PERGAMON

Control Engineering Practice 9 (2001) 1163–1175

CONTROL ENGINEERING PRACTICE

www.elsevier.com/locate/conengprac

The future of PID control

K.J. Åström*, T. Hägglund

Department of Automatic Control, Lund University of Technology, Box 118, S-22100, Lund, Sweden

Received 6 April 2001; accepted 6 April 2001

Abstract

This paper presents the state of the art of PID control and reflects on its future. Particular issues discussed include specifications, stability, design, applications, and performance of PID control. The paper ends with a discussion of alternatives to PID and its future.

Keywords: PID control; Stability; Performance; Application; Design; Tuning

1. Introduction

Feedback is a very powerful idea. Its use has often had revolutionary consequences with drastic improvements in performance, see e.g. Bennett, (1979, 1993). Credit is often given to a particular form of feedback although it is frequently feedback itself that gives the real benefits and the particular form of feedback used is largely irrelevant. The PID controller is by far the most dominating form of feedback in use today. More than 90% of all control loops are PID. Most loops are in fact PI because derivative action is not used very often. Integral, proportional and derivative feedback is based on the past (I), present (P) and future (D) control error. It is surprising how much can be achieved with such a simple strategy. A strength of the PID controller is that it also deals with important practical issues such as actuator saturation and integrator windup. The PID controller is thus the bread and butter of automatic control. It is the first solution that should be tried when feedback is used.

The PID controller is used for a wide range of problems: process control, motor drives, magnetic and optic memories, automotive, flight control, instrumentation, etc. The controller comes in many different forms, as standard single-loop controllers, as a software component in programmable logic controllers and distributed control systems, as a built in controller in robots and CD players.

Although the PID controller has always been very important, practically it has only received moderate interest from theoreticians. Therefore, many important issues have not been well documented in the literature. A result of this is that many mistakes have been repeated when technology shifted from pneumatic, via electrical to digital. There has, however, been an increased interest in the last ten years. One reason is the emergence of automatic tuning, another is the increased use of model predictive control which requires well tuned PID controllers at the basic level. Still most papers on single loop control use PID controllers with Ziegler–Nichols tuning as a benchmark. This is a very unsatisfactory situation because the Ziegler–Nichols rules are known to give very poor results in many cases.

This paper treats the future of PID control. It will try to answer questions like: Will the PID controller continue to be used or will it be replaced by other forms of feedback? What additional features are desirable in a PID controller? Are there any essential research issues in PID control? Our prediction will be based on a consideration of the current state of research and practice in PID control.

2. Specifications

Before going into details it is important to be aware that there is a wide range of control problems. A few examples are given below.

- Design a robust controller that keeps process variables reasonably close to desired values without strong demands on specifications.

*Corresponding author. Tel:.00-46-461-08-781; fax: 00-46-461-38-118.

E-mail address: karl-johan.astrom@control.lth.se (K.J. Åström).

0967-0661/01/$ - see front matter
PII: S0967-0661(01)00062-4

- Design a controller that keeps process variables as close as possible to desired specifications.
- Design a controller where the process variables can follow variations in set points.
- Design a controller that keeps process variables within a range.

The last situation is typical for level control in surge tanks where it is desired that the level changes but it is not permitted either to have the tanks flooded or to have them empty.

Another factor that has a strong influence is the effort that can be devoted to design and tuning of a system. One extreme case is found in process control where one process engineer may be responsible for several hundred loops. In such a case it is not possible to devote much effort to each loop. Simplicity of handling and robustness are then primary requirements. Another case is a dedicated system that is manufactured in large quantities, for example, a feedback loop in a CD player. In this case it is possible to devote a substantial effort to design a single control loop.

2.1. Formalization

To describe a design problem the process, the environment, and the requirements on the control have to be characterized. A typical situation is illustrated in Fig. 1. The process is described as a linear system with transfer function $G(s)$. The controller is also linear with two degrees of freedom. The transfer function $G_c(s)$ describes the feedback from process output y to control signal u, and the transfer function $G_{ff}(s)$ describes the feed forward from set point y_{sp} to u. For PID control, one typically has

$$G_c(s) = k + \frac{k_i}{s} + k_d s,$$

$$G_{ff}(s) = bk + \frac{k_i}{s} + ck_d s. \tag{1}$$

which means that the input–output relation for the controller can be described as

$$\begin{aligned} u(t) = {} & k(by_{sp}(t) - y(t)) \\ & + k_i \int_0^t (y_{sp}(\tau) - y(\tau))\, d\tau \\ & + k_d \left(c\frac{dy_{sp}(t)}{dt} - \frac{dy(t)}{dt} \right), \end{aligned} \tag{2}$$

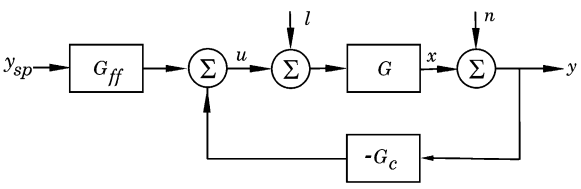

Fig. 1. A block diagram describing a typical control problem.

Notice that a real PID controller uses a filtered derivative dy_f/dt instead of dy/dt where

$$\frac{T_d}{N}\frac{dy_f}{dt} + y_f = y.$$

Additional filters may also be used. The controller should make sure that the integral part does not wind up when the actuator saturates. This is discussed in great detail in Åström and Hägglund (1995).

It is essential that the controller is implemented as (2) with one integrator only. Parameters b and c are called set-point weights. They have no influence on the response to disturbances but they have a significant influence on the response to set-point changes. Set-point weighting is a simple way to obtain a structure with two degrees of freedom (Horowitz, 1963). It is also worthwhile to observe that the so-called PI-PD controller (Atherton, 1999) is equivalent to set-point weighting. See Taguchi and Araki (2000).

Three external signals act on the control loop, namely set point y_{sp}, load disturbance l and measurement noise n. The load disturbance drives the process variables away from their desired values and the measurement noise corrupts the information obtained from the sensors.

The design objective is to determine the controller parameters in $G_c(s)$ and $G_{ff}(s)$ so that the system behaves well. This means that the effect of load disturbances should be reduced, that too much measurement noise should not be fed into the system and that the system should be robust towards moderate changes in the process characteristics. Therefore, the specification will express requirements on

- Load disturbance response,
- Measurement noise response,
- Set point response,
- Robustness to model uncertainties.

The relations between the external input signals, y_{sp}, l and n, process output x measured signal y and control signal u are

$$\begin{aligned} x &= \frac{G}{1+GG_c}l - \frac{GG_c}{1+GG_c}n + \frac{GG_{ff}}{1+GG_c}y_{sp}, \\ y &= \frac{G}{1+GG_c}l + \frac{1}{1+GG_c}n + \frac{GG_{ff}}{1+GG_c}y_{sp}, \\ u &= -\frac{GG_c}{1+GG_c}l - \frac{G_c}{1+GG_c}n + \frac{G_{ff}}{1+GG_c}y_{sp} \end{aligned} \tag{3}$$

2.2. Criteria

Regulation performance is often of primary importance since most PID controllers operate as regulators. Regulation performance is often expressed in terms of

the control error obtained for certain disturbances. A disturbance is typically applied at the process input. Typical criteria are to minimize a loss function of the form

$$I = \int_0^\infty t^n |e(t)|^m \, dt, \tag{4}$$

where the error is defined as $e = y_{sp} - y$. Common cases are IAE ($n = 0$, $m = 1$) or ITSE ($n = 1$, $m = 2$). Quadratic criteria are particularly popular since they admit analytical solutions. Criteria which put a penalty on the control signal such as

$$I = \int_0^\infty (e^2(t) + \rho u^2(t)) \, dt \tag{5}$$

have also been used.

2.3. Robustness

It turns out that it is not sufficient to minimize criteria such as (4) or (5) because the solutions obtained may not be sufficiently robust to uncertainties in the process model. Excellent insight into the robustness problem has been provided by the H_∞ theory, see Panagopoulos and Åström (2000). According to this theory it follows that the closed-loop system is robust to perturbations if the H_∞-norm of the transfer function

$$\frac{1}{1 + GG_c} \begin{pmatrix} 1 & G \\ G_c & GG_c \end{pmatrix},$$

is sufficiently small. The norm is given by

$$\gamma = \max \frac{1 + |G(i\omega)G_c(i\omega)|}{|1 + G_c(i\omega)G_c(i\omega)|}. \tag{6}$$

In Panagopoulos and Åström (2000) it is shown that the norm is always less than M if the Nyquist curve of the loop transfer function of GG_c is outside a circle with diameter on the interval $-M/(M-1), -M/(M+1)$.

3. Stability regions

Instability is the disadvantage of feedback. When using feedback there is always a risk that the closed-loop system will become unstable. Stability is therefore a primary requirement on a feedback system. It turns out that much insight into PID control can be obtained by analyzing the stability region, which is the set of controller parameters that give stable closed-loop systems. This is also an area where there are interesting recent results, see Shafiei and Shenton (1994), Shenton and Shafiei (1994) and Anon (1999). These results are based either on the Hermite–Bieler Theorem, see Ho, Datta and Bhattacharyya (1996, 1997), or on elaborate polynomial calculations, see Munro, Söylemez and Baki (1999) and Söylemez, Munro and Baki (1999). In this section, it will be shown that the results can be obtained from the Nyquist stability theorem. An advantage with this approach is that time delays are easy to deal with.

3.1. Constant derivative gain

Consider linear systems with the transfer function $G(s)$. It is assumed that the transfer function has no poles in the right half-plane apart from possibly a pole at the origin. Furthermore, it is assumed that $G(0) > 0$ if the system is stable or that $\lim_{s\to 0} sG(s) > 0$ when the system has a pole at the origin. It is also assumed that the controller is parameterized in terms of k, k_i and k_d. The closed-loop characteristic equation is then

$$1 + \left(k + \frac{k_i}{s} + k_d s\right) G(s) = 0. \tag{7}$$

By analyzing this equation for small s it is seen that $k_i > 0$ to have a stable system. Next, conditions for roots on the imaginary axis will be investigated. This will represent a stability boundary. Introducing

$$G(i\omega) = r(\omega) e^{i\phi(\omega)},$$

after some calculations gives the following boundary of the stability region:

$$k = -\frac{\cos\phi(\omega)}{r(\omega)},$$

$$k_i = \omega^2 k_d - \frac{\omega \sin\phi(\omega)}{r(\omega)}. \tag{8}$$

Fig. 2 shows the stability region for a system with the transfer function $G(s) = 1/(s+1)^4$ for different values of k_d. The figure shows that in this particular case the integral gain can be increased substantially by using derivative action. The figure also shows that the stable

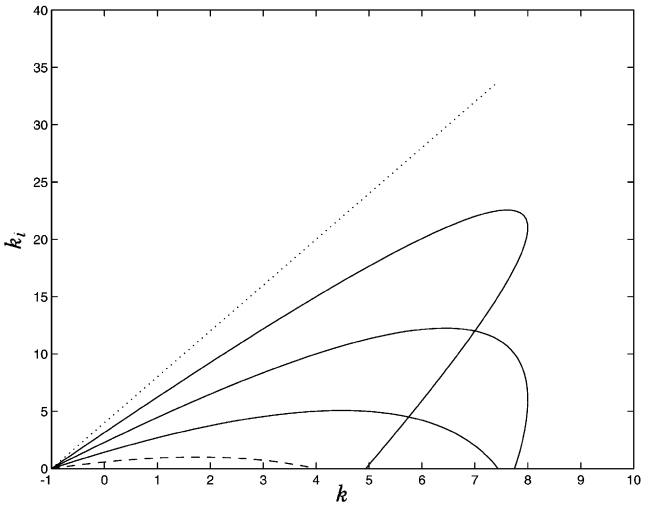

Fig. 2. Stability regions for $G(s) = 1/(s+1)^4$ for $k_d = 0$ (dashed), 5, 10, 15 and 20 (dotted).

controller that maximizes the integral gain has parameters which lie in a sharp corner. For systems with monotone transfer functions the stability region is a convex set but for other transfer functions the stability region may consist of several disjoint sets.

The robustness region lies inside the stability region and has a similar shape. This state of affairs is one explanation why it is difficult to find appropriate values of the derivative action. Other constraints can be introduced to deal with the problem, see Panagopoulos (2000). Another possibility is to postulate a given value of the ratio of integral time to derivative time. Classically it was quite common to postulate $T_i = 4T_d$. Postulating that $T_i = mT_d$ it is found that $k^2 = k_i k_d/m$. Inserting this in Eq. (8) gives

$$k = -\frac{\cos\phi(\omega)}{r(\omega)},$$

$$k_i = \frac{\sqrt{1 + (4/m - 1)\cos^2\phi(\omega)} - \omega\sin\phi(\omega)}{2r(\omega)} \tag{9}$$

Fig. 3 shows the stability regions for the system $G(s) = 1/(s+1)^4$ and PID controllers with $m = 4$, 6.25 and 10. The figure shows clearly that the stability region may be increased by using derivative action. But it also shows that the stability region becomes quite sharp when the parameter m is small. A comparison with Fig. 2 shows that the largest value of integral gain k_i is substantially reduced by the requirement that $T_i = 4T_d$.

3.2. Constant proportional gain

The stability region is a subset of R^3. The calculations performed gives the two-dimensional intersections with constant derivative gain. Additional insight can be obtained from another representation of the stability regions. To investigate the stability, the Nyquist criterion is used and the locus of the loop transfer function $L(i\omega) = G(i\omega)G_c(i\omega)$ is plotted. First a fixed value $k > 0$ of the proportional gain is picked and the frequencies ω_n, where the Nyquist curve of $kG(i\omega)$ intersects the circle with the line segment $(-1, 0)$ as a diameter are determined. See Fig. 4, where the critical point $s = -1$ is denoted by C. First, the case when the intersection of the Nyquist curve and the circle occur in the lower half-plane is considered. The controller transfer function can be written as

$$G_c(i\omega) = k + i\left(-\frac{k_i}{\omega} + k_d\omega\right) = k - i\left(\frac{k_i}{\omega} - k_d\omega\right).$$

When proportional and derivative gains are changed the Nyquist curve moves from A along the line AC. To avoid reaching the critical point it must be required that

$$\left(\frac{k_i}{\omega_n} - k_d\omega_n\right)|G(i\omega_n)| < |1 + G(i\omega_n)|. \tag{10}$$

The same analysis can be made when the intersection of the Nyquist curve and the circle occur in the upper half-plane. To have a stable system the point A must then be moved beyond the critical point. Combining the conditions give the stability regions

$$k_i > 0,$$

$$k_i < \omega_n^2 k_d + \omega_n \frac{|1 + kG(i\omega_n)|}{|G(i\,\omega_n)|} \text{ for Im } G(i\omega_n) < 0,$$

$$k_i > \omega_n^2 k_d - \omega_n \frac{|1 + kG(i\omega_n)|}{|G(i\omega_n)|} \text{ for Im } G(i\omega_n) > 0 \tag{11}$$

for all ω_n such that

$$\left|kG(i\omega_n) + \tfrac{1}{2}\right| = \tfrac{1}{2}. \tag{12}$$

It is thus concluded that for constant proportional gain, the stability region is represented by several convex polygons in the k_i–k_d plane. In general, there may be

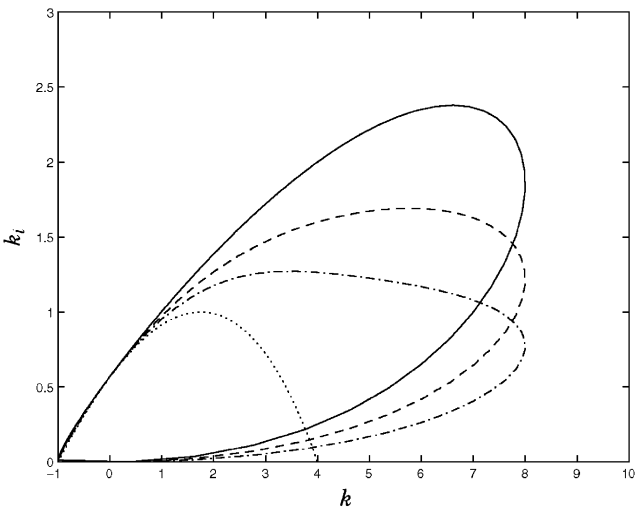

Fig. 3. Stability regions for $G(s) = 1/(s+1)^4$ and controllers with $T_d = 0$ (dotted), and $T_i = mT_d$, where $m = 10$ (dash–dotted), 6.25 (dashed) and 4 (solid). The stability region for pure PI control corresponds to the dotted curve.

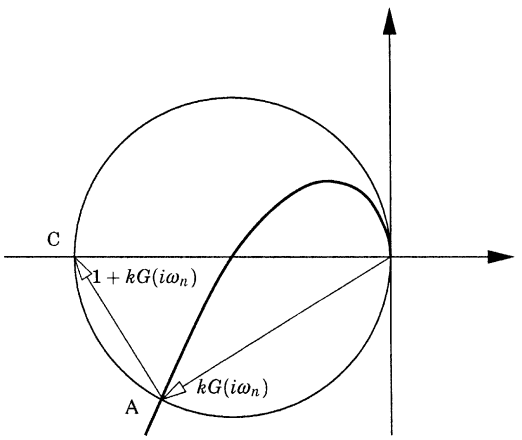

Fig. 4. Nyquist curve for the transfer function $kG(s)$.

several polygons and each may have many surfaces. The number of surfaces of the polygons is determined by the number of roots of Eq. (12). In many cases, the polygons are also very simple, as is illustrated with the following example.

Example 1 (Four equal poles). Consider a system with the transfer function

$$G(s) = \frac{1}{(s+1)^4} = \frac{1}{s^4 + 4s^3 + 6s^2 + 4s + 1} = \frac{1}{s^4 + 6s^2 + 1 + 4s(s^2+1)}.$$

In this case Eq. (12) becomes

$$\omega^4 - 6\omega^2 + 1 + k = 0.$$

This equation has only two positive solutions

$$\omega^2 = 3 \pm \sqrt{8-k}$$

and it is thus concluded that the stability region is determined by the condition $k_i > 0$ and two lines. Simple calculations show that the stability conditions (11) become

$$\begin{aligned} k_i &< (3 - \sqrt{8-k})k_d + 4k - 56 + 20\sqrt{8-k}, \\ k_i &> (3 + \sqrt{8-k})k_d + 4k - 56 - 20\sqrt{8-k}, \\ k_i &> 0. \end{aligned} \tag{13}$$

The integral gain has its maximum $k_i = 36$ at the boundary of the stability region for $k = 8$ and $k_d = 20$. Notice that the stability region is needle shaped near this point. This is more clear in the 3D plot of the stability region shown in Fig. 5.

Analyzing systems with time delay it follows from the equations that if $G(s)$ approaches a constant for large s then stability requires that $k_d = 0$. If $\lim sG(s) \to \infty = K_v$ then the integral gain is limited to $|k_d| < 1/K_v$.

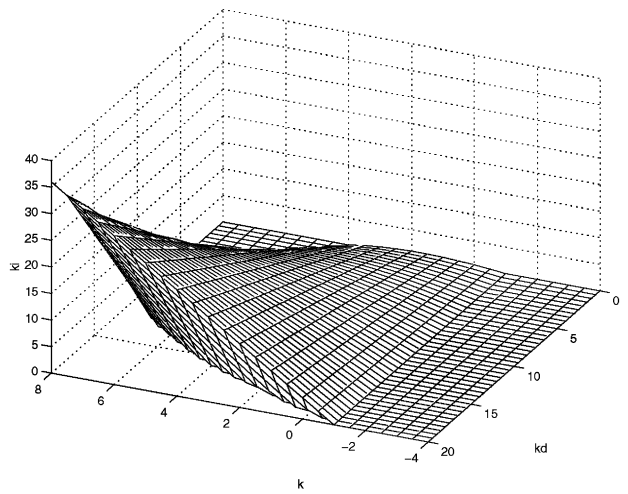

Fig. 5. The 3D plot of the stability region for the system $G(s) = (1+s)^{-4}$.

4. Design and tuning

Design and tuning of PID controllers have been a large research area ever since Ziegler and Nichols presented their methods in 1942. There are many aspects that should be taken into account when designing a PID controller. Desirable features of a design procedure are:

- It should give a controller that meets the design specifications.
- It should be based on the available/obtainable process knowledge.
- It should meet limitations on computational power and resources available for design.

Therefore, there is a need for several different design procedures with varying objectives and complexity. Some recent results are Zhuang and Atherton (1993), Lieslehto (1999), Anon (1999), Yamamoto, Fujii and Kaneda (1999), Panagopoulos (2000) and Wallen (2000).

4.1. Specifications for design

The design specifications were discussed in Section 2. Basically, one has to take care of specifications on responses to the three external signals y_{sp}, l, and n, as well as robustness with respect to changes in the process G.

Since the controller has two degrees of freedom, it is possible to separate the specifications. The response to measurement noise can be treated by designing a low-pass filter for the measurement signal. The desired set-point response can be obtained by feedforward. One way is to choose a proper value of the set-point weight b in Eq. (2). It is also possible to feed the set point through filters or ramping modules.

Load disturbances are often the most common disturbances in process control. See Shinskey (1996). Most design methods should therefore focus on load disturbances, and try to find a suitable compromise between demands on performance at load disturbances and robustness. It is a great advantage if this compromise can be decided by the user by a tuning parameter. There are two types of tuning parameters, those where the robustness is specified, and those where the performance is specified. Gain and phase margins are often used as robustness-related design parameters. See Tan, Wang, Hang, and Hägglund (1999). In Åström, Panagopoulos, and Hägglund (1998) and Panagopoulos, Åström and Hägglund (1999), the maximum sensitivity and the complementary sensitivity are used. In Panagopoulos and Åström (2000) the H_∞ norm is

used. A common performance-related tuning parameter is the apparent closed-loop time constant. It is used in lambda-tuning and in many other analytical tuning methods.

The design procedure also includes selection of controller structure. For example, if there are no requirements on zero steady-state error, there is no need for integral action. If the dead-time is long or if the process dynamics is close to first order, no derivative action should be used.

The design procedure should also take limitations in signals into account. There are always limitation in the control signal, and often also in the rate of the control signal.

In Åström et al. (1998) it is shown that PI control can be very well captured so as to maximize integral gain k_i, which is equivalent to minimizing the integrated error IE at load disturbances, subject to a robustness constraint. The reason is that the parameters k and k_i that satisfy the robustness constraint is a good convex set. This does not work very well for PID control because the set of parameters k, k_i and k_d that satisfy the robustness constraint can have narrow ridges. See Panagopoulos (2000). This is illustrated in Fig. 5 which shows the stability region for a problem that appears very well behaved.

The design method given in Åström et al. (1998) can also be used to evaluate simpler design methods. Fig. 6 shows controller parameters obtained for a large test batch of process models. The figure also shows the Ziegler–Nichols design. The controller gain is normalized with the ultimate gain, and the integral time is normalized with the ultimate period of the process. The parameters are plotted versus the relative gain

$$\kappa = \frac{|G(i\omega_{180})|}{G(0)} = \frac{1}{K_p K_u}. \tag{14}$$

Results are given for two values of the design parameter M_s (the maximum sensitivity gain), $M_s = 1.4$ and $M_s = 2$.

The figure shows that the Ziegler–Nichols method gives too high a gain and for most processes an integral time that is too long. It also shows that it is impossible to obtain good tuning rules that are based only on two process parameters, but that fairly accurate methods may be obtained using three parameters, where the relative gain κ may be a third parameter.

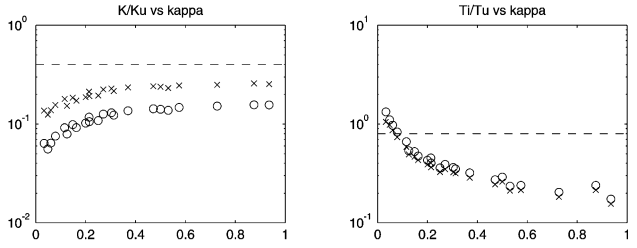

Fig. 6. Normalized controller parameters plotted versus relative gain κ for $M_s = 1.4$ (circles) and $M_s = 2$ (crosses). The Ziegler–Nichols rule is shown in dashed lines

4.2. Process knowledge

The process knowledge required by design procedures varies from simple characteristics described by two parameters (a gain and a time) as in the Ziegler–Nichols rules, to higher order transfer functions.

In process control, there is normally little time available to derive detailed models of the processes. The early design methods were therefore based on very simple models. Nowadays, there are often equipments available that makes it possible to obtain the models fast and simple. In automatic tuning procedures, the process models are obtained fast with very little effort from the user.

Many design methods assume that the process is known completely in terms of step responses, frequency responses or transfer functions. Actually, it is sufficient to know the frequency response in a fairly narrow frequency range. For the design method in Panagopoulos (2000) the transfer function should be known in a region around ω_{pi}^* and ω_{pd}^*, where

$$\arg G(i\omega_{pi}^*) = -\pi + \arcsin\frac{1}{M_s},$$

$$\arg G(i\omega_{pd}^*) = -\pi - \arccos\frac{1}{M_s}.$$

4.3. Computational aspects

The equipment, the computational power, and the time available for the design are also important aspects that must be taken into account. The early methods were, for obvious reasons, based on simple registrations of process characteristics and calculations that could easily be performed by hand. There is still an interest in these methods for fast controller settings. These methods are also useful for the understanding and intuition needed for manual controller adjustments.

Nowadays, there is often a computational power available that admits more advanced process identification as well as controller design. In automatic tuning procedures, the identification and design are both made automatically. The process models used in these procedures are normally still rather simple. The design procedures are, however, often much more sophisticated. It is, for example, possible to use optimization routines to obtain optimal solutions stated by the design criteria given in Section 2.

5. Performance assessment

It is useful to have tools to make an assessment of the performance that can be achieved and the factors that limit the performance. This is a research topic that has recently attracted much attention, see Seron, Braslavsky and Goodwin, (1997) and Åström (2000).

For a control system the achievable performance is typically limited by

- Process dynamics.
- Nonlinearities.
- Uncertainties.
- Disturbances.

It is essential to be aware of the factors that are crucial for a specific application. Most of the conventional design methods for PID control focus on process dynamics and uncertainties.

5.1. A preliminary assessment

To make a preliminary assessment, the gain crossover frequency ω_{gc} will be used to characterize performance. Let φ_m be the desired phase margin, the crossover frequency is the smallest frequency such that

$$\arg G_p(i\omega_{gc}) + \arg G_c(i\omega_{gc}) = -\pi + \varphi_m. \quad (15)$$

A PD controller has a maximum phase lead of about 60 degrees, a proportional controller has zero phase lag, and a PI controller has a phase-lag of about 45 degrees and a PID controller can have a phase lead of 45 degrees. If a phase margin of 45 degrees is desired it follows from Eq. (15) that crossover frequencies for PI, PID and PD control are the frequencies where the process has phase-lags of 90 degrees, 135 degrees and 195 degrees, respectively. These frequencies are denoted as ω_{90}, ω_{135}, and ω_{195}. These estimates are valid only if the non minimum phase parts of the dynamics are not dominating.

In Åström (2000) it is shown that for minimum phase systems the phase lag of the minimum phase parts has to be less than 60 degrees at the crossover frequency. The system will have poor robustness if the phase lag is larger. For a system with dead time this implies that the crossover frequency is approximately limited to $\omega_{gc}L<1$. This result can be used to determine when the PI control can be used for processes with time delay.

5.2. Process uncertainties

Process uncertainties impose severe limitations on control performance. Traditionally, this has been taken into account using gain and phase margins. The design methods developed by Kessler for motor drives used the slowest neglected time constant as an important part of the specifications. See Kessler (1958a,b).

Process uncertainty is used explicitly in design methods such as H_∞ (Zhou, Doyle, & Glover, 1996) and QFT (Horowitz, 1993). Some consequences of this for PID control are discussed in Panagopoulos and öström (2000) and Fransson, Lennartsson, Wik, and Gutman (2000).

5.3. When is derivative action useful?

In Section 3, it was found that adding derivative action to a PI controller increases the complexity of the design considerably. There is much folklore concerning derivative action. Why is the derivative term hard to tune? Why should one choose $T_i = 4T_d$? What is the use of complex zeros, i.e. $T_i<4T_d$? Why is the derivative action not suitable for a pure dead-time process. To get some insight into these problems, the results on performance assessment will be used. First, a specific example is discussed.

Example 2 (Process with lag and delay.). Consider a process with a first-order lag and a time delay, i.e.

$$G(s) = \frac{1}{1+sT}e^{-sL}. \quad (16)$$

Conditions for when the process can be successfully controlled by a PI controller will first be determined. It follows from the discussion above that the crossover frequency is limited by ω_{gc}. It follows from Eq. (15) that

$$\arctan\omega_{gc}T + \omega_{gc}L - \arg G_c(i\omega) = \pi - \varphi_m.$$

Assume that a phase margin of 60 degrees is desired and that PI control is used. Since $\omega_{gc}L<1$ it follows that $\arctan\omega_{gc}T<2\pi/3-\pi/4-1=0.3090$. It can thus be

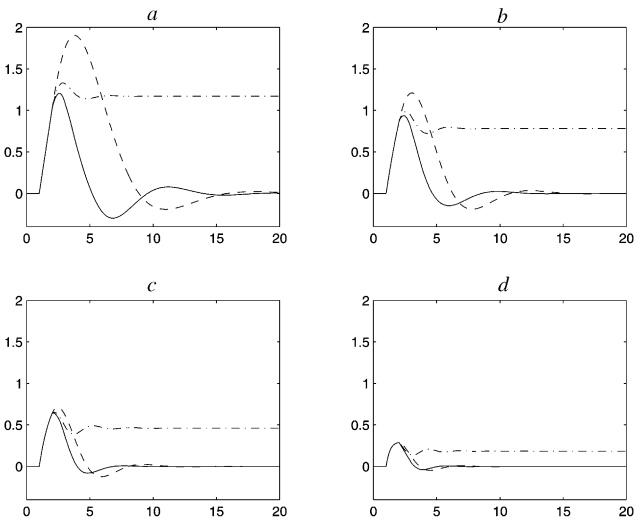

Fig. 7. Responses in process output to steps in load disturbances for PD, PI and PID controllers for systems with the transfer function $G(s) = Te^{-s}/(1+sT)$ with (a) $T = 10$, (b) $T = 3$ (c) $T = 1$ and (d) $T = 0.3$.

concluded that PI control can be used if $T<0.3L$ or if $L>3T$.

Fig. 7 shows responses to load disturbances for PD, PI and PID controllers for four different systems with different ratios of dead-time to time constant. Notice that derivative action gives substantial benefits for systems having large values of T. Due to the predictive action the error starts to decrease much faster. Notice that the peak disturbance obtained for PID control is lower than for PD control.

All controllers simulated in Fig. 7 are designed by maximizing integral gain subject to the robustness constraint $M_s<2$ using the method described in Panagopoulos (2000), where the constraint $T_i=4T_d$ was imposed on the PID controller, see Wallen, (2000).

The figure shows that there is a significant improvement by using derivative action for small values of L/T but that the benefit of derivative action decreases when the ration L/T increases.

To get more insight into the question, when derivative action is useful the extreme case of an integrator with delay will be considered.

Example 3 (Integrator with delay.). Consider a system with the transfer function

$$G(s)=\frac{1}{s}e^{-s}.$$

For this system $\omega_{90}=0$, $\omega_{135}=0.78$, and $\omega_{195}=1.8$. These values indicate that the crossover frequency can be doubled by using integral action. To get more insight PD, PI and PID controllers were designed to maximize k_i subject to $M_s<2$ using the method described in Panagopoulos (2000), where the constraint $T_i=4T_d$ was imposed on the PID controller, see Wallen (2000). The controller parameters obtained are given in Table 1. Notice that the integrated error IE is reduced from 7.6 for PI control to 2.2 for PID control and that the integrated absolute error IAE is reduced from 9.2 for PI control to 3.8 for PID control.

Fig. 8 shows the responses to a step change in load disturbance and the control signal for PD, PI and PID controllers designed to maximize k_i subject to $M_s<2$. The figure shows that the major benefit in reducing the effect of the load disturbance is that the control signal reacts much faster initially for controllers with derivative action. This shows clearly how important it is to react fast in order to achieve good control. Also notice that at a casual inspection the differences in the control signal do not appear very large but that they have a profound effect on the output signal.

Table 1
Parameters of PD, PI and PID controllers for a system with the transfer function $G(s)=e^{-s}/s$.

Type	k	k_i	k_d	T_i	T_d
PD	0.854		0.255		0.299
PI	0.488	0.131		3.739	
PID	0.864	0.462	0.404	1.869	0.467

The examples support the folklore that derivative action is useful for processes that are lag dominated and that it is less useful for systems that are dead-time dominated, see Shinskey, (1996).

6. Competing strategies

This section will discuss some situations where there are competing strategies that can do better than PID control. Four different cases will be discussed dead-time dominant systems, oscillatory systems, multivariable systems, and nonlinear systems. It has attempted to answer the question when it is possible to obtain drastic improvements over PID control.

6.1. General linear controllers

The PID controller with only a few parameters is certainly a restricted complexity controller. An alternative would be to replace it by a general linear controller. A general two degree controller can be represented as

$$R(s)U(s)=T(s)Y_{sp}(s)-S(s)Y(s),$$

where R, S and T are polynomials of arbitrary order. Another alternative would be to use a state space representation consisting of an observer with state

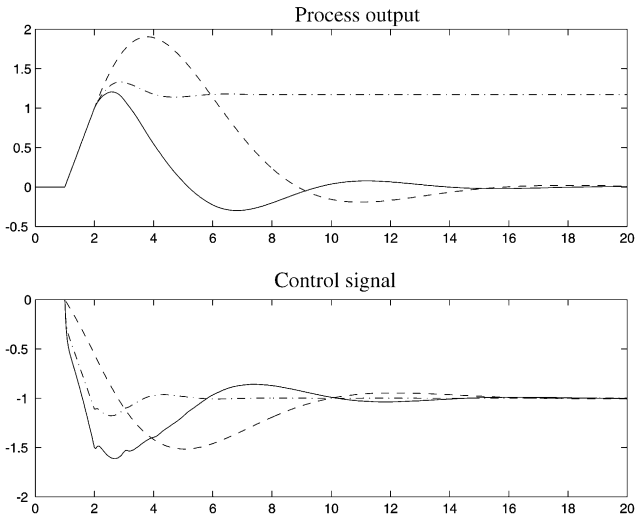

Fig. 8. Responses in process output to a step in the load disturbance applied at time t = 0 for PD (dash–dotted), PI (dashed), and PID (solid) control of a system with the transfer function $G(s)=e^{-s}/s$.

feedback and model following. Such controllers are straight forward extensions to PID control. Some problems with these general controllers are parameterization, design and tuning. General linear controllers include PID control as a special case and can thus always outperform a PID controller. It is however a nontrivial question to find good man-machine interfaces so that the users can easily specify structure and parameters. Tuning tools are also necessary. There are a number of commercial linear adaptive controllers but they are at present much less common than PID controllers, see Bengtsson and Egardt (1984) and Modén (1995).

6.2. Set point response

It has been emphasized several times that the problem of regulation and set point response should be separated, see Fig. 1. Our approach was to first design the feedback G_c to obtain a system that will balance the effects of load disturbances and measurement noise and which is robust to process uncertainties. The feedforward path G_{ff} is then designed to obtain the desired response to set point changes. This results in the two-degree of freedom structure advocated in Horowitz (1963). For set point changes of moderate size good results can often be obtained by using set point weighting, see Åström and Hägglund (1995). In some cases it may be desirable to add linear filters to reduce the overshoot, see Panagopoulos (2000). Rapid set point responses are often limited by the nonlinearities of the process. Recent work by Wallen (2000) indicates that there are substantial benefits by using nonlinear control strategies. For systems with time delays the response time can be decreased substantially by using Smith predictors.

6.3. Processes with time delays

There is much folklore concerning systems with time delay. Such systems can be controlled quite well with PID control. However, the traditional tuning rules often give very poor results. Derivative action is quite useful for systems which also has lags as was found in Section 5. Derivative action, however is of limited value for systems that are dead time dominant. The reason for this is that prediction of the output based on the linear extrapolation is not effective. It is much better to make predictions based on inputs that were fed into the system than those which have not yet shown up in the output. The Smith predictor accomplishes this, see Kaya and Atherton (1999), Kristiansson and Lennartson (1999) and Jiya, Shao and Chai, (1999).

Let the process transfer function be $G = \tilde{G}e^{-sL}$. A controller with a Smith predictor can be interpreted as a controller of the type shown in Fig. 1 with $G_c = \tilde{G}_c G_{sp}$, where G_{sp} is the Smith predictor given by

$$G_{sp} = \frac{1}{1 + \tilde{G}\tilde{G}_c(1 - e^{-sL})}. \tag{17}$$

If $\tilde{G}\tilde{G}_c \approx -1$ then $G_{sp} \approx e^{sL}$ which indicates that G_{sp} acts as a predictor in a certain frequency range, see Åström (1977).

The transfer function from the set point to the process output is given by

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{\tilde{G}\tilde{G}e^{-sL}}{1 + \tilde{G}\tilde{G}_c} \tag{18}$$

The transfer function from load disturbance to the process output is

$$\frac{Y(s)}{L(s)} = \frac{1 + \tilde{G}\tilde{G}_c(1 - e^{-sL})}{1 + \tilde{G}\tilde{G}_c}\tilde{G}e^{-sL}. \tag{19}$$

A particularly simple case is when $\tilde{G}G_{pi} = k_v/s = 1/(sT)$. To see that G_{sp} acts like a predictor, a series expansion in s is made

$$G_{sp} \approx \frac{1}{1 + k_v L}\left(1 + \frac{k_v L}{1 + k_v L}\frac{L}{2}s\right).$$

For large k_v the predictor thus predicts over a time horizon of $L/2$. Assuming that the process has a unit gain at low frequencies, the effective integral gain of a PI controller with a Smith predictor becomes

$$k_i = \frac{k_v}{1 + k_v L} = \frac{1}{T + L}. \tag{20}$$

It follows from this equation that it is desirable to have a large value of k_v or small values of T. The largest value that can be used is limited by the requirement on robustness.

The sensitivity function is

$$G_s = \frac{1 + \tilde{G}\tilde{G}_c(1 - e^{-sL})}{1 + \tilde{G}\tilde{G}_c} = 1 - \frac{\tilde{G}\tilde{G}_c e^{-sL}}{1 + \tilde{G}\tilde{G}_c}.$$

For the particular design the sensitivity function becomes

$$G_s = 1 - \frac{k_v e^{-sL}}{s + k_v} = 1 - \frac{e^{-sL}}{sT + 1}.$$

The maximum values of the sensitivity and the complementary sensitivity are thus less than 2 for all values of T. A sensitivity analysis thus indicates that arbitrary small values of T can be used if sensitivities less than 2 are admitted. This however does not tell the full story.

Fig. 9 shows the Nyquist curves for the loop transfer function for different values of the ratio T/L. Notice that there are loops in the Nyquist curves that increase with decreasing values of T. Because of these loops the system will be sensitive to variations in the time delay. Systems of this type are classic cases where the sensitivity functions do not give good insight into the robustness issues. It is necessary to also ensure that the

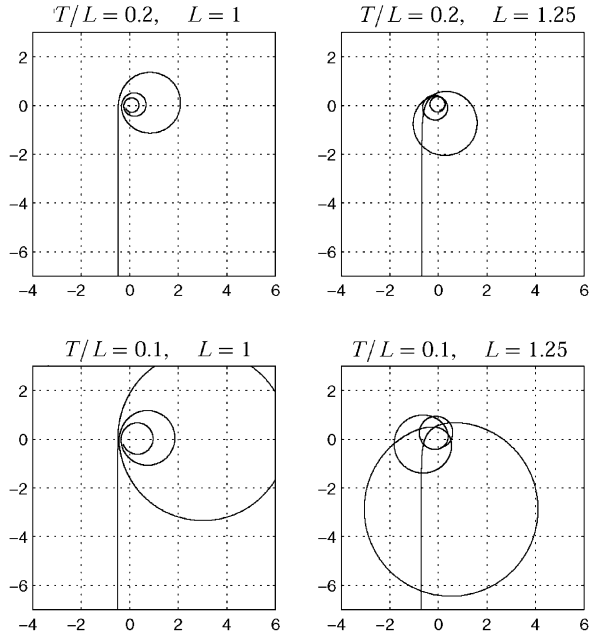

Fig. 9. Nyquist curves for the loop transfer functions for $T/L = 0.1$ and 0.2 for the nominal case (left) and for a system with a 25% increase of the time delay (right).

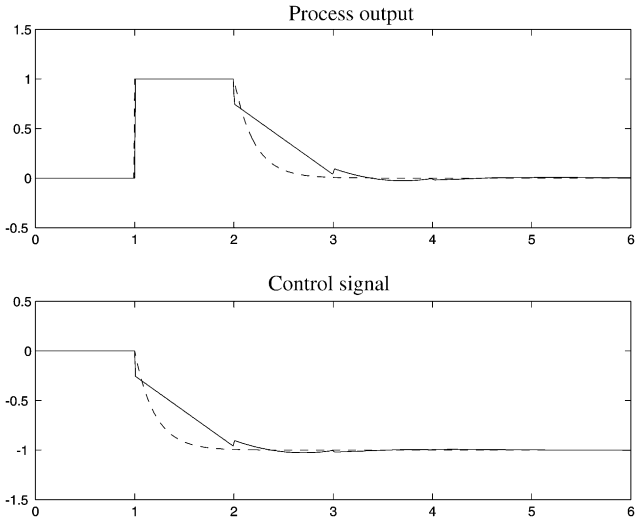

Fig. 10. Responses to a step in load disturbance for a system with a Smith predictor with $k_v = 5$ (dashed) and PI control with $k = 0.25$ and $T_i = 0.35L$ (solid).

delay margins are satisfactory. Fig. 9 shows, for example, that a system with $T/L = 0.1$ becomes unstable if the time delay is increased by 25% from $L = 1$ to 1.25.

It is of interest to compare the results that can be obtained with simple PI control. This is made in the following example.

Example 4 (A pure delay process). Consider a pure delay process which has the transfer function $G = e^{-sL}$. The integral gain obtained with Smith predictor control is given by Eq. (20). It gives $k_i = 0.8/L$ for $k_v = 5$. Designing a PI controller to maximize k_i subject to a robustness constraint $M_s < 2$ gives $k = 0.25$ and $T_i = 0.35L$, hence $k_i = 0.71/L$. The Smith predictor thus gives only a very small increase of k_i compared to the PI controller. Fig. 10 compared the responses to load disturbances of systems with PI control and with PI control and a Smith predictor. Notice that the control signal for the Smith predictor is smoother. This is an advantage because it means that modes of higher frequencies are not excited.

The results indicate that for systems with pure time delays the Smith predictor only gives marginal improvements in regulation performance. These improvements may be relevant for control of important quality control problems but not for ordinary problems. The Smith predictor may, however, give substantial improvements in the response to set point changes.

6.4. Systems with oscillatory modes

Systems with oscillatory modes can be controlled with PI controllers if the requirements are not too high. The performance of a PID controller can, however, be improved by adding filters to the control system.

6.5. Multivariable systems

Many processes are multivariable. Much research has lately been devoted to control schemes for such systems, see e.g. Morari and Lee (1991), Camacho and Bordons (1995). Most multivariable schemes do however operate in cascade mode, where the multivariable controller provides set points to local PID controllers, see Lee, Park and Lee, (1999). The PID controller is thus an essential building block, also in this case, tuning can often be done iteratively, see Vásquez, Morilla and Dormido (1999). It should also be emphasized that PID controllers with static decoupling can do very well for multivariable systems if the requirements are not too high. This is illustrated by an example.

Example 5 (Rosenbrock's system). A seemingly simple multivariable system proposed by Rosenbrock has the transfer function

$$G(s) = \begin{pmatrix} \dfrac{1}{s+1} & \dfrac{1}{s+1} \\ \dfrac{2}{s+3} & \dfrac{1}{s+1} \end{pmatrix}.$$

By static decoupling a system with the transfer function

$$Q(s) = G(s)G^{-1}(0) = \begin{pmatrix} \dfrac{1}{s+1} & 0 \\ \dfrac{4s}{(s+1)(s+3)} & \dfrac{3(1-s)}{(s+1)(s+3)} \end{pmatrix},$$

is obtained. This shows that the system has a right half-plane zero which explains why it is difficult to control. It is, however, possible to control this system by PID controllers provided that the demands are not excessive. This is illustrated in the step responses shown in Fig. 11. It is also possible to provide simple estimates for how fast the system can be made while keeping the interactions at a reasonable level.

6.6. Nonlinear systems

The PID controller being linear is not suited for strongly nonlinear systems. Excellent results can however be obtained by combining the PID controller with gain scheduling. A particularly attractive feature is the use of auto-tuning which drastically reduces the effort required to build up the gain schedule.

7. Does PID control have a future?

It is quite reasonable to predict that PID control will continue to be used in the future. Feedback has had a revolutionary influence in practically all areas where it has been used and will continue to do so. PI(D) control is perhaps the most basic form of feedback. It is very effective and can be applied to a wide range of problems. The emerging features of automatic tuning have greatly simplified the use of PID control.

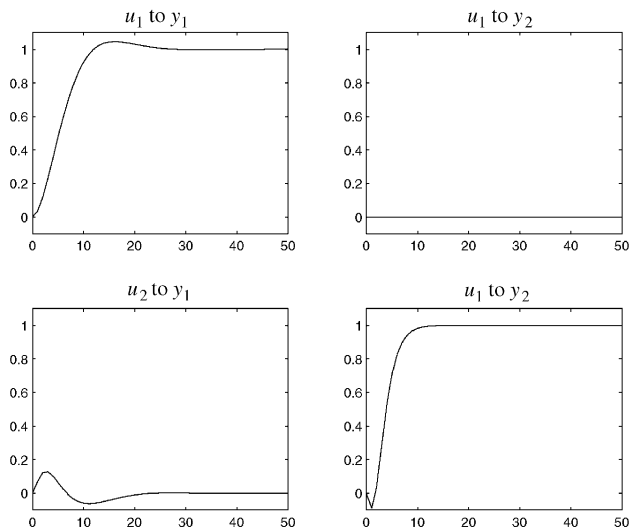

Fig. 11. Simulation of PI control with static decoupling of Rosenbrock's system. The figure shows the response of the outputs to steps in the command signals. All controllers have set point weighting with $b = 0$.

More knowledge about PID control has been available for a long time. Unfortunately, it has been buried in proprietary information of suppliers. There was a strong resurgence in the interest in PID control over the last 10 years. Many publications have appeared and it is typical that IFAC organized a workshop on PID control in the year 2000.

The alternatives to PID control are:

RST: Discrete-time linear MISO controllers.
SFO: State feedback and observers.
MPC: Model predictive control.

Fuzzy control is often mentioned as an alternative to PID control, see Passino and Yurkovich (1998). Most fuzzy controllers used in industry have the same structure as incremental PI or PID controllers. The parameterization using rules and fuzzy membership functions makes it easy to add nonlinearities, logic, and additional input signals to the control law.

The main advantages claimed by fuzzy control are that it is easy to use and that it is nonlinear. Many of the comparisons between PID and fuzzy control made in the literature are, however, very sloppy. A textbook PID with Ziegler–Nichols tuning is typically used as a reference. Furthermore, if nonlinear behavior is desired gain scheduling can be added to a PID controller. Many fuzzy controllers are also used in a cascade structure using PID controllers at the lower level. An advantage of fuzzy control is that very good software is available. Many fuzzy controllers are, however, used in a cascade structure using PID controllers at the lower level.

To assess the different alternatives it is important to consider.

- Performance.
- Tuning.
- Ease of use.
- Maintenance.

All alternatives offer potential improvements in the linear behavior of the system. This is particularly useful when dealing with systems having poorly damped oscillatory modes. It is, however, necessary to take actuator saturation and windup into account. This is done very well for MPC but it requires special consideration for RST and SFO, particularly in combination with mode changes.

The tuning problem is a major difficulty with all alternatives. Much work is needed in order to develop appropriate tuning tools.

The RST controller is beginning to appear in applications for motor drives. The improvements in

performance are particularly important for high performance systems.

Controllers based on state feedback and observers are used in special applications where the cost of the modeling effort can be justified.

Model predictive control is typically used in supervisory mode with PID controllers at the base level. Much of the improvement accredited to MPC in the process industry actually comes from the improved tuning of the basic loops. The MPC does, however, offer drastic improvements in set point responses for multivariable systems because of the coordination it provides. There are several interesting problems related to the integration of MPC and PID. Tuning of the PID controller in the inner loop gives valuable modeling information for the MPC. Since the MPC operates in a supervisory mode it can deal with slow interactions very well. The lower level PID loops still have to manage fast interactions.

Another issue is the general philosophy of approaching design of a complex system. Top–down approaches to system design will clearly favor SFO and MPC. A bottom–up approach favors the use of simple building blocks such as PID controllers. In this context, it is interesting to see how the PID controller can be augmented. The set point response can be improved substantially by exploiting a controller structure with two degrees of freedom as illustrated in Fig. 1 in this paper. Special considerations have to be given to a PID controller that will be used effectively in this way. It is particularly important that the controller output is available for feedforward in such a way that saturation and windup are handled properly.

More improvements can be made through proper use of feedforward. Other possibilities are to add filters and blocks for dealing with measurement noise and systems having oscillatory modes. There are many good research problems in developing good approaches to bottom up design techniques. The problem of exploring system interactions is an important one.

Even if the applications of other control strategies increase, PID control will certainly continue to be used. When correctly used it is a very effective way of using feedback. Good results can often be obtained if the performance requirements are not extreme. The PID controller will also serve as a building block in more complex controllers. Most DMC controllers in fact deliver set points to PID controllers. Good performance of these PID controllers are essential. Much commissioning work for DMC control actually consists of tuning up the underlying PID loops. There are also useful augmentations for PID control in the form of Smith predictors, gain scheduling and filters for oscillatory systems.

It is important to realize that there is a very wide range of control problems and consequently also a need for a wide range of tuning techniques. There are already many tuning methods available, but a replacement of the Ziegler–Nichols method is long overdue. It is very easy to demonstrate that any controller with reasonable tuning will outperform a PID with Ziegler–Nichols tuning. Many strategies proposed can easily be eliminated if they are compared with a well-tuned PID.

Development of suitable software is another area that has to be developed. PID control is quite underdeveloped in that respect compared, for example, to fuzzy control. It would be highly desirable to have software so that persons with a moderate knowledge can experiment with PID control. Tools for modeling and methods for automatic tuning should be a part of such a software.

On the research side, it appears that the development of design methods for PID control is approaching the point of diminishing returns. There are some difficult problems that remain to be solved. For example, there is no characterization of the processes where PID control is useful.

References

Anon, A. (1999). Special edition on PID tuning methods. *Computing & Control Engineering Journal*, *10*(2).

Åström, K. J. (1977). Frequency domain properties of Otto Smith regulators. *International Journal of Control, 26*, 307–314.

Åström, K. J. (2000). Limitations on control system performance. *European Journal on control*.

Åström, K. J., & Hägglund, T. (1995). *PID controllers: Theory, design and tuning*. Research Triangle Park, N.C. Instrument Society of America.

Atherton, D. (1999). PID controller tuning. *Computing & Control Engineering Journal*, 44–50.

Bengtsson, G., & Egardt, B. (1984). Experiences with self-tuning control in the process industry. *Preprints 9th IFAC world congress,* Budapest, Hungary (pp. XI: 132–140).

Bennett, S. (1979). *A History of Control Engineering 1800–1930*. London: Peter Peregrinus.

Bennett, S. (1993). *A History of Control Engineering 1930–1955*. London: Peter Peregrinus.

Camacho, E. F., & Bordons, C. (1995). *Model prediction control in the process industry*. *Advances in industrial control*. Berlin: Springer.

Fransson, C., Lennartsson, B., Wik, T., & Gutman, P. (2000). On optimizing PID controllers for uncertain plants using Horowitz bounds. *IFAC workshop on digital control—past, present, and future of PID Control,* Terrassa, Spain.

Ho, M. T., Datta, A., & Bhattacharyya, S. P. (1996). A new approach to feedback stabilization. *In Proceedings of the 35th IEEE conference on decision and control, IEEE*, vol. 4 (pp. 4643–4648).

Ho, M. T., Datta, A., & Bhattacharyya, S. P. (1997). A linear programming characterization of all stabilizing PID controllers. *Proceedings of the American control conference, IEEE*, Albuquerque, NM (pp. 3922–3928).

Horowitz, I. (1993). *Quantitative feedback theory* (*QFT*). Boulder, Co: QFT Publications.

Horowitz, I. M. (1963). *Synthesis of feedback systems*. New York: Academic Press.

Jiya, J., Shao, C., & Chai, T. Y. (1999). Comparison of PID and PPI design techniques for a process with time delay. *Preprints. 14th world congress of IFAC*, Beijing, China (pp. 391–396).

Kaya, I., & Atherton, D. P. (1999). A new pipd smith predictor for control of processes with long time delays. *Preprints. 14th world congress of IFAC*, Beijing, China (pp 283–288).

Kessler, C. (1958a). Das symmetrische Optimum, Teil I. *Regelungstechnik*, *6*(11), 395–400.

Kessler, C. (1958b). Das symmetrische, Optimum Teil II. *Regelungstechnik*, *6*(12), 432–436.

Kristiansson, B., & Lennartson, B. (1999). Optimal PID controllers including roll off and smith predictor structure. *Preprints. 14th world congress of IFAC*, Beijing, China (pp. 297–302).

Lee, Y., Park, S., & Lee, J. H. (1999). On interfacing model predictive controllers with low-level loops. *Preprints. 14th world congress of IFAC*, Beijing, China (pp. 313–318).

Lieslehto, J. (1999). Collection of java applets for PID controller design. *Preprints. 14th world congress of IFAC*, Beijing, China (pp 421–426).

Modén, P. E. (1995). Experiences with adaptive control since 1982. *Proceedings of the 34th IEEE conference on decision and control*, New Orleans, LA, pp. 667–672.

Morari, M., & Lee, J. H. (1991). Model predictive control: The good, the bad, and the ugly. *Chemical process control, CPCIV*, Padre Island Tx (pp 419–442).

Munro, N., Söylemez, M. T., & Baki, H. (1999). Computation of D-stabilizing low-order compensators. *IEEE Trans. on Automatic Control*, Submitted for publication.

Panagopoulos, H. (2000). *PID control design, extension, application*. Ph.D. thesis, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

Panagopoulos, H., & Åström, K. J. (2000). PID control design and H_∞ loop shaping, *International Journal of Robust and Nonlinear Control, 10*, 1249–1261.

Panagopoulos, H., & Åström, & Häglund, T. (1999). Design of PID controllers based on constrained optimization. *Proceedings of the 1999 American control conference* (*ACC'99*), San Diego, CA. Invited paper.

Panagopoulos, H., Åström, K. J., and Hägglund, T. (1998). Design of PI controllers based on non-convex optimization. *Automatica*, *34*(5) 585–601.

Passino, K. M., & Yurkovich, S. (1998). *Fuzzy control*. Menlo Park, CA. Addison-Welsley-Longman.

Seron, M. M., Braslavsky, J. H., Goodwin, G. C. (1997). *Fundamental limitations in filtering and control*. Berlin: Springer.

Shafiei, Z., & Shenton, A. T. (1994). Tuning of PID-type controllers for stable and unstable systems with time delay. *Automatica*, *30*(10), 1609–1615.

Shenton, A. T., Shafiei, Z. (1994). Relative stability for control systems with adjustable parameters. *Journal of Guidance Control and Dynamics, 17*(2) 304–310.

Shinskey, F. G. (1996). *Process control systems. Application, design and tuning* (4th ed.) New York: McGraw-Hill.

Söylemez, M. T., Munro, N., Baki, H. (1999). Fast calculation of stabilizing PID controllers. Automatica, submitted for publication.

Taguchi, H., & Araki, M. (2000). Two-degree-of-freedom PID controllers–their functions and optimal tuning. *IFAC workshop on digital control—Past, present and future of PID Control*, Terrassa, Spain.

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

Vásquez, F., Morilla, F., & Dormido, S. (1999). An iterative method for tuning decentralized PID controllers. *Preprints. 14th world congress of IFAC*, Beijing, China (pp. 491–496).

Wallen, A. (2000). Tools for autonomous process control. Ph.D. thesis, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

Yamamoto, T., Fujii, K., & Kaneda, M. (1999). A design of self-tuning PID controllers based on a pole-assignment scheme. *Preprints. 14th world congress of IFAC*, Beijing, China (pp. 259–264).

Zhou, J., Doyle, J., & Glover, K. (1996). *Robust and optimal control*. Englewood Cliffs, NJ: Prentice-Hall.

Zhuang, M., & Atherton, D. P. (1993). Tuning of optimum PID controllers. *Proceedings of IEE*, *140*, 216–224.