Dead-Time Compensation and Performance Monitoring in Process Control

Ingimundarson, Ari

2003

Document Version:
Publisher's PDF, also known as Version of record

Dead-Time Compensation and Performance Monitoring in Process Control

Dead-Time Compensation and Performance Monitoring in Process Control

Ari Ingimundarson

Department of Automatic Control
Lund Institute of Technology
Lund, January 2003

To Bea

Department of Automatic Control
Lund Institute of Technology
Box 118
SE-221 00 LUND
Sweden

ISSN 0280–5316
ISRN LUTFD2/TFRT–1064–SE

Printed in Sweden by Bloms i Lund Tryckeri AB.
Lund 2003

Contents

Preface

This thesis addresses two topics in process control. The topics are dead-time compensation and closed-loop performance monitoring. The first is concerned with the control of processes with long dead-time. The second is about monitoring of feedback controllers to know how well they are performing. The thesis is divided into two parts accordingly. The first part is composed of two published articles with an introduction while the second part is composed of 4 chapters.

As an engineering discipline, process control has many exiting challenges to offer researchers within the academic community. One of the most difficult challenge researchers face is to transmit results of their research to the practitioners within the field. It has been the hope of the author when writing this thesis that the results might be of relevance to practitioners.

The thesis is the result of a few years of Ph.D. studies. During the years some of the material has been published at other occasions. The research on dead-time compensation was published in the following articles:

Ingimundarson, A. and T. Hägglund (2000a): "Closed-loop identification of first-order plus dead-time model with method of moments." In *ADCHEM 2000, IFAC International Symposium on Advanced Control of Chemical Processes*. Pisa, Italy.

Ingimundarson, A. and T. Hägglund (2000b): "Robust automatic tuning of an industrial PI controller for dead-time systems." In *IFAC Workshop on Digital Control – Past, present, and future of PID Control*. Terrassa, Spain.

The two articles included in this thesis are:

Ingimundarson, A. and T. Hägglund (2001): "Robust tuning procedures for dead-time compensating controllers." *Control Engineering Practice*, **9**, pp. 1195–1208.

Ingimundarson, A. and T. Hägglund (2002): "Performance comparison between PID and dead-time compensating controllers." *Journal of Process Control*, **12**, pp. 887–895.

The second part of the thesis has been partially published in :

Ingimundarson, A. (2002): "Performance monitoring of PI controllers using a synthetic gradient of a quadratic cost function." In *IFAC World Congress*. Barcelona, Spain.

Work that has not been included in this thesis but was performed during the Ph.D. studies was published in

Solyom, S. and A. Ingimundarson (2002): "A synthesis method for robust PID controllers for a class of uncertain systems." *Asian Journal of Control*, **4:4**.

Also the following article has been submitted.

Ingimundarson, A. and S. Solyom (2002): " On a synthesis method for robust PID controllers for a class of uncertainties", Submitted for publication in *European Control Conference ECC2003.*

Acknowledgments

It is a privilege to be a Ph.D. student at the Department of Automatic Control at Lund Institute of Technology. The department is second to none in providing opportunities for Ph.D. students to realize any ambition they might have in research on automatic control. The nice atmosphere there is due to the staff and my thanks go out to all of them.

I would like to thank the Center for Chemical Process Design and Control (CPDC) and The Swedish Agency for Innovation Systems (Vinnova) for financial support.

I would like to thank my supervisor, Tore Hägglund, for always having time and a positive attitude to most ideas. The people that proof read versions of the manuscript, Björn Wittenmark and Johan Eker are thanked for their comments.

Finally, I would like to thank the people that have indirectly contributed to making this thesis a reality. Friends here in Sweden have made the stay much more enjoyable. My family in Iceland are thanked for support and encouragement. At last, I would like to thank the person who made the whole thing possible, my beloved Bea.

Ari

Part I

Topics in Dead-Time Compensation

Introduction to Dead-Time Compensation

1. Background and Motivation

Dead time is a frequently quoted reason for increased loop variability within the process industry, see [Bialkowski, 1998]. A control structure specially designed to deal with long dead times is called a dead-time compensator (DTC). One of the earliest papers dealing specifically with dead time was [Smith, 1957]. The dead-time compensator presented there, has been referred to as the Smith predictor in the literature and that name has actually become a synonym for a dead-time compensator. Smith showed how the design problem of a plant with dead time could be reduced to a design problem of the plant without the dead time. This idea has been used many times since the original publication.

Reasons for the dead time can be many. The most frequent in the process industry is dead time due to transportation time of material between actuator and sensor. There are other reasons for dead time. It might be caused by computation and communication delays or it might appear when a higher order model is approximated with a low order model. Dead time sets a fundamental limit on how well a controller can fulfill design specifications since it limits how fast a controller can react to disturbances.

The work presented in [Smith, 1957] gained considerably in value with the advent of computer control. The reason being that to implement a dead-time element in the control structure was difficult using only analog components. This problem was simplified with computer control. Now DTCs are offered as standard modules in commercial control systems.

2. Research on Dead-Time Compensation

The literature on dead-time compensators covers topics from the more mathematical infinite dimensional system theory to practical issues such as the commissioning and tuning of common dead-time compensators.

Most DTCs are model based controllers. The control structure contains a model of the plant. For a discussion on the role of the model in dead-time compensation, see [Watanabe and Ito, 1981]. A simple explanation of the role of the model is that it is used to predict the effect of the control signal on the output. By applying feedback from the model output the dead time is taken into account when the control signal is decided.

A large portion of the work in DTCs in the literature is for fixed models, that is, the dynamics in addition to the dead time are of specific order and form. The most common models are

$$\frac{K_p}{Ts+1}e^{-sL} \qquad \frac{K_v}{s}e^{-sL}$$

or first-order plus dead time (FOPDT) transfer functions. The reason for this is that these are the models most commonly used within process control. One of the main application areas for DTCs is within the process industry.

The most common controller in the process industry is the PID controller. One of the key difference between a DTC and PID is the inherent model in the DTC. The model adds to the complexity of the DTC structure. As an example, if the model in the DTC is a stable first-order plus dead time transfer function and the controller is a PI, the number of parameters for the DTC is five. This corresponds to a drastic increase from the PID which has three. Even though an initial tuning is obtained from an experiment, if the performance deteriorates and maintenance is needed, the PID has the advantage that it can be manually tuned. In [Normey-Rico and Camacho, 2002] it was pointed out that often when new structures are introduced and compared to existing structures the complexity of each structure is not taken into account which can lead to unfair comparisons.

The additional complexity of DTCs has been noticed by authors and often an effort is made to keep structures simple. This is done by keeping the number of adjustable parameters few and with a clear interpretation. In [Hägglund, 1996] a DTC was presented which reduced the number of parameters from five to three by fixing together some parameters of the model and those of the controller. It is still quite common that complexity is ignored.

The robustness of DTCs has been investigated in a number of articles. In [Palmor, 1980] it was shown that conventional approaches to design Smith predictors could lead to closed-loop systems with an arbitrary

small dead time margin. In [Laughlin *et al.*, 1987] the robust performance of Smith predictors was investigated. Conditions for robust performance were presented for the stable FOPDT case assuming that each of the three model parameters would lie in an interval. A very similar problem was addressed in [Lee *et al.*, 1996]. The robustness of the structure presented in [Hägglund, 1996] was further investigated in [Normey-Rico *et al.*, 1997] and a new structure proposed. For a more theoretical approach to the robustness problem see [Meinsma and Zwart, 2000] where a mixed sensitivity $\mathcal{H}_\infty$ problem is solved for a linear system with delay.

The Smith predictor does not yield zero steady state error to a load disturbance when the plant has an integrator. A dead-time compensation structure which had this very desirable property even for integrating plants was presented in [Watanabe and Ito, 1981]. A number of publications have followed and treated this problem. The solutions vary in complexity and disturbance rejection capability. Some references are [Åström *et al.*, 1994; Matausek and Micic, 1996; Matausek and Micic, 1999; Normey-Rico and Camacho, 1999].

3. Outline and Summary of Contribution

Paper 1

The first paper deals with tuning of simple dead-time compensators. Two DTCs are considered, one for self regulating processes and one for integrating processes. The DTC structure for self regulating processes was first presented in [Normey-Rico *et al.*, 1997] but the parameters of the structure are selected differently in the current work. The DTC for integrating processes is the one presented in [Matausek and Micic, 1996].

A new method to identify the first-order plus dead time models shown before, is presented. The identification procedure consists of two step responses, one in closed loop and one in open. The design procedure for the two DTCs result in a first order set-point response, corresponding to the model

$$\frac{1}{T_r s + 1} e^{-L}$$

It is shown how a suitable lower bound on the closed-loop time constant, T_r, can be found by considering the area between the plant output and the model output when a step is applied to both. In the integrating case it is actually two steps, one up and one down, to limit the change in the plant output. It is also shown how dead time margin depends on T_r.

Paper 2

A common sight in the literature is an introduction of a new structure which is shown to outperform other structures in a few simulation examples. Robustness is frequently not taken into account in the comparison even though it is well known that a robustness/performance tradeoff is always present in controller design.

The second paper is concerned with the use of DTCs or more specifically, when they should be used. Recognizing that the control strategy that the DTC probably would replace would be a PI or PID, the performance of the DTC is compared to that of PI(D) under a robustness constraint. Typical DTCs for both stable and integrating processes are compared to the best PI and PID which fulfill the robustness constraint.

Paper 1

Robust tuning procedures of dead-time compensating controllers

Ari Ingimundarson and Tore Hägglund

Abstract

This paper describes tuning procedures for dead-time compensating controllers (DTC). Both stable and integrating processes are considered. Simple experiments are performed to obtain process models as well as bounds on the allowable bandwidth for stability. The DTCs used have few parameters with clear physical interpretation so that manual tuning is possible. Furthermore, it is shown how the DTCs can be made robust towards dead-time variations.

Keywords Automatic tuning, Dead-time compensation, Robustness, PID control

Reproduced with permission from: Ingimundarson, A. and T. Hägglund (2001): "Robust tuning procedures of dead-time compensating controllers", *Control Engineering Practice 9*, p.1195-1208.

1. Introduction

Most control problems in the process industry are solved using PID controllers. There are several reasons for this. One is that the PID controller can be tuned manually by "trial-and-error" procedures, since it only has three adjustable parameters. The possibility to make manual adjustments of the controller parameters is important even when automatic tuning procedures are available.

When there are long dead times in the process, the control performance obtained with a PID controller is, however, limited. For these processes, dead-time compensating controllers (DTCs) may improve the performance considerably. These controllers require a process model to provide model-predictive control. This usually means a significant increase in controller parameters.

The use of DTCs also brings into existence new robustness problems connected to the dead-time. The classical ways to characterize robustness, phase margin and amplitude margin are not sufficient. In this paper, the delay margin which is the greatest variation in dead time that can occur in the process before the closed-loop system becomes unstable, will be used as well.

The aim of this paper is to show how it is possible with simple experiments to find parameters for the DTCs that give good performance while remaining robust. The experiments are composed of an identification of simple process models and then an experiment to determine an upper limit on closed-loop bandwidth. The latter is performed in open loop while the former is partially performed in closed loop. As a measure of closed-loop bandwidth, the reciprocal of the time constant of the set point response is used. This can then be related to other measures such as the loop-gain crossover frequency. The DTCs used in this paper have certain PID qualities, i.e. few parameters that can be tuned manually and have good interpretation in terms of classical control theory concepts. It will also be shown how the DTCs can be given a guaranteed delay margin.

For the identification of the simple process models in the DTCs an identification method first presented in [Ingimundarson and Hägglund, 2000a] is used. In this paper only the main equations and results are presented.

The paper is arranged in the following manner. In Section 2 the identification method is introduced. In Section 3, dead-time compensating controllers are discussed. In Section 4 the tuning procedure for stable processes is presented. This is followed by the procedures for integrating processes in Section 5. Finally conclusions are drawn in Section 6.

2. Identification

The two processes that are identified are the first-order plus dead time (FOPDT)

$$P_n(s) = \frac{K_n}{T_n s + 1} e^{-L_n s} \tag{1}$$

and the two-parameter model

$$P_n(s) = \frac{K_n}{s} e^{-L_n s} \tag{2}$$

These models are frequently used in the process industry and are considered to capture dynamics of real plants sufficiently well for many applications.

The identification method presented in this paper can be divided into two phases. First, the average residence time, $T_{ar} = L_n + T_n$ and the gain K_n are estimated with a change in operating levels. This change can be accomplished by a change in set point while operating in closed loop. The approach is based on the method of moments, see [Åström and Hägglund, 1995] for a general input signal applied to a linear system initially at rest. Second, the apparent time constant T_n is determined with an open-loop experiment where the input signal is a step or a ramp. In the case of the two-parameter model given by Eq. (2) only the first part of the experiment is necessary.

The method of moments

The method of moments can be explained with the following equations. For a general transfer function $G(s)$ an arbitrary input signal $U(s)$ results in an output signal given by

$$Y(s) = G(s)U(s) \tag{3}$$

By derivating $Y(s)$ with regard to s one gets

$$Y'(s) = G'(s)U(s) + G(s)U'(s) \tag{4}$$

The transfer function G(s) and its derivative can be evaluated at an arbitrary point α by calculating

$$Y^{(m)}(\alpha) = (-1)^m \int_0^\infty t^m e^{-\alpha t} y(t) dt$$
$$U^{(m)}(\alpha) = (-1)^m \int_0^\infty t^m e^{-\alpha t} u(t) dt$$

and solving Eqs. (3) and (4). Notice that if $\alpha = 0$ it is necessary for the signals considered to go to zero as time goes to infinity. Otherwise, the

integrals will not converge. Typically an input signal of the sort $\bar{u}(t) = u(\infty) - u(t)$ is selected. $u(\infty)$ is the value of the input signal after steady state has been reached again. The corresponding output signal is then $\bar{y}(t) = y(\infty) - y(t)$. Then it is only necessary to integrate for a finite time interval. This interval is denoted $[t_b, t_f]$ in the following.

In the case of the FOPDT model, (Eq. (1)) it is easy to get the following expression

$$\frac{P_n'(s)}{P_n(s)} = -L_n - \frac{T_n}{1 + T_n s}$$

By evaluating the transfer function at $\alpha = 0$, T_{ar} can be written as

$$T_{ar} = -\frac{P_n'(0)}{P_n(0)} \tag{5}$$

To evaluate $P_n'(0)$ it would be necessary to calculate the first moment of $\bar{y}(t)$ and $\bar{u}(t)$ for signals for which the integrals converge. These integrals have bad noise properties because of the factor t. Values at the end of the experiment have much higher weight than the ones in the beginning of the experiment. Therefore it is beneficial to consider the artificial signals

$$\begin{aligned} y_d(t) &= \tfrac{\mathrm{d}}{\mathrm{d}t}\bar{y}(t) \\ u_d(t) &= \tfrac{\mathrm{d}}{\mathrm{d}t}\bar{u}(t) \end{aligned}$$

The novelty of the method is the use of these signals. Denoting the mth moment of y_d and u_d with y_m and u_m respectively, it is possible to evaluate $P_n'(0)$ as

$$P_n'(0) = \frac{y_1 - P_n(0)u_1}{u_0} \tag{6}$$

Evaluating the moment integrals y_0 and y_1 gives

$$\begin{aligned} y_0 &= \int_0^\infty y_d(t)dt = [\bar{y}(t)]_0^\infty \\ &= -\bar{y}(0) \qquad (7) \\ y_1 &= -\int_0^\infty t y_d(t)dt \\ &= -[t\bar{y}(t)]_0^\infty + \int_0^\infty \bar{y}(t)dt \\ &= \int_0^\infty \bar{y}(t)dt \qquad (8) \end{aligned}$$

By evaluating Eq. (5) with moment equations Eqs. (7) and (8) the following expression is derived

$$\begin{aligned} T_{ar} &= -\frac{P_n'(0)}{P_n(0)} = \frac{-y_1 + P_n(0)u_1}{u_0 P_n(0)} \\ &= \int_{t_b}^{t_f} \frac{\bar{y}(t)}{\bar{y}(0)} dt - \int_{t_b}^{t_f} \frac{\bar{u}(t)}{\bar{u}(0)} dt \end{aligned} \tag{9}$$

This can be rewritten as

$$T_{ar} = \int_{t_b}^{t_f} \frac{u(t) - u(t_b)}{u(t_f) - u(t_b)} dt - \int_{t_b}^{t_f} \frac{y(t) - y(t_b)}{y(t_f) - y(t_b)} dt \tag{10}$$

The limits of the integrals have been changed to t_b and t_f. Eq. (10) has a nice graphical explanation. The functions under the integrals have beginning value equal to zero and final values equal to one. T_{ar} is simply the area between the two signals when they go from 0 to 1. This is shown in Fig. 1.

The method presented requires a change in the process levels. If this is accomplished by changing the set point in closed loop a stable controller is assumed to be present. The form of this change is not important but a step or a ramp would be easiest to implement in practice. Since the identification is based on integrals of the input and output signals it is preferable that the experiment is as short as possible in the presence of measurement noise.

If an integrator is present in the process the method presented can be used in the same way by considering the integral of the process input signal

$$u_i(t) = \int_{t_b}^{t} u(t) dt$$

and replacing $u_i(t)$ with the input signal in the previous paragraphs. Dead time, L_n, in Eq. (2) is then given by T_{ar}

Process gain K_n is trivially estimated as the ratio between input and output signal change, $K_n = \bar{y}(0)/\bar{u}(0)$.

Open-loop experiment to identify T_n

If a step or a ramp is applied to the FOPDT model, analytical expressions for the output signal are easily obtained. By integrating these expressions from 0 to T_{ar} it is possible to get expressions with only T_n as the unknown.

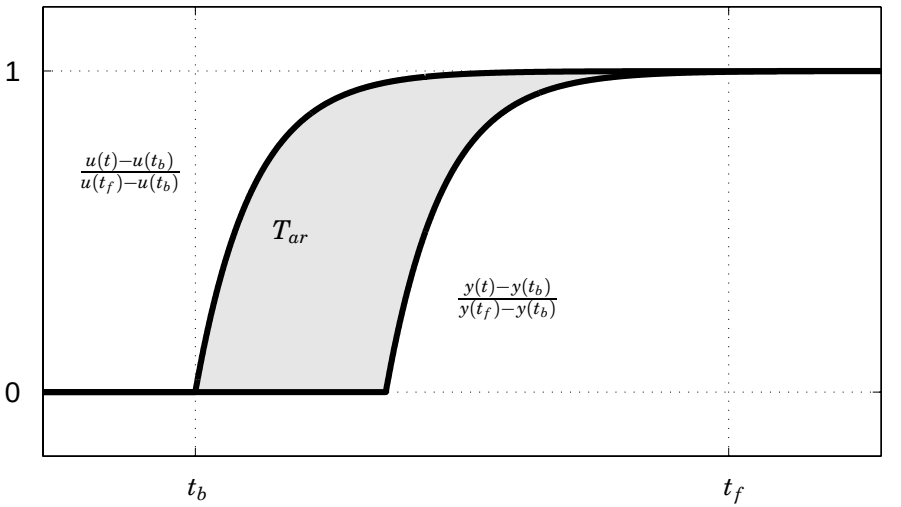

Figure 1. Graphical explanation for Eq. (10)

In Eqs. (11) and (12) T_n is given for a step or ramp input signal.

$$\text{Step}: \quad T_n = \frac{Ae^1}{hK_n} \tag{11}$$

$$\text{Ramp}: \quad T_n = \sqrt{\frac{A}{hK_n(1/2 - e^{-1})}} \tag{12}$$

Parameter A is the integral of $y(t)$ from 0 to T_{ar}. The parameter h is the amplitude for a step signal or the rate for a ramp signal. The length of the open-loop experiment is always T_{ar}.

Notice that Eq. (12) can be used for a FOPDT model with integrator as well. Sending a step of height h to an integrating process is the same as sending a ramp with rate h to the process without an integrator. In both cases Eq. (12) can be used to find T_n. The problem with the open-loop ramp experiment is that if the dead time L_n is sufficiently larger than time constant T_n, one has to select a large h for a good signal to noise ratio in integral A. But after time T_{ar}, $y(t)$ will continue to rise since the process contains an integrator. Assuming that after time T_{ar} precautions are taken to reverse the direction of $y(t)$ the maximum value of $y(t)$, will still be around hK_nT_{ar}. This gives then an upper limit on rate h.

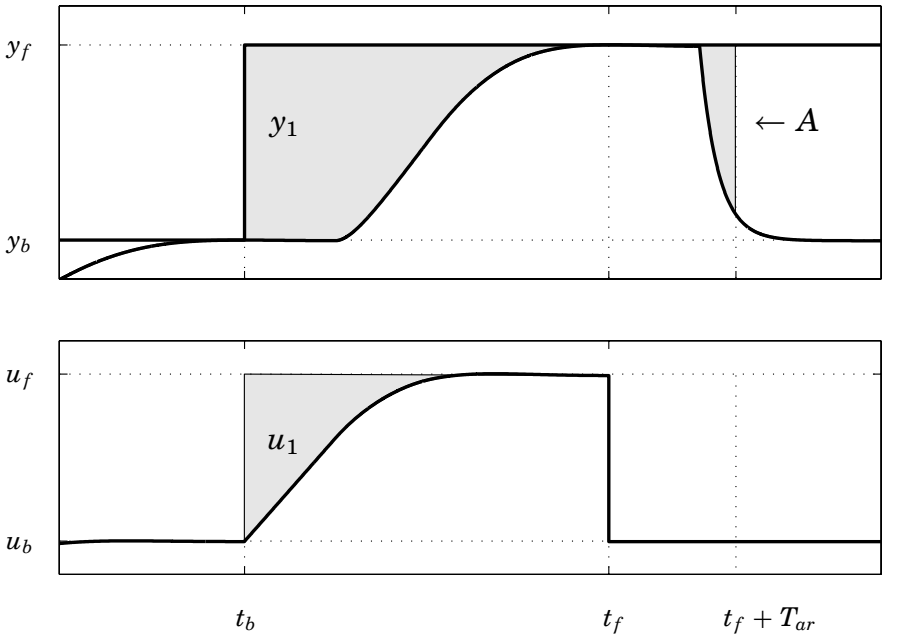

Figure 2. Identification experiment for $G(s) = \frac{1}{s+1}e^{-5s}$

Identification procedure

The basic steps in the identification procedure can now be presented. In Fig. 2, the input and output signals are shown for a specific FOPDT process. For simplicity the change in set point for the closed-loop experiment is a step. The open-loop experiment is also a step.

1. Control the process to a steady state initial level y_b. Record the signal levels u_b and y_b.
2. Apply a step in the reference signal $y_{sp}(t)$ at time t_b
3. Integrate $y(t)$ and $u(t)$ until process reaches steady state again. This occurs at time t_f. Again record the signal levels u_f and y_f.
4. Determine process gain K_n by observing the signal levels and T_{ar} from Eq. (10).
5. Apply a step in open loop and integrate the area A using the estimate of T_{ar} obtained from previous step.
6. Estimate time constant T_n from Eq. (11) and dead time by $L_n = T_{ar} - T_n$.

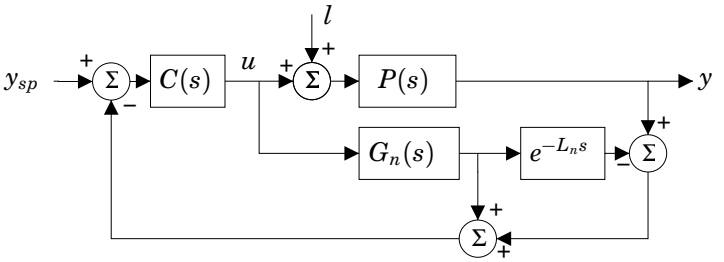

Figure 3. Block diagram of a Smith predictor.

The method requires a decision criterion on when steady state has been achieved. This occurs at times t_b and t_f. If integral action is present in the controller the steady-state error should be zero. This reduces the task to determining when $y(t)$ is the same as the set point $y_{sp}(t)$.

In the presence of noise the steady state values u_b, u_f, y_b and y_f can be determined by averaging the signals over a period of time.

In the presence of measurement noise the numerical evaluation of the integrals might have large variance. Less variance would be assured by proper noise filtering. The noise filter would then of course be considered as part of the controlled process.

3. Dead-time compensation

The most common dead-time compensating controller is the Smith predictor (SP) [Smith, 1957]. The structure of this controller is shown in Fig. 3. The controller output u is fed through a model of the process and through the same model without dead time. In this way, the controller acts, in the ideal situation of perfect modeling, on a simulated process which behaves as if there was no dead time in the process.

The real process is assumed to be linear time invariant and is denoted as $P(s)$. The model is denoted $P_n(s) = G_n(s)e^{-L_n s}$. $G_n(s)$ is delay free.

A SP which uses the simple FOPDT model given by Eq. (1) combined with a PI controller

$$C(s) = K\left(1 + \frac{1}{sT_i}\right) \tag{13}$$

requires five parameters to be determined, namely K, T_i, K_n, T_n, and L_n. These parameters may be obtained from a systematic process identification experiment. However, it is practically impossible to tune this controller manually by trial and error procedures. Therefore, replacing a PID

controller with a standard SP gives a drastic increase in operational complexity. This increase is present in both the commissioning and maintenance of dead-time compensating controllers.

A common way to deal with this complexity is to automize the tuning procedure. Automatic tuning of DTCs has received some attention in the literature, some references are [Palmor and Blau, 1994], [Lee *et al.*, 1995] and [Vrancić *et al.*, 1999]. But even when automatic tuning procedures are available simpler structures are advantageous since it provides a possibility for the user to make the last final adjustments manually or manually retune the controller later.

A few papers have been written that emphasize the importance of less complex DTCs. In the stable case, [Hägglund, 1996] is one example and in the case of integrating processes ,[Matausek and Micic, 1999] have addressed the problem.

The bandwidth of DTCs is usually related to the model parameters which are assumed to be available when the DTCs are initially tuned. In [Palmor and Blau, 1994], the closed-loop time constant was set proportional to the apparent dead-time of the process. In [Hägglund, 1996] it was related to the open-loop apparent time constant. In [Normey-Rico *et al.*, 1997] the closed-loop bandwidth was related to both of these.

In the case of integrating processes, it has been more common that the initial bandwidth is supposed to be manually tuned. Guidelines are given from where a starting point can be obtained. In [Normey-Rico and Camacho, 1999] the closed-loop time constant was related to an assumed dead-time error between the model and process. In [Matausek and Micic, 1999] it was suggested that the closed-loop time constant should be set equal to the apparent time constant of the dynamics additional to the integrator.

In this paper, a new approach is taken to determine the closed-loop time constant in the initial tuning. Given the model parameters it is possible to calculate the uncertainty norm boundary of the DTCs. The uncertainty norm boundary tells how much the real process can deviate from the model at each frequency without the closed-loop system becoming unstable. Then it is shown how it is possible with simple experiments to obtain frequency dependent inequalities bounding the model uncertainty. A lower bound on the closed-loop time constant is then found by making sure the model uncertainty found is always less than the uncertainty norm boundary of the DTC's. The goal is that this initial tuning is, when the model is close to the process, less conservative but robustly stable.

It was mentioned in the introduction that classical measures of robustness such as gain and phase margin are not sufficient when dealing with dead-time systems. This is discussed in [Palmor, 1980]. In addition to these classical ones it is proposed that a third one is used, namely the

delay margin. The delay margin of a closed-loop system can be defined in the following way (modifying slightly the definition in [Landau *et al.*, 1995]). If the Nyquist curve intersects the unit circle at frequencies ω_i with the corresponding phase margins Φ_i then the delay margin can be defined as

$$D_M = \min_i \frac{\Phi_i}{\omega_i}$$

Most of the tuning rules for DTCs presented in the literature provide a certain delay margin. The initial tuning, which results from the procedures in this paper, can have an arbitrary small delay margin if the model describes the process well. Therefore, it is shown how the DTCs can be retuned with a guaranteed delay margin. This can have a practical value when it is known how much the dead time might vary around the operating point. Finally, it is also shown what delay margin can be expected from the initial tuning in the nominal case, i.e., when the model and process are equal.

4. Stable Case

In [Hägglund, 1996], a dead-time compensating controller with only three adjustable parameters was presented. The controller can be viewed as a PI controller with model-based prediction. The abbreviation PPI stands for "Predictive PI". The controller can be tuned manually in the same way as a PID controller.

The structure of the PPI controller is the same as for the Smith predictor, with the FOPDT model (1) combined with the PI controller (13). The only difference is the parameterization. The five adjustable parameters are reduced to three by introducing constraints between the controller parameters and the model. These constraints are

$$\begin{aligned} T_n &= T_i \\ K_n &= \kappa / K \end{aligned} \tag{14}$$

κ is a constant to be determined later. The identification method presented before only provides a good approximation of the real process at low frequencies. Robustness problems for PPI can occur because of model error at high frequencies. In [Normey-Rico *et al.*, 1997] a filter was proposed to provide robustness towards high frequency model errors. The resulting controller was abbreviated FPPI. The proposed controller structure can be seen in Fig. 4. The filter $F(s)$ is typically a one parameter low-pass filter

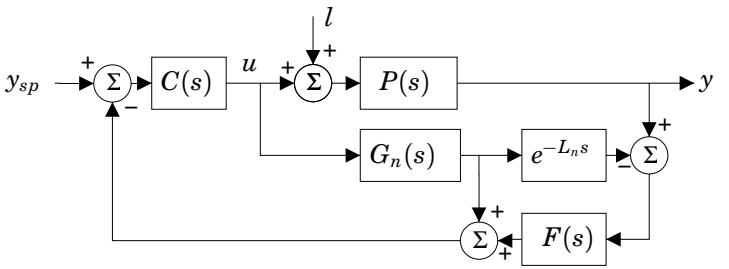

Figure 4. The FPPI

and it is shown later that it is sufficient for the purpose of this paper to assume it is of first order

$$F(s) = \frac{1}{T_f s + 1}$$

In the nominal case, that is when the model describes the process perfectly, $P = P_n$, the controller parameterization given by Eq. (14) results in a closed-loop set point response

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{1}{T_r s + 1} e^{-L_n s} \tag{15}$$

where $T_r = T_n/\kappa$. This is a familiar result regarding Smith-predictors. The time constant of the closed loop system can be reduced by increasing κ. Notice that $\kappa = 1$ places the closed-loop pole in the same place as the open-loop one. A proper selection of T_r will be the main subject of the next subsection. It is shown how a lower limit on T_r can be obtained by performing a simple open-loop step experiment.

Selection of T_r

The closed loop characteristic equation for the FPPI is

$$1 + \frac{C}{1 + CG_n} F(P - P_n) = 0 \tag{16}$$

Denoting the difference between model and process as $\delta P = P - P_n$ the maximum value of $|\delta P|$ allowable while maintaining closed loop stability, or the uncertainty norm boundary can be obtained from Eq. (16) by solving for δP

$$|\Delta P|_{\text{FPPI}} = K_n \left| \frac{(i\omega T_r + 1)(i\omega T_f + 1)}{(i\omega T_n + 1)} \right| \tag{17}$$

Notice that if an inequality of type

$$|\delta P| \leq A\omega \tag{18}$$

is available the system can be made stable by chosing appropriately T_r and T_f. This follows from the fact that the degree of the denominator polynomial is one higher than the numerator polynomial. The condition for stability would then be

$$|\delta P| \leq A\omega < |\Delta P|_{\text{FPPI}} \quad \forall\omega \tag{19}$$

An uncertainty bound of type (18) can be obtained with a simple open-loop step experiment. The step response of the real system is denoted by $y(t)$. After an identification experiment the FOPDT model response is available and given by

$$y_n(t) = \begin{cases} 0 & \text{for } t < L_n \\ K_n(1 - e^{-(t-L_n)/T_n}) & \text{for } t > L_n \end{cases} \tag{20}$$

Denoting the difference between the two responses $f(t) = y(t) - y_n(t)$, the following expression is the definition of the Laplace transform

$$\frac{P(s) - P_n(s)}{s} = \int_0^\infty e^{-st} f(t) dt \tag{21}$$

Putting $s = i\omega$ the following equation is obtained

$$\begin{aligned} \left| \frac{P(i\omega) - P_n(i\omega)}{i\omega} \right| &= \left| \int_0^\infty e^{-i\omega t} f(t) dt \right| \\ &\leq \int_0^\infty |f(t)| dt \\ &= A \end{aligned} \tag{22}$$

Note that the error is weighted with one over ω. At stationarity therefore there can be no error. Therefore $P_n(s)$ and $P(s)$ have to have the same steady state gain K_n.

If a time constant $\bar{T}_f = A/K_n$ is defined the relevant areas can be graphically displayed on a normalized step response. This is shown in Fig. 5.

Inequality (19) can now be restated the following way

$$\left\| \frac{\bar{T}_f \omega (iT_n\omega + 1)}{(iT_r\omega + 1)(iT_f\omega + 1)} \right\|_\infty < 1 \tag{23}$$

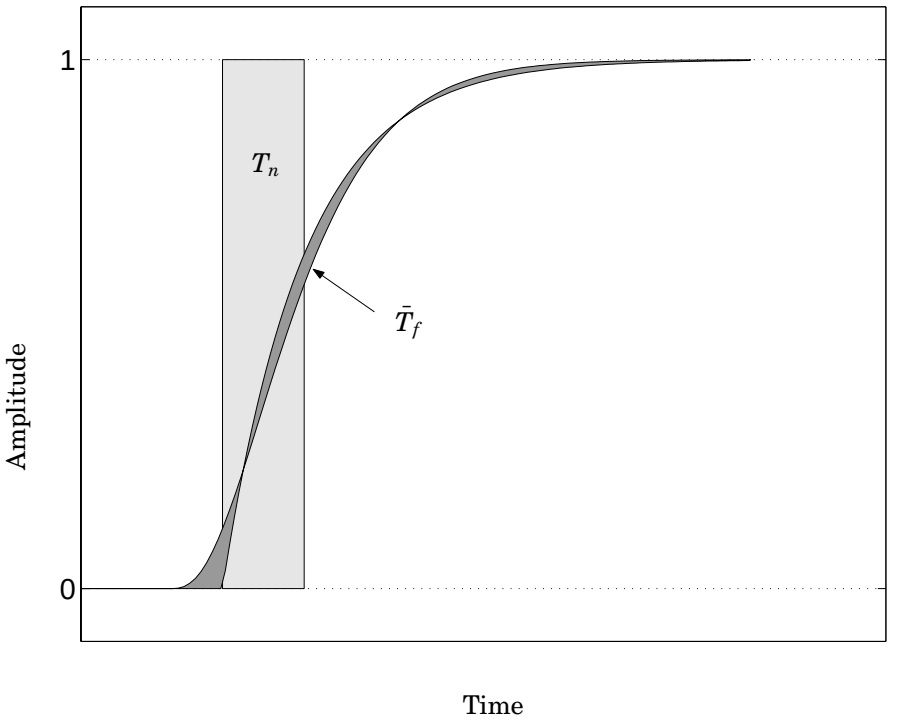

Figure 5. Step responses $y(t)$ and $y_n(t)$ normalized to 1.

Notice that $\bar{T}_f$ and T_n are assumed to be known while T_r and T_f are design parameters. The latter two should be chosen to minimize some performance criteria while fulfilling the above inequality. For a fast set point response, T_r could be chosen small while T_f would be used to fulfill the above inequality. Commonly in process control, regulatory performance is considered more important. The performance criteria recommended here is to minimize the integrated error when load disturbance l is a unit step. Using the final value theorem the following expression is obtained.

$$\int_0^\infty e(t)dt = (T_r + T_f + L_n)K_n \tag{24}$$

The design problem can be set up as a minimization problem where Eq. (24) is the cost function and Eq. (23) is the constraint. This problem can be further simplified. Notice that T_r and T_f enter the cost function and the constraint the same way. Using this fact it is possible to obtain

necessary conditions that show that at the optimum, T_r is equal to T_f. Still the analytic solution to this problem quickly becomes rather involved. A necessary condition for the constraint in Eq. (23) to be fulfilled is that the transfer function has a direct term less than 1. This gives the following condition

$$T_r > \sqrt{T_n \bar{T}_f} \tag{25}$$

A further simplification of the problem can be obtained by normalizing the frequency in inequality (23) with T_r. If the following quantities are defined

$$\bar{\omega} = T_r \omega \quad \gamma = \bar{T}_f / T_r$$

inequality (23) can be written as

$$\left\| \frac{\gamma \bar{\omega}(i\kappa\bar{\omega} + 1)}{(i\bar{\omega} + 1)^2} \right\|_\infty < 1 \tag{26}$$

κ was defined following Eq. (15). Using a bisection algorithm, an upper limit on κ for which inequality (26) holds, was calculated as a function of γ. This is shown in Fig. 6. Any pair of κ and γ that lies within the shades area fulfills inequality (26). Using the above figure the following design rules are proposed

$$\gamma\kappa \leq 1 \qquad \gamma \leq 1 \tag{27}$$

In terms of the time constants this becomes

$$T_r \geq \sqrt{\bar{T}_f T_n} \tag{28}$$

$$T_r \geq \bar{T}_f \tag{29}$$

Remark. The use of an equality in Eq. (28) requires justification. Inequality (22) allows infinite error when $\omega \to \infty$. Most normal processes are on the other hand of low-pass character. This means the inequality could be replaced with a strict inequality at high frequencies. When $\kappa\gamma = 1$ the supremum of the norm in Eq. (26) is achieved when $\omega \to \infty$. So using additional information about inequality (22), Eq. (28) can be justified.

Sensitivity to dead-time errors

Robustness of DTCs has been analyzed by many authors. Some references are [Morari and Zafiriou, 1989], [Palmor, 1980] and [Lee *et al.*, 1996]. Usually, most attention is devoted to analyzing the sensitivity towards errors in the dead time. The reason for this is that it is often towards these errors dead-time compensators are most sensitive.

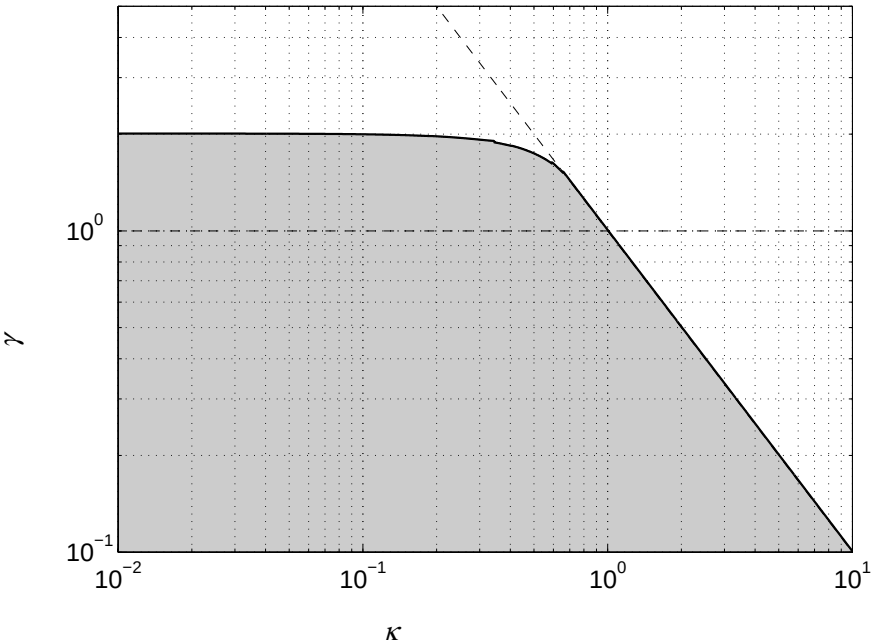

Figure 6. κ and γ within the shaded area fulfill inequality (26). The edge of the shaded area corresponds to the equality. Also shown are design lines $\kappa = 1/\gamma$ and $\gamma = 1$

With this in mind it is important to give easy ways to reduce the dead-time sensitivity and be able to set a desired delay margin without limiting the bandwidth, i.e. sacrificing performance, to much. If the dead time is expected to vary an amount ΔL, the increase in integral A in Eq. (22) is at most $K_n \Delta L$. So an arbitrary delay margin can be set by adding the expected dead-time variation to $\bar{T}_f$ and recalculating T_r with Eq. (28).

Using the result of [Palmor, 1980] one can see that with the tuning presented in this section, the resulting DTC has phase margin 60^o and gain margin equal to 2 in the nominal case. The way to see this is to note that the loop gain is

$$L(i\omega) = \frac{e^{-iL_n\omega}}{(iT_r\omega + 1)^2 - e^{-iL_n\omega}} \tag{30}$$

which can be written as

$$L(i\omega) = \frac{N(i\omega)}{1 - N(i\omega)}$$

where $N(i\omega)$ is a frequency response with amplitude less than 1 for all ω. Loop gain $L(s)$ is always written as a function of a complex variable to distingish it from dead time L.

Further insight can be obtained into the relation between the delay margin and closed-loop time constant T_r by normalizing the variables of the loop gain with L_n. This is the same approach as was taken in [Palmor and Blau, 1994]. If the process is equal to the model except for an error in the dead-time, $L = L_n + \delta L$, the loop gain becomes

$$L(s) = \frac{e^{-Ls}}{(T_r s + 1)^2 - e^{-L_n s}} \tag{31}$$

If the variables are normalized with L_n the following dimensionless variables are obtained.

$$\delta_L = \delta L / L_n \quad \bar{\omega} = \omega L_n \quad \bar{T}_r = T_r / L_n \tag{32}$$

The normalized loop gain is then

$$\bar{L}(i\bar{\omega}) = \frac{e^{-i\bar{\omega}(1+\delta_L)}}{(i\bar{T}_r\bar{\omega} + 1)^2 - e^{i\bar{\omega}}} \tag{33}$$

The relationship between the two loop gains is $\bar{L}(i\omega L_n) = L(i\omega)$. By using the Nyquist criterion it is possible to calculate a lower limit on $\bar{T}_r$ for which the closed loop is stable, as a function of δ_L. This is shown in Fig. 7. The figure shows that if $\bar{T}_r \geq 0.22$, δ_L can vary from -1 to 1 without the closed-loop system becoming unstable. It also shows that very aggressive tuning, $\bar{T}_r = 0.01$, would give a relative delay margin from -0.2 to 0.15.

It is of interest to see what time constant is obtained by specifying a desired delay margin, ΔL and calculating $\bar{T}_r$ from Eq. (28). This desired time constant is referred to as $T_{r,d}$. Dividing by L

$$\bar{T}_{r,d} = \sqrt{\frac{\Delta L}{L}\frac{T_n}{L}}$$

In Fig. 7 $\bar{T}_{r,d}$ is plotted as a function of $\Delta L/L$ for different values of T_n/L. Also shown is $\bar{T}_{r,d}$ calculated with Eq. (29). The recommended $\bar{T}_{r,d}$ is well above the stability limit so a certain degree of robust performance should be assured. Notice that when T_n is close to L_n the tuning can be quit conservative. Increasing dead time results in lower delay margins.

The lower limit on $\bar{T}_r$ in Fig. 7 is calculated for the nominal case. Experience from simulations indicate that this lower limit is not a bad approximation for other processes, specially if $\bar{T}_f/T_n$ is little. Notice that the method of setting a desired delay margin by adding ΔL to $\bar{T}_f$ is valid for any stable process.

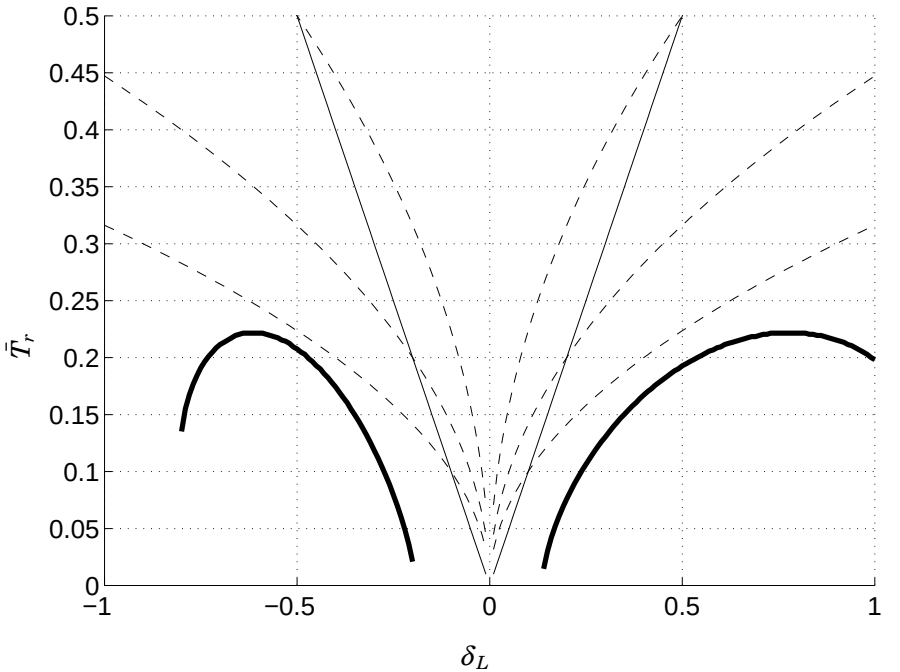

Figure 7. Thick line is the lower bound on $\bar{T}_r$ for a given error δ_L. The dashed lines are the design recommendation, $\bar{T}_{r,d}$ for T_n/L_n equal to 0.5, 0.2 and 0.1 counting from above. Finally the thin solid line is the design rule given by Eq. (29).

Practical issues

Some comments are in order relating to the practical use of the FPPI and specially the selection of T_r.

- If the model is close to the real process, $\bar{T}_f$ will be small resulting in a small T_r. Other limiting factors such as actuator saturation will then come into the picture.
- The performance of the FPPI is always limited by the dead time of the process. Therefore, tuning aggressively by selecting a small T_r might only give marginal improvements. In the nominal case this is apparent from Eq. (24). This fact can be used to motivate tuning rules. One can decide how large portion of the integrated error comes from L_n, which one can do nothing about. From this and equation relating L_n, T_r can easily be obtained.
- Looking at Fig 7 while keeping Eq. (24) in mind, the tradeoff be-

Table 1. FOPDT model parameters as well as A and κ

i	T_n	L_n	$\bar{T}_f$	T_r	κ
1	10.4	6.8	0.27	1.7	6.2
2	2.0	6.2	0.27	0.7	2.7
3	3.0	9.1	1.06	1.8	1.7
4	1.7	8.4	1.18	1.4	1.2
5	0.8	5.4	0.17	0.4	2.4
6	7.6	2.6	1.32	3.2	2.4
7	1.3	5.7	0.17	0.5	2.7

tween performance and robustness is apparent. Robustness in terms of delay margin costs in terms of increased T_r.

Simulation examples

Simulation results for the PPI and the FPPI have been presented in [Hägglund, 1996; Normey-Rico *et al.*, 1997; Ingimundarson and Hägglund, 2000b]. To give an idea of how conservative the design method presented is, a FPPI controller was designed for a collection of processes. They are shown here without the dead time.

$$
\begin{aligned}
P_1(s) &= \tfrac{1}{(10s+1)(2s+1)} & P_2(s) &= \tfrac{1}{(s+1)^3} \\
P_3(s) &= \tfrac{-s+1}{(s+1)^5} & P_4(s) &= \tfrac{-2s+1}{(s+1)^3} \\
P_5(s) &= \tfrac{9}{(s+1)(s^2+2s+9)} & P_6 &= \tfrac{0.5}{(s+1)} + \tfrac{0.05}{s+0.1} \\
P_7(s) &= \tfrac{64}{(s+1)(s+2)(s+4)(s+8)}
\end{aligned}
$$

The dead time was always equal to 5 making the total process equal to

$$P(s) = P_i(s)e^{-5s}$$

for i ranging from 1 to 7.

The results can be seen in Table 1. Notice that T_r is in all cases smaller than T_n which means that the closed-loop system has a faster step set-point response than the open-loop one. Notice that the two processes with smallest T_n/T_r ratio are non minimum phase and not monotonically decreasing. Since the response goes in the wrong direction in the beginning the area A becomes quite large in those cases. This reduces the bandwidth through $\bar{T}_f$.

Summary of tuning procedure

The tuning procedure starts with an identification of a FOPDT model resulting in parameters K_n, T_n and L_n. The open-loop step response in the final stage of the identification is used to calculate area A in Eq. (22). With this information all parameters of the controller can be found. First, $\bar{T}_f$ is found as $\bar{T}_f = A/K_n$. Given a wanted delay margin ΔL, κ is calculated as

$$\kappa = \sqrt{\frac{T_n}{\bar{T}_f + \Delta L}} \tag{34}$$

The value of controller parameter T_f should be equal to

$$T_f = T_n/\kappa = T_i/\kappa \tag{35}$$

Using Eq. (14) the FPPI has then 3 parameters to tune, namely K,T_i and L_n. The controller parameters T_i and K respond as in a normal PI controller.

Application to a tank lab process

The above mentioned methods were applied to a tank laboratory process at the Department of Automatic Control in Lund. It consists of a tank with free outflow and a level sensor as well as a pump. The process has a long dead-time because the pump pumps the water into an open channel with a small slope. Obvious nonlinearities in the process are the relation of outflow to the height in the tank as well as nonsymmetry because the pump cannot remove water from the tank. A PI controller was tuned and a closed-loop experiment performed followed by a an open-loop step experiment. The result of the open-loop experiment as well as the corresponding FOPDT response are shown in Fig. 8. The sampling time was 1 second. The identified FOPDT model was

$$P_n(s) = \frac{5.6}{40.2s + 1} e^{-93.9s}$$

From the area between the responses $\bar{T}_f$ was found to be equal to 4.4. This gave $T_r = 13.3$. In Fig. 9 a set-point step response can be seen. The gain in the proportional part of the PI controller was set to 0 from the set point to achieve a smoother response. The controller was started with a reference value of 4 cm. Then an additional step was applied at time 1000 taking the level to 8cm. The over shoot is caused by unexplained nonlinearities. Reducing κ considerably did not eliminate it. Otherwise, a smooth control is observed.

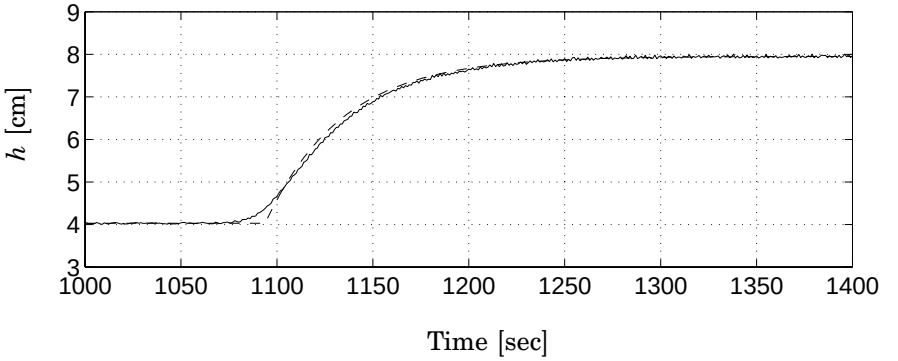

Figure 8. Open loop step response of the real tank (solid line), and the model (dashed line).

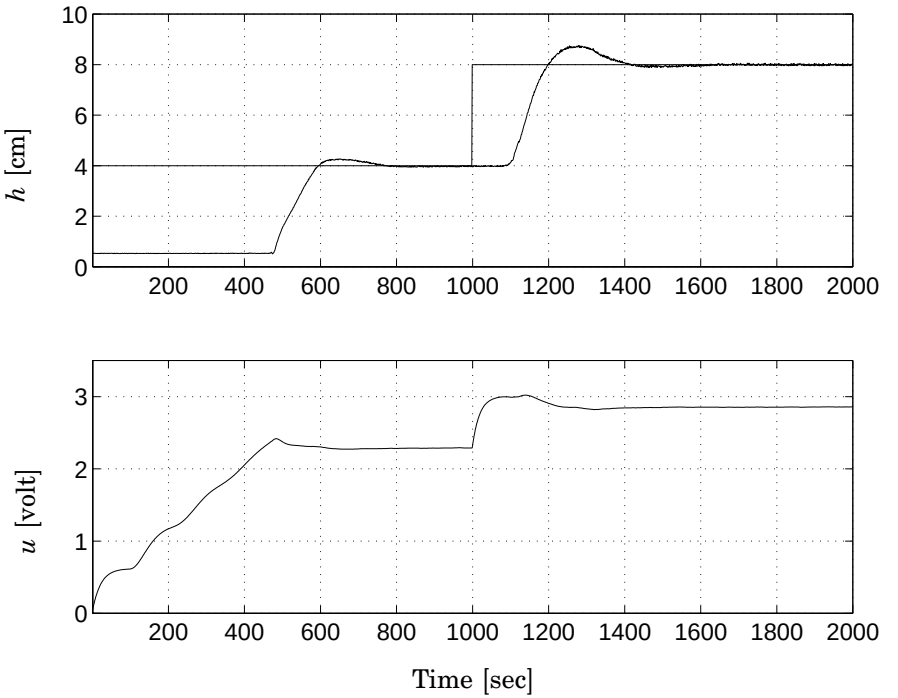

Figure 9. Closed loop step response of the real tank with a FPPI controller.

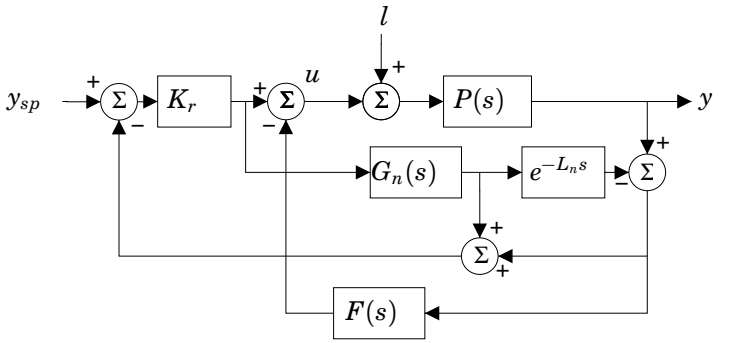

Figure 10. The improved dead-time compensating controller for integrating processes.

5. Unstable case, integrating processes

The extension of the Smith predictor to the unstable plant case was first presented in [Watanabe and Ito, 1981]. The compensator used here is the Modified Smith Predictor (MSP) presented in [Matausek and Micic, 1996]. There, care was taken to make sure all parameters could be related to identified parameters of the process model or to classical control theory concepts.

The block diagram of the dead-time compensator can be seen in Fig. 10. The model of the plant is the simple two-parameter model

$$P_n(s) = G_n(s)e^{-L_n s} = \frac{K_n}{s}e^{-L_n s} \tag{36}$$

The transfer function $F(s)$ is a constant K_0 which is related to the two-parameter model parameters as

$$K_0 = \frac{1}{2L_n K_n} \tag{37}$$

Introducing

$$T_r = \frac{1}{K_n K_r} \tag{38}$$

the transfer function from set point value $y_{sp}(t)$ is in the nominal case given by

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{1}{T_r s + 1}e^{-L_n s} \tag{39}$$

T_r has therefore the nice interpretation of being the time constant from set point to the output signal.

From the above equations it is clear that given a process model $P_n(s)$ the only parameter left to determine is T_r. That is the subject of the next section.

Determining T_r

In [Normey-Rico and Camacho, 1999] a DTC for integrating processes was proposed whose closed-loop time constant was related to asymptotes of the uncertainty norm-bound of the DTC. The same approach is taken here.

The error between the plant and the model is

$$\delta P(s) = P(s) - P_n(s) \tag{40}$$

The uncertainty norm-bound for the MSP or the maximum value of $|\delta P|$ allowed while keeping closed loop stability is

$$|\Delta P|_{\text{MSP}} = \frac{|K_n(j\omega T_r + 1)(2L_n j\omega + e^{-jL_n\omega})|}{|-(2L_n + T_r)\omega^2 + j\omega)|} \tag{41}$$

This equation is obtained from the characteristic equation in a similar way as Eq. (17).

As pointed out in [Normey-Rico and Camacho, 1999] this bound depends almost entirely on T_r. The bound is shown in Fig. 11. For low frequencies the bound behaves as

$$|\Delta P(i\omega)| \approx \frac{K_n}{\omega}$$

For high frequencies the bound has an asymptote given by

$$\lim_{\omega\to\infty} = \frac{2K_nL_nT_r}{2L_n + T_r} \tag{42}$$

To evaluate the minimum of the uncertainty norm-bound it is fruitful to consider it a product of two factors, one of which is monotonically decreasing. The one that has local minima is

$$f(\omega) = \left|\frac{2jL_n\omega + e^{-jL_n\omega}}{\omega}\right|$$

The minimum value of this function can be approximated with L_n. This gives a lower bound of $|\Delta P(i\omega)|$

$$\beta = \frac{K_nL_nT_r}{2L_n + T_r} \leq \min_{\omega} |\Delta P(i\omega)| \tag{43}$$

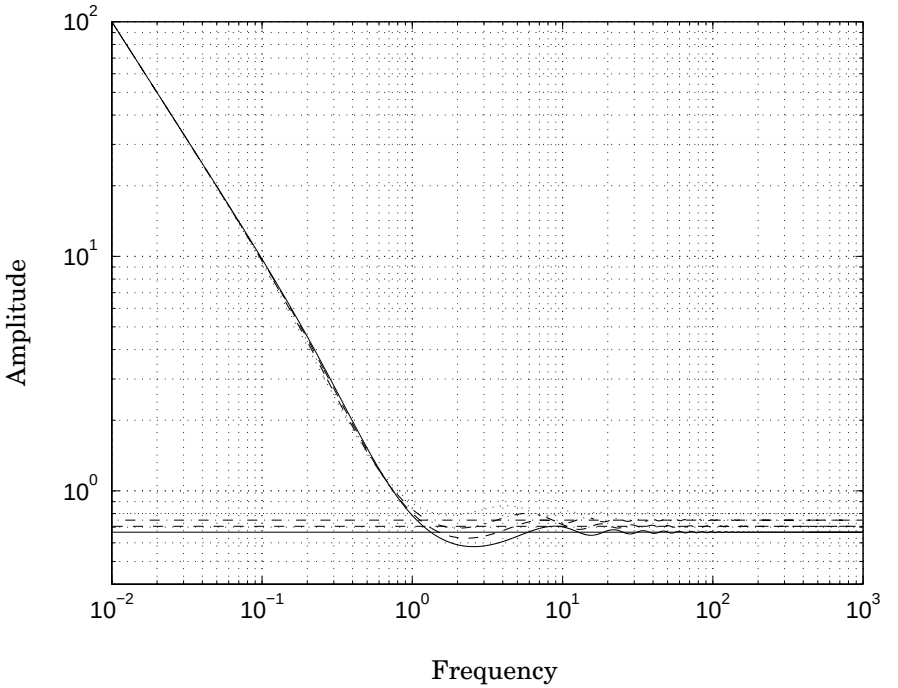

Figure 11. Uncertainty norm boundary for $T_r = K_n = 1$ for varying L_n equal to 1 (−), 3 (−−), 5 (− · −) and 10 (· · ·).

Taking this into account when solving for T_r gives

$$T_r = \frac{2L_n\beta}{K_nL_n - \beta} \tag{44}$$

From closer inspection of Fig. 11 it can be concluded that for L_n larger than 1, the bound given by Eq. (43) is conservative, and that the real minimum is much closer to the value given by Eq. (42).

To evaluate the value of $|\delta P(i\omega)|$ for each frequency, the assumption is made that the real process is given by a first order plus dead time transfer function with an integrator

$$P(s) = \frac{K_p}{s(Ts+1)}e^{-Ls}$$

The absolute value of $\delta P(i\omega)$ is then

$$|\delta P(i\omega)| = \left| \frac{K_p e^{-iL\omega} - K_n(iT\omega+1)e^{-iL_n\omega}}{(iT\omega+1)i\omega} \right|$$

As pointed out in [Normey-Rico and Camacho, 1999] this function can be characterized by three main frequency intervals. For small ω this function behaves as

$$|\delta P(i\omega)| \approx \left| \frac{K_p - K_n}{\omega} \right|$$

For large ω this function has a slope of 20dB/dec. For sufficiently high and low frequencies $|\delta P|$ will be smaller than $|\Delta P|$. It is the error at intermediate frequencies that is of most interest. This will be referred to as δP_0 in the following.

In [Normey-Rico and Camacho, 1999] the value at intermediate frequencies was approximately calculated by assuming that the velocity gains of the model and process were equal, $K_p = K_n$. Then δP_0 could be estimated as

$$\delta P_0 = \lim_{s\to 0} \delta P(s) = K_n(L + T - L_n) \tag{45}$$

In [Normey-Rico and Camacho, 1999] δP_0 is viewed as a tuning parameter from which the closed-loop time constant is calculated. The methodology is therefore to assume an error between $L+T$ and L_n and from there get the initial tuning.

Here the approach is to perform a simple open-loop experiment from where a upper bound is found on δP_0. The area between the actual response and the response of the model is calculated when the input is an impulse of height h_p and duration τ_p. Denoting as before $f(t) = y(t) - y_n(t)$ the following equation is the definition of the Laplace transform.

$$\delta P(s)\frac{h_p(1-e^{-\tau_p s})}{s} = \int_0^\infty e^{-st} f(t)dt \tag{46}$$

From this equation, by replacing the argument s with $i\omega$ the following inequalities are obtained.

$$\begin{aligned} \left| \delta P(i\omega)\frac{h_p(1-e^{-\tau_p i\omega})}{i\omega} \right| &= \left| \int_0^\infty e^{-i\omega t} f(t)dt \right| \\ &\leq \int_0^\infty |f(t)|dt \end{aligned}$$

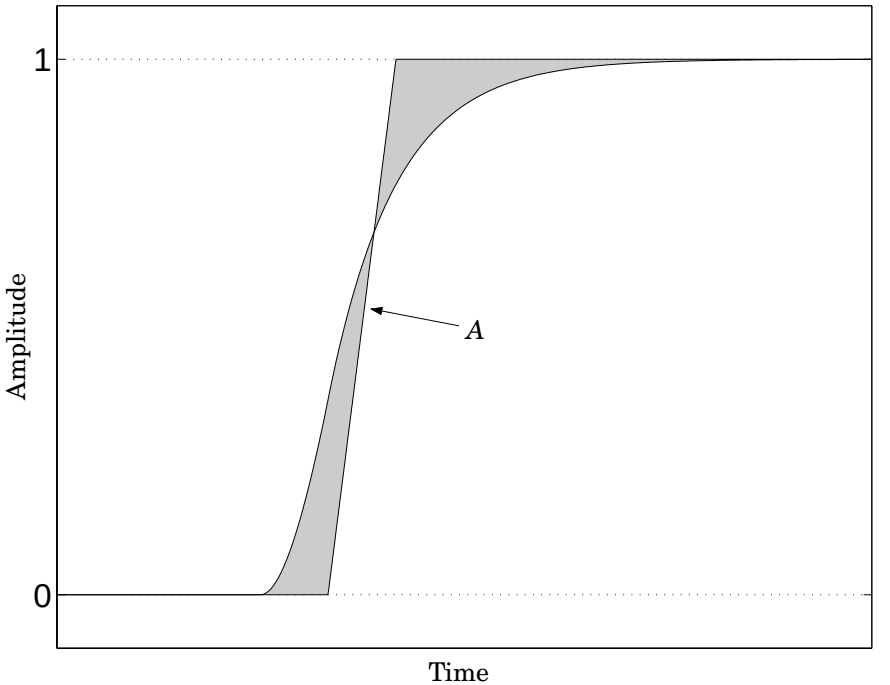

Figure 12. The area A for the integrating case.

Notice that an upper bound on δP_0 can be obtained as

$$\delta P_0 \leq \frac{\int_0^\infty |f(t)|dt}{h_p \tau_p} = A \tag{47}$$

The area A has a nice graphical interpretation if the impulse response of the model and real process is normalized to 1 and plotted on the same graph. A is then the area between the curves. This is shown in Fig. 12.

Notice that this bound on δP_0 is always larger and therefore more conservative than the value obtained by Eq. (45). Since $\delta P(i\omega)$ is weighted with $(1 - e^{-i\tau_p\omega})/\omega$ in the inequality, $\delta P(i\omega)$ can be arbitrary large when $\tau_p\omega = 2\pi$. The above method therefore does not guarantee stability but should work well for well-behaving processes. The gain is that it eliminates the need to tune the last parameter. Substituting β with A in Eq. (44) gives then the time constant T_r.

Sensitivity to dead-time errors

Given an error ΔL in the dead-time of the real process, the increase in integral A would be maximum $\Delta L K_n$. Suspected variations in the dead time can therefore be taken into account by increasing β in Eq. (44) by $\Delta L K_n$.

If it is assumed that the process has the same structure as the model but a different dead time, $L = L_n + \delta L$ the loop gain of the MSP is

$$L(s) = \frac{1}{2} \frac{e^{-Ls}((2L_n + T_r)s + 1)}{sL_n(T_r s + 1 - e^{-L_n s})} \tag{48}$$

Using the same approach as in the previous section the following dimensionless variables are defined

$$\delta_L = \Delta L / L_n \qquad \bar{\omega} = \omega L_n \qquad \bar{T}_r = T_r / L_n \tag{49}$$

This gives the following dimensionless loop gain

$$\bar{L}(i\bar{\omega}) = \frac{1}{2} \frac{e^{-i(1+\delta_L)\bar{\omega}}(i(2 + \bar{T}_r)\bar{\omega} + 1)}{i\bar{\omega}(i\bar{T}_r\bar{\omega} + 1 - e^{i\bar{\omega}})} \tag{50}$$

Fig. 13 shows the lower bound on $\bar{T}_r$ to maintain stability as a function of δ_L. The bound is not symmetric around 0. Rather it is shown that an increase in dead time is more likely to cause instability than a decrease. If $\bar{T}_r$ is larger than 0.4 it will be stable for any decrease in δ_L down to -1. For an initial tuning which would give $\bar{T}_r$ equal to 0.01, the controller would be stable for $\delta_L \in [-0.22, 0.14]$.

The design rule given by Eq. (44) can be rewritten relating $\bar{T}_r$ to an assumed error ΔL by using $\beta = \Delta L K_n$. This results in

$$\bar{T}_{r,d} = \frac{2\delta_L}{1 - \delta_L} \tag{51}$$

This function is also shown in Fig. 13. The suggested $\bar{T}_{r,d}$ is well above the stability limit.

Simulation examples

To get an idea about what closed-loop time constant one can expect to obtain with the presented methodology, a two-parameter model was found for a collection of processes. The dynamics additional to the integrator and

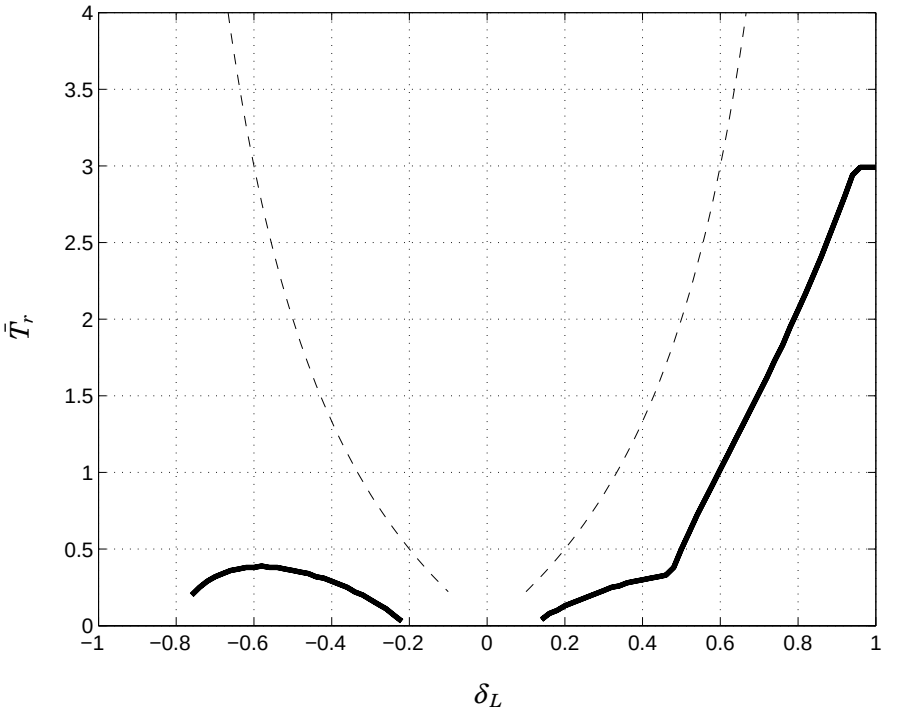

Figure 13. Thick line is the lower bound on $\bar{T}_r$ for a given error δ_L. The dashed line is the design recommendation, $\bar{T}_{r,d}$ according to Eq. (44).

dead time are shown below as $P_1(s) - P_8(s)$

$$P_1 = \frac{1}{s+1} \qquad P_2 = \frac{1}{(0.1s+1)(s+1)}$$
$$P_3 = \frac{1}{(s+1)^3} \qquad P_4 = \frac{-s+1}{(s+1)^5}$$
$$P_5 = \frac{-2s+1}{(s+1)^3} \qquad P_6 = \frac{1}{(s+1)(s^2+2s+9)}$$
$$P_7 = \frac{0.5}{(s+1)} + \frac{0.05}{(s+0.1)}$$
$$P_8 = \frac{64}{(s+1)(s+2)(s+4)(s+8)}$$

The dead time, L, was equal to 5 in all simulations. The total process was therefore

$$P(s) = P_i(s)\frac{1}{s}e^{-5s}$$

The resulting T_r is shown in Table 2. In [Matausek and Micic, 1996] it was suggested that a suitable value of T_r, given that this value would be

Table 2. Identification and tuning results for integrating processes.

i	T_n	T_r	A	L_n
1	1	0.8	0.4	6.0
2	1.0	0.8	0.4	6.1
3	2.0	2.1	0.9	8.0
4	3.0	2.8	1.2	11.0
5	1.7	3.2	1.4	10.0
6	0.8	0.7	0.3	6.2
7	7.6	21.5	5.3	10.6
8	1.3	1.1	0.5	6.9

available, could be the apparent time constant of the dynamics additional to the integrator or $P_i(s)e^{-5s}$. This is shown as well in the table as T_n. Notice that L_n is the dead time of the two-parameter model. Comparing T_r and T_n one can see that usually they are quite close. Often T_r is a little bit smaller than T_n. Exceptions to this are processes 5 and 7. Process 5 is non-minimum phase while process 7 has a slow zero giving a large area A compared to T_n. In the case of process 7, A is also very large compared to L_n. This results in a large T_r according to Eq. (44).

Closed-loop set point and load disturbance responses are shown in Figs. 14 to 16 for a selection of processes when T_r is found from Eq. (44). The amplitude of the load disturbance was -0.03. Also shown (dashed line) are simulations when T_r is set equal to T_n.

Practical issues

Some remarks on the practical use of the methodology presented are in order. Most of the remarks made in Section 4 apply here as well. Notice that when A is close to K_nL_n, T_r becomes very large calculated with Eq. (44).

Extensions

In [Matausek and Micic, 1999] an extension to the MSP was proposed. To improve load disturbance rejection the transfer function $F(s)$ should have the form

$$F(s) = K_0\frac{sT_d + 1}{sT_d/10 + 1} \tag{52}$$

where $T_d = 0.4L_n$ and $K_0 = 0.6/(L_nK_n)$. The form of the transfer function is similar to a PD controller. It's purpose is also to predict the load

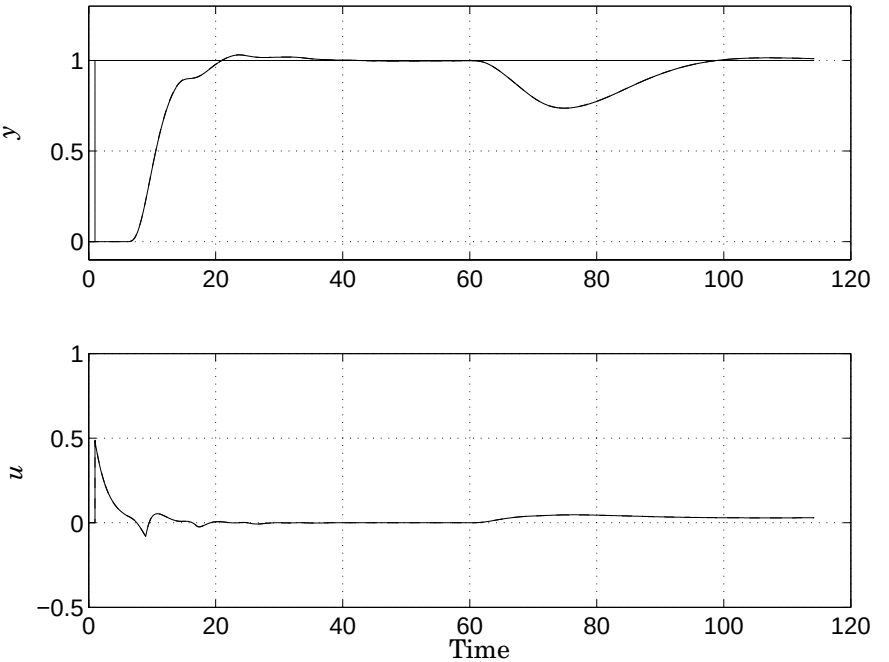

Figure 14. Simulation example. Process 3.

disturbances l better which in turn leads to better disturbance rejection. Simulation experience indicates it is possible to use this extension with the tuning procedure presented. Sometimes the increase in conservativeness associated with the procedure is welcomed. In Fig. 17 the Nyquist diagram is shown for process 5 when T_r is found by the procedure presented and as recommended in [Matausek and Micic, 1996]. It can be seen that the latter results in an unstable closed-loop system.

Summary of tuning procedure

As in the stable case, the tuning procedure starts with an identification experiment which results in parameters K_n and L_n. K_0 can then be found from Eq. (37). Then, area A in Eq. (47) is found by performing an open-loop inpulse response on the process. If a delay margin ΔL is desired, it is added to A and then T_r is found from Eq. (44). Then K_r can be found from Eq. (38).

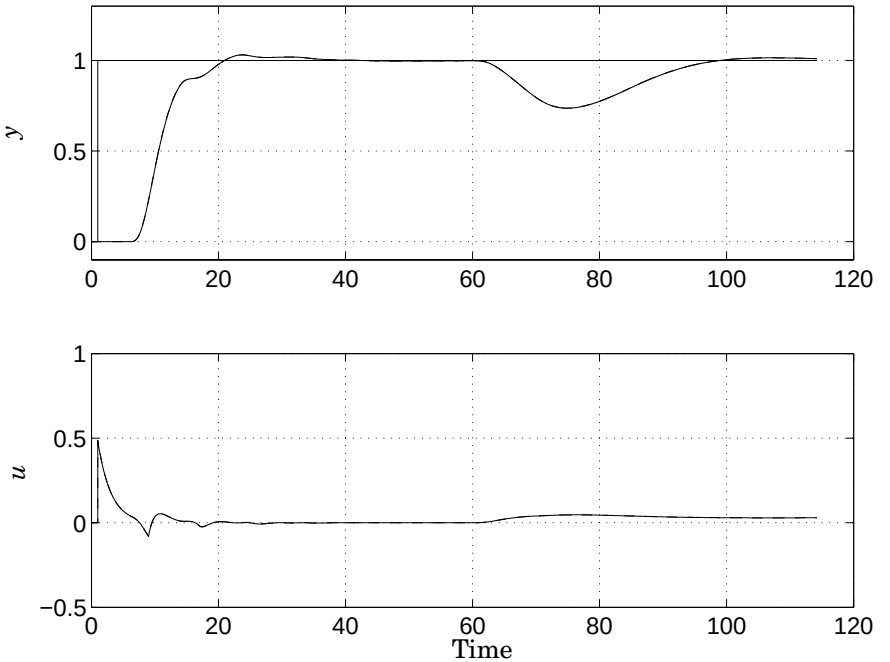

Figure 15. Simulation example. Process 5.

Application to a lab process

The method presented was used on the tank laboratory process presented in Section 4 after some modification of the process. To make an integrating process a tank without an outflow hole was added under the first one. The controlled variable was therefore the height in the second tank. To avoid nonlinearities at low flow levels, a second pump was installed which pumped with a constant flow rate out of the lower tank. In this way the first pump was set to work around a constant flow rate corresponding to 3 V.

The identification experiment was performed manually because it was difficult to tune a suitable PI controller. An impulse of height 1 Volt and duration 60 seconds was added to the equilibrium value of 3 Volts. The response of the tank and the identified two-parameter model are shown in Fig. 18. The identified two-parameter model was

$$P_n(s) = \frac{0.07}{s} e^{-132.5s}$$

The area A between the responses was 1.6. This resulted in a closed-loop

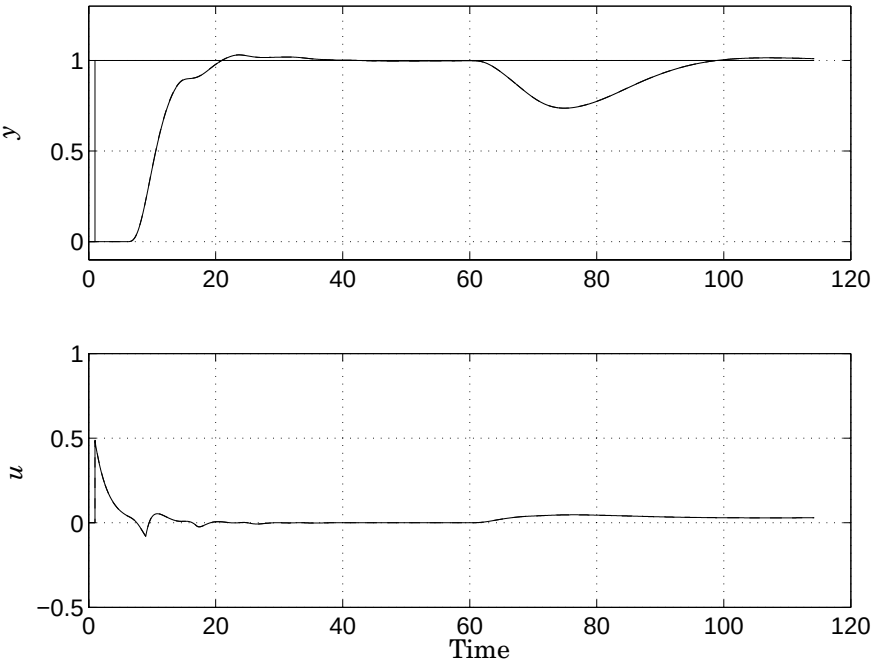

Figure 16. Simulation example. Process 6.

time constant $T_r = 54.8$ seconds. In Fig. 19 a closed-loop step response experiment using the MSP is shown. The controller was started at time 1000 while a step was applied at time 2000 taking the level in the tank from 4 cm to 8 cm. The controller performs as expected. The step response looks similar to what was seen in simulation.

6. Conclusions

In this paper tuning procedures for dead-time compensators have been presented. Dead-time compensators for both stable and integrating processes are considered. The closed-loop time constant is found by comparing the model output to the process output when a simple open-loop experiment is performed.

In the case of integrating processes the procedure eliminates the need to manually tuning one parameter.

The DTCs are simple and can be manually fine-tuned or re-tuned. It is also shown how the DTCs can be given a guaranteed delay margin.

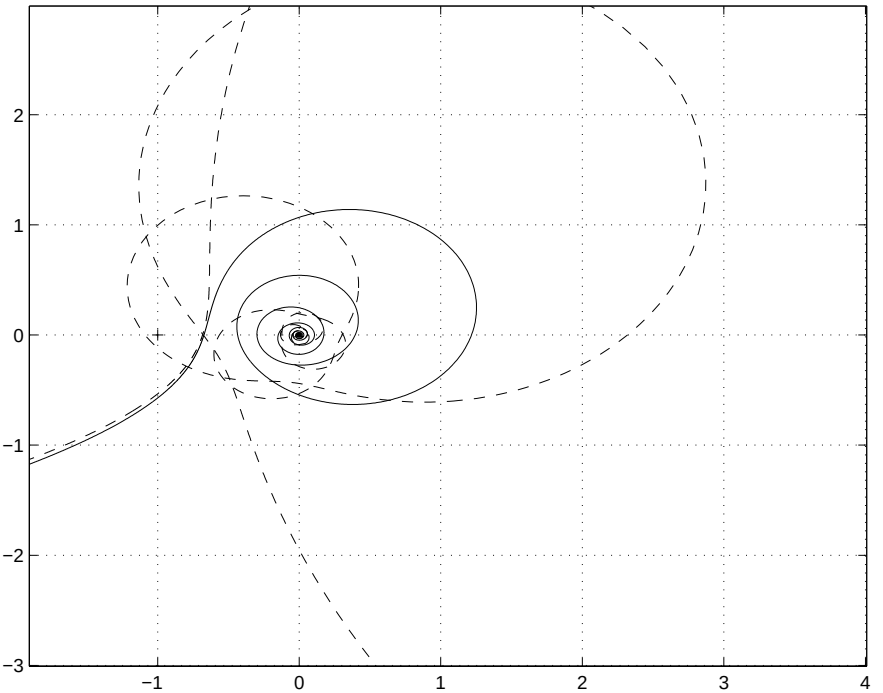

Figure 17. Nyquist curve for process 5, $L = 10$. Extended MSP with T_r found from Eq. (44), (−). Extended MSP with T_r set equal to apparent time constant, $T_r = 1.7$, (−−).

Finally, the methodology has been applied to a laboratory process with success.

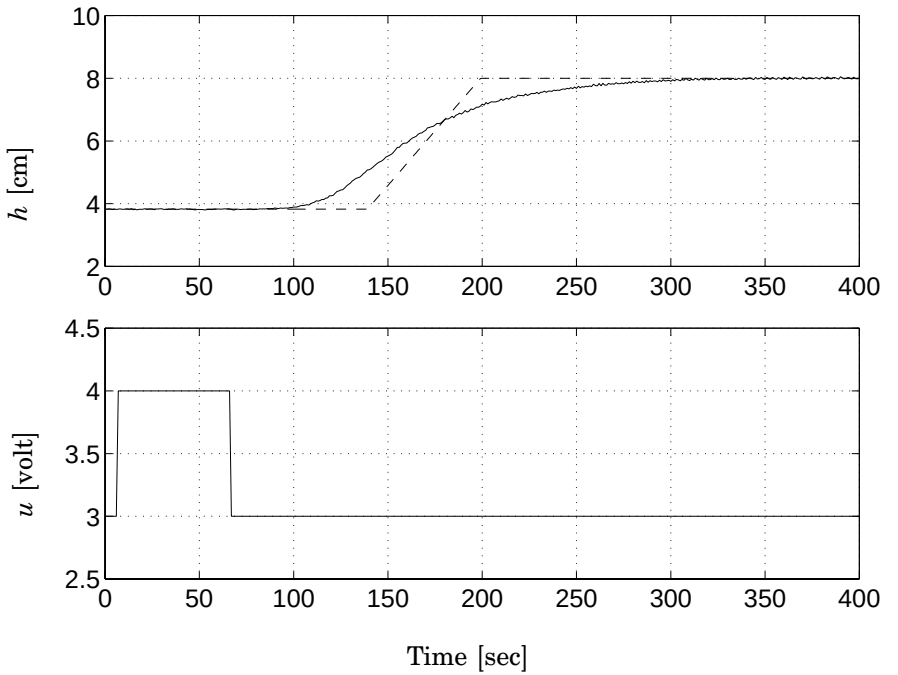

Figure 18. Identification experiment for integrating case. The upper graph shows the impulse response of real tank (−) and that of the two-parameter model (−−).

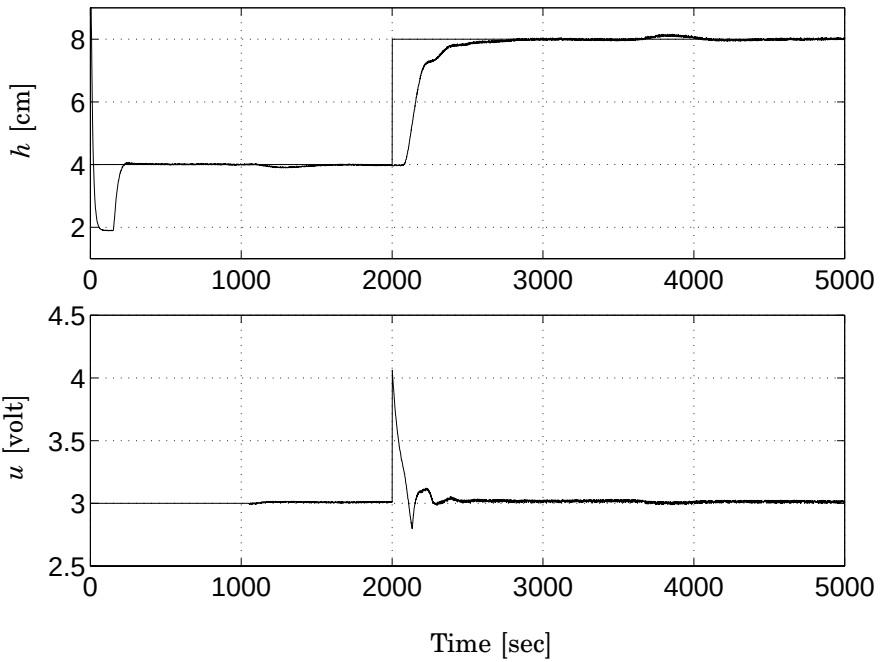

Figure 19. Step response of the real tank.

Paper 2

Performance comparison between PID and dead-time compensating controllers

Ari Ingimundarson and Tore Hägglund

Abstract

This paper is intended to answer the question: "When can a simple dead-time compensator be expected to perform better than a PID?". The performance criterion used is the integrated absolute error (IAE). It is compared for PI and PID controllers and a simple dead-time compensator (DTC) when a step load disturbance is applied at the plant input. Both stable and integrating processes are considered. For a fair comparison the controllers should provide equal robustness in some sense. Here, as a measure of robustness, the $\mathcal{H}_\infty$ norm of the sum of the absolute values of the sensitivity function and the complementary sensitivity function is used. Performance of the DTCs is given also as a function of dead-time margin (D_M).

Keywords Performance Comparison, PID Control, Dead-time Compensators

Reprinted from: Ingimundarson, A. and T. Hägglund (2002): "Performance comparison between PID and dead-time compensating controllers", *Journal of Process Control 12*, p.887-895, with permission from Elsevier Science.

1. Introduction

In [Smith, 1957] a control structure was presented which became one of the main solutions to deal with processes with long dead time. Recently, a renewed interest in dead-time compensators has been noted in the control literature. Extensions to integrating and unstable processes have been presented, (see for example [Matausek and Micic, 1996] and [Majhi and Atherton, 2000]).

Despite this, little has been written about when DTCs should be used. Commonly in textbooks on process control a few pages are devoted to explain the control structures of common DTCs but recommendations and guidelines about when to use DTCs are very rare.

The most common control structure in the process industry is the PI. The derivative gain is often turned off. This applies specially when long dead times are present. Therefore, given that the current control structure is a PI two options should be compared. One is to add the derivative part to the PI making it a PID. The other is to implement a DTC.

One reference where a similar comparison was done is [Rivera *et al.*, 1986]. There, within the IMC framework, a PID controller was designed for a first-order plus dead time (FOPDT) transfer function and compared to best achievable performance of a DTC, that is when it has infinite gain. The performance criterion used was the Integrated Squared Error (ISE). DTCs with very high gains can have arbitrary small robustness toward dead-time errors, see [Palmor, 1980], even though their robustness measured with traditional measures like amplitude margin, phase margin or maximum sensitivity, can be good. Therefore, in the current paper, performance of DTCs is given as a function of dead-time error sensitivity, measured with the dead-time margin (D_M) or the smallest error in the dead time which causes instability. An other reference where the subject is treated is [Meyer *et al.*, 1976]. There the robustness of the two structures is not treated specially and the range of process dynamics for which the comparison is made is small.

The layout of this paper is the following. The comparison criteria is treated in Section 2. The DTCs and their tuning is presented in Section 3. In Section 4 the results are presented. This is followed by a discussion in Section 5. Finally conclusions are drawn in Section 6.

2. Comparison criteria

The purpose of this paper is to give insight into the problem of choosing between a DTC and a PI(D) within a typical process control environment.

While providing results general enough to be useful in that environment, certain assumptions are necessary to limit the size of the paper.

One question that arises is what information should be assumed to be available to select between control structures. Detailed knowledge can not be assumed to be available since this is usually not the case within the process industry. In connection with the usage of DTCs, plants are often referred to as being dead-time dominated. What is meant by this is that the ratio between dead-time and other dynamics is large and accordingly a decision to implement a DTC is taken. In this article it is assumed that process information is available in terms of an first-order plus dead time (FOPDT) model, denoted with

$$P(s) = \frac{K_p}{Ts+1}e^{-Ls} \tag{1}$$

This is the simplest way to represent the division of process dynamics into pure dead-time, and other dynamics. In the case of integrating processes it is assumed that the FOPDT is in serial with an integrator

$$P(s) = \frac{K_p}{s(Ts+1)}e^{-Ls} \tag{2}$$

In the process industry, processes are commonly modeled with these transfer functions so most control engineers are familiar with their parameters. The comparison is made for these processes only. By making sure controllers fulfill a robustness constraint, the results should be valid for plants sufficiently close to these models.

Another question that arises is the complexity of the controller structures. The most commonly used structure in the process industry is the PI which has 2 parameters. A more complex structure with more parameters might show better performance but still it might never be implemented since this requires more expertize and advanced maintenance than the PI. In this paper, an effort has been made to keep the DTCs as simple as possible. The DTC structures considered all contain a model of the process. Most parameters are related to this model. When the plant is equal to the model, the set-point response is given by a FOPDT transfer function with unit gain.

$$\frac{1}{T_r s+1}e^{-Ls} \tag{3}$$

T_r is selected with dead-time sensitivity in mind. This applies to the stable and integrating plant case. The models and the tuning of the DTCs will be introduced in Section 3.

Within the process industry regulatory performance is usually more important that set-point response. Therefore, the performance criterion

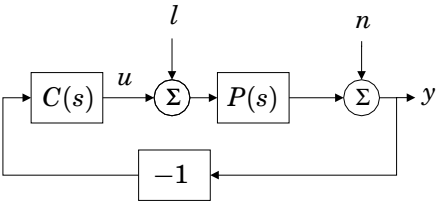

Figure 1. Block diagram of disturbance signals.

used was the IAE when a step load disturbance is applied at the plant input. This comparison was performed for the FOPDT model for a griding of the interval

$$T \in [0.01, 10] \tag{4}$$

while L and K_p were kept equal to 1. This range includes "almost" pure dead-time processes to "almost" dead-time free processes.

It should be pointed out that while the comparison made in this paper is for load disturbances only, the main strength of DTCs has on the other hand been its set point response. Further more it should be pointed out that it has been a subject of many papers to improve the load disturbance response of DTCs. As was said before, the DTCs presented here are simple and there is much room for improvement.

The block diagram of the loop is shown in Fig. 1. $C(s)$ is the controller while $P(s)$ is the process to be controlled. l is the load disturbance affecting the system while n is the measurement noise. For a FOPDT model on interval (4), the PI(D) with the lowest IAE and with equal or better robustness was compared to the IAE of the DTC with the tuning obtained from the FOPDT model. The robustness condition will be introduced in the next section.

Since the PI(D) controller parameters are only subject to a robustness constraint, the comparison presented is not dependent of a specific tuning rule of the PI(D).

Robustness constraints

Caution has to be shown when comparing performance of control structures because of the ever present tradeoff between robustness and performance. The comparison should be made under the assumption that robustness of the control structures is similar. A small deviation in the process should not result in a great difference in performance between the structures.

The robustness measure used here is the $\mathcal{H}_\infty$ norm of the sum of the absolute values of the sensitivity function and the complementary

sensitivity function

$$\gamma = \sup_{\omega} \left(|S(i\omega)| + |T(i\omega)| \right) \tag{5}$$

where

$$S = \frac{1}{1+CP} \qquad T = \frac{CP}{1+CP} \tag{6}$$

This robustness measure or similar measures have appeared in various places in the control literature. In [Panagopoulos and Åström, 2000] it was shown how the Nyquist curve of a closed-loop system with a given γ is guaranteed to stay out of the region given by the contours of

$$\frac{1+|L|}{|1+L|} = \gamma \tag{7}$$

, $L(s)$ is the loop gain, $L(s) = C(s)P(s)$. Furthermore it was shown how this parameter, with a suitable selection of weights for l and n is equivalent to the generalized robustness margin, see [Zhou, 1998].

In [Morari and Zafiriou, 1989] and [Doyle *et al.*, 1992] the following condition for robust performance appears:

$$W_1|S(i\omega)| + W_2|T(i\omega)| < 1 \ \forall \ \omega \tag{8}$$

For a typical process control problem the weights W_1 and W_2 should have a special shape. Weight W_1 should be large at low frequencies to assure good load-disturbance rejection. Weight W_2 usually increases with frequency to guarantee robustness toward model perturbations at high frequencies. The larger W_1 can be, the better disturbance rejection. Larger W_2 means better robustness toward multiplicative uncertainties is assured.

The control structures presented in this article have certain inherited qualities. All of them have infinite gain at low frequencies resulting in asymptotic rejection of a step load disturbance. In the nominal case it can be shown with simple analysis that $|S(i\omega)| \to 0$ as ω when $\omega \to 0$. At high frequencies the controllers have constant gain which means $|T(i\omega)| \to 0$ as $1/\omega$ when $\omega \to \infty$. Given two controllers with these qualities and equal γ there is a pair of weights, W_1 and W_2, such that condition (8) is fulfilled for both controllers. These weights can be constructed by choosing them to be $1/\gamma$ for the frequency for which the supremum is achieved in Eq. (5). For other frequencies, W_1 would be put equal to the smaller value of $|1/S(i\omega)|$ for the two controllers. W_2 would be chosen similarly to be the smaller value of $|1/T(i\omega)|$. The implication of this is that there is a set of plants, around the nominal one, for which both controllers satisfy the robust performance condition given by Eq. (8).

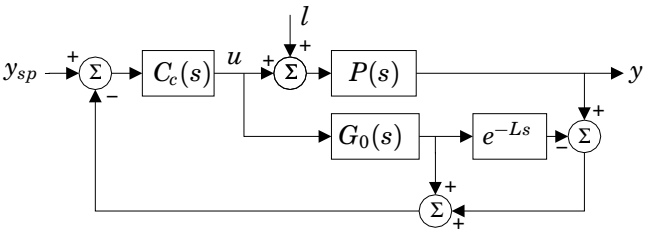

Figure 2. Block diagram of the simple DTC.

Dead-time sensitivity

A useful concept when dealing with DTCs is the dead-time margin (D_M) or the smallest change in the dead time which will cause instability. The robustness measure presented in the previous chapter does not capture dead-time sensitivity of DTCs. The reason is that it is caused by large loop gain in the right-half plane of the Nyquist diagram, see [Åström and Hägglund, 2001; Palmor and Blau, 1994], while the robustness measure is related to regions on the left side of the Nyquist diagram. Dead-time sensitivity can be taken into account by proper tuning of the DTCs. The general principle in this article was to select a tuning so that the amplitude of the loop gain was strictly smaller than 1 for frequencies larger than the crossover frequency, ω_c (smallest frequency where $|C(i\omega)P(i\omega)| = 1$).

3. Controllers

Stable case

The block diagram for the DTC can be seen in Fig. 2. $P(s)$ is the real process to be controlled. $P_0(s)$ is the model of the process with dead time, $G_0(s)$ is the model without. This way the controller for the DTC as indicated by Fig. 1 can be written as

$$C(s) = \frac{C_c(s)}{1 + C_c(s)G_0(s)(1 - e^{-Ls})} \tag{9}$$

The model $P_0(s)$ was the FOPDT model given by Eq. (1) and $C_c(s)$ was a PI controller set to

$$C_c(s) = \frac{1}{T_r s} G_0^{-1}(s) = \frac{Ts + 1}{T_r K_p s} \tag{10}$$

where T_r is a design parameter. This parametrization is related to H_2 optimal control. The actual criteria it minimizes is the integrated squared error (ISE) but because of its simplicity it is used even here. For references about this parametrization, see [Laughlin *et al.*, 1987].

In the nominal case, when there is no model error, $P_0(s) = P(s)$, then T_r is the time constant from set point y_{sp} to output y.

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{1}{T_r s + 1} e^{-Ls} \tag{11}$$

Furthermore, the loop gain $L(s) = C(s)P(s)$ is then given by

$$L(s) = \frac{e^{-Ls}}{T_r s + 1 - e^{-Ls}} \tag{12}$$

The robustness properties of this structure were analyzed in [Palmor and Blau, 1994]. There it was shown that $L(i\omega)$ never has a real part smaller than $-1/2$. This means the phase margin and amplitude margin are minimum 60^o and 2 respectively. It was also shown how the dead-time margin depends on T_r/L. Some values are given in Table 1. For $T_r/L \geq 0.34$ the dead-time margin is larger than 100%. This was the value used for the comparison. Again, for this value the amplitude of the loop gain is always less than 1 for frequencies larger than the crossover frequency.

In the nominal case the transfer function from l to y is given by

$$y(s) = P_0(s) \left(1 - \frac{e^{-Ls}}{T_r s + 1} \right) l(s) \tag{13}$$

This is easily seen to have an over damped response if l is a step. This means that the IAE is equal to the integrated error (IE) and an analytical expression can be obtained by a Taylor series expansion of the exponent in Eq. (13) and applying the final-value theorem. This gives

$$IAE = \lim_{t \to \infty} \int_0^\infty e(t)dt = \lim_{s \to 0} s \frac{S(s)}{s^2} = K_p(T_r + L) \tag{14}$$

Table 1. Dead time margin in percentage with the corresponding minimum T_r/L

D_M %	5	15	25	35	>100
T_r/L	0.03	0.10	0.16	0.24	0.34

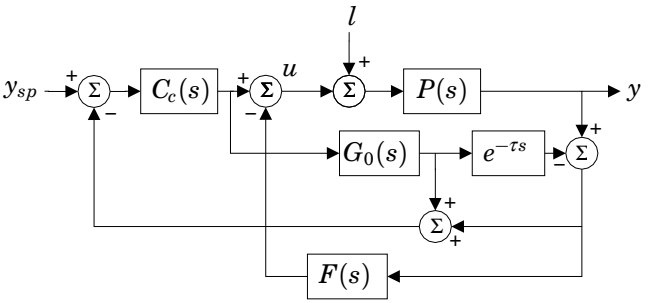

Figure 3. The improved dead-time compensating controller for integrating processes.

For example a DTC with DM equal to 35% will have IAE from a step load disturbance equal to $1.24K_pL$. With a very high gain DTC, obtained by $T_r \approx 0$, one could get close to $IAE = K_pL$. That is the lower limit with the presented tuning.

Notice also that even if the IAE is independent of T, the response is not. With a large T, settling time will also be very large.

The Modified Smith Predictor, integrating case

DTCs for integrating plants have been presented by a number of authors, see for example [Matausek and Micic, 1999; Normey-Rico and Camacho, 1999]. The one chosen for the comparison in this article is the one presented in [Matausek and Micic, 1999], see Fig. 3.

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{1}{T_r s + 1} e^{-\tau s} \tag{15}$$

The reason for this is that this structure is very simple and replacement from a PI should be straight forward. This structure is an improvement of a structure presented in [Matausek and Micic, 1996] which will be considered for comparison as well. Tuning of this structure has been presented in [Ingimundarson and Hägglund, 2001]. These controllers in [Matausek and Micic, 1996] will be referred to as the MSP (modified Smith predictor) and the one in [Matausek and Micic, 1999] as the IMSP (improved MSP).

As with most DTCs, these controllers contain a model of the plant to be controlled. This is the two-parameter model

$$P_0(s) = G_0(s)e^{-\tau s} = \frac{K_p}{s} e^{-\tau s} \tag{16}$$

As was explained before, the comparison is made for the collection of FOPDT models in serial with an integrator as given by Eqs. (2) and (4). A two-parameter model approximation of the model given by Eq. (2) was obtained by putting $\tau = T + L$

The equation for the controller according to Fig. 1 is

$$C = \frac{C_c + F + FC_cG_0}{1 + C_c(G_0 - P_0)} \tag{17}$$

The transfer function $F(s)$ is given by

$$F(s) = \begin{cases} \dfrac{1}{2\tau K_p} & : \text{MSP} \\ \\ \dfrac{0.72}{\tau K_p}\dfrac{0.4\tau s + 1}{0.04\tau s + 1} & : \text{IMSP} \end{cases} \tag{18}$$

Finally C_c is a constant given by

$$K_r = \frac{1}{K_n T_r} \tag{19}$$

where T_r is the time constant of the transfer function from set point in the nominal case when $P = P_0$,

Then the loop gain can be expressed as a function of only dead time τ and the time constant from reference value to the output, T_r

$$L(s) = \frac{e^{-\tau s}(s + F(s)T_r s + F(s)K_p)}{s(T_r s + 1 - e^{-\tau s})} \tag{20}$$

It can be seen that the loop gain depends only on the ratio T_r/τ and therefore the D_M as well. In accordance with the tuning of the DTC for stable processes, this ratio was chosen so that the amplitude of $L(i\omega)$ for frequencies larger than the cross-over frequency would be strictly smaller than 1. This gave $T_r/\tau = 0.534$.

The tuning recommended in [Matausek and Micic, 1996] was to set T_r equal to T in the model (2). This parameter setting is included in the simulation study that follows. Notice though that when T is small this tuning can give arbitrary small D_M.

The PID parametrization

The PID was parametrized on parallel form

$$C(s) = K\left(1 + \frac{1}{T_i s} + \frac{T_d s}{T_d/N s + 1}\right) \tag{21}$$

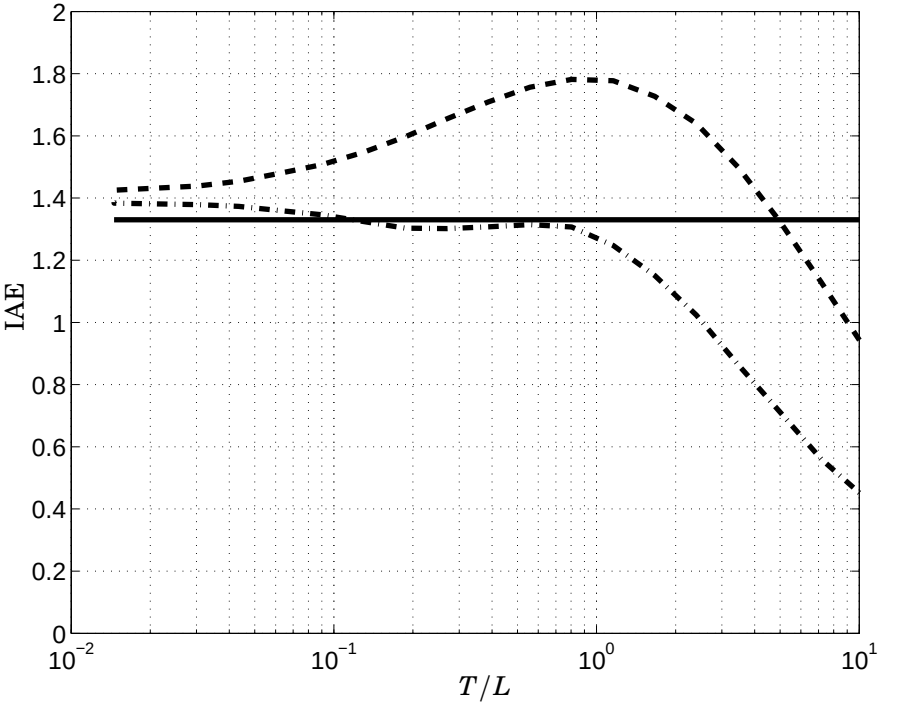

Figure 4. IAE results for the different control structures in the stable case. DTC(−),PI (−−), PID (· − ·)

with $N = 10$. N is sometimes considered a tuning parameter in the study of PID controllers. To limit the complexity of the PID controller it was fixed to a typical value as recommended in textbooks on process control. For PI, T_d was set to 0.

4. Results

Stable case

The main result of the paper for the stable case is shown in Fig 4. IAE is displayed for PI, PID and the DTC for different values of T/L. The performance of the DTC is only dependent on the T_r/L ratio determined by the dead-time margin and it is therefore constant over the entire range of the T/L ratio.

To simplify discussions the range of T/L is split into three regions.

Top region. $T/L \in [5, 10]$**.** Here the performance of the PI is better than the DTC. This is not unexpected since when $L = 0$ the poles of the closed loop system can be placed arbitrarily with a PI controller. The PID performs best of the controller structures.

Middle region. $T/L \in [0.1, 5]$**.** In this region the performance of the DTC is better than the PI. The largest difference is when $T/L \approx 1$. There the reduction in IAE following a switch to a DTC is at least 25%.

For $T/L < 1$ the PID performance is similar to the DTC. To obtain a DTC with better performance than a PID, it would have to be tuned more aggressively with lower D_M.

Bottom region. $T/L = [0.01, 0.1]$**.** The PI has higher IAE over the region but the performance is quite similar. Notice that the performance curve for the PI is leveling out and approaching a constant value as the FOPDT process approaches a pure dead-time process. Only a DTC with smaller D_M would have superior performance in this interval.

The performance of the PID approaches the performance of the PI when the ratio is decreasing. Investigation of the derivative gain showed it approaches zero for smaller T/L.

Integrating Case

The results in the integrating case are shown in Fig. 5. As the IAE varies considerably over the T/L range the IAE of all control structures was normalized with the IAE of the IMSP with $T_r/L = 0.534$.

The performance of the DTCs is better than the PI for all values of T/L. Largest difference is for $T/L \approx 3$. There the IAE of the PI is 3.3 times larger than the IAE for the DTC. For small T, when the process approaches a pure dead-time process with an integrator, the ratio of the IAE between PI and DTC approaches a constant value. The PI has around 2.5 times higher IAE than the DTC in this case.

If the IMSP and MSP are tuned as recommended with $T_r = T$, the performance depends strongly on T/L but is always superior to the PI.

The PID performs much better than the other structures for large T/L. This it not surprising since if $L = 0$ in Eq. (2) the poles of the closed-loop system can be placed arbitrarily with a PID. For small T/L the PID has a performance very similar to the IMSP.

Sensitivity analysis

The PI(D) controllers were compared to the DTCs under the assumption that they fulfilled the robustness constraint in Eq. (5) for equal or less

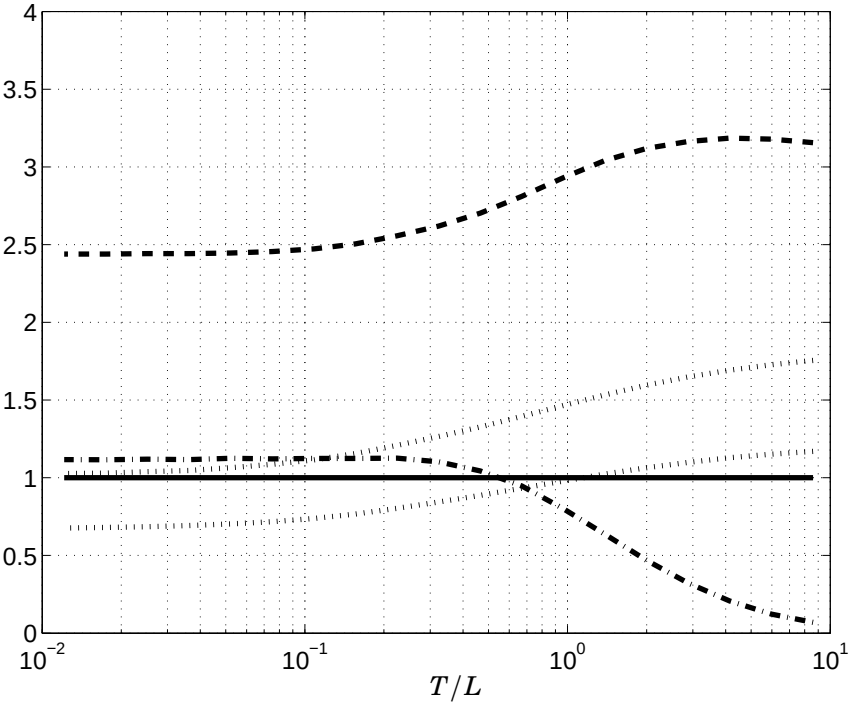

Figure 5. Results for the integrating case. What is shown is the IAE normalized with the IAE of IMSP with $T_r/L = 0.534$ $(-)$, PI $(--)$, PID $(-\cdot)$,IMSP with $T_r = T$ $(\cdots$ upper),MSP with $T_r = T$ $(\cdots$ lower).

γ. Yet it is interesting to see how IAE changes when the true process is not the same as the nominal model. As there are only three parameters, a sensitivity star was determined for each parameter, for the stable and integrating case. This was done for T/L ratio equal to 1. The sensitivity stars show how IAE changes, relative to the IAE for the correct model, as a function of parameter variation.

Stable case In Fig 6. the Nyquist diagram is shown. Also shown, on the left half plane is the γ contour (encircling -1). It can be seen that all controllers touch the contour meaning they all have equal γ. Notice that the PID seems to touch the contour in more than one place.

In Figs. 7 to 9 the sensitivity stars for the different parameters are shown. Notice that the scale for IAE change is different on each figure. Generally speaking one can say that none of the structures seemed significantly more sensitive to parametric changes than the others. For varying

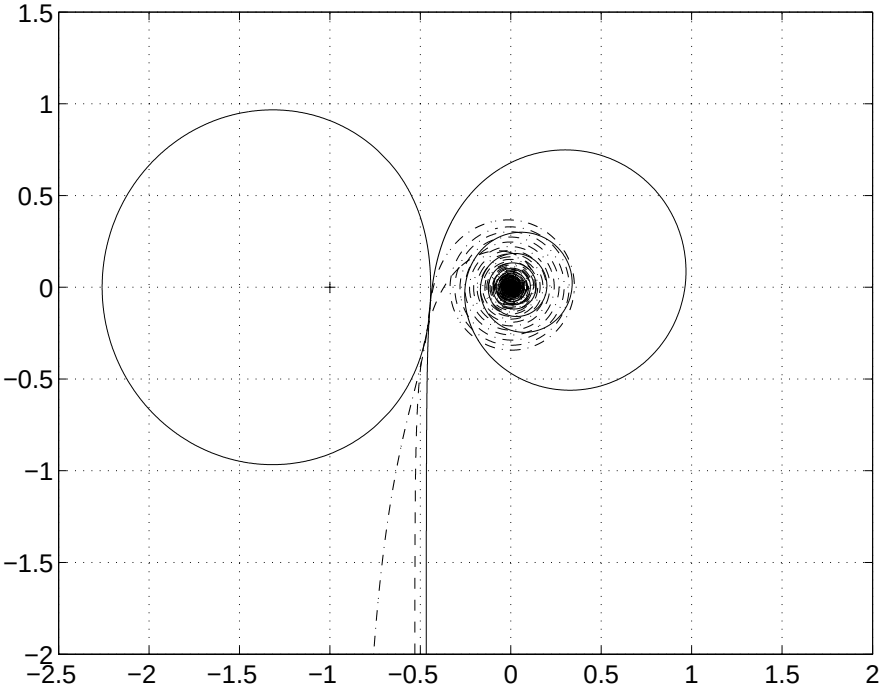

Figure 6. Nyquist diagram for the stable case, $T/L = 1$. DTC $(-)$, PI $(--)$, PID $(\cdot - \cdot)$

K_p the surprising thing is the increase in IAE for the PI on the right side. Comparing the load disturbance responses, one can see that the damping in the PI response decreases more as a function of ΔK_p, than for the PID and DTC. For varying T the behavior is rather good for all structures when the change is larger than -0.5. For smaller values, the performance of the DTC and the PID increases very quickly before the loop becomes unstable ($\Delta T < -0.65$). The PI on the other hand remains stable with a small increase in IAE on that portion of the interval. The PID is most sensitive toward change in dead time, L. The IAE increases for all structures for positive change in dead time.

Integrating case The dead-time compensator for the sensitivity study was the IMSP with $T_r/L = 0.543$. The Nyquist diagram for the integrating case is shown in Fig. 10. Again all controllers lie tight up to the robustness constraint. The sensitivity stars are shown in Figs. 11 to 13. Again one can say that no structure is significantly more sensitive than the others.

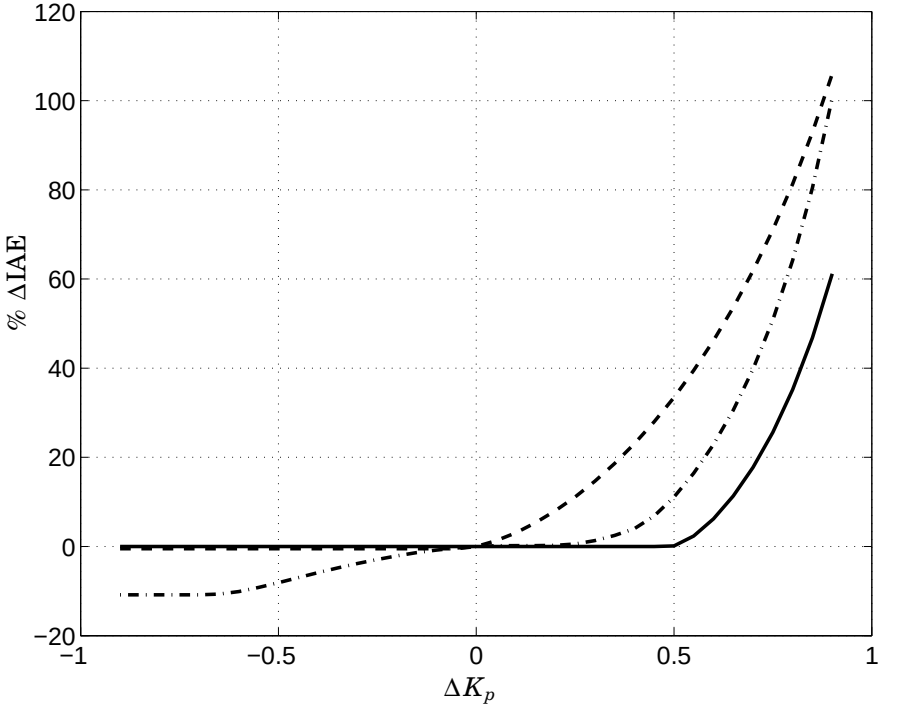

Figure 7. Sensitivity star for K, stable case. DTC (−), PI (−−), PID ($\cdot - \cdot$).

For a change in K_p it is obvious that the change in IAE for the DTC has a dependence that is very close to linear in ΔK_p and the dependence is opposite to the other structures.

5. Discussion

There are many aspects that have not been considered in this comparison. The PID and DTC control structures each have advantages that should be taken into account. The PID is a well known structure frequently found in distributed control systems as standard modules. One advantage of the DTC is that the dead time enters the control law directly, opening the possibility for gain scheduling if the dead time is measurable. Furthermore, DTCs can be designed with certain dead-time margin if variations in the dead time are known, see [Ingimundarson and Hägglund, 2001].

An other aspect ignored here is the control signal amplitude or energy.

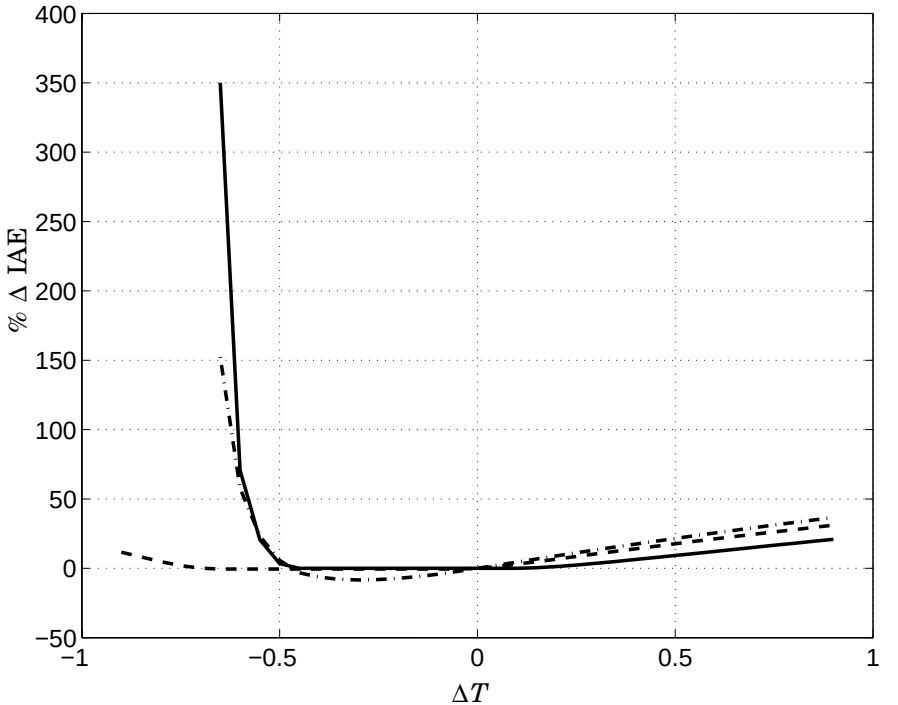

Figure 8. Sensitivity star for T, stable case. DTC $(-)$, PI $(--)$, PID $(\cdot - \cdot)$.

This could be checked by looking at the transfer function from measurement noise, n, to the control signal, $C(s)S(s)$. The performance of the PID for large T/L was often followed by large values of this transfer function.

It was said in the introduction that recommendations for when to use DTCs are scares in the literature. An exception is [Marlin, 1995]. There, a DTC is recommended for stable processes if the feedback fraction dead time, $L/(T+L)$ is larger than 0.7. This translates to a T/L ratio smaller than 0.4. As indicated in Fig. 4 the benefits of replacing a PI start for larger values of T/L.

PID considerations

The DTC is a model based control strategy. Implementing it necessitates an identification of a process model. Having said that a PID performs similar to a DTC it should be noted that to obtain this performance with a PID some kind of identification of the process would be necessary. Achieving similar performance with a manual tuning would be difficult. And this

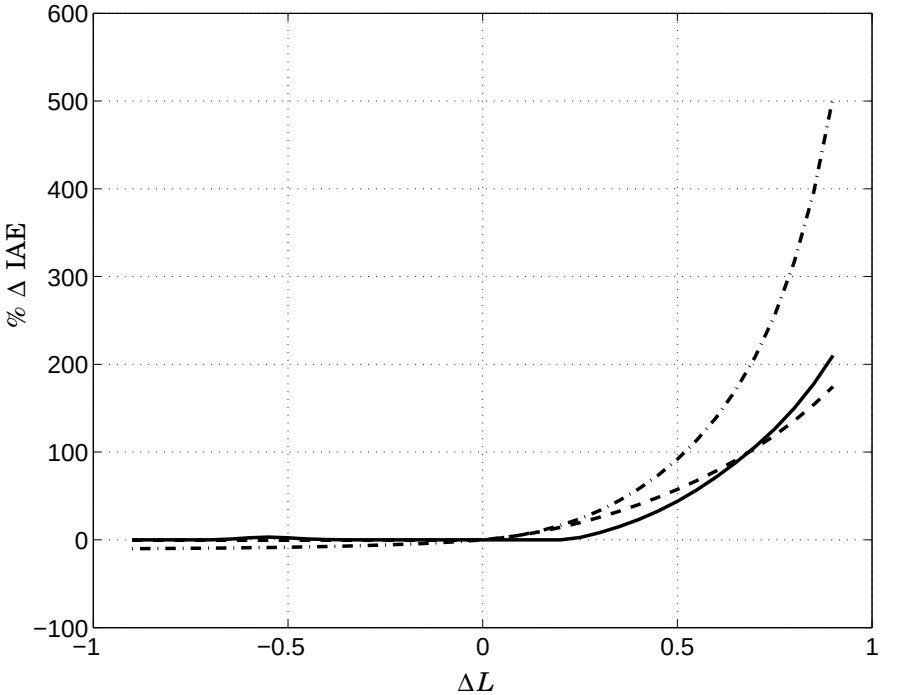

Figure 9. Sensitivity star for L, stable case. DTC (−), PI (−−), PID ($\cdot - \cdot$).

is not only due to the increase from two to three parameters. To support this claim, some contour lines in the PID parametric space is shown in Fig. 14 for the integrating case when $T/L = 1$. For several values of T_d the contour lines in K and T_i are shown for the same γ as the IMSP has when tuned with fixed T_r/τ ratio. It can be seen that the surface has sharp edges on it. Sharp edges of the parametric surfaces of PID's has been reported elsewhere, see [Åström and Hägglund, 2001]. To reach this optimal point by manual tuning might be difficult.

6. Conclusions

The paper has focused on comparing the performance of the PID control structure with simple DTCs. The purpose has been to gain insight about when each control structure should be used. IAE performance was determined as a function of the ratio between time constant and dead time, T/L, for a collection of FOPDT models.

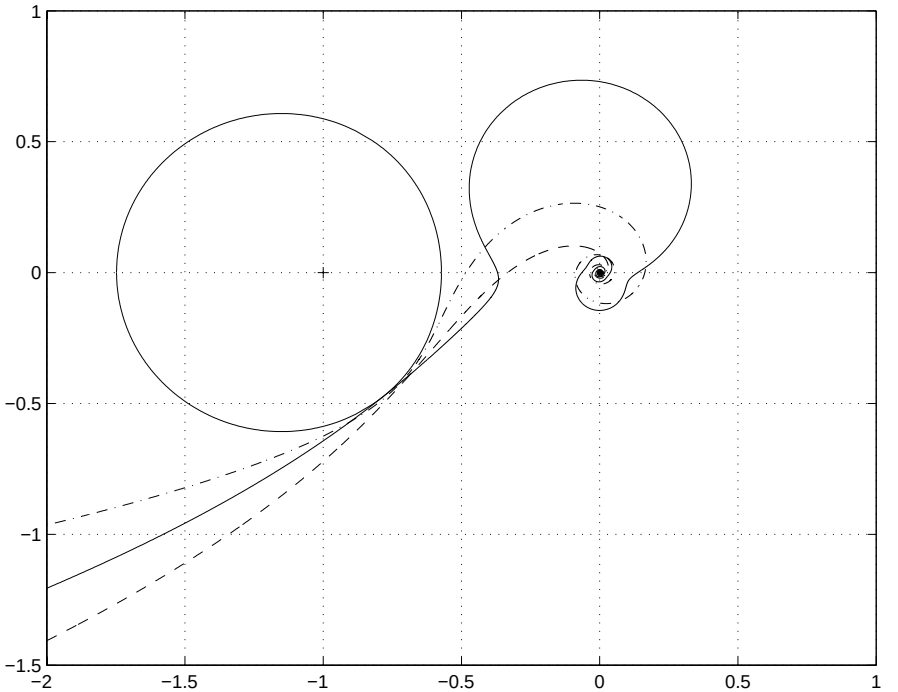

Figure 10. Nyquist diagram for the integrating case, $T = 1$. IMSP $(-)$, PI $(--)$, PID $(\cdot - \cdot)$

In the stable case, the main result is that for T/L in the interval $[0.1, 5]$ a substantial decrease in IAE can be obtained by switching from a PI to a DTC. The performance of a PID is better than a DTC for T/L on $[1, 10]$, otherwise it is similar to the DTC. A superior performance of the DTC on the interval $[0.1, 1]$ could be obtained by accepting a smaller D_M.

In the integrating case, the performance difference is much greater between PI and the DTC. A switch to a DTC might reduce IAE to about one third of the IAE for PI. The PID performs best over a large portion of the area but it has been shown it might be difficult to obtain this optimal performance with manual tuning.

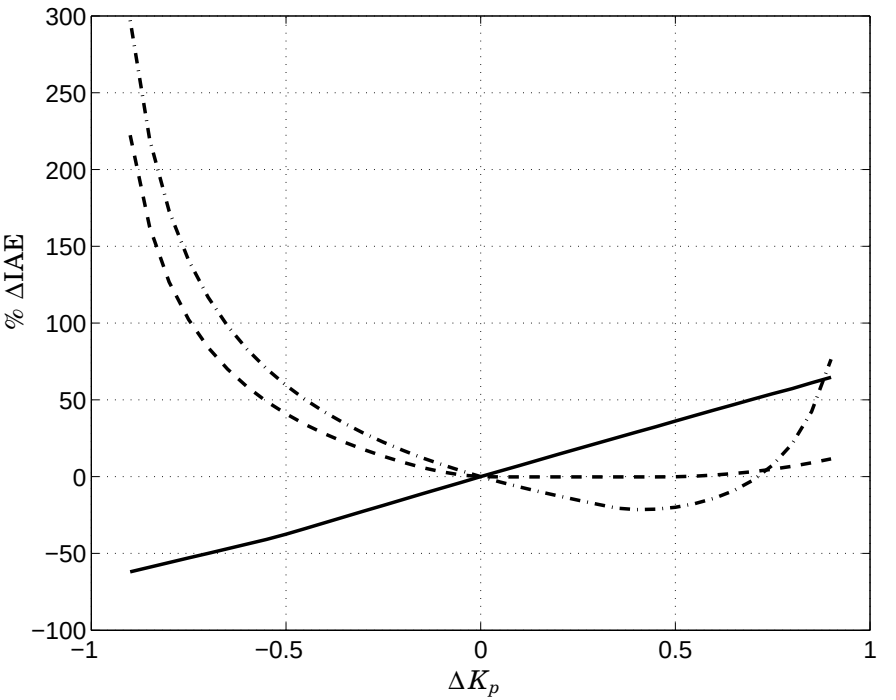

Figure 11. Sensitivity star for K, integrating case. IMSP (−), PI (−−), PID (·−·).

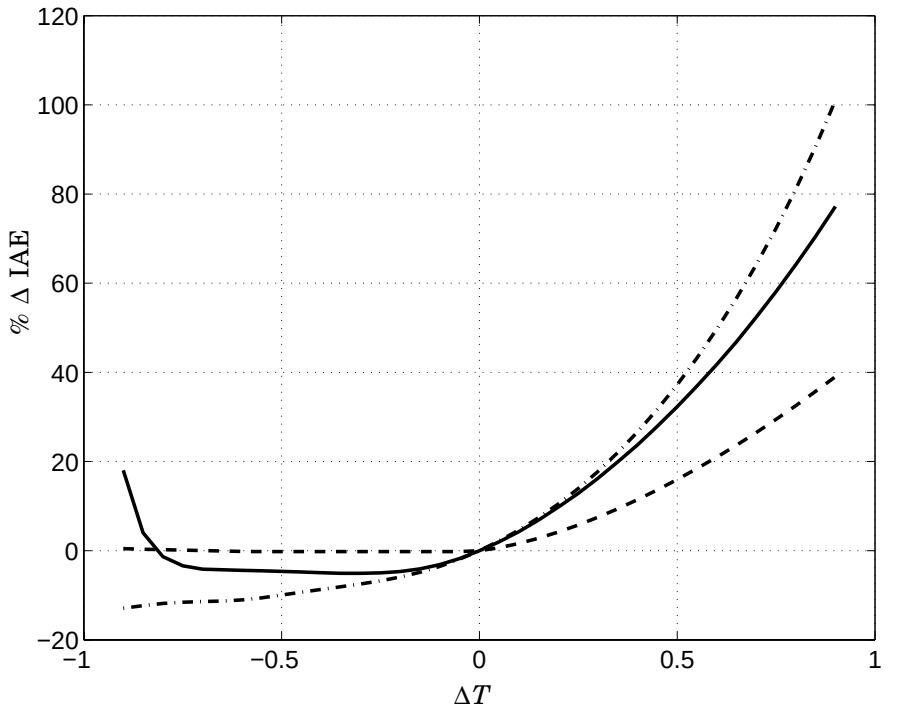

Figure 12. Sensitivity star for T, integrating case. DTC $(-)$, PI $(--)$, PID $(\cdot - \cdot)$.

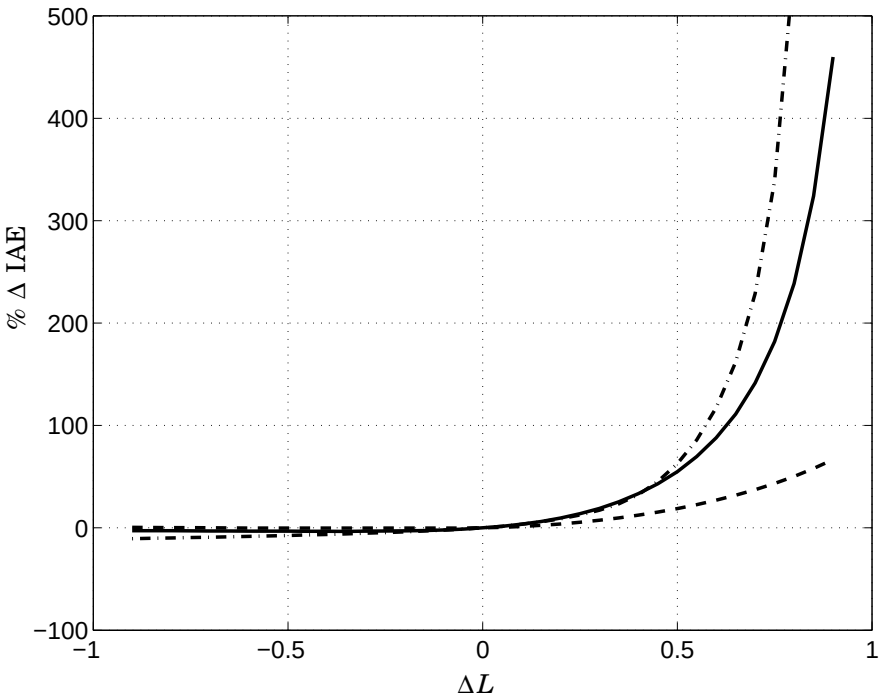

Figure 13. Sensitivity star for L, integrating case. IMSP (−), PI (−−), PID ($\cdot-\cdot$).

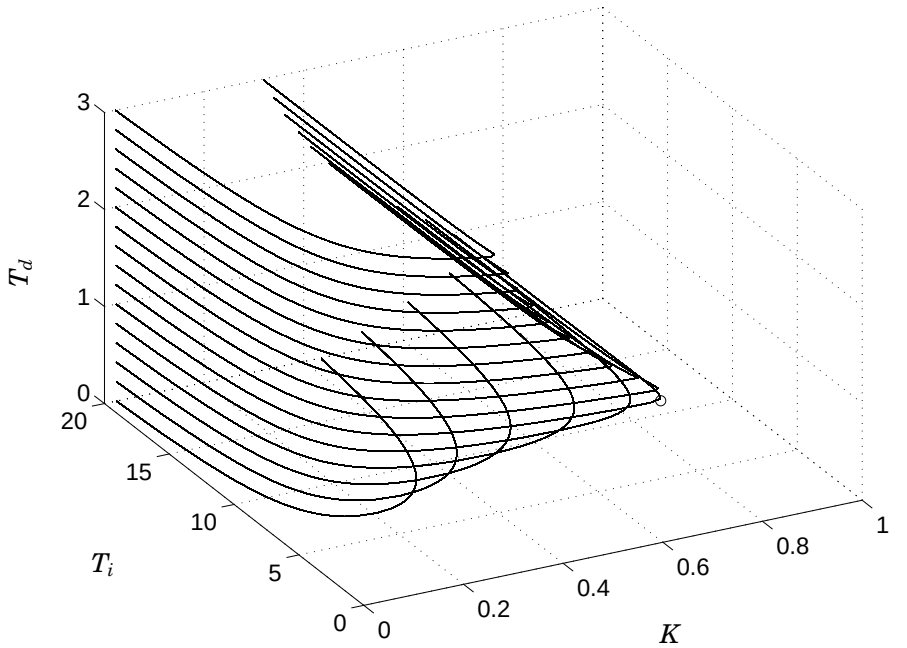

Figure 14. Contour lines for fixed γ. Optimal IAE solution is marked with ^{o}.

Part II

Performance Monitoring of λ-Tuned Controllers

1

Introduction to Performance Monitoring

Much work has been done in the field of closed-loop performance monitoring and diagnostics (CLPM&D) in the last decade. This branch of automatic control research has matured enough so that a few survey articles have appeared, see for example [Qin, 1998; Harris *et al.*, 1999].

The purpose of the CLPM&D methods is to assist plant staff to interpret plant data. The goal is to run the plant as effectively as possible. The monitoring algorithm should sound an alarm when controllers are not performing as expected. The diagnosis should give decision support to help with the overall maintenance of the closed loops in a plant.

The terms monitoring and assessment are used somewhat interchangeably in the literature, usually referring to very similar algorithms. In this thesis, what is meant with closed-loop performance monitoring is the action of watching for changes in a statistic reflecting performance over time. Assessment in this thesis refers to the action of assessing or evaluating a statistic that reflects performance at a certain point in time. The purpose of diagnosis is usually to determine the reason of poor performance. This might focus on one loop or a group of loops.

1.1 Control Performance in the Process Industry

Automated process control is an enabling technology to deal with the natural variations of process variables from their desired values in a plant. The performance of a control system relates to its ability to deal with this variability.

The basic indication that something is wrong with a process is that the variability becomes too large. The deviation might be measured as variance or standard deviation of the control error or estimated by visual

inspection of process data. Operation targets are usually given in terms of deviation from set point, either as the absolute maximum or average over time. There are many measures of variability, the one most often used in this thesis is variance.

Numerous investigations have shown that performance of feedback controllers is not good in the process industry, see [Bialkowski, 1993; Ender, 1993]. Typical numbers quoted are that 80% of loops have performance problems, 30% of loops actually increase variability in the short term over manual control.

This might be surprising in the light of the fact that about 95% of feedback controllers in the process industry are PI controllers. It is fair to say that with the amount of research effort put into control theory the last decades, the problem of designing and commissioning PI controllers is solved. There are many ways to design good PI controllers but the problem is that the performance of a controller that once worked well is prone to degrade over time. Therefore, what is lacking is the maintenance of the controller after commissioning. Some of the reasons quoted for the lack of maintenance are

- Staff have limited time. The maintenance of controllers gets little time, sometimes most of it is done on a "fire-fighting" basis. One reason for the lack of time is that the operators have to many loops to maintain.
- Lack of education and understanding of process control. Sometimes the poor performance becomes the norm and people accept it as normal.

The overall result is that companies are frequently not getting the best possible return on their investments into automated process control. In light of these facts the need for tools to assist staff to discover and fix process control problems is apparent. This need has generated a fair amount of research.

1.2 Why is Performance Poor?

There is a wide variety of reasons why loop performance might be poor. The equipment that the loop depends on can become faulty. Valves have excessive friction or stiction or they are not correctly dimensioned. Input saturation breaks the feedback path temporarily. Varying transportation times causes changes in dead time. The control structure or algorithm might be inappropriate. The controller might be badly tuned or disturbance characteristics for which the controller once was suitable have

changed. The sampling time might be poorly selected. Finally, if a loop has performance problems, this might show up in neighboring loops.

This wide spectrum of root causes for performance deterioration has given rise to a wide variety of methods to monitor and diagnose performance. There are a number of methods to monitor and diagnose oscillations, see [Horch, 2000; Hägglund, 1995]. Detection of sluggish control loops has been treated in [Hägglund, 1999]. There are methods specialized to diagnose valve problems such as stiction and backlash, see [Horch, 2000].

Efficient diagnosis between root causes is a difficult problem. To implement a number of methods and run them on-line in a distributed control system (DCS) might be considered troublesome. An approach to this problem is to have a performance monitoring algorithm running on-line which detects problem loops but does not distinguish between root causes. Then the root-cause diagnosis can be performed off-line. The performance monitoring algorithm presented in Chapter 3 is well suited as the algorithm running on-line.

The performance of feedback loops is subjected to limitations. These arise from plant dynamics such as dead time or right-half plane zeros. Saturation and dynamics of actuators pose limitations as do noise characteristics. This is important information when monitoring performance. Performance is hardly bad if it is close to what is optimally achievable considering the limitations. Even though it is close to optimal it might still not be acceptable. Many methods monitor performance by comparing the actual performance to the best achievable subjected to the known limitations. One of the main advantages of this approach is that only information about the limitations is required. One of the disadvantage is that there are many types of limitations and it is difficult to know them at all operating levels. Furthermore, performance far away from optimal performance might be perfectly acceptable in some situations.

1.3 Desired Properties of CLPM&D Methods

In [Vaught and Tippet, 2001] a list of desirable properties a CLPM&D method should have, was presented. Some of these properties are listed below and new ones have been added.

- *Non-invasive.* The method should not disturb normal operation of the loop. A few researchers have suggested methods where this condition is not fulfilled. For instance, in [Kendar and Cinar, 1997] a pseudo random binary sequence is added to the reference to find the sensitivity function and complementary sensitivity function. In

[Gustafsson and Graebes, 1998] a method which uses an external probing signal to distinguish between performance changes due to system changes and disturbances was presented. In [Huang and Shah, 1998] a dither excitation is used to identify a disturbance model which was then used to calculate achievable performance for a specific IMC design.

- *Low error rate.* False alarms occur when the algorithm signals bad performance eventhough the performance is actually good. Missed detections are those situations when the algorithm should give alarm but it doesn't. Too many false alarms or missed detections result in a reduced or no use of the method. The problem is to characterize exactly, in terms of the loop condition, when either will happen.

- *Automated operation.* The method should run with little or no manual intervention.

- *No history.* The method should result in an absolute measure of performance. It should not need to run for a time to be effective. Frequently the indices of today have to run for some time after commissioning to create a "normal" operation level. Deviations from this level is then what is monitored. This means some time is required to start using the indices. If the loop is changed in some way it might be necessary to repeat the startup procedure.

- *No process information.* One should need little knowledge of the process or the controller. The reason being that to obtain this information and keep it uppdated would be expensive.

- *Implementation in the DCS.* It is preferable that a method is implemented directly in the DCS since information about parameters and the operating condition (auto/manual) usually resides there, see a discussion in [Hägglund, 2002].

Tradeoffs

The desirable properties are not all attainable or realistic. An actual implementation of a CLPM&D method would mean a tradeoff between these conditions. New methods should also be viewed with this in mind as well. Probably not all prefer the same set of characteristics in a method.

Performance monitoring is a form of data compression where raw plant data is compressed to present information for plant staff for maintenance purposes. As always when data is compressed, information is lost. Usually with an increase in information content an increase in complexity follows. To interpret more detailed information requires more knowledge.

1.4 The Following Chapters

In Chapter 2 a review of previous work is presented. Most emphasis will be on the work on which the performance monitoring algorithm presented in Chapter 3, is built. The main characteristics of the monitoring algorithm presented in Chapter 3, separating it from other methods, is that specifications about performance from the tuning phase are used to set all parameters of the monitoring algorithm. This should reduce the commissioning time since no historical information is required to implement the index. The drawback is that the tuning information is needed for the commissioning. In Chapter 4 a method will be presented which estimates a gradient of the variance with regard to a controller parameter. The purpose of this statistic is to assist with parameter adjustment of controllers.

2

Previous Work

2.1 Introduction

A short review of the literature will be given. Most emphasis will be put on work related to λ-monitoring, the method presented in the next chapter, but other research directions are noted in the passing. Some notation will be introduced as well. To begin with a short note on industrial practice that has not received much academic attention will be presented.

Industrial Practice

Despite poor performance of controllers in the process industry maintenance of controllers has always taken place. Operation targets are specified in terms of measures of variability such as variance or standard deviation. The most basic trigger for maintenance is unacceptable variability in the loop. To monitor some measure of variability is obviously of fundamental importance for efficient operation. It is unlikely that maintenance efforts will be devoted to a loop that fulfills its operation targets even though the controller is performing badly according to some statistic. In this thesis variance is the main measure of variability.

A commonly used statistic for the overall performance of a controller is the uptime or service factor. This is the portion of time spent in auto, that is, the portion of the time the controller is on. [Kozub, 2002] argues that uptime will continue to be "the work horse" of performance monitoring since low uptime is a sure signal that a loop needs maintenance.

To rely on variance alone as a measure of performance gives false alarms regarding controller performance. The controller might be doing its job perfectly but still the variance might be unacceptable. With regard to uptime as a performance monitor it is noted in [Desbourough and Miller, 2001] that high uptime does not mean good performance.

2.2 Academic Work

A rough classification of approaches to performance monitoring and diagnosis presented in the literature is the following. Firstly, a number of single number performance statistics, referred to as indices or potentials, have been suggested. The other approach can be explained as graphical inspection of various functions, usually related to the controller error. Both approaches aim at compressing the complex plant data and presenting the result to the operators. The first one is easier to automate, making it a more valuable "first indicator" of bad performance.

The method presented in this thesis is strictly SISO. Research into performance monitoring of multi-variable controllers has lagged behind the SISO case but lately there has been much activity in this field. Extensions from the SISO case where minimum variance performance is compared to the actual performance has been advocated in [Harris *et al.*, 1996; Huang *et al.*, 1997]. An industrial perspective of the research is offered in [Kozub, 2002].

An other problem receiving more attention is to diagnose which loop is responsible when bad performance is exhibited by a group of loops. Typically these disturbances are oscillatory. Some references dealing with these problems are [Thornhill *et al.*, 2002; Thornhill *et al.*, 2001].

Most control problems in the process industry are solved using SISO controllers. To deal with interactions, loops are commonly configured with different dynamics. The impact of well working "fast" loops on the "slower" loops should be small as their variability can be assumed to be of frequencies above the bandwidth of the "slower" loops. As the method presented relates to how well a controller fulfills the assumptions from the tuning stage, it should be useful to validate whether the assumptions about the configurations of dynamics hold.

Multi-variable controllers are considerably less frequent in the process industry but their use is increasing considerably. Perhaps the advent of monitoring and diagnosis technology of these controllers will make their implementation feasible.

Performance Index Monitoring

Performance index monitoring refers to monitoring algorithms where a single number reflecting performance is calculated repeatedly over time and compared to an alarm limit. The alarm limit can be decided from statistical characteristics of the index or by some other criteria. One of the main advantage of this approach is that it is easy to automate.

An early reference based on this approach is [Devries and Wu, 1978]. The index presented by Harris (see [Harris, 1989]) received much attention and many improvements followed. The idea was to compare the vari-

ance of the controller error to the minimum variance when the limitation to performance due to dead time are taken into account. The minimum variance was found by identifying a time-series model of the error and solving a Diophantine equation. The ratio of the minimum variance and the actual variance has been called the Harris index in the literature.

Time-series modeling continued to be a cornerstone of much of the work following after [Harris, 1989]. In [Desborough and Harris, 1992] an alternative form of the Harris index was presented as well as approximate probability distributions which are useful for designing alarm limits. The index presented in this thesis is based on this alternative form of the Harris index. Hence, a special attention will be devoted to the indices calculated from time series.

Graphical Inspection

Some of the earliest mentioning of the idea of performance monitoring is in [Åström, 1967]. There it is pointed out that an assessment of how close the performance of a controller is to the optimal one could be estimated by calculating and inspecting the covariance function. Under minimum variance control, the control error should be a moving average of order d where d is the pole-zero excess of the plant in discrete time. Calculating the covariance function of such a time series for arguments larger than d should result in zero.

In [Kozub, 1997; Kozub, 2002], a function deemed easier to interpret than the covariance function was the impulse response from an identified time-series model of the control error. The value of the impulse response for graphical inspection is also pointed out in [Huang, 1999].

In [Stanfelj *et al.*, 1993] the cross correlation function between measurable disturbances and controller error was suggested as a tool to determine influence of disturbances. Finally, different spectra have been mentioned. An extension to the multi-variable case has been presented in [Seppala *et al.*, 2002].

In [Kozub, 2002] it is noted that graphical interpretation of the impulse response requires a critical threshold of training for plant staff. Graphical inspections take time and therefore they are not a solution to the problem of plant wide performance monitoring because of the limited time of plant staff. They are an important part of the secondary tool set to be used after a more basic algorithm triggers the auditing of the loop. Use of expert systems to package the information and provide decision support has been presented in [Harris *et al.*, 1996]. See [Paulonis and Cox, 2003] for an idea of large scale implementation of graphical methods.

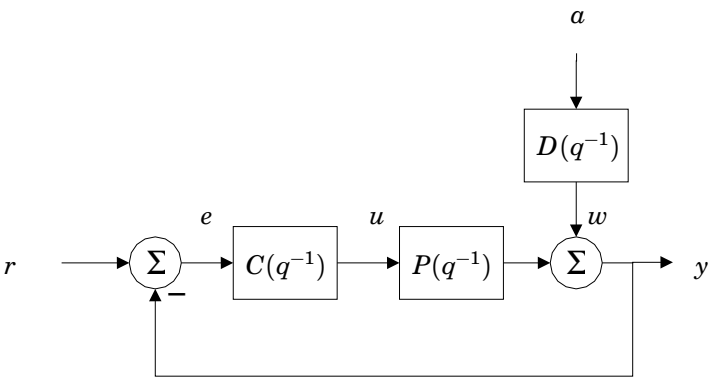

Figure 2.1 Basic block diagram of closed-loop system.

2.3 Indices Calculated from Time-Series Models

A set of assumptions about the process and the disturbances are usually made when the indices are derived. The central one is that the control error, $e(k) = r(k) - y(k)$, is a stochastic process resulting when a linear model is driven by white noise, see Fig. 2.1. The plant output is described by

$$y(k) = P(q^{-1})u(k) + w(k) \tag{2.1}$$

where q^{-1} is the backward shift operator. $P(q^{-1})$ is linear time invariant and can be written as

$$P(q^{-1}) = \frac{B_u(q^{-1})}{A_u(q^{-1})}q^{-d}$$

where the pole-zero excess is d. Notice that sometimes d is called dead time while other times it is $d-1$. Following [Harris, 1989] the dead time is defined as d, the pole-zero excess. The controller is assumed to be one degree of freedom

$$u(k) = C(q^{-1})(r(k) - y(k))$$

The disturbance $w(k)$ is thought to be generated by white noise $a(k)$ filtered through the disturbance model $D(q^{-1})$. The transfer function $D(q^{-1})$ therefore determines the character of the disturbance affecting the system. A common assumption is that it is of ARIMA type or

$$D(q^{-1}) = \frac{B_w(q^{-1})\nabla^l}{A_w(q^{-1})} \qquad \nabla = \frac{1}{1-q^{-1}} \tag{2.2}$$

The dependence of the controller error on disturbance $w(k)$ and the set point $r(k)$ is given by the following equations

$$\begin{aligned} e(k) &= r(k) - y(k) \\ &= \frac{1}{1 + C(q^{-1})P(q^{-1})}(r(k) - w(k)) \end{aligned} \tag{2.3}$$

Notice that the way set point and the disturbance affect $e(k)$ differs only by a sign.

Under regulatory control, $r(k)$ is constant. For simplicity it can be assumed that $e(k)$ has mean zero. Otherwise, it is the deviation from a mean value which is modeled. For a treatment when the mean is not zero see [Desborough and Harris, 1992]. If $r(k)$ is a stochastic process, by means of spectral factorization $e(k)$ can be described by an ARMA type linear model driven by a single white noise source. In this section regulatory control is assumed unless otherwise specified.

It is assumed that the data available for monitoring is

$$\{r(k), y(k)\} \quad \text{for} \quad k = 1 \ldots N$$

or only at the latest time instant, k, for recursive implementation, see Fig. 2.1. This allows us to calculate the error, $e(k) = r(k) - y(k)$.

Assuming that the plant and the disturbance model are not known, given $e(k)$, a time-series model can be identified giving valuable information about the characteristics of the closed-loop system and the disturbances affecting it. The identified time-series model is denoted

$$e(k) = \frac{B(q^{-1})}{A(q^{-1})}a(k) \tag{2.4}$$

Deterministic Disturbances

Notice that deterministic disturbances can be modeled in the same framework by assuming $a(k)$ is zero except at isolated points in time. Disturbance scenarios more common in process control can then be taken into account with assumptions on the disturbance model $D(q^{-1})$. When $D(q^{-1}) = P(q^{-1})\nabla$ then the disturbance will be referred to as a load disturbance. If $D(q^{-1}) = \nabla$ then it will be referred to as step disturbance. Since the disturbance w affects the error in the same way as r, a step disturbance can be attributed to a change in set point or a step change in $w(k)$.

The Impulse Response

A useful and informative representation of the time series is the impulse response. It can be obtained by a series expansion in q^{-1}, of the time-series model in Eq. (2.4). The equation can then be written as

$$e(k) = \sum_{j=0}^{\infty} f_j q^{-j} a(k) \tag{2.5}$$

The first coefficient of the impulse response, f_0, is often normalized to be equal to 1. The variance of $e(k)$ can be expressed with the impulse response as

$$\sigma_e^2 = \mathrm{var}(e(k)) = \sum_{j=0}^{\infty} f_j^2 \sigma_a^2 \tag{2.6}$$

Considering Eqs. (2.5) and (2.6) together the subscript of the impulse response coefficients gives an idea from what time the contributions to the variance come from. The impulse response coefficients f_j can be plotted as a function of j. For finite variance it is clear that f_j go to zero with j. A plot gives therefore an idea on how fast disturbances are rejected. In [Tyler and Morari, 1996] the relation of impulse response coefficients to common performance specifications is shown.

The pole excess d places a limit on how fast a controller can reject disturbances. The effect of a disturbance acting at time k will be reacted on earliest at time $k+d$. The controller that manages to cancel the noise perfectly as soon as it can is called a minimum variance (MV) controller, see [Åström, 1970]. To calculate the MV controller, full information about the disturbances and the plant is required. Under MV control the error $e(k)$ is a moving average given by the first d terms of the impulse response which the controller can not affect. The variance of $e(k)$ under MV control, denoted σ_{MV}^2 is therefore

$$\sigma_{MV}^2 = (1 + f_1^2 + \cdots + f_{d-1}^2)\sigma_a^2$$

The Harris Index

It was noted in the introduction that variance alone is a poor indication on how well the controller is doing. The problem is that the variance as an absolute statistic is not very informative. It has to be compared with something.

Harris noted that to assess control performance the variance of $e(k)$ could be compared with the MV performance, σ_{MV}^2. This would be more informative on how well the loop was actually doing. If variance of $e(k)$ is close to σ_{MV}^2, better performance can only be obtained by changing the

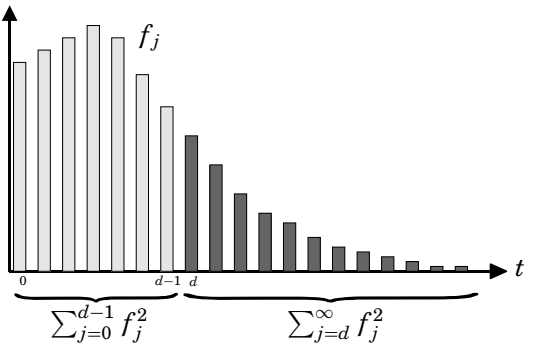

Figure 2.2 An impulse response to show the impulse response coefficients used to calculate the Harris index.

process since σ_{MV}^2 is invariant to feedback. An index of performance could therefore be calculated as the ratio between σ_{MV}^2 and σ_e^2.

Harris also showed that a time-series model identified from $e(k)$ could serve to find σ_{MV}^2. As the first d terms of the impulse response are invariant to feedback they should be equal in the actual system and the identified time series model. A number of single number monitoring algorithms followed the work of Harris.

It was noted in the introduction that a graphical inspection of the impulse response is considered to be one of the most informative performance inspections that can be done on data. Most of the indices calculated from time-series models can be thought to be attempts to automate graphical inspections. The Harris index, for example, can be written as

$$I = \frac{\sum_{j=0}^{d-1} f_j^2}{\sum_{j=0}^{\infty} f_j^2} \tag{2.7}$$

It compares the sum of the d first impulse response coefficients squared to the total sum, see Fig. 2.2.

The Normalized Performance Index

In [Desborough and Harris, 1992] the normalized performance index was introduced. It extended ideas presented in [Harris, 1989] by giving an alternative way of estimating an index related to the Harris index. Assume that a predictor is created to predict b steps into the future for the time series in Eq. (2.4). The best predictor in the mean square sense will be a linear predictor, see [Åström and Wittenmark, 1997]. The predictor can

be obtained by solving a Diophantine equation or equivalently by long division of $B(q^{-1})$ by $A(q^{-1})$.

$$B(q^{-1}) = A(q^{-1})F(q^{-1}) + q^{-b}G(q^{-1}) \tag{2.8}$$

$q^{-b}G(q)$ is the remainder of the division while $F(q)$ is the quotient. The polynomial $F(q)$ is also equal to the first b terms of the impulse response in Eq.(2.5). The predictor is then given by

$$\hat{e}(k|k-b) = \frac{G(q^{-1})}{B(q^{-1})}e(k-b) \tag{2.9}$$

The variance of the prediction error, $\varepsilon_b(k) = e(k) - \hat{e}(k|k-b)$, is

$$\text{var}(\varepsilon_b(k)) = (1 + f_1^2 + \cdots + f_{b-1}^2)\sigma_a^2 \tag{2.10}$$

where σ_a^2 is the variance of the noise $a(k)$.

Notice that when $a(k)$ is a deterministic disturbance, the predictor is found by solving Eq. (2.8) in the same way as when the disturbance is stochastic, see [Åström and Wittenmark, 1997]. The expression to calculate the normalized performance index is the following

$$\eta(b) = 1 - \frac{\text{var}(e(k) - \hat{e}(k|k-b))}{\text{var}(e(k))} = 1 - \frac{\text{var}(\varepsilon_b(k))}{\text{var}(e(k))} \tag{2.11}$$

If b is the dead time of the system, $\eta(b)$ is closely related to the Harris index. When η is calculated for a number larger than dead time, it is referred to as an extended horizon performance index. Assuming η is calculated for b larger than d, then an interpretation is that $\eta(b)$ gives the portion of the variance that can be accounted for with a b step ahead predictor. This allows the normalized performance index to be written as

$$\eta(b) = 1 - \frac{\sum_{j=0}^{b-1} f_j^2 \sigma_n^2}{\sum_{j=0}^{\infty} f_j^2 \sigma_n^2} = \frac{\sum_{j=b}^{\infty} f_j^2}{\sum_{j=0}^{\infty} f_j^2} \tag{2.12}$$

The last equation shows the importance of the impulse response for $\eta(b)$. For a general b the index tells how large portion of the variance of the control signal comes from noise older than b samples. Fig. 2.2 can again be used for a visual interpretation of the contributions to $\eta(b)$. A rise in $\eta(b)$ means that impulse response coefficients older than b have increased compared to the first b ones.

Under the assumption of closed-loop stability, it was noted in [Desborough and Harris, 1992] that the transfer function in Eq. (2.9) can be

written as a convergent series in the backward shift operator. This allows $\hat{e}(k|k-b)$ to be written as

$$\hat{e}(k|k-b) = \sum_{i=1}^{\infty} \alpha_i e(k-b-i+1) \tag{2.13}$$

An estimate of the prediction error can then be attained by truncating the infinite series. Keeping only m terms, this results in the following linear regression model for an estimate of the prediction error $\varepsilon_b(k)$

$$e(k) - \hat{e}(k|k-b) \approx e(k) - \sum_{i=1}^{m} \alpha_i e(k-b-i+1) = \hat{\varepsilon}_b(k) \tag{2.14}$$

The parameter vector $\bar{\alpha}^T = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_m]$ can be estimated with the least squares method by forming the linear regression model

$$e(k) = \hat{\varepsilon}_b(k) + \phi(k)^T \bar{\alpha} \tag{2.15}$$

where $\phi(k)$ is given by

$$\phi^T(k) = [e(k-b) \ e(k-b-1) \ \cdots \ e(k-b-m+1)]$$

An estimate of $\eta(b)$ can be found from the LS solution as

$$\hat{\eta}(b) = 1 - \frac{s^2_{\hat{\varepsilon}_b}}{s^2_y} \tag{2.16}$$

where $s^2_{\hat{\varepsilon}_b}$ is an estimation of the variance of $\hat{\varepsilon}_b$ from the LS solution and s^2_y is an estimate of σ^2_y.

One of the biggest advantages of this approach is that solving a Diophantine equation to find the impulse response coefficients is not necessary. Furthermore, as it is based on least squares, a recursive algorithm to find $\hat{\eta}(b)$ is readily available. Later it is shown how a stochastic gradient algorithm which does not need matrix computations can be used as well with success.

Disturbance Rejection and $\eta(b)$

A common measure of performance in process control is the settling time. Defined from a step disturbance or load disturbance, it refers to the time after which the plant output stays within an interval of width 2δ centered around the final value of the plant output. δ is typical $2-5\%$ of the total change in the plant output. See for example [Seborg *et al.*, 1989].

In [Harris *et al.*, 1999] the relation between $\eta(b)$ and settling time was presented. Settling time is the time b_s when for all $b > b_s$

$$\eta(b) - \eta(b+1) \leq \frac{\delta^2}{\sum_{j=0}^{\infty} f_j^2} \tag{2.17}$$

As a number reflecting disturbance rejection characteristics of the closed loop system, settling time is well known. The relation to $\eta(b)$ is therefore valuable. On the other hand, to check the condition given by Eq. (2.17) is a rather involved task. A good replacement to represent disturbance rejection in a loop ,is $\eta(b)$ itself. $\eta(b)$ has nice interpretations related to how long time it takes to reject a disturbance and reflects predictability in the time series. If $\eta(b)$ is close to 0 for a specific b it means that the effect of a disturbance dies out in at least b samples. $\eta(b)$ can also be more directly related to settling time. For example, defining η-settling time, b_s, as the time when $\eta(b_s) < 0.01$ then the following inequalitys hold

$$\sum_{j=b_s}^{\infty} f_j^2 < 0.01 \sum_{j=0}^{\infty} f_j^2 \quad \Rightarrow \quad f_{b_s}^2 < 0.01 \sum_{j=0}^{\infty} f_j^2$$

If it is assumed f_j are nonnegative decreasing then

$$f_{b_s} < 0.1 \sqrt{\sum_{j=0}^{\infty} f_j^2} < 0.1 \sum_{j=0}^{\infty} f_j$$

Since $\sum_{j=0}^{\infty} f_j$ is the total change if a step disturbance occurs, the inequality gives that b_s is the settling time defined from an interval 10% around the final value. This is a very conservative estimate of the settling time as it ignores all larger impulse response terms with time argument larger than b_s.

2.4 User Defined Benchmarks

In [Kozub and Garcia, 1993] it was noted that often the minimum variance control is neither desirable nor practically achievable. What is considered to be a well functioning loop in the process industry will frequently have variance well above the minimum variance. The Harris index is therefore often too pessimistic and gives false alarms. As a consequence, it is not always useful as an absolute measure of performance but becomes a relative measure which is watched over time for trends. This use of the Harris

index was also noted in [Jofriet and Bialkowski, 1996]. The bottom line is the Harris index has to be compared to historical values. When installed it needs to run an initial period when a "normal" value is registered to which later values are compared.

One reason why the Harris index is pessimistic is that it compares to optimal performance considering all linear controllers. One approach to make it less pessimistic is to limit the set of controllers to the type of controllers commonly used. Typical choices would be PI, PID or common dead-time compensators. Authors addressing this problem are [Ko and Edgar, 1998] and [Isaksson, 1996].

On the other hand, these controllers are seldom tuned for optimal stochastic control. Rather they are tuned for good performance when disturbances are deterministic. So even if the set of allowable controllers is limited, a pessimistic estimate can be obtained. For a nice discussion about this issue see [Isaksson, 1996].

One approach to the problem of obtaining a more absolute measure of performance is to forget about all the limitations and corresponding optimal performance and try to define in advance what performance is acceptable. Alongside with the definition of acceptable performance there should be a condition of acceptable performance which can be checked.

A few authors have addressed this problem. The indices that have been presented have been referred to as user defined benchmarks in contrast to benchmarks defined from process limitations. When starting the monitoring procedure the user specifies with some parameter what is acceptable performance to which actual performance is compared. The drawback of these methods is that they need more information from the user.

Following the observations in [Kozub and Garcia, 1993] the following approach to the problem was presented. An acceptable dynamic response could be specified for Eq. (2.3). As an example, [Kozub and Garcia, 1993] specified a first order filter,

$$e(k) = \frac{1}{1-\mu q^{-1}} a(k)$$

where μ is the decay rate for disturbances entering the system, typically selected from the dominant pole or time constant. Other types of filters could be considered as well. Assuming a noise variance σ_a^2 the variance of $e(k)$ will be

$$\bar{\sigma}_e^2 = \frac{1}{1-\mu^2}\sigma_a^2 \tag{2.18}$$

$\bar{\sigma}_e^2$ denotes variance of $e(k)$ with the specified response. As no delay is assumed $\sigma_{MV}^2 = \sigma_a^2$. These equations could be used to design an alarm limit for the Harris index. If the Harris index, given by Eq. (2.7) was found

statistically to be lower than $1-\mu^2$ then performance could be considered poor. Notice that if there is dead time in the loop this approach can lead to many missed detections of poor performance.

In [Huang and Shah, 1998] a similar idea of a more practical benchmark was the topic. There the IMC framework was used to express user-specified closed loop dynamics. By assuming a disturbance model *a priori* or by estimating it with dither excitation, the response of $e(k)$ could be written as

$$e(k) = (f_0 + f_1 q^{-1} + \cdots + f_{d-1} q^{-d+1} + q^{-d} G_R(q^{-1}))a(k) \tag{2.19}$$

where $G_R(q^{-1})$ depends on the specified dynamics and the noise model. By calculating the variance of $e(k)$ from Eq. (2.19) and comparing with the actual variance an index could be formed as

$$\eta = \frac{\bar{\sigma}^2_{MV} + \bar{\sigma}^2_R}{\sigma^2_e} \tag{2.20}$$

where $\bar{\sigma}^2_R$ is the variance from the term $G_R(q^{-1})$ in Eq. (2.19).

This index should be equal or greater than 1 if performance is satisfactory. For on-line monitoring, the repeated identification of the disturbance model using dither excitation would be considered troublesome. Using an *a priori* decided disturbance model the only thing that would need to be decided in each monitoring instance is σ^2_a. This could be estimated as the variance of a predictor predicting one step ahead. But a test depending only on σ^2_a ignores all temporal characteristics in the impulse response.

In [Tyler and Morari, 1996] it is shown how common design specifications can be expressed as linear inequalities of the impulse response coefficients and how an statistical test can be designed to test if the inequalities are fulfilled.

In [Horch and Isaksson, 1999] an index was presented which can be considered as an extension of the index presented in [Kozub and Garcia, 1993] taking into account the dead time of the plant. Under MV control, all poles of the closed-loop system are placed in the origin. A more realistic benchmark could be obtained by placing one pole in $q = \mu$ and the rest in the origin. The impulse response is therefore assumed to decay exponentially with μ after dead time d. The index can be expressed in terms of the impulse response coefficients as

$$I = \frac{\sum_{j=0}^{d-1} f_j^2 + f_{d-1}^2 \sum_{j=1}^{\infty} \mu^{2j}}{\sum_{j=0}^{\infty} f_j^2} = \frac{\sum_{j=0}^{d-1} f_j^2 + f_{d-1}^2 \mu^2/(1-\mu^2)}{\sum_{j=0}^{\infty} f_j^2} \tag{2.21}$$

The index presented in Eq. (2.21) is the reciprocal of the index presented in [Horch and Isaksson, 1999] to keep consistency with the indices presented earlier. If the index is greater or equal to 1 then the performance

is good. This is equivalent to the condition

$$\sum_{j=d}^{\infty} f_j^2 < f_{d-1}^2 \mu^2 / (1 - \mu^2) \tag{2.22}$$

A graphical interpretation of inequality (2.22) in Fig. 2.2 is that the sum of the squared impulse response coefficients larger than d (the dark ones) should be less than an exponentially decaying response starting from f_{d-1}. Notice that to evaluate this index it is necessary to find one specific impulse response coefficient, namely f_{d-1}.

In [Kozub, 1997] it is mentioned that a possible use of $\eta(b)$ is to check if a settling time specification is satisfied. $\eta(b)$ should be zero for the desired settling time and a deviation from zero indicates that the specification is not fulfilled. The author notes that the statistic can give misleading results when bad initial transients are present in the impulse response. This approach will be noted on later.

Finally, in [Thornhill *et al.*, 1999] it was shown how parameters for the extended horizon performance index could be used for plant wide performance monitoring by selecting prediction horizon and sampling rate according to type of loop. In [Paulonis and Cox, 2003] the work in [Thornhill *et al.*, 1999] is considered a significant advance to lower the barrier towards large scale implementation of a performance index.

Comments on "User Defined Benchmark" Methods

The original argument for user defined benchmark methods was that MV control often was not desirable nor achievable. As a result, false alarms might be sounded. The Harris index might indicate poor performance even though the loop performance was such that maintenance action was not necessary. In [Horch, 2000] it was shown that the Harris Index might miss alarms when limit cycles are present.

The mismatch between minimum variance performance and more normal process control performance requirements was highlighted in [Eriksson and Isaksson, 1994]. For example it was shown that to incorporate integral action in controllers often is guaranteed to result in a worse Harris index for a controller. Still integral action is considered important in process control. Therefore a good stochastic controller which would do well measured by the Harris index will perform poorly when deterministic disturbances affect the loop. The problem does not become easier when entering the realm of deterministic disturbances. It was pointed out that it is a well known fact that a controller tuned for load disturbances is not guaranteed to perform well for set-point changes.

The dilemma is that disturbances that affect a loop vary considerably over time. Most of the time perhaps a stochastic description is most ap-

propriate but when a load disturbance does occur it is vital to compensate for it quickly since it might draw the plant output far away from its set point.

A possible solution though is to try to take changing disturbances into account. Performance assessment for abruptly changing disturbances was the topic in [Huang, 1999]. The algorithm includes the detection of the change. Options on benchmarks are discussed, some relying on selecting the “most representative” disturbance dynamics.

3

λ-Monitoring

3.1 Introduction

The starting point of the work presented in this thesis was the idea to integrate a tuning method for controllers with a performance monitoring method. The initial tuning method chosen was λ-tuning of PI controllers, a method more and more widely used within for instance the pulp and paper industry. The monitoring algorithm that will be presented can be applied to more general controllers. The only assumption on the controller structure is that it results in first-order plus dead time dynamics from set point to plant output. The time constant of the response is denoted λ, and thus the name of the monitoring method. For this case some easy design rules for the monitoring algorithm will be presented.

The potential benefits with this approach are:

- Information from the tuning stage is used directly to commission the monitoring method. This should reduce the commissioning effort. Little or no historical information should be needed.
- If a tuning method is used consistently in a plant one can assume that most operators and plant staff working on control are familiar with the tuning method. To be able to present an integrated approach to monitoring and tuning of controllers is beneficial for acceptance and understanding of the monitoring method.

The monitoring algorithm is based on monitoring a single number statistic, the extended horizon performance index presented in Section 2.3 but with prediction horizon and alarm level chosen with regard to the tuning of the loop.

Unlike the Harris index the method is not supposed to give information about how close the current performance is to optimal performance. The

index calculated rather reflects how well the loop is doing compared to design specifications. The most related methods are the ones presented in Section 2.4.

The method is primarily thought to be a “first indicator” of bad performance after which other methods could be applied to diagnose the root cause. The performance index does not give any diagnosis, it is rather supposed to detect all problem situations but not distinguish between them. The synthetic gradient monitoring that is presented in the next chapter does include some diagnosis elements.

In addition to simulations, the method has been tested on data from 19 λ-tuned PI loops from a Swedish pulp and paper mill. Since most controllers in the process industry are of PI-type, this industrial validation should be of relevance. Furthermore, it was found that many loops had performance problems, increasing the value of the validation. With further analysis and consultation with plant staff a probable root cause was established for most loops. Most of the root causes were rather typical such as limit cycles because of valve stiction and aggressive tuning. It is important to have a good idea how the method will react to typical problem loops and when it will sound an alarm. The application of the method on the problem loops will be presented in a later section.

The assumption that a single tuning method is used consistently does not hold in a large portion of the process industry. Reports that many controllers stand on their default values when audited are common. On the other hand it should be of interest what are the possibilities for monitoring when the tuning practice is improved.

3.2 Assumptions on Tuning

The main condition on the design procedure is that it results in a model of the closed-loop system, from set point to plant output. An example of a design methodology that fulfills this is the Internal Model Control (IMC) methodology, see [Morari and Zafiriou, 1989].

The industrial data that was investigated was from loops where the controller was PI tuned with the λ-tuning method, see for example [Åström and Hägglund, 1995]. The λ-tuning method can be thought of as a special case of IMC tuning. The aim with the method is for a set-point response of first-order plus dead time (FOPDT).

$$T_d(s) = \frac{1}{\lambda s + 1} e^{-Ls} \tag{3.1}$$

The subscript d on T_d stands for design. The main design specification is the time constant, λ, of the set-point response. Even though a monitoring

algorithm devoted to PI controllers only would be justified in light of the popularity of the PI structure, in what follows it will be assumed that the aim of the controller design is to have set-point dynamics described by a FOPDT transfer function like the one given by Eq. (3.1). The controller is assumed to be of one degree of freedom so that associated with the closed-loop transfer function T_d there is a sensitivity function S_d so that $T_d + S_d = 1$. A loop where this is the case after the controller design will be referred to as a λ-tuned loop. A number of common control structures within process control result in λ-tuned loops. Some of these will be reviewed below. All of these can be viewed as IMC controller designs. Furthermore, IMC designs for models found in the process industry commonly result in PID controllers where the set-point response is FOPDT, see [Morari and Zafiriou, 1989].

λ-Tuning of PI Controllers In a document published by the SSG, the Swedish Pulp and Paper Industries' Engineering Co., see [Sko, 1997], the following tuning method is recommended for feedback loop optimization. The controller is a PI

$$C(s) = K\left(1 + \frac{1}{T_i s}\right) \tag{3.2}$$

Given a (FOPDT) model of the plant

$$P(s) = \frac{K_p}{Ts + 1} e^{-Ls} \tag{3.3}$$

the controller parameters are chosen as

$$K = \frac{1}{K_p} \cdot \frac{T}{L + \lambda} \tag{3.4}$$

$$T_i = T \tag{3.5}$$

The resulting closed loop transfer function is given by Eq. (3.1).

This design is an IMC design obtained from approximating the dead time with a first order Taylor approximation, $\exp(-Ls) \approx 1 - Ls$. The disturbance is assumed to be a step disturbance. Parameter λ is typically chosen to be between $\lambda \in [T, 3T]$. If $\lambda = T$ the tuning is considered aggressive while $\lambda = 3T$ is referred to as robust.

λ tuning of a PI controller for an integrating plant with dead time

$$P(s) = \frac{K_v}{s} e^{-Ls}$$

results in the parameters

$$K = \frac{T_i}{K_v(\lambda + L)^2} \tag{3.6}$$

$$T_i = 2\lambda + L \tag{3.7}$$

The closed loop transfer function is then given by

$$T_d(s) = \frac{(2\lambda + L)s + 1}{(\lambda s + 1)^2} e^{-Ls}$$

This is not a λ-tuned loop but it will still be mentioned specially in when the monitoring algorithm is introduced. The disturbance assumed in the IMC design is a ramp at the plant input, see [Rivera *et al.*, 1986].

Simple Dead-Time Compensators In [Morari and Zafiriou, 1989] it was shown how the Smith predictor could be interpreted as an IMC controller. If the model used in the Smith predictor control structure is a FOPDT model and the controller is a PI with parameters selected as recommended within the IMC methodology, the closed-loop transfer function from set point to output becomes equivalent to Eq. (3.1).

The Dahlin Controller In [Dahlin, 1968] a controller structure was presented which is frequently used for machine direction control loops in the pulp and paper industry. Assuming that the plant is given by Eq. (2.1) is stable, that it has no zeros outside the unit circle and furthermore it has no badly damped zeros, the Dahlin-controller is

$$C(q^{-1}) = \frac{[1 - \exp(-h/\lambda)]}{1 - \exp(-h/\lambda)q^{-1} - [1 - \exp(-h/\lambda)]q^{-d}} \cdot \frac{A_u(q^{-1})}{B_u(q^{-1})} \tag{3.8}$$

where h is the sampling time and λ is the time constant given in continuous time. This form is equivalent to an IMC design. The set-point response is of first order with dead time equal to that of the plant

$$T_d(q^{-1}) = \frac{1 - \exp(-h/\lambda)}{1 - \exp(-h/\lambda)q^{-1}} q^{-d} \tag{3.9}$$

In [Laughlin *et al.*, 1987] is was shown how the desired response can be related to the low pass filter F in IMC designs. In [Qin, 1998] it was shown that when the disturbance is an integrating moving average, IMA(1,1),

$$D(q^{-1}) = \frac{1 - c_1 q^{-1}}{1 - q^{-1}}$$

the Dahlin controller is equivalent to a minimum variance controller with closed loop time constant $c_1 = \exp(-h/\lambda)$.

Filter Parameter λ

The parameter λ has a direct and intuitive interpretation as the closed-loop time constant making it suitable for adjustment by plant personnel. An other interpretation within the IMC design methodology is that it is the time constant of the low-pass filter used to provide robustness towards high-frequency model uncertainties.

Continuous versus Discrete Time

With the exception of the Dahlin controller the control structures so far have been presented in continuous time. In the next section when the algorithm will be presented, it is assumed that a discrete model is available of the closed loop. Whether this comes from a sampled version of a continuous time design or from a discrete design is of importance but usually an approximation will be sufficient. The reason for staying faithful to continuous time is that it has a nicer interpretation with regard to dynamic behavior with its time constants and dead times measured in seconds. What is used in the monitoring algorithm is a discrete transfer function of the closed loop.

Disadvantages of λ-Tuning

All of the control structures presented that result in a λ-tuned loop are known to have a good set-point response. It is also a well known fact that their response from load disturbances suffers because of the cancellation of the dominant plant pole T. If λ has been chosen much smaller than T to speed up the set-point response of the system then the load disturbance response will be disappointing since the open loop pole appears in the load disturbance dynamics. However, when λ is chosen in the interval $[T, 3T]$ where T is the dominant plant time constant, this effect matters less.

3.3 The Monitoring Algorithm: λ-Monitoring

It is assumed that the loop under consideration is λ-tuned. In discrete time the resulting FOPDT set-point response is denoted $T_d(q^{-1})$, see for example Eq. (3.9). It is obtained directly if the design is done in the discrete domain, otherwise it is obtained after sampling the continuous time equivalent.

Here it is assumed that L and λ in Eq. (3.1) are an integer multiple of the sampling time. If this is not the case an exact sampling of a continuous time FOPDT model can be found in [Åström and Wittenmark, 1997]. Usually rounding to the closest integer will be sufficient. The relation between the continuous dead time, L, and discrete dead time, d, is $L = (d-1)h$.

The Proposed Algorithm

In the spirit of [Thornhill *et al.*, 1999] and [Kozub, 1997] it is proposed that the extended horizon performance index, $\eta(b)$, should be monitored, see Section 2.3. The prediction horizon, b, is selected as an engineering criterion, not from type of loop as in [Thornhill *et al.*, 1999] but from the tuning of the loop. The monitoring algorithm monitors that the loop has certain disturbance rejection capabilities reflected by $\eta(b)$. The selection of b and the alarm limit depends in addition to the tuning on the disturbance model. As the index is supposed to reflect how a loop is fulfilling the design specifications, the disturbance model is set equal to the one in the controller design. For λ-tuning this is a step disturbance

$$D(q^{-1}) = \frac{1}{1-q^{-1}}$$

The controller error is according to design specifications equal to

$$e(k) = [1 - T_d(q^{-1})]D(q^{-1})a(k) \tag{3.10}$$

where $a(k)$ can be thought of as a deterministic or stochastic disturbance. The impulse response of $e(k)$ is then

$$e(k) = (1 + q^{-1} + \cdots + q^{-d} + \alpha q^{-d-1} + \alpha^2 q^{-d-2} + \cdots)a(k) \tag{3.11}$$

where $\alpha = \exp(-h/\lambda)$.

Prediction Horizon and Control Limit In Fig. 3.1 $\eta(b)$ is shown for λ-tuned loops when the disturbance model is an integrator. In that case an exact equation of $\eta(b)$ when $b > d$ can be found by considering the impulse response in Eq. (3.11). The equation for $\eta(b)$ is

$$\eta(b) = \frac{\alpha^{2(b-d)}}{1 + d(1-\alpha^2)} \tag{3.12}$$

In selecting the prediction horizon and the corresponding alarm limit the following guidelines were observed. As the monitoring should reflect how well the controller is doing fulfilling the design specifications, the main characteristics of the curves in Fig. (3.1) should be captured. But as $\eta(b)$ is monitored for only one value of b one could say that only one point is checked on the curves in the figure. Aiming for a similar disturbance rejection as measured by $\eta(b)$, the curve should have fallen to levels close to 0.1 when the prediction horizon is equal to $(\lambda + L)/h$ samples. The

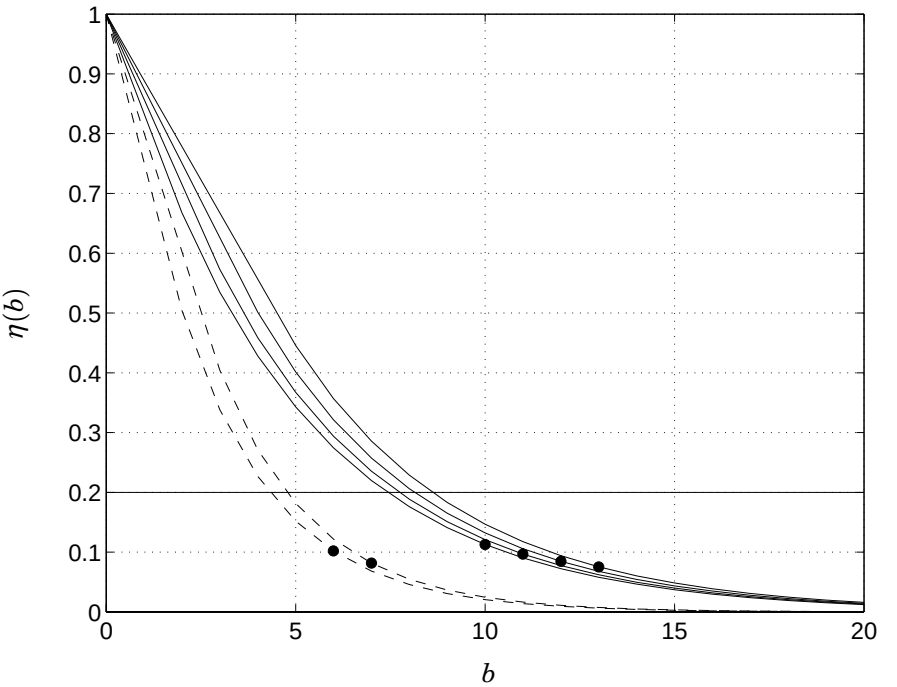

Figure 3.1 $\eta(b)$ for λ-tuned loops with disturbance model as an integrator. The cases shown are $\lambda = 5$ for $L = 1, 2$ (dashed) and $\lambda = 9$ for $L = 1, 2, 3, 4$ (solid). The smaller the dead time the lower is $\eta(b)$. Dots are $\eta(b)$ when $b = (\lambda + L)/h$.

prediction horizon is therefore placed in the "knee" of the curve when it has fallen significantly and is about to level out.

The alarm limit should be selected in accordance with the prediction horizon so that when the alarm is sounded the real $\eta(b)$ should have deviated from the desired value. To make false alarms sufficiently rare the statistical properties of the estimate $\hat{\eta}(b)$ should be taken into account. The frequency of false alarms can then be estimated. The approximate first two moments of $\hat{\eta}(b)$ were given in [Desborough and Harris, 1992]. The variance of $\hat{\eta}(b)$ depends on b and size of the data series that is used.

The basic tradeoff to consider when selecting the prediction horizon and the alarm limit is the number of false alarms versus the number of missed detections of poor performance. But the idea of poor performance varies across the process industry and therefore the prediction horizon should be considered an engineering criterion to be set by plant engineers. The recommended prediction horizon, $(\lambda+L)/h$ should in most cases serve as a good starting value.

A large portion of this chapter is devoted to demonstrating for what situations the index presented here indicates poor performance in an industrial environment. The industrial data mentioned in the introduction will be used for this purpose. This is the main justification for the suitability of the recommended prediction horizon and alarm limit.

In [Thornhill *et al.*, 1999] the recommended length of the data series was given as 1500 samples. For this data series size it was found that the standard deviation of $\hat{\eta}(b)$ was between $0.04 - 0.8$ for the industrial data used. Actually a value down to 1000 produced a satisfactory result. More variation was on the standard deviation between loops than when the data series length was changed from, for example, 1000 to 1500.

Selecting the control limit approximately two standard deviations above the target value the control limit is suggested to be 0.2. The control limit is shown in Fig. 3.1 as a solid horizontal line. For λ-tuning of PI controllers for integrating processes, a suitable prediction horizon found in a similar way was $b = 2\lambda + L$.

In [Thornhill *et al.*, 1999], a strategy for plant wide implementation of a similar performance index was presented which could be adopted for the index presented here.

Sampling Time and Model Order for Estimation of $\eta(b)$ It is beneficial to select the sampling time of monitoring as long as possible to relive load on DCS computers. It is proposed that the sampling time is chosen in relation to λ so that there are $5 - 9$ sampling intervals within each λ. It is not uncommon that a single sampling time is chosen for a whole plant even though the loops in the plant have time constants with orders of magnitude difference. If the recommended sampling time requires down sampling the output data from the DCS, the data should be filtered before the index is calculated. Following recommendations in [Sko, 1997] for filtering when λ-tuning PI controllers, prior to down sampling the data should be filtered with a first order filter with time constant $1.3h$ where h is the new sampling time.

In [Thornhill *et al.*, 1999] the recommended length of the truncated autoregressive model to estimate $\eta(b)$ was $m = 30$. This was found to be a sound recommendation but frequently good performance could be achieved with lower model order.

Short Summary of Monitoring Method

For a λ-tuned loop it is recommended that $\eta(b)$ is monitored where the prediction horizon is set as $b = (\lambda + L)/h$. The data series length should be between $1000-1500$ samples. The sampling time, h, should be selected so that λ equals $(5-9)\cdot h$. The model order should be 30 but often a lower

model order will be sufficient. The index that results will in the following be referred to as the λ-index and denoted as I_λ.

3.4 Recursive Implementation of λ-Monitoring

In [Desborough and Harris, 1992] it was noted that the linear regression approach used to find an estimate of $\eta(b)$ could be implemented in a recursive manner for online implementation. But even the recursive least square (RLS) algorithm can be considered burdensome, specially when model order m is large as it involves calculations with matrixes of size $m \times m$. A simpler recursive algorithm is the LMS or stochastic approximation, see for example [Ljung, 1999]. The difference lies in the update direction, RLS uses a Gauss-Newton update direction while the LMS uses a steepest descent update direction. The simplicity is paid for by slower convergence rate.

The forgetting factor for a recursive algorithm can be chosen as

$$l = 1 - 1/N \tag{3.13}$$

The length of the data series affects the reaction time of the recursive version of the index to changes in disturbance spectra. When λ equals 5 sampling periods, the above recommended data series length was sometimes to long. Then a data series length between $1000 - 1200$ was more appropriate.

The complete recursive algorithm to estimate I_λ is now given. Two vectors of length m are needed, one data vector containing $e(k)$ for times $k-b$ to $k-b-m+1$ and the parameter vector $\alpha^T = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_m]$.

$$\begin{aligned}
\phi^T(k) &= [\ e(k-b) \quad e(k-b-1) \quad \cdots \quad e(k-b-m+1)\ ] \\
R(k) &= lR(k-1) + \phi(k)^T\phi(k) \\
\varepsilon(k) &= e(k) - \phi^T(k)\alpha(t-1) \\
\alpha(k) &= \alpha(k-1) + \phi(k)\varepsilon(k)/R(k) \\
E(k) &= lE(k-1) + \varepsilon^2 \\
S(k) &= lS(k-1) + e(k)^2 \\
I_\lambda(k) &= 1 - E(k)/S(k)
\end{aligned}$$

where l is again the forgetting factor. $S(k)$ is equal to the sum

$$S(k) = \sum_{i=0}^{\infty} l^i e^2(k-i)$$

With a normalization of $1/(1-l)$ it is an estimate of the variance. In the same way $E(k)$ is an estimate of the prediction error variance, $\mathrm{var}(\varepsilon_b(k))$, with the normalization. The starting values, $S(0)$ and $E(0)$, should be chosen with this in mind. The starting value of $R(k)$ can be set equal to $R(0) = m \cdot \mathrm{var}(e(k))$.

3.5 Validation on Industrial Data

A monitoring algorithms acceptance in an industrial environment depends on a variety of factors. As was mentioned before, a very important one is the rate of false alarms and missed detections. It is therefore important to characterize for what loop conditions alarm will be sounded. For industrial relevance the method should be tested on industrial data.

For validation of the monitoring method, it was applied to data from 19 loops in a pulp and paper mill. The loops were all controlled with PI-controllers tuned with the λ-tuning method. The time constants are shown in Table 3.1. Most of the data was from self-regulating pressure and flow loops and level control loops. The loops had been tuned up to one year before the data was collected. The data series were from 6 hours to 28 hours. In some series the disturbance characteristics changed notably over time.

The value quoted for T in the level control cases was the integral time from the PI controller. Because of the size of λ for the level control loops I_λ could only be estimated for very few batches making the evaluation not very informative.

The application of the monitoring method to the industrial data indicated that several of the loops had performance problems. With inspection of the data and consultation with plant staff a probable root cause was established for most of the problem loops. The root causes were rather typical and could be classified into categories frequently cited in the literature. These classifications are shown in Table 3.1.

To demonstrate some properties of the monitoring method examples will be shown from the industrial data for some of the root cause categories. The signals and indices presented in this chapter and the next one are shown in Appendix A for a selection of loops. All data was originally sampled with 1 second sampling time. For the calculation of the index the data series were down sampled as presented earlier according to the value of λ in the loop.

The plant staff generally agreed that the method indicated poor performance when it should have.

Table 3.1 Data on industrial loops. T and L are values of the open loop time constant and dead time.

Loop	T	L	λ	Performance
831FC001	4	1.5	5.4	Good/Sluggish
831FC002	5.4	1.2	5.4	Limit cycle, stiction
831FC028	6	1.4	6	Bad, reason unknown
831FC030	4.7	1.6	5.4	Limit cycle, stiction
831FC039	3.5	1.3	10.6	Saturates
890FF050	5	1.5	10	Limit cycle, stiction
890FF051	5	1.5	10	Load disturbances/Aggressive tuning
831FF160	1.3	1.8	5.4	Good
831FF176	1.5	1.6	4.9	Bad step response
831FF177	1.2	1.5	4.5	Good
890QC040	5.1	2.3	15.3	In manual
831LC361	614	2	304	Good
LC57306	1135	10	562	Good
831LC031	455	25	215	Good
832PC108	5	2	5	Good
832PC109	6	3	6	Aggressive tuning
834PC062	10	1	30	Quantization
834PC068	10	2	45	Quantization
836TC021	41.0	123	246	Saturates

Oscillations

Oscillations in loops is a frequent phenomena within the process industry. Studies have shown that up to 30% of all loops within a typical pulp and paper plant may be oscillating. The reason for the oscillations can be many. Typical ones are problems associated with valves such as valve stiction and backlash. Inappropriate tuning or change in loop gain following a change in operating region can cause the loop to become unstable. The amplitude of the resulting oscillation increases until nonlinear effects such as saturation limit it. But the output continuous to oscillate persistently.

A number of loops were found to oscillate. Some examples will now be presented for two different root causes, valve stiction and aggressive

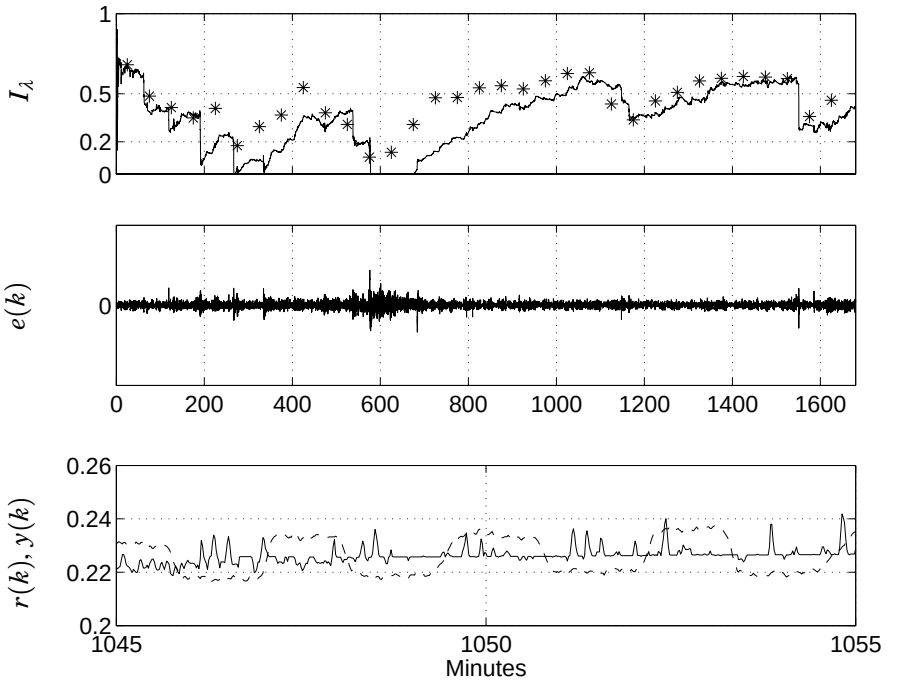

Figure 3.2 Loop 890FF050. Loop with valve stiction. Top graph: I_λ. Middle graph: $e(k)$. Bottom graph: set point $r(k)$ (solid) and output $y(k)$ (dashed).

tuning.

Valve Stiction The most common reason for oscillations was valve stiction. Three flow loops were found to have this problem. In Fig. 3.2 the λ-index is shown for one of the loops (loop 890FF050). The index lies above 0.2 for most of the series length. The lowest graph shows the reference value and the output signals for an interval when I_λ was close to 0.5. An easily recognizable square-shaped signal form can be noticed. The control signal had the familiar sawtooth shape as well.

Loops 831FC002 and 831FC030 also seemed to have stiction in the valve. The indices for these loops are shown in Appendix A.

Remark: Dynamic Properties of Recursive Estimator The data from the current loops serves well to demonstrate some dynamic properties of the recursive algorithm presented. The original data was sampled with a sampling time of 1 second. Since λ was equal to 10 seconds, the data series was down sampled with a sampling time of 2 seconds. With a data

series length of 1500 data points, each data series covered 50 minutes. The recursive algorithm can be thought to have a similar time constant since the forgetting factor is chosen with Eq. (3.13). But that does not always coincide with the observed behavior.

Notice a sharp fall in the recursively estimated index at time 580. As the batch estimated index falls as well for this time, the observed fall in the recursively estimated index is correct. But it takes very long time for the recursively estimated index to catch up with the batch estimated index. At time 900 the recursively estimated index is finally closing in on the batch estimated one.

The reasons for this difference between the batch estimated and the recursively estimated index are mainly two. The stochastic gradient method has a certain convergence speed which is considerably slower than than the convergence speed for the recursive least square method, for example. Secondly, the recursive estimator contains an estimation of the variance of the control error $e(k)$ denoted $S(k)$ and of the prediction error $\varepsilon_b(k)$ denoted $E(k)$.

Notice that there is extra variability in the control error around the time when the recursive estimate falls. Closer inspection of the error reveals a change in the dynamic characteristics. This causes an increase in prediction error variance as the prediction of the error inherent in the estimation of the index does not work well for the new dynamic characteristics. So at this time both control error variance and prediction error variance increase sharply. This causes a drop in the estimated index as the ratio $E(k)/S(k)$ becomes very close to 1. When the variability decreases again the recursively estimated variances decay exponentially. But as the λ-index depends on the ratio of the two variances it takes longer time for the index to recover.

Aggressive Tuning Aggressive tuning was thought to be the root cause of performance deterioration when the output signal oscillated at a frequency close to the ultimate frequency of the loop or in the frequency range where the sensitivity function was greater than 1. Also a visual inspection of the wave form was used to rule out other types of oscillatory disturbances such as valve stiction. To determine if the oscillation was originated internally or externally was difficult since no experiments could be performed on the loop for verification. An indication that the oscillation originated internally was that the control signal would be 180^o out of phase with the output signal. The loops in Table 3.1 that are denoted aggressively tuned fulfilled this condition.

The amplitude of the sensitivity function is larger than 1 for frequencies close to the ultimate frequency. This amplitude can usually be lowered by reducing the loop gain. To call the loops aggressively tuned is not wrong

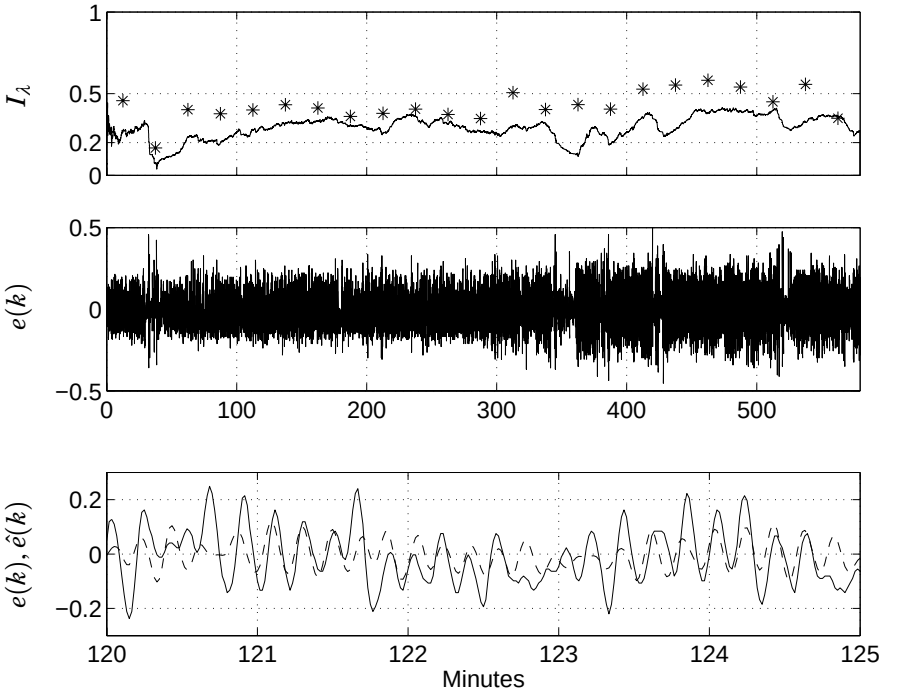

Figure 3.3 Loop 832PC109. Aggressively tuned loop. Top graph: I_λ recursively estimated (solid line) and batch wise estimated ($*$). Middle graph: $e(k)$. Bottom graph: $e(k)$ and the predicted error, $\hat{e}(k|k-b)$.

even if the oscillation originates externally since the signal oscillates at a frequency that the sensitivity function amplifies.

In Fig. 3.3 λ-monitoring is applied to the data from a pressure control loop (Loop 832PC109). In the upper graph, I_λ, is shown calculated batch wise and with the recursive algorithm in Section 3.4. The error is oscillating considerably. The period of the oscillation is about 10 sec which is around 0.6 rad/sec. The ultimate frequency of the loop with the λ-tuning is 0.7 rad/sec.

In the lowest graph of the figure the error $e(k)$ and the b-step ahead predicted error, $\hat{e}(k|k-b)$ is shown for some minutes. b is equal to $(\lambda + L)/h$. It can be seen that the predictor manages to predict parts of the oscillations thus reducing the prediction error. This in turn will cause the I_λ to increase.

It is of interest to try to characterize how oscillatory the output signal will have to become before the λ-index indicates poor performance. In

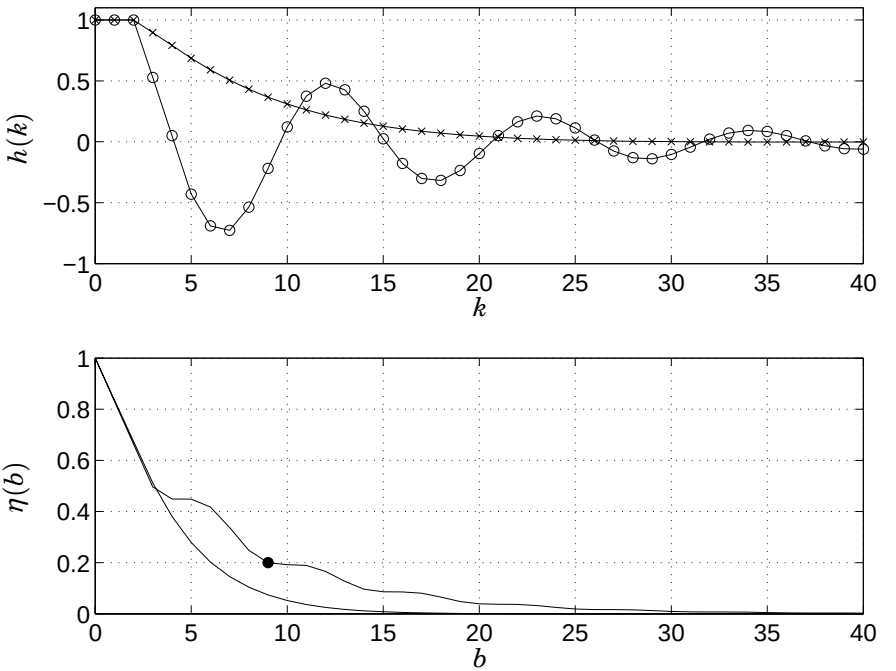

Figure 3.4 Normal and oscillatory impulse response and $\eta(b)$. The dot shows the alarm limit, 0.2, for the prediction horizon $b = 9$

Fig. 3.4 two impulse responses are shown. Both are from a loop where a FOPDT process is controlled with a PI controller. What is shown is the impulse response of the error when the disturbance model is an integrator. The damped response corresponds to when the PI controller is λ-tuned normally with $\lambda = 7$ and $L = 2$ assuming unit sampling. For the oscillatory response, the loop gain had been increased by a factor of 4. Also shown in the figure is $\eta(b)$ for both cases. For the oscillatory response, $\eta(b)$ equals 0.2 for the prediction horizon $b = 7 + 2$. As expected, when the impulse response coefficients start to become larger for values larger then λ, $\eta(b)$ rises the most for values of b close to λ.

It can be thought as an disadvantage of the monitoring method that the response can become this oscillatory before an alarm is sounded. The gradient method presented in the next chapter will complement this weakness of the λ-monitoring method. It should be pointed out that the variance of the oscillatory response in Fig. 3.4, calculated by summing the squared impulse response coefficients, is actually lower than the nominal case so performance measured in variance has not started to suffer yet.

A few loops seemed to display the characteristics of aggressive tuning, originating inside the loop, for short intervals only. In Appendix A loops 890FF050 and 890FF051 showed this behavior for very short time intervals.

Remark In [Kozub, 1997] a monitoring approach was suggested where the index to monitor was $\eta(b)$ but with b equal to the desired settling time. It is said that the statistic is valuable when only settling time is of importance but that this approach suffers from the fact that large initial transients in the impulse response go unnoticed. In [Kozub, 1997] the definition of the settling time is not given but for a λ-tuned loop the settling time is usually considered to be equal to 4λ. In the current case this would mean $b = 28$. The problem with this approach is then seen in Fig. 3.4. When performance deteriorates $\eta(b)$ changes first for small b's. Before a noticeable change occurs for larger b the performance has deteriorated severely.

Sluggish Tuning and Slow Disturbances

If a loop works at many operating regions it is frequent that the tuning ends up being the most conservative one or the most robust one over all operating regions. The reason for this is that as performance problems occur at any operating region it is common to detune the controller, typically by lowering the gain, K, in the PI controller. Because of lack of time of operators the old parameter values are not put back in the controller when the loop operates at the more normal operating region again.

The consequence of this is that many controllers are detuned. When load disturbances or set-point changes occur the error signal tends to stay away from zero for a long time. In [Hägglund, 1999] a method to detect sluggish control loops is presented. Both controller output and plant output are used to determine if the loop is sluggish. The index describes the relation between times of positive and negative correlation between the control and plant output increments.

If the error signal stays away from zero for a long period of time, the λ-index will give an alarm indicating bad performance, as it will quickly be able to predict the error. A sluggish control loop will therefore be discovered by the λ monitoring method. On the other hand, if a slow load disturbance affects the loop, for example a step or a ramp on the plant input, the error signal will also stay away from zero for long periods of time. This will cause the λ-index to indicate poor performance even though the loop might be working well. So it might be claimed that the λ-monitoring method will give false alarms when slow disturbances affect the loop. It might also be claimed that these alarms will not be false alarms since they indicate that the disturbance model used in the design, i.e. a step

disturbance, is incorrect. The problem is that load disturbances might be quite frequent and the operators might not be interested in these alarms.

A remedy would be to redesign the prediction horizon for a different disturbance model. It is for these situations that the prediction horizon and alarm level of the λ-index should be considered an engineering criterion. On the other hand there is no clear upper limit that should be put on the prediction horizon. If the prediction horizon is designed for load disturbance it will give alarm when a temporary ramp is applied to the plant input in stead of a step. Furthermore, longer prediction horizon means less sensitivity to other faults like aggressive tuning.

The final suggestion is that the prediction horizon should be treated as an engineering criterion and adjusted if slow disturbances affect the loop. If the slow disturbances are less frequent, an other option would be to expect occasional alarms. When the alarms occur they could trigger inspection or application of the method presented in [Hägglund, 1999] to determine if tuning is sluggish.

Once again the gradient method will complete the weakness of the λ-monitoring method since it will give information about the frequency content of the disturbances.

In Fig. 3.5 data from loop 831FC001 is shown. The batch wise estimated λ-index is constantly far above 0.2, actually close to 0.7 for the whole period. The recursively estimated λ-index is always below the batch wise estimated one but does rise quickly above 0.2 so that an alarm is sounded. In the bottom graph of Fig. 3.5 the error is shown for a shorter interval so that the dynamics can be inspected more carefully. It is seen that the error sometimes stays minutes away from 0, not consistent with the λ-tuning which is 10 sec. The reason for the sluggishness was a combination of bad tuning, the controller gain had been decreased but also that the reference value was changing in a ramp like manner.

In Fig. 3.6 the impulse response and $\eta(b)$ are shown for two systems. Both of them are λ-tuned loops with a integrator as the disturbance model. The faster response has $\lambda = 7$ and $L = 2$. The slower response has $\lambda = 13$. It can be seen that $\eta(b)$ for the slower response intersects 0.2 for prediction horizon $b = 9$. The conclusion is that if this loop would have been monitored assuming $\lambda = 7$ and $L = 2$, and later detuned, the monitoring would indicate poor performance when the loop would have a time constant of $\lambda = 13$. So in this case λ can be almost doubled before the λ-index sounds an alarm.

Remark: Stochastic and Deterministic Disturbances Notice that with the current formulation no difference is made between stochastic and deterministic disturbances. In both cases the problem can be interpreted as finding the best b-step ahead predictor and the corresponding

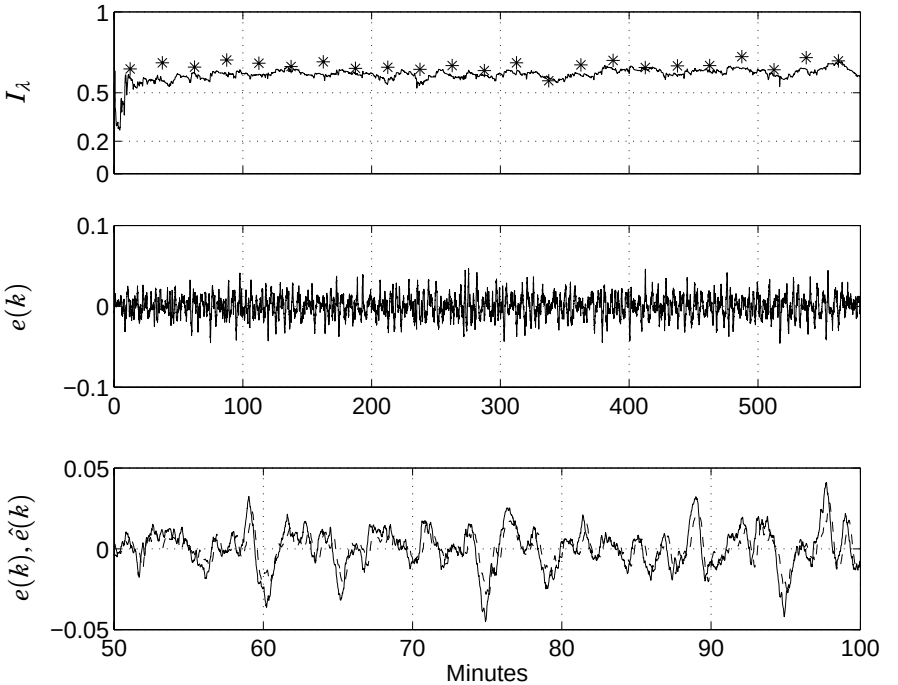

Figure 3.5 Loop 831FC001. Sluggish loop. Top graph: I_λ recursively estimated (solid line) and batch wise estimated ($*$). Middle graph: $e(k)$. Bottom graph: $e(k)$ and the predicted error, $\hat{e}(k|k-b)$.

prediction error variance which is compared to the variance of the error signal $e(k)$. It is considered a strength that all types of disturbances are treated in the same way.

In Fig. 3.7 a step change in the set point is shown for the loop 831FF176. This loop was tuned with $\lambda = 4.9$ and $L = 1.6$. As can be seen in the figure the step response does not fulfill the specifications from the tuning phase. According to the tuning the settling time should be no more than 30 sec. The settling time of the step response in the figure is around 50 sec. The recursively estimated performance index rises above 0.2 as it should. An other example of a set-point response which does not fulfill design specifications is shown in Fig. A.3 in Appendix A. An example of a loop shown in the Appendix that did fulfill design specifications in terms of set-point response was loop 831FF177.

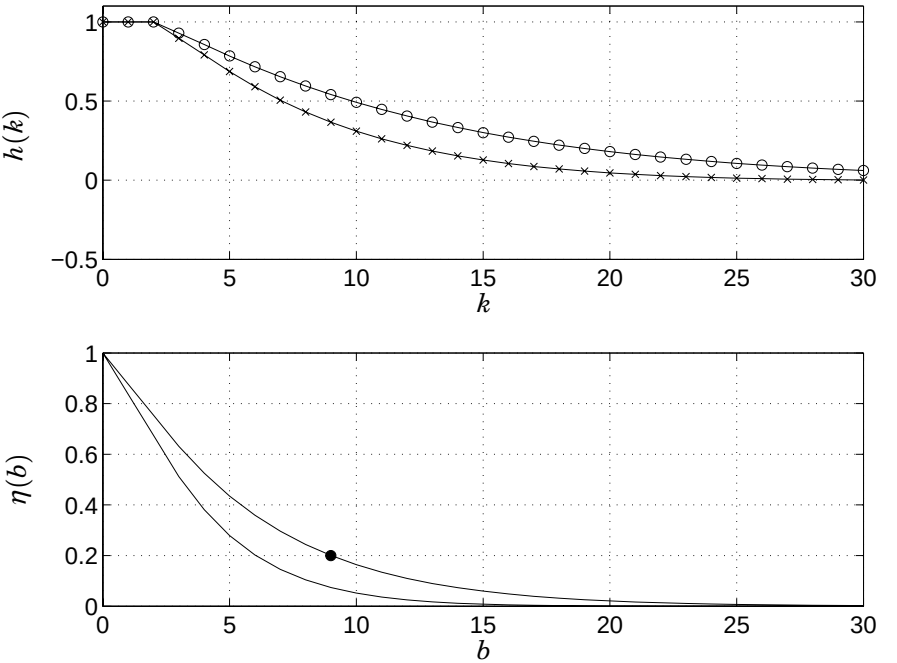

Figure 3.6 Top graph, normal λ-tuned impulse response (x) and sluggish impulse response (o). Lower graph, $\eta(b)$ for normal and sluggish impulse response.

3.6 Conclusions

A nonintrusive performance monitoring method for λ-tuned loops has been described. λ-tuned loops are those loops with an one-degree-of freedom controller whose set-point response is of first-order plus dead time. λ is the time constant of the set-point response.

The method is thought to be a first indicator of bad performance after which existing diagnosis algorithms could be applied. For online implementation in distributed control systems, a simple recursive algorithm to estimate the index has been presented. All parameters of the monitoring method are set by using the tuning. The need for historical information should be none or little.

The method applies equally to stochastic or deterministic disturbances. Bad performance is related to low variance of prediction error when the control error is predicted over a time horizon related to the tuning. Low prediction error variance means high predictability which in turn means that the controller has not done a good job rejecting the disturbances

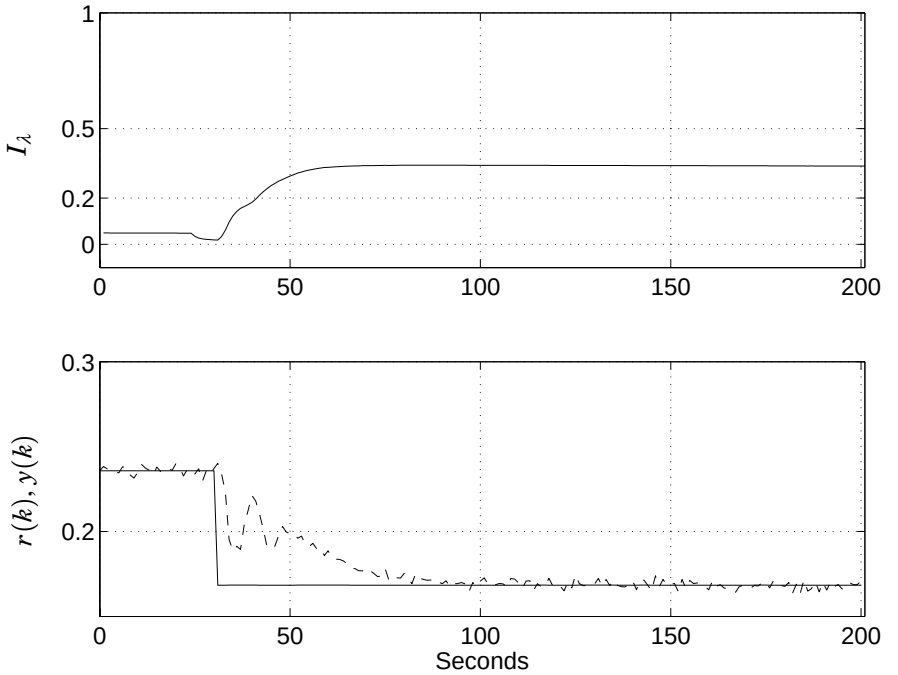

Figure 3.7 Loop 831FF176. Set-point response. Top graph: λ-index recursively estimated. Bottom graph: set point $r(k)$ (solid) and output $y(k)$ (dashed). A step down occurs in the set point at time 30 sec.

within a suitable time horizon.

The method has been tested on 19 λ-tuned PI loops from a Swedish pulp and paper mill. The method indicated poor performance in many of the loops. In most of the problem loops a typical root cause could be established explaining the poor performance. The method has therefore been tested on typical problem loops from the process industry.

4

Gradient Monitoring

4.1 Introduction

One of the more frequent acts of maintenance of PID controllers is to adjust the controller gain. This is often done without a systematic approach, often following the preferences (or whims) of a particular operator. The reason for the parameter adjustment is usually abnormal variability that the operator thinks can be fixed by an adjustment. Abnormal disturbances can for example occur when the loop is working in a new operating region.

What is shown in this chapter is how to estimate in a simple way a gradient of a quadratic cost function with respect to a controller parameter. The purpose is to make this already existing maintenance approach, more systematic. By using the gradient, decisions to adjust controller parameters can be made with more information. The plant staff should be familiar with the controller parameter for which the gradient is calculated. It should also be strongly correlated with plant variability.

It will be shown that in many cases the equations for the gradient become particularly simple allowing recursive implementation in most DCSs. The focus will be on λ-tuned loops but as long as the controller is available and a model exists of the closed-loop system, an estimate of the gradient will be possible.

For λ-tuned loops the parameter for which the gradient is calculated is λ. The reason for this is to keep consistency with the previous chapter and because λ has a nice interpretation as the closed-loop time constant. For loops controlled with a PID an other choice of controller parameter might be the gain K.

The gradient will be interpreted in a simple way from the sensitivity function. Furthermore, in the λ-tuning case a normalization of the gradient will be presented so that it becomes independent on variance and

parameter value.

The synthetic gradient will be introduced through the IFT (Iterative Feedback Control) framework. It should be noted that similar gradient calculations are frequent within the adaptive control community, see [Åström and Wittenmark, 1995] and as within the IFT framework the gradients are used to update the controller parameters to lower the cost function.

It is possible to use the gradient presented here to initiate adjustment of the controller parameters to lower the cost function. But no automatic adjustment is considered. An operator is always thought to be "in the loop" of parameter adjustment. The gradient gives information about the spectral content of the disturbances affecting the loop. This information is compressed into one number related to a controller parameter. As this information can be collected and monitored over time it might provide a valuable overview over the state of the loop and the disturbances affecting it. For example if the gradient is watched over many operating regions it might indicate opportunities for gain scheduling.

4.2 Iterative Feedback Tuning

Iterative feedback tuning has recently emerged as a technology to tune fixed order controllers like the PID by performing experiments on the closed-loop system. The tuning is performed by calculating the controller parameter gradient of a quadratic cost function and modifying the parameters in the descent direction of this cost function. A reference covering most aspects of IFT is [Hjalmarsson *et al.*, 1998].

The method deals with SISO linear systems on the form

$$y(k) = P(q^{-1})u(k) + w(k) \tag{4.1}$$

$w(k)$ is the process disturbance. The controller is assumed to be of one degree of freedom.

$$u(k) = C(\rho)(r(k) - y(k)) \tag{4.2}$$

The dependence of transfer functions on the back shift operator q^{-1} will be omitted in what follows. Instead the dependence on controller parameters ρ will be written. The closed-loop system is then given by

$$\begin{aligned} y(k) &= \frac{C(\rho)P}{1 + C(\rho)P} r(k) + \frac{1}{1 + C(\rho)P} w(k) \\ &= T(\rho)r(k) + S(\rho)w(k) \end{aligned} \tag{4.3}$$

$T(\rho)$ is the complementary sensitivity function while $S(\rho)$ is the sensitivity function. It is easy to check that $T + S = 1$. The time argument of the signals will now be omitted as well but again they will be written as a function of controller parameters ρ.

Let y_d be the desired response to the reference signal r. Given that the design method resulted in a closed loop transfer function T_d then $y_d = T_d r$. In the case of λ-tuning T_d is the first order response with λ as the time constant. Putting $\tilde{y} = y - y^d$ the cost function that is monitored is of quadratic type,

$$J(\rho) = \frac{1}{N} E\left[\sum_{t=1}^{N} \tilde{y}(\rho)^2\right] \tag{4.4}$$

For large N, J is an estimate of the variance of $\tilde{y}$. Notice that $\tilde{y}$ is not equal to the error $e(k)$ in the previous chapter. In the IFT framework a term containing the square of the input signal can be included. To simplify the derivation this term is excluded here. The expectation is with respect to the weakly stationary disturbance w.

The question that arises is how $\partial J/\partial \rho$ is obtained. It was shown in [Hjalmarsson *et al.*, 1994] that an estimate of this gradient could be computed from signals obtained from closed-loop experiments. Assume that N measurements of r and $y(\rho)$ have been logged. Measurements from this series are denoted with a subscript 1. Elementary calculus gives

$$\frac{\partial J}{\partial \rho} = \frac{2}{N} E\left[\sum_{t=1}^{N} \tilde{y}(\rho) \frac{\partial \tilde{y}(\rho)}{\partial \rho}\right] \tag{4.5}$$

It is assumed that

$$\frac{\partial \tilde{y}(\rho)}{\partial \rho} = \frac{\partial y(\rho)}{\partial \rho}$$

From Eq. (4.3) one can see that (dropping the argument of ρ)

$$\begin{aligned} \frac{\partial y}{\partial \rho} &= \left(\frac{P}{1+PC}\frac{\partial C}{\partial \rho} - \frac{P^2 C}{(1+PC)^2}\frac{\partial C}{\partial \rho}\right) r \\ &- \frac{P}{(1+CP)^2}\frac{\partial C}{\partial \rho} w \\ &= \frac{1}{C}\frac{\partial C}{\partial \rho} T\ r - \frac{1}{C}\frac{\partial C}{\partial \rho}\left(T^2\ r + TS\ w\right) \end{aligned}$$

Assuming the data is from data series 1, notice now that

$$T^2\ r_1 + TS\ w_1 = T\ y_1$$

Using this in Eq. (4.6) results in the following equation for $\partial y(\rho)/\partial \rho$

$$\frac{\partial y}{\partial \rho} = \frac{1}{C}\frac{\partial C}{\partial \rho} T \, (r_1 - y_1) \tag{4.6}$$

The above equation implies that it is possible to calculate $\partial y/\partial \rho$ by filtering $r_1 - y_1$ through the complimentary sensitivity function. One way to obtain this term is to apply $r_1 - y_1$ as a reference value to the real process and sampling the output yielding measurement series 2. This gives an estimate with an error $S\, w_2$ since noise would be added. This is the approach within the IFT methodology. An other approach is to estimate T with system identification techniques, and filtering $r_1 - y_1$ with this identified transfer function. This is the approach taken in [Bruyne and Carrette, 1997]. A third approach, and the one used in this paper, is to use $T = T_d$ obtained from a design method. The gradient calculated by using the last two approaches is referred to as synthetic.

In the synthetic gradient case, an estimate of $\frac{\partial J}{\partial \rho}$ is formed by using one measurement series and evaluating Eq. (4.5). A perfect realization of $\tilde{y}$ is $y_1 - T_d r_1$. Eq. (4.5) can now be written as

$$\frac{\partial J}{\partial \rho} = \frac{2}{N}\sum_{k=1}^{N} (y(k) - T_d r(k)) \, \frac{1}{C}\frac{\partial C}{\partial \rho} T_d \, (r(k) - y(k)) \tag{4.7}$$

The subscript 1 is omitted since all signals come from the same data series.

4.3 Monitoring the Gradient

It is suggested that Eq. (4.7) is used to calculate the gradient with respect to the desired control parameter. A recursive algorithm will be presented in a later section. The number of data points to form the estimate is between 1000-1500 as in the previous chapter on the λ-index. T_d is obtained from the design specification. In the case of a λ-tuned loop the controller parameter of choice is λ with the corresponding set point response. The filter $\frac{1}{C}\frac{\partial C}{\partial \rho}$ will now be presented for a few controllers.

In what follows it will be assumed that λ and L are multiples of the sampling time h. If this is not the case, the closest integer usually is a good enough approximation.

The Gradient for Common Controllers

The equation for the synthetic gradient is now further evaluated for common controllers. The main transfer function to evaluate is the filter

$$\frac{1}{C}\frac{\partial C}{\partial \rho}(q^{-1}) \tag{4.8}$$

λ-Tuned PI Controllers The λ-tuning rule for PI-controllers was introduced in Section 3.2. In discrete time form, with the integral approximated with a forward approximation, the λ-tuned PI controller becomes

$$C(q^{-1}) = \frac{T}{K_p(L+\lambda)}\left(1 + \frac{hq^{-1}}{T(1-q^{-1})}\right)$$

This assumes that the parameters of a FOPDT model of the plant are available, see Eq. (3.3). Then the filter in Eq. (4.8) becomes

$$\frac{1}{C}\frac{\partial C}{\partial \lambda}(q^{-1}) = \frac{-1}{L+\lambda} \tag{4.9}$$

In the case of integrating processes, λ enters the controller in a more complex way. The controller is

$$C(q^{-1}) = \frac{2\lambda + L}{K_p(L+\lambda)^2}\left(1 + \frac{hq^{-1}}{(2\lambda+L)(1-q^{-1}}\right)$$

which yields

$$\frac{1}{C}\frac{\partial C}{\partial \lambda}(q^{-1}) = \frac{-2}{L+\lambda}\frac{\lambda - (\lambda - h)q^{-1}}{2\lambda + L + (2\lambda + L - h)q^{-1}} \tag{4.10}$$

This is a first-order filter and very simple to implement.

The Dahlin Controller With the expression of the Dahlin controller given by Eq. (3.8) the filter becomes

$$\frac{1}{C}\frac{\partial C}{\partial \lambda}(q^{-1}) = -\frac{1-q^{-1}}{1-aq^{-1}-(1-a)q^{-d}}\frac{ha}{(1-a)\lambda^2} \tag{4.11}$$

where $a = \exp(-h/\lambda)$.

General PID A general form for a discrete PID controller on serial form, (see [Åström and Hägglund, 1995]) can be written as

$$u(k) = K\left(1 + I(k) + D(k)\right) \tag{4.12}$$

where $I(k)$ and $D(k)$ are the derivative and integrating terms. The parameter for which the gradient could be calculated is the gain K. The equation for the gradient becomes particularly simple in this case as

$$\frac{1}{C}\frac{\partial C}{\partial K}(q^{-1}) = \frac{1}{K} \tag{4.13}$$

Both equations (4.9) and (4.13) are constants which simplifies the evaluation of Eq. (4.7).

In Eq. (4.11) the filter has derivative action. The filter in Eq.(4.10) is close to have derivative action as well when λ is much larger than h. But notice that the signal to which $\frac{1}{C}\frac{\partial C}{\partial \rho}$ is applied in Eq. (4.7) is first filtered with T_d which always has low-pass character.

4.4 Interpretation

Insight into what information the gradient gives can be obtained from Parseval's relation which relates the cost function in Eq. (4.4) with a frequency domain expression

$$J(\rho) \approx \frac{1}{2\pi}\int_{-\pi}^{\pi} |S(\rho)|^2 \Phi_{ww} \tag{4.14}$$

where S is the sensitivity function. Here it is assumed that the reference value is constant and the error is explained solely by the disturbance $w(k)$. Φ_{ww} is the spectra of the disturbance. The controller parameters affect the sensitivity function by lifting up or dragging down the amplitude on specific frequency intervals. If the power spectrum of the disturbances is concentrated in a frequency region where an increase in a parameter increases the amplitude of the sensitivity function, the gradient with regard to this parameter will be positive. If the sensitivity function amplitude is reduced with a positive change in the parameter the gradient will be negative.

The sensitivity function is shown in Fig. 4.1 for a λ-tuned loop. As seen in the figure, a decrease in λ pulls down the low frequency gain while the gain above approximately $1/\lambda$ will be larger than 1. Above $1/\lambda$, a decrease in λ pushes the amplitude up. If the disturbances would be concentrated in this frequency region the gradient would be negative.

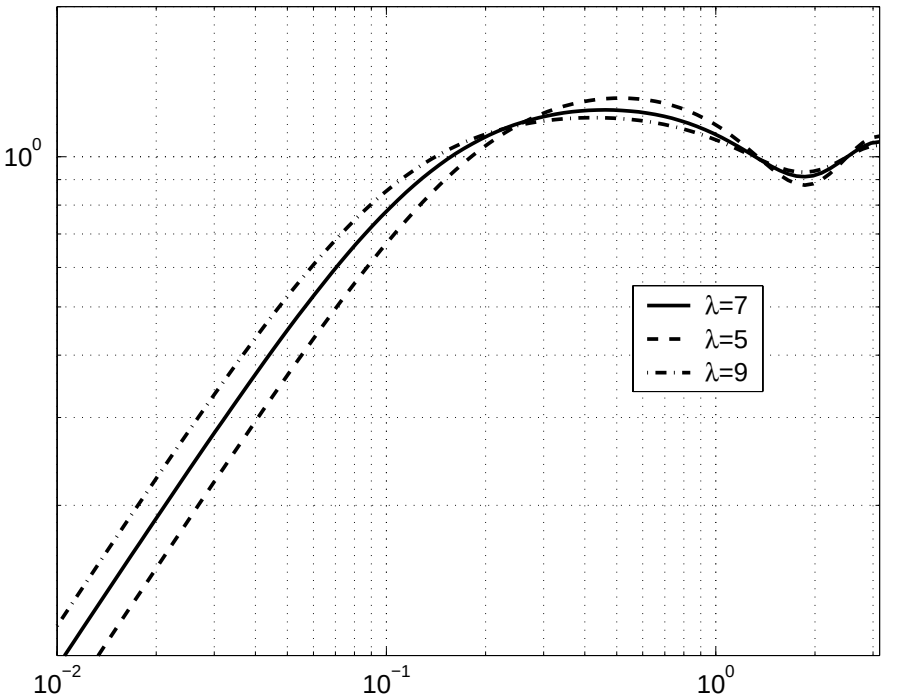

Figure 4.1 Sensitivity function for a λ-tuned loop with $L = 2$ and varying λ.

Normalized Gradient

The number returned from the gradient estimation in Eq. (4.7) depends on the process and also on the variance of the loop. It is of interest to obtain a more absolute number which does not depend on the process characteristics as such but rather, when calculated for a new process, similar numerical values should mean the same thing. One way to do this is to normalize the gradient obtained with Eq. (4.7). Denoting the normalized gradient as β, a straight forward way of doing this in the case of λ-tuned controllers is

$$\beta = \frac{\partial J}{\partial \lambda} \frac{\lambda}{J} \tag{4.15}$$

As a dimensionless number it is informative to know what numerical values this number attains and what they mean. To investigate this an assumption has to be made about the disturbances affecting the system. Assume now that the disturbance is concentrated at one dominating frequency. Approximately the disturbance can then be represented with one sinusoidal of frequency ω_0. As a sampled system is considered it is as-

sumed that ω_0 is smaller than the Nyquist frequency. The spectra of the disturbance in this case is a sum of delta functions

$$\Phi_{ww} = \pi \left(\delta(\omega - \omega_0) + \delta(\omega + \omega_0) \right)$$

see for example [Ljung, 1987]. Evaluating the integral in Eq. (4.14) with the delta functions as the disturbance spectra gives $J = |S(\rho, e^{i\omega_0 h})|^2$. If the loop is λ-tuned this results in a J dependent on ω_0, λ, L

$$J(\lambda, L, \omega_0) = |S_d(\lambda, L, e^{-i\omega_0 h})|^2 = \left| 1 - \frac{1 - \exp(-h/\lambda)}{e^{i\omega_0 h} - \exp(-h/\lambda)} e^{-iL\omega_0} \right|^2$$

where it has been used that $S_d = 1 - T_d$ where T_d is the set-point response of the λ-tuned loop. Notice that h can be assumed to be 1 since all parameters with units in time are scaled with h. This expression can be analytically derivated with regard to λ and the normalized index calculated as a function of the three parameters. The expressions become rather involved but in Fig. 4.2, β is plotted for a range of λ/L ratios.

The normalized gradient takes values between -0.3 and 1.7 for the cases presented in the figure. For λ between 5 and 9 and with varying L the curves look very similar. For low frequencies, around 0.01 rad/sec the curves level out above $\beta = 1$. The smallest value of β is obtained below zero. For low frequencies the system with the highest λ/L ratio attains the highest value of the gradient. For high frequencies the system with the lowest λ/L ratio attains the lowest value of the gradient.

It is of value to compare the sensitivity functions in Fig. 4.1 with the normalized gradient in Fig. 4.2. Both contain the case when $\lambda = 7$ and $L = 2$. The gradient crosses 0 for the frequency where the amplitude of the sensitivity function is invariant to small changes in λ. In Fig. 4.1 this is where the sensitivity functions seem to intersect.

If the gradient is zero it also implies optimality of variance for the parameter for which the gradient is calculated. Stepping away from the scenario where the disturbance contains one frequency only, for λ-tuned loops this is the point where any reduction in variance due to lowering of the amplitude curve in one frequency region is compensated by a rise in the amplitude in an other frequency region. But it should be kept in mind that the gradient is calculated for the sensitivity function obtained from the λ-tuning. Even though it is useful to think in terms of the sensitivity function to interpret the gradient it is based on the nominal situation where the controller and plant are known exactly and all disturbances enter according to Eq. (4.1). The actual situation might be different. For example a negative gradient is interpreted as a high concentration of energy in the disturbance spectra above the crossover frequency. The real

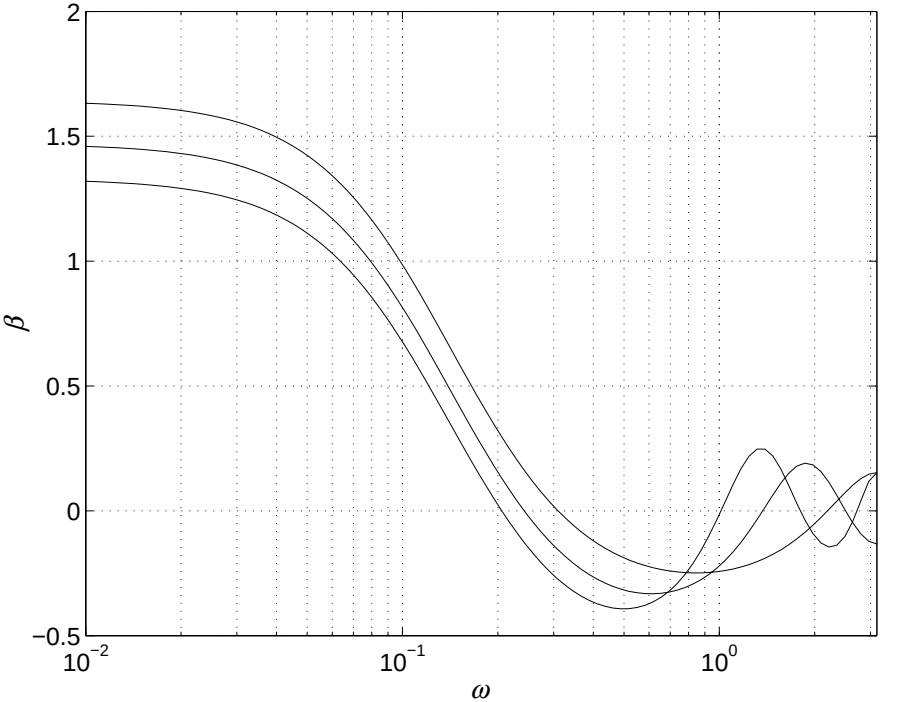

Figure 4.2 Normalized gradient β for $\lambda = 7$ and $L = 1, 2, 3$. For low frequencies the lowest curve has the largest L.

reason is not unlikely aggressive tuning of the loop. Aggressive tuning with a slight model-plant mismatch will result in a real sensitivity function that has a very large amplitude peak above but close to the crossover frequency. This high peak will amplify frequencies in the disturbance spectra which will become dominant in the output.

The normalized gradient should be thought of as a coarse spectral analysis where the disturbance spectra is compared to the sensitivity function from the controller design stage and the result squeezed into one number related to the most important tuning parameter. Considering again the two figures, Figs. 4.1 and 4.2 a positive gradient close to or above 1 means that most of the disturbances are of low frequency character. For such a loop there might be opportunities to increase the gain of the controller and make it more aggressive without amplifying too much high frequency noise. A negative gradient means variability should decrease if the controller is made less aggressive. To give a point of reference for what would be a "normal" gradient, for a λ-tuned loop with a disturbance

model which is an integrator the gradient lies around $0.5 - 0.7$.

4.5 Recursive Implementation of the Normalized Gradient

A recursive implementation will now be presented for the case of λ-tuned PI controllers. The difference between this case and other cases lies in the filter $\frac{1}{C}\frac{\partial C}{\partial \rho}$ and perhaps the specified set-point response T_d. To create the recursive version of the gradient, it is the sum in Eq. (4.7) which is expressed on recursive form with a forgetting of older values. It is assumed that the available signals are $\{y(k), r(k)\}$.

Some of the signals need to be filtered with the set-point response obtained from the λ tuning. A discrete form of this transfer function is

$$T_d(q) = \frac{1 - \exp(-h/\lambda)}{1 - \exp(-h/\lambda)q^{-1}} q^{-d}$$

see Section 3.2 where the λ-tuning is introduced. A filtered signal, $e(k)$ is denoted $e^f(k)$. The recursive form is

$$\begin{aligned}
\tilde{y}(k) &= (y(k) - r^f(k)) \\
\gamma(k) &= \frac{-2}{\lambda + L}\tilde{y}(k)(r^f(k) - y^f(k)) \\
G(k) &= lG(k-1) + \gamma(k) \\
S(k) &= lS(k-1) + \varepsilon^2(k) \\
\beta(k) &= G(k)/S(k)\lambda
\end{aligned}$$

$G(k)$ denotes the gradient sum in Eq. (4.7). $S(k)$ is the variance sum. Notice it is not the same as the one that was calculated in the recursive equations for the λ-monitoring algorithm. l is the forgetting factor. If the set point is not changing, $\varepsilon(k)$ is the controller error, $e(k)$. If the set point is changing, $\varepsilon(k)$ is the error between $y(k)$ and $r(k)$ taking the response of the closed-loop system into account.

4.6 Use of the Gradient

In this section some of the possible uses of the gradient will be demonstrated. To begin with, a motivating example based on simulations will be presented. Then two applications of the gradient on industrial data will be considered. One will deal with detecting aggressive tuning of loops. The other addresses detection of sluggish loops. The gradient is shown for the loops presented in Appendix A. Other possible uses of the gradient will be discussed in the section on future work.

A Motivating Simulation Example

The example is supposed to give insight into the behavior of the gradient when the loop gain depends on the operating region. Consider the system

$$P(q^{-1}) = K_p \frac{1 - \exp(-1/7)}{1 - \exp(-1/7)q^{-1}} q^{-3}$$

controlled with a PI controller which is λ-tuned for $K_p = 1$ with $\lambda = 7$. The disturbance model is a slow load disturbance,

$$D(q^{-1}) = \frac{1}{1 - \exp(-1/9)q^{-1}} \nabla$$

see Fig. 2.1 for a block diagram. The input to the disturbance model was normally distributed white noise. This system has an amplitude margin of 5.9. Assume now that the gain K_p depends on the output so that when $y = 0$, $K_p = 1$ and when $y = 1$, $K_p = 5.8$. A simulation of this system is shown in Fig. 4.3. From 0 to 400 minutes the set point stays at 0. Then a slow ramp is added to the set point so that it reaches 1 at 1000 minutes. The set point stays at 1 for the rest of the time. So the system goes from a condition where a slow disturbance is affecting it to a condition where it is marginally unstable but with the same slow disturbance. The marginally unstable system will have a large peak in the sensitivity function for a frequency where the nominal one is larger than 1. Because of the peak in the sensitivity function the frequency content of $e(k)$ will change. After 1000 minutes the system is quite oscillatory. This will be picked up by the gradient.

In Fig. 4.4 the normalized gradient, the λ-index from the previous chapter, and a recursively estimated variance are shown. The normalized gradient was calculated batch wise with Eq. (4.7) and with the recursive algorithm. The λ-index is also shown for both cases. The λ-index indicates bad performance in the beginning and in the end of the time interval. In the beginning it is the slow disturbances which causes I_λ to rise. In the end it is the oscillatory behavior due to the increase in loop gain.

The gradient takes the largest and smallest value in the beginning and in the end of the time period. As expected for the slow disturbances the gradient is large and positive. This would indicate that opportunities to improve the tuning might be at hand. The interval with the lowest variance is roughly between 500 and 900 minutes. The λ-index lies between 0.2 and zero on this interval, indicating that the controller is rejecting disturbances within the correct time horizon. On this interval the gradient lies between 0.7 and zero. Above 1000 minutes the variance increases drastically because the system is marginally stable. The gradient

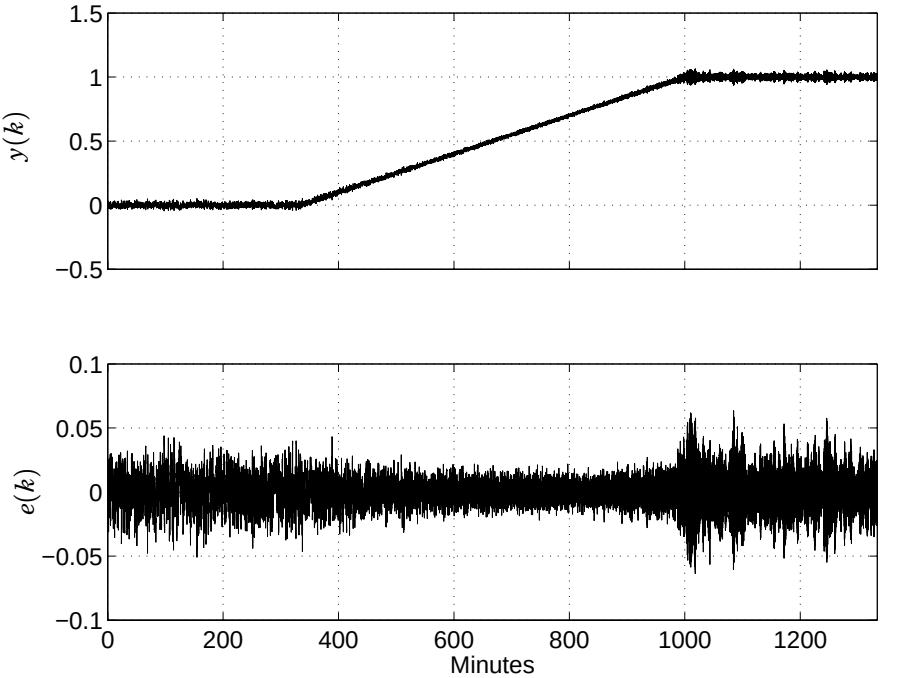

Figure 4.3 Simulation example. Bottom graph: control error $e(k)$. Top graph: plant output $y(k)$.

is negative on this interval indicating that variance could be reduced by increasing λ.

This example shows that valuable information can be obtained by monitoring the gradient. With regard to the previous mentioned praxis of parameter adjustment, the gradient would indicate the following parameter adjustment opportunities. In the beginning of the time interval, the way to reduce variability would be to decrease λ. Since the gradient is large and positive, the loop is mostly affected by low frequency disturbances. A decrease in λ would most probably reduce the effect of these disturbances without increasing the amplitude of the high frequency disturbances too much. In the end of the time interval the disturbances affecting the loop have most of their energy concentrated in the region where the sensitivity function amplifies the disturbances. λ should be increased to make the tuning less aggressive.

As mentioned before, it is common that the tuning of a loop ends up being the most conservative over all operating regions. In the current example, if the loop is expected to work at $y = 1$ the tuning of the loop

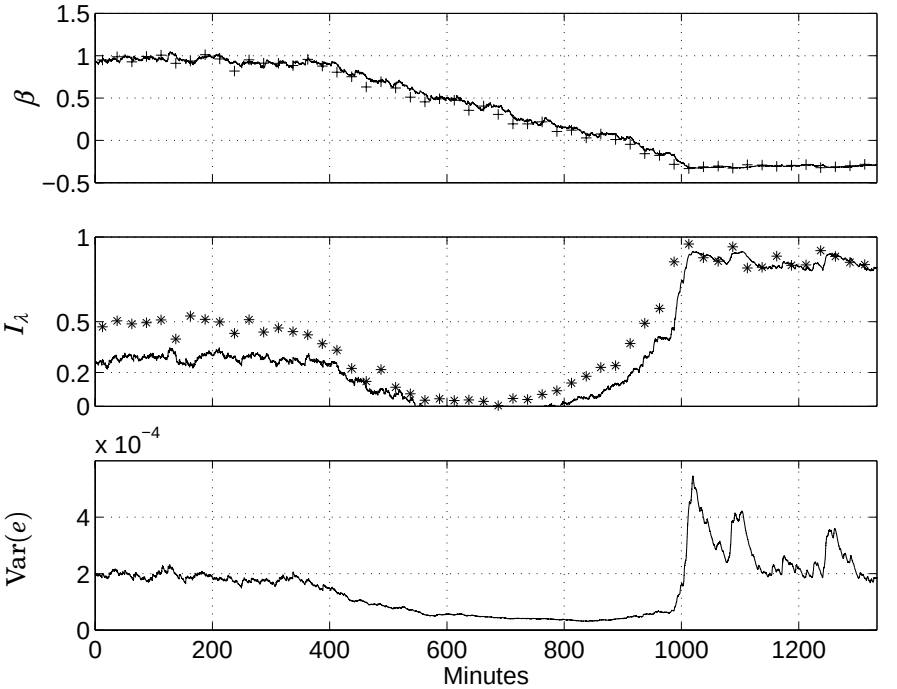

Figure 4.4 Simulation example. Top graph: normalized gradient β calculated batch wise (+) and recursively (−). Middle graph: λ-index calculated batch wise (*) and recursively (−). Bottom graph: the variance estimated recursively.

is too aggressive since the gradient is negative at this operating region. On the other hand it is obvious that increasing λ to make the loop less aggressive would result in increased variability at operating region $y = 0$ as the disturbance affecting the loop there is very slow. With this in mind perhaps a more suitable solution would be gain scheduling. This is one area where monitoring the gradient might be of assistance, to discover opportunities for gain scheduling.

The gradient will be between $0.7 - 0.5$ in the nominal case when a λ-tuned loop is affected white noise filtered through an integrator. Considering this fact and keeping the current example in mind a suitable interval for the normalized gradient is between 0 and 0.7.

Detecting Aggressively Tuned Loops

In Fig. 4.5 the gradient is shown, calculated batch wise and recursively for the pressure loop 823PC109, previously discussed in Section 3.5. Also shown in the figure is the λ-index and the controller error.

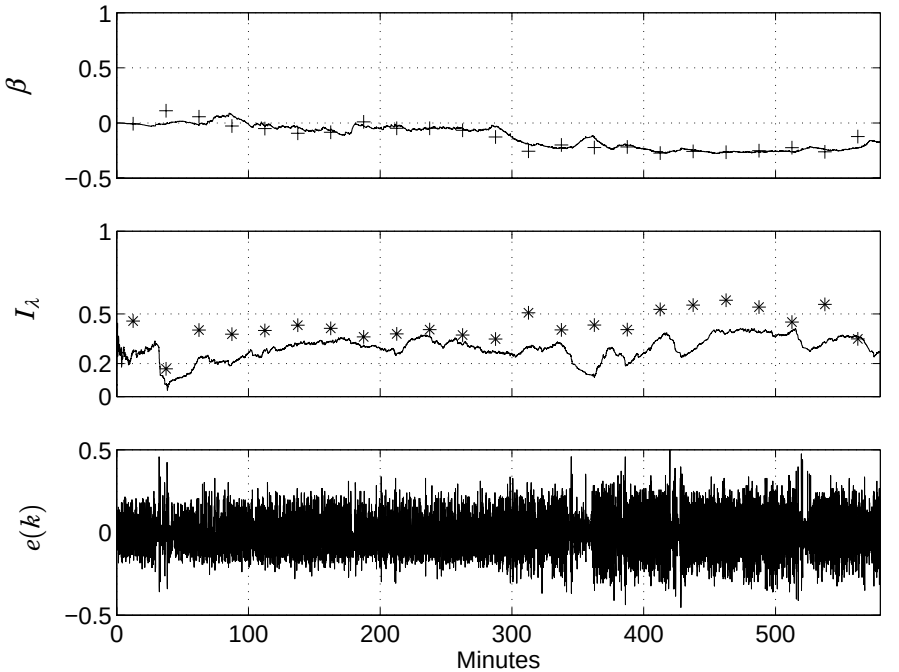

Figure 4.5 Loop 823PC109. Top graph: β estimated batch wise and recursively. Middle graph: λ-index. Bottom graph: $e(k)$.

The gradient is close to zero or negative over the whole interval. A loop with negative gradient presents a very clear opportunity to reduce variability in the plant by adjusting the relevant parameter. In this case the λ could have been increased. The fact that the λ-index is well above 0.2 for most of the interval also indicates that the loop has problems. Neither the λ-index nor the gradient indicates that the oscillation observed is due to aggressive tuning of the loop. But in either case an adjustment of the controller parameters is justified since even if the disturbances originate outside the loop, their spectral composition is such that the current loop is increasing the variability instead of decreasing it. The appropriate action is still to increase λ.

Detecting Sluggish Loops

In Fig. 4.6 the normalized gradient is shown for the flow loop 831FC001, previously discussed in Section 3.5.

It takes some time for the recursively estimated gradient to settle at the same level as the gradient estimated batch wise. But it settles at a

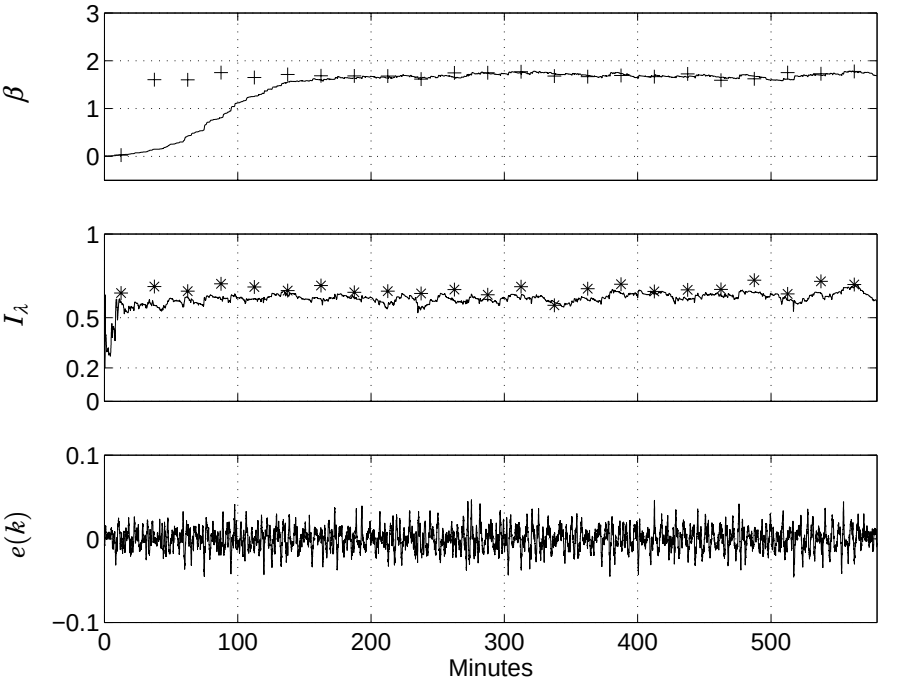

Figure 4.6 Loop 831FC001. Top graph: β estimated batch wise and recursively. Middle graph: λ-index. Bottom graph: $e(k)$.

similar level around 1.7. This is a very large value of the gradient and indicates that most disturbances are of low frequency character. In the case of this loop, there might have been opportunities to increase the aggressiveness of the tuning by reducing λ.

Notice that the loops shown in Figs. 4.5 and 4.6 have similar λ values or 6 and 5.4 respectively. Their data is also shown over the same time interval.

Monitoring Over Operating Ranges

In Fig. 4.7 the statistics are shown for loop 831FC002 where two grade changes took place during logging of the data. The set point changes at time 50, 550 and 1550 minutes. The variance changes considerably following each of these set point changes.

The λ-index is above 0.2 for most of the interval indicating that the loop is not rejecting the disturbances as efficiently as required by the λ-tuning. It starts at very high values close to and above 0.5. A quick view of the control signal and plant output indicated strongly valve stiction

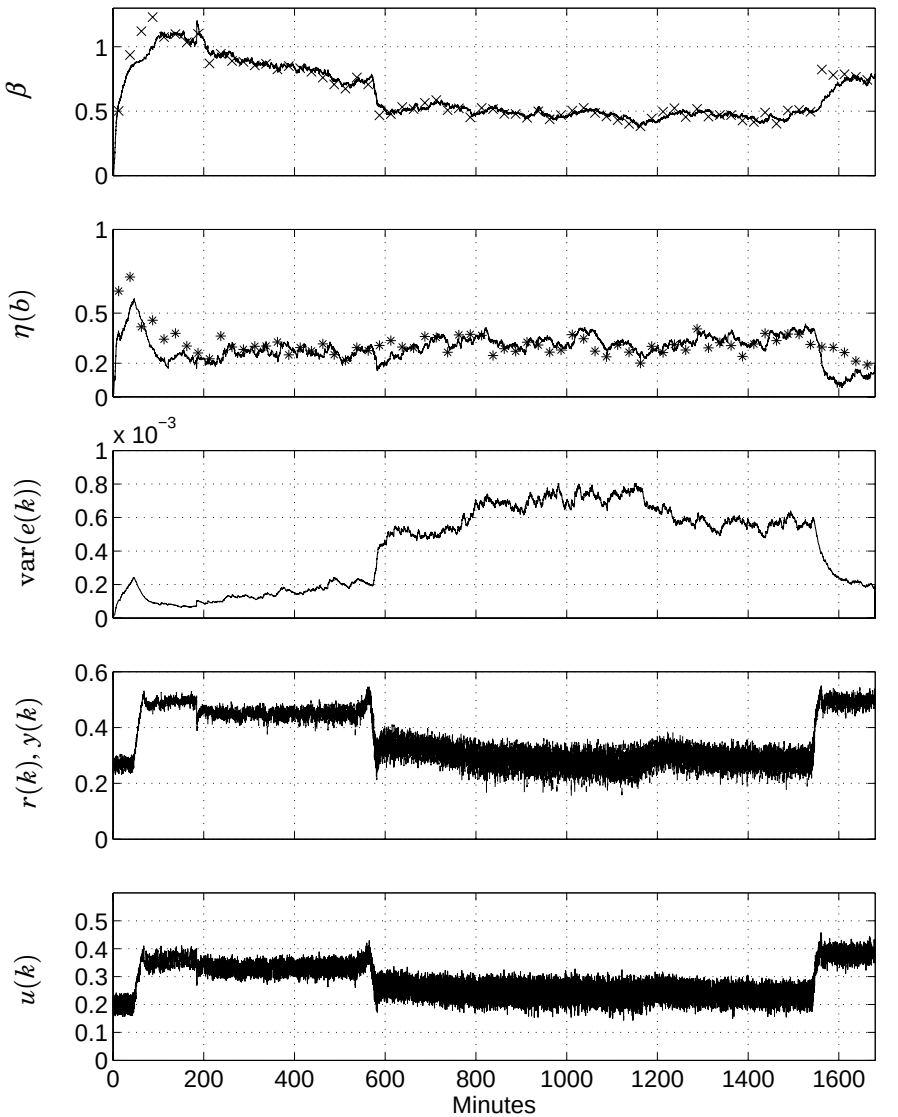

Figure 4.7 Loop 831FC002. Changes in gradient over operating ranges. Graphs from top to bottom: β estimated batch wise and recursively, the λ-index, variance of $e(k)$, set point and plant output, control signal $u(k)$.

as the root cause of the problem. This is shown in greater detail in Fig. A.6. The assumption of linearity previously discussed does therefore not hold for this loop. This fact might explain other unexpected observations within the data series.

The set point is changed at time 50 from 0.25 to 0.5 approximately. Following the change there is a notable drop in variance. The gradient settles around 1 indicating slow disturbances at the new set-point level. The slow disturbances seem to be caused by the stiction but the frequency of the limit cycle is considerably reduced from the time interval $0 - 50$. This indicates lower integral gain. As the integral gain was not changed specially in the PI controller over the time the lower integral gain can only be explained by a lower process gain for the new operating region. Then from time 50 to 550 minutes the gradient is decreasing as variance is increasing. The stiction pattern is still apparent but more difficult to distinguish as the set point is changed as well. The decreasing gradient implies that the spectral composition of the control error is moving up to higher frequencies. This in turn indicates increasing loop gain over the interval. Notice that in a flow loop with a valve actuator the gain can easily change when pressure difference over the valve changes. Then at time 550 the set point is changed back to 0.3, close to the previous level of 0.25. This is followed by a very big increase in variance and significant drop in the gradient down to 0.5. Again one can suspect that the reason for increased variance is an increase in loop gain.

Some conclusions from these observations will now be summarized. The change in loop gain at the different operating regions is picked up by the gradient as there is a change in the spectral composition of the error. However, the linearity assumption which is important for any spectral analysis, does not hold for this loop. As can be expected there are inconsistencies in the information obtained from the gradient. The largest value of the gradient is in the beginning of the interval 50 to 550 minutes. There it is above 1 indicating a sluggish loop. But this is the point where the loop has lowest variability. The gradient has a value of 0.5 at the operating level with the largest variance, between 550 and 1550 minutes. The numerical value of 0.5 for the gradient should be acceptable. But the fact that the λ-index is above 0.2 for the whole interval should indicate that something is wrong with the loop. A conclusion is that if there are "strong" nonlinearities in the loop the information from the gradient will not be reliable.

4.7 Discussions and Future Work

Missing in the treatment are results from experiments in an industrial environment where gradient data has been used in a more proactive way to reduce variability and optimize performance. If it is of interest to keep certain loops tightly tuned, λ could be trimmed until the gradient would be on the interval $0 - 0.7$. As long as variability would drop, the closer to zero the better.

For loops known to have different disturbance characteristics for different operating regions the gradient might give information valuable for implementation of gain scheduling.

For most of the industrial data, when the λ-index indicated good performance, i.e. I_λ was below 0.2, the normalized gradient was found to lie on the interval $[0, 0.7]$. An example of this can be seen in Appendix A for two of the loops. Consider the loops 831FF176 and 831FF177 which have long intervals of good performance. Both loops have intervals where the control signal saturates but outside of these intervals I_λ is below 0.2. When I_λ is below 0.2 the gradients are positive and close to zero. Loop 890FF051 shows similar behavior if the batch wise estimated statistics are considered. The gradient is actually above 0.7 but always below 1.

4.8 Conclusions

A statistic has been presented which gives useful information regarding the assumed sensitivity function and the disturbances affecting the loop. The statistic is an estimate of the partial derivative of the variance, with regard to a control parameter.

It has been shown that the evaluation of this gradient is very simple for many common controllers. A recursive version has been presented that should be implementable in most distributed control systems. In the case of λ-tuning the gradient has been shown to have a nice interpretation by considering the sensitivity function of the λ-tuned loop. A normalization of the gradient has been presented as well and it has been shown that its numerical values should lie between 0 and 0.7. The gradient was calculated for the industrial data presented in Chapter 3. For most of the industrial data, the condition of good performance according to the λ-index was almost always followed by a value of the gradient on the desirable interval.

The fact that the statistic is related to a simple tuning rule should facilitate its acceptance in an industrial environment.

A

Industrial data

Data from some of the loops in Table 3.1 is shown on the following pages along with the λ-index, I_λ, and the normalized gradient, β. Most loops shown here were found to have some kind of performance problem. Some good loops where the λ-index was below 0.2 most of the time are included as well.

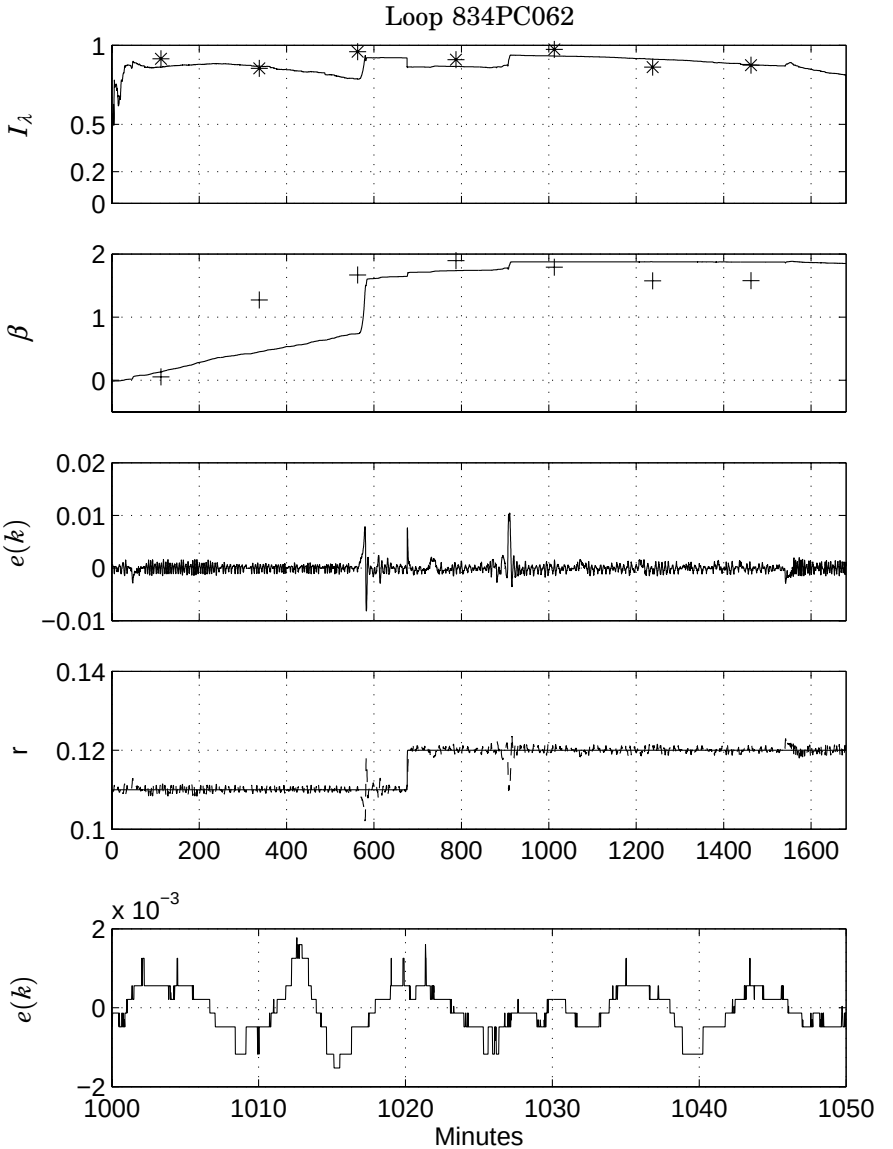

Figure A.1 Pressure loop with quantization.

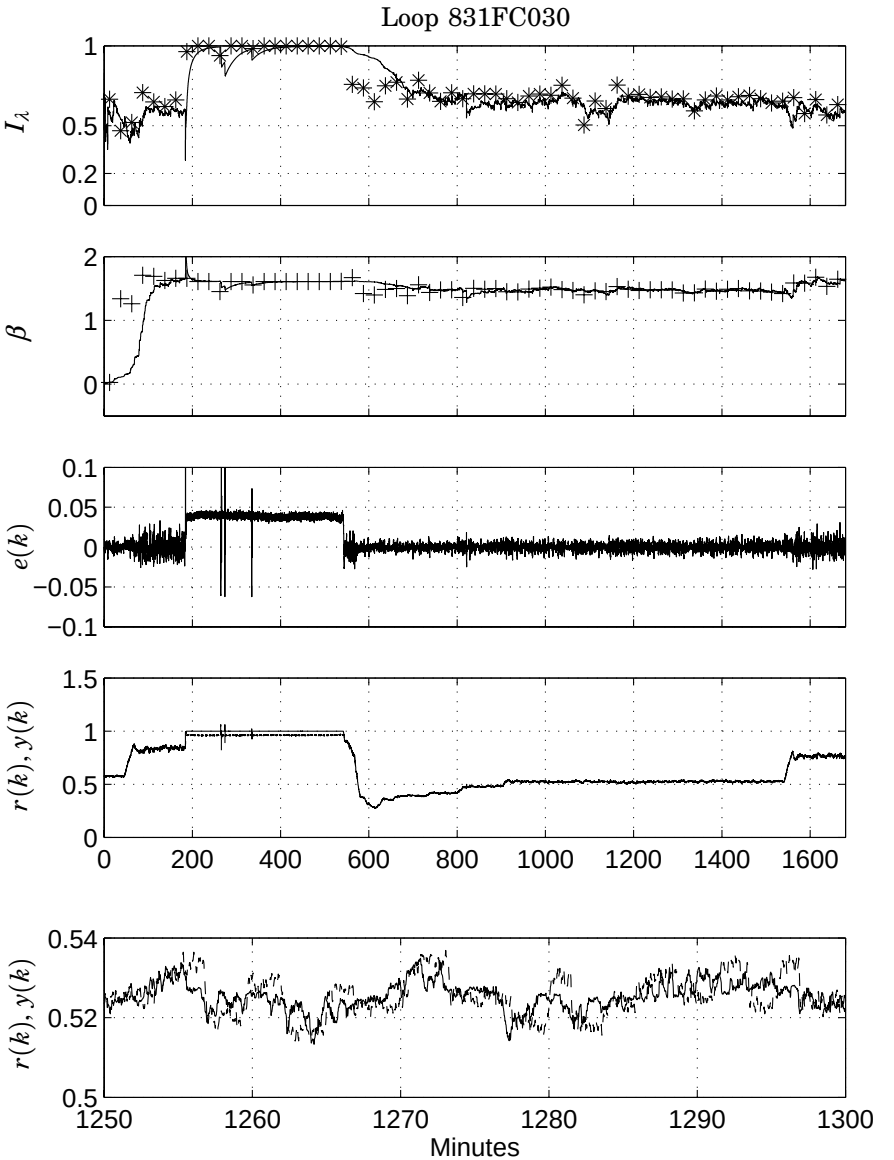

Figure A.2 Flow loop with stiction. The lowest graph shows the set point and output on an interval where the stiction square wave pattern was apparent.

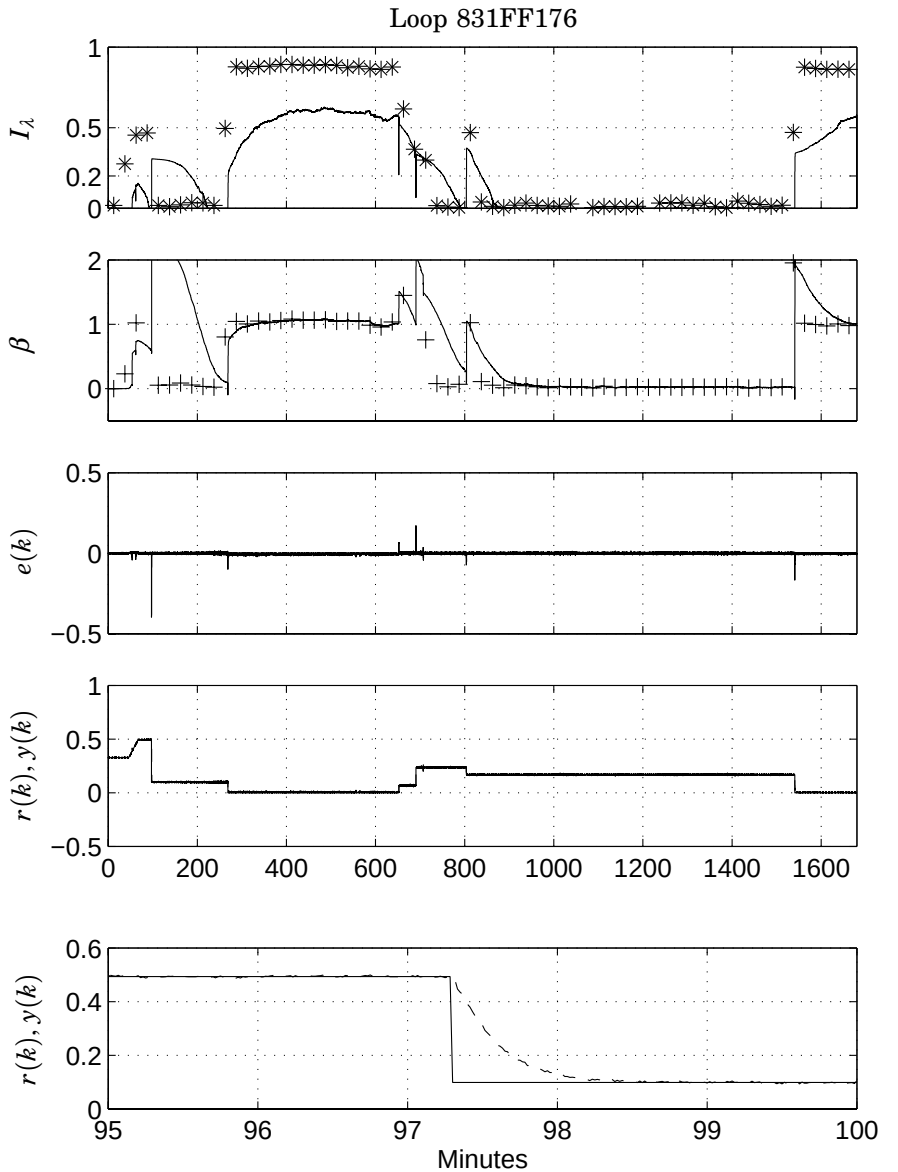

Figure A.3 Flow loop with good performance except when set point is changed. The input saturates at time 250. The rise in the indices at time 97 and 800 is due to steps in the set point where the response of $y(k)$ is not adequate. The lowest graph shows one set point response. The loop was tuned for $\lambda = 4.9$ sec and $L = 1.6$ sec.

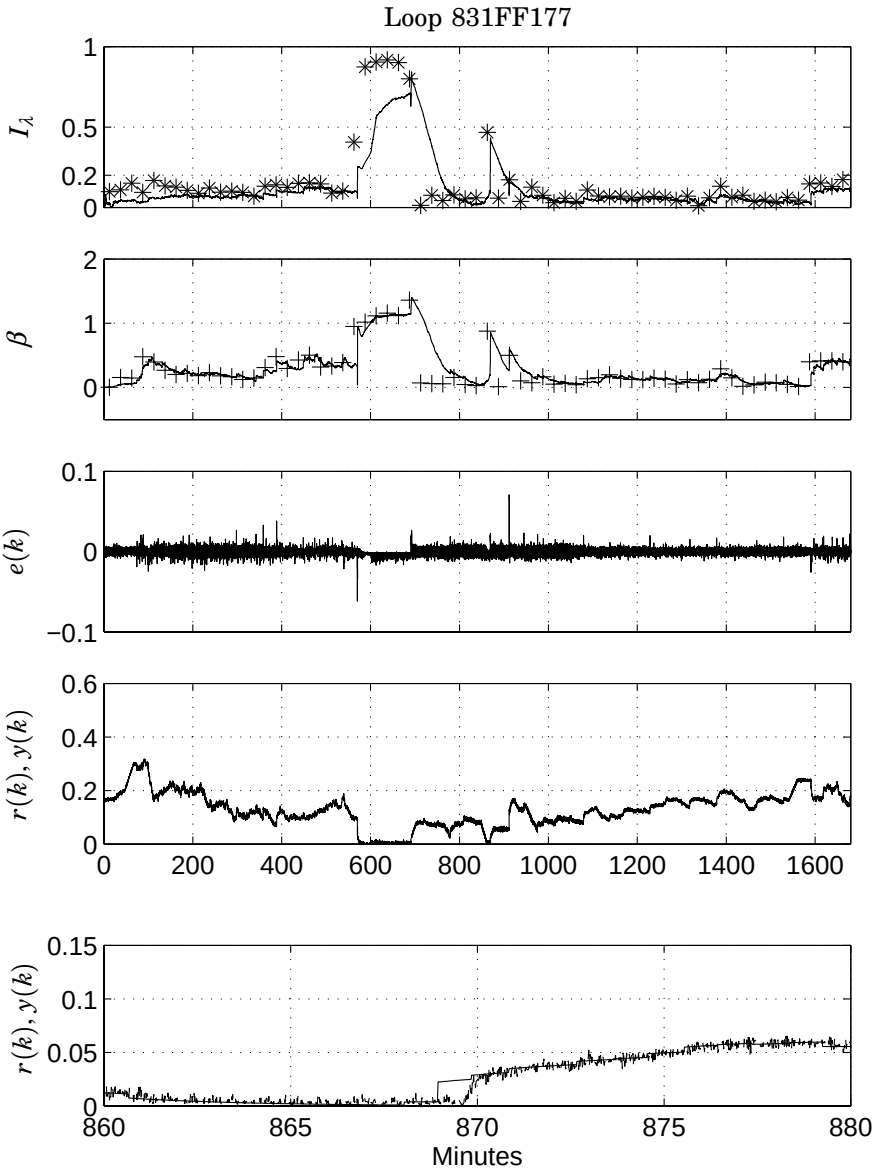

Figure A.4 Flow loop with good performance. The rise of the indices on the interval 560 to 750 minutes was due to saturation in the control signal. Notice that the output lies close to 0 for this interval.

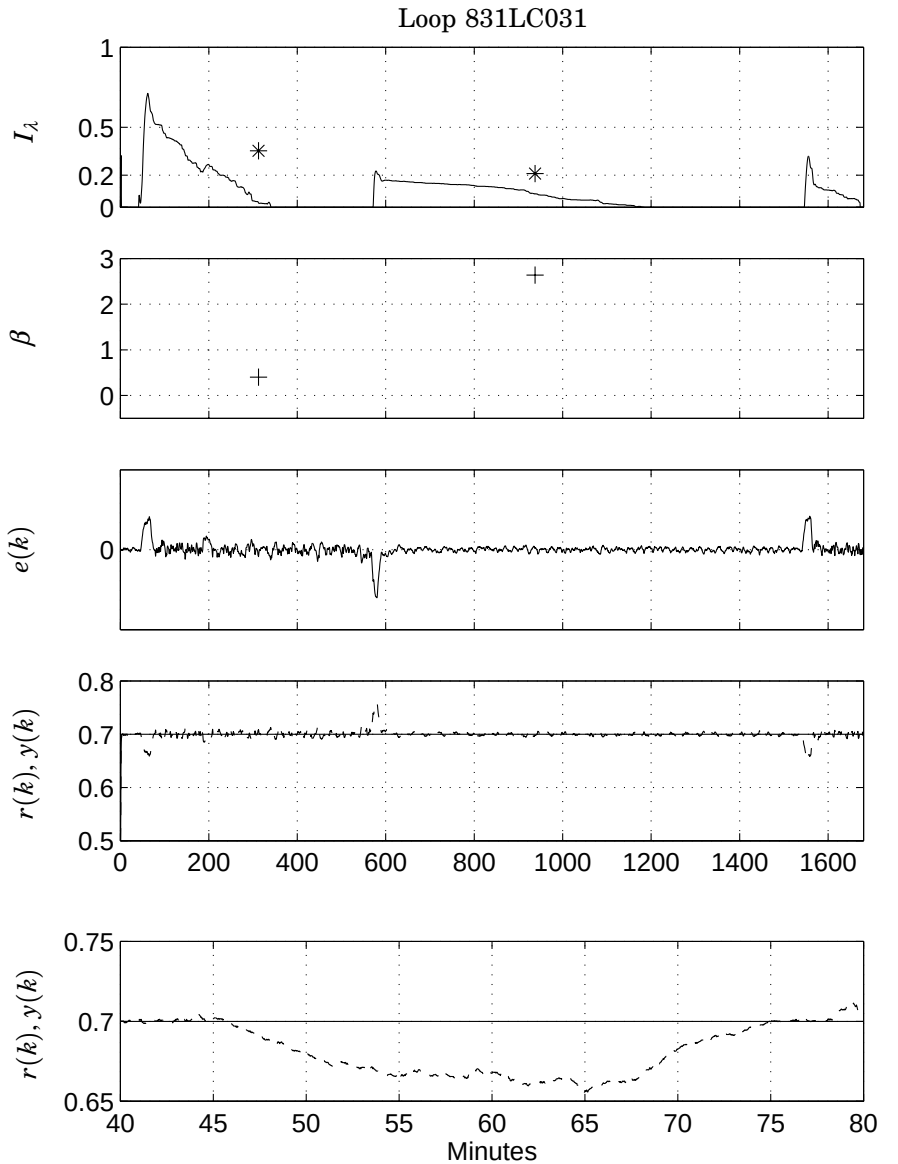

Figure A.5 Level control loop. Due to large λ (215 seconds) only two batch estimates were calculated. Load disturbance at time 45 minutes causes a rise in I_λ.

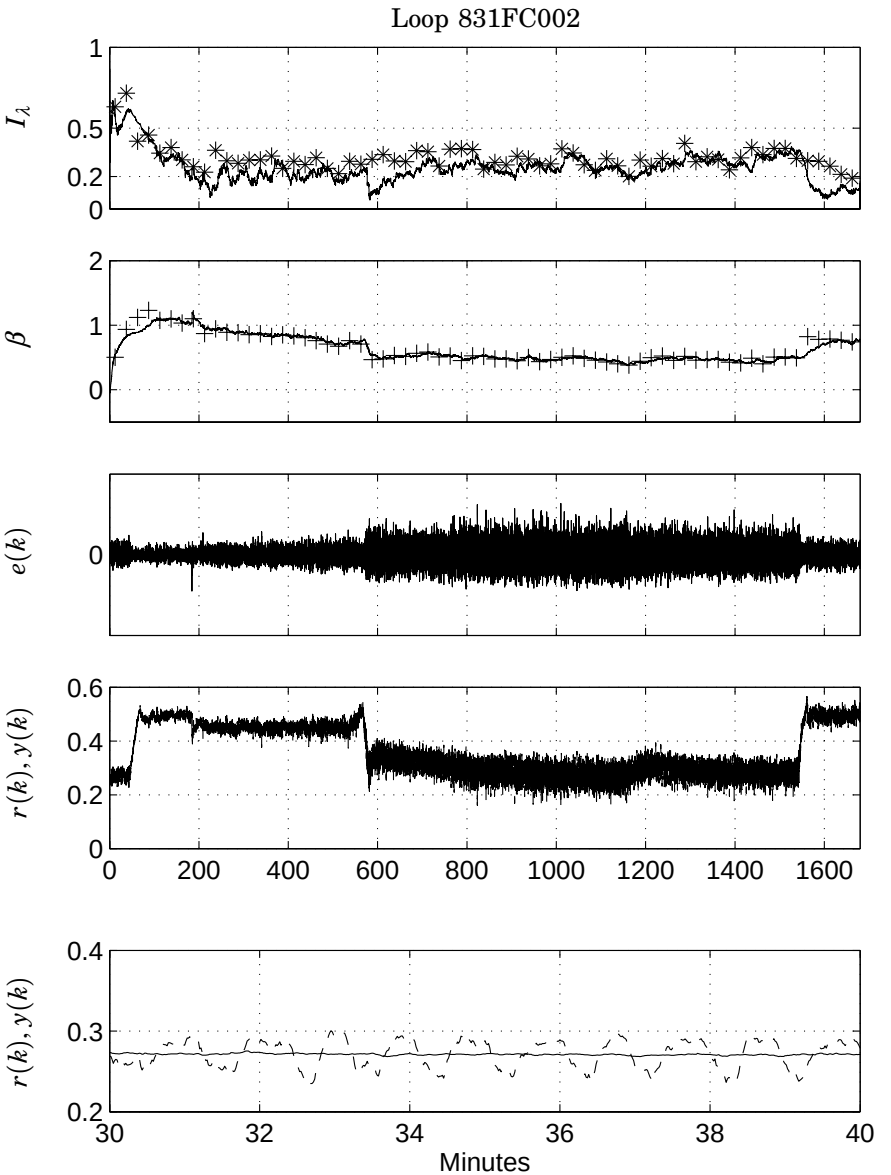

Figure A.6 Flow loop with stiction. The lowest graph shows the set point and output on an interval where the stiction square wave pattern was apparent.

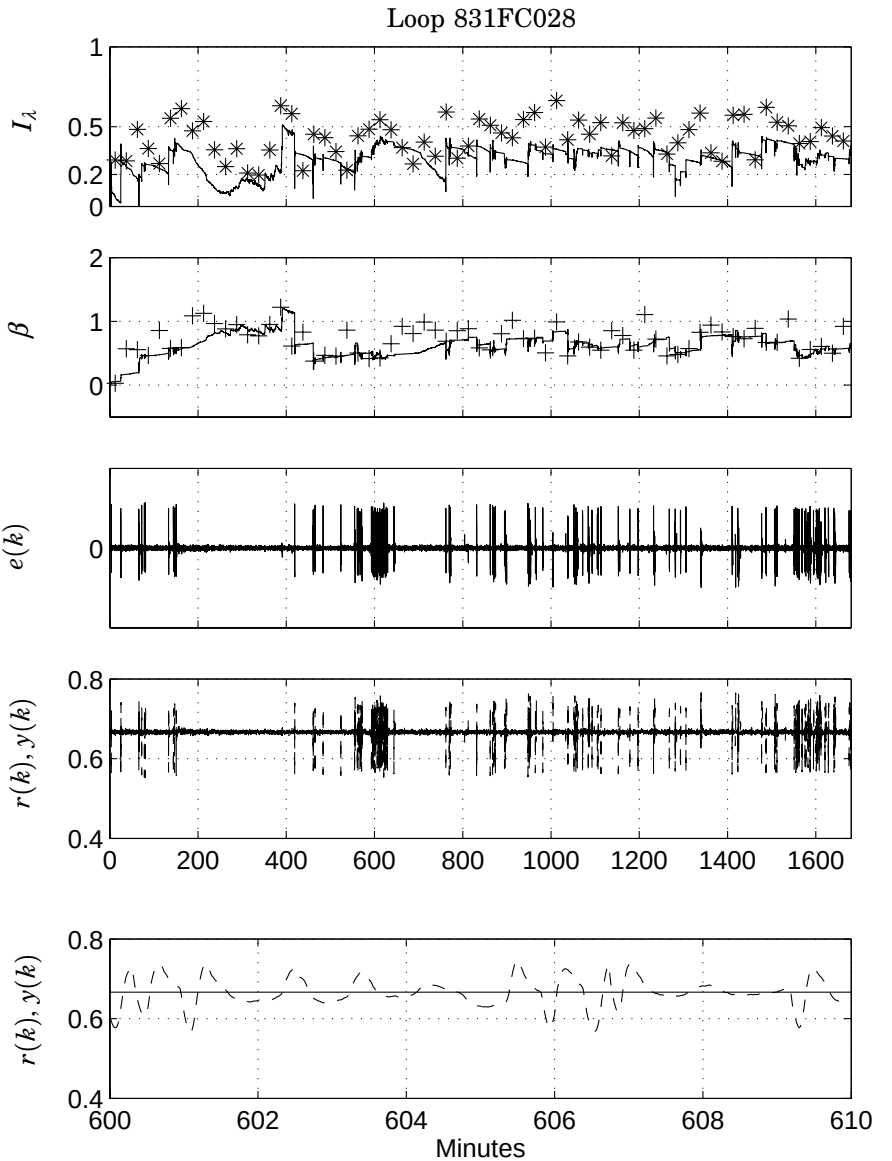

Figure A.7 Flow loop with poor performance. Notice the sporadic disturbances affecting the loop. Reason for poor performance unknown.

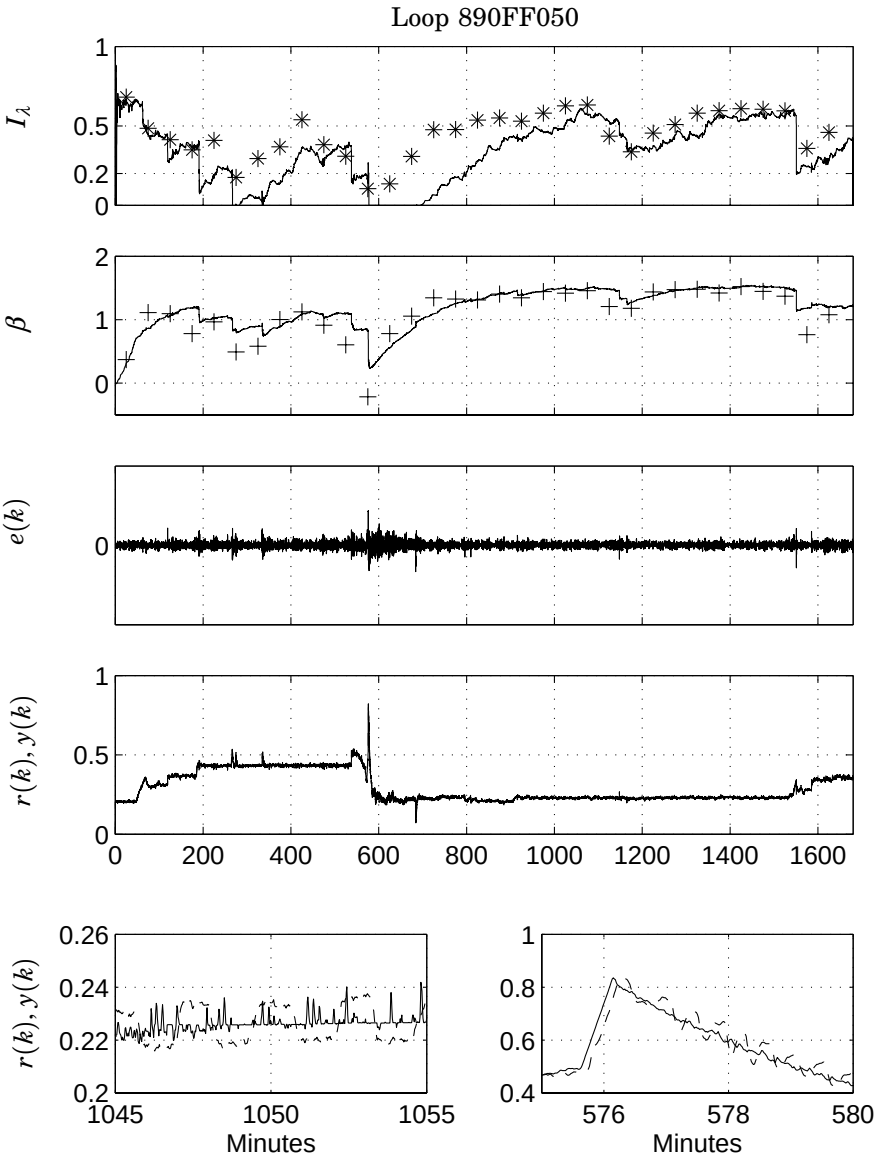

Figure A.8 Flow loop with stiction. The lowest graph to the left shows the set point and output on an interval where the stiction square wave pattern was apparent. Notice that the gradient dives at time 576. The graph to the right shows an interval where the output $y(k)$ displayed oscillatory behavior with the characteristics of instability.

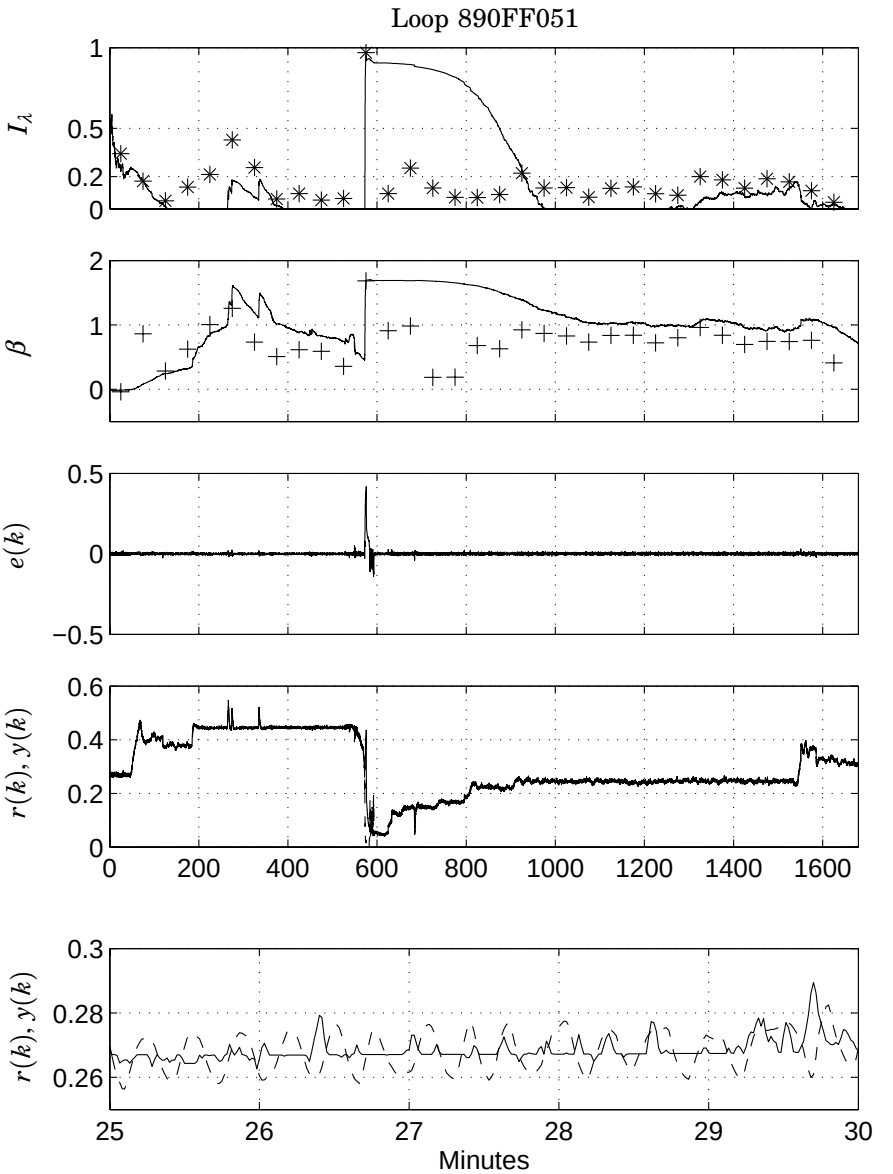

Figure A.9 Loop with good performance over most of the interval. In the beginning the loop displayed oscillatory behavior with the characteristics of instability. This is shown in the lowest graph. At time 600 the input saturated which caused the recursively estimated indices to become very large. The rise in I_λ around time 300 is due to load disturbances.

Bibliography

Åström, K. J. (1967): "Computer control of a paper machine—An application of linear stochastic control theory." *IBM Journal of Research and Development*, **11:4**, pp. 389–405.

Åström, K. J. (1970): *Introduction to Stochastic Control Theory*. Academic Press, New York.

Åström, K. J. and T. Hägglund (1995): *PID Controllers: Theory, Design, and Tuning*. Instrument Society of America, Research Triangle Park, North Carolina.

Åström, K. J. and T. Hägglund (2001): "The future of PID control." *Control Engineering Practice*, **9**, pp. 1163–1175.

Åström, K. J., C. C. Hang, and B. C. Lim (1994): "A new Smith predictor for controlling a process with an integrator and long dead-time." *IEEE Transactions on Automatic Control*, **39:2**.

Åström, K. J. and B. Wittenmark (1995): *Adaptive Control*. Addison-Wesley, Reading, Massachusetts.

Åström, K. J. and B. Wittenmark (1997): *Computer-Controlled Systems*. Prentice Hall.

Bialkowski, W. (1993): "Dreams versus reality: a view from both sides of the gap." *Pulp and Paper Canada*.

Bialkowski, W. (1998): "Mill audits can cut costs by reducing control loop variability." *Pulp & Paper*.

Bruyne, F. and P. Carrette (1997): "Synthetic generation of the gradient for an iterative controller optimization method." In *4th European Control Conference*.

Dahlin, E. B. (1968): "Designing and tuning digital controllers." *Instruments and Control Systems*, **42**, June, pp. 77–83.

Desborough, L. and T. Harris (1992): "Performance assessment measures for univariate feedback control." *The Canadian Journal of Chemical Engineering*, **70**, pp. 1186–1197.

Desbourough, L. and R. Miller (2001): "Increasing customer value of industrial control performance monitoring - honeywell's experience." In *Sixth International Conference on Chemical Process Control*, vol. 98 of *AIChE Symposium Series*.

Devries, W. and S. Wu (1978): "Evaluation of process control effectiveness and diagnosis of variation in paper basis weight via multivariate time-series analysis." *IEEE Transaction on Automatic Control*, **23:4**, pp. 702–708.

Doyle, J., B. Francis, and A. Tannenbaum (1992): *Feedback Control Theory*. Macmillan Publishing Company: New York.

Ender, D. (1993): "Process control performance: not as good as you think." *Control Engineering*, September.

Eriksson, P.-G. and A. Isaksson (1994): "Some aspects of control loop performance monitoring." In *3rd IEEE conference on Control Applications*. Glasgow, Scotland.

Gustafsson, F. and S. F. Graebes (1998): "Closed-loop performance monitoring in the presence of system changes and disturbances." *Automatica*, **34:11**, pp. 1311–1326.

Hägglund, T. (1995): "A control-loop performance monitor." *Control Engineering Practice*, **3**, pp. 1543–1551.

Hägglund, T. (1996): "An industrial dead-time compensating PI controller." *Control Engineering Practice*, **4**, pp. 749–756.

Hägglund, T. (1999): "Automatic detection of sluggish control loops." *Control Engineering Practice*, **7**, pp. 1505–1511.

Hägglund, T. (2002): "Industrial applications of automatic performance monitoring tools." In *IFAC World Congress*. Barcelona, Spain.

Harris, H. (1989): "Assessment of control loop performance." *The Canadian Journal of Chemical Engineering*, **67**, pp. 856–861.

Harris, T., F. Boudreau, and J. MacGregor (1996): "Performance assessment of multivariable feedback controllers." *Automatica*, **32**, pp. 1503–1517.

Harris, T., C. Seppala, and L. Desborough (1999): "A review of performance monitoring and assessment techniques for univariate and multivariate control systems." *Journal of Process Control*, **9**, pp. 1–17.

Harris, T., C. Seppala, P. Jofriet, and B. Surgenor (1996): "Plant-wide feedback control performance assessment using and expert-system framework." *Control Engineering Practice*, **4:9**, pp. 1297–1303.

Hjalmarsson, H., M. Gevers, S. Gunnarsson, and O. Lequin (1998): "Iterative feedback tuning: Theory and applications." *IEEE Control Systems*, August.

Hjalmarsson, H., S. Gunnarsson, and M. Gevers (1994): "A convergent iterative restricted complexity control design scheme." In *Proc. 33rd IEEE Conference on Decision and Control*, pp. 1735–1740. Orlando, Florida.

Horch, A. (2000): *Condition Monitoring of Control Loops*. PhD thesis, Department of Signals, Sensors and Systems, Royal Institute of Technology.

Horch, A. and A. Isaksson (1999): "A modified index for control performance assessment." *Journal of Process Control*, **9**, pp. 475–483.

Huang, B. (1999): "Performance assessment of processes with abrupt changes of disturbances." *The Chanadian Journal of Chemical Engineering*, **77**, pp. 1044–1054.

Huang, B. and S. L. Shah (1998): "Practical issues in multivariable feedback control performance assessment." *Journal of Process Control*, **8**, pp. 421–430.

Huang, B., S. L. Shah, and E. Kwok (1997): "Good, bad or optimal? performance assessment of multivariable processes." *Automatica*, **33**, pp. 1175–1183.

Ingimundarson, A. and T. Hägglund (2000a): "Closed-loop identification of first-order plus dead-time model with method of moments." In *ADCHEM 2000, IFAC International Symposium on Advanced Control of Chemical Processes*. Pisa, Italy.

Ingimundarson, A. and T. Hägglund (2000b): "Robust automatic tuning of an industrial PI controller for dead-time systems." In *IFAC Workshop on Digital Control – Past, present, and future of PID Control*. Terrassa, Spain.

Ingimundarson, A. and T. Hägglund (2001): "Robust tuning procedures for dead-time compensating controllers." *Control Engineering Practice*, **9**, pp. 1195–1208.

Isaksson, A. (1996): "PID controller performance assessment." In *Proc. Control Systems, Halifax, NS*.

Jofriet, P. and W. Bialkowski (1996): "Process knowledge: The key to on-line monitoring of process variability and control loop performance." In *Control Systems '96*. Halifax.

Kendar, S. J. and A. Cinar (1997): "Controller performance assessment by frequency domain techniques." *Journal of Process Control*, **7:3**, pp. 181–194.

Ko, B.-S. and T. F. Edgar (1998): "Assessment of achievable PI control performance for linear processes with dead time." In *Proc. American Control Conference, Philadelphia, Pennsylvania*.

Kozub, D. and C. Garcia (1993): "Monitoring and diagnosis of automated controllers in the chemical process industries." In *Proc. AIChE*. St.Louis.

Kozub, D. J. (1997): "Controller performance monitoring and diagnosis: Experiences and challenges." In *AIChE Symposium Series*. AIChE.

Kozub, D. J. (2002): "Controller performance monitoring and diagnosis. industrial perspective." In *Proceedings of the 15th IFAC World Congress, Barcelona, Spain*.

Landau, I., D. Rey, A. Karimi, A. Voda, and A. Franco (1995): "A flexible transmission system as a benchmark for robust digital control." *Proceedings of the Third European Control Conference, Rome, Italy*.

Laughlin, D., D. Rivera, and M. Morari (1987): "Smith predictor design for robust performance." *Int. J. Control*, **46**, pp. 477–504.

Lee, T., Q. Wang, and K. Tan (1995): "Automatic tuning of the Smith predictor controller." *Journal of Systems Engineering*, **5**.

Lee, T., Q. Wang, and K. Tan (1996): "Robust Smith-predictor controller for uncertain delay systems." *AIChE Journal*, **42**.

Ljung, L. (1987): *System Identification—Theory for the User*. Prentice Hall, Englewood Cliffs, New Jersey.

Ljung, L. (1999): *System Identification—Theory for the User, Second Edition*. Prentice Hall.

Majhi, S. and D. Atherton (2000): "Obtaining controller parameters for a new Smith predictor using autotuning." *Automatica*, **36**, pp. 1651–1658.

Marlin, T. (1995): *Process Control. Designing Processes and Control Systems for Dynamic Performance*. McGraw-Hill International Editions.

Matausek, M. and A. Micic (1996): "A modified smith predictor for controlling a process with an integrator and long dead-time." *IEEE Transaction on Automatic Control*, **41**, pp. 1199–1203.

Matausek, M. and A. Micic (1999): "On the modified smith predictor for controlling a process with an integrator and long dead-time." *IEEE Transaction on Automatic Control*, **44**, pp. 1603–1606.

Meinsma, G. and H. Zwart (2000): "On $\mathcal{H}_\infty$ control for dead-time systems." *IEEE Transaction on Automatic Control*, **45**.

Meyer, C., D. E. Seborg, and R. K. Wood (1976): "A comparison of the Smith predictor and conventional feedback control." *Chemical Engineering Science*.

Morari, M. and E. Zafiriou (1989): *Robust Process Control*. Prentice-Hall, Englewood Cliffs, New Jersey.

Normey-Rico, J., C. Bordons, and E. Camacho (1997): "Improving the robustness of dead-time compensating PI controllers." *Control Engineering Practice*, **5**, pp. 801–810.

Normey-Rico, J. and E. Camacho (1999): "Robust tuning of dead-time compensators for processes with an integrator and long dead-time." *IEEE Transaction on Automatic Control*, **44**, pp. 1597–1603.

Normey-Rico, J. and E. Camacho (2002): "A unified approach to design dead-time compensators for stable and integrative processes with dead-time." *IEEE Transaction on Automatic Control*, **47**, pp. 299–305.

Palmor, Z. (1980): "Stability properties of smith dead-time compensator controllers." *Int. J. Control*, **32**, pp. 937–949.

Palmor, Z. and M. Blau (1994): "An auto-tuner for smith dead time compensator." *Int. J. Control*, **60**, pp. 117–135.

Panagopoulos, H. and K. J. Åström (2000): "PID control design and H_∞ loop shaping." *Int. J. Robust Nonlinear Control*, **10**, pp. 1249–1261.

Paulonis, M. A. and J. W. Cox (2003): "A practical approach for large-scale controller performance assessment, diagnosis and improvement." *Journal of Process Control*, **13**, pp. 155–168.

Qin, S. (1998): "Control performance monitoring - a review and assessment." *Computers and Chemical Engineering*, **23**, pp. 173–186.

Rivera, D. E., M. Morari, and S. Skogestad (1986): "Internal model control—4. PID controller design." *Ind. Eng. Chem. Proc. Des. Dev.*, **25**, pp. 252–265.

Seborg, D. E., T. F. Edgar, and D. A. Mellichamp (1989): *Process Dynamics and Control*. Wiley, New York.

Seppala, C., T. Harris, and D. Bacon (2002): "Time series methods for dynamic analysis of multiple controlled variables." *Journal of Process Control*, **12**, pp. 257–276.

Skogsindustriernas Teknik AB (1997): *Regleroptimering*, ssg 5253 edition.

Smith, O. J. M. (1957): "Closed control of loops with dead time." *Chemical Engineering Progress*, **53**, May, pp. 217–219.

Stanfelj, N., T. E. Marlin, and J. F. MacGregor (1993): "Monitoring and diagnosing process-control performance - the single-loop case." *Industrial & Engineering Chemistry Research*, **32:2**, pp. 301–314.

Thornhill, N., M. Oettinger, and P. Fedenczuk (1999): "Refinery-wide control loop performance assessment." *Journal of Process Control*, **9**, pp. 109–124.

Thornhill, N. F., S. Shah, and B. Huang (2001): "Detection of distributed oscillations and root cause diagnosis." In *Preprints of CHEMFAS-4, IFAC,Korea*.

Thornhill, N. F., C. Xia, J. Howell, J. Cox, and M. Paulonis (2002): "Analysis of plant-wide disturbances through data-driven techniques and process understanding." In *Proceedings of the 15th IFAC World Congress, Barcelona, Spain*.

Tyler, M. and M. Morari (1996): "Performance monitoring of control systems using likelihood methods." *Automatica*, **32**, p. 1145.

Vaught, R. and J. Tippet (2001): "Control performance monitoring: shaman or saviour." *Pulp & Paper Canada*, **102:9**, pp. 26–29.

Vrancić, D., D. Vrecko, D. Juricić, and S. Strmcnik (1999): "Automatic tuning of the flexible Smith predictor controller." *Proceedings of the American Control Conference, San Diego, California*.

Watanabe, K. and M. Ito (1981): "A process-model control for linear systems with delay." *IEEE Transaction on Automatic Control*, **26**.

Zhou, K. (1998): *Essentials of Robust Control*. Prentice Hall: New Jersey.