Automatic Tuning of Digital Controllers with Applications to HVAC Plants

Wallenborg, Anders; Åström, Karl Johan; Hägglund, Tore

Published in:
Automatica

DOI:
10.1016/0005-1098(93)90053-V

1993

Automatica, Vol. 29, No. 5, pp. 1333–1343, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Automatic Tuning of Digital Controllers with Applications to HVAC Plants*

K. J. ÅSTRÖM,† T. HÄGGLUND† and A. WALLENBORG‡

A tuning method based on relay feedback for a general digital controller is presented and applied to heating, ventilation and air-conditioning plants.

Key Words—Automatic tuning; adaptive control; digital control; air conditioning, ventilation.

Abstract—It has been demonstrated that PID controllers can be tuned effectively based on an experiment with relay feedback. This paper describes a tuning method for a general digital controller based on relay feedback. The control design method is based on pole placement. An interesting feature is that the sampling period and the desired closed-loop poles are determined from the experiment. The method is also suited for pretuning of adaptive algorithms. The paper describes the basic ideas, which are illustrated by simulations. Results from tests on HVAC (heating, ventilation and air-conditioning) plants are also reported.

1. INTRODUCTION

IN SPITE OF many advances in control theory, simple controllers of the PID type are still used in the majority of control loops (see Deshpande and Ash, 1981; McMillan, 1983). Lately there have been significant efforts to give PID controllers added capabilities by providing facilities for automatic tuning, gain scheduling and adaptation (see Bristol, 1977; Hägglund and Åström, 1991). This has drastically simplified the use of PID controllers and significantly improved their performance.

One procedure for automatic tuning was proposed in Åström and Hägglund (1984). It is based on determination of the ultimate period and the ultimate gain from a simple experiment with relay feedback. Several industrial products based on this idea are now available on the market (Åström and Hägglund, 1988a, 1990). An attractive feature of relay tuning is that it is easy to use. It can be implemented in such a way that tuning is done simply by pushing a button.

The PID controller does, however, have some drawbacks. It performs poorly for processes with long dead-time and it requires unnecessary fast sampling. It is thus of interest to use other control algorithms to cope with processes with time delay and to provide those controllers with some tuning facility.

In this paper we take a different approach to single loop control. The key idea is to choose a general digital control algorithm and to develop a method for tuning such a controller. By choosing a digital controller the problem of finding an appropriate discretization of a continuous time controller is avoided. The sampling period is determined with respect to the process dynamics which means that the computing power is used economically. The chosen controller structure admits dead-time compensation.

2. PULSE TRANSFER FUNCTION IDENTIFICATION

To design digital control laws it is necessary to know the sampling period and a discrete time process model. To determine a process model experimentally, a sampling period is first chosen, perturbation signals are then introduced and the process model is finally obtained from some parameter estimation method (Ljung and Söderström, 1983). In self-tuning control, the perturbations are generated by conventional feedback and the parameters are estimated recursively. The sampling period is a crucial parameter both in conventional parameter estimation and in adaptive control. Prior knowledge about the timescale of the process and the closed-loop system is required to determine the sampling period. This fact has for a long time been a stumbling block for automatic modeling and adaptive control.

* Received 20 April 1992; revised 19 January 1993; received in final form 20 February 1993. The original version of this paper was presented at the IFAC Symposium on Adaptive Control of Chemical Processes which was held in Lyngby, Denmark during August 1988. The Published Proceedings of this IFAC Meeting may be ordered from: Pergamon Press Ltd, Headington Hill Hall, Oxford OX3 0BW, U.K. This paper was recommended for publication in revised form by Editor C. C. Hang.

† Department of Automatic Control, Lund Institute of Technology, S-221 00 Lund, Sweden.
‡ Tour & Andersson AB, Controls Division, Jägershillgatan 18, S-213 75 Malmö, Sweden.

A tuning method which bypasses the difficulties discussed above was proposed in Åström and Hägglund (1984). The idea is that most plants will exhibit a periodic oscillation under relay feedback. The amplitude and the period of the oscillation can then be used to tune a controller based on conventional design methods of the Ziegler–Nichols type. The tuning methods discussed in Åström and Hägglund (1984) are only based on knowledge of the amplitude and the period of the oscillation. We will here show that conventional sampled data models can be determined using the wave-form of the oscillation. The parameters can of course be determined by any standard estimation method. A significant simplification can, however, be obtained by exploiting the periodic nature of the system.

The starting point is that a relay feedback experiment in stationarity gives periodic input and output signals as shown in Fig. 1. The period of the oscillation is approximately the ultimate period under proportional feedback. This period can be used as a basis for selecting the sampling period. It will now be shown how a model can be fitted to the results of an experiment with relay feedback.

Consider a system with a periodic input signal. Assume that the system is sampled with a period corresponding to N samples per period. The input signal is then given by the values

$$\begin{aligned} &u_0, u_1, u_2, \ldots, u_{N-1} \\ &u_{N+i} = u_i \quad \text{for} \quad i \geq 0 \end{aligned} \tag{1}$$

where u denotes deviations from the mean value of the oscillation. Let the steady-state output be

$$\begin{aligned} &y_0, y_1, y_2, \ldots, y_{N-1} \\ &y_{N+i} = y_i \quad \text{for} \quad i \geq 0 \end{aligned} \tag{2}$$

where y also denotes deviations from the mean value.

A standard input–output model for a linear system can be written as

$$A(q)y(k) = B(q)u(k) \tag{3}$$

where $A(q)$ and $B(q)$ are polynomials in the forward shift operator. The process dead-time d

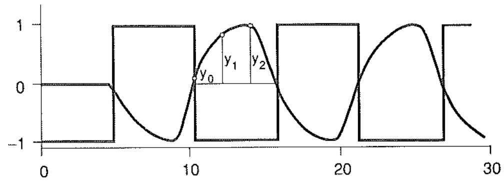

FIG. 1. Input and output signals obtained from an experiment with relay feedback.

is included in the A polynomial. Hence

$$d = \deg A - \deg B.$$

The estimation problem is thus to determine the model (3) from the data (1) and (2). A simple scheme for doing this was described in Åström and Hägglund (1988b). The z-transforms of the signals (1) and (2) are

$$\begin{aligned} U(z) &= \frac{u_0 z^N + u_1 z^{N-1} + \cdots + u_{N-1} z}{z^N - 1} \\ &= \frac{E(z)}{z^N - 1} \\ Y(z) &= \frac{y_d z^N + y_{d+1} z^{N-1} + \cdots + y_{d+N-1} z}{z^d (z^N - 1)} \\ &= \frac{D(z)}{z^d (z^N - 1)}. \end{aligned}$$

The particular representation of the z-transform of the periodic function y is chosen to be causally compatible with the model (3). It follows from equation (3) that

$$Y(z) = \frac{B(z)}{A(z)} U(z) + \frac{Q(z)}{A(z)}$$

where the polynomial $Q(z)$ corresponds to initial conditions which give the steady-state periodic output. Introducing the expressions for $Y(z)$ and $U(z)$ given above we find

$$\frac{D(z)}{z^d (z^N - 1)} = \frac{B(z)E(z)}{(z^N - 1)A(z)} + \frac{Q(z)}{A(z)}$$

or

$$A(z)D(z) - z^d B(z)E(z) = z^d (z^N - 1)Q(z). \tag{4}$$

This is a Diophantine equation for determining the polynomials A, B and Q. It can be solved if $D(z)$ and $E(z)$ are relatively prime.

Relay feedback oscillations

The special case when the input is a symmetric square wave oscillation is of particular interest. It is then sufficient to consider only a half period. Without loss of generality we can assume that the relay gives an output with unit amplitude (see Fig. 1). If $n = N/2$, the input signal is then given by

$$\begin{aligned} &u_0 = u_1 = \cdots = u_{n-1} = -1 \\ &u_{n+i} = -u_i \quad \text{for} \quad i \geq 0. \end{aligned} \tag{5}$$

The output signal is given by

$$\begin{aligned} &y_0, y_1, \ldots, y_{n-1} \\ &y_{n+i} = -y_i \quad \text{for} \quad i \geq 0. \end{aligned} \tag{6}$$

The z-transforms of the signals are

$$U(z) = -\frac{z^n + z^{n-1} + \cdots + z}{z^n + 1} = \frac{E(z)}{z^n + 1}$$

$$Y(z) = \frac{y_d z^n + y_{d+1} z^{n-1} + \cdots + y_{d+n-1} z}{z^d (z^n + 1)} = \frac{D(z)}{z^d (z^n + 1)}.$$

Equation (4) then becomes

$$A(z)D(z) - z^d B(z)E(z) = z^d (z^n + 1)Q(z). \quad (7)$$

Equating coefficients of equal powers of z we get $n + \deg A + 1$ linear equations to determine the coefficients in the model. The number of unknown parameters in the polynomials A, B and Q is $3\deg(A) - 2d + 2$. To determine these parameters, it is therefore necessary to have $n \geq 2(\deg(A) - d) + 1$. If $n > 2(\deg(A) - d) + 1$, the number of equations is larger than the number of unknown parameters. It is therefore suggested to use some kind of minimization technique to determine the parameters in this case.

The identification procedure can be summarized as follows:

1. Introduce relay feedback and wait for steady-state conditions.
2. Determine a suitable sampling period $h = T_u/2n$ such that $n \geq 2(\deg(A) - d) + 1$, *i.e.* based on the oscillation period T_u and the model complexity.
3. Determine $u_0, u_1, \ldots, u_{n-1}$ and $y_0, y_1, \ldots, y_{n-1}$. Compute averages over several periods to compensate for noise.
4. Calculate the process model from (7).

This identification procedure both provides a suitable sampling period of the controller and gives information about reasonable choices of closed-loop poles. This will be further discussed in the next section.

Example—first-order system with time delay

The identification procedure will now be applied to a process with the transfer function

$$G(s) = K\frac{e^{-sL}}{1 + sT} \quad (8)$$

i.e. a first-order process with static gain K, time constant T and time delay L. The corresponding pulse transfer function is

$$H(z) = \frac{b_1 z + b_2}{z^d (z - a)}$$

The model has four parameters, the integer d and the real numbers a, b_1, and b_2. The integer d must be less than or equal to n. To determine these parameters it is thus necessary to have at least three measurements per half period, *i.e.* $n \geq 3$. The following examples show how the models are obtained using the identification method described above. The cases $d = 1$, 2 and 3 will be considered separately.

Example 1. Time delay $d = 1$. Consider the process model

$$y(t+1) = ay(t) + b_1 u(t) + b_2 u(t-1) \quad (9)$$

which corresponds to sampling the model (8) with a time-delay L less than one sampling period. In this case we have

$$A(z) = z(z - a)$$
$$B(z) = b_1 z + b_2.$$

For $n = 3$, equation (7) becomes

$$z(z-a)(y_1 z^3 + y_2 z^2 - y_0 z) + z(b_1 z + b_2)(z^3 + z^2 + z) = z(z^3 + 1)(q_0 z + q_1).$$

It follows that $q_1 = 0$, since the left-hand side lacks first-order terms. The equation can therefore be reduced to

$$(z - a)(y_1 z^2 + y_2 z - y_0) + (b_1 z + b_2)(z^2 + z + 1) = q_0(z^3 + 1).$$

Equating coefficients of equal powers of z gives four equations to determine the unknowns a, b_1, b_2 and q_0. These equations have a solution if $y_1 \neq y_2$. It is given by

$$\begin{aligned} a &= \frac{y_0 + y_2}{y_1 - y_2} \\ b_1 &= \frac{y_0^2 - y_1^2 + y_2^2 + y_0 y_1 + y_0 y_2 + y_1 y_2}{2(y_1 - y_2)} \\ b_2 &= \frac{-y_0^2 + y_1^2 + y_2^2 + y_0 y_1 - y_0 y_2 - y_1 y_2}{2(y_1 - y_2)} \\ q_0 &= \frac{y_0^2 + y_1^2 + y_2^2 + y_0 y_1 + y_0 y_2 - y_1 y_2}{2(y_1 - y_2)}. \end{aligned} \quad (10)$$

Example 2. Time delay $d = 2$. Consider a first-order system where the time delay L is between one and two sampling intervals. The sampled model of such a system is

$$y(t+1) = ay(t) + b_1 u(t-1) + b_2 u(t-2). \quad (11)$$

If we observe that this case is the same as shifting the output one sampling interval to the right in Example 1, we find that the result is obtained from the previous case by making the cyclic permutation

$$y_0 \rightarrow y_1, \quad y_1 \rightarrow y_2 \quad \text{and} \quad y_2 \rightarrow -y_0.$$

If $y_0 \neq -y_2$, parameter a is e.g. given by

$$a = \frac{y_1 - y_0}{y_0 + y_2}. \quad (12)$$

Example 3. Time delay $d = 3$. Consider a first-order system where the time delay L is between two and three sampling intervals. The sampled model of such a system is

$$y(t+1) = ay(t) + b_1 u(t-2) + b_2 u(t-3). \quad (13)$$

If we observe that equation (13) is obtained from (11) by shifting the output one sampling interval to the right we find that the result is obtained from Example 2 by making cyclic permutation

$$y_0 \rightarrow y_1,\ y_1 \rightarrow y_2 \quad \text{and} \quad y_2 \rightarrow -y_0.$$

If $y_0 \neq y_1$, parameter a is e.g. given by

$$a = \frac{y_1 - y_2}{y_0 - y_1}. \quad (14)$$

Three different models were obtained above. They correspond to the cases that the dead-time L of the process is in the ranges

$$0 \leq L < \tfrac{1}{6} T_p$$
$$\tfrac{1}{6} T_p \leq L < \tfrac{1}{3} T_p$$
$$\tfrac{1}{3} T_p \leq L < \tfrac{1}{2} T_p$$

where T_p is the period of the limit cycle. Notice that we cannot have $L > T_p/2$, i.e. that the dead-time cannot be longer than half the oscillation period. Also notice that

$$a^1 a^2 a^3 = -1$$

where a^i is the a coefficient obtained for $d = i$. This means that at least one of the numbers a^i is negative. Such a model cannot correspond to a system with the transfer function (8). In a specific situation we thus only have to consider two values of d.

The identification procedure can of course also be applied to systems of higher order. It is then necessary to use more measurements per sampling period. With the accuracy that is normally obtained in practical situations, it is not realistic to consider models having very high order. In Appendix A1, the identification procedure is applied to a second-order process with time delay.

3. CONTROL DESIGN

A general linear control algorithm for a system with measured signal y and set point y_{sp} can be described by the difference equation

$$\begin{aligned} u(t) + r_1 u(t-1) + \cdots + r_k u(t-k) \\ = t_0 y_{sp}(t) + t_1 y_{sp}(t-1) + \cdots + t_m y_{sp}(t-m) \\ - s_0 y(t) - s_1(t-1) - \cdots - s_l y(t-l) \end{aligned}$$

where u is the control variable and the sampling period h has been chosen as the time unit. In shorthand notation this equation can be described by

$$R(q)u(t) = T(q)y_{sp}(t) - S(q)y(t)$$

where R, S and T are polynomials in the forward shift operator q. Since integral action is required in most process control applications the polynomial $R(q)$ must have $q - 1 = \Delta$ as a factor. The control law then becomes

$$R_1(q)\Delta u(t) = T(q)y_{sp}(t) - S(q)y(t). \quad (15)$$

Equation (15) will therefore be the standard form that we will use.

The parameter estimation procedure presented in the previous section gives information that is very useful to assess the control problem and to select a suitable controller. An estimate of the ratio of the apparent dead-time to the apparent time constant can be obtained from the model. This makes it possible to choose a suitable sampling interval and to judge the difficulty of the control problem. The experiment with relay feedback also gives the ultimate frequency ω_u which is a good indication of the closed-loop bandwidth that can be achieved. The estimation procedure also gives an approximation of the process dynamics in terms of a low-order pulse transfer function.

When a process model of the form (3) is available there are many design methods that can be used to obtain a control law. A pole placement design where natural frequency ω and relative damping ζ of the dominant poles are specified is one alternative. The design parameters can be chosen automatically. Parameter ζ can be fixed and frequency ω can be chosen as $\omega = \omega_u$ for systems with low-order dynamics. For systems with a large pole excess this value of ω is, however, too large. In those cases where ω is reduced we may also consider a new experiment at a lower frequency. Otherwise, the input signal is not ideal for thle parameter estimation (Åström and Wittenmark, 1990).

There are many ways to carry out the design. Since the model is obtained in terms of a rational function it is convenient to use polynomial methods. We will now show how such a method can be applied to the system in Examples 1, 2 and 3.

Example 4. Consider the system (9) obtained in Example 1, i.e.

$$A(z) = z(z-a)$$
$$B(z) = b_1 z + b_2.$$

Assume that the desired behavior of the

closed-loop system is described by a reference model with the pulse transfer function $B_m(z)/A_m(z)$, where

$$A_m(z) = z^2 + p_1 z + p_2$$
$$B_m(z) = \beta B(z)$$

and β is chosen so that $B_m(z)/A_m(z)$ has unit static gain. The polynomial $A_m(z)$ is chosen so that

$$p_1 = -2e^{-\zeta\omega h} \cos(\zeta\omega h\sqrt{1-\zeta^2})$$
$$p_2 = e^{-2\zeta\omega h}.$$

This corresponds to a second-order response with relative damping ζ and frequency ω. The observer polynomial must also be given to obtain a complete specification see (Åström and Wittenmark, 1990). For simplicity we choose a dead-beat observer of the same degree as the polynomial $A(z)$. The observer polynomial is thus

$$A_o(z) = z^{\deg(A)} = z^2.$$

Using the controller (15), the closed-loop transfer function becomes

$$\frac{B(z)T(z)}{A(z)(z-1)R_1(z) + B(z)S(z)}.$$

Comparing this with the desired closed-loop transfer function

$$\frac{A_o(z)B_m(z)}{A_o(z)A_m(z)} = \frac{A_o(z)\beta B(z)}{A_o(z)A_m(z)}$$

gives the following Diophantine equation to determine the polynomials R_1 and S

$$A(z)(z-1)R_1(z) + B(z)S(z) = A_o A_m(z).$$

The polynomial T is determined from the condition that the closed-loop transfer function should be equal to one for $z = 1$. This gives

$$T(1) = S(1).$$

It is straightforward to find the following minimal degree solution

$$R_1(z) = z + r_1$$
$$S(z) = s_0 z^2 + s_1 z$$
$$T(z) = t_0 z^2$$

where

$$s_0 = \frac{1}{1-a}\left(\frac{A_m(1)}{B(1)} - a\frac{A_m(a)}{B(a)}\right)$$
$$s_1 = \frac{a}{a-1}\left(\frac{A_m(1)}{B(1)} - \frac{A_m(a)}{B(a)}\right)$$
$$r_1 = -\frac{b_2}{a}s_1$$
$$t_0 = s_0 + s_1.$$

The control law is then given by

$$u(t) = t_0 y_{sp} - s_0 y(t) - s_1 y(t-h) + (1-r_1)u(t-h) + r_1 u(t-2h).$$

The design procedure applied to the process models obtained in Examples 2 and 3 is given in Appendix A2. Notice that the complexity of the controllers obtained depends on the delay structure of the system. A dead-beat observer is used in all the examples. A more robust design is obtained by choosing observer poles that are removed from the origin, *e.g.* to $z = 0.2$ (Lennartson, 1987).

4. MODEL VALIDATION

In all modeling schemes it is necessary to validate the model obtained. In the schemes discussed in Section 2 it is necessary to determine that the model order used is reasonable. It is also necessary to choose among models obtained for different values of d. This means in general to choose among n different process models. One effective way to do this is to calculate the waveform in several intermediate points and compare this with measurements. This is illustrated by an example.

Example 5. Model validation. To illustrate the valuation procedure we will show the results when the identification procedure is applied to a process with the transfer function

$$G(s) = \frac{e^{-4s}}{(s+1)^2}. \tag{16}$$

The results of a relay experiment with the process is shown in Fig. 1. The relay amplitude was 1 and the hysteresis level 0.1. With six samples per period we get

$$y_0 = 0.106$$
$$y_1 = 0.782$$
$$y_2 = 0.956.$$

The reason why y_0 is slightly greater than the hysteresis level (0.1) is due to the sampling. Since $n = 3$, three first-order process models are obtained. The model parameters are given in Table 1.

The model for $d = 1$ can immediately be excluded since the parameter a is negative. For

TABLE 1. PARAMETERS OF MODELS HAVING DIFFERENT DELAY

d	a	b_1	b_2	K	T	L
1	−6.11	−3.59	−2.16			
2	0.636	0.128	0.586	2.0	4.3	3.6
3	0.257	0.554	0.202	1.0	1.4	4.7

the other two models, corresponding continuous time models with the structure

$$G(s) = K\frac{e^{-sL}}{1+sT} \tag{17}$$

have been computed from the equations

$$K = \frac{b_1 + b_2}{1 - a}$$

$$T = -\frac{h}{\ln a}$$

$$L = hd + T\ln\frac{ab_1 + b_2}{b_1 + b_2}$$

where $h = 1.94$ is the sampling period. The numerical values are given in Table 1.

Figure 2 shows the process output obtained from the relay feedback experiment together with the outputs obtained from the continuous time models corresponding to $d = 2$ and $d = 3$. All outputs will of course coincide at the sampling instants. There are, however, significant deviations between the sampling instants.

From Fig. 2 it is obvious that the model corresponding to time delay $d = 3$ gives the best fit to the measured data. The integral of the absolute errors (IAE) between the model and the measured data is 2.4 times larger for the model with $d = 2$ than for the model with $d = 3$.

The model (17) is simpler than the transfer functions (16). The dead-time L is overestimated and the two lags of 1 sec in (16) are replaced by a single lag of 1.4 sec. Notice that the sums of the dead-times and the lags are quite close, 6.0 for (16) and (6.1) for (17).

In Fig. 3 we show the Nyquist curve of the continuous time model (16), and the estimated continuous time model (17) with $d = 3$. Notice

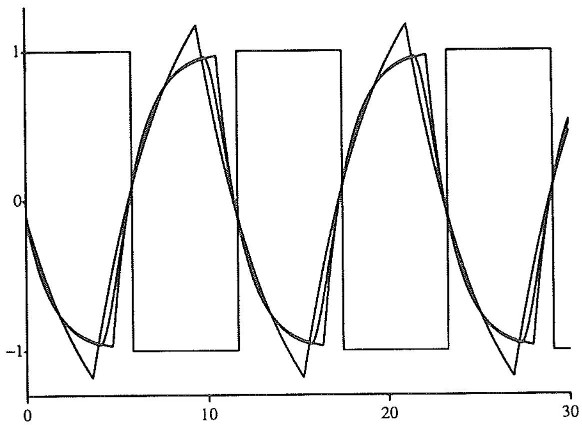

FIG. 2. The process input and output obtained from the relay feedback experiment together with the outputs obtained from the continuous time models corresponding to $d = 2$ and $d = 3$. (The model corresponding to $d = 3$ is the one closest to the process output.)

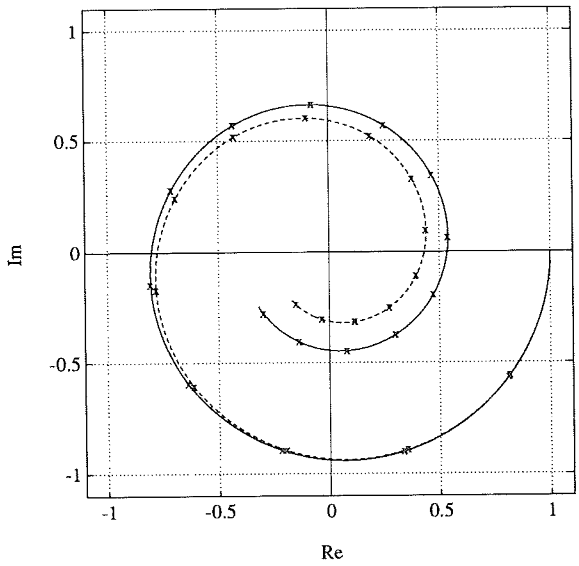

FIG. 3. Comparison of the estimated transfer function (solid line) and the true transfer function (dashed line).

the remarkably good agreement, particularly below the ultimate frequency. This is the frequency range of importance for design of simple controllers.

The controller becomes

$$\Delta u(t) + r_1\Delta u(t-1) + \cdots + r_k\Delta u(t-k) = t_0 y_{sp}(t) - s_0 y(t) - s_1 y(t-1) \tag{18}$$

where

$$t_0 = s_0 + s_1$$

to have the correct steady state. This controller can be interpreted as a PI controller with dead-time compensation. If the undamped closed-loop frequency is chosen to correspond to the period of the relay oscillation and if the relative damping is specified to $\zeta = 0.707$, the controller parameters are

$$s_0 = 0.925 \quad r_2 = 0.665$$

$$s_1 = -0.232 \quad r_3 = 0.182$$

$$r_1 = 0.553.$$

Figure 4 shows how the closed-loop system

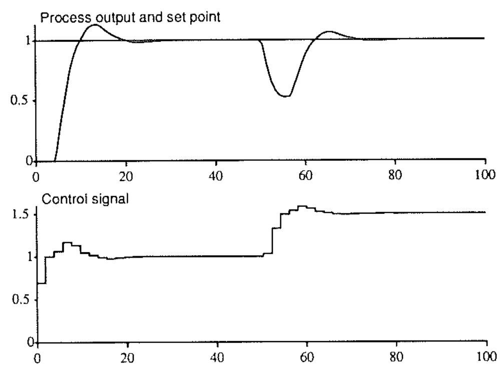

FIG. 4. Response of the closed-loop system obtained when the design procedure is applied to a system with the transfer function $G(s) = e^{-4s}/(s+1)^2$.

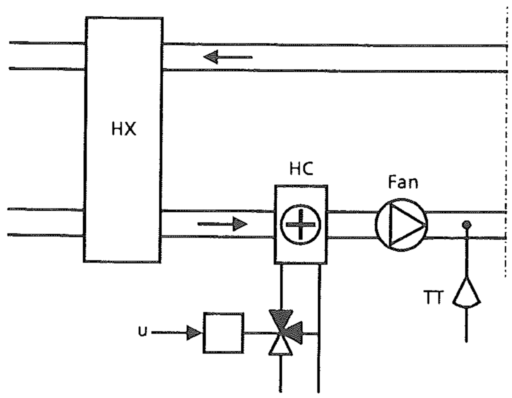

FIG. 5. Schematic diagram of air-handling unit with supply air temperature control. HX = rotary heat exchanger, HC = heating coil, TT = temperature transducer, and u = control signal to valve actuator.

responds to step changes in setpoint and load. The controller gives good performance both with respect to setpoint changes and load disturbance rejection.

5. APPLICATIONS TO HVAC PLANTS

During the last decade, there has been a breakthrough in digital control and modern computer-based building management systems for supervision, monitoring and control of HVAC (heating, ventilation and air conditioning) plants (Wallenborg, 1991). With these tools, it is much easier to evaluate the control performance than before. Therefore, in the future it can be expected that more emphasis will be put on control performance in HVAC system specifications. Consequently, more effort must be put into controller tuning in order to fulfill the increased demands. Both time and money can thus be saved in the commissioning procedure by introducing controllers with automatic tuning facilities.

The new algorithm has been successfully tested on a number of different HVAC plants (Wallenborg, 1991). We will present here two examples: control of supply air temperature and of air duct pressure. Some of the practical issues that were considered during the tests will also be presented.

Control of supply air temperature

The first example is taken from an air-handling unit where outdoor air is heated before being distributed to the interior of a building. The air is heated in two stages. First, the incoming fresh air is preheated by a rotating heat exchanger that recovers excess heat energy from the return air before it leaves the building. Then the supply air is heated to the desired temperature with a heating coil, *i.e.* a water-to-air heat exchanger. Figure 5 shows a schematic diagram of the process. The control signal is the position of the actuator for the valve that controls the hot water supply to the heating coil. The measured process output is the supply air temperature beyond the heating coil and the fan.

Figure 6 shows results of a tuning experiment. The upper graph shows the measured supply air temperature, T(°C), and the lower graph shows the control valve actuator position, u(%). The relay amplitude is automatically adjusted to obtain a desired limit cycle amplitude.

Figure 7 shows the step response of the closed-loop system with controller parameters obtained from the tuning experiment in Fig. 6. The graphs show the setpoint (dashed line), the measured supply air temperature, T(°C), and the control valve actuator position, u(%). The

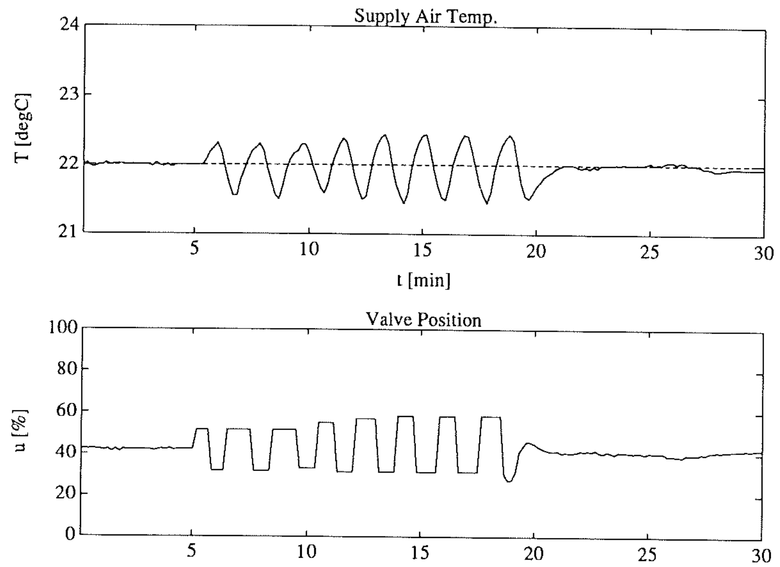

FIG. 6. Tuning experiment on air-handling unit with supply air temperature control.

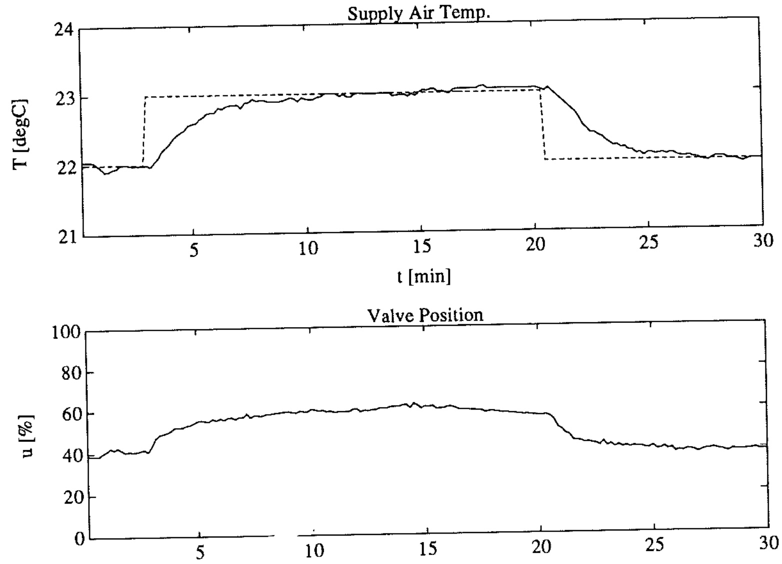

FIG. 7. Closed-loop supply air temperature step response with controller parameters from the tuning experiment in Fig. 6.

step response is well behaved without overshoot. This type of response is the most requested in HVAC systems.

Control of air duct pressure

In the second example, the autotuner is used for control of supply air duct pressure. Figure 8 shows a schematic diagram of the process. The static pressure in the main duct is controlled by adjusting the inlet guide vanes (pitch control) at the fan entry. The control signal is the position of the guide vane actuator (0–100%). The process output is the static air pressure, measured with a pressure sensor located downstream of the fan.

Figure 9 shows a tuning experiment performed on the main duct pressure control loop. The graphs show the measured static pressure, p (Pa), and the guide van actuator position, u (%). The wave forms of both the measured duct pressure and the control signal during the tuning experiment are distorted due to the rather slow sampling interval (10 sec) of the trend-logging program.

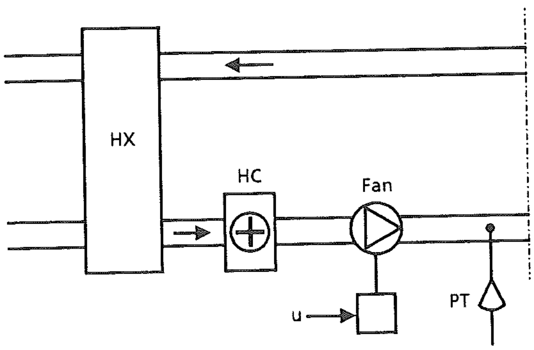

FIG. 8. Schematic diagram of air-handling unit with main duct static air pressure control. HX = rotary heat exchanger, HC = heating coil, PT = pressure transducer, and u = control signal to inlet guide vanes.

In this experiment it was more difficult to obtain a stable oscillation because the selected value of the oscillation amplitude, 10 Pa, is quite small compared with normal pressure disturbances.

Figure 10 shows the step response of the closed-loop system with controller parameters obtained from the tuning experiment in Fig. 9. The graphs show the pressure setpoint (dashed line), the measured duct static pressure and the guide vane actuator position. The step responses are also well behaved here, with only a minor overshoot.

Practical issues

Several practical problems were encountered and solved during the tests on different HVAC plants. Some of these will now be discussed.

Bias adjustment. In order to counteract load disturbances during the tuning procedure and to ensure a symmetric oscillation, a bias signal may be added to the relay output. The offset level should correspond to the mean value of the relay output control signal. It may be calculated on-line during the tuning procedure as

$$b = d\frac{t_{pos} - t_{neg}}{t_{pos} + t_{neg}}$$

where b is the bias level, d is the relay amplitude, t_{pos} is the length of the positive half period, and t_{neg} is the length of the negative half period. For further details, see Hang and Åström (1988).

Automatic relay amplitude adjustment. In some

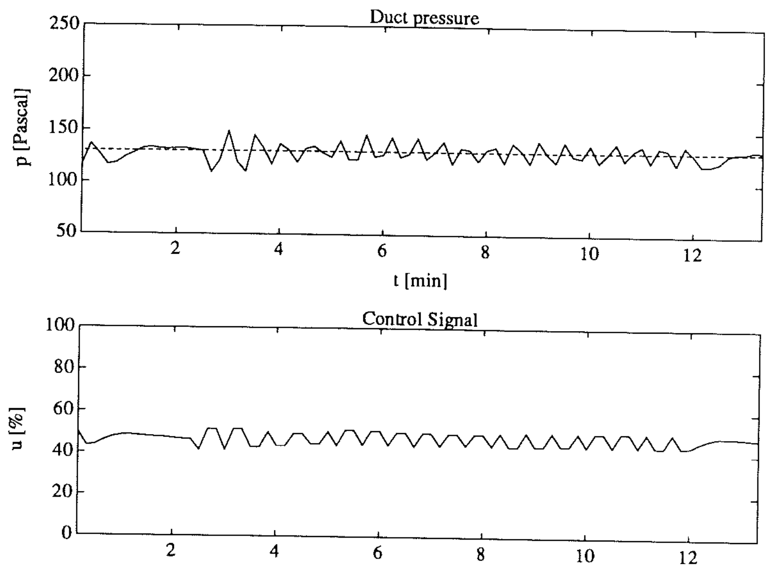

FIG. 9. Tuning experiment on air-handling unit with static air pressure control.

applications, it may be desired to keep the fluctuations in the measured signal within certain limits, so that the tuning procedure can be carried out without disturbing the process too much. This can be achieved by adjusting the relay amplitude on-line during the tuning procedure. A simple algorithm for automatic adjustment of the relay amplitude is

$$d_{\text{new}} = d\frac{a_{\text{ref}}}{a}$$

where a_{ref} is the desired amplitude setpoint. This adjustment rule is actually equivalent to an integrating controller.

It normally takes a few periods before a steady state oscillation is obtained after a change in relay amplitude. It is therefore recommendable to either low-pass filter d_{new} or to apply the adjustment rule at such a low rate that steady state is obtained after each adjustment.

Pretune control design. In complex plants with many control loops, *e.g.* large HVAC systems, it may be cumbersome to use manual control to bring the plant to normal steady-state operating conditions suitable for the tuning procedure. The autotuner should therefore be provided with some initial parameter values which allow one to close the loops before the first tuning procedure has been carried out. Quite often PI control parameters are available which at least stabilize the loops, even if the control performance is poor. These parameters can be used to initialize the autotuner, *e.g.* in the following way.

Consider a discrete time PI controller with

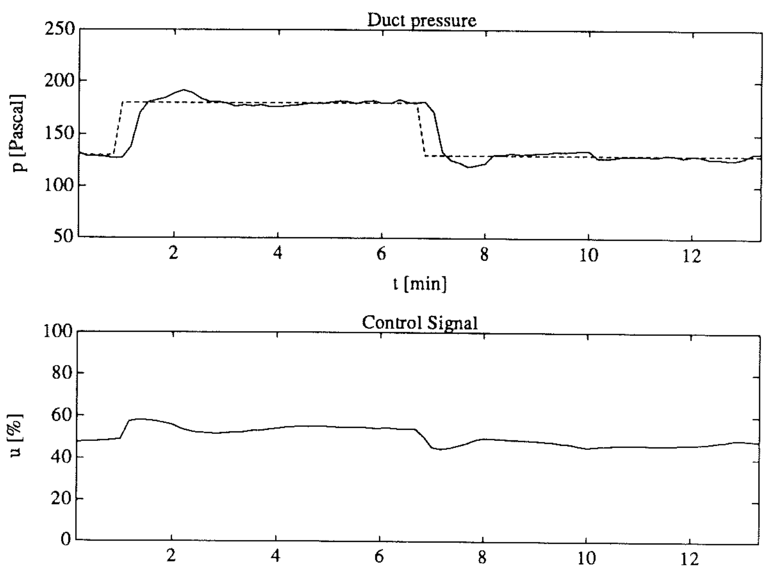

FIG. 10. Closed-loop static air pressure step response with controller parameters from the tuning experiment in Fig. 9.

proportional gain K, integral time T_i, and the sampling interval h. The discrete time relation between the control error e and the output u of the PI controller is given by

$$u(t) = \frac{K(q - (1 - h/T_i))}{q - 1} e(t).$$

Comparing this with the general discrete time controller

$$R(q)u(t) = T(q)y_{sp}(t) - S(q)y(t)$$

gives the following controller polynomials:

$$R(q) = q - 1$$
$$S(q) = Kq - K(1 - h/T_i)$$
$$T(q) = Kq - K(1 - h/T_i).$$

These formulae can be used in a pretune design mode to select initial values for the control parameters.

6. CONCLUSIONS

This paper proposes a simple method for tuning digital control laws directly. It is believed that this method is superior to the techniques based on continuous time PID algorithms for the following reasons: these are fewer approximations involved, because a discrete time model is fitted directly; information about the full wave-form is used, not only amplitude and frequency; it is easy to include an adjustment of the response rate in the system simply by letting the operator choose ζ and ω; the algorithm can cope with systems having time delays, and also allows adaptive prefiltering.

Extensive simulations have shown that the method works very well for low-order systems with time delay. This can of course be expected. For systems with a large pole excess, the direct approach does not work so well. There are several reasons for this. The output signal is almost sinusoidal which means that only two parameters can be determined. It is, however, possible to arrange the relay experiment so that the steady state gain can also be determined. The model (17) can then be determined also in this case.

These are some key design issues that require further investigation. A major issue is the model complexity required. It is our guess that the simple model used in the examples will be sufficient for many applications. It is a good idea to introduce some observer dynamics. These may be chosen as a function of the noise level. A design based on predictive control may also be considered.

The algorithm is also ideally suited for initialization of adaptive controllers. In this case the initialization is executed under tight feedback conditions. The algorithm gives initial parameter estimates as well as estimates of sampling periods and an estimate of the achievable bandwidth.

The algorithm has been tested on several HVAC plants with good results. Two examples are reported in this paper.

REFERENCES

Åström, K. J. and T. Hägglund (1984). Automatic tuning of simple regulators with specifications on phase and amplitude margins. *Automatica*, **20,** 645–651.

Åström, K. J. and T. Hägglund (1988a). *Automatic Tuning of PID Controllers*. ISA, Research Triangle Park, NC.

Åström, K. J. and T. Hägglund (1988b). A new auto-tuning design. *IFAC Symp. on Adaptive Control of Chemical Processes*, Copenhagen.

Åström, K. J. and T. Hägglund (1990). Practical experiences of adaptive techniques. 1990 *American Control Conf.*, San Diego, California.

Aström, K. J. and B. Wittenmark (1990). *Computer Controlled Systems*, 2nd ed. Prentice-Hall, Englewood Cliffs, NJ.

Bristol, E. H. (1977). Pattern recognition: An alternative to parameter identification in adaptive control. *Automatica*, **13,** 197–202.

Deshpande, P. B. and R. H. Ash (1981). *Elements of Computer Process Control with Advanced Control Applications*. ISA, Research Triangle Park, NC.

Hägglund, T. and K. J. Åström (1991). Industrial adaptive controllers based on frequency response techniques. *Automatica*, **17,** 509–609.

Hang, C. C. and K. J. Åström (1988). Practical aspects of PID auto-tuners based on relay feedback. *IFAC Symp. on Adaptive Control of Chemical Processes*, Copenhagen.

Lennartson, B. (1987). On the choice of controller and sampling period for linear stochastic control. *Preprints 10th IFAC World Congress on Automatic Control*, Munich, Vol. 9, pp. 241–246.

Ljung, L. and Söderström (1983). *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, MA.

McMillan, G. K. (1983): Tuning and control loop performance. ISA Monograph 4, 15A, Research Triangle Park, NC.

Wallenborg, A. (1991). A new self-tuning controller for HVAC systems. *ASHRAE Trans.* **97,** pt 1.

APPENDIX A1: IDENTIFICATION PROCEDURE EXAMPLE

The identification procedure presented in Section 2 is here illustrated for a second-order process with time delay.

Consider the process model

$$Y(z) = \frac{1}{z^d} \frac{b_0 z^2 + b_1 z + b_2}{z^2 + a_1 z + a_2} \tag{A1.1}$$

which corresponds to sampling of a second-order system with a time delay. The model has five parameters. To determine there it is thus necessary to have $n \geq 5$. For $n = 5$ the equation (7) becomes

$$z^d(z^2 + a_1 z + a_2)D(z) - z^d(b_0 z^2 + b_1 z + b_2)E(z) = z^d(z^5 + 1)(q_0 z^2 + q_1 z + q_2). \tag{A1.2}$$

The factor z^d can be cancelled in the equation. For $d = 1$, $D(z)$ becomes

$$D(z) = (y_1 z^5 + y_2 z^4 + y_3 z^3 + y_4 z^2 - y_0 z).$$

For other values of d, $D(z)$ can easily be obtained by cyclic substitutions of y_i in the same way as in Examples 1–3. The polynomial $E(z)$ is the same for all values of d

$$E(z) = -(z^5 + z^4 + z^3 + z^2 + z).$$

The parameter q_2 must be zero, since both $D(z)$ and $E(z)$ lack zero-order terms. The equation (A1.2) can thus be reduced to

$$(z^2 + a_1 z + a_2)D(z) - (b_0 z^2 + b_1 z + b_2)E(z) = (z^5 + 1)(q_0 z^2 + q_1 z). \quad \text{(A1.3)}$$

The equation can be solved for all five values of d. For $d = 1$, the seven unknown parameters are obtained from the following set of equations:

$$\begin{aligned}
b_0 - q_0 &= -y_1 \\
b_0 + b_1 + y_1 a_1 - q_1 &= -y_2 \\
b_0 + b_1 + b_2 + y_2 a_1 + y_1 a_2 &= -y_3 \\
b_0 + b_1 + b_2 + y_3 a_1 + y_2 a_2 &= -y_4 \\
b_0 + b_1 + b_2 + y_4 a_1 + y_3 a_2 &= y_0 \\
b_1 + b_2 - y_0 a_1 + y_4 a_2 - q_0 &= 0 \\
b_2 - y_0 a_2 - q_1 &= 0.
\end{aligned}$$

These equations have the solution

$$\begin{aligned}
a_1 &= [y_0 y_1 - y_0 y_2 + y_1 y_4 - y_2 y_3 + y_3^2 - y_3 y_4]/N \\
a_2 &= [-y_0 y_2 + y_0 y_3 - y_2 y_4 + y_3^2 - y_3 y_4 + y_4^2]/N \\
b_0 &= [y_0(y_0(-y_1 + y_2) - y_1(y_3 + y_4) \\
&\quad + y_2(y_2 + y_3 - y_4) \\
&\quad + y_3(-y_3 + 2y_4)) + y_1(y_1(y_3 - y_4) \\
&\quad - y_2(y_2 - y_3 - y_4) - y_3^2 - y_4^2) \\
&\quad + y_2(2y_3 y_4 - y_4^2) - y_3(y_3^2 - y_3 y_4 + y_4^2) + y_4^3)]/2N \\
b_1 &= [y_0(y_0(y_1 - y_3) - y_1(y_1 - y_2 - y_4) \\
&\quad - y_2(y_3 - 2y_4) - y_3 y_4 - y_4^2) \\
&\quad - y_1(y_1 y_3 - y_2(y_2 + y_3 - 2y_4) - y_3 y_4) \\
&\quad + y_2(y_2(-y_2 + y_3 + y_4) - y_3^2 - y_4^2) \\
&\quad + y_3(-y_3 y_4 + y_4^2) - y_4^3]/2N \\
b_2 &= [y_0(y_0(-y_2 + y_3) + y_1(y_1 - y_2 - y_3) + y_2(y_2 - y_4) \\
&\quad + y_3(y_3 - y_4) + y_4^2) + y_1(y_1 y_4 + y_2(-2y_3 + y_4) \\
&\quad + y_3(y_3 - y_4) - y_4^2) \\
&\quad + y_2(y_2(y_2 - y_3 - y_4) + y_3(y_3 + 2y_4)) - y_3^3]/2N \\
N &= -y_1 y_3 + y_1 y_4 + y_2^2 - y_2 y_3 - y_2 y_4 + y_3^2.
\end{aligned}$$

Since the dead-time cannot be longer than half the period, parameter d takes values from one to five corresponding to the following situations:

$$\begin{aligned}
0 \le L &< \tfrac{1}{12}T_p \\
\tfrac{1}{12}T_p \le L &< \tfrac{1}{6}T_p \\
\tfrac{1}{6}T_p \le L &< \tfrac{1}{4}T_p \\
\tfrac{1}{4}T_p \le L &< \tfrac{1}{3}T_p \\
\tfrac{5}{12}T_p \le L &< \tfrac{1}{2}T_p.
\end{aligned}$$

To reduce the number of cases to be considered, we can proceed iteratively starting with models of lower order.

APPENEIX A2: DESIGN PROCEDURE EXAMPLE

The design procedure presented in Section 3 is here applied to the process models obtained in Examples 2 and 3.

Process model obtained in Example 2

Consider first the system (11) obtained in Example 2, *i.e.*

$$A(z) = z^2(z - a)$$
$$B(z) = b_1 z + b_2.$$

The system is of third order. Choosing the desired closed-loop characteristic polynomial as

$$A_o(z)A_m(z) = z^4(z^2 + p_1 z + p_2)$$

we find that polynomials R_1 and S satisfy the Doophantine equation

$$A(z)(z - 1)R_1(z) + B(z)S(z) = z^4(z^2 + p_1 z + p_2).$$

The minimal degree solution becomes

$$R_1(z) = z^2 + r_1 z + r_2$$
$$S(z) = s_0 z^3 + s_1 z^2$$

where

$$\begin{aligned}
s_0 &= \frac{1}{1-a}\left(\frac{A_m(1)}{B(1)} - a^2\frac{A_m(a)}{B(a)}\right) \\
s_1 &= \frac{a}{a-1}\left(\frac{A_m(1)}{B(1)} - a\frac{A_m(a)}{B(a)}\right) \\
r_2 &= -\frac{b_2}{a}s_1 \\
r_1 &= \frac{1+a}{a}r_2 - \frac{b_1}{a}s_1 - \frac{b_2}{a}s_0.
\end{aligned}$$

The control law is then given by

$$u(t) = t_0 y_{sp} - s_0 y(t) - s_1 y(t-h) + (1 - r_1)u(t-h) + (r_1 - r_2)u(t-2h) + r_2 u(t-3h)$$

where

$$t_0 = s_0 + s_1$$

to obtain the correct steady state.

Process model obtained in Example 3

Consider the fourth-order system (13) obtained in Example 3, *i.e.*

$$A(z) = z^3(z - a)$$
$$B(z) = b_1 z + b_2.$$

Choosing the following closed-loop characteristic polynomial:

$$A_o(z)A_m(z) = z^6(z^2 + p_1 z + p_2)$$

R_1 and S satisfy the following Diophantine equation:

$$A(z)(z - 1)R_1(z) + B(z)S(z) = z^6(z^2 + p_1 z + p_2).$$

The minimal degree solution is given by

$$R_1(z) = z^3 + r_1 z^2 + r_2 z + r_3$$
$$S(z) = s_0 z^4 + s_1 z^3$$

where

$$\begin{aligned}
s_0 &= \frac{1}{1-a}\left(\frac{A_m(1)}{B(1)} - a^3\frac{A_m(a)}{B(a)}\right) \\
s_1 &= \frac{a}{a-1}\left(\frac{A_m(1)}{B(1)} - a^2\frac{A_m(a)}{B(a)}\right) \\
r_3 &= -\frac{b_2}{a}s_1 \\
r_2 &= \frac{1+a}{a}r_3 - \frac{b_1}{a}s_1 - \frac{b_2}{a}s_0 \\
r_1 &= \frac{(1+a)}{a}r_2 - \frac{1}{a}r_3 - \frac{b_1}{a}s_0.
\end{aligned}$$

The control law is then given by

$$u(t) = t_0 y_{sp} - s_0 y(t) - s_1 y(t-h) + (1 - r_1)u(t-h) + (r_1 - r_2)u(t-2h) + (r_2 - r_3)u(t-3h) + r_3 u(t-4h)$$

where

$$t_0 = s_0 + s_1$$

to obtain the correct steady state.