CONTRIBUTION TO THE THEORETICAL ANALYSIS OF THE ZIEGLER–NICHOLS METHOD

Mikuláš Huba — Katarína Žáková *

The high practical significance of the \( I_1T_d \)-plant approximations has been shown by the perhaps most frequently used method for controllers tuning by Ziegler and Nichols, modified later for sampled data systems by Takahasi et al. This paper extends previous works by analysing also control structures that include \( I_1T_1 \) and more general \( I_1T_1T_d \)-system. All these approximations are used for tuning of the \( P \) and \( PI \) controllers designed for single integrator system with constrained input.

**Key words:** \( P \) (PD) controllers, controller tuning, Ziegler-Nichols method

1 INTRODUCTION

One of the most widespread tuning rules was introduced by [13] already at the beginning of 40’s in the last century. They have presented two different methods of the controller setting. While in the first one, the controller parameters are determined on the basis of the transient response measurement, in the “ultimate sensitivity” method it is necessary to bring the system to the stability boundary by raising the gain of a \( P \) controller. Ziegler and Nichols have designed controllers to guarantee the “quarter amplitude ratio” (quarter amplitude damping). This criterion was also followed by eg [9] and [2]. Ziegler-Nichols method creates a starting point for numbers of other tuning rules.

2 SETTING OF THE PROPORTIONAL ACTION

For the sake of simplicity, Ziegler and Nichols have not analyzed the types of time lag involved in the control loop. In order to show the possible consequences, the behavior of an integrator system combined with two basic parasitic delays: the dead time and the time constant delay will be analyzed. A similar approach was used previously by [8], [4] or [12]. The derived gain can be used for setting of \( P \) controller in the structures with constrained control signal (Fig. 1), too. The blocks “Plant” and “Delay” together represent any process that can be approximated by \( I_1T_d \) model. The dominant 1st order dynamics presented by a single integrator is supposed to be deployed in the feedforward path. The transport delay can be present at any place within the feedback.

2.1 \( I_1T_d \) system

2.1.1 Optimal setting of \( P \) controller

Let us consider the integral + dead time system

\[
F(s) = \frac{K_S e^{-Ts_1}}{s}
\]

that is going to be controlled by an analog \( P \) controller. The suggested design of the optimal controller setting that corresponds to the multiple dominant pole follows a requirement of monotonic transient response at the presence of possible maximal value of controller gain. Therefore, its setting can be found by looking for a controller gain extreme. It means the optimal gain of the \( P \) controller \( K_C \) has to satisfy the condition

\[
\frac{dK_C}{ds} = 0
\]

 whereby \( K_C \) can be expressed from the characteristic equation of the system closed loop:

\[
K_C K_S e^{-Ts_1} + s = 0.
\]

Solving (2) and (3) leads to the optimal controller gain

\[
K_{Copt} = \frac{1}{eK_S T_d}
\]

and dominant real pole

\[
s_{opt} = -1/T_d.
\]

These formulae are known for long time — see eg [8], [4] or [12]. The derived gain can be used for setting of \( P \) controller in the structures with constrained control signal (Fig. 1), too. The blocks “Plant” and “Delay” together represent any process that can be approximated by \( I_1T_d \) model. The dominant 1st order dynamics presented by a single integrator is supposed to be deployed in the feedforward path. The transport delay can be present at any place within the feedback.

2.1.2 Critical controller gain

For comparison, in addition to the optimal gain of the \( P \) controller there was analytically derived also a gain at the stability margin. For this one a term “critical controller gain” was introduced. To derive the critical closed loop gain we can substitute

\[
s = j\omega
\]

* Department of Automation and Control, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovakia Email: [huba, katarina.zakova]@elf.stuba.sk

ISSN 1335-3632 © 2003 FEI STU
into (3). In this way, there arises a set of two equations
\[ \begin{align*}
KC K_S \cos(T_d \omega) &= 0 \\
KC K_S \sin(T_d \omega) &= \omega
\end{align*} \]  \hspace{1cm} (7)
whose solution leads to the critical controller gain. A trivial solution of (7) \( \omega = 0 \) leads to the minimum critical value
\[ K_{Ccr1} = 0. \] \hspace{1cm} (8)
The second solution gives the maximum critical value of the stability area
\[ K_{Ccr2} = \frac{\pi}{2K_S T_d}. \] \hspace{1cm} (9)
The corresponding period of oscillation is
\[ P_u = \frac{2\pi}{\omega} = 4T_d. \] \hspace{1cm} (10)

2.2 \( I_1T_1 \) system

2.2.1 Optimal controller gain

In the case of \( I_1T_1 \)-system
\[ F(s) = \frac{K_S}{s(T_1s + 1)} \] \hspace{1cm} (11)
whose dynamics is deployed in the control structure according to Fig. 2, the optimal controller gain can be derived from the requirement of double real pole of the closed loop characteristic equation
\[ T_1s^2 + s + KC K_S = 0. \] \hspace{1cm} (12)
This requirement corresponds to
\[ K_{Copt} = \frac{1}{4K_S T_1} \] \hspace{1cm} (13)
whereby
\[ s_{opt} = -\frac{1}{2T_1}. \] \hspace{1cm} (14)
The optimal controller gain (13) can also be applied in the structure shown in Fig. 1. Here, the block “Plant” represents the dominant single integrator dynamics and the block “Delay” represents the identified time constant \( T_1 \).

2.2.2 Critical controller gain

The critical controller gain for this system is going to infinity, since all coefficients of the characteristic polynomial are positive and according to Routh-Schur test of stability it means that the system is stable.
\[ K_{Ccr} = \infty. \] \hspace{1cm} (15)

2.3 \( I_1T_1T_d \) system

2.3.1 Optimal controller gain

The last integral model that will be considered is an integral system with both dead time and time constant
\[ F(s) = \frac{K_S}{s(T_1s + 1)} e^{-T_d s}. \] \hspace{1cm} (16)
The optimal control gain can be determined from the characteristic equation
\[ s(T_1s + 1) + KC K_S e^{-T_d s} = 0. \] \hspace{1cm} (17)
After taking a derivative of (17) according to the complex variable \( s \)
\[ 2T_1s + 1 - KC K_S T_d e^{-T_d s} = 0 \] \hspace{1cm} (18)
and expressing
\[ KC K_S T_d e^{-T_d s} = -(T_1 s^2 + s) \] \hspace{1cm} (19)
we get the double real pole value by means of a simple substitution (19) into (18):
\[ s_{1,2} = -\frac{1}{2} \frac{T_d + 2T_1 \pm \sqrt{T_d^2 + 4T_1^2}}{T_1 T_d}. \] \hspace{1cm} (20)
Substituting the value of \( s_1 \) corresponding to the sign “+” into (19) we get the optimal gain of the proportional controller
\[ K_{Copt} = \frac{-2T_1 + \sqrt{T_d^2 + 4T_1^2}}{K_S T_d^2 e^{2T_d + 4T_1^2} - 2T_1}. \] \hspace{1cm} (21)
After denoting the total time lag
\[ L = T_d + T_1 \] \hspace{1cm} (22)
2.3.2 Critical controller gain

The derivation of the critical closed loop gain comes out from the characteristic equation (17). Substitution $s = j\omega$ leads to the pair of equations

$$K_C K_S \cos (\omega T_d) = \omega^2 T_1$$
$$K_C K_S \sin (\omega T_d) = \omega$$

Unfortunately, it is not possible to compute the value of critical gain analytically and therefore only a graphical representation of $K_{CCr}$ corresponding to $\varepsilon = 1$ is introduced in Fig. 3. This is close to that corresponding to $\varepsilon \to \infty$.

3 SETTING OF THE PROPORTIONAL AND INTEGRAL ACTIONS

Up to now, only the design of the $P$ controller for the single integrator has been considered. The additional time lags were compensated by decreasing the optimal controller gain. Due to this, an increase of time delays causes an increase of the possible permanent control error. One possibility is to compensate the permanent control error by the feedforward control. The other possibility is to use a windupless integral action. The basic control loop (1st order system + delay + $P$ controller) can be extended by blocks that serve for reconstruction and compensation of an input system disturbance $v_I$. It creates the integral action of the controller.

The new control structure (see also [5], [7] and [6]) can be seen in Fig. 4. The included filter serves for a reduction of possible noise in the structure. For simplicity, it is connected with reconstruction blocks. In the case when only a linear part of saturation is considered, the whole control structure can be modified to the form that is presented in Fig. 5.

It is evident that the control corresponds to $PI$ controller with the transfer function

$$C(s) = K_C \left( 1 + \frac{1}{T s} \right).$$

Fig. 3. Optimal and critical $P$ controller gain. Solid line: $I_1 T_1 T_d$ system with $\varepsilon = 1$; dashed line: $I_1 T_d$ system ($\varepsilon \to \infty$); dotted line: $I_1 T_1$ system ($\varepsilon = 0$).

Fig. 4. Reconstruction and compensation of the input system disturbance (windupless $PI$ controller).

and the ratio

$$\varepsilon = \frac{T_d}{T_1}$$

there follow

$$T_d = \frac{L \varepsilon}{\varepsilon + 1}; \quad T_1 = \frac{L}{\varepsilon + 1}.$$  \hspace{1cm} (23)

Then

$$K_{Copt} = \frac{-2 + \sqrt{\varepsilon^2 + 4}}{K_S \varepsilon^2 L} \frac{\varepsilon + 1}{\varepsilon^{2+\sqrt{\varepsilon^2+4}}}.$$  \hspace{1cm} (24)

The graphical dependence of the normed gain $K_{opt} = K_{Copt} K_S$ on the total lag $L$ for $\varepsilon = 1$ is presented in Fig. 3. It is interesting to note that this optimal value lies between the values corresponding to the limit situations with $\varepsilon = 0$ and $\varepsilon \to \infty$. This conclusion regarding $I_1 T_1 T_d$ system can be generalized also for other values $\varepsilon \in (0, \infty)$.

The dependence of the optimal and critical gain on the ratio $T_d/T_1$ can be found e.g. in [8].
where the achieved solutions are complex numbers and
with the windupless PI controller according to Fig. 4 (\(K_S = 3\),
\(T_d = 1.5\))

### 3.1 Structures with transport delay

The controller design follows the control structure in
Fig. 4 where the block of delay represents the transport
delay in the loop. ARISING FROM THE CHARACTERISTIC POLYNOMIAL OF THE CLOSED LOOP

\[ chp(s) = T_f s^2 + (1 + K_S K_C T_d) s e^{-T_d s} + K_S K_{C_{re}} e^{-T_d s} \]  \tag{26} 

the setting of controller and filter can be determined from
the requirement of its dominant triple real pole. it can be
achieved by solving a following set of equations

\[ chp(s) = 0; \quad \frac{d}{ds} chp(s) = 0; \quad \frac{d^2}{ds^2} chp(s) = 0. \]  \tag{27} 

Hence

\[ s_{opt} = -2 + \frac{\sqrt{2}}{T_d}, \]  \tag{28} 

\[ K_{C_{opt}} = \frac{\sqrt{2} - 1 + I(5\sqrt{2} - 7)\tau}{K_S T_d} e^{-2 + \sqrt{2}}, \]  \tag{29} 

where

\[ \tau = \sqrt{(14 + 10\sqrt{2})e^{-2 + \sqrt{2}} - 12\sqrt{2} - 17}. \]

The time constant of the filter is

\[ T_{f_{opt}} = \frac{3 + 2\sqrt{2} + I\tau}{2} T_d. \]  \tag{30} 

However, the achieved solutions are complex numbers and
the controller and the filter as well are supposed to work
with real parameters. Experimentally, it can be shown
that the loop dynamics that corresponds to parameters,
which are derived by approximating the computed com-
plex values by the real part or by the module, is close to
the required one. Then

\[ T_{f_{re}} = \frac{2\sqrt{2} + 3}{2} T_d \approx 2.914T_d, \]  \tag{31} 

\[ T_{f_{mod}} = T_d \frac{\sqrt{10\sqrt{2} + 14}}{2} e^{-2 + \sqrt{2}} \approx 3.555T_d. \]  \tag{32} 

For comparison, the controller gain of the P controller
was \(K_{C_{opt}} \approx 0.367/K_S T_d\) (4), \(\approx\) approximately 1.6
times bigger that the gain \(K_{C_{re}}\) of the PI controller.

In Fig. 6 it is possible to compare transient responses
with the corresponding control signals for both types of
approximations. Simulations are done for \(I_1 T_d\) system
controlled by PI controller according to the structure in
Fig. 4. It can be seen that there is not a big difference
between both approximations and therefore the simpler
approximation by the real part will be used.

The correctness of the chosen approximation is also
observable from Fig. 7. Let us follow the approximation
by the real part. Firstly, the simulation was accomplished
with the controller parameters that were set according to
formulae given by (31) and (32). Then, the simulation
was compared with simulations where controller param-
eters were increased or decreased by 20%. The quality of
these modified transient responses and the corresponding
control signals is slightly going down.

Finally, it is to note that resulted formulae (31) and (32)
remind of formulae achieved by Ziegler and Nichols. Their
form is the same, they slightly differ only in constants.
3.2 Structures with time constant delay

The controller design follows the control structure in Fig. 4 where the block of delay represents the time constant delay in the loop. The optimal setting of the controller can be found from the closed loop characteristic polynomial

\[ chp(s) = T_1 T_f s^3 + T_f s^2 + [(1 + K_C K_S T_f) + K_C K_S] e^{-T_d s} \]

Following (27) one gets the optimal controller setting

\[
\begin{align*}
K_{Copt} &= \frac{3 \pm i \sqrt{3}}{18 K_S T_1} \\
T_f opt &= \frac{3T_1}{2}(3 \pm i \sqrt{3})
\end{align*}
\]

that correspond to the dominant triple pole

\[ s_{opt} = -\frac{1}{3T_1} \]

(34)

After approximation of the complex solution by the real part or by the module one gets the following PI controller settings

\[
\begin{align*}
K_{C re} &= \frac{1}{6K_S T_1} = 0.1667 \\
K_{C mod} &= \frac{\sqrt{3}}{9K_S T_1} = 0.192 \\
K_{T re} &= \frac{9}{2} T_1 = 4.5 T_1 \\
K_{T mod} &= 3 \sqrt{3} T_1 = 5.196 T_1
\end{align*}
\]

(35)

and

\[
\begin{align*}
T_f re &= 9 T_1 = 4.5 T_1 \\
T_f mod &= 3 \sqrt{3} T_1 = 5.196 T_1
\end{align*}
\]

(35)

For comparison the \( P \) controller gain for the \( I_1 T_d \) system was \( K_{Copt} = 0.25/K_S T_1 \).

The graphical presentation of the achieved solution can be found in Fig. 9.

3.3 Structures with transport and time constant delay

Following the control structure (Fig. 4) with a transport and a time constant delay (\( I_1 T_1 T_d \) system) one gets the characteristic polynomial

\[ chp(s) = T_1 T_f s^3 + T_f s^2 + [(1 + K_C K_S T_f) + K_C K_S] e^{-T_d s} \]

(37)

As it was mentioned already before, the setting of controller and filter can be determined from the requirement of its dominant triple pole that can be found from equations (27). Unfortunately the analytical solution is not available. The controller parameters can be computed only numerically as depending on the total lag \( L \) (21) and the ratio \( \varepsilon \) (22). For example for \( \varepsilon = 0.5 \) one gets \( K_{C opt} = 0.193/K_S L \) and \( T_f opt = 3.717 L \) and for \( \varepsilon = 2 \) it can be found that \( K_{C opt} = 0.219/K_S L \) and \( T_f opt = 3.127 L \). The controller parameters for \( \varepsilon = 1 \) are illustrated in Fig. 9 and there it holds \( K_{C opt} = 0.207/K_S L \) and \( T_f opt = 3.386 L \).

4 EXAMPLE

Let us consider the system described by the transfer function

\[ F(s) = \frac{1}{(s+1)^3} \]

(38)

Firstly, the controller design is based on the \( I_1 T_d \), \( I_1 T_1 \) and \( I_1 T_1 T_d \) approximations. Experiments show that such integrator based approximation model should concentrate mainly on the initial part of the step response [14]. Following this requirement one can receive

<table>
<thead>
<tr>
<th>model</th>
<th>transfer function</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I_1 T_d )</td>
<td>( \frac{0.257}{s} e^{-0.745 s} )</td>
</tr>
<tr>
<td>( I_1 T_1 )</td>
<td>( \frac{0.429}{s(1.95 s + 1)} )</td>
</tr>
<tr>
<td>( I_1 T_1 T_d )</td>
<td>( \frac{0.32}{s(0.9 s + 1)} e^{-0.215 s} )</td>
</tr>
</tbody>
</table>

The method described in [11], [3] and [13] arising from \( I_1 T_d \) approximation are compared with the proposed \( PI \) controller. All controller parameters can be found below

<table>
<thead>
<tr>
<th>Controller parameters</th>
<th>( K_C T_f )</th>
<th>( (T_f) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I_1 T_d )</td>
<td>1.2065</td>
<td>2.1709</td>
</tr>
<tr>
<td>( I_1 T_1 )</td>
<td>0.1993</td>
<td>8.7750</td>
</tr>
<tr>
<td>( I_1 T_1 T_d )</td>
<td>0.5093</td>
<td>4.4964</td>
</tr>
<tr>
<td>Tyreus-Luyben</td>
<td>2.5435</td>
<td>6.5187</td>
</tr>
<tr>
<td>Frechauf et al.</td>
<td>2.6114</td>
<td>3.7250</td>
</tr>
<tr>
<td>Ziegler-Nichols</td>
<td>4.7006</td>
<td>2.4809</td>
</tr>
</tbody>
</table>

The control signal was limited to \( U_{min} = -1.25 \) and \( U_{max} = 1.25 \) and at \( t = 30 \) there was considered an input disturbance with the value \( v_i = 0.5 \).

The simulation results are presented in Fig. 8. The proposed \( PI \) controller based on \( I_1 T_d \) approximation was able to reach the required value without an overshoot. The non presented simulations following \( I_1 T_1 \) and \( I_1 T_1 T_d \) approximations had a too long settling time whereby the settling time of transient response based on \( I_1 T_1 \)
Following the control structure with the 1st order dynamics, delays in the feedback. On the basis of such distribution of the dynamical elements, $P$ controller following the dominant pole design was derived. The design procedure also takes into account always present limitation in the control signal. The disturbance reconstruction and compensation is solved by introducing an $I$ action to the control structure. It leads to the establishment of $PI$ controller.

In contrast to the “quarter amplitude damping” introduced by Ziegler and Nichols, the controller setting corresponds to the double (triple) real dominant pole and guarantees the fastest possible transients without overshoot.

The comparison of the optimal $P$ controller gains (4) and (13) with that one given by the Ziegler-Nichols “transient response method” shows that the approximations by the integral+dead time system or integral+time constant system yield the gains $K_{Copt} = 0.368/K_S L$ ($L = T_d$) and $K_{Copt} = 0.25/K_S L$ ($L = T_1$) which are 2.71 or 4 times smaller than the values recommended by Ziegler and Nichols. This is caused by the different performance criteria: The double real pole corresponds to transients without overshoot, while Ziegler and Nichols tried to set the gain as high as possible in order to reduce the offset and influences of possible load changes. However, this also creates a closed loop system that is very poorly damped and has very poor stability margins. Then, as also mentioned in [1], the closed loop gain is typically 2–3 times higher. It is to note that analytical solutions for optimal and critical gain of $P$ controller were discussed for $I_1T_1$ and $I_1T_d$ model already by Oldenbourg and Sartorius, [8].

For the $PI$ controller it is possible to compare parameter $K_i = 1/T_i$ that for the approximation by $I_1T_d$ model with the dead time $L = T_d$ gives

$$K_{Iopt} = 0.343/L.$$  \hspace{0.5cm} (39)

For $I_1T_1$ model where $L = T_1$ one gets

$$K_{Iopt} = 0.222/L.$$ \hspace{0.5cm} (40)

The above computed parameters are very close to the value determined by Ziegler and Nichols that is $K_i = 0.3/L$. As far as the setting of the proportional gain of the $PI$ controller is considered, Ziegler and Nichols proposed to decrease its value in comparison to the “pure” $P$ controller. This was due to the increased inclination of the loop to oscillations in the presence of the $I$ part of the controller. The tendency of the proportional parameter decrease can also be seen at the presented controller design.

Following the previous analysis, to accept only one type of the dominant time lag could lead to not sufficiently precise results. A broad class of dynamical systems can be approximated by a model containing both types of the dominant time lags (16). Then, new relations for the optimal and the critical controller gain depending on the ratio $T_d/T_1$ can be found. The previous analysis shows that the $P$ controller optimal gain computed on the basis of $I_1T_1T_d$ model lies between values that were computed for above mentioned $I_1T_2$ and $I_1T_1$ models. For example, for $T_d/T_1 = 0.5$ the optimal gain is $K_{Copt} = 0.297/K_S L$.
(L = Tₐ + T₁), for Tₐ/T₁ = 1, \(K_{\text{Copt}} = 0.322/K_S L\) and for \(Tₐ/T₁ = 2, K_{\text{Copt}} = 0.346/K_S L\). The gains are 3.371, 3.103 and 2.891 times smaller than the values recommended by Ziegler and Nichols. The conclusion can be confirmed by the graphical dependence in Fig. 3 that is drawn for \(\varepsilon = 1\).

Similar results were also achieved for \(PI\) controller setting. As it can be seen from Fig. 9 both controller parameters (\(K_{\text{Copt}}\) and \(K_{\text{Iopt}}\)) computed on the basis of \(I₁T₁Tₐ\) model are again situated between values corresponding to \(I₁Tₐ\) and \(I₁T₁\) models. In the figure the optimal values for selected models are compared with the values determined by Ziegler and Nichols. The analytically computed parameter \(K_I\) that follows \(I₁T₁Tₐ\) model is practically equivalent to the value that was found by Ziegler and Nichols. The coincidence is reached for \(\varepsilon = Tₐ/T₁ \approx 1.13\).

However, it is necessary to mention that Ziegler and Nichols used the classical structure of \(PI\) controller, whereas the \(PI\) controller introduced here arose on the basis of reconstruction and compensation of the input system disturbance. Such modification of control structure ensures smaller overshoot of the transient response than the parallel \(PI\) controller does.

One of the main advantages of this concept of the windupless \(PI\) controller and its tuning is that they can be generalized also for the higher order approximative models, which allow to achieve higher control dynamics.

Acknowledgements

The work has been partially supported by the Slovak Scientific Grant Agency, Grant No. VG-0145.

REFERENCES


Received 18 August 2003

Mikuláš Huba (Doc, Ing, CSc) born 1951 in Ružomberok, graduated from the Slovak University of Technology in Technical Cybernetics in 1974. From 1989 an associated professor at the Department of Control and Automation, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, appointed from 1996 also by leading the Local Centre of Distance Education. Active in the field of nonlinear control and modern learning technologies, like Internet based interactive education, computer aided/based control design, multimedia, etc.

Katárina Žáková (Ing, PhD) graduated from the Slovak University of Technology in Technical Cybernetics in 1991. Since 1991 she is with the Department of Automation and Control, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology. Her current research activities are connected with synthesis of nonlinear systems, computer aided control design and e-learning.