PID Controller Design

Panagopoulos, Helene

1998

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

Link to publication

Citation for published version (APA):
Panagopoulos, H. (1998). *PID Controller Design*. [Licentiate Thesis, Department of Automatic Control]. Department of Automatic Control, Lund Institute of Technology (LTH).

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

PID Controller Design

Hélène Panagopoulos

Department of Automatic Control
Lund Institute of Technology

December 1998

To Mom, Dad, and Elias

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

ISSN 0280–5316
ISRN LUTFD2/TFRT–3224–SE

Printed in Sweden by Reprocentralen, Lunds Universitet
Lund 1999

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

Document name
LICENTIATE THESIS

Date of issue
December 1998

Document Number
ISRN LUTFD2/TFRT--3224--SE

Author(s)
Hélène Panagopoulos

Supervisor
Karl Johan Åström and Tore Hägglund

Sponsoring organisation

Title and subtitle
PID Controller Design

Abstract

The PID controller is the most commonly used control algorithm today. In spite of its wide-spread use there exist no generally accepted design method. There are several reasons to look for better methods to design PID controllers. One reason is the significant impact it may give because of the wide-spread use of the controllers. Another reason is the significant benefit improved design methods will give emerging auto tuners and tuning devices.

This thesis presents two new design methods: one for PI controllers and the other one for PID controllers. Also, a number of considerations in PID controller design is discussed.

To begin with the two design methods are presented. The specifications capture demands on load disturbances rejection, set point response, measurement noise, and model uncertainty. The primary design goal is to obtain good load disturbance responses. This is done by minimizing the integrated error, IE. Robustness is guaranteed by requiring that the maximum sensitivity is less than a specified value. Good set point response is obtained by using a structure with two degrees of freedom. Measurement noise is dealt with by filtering. The design procedures have been applied to a variety of systems.

The thesis ends by showing how the the two design methods for PI and PID control establish a nice connection with $\mathcal{H}_\infty$ control.

Key words
PID control. Design. Optimization. Specifications. Load disturbance rejection. Measurement noise filtering. Set point response. Robustness. Sensitivity. $\mathcal{H}_\infty$ control.

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title

ISBN

Language
English

Number of pages
82

Security classification

Recipient's notes

The report may be ordered from the Department of Automatic Control or borrowed through:
University Library 2, Box 3, S-221 00 Lund, Sweden
Fax +46 46 222 44 22 E-mail ub2@ub2.lu.se

Contents

Acknowledgment

Since I started my research in design of PID controllers in 1995, I have had the pleasure, and privilege, to work with Karl Johan Åström, professor, and Tore Hägglund, associate professor, who have been my supervisors. They have contributed with valuable ideas, suggestions, and examples to accomplish this thesis. Furthermore, I would like to thank colleagues, family, and friends for all help and encouragement they have given me during this period of research.

1

Introduction

This thesis deals with design of proportional integral derivative (PID) controllers. PID control is by far the most common control algorithm and very much has been written about it, see e.g. Åström and Hägglund (1995b). In spite of this there are no uniformly accepted design methods. There are several reasons to look for better methods to design PID controllers. One reason is the significant impact it may give because of the widespread use of the controllers. In a plant with several hundred subprocesses and control loops there is no way to model each of them and custom design controllers for each one. There is a need for a single controller structure with few parameters to tune. Another reason is that auto tuners and tuning tools can benefit significantly from improved design methods.

In the thesis new design methods for PID controllers will be presented and a number of considerations in PID controller design will also be discussed.

The thesis consists of the following three papers:

I. ÅSTRÖM, K.J., H. PANAGOPOULOS, AND T. HÄGGLUND (1998): "Design of PI controllers based on non-convex optimization." *Automatica,* **35:5**.

II. PANAGOPOULOS, H., K. J. ÅSTRÖM, AND T. HÄGGLUND (1998): "Design of PID controllers based on constrained optimization." Department of Automatic Control, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden.

III. PANAGOPOULOS, H., AND K. J. ÅSTRÖM (1998): "PID controller design and $\mathcal{H}_\infty$ loop shaping." Department of Automatic Control, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden.

1.1 A Perspective on Control Design

Development of design methods has been a goal for control theory for a long time. The classical frequency domain methods were developed in the thirties and forties starting with the breakthrough in stability theory made by Nyquist and continuing with the work on feedback amplifiers by Black and Bode.

In the classical approach the main concern was to design feedback compensators such that a certain stability margin was achieved, see e.g. Truxal (1955). The emphasis was then on model uncertainties. Feedback was used to decrease sensitivity to disturbances and model errors. The compensator design was done mainly by graphical methods evolving from the Nyquist stability criterion.

The development of analytical design methods in the fifties, see e.g. Newton *et al.* (1957), made it possible to give specifications on the transient performance by giving a process model together with a closed loop specification. At the same time, less attention was given to robustness and sensitivity issues. In the last two decades analytical methods have been developed in which robustness has regained its importance.

In the sixties the development of methods for control design based on optimization techniques had the advantage to capture many different aspects of the design problem. During this time efficient computer methods were developed to solve these optimal control problems. A general discussion of the use of optimization for control design is found in Boyd and Barratt (1991) and Mayne and Polak (1993).

What about PID Controller Design?

Despite the development in control theory the PID controllers are the most used controllers in industry, see, for instance, Yamamoto and Hashimoto (1991). There are a number of reasons for this: To begin with PID controllers are well understood by industrial operational, technical, and maintenance individuals. Secondly, they have a long history of proven operation. As a matter of fact, in many applications a properly designed and well tuned PID controller meets or exceeds the control objectives. Finally, there are many extensions which make an industrial PID controller practical for operating a process. For example, it has automatic and manual switching, set point tracking, and emergency manual modes.

In spite of the wide spread use of PID controllers there is a lack of a universally accepted tuning method. Many design methods for PID controllers have been proposed in the literature, see e.g. Åström and Hägglund (1995b) and references therein. Finding design methods which lead to the optimal operation of PID controllers is therefore of significant interest. One reason is the significant impact it may give because of the widespread use of the controllers. Another reason is that emerging auto tuners and tuning devices can benefit significantly from improved design methods, see e.g. Van Overshee *et al.* (1997).

The simple tuning rules enjoy considerable success because of their computational simplicity and the moderate process knowledge required. The simplicity is, however, obtained at the expense of restrictive conditions on the plant and the lack of tuning parameters. A tuning method as Ziegler Nichols is simple and widely used, but it is not applicable to a wide range of systems, where it is necessary to have more information. Consequently, more elaborate methods which are based on process modeling, formulation of specifications, and control design gives better performance for a wide range of systems. In these cases optimization is a powerful tool.

1.2 Automatic Tuning of PID Controllers

The original scheme for auto tuning in Åström and Hägglund (1984) used a very simple design method based on a modified Ziegler-Nichols method. An improved method, dominant pole design, was proposed in Hägglund and Åström (1985) and further developed in Persson (1992). This method is based on knowledge of the process transfer function. The relative damping of the dominant poles is specified and their distance to the origin is chosen to optimize integral gain. A drawback with the method is that it is difficult to find a good universal choice

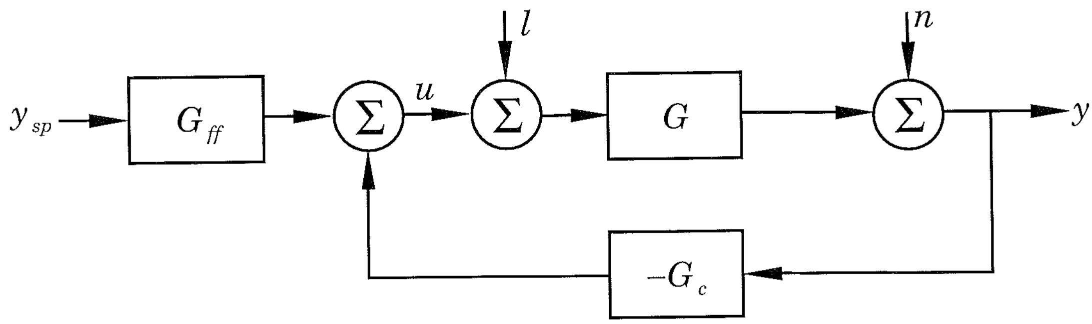


Figure 1.1 Block diagram describing the design problem.

of the relative damping. A search method which sweeps over the relative damping to obtain a given value of M_s was proposed in Persson (1992) where extension to PID control were also given. A commercial product PIDCona, used in the ABB Master distributed control system, based on a discrete time version of dominant pole design is described in Modén (1995). The main drawback of the dominant pole design is that we must either specify the relative damping or resort to elaborate sweeping methods. It is difficult to obtain reliable numerical methods for the sweeping. The present work can be viewed as a continuation that is based on a direct solution of an optimization problem.

1.3 Contribution

There are several requirements on an efficient design method. It should be applicable to a wide range of systems and it should have the possibility to introduce specifications that capture the essence of real control problems. The method should also give all parameters of the PID controller including set point weight and filters for set point and measurement noise if such filters are desired. Furthermore, the method should be robust in the sense that it provides controller parameters if they exist, or if the specifications can not be met an appropriate diagnosis should be presented. These requirements are satisfied by the design methods for PI and PID controllers presented in the thesis.

First a general formulation of the design problem is given before the two design methods for the PI and PID controller are presented. Consider the design problem illustrated in Figure 1.1. A process with transfer function $G(s)$ is controlled with a PID controller with two degrees of freedom. Transfer function $G_c(s)$ describes the feedback from process output y to control signal u, and $G_{ff}(s)$ describes the feed forward from set point y_{sp} to u. Three external signals act on the control loop, namely set point y_{sp}, load disturbance l and measurement noise n.

The design objective is to determine the controller parameters in $G_c(s)$ and $G_{ff}(s)$ so that the system behaves well with respect to changes in the three signals y_{sp}, l and n as well as in the process model $G(s)$. Therefore, the specification will express requirements on

- Load disturbance response
- Robustness with respect to model uncertainties
- Measurement noise response
- Set point response

Consequently, the formulation of the design problem takes care of rejection of load disturbances and measurement noise at the same time as it gives good set point response. Note that, even if it has been emphasized many times by practitioners such as Shinskey (1988) and Shinskey (1990), a lot of papers still focus primarily on the set point response only. Also, the formulation takes into account the sensitivity of model uncertainties, which was one of the major drawbacks of the classical Ziegler Nichols method.

In virtually all applications it is useful to have a tuning parameter that permits adjustment of the trade off of aggressiveness versus robustness. The effects of the tuning parameter should be transparent to the user. The design problems to be presented have a tuning parameter to specify the sensitivity to model uncertainties. A nice feature of the methods is that good default values of the tuning parameter can be found. Therefore, the user can simply supply the process transfer function and the method gives the parameters of the controller. Modifying the tuning parameters give a flexible way to change the main characteristics of the system.

In the formulation of the design problem it is expected that the process is given and the properties of the input signals of the system. The process is assumed to be linear, time invariant, and specified by the transfer function $G(s)$ which is analytic with finite poles and possibly an essential singularity at infinity. The description covers finite dimensional systems with time delays and infinite dimensional systems described by linear partial differential equations. Furthermore, this is very useful as it admits design of controllers for a very wide range of systems.

To begin with a short resumée of the paper Åström *et al.* (1998) is given.

Design of PI Controllers based on Non-Convex Optimization

This paper presents an efficient numerical method for designing PI controllers. The specifications capture demands on load disturbance rejection, set point response, measurement noise and model uncertainty. The primary design goal is to obtain good load disturbance responses. This is done by minimizing the integrated control error IE. Robustness is guaranteed by requiring that the maximum sensitivity is less than a specified value M_s. Good set point response is obtained by using a structure with two degrees of freedom. This requires an extra parameter, the set point weighting b, in the algorithm. The primary design parameter is the maximum sensitivity, M_s, but auxiliary design parameters such as the maximum of the complementary sensitivity, M_p, can be added. Thus, the formulation of the design problem captures three essential aspects of industrial control problems, leading to a non-convex optimization problem.

The proposed method formulates the design problem as a constrained optimization problem: optimize the load disturbance rejection with a constraint on the maximum sensitivity. By exploiting the structure of the optimization problem it is reduced to the solution of nonlinear algebraic equations. Efficient iterative methods are given together with good methods for finding initial values.

There are unique solutions for special classes of systems but very complicated situations may occur for complicated systems. The method will give a solution if one exists and it will indicate when there is no PI controller that satisfies the specifications.

The design procedure has been applied to a variety of systems; stable and integrating, with short and long dead times, with real and complex poles, and with positive and negative zeros.

Parts of the proposed method builds on previous works. The use of optimization was discussed in Hazebroek and van der Waerden (1950). In this and in other early works the emphasis was on criteria that admitted analytical solutions. The idea of optimizing load disturbance rejection with sensitivity constraints was suggested by Shinskey (1990). He used a constraint in terms of a rectangle around the critical point but the idea to use a constraint on the maximum sensitivity, M_s, was proposed by Persson (1992) and Persson and Åström (1992). The use of both M_s and the maximum complementary sensitivity, M_p, as design parameters was suggested by Schei (1994).

The new contributions in this paper are: the analysis of the nature of the sensitivity constraints, the efficient numerical procedures, and the method for determining set point weighting.

A natural step would be to extend the work on designing PI controllers presented in Åström *et al.* (1998) to the design of PID controllers. The outcome of this idea is presented in the paper Panagopoulos *et al.* (1998), for which a short resumée follows.

Design of PID Controllers based on Constrained Optimization

This paper describes a new design method for PID controllers. The PID controller to be discussed has four primary parameters, gain k, integral time T_i, derivative time T_d and set point weight b. In addition there is filtering of the measured signal and sometimes also of the set point. The design method gives all parameters and filters required.

A design method for PI controllers based on maximization of integral gain subject to a robustness constraint was developed in Åström *et al.* (1998). In the present paper it is shown that this method cannot be extended directly to PID control. The reason for this is that the optimization problem in most cases has ridges which result in poor robustness and thus also poor control. Having developed an understanding for the problem it is possible to introduce additional constraints. The result is that design of PID controllers can be formulated as a constrained optimization problem which can be solved iteratively. Initial conditions are very important since the problem is non convex. A good way to find initial conditions is also presented.

The solution of the optimization problem gives a PID controller with a pure derivative. Simple rules for choosing a filter for the measured signal are then presented. Adjusted controller parameters are obtained simply by repeating the design with the process replaced by the combination of the original process transfer function and the transfer function of the chosen filter. This gives a very flexible way of choosing the filter.

Having obtained a controller able to deal with load disturbances, measurement noise and plant uncertainty it is then designed for good set point response. First it is attempted to determine the set point weighting so that the maximum of the transfer function from the set point to the output is less than a given value. Sometimes this cannot be accomplished and a filter for the set point is then designed. The advantage of this approach is that a filter is only introduced if it is necessary.

The design procedure has been applied to a variety of systems: stable and integrating, with long dead times and with right half plane zeros.

The new contributions of this paper are: the analysis of the nature of the derivative part of the PID controller, the additional constraints needed to solve the design problem, the methods for determining filters for measurement noise and set point weighting.

All the robustness constraints in Åström *et al.* (1998) and Panagopoulos *et al.* (1998) can be encapsulated in a constraint that the loop transfer function avoids a circle in the Nyquist diagram. This establishes also a nice connection between traditional design of PID controllers and $\mathcal{H}_\infty$ control, which is described in Panagopoulos and Åström (1998). A short resumée of the paper is given next.

PID Control Design and $\mathcal{H}_\infty$ Loop Shaping

The paper shows how the specifications for the PID design in Åström *et al.* (1998) and Panagopoulos *et al.* (1998) should be chosen to guarantee that the weighted $\mathcal{H}_\infty$ norm, see Glover and McFarlane (1989), of the transfer function from load and measurement disturbances to process inputs and outputs is less than a specified value γ. Also, a new way to determine for which class of systems a PID controller will be stabilizing is presented.

Many different methods have been proposed to design PID controllers. In Glover and McFarlane (1989) and Vinnicombe (1998) a loop shaping method was developed and in Åström *et al.* (1998) and Panagopoulos *et al.* (1998) a method for designing PID controllers were presented. In this paper it is shown that these two methods are closely related to the $\mathcal{H}_\infty$ loop shaping method developed in Glover and McFarlane (1989). In particular it is shown how the specifications for the

PID design should be chosen to guarantee that the weighted $\mathcal{H}_\infty$ norm of the transfer function from load disturbances to process inputs and outputs is less than a specified value γ.

Also, a new way to determine for which class of systems a PID controller will be stabilizing is presented. By ensuring that a PID design is such that $\gamma \leq \gamma_0$ a good guarantee for robustness of the closed loop system is given.

The new contributions of this paper are: If the robustness measure of the $\mathcal{H}_\infty$ design γ is accepted as a good performance measure the results of this paper suggest that the robustness constraint of PID controllers should be chosen as the combined M_sM_p-circle. In this way it is guaranteed that a design will automatically satisfy the $\mathcal{H}_\infty$ robustness constraint.

1.4 Future Work

The method for design of PI control is very straightforward, maximize the integral gain, k_i, subject to a robustness constraint. The work shows clearly the importance of using a robustness constraint. Analysis of PID control shows that maximization of k_i subject to the robustness constraint does not necessarily result in good controllers. Additional constraints are required in this case. In our work we have introduced these in turns of constraints on the shape of the Nyquist curve. It may be worth while to consider some alternatives.

Two design methods have been presented, one for PI control and one for PID control. A comparison between the results of PI and PID control shows that in some cases derivative action gives substantial improvements but in other cases the improvements are very marginal. It would be interesting to explore this further and to try to characterize the different cases.

The design method require that the transfer function of the process is known. Results for particular examples show however that it is sufficient to know the transfer function in certain frequency ranges. It would be interesting to explore this further and develop some design rules. In Åström and Hägglund (1995a), it was, e.g., shown that good results could be obtained using process models with only three parameters.

The design methods are based on non-convex optimization. Local optimization may occur in such problems. This has not been encountered in the particular examples. It would be interesting to investigate it. Conditions for uniqueness can be found for particular classes of systems.

Finally, it would be useful to package the design procedures so that a user-friendly interface is obtained.

References

ÅSTRÖM, K. J. and T. HÄGGLUND (1984): "Automatic tuning of simple regulators with specifications on phase and amplitude margins." *Automatica*, **20**, pp. 645–651.

ÅSTRÖM, K. J. and T. HÄGGLUND (1995a): "New tuning methods for PID controllers." In *European Control Conference*, pp. 2456–2462. Rome, Italy.

ÅSTRÖM, K. J. and T. HÄGGLUND (1995b): *PID Controllers: Theory, Design, and Tuning*, second edition. Instrument Society of America, Research Triangle Park, NC.

ÅSTRÖM, K. J., H. PANAGOPOULOS, and T. HÄGGLUND (1998): "Design of PI controllers based on non-convex optimization." *Automatica*, **35:5**.

BOYD, S. P. and C. H. BARRATT (1991): *Linear Controller Design – Limits of Performance.* Prentice Hall Inc., Englewood Cliffs, New Jersey.

GLOVER, K. and D. MCFARLANE (1989): "Robust stabilization of normalized coprime factor plant descriptions with H_∞-bounded uncertainty." *IEEE Transactions on Automatic Control*, **34:8**, pp. 821–830.

HÄGGLUND, T. and K. J. ÅSTRÖM (1985): "Automatic tuning of PID controllers based on dominant pole design." In *Proceedings of the IFAC Conference on Adaptive Control of Chemical Processes.* Frankfurt, Germany.

HAZEBROEK, P. and B. L. VAN DER WAERDEN (1950): "Theoretical considerations on the optimum adjustment of regulators." *Trans. ASME*, **72**, pp. 309–322.

MAYNE, D. Q. and E. POLAK (1993): "Optimization based design and control." In *Preprints IFAC 12th World Congress*, vol. 3, pp. 129–138. Sydney, Australia.

MODÉN, P. E. (1995): "Advanced PID autotuning, easy to use." In *Preprints European Control Conference, ECC '95*, pp. 2483–2487. Rome, Italy.

NEWTON, JR, G. C., L. A. GOULD, and J. F. KAISER (1957): *Analytical Design of Linear Feedback Controls.* John Wiley & Sons.

PANAGOPOULOS, H. and K. J. ÅSTRÖM (1998): "PID control design and $\mathcal{H}_\infty$ loop shaping." Department of Automatic Control, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden.

PANAGOPOULOS, H., K. J. ÅSTRÖM, and T. HÄGGLUND (1998): "Design of PID controllers based on constrained optimization." Department of Automatic Control, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden.

PERSSON, P. (1992): *Towards Autonomous PID Control.* PhD thesis ISRN LUTFD2/TFRT--1037--SE, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

PERSSON, P. and K. J. ÅSTRÖM (1992): "Dominant pole design—A unified view of PID controller tuning." In *Preprints 4th IFAC Symposium on Adaptive Systems in Control and Signal Processing*, pp. 127–132. Grenoble, France.

SCHEI, T. S. (1994): "Automatic tuning of PID controllers based on transfer function estimation." *Automatica*, **30:12**, pp. 1983–1989.

SHINSKEY, F. G. (1988): *Process-Control Systems. Application, Design, and Tuning*, third edition. McGraw-Hill, New York.

SHINSKEY, F. G. (1990): "How good are our controllers in absolute performance and robustness?" *Measurement and Control*, **23**, May, pp. 114–121.

TRUXAL, J. (1955): *Automatic Feedback Control System Synthesis.* McGraw-Hill, New York.

VAN OVERSHEE, P., C. MOONS, W. VAN BREMPT, P. VANVUCHELEN, and B. DE MOOR (1997): "RaPID: The end of heuristic PID tuning." Paper. Intelligent System and Modeling and Control nv, Kardinaal Mercierlaan 94, 3000 Leuven, Belgium.

VINNICOMBE, G. (1998): *Uncertainty and Feedback. H_∞ Loop-Shaping and the ν-Gap Metric.* To be published.

YAMAMOTO, S. and I. HASHIMOTO (1991): "Present status and future needs: The view from Japanese industry." In ARKUN AND RAY, Eds., *Chemical Process Control—CPCIV.* Proceedings of the Fourth Internation Conference on Chemical Process Control, Texas.

Paper I

Pergamon

PII: S0005-1098(98)00011-9

Automatica, Vol. 34, No. 5, pp. 585–601, 1998

Printed in Great Britain
0005-1098/98 $19.00 + 0.00

Design of PI Controllers based on Non-Convex Optimization*

K. J. ÅSTRÖM†, H. PANAGOPOULOS† and T. HÄGGLUND†

Key Words—PI control; design; optimization; specifications; load disturbance rejection; set point response; robustness; sensitivity.

Abstract—This paper presents an efficient numerical method for designing PI controllers. The design is based on optimization of load disturbance rejection with constraints on sensitivity and weighting of set point response. Thus, the formulation of the design problem captures three essential aspects of industrial control problems, leading to a non-convex optimization problem. Efficient ways to solve the problem are presented. © 1998 Elsevier Science Ltd. All rights reserved.

1. INTRODUCTION

The PI controller is unquestionably the most commonly used control algorithm; see Yamamoto and Hashimoto (1991). In spite of its wide spread use there exists no generally accepted design method for the controller. PI controllers have traditionally been tuned empirically, e.g. by the method described in Ziegler and Nichols (1942). This method has the great advantage of requiring very little information about the process. There is, however, a significant disadvantage because the method inherently gives very poor damping, typically $\zeta \approx 0.2$, see Åström and Hägglund (1995b).

There are several reasons to look for better methods to design PI controllers. One reason is the significant impact it may give because of the widespread use of the controllers. Another reason is that emerging auto-tuners and tuning devices can benefit significantly from improved design methods.

There are several requirements on an efficient design method. It should be applicable to a wide range of systems and it should have the possibility to introduce specifications that capture the essence of real control problems. Furthermore, the method should be robust in the sense that it provides controller parameters if they exist, or if the specifications cannot be met an appropriate diagnosis should be presented. We believe these requirements are satisfied by the method presented in this paper. The approach gives a simple way to solve simple control problems and more difficult ones where more efforts are needed. The method will give a PI controller which satisfies the specifications, provided that such a controller exists. If not, the reasons for failure will be indicated. It can also be used to develop simpler methods for restricted classes of systems as was done in Åström and Hägglund (1995a).

The method proposed in this paper formulates the design problem as an optimization problem: optimize the load disturbance rejection with a constraint on the maximum sensitivity. By exploiting the structure of the optimization problem it is reduced to the solution of algebraic equations. Efficient iterative methods are given together with good methods for finding starting values.

The method presented assumes a linear process whose dynamics is characterized in terms of a transfer function, which does not have to be rational. Thus, it can be applied to systems described by partial differential equations. If the transfer function is not known it can be obtained by system identification.

Parts of the proposed method builds on previous works. The use of optimization was discussed in Hazebroek and van der Waerden (1950). In this and other early works the emphasis was on criteria that admitted analytical solutions. The idea of optimizing load disturbance rejection with sensitivity constraints was suggested by Shinskey (1990). He used a constraint in terms of a rectangle around the critical point but the idea to use a constraint on the maximum sensitivity, M_s, was proposed by Persson (1992) and Persson and Åström (1992). The use of both M_s and the maximum complementary sensitivity, M_p, as design parameters was suggested by Schei (1994). The new contributions in this paper are: the analysis of the nature of the sensitivity constraints, the efficient numerical procedures, and the method for determining set point weighting.

2. FORMULATION OF THE DESIGN PROBLEM

The formulation of a design problem includes a characterization of the process and its

*Received 24 March 1997; received in final form 10 December 1997. This paper was recommended for publication by Associate Editor C. P. Jobling under the direction of Editor Sigurd Skogestad. *Corresponding author* Professor K. J. Åström. Tel. +46-46 2228781; Fax +46-46 138118; E-mail kja@control.lth.se.

† Department of Automatic Control, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden.

environment, the controller structure, and specifications on the performance of the closed-loop system.

Requirements: Before going into details we will first discuss the requirements. The method should be applicable to a wide range of systems and it should be based on specifications that reflect the essence of real control problems. Furthermore, the method should give a simple solution to simple problems. With more efforts and more skills from the designer it should also be possible to sharpen the specifications.

The process: It is assumed to be linear, and specified by a transfer function $G(s)$ which is analytical with finite poles and possibly an essential singularity at infinity. This description covers finite dimensional systems with time delays and infinite dimensional systems described by linear partial differential equations.

The controller: It is a PI controller described by

$$u(t) = k(by_{sp}(t) - y(t)) + k_i \int_0^t (y_{sp}(\tau) - y(\tau))\,d\tau, \quad (1)$$

where $u(t)$ is the control signal, $y(t)$ is the process output, $y_{sp}(t)$ is the set point, and k, k_i, and b are controller parameters.

When there are substantial measurement noise, it is customary to filter the measurement signal with a filter, typically of the form

$$G_f(s) = \frac{1}{1 + sT_f}, \quad (2)$$

where T_f is the filter time constant.

Thus, the controller has three or possibly four parameters, k, k_i, b, and T_f. It is industry practice to use integration time, defined as $T_i = k/k_i$, instead of parameter k_i. However, for the computations it is more convenient to use k_i. Industrial controllers typically use either $b = 0$ or $b = 1$, but lately it has been recognized that it is advantageous to use full range of b-values, that is $0 \le b \le 1$. The controller given by equation (1) is said to have two degrees of freedom when $b \ne 1$. The advantage of such structures has been pointed out by Horowitz (1963) and their use in PID controllers is discussed in Shigemasa *et al.* (1987) and Åström and Hägglund (1995b).

Specifications: They express requirements on

- load disturbance response,
- set point response,
- robustness with respect to model uncertainties.

In process control applications efficient rejection of load disturbances is of primary concern, whereas set point responses are typical of secondary importance. However, set point response may be of primary importance, for example, in motion control systems. Although it has been frequently pointed out by engineers that load disturbances is of primary concern, it is interesting to note that papers on PI control traditionally focus on set point response, see e.g. Shinskey (1990).

The sensitivity to model uncertainty is of primary significance, and observe that the poor sensitivity is one of the major drawbacks of the classical Ziegler–Nichols method.

In virtually all applications it is useful to have a tuning parameter that permits adjustment of the trade off of aggressiveness versus robustness. The effects of the tuning parameter should be transparent to the user. In the method presented, we have a tuning parameter to specify the sensitivity to model uncertainties.

2.1. *A formal description*

In order to use a formal design method it is necessary to capture specifications in a suitable mathematical form. This is extensively discussed in Åström and Hägglund (1995b).

2.1.1. *Load disturbance rejection.* It can be conveniently expressed in terms of the integrated absolute error due to a load disturbance in the form of a unit step at the process input, i.e.

$$IAE = \int_0^\infty |e(t)|\,dt. \quad (3)$$

This criterion is difficult to deal with analytically because the evaluation requires computation of time functions. The integrated error defined by

$$IE = \int_0^\infty e(t)dt, \quad (4)$$

is much more convenient. In Åström and Hägglund (1995b) it is shown that $IE = 1/k_i$. Thus, the criterion IE is directly given by the integrating gain of the controller. Remember that, $IE = IAE$ if the error is positive. Furthermore, if the system is well damped the criteria will be close which, in our case, will be ensured by the sensitivity constraints.

2.1.2. *Sensitivity to modeling errors.* They can be expressed in terms of the largest value of the sensitivity function. Let the loop transfer function be $L(s) = G(s)G_c(s)$ where G_c is the controller transfer function, and let the sensitivity function be $S(s) = 1/(1 + L(s))$. The maximum sensitivity is then given by $M_s = \max|S(i\omega)|$. Keep in mind that the quantity M_s is the inverse of the shortest distance from the Nyquist curve of the loop transfer function to the critical point -1. Typical values of M_s are in the range of 1.2–2.0.

With a constraint on M_s it follows from the circle criterion that the closed loop system will also remain stable with a static nonlinearity in the loop, provided that the nonlinearity is bounded by two straight lines with slopes $M_s/(M_s+1)$ and $M_s/(M_s-1)$, see Khalil (1992).

Let $T(s) = 1 - S(s) = L(s)/(1 + L(s))$ be the complementary sensitivity function. The sensitivity can also be expressed by the largest value of the complementary sensitivity function, i.e. $M_p = \max|T(i\omega)|$. The value M_p is the size of the resonance peak of the closed loop system obtained with $b = 1$, see equation (1). Typical values of M_p are in the range of 1.0–1.5.

2.1.3. *Set point response.* The design has so far focused on the response to load disturbances, which is of primary concern. However, it is also important to have a good response to set point changes. The transfer function from set point to process output is given by

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{k_i + bks}{k_i + ks}\frac{L(s)}{1 + L(s)} = G_{sp}(s). \tag{5}$$

One way to give specifications on the set point response is to specify the resonance peak of the transfer function $G_{sp}(s)$, i.e.

$$M_{sp} = \max|G_{sp}(i\omega)|. \tag{6}$$

Consequently, the b-value is determined such as it fulfills equation (6). Notice that $M_{sp} < M_p$ if $b < 1$.

2.1.4. *Measurement noise filtering.* The PI controller has a high-frequency gain k. The effects of measurement noise can be reduced, substantially, by filtering the signal with the filter given by equation (2). The specifications can be expressed in terms of the magnitude of the high-frequency gain of the controller. This specification is optional and only used in exceptional cases.

2.2. *Design parameters*

The tradeoff between performance and robustness varies between different control problems. Therefore, it is desirable to have a design parameter to change the properties of the closed-loop system. Ideally, the parameter should be directly related to the performance of the system, it should not be process oriented. There should be good default values so a user is not forced to select some value. This is of special importance when the design procedure is used for automatic tuning. The design parameter should also have a good physical interpretation and natural limits to simplify its adjustment.

The variables M_s and M_p are both possible candidates for design variables. The influence of these design parameters is illustrated in Fig. 1. The curves to the left show that M_s is a suitable design parameter. Decreasing values of M_s results in time responses that are slower but less oscillatory. The curves to the right show that M_p is not suitable. Although M_p is varied between 1.2 and 2.0, no significant variation in the time responses is noticed. The reason is, that the M_p circles are too close to each other in the frequency region around $-180°$. Therefore, we will choose M_s as a design variable.

On the other hand, it is important that the resulting M_p value is not too large. We will therefore also calculate M_p when the design is completed. If M_p is too large there are several possibilities. One is to repeat the design with a smaller M_s value. Another is to use constraints on both M_s and M_p. This will give rise to difficulties in the optimization because the set enclosed by M_s and M_p circles is not convex. This difficulty can be avoided by constructing a circle that has the M_s and M_p circles in its interiors. A straightforward calculation shows that this is a circle with a center C and a radius R where

$$\begin{aligned} C &= \frac{M_s - M_sM_p - 2M_sM_p^2 + M_p^2 - 1}{2M_s(M_p^2 - 1)}, \\ R &= \frac{M_s + M_p - 1}{2M_s(M_p - 1)}. \end{aligned} \tag{7}$$

An advantage with this is that the same optimization procedure can be used because the constraint set is a circle.

3. THE OPTIMIZATION PROBLEM

The design problem discussed in the previous section can be formulated as an optimization problem: Find controller parameters that maximize k_i subject to the constraints that the closed-loop system is stable and that the Nyquist curve of the loop transfer function satisfies the encirclement condition and that it is outside a circle with center at $s = -C$ and radius R.

Introduce $L(s) = (k + k_i/s)G(s)$ and the function

$$f(k, k_i, \omega) = |C + \left(k - i\frac{k_i}{\omega}\right)G(i\omega)|^2. \tag{8}$$

The sensitivity constraint can then be expressed as

$$f(k, k_i, \omega) \geq R^2 \tag{9}$$

and the optimization problem is to maximize k_i subject to the sensitivity constraint (9).

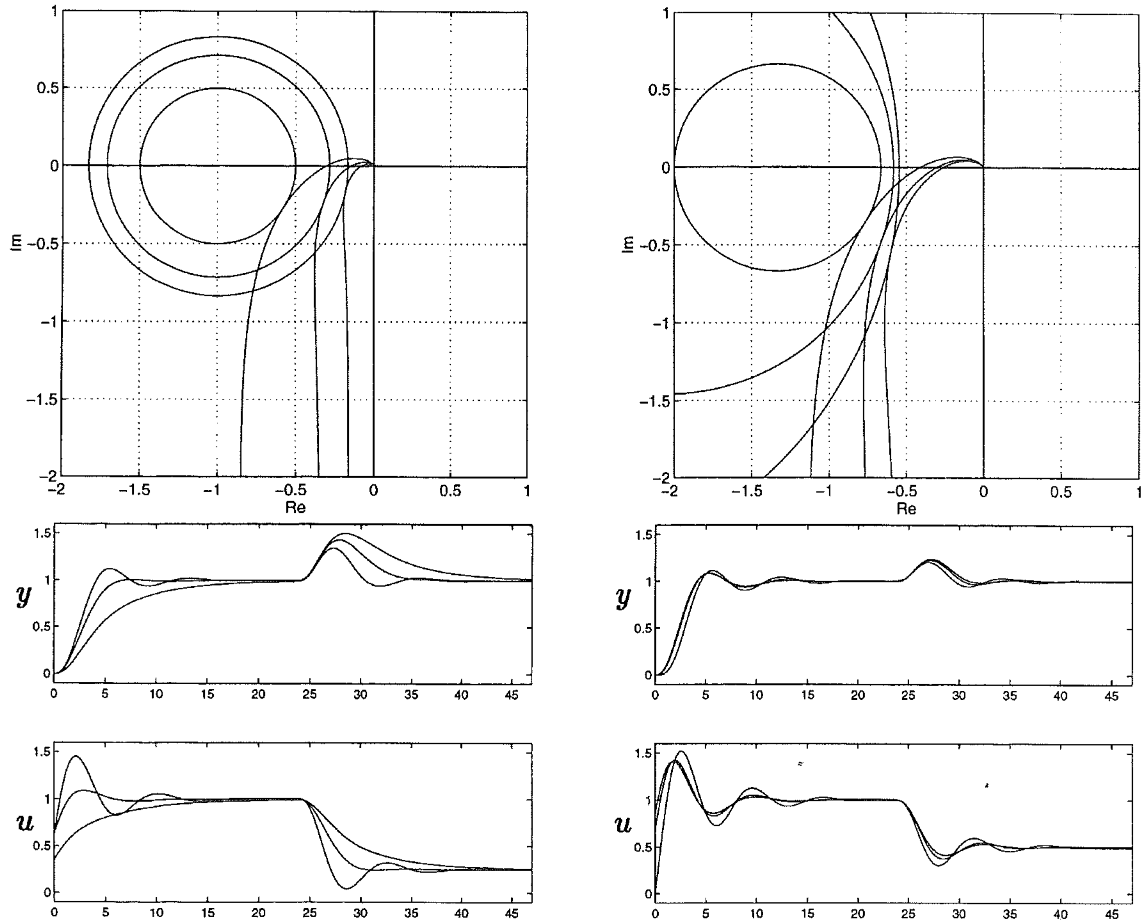


Fig. 1. Illustrates the effects of using M_s (left) and M_p (right) as a design parameter. The upper curves show the Nyquist curves of the loop transfer functions together with the M_s and M_p circles for $M_s = 1.2, 1.4, 2.0$ and $M_p = 1.2, 1.4, 2.0$. The lower curves show process outputs and control signals for the different design parameters. They give responses to a set point change followed by a load disturbance.

Let $\alpha(\omega)$ and $\beta(\omega)$ be the real and imaginary parts of the process transfer function. Hence

$$G(i\omega) = \alpha(\omega) + i\beta(\omega) = r(\omega)e^{i\varphi(\omega)},$$

where

$$\alpha(\omega) = r(\omega)\cos\varphi(\omega),$$
$$\beta(\omega) = r(\omega)\sin\varphi(\omega).$$

The function f can then be written as

$$f(k, k_i, \omega) = C^2 + 2C\alpha(\omega)k + 2C\frac{\beta(\omega)}{\omega}k_i + r^2(\omega)k^2 + \frac{r^2(\omega)}{\omega^2}k_i^2. \tag{10}$$

In the following, we will occasionally drop the argument ω in α, β, r, and φ in order to simplify the writing.

The optimization problem is nontrivial because the constraint, which is infinite dimensional, defines a set in parameter space which is not convex. There are also other subtleties which may cause problems. For a specific problem it is not difficult to solve the problem numerically with standard optimization routines because the search range can often be limited and if the optimization fails it is possible to interfere manually. Since PI controllers are very common it is, however, worthwhile to make special algorithms which are tailored for the problem. Such procedures are also required for automatic tuning where manual interaction is very inconvenient.

Before discussing the solution to the optimization problem we will investigate the sensitivity constraint which is a key difficulty of the problem.

3.1. *The sensitivity constraint*

The sensitivity constraint given by equation (9) has a nice geometric interpretation. For fixed ω, equation (9) represents the exterior of an ellipse in the k–k_i plane. The ellipse has its axes parallel to the coordinate axes. For $0 \le \omega < \infty$ the ellipses generate envelopes that define the boundaries of the sets of parameters which satisfy the sensitivity constraints. It can be assumed that the process transfer function is such that a stable closed system is obtained with positive k_i. It is thus sufficient to consider the upper-half of the k–k_i plane. The center of

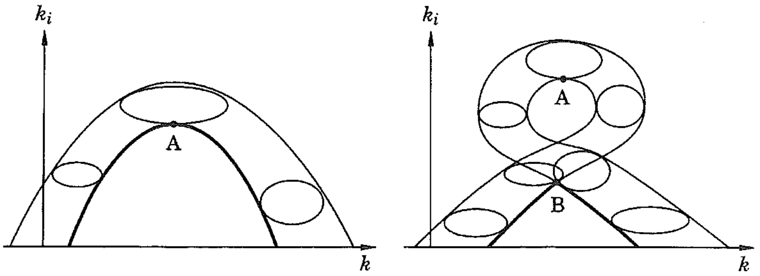


Fig. 2. Geometrical illustration of the sensitivity constraint (9) and the envelope generated by it. The envelope to the left has a continuous derivative, but the one to the right has a corner.

the ellipses generate the boundary of the stability regions. The envelopes for different systems may have different characters as shown in Fig. 2. Notice in particular that the envelope may have a corner as is illustrated in Fig. 2b.

3.1.1. *Stability*. The stability region can be expressed in terms of a condition on k and k_i which is obtained by setting $C = 1$ and $R = 0$ in equation (9), i.e.

$$k\alpha(\omega) + k_i \frac{\beta(\omega)}{\omega} + i\left(k\beta(\omega) - k_i \frac{\alpha(\omega)}{\omega}\right) = -1. \tag{10}$$

Hence

$$k\alpha(\omega) + k_i \frac{\beta(\omega)}{\omega} = -1,$$

$$k\beta(\omega) - k_i \frac{\alpha(\omega)}{\omega} = 0.$$

Solving these equations for k and k_i gives the following parametric description of the boundary of the stability region:

$$k = -\frac{\alpha(\omega)}{r^2(\omega)},$$
$$k_i = -\omega \frac{\beta(\omega)}{r^2(\omega)}, \tag{11}$$

where $r^2 = \alpha^2(\omega) + \beta^2(\omega)$, and the parameter ω ranges from zero to infinity. Notice that the stability region may consist of disjoint sets.

3.2. *Optimization*

Having understood the nature of the constraints it is now conceptually clear how to solve the optimization problem. It is simply a matter of finding the largest value of k_i on the envelope. The difficulties that may occur are due to the fact that there may be several local maxima and that the maximum may occur at a corner.

Since it is quite time consuming to generate the envelope it is desirable trying to find algorithms that can give a more effective solution. It is also of interest to characterize the situations when there is only one local minimum.

The envelope is given by

$$f(k, k_i, \omega) = R^2,$$
$$\frac{\partial f}{\partial \omega}(k, k_i, \omega) = 0. \tag{12}$$

Since the function f is quadratic in k_i the envelope has two branches. Only one branch corresponds to stable closed-loop systems. Compare with Fig. 2.

Instead of generating the envelopes and searching for the largest value of k_i on the envelope, we will characterize the points where k_i has its largest value. From the discussion of the constraint it is clear that there are two cases. The simplest case is when the largest value of k_i occurs at a point where the envelope has a continuous derivative. The other case is when the envelope has a corner.

The envelope given by equation (12) defines implicitly k_i as a function of k. To find the maximum of this function we observe that

$$df = \frac{\partial f}{\partial k} dk + \frac{\partial f}{\partial k_i} dk_i + \frac{\partial f}{\partial \omega} d\omega = 0. \tag{13}$$

It follows from equation (12) that $\partial f/\partial \omega = 0$ on the envelope. At a local extremum we have $dk_i = 0$. For arbitrary variations of dk we must thus require that $\partial f/\partial k = 0$. Combining this with the envelope conditions (12) we get

$$\frac{\partial f}{\partial k}(k, k_i, \omega) = 0,$$
$$\frac{\partial f}{\partial \omega}(k, k_i, \omega) = 0, \tag{14}$$
$$f(k, k_i, \omega) = R^2,$$

which corresponds to the situation when the maximum occurs at a point on the envelope where it has continuous derivatives, see Fig. 2a. In the Nyquist diagram this agrees with the case when the loop transfer function is tangent to the circle at one point.

We will now consider the case when the largest value of k_i occurs at a point where the envelope has a corner. This occurs at the intersection of ellipses corresponding to two different frequencies, ω_1 and ω_2, see Fig. 2b. The envelope condition (12) is then satisfied for both frequencies. This gives the condition

$$\begin{aligned} f(k, k_i, \omega_1) &= R^2, \\ \frac{\partial f}{\partial \omega}(k, k_i, \omega_1) &= 0, \\ f(k, k_i, \omega_2) &= R^2, \\ \frac{\partial f}{\partial \omega}(k, k_i, \omega_2) &= 0. \end{aligned} \tag{15}$$

In the Nyquist diagram this corresponds with the case when the loop transfer function is tangent to the circle at two points.

It is thus possible to characterize the point where k_i has its largest value by algebraic equations. This means that the design problem is reduced to solving algebraic equations (14) or (15), and that elaborate search procedures are avoided. Both equations can be solved by using the Newton–Raphson method. Equation (14) which represents the most common situation can, however, be simplified substantially.

3.3. *A simplification*

Equation (14) is a nonlinear equation in three variables, k, k_{ii} and ω. It is possible to solve this equation directly, but a much more efficient algorithm will be obtained by eliminating some variables.

Inserting expression (10) into equation (14), gives

$$\begin{aligned} \frac{\partial f}{\partial k} &= 2C\alpha + 2r^2 k = 0, \\ \frac{\partial f}{\partial \omega} &= 2C\left(\frac{\beta}{\omega}\right)' k_i + 2C\alpha' k \\ &\quad + \left(\frac{r^2}{\omega^2}\right)' k_i^2 + 2rr'k^2 = 0, \\ f &= R^2, \end{aligned} \tag{16}$$

where prime means differentiation with respect to ω. Solving k and k_i from the first and last equations gives

$$\begin{aligned} k &= -C\frac{\alpha}{r^2} = -C\frac{1}{r}\cos\varphi, \\ k_i &= -\frac{\omega\beta C}{r^2} - \frac{\omega R}{r} = -\frac{\omega}{r}(C\sin\varphi + R), \end{aligned} \tag{17}$$

where the positive sign is chosen to satisfy the encirclement criterion. The following condition is obtained by inserting the expressions of k and k_i in the second equation of equation (16),

$$\begin{aligned} h(\omega) &= 2R\left(\left(C\frac{\beta}{r} + R\right)\left(\frac{r'}{r} - \frac{1}{\omega}\right) - C\left(\frac{\beta}{r}\right)'\right) \\ &= 2R\left((R + C\sin\varphi)\left(\frac{r'}{r} - \frac{1}{\omega}\right) - C\varphi'\cos\varphi\right) \\ &= 0. \end{aligned} \tag{18}$$

Thus, the solution to equation (16) is reduced to a single algebraic equation (18) in ω. Solving it gives the frequency ω_0 for which we can compute the controller gains k and k_i given by equation (17).

The condition (16) does not tell if the extremum is a minimum, a maximum or saddle point, but constraint (9) implies that the function should be a minimum with respect to ω. This gives the following local condition:

$$\frac{\mathrm{d}^2 f}{\mathrm{d}\omega^2}(\omega_0) > 0. \tag{19}$$

Equation (18) can be solved iteratively with the Newton–Raphson method which converges very fast if suitable initial conditions are given. Notice, however, that in general there may be several solutions which can be found by starting the iteration from different initial conditions.

For special classes of systems, for example systems with monotonic transfer function, it is possible to provide good initial conditions. The following result is useful.

Theorem 1. Let ω_φ denote the frequency where the process has a phase of φ. Assume that the transfer function $G(s)$ has positive low-frequency gain and that

$$\begin{aligned} \frac{\mathrm{d}\arg G(\mathrm{i}\omega)}{\mathrm{d}\omega} &< 0, \\ \frac{\mathrm{d}\log_{10}|G(\mathrm{i}\omega)|}{\mathrm{d}\log_{10}\omega} &< 1. \end{aligned} \tag{20}$$

Then there exists a solution to equation (18) in the interval

$$\omega_{90} < \omega < \omega_\dagger = \omega_{180-\arcsin R/C}. \tag{21}$$

Proof. It is assumed that the low-frequency gain of the process is positive. Then the integral gain k_i must be positive if the closed-loop system should be stable. For this reason equation (17) implies that $R + C\sin\varphi < 0$. It follows from the assumptions (20) that $\varphi' < 0$ and $r'/r - 1/\omega < 0$. As a result equation (18) and the assumptions imply that $h(\omega_{90}) > 0$ and $h(\omega_\dagger) < 0$. But h is a continuous function, therefore equation (18) must have a solution in the interval (21). □

Remark 1. Notice that since a PI controller has negative phase for all frequencies, the solution must be found in the interval $\omega_{90} < \omega < \omega_{\dagger}$. Thus, the monotonicity condition has to be valid in that interval.

Remark 2. It follows from equation (17) that the condition (21) implies that both k and k_i are positive. Systems with monotone transfer functions satisfy the condition (20). For such systems it is thus straightforward to find a good range of frequencies to solve equation (18). Also notice that as long as the phase decreases monotonically it is possible to have an increase in the amplitude curve, provided that the slope is never larger than one. Most systems encountered in process control satisfy these conditions.

A necessary condition for stability of the closed-loop system is that the parameter k_i is positive which implies that $\omega_0 > \omega_{90}$.

3.4. *Initial conditions*

Good initial conditions is a crucial factor for the computational efficiency of the Newton–Raphson iterations when solving the optimization problem. The two cases when the envelope has a continuous derivative or a corner has to be treated separately.

First we consider the case when the envelope has no corner. The solution to the optimization problem is obtained by solving the algebraic equation (18) with Newton–Raphson for a suitable search range. According to Theorem 1 the search range is chosen as $[\omega_{90}, \omega_{\dagger}]$. To obtain better initial conditions we narrow it by applying interval bisection in equation (18).

Finally, the case when the envelope has a corner is considered. This problem is more difficult to solve, because it is necessary to solve a system of equations (15), which in this case requires four initial conditions of ω_1, ω_2, k and k_i. They can be obtained with the following procedure.

For fixed values of ω the sensitivity constraint, equation (12), represents ellipses in gain space which generates envelopes. It is quite complicated to compute the envelopes, however, they can be approximated by the loci of the vertices of the ellipses. The horizontal vertices are given by

$$k = -\frac{\alpha C}{r^2} \pm \frac{R}{r}, \qquad k_i = -\frac{\omega \beta C}{r^2}, \tag{22}$$

where the left vertex corresponds to a minus sign and the right vertex to a plus sign. The loci of the vertices define curves in the gain space that enclose the envelope. Good initial values for the Newton–Raphson iteration can be obtained by finding the point on the loci where k_i has its largest value. When there is a corner this value is obtained as the intersection of the vertices given by equation (22), see Fig. 2b, giving ω_1 and ω_2. Thus, k can be computed from equation (22) for either ω_1 or ω_2. Because of the construction this value overestimates the integral gain.

Note that the same technique can be used as an alternative way of finding initial conditions in the case of no corners. In this case the envelope can be approximated by the loci of the lowest vertex of the ellipse, that is

$$k = -\frac{\alpha C}{r^2}, \qquad k_i = -\frac{\omega \beta C}{r^2} - \frac{\omega R}{r}. \tag{23}$$

Notice that this equation is identical to equation (17). The upper vertex is not of any interest because it gives an unstable closed-loop system.

4. THE DESIGN PROCEDURE

We have thus found efficient procedures to determine feedback gains k and k_i, by optimizing load disturbance rejection subject to constraints on sensitivity to model uncertainties. To complete the design procedure, it remains to determine the set point weighting, i.e. parameter b in equation (1).

4.1. *Set point weighting*

The set point response is governed by the transfer function G_{sp} given by equation (5). In order to have a small overshoot in set point response, set point weighting b will be determined so that $M_{sp} = \max|G_{sp}(i\omega)|$ is close to one. It follows from equations (5) that $M_{sp} \le M_p$ when $0 \le b \le 1$. A bound of M_{sp} is thus given indirectly through M_p.

We will make the approximation that maximum of $|G_{sp}(i\omega)|$ occurs for ω_{mp}, where $\omega = \omega_{mp}$ is the frequency where the maximum of $|L(i\omega)/(1 + L(i\omega))|$ occurs. Parameter b will be determined so that

$$|G_{sp}(i\omega_{mp})| = 1 \tag{24}$$

with the constraint $0 \le b \le 1$. Using equations (5) and (24) this implies that

$$b = \begin{cases} \dfrac{\sqrt{k^2\omega_{mp}^2 - k_i^2(M_p^2 - 1)}}{k\omega_{mp}M_p} & \text{if } (\omega_{mp}k/k_i)^2 \ge M_p^2 - 1, \\ 0 & \text{if } (\omega_{mp}k/k_i)^2 < M_p^2 - 1. \end{cases} \tag{25}$$

If $b = 0$, it is not sure that the design objective (24) will be obtained. If the set point response is important and the M_p value is large, the design can then be repeated with a lower value of M_s or using the constraint given by the circle (7).

4.2. *Measurement noise filtering*

Having performed a design of the controller we obtain the controller gain k, which is also the high-frequency gain of the controller. Combined with the specifications on roll-off we can then determine the order and the bandwidth of a suitable noise filter.

Here we will only consider the first-order filter given by equation (2). The choice of filter-time constant T_f in equation (2) is a tradeoff between filtering capacity and loss of performance. A large value of T_f provides an effective noise filtering, but it will also change the control performance. A small value of T_f means that the control performance is retained, but the noise filtering is less effective.

A nice feature of the new design procedure is that it provides a systematic way to determine T_f. The choice

$$T_f = \frac{1}{m\omega_0}, \tag{26}$$

makes it possible to determine the effects of the filter at the frequency ω_0. We get

$$|G_f(\mathrm{i}\omega_0)| = \frac{1}{\sqrt{1 + 1/m^2}},$$

$$\arg G_f(\mathrm{i}\omega_0) = -\arctan(1/m).$$

Reasonable values of m are in the interval 5–10. For $m = 5$ we get $|G_f(\mathrm{i}\omega_0)| = 0.981$ and $\arg G_f(\mathrm{i}\omega_0) = -11°$. For $m = 10$ we get $|G_f(\mathrm{i}\omega_0)| = 0.995$ and $\arg G_f(\mathrm{i}\omega_0) = -5.7°$. These modifications of the loop transfer functions give normally only minor changes in control loop performance. If needed, it is also possible to recalculate the parameters by determining the filter as described above, and then perform a regular design of the PI controller with the transfer function $G_f(s)G(s)$.

4.3. *The procedure*

To sum up we find that the design problem can be solved by the following procedure:

(1) Choose the design parameters M_s and/or M_p and compute C and R.
(2) Determine the search range ($\omega_l < \omega < \omega_h$). We have $\omega_l = \omega_{90}$ and $\omega_h = \omega_{180}$. For systems that satisfy the monotonicity condition (20) we have $\omega_h = \omega_{180\text{-arcsin}\,R/C}$. Notice that there may be several search intervals. Narrow the search range by applying interval bisection to equation (18).
(3) *Normal case*: Use the initial values from Step 2 and solve equation (18) by Newton–Raphson. Evaluate the condition (19) and compute M_s and M_p. If both are satisfactory go to Step 5 otherwise compute new C and R and go to Step 2.
(4) *Corner case*: Compute the approximate envelope for values in the range $\omega_l < \omega < \omega_h$. Determine largest value of k_i, the frequencies of tangency ω_1 and ω_2 and compute k. These are used as initial values too solve equation (15) with Newton–Raphson. Compute M_s and M_p. If both are satisfactory go to Step 5 otherwise compute new C and R and go to Step 2.
(5) Determine the parameter b from equation (25).
(6) Evaluate the design including noise sensitivity. Modify the design parameters if required.

For simple systems it is sufficient to choose only the design parameter M_s. For special classes of systems all choices can be made automatically.

5. EXAMPLES

The design method has been tested on a large number of examples. In this section we will give a number of examples illustrating its properties.

5.1. *Typical process control problems*

To start with we will consider some representative systems which are normally encountered in process control. They have the following transfer functions,

$$G_1(s) = \frac{1}{(s+1)^3},$$

$$G_2(s) = \frac{1}{(s+1)(1+0.2s)(1+0.04s)(1+0.008s)},$$

$$G_3(s) = \frac{e^{-15s}}{(s+1)^3}, \qquad G_4(s) = \frac{1}{s(s+1)^2},$$

$$G_5(s) = \frac{1-2s}{(s+1)^3}, \qquad G_6(s) = \frac{9}{(s+1)(s^2+2s+9)}.$$

Systems G_1 and G_2 represent processes that are relatively easy to control. System G_3 has a long dead time, and G_4 models an integrating process. System G_5 has a zero in the right half plane, and system G_6 has complex poles with relative damping 0.33. Systems of type G_5 and G_6 are not common in process control, but they have been included to demonstrate the wide applicability of the design procedure. For all systems, except G_6, we can, by inspection, verify that they satisfy the monotonicity assumption. The system $G_6(s)$ will be discussed further on in Example 5 where it is shown that it also satisfies this assumption. We thus know that there exists a solution where the envelope has a continuous derivative.

Figure 3 shows the Nyquist curves of the loop transfer functions obtained for two values of the

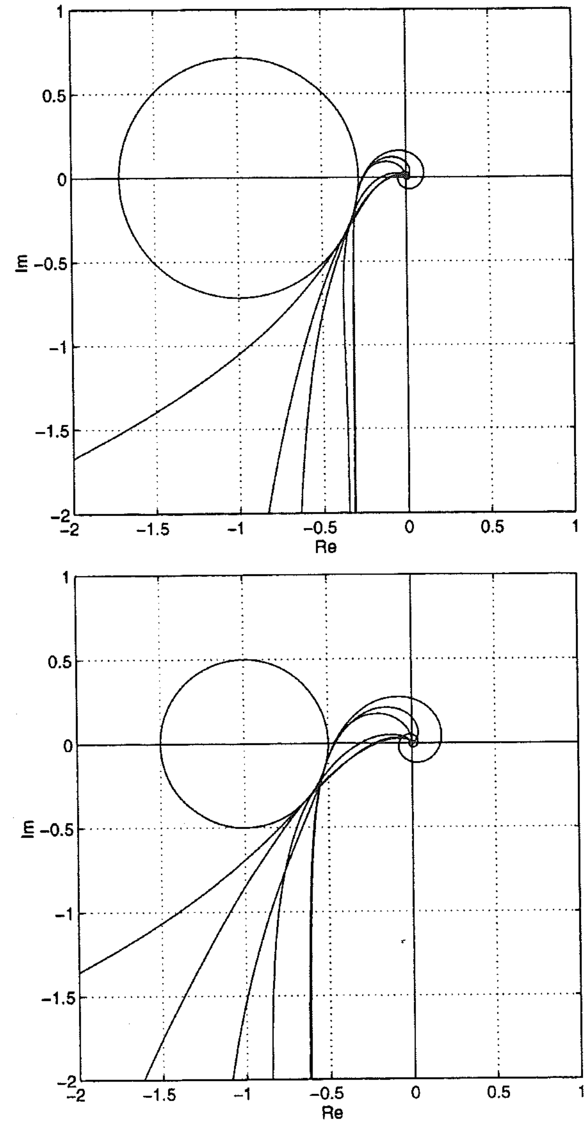


Fig. 3. The Nyquist plots of the open-loop frequency response for the systems $G_j(s)$, $j = 1, \ldots, 6$ for $M_s = 1.4$ in the upper figure and $M_s = 2.0$ in the lower one.

design parameter M_s. The responses to changes in set point and load are shown in Fig. 4, and the details of the design calculations and simulations are summarized in Table 1.

Even though the systems G_1–G_6 represent processes with large variations in process dynamics, Figure 4 shows that the resulting closed-loop responses become similar for each value of M_s. This is important because it means that the proposed design procedure gives closed-loop systems with desired and predictable properties. The fact that even integrating processes can be treated in the same way as stable processes is interesting. In many other design approaches, stable and integrating processes have to be treated separately, see Åström and Hägglund (1995b).

There is also a large similarity between the responses obtained with the different values of the tuning parameter M_s. This shows that the M_s-value is a suitable tuning parameter. Responses obtained with $M_s = 1.4$ show little or no overshoot. This is normally desirable in process control. Responses obtained with $M_s = 2.0$ give faster responses. The settling time at load disturbances, t_s, is significantly shorter with this larger value of M_s. On the other hand, these responses are oscillatory with a larger overshoot. This can be seen from the comparison between IE and the integrated absolute error IAE in Table 1.

The controller gain k varies significantly with the design parameter M_s. However, integral time T_i is fairly constant for the stable processes, i.e. all processes except G_4. This means that, for PI control, the different design specifications are mainly obtained by adjusting only the gain. This observation is made earlier, see Åström and Hägglund (1995b).

For the systems in Table 1 the values of M_p are smaller than the M_s values except for the system G_4 with $M_s = 1.4$ where M_s and M_p are equal. The constraint on the sensitivity function is thus the critical constraint for these systems.

Except for the integrating process G_4, the M_p values obtained for $M_s = 1.4$ are all close to one. Consequently, parameter b is also close to one. For $M_s = 2.0$, the M_p values are, however, larger. This means that the overshoots would have been significant if the set point weighting were chosen to $b = 1$. However, acceptable set point responses are obtained by using small values of b. In some cases, $b = 0$. It means that the procedure has failed to obtain $M_{sp} = 1$. If set point responses are important and if the overshoots are unacceptable, a redesign may be done using smaller values of M_s or optimization with constraint on M_p, see Section 4.

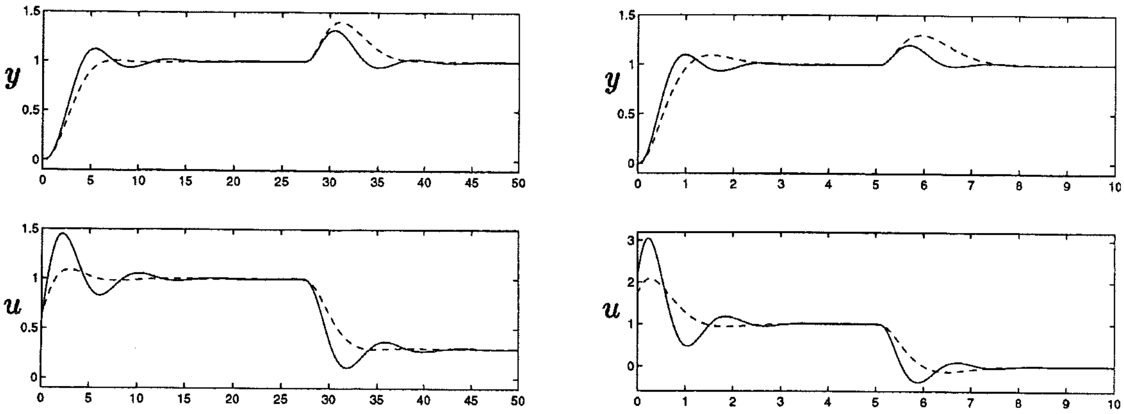
5.2. *More complex systems*

We will now discuss several examples with more complex dynamics.

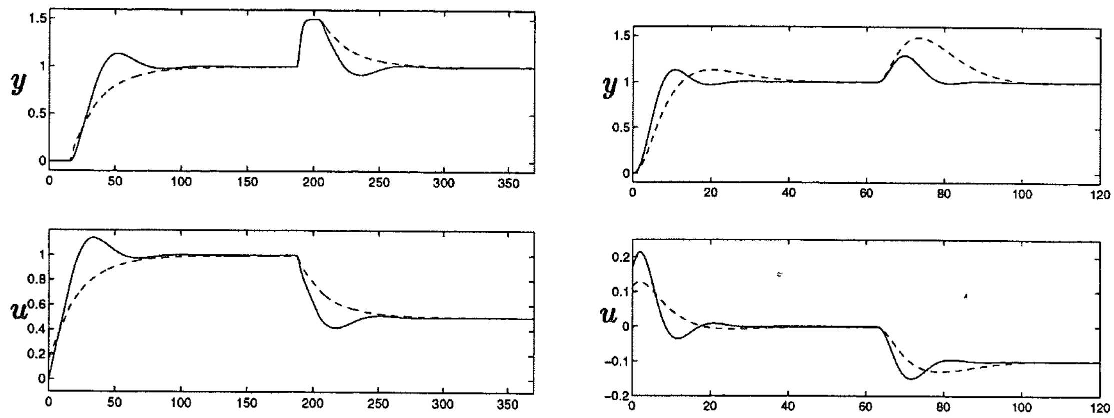
Example 1. *Pure time delay* (see Example 2). Many design methods for PI controllers perform poorly for systems with long relative time delays. The system G_3 is of this type because the ratio between the dead time and the dominating time constant is 15/3. A more extreme case is

$$G_7(s) = e^{-s}.$$

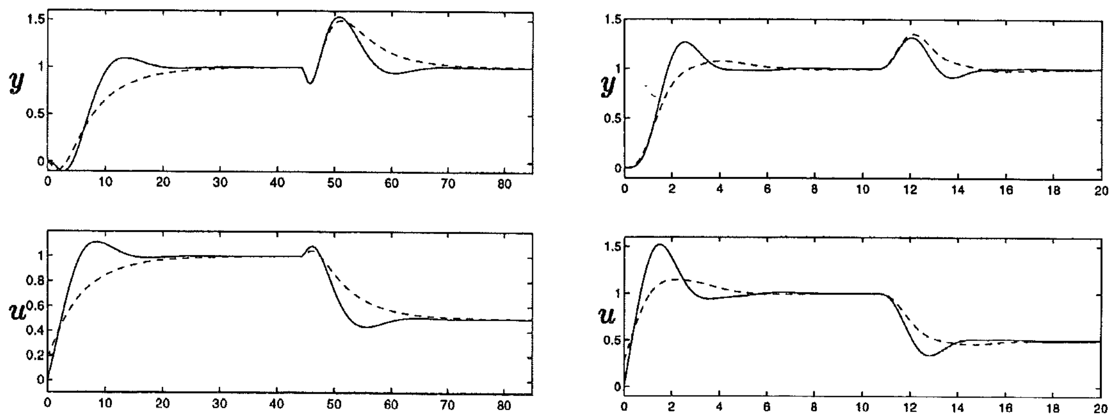
This system has $\omega_{90} = \pi/2$ and $\omega_{180} = \pi$. It satisfies the monotonicity condition (20). Making the design for $M_s = 1.4$ and 2.0, the design procedure gives the controller parameters which can be found in Table 2. The Nyquist diagrams and the set point and load disturbance responses are shown in Fig. 5. The figure shows that the design procedure manages to obtain controller parameters even for systems with extreme dead times.



Systems G_1 (left) and G_2 (right)



Systems G_3 (left) and G_4 (right)



Systems G_5 (left) and G_6 (right)

Fig. 4. Comparison between the PI controllers for $M_s = 1.4$ and 2.0. The graphs shows a step response followed by a load disturbance of the closed-loop system when designing for $M_s = 1.4$ (dashed line) and 2.0 (full line).

Example 2 (*Pure integrator with time delay*). A pure integrator with a time delay is another common model. The transfer function for such a system is

$$G_8(s) = \frac{e^{-s}}{s}.$$

This system has $\omega_{90} = 0$ and $\omega_{180} = \pi/2$. It fulfills the monotonicity condition (20). Making the design for $M_s = 1.4$ and 2.0, the design procedure gives the controller parameters which can be found in Table 2. The Nyquist diagrams and the set point and load disturbance responses are shown in Fig. 5, which demonstrates that the design procedure produces suitable controller parameters in this example too.

Table 1. Properties of controllers obtained for system G_1–G_7 for different values of the design parameter M_s

Process	M_s	k	T_i	b	IE	IE/IAE	ω_0	t_s	M_p
$G_1(s)$	1.4	0.633	1.95	1.00	3.07	1.00	0.74	10.3	1.00
	1.6	0.862	1.87	0.93	2.17	0.95	0.79	7.87	1.05
	1.8	1.06	1.82	0.70	1.72	0.86	0.82	6.77	1.24
	2.0	1.22	1.78	0.50	1.45	0.77	0.85	6.27	1.45
$G_2(s)$	1.4	1.93	0.745	0.89	0.387	1.00	3.33	2.25	1.10
	1.6	2.74	0.672	0.75	0.245	1.00	3.83	1.68	1.27
	1.8	3.47	0.625	0.63	0.180	0.96	4.25	1.39	1.46
	2.0	4.13	0.591	0.52	0.143	0.91	4.40	1.21	1.66
$G_3(s)$	1.4	0.164	6.16	1.00	37.5	1.00	0.096	91.3	1.00
	1.6	0.208	5.87	1.00	28.2	1.00	0.099	51.2	1.00
	1.8	0.241	5.66	0.92	23.5	0.88	0.101	39.8	1.02
	2.0	0.266	5.51	0.00	20.8	0.75	0.102	35.9	1.17
$G_4(s)$	1.4	0.167	14.0	0.70	84.0	0.97	0.29	35.3	1.40
	1.6	0.231	10.7	0.64	46.2	0.99	0.34	24.1	1.49
	1.8	0.286	9.00	0.57	31.5	0.99	0.38	18.5	1.62
	2.0	0.333	8.00	0.50	24.0	0.96	0.41	15.7	1.77
$G_5(s)$	1.4	0.179	1.78	1.00	9.90	0.90	0.38	28.6	1.00
	1.6	0.228	1.69	1.00	7.43	0.87	0.40	18.6	1.00
	1.8	0.265	1.64	0.87	6.18	0.80	0.41	14.7	1.04
	2.0	0.294	1.60	0.00	5.42	0.70	0.41	13.5	1.20
$G_6(s)$	1.4	0.313	0.373	0.88	1.19	0.87	1.98	4.13	1.04
	1.6	0.387	0.344	0.51	0.891	0.79	2.05	2.94	1.15
	1.8	0.441	0.325	0.00	0.739	0.70	2.05	2.69	1.26
	2.0	0.482	0.313	0.00	0.648	0.64	2.12	2.56	1.37

Table 2. Details of the design calculations of the systems G_7–G_{11} for different values of the design parameter M_s

Process	M_s	k	k_i	b	ω_0	M_p
$G_7(s)$	1.4	0.158	0.472	1.00	1.73	0.99
	2.0	0.255	0.854	0.00	1.83	1.17
$G_8(s)$	1.4	0.282	0.0418	0.66	0.54	1.45
	2.0	0.488	0.131	0.46	0.73	1.82
$G_9(s)$	1.4	2.94	11.5	0.81	7.89	1.17
	2.0	5.31	27.0	0.48	9.68	1.59
$G_{10}(s)$	1.4	1.25	1.62	0.78	3.49	1.23
	2.0	2.48	4.43	0.51	4.59	1.68
$G_{11}(s)$	1.4	1.30	2.03	0.86	3.75	1.13
	2.0	2.59	5.24	0.51	4.82	1.64

Example 3 (*A distributed parameter system*). The method also applies directly to systems described by partial differential equations. To illustrate this we consider a system described by a linear heat equation. Such a system has the transfer function

$$G_9(s) = \mathrm{e}^{-\sqrt{s}}.$$

Hence

$$G_9(\mathrm{i}\omega) = \mathrm{e}^{-\sqrt{\omega/2}}\mathrm{e}^{-\mathrm{i}\sqrt{\omega/2}}.$$

The system satisfies the monotonicity condition (20) and we have $\omega_{90} = 4.93$, $\omega_{150} = 13.7$, and $\omega_{180} = 19.7$. Making the design for $M_s = 1.4$ and 2.0, the design procedure gives the controller parameters which can be found in Table 2. Notice that in this case there is a significant increase in performance when changing from $M_s = 1.4$ to $M_s = 2.0$.

Example 4 (*Fast and slow mode*). Consider a system with the transfer function

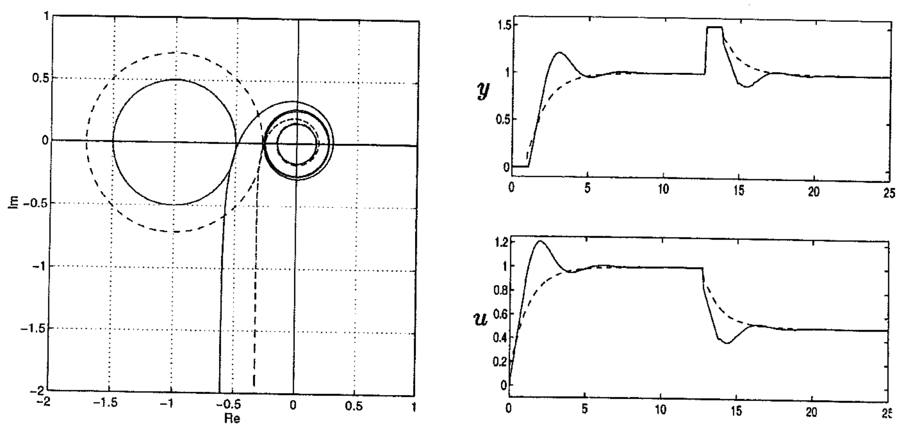
$$G_{10}(s) = \frac{100}{(s+10)^2}\left(\frac{1}{s+1} + \frac{0.5}{s+0.05}\right). \quad (27)$$

This system has two fast modes with time constants 0.1 s, one mode with a time constant of 1 s, and a slow mode with time constant 20 s. The static behavior is dominated by the slow mode which has a low-frequency gain of 10. The step response is dominated by the slow time constant, but it is the faster modes that are critical for the closed-loop system. The properties that are important for control are thus hidden in the step response. This means that most attempts to tune the system based on step response data will give poor results. Making the design for $M_s = 1.4$ and 2.0, the design procedure gives the controller parameters which can be found in Table 2. The properties of the closed loop systems obtained are illustrated in Fig. 5.

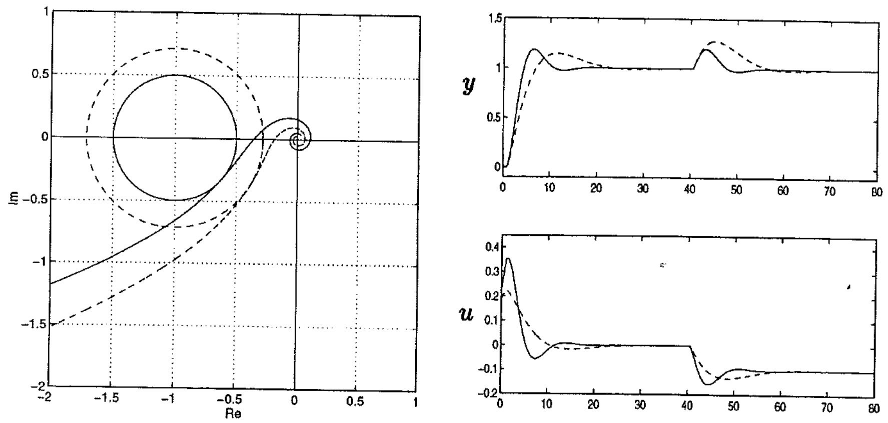
It is of interest to compare with the controller parameters obtained for the system

$$G_{11}(s) = \frac{150}{(s+10)^2(s+1)}, \quad (28)$$

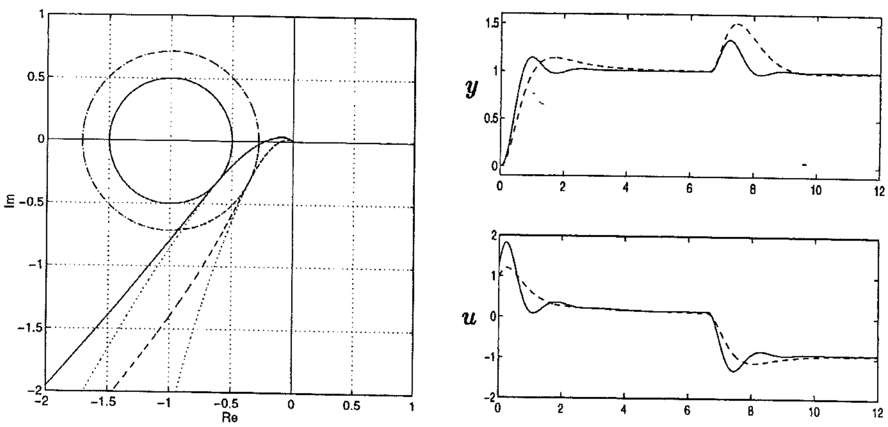
which is obtained by removing the slowest mode from the system (27). In this case when making the



System $G_7(s)$.



System $G_8(s)$.



Systems G_{10} and G_{11}.

Fig. 5. Nyquist curves of the loop transfer function and time responses of the closed-loop system when designing for $M_s = 1.4$ (dashed line) and 2.0 (full line) for systems $G_7(s)$, $G_8(s)$, $G_{10}(s)$ and $G_{11}(s)$.

design for $M_s = 1.4$ and 2.0, the design procedure gives the controller parameters which can be found in Table 2. These parameters are close to those obtained for the system (27), which shows that the design procedure manages to disregard the slow mode. The Nyquist curves for the two systems are compared in Fig. 5. Here it can be noticed that the two curves coincide for frequencies above ω_0.

Example 5 (*An oscillatory system*). Consider the system with the transfer function

$$G_{12}(s) = \frac{9}{(s+1)(s^2+as+9)}.$$

This is an interesting system from two points of view. First, the system has two oscillatory poles with relative damping $\zeta = a/6$. When parameter a is decreased it becomes more and more difficult to control. Second, depending on the value of parameter a the envelope may have a continuous derivative, $a \geq 1.0653$, or a corner, $a < 1.0653$.

For the case when the envelope has a continuous derivative a controller has already been designed for $a = 2$, see process G_6 at the beginning of this section. According to Fig. 4 the design gives a controller with good performance.

For the case when the envelope has a corner a controller was designed for $M_s = 2.0$. In Fig. 6 the Nyquist curves and the time responses are shown for the cases $a = 0.2, 0.5, 1.0$. The controller behaves reasonably well in spite of the poorly damped poles.

In Table 3 the controller parameters and the frequencies at which the loop transfer function is tangent to the M_s-circle are shown. Notice in Table 3 how the proportional gain is negative for small values of a. This is the only way to increase the damping of the oscillatory poles with a PI controller.

Finally, we illustrate how our design method will provide a reasonable PI controller for the extreme case $a = 0$. With the design parameter $M_s = 1.4$ we obtain the following controller parameters: $k = -0.183, k_i = 0.251$ and $b = 0$. The time responses are shown in Fig. 7. We observe that the set point response is quite reasonable, even if there is a trace of poorly damped modes. The load disturbance will, however excite the oscillatory modes. The fact that the PI controller is unable to provide damping of these modes is clearly noticeable in the figure.

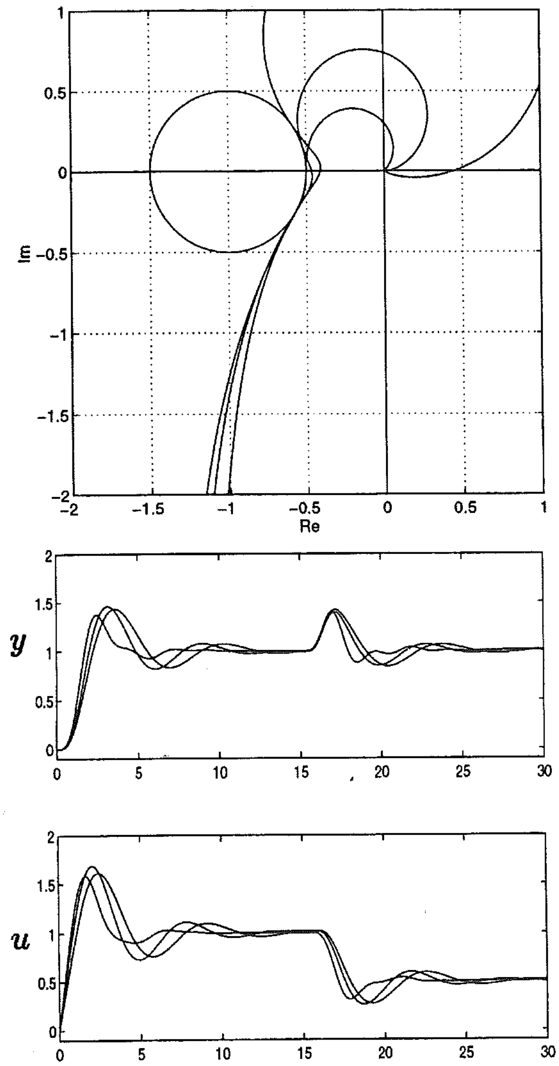


Fig. 6. Nyquist curves of the loop transfer function and time responses for Example 5 with $a = 0.2, 0.5, 1.0$, when designing for $M_s = 2.0$.

Table 3. Interesting parameters when designing a controller for $M_s = 2.0$ and different values of a in Example 5

a	k	k_i	ω_1	ω_2
0.0	− 0.29	0.68	0.97	2.75
0.1	− 0.25	0.82	1.08	2.71
0.2	− 0.20	0.93	1.16	2.67
0.5	− 0.09	1.17	1.37	2.55
1.0	0.09	1.38	1.65	2.30
2.0	0.48	1.54	2.79	2.79

Example 6 (*A conditionally stable system*). Consider the system with the transfer function

$$G_{13}(s) = \frac{(s+6)^2}{s(s+1)^2(s+36)}.$$

This system does not satisfy the monotonicity assumption because the phase lag is not monotonic. The system is conditionally stable, since the Nyquist curve crosses the negative real axis at points $s = -0.0191$ and $s = -0.1656$. We have $\omega_{90} = 0$, and $\omega_{180} = 1.69$ and 4.17. With proportional feedback the system is stable for $k < 6.04$ or $k > 52.26$.

There are two solutions to the optimization problem for $M_s = 2.0$: $k = 0.47$, $k_i = 0.067$, $b = 0.52$, and $k = 921, k_i = 1098, b = 0.50$. The first solution gives $\omega_0 = 0.5196$ rad/s and the second gives $\omega_0 = 25.93$ rad/s. Nyquist curves and time responses to set point changes and load disturbances are shown in Fig. 8. Notice the similarities of the Nyquist curves and the differences in response speed for the solutions. The frequency scalings of the Nyquist curves are different.

Only one solution $k = 0.214$, $k_i = 0.0178$ and $b = 0.710$ is obtained for $M_s = 1.4$ with $w_0 = 0.3531$ rad/s. This illustrates that the method indicates that there is no controller that satisfies the specifications.

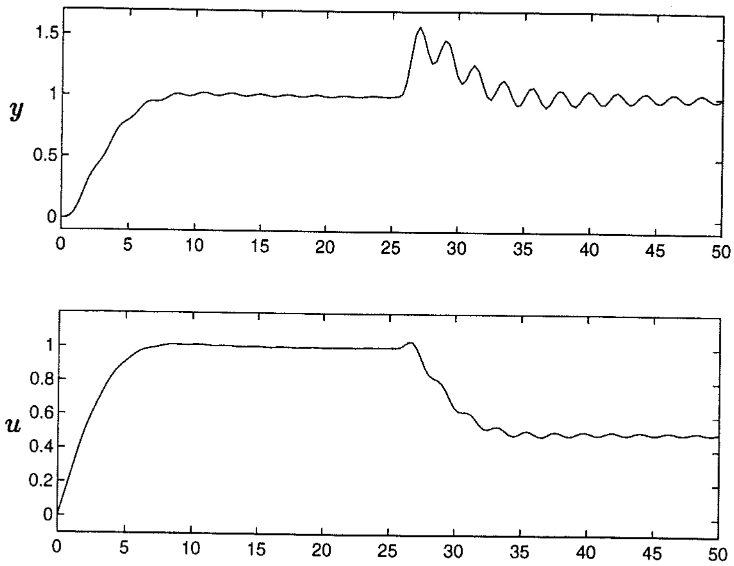


Fig. 7. Time response of the closed-loop system of Example 5 obtained for $a = 0$, when designing the PI controller for $M_s = 1.4$.

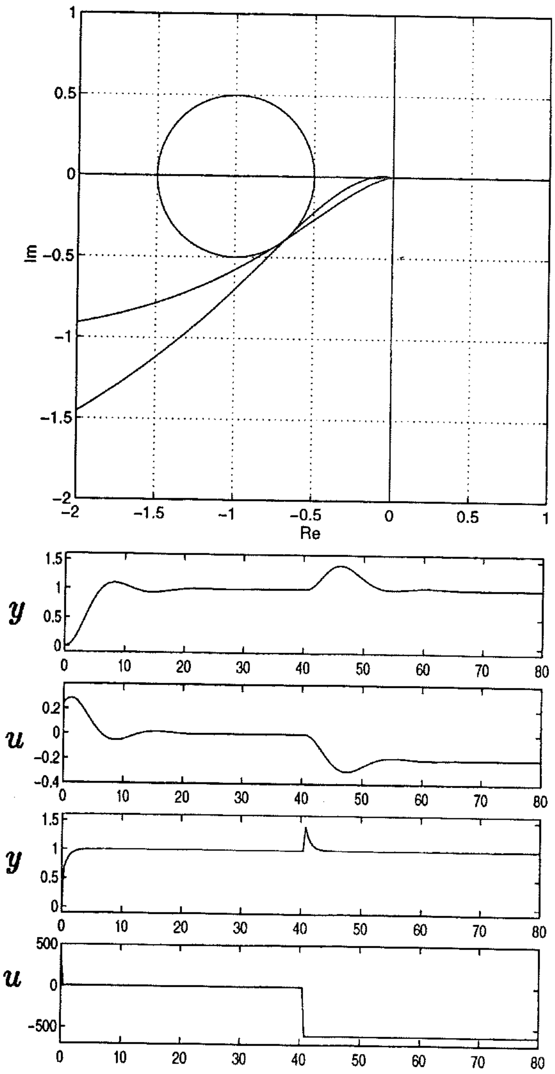


Fig. 8. Nyquist curves of the loop transfer function and time responses for the closed-loop system of Example 6 when designing for $M_s = 2.0$.

Example 7 (*An unstable system*). To show that PI control can also be used for unstable systems we consider a system with the transfer function

$$G_{14}(s) = \frac{a}{(s+a)(s-1)}.$$

This system does not satisfy the monotonicity assumption (20) because its phase does not decrease monotonically. The phase lag is π both at very low and at very high frequencies. The smallest phase lag is

$$\pi\text{-arctan}\,\frac{(a-1)\sqrt{a}}{2a}$$

which occurs for $\omega = \sqrt{a}$. To have a loop transfer function that lies outside a circle with center C and radius R it must be required that

$$\frac{(a-1)\sqrt{a}}{2a} \geq \frac{R}{\sqrt{C^2 - R^2}},$$

which implies that

$$a \geq \frac{(C+R)^2}{C^2 - R^2}.$$

For $C = 1$ and $R = 0.5$ we get $a \geq 3$. To solve the design problem we will therefore start with an interval of frequencies around $\sqrt{a}$. Figure 9 shows the Nyquist curves and the set point and load disturbance responses obtained for $a = 4$ and 8, respectively. The design is made for $M_s = 2.0$. For $a = 4$, the design procedure gives the controller parameters $k = 3.31$, $k_i = 0.82$, $b = 0.50$, and the frequency $w_0 = 3.04$ rad/s. For $a = 8$, we get $k = 8.70$, $k_i = 10.4$, $b = 0.5$, and the frequency $w_0 = 7.85$ rad/s. The M_p-values becomes quite large,

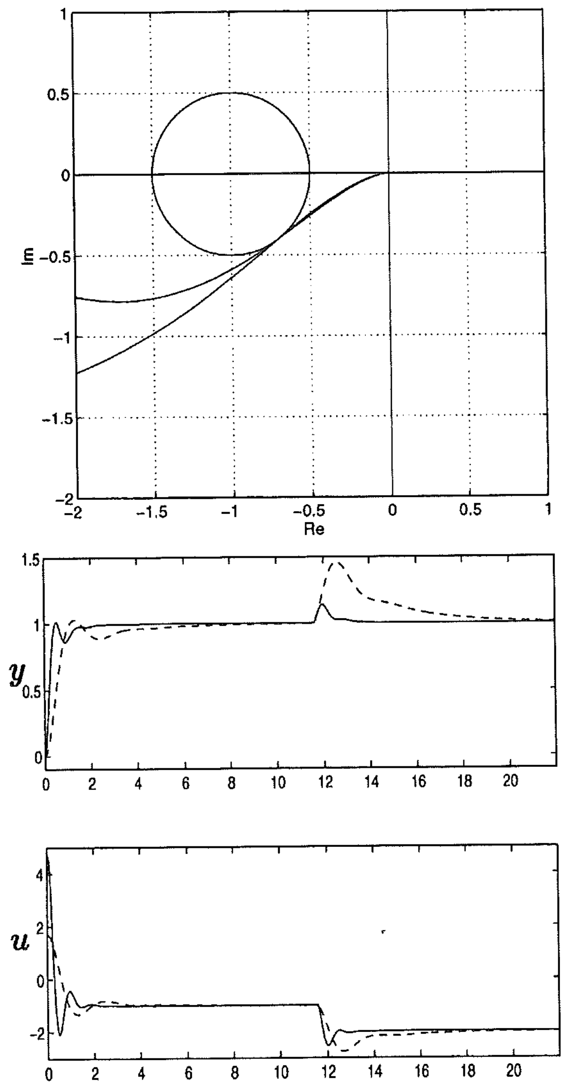


Fig. 9. Nyquist curves of the loop transfer function and time response of the closed-loop system of Example 7 with $a = 4$ and 8. The design is done for $M_s = 2.0$ in both cases.

$M_p = 1.98$ for $a = 4$ and $M_p = 1.87$ for $a = 8$. Still, we obtain set point responses with acceptable overshoots because of the set point weighting b.

Example 8 (*Filtering of measurement noise*). In most cases in PI control it is not a primary design consideration that measurement noise is fed into the control system through the controller. With reasonable sensors it is often sufficient to check the fluctuations of the control signal generated by the noise and introduce a filter on the measured signal if necessary. The design of such a filter was discussed in Section 2 and is illustrated here with the system in Example 1. Measurement noise is particularly difficult in this case because the process has constant gain for high frequencies.

We consider the case of a first-order noise filter with the transfer function

$$G_f(s) = \frac{1}{1 + sT_f}.$$

A PI controller is first designed without considering the noise filter. The design gives the frequency ω_0 from which the parameter T_f is determined as $T_f = 1/(m\omega_0)$. The design is then repeated for a system with the transfer function $G_p G_f$.

To evaluate the effectiveness of the noise filter we observe that the transfer function from measurement noise to control signal is given by

$$U(s) = \frac{G_c G_f}{1 + G_p G_c} N(s) \approx M_s G_c G_f N(s),$$

where N is the measurement noise. The approximation is obtained by replacing the sensitivity function with its maximum M_s. This will overestimate the control signal by at most a factor M_s. Furthermore, assume that the measurement noise is obtained by filtering white noise by a system with the transfer function

$$G_n(s) = \frac{s}{s + a}.$$

The control signal is then white noise filtered by

$$G_u(s) \approx M_s G_c G_f G_n = M_s \frac{ks + k_i}{(s + a)(1 + sT_f)}.$$

Assuming that the spectral density of the white noise is ϕ we find that the variance of the control signal is

$$\sigma_u^2 = M_s^2 \phi \frac{ak^2 + k_i^2 T_f}{2aT_f(1 + aT_f)}.$$

Table 4 shows the results of the PI controllers designed by the method mentioned above for different values of m. It has been computed for $M_s = 2.0$, $\phi = 1$, $a = 2$ and $\zeta = 0.707$. In the table we have also shown the ratio between integration time T_i and the filter time constant T_f. The tradeoffs are clear from the table. Decreasing the value of m makes the system less sensitivity to measurement noise but more sensitive to load disturbances. It is possible to introduce loss functions which capture this compromise but it is rare that the data required is available with sufficient precision to justify the analysis.

Table 4. Parameters of PI controllers with filtering of the process output discussed in Example 8

m	k	k_i	ω_0	IE	σ	T_i/T_f
2	0.31	0.73	1.48	1.36	0.89	1.25
5	0.27	0.78	1.66	1.26	1.25	2.82
10	0.26	0.81	1.74	1.21	1.67	5.51
20	0.26	0.83	1.78	1.17	2.28	11.0
∞	0.26	0.85	1.83	1.12	∞	—

6. COMPARISONS WITH OTHER METHODS

Many different methods have been proposed for tuning PI controllers. A comprehensive presentation which includes many comparisons are given in Åström and Hägglund (1995b). In this section, the new design procedure is compared with two other common methods, namely the Ziegler–Nichols frequency response method, see Ziegler and Nichols (1942), and the Lambda tuning procedure, see Rivera *et al.* (1986).

All methods differ with respect to the process knowledge required and the design specifications. In the Ziegler–Nichols method, the process model is specified in terms of the ultimate gain and the ultimate frequency. No design specification can be made. In Lambda tuning, the process model is specified in terms of the static gain, time constant, and time delay. The design variable is the desired closed-loop time constant. A common rule-of-thumb that is used in this section is that the desired closed-loop time constant should be three times the open-loop time constant, see EnTech (1993). The new design method requires that the transfer function of the process is specified, and the design is given in terms of the M_s-value. The value $M_s = 1.6$ is used in this comparison.

The design procedures have been applied to the processes G_1, G_2, and G_3 given in Section 5. The controller parameters and some performance measures are presented in Table 5.

For process G_1, the Ziegler–Nichols method gives a very oscillatory response with an overshoot of 36% and a high M_s value. This is due to the high controller gain. The Lambda tuning method gives a well damped but sluggish response. The settling time and the *IAE* values are three times larger than those obtained in the new design. This is due to the low controller gain.

For process G_2, the Ziegler–Nichols method gives a control loop that is close to the stability boundary. The overshoot is 47% and the M_s-value is $M_s = 11$. Again this is mainly due to the high controller gain. The Lambda tuning method gives a very sluggish response, were the *IAE* value and the settling time are almost a magnitude larger than in the new design method.

Process G_3 has a long dead time. For this process, the Ziegler–Nichols design results in a controller with too high gain and too long integral time. This gives a control loop with a large M_s value, and a very long settling time. Lambda tuning gives a low controller gain and an integral time that is too short. This results in oscillatory behavior with an overshoot of 21%. The rule of thumb for choosing the desired closed-loop time constant is not suitable for processes with a long dead time. Therefore, it is sometimes suggested to relate the desired closed-loop time constant to the dead time L instead of the open-loop time constant T when the dead time is long.

The Ziegler–Nichols method gives systems which are inherently poorly damped. The examples demonstrate that it is not trivial to choose the desired closed-loop time constant in Lambda tuning. The new design procedure manages to obtain consistent behavior for a wide range of processes. On the other hand, the procedure assumes that the complete transfer function of the process is known, whereas the other two methods only uses simpler models.

Many traditional methods fail to give acceptable control for several of the more difficult control problems discussed in the paper. For example, both the Ziegler–Nichols method and the Lambda tuning method gives controller gain $k = 0$ for the pure delay process $G(s) = e^{-s}$.

7. CONCLUSIONS

This paper describes a design method for PI controllers. The method assumes that the transfer function of the process is given. The specifications capture demands on load disturbance rejection, set point response, measurement noise and model uncertainty. The primary design goal is to obtain good load disturbance responses. This is done by

Table 5. Comparison between the three design procedures. The table shows controller parameters k, T_i, and k_i, and performance measures M_s, M_p, *IAE*, overshoot o, and settling time t_s

Process	Design	k	T_i	k_i	M_s	IAE	o %	t_s	M_p
G_1	Ziegler–Nichols	3.60	3.02	1.19	4.93	1.40	35.9	18.0	4.39
	Lambda tuning	0.278	1.92	0.145	1.17	6.90	0	25.8	1.00
	New design	0.862	1.87	0.461	1.60	2.28	8.80	8.12	1.05
G_2	Ziegler–Nichols	13.6	0.468	29.1	11.4	0.098	47.1	2.28	11.1
	Lambda tuning	0.312	1.05	0.299	1.06	3.35	0.00	14.0	1.00
	New design	2.74	0.672	4.08	1.60	0.246	9.95	1.83	1.27
G_3	Ziegler–Nichols	0.471	30.0	0.0157	1.86	63.2	0.00	237	1.00
	Lambda tuning	0.081	1.73	0.0465	2.15	37.3	21.5	118	1.39
	New design	0.208	5.87	0.0355	1.60	28.2	0.00	51.2	1.00

minimizing the integrated control error IE. Robustness is guaranteed by requiring that the maximum sensitivity is less than a specified value M_s. Good set point response is obtained by using a structure with two degrees of freedom. This requires an extra parameter, the set point weighting b, in the algorithm. The primary design parameter is the maximum sensitivity, M_s, but auxiliary design parameters such as the maximum of the complementary sensitivity, M_p, can be added.

The design problem can be formulated as a constrained optimization problem. It is shown that this problem can be reduced to a solution of nonlinear algebraic equations. Efficient ways of solving these equations are presented. There are unique solutions for special classes of systems but very complicated situations may occur for complicated systems. The method will give a solution if one exist and it will indicate when there is no PI controller that satisfies the specifications.

The design procedure has been applied to a variety of systems; stable and integrating, with short and long dead times, with real and complex poles, and with positive and negative zeros.

REFERENCES

Åström, K. J. and Hägglund (1995a). New tuning methods for PID controllers. *European Control Conference*, Rome, Italy, 2456–2462.

Åström, K. J. and T. Hägglund (1995b). *PID Controllers: Theory, Design, and Tuning*, 2nd edn. Instrument Society of America, 1995, Research Triangle Park, NC.

EnTech (1993). Automatic Controller Dynamic Specification. Number Version 1.0, 11/93. EnTech Control Engineering Inc.

Hazebroek, P. and B. L. van der Waerden (1950). Theoretical considerations on the optimum adjustment of regulators. *Trans. ASME*, **72**, 309–322.

Horowitz, I. M. (1963). *Synthesis of Feedback Systems*. Academic Press, New York.

Khalil, H. K. (1992). *Nonlinear Systems*. MacMillan, New York.

Persson, P. (1992). Towards autonomous PID control. PhD Thesis, ISRN LUTFD2/TFRT-1037-SE, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

Persson, P. and K. J. Åström (1992). Dominant pole design—a unified view of PID controller tuning. Preprints 4th IFAC Symposium on Adaptive Systems in Control and Signal Processing, Grenoble, France, pp. 127–132.

Rivera, D. E., M. Morari and S. Skogestad (1986). Internal model control—4. PID controller design. *Ind. Engng. Chem. Process. Des. Dev.*, **25**, 252–265.

Schei, T. S. (1994). Automatic tuning of PID controllers based on transfer function estimation. *Automatica*, **30**(12), 1983–1989.

Shigemasa, T., Y. Iino and M. Kanda (1987). Two degrees of freedom PID auto-tuning controller. *Proc. ISA Annual Conf.*, 703–711.

Shinskey, F. G. (1990). How good are our controllers in absolute performance and robustness? *Measurement and Control*, **23**, 114–121.

Yamamoto, S. and I. Hashimoto (1991). Present status and future needs: the view from Japanese industry. In Arkun and Ray, Eds., *Chemical Process Control—CPCIV*. *Proc. 4th Inter. Conf. on Chemical Process Control*, TX.

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

Erratalist for "Design of PI Controllers based on Non-Convex Optimization"

p. 587 Equation (7): The expressions C and R should be,

$$C = \frac{M_s(2M_p - 1) - (M_p - 1)}{2M_s(M_p - 1)},$$
$$R = \frac{M_s + M_p - 1}{2M_s(M_p - 1)}.$$

p. 589 Equation (10): Equation (10) should be labelled out.

p. 597 Figure 6: The time response in Figure 6 is wrong. The correct one is as follows,

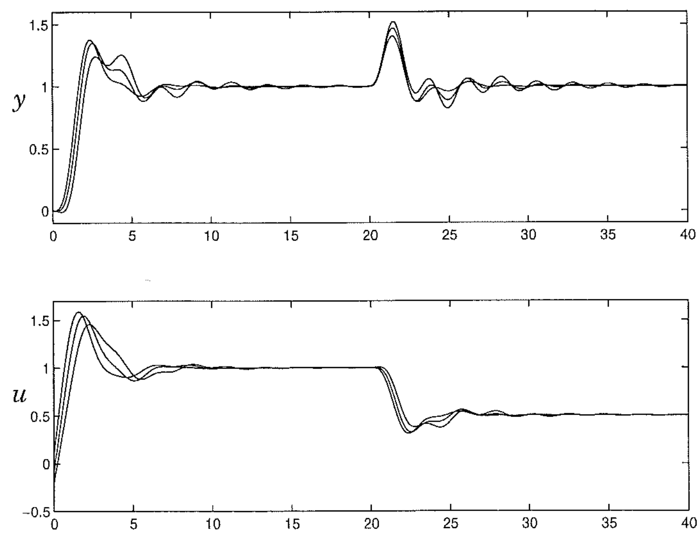


Figure 1 The time responses for Example 5 with $a = 0.2,\ 0.5,\ 1.0$, when designing for $M_s = 2.0$.

Paper II

Design of PID Controllers based on Constrained Optimization

H. Panagopoulos, K. J. Åström, and T. Hägglund

Department of Automatic Control
Lund Institute of Technology
Box 118
S-221 00 Lund, Sweden
Phone: +46-46 2228789, Fax: +46-46 138118
Email: hp@control.lth.se

Abstract This paper presents a new design method for PID controllers based on optimization of load disturbance rejection with constraints on robustness to model uncertainties. The design also delivers parameters to deal with measurement noise and set point response. Thus, the formulation of the design problem captures four essential aspects of industrial control problems, leading to a constrained optimization problem which can be solved iteratively. A straightforward way to find initial conditions is also presented.

Keywords PID control. Design. Optimization. Specifications. Load disturbance rejection. Measurement noise filtering. Set point response. Robustness. Sensitivity.

1. Introduction

The PID controller is todays most commonly used control algorithm, see Yamamoto and Hashimoto (1991). At the moment, there exits many different methods to find suitable controller parameters. The methods differ in complexity, flexibility, and in the amount of process knowledge used. Depending on the application, there is a need to have several types of tuning methods. There are simple, easy to use methods which require little information, e.g. the method described in Ziegler and Nichols (1942), and on the other hand more sophisticated methods which requires more information and more computations. But with todays computational capacity this is no disadvantage.

There are several reasons to look for better methods to design PID controllers. One is the significant impact it may give because of the widespread use of the controllers. Another is the benefit emerging auto-tuners and tuning devices can derive from improved design methods.

There are several requirements on an efficient design method. It should be applicable to a wide range of systems and it should have the possibility to introduce specifications that capture the essence of real control problems. Furthermore, the method should be robust in the sense that it provides controller parameters, if they exist, or if the specifications can not be met, an appropriate diagnosis should be presented. We believe these requirements are satisfied by the method presented in this paper. The approach gives a simple means of solving both simple control problems and more difficult ones. The method provides a PID controller which satisfies the specifications, provided that such a controller exists. If not, the reasons for failure will be indicated. It can also be used to develop simpler methods for restricted classes of systems, as was done in Åström and Hägglund (1995a).

This paper describes a new design method for PID controllers based on the assumption that the process transfer function is known. The primary design goal is to obtain good load disturbance responses. This is done by minimizing the integrated control error IE. Robustness is guaranteed by requiring that the maximum sensitivity be less than a specified value M_s. Measurement noise is dealt with by filtering. Good set point response is obtained by using a structure with two degrees of freedom. Such response can often be achieved by set point weighting, see Åström and Hägglund (1995b), p. 73–76. Filtering can be used when this is not sufficient. Consequently the presented method captures demands on load disturbance rejection, set point response, measurement noise and model uncertainty. The design method for PID controllers is based on earlier work of designing PI controllers presented in Åström *et al.* (1998).

The PID controller we discuss has four primary parameters: controller gain k, integral time T_i, derivative time T_d and set point weight b. In addition there is filtering of the measured signal and sometimes also of the set point. The design method gives all parameters required.

The specifications are expressed in terms of a number of parameters for which good default values can be found. In the simplest case good default values can be given to all parameters. The primary design parameter is the maximum sensitivity, M_s; auxiliary design parameters are the maximum of the complementary sensitivity, M_p, and the largest magnitudes of the transfer functions from set point to output M_{sp} and from measurement noise to control signal M_n.

The user simply supplies the process transfer function and the method provides all the parameters of the PID controller, controller gain k, integral time T_i, derivative time T_d and set point weight b. In addition the filters of the measured signal and the set point are delivered. A natural next step is to modify the maximum sensitivity M_s. This gives a flexible way of modifying the main characteristics of the system. If the user so desires the specifications can gradually be modified for fine tuning.

The primary design goal, i.e. ensuring good rejection of load disturbances, can be formulated as constrained optimization of the integral gain k_i. Constraints are introduced to ensure robustness, which can be encapsulated in a constraint that the loop transfer function should avoid a circle in the Nyquist diagram. This establishes a nice connection between traditional design of PID controllers and $\mathcal{H}_\infty$ control, see Panagopoulos and Åström (1998). Simple measures of the robustness are the maximum sensitivity, the maximum complementary sensitivity alternately the $\mathcal{H}_\infty$-norm of the multi variable transfer function of the system, see Vinnicombe (1998).

A design method for PI controllers based on maximization of integral gain subject to a robustness constraint was developed in Åström *et al.* (1998). In the present paper it is shown that this method cannot be extended directly to PID control. The reason for this is that the optimization problem in most cases has ridges which result in poor robustness and thus also poor control.

Having developed an understanding for the problem, it is possible to introduce additional constraints. The result is that design of PID controllers can be formulated as a constrained optimization problem which can be solved iteratively. Initial conditions are very important since the problem is non convex. A good way to find initial conditions is also presented.

The solution of the optimization problem gives a PID controller with a pure derivative,thus providing a very flexible way of choosing the filter. Simple rules for choosing a filter for the measured signal are presented. Adjusted controller parameters are obtained simply by repeating the design with the process replaced by a combination of the original process transfer function and the transfer function of the chosen filter.

Having created a controller able to deal with load disturbances, measurement noise and plant uncertainty, the system is then designed for good set point response. First the set point weighting must be determined so that the maximum of the transfer function from the set point to the output will be less than a given value. Sometimes this cannot be accomplished and a filter for the set point must be designed. The advantage of the approach is that a filter is only introduced if necessary.

2. Problem Formulation

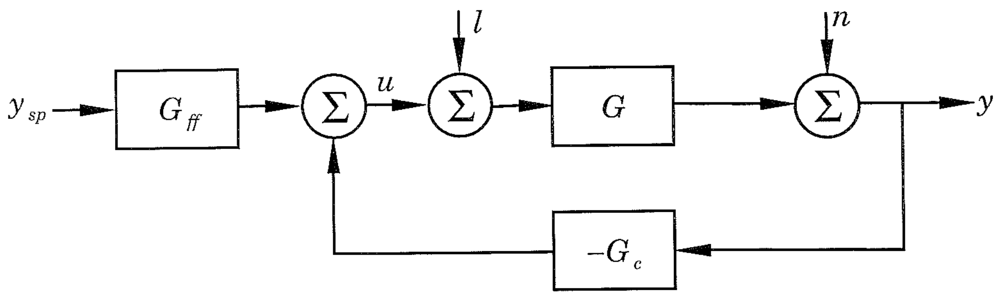


Figure 1 Block diagram describing the design problem.

The design problem is illustrated in Figure 1. A process with transfer function $G(s)$ is controlled with a PID controller with two degrees of freedom. Transfer function $G_c(s)$ describes the feedback from process output y to control signal u, and $G_{ff}(s)$ describes the feed forward from set point y_{sp} to u. Three external signals act on the control loop, namely set point y_{sp}, load disturbance l and measurement noise n.

The design objective is to determine the controller parameters in $G_c(s)$ and $G_{ff}(s)$ so that the system behaves well with respect to changes in the three signals y_{sp}, l and n as well as in the process model $G(s)$. Therefore, the specification will express requirements on

- Load disturbance response
- Measurement noise response
- Set point response
- Robustness with respect to model uncertainties

Process and controller structures

The design problem was formulated so as to apply to a wide variety of systems. Thus it is assumed that the process is given as well as the properties of the input signals of the system. This is very useful because it admits design of controllers for a very wide range of systems. Later it is seen that only certain properties of the processes are needed for the design. Consequently, the process is assumed to be linear, time invariant, and specified by a transfer function $G(s)$, which is analytic with finite poles and, possibly, an essential singularity at infinity. The description covers finite dimensional systems with time delays and infinite dimensional systems described by linear partial differential equations.

Initially it is assumed that the controller can be described by

$$u(t) = k(by_{sp}(t) - y(t)) + k_i \int_0^t (y_{sp}(\tau) - y(\tau))d\tau + k_d(-\frac{dy(t)}{dt}), \tag{1}$$

where k, k_i, k_d and b are controller parameters.

It proved beneficial to replace the signals y and y_{sp} with their filtered values y^f and y^f_{sp}. The filtered signals are generated by

$$\begin{aligned} Y^f(s) &= F_y(s)Y(s), \\ Y^f_{sp}(s) &= F_{sp}(s)Y_{sp}(s), \end{aligned} \tag{2}$$

where the filters are low pass filters of first or second order, for example

$$F_y(s) = \frac{1}{(1+sT_f)^n} \tag{3}$$

with $n = 1$ or $n = 2$.

The controller can thus be characterized by either four parameters k, k_i, k_d, b, and two filters F_y, F_{sp} or by two transfer functions: the feedback G_c and the feed forward G_{ff} transfer functions where

$$\begin{aligned} G_c(s) &= (k + \frac{k_i}{s} + k_d s)F_y(s), \\ G_{ff}(s) &= (bk + \frac{k_i}{s})F_{sp}(s). \end{aligned}$$

Consequently, the relations among the three external input signals, y_{sp}, l and n, on the one hand, and process output y and control signal u, on the other, become

$$y = \frac{G}{1+GG_c}l + \frac{1}{1+GG_c}n + \frac{GG_{ff}}{1+GG_c}y_{sp}, \tag{4}$$

$$u = -\frac{GG_c}{1+GG_c}l - \frac{G_c}{1+GG_c}n + \frac{G_{ff}}{1+GG_c}y_{sp}. \tag{5}$$

Load Disturbance Attenuation

The primary design goal is to achieve good rejection of load disturbances where no detailed assumptions are made about the load disturbances except that they are low frequent. Assuming that the transfer function of the process does not go to zero for small s, the transfer function between l and y, see Equation (4), can then be approximated according to

$$\frac{G}{1+GG_c} \approx \frac{s}{k_i}$$

for small s. Thus, by maximizing the integral gain k_i the effect of the load disturbance l on the output y is minimized. According to Åström *et al.* (1998) it is shown that this is equivalent to minimizing the integrated error (IE) for a step change in the load disturbance.

Measurement Noise

Assuming that the process has such a high roll off that GG_c goes to zero for large s, the transfer function between n and u, see Equation (5), can be approximated according to

$$\frac{G_c}{1+GG_c} \approx k + k_d s$$

for large s.

This shows that the transfer function from measurement noise to controller output goes to infinity for high frequencies. This can be avoided by filtering the process output.

In traditional PID controllers the filters are often chosen in an ad hoc manner. Typically the filters are only applied on the derivative term. A common choice is to give the derivative term

$$D(s) = -\frac{kT_d s}{1+s\frac{T_d}{N}}Y(s) = -\frac{k_d s}{1+s\frac{k_d}{kN}}Y(s),$$

where N is a number in the range 2-10. This will reduce the high frequency gain to $k(1+N)$. Since the filter is only applied on the derivative term and not on the proportional term, the high frequency gain can never be made smaller than k.

In this study all terms in the controller were filtered. For first order filter with filter time constant T_f, the high frequency gain becomes k_d/T_f. For second order filters the high frequency gain goes to zero as the frequency goes to infinity. Therefore it seems reasonable to use a second order filter.

The largest gain from measurement noise to controller output

$$M_n = \max_{\omega} \left| \frac{G_c(i\omega)}{1+GG_c(i\omega)} \right| \tag{6}$$

was used as a simple quantitative measure of the noise rejection.

Set Point Response

The transfer function relating set point to process output is given by

$$G_{sp}(s) = \frac{GG_{ff}}{1+GG_c} = \frac{k_i + bks}{k_i + ks + k_d s^2} \frac{GG_c}{1+GG_c} \frac{F_{sp}}{F_y}, \tag{7}$$

see Equation (4). When the controller has been designed to give good attenuation of disturbances, the parameter b and the filter F_{sp} can be chosen to give an appropriate set point response. The maximum of G_{sp} is chosen as a simple criterion, i.e.

$$M_{sp} = \max_{\omega} |G_{sp}(i\omega)|. \tag{8}$$

Robustness

Sensitivity to modeling errors can be expressed in terms of the largest value of the sensitivity function, i.e.

$$M_s = \max_{\omega} \left| \frac{1}{1+GG_c(i\omega)} \right|. \tag{9}$$

Keep in mind that the quantity M_s is simply the inverse of the shortest distance from the Nyquist curve of the loop transfer function to the critical point -1. Typical values of M_s are in the range 1 to 2.

The sensitivity can also be expressed by the largest value of the complementary sensitivity function, i.e.

$$M_p = \max_{\omega} \left| \frac{GG_c(i\omega)}{1+GG_c(i\omega)} \right|. \tag{10}$$

Typical values of M_p are in the range 1.0 to 1.5.

Another possibility is to use the $\mathcal{H}_\infty$-norm

$$\gamma = \max_{\omega} \left| \frac{1+|GG_c(i\omega)|}{1+GG_c(i\omega)} \right|, \tag{11}$$

which is discussed in detail in Panagopoulos and Åström (1998).

In Åström *et al.* (1998) it was shown that the condition that M_s and M_p are sufficiently small can be guaranteed by the condition that the Nyquist curve of the loop transfer function should be outside a circle with center at $s = -C$ and radius R. Such a condition can be expressed by the inequality $f(k, k_i, k_d, \omega) \geq R^2$, where

$$f(k, k_i, k_d, \omega) = \left| C + \left(k - \frac{i}{\omega}(k_i - \omega^2 k_d)\right) G(i\omega) \right|^2. \tag{12}$$

In Panagopoulos and Åström (1998) it is shown that the condition on the $\mathcal{H}_\infty$-norm (11) can be expressed by a similar condition.

Tuning Parameter

The tradeoff between performance and robustness varies among different control problems. Therefore, it is desirable to have a design parameter to change the properties of the closed-loop system. Ideally, the parameter should be directly related to the performance of the system; it should not be process oriented. There should be good default values so a user is not forced to select a value. This is of special importance when the design procedure is used for automatic tuning. The design parameter should also have a good physical interpretation and natural limits to simplify its adjustment.

For the proposed design method the robustness constraint is a good measure of the performance of the system. It has been shown in Åström *et al.* (1998) that the variable M_s fulfills all the requirements of a good design parameter. Another choice is to use both M_s and M_p as a design parameter, see Åström *et al.* (1998).

The advantages of the chosen tuning parameter is that it gives a good way to determine the performance of the closed loop system. At the same time it is directly related to the robustness and stability. Furthermore, the chosen tuning parameter is dimension free which makes it suitable for automatic tuning.

The disadvantage of the chosen design variable may be that it is dimension free, compared with the Lambda tuning procedure, see Rivera *et al.* (1986), which has dimension time. It might be an advantage to have a dimension on the tuning parameter as in the special case of blending. But in general it is not necessary, then a tuning chart presenting the overall behavior of the closed loop system for different values of the tuning parameter is enough, see Figure 2. Here the time response to changes in set point and load for different values of M_s as a design variable has been presented. According to Figure 2 the expected responses to changes in load

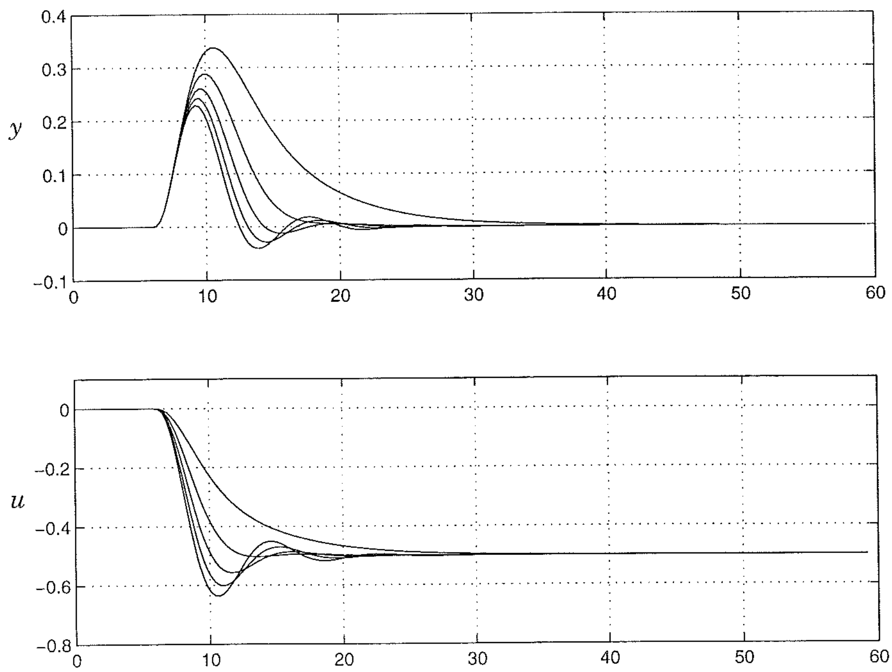


Figure 2 Tuning chart showing the response to a load disturbance for the tuning parameter M_s=1.2, 1.4, 1.6, 1.8, 2.0.

will be transparent to the user for each values of the design variable.

3. The Design Procedure

According to the previous section design of a PID controller can be made using the following procedure:

> **Step 1:** Find controller parameters k, k_i and k_d which maximize k_i subject to the constraints that the closed loop system is stable and the constraints on sensitivity expressed by M_s, M_p or some other norm.
>
> **Step 2:** Determine the filter F_y and repeat step 1, iterate if necessary.
>
> **Step 3:** Determine parameter b and the filter F_{sp} so that M_{sp} is less than a specified value.

The details of the procedure will be discussed in Section 5.

4. A Difficulty

The design problem discussed in the previous section can be formulated as an optimization problem: Find controller parameters maximizing k_i subject to the constraints that the closed loop system is stable and the Nyquist curve of the loop transfer function is outside a circle with center at $s = -C$ and radius R. Another way to express the optimization problem is

$$\begin{array}{ll} \max & k_i \\ \text{such that} & f(k, k_i, k_d, \omega) \geq R^2 \end{array} \tag{13}$$

where f is defined in Equation (12). This formulation has been shown to work very well for PI controllers, see Åström *et al.* (1998), but not for PID controllers where difficulty arises.

The constraint (12) conceptually defines k_i as a function of k and k_d. Thus, the design problem is simply to maximize the function $k_i(k, k_d)$. The difficulty is that in most cases the function will have a discontinuous derivative. This is illustrated in Figure 3 which shows a graphical illustration of the function. Consider for example the contours for constant k_d.

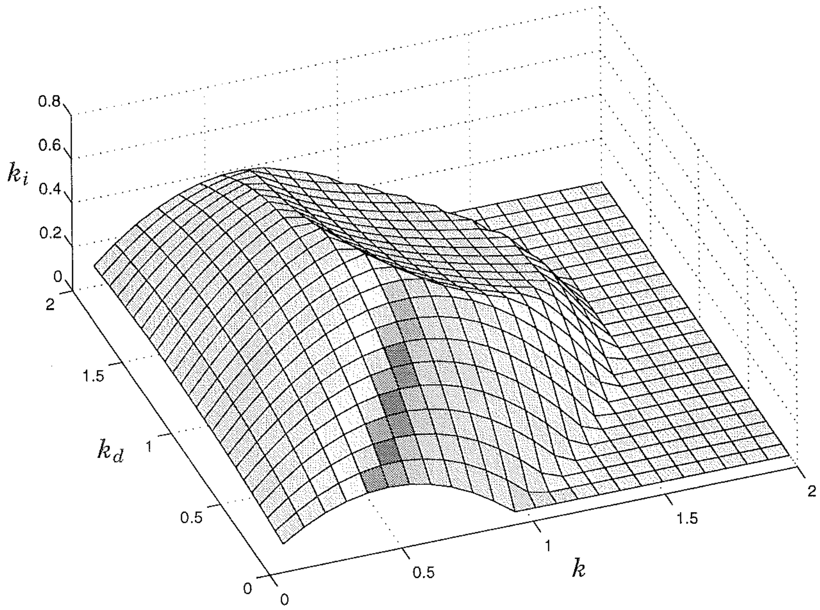


Figure 3 Geometric illustration of the sensitivity constraint.

For small values of k_d the curves are smooth with a regular optimum. For larger values of k_d the curves have a discontinuous derivative at the peak. The discontinuous derivative causes problems in optimization but more seriously it implies that the optimal solution is very sensitive to changes in the controller parameters, because small changes in k and k_d may give large changes in k_i. In the following these difficulties will be more closely investigated. The insight gained is used to introduce additional constraints for the optimization problem.

Geometric Interpretation of the Sensitivity Constraint

The sensitivity constraint (12) has a nice geometric interpretation which will be exploited to gain some insight into the problem. Introduce

$$G(i\omega) = r(\omega)e^{i\varphi(\omega)} = \alpha(\omega) + i\beta(\omega).$$

The sensitivity constraint (12) can then be written as

$$C^2 + 2C\alpha(\omega)k + 2C\frac{\beta(\omega)}{\omega}(k_i - \omega^2 k_d) + r^2(\omega)k^2 + \frac{r^2(\omega)}{\omega^2}(k_i - \omega^2 k_d)^2 \geq R^2. \tag{14}$$

In the following, the argument ω in α, β, r, and φ will be dropped in order to simplify the writing. Then Equation (14) can be written as

$$\frac{r^2}{R^2}\Big(k + \frac{\alpha C}{r^2}\Big)^2 + \frac{r^2}{\omega^2 R^2}\Big(k_i - \omega^2 k_d + \frac{\omega\beta C}{r^2}\Big)^2 \geq 1, \tag{15}$$

which is for fixed ω the exterior of a cylinder in R^3. It will be shown that the intersection of the cylinder with the k-k_i-plane is an ellipse with axes parallel to the coordinate axes. When ω sweeps over the range $0 \leq \omega < \infty$ the ellipsoidal cylinders generate a volume that defines the boundaries of the sets of parameters which satisfy the sensitivity constraints. It will be shown that the boundary of this set may have discontinuous derivatives even for very simple processes. Recall that in Åström *et al.* (1998) it was shown that a similar situation may also occur for PI control. However in that case it was not so common.

An Example

Now it will be shown that the boundaries of the sets of parameters which satisfy the sensitivity constraint has discontinuous gradients even in the simple case when the process has the transfer function

$$G(s) = \frac{1}{(s+1)^4}. \tag{16}$$

Consider the case of constant ω and k_d, then the constraint (15) is an ellipse in the k-k_i plane. As ω changes the ellipses form an envelope. Since it is quite complicated to compute the envelopes they are approximated by the envelopes of the ellipses vertices. It follows from Equation (15) that the horizontal vertices are given by

$$\begin{aligned} k &= -C\frac{\alpha}{r^2} \pm \frac{R}{r}, \\ k_i &= -C\frac{\omega b}{r^2} + \omega^2 k_d, \end{aligned}$$

where the left vertex corresponds to a minus sign and the right vertex to a plus sign. The vertical vertices are given by

$$\begin{aligned} k &= -C\frac{\alpha}{r^2}, \\ k_i &= -C\frac{\omega b}{r^2} + \omega^2 k_d \pm \frac{\omega R}{r}. \end{aligned}$$

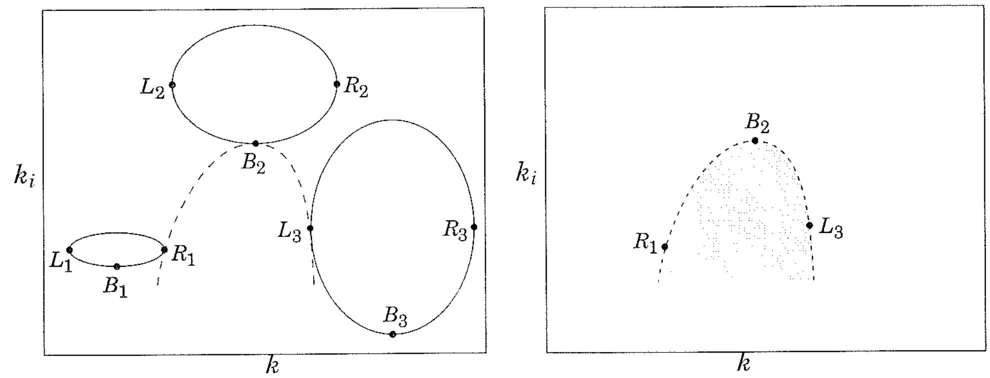
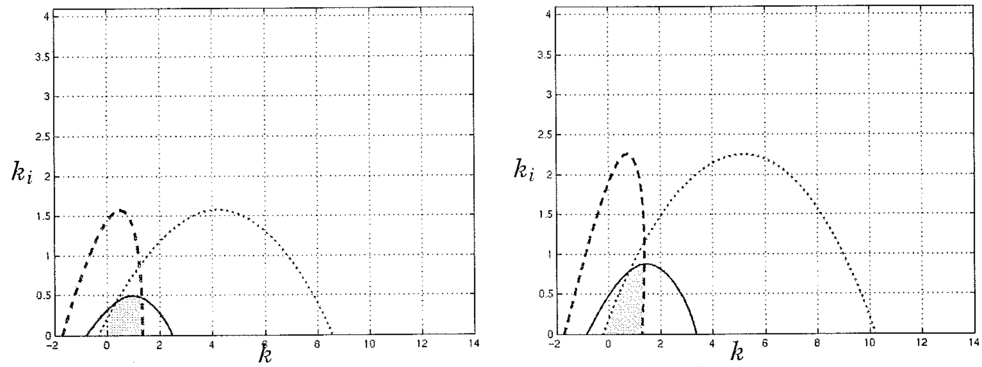
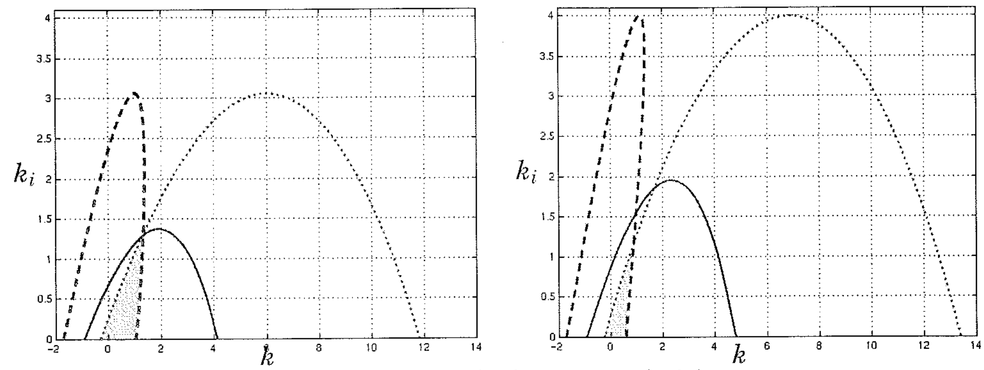


Figure 4 How to derive the set of (k, k_i)-parameters which satisfy the sensitivity constraint.



Systems $k_d = 1$ (left) and $k_d = 2$ (right)



Systems $k_d = 3$ (left) and $k_d = 4$ (right)

Figure 5 The set of (k, k_i)-parameters which satisfy the sensitivity constraint for $k_d = 1,\ 2,\ 3,\ 4$ with $G(s) = 1/(s+1)^4$ for $M_s = 1.4$.

How to derive the approximated envelopes of the ellipses is explained in Figure 4. To begin with consider the left figure, where the loci of the left (L_i), right (R_i), and bottom (B_i) vertices, have been plotted respectively, for fixed values of k_d. When ω is changed a new set of vertices $(L_{i+1}, R_{i+1}, B_{i+1})$ are given. Finally, when enough sets of vertices (L_i, R_i, B_i) have been plotted, they will form the envelopes of the ellipses vertices in the k-k_i-plane. Consequently, it will be possible to derive the set of (k, k_i)-parameters which satisfy the sensitivity constraint of Equation (15). In the left figure of Figure 4 it corresponds to the shaded area.

Table 1 The obtained controller parameters when maximizing k_i for fixed k_d.

k_d	k	k_i	ω_1	ω_2	IAE
1.0	0.96	0.5	0.59	-	2.82
1.5	1.21	0.97	0.51	1.37	2.52
2.77	1.1	0.98	0.47	1.46	2.72
2.897	1.05	0.99	0.46	1.42	2.83

In Figure 5 the envelopes of the ellipses vertices have been plotted for $k_d = 1, 2, 3, 4$ with the process in Equation (16) when $M_s = 1.4$. The shaded areas correspond to the set of (k, k_i)-parameter which satisfy the sensitivity constraint. The following conclusions are made: for small values of k_d the maximum of k_i occurs on one curve. As k_d increases the maximum occurs at a corner where two curves intersect. The corner becomes sharper as the value of k_d increases. This is clearly seen in the curve for $k_d = 4$. The presence of the corner explains the difficulties encountered in optimization of k_i, see for example Persson (1992).

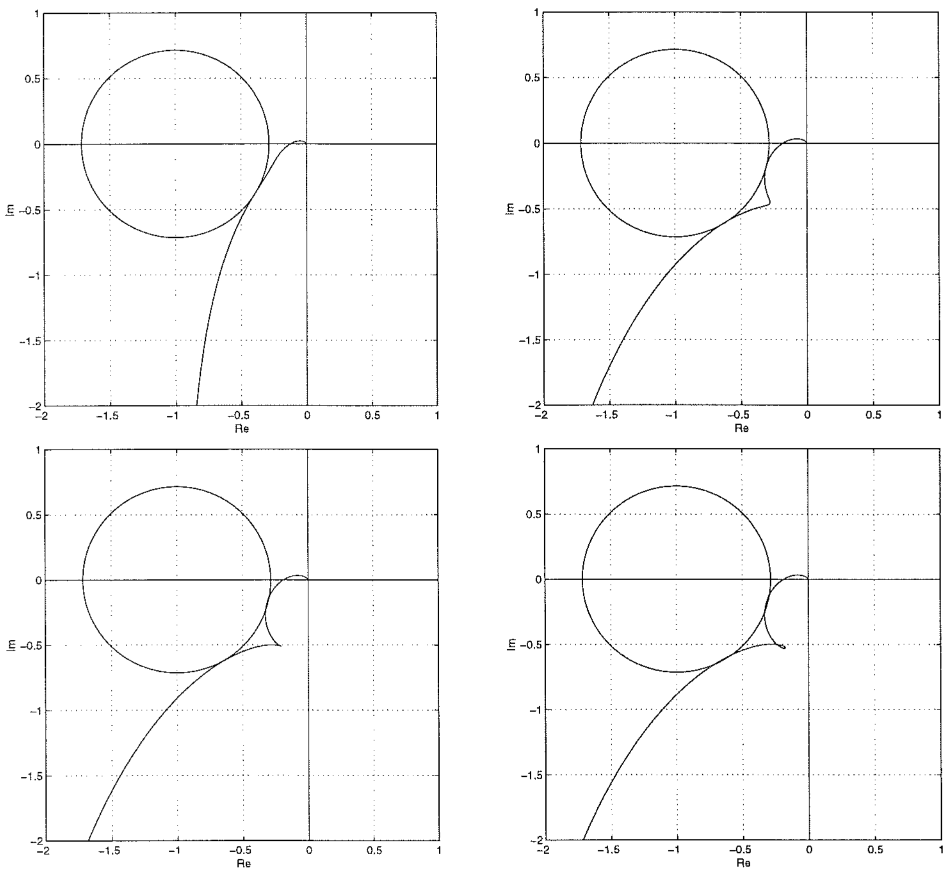


Figure 6 The Nyquist curves of loop transfer functions for designs with $k_d = 1$ (upper left), 1.5 (upper right), 2.77 (lower left) and 2.897 (lower right).

Bode and Nyquist Diagrams More insight into the problem can be obtained by investigating the Nyquist curves of the loop transfer functions

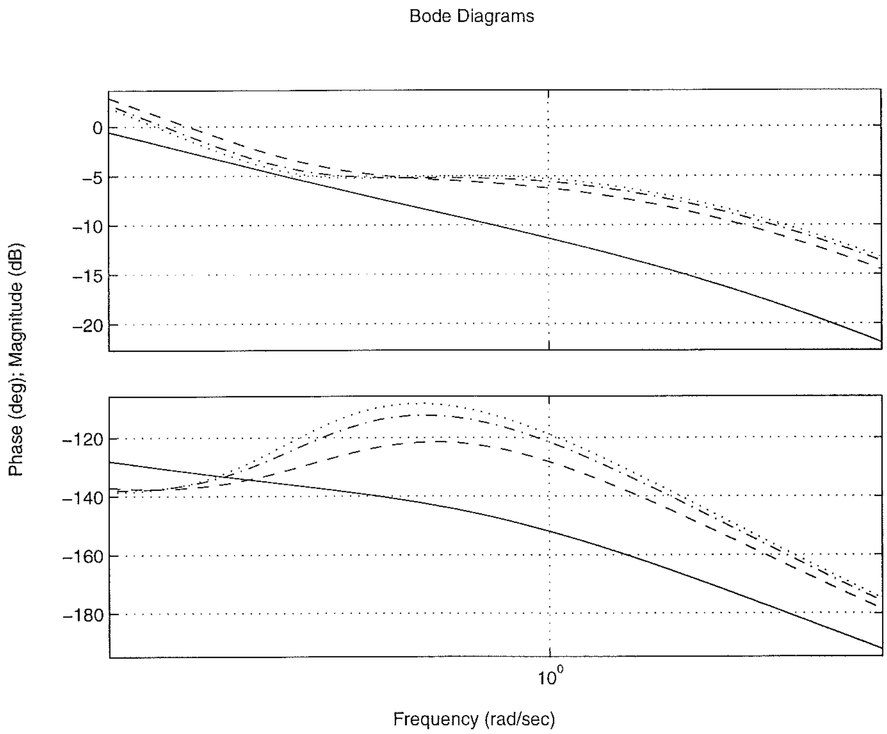


Figure 7 The Bode diagrams of the loop transfer functions of the designs with $k_d = 1$ (full line), 1.5 (dashed line), 2.77 (dotted line) and 2.897 (dashed dotted line).

for systems with different controller parameters. Again the system with the transfer function (16) will be considered. The parameter k_d will be fixed and controllers that maximizes k_i subject to the sensitivity constraint will be determined, by using essentially the same code when designing PI controllers. The design parameter is chosen as $M_s = 1.4$.

Calculations are performed for $k_d = 1$, 1.5, 2.77 and 2.897. The values obtained are summarized in Table 1.

The results are shown in Figure 6. For small values of k_d the loop transfer function has one point tangent to the sensitivity circle. As k_d increases two points will be tangent to the circle. As k_d is increased even further the Nyquist curve will actually have a cusp. The curve with a cusp has the largest value of k_i which should be the preferred solution if k_i was the only concern.

Another illustration of the observed phenomena in Figure 6 is given in Figure 7 which shows the Bode diagrams of the loop transfer functions. The figure indicates that an increase of k_d increases the phase advance which makes it possible to increase k_i a little. The increase of k_i is however marginal and it may be questioned if the designs with the larger values of k_d are reasonable.

Time Responses Additional insight into the problem is obtained by investigating the time responses. Figure 8 shows the responses to load disturbances for the different systems and in Table 1 details of the simulation are summarized.

According to Figure 8 the elimination of a load disturbance is best for Nyquist curves with a cusp but not with any remarkable amelioration of the control performance given by the integrated error IAE in Table 1. Thus, to have an overall good performance it is preferable to choose the controller for which the Nyquist curve has no cusp.

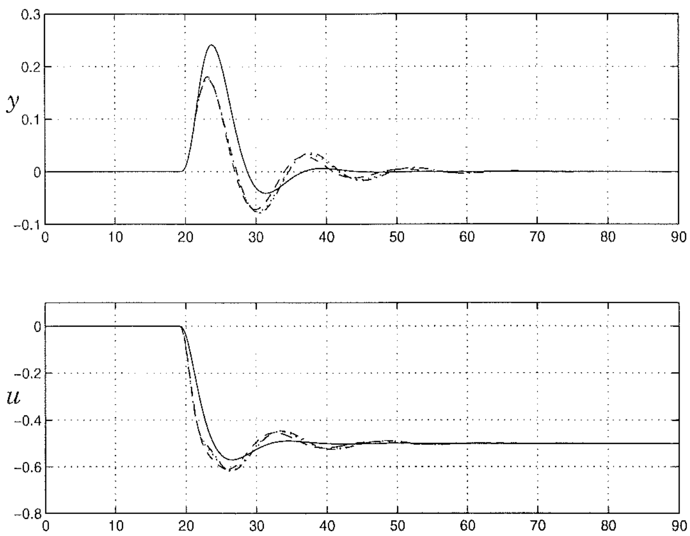


Figure 8 Comparing load responses for $k_d = 1$ (full line), 1.5 (dashed line), 2.77 (dotted line) and 2.897 (dashed dotted line).

Conclusion

This section has shown that a direct generalization of the method used to design PI controllers in Åström *et al.* (1998) where k_i is maximized subject to a sensitivity constraint is not a good formulation of the design problem, which can be explained geometrically. The sensitivity constraint can be represented as manifold in parameter space. This manifold has unfortunately not a smooth tangent even for the simple well behaved system discussed in the section. One consequence of the discontinuous tangents is that the optimal solution may occur at a corner which makes it quite sensitive to parameter variations.

Also, the consequences in the frequency domain have been investigated to show that the Nyquist curve of the loop transfer function may have a cusp. This cusp is associated with a significant increase of the phase advance. This type of phenomena was also encountered for design of PI controller for systems with resonant poles. However, for most systems the problem does not occur for PI control.

To have a good method for PID control the design problem must therefore be reformulated, which will be done in the following section.

5. PID Design

From the insights of the previous section, it is now possible to formulate at good method for designing PID controllers. It has been shown that only the sensitivity constraint in the formulation of the constrained optimization problem in Equation (13) is not enough to obtain a design which is insensitive to parameter variations. Thus, more constraints are needed to obtain the desired shape of the Nyquist curve of the loop transfer function in the upper left figure of Figure 6. A good method for designing PID controllers is obtained if the following two constraints are added in the design formulation of Equation (13), i.e.

$$\begin{aligned} \text{max} \quad & k_i \\ \text{such that} \quad & f \geq R^2, \\ & \kappa < 0, \\ & \delta < 0, \end{aligned} \tag{17}$$

where κ is the curvature of the loop transfer frequency function, $L(i\omega)$, and δ is the difference in phase change of $L(i\omega)$ at two consecutive frequency points. Consequently, the first constraint in Equation (17) express the sensitivity condition, the second constraint specifies that a negative curvature of $L(i\omega)$ should be obtained and the third constraint prevents $L(i\omega)$ to have undesirable phase leads.

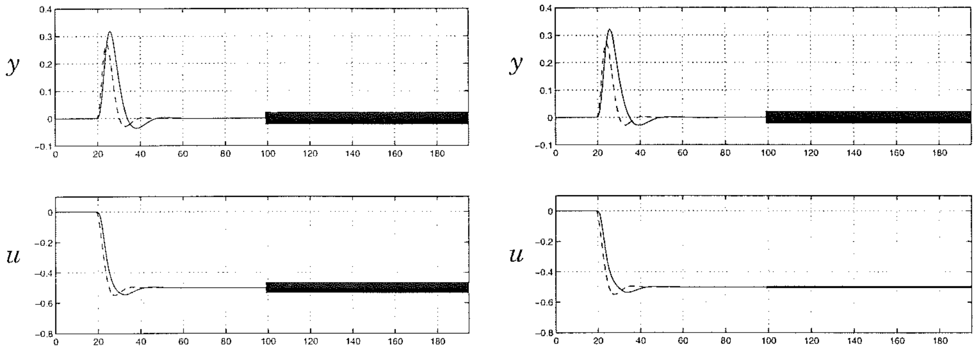
For systems with integral action and close to integral action, i.e. a pole relatively close to zero, the second constraint in Equation (17) is, however, too severe. Omitting it will in these cases be appropriate. Consequently, the design of PID controllers is separated into two cases depending on the considered system. This separation between integrating and non integrating processes is done in several previous design methods, see Åström and Hägglund (1995b).

Measurement noise filtering

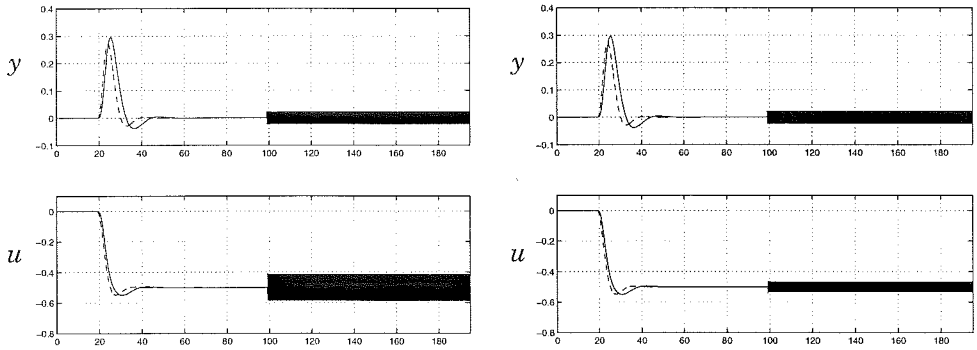
When there are substantial measurement noise, it is customary to filter the measurement signal with a filter given by Equation (3), see the discussion in Section 2. The choice of filter time constant T_f in Equation (3) is a tradeoff between filtering capacity and loss of performance. A large value of T_f provides an effective noise filtering, but it will also change the control performance, on the contrary to a small value of T_f for which control performance is retained, but with less efficient noise filtering. When comparing different kinds of filters the effectiveness of the noise filtering measured as M_n, defined in Equation (6), is to be compared with the control performance measured as the integrated absolute error (IAE), defined in Åström and Hägglund (1995b).

A nice feature of the new design procedure is to provide a systematic way to determine T_f. The choice

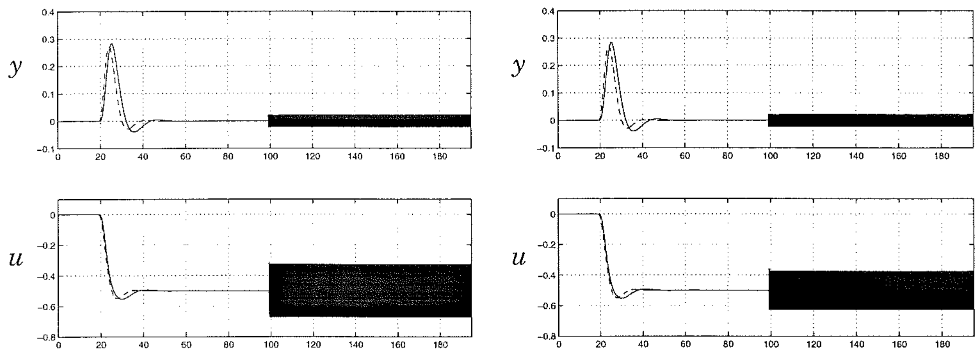
$$T_f = \begin{cases} \dfrac{1}{N\omega_0} & \text{for first order filter,} \\ \dfrac{1}{2N\omega_0} & \text{for second order filter,} \end{cases} \tag{18}$$



Filter constant $N = 2$.



Filter constant $N = 5$.



Filter constant $N = 10$.

Figure 9 The response to a load disturbance followed by measurement noise using first (left) and second (right) order filters with different values of N. The graphs show responses from F_yG (full line) and load disturbance responses from G (dashed line).

makes it possible to determine the effects of the filter at the frequency ω_0, where ω_0 is the frequency at which the sensitivity function is maximal. Reasonable values of N are in the interval 2–10. In Table 2 the amount of modification by the filter on the loop transfer function at the critical frequency ω_0 is displayed by calculating the $arg F_y(i\omega_0)$. Because of the special choice of T_f the first and second order filter has the same amount of modification on the loop transfer function at ω_0.

Inserting a filter modifies the loop transfer function which gives minor changes in control loop performance. Consequently, adjusted controller pa-

Table 2 Properties of the PID controllers obtained for system $F_y G$ with $M_s = 1.4$ and different orders of filter n.

n	1	1	1	2	2	2
N	2	5	10	2	5	10
$arg F_y(i\omega_0)$	$-26.6°$	$-11.3°$	$-5.71°$	$-28.1°$	$-11.4°$	$-5.72°$
T_f/T_d	0.4432	0.1933	0.0988	0.2053	0.0972	0.0495
k_i	0.25	0.30	0.32	0.23	0.30	0.32
IAE	5.05	4.41	4.12	5.20	4.44	4.13
M_n	1.59	3.75	6.82	1.62	3.89	7.84

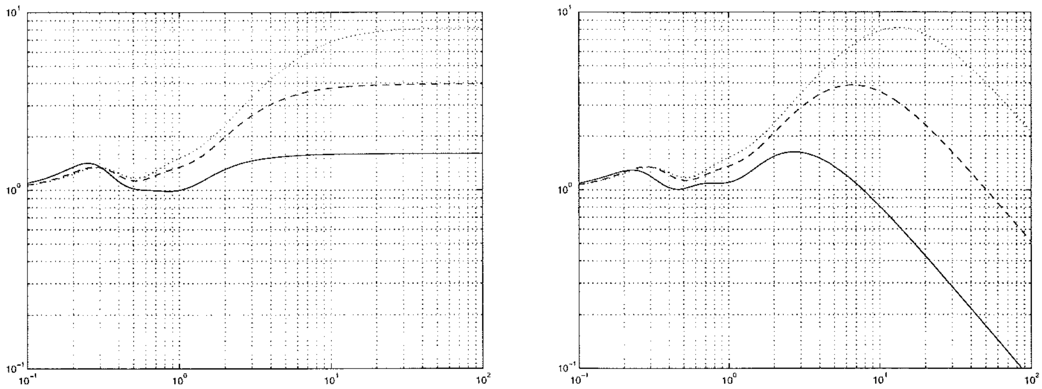


Figure 10 The amplitude curve of the transfer function from measurement noise n to control signal u for the process $F_y(s)G(s)$. First (left) and second (right) order filter for filter constant $N = 2$ (full line), 5 (dashed line) and 10 (dotted line).

rameters are obtained simply by repeating the design with the process G replaced by the transfer function $F_y G$.

The process $G(s) = 1/(s+1)^5$ has been used to demonstrate the effects of the tradeoff between filtering capacity and loss of control performance. Figure 9 shows the responses to changes in load disturbances and measurement noise of the measurement signal and the controller output. It is also of interest to consider the response of the process output to measurement noise. But with the actual choice of measurement noise as a sinusoidal signal with frequency 40 rad/s, the effects will not be visible why it is not plotted in Figure 9. In Figure 10 the amplitude curve of the transfer function from measurement noise n to control signal u is presented, see Equation (5). The details of the design calculations and simulations are summarized in Table 2. Note that in Figure 9 the obtained PID controller designed for $F_y G$ is compared to the one designed for G to show the loss of control performance.

Figure 9 show that the incorporation of a measurement filter gives only a small deterioration in control performance, compared to the case with no filter. According to Table 2, the control performance decreases for decreasing values of N, which is shown by the performance measure IAE.

Figure 9 demonstrates the efficiency of filtering measurement noise with second order filters compared to first order ones. Notice how the filtering capacity is decreased for increasing values of N, which can also be seen in the amplitude curves of Figure 10 and in the values of the performance measure M_n in Table 2.

According to Table 2, the difference in control performance and in filtering capacity between the first and second order filter are almost the same because of the choice of T_f in Equation (18). Only in the case of filtering a measurement noise around the frequency $1/T_f$ it is preferable to choose a first order filter because of the large values of the amplitude curve in Figure 10 for second order filters.

Consequently, the proposed design method determines measurement noise filters in a systematic way, such that the noise level is reduced. People working with traditional PID controllers, such that

$$k + \frac{ki}{s} + \frac{k_d s}{1 + s\frac{k_d}{kN}} \tag{19}$$

would prefer that the proposed design method should optimize on the controller parameters k, k_i, k_d and N. This approach would also determine the filter applied on the derivative term automatically, but at the expense that the filter constant N in Equation (19) would be optimized with respect to the load disturbance and not the measurement noise.

Set point weight

The design has so far focused on the response to load disturbances, which is of primary concern. However, it may also be important to have a good response to set point changes. One way to give specifications on the set point response is to consider the transfer function from set point to process output given by Equation (7).

In order to have a small overshoot in set point response, set point weight b and filter F_{sp} will be determined so that the resonance peak of the transfer function $G_{sp}(s)$, i.e.

$$M_{sp} = \max \left| G_{sp}(i\omega) \right|, \tag{20}$$

is close to one.

As Equation (20) is a function of the set point weight b and filter F_{sp}, the following approximation is made: the maximum of $|G_{sp}(i\omega)|$ occurs for $\omega = \omega_{mp}$, where ω_{mp} is the frequency for which the maximum of $|L(i\omega)/(1 + L(i\omega))|$ occurs. First, try to solve the problem without filtering the set point, i.e. with $F_{sp} = 1$. Attempting to choose parameter b such that

$$|G_{sp}(i\omega_{mp})| = |\frac{k_i + ibk\omega_{mp}}{k_i + ik\omega_{mp} - k_d\omega_{mp}^2}|M_p = 1, \tag{21}$$

implies that

$$b = \begin{cases} \dfrac{\sqrt{k^2\omega_{mp}^2 + (k_i - \omega_{mp}^2 k_d)^2 - k_i^2 M_p^2}}{k\omega_{mp} M_p} & \text{if } \dfrac{k^2\omega_{mp}^2 + (k_i - \omega_{mp}^2 k_d)^2}{k_i^2} \geq M_p^2, \\ 0 & \text{if } \dfrac{k^2\omega_{mp}^2 + (k_i - \omega_{mp}^2 k_d)^2}{k_i^2} \leq M_p^2. \end{cases} \tag{22}$$

Only positive values of b are allowed, since negative values of b may result in inverse step responses in the control signal. This is an undesirable way to reduce overshoots.

In those case when $b > 0$, Equation (21) holds, but it may happen that $|G_{sp}|$ becomes large for values of $\omega \neq \omega_{mp}$. To avoid this, b is restricted to values such that

$$|\frac{k_i + ibk\omega}{k_i + ik\omega - k_d\omega^2}| \leq 1, \qquad (23)$$

for all frequencies ω. This gives the following additional constraints on b

$$b \leq \begin{cases} \sqrt{\dfrac{T_d}{T_i}} & T_i \leq 4T_d, \\ \dfrac{2T_d}{T_i}\dfrac{1}{1-\sqrt{1-4T_d/T_i}} & T_i \geq 4T_d. \end{cases} \qquad (24)$$

Notice, that for set point weight of PI controllers, $(k_d = 0)$, it follows from Equation (23) that the corresponding upper limit of b is $b \leq 1$.

If $b = 0$, it is not sure that the design objective (21) will be obtained. If the set point response is important and the M_p value is large, the output of $G_{sp}(s)$ may be filtered by a first order filter, i.e.

$$F_{sp}(s) = \frac{1}{1 + sT_{sp}}. \qquad (25)$$

If the filter time constant T_{sp} is chosen as

$$T_{sp} = \frac{1}{\omega_{mp}}\sqrt{\frac{k_i^2 M_p^2}{k^2\omega_{mp}^2 + (k_i - \omega_{mp}^2 k_d)^2} - 1}, \qquad (26)$$

then $|F_{sp}(i\omega_{mp})G_{sp}(i\omega_{mp})|^2 = 1$.

Implementation Aspects of Matlab

Finally a brief presentation of the numerical solver and some implementation aspects are given. According to Equation (13) and (17) the design of PID controllers is formulated as a constrained optimization problem. Thus a numerical optimization problem is to be solved as it is not possible to solve it analytically. In this project the Optimization Toolbox in Matlab 5 has been used.

As for most numerical optimization routines it is of great importance to have good initial conditions, and a suitable search interval. Here a natural choice of initial conditions are the controller parameters k and k_i from the PI design, i.e.

$$[k^0\ k_i^0\ k_d^0] = [\bar{k}\ \bar{k}_i\ 0],$$

and a suitable search interval is given by,

$$\omega_{start} = \bar{\omega}_0/2,$$
$$\omega_{stop} = (\bar{\omega}_{180} + \bar{\omega}_{270})/2,$$

where $\bar{\omega}_0$ is the frequency at which the sensitivity function from the PI design is maximum. $\bar{\omega}_{180}$ and $\bar{\omega}_{270}$ are the frequencies where the argument of the loop transfer function from the PI design is $-180°$ and $-270°$ respectively.

The design problem can be solved by the following procedure:

(1) Give the transfer function of the process. Choose the design parameter expressed by M_s, M_p or some other norm.

(2) Determine the number of constraints, as it differs depending on the considered system.

(3) Make a PI design to obtain the initial values $[\bar{k}\ \bar{k}_i\ 0]$ and the frequencies $[\bar{\omega}_0\ \bar{\omega}_{180}\ \bar{\omega}_{270}]$.

(4) Solve the design problem with the Optimization Toolbox in Matlab 5.

(5) Verify that the resulting controller parameters fulfills the constraints. If not adjust the initial values or settings in the accuracy of the numerical routine.

6. Examples

The design method has been tested on a number of examples which illustrate its properties. The following transfer functions have been considered,

$$
\begin{aligned}
G_1(s) &= \frac{1}{s(s+1)^3}, \\
G_2(s) &= \frac{e^{-5s}}{(s+1)^3}, \\
G_3(s) &= \frac{1}{(s+1)(1+0.2s)(1+0.04s)(1+0.008s)}, \\
G_n(s) &= \frac{1}{(s+1)^n}, \quad n = 4,\ 5,\ 6,\ 7, \\
G_8(s) &= \frac{1-2s}{(s+1)^3}.
\end{aligned}
$$

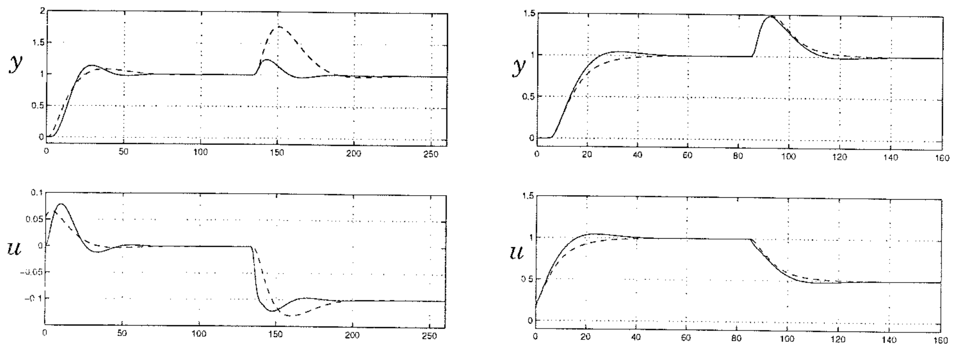
The first seven models capture typical dynamics encountered in the process industry. Model G_1 is an integrating process and G_2 models a process with long dead time. Model G_8 has a zero in the right half plane, which is uncommon in process control, but it has been included to demonstrate the wide applicability of the design procedure.

Figures 11 and 12 show the responses to changes in set point and load. The details of the design calculations and simulations are summarized in Table 3. Note that the PID controller obtained is compared to the corresponding PI controller to show the amelioration of the PID design. Although models $G_1 - G_8$ represent processes with large variations in process dynamics, Figure 11 and 12 show that the resulting closed loop responses for a load disturbance become similar for each value of M_s. This is important because it means that the proposed design procedure gives closed loop systems with desired and predictable properties.

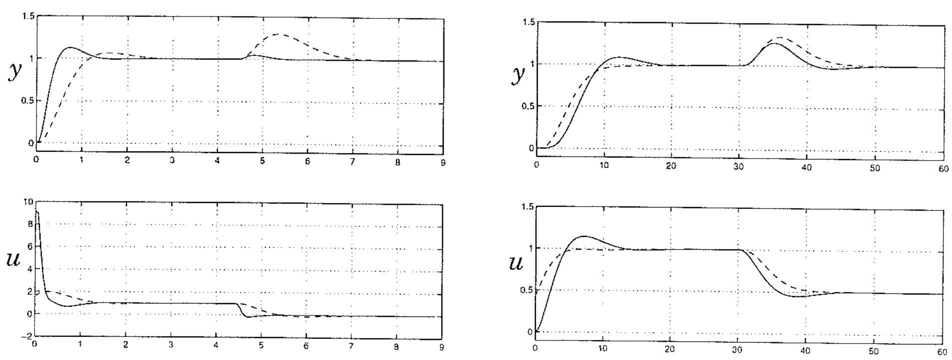
There is also a clear similarity between the responses obtained with the different values of the tuning parameter M_s, thus indicating the suitability of the M_s-value as a tuning parameter. Responses obtained with M_s=1.4 show little or no overshoot, as is normally desirable in process control. Faster responses are obtained with M_s=2.0. The settling time at load disturbances, t_s, is significantly shorter with the larger value of M_s. On the other hand, these responses are oscillatory with larger overshoots. This can be seen from the comparison between IE and the integrated absolute error IAE in Table 3. Notice the agreement with the conclusions made for design of PI controllers in Åström *et al.* (1998).

The controller gain k varies significantly with the design parameter M_s: it is larger for designs when $M_s = 2.0$ than for those when $M_s = 1.4$. However, integral time T_i is fairly constant for the stable processes, i.e., all processes except G_1. The derivative time T_d is usually larger for designs with $M_s = 1.4$ than for those with $M_s = 2.0$. In all cases the PID design generates a controller with complex zeros for $M_s = 2.0$. Thus, the controller will not be realizable in serial form.

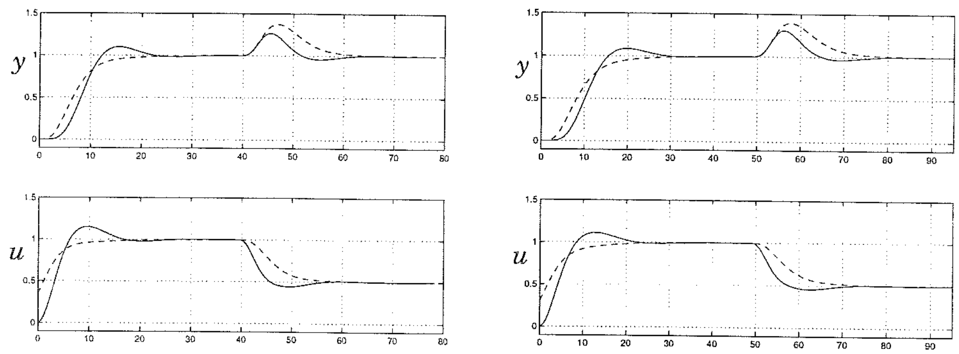
For M_s=2.0, the M_p-values are large. Consequently, the overshoots would be significant if the set point weight is $b = 1$. However, acceptable set point responses are obtained by suitable choices of either the set point weight b or the filter F_{sp}. According to Table 3, it is not always enough to set $b = 0$ to obtain a small overshoot filtering may also be needed.



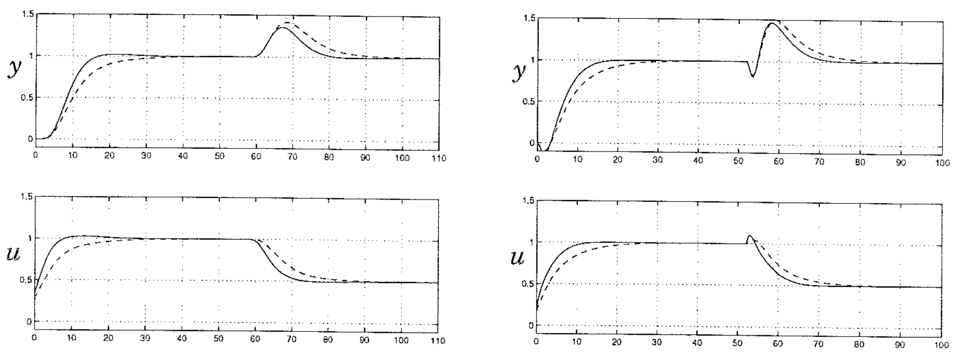
Systems G_1 (left) and G_2 (right)



Systems G_3 (left) and G_4 (right)

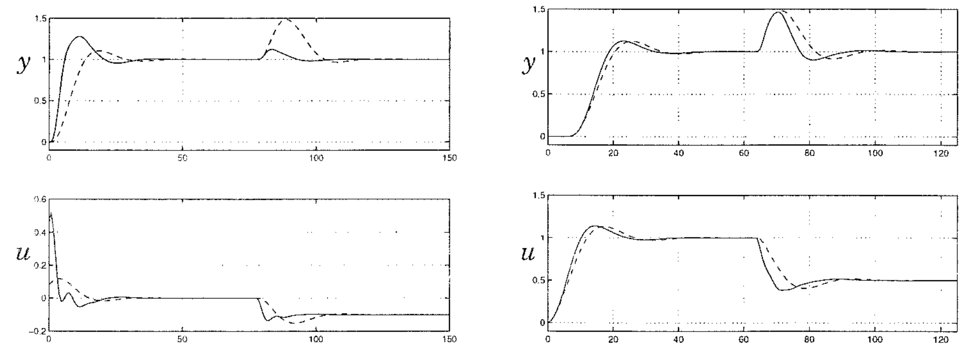


Systems G_5 (left) and G_6 (right)

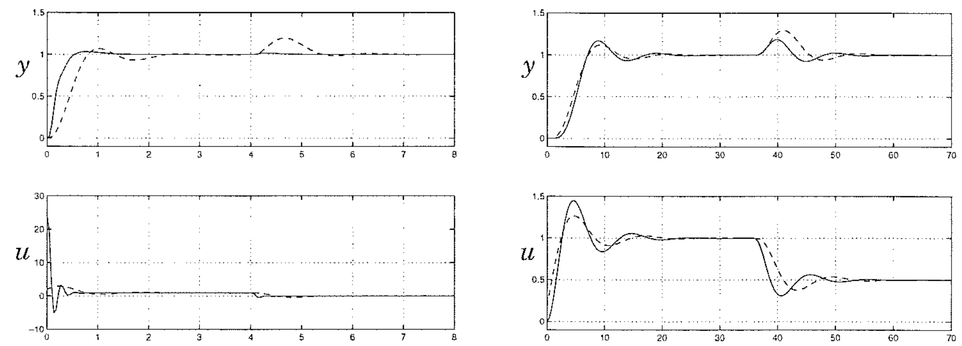


Systems G_7 (left) and G_8 (right)

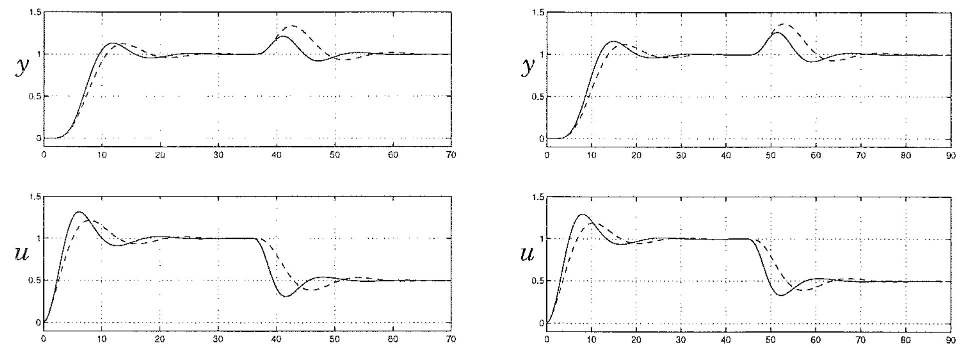
Figure 11 Comparison between the PID (full line) and the PI controller (dashed line) for $M_s = 1.4$. The graphs shows a step response followed by a load disturbance of the closed loop system.



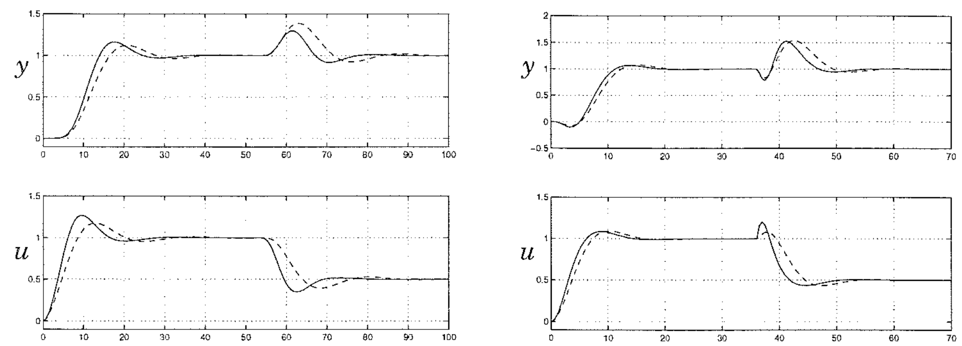
Systems G_1 (left) and G_2 (right)



Systems G_3 (left) and G_4 (right)



Systems G_5 (left) and G_6 (right)



Systems G_7 (left) and G_8 (right)

Figure 12 Comparison between the PID (full line) and the PI controller (dashed line) for $M_s = 2.0$. The graphs shows a step response followed by a load disturbance of the closed loop system.

Table 3 The properties of the obtained controllers for system G_1-G_8 with different values of the design parameter M_s.

Process	M_s	k	T_i	T_d	b	T_{sp}	IE	$\frac{IE}{IAE}$	ω_0	t_s	M_p
$G_1(s)$	1.4	0.291	8.03	3.15	0.00	3.58	27.6	0.71	0.90	23.5	1.37
	2.0	0.645	5.20	2.66	0.72	0.00	8.05	0.72	0.99	14.0	1.40
$G_2(s)$	1.4	0.180	2.33	2.04	0.94	0.00	12.9	0.92	0.18	27.2	1.00
	2.0	0.618	3.39	1.95	0.00	3.98	5.47	0.62	0.35	5.47	1.29
$G_3(s)$	1.4	15.92	0.322	0.15	0.39	0.00	0.020	0.81	19.4	0.72	1.26
	2.0	42.81	0.254	0.13	0.71	0.00	0.0059	0.86	26.1	-	1.64
$G_4(s)$	1.4	0.758	2.07	0.82	0.00	0.86	2.74	0.83	0.55	10.8	1.05
	2.0	1.74	1.69	0.96	0.00	1.91	0.97	0.48	0.73	15.4	1.75
$G_5(s)$	1.4	0.870	2.41	1.50	0.00	2.27	2.77	0.73	0.66	12.1	1.12
	2.0	1.57	2.13	1.42	0.00	2.85	1.37	0.50	0.68	9.05	1.63
$G_6(s)$	1.4	0.703	2.93	1.77	0.00	2.39	4.17	0.79	0.55	16.1	1.06
	2.0	1.23	2.72	1.62	0.00	2.78	2.20	0.55	0.53	11.8	1.49
$G_7(s)$	1.4	0.553	3.90	1.50	0.62	0.00	7.05	0.97	0.37	22.9	1.00
	2.0	1.067	3.26	1.85	0.00	2.85	3.06	0.60	0.45	14.1	1.40
$G_8(s)$	1.4	0.330	2.24	0.90	0.64	0.00	6.78	0.85	0.89	19.3	1.00
	2.0	0.569	2.08	0.86	0.00	2.01	3.66	0.61	0.64	17.3	1.33

7. Conclusions

PID controllers were designed to capture demands on load disturbance rejection, set point response, measurement noise and model uncertainty. Good load disturbance responses were obtained minimizing the integrated control error IE. Robustness is guaranteed by requiring a maximum sensitivity of less than a specified value M_s. Measurement noise is dealt with by filtering. Good set point response is obtained by using a structure with two degrees of freedom, which requires an extra parameter in the algorithm: the set point weighting b, and a filter for filtering the set point. The primary design parameter is the maximum sensitivity, M_s, auxiliary design parameters such as the maximum of the complementary sensitivity, M_p, can be added.

The design problem is formulated as a constrained optimization problem, where constraints are introduced to ensure robustness. A design method for PI controllers based on the same problem formulation was developed in Åström *et al.* (1998). In the present paper it is shown that this method cannot be extended directly to PID control, because it leads to an optimization problem which in most cases, has ridges, thus resulting in poor robustness and also poor control. However, it is possible to introduce additional constraints. The result is that design of PID controllers can be formulated as a constrained optimization problem which can be solved iteratively. Initial conditions are very important since the problem is non convex. A good way to find initial conditions is also presented.

The solution of the optimization problem gives a PID controller with a pure derivative. Simple rules for choosing a filter for the measured signal are then possible.

The design procedure has been applied to a variety of systems: stable

8. References

ÅSTRÖM, K. J. and T. HÄGGLUND (1995a): "New tuning methods for PID controllers." In *European Control Conference*, pp. 2456–2462. Rome, Italy.

ÅSTRÖM, K. J. and T. HÄGGLUND (1995b): *PID Controllers: Theory, Design, and Tuning*, second edition. Instrument Society of America, Research Triangle Park, NC.

ÅSTRÖM, K. J., H. PANAGOPOULOS, and T. HÄGGLUND (1998): "Design of PI controllers based on non-convex optimization." *Automatica*, **35:5**.

PANAGOPOULOS, H. and K. J. ÅSTRÖM (1998): "PID control design and $\mathcal{H}_\infty$ loop shaping." Department of Automatic Control, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden.

PERSSON, P. (1992): *Towards Autonomous PID Control.* PhD thesis ISRN LUTFD2/TFRT--1037--SE, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

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.

VINNICOMBE, G. (1998): *Uncertainty and Feedback. H_∞ Loop-Shaping and the ν-Gap Metric.* To be published.

YAMAMOTO, S. and I. HASHIMOTO (1991): "Present status and future needs: The view from Japanese industry." In ARKUN AND RAY, Eds., *Chemical Process Control—CPCIV.* Proceedings of the Fourth International Conference on Chemical Process Control, Texas.

ZIEGLER, J. G. and N. B. NICHOLS (1942): "Optimum settings for automatic controllers." *Trans. ASME*, **64**, pp. 759–768.

Paper III

PID Control Design and $\mathcal{H}_\infty$ Loop Shaping

Hélène Panagopoulos and Karl J Åström

Department of Automatic Control
Lund Institute of Technology
Box 118
S-221 00 Lund, Sweden
Phone: +46-46 2228789, Fax: +46-46 138118
Email: hp@control.lth.se

Abstract The present paper relates traditional methods for designing PID controllers to robust $\mathcal{H}_\infty$ control. In particular it is shown how the specifications for the PID design in Åström *et al.* (1998) and Panagopoulos *et al.* (1998) should be chosen to guarantee that the weighted $\mathcal{H}_\infty$ norm, see Glover and McFarlane (1989), of the transfer function from load and measurement disturbances to process inputs and outputs is less than a specified value γ. Also, a new way to determine for which class of systems a PID controller will be stabilizing is presented.

Keywords $\mathcal{H}_\infty$ Controller Design. PID Controller Design. Specifications. Robustness.

1. Introduction

Traditional methods to design PID controllers make a compromise between robustness and performance. For example, in Shinskey (1990) robustness is expressed by requiring that the Nyquist curve is outside a square which encloses the critical point, and performance is expressed as maximization of integral gain. The present paper relates traditional methods for design of PID controllers to robust $\mathcal{H}_\infty$ control, in particular the loop shaping methods developed in Vinnicombe (1998). The idea of robust control is to design a controller that minimizes the signal transmission from load disturbances and measurement noise to process input and output. This can be expressed by the $\mathcal{H}_\infty$ norm, γ, of a two-by-two transfer function matrix. It is shown that the requirement that γ is sufficiently small can also be expressed such that the Nyquist curve should be outside a certain region.

It is also shown that this region can be bounded internally and externally by circles that are closely related to the circles of constant sensitivity and constant complementary sensitivity. The ideas can also be used for efficient computations, see Åström *et al.* (1998).

Many different methods have been proposed to design PID controllers. In Glover and McFarlane (1989) and Vinnicombe (1998) a loop shaping method was developed. In Åström *et al.* (1998) and Panagopoulos *et al.* (1998) a method for designing PID controllers were presented, where the parameters of the controller were determined to maximize the gain of the integral term subject to a robustness constraint. The methods give a straight forward approach which is quite flexible. It is easy to introduce various forms of filtering and it can also find a proper set point weighting that also gives a good set point response. In this paper it is shown that these two methods are closely related to the $\mathcal{H}_\infty$ loop shaping method developed in Glover and McFarlane (1989). In particular it is shown how the specifications for the PID design should be chosen to guarantee that the weighted $\mathcal{H}_\infty$ norm of the transfer function from load disturbances to process inputs and outputs is less than a specified value γ.

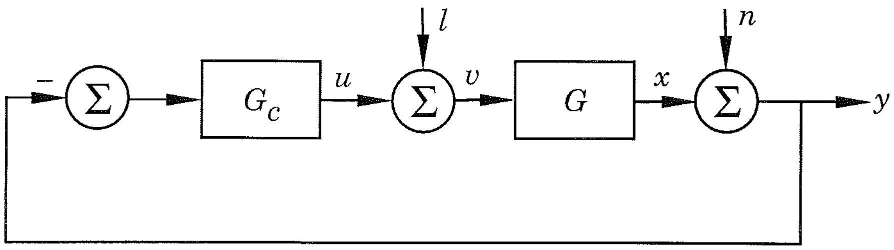


Figure 1 Block diagram describing the design problem.

2. $\mathcal{H}_\infty$ Control

First a brief overview of $\mathcal{H}_\infty$ design is given, emphasizing the parts relevant for the understanding of the paper. Typically a control problem has specifications for

- Attenuation of load disturbances
- Effects of measurement noise
- Robustness to modeling errors
- Set point response

Load disturbances, measurement noise and robustness are primary issues for process controllers which mostly operate as regulators. In the case of $\mathcal{H}_\infty$ control a multi variable transfer function is first derived describing how all disturbances influence the deviations in interesting variables. Then, it is attempted to find a controller which minimizes the $\mathcal{H}_\infty$-norm of this transfer function. This approach is well suited for dealing with the first three design criteria mentioned above. When a controller satisfying these requirements has been found a good set point response is obtained by using a structure with two degrees of freedom, see Horowitz (1963).

As an illustration consider the block diagram shown in Figure 1. The major inputs are the load disturbance l and the measurement noise n. The outputs of interest are the process output x and the signal v which represents the combined action of the load disturbance and the control signal. The signals are related through

$$\begin{bmatrix} x \\ v \end{bmatrix} = H \begin{bmatrix} n \\ l \end{bmatrix}$$

where

$$H = \begin{bmatrix} G \\ I \end{bmatrix} (I + GG_c)^{-1} [-G_c \quad I\,].$$

Notice, that the multi variable transfer function H, includes also the transfer functions relating l and n to u and y in Figure 1. Furthermore, the block diagonal elements of H are

$$T = G(I + GG_c)^{-1}G_c$$

and

$$S = (I + GG_c)^{-1}$$

which represent the complementary sensitivity function T and the sensitivity function S. The off-diagonal elements are the transfer functions $G(I + GG_c)^{-1}$ and $(1 + GG_c)^{-1}G_c$. Good control requires both x and v to be small in spite of the disturbances l and n. One way to express this is to require the $\mathcal{H}_\infty$ norm

$$\gamma = \|H\|_\infty \tag{1}$$

to be small. To use the parameter γ as a criterion for loop shaping was suggested by Glover and McFarlane (1989) and McFarlane and Glover (1992), where recommended values of γ were in the range $[2, 10]$. Also, it is shown how this design method gives many nice properties.

- It gives a good controller if such controllers exist.
- It provides robust stability against coprime factor uncertainties, Vidyasagar (1985).
- There is a good design variable, γ.

Also, in Vinnicombe (1998) a nice interpretation of Equation (1) is given, i.e. as the metric $\|G, -1/G_c\|$.

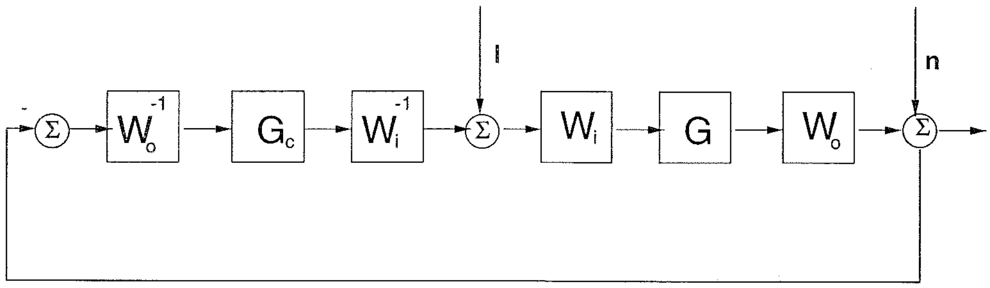


Figure 2 Block diagram describing the design problem.

Frequency Weighting

Frequency weighting may be introduced to emphasize the response at certain frequencies, see Figure 2. Then the problem is solved for the transformed system $\bar{G}$ and the transformed controller $\bar{G}_c$, with

$$\bar{G} = W_o G W_i,$$
$$\bar{G}_c = W_i^{-1} G_c W_o^{-1},$$

where W_i is the input weight and W_o is the output weight. The transformation is equivalent to design for the disturbances

$$\bar{l} = W_i l,$$
$$\bar{n} = W_o^{-1} n.$$

Consequently, the transformed system becomes

$$\bar{H} = \begin{bmatrix} W_o G W_i \\ I \end{bmatrix} W_o (I + GG_c)^{-1} W_o^{-1} [-W_i^{-1} G_c W_o^{-1} \quad I]. \tag{2}$$

Single Input Single Output Systems

For single input single output systems the transfer function H becomes

$$H = \frac{1}{1+GG_c} \begin{bmatrix} -GG_c & G \\ -G_c & I \end{bmatrix} = \frac{1}{1+GG_c} \begin{bmatrix} G \\ I \end{bmatrix} \begin{bmatrix} -G_c & I \end{bmatrix}. \tag{3}$$

As the matrix H is of rank 1, it is easy to compute its largest singular value,

$$\sigma^2(H) = \frac{(1+G_cG_c^*)(1+GG^*)}{(1+GG_c)(1+G^*G_c^*)}.$$

Thus the parameter γ is,

$$\gamma^2 = \sup_{\omega} \frac{(1+G_cG_c^*)(1+GG^*)}{(1+GG_c)(1+G^*G_c^*)}. \tag{4}$$

Also frequency weighting of single input single output systems will be simplified. In this case it is sufficient to use only one weight, i.e. $W_i = W$ and $W_o = 1$. Then the transformed system matrix in Equation (2) becomes

$$\bar{H} = \frac{1}{1+GG_c} \begin{bmatrix} -GG_c & GW \\ -G_cW^{-1} & I \end{bmatrix} = \frac{1}{1+GG_c} \begin{bmatrix} GW \\ I \end{bmatrix} \begin{bmatrix} -G_cW^{-1} & I \end{bmatrix}. \tag{5}$$

3. PID Control

To show that the methods for designing PID controllers found in Åström *et al.* (1998) and Panagopoulos *et al.* (1998) are closely related to the $\mathcal{H}_\infty$ loop shaping method developed in Glover and McFarlane (1989), a brief overview of the PID controller designs are given. The PID controller is described by the transfer function

$$G_c(s) = k + \frac{k_i}{s} + k_d s, \tag{6}$$

where k is the controller gain, k_i is the integral gain and k_d is the derivative gain. In reality its structure is more complicated, because of filtering of the derivative term and set point weighting, see Åström and Hägglund (1995). For the purpose of this paper, both methods for designing PID controllers are based on an idea to maximize the integral gain k_i, subject to the constraint on the Nyquist curve of the loop transfer function to lie outside a specified circle. This idea, which goes back to Shinskey (1990) is discussed in detail in Åström *et al.* (1998) and Panagopoulos *et al.* (1998), where the robustness is measured in the classical terms of the maximum of the sensitivity function, M_s, and the maximum of the complementary sensitivity function, M_p. Thus, the robustness measure will be a transparent design variable. The robustness constraint is expressed as either

$$M_s = \|(I + GG_c)^{-1}\|_\infty,$$

which implies that the Nyquist curve of the loop transfer function avoids the circle with center at $C = -1$ and radius $R = 1/M_s$, see Figure 3, or it is expressed as

$$M_p = \|GG_c(I + GG_c)^{-1}\|_\infty,$$

which implies that the Nyquist curve of the loop transfer function avoids the circle with center at $C = -M_p^2/(M_p^2 - 1)$ and radius $R = |M_p/(M_p^2 - 1)|$, see Figure 3. It is possible to replace the constraint on M_s and M_p with a combined constraint which reduces the computational effort substantially, see Åström *et al.* (1998). The combined constraint implies that the Nyquist curve of the loop transfer function avoids the circle with center at

$$C = -\frac{2M_sM_p - M_s - M_p + 1}{2M_s(M_p - 1)},$$

and radius at

$$R = \frac{M_s + M_p - 1}{2M_s(M_p - 1)},$$

see Figure 3. For practical purposes the combined constraint is not much more stringent than combining the two constraints on M_s and M_p respectively.

The combined constraint is simplified if both the sensitivity and the complementary sensitivity function are less than or equal to M, which implies that the Nyquist curve of the loop transfer function avoids the circle with center

$$C = -\frac{2M^2 - 2M + 1}{2M(M-1)},$$

and radius

$$R = \frac{2M - 1}{2M(M-1)}.$$

The circle has its diameter on the interval $[-M/(M-1), -(M-1)/M]$.

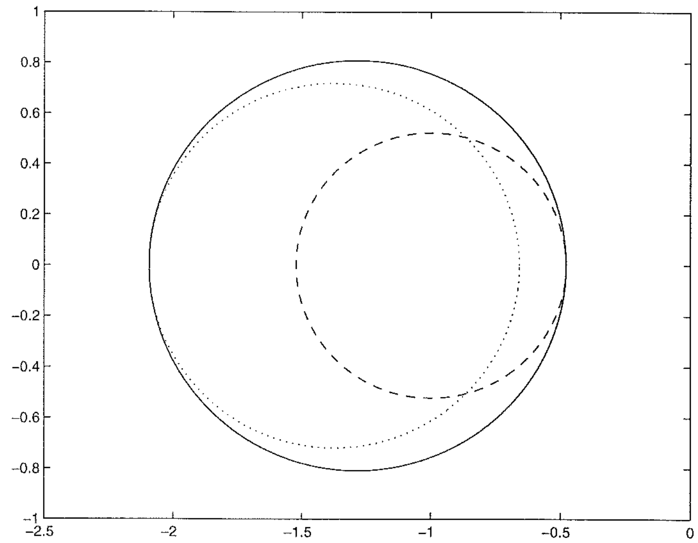


Figure 3 The M_s-circle (dashed line), the M_p-circle (dotted line) and the combined M_s- M_p circle (full line) for $M_s = M_p = 2.0$.

4. Comparisons

In this section it is shown that the $\mathcal{H}_\infty$ design problem for single input single output PID controllers presented in Section 2 is closely related to the methods for designing PID controllers developed in Åström *et al.* (1998) and Panagopoulos *et al.* (1998). In particular it is shown how the specifications for the PID design should be chosen to guarantee that the weighted $\mathcal{H}_\infty$ norm of the transfer function from load disturbances and measurement noise to process inputs and outputs is less than a specified value γ.

The minimization of the robustness measure γ in the $\mathcal{H}_\infty$ design gives a controller that compromises between attenuation of the disturbances l and n. By introducing frequency weighting it is possible to emphasize the weighting of the two disturbances by a proper choice of the weighting function W, which will serve as a design variable. For reasonable choices of W the largest value of γ will occur in the neighborhood of the crossover frequency. Note that the rejection of low frequency disturbances can be influenced by the weighting function but it is not particularly critical.

What will the best choice of W be regarded as a design variable? For the frequency weighted transfer function $\bar{G}$ in Equation (5), the robustness measure γ is given by

$$\gamma^2 = \|\bar{H}\|_\infty^2 = \frac{(1 + G_c G_c^* W^{-1} W^{-1*})(1 + GG^* WW^*)}{(1 + GG_c)(1 + G^* G_c^*)}. \tag{7}$$

The most favorable frequency weighting is the one that minimizes the numerator of Equation (7). Let $X = WW^*$, then the numerator of Equation (7) becomes

$$1 + GG^* X + G_c G_c^* X^{-1} + GG^* G_c G_c^*.$$

which has its minimum for $X = \sqrt{G_c G_c^* / GG^*}$. Thus, the weighting factor becomes,

$$W = \sqrt[4]{G_c G_c^* / GG^*}. \tag{8}$$

The weighting will typically enhance low and high frequencies. With PID control the low frequency weight is proportional to $1/\sqrt{\omega}$. Introducing the weight W given by (8) into (7) gives

$$\gamma^2 = \|\bar{H}\|_\infty^2 = \frac{(1 + |GG_c|)^2}{|1 + GG_c|^2}, \tag{9}$$

which is equivalent to

$$\gamma = \|\bar{H}\|_\infty = \max_\omega (|S(i\omega)| + |T(i\omega)|). \tag{10}$$

Notice that even if the weight (8) depends on the transfer function of the controller and the process, the quantity $\|\bar{H}\|_\infty$ depends only on the loop transfer function GG_c. Consequently, Equation (10) shows that the robustness measure γ is related to the values M_s and M_p which are also used as robustness constraints in the PID design presented in Section 3.

Although the robustness measure of the $\mathcal{H}_\infty$ design and the robustness constraint of the PID design are related, there are some fundamental differences between the two designs. For example, in the $\mathcal{H}_\infty$ design γ is insensitive to k_i for low frequencies and the requirements on the transfer functions $G_c/(1+GG_c)$ and $G/(1+GG_c)$ appear explicitly in Equation (3) compared to the PID design.

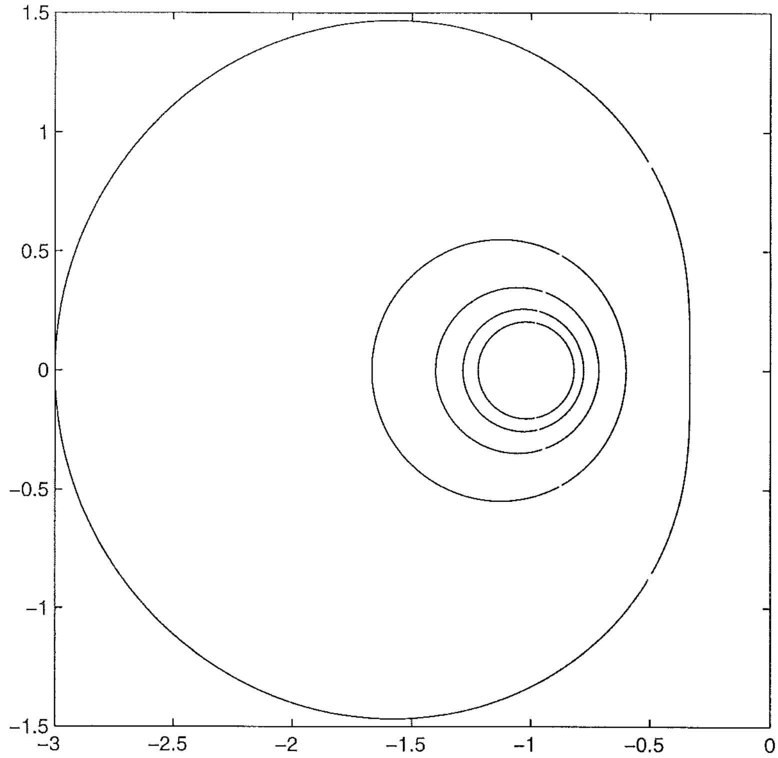


Figure 4 The loci of $(1 + |L|)/(|1 + L|) = \gamma$ for $\gamma = 2$ (outer curve), 4, 6, 8, 10 (inner curve).

The Gamma Contour

A graphical interpretation will be given of the robustness measure γ in Equation (9). Let $L = GG_c$ be the loop transfer function. It follows from Equation (9) that

$$\gamma = \max \frac{1 + |L(i\omega)|}{|1 + L(i\omega)|}, \tag{11}$$

The condition that $\gamma \leq \gamma_0$ can thus be interpreted graphically. That is, the Nyquist curve of the loop transfer function should be outside the contour

$$\frac{1 + |L(i\omega)|}{|1 + L(i\omega)|} = \gamma_0.$$

Therefore it is of interest to investigate the level curves of the function

$$f(L) = \frac{1 + |L|}{|1 + L|},$$

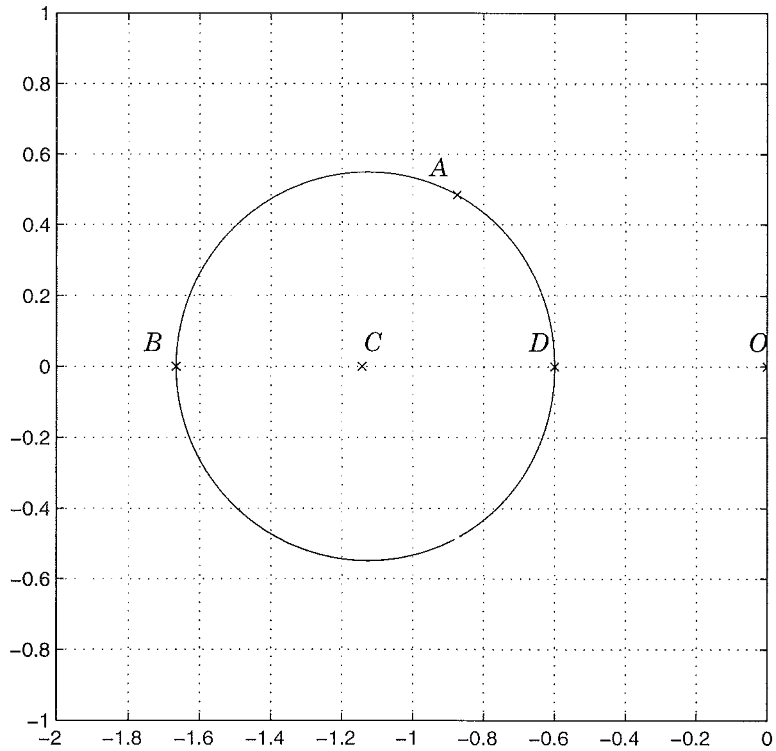


Figure 5 The γ-curve.

in the complex plane. Introduce $L = re^{i\varphi}$ then Equation (11) becomes

$$\gamma = \frac{1+r}{|1+re^{i\varphi}|} = \frac{1+r}{\sqrt{1+r^2+2r\cos\varphi}}.$$

Thus, the γ-curve can be represented as

$$r(\varphi) = -\frac{\gamma^2\cos\varphi - 1}{\gamma^2 - 1} \pm \sqrt{\left(\frac{\gamma^2\cos\varphi - 1}{\gamma^2 - 1}\right)^2 - 1}.$$

A few examples are given in Figure 4. Straight forward calculations show that

$$\begin{aligned}
OA &= 1, \\
OB &= -\frac{\gamma+1}{\gamma-1}, \\
OC &= \frac{\gamma^2}{2-\gamma^2}, \\
OD &= -\frac{\gamma-1}{\gamma+1}, \\
AC &= \frac{2\sqrt{\gamma^2-1}}{\gamma^2-2},
\end{aligned}$$

which is shown in Figure 5. Note that $OB \cdot OD = 1$.

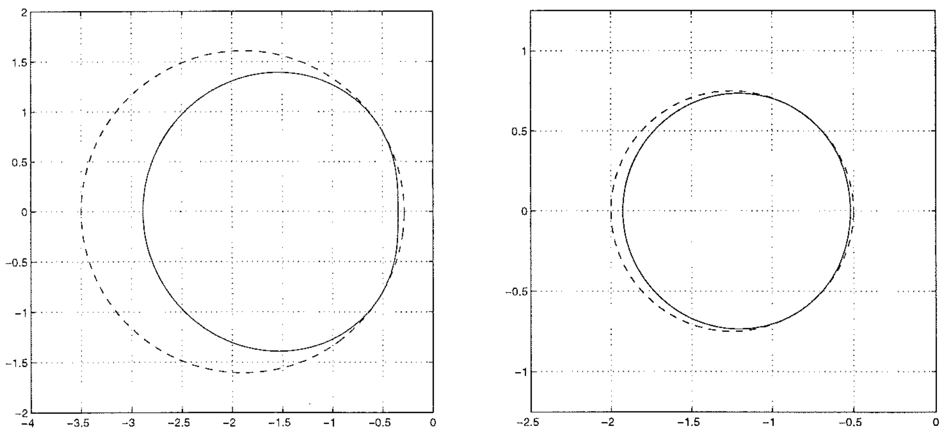


Figure 6 γ-curve (full line) for $\gamma = 2.06$ (left), 3.16 (right) enclosed by the combined M-circle (dashed line) for $M = 1.4$ (left), 2.0 (right).

Relations Between M and γ

In the following several relations will be established between M and γ which will give further insight into the relations between the two design methods. According to Equation (10),

$$\gamma = \max(|S| + |T|) \leq \max|S| + \max|T| = 2M,$$

and

$$\gamma = \max(|S| + |T|) = \max(|S| + |1 - S|) \geq \max(2|S| - 1) = 2M - 1,$$

where the second equality follows from $S + T = 1$ and the inequality from the triangular inequality. Consequently, the following inequality has been established

$$2M - 1 \leq \gamma \leq 2M,$$

which gives an indication of how to chose M to guarantee a certain γ.

Sharper results can be obtained from simple but tedious calculations which show that the γ-curve is inside the combined $M_s M_p$-circle with

$$M = \frac{\gamma^2 + 2\sqrt{\gamma^2 - 1}}{2 + 2\sqrt{\gamma^2 - 1}}. \tag{12}$$

See Figure 6. Solving Equation (12) for γ gives

$$\gamma = \sqrt{4M^2 - 4M + 2}. \tag{13}$$

Accordingly, a controller designed for the combined constraint $M_s = M_p = M$ guarantees that γ is larger than the value given by Equation (13). Table 1 and 2 gives numerical values of corresponding values of M and γ. The choice $M = 2.0$ guarantees a γ is less than $\sqrt{10}$ which according to Vinnicombe (1998) gives good robustness. Lower values of M give even better robustness. Figure 6 shows corresponding γ- and M-curves for $M = 1.4$ and $M = 2.0$.. The figure indicates that the design based on the combined M-curves are not much more conservative than design based on γ particularly for $M = 2.0$. However, the calculations for constraint on γ are much more complicated.

Table 1 Numerical values of corresponding M and γ.

M	1.4	1.5	1.6	1.7	1.8	1.9	2.0
γ	2.06	2.24	2.42	2.60	2.79	2.97	3.16

Table 2 Numerical values of corresponding M and γ.

γ	2	2.5	3	3.5	4	4.5	5
M	1.37	1.65	1.91	2.18	2.44	2.69	2.95

Classification of Stabilizable Systems

A way to determine for which class of systems a PID controller will be stabilizing is presented. In Vinnicombe (1993) and Vinnicombe (1998) the parameter b defined as

$$b := \begin{cases} \dfrac{1}{\gamma} & \text{if } [G, G_c] \text{ is stable,} \\ 0 & \text{otherwise,} \end{cases}$$

has been introduced as a performance measure, also called the generalized stability margin. The parameter is in the range $0 \leq b \leq 1$. It follows from the definition that the system is unstable for $b = 0$ and performance improves with increasing values of b. Solutions of many design examples, see Vinnicombe (1998) and McFarlane and Glover (1992), indicate that the value should be larger than $1/\sqrt{10}$ to have reasonable robustness and performance.

In Vinnicombe (1993) and Vinnicombe (1998) the parameter b has been used to derive some very interesting results relating robustness to model uncertainty. To do this the following norm is introduced to measure the distance between two stable systems P and Q

$$\delta(P, Q) = |(I + Q^*Q)^{-1/2}(Q - P)(1 + P^*P)^{-1/2}|_\infty \tag{14}$$

provided that a certain winding number constraint is satisfied, see Vinnicombe (1993). It is shown that if the closed loop system (G, G_c) is stable with $b(G, G_c) \geq \beta$ then the closed loop system $(\bar{G}, G_c)$ is stable for all $\bar{G}$ such that $\delta(G, \bar{G}) \leq \beta$. Due to symmetry the system $(G, \bar{G}_c)$ is also stable for all $\bar{G}_c$ such that $\delta(G, \bar{G}_c) \leq \beta$.

By ensuring that a PID design is such that $\gamma \leq \gamma_0$ gives a good guarantee for robustness of the closed loop system, i.e. it is possible to determine a set of systems which are stabilized by the controller.

5. Conclusions

This paper describes how traditional methods for designing PID controllers are related to robust $\mathcal{H}_\infty$ control. In particular it is shown how the specifications for the PID design in Åström *et al.* (1998) and Panagopoulos *et al.* (1998) should be chosen to guarantee that the weighted $\mathcal{H}_\infty$ norm, see Glover and McFarlane (1989), of the transfer function from load disturbance and measurement noise to process inputs and outputs is less than a specified value γ. It is shown that the requirement that γ is sufficiently small can be expressed such that the Nyquist curve should be outside a certain region. It is also shown that this region can be bounded internally and externally by circles that are closely related to the circles of constant sensitivity and constant complementary sensitivity. The ideas can also be used for efficient computations, see Åström *et al.* (1998).

A new way to determine for which class of systems a PID controller will be stabilizing is presented. By ensuring that a PID design is such that $\gamma \leq \gamma_0$ a good guarantee for robustness of the closed loop system is given. If γ is accepted as a good performance measure the results of this paper suggest that the robustness constraint of PID controllers should be chosen as the combined M_sM_p-circle. In this way it is guaranteed that a design will automatically satisfy the $\mathcal{H}_\infty$ robustness constraint.

6. References

ÅSTRÖM, K. J. and T. HÄGGLUND (1995): *PID Controllers: Theory, Design, and Tuning*, second edition. Instrument Society of America, Research Triangle Park, NC.

ÅSTRÖM, K. J., H. PANAGOPOULOS, and T. HÄGGLUND (1998): "Design of PI controllers based on non-convex optimization." *Automatica*, **35:5**.

GLOVER, K. and D. MCFARLANE (1989): "Robust stabilization of normalized coprime factor plant descriptions with H_∞-bounded uncertainty." *IEEE Transactions on Automatic Control*, **34:8**, pp. 821–830.

HOROWITZ, I. M. (1963): *Synthesis of Feedback Systems.* Academic Press, New York.

MCFARLANE, D. and K. GLOVER (1992): "A loop shaping design procedure using H_∞ synthesis." *IEEE Transactions on Automatic Control*, **37:6**, pp. 759–769.

PANAGOPOULOS, H., K. J. ÅSTRÖM, and T. HÄGGLUND (1998): "Design of PID controllers based on constrained optimization." Department of Automatic Control, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden.

SHINSKEY, F. G. (1990): "How good are our controllers in absolute performance and robustness?" *Measurement and Control*, **23**, May, pp. 114–121.

VIDYASAGAR, M. (1985): *Control System Synthesis: A Factorization Approach.* MIT Press, Cambridge, Massachusetts.

VINNICOMBE, G. (1993): "Frequency domain uncertainty and the graph topology." *IEEE Transactions on Automatic Control*, **AC-38**, pp. 1371–1383.

VINNICOMBE, G. (1998): *Uncertainty and Feedback. H_∞ Loop-Shaping and the ν-Gap Metric.* To be published.