Available online at www.sciencedirect.com
SCIENCE DIRECT
Control Engineering Practice 14 (2006) 149–163
CONTROL ENGINEERING PRACTICE
www.elsevier.com/locate/conengprac

On the automatic tuning and adaptation of PID controllers

I.J. Gyöngy, D.W. Clarke*

Department of Engineering Science, Oxford University, Parks Road, Oxford OX1 3PJ, UK

Received 16 January 2004; accepted 20 January 2005
Available online 5 March 2005

Abstract

A simple approach to the automatic tuning of PID process controllers is proposed. Like the relay-based autotuner, its objective is to attain a design-point on the Nyquist diagram. By injecting sinewaves and employing a phase/frequency estimator, closed-loop adaptive tuning is possible and there is exact convergence to the design-point without the approximations of describing-function theory. The variant discussed here achieves a required phase margin and imposes a carefully chosen constraint on the controller parameters, leading to consistent behaviour for a wide variety of generic test-cases. A real-life demonstration on a non-linear flow rig is provided.

Keywords: PID tuning; Autotuning; Adaptive control; Process control; Phase/frequency estimator

1. Introduction

Automatic tuning has been extensively studied in literature and is a well-established feature in industrial PI(D) controllers (Hang, Lee, & Ho, 1995; Åström, Hägglund, Hang, & Ho, 1993). Nevertheless, the desire to improve the applicability and robustness of existing autotuners still fuels much research (e.g. Ho, Hong, Hansson, Hjalmarsson, & Deng, 2003; Tan, Ferdous, & Huang, 2002; Yu, 1999).

Since the exact dynamics of the plant is generally unknown, the basic feature of autotuners is some experimental procedure by which plant information is obtained in order to compute the controller parameters. Autotuning techniques can therefore be classified according to this experimental procedure. Three main categories that can be established are step response, periodic excitation and relay-based schemes. The first of these approaches relies on simple open-loop step testing to characterise the process dynamics. The drawback of such methods is their sensitivity to disturbances (due to the experiment being performed open-loop) and, in many cases, the use of a parameterised process model (e.g. first-order/dead-time), which can result in a lack of generality. Alternatively, the open-loop plant may be excited using more complex (binary or multi-level) pseudo-random or multisine signals, with the frequency spectrum tailored to the particular process in question (Braun, Ortiz-Mojica, & Rivera, 2001; Barker & Godfrey, 1999). These type of signals, combined with appropriate signal processing tools, can yield accurate process models or frequency response estimates in which the effects of noise and process non-linearities are suppressed (the signals may also be used to highlight non-linearities).

By contrast, relay feedback tuning schemes use a closed-loop (CL) experiment. The test involves the replacement of the PI(D) controller by a relay with hysteresis, which for a wide range of processes induces a limit cycle in the loop. Describing-function analysis allows the frequency response corresponding to a particular phase shift (set by the size of the hysteresis) to be estimated from the frequency and amplitude of

*Corresponding author. Tel.: +44 1865 273021; fax: +44 1865 273906.
E-mail addresses: istvan.gyongy@eng.ox.ac.uk (I.J. Gyöngy), david.clarke@eng.ox.ac.uk (D.W. Clarke).

0967-0661/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.conengprac.2005.01.007

this oscillation. The method is simple (requiring no prior information) and fast but, especially in its simplest form, not without limitations. Firstly, the approximations inherent in describing-function analysis can, in the cases of some industrially significant types of processes, lead to errors in the estimate of the frequency point; this in turn can result in less effective PI(D) parameters. Load disturbances also influence the estimation accuracy.

A further disadvantage of the relay autotuner (as well as the aforementioned open-loop schemes) is that it is only suitable for off-line tuning (i.e. no control is provided during tuning). Although modifications have been proposed in which the experiment is performed on-line (e.g. Tan, Lee, & Jiang, 2001), the upset caused by the relay is rather large. The latest developments in relay autotuning also include the simultaneous estimation (with no describing-function approximations) of multiple frequency points, at the cost of increased computation complexity (Hang, Åström, & Wang, 2002; Sung & Lee, 2000).

Here we propose a different autotuning method that retains the PID controller in the loop and uses information obtained by injecting a variable-frequency sinewave into the loop (normally undesirable, but in this case the amplitude can be reduced to match that of noise, so output quality is maintained). The concept of using sinusoidal perturbation for CL controller tuning/adaptation is by no means new. One of the early explorations of the technique was carried out by Smyth and Nahi (1963).[1] In their system, the phase and amplitude response information used to adapt the parameters of a feedback compensation network is acquired via a fixed frequency test-signal. The adaptive PID controller of Glattfelder (1969), on the other hand, applies three (fixed) frequencies and seeks to attain a 'well-shaped' CL phase versus frequency curve. More recent work includes the PI tuner of Crowe and Johnson (2002), which uses two sequentially injected, variable-frequency test signals to achieve phase margin and gain margin specifications. However, despite the long history of perturbation methods, autotuners (or indeed adaptive controllers) based on such techniques have so far failed to gain wide-scale adoption in the process control world. The aim here is therefore to have a tuner comparable to relay methods in terms of its ease of use and speed, yet, thanks to its greater flexibility, giving superior PID tuning. Robustness to noise, set-point changes and disturbances is also an important factor.

In addition to providing 'tuning on demand', the method is also capable of keeping the loop continuously in tune, at the expense of a low-amplitude perturbing sinewave signal being added to the loop's set-point. Although effective adaptive algorithms exist (e.g. Huang, Roan, & Jeng, 2002) which use just the normal operating signals in the loop to monitor the process (obtaining process models, frequency points or characteristics such as damping), these require supervision (Hägglund & Åström, 2000).

1.1. Basics

A schematic diagram of the system is shown in Fig. 1. The objective is to adapt the controller so as to achieve a carefully chosen design-point on the Nyquist diagram.

The key components are phase/frequency and plant gain estimators (PFE, GE), described in detail in (Clarke & Park, 2003; Clarke, 2003). In essence a PFE injects a test sinewave into a system and continuously adapts its frequency ω_1 until its phase shift attains a desired value θ_d (in this case that of the design-point). When applied to the CL system, we can describe this process using the equation:

$$\frac{\mathrm{d}\omega_1(t)}{\mathrm{d}t} = K_a(\mathrm{Arg}\{G_c(\mathrm{j}\omega_1)\} - \theta_d), \tag{1}$$

where K_a is the adaptive gain of the PFE and $G_c(s) = CG(s)/(1 + CG(s))$ is the CL transfer function. In the above equation a 'frozen parameter' assumption has been made, in that the frequency ω_1 (and $C(s)$) are taken to vary slowly enough to allow transients to be neglected (so that the response of $G_c(s)$ to the test signal is a quasi-steady-state sinusoid).

In a concurrent operation, the instantaneous values of the test frequency and the corresponding plant gain $|G(\mathrm{j}\omega_1)|$ (as provided by the GE) are used to adjust the controller parameters so that convergence is attained.

Also forming important parts of the tuner, but not shown in Fig. 1, are variable band-pass filters (VBPF) at the inputs of the PFE and GE. These are second-order filters centred on the current value of the test frequency.

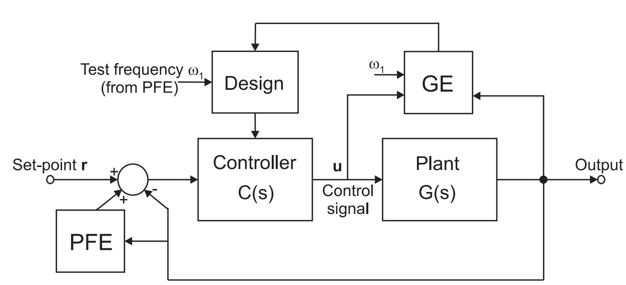

Fig. 1. Schematic diagram of system.

[1]The jet engine fuel control system of Vasu (1957) also deserves a mention. The system tries to maintain an optimal (average) fuel flow (giving maximum thrust, or equivalently, maximum engine pressure ratio) by introducing a slow sinusoidal variation (frequency f_o) in the flow. Coherent detection is used to detect the component at f_o of the resulting fluctuation in the engine pressure ratio. The detected amplitude, which should be zero at optimum fuel flow, is passed through a compensation network to generate the drive signal for the fuel servo.

They are used to isolate the probing signal from the other signals circulating in the loop (such as noise, set-point changes and load disturbances).

The algorithm is initialised using a first-order/dead-time (FODT) approximation $G_a(s)$ for the plant, obtained through a simple step test. The initialisation involves the computation of suitable values for K_a, $\omega_1(0)$ and for other parameters associated with the GE, PFE and the controller.

Note that the tuner is not significantly more complex than the relay method from a signal processing perspective. The basic form of that method requires the amplitude and frequency of a waveform to be accurately determined (and it too uses filters to get better process estimates); here we essentially rely on two amplitude estimators (to compute the gain of the process) and a phase detector.

1.2. PID tuning

A PI tuner based on the approach has already been proposed and shown to be uniformly effective in a wide range of test cases (Clarke, 2003). The aim here is to devise a similar tuner for PID controllers, as for many plants such controllers can offer better control in terms of speed and disturbance rejection (this is especially true in the case of Type 1 processes).

The design approach adopted uses Nyquist ideas, with an aim to choose a design-point D_1 so that the compensated Nyquist plot exhibits certain characteristics and as a result favourable CL behaviour is obtained. This is illustrated (for a typical Type 0 plant) in Fig. 2. Ideally, we would like the Nyquist curve to be the vertical line $\mathrm{Re}\{CG(\mathrm{j}\omega)\} = -0.5$, as that would mean unity CL gain at all frequencies and an infinitely fast response. However, it is evident that this is not achievable in practice (for a start, practical systems exhibit a finite bandwidth and slew rate, so $CG(\mathrm{j}\omega)$ must tend to 0 as $\omega \to \infty$). Hence the best feasible shape is something alike to curve B, with a high ω_d to ensure a fast CL response. By contrast, a compensated Nyquist contour similar to A, where the curve 'bulges out' towards the critical point $(-1 + \mathrm{j}0)$ is highly undesirable, as this indicates an oscillatory response. Likewise, design rules which give a curve such as C (having a low-frequency asymptote away from $\mathrm{Re} = -0.5$) should be avoided, as they result in a slow CL pole.

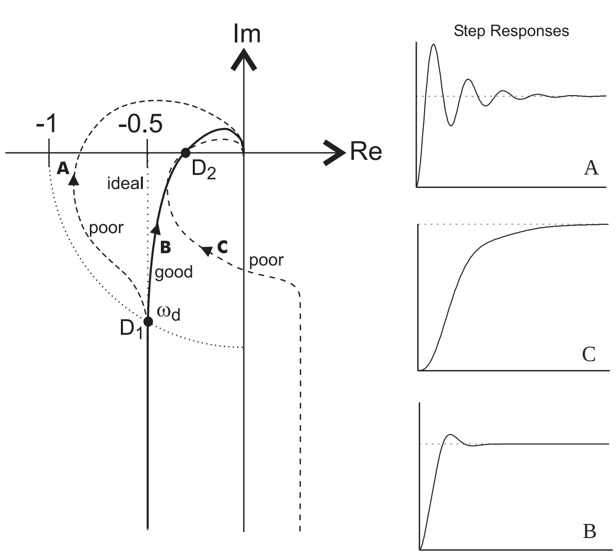

Fig. 2. Desired shape for compensated Nyquist plot.

A natural choice is to make the design-point D_1 a phase margin (PM) point and to use an extra constraint on the controller parameters to ensure that the Nyquist contour does not protrude towards the critical point. This is the design rule implemented in the PID tuner presented here. An alternative option that has also been explored is the use of a second design-point D_2 so that both a prescribed PM and gain margin (GM) may be achieved (Tan, Wang, Hang, & Hägglund, 1999). Indeed, a successful tuner based on this principle and using two *simultaneous* probing signals was designed (Gyöngy, 2003). Space precludes full discussion of the latter algorithm, but we note that the flexibility gained probably does not warrant the increased complexity.

The controller structure used in the PM tuner is as depicted in Fig. 3. The controller is implemented in serial form and is split into two components: a PI and a PD unit, with the plant in between. Hence derivative action is applied on the output of the plant and 'derivative kicks' resulting from set-point changes are avoided. The PD module is filtered to prevent excessive amplification of measurement noise and the configuration used for the PI module protects the controller from integral wind-up (Clarke, 1984).

The effective transfer function of the overall controller (ignoring the filter) takes on the standard form for a PID compensator:

$$C(s) = K\left(1 + \frac{1}{sT_i} + sT_D\right) = K\frac{(s^2T_iT_D + sT_i + 1)}{sT_i}, \tag{2}$$

where $K = K'(T_1 + T_2)/T_1$, $T_i = (T_1 + T_2)$ and $T_D = T_1T_2/(T_1 + T_2)$.

Returning to the schematic diagram of the tuner in Fig. 1, we note that the GE actually monitors the combination plant/filter. Although the effect of the filter is generally small, accounting for it in the adaptation of the controller allows for more accurate convergence to the design-point.

The paper is organised as follows. Section 2 deals with the design approach incorporated in the tuner. In Section 3, the operation and initialisation are outlined. Section 4 discusses the simulations that were carried out

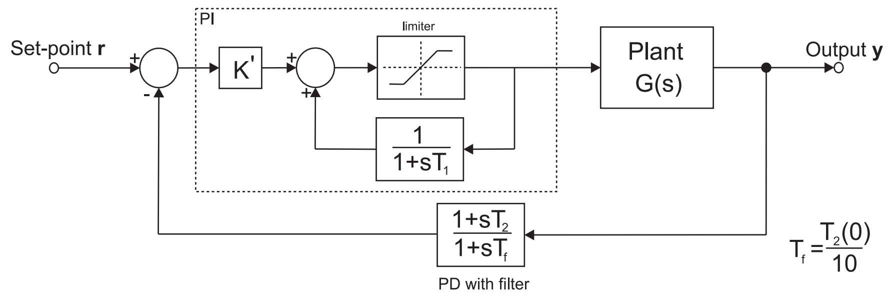

Fig. 3. PID controller structure used in tuner.

on a range of test plants to validate the method's effectiveness. Comparisons are made with existing tuning rules. Section 5 reports real-life results obtained on a non-linear flow rig.

2. Design approach

The aim of the tuner is to achieve a desired phase margin ϕ_m. Denoting the compensated open-loop by $G_o(\mathrm{j}\omega)$, we may express the design problem as

$$G_o(\mathrm{j}\omega_\phi) = K\left[1 + \mathrm{j}\left(\omega_\phi T_D - \frac{1}{\omega_\phi T_i}\right)\right] G(\mathrm{j}\omega_\phi) = -\exp(\mathrm{j}\phi_m), \tag{3}$$

where ω_ϕ is the PM frequency of the compensated plant. There are four variables in total $(K, T_i, T_D, \omega_\phi)$, whereas the above complex expression yields two real equations. Thus, to have a unique solution, two additional constraints must be introduced. Firstly, the ratio $r = T_D/T_i$ is fixed at 0.25 (meaning that the controller has a double zero at $s = -2/T_i$), which is found to be suitable for the typical plant dynamics encountered in process control.

The second condition is better appreciated if we consider a graphical method for devising PID controllers for a given design-point D on the Nyquist diagram Hind, 1980. The approach requires knowledge of $G(\mathrm{j}\omega)$ and is illustrated in Fig. 4. Expressing the controller as

$$C(\mathrm{j}\omega) = KC'(\mathrm{j}\omega) = K(1 + \mathrm{j}a) \quad \text{where } a = \omega r T_i - \frac{1}{\omega T_i},$$

we can write

$$\frac{1}{C'(a)} = \frac{1}{1 + \mathrm{j}a}.$$

As a goes from $-\infty$ to ∞, the Argand diagram of $1/C'$ describes a clockwise circle about $0.5 + \mathrm{j}0$, starting and finishing at O and getting to the halfway-point $1 + \mathrm{j}0$ when $a = 0$. Rotating and scaling this circle so that its diameter is OD', where D' is any point on the line OD, we arrive to the situation depicted in Fig. 4.

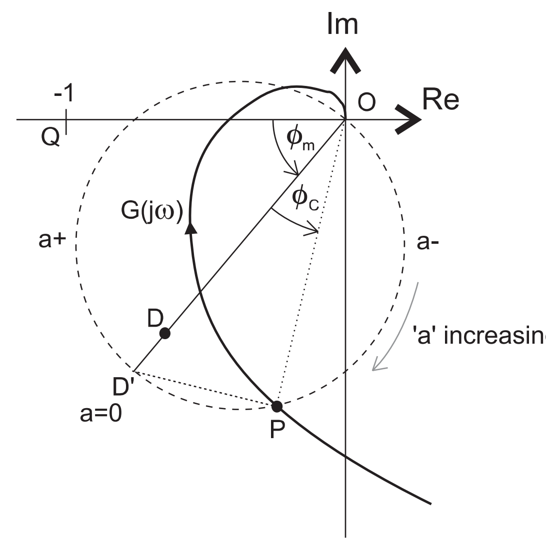

Fig. 4. Graphical PID design.

Assume now that $KC'(a)$ is a controller giving the desired design-point D at a frequency ω_ϕ. Then the corresponding point P on the uncompensated Nyquist contour may be expressed as

$$G(\mathrm{j}\omega_\phi) = \frac{1}{KC'(a)} G_o(\mathrm{j}\omega_\phi) = \frac{1}{K}\frac{\exp(\mathrm{j}\phi_c)}{|C'(a)|} G_o(\mathrm{j}\omega_\phi),$$

where $\phi_c = \mathrm{Arg}\{1/C\} = -\tan^{-1} a$. Returning to Fig. 4, if we let $K = OD/OD'$ then the inverse controller will scale D to D' and rotate it along the circle, so that P will correspond to the intersection of the circle with the Nyquist plot of $G(\mathrm{j}\omega)$. The remaining controller parameters may be easily deduced from the diagram. Firstly, we find the frequency ω_ϕ associated with P and determine $a = \tan P\hat{O}D = DP/OP$ (N.B. positive a is for P above OD and negative a for below). Then, from the definition of a we have

$$T_i = \frac{a + \sqrt{a^2 + 4r}}{2r\omega_\phi}, \tag{4}$$

and T_D is computed from rT_i.

We note however, that the freedom in the diameter of the design circle (OD' can be anything as long as the resulting circle intersects the $G(\mathrm{j}\omega)$ plot) means that there is a wide range of possible controller designs giving the required design-point (even with r fixed). Thus we fix the value of the parameter a, or equivalently, we specify the position at which the point P should lie along the design circle. The question which then remains concerns the ideal value of a.

2.1. Choosing the design parameter a

There are a number of factors which need to be taken into account in the choice of a. For instance, the design frequency ω_ϕ, which provides a good indicator of CL bandwidth, must be reasonably high to ensure a fast speed of response, but not excessively so to avoid plant saturation and noise problems. In addition, we would like a small diameter OD' for the design circle as that results in a high controller gain K and therefore good disturbance rejection.

We can see from Fig. 4 that higher (more positive) values of a (P further along the design circle) will correspond to higher K (shorter OD'), until we get to the situation whereby the design circle touches the $G(\mathrm{j}\omega)$ locus, which maximises K. Any additional increase in a leads to a reduction in K. Clearly, the design frequency ω_ϕ and the associated phase lead (and derivative action $\mathrm{j}\omega_\phi T_D$) also increase with a. We therefore find that for overly high, positive a, the frequency ω_ϕ is too high, and we have a low K. With large negative a, the controller gain is again small, but this time ω_ϕ is low, leading to a slow CL.

It turns out however that due to the differing high-frequency Nyquist diagram sections exhibited by different plants, there is no single value of a that is universally suitable. Specifically, for lower-order plants, higher (or more positive) values of a are appropriate (for high K and fast response), whereas higher-order plants and plants with dead-time require lower (more negative) a if oscillatory CL behaviour is to be avoided. This follows from the fact that the latter family of processes exhibit a slow roll-off of gain with phase shift. As a result, the increased phase lead (at high ω) associated with a high a does not help to move the high-frequency section of the Nyquist contour away from the critical point. Instead, the derivative action behind the phase lead actually makes matters worse by combining with the high K to amplify the high-frequency section.

2.1.1. Use of FODT model

Fortunately, the FODT model (see Eq. (8)) obtained from the prior step test allows the parameter a to be effectively chosen. We find that the FODT models of higher-order plants (and plants with a dead-time) are characterised by a longer relative dead-time:

$$\lambda = \frac{T_d}{T}.$$

By selecting a according to λ, good tuning can be obtained. In practice the exact value of λ obtained for a given plant will depend on the duration of the step test, amongst other factors. Hence we look to cover the likely range of λ by just a few values of a, each being suitable for a wide range of λ. Taking FODT plants with λ ranging from 0.1 to 4 and using $PM = 60^\circ$, we can obtain a fairly consistent step response (similar overshoot and no oscillatory behaviour or slow pole) and robustness (gain margin >2) by choosing between just three different values for a, as shown in Fig. 5. Moreover, these values include some leeway for inaccuracies in λ and are also appropriate for $PM < 60^\circ$. For $PM > 60^\circ$ the slightly altered values given in Table 1 must be used.

The values of a obtained this way also turn out to be suitable for a wide range of non-FODT plants, as it is demonstrated later. The main exceptions are Type 1 plants and highly underdamped plants (not common in process control and cannot be approximated well by FODT models).[2] Provided there is negligible dead-time, a value of around $a = 1.5$ (so that P is roughly $\frac{3}{4}$ the way around the design circle) is found to be appropriate for both types of processes (from the perspective of achieving the fastest response without compromising on robustness). For Type 1 plants having considerable dead-time a should be lowered to 0. Unfortunately, in the case of underdamped systems with dead-time the appropriate value of a is not well defined. However it is found that with high λ processes (>2.5), whether or not they involve oscillatory dynamics, it is often useful to use $a < -3$ if good robustness is to be ensured.

With a fixed a, the resulting design problem is guaranteed to yield a solution provided that the open-loop phase shift of the plant reaches that corresponding to P as the frequency is increased from zero. We can express this condition as

$$\mathrm{Arg}\{G(\mathrm{j}\omega)\} = -\pi + \phi_m - \tan^{-1} a \quad \text{for some } \omega = \omega_\phi, \tag{5}$$

where ϕ_m is the desired phase margin.

Comparing the overall design strategy with that implemented in relay-based methods we note that in the latter case the point P (identified by the methods) is normally of a fixed phase-shift, though the effective design-point D, which is not generally a phase margin one, is in some rules dependent on λ.[3]

[2]Underdamped processes are perhaps best controlled by pole cancellation. With our controller this is not possible as it has real zeros, so a high controller gain K is required to pull down the complex poles towards the real axis and hence increase the damping.

[3]E.g. in the refined Ziegler and Kappa-Tau tuning rules (Hang et al., 1995; Åström & Hägglund, 1995).

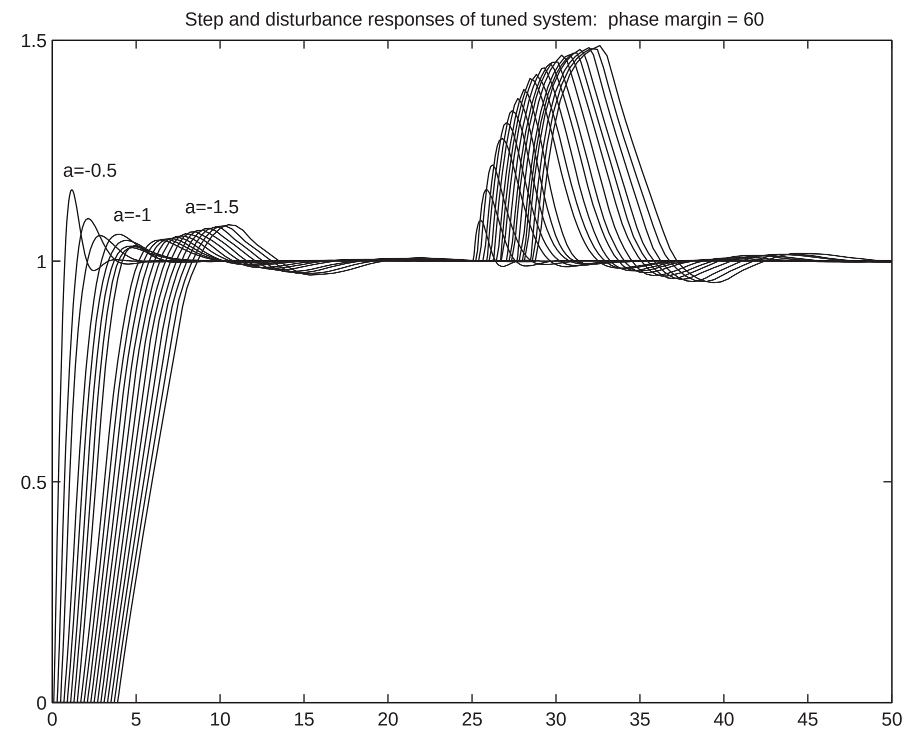

Fig. 5. Choosing a for FODT systems.

Table 1
Choosing a according to λ

PM	$\lambda \leqslant 0.5$	$0.5 < \lambda \leqslant 1.5$	$\lambda > 1.5$
$\leqslant 60^\circ$	-0.5	-1	-1.5
$> 60^\circ$	-1	-1.5	-2.5

3. Tuner operation

Although the design-point methodology is described above in an open-loop context, it can be easily related to the CL case. In particular, the CL frequency response associated with the design-point D may be expressed using the triangle QDO as

$$\text{Gain}: \; K_d = \frac{OD}{QD} = \frac{\sin(\pi - \phi_m)/2}{\sin \phi_m}$$

$$\text{Phase shift}: \; \theta_d = -P\widehat{O}D = -\frac{(\pi - \phi_m)}{2}. \tag{6}$$

To implement the design rules in the tuner, θ_d is specified in the PFE as the desired phase shift for the test frequency ω_1. Thus the PFE will vary the frequency ω_1 of the probing signal such that its phase shift approaches that of the design-point. At the same time, we seek to adapt the controller parameters so that the open-loop gain at ω_1 corresponds to that dictated by the design-point ($OD = 1$), whilst ensuring that the design parameter a is met, and so convergence is obtained to the specified PM ($\omega_1 \to \omega_\phi$). Thus from Eqs. (3) and (4) the required adaptation are

$$K = \frac{1}{\sqrt{1 + a^2}|G(\mathrm{j}\omega_1)|}, \quad T_i = \frac{2}{\omega_1}(a + \sqrt{a^2 + 1}),$$

$$T_D = 0.25 T_i. \tag{7}$$

Note that in the serial controller realisation shown in Fig. 3 we have

$$K' = \frac{K}{2}, \quad T_1 = T_2 = \frac{T_i}{2}.$$

3.1. Initialisation

The tuner comprises numerous parameters which must be well-chosen if effective operation is to be ensured. For instance, the initial value for the PFE frequency must be reasonably close to the actual design frequency ω_ϕ. Equally, the adaptive gain K_a of the PFE needs to be large enough to ensure a reasonably fast rate of convergence, yet excessively high values lead to instability. Hence a procedure must be devised that initialises the algorithm in a semi-automatic manner.

A similar route is followed as in the PI tuner of Clarke (2003), the initialisation being based on a FODT model

$G_a(j\omega)$ for the plant, obtained through a prior step test:[4]

$$G_a(s) = \exp(-sT_d)\frac{K_p}{1+sT}. \tag{8}$$

The model does not have to be particularly accurate or up-to-date so in most cases the step test will only need to be carried out the first time the algorithm is run on a specific plant.

To begin with, we select the PID design parameter a according to the FODT model, as detailed in the previous section (or set $a = 1.5$ if plant is Type 1). We then estimate the frequency ω_ϕ corresponding to the design-point by carrying out the PID design on the FODT model $G_a(s)$ (we denote the estimate by ω_ϕ^e). This merely involves solving for ω the equation:

$$\text{Arg}\{C(j\omega)G_a(j\omega)\} = -\pi + \phi_m$$

$$\Rightarrow -\omega T_d - \tan^{-1}(\omega T) + \pi - \phi_m + \tan^{-1} a = 0, \tag{9}$$

either numerically, or using the approximate analytical solution derived in Clarke and Park (2002) for an equation of the same form. Next, the initial frequency $\omega_1(0)$ of the PFE is set to ω_ϕ^e while the initial output of the GE to $|G_a(j\omega_\phi^e)|$, allowing start-up controller parameters to be generated. The time constant of the controller filter is fixed at $T_f = T_2(0)/10 = T_i(0)/20$ (where $T_i(0)$ is as given by Eq. (7)). The adaptive gain K_a for the PFE is then calculated from the design-point frequency estimate and the FODT model as specified in Clarke (2003). All the remaining parameters of the PFE and GE are also chosen exactly the same way as in that article.

The VBPF bandwidths $\beta = \Delta\omega/\omega_1$ are given values of 1.5, 0.15 (PFE, GE), and the time constant of the GE post filter is set to $T_F = 10/\omega_1(0)$.

The initialisation procedure culminates in the PFE being switched on and starting to inject the sinewave into the CL system. To allow the responses of the estimators to settle before they embark on adjusting the controller parameters, the adaptation of the test frequency and the plant gain estimate is inhibited for a period:

$$2(T_d + \tau), \tag{10}$$

where τ relates to the dead-time in the PFE and GE resulting from the use of Hilbert transformers (Clarke, 2003). This way, erratic adaptation of the controller at start-up is avoided. The inhibition period can also be applied (optionally) after large step changes in set-point for increased robustness in the adaptation.

[4]Note that the model can provide a crude representation for Type 1 plants since for high K_p, T we have $G_a(s) \approx K_p \exp(-sT_d)/(sT)$.

4. Simulations

Three series of simulations were performed using Matlab and Simulink. The first set involved noise-free plants and the emphasis was on assessing the initialisation and convergence of the algorithm, as well as the quality of the final tuning. The tuning performance of the algorithm was then contrasted to that of popular tuning rules employed by industrial autotuners and a recently proposed iterative feedback tuning technique (Lequin, Gevers, Mossberg, Bosmans, & Triest, 2003). The third set of simulations were devised to evaluate the convergence and tracking ability of the tuner under real-world operating conditions such as measurement noise, load disturbances and set-point changes, so that the method's practical potential could be established. The favourable performance of the tuner in these last series of tests was found to broadly correspond to that of its PI counterpart (Clarke, 2003) and will not be discussed here due to space restrictions.

A small selection of the results from the first two set of trials is presented below, together with summaries of the overall findings; the phase margin is set at the commonly used value of $PM = 60^\circ$.

4.1. Convergence to design-point

The following test procedure was adopted. Firstly, a step test was carried out on the open-loop plant from which the FODT model was obtained. Next, the algorithm was initialised and left to operate for a sufficiently long period for convergence to be attained. This was followed by step tests on both the initial CL, and that with the final, tuned, controller. Half-way through the step tests, the loop was subjected to a load disturbance of magnitude 0.5. (0.1 in the case of Type 1 plants).

As a benchmark for the tuning performance, the results obtained by the optimal PID design method of Panagopoulos, Åström, and Hägglund (2002) were used. We note however that the method requires full knowledge of the plant transfer function $G(s)$, is rather computationally demanding. Two sets of results of differing robustness (or maximum sensitivity[5] M_s) are given in Panagopoulos et al. (2002). We compare our tuning with the first, more robust set ($M_s = 1.4$), as they are found to correspond in robustness to our results.

The first example is the Type 1 process

$$IG_3(s) = \frac{1}{s(1+s)^3},$$

leading to the results of Fig. 6. The first, top-left, plot displays the progress of the adaptation in terms of the normalised PFE frequency $\omega_1(t)/\omega_1(0)$ (full line) and the

[5]$1/M_s$ is the shortest distance from Nyquist plot to the critical point.

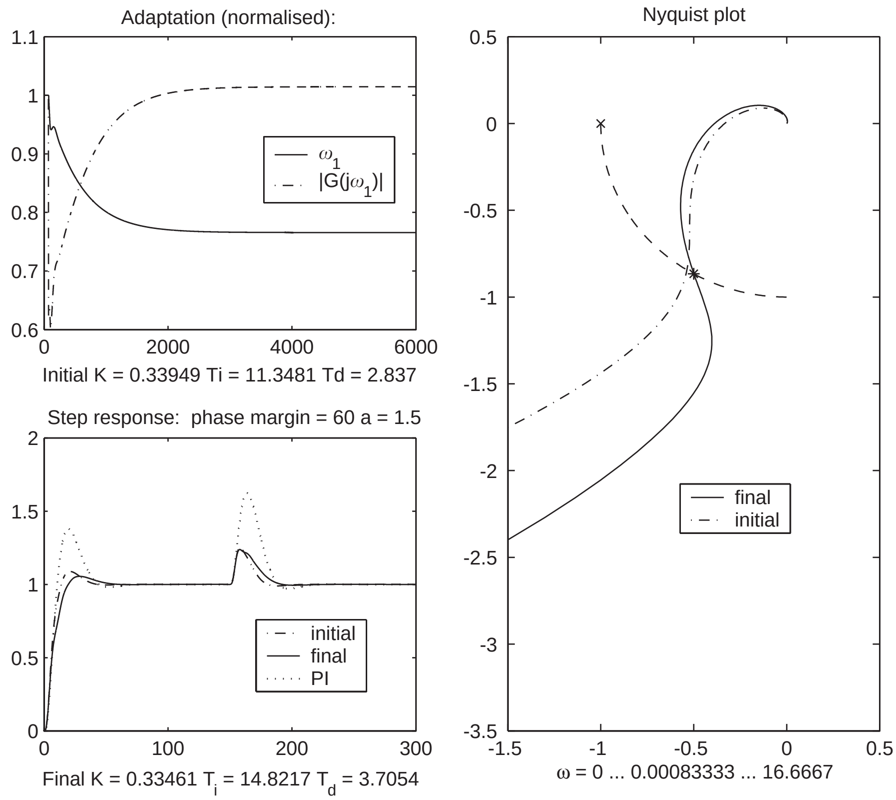

Fig. 6. Simulation results for $IG_3(s)$.

corresponding plant gain $|G(\omega_1)|(t)/|G(\omega_1)|(0)$ (dash-dot). Also shown are the initial controller parameters. To the right are the Nyquist plots for the open-loop system with the initial (dash-dot) and with the final, tuned (full line), controller. The design-point (star) and the critical point (cross), together with the phase margin line (dashed), are also indicated on this diagram. Finally, the bottom plot shows the initial and final CL step responses, together with the tuned controller parameters. Shown too is the tuning obtained by the PFE PI tuner of Clarke (2003) (with the desired damping factor set to $\zeta = 0.6$ so that a similar level of robustness is achieved as by the PFE PID) to assess the improvement gained through the extra derivative term.

Despite integrating plants not being represented well by FODT models, the initialisation is adequate and the tuner successfully converges to the design-point, albeit quite slowly. The final set-point and disturbance responses are almost identical to those in Panagopoulos et al. (2002). Compared to the PI tuning, we get a vastly reduced disturbance response and a much smaller overshoot to the set-point change.

Fig. 7 shows the operation of the tuner for the seventh-order plant:

$$G_7(s) = \frac{1}{(1+s)^7}.$$

For this plant, the initial PID deduced from the FODT model requires virtually no retuning. The reason for it is that the FODT model gives an excellent fit for high-order plants (and also for plants with significant dead-time). The tuned CL is again very similar to that in Panagopoulos et al. (2002). This time however, the performance improvement offered over the PI controller is more modest.

Remarks Overall, the simulations confirmed the effectiveness of the simple PID design approach implemented in the tuner, the tuned CL comparing very well in most cases to that obtained by the optimal design method of Panagopoulos et al. (2002). Notably, our design objective for the Nyquist contour to be situated close to $\mathrm{Re} = -0.5$ at low frequencies was achieved for all but two of the test cases in that article (the other exception in addition to IG_3 was a fourth-order plant dominated by a single pole). The following additional observations were made:

- Convergence takes about 10× the system's settling time T_s for Type 0 plants, though longer for Type 1 plants (by contrast in relay-based methods the experiment time is in the region of 2–4 × T_s).
- For many plants the initial tuning may be considered to be entirely acceptable (in terms of the step and disturbance responses).

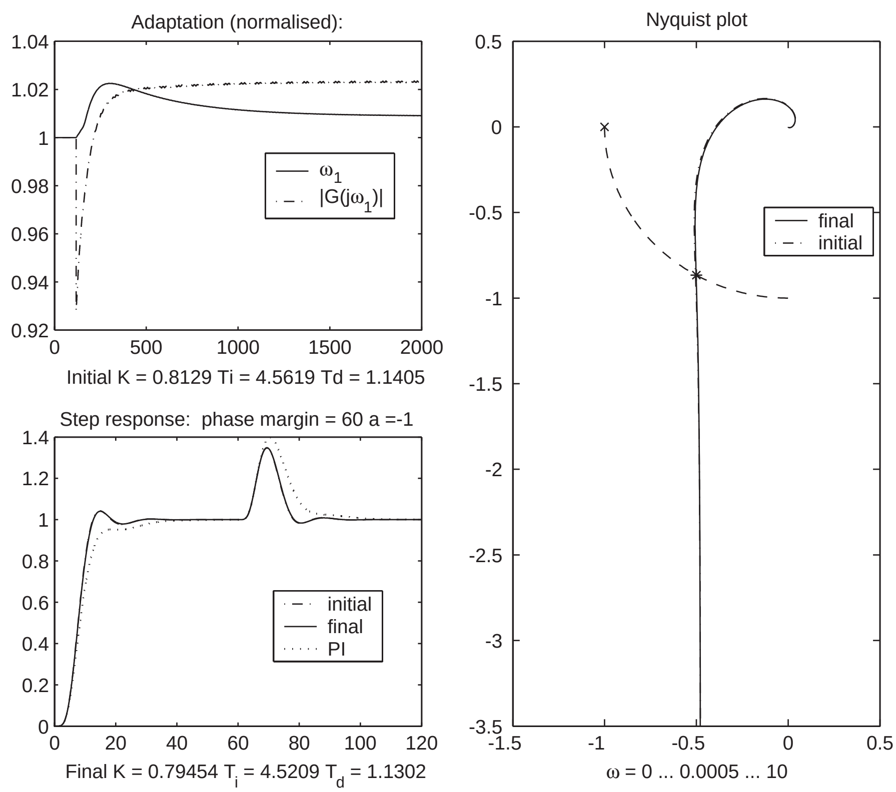

Fig. 7. Simulation results for $G_7(s)$.

- The tuner works with a wide range of plants. In fact, effective behaviour was obtained in all test cases representing typical process dynamics (including non-minimum phase and conditionally stable systems), and the tuner successfully converged for a series of underdamped test plants.
- The extent of CL performance gains over the PI method reflected the discussion in Hägglund and Åström (2002) on the relative merits of PI versus PID. Namely, PID control only offers substantial benefits in comparison to PI in the case of lag dominated (apparent dead-time $\lambda < 1$) processes (this includes Type 1 plants), while smaller improvements may also be achieved for balanced ($\lambda \approx 1$) processes.[6]

4.2. Comparison with other tuning rules

Many autotuning techniques, like that of Panagopoulos et al. (2002) considered above, rely on accurate process models or frequency response estimates to optimise the controller parameters (e.g. Sung, Lee, & Park, 2002; Grassi et al., 2001). However, some of the most common schemes in industry are (computationally less intensive) tuning rules that make use of simple characterisations of the plant (e.g. FODT model as obtained from a step test or frequency point from relay feedback). Well-known examples in the latter category include:

- The Ziegler and Nichols (1942) method (and its variants, e.g. Åström & Hägglund, 1984; Chien, Hrones, & Reswick, 1952). The rules give the controller settings as simple functions of the system ultimate data (plant gain K_u and frequency f_u corresponding to $\mathrm{Arg}\{G(\mathrm{j}\omega)\} = -\pi$). A time domain version, based on a two parameter description of the open-loop step response, also exists. The tuning attained is known to be rather inferior, partly due to the limitations in the ways the process is characterised and also because the rules were derived solely for a disturbance rejection criteria. The resulting CL systems tend to have poor damping and, in case of delay dominated (high λ) processes, a sluggish response.
- In the Cohen and Coon (1953) method the plant is described by an FODT model, but the tuning given is still rather oscillatory.
- Some more recent schemes ensure good robustness by using two frequency points (or a FODT model) to achieve (in some cases approximately) a prescribed phase margin and gain margin (Tan, Lee, & Wang,

[6]The reason behind this is related to the need for lowering the parameter a for higher λ plants (see Section 2.1). Essentially, the phase lead associated with derivative action cannot be used effectively with a high controller gain for speeding up the response (and improving the disturbance rejection) without compromising the stability margins.

1996; Ho, Hang, & Cao, 1995). However, the tuning is often less than favourable in other respects (e.g. speed of response) and a degree of arbitrariness exists regarding the choices of PM and GM (Tan et al., 1999).

- Internal model control or IMC (Rivera, Morari, & Skogestad, 1986). The approach uses an FODT model for the process and features a design parameter b representing the desired CL time constant. A first-order filter is added in series with the controller to smoothen the CL response.
- Kappa-Tau Tuning (Åström & Hägglund, 1995). The scheme is based on the system ultimate data K_u,f_u and the static gain K_p (though as in the case of Z–N, a step response version is also given). This time the design parameter is the desired maximum sensitivity M_s.
- The integral (time-weighted) square error and absolute error, abbreviated by I(T)SE and I(T)AE, are other widely used methods where the aim is to minimise a cost function of the error signal associated with a step change in set-point or disturbance. Traditionally, the process is approximated by a FODT model (e.g. Zhuang & Atherton, 1993).

Here we contrast the PFE tuner to the Kappa-Tau and IMC methods. In the former method, M_s was chosen to be 1.4, which corresponds (roughly) to the sensitivity resulting from PFE tuning (with the phase margin set to 60°). The design parameter b of the IMC method was set to the lowest recommended value (giving the fastest CL).

The comparative tests were carried out as follows. For each test plant, the PFE tuning algorithm was run and the final controller settings recorded. The FODT model obtained by the initialisation procedure was then used to compute PID parameters by the IMC method. Next a relay experiment was performed to provide the Kappa-Tau PID design. The CL (set-point and disturbance) responses obtained by the three methods were subsequently examined.

Before moving on to the results, it is important to note that the controller structure used by the PFE method differs from those in the IMC and Kappa-Tau designs. In the PFE approach, following good engineering practice, the derivative action is in the feedback path and is filtered (see Fig. 3), whereas the latter schemes implement the controller in standard parallel form in the forward path (though the IMC controller does feature a filter). The modified structure gives a slightly slower set-point response (and lower overshoot), so it is not possible (in the strict sense) to make exact comparisons between the set-point responses (though we can make such comparisons with the disturbance responses).

Figs. 8 and 9 show the results for FODT plants of differing dead-time: $\lambda = 0.5, 2$. It is seen that the PFE method gives consistent set-point responses, while in the Kappa-Tau tunings the overshoot varies markedly. Although IMC provides the fastest step response for both plants, the tuning is rather oscillatory and so the actual settling time is only quicker in the first case (the damping can be improved by raising the value of the design parameter b). The plots show similar disturbance responses for the IMC and PFE approaches, the disturbance rejection given by the Kappa-Tau method being inferior.

It must be noted that the fast (albeit oscillatory) tuning obtained by the IMC rules for low λ processes comes at the price of excessive control action. In the case of $\lambda = 0.1$, the peak of the control signal during a step input is about $40\times$ the step magnitude and in real life this would cause saturation and wear of the actuator (for the PFE tuning the ratio is around 2.2, which is far more acceptable). The problem can be remedied by increasing the IMC design parameter b but this slows down the response (and worsens the disturbance rejection).

Fig. 10 gives the results for the non-minimum phase plant:

$$NG = \frac{1-5s}{(1+10s)(1+20s)}.$$

This time the PFE approach results in a somewhat slower set-point response (though smaller overshoot) than that provided by the IMC and Kappa-Tau PID designs. Nevertheless, we get a well-behaved response and all three methods give a similar disturbance rejection.

Remarks The tuning performance of the PFE PID method was found to compare well to that of the Kappa-Tau and IMC rules. Although the method produces slower tuning than IMC in the case of low λ processes, it avoids excessive control activity. The main attribute of the PFE approach is the *consistent* set-point response it provides, much more so than the Kappa-Tau rules (and is designed to do so even if the value of λ is inaccurate). Of course, a large overshoot can always be reduced using set-point filters, or by raising b in the IMC case, but with the PFE tuner there is no need to be concerned about this. In addition, if a faster response is desired, one can increase the design parameter a or reduce the phase margin (e.g. to 45°), at the possible cost of inferior set-point behaviour (and robustness).

One additional point worth making is that design methods based on a step response experiment (such as IMC, ISE) are dependent on a good model being obtained for the plant. By contrast, the design rules incorporated into the PFE algorithm only require a rough initial value for λ for good subsequent tuning, and the desired phase margin is always exactly met. This does however come at the price of increased complexity with respect to one-shot methods.

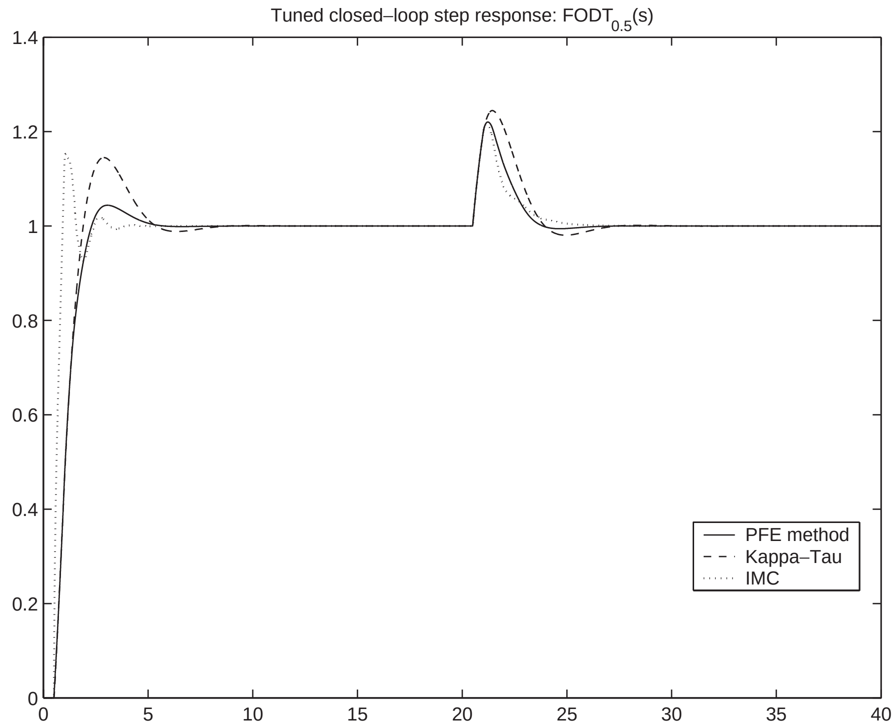

Fig. 8. Comparative results for FODT with $\lambda = 0.5$.

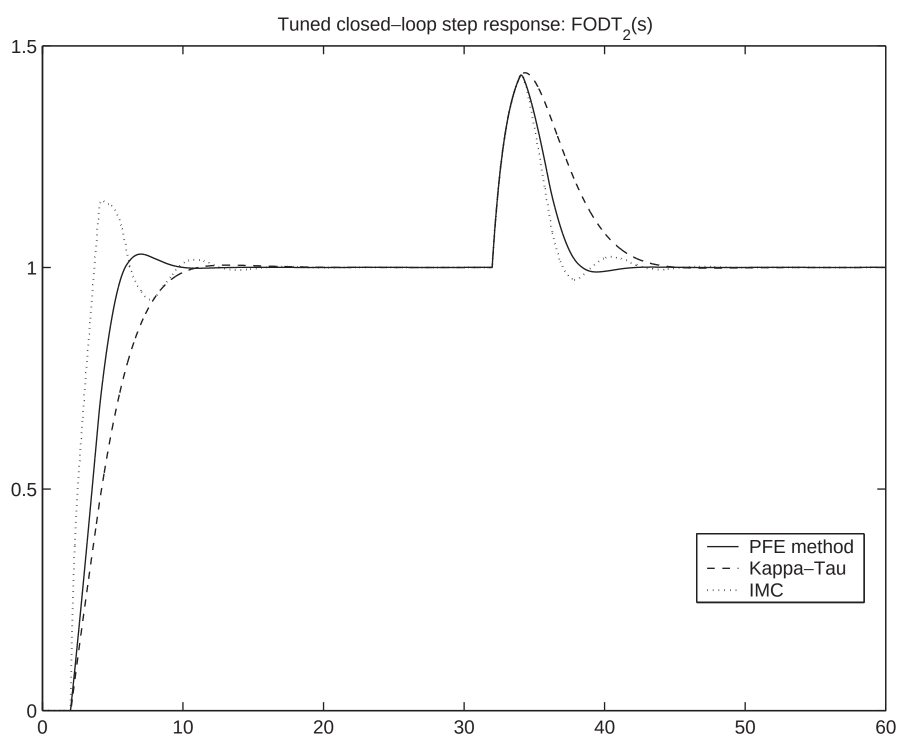

Fig. 9. Comparative results for FODT with $\lambda = 2$.

It is also interesting to compare the PFE approach to the recently proposed Iterative Feedback Tuning (IFT) method (Lequin et al., 2003). That scheme involves repeated CL step testing, the controller parameters being adapted iteratively such as to minimise a cost function based on the error and the control signal

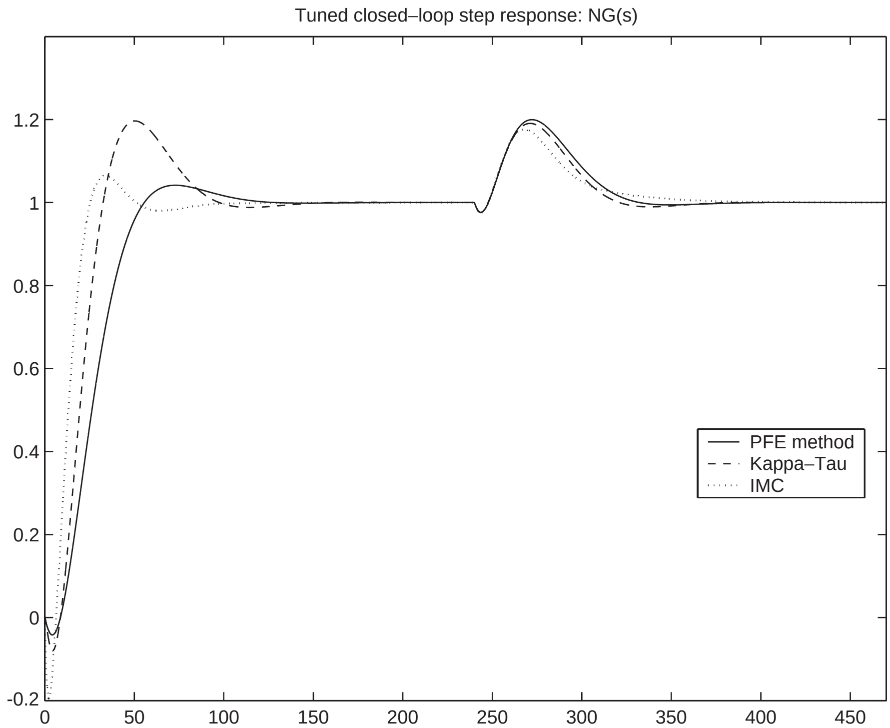

Fig. 10. Comparative results for $NG(s)$.

(no consideration is given to the disturbance response). It is shown to produce well tuned set-point responses, comparable to IMC with increased b. The method is especially effective for noisy plants due to the in-built noise-rejection objective. However, in its present form it requires a high degree of user intervention. For a start, it does not come with an automatic initialisation procedure (instead it requires the user to give initial values for the controller parameters). In addition, no specific guidelines are given on how the step length should be selected and varied, even though it is emphasised that these choices are important if optimal tuning is to be obtained.

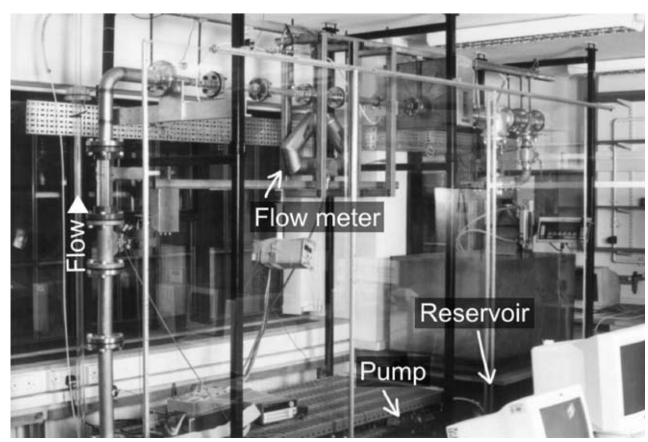

Fig. 11. Picture of the flow rig.

5. Trials on a flow rig

This section outlines the application of the algorithm to a real-world control problem: flow control in a flow rig. The system is pictured in Fig. 11. Water is driven round the rig by a variable speed pump (with an in-built controller so that the speed is proportional to the applied voltage); a Coriolis mass-flow meter provides the flow measurement.

Adaptive techniques prove to be useful because the plant dynamics are subject to signification variations, which may be classified into two types:

- *Predictable process variations*: Dynamics change with pump speed demand, so controller settings must be adjusted accordingly if effective flow control is to be ensured over the whole operating range.
- *Unpredictable*: Dynamics are also liable to vary over time, for instance due to wear and adjustments (or alterations) made to the rig.

The former type of variation is best addressed by running the tuner at several set-points and building up a gain schedule (Åström et al., 1993), a table specifying a set of controller parameters for each operating point. This subsequently allows the controller parameters to be instantaneously given suitable values as the set-point changes. To accommodate unpredictable process

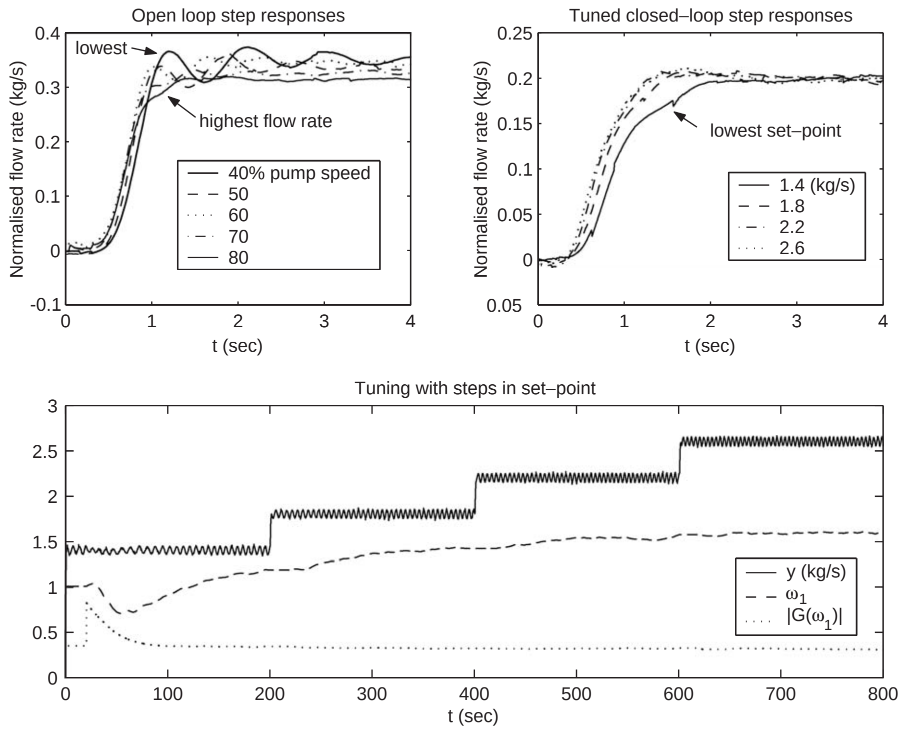

Fig. 12. Results on flow rig.

changes, gain scheduling can be combined with adaptive control (e.g. when set-point is constant, algorithm adapts the controller in the 'background' and updates the corresponding entry in the gain schedule).

The algorithm was set-up using the xPC Target environment from Mathworks. This involved the use of a 'Host' PC, containing a Simulink model of the algorithm. The model is converted into real-time code and transferred through the Ethernet network to a second, 'Target' machine, where it is executed. The latter is a dedicated PC (running a real-time OS), interfaced to the pump and the flow meter (the output of which is considerably noisy and must therefore be filtered). The set-point and key parameters of the algorithm may be adjusted on-line through the Host PC. The environment also allows data to be passed back from the Target PC (continuously or in a batch).

The top-left subplot of Fig. 12 shows the open-loop step response of the system at various pump speeds (superimposed on the same graph). We can see that the response varies quite significantly and is rather oscillatory at low pump speeds. This is actually due to the recent addition of two new arms to the flow rig (not shown in Fig. 11). When the arms are shut off (as in the tests), any trapped air leads to compression effects. In fact it turns out that at lower pump speeds the resulting dynamics can be approximated by a model involving a highly underdamped pole pair, a nearby complex zero pair plus a real pole (as well as a dead-time that is consistent over the various pump speeds and arises mostly from the response of the flow meter).

The oscillatory nature of the system makes for a stressful test platform, as the tuning rules incorporated into the algorithm were not designed with such dynamics in mind. Moreover, any induced oscillations could potentially swamp the GE and PFE.

The test procedure was as follows. The tuner was initialised using the open-loop step response in Fig. 12 corresponding to the lowest pump speed (the tuner parameters were unchanged from those used in the simulations). The set-point, initially relatively low, was increased in steps at fixed intervals and at each stage the tuned controlled settings were recorded. Next, the gain schedule was used to carry out CL step tests at the various set-points. Fig. 12 gives the results of the tuning. The bottom sub-plot displays the adaptation of the controller settings, together with the output flow rate. The top-right plot shows the tuned step-responses for the different set-points.

We note that despite the fact that the plant dynamics cannot be represented well by an FODT model, the initialisation is adequate. The tuner converges successfully at the initial set-point, and adapts well to the plant changes brought on by the subsequent steps in demand. Good flow control is maintained throughout the tuning test. The tuned CL step responses are consistent except the one at the lowest set-point, which shows a slow mode. The oscillations present in the open-loop plots have been largely damped out. It is argued that the results are quite favourable given the troublesome dynamics exhibited by the plant.

6. Conclusions

An alternative PID autotuning approach has been presented to the popular step response and relay-based methods. The approach involves the injection of a variable-frequency probing signal into the CL. Crucially, it differs from most existing methods in that the tuning is performed on-line, that is whilst the controller is undertaking CL control. As a result, it cannot only provide single-shot autotuning but also, subsequently, continuous adaptation of the controller. Ease-of-use is ensured by a semi-automatic initialisation procedure, which employs the results of a prior step-test. Consequently, in many cases satisfactory tuning can be obtained straight away.

The tuner appears to produce uniformly good results for a wide range of typical process dynamics. Compared to standard design rules, the tuning is characterised by a consistent set-point response and is similar to that achieved by 'optimal' PID design (which requires knowledge of the plant transfer function). The approach compares well in speed and simplicity with the relay experiment. In addition, it is able to converge and track plant changes even with the probing signal 'buried' in noise. The real-life results produced by trials on a non-linear flow rig are also encouraging.

There is significant room for the further exploration of the tuning technique. For instance, multiple frequency points in the CL transfer function could be monitored using a single, square wave test signal. Appropriate tuning rules could be devised. In terms of enhancing the existing tuner, recent research suggests that the dead-time incurred in observing the phase and amplitude of the probing sinewave can be circumvented. This would allow for faster rates of convergence, perhaps matching that of the relay method. Moreover, to establish fully the industrial potential of the tuner, in particular its ability to handle actuator non-linearities, more trials should be carried out on practical systems.

References

Åström, K. J., & Hägglund, T. (1984). Automatic tuning of simple regulators with specifications on phase and amplitude margins. *Automatica*, *20*(5), 645–651.

Åström, K. J., & Hägglund, T. (1995). *PID control—theory, design and tuning*. Research Triangle Park, NC: Springer-Verlag.

Åström, K. J., Hägglund, T., Hang, C. C., & Ho, W. K. (1993). Automatic tuning and adaptation for PID controllers—a survey. *Control Engineering Practice*, *1*(4), 699–714.

Barker, H. A., & Godfrey, K. R. (1999). System identification with multi-level periodic perturbation signals. *Control Engineering Practice*, *7*, 717–726.

Braun, M. W., Ortiz-Mojica, R., & Rivera, D. E. (2001). Application of minimum crest factor multisinusoidal signals for 'plant-friendly' identification of nonlinear process systems. *Control Engineering Practice*, *101*, 301–313.

Chien, K. L., Hrones, J. A., & Reswick, J. B. (1952). On the automatic control of generalized passive systems. *Transactions ASME*, *64*, 759–768.

Clarke, D. W. (1984). PID algorithms and their computer implementation. *Transactions of the Institute of Measurement and Control*, *6*(6), 305–316.

Clarke, D. W. (2003). Pretuning and adaptation of PI controllers. *IEE Proceedings, Control Theory Applications*, *150*(6), 585–598.

Clarke, D. W., Park, J. W. (2002). *Initialising an adaptive phase/frequency estimator*. Technical report, Department of Engineering Science, Oxford University.

Clarke, D. W., & Park, J. W. (2003). Phase-locked loops for plant tuning and monitoring. *IEE Proceedings, Control Theory Applications*, *150*(2), 155–169.

Cohen, G. H., & Coon, G. A. (1953). Theoretical considerations of retarded control. *Transactions ASME*, *75*, 827–834.

Crowe, J., & Johnson, M. A. (2002). Towards autonomous PI control satisfying classical robustness specifications. *IEE Proceedings, Control Theory Applications*, *149*(1), 26–31.

Glattfelder, A. H. (1969). *Entwurf und Erprobung eines Verfahrens zur geregelten Adaptierung eines Regelkreises mit simultanen harmonischen Pruefsignalen*. PhD thesis, diss. no. 4275, ETH, Zurich.

Grassi, E., Tsakalis, K. S., Dash, S., Gaikwad, S. V., MacArthur, W., & Stein, G. (2001). Integrated system identification and PID controller tuning by frequency loop-shaping. *Transactions IEEE, Control Systems Technology*, *9*(2), 285–294.

Gyöngy, I. J. (2003). *Adaptive control by tracking critical frequencies*. Technical Report, Department of Engineering Science, Oxford University.

Hägglund, T., & Åström, K. J. (2000). Supervision of adaptive control algorithms. *Automatica*, *36*, 1171–1180.

Hägglund, T., & Åström, K. J. (2002). Revisiting the Ziegler–Nichols tuning rules for PI control. *Asian Journal of Control*, *4*(4), 364–380.

Hang, C. C., Åström, K. J., & Wang, Q. G. (2002). Relay feedback auto-tuning of process controllers—a tutorial overview. *Journal of Process Control*, *12*(1), 143–162.

Hang, C. C., Lee, T. H., & Ho, W. K. (1995). *Adaptive control*. North Carolina: Instrument Society of America, Research Triangle Park.

Hind, E. C. (1980). Controller selection and tuning. *Transactions of the Institute of Measurement and Control*, *2*(1), 46–56.

Ho, W. K., Hang, C. C., & Cao, L. S. (1995). Tuning of PID controllers based on gain and phase margin specifications. *Automatica*, *31*, 497–502.

Ho, W. K., Hong, Y., Hansson, A., Hjalmarsson, H., & Deng, J. W. (2003). Relay feedback auto-tuning of PID controllers using iterative feedback tuning. *Automatica*, *39*(1), 149–157.

Huang, H.-P., Roan, M.-L., & Jeng, J.-C. (2002). On-line adaptive tuning for PID controllers. *IEE Proceedings, Control Theory Applications*, *149*(1), 60–67.

Lequin, O., Gevers, M., Mossberg, M., Bosmans, E., & Triest, L. (2003). Iterative feedback tuning of PID parameters: comparison with classical tuning rules. *Control Engineering Practice*, *11*, 1023–1033.

Panagopoulos, H., Åström, K. J., & Hägglund, T. (2002). Design of PID controllers based on constrained optimisation. *IEE Proceedings, Control Theory Applications*, *149*(1), 32–40.

Rivera, D. E., Morari, M., & Skogestad, S. (1986). Internal model control. 4 PID controller design. *Industrial Engineering and Chemistry Research*, *25*(1), 252–265.

Smyth, R., & Nahi, N. (1963). Phase and amplitude sinusoidal dither adaptive control system. *Transactions IEEE, Automatic Control*, *8*(4), 311–321.

Sung, S. W., & Lee, I.-B. (2000). An improved algorithm for automatic tuning of PID controllers. *Chemical Engineering Science*, *55*, 1883–1891.

Sung, S. W., Lee, T.-Y., & Park, S. (2002). Optimal PID controller tuning method for single-input/single-output processes. *AIChE Journal*, *48*(6), 1358–1361.

Tan, K. K., Ferdous, R., & Huang, S. (2002). Closed-loop automatic tuning of PID controller for non-linear systems. *Chemical Engineering Science*, *57*, 3005–3011.

Tan, K. K., Lee, T. H., & Jiang, X. (2001). On-line relay identification, assessment and tuning of PID controller. *Journal of Process Control*, *11*, 483–496.

Tan, K. K., Lee, T. H., & Wang, Q. G. (1996). An enhanced automatic tuning procedure for PI/PID controllers for process control. *AIChE Journal*, *42*(9), 2555–2562.

Tan, K. K., Wang, Q.-G., Hang, C. C., & Hägglund, T. (1999). *Advances in PID control*. London: Springer.

Vasu, G. (1957). Optimalizing control applied to control of engine pressure. *Transactions ASME*, *79*, 481–488.

Yu, C. C. (1999). *Autotuning of PID controllers*. London: Springer.

Zhuang, M., & Atherton, D. P. (1993). Automatic tuning of optimal PID controllers. *IEE Paroceedings D*, *140*(3), 216–224.

Ziegler, J. G., & Nichols, N. B. (1942). Optimum settings for automatic controllers. *Transactions ASME*, *64*, 759–768.