Structure-specific analytical PID tuning
for load disturbance rejection

Alberto Leva* Silvano Seva**

* Dipartimento di Elettronica e Informazione e Bioingegneria, Politecnico di Milano, Italy (e-mail: alberto.leva@polimi.it).
** Graduate student at the Dipartimento di Elettronica, Informazione e Bioingegneria (e-mail: silvano.seva@mail.polimi.it).

Abstract This paper addresses the important and well studied problem of synthesising PID controllers for load disturbance rejection. The tuning rationale, on which some general words are spent in connection to literature research, is to shape the disturbance-to-output frequency response, together with conveniently assigning the poles of the corresponding transfer function. Analytical tuning formulæ are derived, to maximise simplicity and make the presented method applicable on any device. Simulation results support the proposal.

Keywords: PID control; controller tuning; model-based tuning; load disturbance rejection.

1. INTRODUCTION

In process control, “load disturbances are often the major consideration” (Aström and Hägglund, 2004). As a consequence, a lot of research effort has been spent on their effective rejection. However, if one restricts the focus to model-based tuning and explicit rules, two characteristics that are often considered beneficial for the industrial acceptance of an (auto)tuning procedure, the panorama becomes quite narrower.

In this paper we address the type of tuning technique just sketched, and with respect to a major reference, hence a good representative of the state of the art, we propose a solution that is simpler, in that it adopts a uniform controller structure, and produces comparable or better results. It has to be noted that our technique – like its reference counterpart – considers delay-free processes: this may or may not be a limitation, as discussed in Leva and Maggio (2010); anyway, addressing the delay (dominated) case will be the subject of future research.

The paper is organised as follows. Section 2 provides a minimal review of related work, and motivates the particular research here presented. In Section 3 the proposed tuning rules are derived, and in Section 4 they are applied to some process structures that are relevant according to the literature. Sections 5 and 6 respectively report a benchmark simulation campaign and a laboratory experiment, to evidence the achieved results and advantages, while Section 7 concludes the paper by drawing some conclusions and outlining future work.

2. BRIEF LITERATURE REVIEW

Many approaches were proposed to PI/PID tuning for load disturbance rejection, historically in an attempt to cure an “overemphasis on the set point response” (Shinskey, 2002), but basically driven by the necessities of process control: Chen and Seborg (2002) resorted to direct synthesis, Vrančić et al. (2004) adapted the magnitude optimum technique, works like Liu and Gao (2008) and Leva (2005) employed relay-based tuning, others such as Shamsuzzoha and Lee (2008) proposed PIDs with additional filters, and many adopted an IMC-like paradigm (Liu and Gao, 2010; Skogestad and Grimholt, 2012). Also, several works concentrate on “balancing” tuning for set point tracking and for disturbance response, see e.g. Arrieta et al. (2010); Alcántara et al. (2013)

In this work we concentrate on model-based tuning, accounting for the form of the process model but at the same time considering different possible structures for it, and aiming at explicit tuning formulæ. According to Scopus, to date the most cited paper encompassing all the characteristics above, is that by Horn et al. (1996). The authors adopt an IMC-centred approach, which is very well suited to achieve model-based explicit tuning for various model structures, and determine a controller in the form of an ideal PID cascaded to a filter up to the second order. In the quoted paper, rules are obtained for five model structures by changing the structure of the filter, interpreted in IMC terms. This results in controllers of different structure themselves, as some parameters of the general form (PID plus filter), for some models are structurally zero.

The method we propose aims at obtaining the same (or better) result with a uniform and completely standard controller structure, i.e., a real PID (more precisely, a PID with filtered derivative action). The closest work we could find to our approach is that by Jin and Liu (2014), that however does not encompass all the characteristics above, and in particular does not directly shape the disturbance-to-output frequency response magnitude the way we do. As for the tracking/rejection tradeoff, here we concentrate on the latter: if tracking is of interest as well, one can recover it for example with method in Leva and Bascetta (2006), as that technique is independent of how the feedback part of the controller is designed.
3. THE PROPOSED TUNING METHOD

The block diagram for the addressed control scheme is shown in Figure 1, where \( P(s) \) and \( C(s) \) are respectively the transfer functions of the process and the controller, \( w(t) \) is the set point, \( y(t) \) the controlled variable, \( u(t) \) the control signal, and \( d(t) \) a load disturbance to be rejected.

![Control loop with load disturbance.](image)

Figure 1. Control loop with load disturbance.

We consider as the controller to tune a real PID with possibly complex zeroes, that we write as

\[
C(s) = \frac{bC_0 + bC_1 s + bC_2 s^2}{aC_1 s + aC_2 s^2}. \tag{1}
\]

The tuning goal is to constrain the disturbance-to-output transfer function \( Q(s) = Y(s)/D(s) \) to take the form

\[
Q_d(s) = Q_N(s) Q_D(s), \quad Q_D(s) = (1 + s\tau_Q)^n Q, \tag{2}
\]

with \( Q_N(0) = 0 \) as inherent to the structure of (1), incidentally to guarantee asymptotic disturbance rejection, \( \tau_Q > 0 \) for stability, and \( nQ \) large enough to ensure high-frequency roll-off (if not for a particular case briefly discussed in Section 4.3).

In fact, \( nQ \) is dictated by the structure of the considered process models, as there must be as tuning equations as controller parameters—whence also the apparent redundancy in the parametrisation of \( C(s) \), as will become clear in Sections 4.1 through 4.5. On the contrary, \( \tau_Q \) is a tuning parameter (the only one for the method, interpreted as a time constant connected to the desired disturbance step response convergence time—i.e., in some sense, pretty much lie the desired closed-loop dominant time constant in lambda or IMC tuning.

For completeness, in the (dominantly frequent) case of a controller with real zeroes, the controller (1) is immediately converted to the standard ISA PID form

\[
C(s) = K \left( 1 + \frac{1}{sT_i} + \frac{sT_d}{1 + sT_d/N} \right) \tag{3}
\]

by

\[
K = \frac{aC_1 bC_1 - aC_2 bC_0}{aC_1^2}, \quad T_i = \frac{aC_1 bC_1 - aC_2 bC_0}{aC_1 bC_0}, \tag{4}
\]

\[
T_d = \frac{aC_1 bC_0 - aC_1 aC_2 bC_1 + aC_2 bC_2}{aC_1 (aC_1 bC_1 - aC_2 bC_0)},
\]

\[
N = \frac{aC_1 bC_0 - aC_1 aC_2 bC_1 + aC_2 bC_2}{aC_2 (aC_1 bC_1 - aC_2 bC_0)}.
\]

Coming to the process model, we assume the general form

\[
P(s) = \frac{bP_0 + bP_1 s + bP_2 s^2}{aP_0 + aP_1 s + aP_2 s^2}. \tag{5}
\]

with all parameters nonnegative; notice that we are assuming positive gain as well, but obviously this does not impair generality.

The rationale behind the tuning approach we adopt, is to prevent \( |Q(j\omega)| \) from exhibiting a plateau, thereby minimising the width of the band where the disturbance is rejected to the least extent (i.e., where \( |Q(j\omega)| \) is near to its maximum), while allowing some control on the maximum value of \( |Q(j\omega)| \) and of the high-frequency control sensitivity.

The structure chosen for \( Q_o(s) \), see (2), allows to easily obtain an estimate on its \( \infty \)-norm by taking the magnitude of its frequency response at \( \omega = 1/\tau_Q \), where the coincident poles are. For example, if \( Q_o(s) \) has four poles (and one zero in the origin, remember) then the only possible mutual locations of zeroes and poles are shown in Figure 2, and clearly the estimate just introduced is sensible; in the figure, asymptotic Bode plots are sketched for simplicity.

![Possible frequency positions of zeroes (Z) and poles (P) of Q(s) in the fourth order case.](image)

Figure 2. Possible frequency positions of zeroes (Z) and poles (P) of \( Q(s) \) in the fourth order case.

As such, in the following we further denote the estimated \( \infty \)-norm of \( Q_o(s) \) and the high-frequency value of the controller (i.e., of the control sensitivity) magnitude, respectively, as

\[
\hat{Q}_{max} = \max_\omega |Q(j\omega)|, \quad C_\infty = \lim_{\omega \to \infty} |C(j\omega)|, \tag{6}
\]

where the hat sign in \( \hat{Q}_{max} \) underlines that this is an estimate; the exact value depends on the detailed shape of the frequency response, but the estimate is representative in any case. The nice fact in all this reasoning is that once the systems of equations that provide the controller parameters are analytically solved, as done in the following sections, both \( Q_{max} \) and \( C_\infty \) — that respectively quantify the disturbance rejection and the high-frequency control activity — are functions of the tuning parameter \( \tau_Q \), and can be used as the basis for a (possibly) automatic selection of that parameter, although in this paper we do not discuss the matter.

Summarising, the main advantage of our rationale is that the disturbance-to-output frequency response is shaped in a very direct and straightforward manner, and for several process structures — like those addressed below — the resulting tuning relationships are explicit.

4. APPLICATION TO RELEVANT PROCESS STRUCTURES

The quoted paper by Horn et al. (1996) reports IMC controllers for five process structures, optimised for “open
loop dynamics slower than the [desired] closed-loop dynamics”, i.e., for what is sometimes called “strong feedback”, i.e., in turn, essentially for disturbance rejection. With the notation of that reference, that we report here for the reader’s convenience, the five structures treated are

\[ P_A(s) = \frac{k_p}{\tau s + 1}, \quad P_B(s) = \frac{k_p(-s + z)}{(\tau_1 s + 1)(\tau_2 s + 1)}, \]

\[ P_C(s) = \frac{k_p(-s + z)}{\tau s + 1}, \quad P_D(s) = \frac{k_p}{(\tau_1 s + 1)(\tau_2 s + 1)}, \]

\[ P_E(s) = \frac{k_p}{s(\tau s + 1)} \]

4.1 Process A

In the form (5), process A in (7) reads

\[ P_A(s) = \frac{b_p}{1 + a_1 p s}, \]

and with controller (1), this structurally yields a third-order \( Q(s) \), i.e., \( n_Q = 3 \). Setting \( Q_D(s) = (1 + s\tau Q)^3 \) we then get

\[ a_{C1} = 1, \quad a_{C2} = \frac{\tau_Q^3}{a_{P1}}, \quad b_{C0} = \frac{1}{b_{P0}}, \]

\[ b_{C1} = \frac{3\tau_Q - 1}{b_{P0}}, \quad b_{C2} = \frac{a_{P1}(3\tau_Q^2 - a_{P1}) - \tau_Q^3}{a_{P1}b_{P0}} \]

and

\[ \tilde{Q}_{\text{max}} = \frac{b_{P0}\sqrt{\tau_Q^4 + a_{P1}^2}}{2^{1/2}a_{P1}\tau_Q} \]

\[ C_{\infty} = \left[ \frac{a_{P1}(a_{P1} - 3\tau_Q^2) + \tau_Q}{b_{P0}\tau_Q^3} \right] \]

4.2 Process B

In the form (5), process B in (7) reads

\[ P_B(s) = \frac{b_p - b_{P1}s}{1 + a_1 p s + a_2 p s^2} \]

and in this case we have to set \( n_q = 4 \), whence

\[ a_{C1} = \frac{4a_{P2}\tau_Q^3 - a_{P1}\tau_Q^4}{a_{P2}}, \quad a_{C2} = \frac{\tau_Q^4}{a_{P2}}, \quad b_{C0} = \frac{1}{b_{P0}}, \]

\[ b_{C1} = \frac{a_{P1}a_{P2}\tau_Q^3 + a_{P2}^2p_{b0} + a_{P1}p_{b0}p_{b1} + 4a_{P2}^2p_{b0}}{a_{P2}b_{P0}\tau_Q^3 + (a_{P1}p_{b0} + a_{P2}b_{P0})\tau_Q}, \]

\[ a_{C2} = \frac{\tau_Q^4}{a_{P2}}, \quad b_{C2} = \frac{a_{P1}a_{P2}\tau_Q^3 + a_{P2}^2p_{b0} + a_{P1}p_{b0}p_{b1} + 4a_{P2}^2p_{b0}}{a_{P2}b_{P0}\tau_Q^3 + (a_{P1}p_{b0} + a_{P2}b_{P0})\tau_Q} \]

while the expressions of \( \tilde{Q}_{\text{max}} \) and \( C_{\infty} \) are too long to be reported here.

4.3 Process C

In the form (5), process C in (7) reads

\[ P_C(s) = \frac{b_p - b_{P1}s}{1 + a_1 p s}. \]

We include this case for completeness, but we have to notice that a first-order, non strictly proper process – and to top, nonminimum-phase – seems a bit unrealistic indeed in practice. Joined to a real PID, in addition, this gives rise to an open loop transfer function with zero relative degree, which has very well known drawbacks—and in the end, adopting a closed-loop pole placement approach like ours is probably among the best ways to handle such a situation. In any case, we apply the technique to show that formally it works. To do this we have to set \( n_q = 3 \), whence

\[ a_{C1} = 1, \quad a_{C2} = \frac{4a_{P2}\tau_Q^3 - a_{P1}\tau_Q^4}{a_{P2}}, \quad a_{C3} = \frac{\tau_Q^4}{a_{P2}}, \quad b_{C0} = \frac{1}{b_{P0}}, \]

\[ b_{C1} = \frac{4a_{P2}\tau_Q^3 - a_{P1}\tau_Q^4}{a_{P2}}, \quad b_{C2} = \frac{4a_{P2}\tau_Q^3 - a_{P1}\tau_Q^4}{a_{P2}}, \quad b_{C3} = \frac{4a_{P2}\tau_Q^3 - a_{P1}\tau_Q^4}{a_{P2}} \]

4.4 Process D

In the form (5), process D in (7) reads

\[ P_D(s) = \frac{b_p}{1 + a_1 p s + a_2 p s^2} \]

and we have to set \( n_q = 4 \), whence

\[ a_{C1} = 4a_{P2}\tau_Q^3 - a_{P1}\tau_Q^4, \quad a_{C2} = \frac{\tau_Q^4}{a_{P2}}, \quad b_{C0} = \frac{1}{b_{P0}}, \]

\[ b_{C1} = \frac{4a_{P2}\tau_Q^3 - a_{P1}\tau_Q^4}{a_{P2}}, \quad b_{C2} = \frac{4a_{P2}\tau_Q^3 - a_{P1}\tau_Q^4}{a_{P2}}, \]

and

\[ \tilde{Q}_{\text{max}} = \frac{b_{P0}\tau_Q^2}{4^{1/2}a_{P2}(4a_{P2} - a_{P1}\tau_Q)^2 + a_{P2}^2}, \]

\[ C_{\infty} = \frac{6a_{P2}^2 - 4a_{P1}a_{P2}\tau_Q + (a_{P1}^2 - a_{P2})\tau_Q^2}{a_{P2}b_{P0}\tau_Q} \]
4.5 Process E

In the form (5), process E in (7) reads

\[ P_{E}(s) = \frac{b_{P_{0}}}{s + a_{P_{2}}s^{2}} \]

and we have to set \( n_{q} = 4 \), whence

\[ a_{C_{1}} = \frac{4a_{P_{2}}^{3} - \tau_{Q}^{4}}{a_{P_{2}}^{2}}, \quad a_{C_{2}} = \frac{\tau_{Q}^{2}}{a_{P_{2}}}, \quad b_{C_{0}} = \frac{1}{b_{P_{0}}}, \]

and

\[ b_{C_{1}} = \frac{4\tau_{Q}}{b_{P_{0}}}, \quad b_{C_{2}} = \frac{6a_{P_{2}}^{2} + 4a_{P_{2}}^{3} + \tau_{Q}^{4}}{a_{P_{2}}^{2}b_{P_{0}}}, \]

and

\[ \hat{Q}_{\text{max}} = \frac{b_{P_{0}}\tau_{Q}^{2}(4a_{P_{2}} - \tau_{Q})^{2} + a_{P_{2}}^{2}}{4a_{P_{2}}^{2}}, \]

\[ C_{\infty} = \frac{6a_{P_{2}}^{2} - 4a_{P_{2}}\tau_{Q} + \tau_{Q}^{2}}{a_{P_{2}}^{3}b_{P_{0}}\tau_{Q}^{3}}.\]

4.6 General remarks

A first thing to notice is that the five process structures considered yield very different results as for the existence of values for \( \tau_{Q} \) that minimise \( \hat{Q}_{\text{max}} \) and/or \( C_{\infty} \), as summarised in table 1. The somewhat opposite effect of the tuning parameter conforms of the opportunity of structure-specific tuning if high rejection performance is required.

A second notable remark is that, although critical cancellations never happen (we omitted computations for brevity), in the case of nonminimum-phase processes the PID can turn out to be unstable, i.e., to have its second pole in the right half plane. In principle this can make sense as high rejection performance is required, but further study is required on this aspect, especially to properly quantify the closed-loop stability robustness, that we cannot undertake in this paper.

<table>
<thead>
<tr>
<th>Proc.</th>
<th>( Q_{\text{max}} )</th>
<th>( C_{\infty} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>min. for ( \tau_{Q} = \sqrt{a_{P_{2}}} )</td>
<td>min. for ( \tau_{Q} = \sqrt{a_{P_{2}}} )</td>
</tr>
<tr>
<td>B</td>
<td>no analytical solution</td>
<td>( \exists ) minimising ( \tau_{Q} ) (complex expression)</td>
</tr>
<tr>
<td>C</td>
<td>no analytical solution</td>
<td>strictly incr. with ( \tau_{Q} )</td>
</tr>
<tr>
<td>D</td>
<td>strictly incr. with ( \tau_{Q} )</td>
<td>strictly decr. with ( \tau_{Q} )</td>
</tr>
<tr>
<td>E</td>
<td>strictly incr. with ( \tau_{Q} )</td>
<td>min. for ( \tau_{Q} = 3\sqrt{a_{P_{2}}} )</td>
</tr>
</tbody>
</table>

Table 1. Summary of the effect of parameter \( \tau_{Q} \) on \( \hat{Q}_{\text{max}} \) and \( C_{\infty} \).

5. SIMULATION RESULTS

In this section we present some samples of an extensive simulation campaign, that we conducted to assess the correct behaviour of the proposed technique, and to verify that the it actually exhibits the expected strengths.

For compactness we introduce a normalised complex variable \( \sigma = \frac{\tau}{\tau} \), where \( \tau \) is the only or the largest time constant of the pole(s) of processes A–E not in the origin. Defining the normalised processes \( P_{A–E}(\sigma) \) as \( P_{A–E}(\sigma/\tau)/\mu \) we get

\[ p_{A}(\sigma) = \frac{1}{1+\sigma}, \quad p_{B}(\sigma) = \frac{1 - \sigma\theta}{(1+\sigma)(1+\sigma\psi)}, \]

\[ p_{C}(\sigma) = \frac{1 - \sigma\theta}{1+\sigma}, \quad p_{D}(\sigma) = \frac{1}{(1+\sigma)(1+\sigma\psi)}, \]

\[ p_{E}(\sigma) = \frac{1}{(1+\sigma)}. \]

where \( 0 < \psi < 1 \) since the “other” pole has to be faster, while in principle \( \theta \) (positive) could be arbitrarily large. Rigorously the numerator of \( p_{B} \) should be \( \tau \), incidentally, but in this case it just plays the role of a scale factor, as the process is type 1, and therefore can be omitted in the normalised transfer function.

5.1 Test 1

First, we show that the proposed rule actually avoids the magnitude plateau in the frequency response or \( |Q(s)| \).

Figure 3 shows an example with processes A and E; the disturbance is least rejected.

Figure 3 shows an example with processes A and E; the disturbance is least rejected.

5.2 Test 2

We now show an example of comparison between our proposal and the reference by Horn et al. (1996). The example uses two processes in class B, namely

\[ p_{B1}(\sigma) = \frac{1 - 2\sigma}{(1 + \sigma)(1 + 0.4\sigma)}, \]

\[ p_{B2}(\sigma) = \frac{1 - 0.4\sigma}{(1 + \sigma)(1 + 0.9\sigma)}, \]

of which the first has a remarkably undershooting step response and a second pole at significantly higher frequency with respect to the dominant one, while the second has two

Figure 3. Test 1 – avoiding the magnitude plateau in the frequency response or \( |Q(s)| \).
almost coincident poles and a less evident overshoot (in some sense, thus, the two processes are “quite extreme” in the class).

Figure 4 reports the load disturbance step responses of (4) with a “low” and a “high” value of the parameter controlling the closed-loop band, i.e., $\lambda$ in the work by Horn et al. (1996) and $\tau_q$ in our proposal.

As can be seen, both parameters act in the respective technique in the desired manner. On one hand this proves that the analogy in their interpretation is correct, but on the other hand, as expected, there is some advantage in governing the disturbance-to-output transfer function directly. For example, with $p_{B1}$ the maximum error might slightly increase, the ISE is comparable but the settling is definitely shorter, while with $p_{B2}$ there is an apparently less oscillatory behaviour.

5.3 Test 3

In this test we again compare our technique to the quoted reference, but this time in terms of the load disturbance step response ISE, defined as

$$ISE = \int_0^\infty e(t)^2 dt,$$

(26)

where $e(t) = w(t) - y(t)$ is the error. More precisely, we compare the relative ISE, i.e., the quantity $R_{ISE}$, as defined by

$$R_{ISE} = \frac{ISE_{prop}}{ISE_{horn}},$$

(27)

$ISE_{prop}$ and $ISE_{horn}$ being the ISE computed with the proposed and the reference controller, respectively.

Figure 5. Test 3 – comparing the load disturbance step response ISE.

In this section we present an experiment in which the proposed method is applied to a laboratory apparatus, and compared to Horn et al. (1996) plus two other PID tuning techniques well known to be suitable for disturbance rejection, namely the AMIGO (Åström and Hägglund, 2004) and the SIMC (Skogestad and Grimholt, 2012). For completeness we notice that AMIGO tunes an ideal PID, that we made proper with $N = 20$ as suggested in the quoted reference. As for the SIMC, given the first-order process, we used the PI rule, with the additional advantage of achieving a zero relative degree controller without choosing the second pole arbitrarily. Finally, we notice that for process class A, the most appropriate for the addressed dynamics, the controller by Horn et al. reduces to a PI. All in all, with the inevitable shortcomings of any experimental setup, we deem this a fair enough comparison.

The used apparatus, described in Leva (2003), is a temperature control system composed of a small metal plate electrically heated by a transistor. The command to that transistor is the control signal, while the plate temperature is the controlled variable. A second transistor is connected to heat the plate, thereby apparently providing a load disturbance. Since the controlled dynamics is dominantly first-order, all the three methods just mentioned fit. The tuning results are in table 2.

Figure 6 shows the response of the controlled plate temperature $T_p$ (top plot) and of the $[0,100]$ command $u_h$ to the
Table 2. Controller parameters in the experimental test.

<table>
<thead>
<tr>
<th>Method</th>
<th>$K$</th>
<th>$T_i$</th>
<th>$T_d$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>88</td>
<td>15</td>
<td>0.35</td>
<td>0.03</td>
</tr>
<tr>
<td>Horn et al.</td>
<td>41</td>
<td>30</td>
<td>0</td>
<td>n/a</td>
</tr>
<tr>
<td>AMIGO</td>
<td>31</td>
<td>30</td>
<td>3.5</td>
<td>20</td>
</tr>
<tr>
<td>SIMC</td>
<td>33.5</td>
<td>56</td>
<td>0</td>
<td>n/a</td>
</tr>
</tbody>
</table>

The proposed technique yields the best disturbance rejection; there is of course a cost in terms of of a more nervous control and a higher sensitivity to measurement noise, but both phenomena appear up to a reasonable extent. The technique can therefore supposed to be also viable in practical process control applications.

7. CONCLUSIONS AND FUTURE WORK

We presented a technique to analytically tune a real PID in a structure-specific manner as for the process dynamics. The underlying approach is based on assigning the closed-loop poles so as to shape the disturbance-to-output frequency response to exhibit no plateau. The technique has a single parameter, interpretable in a way similar to $\lambda$ in IMC-based rules, and for which “optimal” values can be sought (although this was not discussed here in depth).

Future work will proceed along three main directions. The first one is to apply the same rationale to other process structures—including delay-dominated ones. The second consists of posing and solving optimisation problems (possibly not analytically, however) to automatically select the tuning parameter. The third is to integrate the proposed method — that refers to the feedback part of a PID – into the synthesis of two-degree-of-freedom controllers. As a further possibility, then, different structures for the desired disturbance-to-output dynamics can be considered, and analogously, the application of the same idea to other controller structures can be studied.

REFERENCES


