms similar to those of the corresponding equations for the step responses, except that they differ only in weighting factor ($2/(1+e^{-P_u/2\tau})$). If we compare the terms of the expressions of the relay feedback response with those of step response of a process, we see that they differ by a weighting factor of $2/(1+e^{-P_u/2\tau})$. For an FOPDT system, the response starts ($t=0$) from the minimal point, at $y=-a$, and ends ($t=P_u/2$) at the maximal point, at $y=a$. Also note that, for an unstable FOPDT system, stable limit cycles can occur only if $D/\tau < \ln(2)$. For the lead/lag second-order system (No. 6 in Table 3.2), the expression is applicable to systems with left-half plane ($\tau_3 > 0$) or right-half plane ($\tau_3 < 0$) zero.

Analytical expressions of relay feedback output responses for higher order systems are presented in Table 3.3. They are of much interest because, when we see, for example, the expression for fifth-order process, the equation contains mainly five terms (except '1') and each of these terms represents corresponding lower order processes. The first term inside the third bracket of the first line/row appears to be for an FOPDT. The second term (having two terms inside the first bracket) is for an SOPDT (critically damped). The third term (having three terms inside the first bracket) is for a third-order process. The terms in the second row/line (having

Table 3.2. Time response y_n of relay feedback for FOPDT, SOPDT and third-order processes

No	Process Transfer Functions	Time response (y_n) of Relay Feedback (y_n is for ascending part of response; $-y_n$ is for descending part)
1	$\dfrac{K_p e^{-Ds}}{\tau s+1}$	$K_p\left(1-e^{-t/\tau}\left[\dfrac{2}{1+e^{-P_u/2\tau}}\right]\right)$
2	$\dfrac{K_p e^{-Ds}}{\tau s-1}$	$K_p\left(1-e^{t/\tau}\left(\dfrac{2}{1+e^{P_u/2\tau}}\right)\right)$. $(D/\tau < \ln(2))$
3	$\dfrac{K_p e^{-Ds}}{(\tau s+1)^2}$	$K_p\left\{1-e^{-\frac{t}{\tau}}\left[\dfrac{2}{1+e^{-P_u/2\tau}}-2e^{-\frac{t}{\tau}}\left[\dfrac{t/\tau}{1+e^{-P_u/2\tau}}+\dfrac{-(P_u/2\tau)e^{-P_u/2\tau}}{\left(1+e^{-P_u/2\tau}\right)^2}\right]\right]\right\}$
4	$\dfrac{K_p e^{-Ds}}{(\tau_1 s+1)(\tau_2 s+1)}$	$K_p\left\{1-a_1 e^{-\frac{t}{\tau_1}}\left(\dfrac{2}{1+e^{-P_u/2\tau_1}}\right)+b_1 e^{-\frac{t}{\tau_2}}\left(\dfrac{2}{1+e^{-P_u/2\tau_2}}\right)\right\}$. where $a_1 = \dfrac{\tau_1}{\tau_1-\tau_2}$ and $b_1 = \dfrac{\tau_2}{\tau_1-\tau_2}$ $(\tau_1 > \tau_2)$
5	$\dfrac{K_p e^{-Ds}}{\tau^2 s^2 + 2\xi\tau s+1}$	$K_p\left\{1-2\dfrac{e^{-\frac{\xi t}{\tau}}}{\beta}\sin\left(\dfrac{\beta t}{\tau}+\alpha\right)\right\}$. where $\alpha = \tan^{-1}\left(\dfrac{\beta+\beta r\cos(\theta)-\xi r\sin(\theta)}{\xi+\xi r\cos(\theta)+\beta r\sin(\theta)}\right)$. $\beta = \sqrt{1-\xi^2}$, $r = e^{\frac{P_u\xi\tau}{2}}$ and $\theta = P_u\cdot\beta\cdot\tau/2$
6	$\dfrac{K_p(\tau_3 s+1)e^{-Ds}}{(\tau_1 s+1)(\tau_2 s+1)}$	$K_p\left\{1-a_1 e^{-\frac{t}{\tau_1}}\left(\dfrac{2}{1+e^{-P_u/2\tau_1}}\right)+b_1 e^{-\frac{t}{\tau_2}}\left(\dfrac{2}{1+e^{-P_u/2\tau_2}}\right)\right\}$. where $a_1 = \dfrac{\tau_1-\tau_3}{\tau_1-\tau_2}$ and $b_1 = \dfrac{\tau_2-\tau_3}{\tau_1-\tau_2}$ $(\tau_1 > \tau_2)$
7	$\dfrac{K_p}{(\tau s+1)^3}$	$K_p\left\{1-2e^{-t/\tau}\left[\left(\dfrac{1}{1-r}\right)+\left(\dfrac{q}{1-r}+\dfrac{rv}{(1-r)^2}\right)+\dfrac{1}{2!}\left(\dfrac{q^2}{1-r}+\dfrac{2qvr}{(1-r)^2}+\dfrac{v^2 r(1+r)}{(1-r)^3}\right)\right]\right\}$ where $q=t/\tau$, $v=P_u/2\tau$, and $r=-e^{-P_u/2\tau}$
8	$\dfrac{K_p}{(\tau_1 s+1)(\tau_2 s+1)(\tau_3 s+1)}$	$K_p\left\{1+a_1 e^{-\frac{t}{\tau_1}}\left(\dfrac{2}{1+e^{-P_u/2\tau_1}}\right)+b_1 e^{-\frac{t}{\tau_2}}\left(\dfrac{2}{1+e^{-P_u/2\tau_2}}\right)+c_1 e^{-\frac{t}{\tau_3}}\left(\dfrac{2}{1+e^{-P_u/2\tau_3}}\right)\right\}$. where $a_1 = \dfrac{\tau_1^2(\tau_1-\tau_2)}{(\tau_1-\tau_3)}$, $b_1 = \dfrac{\tau_2^2(\tau_3-\tau_2)}{(\tau_1-\tau_2)}$ and $c_1 = \dfrac{\tau_3^2}{(\tau_1\tau_2-\tau_2\tau_3-\tau_3\tau_1+\tau_3^2)}$ $(\tau_1 > \tau_2 > \tau_3)$

Table 3.3. Time response y_n of relay feedback for fourth and high-order processes

No	Process Transfer Functions	Time response (y_n) of Relay Feedback (y_n is for ascending part of response; $-y_n$ is for descending part)
9	$\dfrac{K_p}{(\tau s+1)^4}$	$K_p\left\{1-2e^{-t/\tau}\left[\dfrac{1}{1-r}+\dfrac{q}{1-r}+\dfrac{rv}{(1-r)^2}+\dfrac{1}{2!}\left(\dfrac{q^2}{1-r}+\dfrac{2qvr}{(1-r)^2}+\dfrac{v^2r(1+r)}{(1-r)^3}\right)\right]-\dfrac{2}{3!}e^{-t/\tau}\left[\dfrac{q^3}{1-r}+\dfrac{3q^2vr}{(1-r)^2}+\dfrac{3qv^2r(1+r)}{(1-r)^3}+\dfrac{v^3r(r^2+4r+1)}{(1-r)^4}\right]\right\}$ where $q=t/\tau$, $v=P_u/2\tau$, and $r=-e^{-P_u/2\tau}$
10	$\dfrac{K_p}{(\tau s+1)^5}$	$K_p\left\{1-2e^{-t/\tau}\left[\dfrac{1}{1-r}+\left(\dfrac{q}{1-r}+\dfrac{rv}{(1-r)^2}\right)+\dfrac{1}{2!}\left(\dfrac{q^2}{1-r}+\dfrac{2qvr}{(1-r)^2}+\dfrac{v^2r(1+r)}{(1-r)^3}\right)\right]\right.$ $-\dfrac{2}{3!}e^{-t/\tau}\left[\dfrac{q^3}{1-r}+\dfrac{3q^2vr}{(1-r)^2}+\dfrac{3qv^2r(1+r)}{(1-r)^3}+\dfrac{v^3r(r^2+4r+1)}{(1-r)^4}\right]$ $\left.-\dfrac{2}{4!}e^{-t/\tau}\left[\dfrac{q^4}{1-r}+\dfrac{4q^3vr}{(1-r)^2}+\dfrac{6q^2v^2r(1+r)}{(1-r)^3}+\dfrac{4qv^3r(r^2+4r+1)}{(1-r)^4}+\dfrac{v^4(r^3+11r^2+11r+1)}{(1-r)^5}\right]\right\}$ where $q=t/\tau$, $v=P_u/2\tau$ and $r=-e^{-P_u/2\tau}$
11	$\dfrac{K_p}{(\tau s+1)^n}$	$K_p\left\{1-2e^{-q}\left[\dfrac{1}{0!}\left(\dfrac{1}{1-r}\right)+\dfrac{1}{1!}\left(\dfrac{q}{1-r}+\dfrac{rv}{(1-r)^2}\right)+\dfrac{1}{2!}\left(\dfrac{q^2}{1-r}+\dfrac{2qvr}{(1-r)^2}+\dfrac{v^2(1+r)}{(1-r)^3}\right)+\cdots\right]\right.$ $-\dfrac{2}{3!}e^{-q}\left[\dfrac{q^3}{1-r}+\dfrac{3q^2vr}{(1-r)^2}+\dfrac{3qv^2r(1+r)}{(1-r)^3}+\dfrac{v^3r(r^2+4r+1)}{(1-r)^4}+\cdots\right.$ $-\dfrac{2}{(n-1)!}e^{-q}\left[\dfrac{{}^{n-1}C_0q^{n-1}}{1-r}+\dfrac{{}^{n-1}C_1q^{n-2}vr}{(1-r)^2}+\dfrac{{}^{n-1}C_2q^{n-3}v^2r(1+r)}{(1-r)^3}+\dfrac{{}^{n-1}C_3qv^3r(r^2+4r+1)}{(1-r)^4}+\cdots+{}^{n-1}C_{n-1}v^{n-1}\sum_{d=0}^{\infty}d^{n-2}r^d\right]$ where $q=t/\tau$, $v=P_u/2\tau$, and $r=-e^{-P_u/2\tau}$

four terms inside) are for a fourth-order process. In the third or last row/line there are five terms for a fifth-order process. Hence, the number of terms (size of the series) for a particular order of process is rhythmic. These tables are similar to the tables of inverse Laplace transform and will help in finding an equation for relay feedback responses.
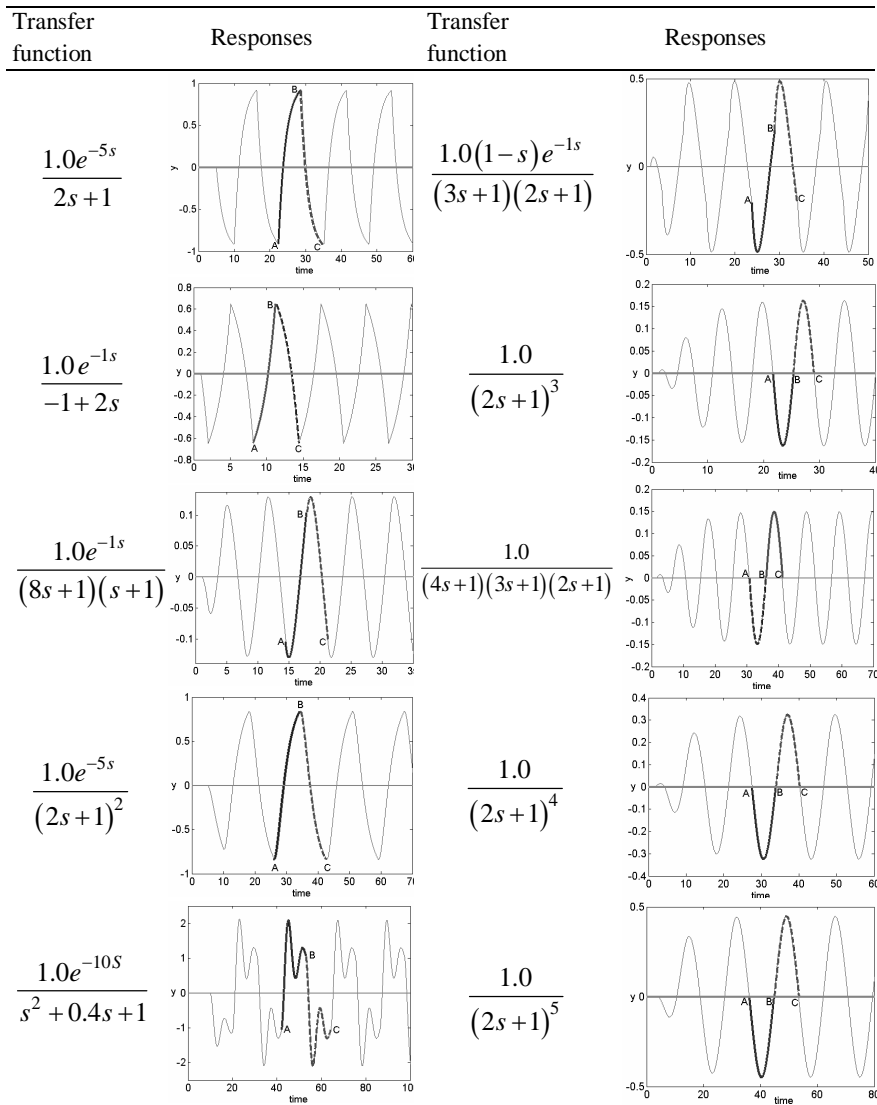
3.3.3 Validation

Two kinds of response can be observed in the analytical expressions in Tables 3.2 and 3.3. These responses are tabulated in Figure 3.10. Systems with serial numbers 1 and 2 in Table 3.2 always produce a monotonic response, where, at $t = 0$, the response from the model starts at the lowermost (or uppermost) point (A or B) and, at $t = P_u / 2$, it ends at the other extreme point (B or C). Processes with serial numbers 3, 4, 5 and 6 in Table 3.2 may give a non-monotonic response, as shown in Figure 3.10. The third type is higher order systems without dead time (*i.e.* $n \geq 3$). For this type of system, this value occurs at the mid-point of the half period, as also shown in Figure 3.10.

Figure 3.10 shows the correctness of the derived mathematical models. If the relay height is other than unity, then the model for the relay output response will be just multiplied by actual value of relay height h.

3.4 Conclusion

In this chapter the relay feedback test is introduced and the steps required to perform the experiment are also given. It can be carried out with or without a commercial autotuner. Once you have obtained the information on the ultimate frequency, the controller settings can be decided using the original or modified Ziegler–Nichols methods. You can also go a step further to find an appropriate transfer function for the process. This can be useful for implementing MPC or dead time compensator (Smith predictor). Better approximation can be achieved using the improved algorithm. Finding transfer functions using the biased relay plus hysteresis was discussed by Wang *et al.* [12]. Finally, analytical expressions for relay feedback responses are tabulated for different types of process. This can be useful if the model structure is known.

Transfer function	Responses	Transfer function	Responses
$\dfrac{1.0e^{-5s}}{2s+1}$		$\dfrac{1.0(1-s)e^{-1s}}{(3s+1)(2s+1)}$	
$\dfrac{1.0\,e^{-1s}}{-1+2s}$		$\dfrac{1.0}{(2s+1)^{3}}$	
$\dfrac{1.0e^{-1s}}{(8s+1)(s+1)}$		$\dfrac{1.0}{(4s+1)(3s+1)(2s+1)}$	
$\dfrac{1.0e^{-5s}}{(2s+1)^{2}}$		$\dfrac{1.0}{(2s+1)^{4}}$	
$\dfrac{1.0e^{-10S}}{s^{2}+0.4s+1}$		$\dfrac{1.0}{(2s+1)^{5}}$	

Figure 3.10. Validation of analytical expressions for relay output of different systems: solid line is relay output and dashed line is model output. (A denotes starting of one cycle that ends at B. Again from B next cycle starts and ends at C).

3.5 References

1. Åström KJ, Hägglund T. Automatic tuning of simple regulators with specifications on phase and amplitude margins. Automatica 1984;20:645.

2. Luyben WL. Derivation of transfer functions for highly nonlinear distillation columns. Ind. Eng. Chem. Res. 1987;26:2490.

3. Seborg DE, Edgar TF, Mellichamp DA. Process dynamics and control. 2nd ed. New York: Wiley; 2004.

4. Luyben WL, Luyben ML. Essentials of process control. New York: McGraw-Hill; 1997.

5. Wang QG, Lee TH, Lin C. Relay feedback. London: Springer-Verlag; 2003.

6. Ogata K. Modern control engineering. Prentice-Hall: Englewood Cliffs; 1970.

7. Tyreus BD, Luyben WL. Tuning PI controllers for integrator/dead time processes. Ind. Eng. Chem. Res. 1992;31:2625.

8. Friman M, Waller KV. Autotuning of multiloop control systems. Ind. Eng. Chem. Res. 1994;33:1708.

9. Wood RK, Berry MW. Terminal composition control of a binary distillation column. Chem. Eng. Sci. 1973;28:1707.

10. Chang RC, Shen SH, Yu CC. Derivation of transfer function from relay feedback systems. Ind. Eng. Chem. Res. 1992;31:855.

11. Li W, Eskinat E, Luyben WL. An improved autotune identification method. Ind. Eng. Chem. Res. 1991;30:1530.

12. Wang QG, Hang CC, Zou B. Low-order modeling from relay feedback. Ind. Eng. Chem. Res. 1997;36:375.

13. Majhi S, Atherton DP. Auto-tuning and controller design for processes with small time delays. IEE Proc. Control Theory Appl. 1999;146(3):415.

14. Kaya I, Atherton DP. Parameter estimation from relay auto-tuning with asymmetric limit cycle data. Process Control 2001;11:429.

15. Panda RC, Yu CC. Analytical expressions for relay feedback responses. J. Process Control 2003;13:48.