Algorithm Solutions in Model-free Controller Autotuning Linked with Evaluation of Frequency

BOHUMIL ŠULC
Department of Instrumentation and Control Engineering
Czech Technical University in Prague, Faculty of Mechanical Engineering
Technická 4, Prague 6 - Dejvice
CZECH REPUBLIC
bohumil.sulc@fs.cvut.cz

Abstract: Controllers, especially PIDs, are extensively used in industrial practice, and there is a continual search for methods that facilitate controller parameter setting, with low requirements on the operator’s knowledge. Ideally, they should be fully automatically, and should require no human participation. One of the methods presented recently is based on an experimentally performed evaluation of excited frequency responses with the aim of achieving recommended values of one or more control quality indicators known from the course of the Nyquist plot. Unlike their linear origin, the indicators can be obtained in control loops involving nonlinearities even in the controller. In this sense, the method has a philosophy similar to that of the popular Ziegler and Nichols method. No mathematical model and theory is required for its implementation. All processing is carried out by a program added to the control algorithm. No additional instrumentation is necessary. This paper presents considerations when dealing with real plants with strong nonlinearity in their behaviour in conjunction with improvements in indicator evaluation.

Key-Words: Model-free tuning controller, autotuning, frequency response indicators, gain margin, phase margin, PID controller, laboratory model

1 Introduction

Out of dozens of methods that have been developed, mostly on the basis of linear theory, only a few have had the opportunity to be widely applied in control practice. Most of the methods require models of the controlled processes, and this often limits their applicability. A routine design of a control circuit is a process that runs according to strict economic rules. It is therefore important to make a proper choice of instrumentation (technical usability of the sensor type, choice of ranges and accuracy, actuator characteristics, etc.). Except in some special applications, the controversial benefit of an optimal controller setting is not the main design goal. In some previous publications [2], we have discussed what standard industrial practice has accepted from the broad range of the methods for optimising the function of control loops. The conclusions have not been very optimistic. With some exceptions, the hidden preference of industrial practice has been to reject — theory, models and expensive experts. Pre-setting the controller and adjusting it on the spot, if necessary, is a popular procedure in intuitive working process based on empiric knowledge and experience. For this reason the Ziegler and Nichols method, and similar methods, have retained their popularity, and any new procedure using a similar philosophy [1] is much better accepted than the other approaches.

A new modification of the methods has been proposed, which utilizes the oscillations excited in the control loop for setting controller. Unlike the Ziegler and Nichols method, where auto-oscillation is not invoked by reaching a critical setting of the controller, or the Aström Relay method, where a relay is inserted, the new modification uses software that adds a harmonic signal to the control error signal processed by the controller. The decision when to start the procedure is fully in the hands of the operator. He or she decides whether to accept changes in the controller setting. The method can be used just as a check on whether controller retuning is desirable, because the change in the indicators shows that the dynamics of the controlled plant has changed excessively, and the values of the indicators differ from those recommended by an expert.

Another major disadvantage of the relay method is that the control function is interrupted, while the critical parameter identifying process is being carried out. During this operation, the controller must be disconnected and reconnected without a bump, and a steady state must already have been achieved. The
amplitude of the oscillation added to the controlled variable can be influenced by the parameters of the relay, but it is difficult to forecast its size in advance. Excitation of the oscillation often requires changes from the manipulated variable that are easy to simulate but difficult to execute technically.

2 Assessment of control performance
While the technically simple implementation is a great advantage of the proposed method, there are several problems in evaluating the signal that must be solved within the programme performing the algorithm. For example, it is necessary automatically to recognize the oscillating steady state in which an evaluation of the magnitude and phase shift can be started. Secondly, the amplitude of the inserted signal must be carefully selected so that it does not considerably disturb the controlled variable, though it must be distinguishable from the noise. To achieve desired values of the indicators, it is necessary to combine the changes in the frequency of the exciting signal and in the controller parameter setting. If more than one indicators are monitored, it is not easy to find a strategy for performing these changes simultaneously. We therefore worked on an indicator based on evaluating maximum sensitivity, which is a compromise between the optimum characterized by the phase margin and the gain margin.

The classical indicators defined by specific points on the Nyquist plot depicted in Fig. 1 are:

2.1. Gain margin
The gain margin, denoted $m_A$, is the factor which, when multiplies the amplitude of the point on the Nyquist plot with the phase angle $-\pi$, causes that the plot passes the critical point $-1+0.j$. It expresses how safe the control loop is against the stability loss. Recommended values of the gain margin ranges between 2 and 2.5.

2.2. Phase margin
The phase margin, denoted $\gamma$, expresses the amount of phase shift that can be tolerated before the control loop becomes unstable. It is defined through the angle $\gamma$ given in degrees (see Fig. 1) appertaining to frequency $\omega_\gamma$. It is sometimes called the gain crossover frequency, because this is the frequency at which the loop gain is one (the Nyquist plot passes the unit circle). Recommended values of the optimal phase margin are quoted in the range from 30° to 60°, but our own experience indicates that a higher upper limit usually suits better.

2.3. Maximum sensitivity
It follows from the block scheme in Fig. 2 that the open loop transfer function $G_o(s)$ (the product of the controller transfer function $G_c(s)$ and the controlled plant transfer $G_p(s)$) allows us to define the transfer function of the disturbance

$$G_d(s) = \frac{1}{1 + G_o(s)}$$

from which it follows that a (load) disturbance is transferred with maximum sensitivity $M_s$ if in the sensitivity function $S(j\omega) = G_d(j\omega)$ the absolute value of the denominator $|1 + G_o(j\omega)|$ achieves its minimum. The reciprocal value $1/M_s$ is equal to the length of the vector sum $-1+j.0$ and $|G_o(j\omega)|$

$$M_s = \max_{\omega \in \mathbb{R}} \left| \frac{1}{1 + G_o(j\omega)} \right| = \max_{\omega \in \mathbb{R}} |S(j\omega)|$$

and it corresponds to the radius of a circle with its centre in the critical point touching the Nyquist plot.

The maximum sensitivity value is connected with other indicators by the following relationships

Fig. 1 Definition of some indicators of optimal controller setting using open loop frequency response (Nyquist plot)

Fig. 2 Block scheme of the transfer function model representing a control circuit
The mutual connection of indicators can be also demonstrated by Fig. 3

\[ m_S \geq \frac{M_S}{M_S - 1} \gamma \geq 2 \arcsin \left( \frac{1}{2M_S} \right) > \frac{1}{M_S} \] (3)

The dependence of radius \( r \) on angle \( \gamma \) has a minimum corresponding to the situation when the circle is touching the Nyquist plot, i.e. maximum sensitivity \( M_s \) is achieved. If this extreme falls into the range of recommended values, this indicates optimal setting of the controller. Another indicator can be checked simultaneously – the phase margin where the angle of the phase margin is given by the crossing of the two curves depicted in Fig. 3.

3 Closed loop evaluation of excited frequency responses

The basic principle followed in the proposed model free technique for setting controllers assumes that the function of the tuned control circuit will not be interrupted during the setting procedure. This requirement implies that the evaluating frequency properties are performed in a closed control circuit though most of the optimal control indicators are based on the Nyquist plot, i.e. on the open loop frequency behaviour. In comparison to the frequency response assessment in an open loop, we can make use of information about the control error while making the evaluation in a closed loop. This provides several advantages. One of the advantages can easily confirmed, when the control circuit is modelled by a linear system. Normally, an output bias occurs in open loop frequency responses due to the expected presence of integral action either in the controller or in the behaviour of the controlled plant. Using the final value theorem, the size of the bias can be computed for unit sinusoidal excitation

\[ \lim_{t \to \infty} y(t) = \lim_{s \to 0} sG_k(s)G_p(s) \frac{\omega}{s^2 + \omega^2} = \lim_{s \to 0} sM(s) \frac{\omega}{s^2 + \omega^2} = \frac{k_o}{\omega} \] (4)

Evidently, the bias is not present in the settled control error courses of the closed loops, because

\[ \lim_{t \to \infty} e(t) = \lim_{s \to 0} \frac{1}{1 + G_o(s)} E_{os}(s) = \lim_{s \to 0} \frac{sM(s)}{sM(s) + N(s)} \frac{\omega}{s^2 + \omega^2} = 0 \] (5)

This is confirmed by the responses depicted in Fig. 4.

Fig. 3 Dependence \(|1 + G_o(j\omega)|\) on the angle \(\pi + \text{arg}(G_o(j\omega))\) and plot of the function \(\sqrt{2(1 - \cos \gamma)}\)

Fig. 4 Oscillations excited in a closed and open control loop modeled by a linear system

Evaluating the magnitude \(|G_o(j\omega)|\), and the phase shift \(\text{arg}(G_o(j\omega))\) from closed responses is not only easier, but also the only executable way of evaluation in case this is done in a control circuit containing strong nonlinearity, e.g. in the controlled plant behaviour. Such a situation occurs quite frequently if the controller output is converted into changes in mass or energy supply by a manipulating device - actor with nonlinear features. This nonlinearity may cause loss of symmetry in the sinusoidally excited signals. This exerts the nonlinear dependence of the controller output offset caused by the integral term in the control algorithm, and what is even worse – non-symmetric shape deformation of the open loop output oscillations. Evaluation algorithms must take this circumstance into account, though it is not apparent in the case of standard process nonlinearity. An attempt has been made to demonstrate this by introducing an unrealistic cubic steady state characteristic in Fig. 5.
The influence of such nonlinearity on an investigation of oscillations in a closed loop is reduced by the feedback, if we perform the evaluation from the control error signal as described below. However, if the influence of a nonlinear actuator is obvious, especially observing its input and output (Fig. 6) in comparison with the same signals in Fig. 5, the question is how we can draw conclusions concerning indicators defined by means of open loop oscillations based on oscillations measured in a closed control loop. In the linear case, the answer to the question is quite simple, because the transfer functions (see Fig. 2) can be used in expressing the Laplace transform

$$E(s) = \frac{-G_o(s)}{1 + G_o(s)} E_{ex}(s)$$

where

$$G_o(s) = \frac{E(s)}{E(s) + E_{ex}(s)} = \frac{E(s)}{E(s)}$$

For a frequency interpretation of (7)

$$|G_o(j\omega)|e^{j\arg(G_o(j\omega))} = \frac{E(j\omega)}{E_o(j\omega)} = \frac{e_0}{e_{R_0}} e^{j(\omega + \phi_0 + \phi_0)}$$

the magnitude $|G_o(j\omega)|$ is determined by the ratio of the amplitudes

$$|G_o(j\omega)| = \frac{e_0}{e_{R_0}}$$

and the phase shift is equal to the argument of $G_o(j\omega)$

$$\arg G_o(j\omega) = \phi_0 - \phi_R + \pi$$

In the nonlinear case we would need to operate with “steady-state transfer functions”, but basically the principle of the evaluation is retained. Dependence on the amplitude size of the exciting signal must be respected.

### 4 Models for testing
For testing all algorithms it is advantageous to have a simulation model where all necessary experiments can

![Fig. 8 Scheme of a three tank cascade laboratory set-up](image-url)
be processed much quickly and more flexibly than on a physical laboratory set-up. However, no simulation model can exactly replace reality. Thus it is good to have available such a laboratory device whose mathematical model can be derived on the basis of sufficiently exact valid physical laws with the possibility to get needed coefficients via measurement. Most of these requirements are fulfilled by the physical model of a three tank cascade a functional scheme of which is depicted in Fig. 8.

In the derivation of the mathematical model, a standard flow rates balancing procedure can be used. Such a deductive identification leads to equations of a nonlinear dynamical model of the cascade

\[
\frac{d}{dt} h_i(t) = \frac{1}{A} q_i(t) - \frac{K_{V_{12}}}{A} \sqrt{h_i(t) - h_2(t)}
\]

\[
\frac{d}{dt} h_2(t) = \frac{K_{V_{12}}}{A} \sqrt{h_1(t) - h_2(t)} - \frac{K_{V_{20}}}{A} \sqrt{h_2(t)}
\]

\[
\frac{d}{dt} h_3(t) = \frac{1}{A} q_3(t) - \frac{K_{V_{32}}}{A} \sqrt{h_3(t) - h_2(t)}
\]

(11)

where \( A \) is the cross section (the same for all three tanks) in dm\(^2\), \( K_{V_{xy}} \) stands for the flow rate coefficient used in the equations quantifying the (volumic) flow rate through the valve from tank \( x \) to tank \( y \), \( h_x \) is the water level in tank \( x \) \( (x = 1, 2, 3) \), \( q_x \) \( (x = 1, 3) \) denotes the supply flow rate (Fig. 8).

In model (11) it is necessary to identify the flow coefficients \( K_{V_{xy}} \). For this, it was necessary to measure the values of the flow rates \( q_1 \) and \( q_3 \) and the corresponding heights of the levels in several steady states. The \( K_{V_{xy}} \) values obtained for each steady state according to the formulas

\[
K_{V_{12}} = \frac{q_{10}}{\sqrt{h_{10} - h_{20}}}
\]

\[
K_{V_{20}} = \frac{q_{10} + q_{30}}{\sqrt{h_{20}}}
\]

\[
K_{V_{32}} = \frac{q_{10} + q_{30}}{\sqrt{h_{20}}}
\]

5 Identification of indicators and controller tuning

To evaluate the magnitude \( |G_c(j\omega)| \) and phase shift \( \arg G_c(j\omega) \) from a frequency response an phase-locked loop (PLL) identifier module was proposed several years ago, and its use was applied in a linear concept of the control circuit representation \([7],[10] \) et al. The idea of applying this module for PID autotuning is expressed by a block scheme in Fig. 9.

In the discrete version when the evaluating of responses can be based on values known only in discrete time instants, there can be problems in finding extreme values. For this reason, a direct frequency response assessment seemed to function better. The algorithm for generating new values of excitation frequency and controller parameter changes using a variable gain in adaptation loops is quicker than the use of loops with pure integration.

The evaluation of the gain and phase shift can be evaluated in a steady state, which starts to be tested after a time interval of length \( T_p \) to override the control circuit dynamics. After this time, samples of periodic signals can be saved into data vectors length of which we can define. In each step of the discrete time these values are actualized by the values the new step. Simultaneously, the new values are tested, whether they can be candidates for those values, in the immediate neighbourhood of which an extreme or zero value can be found. If yes, the value is determined by means of interpolation and used for computing the magnitude and phase shift. Consequently, the value of specified indicator is computed. Ratio of the new and desired value of the indicator is used as a quotient in an iteration process of controller parameter tuning.

6 Conclusion

This paper presents solutions of some problems concerning algorithms used in evaluating of indicators by means of which an optimal controller setting can be found automatically. It focuses on the natural
occurrence of nonlinearities in real control circuits and proves the applicability of the new mechanism for optimal controller parameter tuning. It is a great advantage that methods using responses to a periodic signal and additionally exciting real control circuits for optimal control quality assessment do not need any mathematical models. They can be used for PID controller autotuning in which the achieved optimum is considered from the global viewpoint and not only from the course of the response. An advantage of the maximum sensitivity indicator in the search for an optimal controller setting is that it can serve as an all-in-one criterion combining several aspects bound to the other indicators. In such a way, several indicators can easily be reflected simultaneously in a strategy for controller parameter changes.

Acknowledgement:
This research has been supported by Czech Science Foundation Grant No. GAČR 101/07/1667.

References