TOPICS IN THE DESIGN, ANALYSIS AND IMPLEMENTATION OF FEEDBACK CONTROL SYSTEMS

by

N. LEONARD SEGALL, B.SC., M.ENG.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

Hamilton, Ontario, Canada

DESIGN, ANALYSIS AND IMPLEMENTATION OF CONTROL SYSTEMS

DOCTOR OF PHILOSOPHY (1993) McMaster University
(Chemical Engineering) Hamilton, Ontario

TITLE: Topics in the Design, Analysis and Implementation of Feedback Control Systems

AUTHOR: Nathaniel Leonard Segall, B.Sc. (Queen's University at Kingston), M.Eng. (McMaster University)

SUPERVISORS: Dr. J.F. MacGregor and Dr. J.D. Wright

NUMBER OF PAGES: xii, 276

ABSTRACT

Nathaniel Leonard Segall: Topics in the Design, Analysis and Implementation of Feedback Control Systems, Ph.D. Thesis, McMaster University, Hamilton, Ontario, 1993.

This thesis investigates the design, analysis and implementation of feedback control systems. A new definition of robustness for continuous feedback control systems is advanced: the region of the joint allowable variation in the gain and the dead-time of the process. This region is constructed from the open loop system frequency response. The continuous time Proportional-Integral-Derivative (PID), Internal Model Control (IMC) and Linear Quadratic Optimal Control (LQOC) design procedures for a first order plus dead-time process are compared for Integral Square Error (ISE) performance when tuned for the same level of robustness. The order from highest to lowest performance is: PID, IMC, LQOC. The robustness analysis method is used to explain the adverse effects of dead-time compensation on robustness.

The continuous ISE performance of the discrete LQOC, and State Deadbeat IMC controllers are compared as the size of the control interval is changed; which one is better is shown to depend on the choice of the control interval. A new modified LQOC controller and a new Extended Horizon controller are proposed. The robustness analysis developed for continuous systems is extended to discrete systems as the region of joint allowable variation in the process gain and quasi dead-time. This region is constructed from the discrete open loop frequency response. The state deadbeat IMC with an augmented filter, and modified LQOC controllers are shown to have better ISE performance than the LQOC controller when tuned for the same level of robustness.

A procedure is presented for saturation correction in multivariable one-step optimal controllers. A simultaneous correction is included which adjusts the remaining control inputs if some inputs saturate. The procedure is applicable to all related controllers including IMC. Results on saturation correction for PID controllers are also presented.

ACKNOWLEDGEMENTS

I wish to thank my supervisors Dr. J.F. MacGregor and Dr. J.D. Wright for their guidance, patience, encouragement and many discussions throughout this project.

I also wish to thank:

Dr. C.M. Crowe for his encouragement and many valuable discussions, especially on mathematical concepts,

Dr. F. Mirza for his always cheerful encouragement,

Dr. P.A. Taylor for many helpful discussions,

Ms. S. Gallo-O'Toole for her encouragement and administrative guidance,

The many graduate students, too numerous to name, whose stimulating discussion and friendly company created an exciting environment for research.

I would also like to thank:

Bram Brouwer for his support and encouragement in the final completion of this thesis,

my sister Jacqueline for her constant interest, encouragement, advice and support,

my parents for their interest, encouragement and support.

I would like to give special thanks to Marilyn Mason for her love, encouragement, constant support, unending patience and innumerable discussions on all aspects of this research and thesis. Her contribution is immeasurable and without her this thesis would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The design, analysis and implementation of a feedback control system are three steps in moving from a control problem to a practical solution. The design of a controller requires the selection of a design method and the tuning of the design. In analysis, the controller is tested to determine if it is a suitable solution to the original problem. If a controller design proves unsuitable then retuning of the design or the selection of a different design procedure is necessary. Finally, it is important that the controller be implemented correctly so that the benefits of the design and analysis effort are not lost.

The design of feedback control systems is a well developed field with a rich literature and many textbooks. However, the industrial application of advanced controller designs has lagged behind their academic development. The research in this thesis is directed at this problem: a controller design method must be selected, it must be analyzed to assure that it is appropriate to the problem and that it is well tuned, and it must be implemented properly so that the expected performance is not compromised when the control element saturates.

These topics are tied together because poor controller design, inadequate analysis or faulty implementation will lead to a failure of the application.

It is a difficult problem to select a design procedure from the wide variety available. Usually each design procedure has guidelines for tuning associated with it; however, the guidelines are generally related to only one particular objective for which the controller was derived. For example, the Deadbeat controller design (Bergen and Ragazzini (1954)) drives the process errors to zero at the control instants, the discrete Linear Quadratic Optimal Control (LQOC) controller (Box and Jenkins (1970), Åström (1970)) minimizes an objective function and the Internal Model Controller (IMC) (Garcia and Morari (1982)) design has a set of rules for controller synthesis. Each of these designs allows a tuning parameter: the staleness factor in a Deadbeat design, the penalty on the changes in the control actions in the LQOC design, and the tuning filter time constant in the IMC design.

A designer seeking a practical solution to a control problem usually has several attributes which are required of the closed loop system, such as: good performance, good robustness to process changes and reasonable demands on the manipulated control action. Good performance is required to keep the process at its set point despite set point changes or disturbances entering the system. Good robustness to process changes is required because the model of the process which is used to design the controller is an abstraction of the process: it is not the real process. Furthermore the real process often changes with time due to changes in the process operating conditions. Reasonable

demands on the manipulated control action are required because there is usually a limitation on the allowable magnitude of the control actions.

Robustness of control systems design has long been recognized as a desirable feature. The concepts of gain margin and phase margin have traditionally been used to assess the robustness of a closed loop system to changes in the actual process from that used in the controller design (see for example a textbook such as Coughanowr and Koppel (1965)). However, it has more recently been observed that phase margin is inadequate to measure the robustness of controllers with dead-time compensation (Rivera et al (1986), Palmor (1980)). One approach to measuring the robustness of a control system is to map a stability region in the allowable variation of the parameters of the process (Katbab and Jury (1991), Jury (1990), Romagnoli et al (1988), Bergh (1986), Palmor and Shinnar (1979)). Another method is to prescribe a frequency dependent stability requirement in the frequency domain (Doyle and Stein (1981), Morari and Zafiriou (1989), Lewin (1991)).

This thesis addresses the problem of robustness analysis by proposing a new definition of robustness: the region of joint allowable variation in the process gain and the process dead-time for which the closed loop system will remain stable. This region in the parameter space of the process model is calculated from the open loop frequency response. This definition bridges the gap between definitions in the parameter space of the process model and definitions in the frequency domain because it explains how features of the frequency response affect the robustness region in the parameter space. This definition also measures the effect of increasing frequency on robustness very naturally

through the use of the allowable dead-time variation: the effect of a change in the process dead-time has an effect on the phase of the frequency response which increases linearly with frequency.

Robustness for discrete controllers applied to continuous systems is defined as the region of joint allowable change in the process gain and the quasi dead-time. The concept of quasi dead-time is introduced based on observations of the relationship between continuous process dead-time and the frequency response of the equivalent discrete process. This definition is shown to be useful for comparing the robustness of different controller designs, and also for comparing the same controller design when using different sizes of control interval.

Performance has long been the most commonly analyzed feature of controller designs. This is because most controller design procedures have some level of the system output performance as a goal. However, in discrete controller designs it is usually the discrete performance which is measured: the deviation from set point at the control instants. This discrete analysis ignores the fact that discrete controllers are generally designed for continuous processes using discrete models which are coincident with the continuous process only at the control instants. The continuous time performance of the continuous process is much more relevant to most control system designers even when a discrete controller is used to control the process. Zafiriou and Morari (1985), Lennartson and Söderström (1989) and Lennartson (1990) present studies which examine the continuous time performance of discrete controller designs.

The performance studies in this thesis use the continuous time Integral Square Error (ISE) and occasionally the Integral Absolute Error (IAE) criteria for analysing the performance of continuous systems using both continuous and discrete controllers. These measures are shown to be useful for comparing different controllers and different sizes of the control interval. In addition time plots of the continuous process output in response to a set point change are shown to be useful in comparing different controller designs and in explaining differences in the ISE and IAE performance.

The control actions which are required to achieve the design performance have long been recognized as being important (see for example Bergen and Ragazzini (1954) and the discussion which follows the paper). In fact, for practical reasons the designer is sometimes more interested in the control actions which a controller generates than in the process response. However, few studies focus much attention on the control actions. Exceptions to this include studies such as Bergh and MacGregor (1987) and Lennartson (1990) where the variance of the control actions is used to compare different controller designs and Garcia and Morari (1982) and Zafiriou and Morari (1985) where simulated examples show both the process response and the control actions as controllers are tuned.

In this thesis time plots of the control actions generated by the controller in response to a step change in the set point are shown to be extremely useful in explaining the differences between different controller designs. The analysis also shows that a controller can have

extremely good performance and robustness properties and yet the control actions which are generated may not be realizable in practice.

Finally, having arrived at a suitable design it is important that the desirable properties of the design be preserved in the face of real world limitations such as saturation limits on the manipulated variable. When control actions saturate the system performance will naturally be reduced because less control action is available than was called for. However, performance can be lost beyond this expected level because of controller implementation problems.

A new method of correcting one-step optimal controllers (which include minimum variance controllers and equivalent controllers) is derived in this thesis. The correction uses the past unimplemented portions of the control action in calculating the current control action. For multivariable systems an additional simultaneous correction is derived. This simultaneous correction takes in to account the saturation of some of the control actions to calculate the remaining control actions which do not saturate. The saturation problem of the Proportional Integral Derivative (PID) controller is also examined in this thesis. New methods of reset-windup protection are presented and are compared with some standard methods.

The controller design and analysis problems are addressed in Chapters 2 and 3 of this thesis. Chapter 2 analyzes controller designs for continuous single-input-single-output (SISO) systems. Chapter 3 analyzes discrete controller designs for continuous SISO processes. Problems of implementation are explored in Chapters 4 and 5. Chapter 4 explains how to compensate for saturation in discrete multi-input-multi-

output one-step optimal controllers. Chapter 5 examines the reset windup problem and problems of implementation in discrete Proportional Integral and Derivative (PID) controllers.

In Chapter 2, the method of Tsypkin (1946) for calculating the robustness region of joint allowable variation in the process gain and dead-time parameters is extended to systems which can be made unstable by a reduction in the dead-time. His analysis only included systems which could be destabilized (or stabilized) by adding dead-time. This method of robustness analysis provides a clear indication of the robustness of a closed loop feedback system to changes in the process gain and the process dead-time. Frequency domain robustness specifications of Palmor (1980), Doyle and Stein (1981), Morari and Zafiriou (1989) and Lewin (1991) require that the designer specify a priori the magnitude of the process uncertainty as a function of frequency, paying special attention to the "high" frequencies. However, the robustness region method described in this thesis allows a designer to calculate the allowable gain and dead-time change once a controller has been designed. The frequency domain construction used to calculate the robustness region makes it easy to understand the nature of any robustness limitations. The question of what is a "high" frequency is easily answered since the allowable change in the process dead-time incorporates frequency information in a natural and meaningful way.

The robustness region measure of robustness and the Integral Square Error measure of performance are combined to compare the robustness vs performance tradeoff of the tuning of Linear Quadratic Optimal Control (Palmor (1982)), Internal Model Control and Proportional

Integral Derivative control (Rivera et al (1986)) for a first order plus dead-time process. The comparison shows that for the same level of robustness PID has better performance than IMC which has better performance than LQOC. This is a surprising result as Rivera et al (1986) claimed that IMC gives 10% better performance than PID control, but they do not compare the controllers at the same level of robustness. The reason for the poorer LQOC performance is explained by showing that the LQOC controller, excluding the dead-time compensation, has an additional first order filter. The robustness region is shown to be useful for comparing the robustness of these very different controllers. The frequency domain analysis allows the designer to interpret which features of the controller are responsible for robustness and which features are responsible for performance.

The robustness analysis method is used to analyze the robustness implications of dead-time compensation. The high frequency peaks in the open loop frequency response are shown to cause contraction of the robustness region, and tuning of controllers is shown to reduce these peaks. Although it has previously been observed that high frequency peaks in the frequency response of dead-time compensating controllers cause robustness problems (Palmor (1980), Ioannides et al (1979)), there has been no quantitative analysis of how tuning works to reduce these peaks and improve robustness. The insights from robustness analysis are used in this chapter to explain why the substitution of derivative action for dead-time compensation (Rivera et al (1986)) can provide good robustness and good performance.

In Chapter 3 the first order plus dead-time and second order processes are used to provide a varied range of discrete transfer function models to study controller performance as the size of the control interval is varied. The discrete Linear Quadratic Optimal Control (Box and Jenkins (1970), Åström (1970)), modified Dahlin (Dahlin (1968)), and State Deadbeat IMC (Zafiriou and Morari (1985)) controller designs approximate model inverse are compared for the different discrete process models. Under certain conditions each of these controllers can give poor performance. The comparative performance analysis in this chapter shows that when the process model has a zero near minus one the State Deadbeat IMC controller performs best, but when the process model zeros are not near minus one then the LQOC controller performs best. Continuous time simulations of the process output and the control actions are used to explain the results. Zafiriou and Morari (1985) have provided an excellent study where they use continuous time simulations to show the conditions under which each of these controllers perform more poorly than the State Deadbeat controller. Where the study in this chapter provides an additional contribution is in highlighting that the LQOC controller provides improved performance over the State Deadbeat controller when the process model has a zero which is not near minus one.

The tuning of LQOC, IMC (Garcia and Morari (1982)) and State Deadbeat IMC controllers are compared and analyzed. The analysis leads to the suggestion of two novel controller designs. The Modified LQOC controller calculates a controller for a nonstationary disturbance using the spectral factorization approach as if the disturbance were

stationary. This controller has a tuning filter which is of order one less than the LQOC tuning filter. Through the appropriate choice of the tuning filter gain the resulting controller still has integral action. This approach is different from that of Wilson (1970) and Harris and MacGregor(1987) but appears to be the same as the approach of Kucera (1979). Also, a new extended horizon controller design is proposed which behaves as if it were a state deadbeat controller for the same process but with a longer control interval. This latter controller is an improvement over the extended horizon controller of Ydstie et al (1985) in that if a single set point change is made and no disturbance enters the system then the initial control action sequence will be implemented with no changes.

As a prelude to defining robustness analysis for discrete control of continuous processes, the effect of a change in the dead-time of a continuous process on the frequency of the discrete equivalent process is explored. The modified Z transform of a first order process with a fractional period of dead-time is shown to be a convex combination of the Z transforms of the same process with integer periods of dead-time which bracket the actual dead-time. This is believed to be a new result. Through the use of the example of a second order process it is shown that this result is not general. In fact, for a second order process the frequency response of the process with a fractional period of dead-time can be outside of the boundary formed by the frequency response of the process with integer periods of dead-time which bracket the actual dead-time. This result appears to be new and

has implications for studies which determine stability boundaries by a search on the dead-time parameter.

The definition of the robustness region for continuous processes is adapted for discrete control of continuous processes as the region of joint allowable variation in the process gain and the quasi dead-time. The quasi dead-time is defined to be the phase margin of the discrete open loop frequency response divided by the discrete frequency. When the allowable variation in the quasi dead-time is an integer number of control intervals and the process has no fractional period of dead-time then it provides an exact upper limit on the allowable dead-time variation in the underlying continuous process. The power and the limitations of this definition of robustness are shown by application to examples.

The method is used to show that a minimum variance controller for a process with no dead-time subject to random walk disturbances (equivalent to deadbeat control for step changes in set point) can only tolerate an increase in the process dead-time of at most one control interval. That is, the robustness of minimum variance control is proportional to the size of the control interval. This is believed to be a new result. In addition it is shown that the robustness of minimum variance control is further reduced when the process has dead-time. The limitation of the method is shown to be that when calculating the robustness of discrete control applied to a process with a fractional period of delay the allowable variation in the quasi dead-time can significantly over-estimate the allowable variation in the true dead-time; however, the relative results when comparing different controllers

on the same process are meaningful. Morari and Zafiriou (1989) translate a continuous time frequency response robustness specification to the design of controllers for discrete processes without showing any awareness of the difficulties caused by fractional periods of dead-time on the discrete frequency response. Again, as in the continuous time definition of robustness, the quasi-dead time incorporates frequency information in a natural and meaningful way.

The allowable change in the quasi dead-time measure of robustness and the continuous Integral Square Error measure of performance are combined to compare the robustness vs performance tradeoff of the tuning of LQOC, Modified LQOC, IMC, and State Deadbeat IMC controllers applied to first and second order processes with step changes in set point. The results are explained using continuous time plots of the process output and the control actions. This study shows that a controller with good performance and good robustness can still require totally impractical control actions.

The LQOC controller design is shown to have the poorest performance for the same level of robustness. The poor performance of the LQOC controller is because it has a tuning filter which is of order one higher than the process. The IMC and State Deadbeat IMC controllers with a first order filter applied to a second order process are shown to have good performance and robustness, but can generate control actions which are not realizable. The State Deadbeat IMC controller with a filter order matched to the order of the process has good performance, robustness and realizable control actions. The modified LQOC controller has a filter order which is naturally matched to the order of the

process and so has good performance, robustness and realizable control actions. This study contributes to the understanding of the relationships between tuning filters, performance, robustness and realizability of control actions. Bergh and MacGregor (1987) have studied the improved disturbance rejection ability of the LQOC controller over the IMC controller for autoregressive disturbances. Their study, however, focuses on discrete measures of performance for a process example with a fractional period of dead-time and also does not make a comparison with a State Deadbeat IMC controller which provides improved robustness for this type of process.

In Chapter 4 One-step optimal controllers, which include minimum variance controllers, are shown to suffer severe performance degradation when the control action is subject to a saturation limit. A new saturation correction for one-step optimal Multi-Input Multi-Output controllers is derived which allows past saturation to be compensated for in a straightforward calculation using the difference between the past calculated and past implemented control actions. This derivation extends the SISO derivations of Goodwin (1972) and Bezanson (1984) to produce a formula which can be implemented directly. In addition it corrects the MIMO derivation of Mäkilä (1982). A further simultaneous correction is derived which is only required for MIMO systems. The iterative correction calculates an immediate correction for control actions which do not saturate for those which do saturate. In addition the application of this method to other discrete model based controllers such as the Dahlin controller, IMC and DMC is explained. This material has been published as it was developed in a Technical Report (Segall

et al (1987b)), at conferences (Segall et al (1986), (1987a)) and as a journal article (Segall et al (1991)).

That Proportional Integral Derivative controllers written in velocity form can be subject to severe performance degradation when saturation is encountered has only rarely been observed in the literature (Åström and Wittenmark (1984)). Chapter 5 explains the nature of the problem, how it arises from the mathematics of the velocity form and how to fix it. The problems encountered in the tuning of velocity form controllers are also explained. New anti-reset windup protection algorithms are presented for both velocity and position forms of the PID controller. Examples are used to show the improvement which is obtained from the proposed solution. Most of this work has been published as a journal article (Segall and Taylor (1986)).

Conclusions and recommendations for further research are presented in Chapter 6.

CHAPTER 2

ROBUSTNESS ANALYSIS FOR SINGLE-INPUT-SINGLE-OUTPUT CONTINUOUS FEEDBACK CONTROL SYSTEMS WITH APPLICATIONS

In the design of a feedback control system, robustness is usually gained and performance lost as the controller is detuned. However, how rapidly robustness is gained and performance is lost is not usually available from the design method. In order to quantify robustness and performance, specific definitions must be chosen. In this chapter, a new definition of robustness is proposed which is the region of joint allowable variation in the gain and dead-time of the process for which the closed loop system is stable. A plot of this region, the allowable gain variation vs allowable dead-time variation, gives a detailed description of the robustness of a system. This definition is useful for comparing the robustness of controllers obtained by different design methods. A straightforward procedure for calculating the robustness region from the open loop frequency response is presented.

This definition of robustness extends the method of Tsypkin (1946) for calculating the robustness region of joint allowable variation in the process gain and dead-time parameters to systems which

can be made unstable by a reduction in the dead-time. His analysis only included systems which can be destabilized (or stabilized) by adding dead-time. This method of robustness analysis provides a clear indication of the robustness of a closed loop feedback system to changes in the process gain and the process dead-time. Most of the current parameter uncertainty region methods of robustness analysis (for example Katbab and Jury (1991)) apply only to systems where the characteristic equation is a polynomial; that is, the system does not have dead-time. Frequency domain robustness specifications of Palmor (1980), Doyle and Stein (1981), Morari and Zafiriou (1989) and Lewin (1991) require that the designer specify a priori the magnitude of the process uncertainty as a function of frequency, paying special attention to the "high" frequencies. The robustness region method described here allows a designer to calculate the allowable gain and dead-time change once a controller has been designed. The frequency domain construction used to calculate the robustness region makes it easy to understand the nature of any robustness limitations. The question of what is a "high" frequency is easily answered since the allowable change in the process dead-time incorporates frequency information in a natural and meaningful way. This definition of robustness is used to explore the robustness of a PI controller with a Smith Predictor dead-time compensator.

The use of the robustness region in combined performance and robustness analysis is demonstrated by an example where several different controller design procedures are compared for a first order plus dead-time process. The performance is measured in terms of the Integral Square Error (ISE). The controllers which are compared are the

Proportional-Integral-Derivative (PID) controller design procedure of Rivera et al (1984), the Linear Quadratic Optimal Control (LQOC) design procedure of Palmor (1982) and the Internal Model Control (IMC) design procedure as described in Rivera et al (1984). The comparison shows that for the same level of robustness PID has better performance than IMC which has better performance than LQOC. This is a surprising result as Rivera et al (1986) claimed that IMC gives 10% better performance than PID control for this process, but they do not compare the controllers at the same level of robustness.

This example shows that the robustness region is very useful for comparing the robustness of these very different controllers. The frequency response construction of the robustness region shows exactly which features of the frequency response are responsible for limiting the system robustness. The performance improvement of the IMC controller over the LQOC controller in the first order plus dead-time example is explained by comparing the frequency response of the controllers in their Smith predictor form.

The robustness analysis method is used to analyze the robustness implications of dead-time compensation. The high frequency peaks in the open loop frequency response are shown to cause contraction of the robustness region, and tuning of controllers is shown to reduce these peaks. Although it has previously been observed that high frequency peaks in the frequency response of dead-time compensating controllers cause robustness problems (Palmor (1980), Ioannides et al (1979)) there has been no quantitative analysis of how tuning works to reduce these peaks and improve robustness.

The insights from robustness analysis are used to explain why replacing dead-time with a first order Padé approximation (Rivera et al (1986)) results in a controller design with good robustness and good performance. The first order Padé approximation generates derivative action in the controller which approximates the effect of dead-time compensation with none of the peaks in the frequency response which can cause a reduction in the robustness. This feature is shown to be unique to the first order Padé approximation and is not shared by higher order Padé approximations to the dead-time.

2.1 Robustness Analysis: Construction of the Region of Joint Allowable Parameter Variation in the Process Gain and Dead-time

This section presents a definition of robustness as the region of joint allowable variation of the process gain and dead-time for which the closed loop system remains stable. A straightforward constructive method for calculating the robustness region is presented. The usefulness of this definition for comparing the robustness of very different controllers is demonstrated by comparing the robustness of three controllers for a first order plus dead-time system. The stability region is obtained from the Nyquist plot of the open loop transfer function frequency response by extending the method of Tsypkin (1946); the extensions added here accommodate systems where instability occurs due to decreased process dead-time. The robustness region will be shown to be a more useful measure of robustness than either the allowable change in each parameter individually or the traditional frequency domain measures of phase and gain margin. The plot of the

stability region allows a graphic comparison of the robustness of different controller designs.

The analysis of this section is primarily intended for systems with dead-time. Systems without dead-time are often well characterized by the gain and phase margin. Palmor (1980) demonstrates that the phase margin is not meaningful for systems with dead-time when dead-time compensating controllers are used. In particular he shows an example where a feedback control system has a large phase margin but where the feedback control system becomes unstable for even infinitesimal perturbations of the process dead-time. He terms this type of system a "practically unstable" system or an "infinitely sensitive" system. Rivera et al (1984) show that the M value measure of robustness, another frequency domain measure of robustness, suffers from the same problem: a feedback control system may have a large M value but have only slight tolerance for process dead-time changes.

Palmor (1982) and Rivera et al (1984) both address the robustness of dead-time compensating control in the frequency domain. Palmor (1982) defines an ignorance function I(s) which specifies the allowable perturbation to the process transfer function for which a given closed loop system will remain stable. Rivera et al (1984) and Laughlin et al (1987) use the method of Doyle and Stein (1981) where a bound on the process frequency response $l_m(\omega)$ is specified a priori and is then used in the design of a controller. A common thread in frequency domain robustness specifications (Morari and Zafiriou (1989) and Lewin (1991)) is that they require that the control system designer

specify a priori the magnitude of the process uncertainty as a function of frequency, paying special attention to the "high" frequencies.

Relatively few studies have examined the robustness of feedback control of dead-time compensated systems in terms of the parameters of the process transfer function. Tsypkin (1946) presents a technique for using the Nyquist plot to construct the joint region of allowable increases in the process gain and dead-time. The result is a two dimensional plot indicating the stability envelope of the closed loop system in the space of process gain increases and process dead-time increases. His analysis only included systems which could be destabilized (or stabilized) by an increase in the process dead-time. Ioannides et al (1979) examine the robustness of a proportional controller with a Smith Predictor dead-time compensator applied to a first order plus dead-time integrating process. They use a search technique to establish stability envelopes in terms of the various parameters of the system. Romagnoli et al (1988) analyze the robustness of a dead-time compensating self-tuning controller by finding a stability envelope in the parameter space of the process. Bergh and MacGregor (1986) examine the robustness of discrete dead-time compensating control systems with IMC controllers and LQOC controllers applied to a first order plus dead-time process. They also use a search technique to find the stability envelopes for several of the process parameters and disturbance parameters. Laughlin et al (1987) study Smith Predictor controller performance where uncertainty limits on the process gain, dead-time and time constant are specified a priori. They

convert the parameter uncertainty specification into an unstructured frequency domain uncertainty.

In this section the method of Tsypkin (1946) is extended to systems which can be destabilized by a decrease in the dead-time. This extension is particularly important for systems with dead-time compensating controllers. The method allows the robustness to be calculated directly from the Nyquist plot of the open loop system, as a stability region in the space of the process gain and the process dead-time. This not only allows systems which are practically unstable, as defined by Palmor (1980) to be recognized directly, it also allows systems which can be made practically unstable by a slight increase in process gain to be recognized. The use of the stability region as a measure of robustness allows the relative robustness of different controllers to be compared directly.

The robustness region method described in this chapter allows a designer to calculate the allowable gain and dead-time change once a controller has been designed. A priori specification of robustness is not required. The frequency domain construction used to calculate the robustness region makes it easy to understand the nature of any robustness limitations. Also, the question of what is a "high" frequency is easily answered since the allowable change in the process dead-time incorporates frequency information in a natural and meaningful way.

As a prelude to describing the details of the robustness calculations, it would be helpful to examine the robustness problems

which can occur when using a dead-time compensating controller. As an example, consider a first order plus dead-time process:

$$G_p(s) = \frac{K_p\, e^{-T_d s}}{T_p s + 1} \qquad \text{with model} \qquad G_M(s) = \frac{K_M\, e^{-T_{dM} s}}{T_M s + 1} \tag{2.1}$$

and a PI controller:

$$G_{C0}(s) = \frac{K_C\,(T_I s + 1)}{T_I s} \tag{2.2}$$

combined with a Smith Predictor dead-time compensator as shown in Figure 2.1. The overall controller in a unitary feedback structure, shown in Figure 2.2, is:

$$C(s) = \frac{K_C\,(T_I s + 1) \,/\, T_I s}{1 + K_C\, K_M\, (T_I s + 1)\,(1 - e^{-T_{dM} s})\,/\,T_I s\,(T_M s + 1)} \tag{2.3}$$

The open loop transfer function is then:

$$C(s)G_p(s) = \frac{K_C\,(T_I s + 1)}{T_I s + K_C\, K_M\,(T_I s + 1)\,(1 - e^{-T_{dM} s})\,/\,(T_M s + 1)} \cdot \frac{K_p\, e^{-T_d}}{(T_p s + 1)} \tag{2.4}$$

The model is assumed to be exact, and the numeric values chosen for the process in this example are K_P=K_M=1.0, T_P=T_M=1.0 and T_d=T_{dM}=1.0 minutes. The controller integral time is chosen equal to the process time constant, T_I=1.0 minutes and the controller gain is chosen to be K_C=10.0.

The Nyquist plot for this system is shown in Figure 2.3 and the Bode plot is shown in Figure 2.4. The most noticeable feature of the

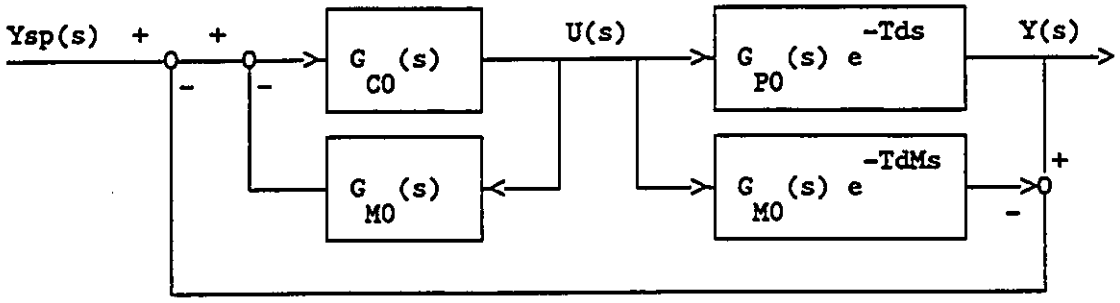


Figure 2.1 Smith Predictor Dead-Time Compensator Structure

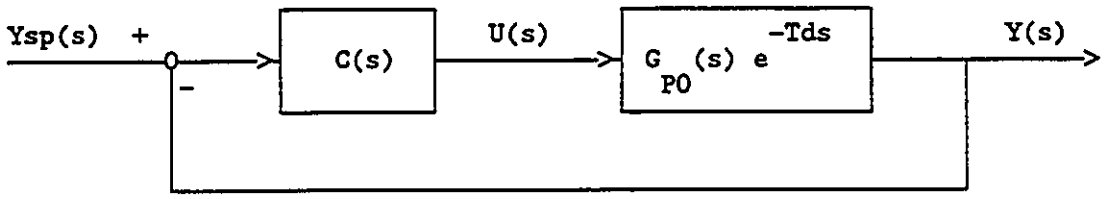


Figure 2.2 Unitary Feedback Structure

Nyquist plot is that although the open loop transfer function does not encircle the critical point (-1,0), there are several segments of the plot where the gain is larger than one. However, as long as the critical point is not encircled the closed loop system will be stable. Robustness is a measure of how much the process can differ from the process model and yet maintain the stability of the closed loop system, or equivalently maintain a Nyquist plot which does not encircle the critical point.

In order to understand robustness it is important to visualize some of the ways in which the process frequency response can vary. Changes in the process gain act to expand the Nyquist plot or to shrink it in the radial direction; that is, a change in the process gain moves

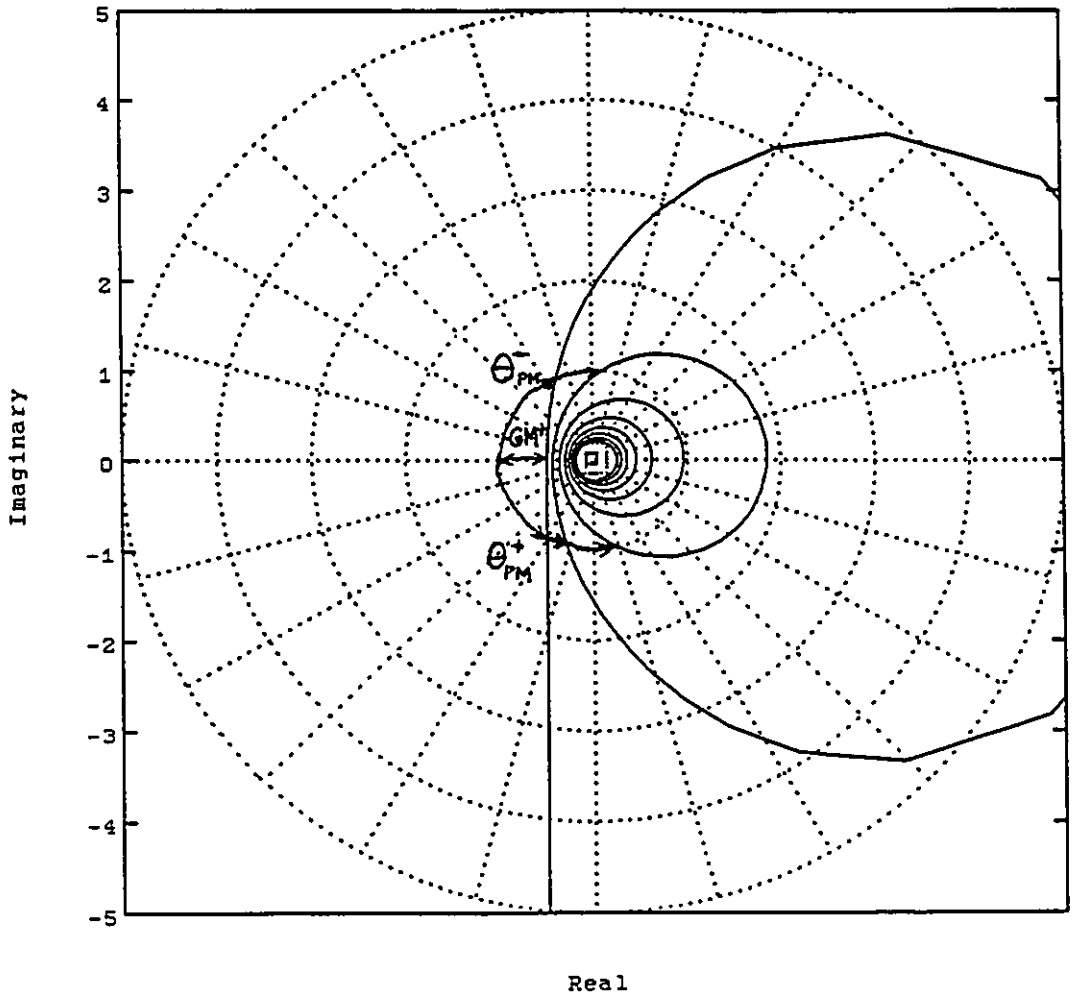


Figure 2.3 Nyquist Plot for Example

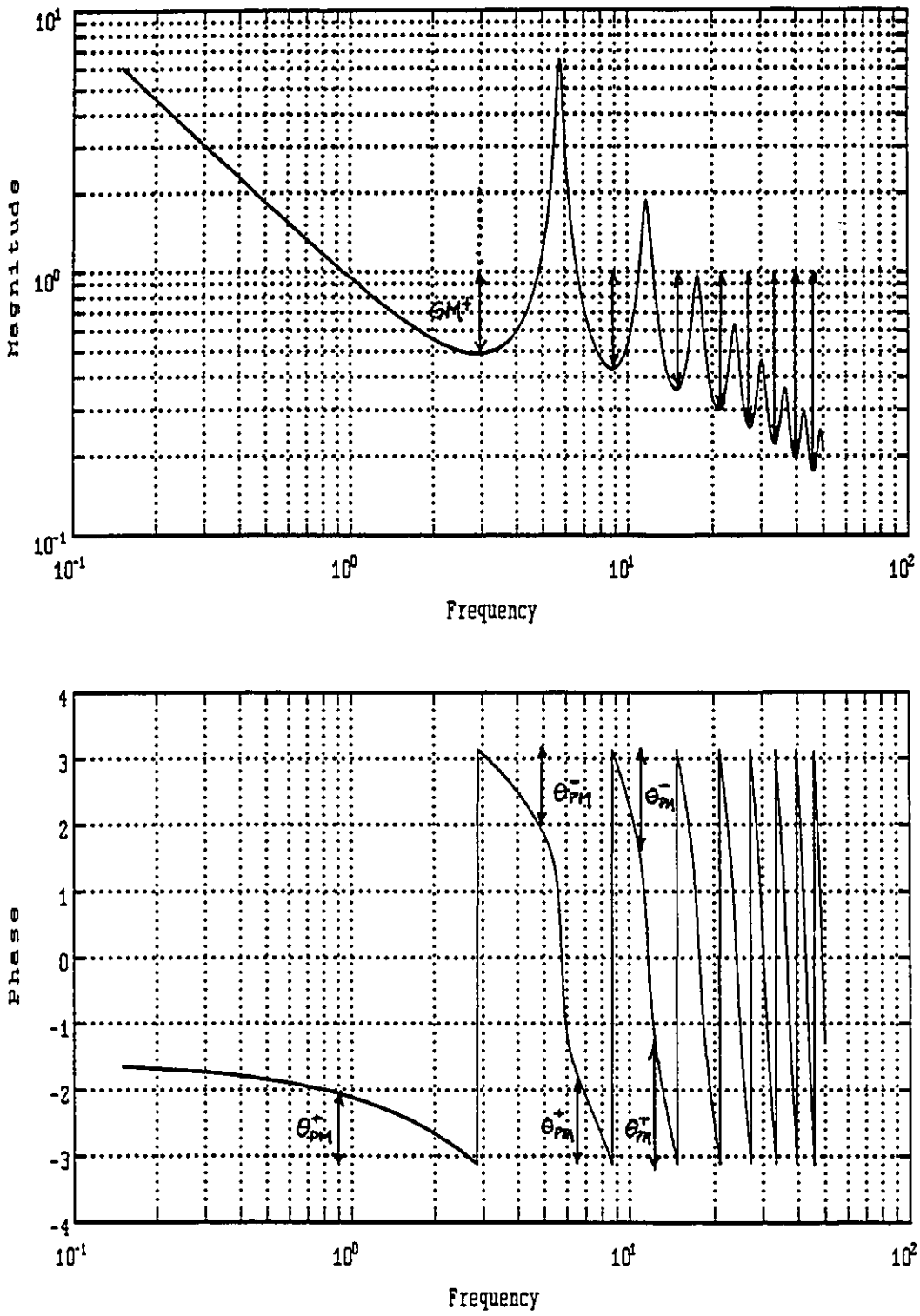


Figure 2.4 Bode Plot for Example

each point on the curve outward or inward along a line from the origin. Changes in the process phase act to rotate the Nyquist plot about the origin; that is, a change in the process phase changes the angular location of a point but does not change its radial distance from the origin. Increasing the process phase rotates the plot in a clockwise direction about the origin, whereas decreasing the process phase rotates the plot in a counter-clockwise direction. The effects of changing process gain and process phase for a particular point are indicated on Figure 2.3. Clearly the Nyquist plot for this example can be made to encircle the critical point by either increasing the process gain or by adding or subtracting phase from the process or by a combination of both. The factor by which the gain alone can be increased before the system becomes unstable is the gain margin. The amount by which the phase alone can be increased or decreased before the system becomes unstable is the phase margin. The gain and phase margins for this example are indicated on Figure 2.4. Note that there are several candidates for the phase margin as there are several lobes of the Nyquist plot which could be rotated about the origin to encircle the critical point, however, the phase margin is generally defined as the smallest among the possible values.

The Bode plot in Figure 2.4 shows the added dimension of frequency: the logarithm of the gain and the phase are each plotted against the logarithm of the frequency. The gain plot clearly shows the regions where the gain of the open loop process is greater than one. Examining the phase plot shows that the phase lag is never equal to $-\pi$ whenever the gain is greater than one; this is the Bode plot equivalent

of the Nyquist plot not encircling the critical point. The effects of changing the process gain or the process phase are easily seen on the Bode plot. Increasing the process gain moves the gain curve up while decreasing it moves the curve down, in both cases leaving the phase curve unchanged. Increasing the process phase moves the process phase curve down while decreasing it moves the curve up. These directions of change are indicated on Figure 2.4, as are the gain and the phase margin. Note that the phase margin is the smallest of several candidate values: there are several regions of frequency where the gain is larger than one and adding phase can cause instability in any of these regions.

These changes in the process frequency response have equivalents in the parameters of the process. Changing the process gain parameter K_p changes the process frequency response gain without affecting the phase. The only parameter which affects the process frequency response phase without affecting the gain is the dead-time T_d. Unlike a change in the gain parameter, the amount of phase added by a change in the process dead-time increases with frequency. An additive perturbation δT_d in the process dead-time adds $\delta T_d w$ to the process phase at each frequency w. So, in order to calculate the amount of dead-time perturbation which is equivalent to a given phase margin one must know the frequency at which the phase margin is calculated.

This last observation that the phase increase due to added dead-time is a linear function of frequency allows a clear understanding of the problem that is caused by the multiple candidate choices for the phase margin. For each of the candidates for phase margin in the example, Table 2.1 shows the size of the phase margin, the frequency at

Table 2.1

Frequency, Phase Margin and Allowable Dead-Time Variation for Example

Frequency (Rad/Min)	Phase Margin (Rad)	Allowable Dead-time Variation (Min)
0.95	1.05	1.11
4.84	-1.17	-0.24
6.63	1.28	0.19
11.02	-1.70	-0.15
12.31	1.86	0.15

which it occurs and the equivalent amount of dead-time which would add that amount of phase.

Note that the process can be destabilized by removing phase, so that decreasing the process dead-time can also lead to instability. From Table 2.1 it is evident that it is the phase margins at high frequency which limit the amount of dead-time which can be added to or subtracted from the process. Unlike the phase margin, the allowable change in the process dead-time incorporates frequency information. So, as a single measure of robustness, the allowable change in the dead-time is more informative than the phase margin. Palmor (1980) has created a pathological example where an open loop frequency response has a large phase margin and yet has an infinitely small allowable dead-time change.

It is also important to consider the effect of simultaneous changes in the process gain and dead-time. It is easy to see that the calculations of Table 2.1 can be repeated for the Nyquist plot where the process gain is increased by any gain factor less than the gain margin. In fact, the gain can be increased from the nominal process gain to the gain margin and the allowable changes in the dead-time can be calculated at each step. This results in a plot of the robustness region of joint

allowable changes in the process gain and dead-time parameters. This region is plotted for the example in Figure 2.5. Note that the allowable gain change is a multiplicative value (like gain margin) and the allowable dead-time change is an additive value (like phase margin).

The region of joint allowable variation in the process gain and dead-time parameters provides a graphic description of the robustness of a feedback control system. The allowable dead-time parameter variation incorporates frequency information into the robustness specification, information which is not contained in the single frequency domain robustness specifications such as phase margin or the M value used by Rivera et al (1984). Furthermore, the robustness is defined in terms with which a control system designer is usually familiar: the gain and dead-time of the process being controlled. In general it is easier to specify robustness requirements in terms of the allowable variation in these parameters than to specify an unstructured frequency domain uncertainty such as the $l_m(\omega)$ of Doyle and Stein (1981). Lastly, the use of the Nyquist plot to construct the robustness region helps give insight into the source of robustness problems. Use of the robustness analysis method to help provide insight into the robustness problems associated with dead-time compensation will be explored in the next section.

The Nyquist stability criterion is a standard tool of stability analysis: given a unitary feedback structure consisting of a controller transfer function C(s) and a process transfer function $G_p(s)$ (Figure 2.2), if the curve generated by the complex frequency response $C(i\omega)G_p(i\omega)$ as ω goes from $-\infty$ to $+\infty$ (excluding poles of $C(i\omega)G_p(i\omega)$ on

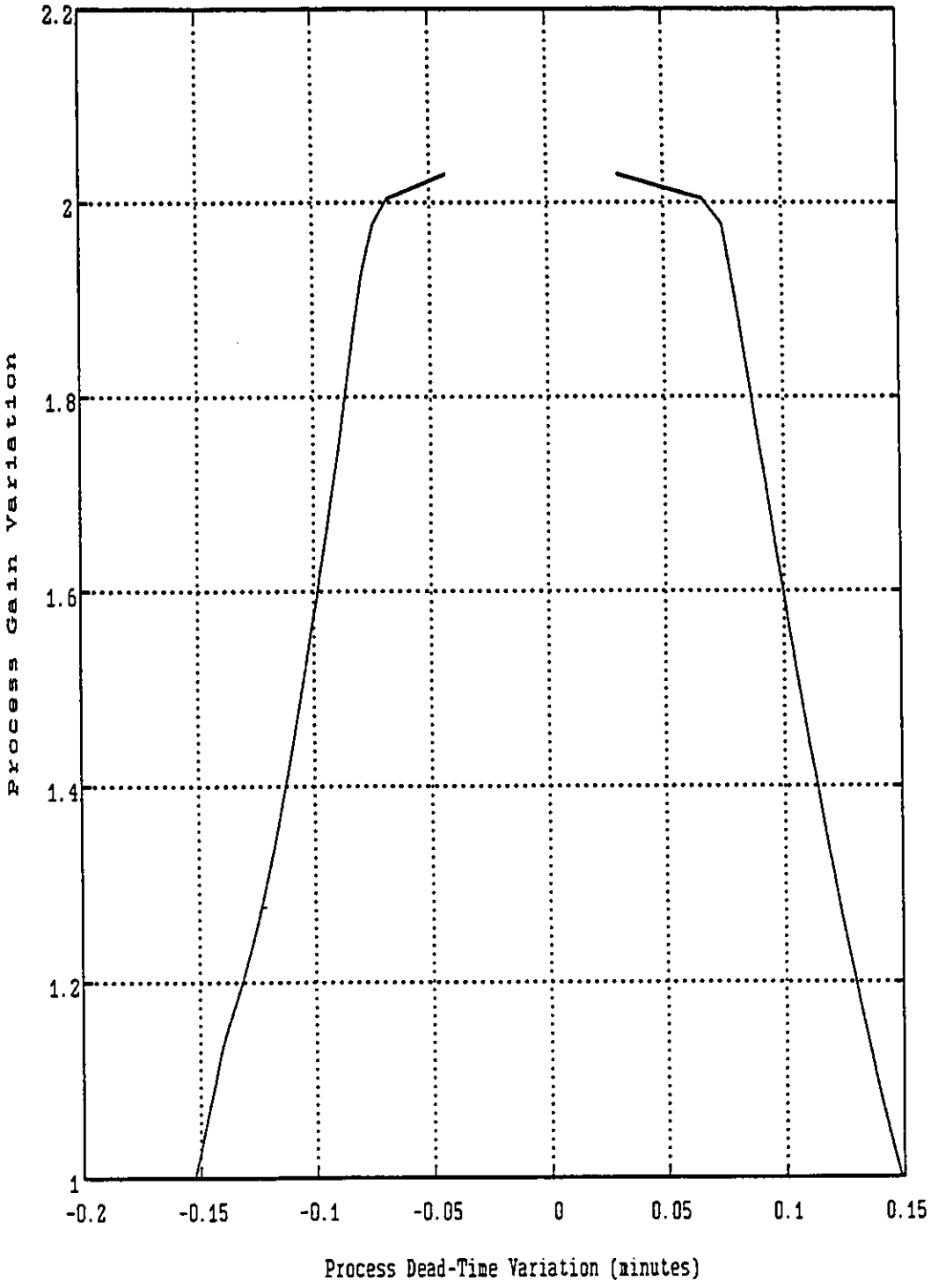


Figure 2.5 Robustness Region for Example

the imaginary axis) does not encircle the critical point -1+i0 then the closed loop system will be stable. This somewhat simplified form of the Nyquist criterion is used here where it is assumed that both C(s) and $G_p(s)$ are open loop stable. The following development could easily be modified for a more rigorous definition of the Nyquist Criterion. The unitary feedback structure is not limiting as it is easily obtained even for quite complex controllers by block diagram manipulations.

Consider an open loop system frequency response:

$$C(i\omega)\ G_p(i\omega) = C(i\omega)\ K_p\ \ G_{p0}(i\omega)\ e^{-(T_d\ i\omega)} \tag{2.5}$$

The open loop transfer function in equation (2.5) is written with the process gain K_p and the process dead-time T_d appearing explicitly. If the process dead-time is varied from its nominal value by an additive perturbation δT_d and the process gain is varied from its nominal value by a multiplicative perturbation δK_p then the transfer function becomes

$$C(i\omega)\ G_p(i\omega) = C(i\omega)\ K_p\ \delta K_p\ G_{p0}(i\omega)\ e^{-([T_d\ +\ \delta T_d]\ i\omega)} \tag{2.6}$$

The dead-time change affects only the phase of the open loop frequency response, shifting it by $\delta T_d\omega$. The phase change thus depends on the size of the dead-time perturbation and increases with frequency. If δT_d is positive then the phase is increased and if δT_d is negative (note: $T_d+\delta T_d>0$) then the phase is decreased. The multiplicative gain change δK_p affects only the gain of the open loop frequency response, increasing it or decreasing it by a factor of δK_p.

The allowable dead-time parameter variation δT_d is defined as the amount of dead-time which if added to the process dead-time will

bring the closed loop system to the limit of instability. It can be calculated directly from the phase margin. The phase margin is defined here as the amount of phase which if added to or subtracted from the Nyquist plot of a closed loop stable system will bring the closed loop system to the verge of instability. Equation (2.7) defines the crossover frequencies w_{co}: the frequencies where the Nyquist Plot crosses the unit circle

$$\text{whenever} \quad \left| C(i\omega)\, G_p(i\omega) \right| = 1 \qquad \text{then} \quad \omega_{co} = \omega \tag{2.7}$$

Equation (2.8) separates the crossovers into two types: positive crossovers where the gain is decreasing, indexed by n, and negative crossovers where the gain is increasing, indexed by m.

$$\text{if} \quad \frac{d \left| C(i\omega)\, G_p(i\omega) \right|}{dw} \quad \begin{cases} < 0 & \text{then} \quad \omega_{co}^{n+} = \omega_{co} \\ > 0 & \text{then} \quad \omega_{co}^{m-} = \omega_{co} \end{cases} \tag{2.8}$$

Equation (2.9) defines the phase margin θ_{pm}; a positive crossover leads to a positive phase margin, indexed by n, and a negative crossover leads to a negative phase margin, indexed by m.

$$\theta_{pm}^{n+} = \pi + \angle\, C(i\omega_{co}^{n+})\, G_p(i\omega_{co}^{n+}) \quad \text{or} \quad \theta_{pm}^{m-} = \angle\, C(i\omega_{co}^{m-})\, G_p(i\omega_{co}^{m-}) - \pi \tag{2.9}$$

The phase operator $\angle$ is assumed to produce positive values in the upper half plane and negative values in the lower half plane with absolute values less than or equal to π. The phase margin may be multiple valued with as many values as there are crossover frequencies.

The allowable dead-time perturbation δT_d which will bring a stable closed loop system to the limit of stability is calculated from the phase margin:

$$\delta T_d^+ = \min_{\text{all } n} \left[\theta_{pm}^{n+} / \omega_{co}^{n+} \right] \quad \text{and} \quad \delta T_d^- = \max_{\text{all } m} \left[\theta_{pm}^{m-} / \omega_{co}^{m-} \right] \tag{2.10}$$

The allowable dead-time perturbation has only two values: one positive and one negative. When there are multiple phase margin values, the smallest value will not necessarily determine δT_d as the calculation depends inversely on the frequency of the crossover. If δT_d^+ does not exist, then it has the value $+\infty$. If δT_d^- does not exist, then it has the value $-T_d$.

Figure 2.6 shows a more complex Nyquist plot which illustrates a system with both a positive phase margin $\theta_{pm}^+ = +2.1$ rad at ω=2.1 rad/min and a negative phase margin $\theta_{pm}^- = -2.0$ at w=1.8 rad/min. The maximum allowable dead-time perturbations while retaining closed loop stability are $\delta T_d^+ = 1.0$ min and $\delta T_d^- = -1.1$ min. One cannot say, however, that the system will be unstable for all dead-time perturbations greater than 1.0 minutes or less than -1.1 minutes. For example δT_d=2.1 min will shift the ω=1.8 rad/min point on the Nyquist curve by 5.6 rad and the ω=2.1 rad/min point by 6.51 rad; that is, the portion of the curve with magnitude greater than 1 would be rotated approximately 2π rad returning very closely to its initial closed loop stable orientation. For robustness analysis considerations, however, it is sufficient to find the smallest values of δT_d which will lead to instability.

The allowable gain parameter variation δK_p is defined as the value which if multiplied by the process gain will bring the closed loop

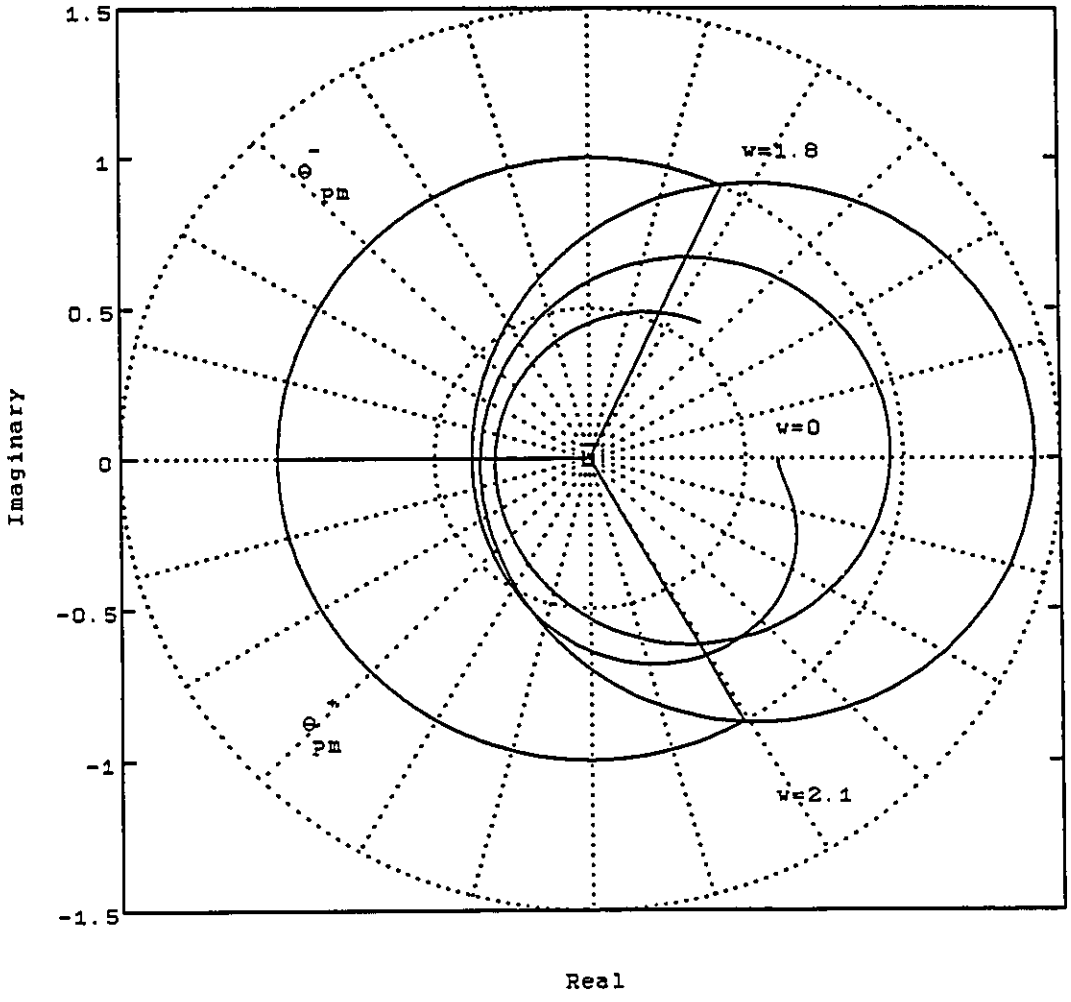


Figure 2.6 Nyquist Plot for Example

system to the limit of instability. The size of the allowable gain parameter variation is identical to the gain margin: the increase or decrease in the process gain which will lead to an encirclement of the critical point -1+0i by the Nyquist plot. The allowable gain parameter variation is determined from the Nyquist plot. Equation (2.11) defines the gain at the points where the Nyquist plot crosses the negative real axis.

$$R = \left| C(i\omega)\, G_p(i\omega) \right| \quad \text{whenever } \angle\, C(i\omega)\, G_p(i\omega) = \pm \pi \tag{2.11}$$

Equation (2.12) separates these gain values into those less than 1, indexed by m, and those greater than 1, indexed by n.

$$R_m = R \quad \text{if} \quad R < 1, \quad R_n = R \quad \text{if} \quad R > 1 \tag{2.12}$$

The positive allowable gain variation δK_p^+, defined in equation (2.13), is the smallest multiplicative gain perturbation larger than 1 which will bring the closed loop system to the limit of stability. The negative allowable gain variation δK_p^- is the largest multiplicative gain perturbation less than 1 which will bring the closed loop system to the limit of stability.

$$\delta K_p^+ = \min R_m^{-1} \qquad \text{and} \qquad \delta K_p^- = \max R_n^{-1} \tag{2.13}$$

The δK_p^+ and δK_p^- are both positive numbers; positive and negative are used here to indicate a gain increase and a gain decrease respectively. If either δK_p^+ or δK_p^- does not exist, then the gain can be increased or decreased indefinitely respectively.

The allowable gain variation δK_p, like the allowable dead-time variation δT_d, is the minimum gain increase or decrease which will make the closed loop system unstable and does not imply that the system will be unstable for all larger or smaller gains respectively. Again, for robustness analysis it is the first region of instability which is most important: other regions of stability beyond the first onset of instability are not considered here. Most systems do not have a value of δK_p, which means they can not be made unstable by decreasing the process gain.

It is important to note that the allowable dead-time variation δT_d and the allowable gain variation δK_p are orthogonal in polar coordinates: δK_p is oriented in the radial direction (as is gain margin) and δT_d is oriented in the angular direction (as is phase margin) as illustrated in Figure 2.3. Thus, any other form of parameter variation can be separated into components representing gain change and dead-time change. However, the size of these components will vary with frequency. It is also important to note that it is the region of joint allowable gain and dead-time variation which determines robustness and not the individual limits. It will be shown in an example in this section that it is quite possible to have a large allowable dead-time parameter variation at the nominal process gain, while the allowable dead-time parameter variation is reduced dramatically when the process gain is increased slightly.

The robustness of a feedback control system is now defined as the region of joint allowable gain and dead-time variation. The procedure for constructing the region of joint allowable gain and dead-

time variation from the Nyquist plot is described here. The allowable multiplicative gain variation is first calculated from equations (2.11) to (2.13). The gain variation δKP is stepped incrementally from δK_P^- (or 1 if δK_P^- does not exist) to δK_P^+, the limit beyond which the closed loop system is unstable. The allowable dead-time variation δT_d^- to δT_d^+ will change as the gain is changed. This change occurs since altering of the gain by a factor of δK_p changes the location of the crossover frequencies. The calculations of equations (2.7) through (2.10) are repeated at each value of the gain to determine the allowable dead-time parameter limits δT_d^- to δT_d^+. This region can be plotted as δKP vs δT_d^- and δKP vs δT_d^+ and represents the robustness of the control system.

Reconsidering our example of a first order plus dead-time process with a Smith Predictor dead-time compensating controller will illustrate the importance of defining robustness as the region of joint allowable dead-time and gain parameter variation. The numeric values chosen for the process model in this example are $K_M=K_P=1.0$, $T_M=T_P=1.0$ and $T_{dM}=T_d=1.0$ minutes. The controller integral time is chosen equal to the process time constant, $T_I=1.0$ minutes. Three different values of the controller gain are used $K_C=3.33$, 2.5 and 1.5.

The Nyquist plot for this system with each of the controller gains is shown in Figure 2.7. The three systems are compared for robustness in Figure 2.8 where the region of joint allowable variation in the gain and dead-time parameters is plotted. All three of these controllers have an allowable gain parameter variation of over 2 (and the gain can be reduced to zero). The system with $K_C=3.33$ has very small both positive and negative allowable dead-time variation limits

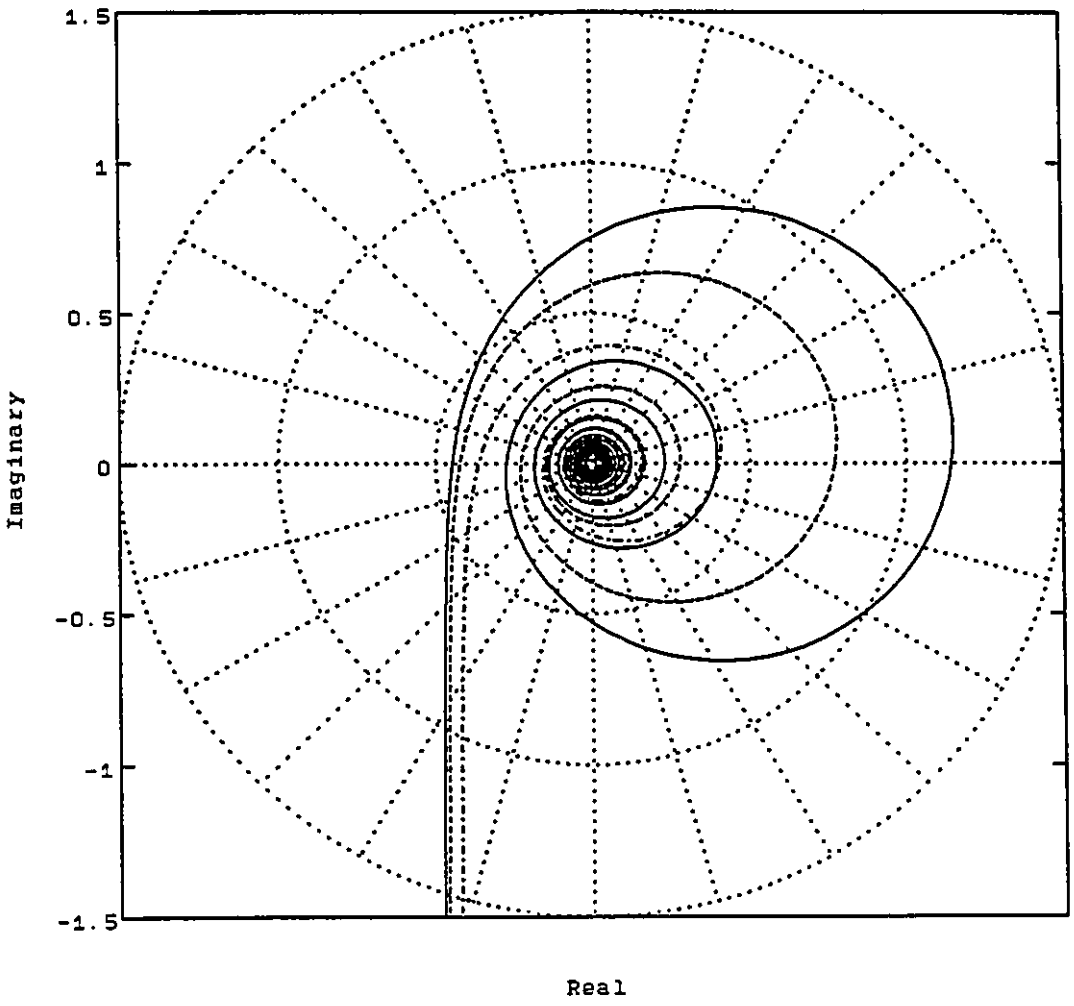


Figure 2.7 Nyquist Plot of PI + Smith Predictor Control
- K_c=3.33, -- K_c=2.5, -. K_c=1.5

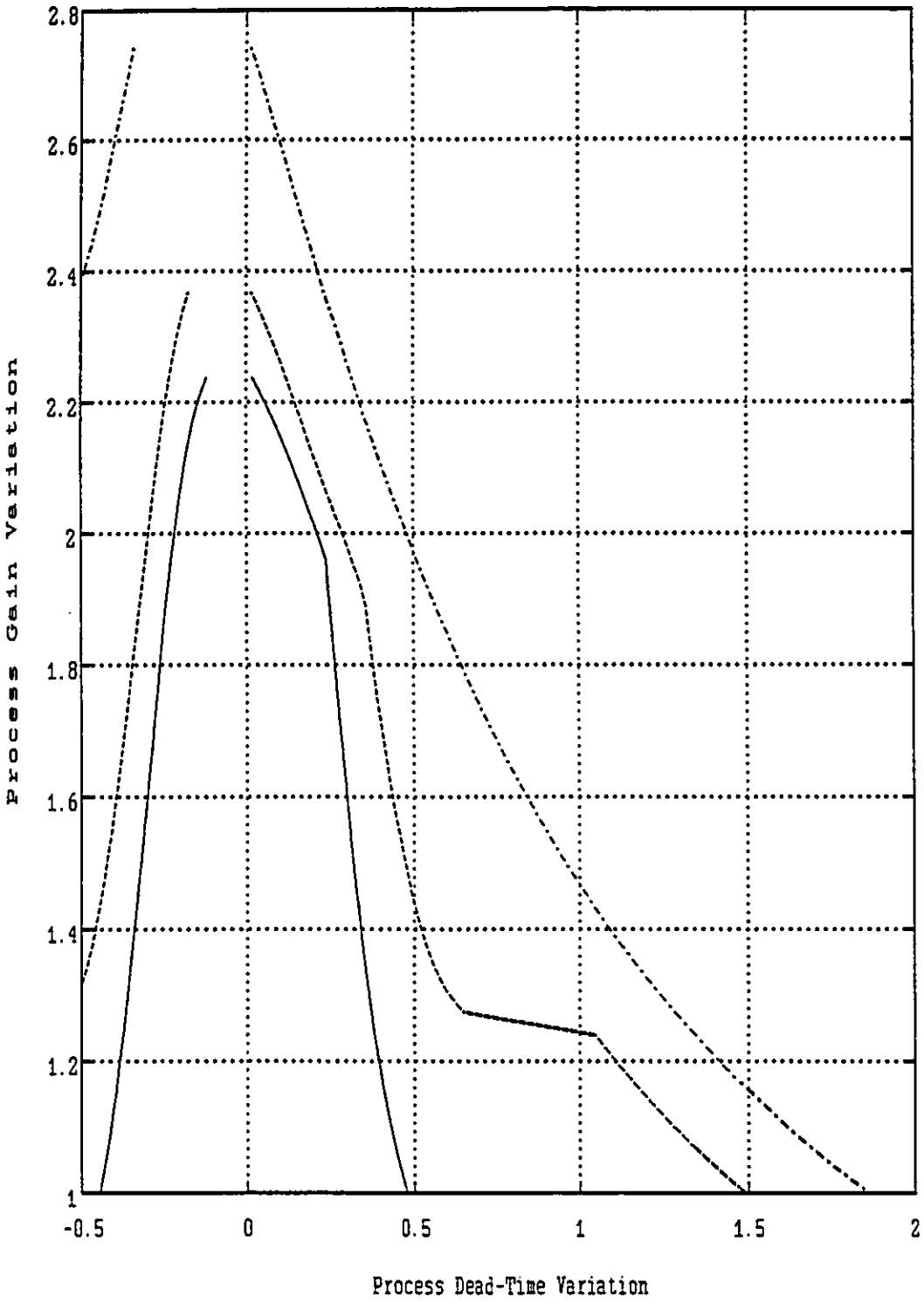


Figure 2.8 Robustness Plot PI + Smith Predictor
– K_c=3.33, -- K_c=2.5, -. K_c=1.5

across the whole range of allowable gains. When the controller gain is set to K_c=2.5 the positive allowable dead-time limit is moderate and there is no negative limit at the nominal gain. However, when the gain is increased slightly from its nominal value the allowable dead-time limits contract rapidly in both directions. This shows the importance of plotting the whole robustness region rather than examining the individual limits alone. It is possible to have acceptable individual limits and yet have significantly tighter parameter variation limits when the parameters are moved jointly. The change in slope of the positive dead-time limit when δK_p=1.9 is due to a change from the high frequency and large phase margin crossover controlling the dead-time variation, to the low frequency and low phase margin crossover controlling it. Lastly, when the controller gain is set to K_c=1.5 the allowable dead-time variation has no negative values until the gain of the process is increased by a factor of 2.4.

All three of these systems have phase margins of greater than $\pi/3$, yet they have dramatically different allowable dead-time parameter variation. This is due to the different frequencies at which the crossovers occur. The important difference between phase margin and the allowable dead-time parameter variation is that δTd incorporates frequency information as shown in equation (2.10). The system with controller gain K_c=3.33 approaches the type of system that Palmor (1980) terms an infinitely sensitive system: a system with significant phase margin but which has near zero tolerance to dead-time changes. Such systems have gain crossovers at near infinite frequency so the allowable dead-time perturbation δT_d is near zero.

In this section a new definition of robustness has been advanced as the region of joint allowable variation in the process gain and dead-time parameters. This definition of robustness is expressed in terms familiar to a control system designer: the gain and dead-time of the process being controlled. A plot of the robustness region provides a graphic description of the robustness of the feedback control system. The robustness region is calculated directly from the Nyquist plot of the frequency response of the open loop system. This construction of the robustness region from the Nyquist plot allows it to deal very naturally with the complicated frequency response of systems with dead-time compensating controllers.

This definition of robustness extends the method of Tsypkin (1946) to systems which can be destabilized by decreasing the process dead-time. The allowable dead-time parameter variation incorporates frequency information into the robustness specification and bypasses the problem with single frequency domain robustness specifications such as phase margin or the M value used by Rivera et al (1984). The frequency response information in the allowable dead-time also obviates the need to specify a priori what is a "high" frequency and what is the shape of an unstructured frequency domain uncertainty such as the $l_m(\omega)$ of Doyle and Stein (1981). Lastly, the use of the Nyquist plot to construct the robustness region helps give insight into the cause of robustness problems. An example was used to show how this robustness analysis can be used to assess the robustness of a feedback control system with a dead-time compensating controller and to assess the improvements in

robustness as the controller is tuned. The use of this new method of robustness analysis method will be explored in the next section.

2.2 Application of Combined Robustness and Performance Analysis to Controller Design for a FOPDT System: Step Set Point Changes

In this section, an illustrative example is used to show the application of the robustness analysis described in the previous section, together with a performance analysis, to the design of a controller for a First Order Plus Dead-Time (FOPDT) system with step changes in set point. The FOPDT system of equation (2.1) is a commonly used example in the chemical engineering control literature. The Integral Square Error (ISE) measure of performance is chosen as it is commonly used in the optimal control literature.

The controller design procedures which are compared are the Proportional-Integral-Derivative (PID) controller design method of Rivera et al (1984), the Internal Model Control (IMC) design procedure described in the same paper, and the Linear Quadratic Optimal Control (LQOC) design procedure of Palmor (1982). The PID controller of Rivera et al (1984) can not be implemented for step changes in set point, because the impulse inputs it generates are not available in most chemical processes, as discussed by Harris and Tyreus (1987). However, it is considered here as it is a design which is treated commonly in the chemical engineering control textbooks. This example provides an interesting comparison of the robustness versus performance tradeoff of different controllers as they are detuned.

2.2.1 PID controller design for a FOPDT system

The Rivera et al (1984) PID design procedure for FOPDT processes produces the controller which approximately minimizes the ISE for a step change in input:

$$C_{PID}(s) = \frac{K_C\,(T_I s + 1 + T_I T_D s^2)}{T_I s} \tag{2.14}$$

where the controller parameters are related to the process parameters through:

$$K_C = (2T_P + T_d) \,/\, (K_P\,(2\epsilon + T_d))$$

$$T_I = T_P + (T_d \,/\, 2) \tag{2.15}$$

$$T_D = T_P T_d \,/\, (2T_P + T_d)$$

and the value of the tuning parameter which is recommended is $\epsilon = 0.8T_d$. With this tuning the ISE for a step change in set point is approximately $1.1T_d\delta r^2$, where δr is the size of the change in the set point. The open loop transfer function is then:

$$C_{PID}(s)\,G_P(s) = \frac{K_C\,(T_I s + 1 + T_I T_D s^2)}{T_I s} \cdot \frac{K_P\, e^{-T_d s}}{T_P s + 1} \tag{2.16}$$

In this section we are assuming that the process model is known so that we can use the parameters in the controller design. It is precisely the robustness of the feedback control system to this assumption that we are investigating.

2.2.2 LQOC controller design for a FOPDT system

The LQOC design procedure of Palmor (1982) produces the optimal controller which minimizes the criterion

$$E\left(y(t)^2 + \lambda \left(\frac{d^r u(t)}{dt^r}\right)^2 \right) \tag{2.17}$$

where E is the expectation operator. The controller design is chosen using the disturbance generating system DGS(0,1,0) (in Box and Jenkins (1970) notation this is an ARIMA(0,1,0) process) which describes a random walk disturbance:

$$DGS(0,1,0)(s) = \frac{\alpha(s)}{s^r \beta(s)} = \frac{1}{s} \tag{2.18}$$

Designing for this disturbance is equivalent to designing for a step set point change. The resulting controller is given as the controller block in a Smith Predictor structure:

$$G_{LQOCO} = \frac{K_C \dfrac{T_P s + 1}{s}}{(C_2/C_1)s + 1} \tag{2.19}$$

where

$K_C = 1 / C_1 K_P$

$C_1 = (\lambda + 2T_P/\lambda)^{1/2}$

and

$C_2 = T_P/\lambda$

This is a PI controller with an additional first order filter where the gain of the controller and the time constant of the filter depend on the value of the tuning parameter λ. The choice of the tuning parameter λ

is left to the discretion of the designer, although guidelines are given for avoiding practically unstable systems.

The open loop transfer function is:

$$G_{LQOC}(s)\,G_p(s) = \frac{K_C\,(T_p s + 1)}{(C_2 / C_1)\,s^2 + s + K_C K_p\,(1 - e^{-T_d s})} \cdot \frac{K_p\,e^{-T_d s}}{(T_p s + 1)} \qquad (2.20)$$

2.2.3 IMC controller design for a FOPDT system

The IMC design procedure of Rivera et al (1984) combines a model inverse $G_C(s)$, a first order filter F(s) with time constant ϵ, and the process model as shown in Figure 2.9. This controller has an equivalent

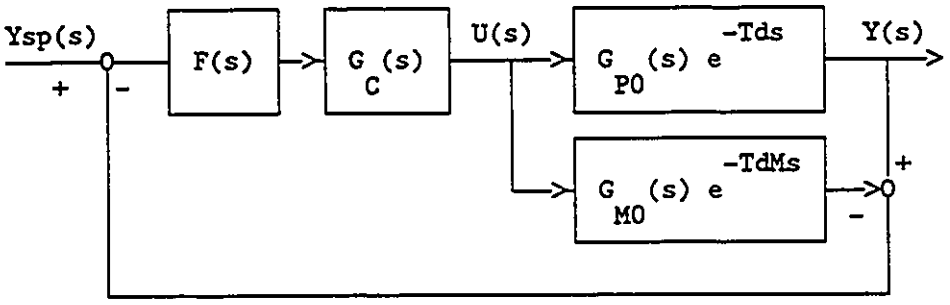


Figure 2.9 IMC Controller Structure

representation as a Smith Predictor where the controller block is given by:

$$G_{IMCO} = K_C \frac{T_p s + 1}{s} \qquad (2.21)$$

where

$K_C = 1 / K_p \epsilon$

This is a PI controller where the gain depends inversely on the tuning parameter ϵ. The recommended tuning is to select ϵ to be the desired closed loop time constant; however, no direct relationship between tuning and performance or robustness is given. The open loop transfer function is:

$$G_{IMC}(s)\,G_p(s) = \frac{K_C\,(T_p s + 1)}{s + K_C K_p\,(1 - e^{-T_d s})} \cdot \frac{K_p\,e^{-T_d s}}{(T_p s + 1)} \qquad (2.22)$$

2.2.4 ISE performance calculations

The ISE performance is calculated by integrating the spectral density of the error across all frequencies from $-i\infty$ to $+i\infty$. The spectral density of the error is the product of the spectral density of the transfer function from the setpoint r(s) to the error e(s) with the spectral density of the setpoint. The integral is:

$$ISE = \int_{-i\infty}^{+i\infty} G_{re}(s)G_{re}(-s)G_r(s)G_r(-s)\,ds = \int_{-i\infty}^{+i\infty} \frac{1}{1+C(s)G_p(s)} \frac{1}{1+C(-s)G_p(-s)} \frac{1}{s} \frac{1}{-s}\,ds \qquad (2.23)$$

2.2.5 Robustness vs performance comparison of PID vs LQOC vs IMC

The numerical values which are chosen for the example are the process gain K_p=1.0, time constant T_p=1.0 minutes, the dead-time T_d=1.0 minutes and the size of the step change in set point is δr=1.0. The minimum possible value of the ISE is $T_d(\delta r)^2$=1.0.

The values of the PID controller parameters are calculated using the formulae of section 2.2.1: the gain K_P=1.1538, the integral time T_I=1.5 and the derivative time T_D=0.3333. The Nyquist plot of the open loop frequency response and the plot of the region of joint allowable gain and dead-time variation are shown in Figures 2.10 and 2.11 respectively. At the nominal gain, the positive allowable additive dead-time variation is 1.4 minutes and there is no negative value. At the nominal dead-time, the allowable multiplicative gain variation is 2. The ISE response to the step set point change is 1.11. Rivera et al (1984) state that with the recommended tuning the ISE performance is at a minimum while the robustness, as measured by the M value, increases monotonically with the tuning parameter ϵ. That is, this tuning represents the minimum ISE performance available from this controller, however more robustness could be obtained at the expense of performance.

The LQOC design procedure allows controller designs for ISE values arbitrarily close to 1.0; however, in order to obtain the limiting value the tuning parameter must be chosen as λ=0 which results in an infinite gain K_C=∞ for the controller of equation (2.19). The frequency response of the open loop transfer function of equation (2.20) with λ=0 and the robustness region are plotted in Figure 2.12 and Figure 2.13 respectively. The frequency response is a straight line, which traces repeatedly from $-.5-i\infty$ to $-.5+i\infty$ as the frequency increases and the ends of these lines are connected by large loops in the right half plane. The open loop transfer function thus has gain crossovers at infinite frequency with phase margin values of $\pm\pi/3$, resulting in an allowable dead-time variation of $\pm$0: an infinitely sensitive system.

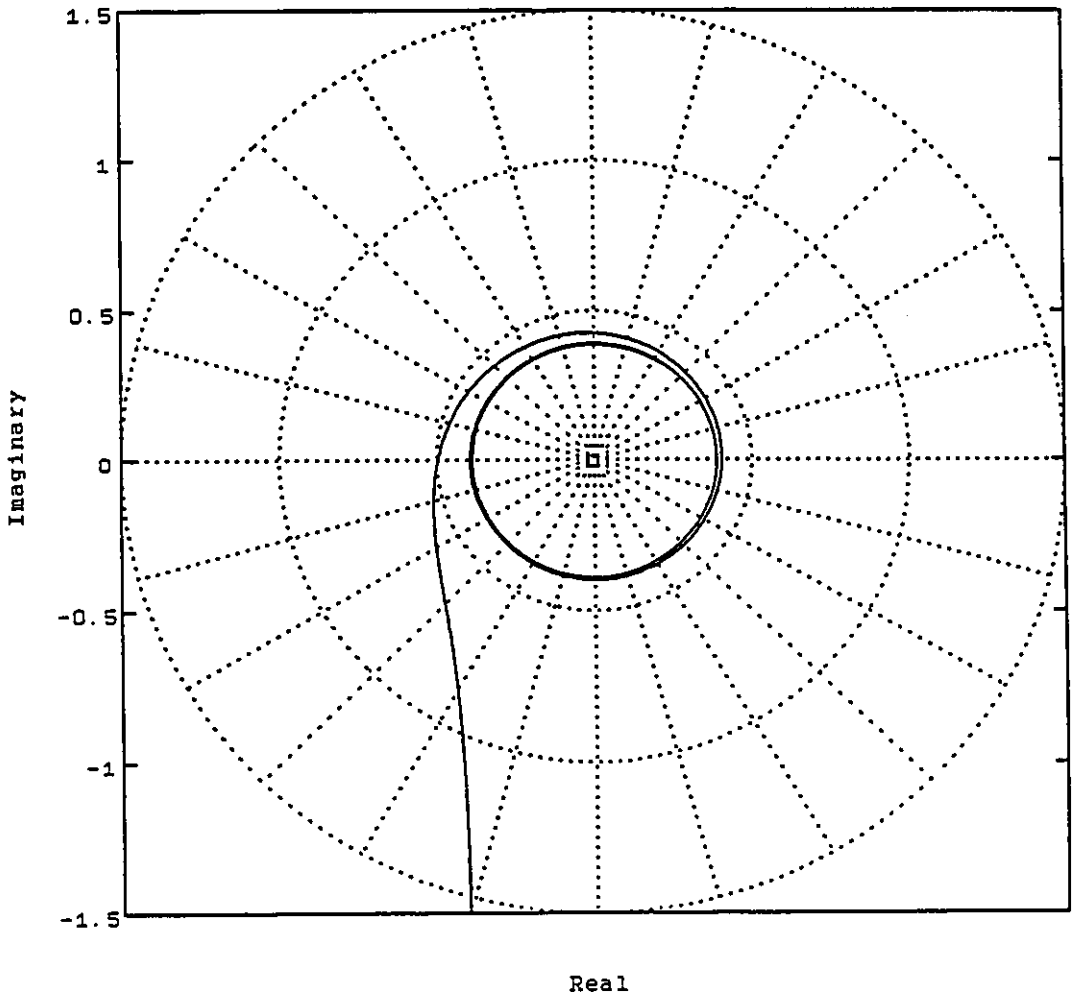


Figure 2.10 Nyquist Plot of PID Control

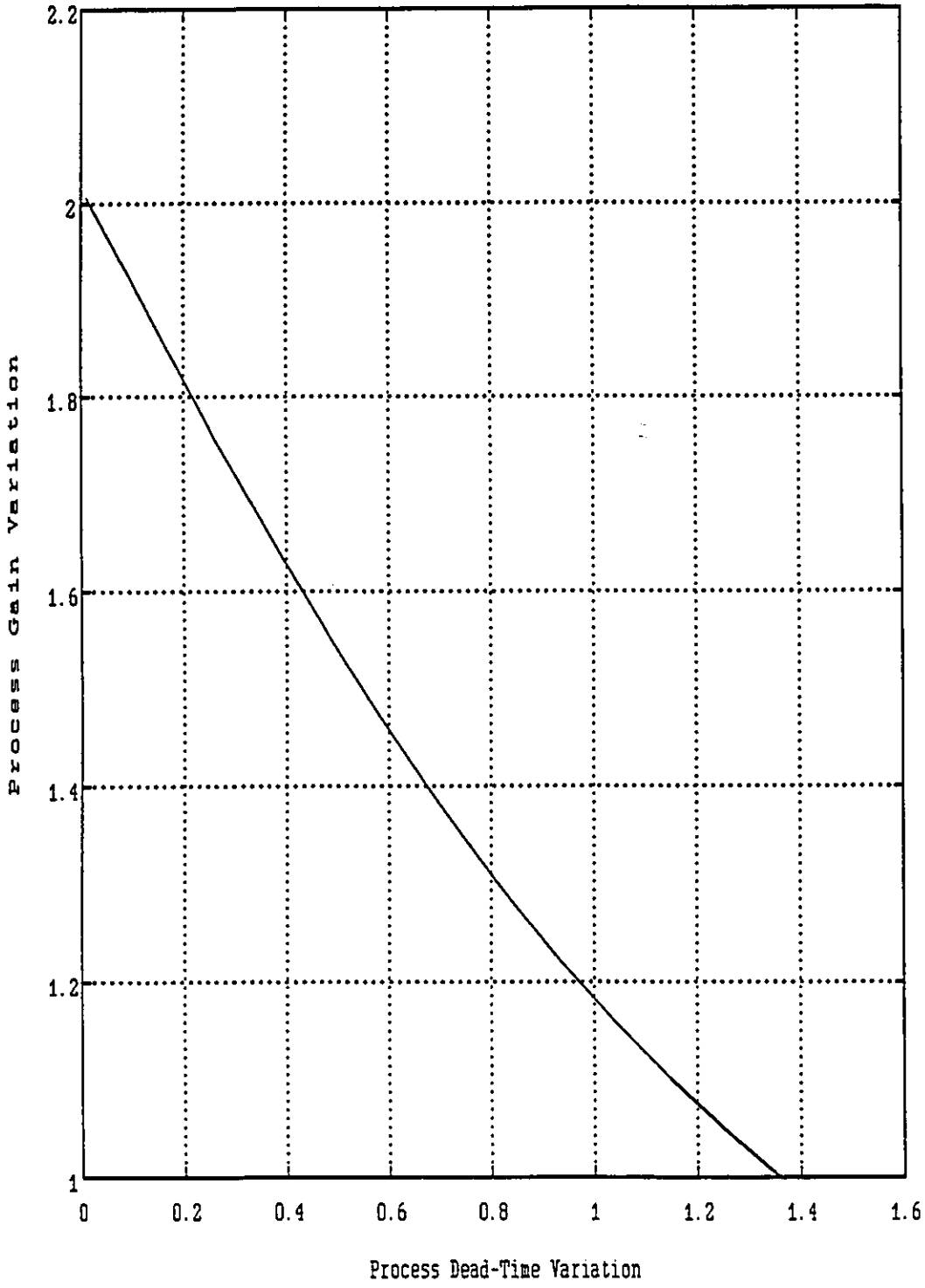


Figure 2.11 Robustness Region PID Control

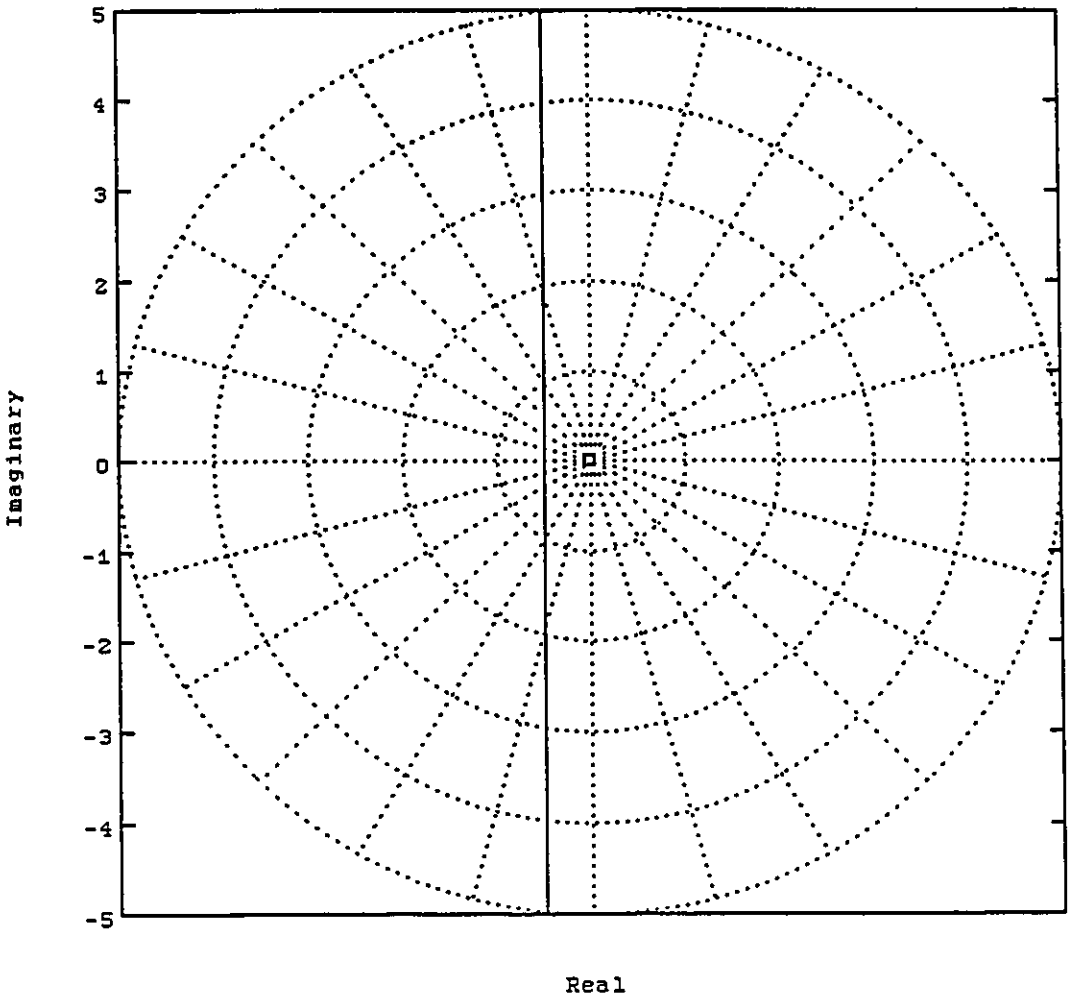


Figure 2.12 Nyquist Plot LQOC Control λ=0

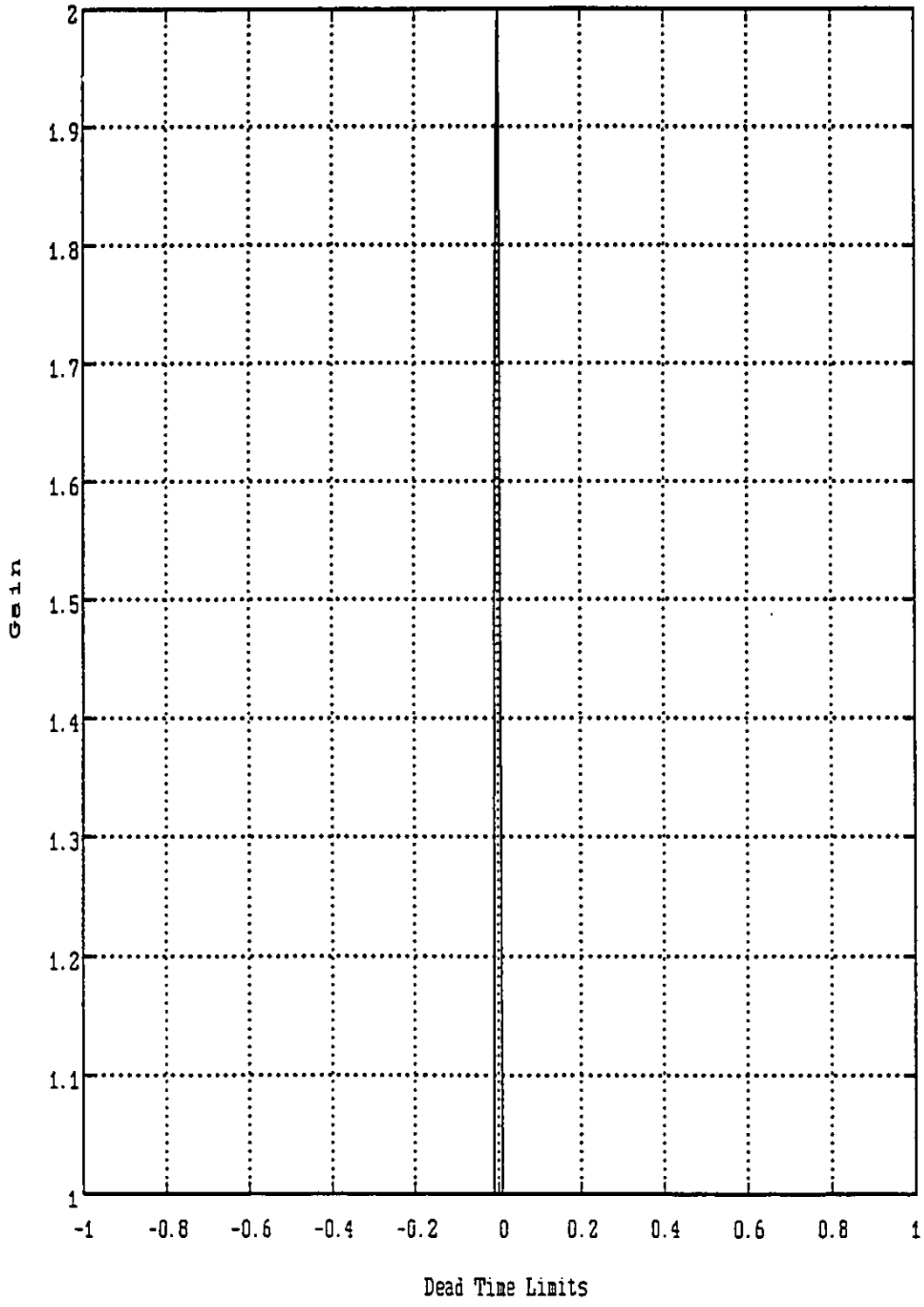


Figure 2.13 Robustness Region Plot LQOC Control λ=0

More simply put: perfect control results in zero robustness despite a significant phase margin. The reason for this shape of open loop frequency response is examined in more detail in section 2.3.

By detuning the LQOC controller the tradeoff between robustness and performance can be examined. Figure 2.14 shows the open loop Nyquist plot of the frequency response and Figure 2.15 shows the robustness region for values of λ=0.0001, 0.01 and 0.04. The ISE values corresponding to these controllers are 1.11, 1.34 and 1.48 respectively. When λ=0.0001 the open loop frequency response has two lobes that extend outside of the unit circle at high frequencies, causing a very narrow robustness region. So, for the tuning which delivers the same ISE performance as the PID controller the robustness region is significantly smaller.

When λ=0.01 the frequency response has a lobe which approaches but does not cross the unit circle. The resulting robustness region has a sudden discontinuity when the gain variation is δK_p=1.07; when the process gain is increased by this factor the lobe in the frequency response at higher frequencies intersects the unit circle creating higher frequency crossovers which lead to a smaller positive allowable dead-time variation and the creation of a negative value. The change in slope of the positive dead-time limit when δK_p=1.6 is due to a change from the high frequency large phase margin crossover controlling the dead-time variation to the low frequency low phase margin crossover controlling it. Note that at the nominal gain the allowable dead-time parameter variation is larger than that of the PID controller with a positive value of 1.5 minutes and no negative value. The allowable gain

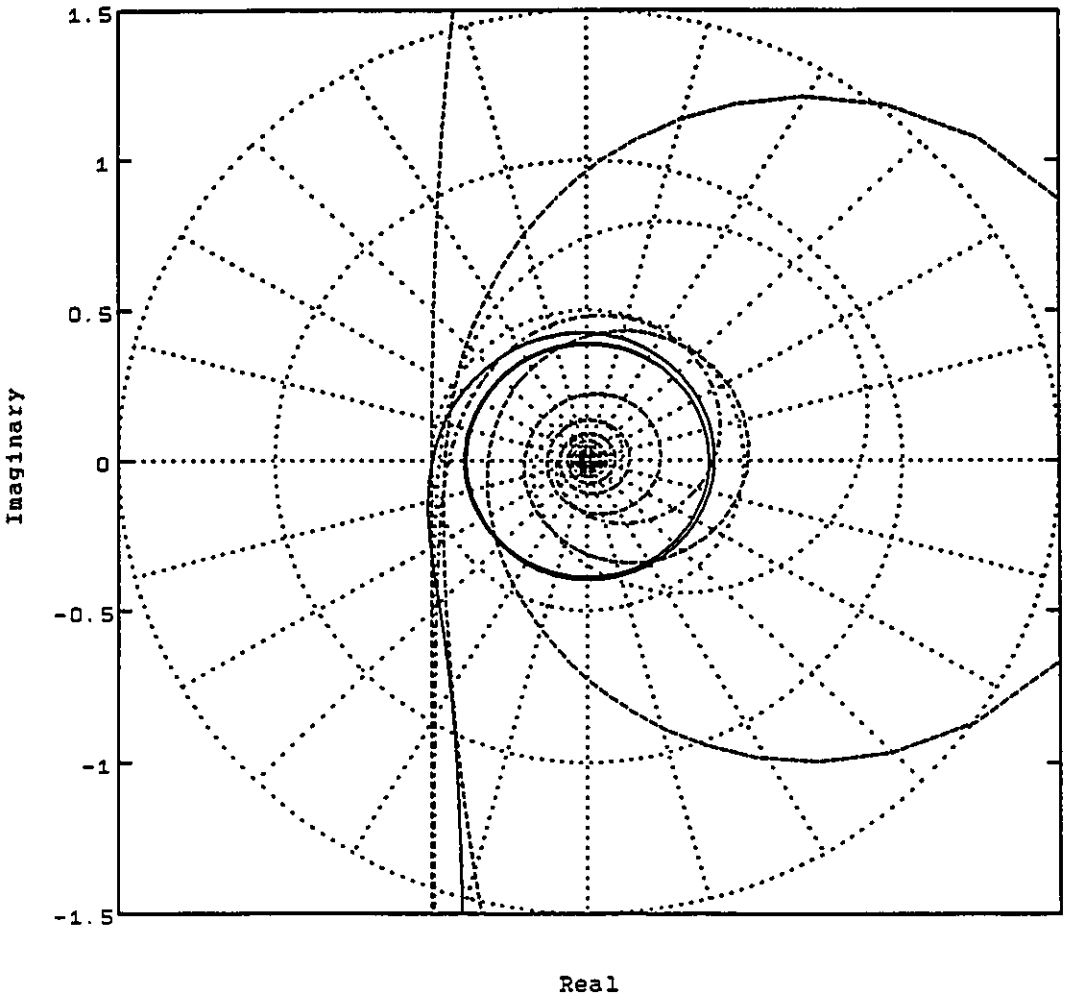


Figure 2.14 Nyquist Plot LQOC and PID Control
- PID, -- LQOC λ=0.0001, .. λ=0.01, -. λ=0.04

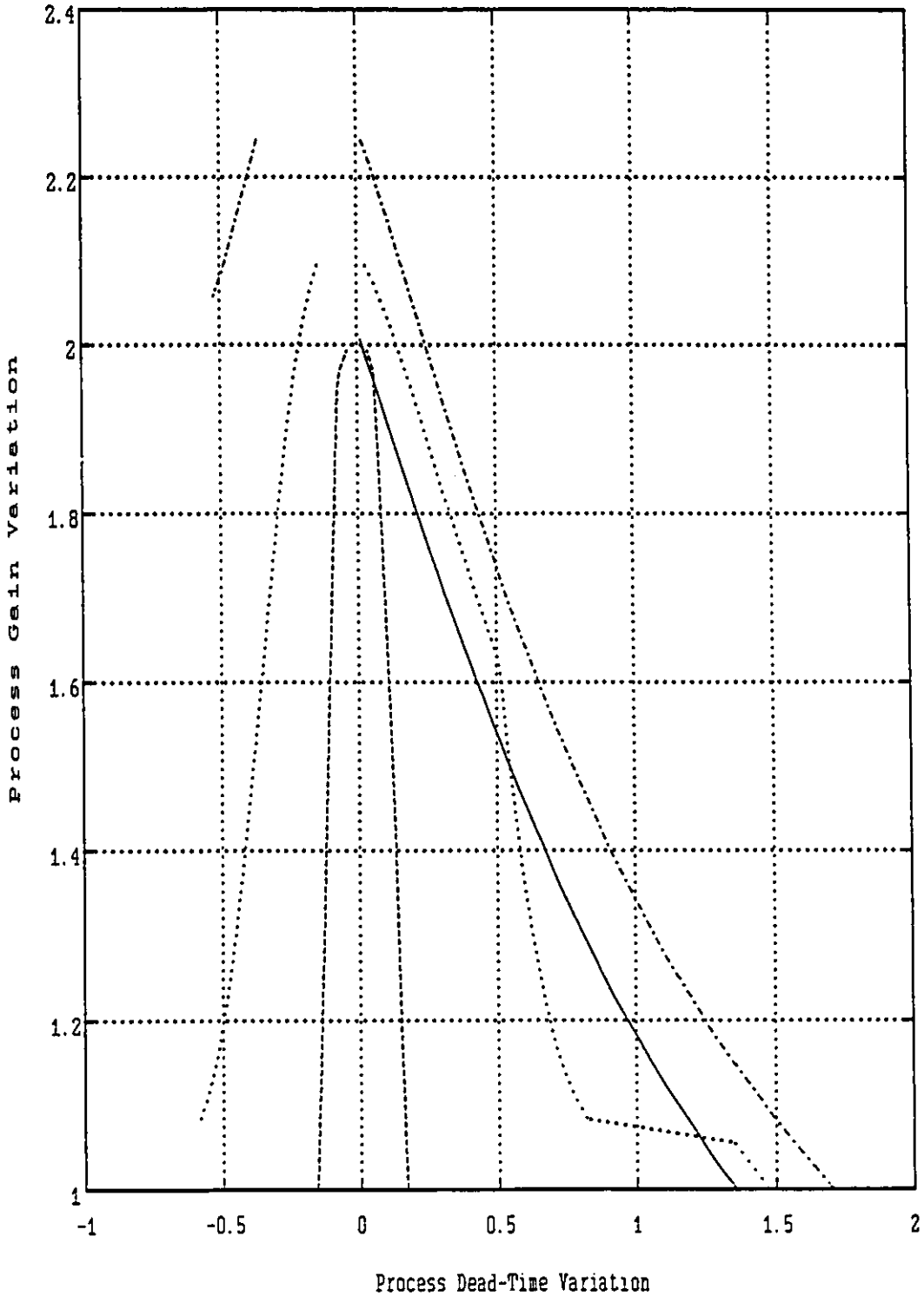


Figure 2.15 Robustness Region Plot LQOC and PID Control
- PID, -- LQOC λ=0.0001, .. λ=0.01, -. λ=0.04

variation at the nominal dead-time is also larger at 2.1. However, the robustness region is significantly smaller than that of the PID controller, in particular there are negative values of the dead-time variation which can cause instability for all values of the gain variation greater than 1.07. This illustrates the importance of defining robustness as the whole region of joint allowable gain and dead-time parameter variation rather than specifying only the individual values. When λ=0.04 the robustness region is as large as for the PID controller with no negative values of the allowable dead-time variation for values of the gain variation less than 2.

Figure 2.15 also illustrates that informed decisions are required when comparing the robustness defined as a region of joint allowable variation. Different controllers can result in very different shapes of robustness regions, so the designer must make a decision as to when two controllers have comparable robustness.

The IMC controller with ϵ=0 is identical to the LQOC controller with λ=0 and so it gives zero ISE and zero robustness. The tradeoff between robustness and performance for the IMC controller is investigated by choosing values of the tuning parameter ϵ=0.3, 0.4 and 0.6. Figure 2.16 shows the Nyquist plot of the open loop frequency response and Figure 2.17 shows the plot of the robustness region for these controller tunings. When ϵ=0.3 the ISE is 1.15 and the frequency response has one lobe which extends outside the unit circle and the robustness region has small values of both positive and negative allowable dead-time variation. When ϵ=0.4 the ISE is 1.20 and the frequency response has a lobe which approaches but does not cross the

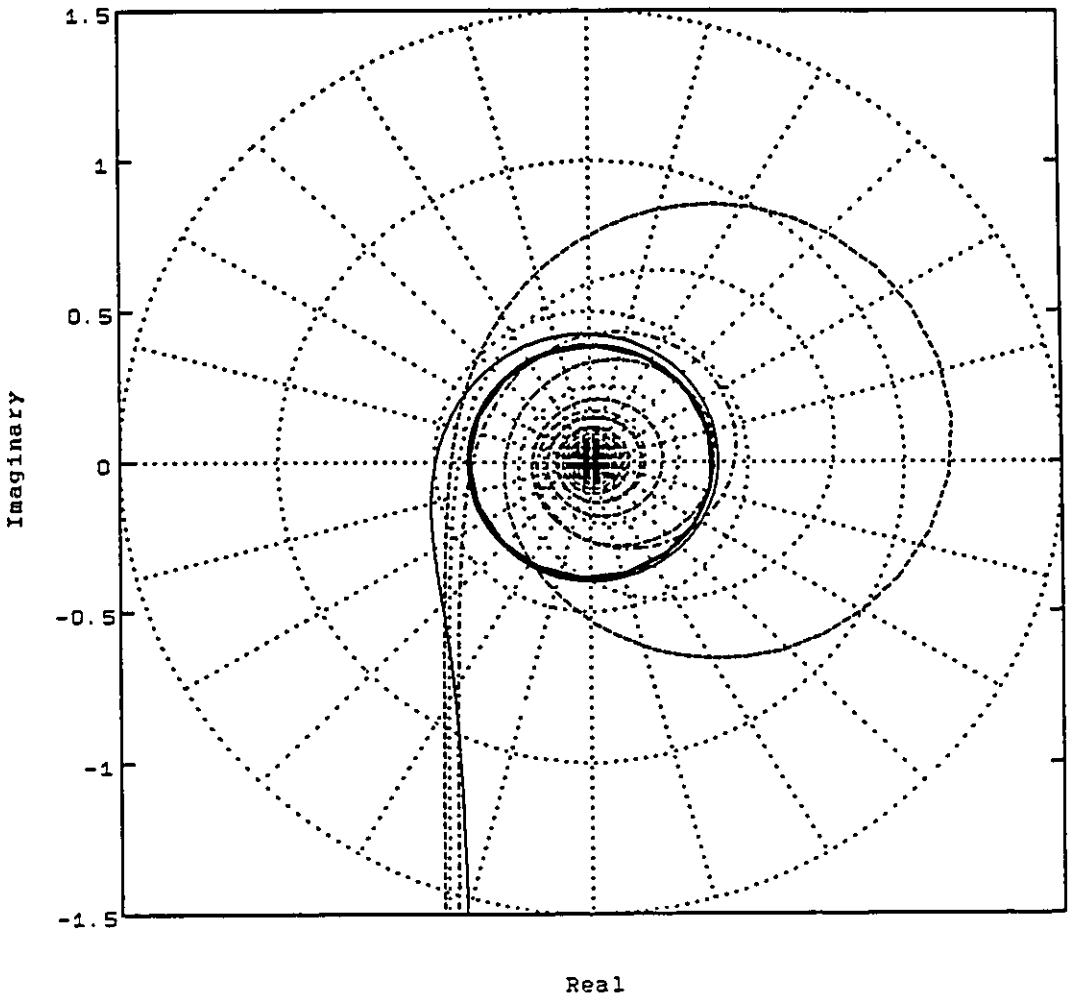


Figure 2.16 Nyquist Plot IMC and PID Control
- PID, -- IMC ϵ=0.3, .. ϵ=0.4, -. ϵ=0.6

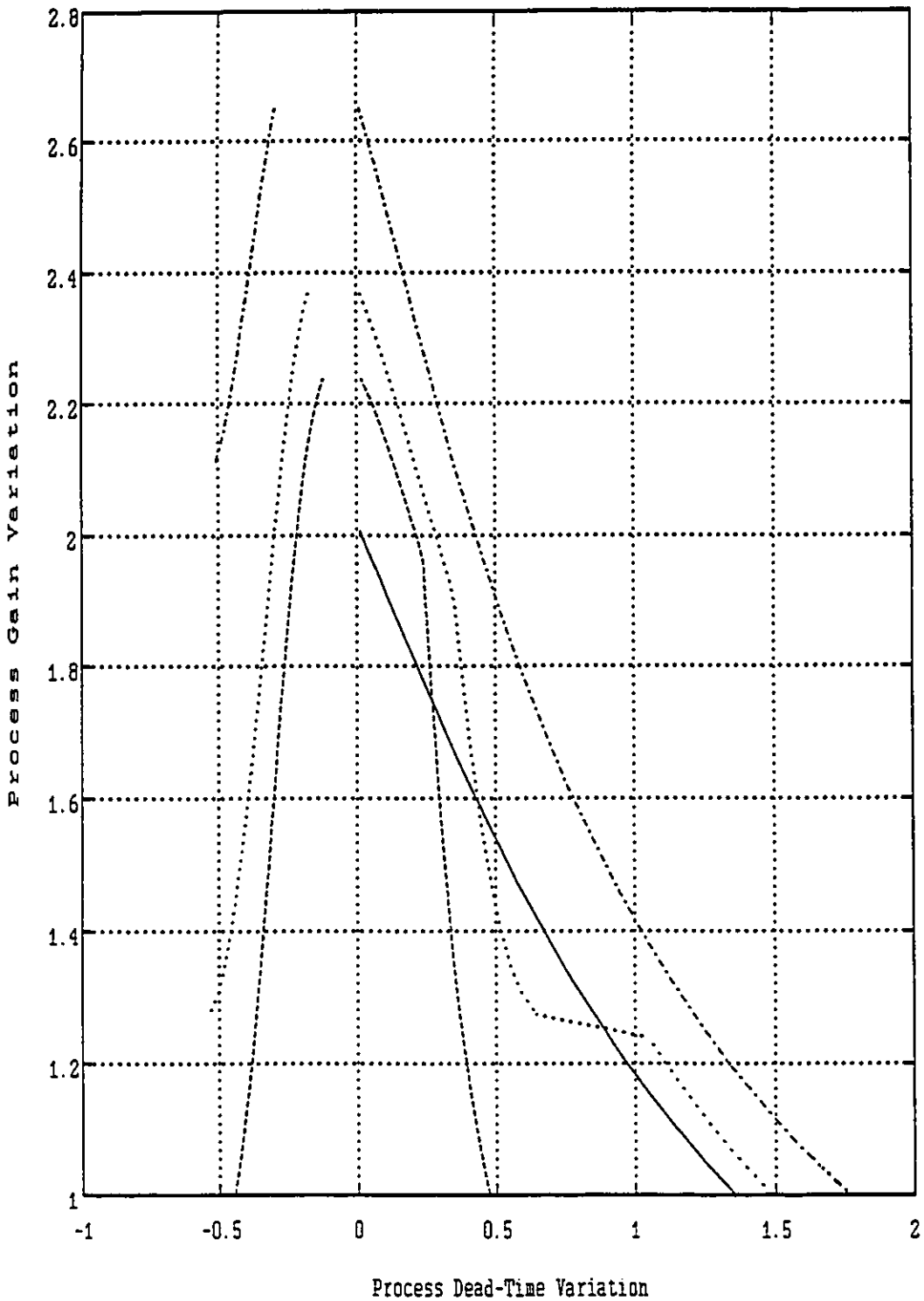


Fig 2.17 Robustness Plot IMC and PID Control
- PID, -- IMC ϵ=0.5, .. ϵ=0.4, -. ϵ=0.6

unit circle. This leads to a large positive allowable dead-time variation of 1.5 minutes and no negative value; however, as the gain is increased by a factor of 1.2 the lobe cuts the unit circle and the allowable dead-time variation contracts in both directions. When ϵ=0.6 the ISE is 1.3 and there are no negative values of the allowable dead-time variation for gains multiplied by less than 2 (the gain margin of the PID controller), the allowable gain variation is 2.5.

For this example, the PID design of Rivera et al (1984) when tuned for its minimum value of ISE, which is 1.11, gives a certain level of robustness. The IMC design of Rivera et al (1984) has an ISE of 1.3 when tuned for a similar level of robustness. The LQOC design of Palmor (1982) has an ISE of 1.48 for a similar level of robustness. These comparisons are interesting in that they show that the different controller designs have significantly different levels of ISE performance for the same level of robustness. It is interesting to note that although the PID design of Rivera et al (1984) is presented as an approximation to the full IMC design, it provides a lower ISE than the full IMC design for the same level of robustness.

The minimum ISE available from the PID design of Rivera et al (1984) is $1.1T_d\delta r^2$. The minimum ISE thus depends on the size of the dead-time. When the dead-time is large, the robustness limits are correspondingly larger. In such situations the robustness provided by the PID controller might be much more than is required so that a better performing IMC or LQOC design might be available with reasonable robustness.

The high level of performance for the given level of robustness of the PID controller is due to the impulse inputs which the derivative action in the controller generates in response to a step change in the set point. This performance is not attainable for most chemical processes as impulse inputs are unavailable, as discussed by Harris and Tyreus (1987). The illusory nature of the ISE performance improvement is illustrated by the simulated step response of Figure 2.18 where both the process input and output are shown. For step changes in set point the ISE penalizes initial large errors most heavily. The PID simulation shows an instantaneous process response which reduces this error immediately. However the dirac delta pulse input which is required is clearly unavailable in most real chemical processes. In situations where the control input will be limited, the low theoretical value of the ISE cannot be obtained.

This section has shown how the robustness region is useful for comparing the robustness of very different controllers. By combining robustness analysis with performance analysis even more interesting results are obtained. The explanation of the difference in performance for the same level of robustness between the IMC controller and the LQOC controller will be explored in the next section. In the subsequent section, the good robustness properties of the PID controller vis a vis the dead-time compensating controllers for this example will be investigated.

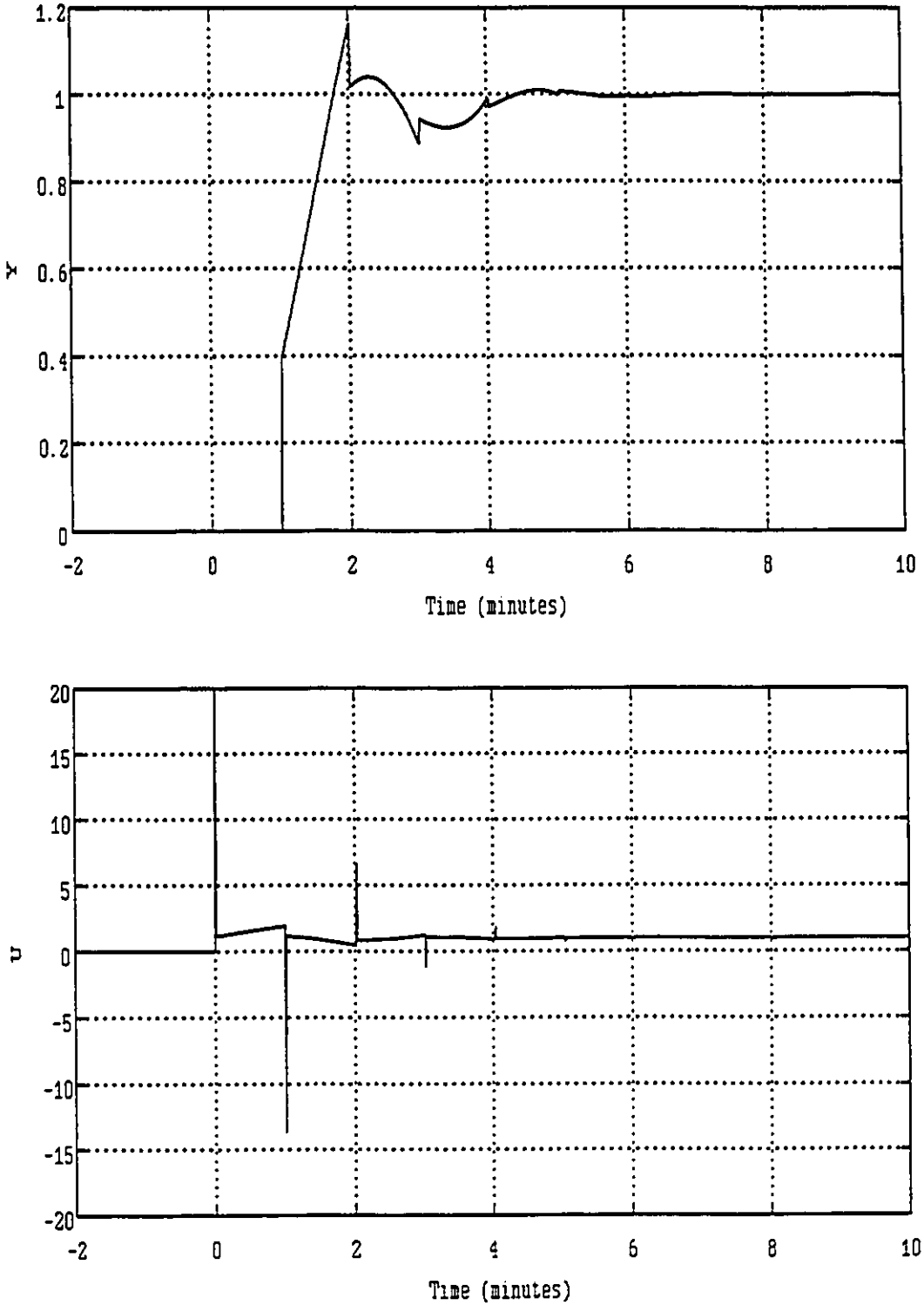


Figure 2.18 PID Response to Step Set Point Change

2.3 IMC vs LQOC Detuning

The results of the last section raise the question of why is it that the IMC controller design gives a lower ISE performance than the LQOC controller for the same level of robustness? Both of these controllers have dead-time compensation which was structured in the form of a Smith Predictor in the previous sections. The Bode plot of the overall IMC controller gain is shown in Figure 2.19 and the Bode plot of the overall LQOC controller gain is shown in Figure 2.20 for the tuning parameter values of section 2.2.

The difference between the two controllers lies in the inner controller block of the Smith Predictor structure described in equation (2.19) for the LQOC design and equation (2.21) for the IMC design. Both of these are basically PI controllers where the level of the gain is determined by the tuning parameter. However, the controller of equation (2.19) is augmented by a first order filter. For the LQOC controller the infinite frequency gain is

$$\lim_{s\to\infty} G_{LQOCO}(s) = \lim_{s\to\infty} \frac{K_C \dfrac{T_P s + 1}{s}}{(C_2/C_1)s + 1} = 0 \qquad (2.24)$$

whereas for the IMC controller the infinite frequency gain is

$$\lim_{s\to\infty} G_{IMCO}(s) = \lim_{s\to\infty} K_C \frac{T_P s + 1}{s} = K_C T_P \qquad (2.25)$$

A gain at infinite frequency allows the IMC controller to have proportional action immediately in response to a step change in set point: that is the controller response to a step discontinuity in the

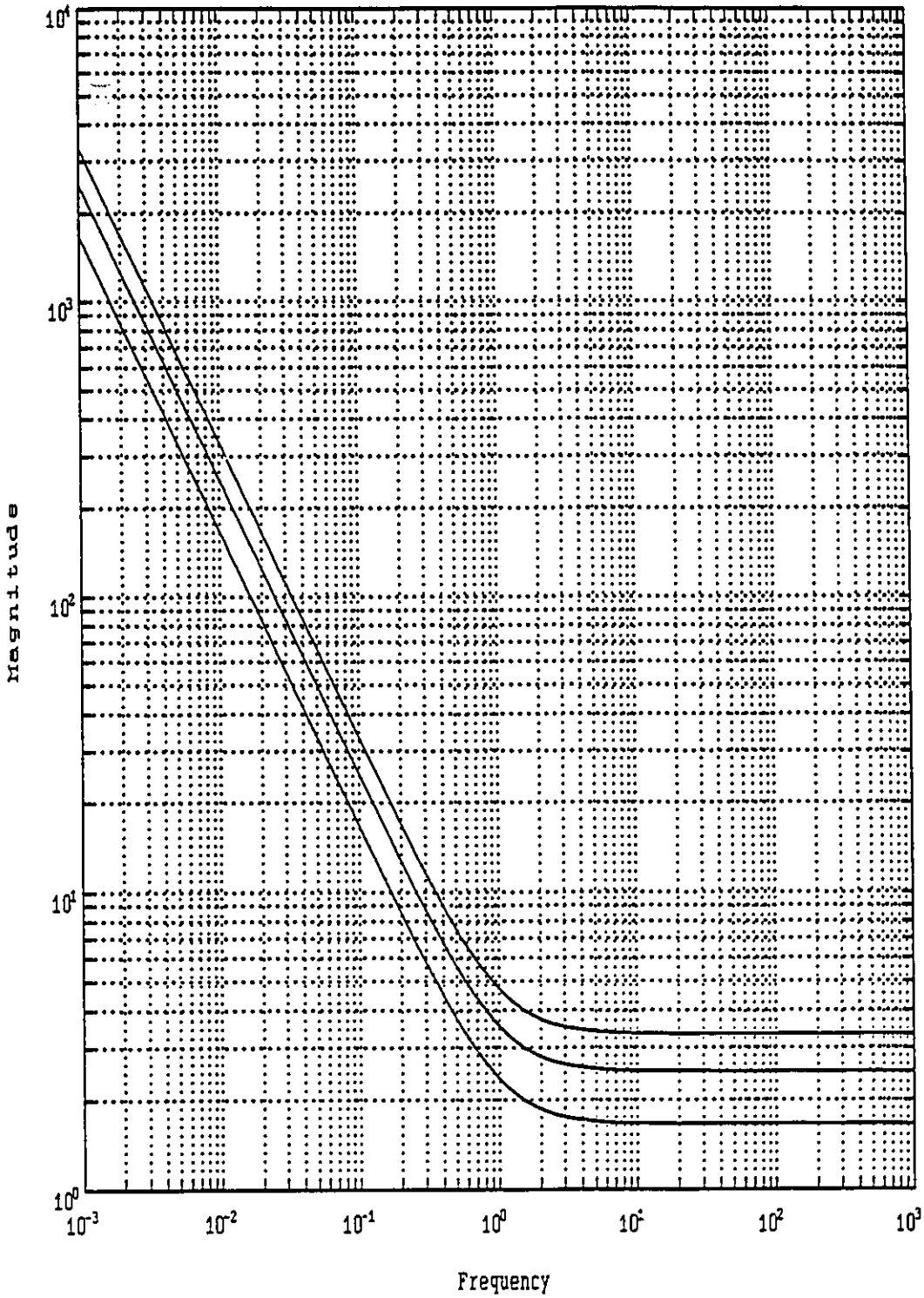


Figure 2.19 Bode Plot of G_{co} for IMC Controller in Smith Predictor Form
ϵ=0.3, ϵ=0.4, ϵ=0.5

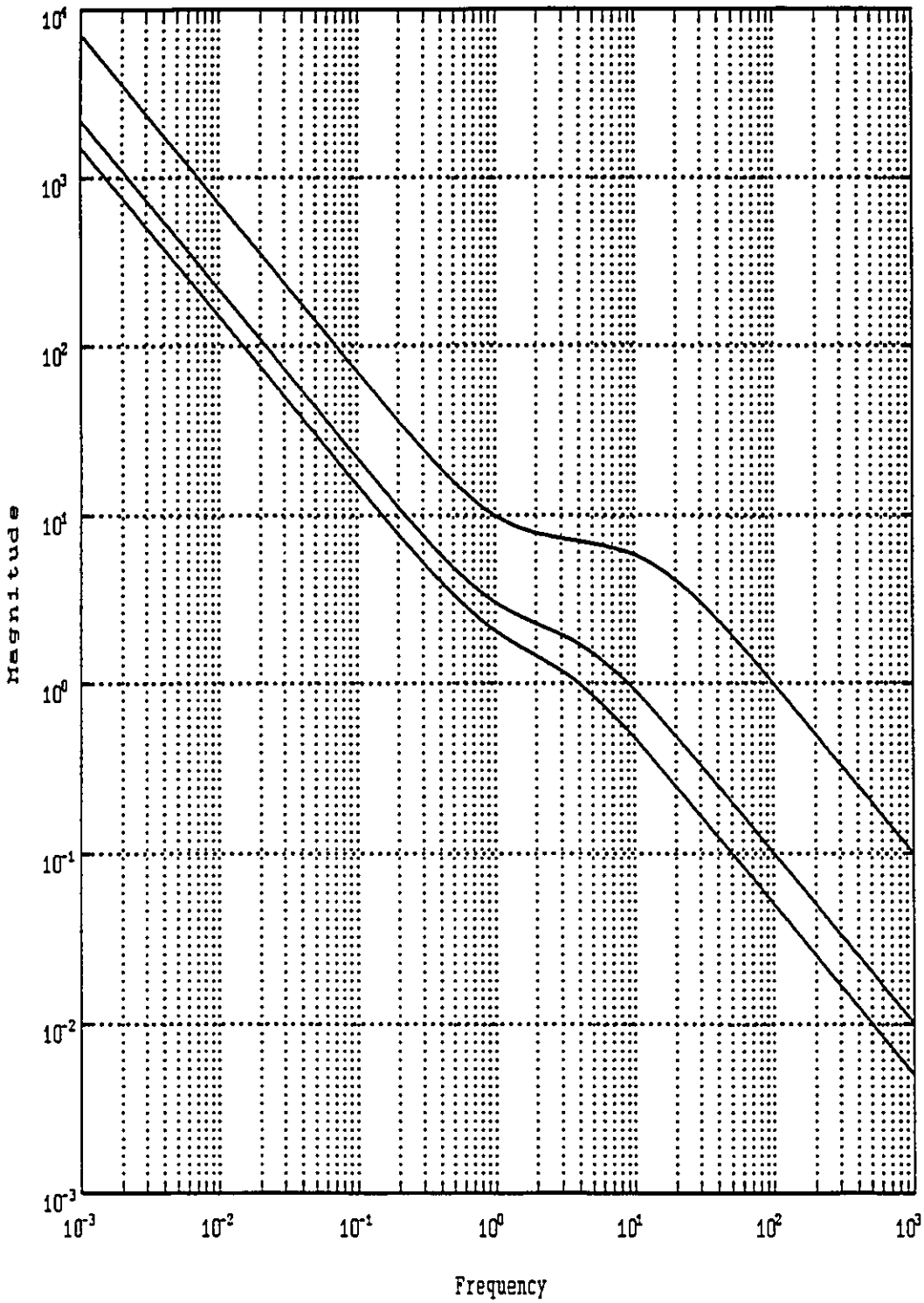


Figure 2.20 Bode Plot of G_{co} of LQOC Controller in Smith Predictor Form
λ=0.0001, λ=0.01, λ=0.04

error produces a step discontinuity in the manipulated control action. Zero gain at infinite frequency means that the LQOC controller response to a step change can only produce a smoothly changing manipulated control action: there is no proportional action. That there is no proportional action in the inner controller LQOC0 at infinite frequency is a direct result of the criterion, equation (2.17), which the LQOC controller design optimizes. A discontinuity in the manipulated control action u(t) would lead to an infinite value of the derivative of the control action and hence an infinite value in the criterion. So, the difference in nature between the controllers obtained through IMC detuning and LQOC detuning is due to the inherent differences in the rationale behind the two methods.

It is interesting to note that the IMC controller maintains the same form as the unconstrained optimal controller but with a reduced gain in the inner controller: the optimal controller has infinite gain whereas the IMC controller has gain $1/K_p\epsilon$. This is only true for the use of a first order filter where for any process $G_p(s)$ the controller block in the Smith Predictor model is

$$G_{IMC0}(s) = \frac{G_p^{-1}(s)}{\epsilon\, s} \tag{2.26}$$

For any process which is improper, where the denominator degree is more than one greater than the numerator degree, this design results in a controller block which will generate impulse inputs to the process in response to a step change in set point.

Rivera et al (1984) recommend using a higher order filter for higher order processes in order to obtain a controller block which is

proper. If the IMC tuning filter, F(s), order is chosen to be of the same order as the process then the IMC controller will not generate impulse inputs to the process and will have a finite gain at infinite frequency. The LQOC controller, on the other hand, will never have a finite gain at infinite frequency because of the nature of the optimization criterion, as described above. So, the result of this section that IMC control has lower ISE for step changes in set point than LQOC control when tuned for the same level of robustness should apply to the control of higher order systems.

2.4 Dead-time Compensation as a Robustness Problem

In section 2.2, the PID controller is shown to have attractive robustness properties relative to the IMC and LQOC controllers which both have dead-time compensation. This section investigates the effect of dead-time compensation on controller robustness and the manner in which the PID design of Rivera et al obtains good performance without dead-time compensation.

There are several good literature studies of the properties of Smith Predictor controllers. Åström (1977) investigates the frequency domain properties of the Smith Predictor structure with particular emphasis on the phase contributions. Ioannides et al (1979) map out a robustness region in the parameter space of an example. Both of these studies mention that the Bode gain plot of a controller with a Smith Predictor has sharp spikes which adversely affect the robustness of the closed loop system. Palmor (1980) examines the stability of systems

with Smith Predictors while paying special attention to infinitely sensitive systems. The contributions of section 2.2 added to this understanding of the stability properties of dead-time compensating controllers. This section investigates the source of the stability problems due to the dead-time compensation and the manner in which they are circumvented by the use of derivative action in the PID design of Rivera et al (1984).

2.4.1 Use of the first order Padé approximation to dead-time as an alternative to dead-time compensation: Frequency domain interpretation

It is useful to consider the LQOC optimal controller for the FOPDT example of section 2.2 when the tuning parameter λ=0; that is, a PI controller as the controller block in a Smith Predictor structure:

$$C_{OPT}(s) = \frac{K_C\,(T_P s + 1)}{s + K_C\,K_P\,(1 - e^{-T_d s})} \tag{2.27}$$

where the controller gain is $K_C=\infty$. The PID design of Rivera et al (1984) is derived by replacing the dead-time in the process transfer function by a 1st order Padé approximation. So,

$$G_P(s) = K_P\,G_{P0}(s)\,e^{-T_d s} \approx K_P\,G_{P0}(s)\,\frac{-(T_d/2)s + 1}{(T_d/2)s + 1} \tag{2.28}$$

Using the design procedure of Palmor (1982) to find the minimum ISE λ=0 controller for this approximate system, the overall controller is:

$$C_{PADE}(s) = \frac{(T_p s + 1)\ ((T_d/2)s + 1)}{T_d s} \tag{2.29}$$

which is a PID controller with a finite gain. Comparing $C_{OPT}(s)$ of equation (2.27) with $C_{PADE}(s)$ of equation (2.29) we see that they have the factor $(T_p s + 1)$ in common. Factoring this from the equations, we are left with the portion of the controller which includes the gain, integral action and the dead-time compensation. For the optimal controller this leaves:

$$DTC_{OPT}(s) = \frac{K_C}{s + K_C K_p (1 - e^{-T_d s})} \tag{2.30}$$

With K_C=∞ the Bode plot of the gain of equation (2.30) will have a peak of height K_C=∞ whenever w=$\pm 2n\pi/T_d$ for all n, and a valley of height 1/2 whenever w=$\pm(2n+1)\pi/T_d$. This shows that the peaks and valleys in the frequency response of the dead-time compensating controller are due to the Smith Predictor dead-time compensation. The dead-time compensation reduces the controller gain to 1/2 whenever the contribution of the dead-time to the phase of the process is π and allows the original process gain whenever the contribution of the dead-time to the phase of the process is 0. This explanation is only exact when K_C=∞; however, it provides a heuristic understanding of the peaks and valleys of the gain Bode plot of a dead-time compensating system. It is the spikes in the gain of the controller frequency response which, when cascaded with the process frequency response, lead to the lobes seen in the open loop

frequency responses of section 2.2. It is these lobes which cause the narrowing of the robustness region.

In comparison, factoring $(T_p s + 1)$ from the Padé approximation controller leaves:

$$DTC_{PADE}(s) = \frac{((T_d/2)s + 1)}{T_d s} \qquad (2.31)$$

This term provides the added derivative action in the controller of equation (2.29). The limit of equation (2.31) as the frequency goes to infinity is 1/2. The Bode gain plot of the gain of $DTC_{OPT}(s)$ together with the gain of $DTC_{PADE}(s)$ is shown in Figure 2.21. The Padé approximation derivative action follows the valleys of the optimal dead-time compensation except for the bottom of the first valley where the derivative action is slightly higher. By avoiding the peaks of the optimal dead-time compensation, the derivative action approximation is able to avoid the robustness problems which they cause. Rivera et. al. (1984) detune the PID controller of equation (2.29) by reducing its gain to find the minimum ISE. Because the Padé approximation gain curve lies above the optimal gain curve in the first valley, the performance increases as the gain of the controller is reduced. The favourable characteristics of this controller result from this combination of higher gain in the low frequency region combined with no peaks in the high frequency region.

This section has shown that although the robustness region of joint allowable variation in the process gain and dead-time parameters is defined in the parameter space of the process model, the way that the region is constructed from the frequency response allows inferences to

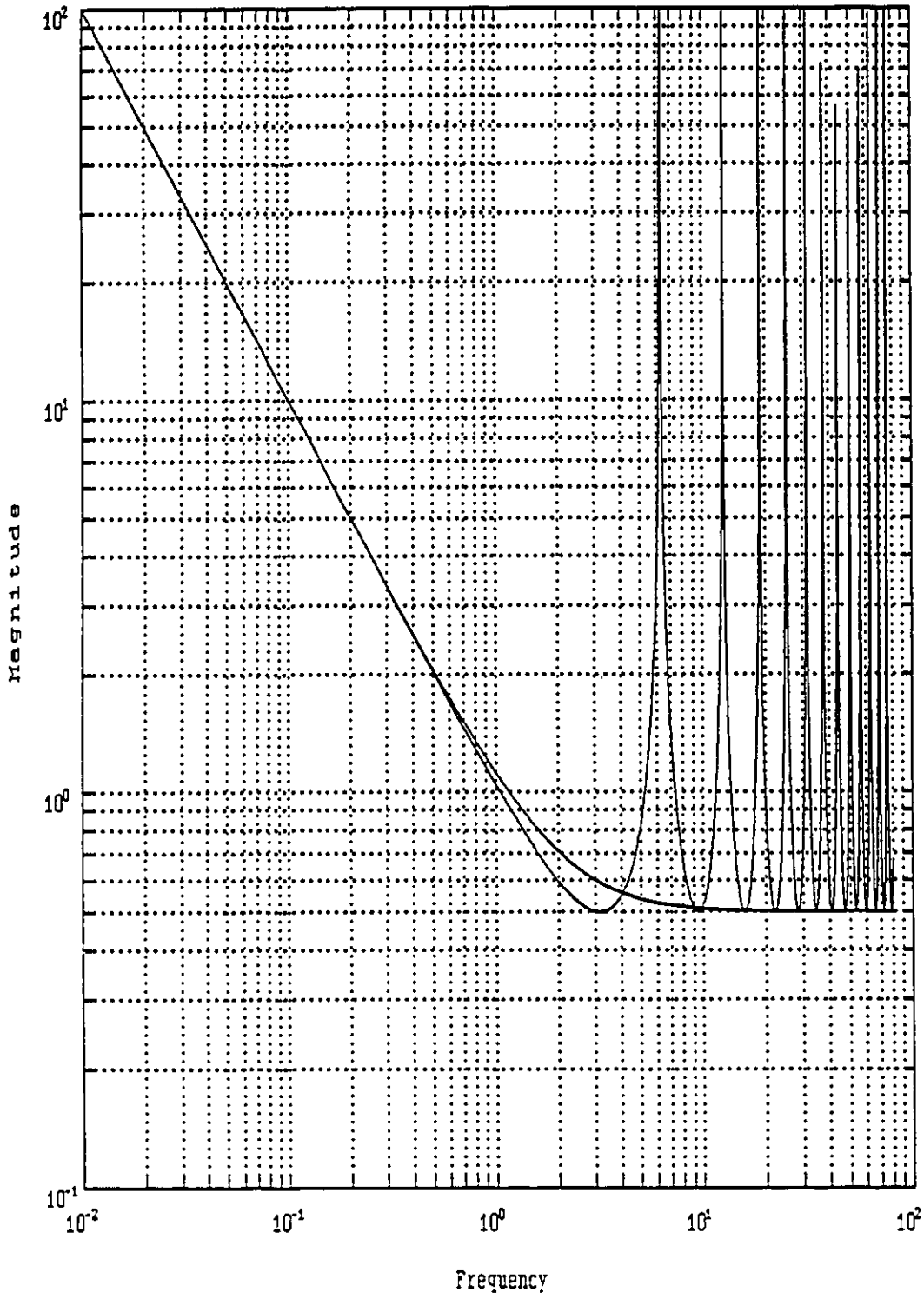


Figure 2.21 Dead-time Compensator Bode Plot for PID and LQOC Control

be made about the effects of features in the frequency response of a controller on system robustness. The use of a first order Padé approximation to replace dead-time in the process model in order to design a controller which has good robustness and performance properties should be general. However, as Rivera et al (1984) have shown, the resulting controller requires tuning. This is because the first order Padé approximation is not an exact representation of the dead-time. Also, for higher order processes the resulting controller will not be a simple PID controller: it will have higher order derivative action.

2.4.2 Use of higher order Padé approximations to dead-time as an alternative to dead-time compensation

If the first order Padé approximation to the dead-time has good properties for controller design, then, are higher order approximations better? The second order Padé approximation is:

$$G_p(s) = K_p\, G_{p0}(s)\, e^{-T_d s} \approx K_p\, G_{p0}(s)\, \frac{-(T_d^2/12)s^2 + (T_d/2)s + 1}{(T_d^2/12)s^2 + (T_d/2)s + 1} \qquad (2.32)$$

The third order Padé approximation is:

$$G_p(s) = K_p\, G_{p0}(s)\, e^{-T_d s} \approx K_p\, G_{p0}(s)\, \frac{-(T_d^3/120)s^3 + (T_d^2/10)s^2 - (T_d/2)s + 1}{(T_d^3/120)s^3 + (T_d^2/10)s^2 + (T_d/2)s + 1} \qquad (2.33)$$

The fourth order Padé approximation is:

$$G_p(s) = K_p\, G_{p0}(s)\, e^{-T_d s} \approx K_p\, G_{p0}(s)\, \frac{(T_d^4/1680)s^4 - (T_d^3/84)s^3 + (3T_d^2/28)s^2 - (T_d/2)s + 1}{(T_d^4/1680)s^4 + (T_d^3/84)s^3 + (3T_d^2/28)s^2 + (T_d/2)s + 1} \qquad (2.34)$$

Extracting the integral and derivative action from the optimal controller designs using these approximations gives

$$DTC_{PADE2}(s) = \frac{(T_d^2/12)s^2 + (T_d/2)s + 1}{T_d s} \tag{2.35}$$

$$DTC_{PADE3}(s) = \frac{(T_d^3/120)s^3 + (T_d^2/10)s^2 + (T_d/2)s + 1}{T_d s\,((T_d^2/60)s^2 + 1)} \tag{2.36}$$

$$DTC_{PADE4}(s) = \frac{(T_d^4/1680)s^4 + (T_d^3/84)s^3 + (3T_d^2/28)s^2 + (T_d/2)s + 1}{T_d s\,((T_d^2/42)s^2 + 1)} \tag{2.37}$$

The even numbered Padé approximations give dead-time compensation approximations which have infinite gain at infinite frequency, compared to the gain of 1/2 at infinite frequency of the odd numbered Padé approximations. The gain Bode plots of $DTC_{PADE1}(s)$, $DTC_{PADE2}(s)$, $DTC_{PADE3}(s)$ and $DTC_{OPT}(s)$ are plotted together in Figure 2.22. The Bode plots for dead-time compensation based on higher order Padé approximations attempt to mimic the gain peaks of the optimal dead-time compensation, but the peaks are clearly misaligned. This result reveals that it is only the first order Padé approximation to the dead-time which has the desirable property of resulting in a controller which approximates the optimal dead-time compensation without having any high frequency peaks.

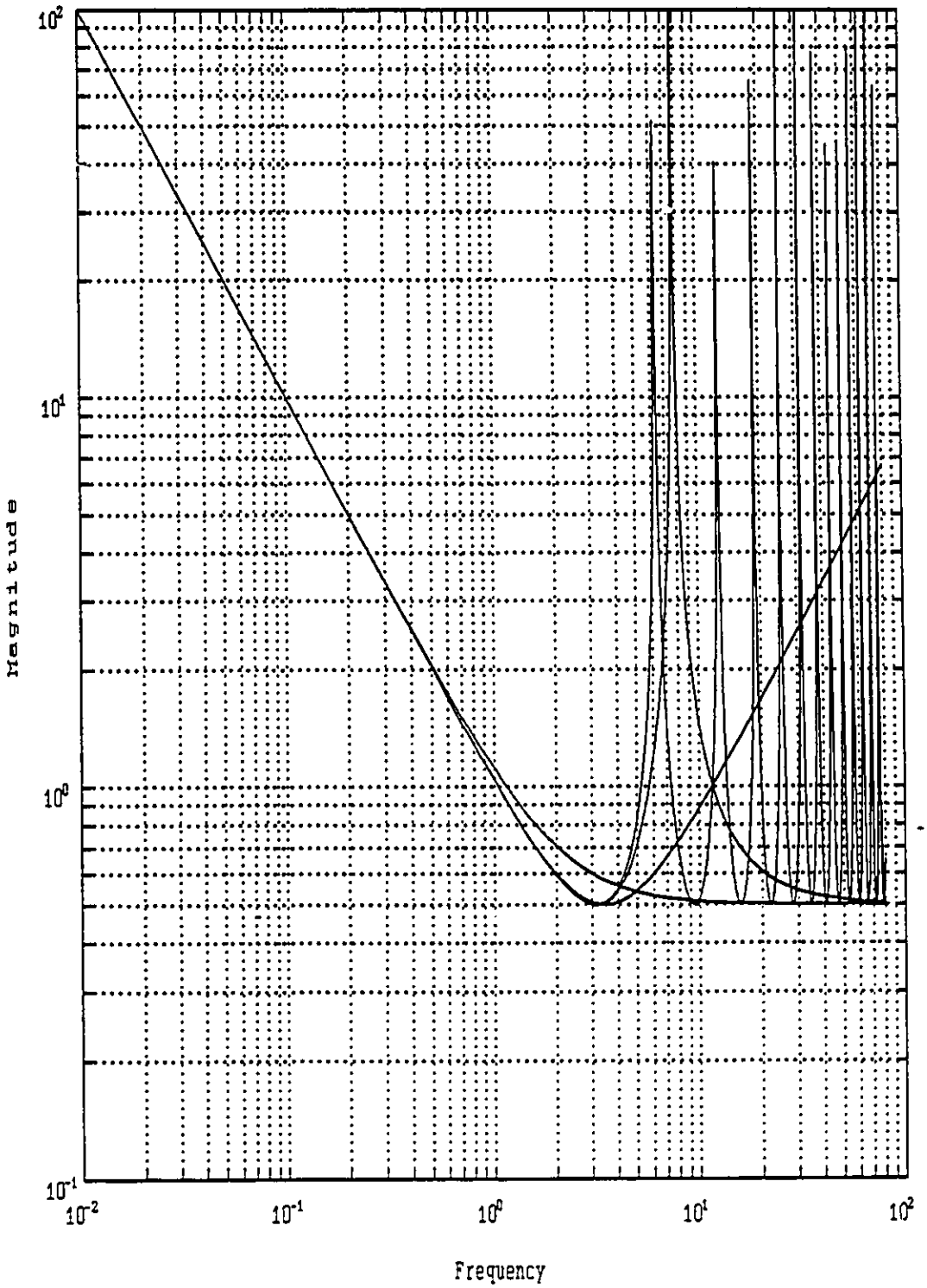


Figure 2.22 Dead-Time Compensator Bode Plot, Padé Orders 1, 2, 3 & LQOC
1^{st} order, 2^{nd} order, 3_{rd} order, LQOC

2.5 Summary and Conclusions

This chapter has presented a new definition of robustness as the region of joint allowable variation in the process gain and dead-time parameters, together with a straightforward method for calculating the robustness region from the Nyquist plot. This chapter extends the method of Tsypkin (1946) to systems which can be made unstable by a reduction in the process dead-time. The main advantage of this definition of robustness over traditional frequency domain definitions, such as the gain and phase margin, lies in the frequency information which is incorporated into the allowable dead-time parameter variation. In addition, the allowable dead-time parameter variation alone is shown to be not as useful for assessing robustness as the region of joint allowable variation in the process gain and dead-time parameters.

This definition of robustness does not require the a priori specification of an unstructured frequency domain uncertainty as do the methods of Morari and Zafiriou (1989). It allows the control system designer to measure the robustness of any particular controller design and to compare the robustness of very different controller designs. Although robustness is defined in the appealingly intuitive parameter space of the process gain and process dead-time parameters, it is calculated from the frequency response of the open loop system. This calculation clarifies which features of the frequency response are responsible for any robustness problems.

The method was used to explore the robustness vs ISE performance tradeoff in the design of controllers for a step change in set point for

a first order plus dead-time process. In this application, the definition of robustness was shown to be useful in comparing the robustness of very different controllers. The controllers which were compared were the PID and IMC designs of Rivera et al (1984) and the LQOC design of Palmor (1982). The favourable robustness and ISE performance of the PID design was explained by comparing the frequency response of the design based on a first order Padé approximation with an optimal controller using dead-time compensation.

The favourable ISE performance and robustness of the IMC design relative to the LQOC design is explained by comparing the controllers in their Smith Predictor form. The inner controller of the LQOC design does not have proportional action at infinite frequency as does the IMC controller; so, the LQOC controller, unlike the IMC controller, cannot produce a step change in the control action in response to a step change in set point.

The shape of the robustness region is related directly to the Nyquist plot from which the region is calculated, and this relationship was exploited in explaining the robustness of the examples in the chapter. This relationship was also used to explain the robustness implications of dead-time compensation. The spikes in the Bode gain plot of controllers with dead-time compensation are shown to be the cause of the contracted robustness region.

CHAPTER 3

DESIGN AND ANALYSIS OF DISCRETE CONTROLLERS FOR THE CONTROL OF CONTINUOUS SINGLE-INPUT-SINGLE-OUTPUT PROCESSES

Discrete controllers, implemented through the use of digital computers, are commonly used for feedback control of continuous-time systems. Performance and robustness to process changes are both important considerations in the design of these discrete feedback controllers. Although discrete process models which are coincident with the continuous process are usually used when designing discrete controllers, it is the performance of and the robustness to changes in the continuous process which are important. With this in mind, this chapter presents a wide ranging analysis of the robustness and continuous time performance of continuous processes with discrete feedback controllers.

The designer of a discrete controller for a continuous process must choose the size of the control interval. Many of the different controller designs which are available do not perform well for some choices of the control interval. This is demonstrated here by comparing the performance of Linear Quadratic Optimal Control (Åström (1970), Box and Jenkins (1970)), the modified Dahlin controller (Dahlin (1968)) and

the State Deadbeat Internal Model Control design (Zafiriou and Morari (1985)) with no detuning as the size of the control interval is varied. The Integral Square Error and Integral Absolute Error of the continuous process output are used to show under what conditions each of the different controllers perform poorly. The results are explained by using continuous time plots of the process output and manipulated control action response to step changes in the set point. This study confirms many of the desirable properties of the State Deadbeat IMC controller observed by Zafiriou and Morari (1985) and Lennartson and Söderström (1989) and extends their studies by showing that LQOC controller outperforms the State Deadbeat IMC controller when the process model numerator has a zero which is not near minus one. The analysis examines the approximate model inverse in each of the different controller designs.

The same discrete controller designs are then compared for the manner in which changes in the tuning parameter effect changes in the tuning filter. Two new controller tuning methods are proposed: a Modified LQOC controller design and an extended horizon controller design. The Modified LQOC controller calculates a controller for a nonstationary disturbance using the spectral factorization approach as if the disturbance were stationary. This controller has a tuning filter which is of order one less than the LQOC tuning filter. Through the appropriate choice of the tuning filter gain, the resulting controller still has integral action. The Modified LQOC controller provides a bridge between the LQOC controller and the State Deadbeat IMC designs.

A new extended horizon controller design is proposed which behaves as if it were a state deadbeat controller for the same process but with a longer control interval. This latter controller is an improvement over the extended horizon controller of Ydstie et al (1985) in that if a single set point change is made and no disturbance enters the system then the initial control action sequence will be implemented with no changes.

As a prelude to exploring robustness analysis for discrete control of continuous systems, the effect of a change in the dead-time of a continuous process on the frequency response of the discrete equivalent model is examined. The modified Z transform of a first order process with a fractional period of dead-time is shown to be a convex combination of the Z transforms of the same process with integer periods of dead-time which bracket the actual dead-time. Through the use of the example of a second order process, it is shown that this result is not general. In fact, for a second order process the frequency response of the process with a fractional period of dead-time can be outside of the boundary formed by the frequency response of the process with integer periods of dead-time which bracket the actual dead-time. This result appears to be new and has implications for studies which determine stability boundaries by a search on the dead-time parameter.

The definition of the robustness region for continuous processes, presented in the previous chapter, is adapted for discrete control of continuous processes as the region of joint allowable variation in the process gain and the quasi dead-time. The quasi dead-time is defined to be the phase margin of the discrete open loop

frequency response divided by the discrete frequency. When the allowable variation in the quasi dead-time is an integer number of control intervals and the process has no fractional period of dead-time then it provides an exact upper limit on the allowable dead-time variation in the underlying continuous process. When the allowable change in the quasi dead-time does not have a direct interpretation in terms of the dead-time of the underlying continuous process it is nonetheless useful for comparing the robustness of different controllers. The power and the limitations of this definition of robustness are shown by application to examples.

The method is used to show that a minimum variance controller for a process with no dead-time subject to random walk disturbances (equivalent to deadbeat control for step changes in set point) can only tolerate an increase in the process dead-time of at most one control interval. That is, the robustness of minimum variance control is proportional to the size of the control interval. This is believed to be a new result. The method is also used to show that dead-time compensation in Minimum Variance control reduces the robustness of the system relative to the undelayed system. The limitation of the method is shown to be that when calculating the robustness of discrete control applied to a process with a fractional period of delay, the allowable variation in the quasi dead-time can significantly over-estimate the allowable variation in the true dead-time. However, the relative results when comparing different controllers on the same process are meaningful. As in the continuous time definition of robustness, the

quasi dead-time incorporates frequency information in a natural and meaningful way.

The allowable change in the quasi dead-time measure of robustness and the continuous Integral Square Error measure of performance are combined to compare the robustness vs performance tradeoff of the tuning of LQOC, Modified LQOC, IMC, and State Deadbeat IMC controllers applied to first and second order processes with step changes in set point. The LQOC controller design is shown to have the poorest performance for the same level of robustness. The results are explained using continuous time plots of the process output and the control actions. These plots show additionally that a controller with good performance and good robustness can still require totally impractical control actions. The IMC and State Deadbeat IMC controllers can require a higher order tuning filter for higher order processes to obtain realizable control actions. The Modified LQOC controller has a filter order which is naturally matched to the order of the process. The poor performance of the LQOC controller is the result of a tuning filter which is of an order one higher than necessary. Bergh and MacGregor (1987) have studied the improved disturbance rejection ability of the LQOC controller over the IMC controller for autoregressive disturbances. Their study, however, focuses on discrete measures of performance for a process example with a fractional period of dead-time and does not make a comparison with a State Deadbeat IMC controller or with the Modified LQOC controller presented here, both of which provide improved robustness for this type of process.

This study contributes to the understanding of the relationships between approximate model inverses, tuning filters, performance, robustness and realizability of control actions.

3.1 Effect of the control interval on the ISE performance of discrete model based controllers applied to a continuous system

This section compares the performance of a number of discrete controllers applied to continuous first and second order plus dead-time systems as a function of the control time. The discrete controller designs are all model based and are designed using a discrete model which is coincident with the continuous process. The design methods are each based on different measures of the discrete time performance such as settling time, variance of the output or variance of the input. However, it is the performance of the continuous-time output which is of interest. So, the ISE and IAE of the continuous output are used as indicators of performance. Although poor performance might be anticipated through examining the discrete measures of input and output performance, the ISE and IAE provide a more direct and physically meaningful measure of the continuous system performance.

The discrete model based controllers are all designed using a discrete z transform model which is coincident with the continuous model at the control instants, where the discrete input is applied through a zero order hold (Ragazzini and Franklin (1958), Box and Jenkins (1970)). In particular, several of the examples of this section are based on a

first order plus dead-time process. The discrete z transform model of this process has a zero in the left half plane when the process dead-time is a non-integer multiple of the control interval. When this zero due to a fractional period of delay is outside the unit circle it is termed unstable or noninvertible. The discrete z transform equivalent to a second order continuous system usually has a stable left half plane zero in its numerator. When the system has a fractional period of delay the z transform has an additional zero which may be stable or unstable. Left half plane zeros also arise naturally in the discrete models equivalent to higher order continuous systems. A zero on the negative real axis can lead to a ringing pole in some controller designs. It has long been known that a ringing pole in a controller can lead to poor continuous output performance despite giving good output performance at the control instants: Ragazzini and Franklin (1958) note that continued excitation of the input to a process will lead to ripple in the output even though the error is zero at the sampling instants.

In discrete model based controller designs, a ringing pole in the controller can arise if there is a left half plane zero in the discrete process model. The inversion of the process model leads to a ringing pole in the controller. Each of the design methods considered here uses a different approximation to the inverse of the process model, and the analysis focuses on this aspect of the controller design procedure. The design methods are applied to step changes in set point, and detuning is not considered; that is, the analysis looks directly at the model inversion in the controller design.

The controller design methods which are considered are the Linear Quadratic Optimal Control (LQOC) controller (Åström (1970), Box and Jenkins (1970)), the Dahlin controller and the modified Dahlin controller (Dahlin (1968)), the Internal Model Control (Garcia and Morari (1982)), and the Internal Model Control (IMC) formulation of the state deadbeat controller (Zafiriou and Morari (1985)). The particular focus is on how each of these design methods deal with left half plane process zeros, without applying any tuning to the controller design.

Under certain conditions each of these controllers can give poor performance. The comparative performance analysis here shows under what conditions each controller can be expected to perform well or to perform poorly. Continuous time simulations of the process output and the control actions are used to explain the results. Zafiriou and Morari (1985) have provided an excellent study where they use continuous time simulations to show the conditions under which each of these controllers perform more poorly than the State Deadbeat controller. Where the study here provides an additional contribution is in highlighting the conditions where the LQOC controller can provide improved performance over the State Deadbeat controller.

3.1.1 The continuous FOPDT system and discrete model

The continuous time transfer function for a first order plus dead-time process is:

$$G_p(s) = \frac{K_p\, e^{-T_d s}}{T_p s + 1} \qquad (3.1)$$

The discrete model which is coincident with the continuous process at the control instants when the input is applied through a zero order hold is (Box and Jenkins (1970), Ragazzini and Franklin (1958)):

$$G_p(z^{-1}) = \frac{w(z^{-1})}{\delta(z^{-1})} = \frac{w_0 + w_1 z^{-1}}{1 + \delta_1 z^{-1}} z^{-b} \qquad (3.2)$$

where

$$\delta_1 = e^{-T_c/T_p}$$

$$w_0 = K_p(1 - \delta_1^{1-c})$$

$$w_1 = K_p(\delta_1^{1-c} - \delta_1)$$

$$b = f + 1$$

$$f = \text{int}(T_d / T_c)$$

and

$$c = (T_d - f\, T_c) / T_c$$

So, the discrete modified z transform model for a first order system contains a left half plane zero whenever the dead-time is not an integer multiple of the control interval, T_c. When $w_1 > w_0$ then the discrete model is non-invertible; that is, the inverse of the process is an unstable system with a pole outside the unit circle in the z domain.

3.1.2 The continuous SOPDT system and discrete model

The continuous time transfer function for a second order plus dead-time process with two real time constants is:

$$G_p(s) = \frac{K_p\, e^{-T_d s}}{(T_1 s + 1)\,(T_2 s + 1)} \tag{3.3}$$

The discrete model which is coincident with the continuous process at the control instants when the input is applied through a zero order hold is (Box and Jenkins (1970), Ragazzini and Franklin (1958)):

$$G_p(z^{-1}) = \frac{w(z^{-1})}{\delta(z^{-1})} = \frac{w_0 + w_1 z^{-1} + w_2 z^{-2}}{1 + \delta_1 z^{-1} + \delta_2 z^{-2}}\, z^{-b} \tag{3.4}$$

where

$$w_0 = K_p\,(T_1 - T_2)^{-1}\,[T_1(1 - S_1^{1-c}) - T_2(1 - S_2^{1-c})]$$

$$w_1 = -K_p\,(T_1 - T_2)^{-1}\,[(S_1 + S_2)(T_1 - T_2) + T_2 S_2^{1-c}(1 + S_1) - T_1 S_1^{1-c}(1 + S_2)]$$

$$w_2 = -K_p\, S_1\, S_2\,(T_1 - T_2)^{-1}\,[T_2(1 - S_2^{-c}) - T_1(1 - S_1^{-c})]$$

$$\delta_1 = S_1 + S_2 \quad , \quad \delta_2 = S_1\, S_2$$

$$S_1 = e^{-T_c/T_1} \quad , \quad S_2 = e^{-T_c/T_2}$$

$$b = f + 1$$

$$f = \text{int}(T_d / T_c)$$

and

$$c = (T_d - f\,T_c) / T_c$$

The discrete modified z transform model for a second order system usually has a stable zero on the negative real axis, and whenever the dead-time is not an integer multiple of the control time there is a second left half plane zero which may be noninvertible. Unlike the

first order system, it is not possible to separate the zero due to the fractional period of delay from the naturally occurring process zero.

3.1.3 Linear Quadratic Optimal Control with λ=0

The Minimum Variance (MV) controller (Åström (1970), Box and Jenkins (1970)) minimizes the objective function E [e_t^2], where E is the expectation operator and e_t is the process output deviation from set point. The MV controller is described here in the IMC form shown in Figure 3.1. In this form a model based controller is separated into a

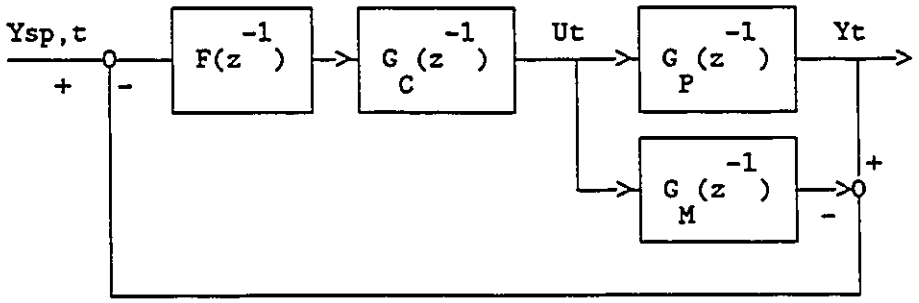


Figure 3.1 Internal Model Control Structure

filter $F(z^{-1})$, an approximate model inverse $G_C(z^{-1})$ and a process model $G_M(z^{-1})$. The MV controller for random walk disturbances or step changes in set point has a filter $F(z^{-1}) = 1$, a controller $G_C(z^{-1}) = G_{P0}^{-1}(z^{-1})$ (the inverse of the process model excluding the dead-time) and a discrete process model $G_P(z^{-1})$ which is the modified z transform of the continuous process $G_P(s)$. The overall feedback controller in a unitary feedback control structure is then

$$C(z^{-1}) = \frac{F(z^{-1})\ G_C(z^{-1})}{1 - F(z^{-1})\ G_C(z^{-1})\ G_p(z^{-1})} = \frac{G_{p0}^{-1}(z^{-1})}{1 - z^{-b}} \tag{3.5}$$

In the MV controller all of the zeros of the process become poles of the controller. When the process is non-invertible this leads to at least one unstable pole in the controller, which is cancelled by the corresponding zero of the process. In this situation the control input in response to a step change in set point will grow exponentially with time; however, the process output deviation from set point will be zero at the sampling instants.

The Linear Quadratic Optimal Controller (Åström (1970), Box and Jenkins (1970)) minimizes the objective function $E\ [\ e_t^2 + \lambda\ (\nabla^d\ U_t)^2\ ]$, where $\nabla^d = 1-z^{-1}$ is the backwards difference operator and d is the degree of nonstationarity in the disturbance or set point model. The only difference between the LQOC design and the MV design for step changes in set point is that the approximate model inverse is:

$$G_C(z^{-1}) = \frac{\delta(z^{-1})}{\gamma(z^{-1})} \tag{3.6}$$

where $\gamma(z^{-1})$ is obtained from the spectral factorization:

$$\gamma(z)\ \gamma(z^{-1}) = w(z)\ w(z^{-1}) + \lambda\ \delta(z)\ \nabla^d(z)\ \nabla^d(z^{-1})\delta(z^{-1}) \tag{3.6a}$$

$\gamma(z^{-1})$ is a polynomial having all of its roots in z inside the unit circle. The resulting controller is:

$$C(z^{-1}) = \frac{F(z^{-1})\ G_C(z^{-1})}{1 - F(z^{-1})\ G_C(z^{-1})\ G_p(z^{-1})} = \frac{\delta(z^{-1})}{\gamma(z^{-1}) - z^{-b}\ w(z^{-1})} \tag{3.7}$$

In particular, if the process is non-invertible so that $w(z^{-1})$ has some roots in z outside the unit circle, then if $\lambda=0$, $\gamma(z^{-1})$ has as its roots all stable roots of $w(z^{-1})$ and the reflection through the unit circle of any unstable roots of $w(z^{-1})$. So, with $\lambda=0$ the LQOC controller provides the same controller as the MV controller for an invertible process, and a stable controller for a non-invertible process.

For a FOPDT system $w(z^{-1})$ has only one root. Whenever this root is stable it becomes a pole of the approximate model inverse $G_c(z^{-1})$ for the LQOC controller with $\lambda=0$. In this case the LQOC controller and the MV controller are identical to the output deadbeat controller of Ragazzini and Franklin (1958) (MacGregor et al (1984)). When the root of $w(z^{-1})$ is unstable, then the approximate model inverse $G_c(z^{-1})$ in the LQOC controller with $\lambda=0$ has as a pole the reflection of the root through the unit circle, $\gamma(z^{-1}) = w_1 + w_0 z^{-1}$. This controller can still produce a ringing manipulated input in response to a set point change. Ringing in the controller can be reduced or eliminated through tuning: choosing an appropriate value of λ. In this section, the focus is on the model inversion step and not on controller tuning, so only the design with $\lambda=0$ is considered.

3.1.4 Minimum Variance control with the controller pole replaced by a gain

The Dahlin controller design (Dahlin (1968)), is identical to the MV controller in its choice of a direct model inverse. The Dahlin controller uses a first order filter $F(z^{-1})$ for tuning. Again, in order to focus on the model inverse only this section considers only $F(z^{-1})=1$. In order to suppress ringing the Modified Dahlin (Dahlin (1968))

controller replaces the ringing pole in the feedback controller, equation (3.5), with its steady state gain.

3.1.5 IMC control with the model inverse pole replaced by a gain: state deadbeat control

In the original formulation of IMC (Garcia and Morari (1982)) a model inverse based on LQOC control with λ=0 was suggested: that the unstable poles in the model inverse be reflected through the unit circle. Zafiriou and Morari (1985) suggest replacing all noninvertible and left half plane model inverse poles with a gain. For this choice the controller block in the IMC structure becomes:

$$G_C(z^{-1}) = \frac{\delta(z^{-1})}{\sum_i w_i} \tag{3.8}$$

In IMC, as in the Dahlin controller, the filter block $F(z^{-1})$ is used to detune the controller. In order to focus on the model inverse this section considers only $F(z^{-1})$=1. The feedback controller is then

$$C(z^{-1}) = \frac{F(z^{-1})\ G_C(z^{-1})}{1 - F(z^{-1})\ G_C(z^{-1})\ G_p(z^{-1})} = \frac{\delta(z^{-1})}{\sum_i w_i - z^{-b}\ w(z^{-1})} \tag{3.9}$$

This controller will be referred to as the state Deadbeat IMC controller. This is a state deadbeat type controller as described by Kalman (1954) and by Jury (1958). In this design, the model inverse is chosen to drive the discrete process output to zero in the minimum number of control intervals such that the discrete input settles at the new steady state in a finite number of steps. The minimum number of steps required for the control actions to settle is the order of the

process model. This can be seen by reference to the IMC structure of Figure 3.1. If the process model is exact then the feedback path will be zero. If the filter is $F(z^{-1})=1$ and the approximate model inverse does not have a pole, that is the $G_c(z^{-1})$ has only a numerator which cancels the process denominator, then the manipulated control actions will settle to the new steady state after a set point change in n intervals, where n is the order of the process model. Also, the process output will settle to the new steady sate after a set point change after m intervals, where m is the order of the process numerator.

3.1.6 Effect of the control interval on the ISE performance of discrete model based controllers applied to a FOPDT system

For a first order system with no dead-time, the three controllers presented above are identical, as the discrete process model has no zero. Consider a first order process with gain $K_p=1$ and time constant $T_p=1$ and with no dead-time. The ISE as a function of the control interval for MV control in response to a step set point change of 1 is plotted in Figure 3.2. As expected, the ISE falls towards zero as the size of the control interval is reduced; in the limit continuous Linear Quadratic Optimal Control would produce zero ISE, but the controller gain would be infinite. This is illustrated by a plot of the variance of the VU_c vs the control interval shown in Figure 3.3. A time plot showing the continuous time process output and the manipulated control actions for a control interval of 1.5 is shown in Figure 3.4.

This example now provides a basis for considering the step response of the same first order system with dead-time $T_d=1$. Because of the time delay, the minimum possible ISE is 1, as the output can not

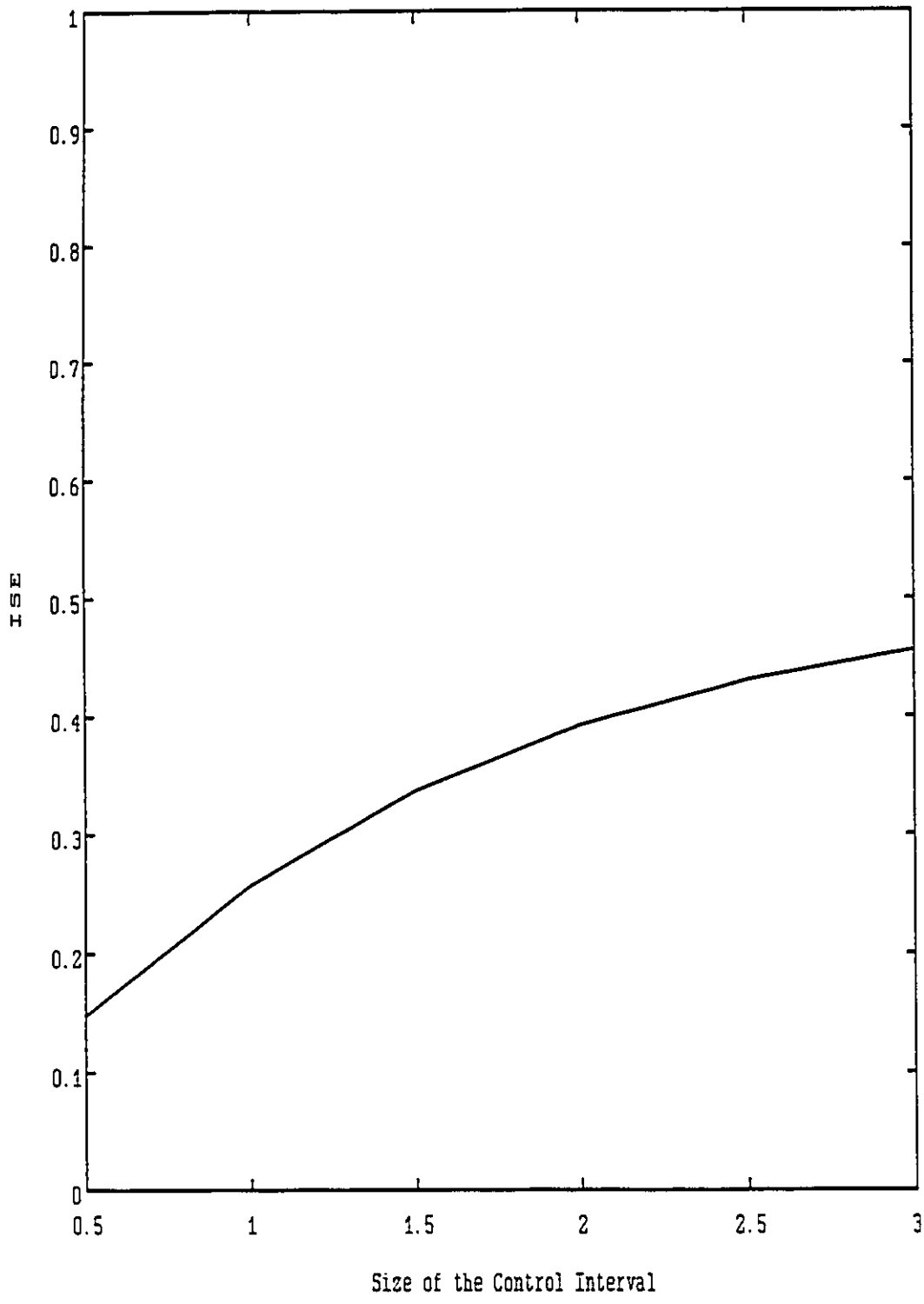


Figure 3.2 FOPDT Step Set Point Change with MV Control ISE vs Tc

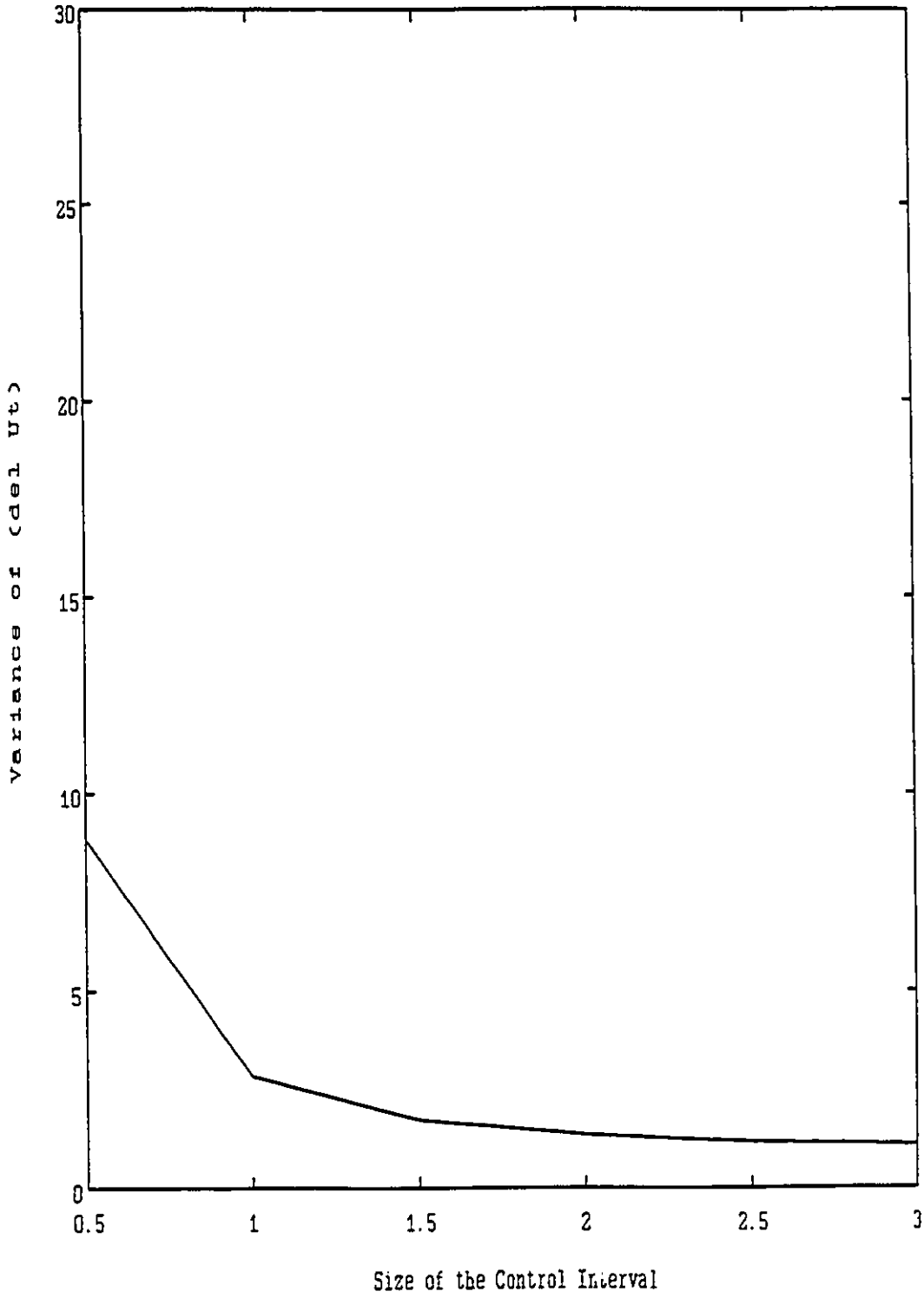


Figure 3.3 FOPDT Step Set Point Change with MV Control Var(VU_t) vs T_c

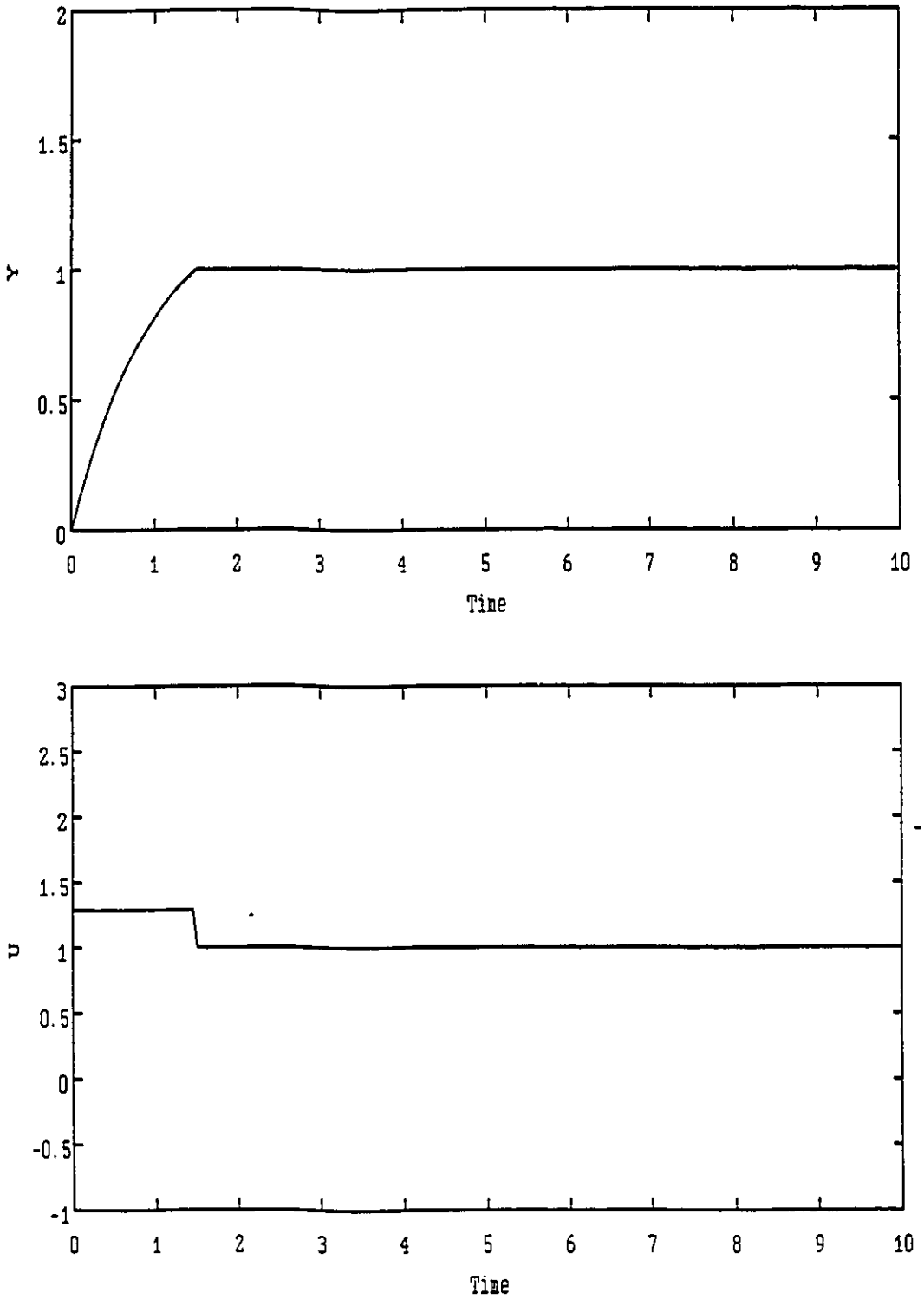


Figure 3.4 FOPDT Step Set Point Change, MV Control, Y & U vs t, T_c=1.5

respond before the dead-time has passed. Figure 3.5 show the ISE vs the size of the control interval for the LQOC controller with λ=0, the State Deadbeat IMC controller and the Modified Dahlin controller. Figure 3.6 shows the variance of the VUt vs the size of the control interval for these three controllers. For the LQOC controller with λ=0, the ISE shows peaks at control intervals of T_c=0.65 and T_c=1.5. In order to explain the peak at the control interval T_c=1.5 the continuous time process output and control actions are plotted in Figure 3.7. This figure shows that the system is stable as the control inputs are gradually decaying, and that the output is at its set point at the discrete control instants beginning the first control instant after the set point change. However, the severe ringing of the manipulated control actions causes large cycles in the continuous output which die out only slowly.

The ringing pole in the controller is due to the left half plane zero in the process model. When T_c=1.5 the discrete model is:

$$G_p(z^{-1}) = \frac{0.3935 + 0.3834\ z^{-1}}{1 - 0.2231\ z^{-1}}\ z^{-1} \qquad (3.10)$$

The ringing could be eliminated by choosing non-zero values of the input penalty parameter λ, but that will be examined in a later section.

The ISE's calculated in this section are based on integrating the continuous time process output over 10 time units for values of the control interval T_c between 0.1 and 5, calculated using a step size of .01 for T_c=.1 and a step size of .05 for all other values of T_c. As Figure 3.9 suggests this means that when the process output does not settle within 10 time units the ISE is underestimated. Thus, the peaks

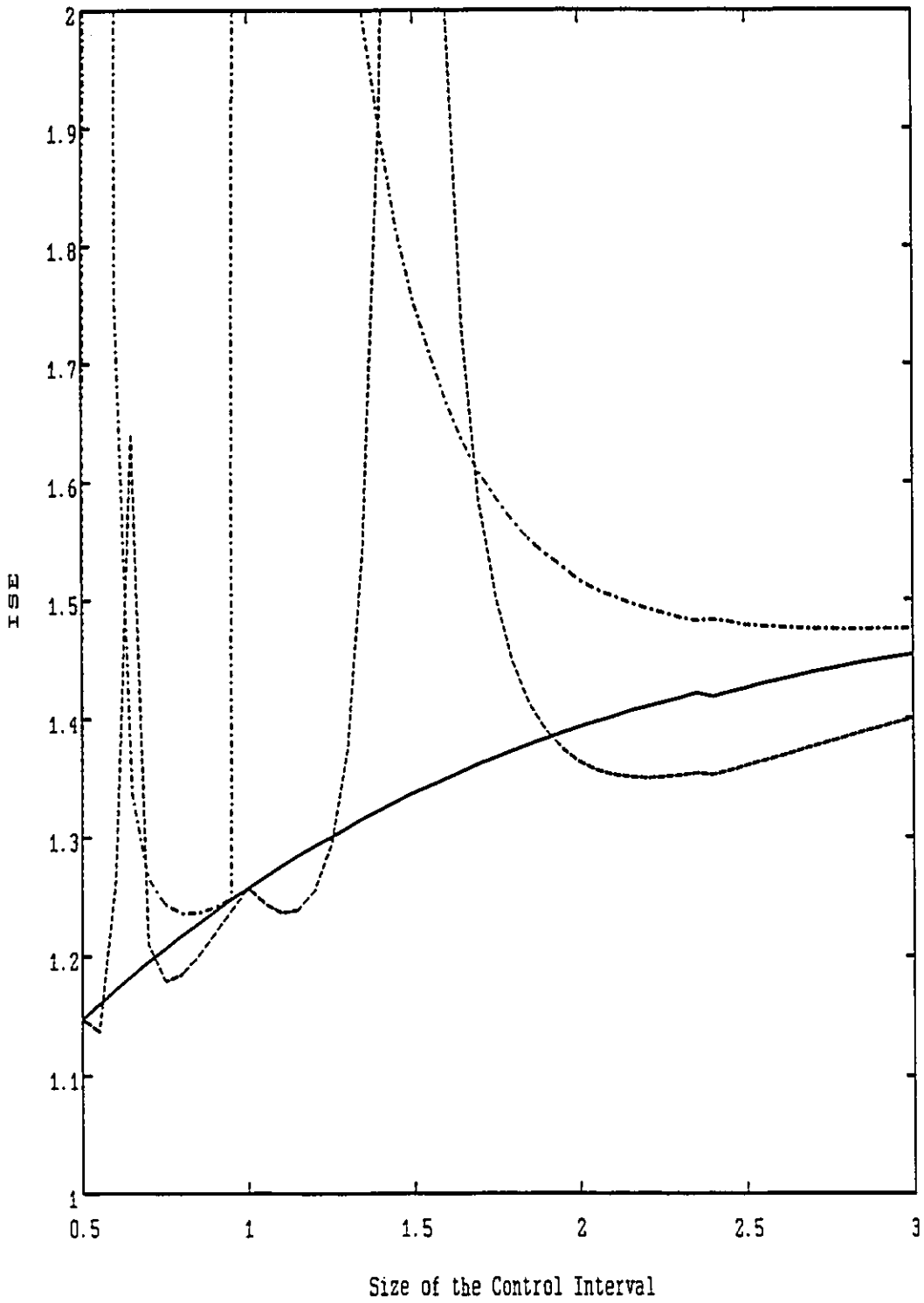


Figure 3.5 FOPDT Step Set Point Response ISE vs T_c
- State Deadbeat IMC, -- LQOC, -. Modified Dahlin

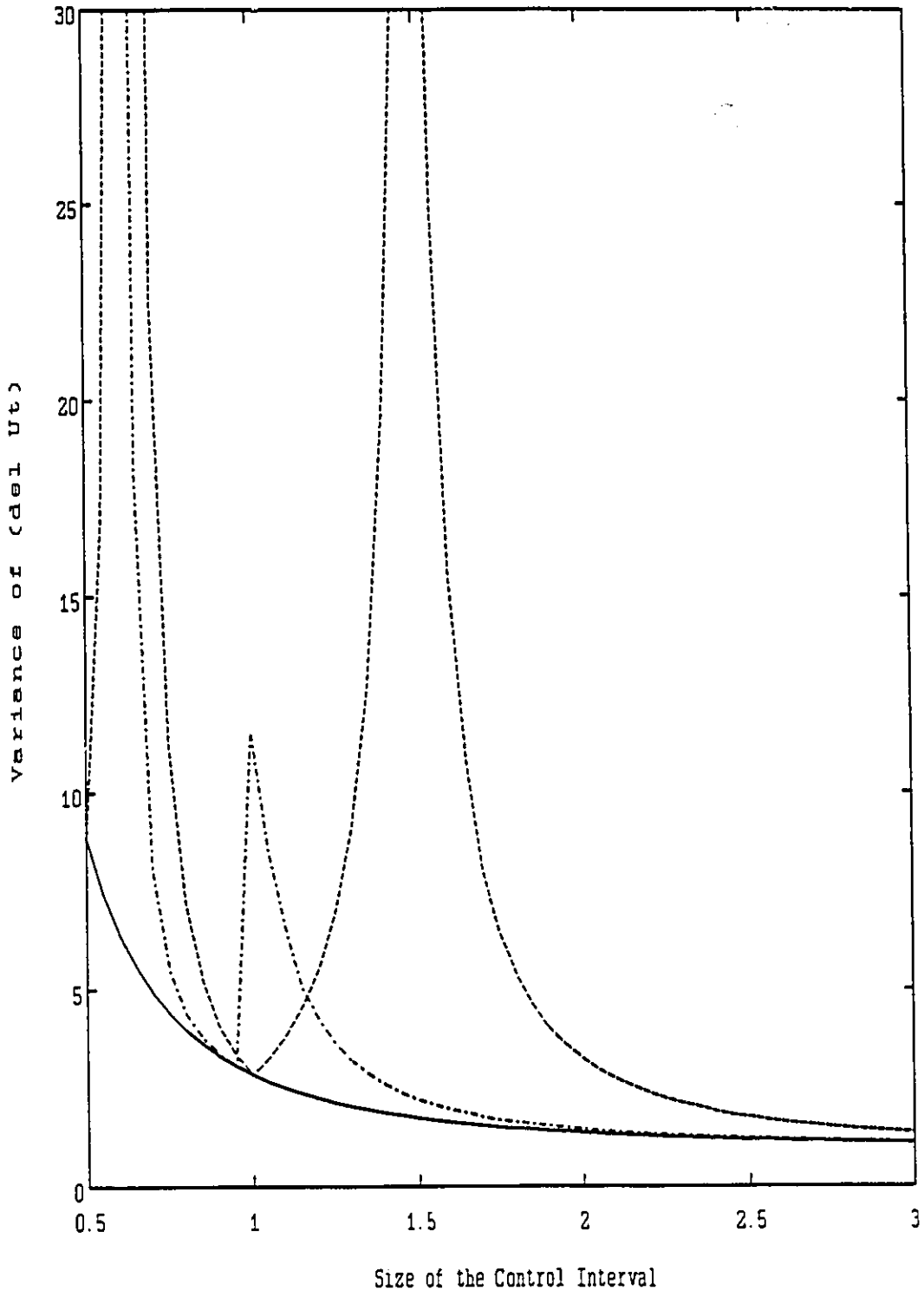


Figure 3.6 FOPDT Step Set Point Change Var(VU_t)
- State Deadbeat IMC, -- LQOC, -. Modified Dahlin

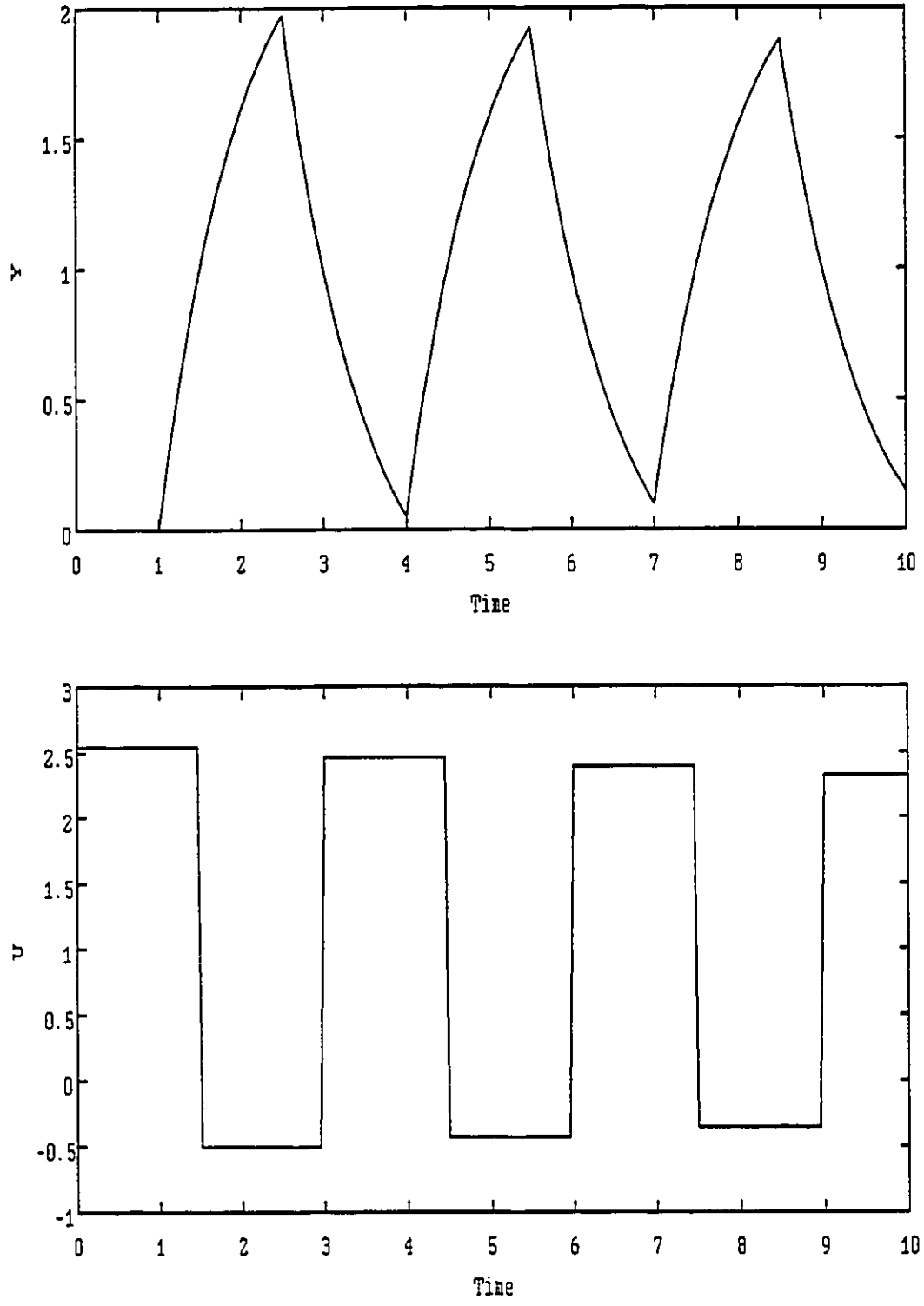


Figure 3.7 FOPDT Step Set Point Change, LQOC λ=0, Y & U vs t, T_c=1.5

on the ISE curve may be underestimated, but the trend is correct.

Figure 3.5 shows a dip in the ISE on either side of the control interval T_c=1.0. For these values of the control time the controller has a left hand plane pole close to the origin. For example, Figure 3.8 shows the continuous time process output and control actions for T_c=1.05. The model inverse for this control interval is

$$G_C(z^{-1}) = \frac{1 - 0.3499\ z^{-1}}{0.6013 - 0.2231\ z^{-1}} \tag{3.11}$$

which has a pole at z=-0.081. As evidenced by the plot of the control actions, this is not considered a ringing pole. However, the output response shows about 10% overshoot, unlike the response of the process without dead-time shown in Figure 3.4. The ISE performance criterion penalizes large errors more than small errors, so the fast rise time which accompanies the overshoot leads to a reduced ISE. This indicates that a pole on the negative real axis in the model inverse which is not close enough to minus one to cause ringing can enhance the ISE performance of the continuous output.

The ISE vs control time plot for the modified Dahlin method of eliminating controller ringing is shown in Figure 3.5 and the variance of VUt vs control time shown in Figure 3.6. It is interesting to note that this design method has eliminated the large ISE peak near the control interval T_c=1.5, where the LQOC with λ=0 controller had a ringing pole, but it has introduced a large peak in the ISE as control intervals slightly larger than 0.5 and than 1. This can be explained by examining the MV control equation (3.3). As the control time is increased through 1, the process model changes from

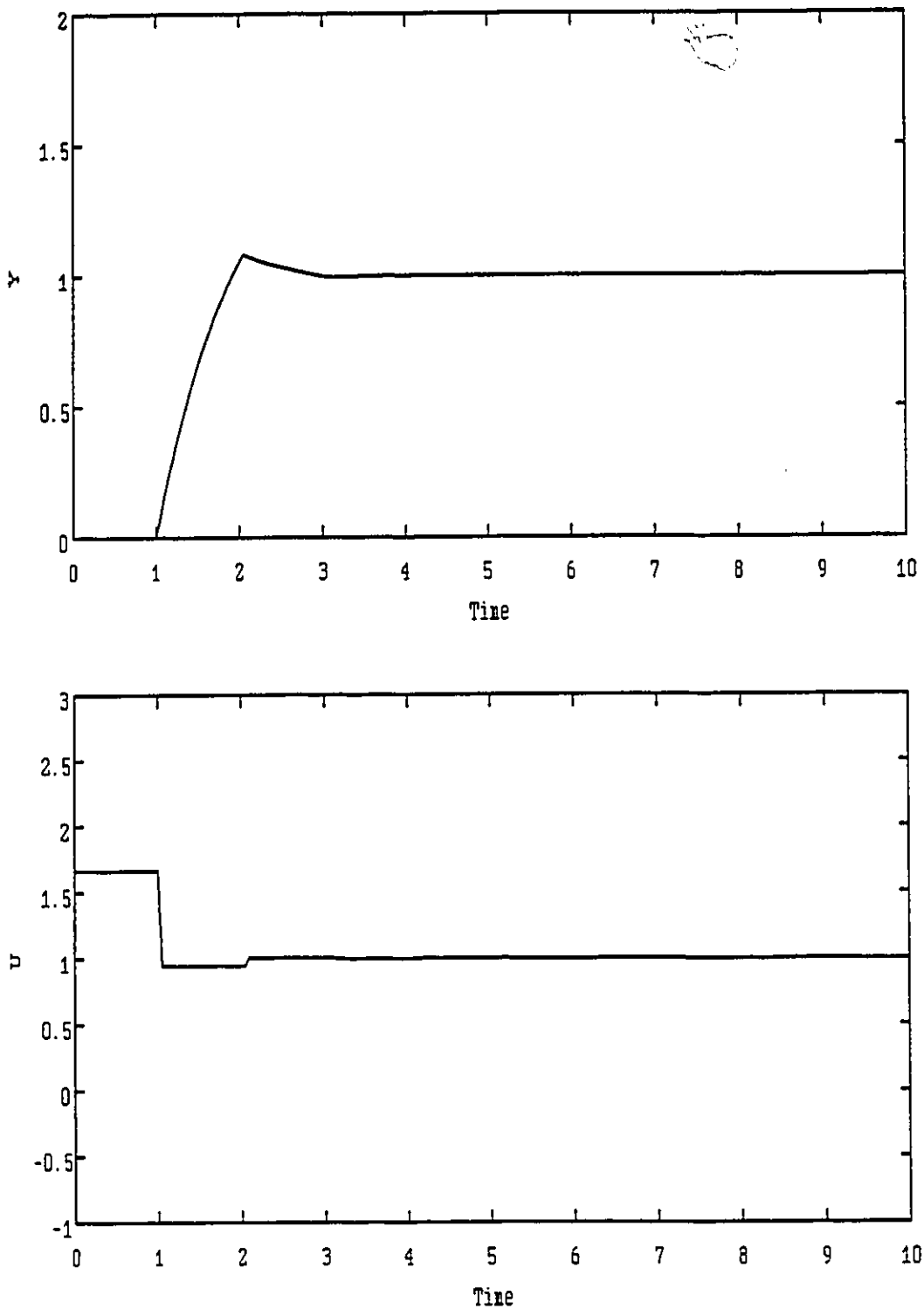


Figure 3.8 FOPDT Step Set Point Change, LQOC λ=0, Y & U vs t, Tc=1.05

$$G_p(z^{-1}) = \frac{0.63\ z^{-1}}{1 + 0.37\ z^{-1}}\ z^{-2} \quad \text{to} \quad \frac{0 + 0.63\ z^{-2}}{1 + 0.37\ z^{-1}}\ z^{-1} \tag{3.12}$$

If the controller pole is replaced by a gain, as the control time is increased through 1, the process model inverse changes only slightly, but the value of b changes from 2 to 1. Thus there is a large mismatch between the controller and the process being controlled.

The process response and the manipulated control action for a control time T_c=1.5 are shown in Figure 3.9. The process output is not very satisfactory showing a slow rise time, overshoot and slow settling. The manipulated control action does not ring, but the controller performance is not very good.

Plots showing the ISE and the variance of VUt vs the control time for the state deadbeat design are shown in Figures 3.5 and 3.6. These plots reproduce exactly those for the undelayed system shown in Figures 3.2 and 3.3, although the ISE is inflated by 1, the error which cannot be removed due to the dead-time. Figure 3.10 shows the process response and the manipulated input for a control interval T_c=1.5. Both the process response and the manipulated control actions behave exactly as in the undelayed response of Figure 3.4. The output reaches the new set point after $T_c + T_d$ units of time, and is constant thereafter. The state deadbeat design reproduces the manipulated control actions and the process output of the undelayed process. This is a very attractive controller design as neither the manipulated control actions nor the process output are affected by the addition of pure dead-time. However, Figure 3.5 shows that the LQOC controller with λ=0 has a lower ISE for many values of T_c. The LQOC controller, as shown above, provides a

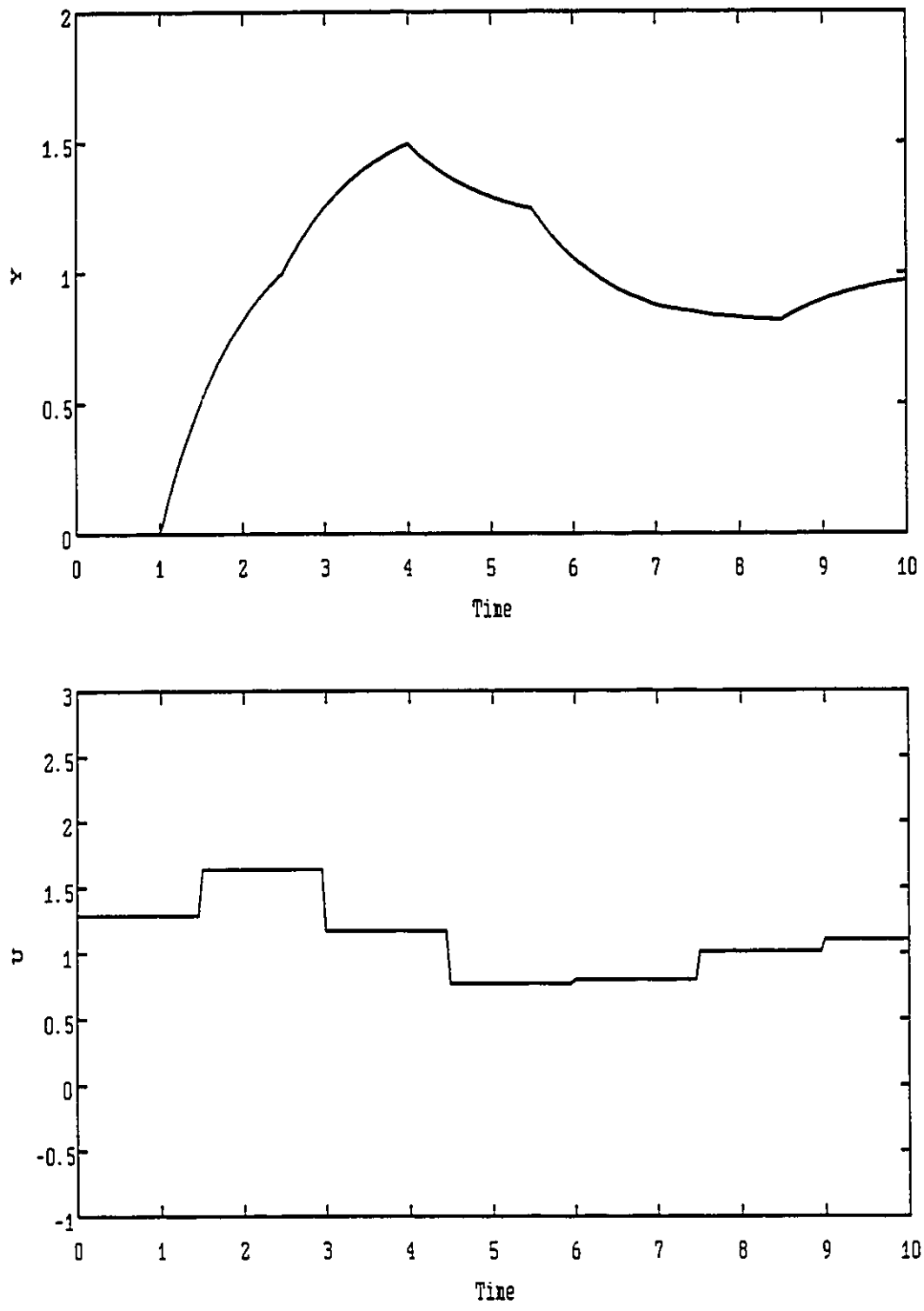


Figure 3.9 FOPDT Step Set Point Change, Modified Dahlin Control Y & U vs t, T_c=1.5

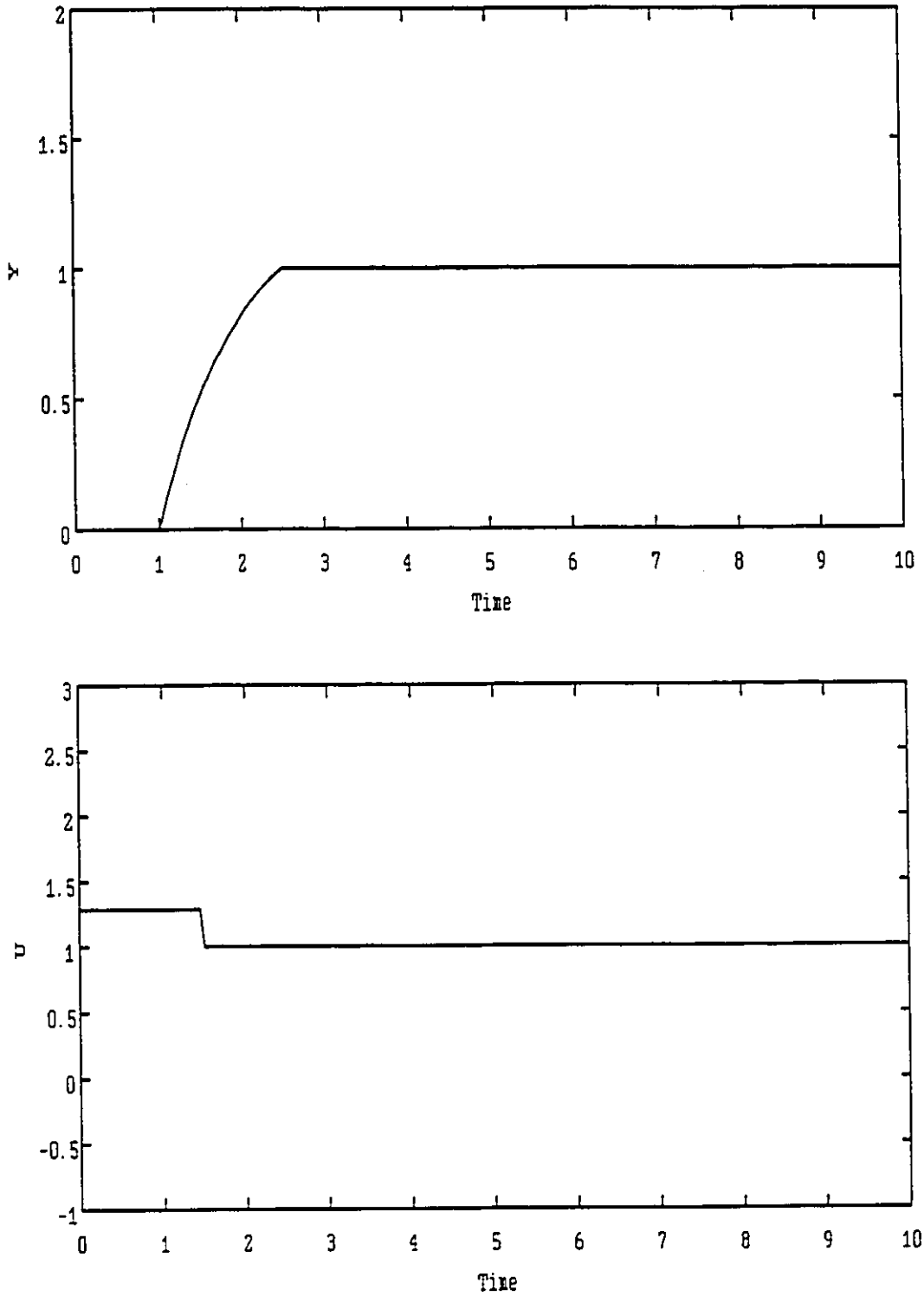


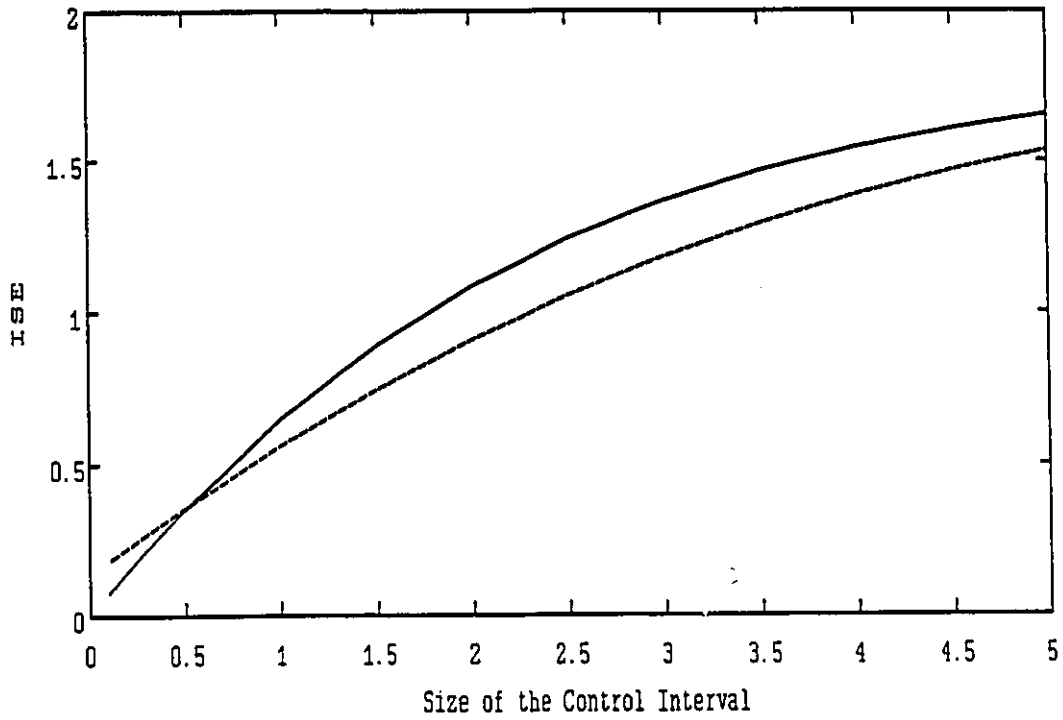
Figure 3.10 FOPDT Step Set Point Change, State Deadbeat IMC
Y & U vs t, T_c=1.5

faster process response when it has a left half plane but non ringing pole in the controller due to the fractional period of delay.

3.1.7 Effect of the control interval on the ISE performance of discrete model based controllers applied to a second order system

The discrete z transform of a second order system with no dead-time has a zero on the negative real axis. Consider a second order process with gain K_p=1 and time constants T_1=1 and T_2=2 and no dead-time. The ISE and IAE as a function of the control interval for LQOC control with λ=0 and for state deadbeat control are shown together in Figure 3.11. For control intervals less than T_c=0.5, the state deadbeat controller produces a lower ISE in the output than the state deadbeat controller. For control intervals less than T_c=2, the state deadbeat controller has a lower IAE in the output. So, for control intervals T_c<0.5 the state deadbeat controller has both a lower ISE and IAE; for 0.5<T_c<2 the state deadbeat controller has a lower IAE but a higher ISE; and for T_c>2 the state deadbeat controller has both higher ISE and IAE than the LQOC controller with λ=0. This shows that neither of these two designs has better continuous time performance across the whole range of control times than the other.

In order to explain why the state deadbeat controller has a lower ISE and IAE performance for short sampling times than the LQOC controller with λ=0, the output responses and control actions for a unit step set point change with T_c=0.1 are plotted in Figure 3.12. The LQOC response has a faster rise time: the output passes through the set point at T=0.1, has a 50% overshoot, and shows a characteristically ringing



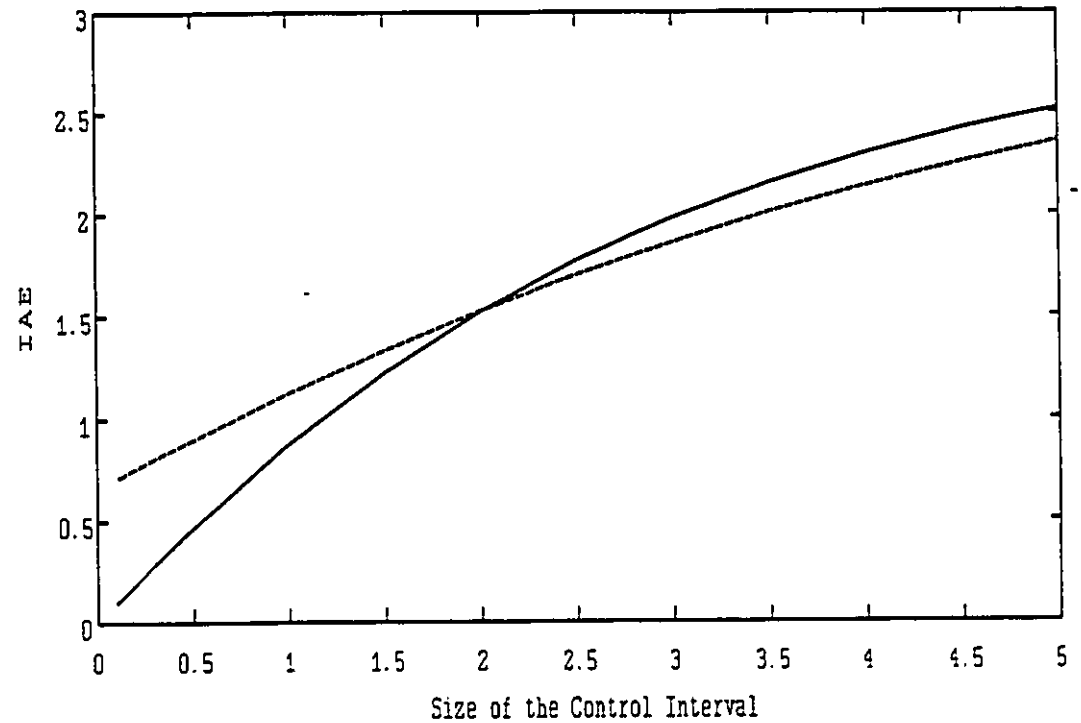


Figure 3.11 Second Order Step Set Point Change ISE and IAE vs Tc
- State Deadbeat IMC, -- LQOC λ=0

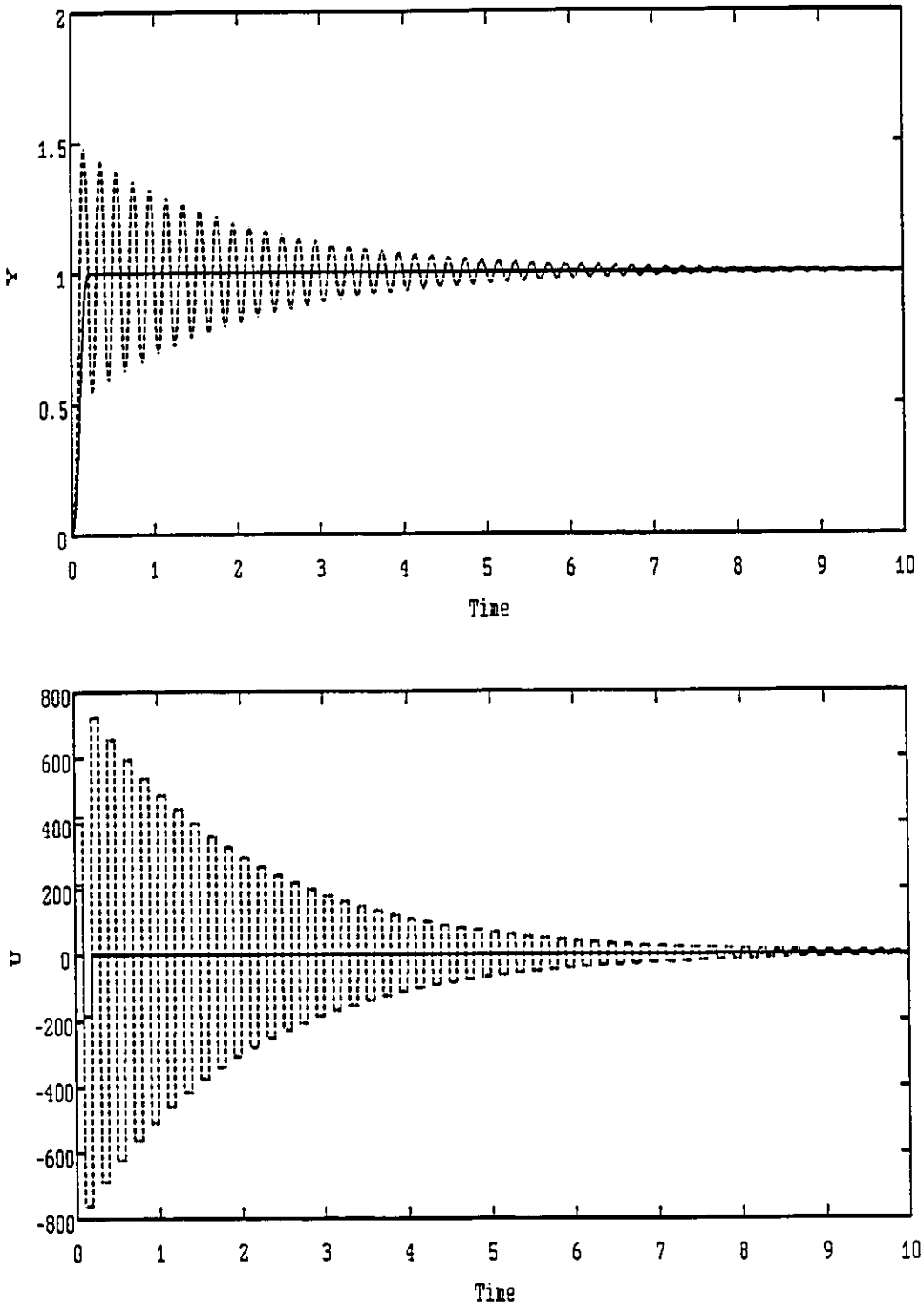


Figure 3.12 Second Order Step Set Point Change with T_c=0.1
- State Deadbeat IMC, -- LQOC with λ=0

response. The state deadbeat response only reaches the set point at T=0.2, but it remains at the set point thereafter. The excessive overshoot and ringing in the LQOC output lead to the higher IAE and ISE.

The state deadbeat controller input settles at its new steady state at Tc=0.2, whereas the LQOC input shows a strong ringing pattern. It is important to note that both of these controllers show extremely large input actions above and below the final steady state value of 1. This is typical of both LQOC and state deadbeat designs when the sampling interval is very short relative to the process time constants when they are not detuned.

Figure 3.13 shows the process outputs and control inputs for control time T_c=0.5. Again the process output with LQOC control shows a rapid rise time with significant overshoot and some oscillation. The state deadbeat control produces a process output which again takes two control intervals to reach the set point. These two responses have approximately equal ISE performance but the state deadbeat response has a lower IAE. That is, although the LQOC response has significant oscillation, its faster response time has a large effect on reducing the ISE error: large errors squared have a larger effect on the ISE than small errors. The oscillatory behaviour of the output with LQOC control is due to the ringing in the manipulated control action.

With a control time T_c=2 both the LQOC control and state deadbeat control produce the same IAE. In this case the output responses shown in Figure 3.14 reveal that the LQOC response has a much faster rise time but now has less than 20% overshoot: the improved rise time is accompanied by only slight oscillation. Examining the

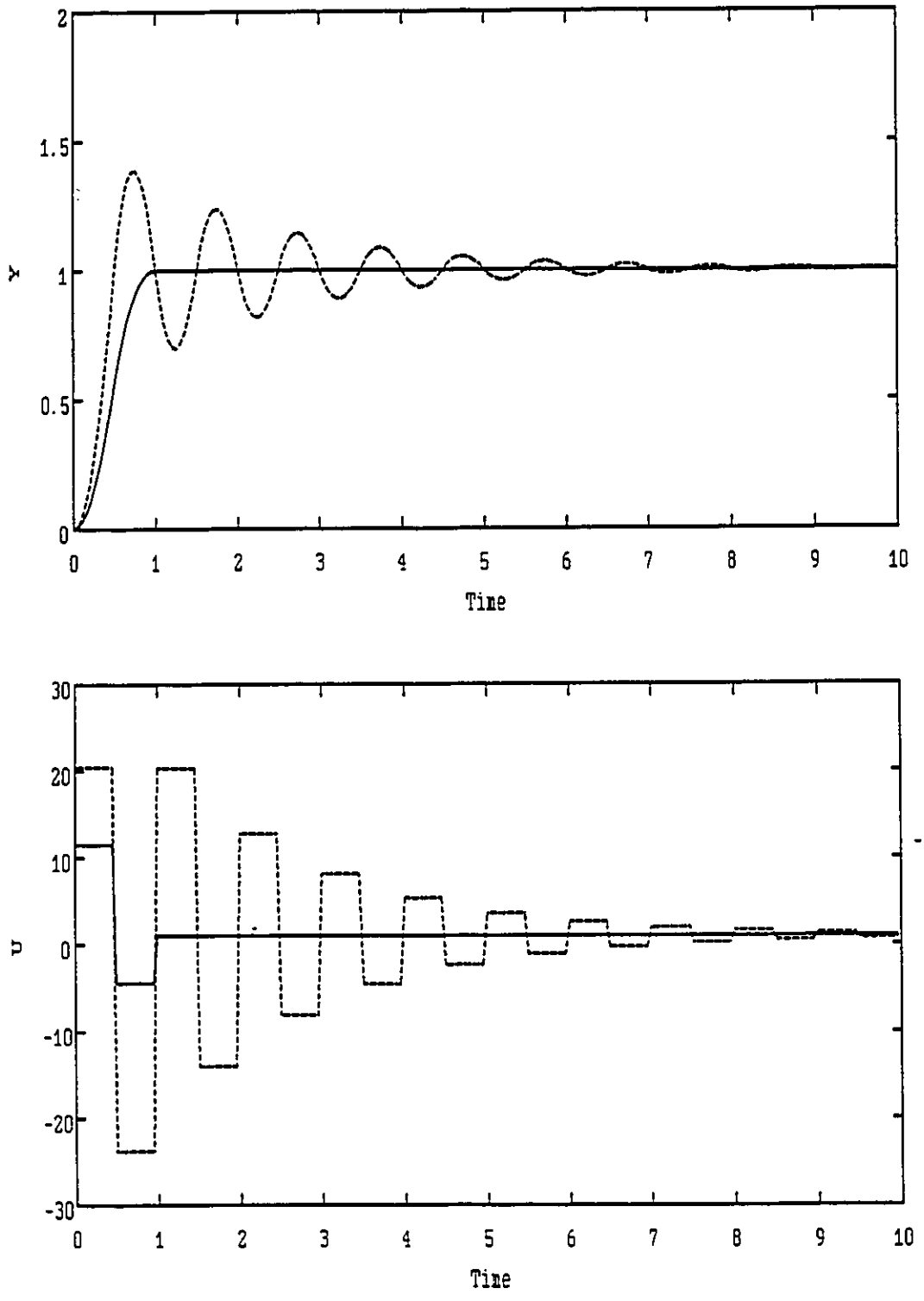


Figure 3.13 Second Order Step Set Point Change with T_c=0.2
- State Deadbeat IMC, -- LQOC with λ=0

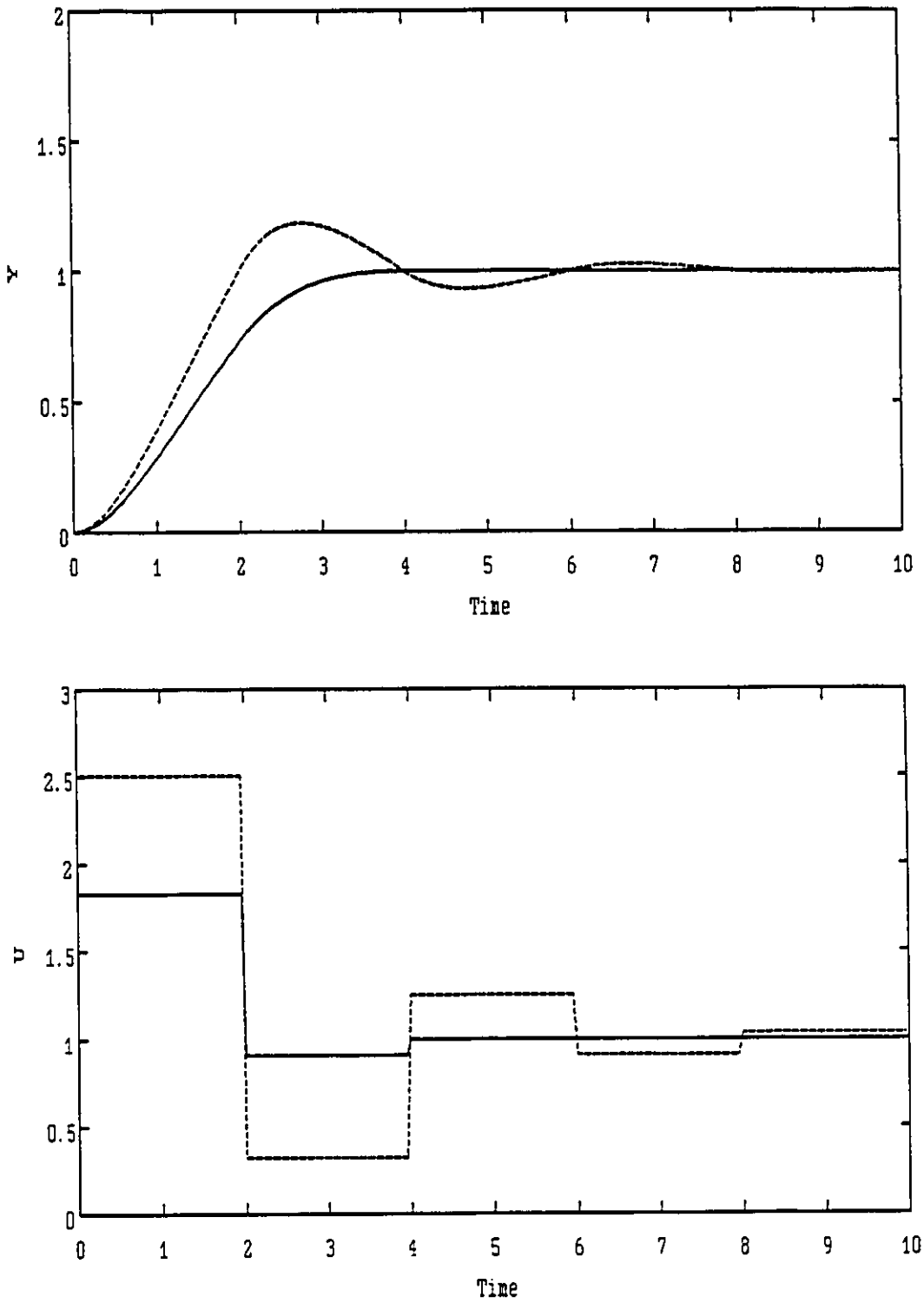


Figure 3.14 Second Order Step Set Point Change with T_c=2
- State Deadbeat IMC, -- LQOC with λ=0

manipulated control actions in Figure 3.14 shows that the LQOC input does not ring significantly. Also, the input manipulations are all much milder than when the control time is shorter. This again illustrates that a pole on the negative real axis in the approximate model inverse which is not close enough to minus one to cause ringing can enhance the ISE performance of the continuous output.

The LQOC design with $\lambda=0$ for an invertible system, such as in this example, always drives the output error to zero after one control action. This ensures a fast response time, however it also leads to overshoot and oscillation. The size of the overshoot and the duration of the oscillation decreases as the control interval is increased. The state deadbeat design requires as many steps as the system order, two in this example, to drive the output error to zero where it remains thereafter. The slow response time is compensated for by no overshoot. The balance between performance improvements from a fast rise time and performance degradation due to oscillation depends on the size of the control interval. For short control intervals the lack of oscillation in the state deadbeat design provides improved performance; whereas, for long control intervals the fast rise time of the LQOC design provides the better performance.

3.1.8 Summary

This section has focused on the effect of the approximate model inverse in the controller design. The questions of how to tune a controller design, and how to design for a particular type of disturbance or set point change have been left for later sections.

Three approaches to the calculation of the approximate model inverse in controller designs were examined with application to a first order plus dead-time process and a second order process. These inverse designs were the Linear Quadratic Optimal Control with λ=0, the modified Dahlin method, and the state deadbeat IMC method. For a first order system with no dead-time these model inverses are all identical, but differ for a first order system plus dead-time. The LQOC approximate model inverse is seen to have a ringing pole causing poor ISE performance whenever the fractional period of delay results in a process with a zero near minus one. However, the LQOC approximate model inverse produces improved ISE performance relative to State Deadbeat IMC when there is a process zero which is not near minus one. The modified Dahlin controller has poor ISE performance when the fractional period of dead-time approaches zero. The state deadbeat IMC calculation of the approximate model inverse produces a controller which gives the same ISE performance for the delayed process (inflated by the dead-time) as for the undelayed system. This is because the state deadbeat design produces the same control actions in response to a set point change for the first order system plus dead-time as for the undelayed first order system, and as a result, the output performance for the delayed system is identical as for the undelayed system.

The LQOC design with λ=0 and the state deadbeat IMC design were also applied to a second order process with no dead-time. This process has an invertible zero which is near minus one when the control interval is small. The LQOC design with λ=0 is seen to have the better ISE and IAE performance at long control intervals because of its more rapid rise

time, whereas it has poorer ISE and IAE performance at small control intervals because of excessive oscillation in the output, caused by the ringing in the control actions.

The state deadbeat approximate model inverse is not subject to ringing in the manipulated control actions which can cause deteriorated ISE and IAE performance. However, when the discrete model of a continuous process has a left half plane zero which is not near minus one (which will not cause a ringing pole in the LQOC design), the LQOC controller produces improved ISE performance than the state deadbeat IMC controller.

3.2 Tuning: LQOC, State Deadbeat and Extended Horizon Designs; Regulatory and Servo Control

This section examines the spectral factorization Linear Quadratic Optimal Control (Åström (1970), Box and Jenkins (1970)), state deadbeat Internal Model Control (Zafiriou and Morari (1985)) and extended horizon (Ydstie et al (1985)) approaches to discrete controller tuning. In addition a new modified Linear Quadratic Optimal Controller and a new technique for extended horizon tuning are presented. In this section the controller tuning problem is separated from the disturbance filtering problem because the choice of the disturbance filter is easily confounded with controller tuning.

Controller tuning is often termed detuning; in the controller design techniques considered here, as with most discrete design techniques, the basic design uses a zero or minimum value of the tuning parameter to obtain the best possible discrete output performance

according to the performance criterion which defines the design method. When the tuning parameter is increased from its minimum value, output performance is compromised, hence, detuning. However detuning usually improves some other system characteristic such as increasing robustness or reducing the manipulated input power requirements.

3.2.1 Controller Structure: Filtering, Prediction and Tuning

The IMC structure of Figure 3.1 is expanded in Figure 3.15. The controller components are now an approximate model inverse $G_C(z^{-1})$, a tuning filter $F(z^{-1})$, a model for prediction $G_M(z^{-1})$, a set point prediction filter $F_{SP}(z^{-1})$, and a disturbance prediction filter $F_N(z^{-1})$. These components are most easily explained in the terminology of optimal stochastic control, following Bergh (1986).

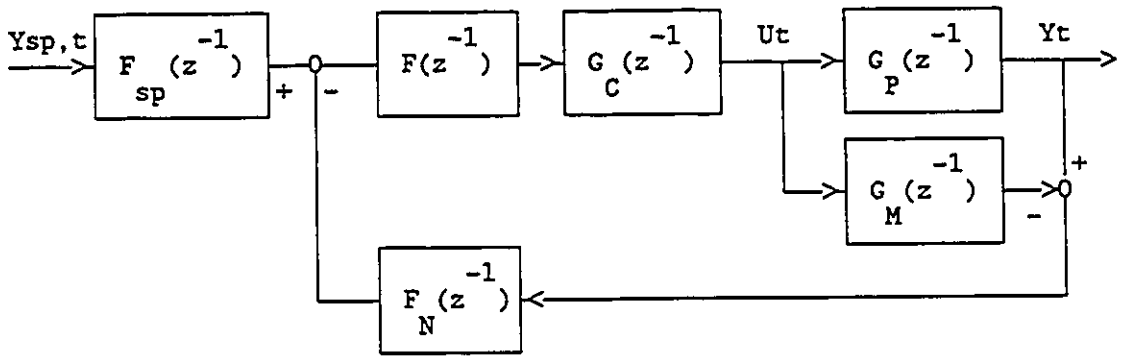


Figure 3.15 Expanded Internal Model Control Structure

The process model is a discrete transfer function (Box and Jenkins (1970)):

$$G_M(z^{-1}) = \frac{w(z^{-1})}{\delta(z^{-1})} z^{-b} \tag{3.13}$$

where the numerator is of order s and the denominator is of order r and the delay is b.

The disturbance prediction filter design assumes that the disturbance can be modelled as an Autoregressive Integrated Moving Average (ARIMA) process (Box and Jenkins (1970)):

$$G_D(z^{-1}) = \frac{\theta(z^{-1})}{\nabla^d \phi(z^{-1})} a_t \tag{3.14}$$

where the moving average portion, the numerator, is of order q and the autoregressive portion, the denominator, is of order p. The degree of nonstationarity in the disturbance is d. The design equation for the prediction filter is then (Box and Jenkins (1970)):

$$\theta(z^{-1}) = \psi_{b-1}(z^{-1}) + T(z^{-1}) \nabla^d \phi(z^{-1}) \tag{3.15}$$

where $\psi_{b-1}(z^{-1})$ is of degree b-1. This equation can be solved by long division of equation (3.14) and noting that $T(z^{-1})$ is the remainder. The filter

$$F_N(z^{-1}) = \frac{T(z^{-1})}{\theta(z^{-1})} \tag{3.16}$$

applied to the disturbance N_t calculates the b step ahead prediction of the disturbance. Note that b is the process dead-time and that no error measured now can be cancelled for b steps; hence, the b step ahead prediction of the error.

Similarly, if the typical set point change can be described as an ARIMA process then a b step ahead prediction filter $F_{SP}(z-1)$ can be designed for the set point (MacGregor et al (1984), Bergh (1986)) using equations (3.14), (3.15) and (3.16). These authors note that the ARIMA model structure includes step changes, ramps and exponential rises to the new set point.

The choice of the approximate model inverse $G_c(z^{-1})$ was discussed in the previous section. Minimum Variance, Deadbeat and IMC controllers are designed using the direct inverse of the process (excluding the dead-time). In the previous section the Linear Quadratic Optimal Control with $\lambda=0$ and State Deadbeat designs for the model inverse were compared. These latter two designs will form the basis of the discussion of controller tuning in this section.

The tuning filter $F(z^{-1})$ appears in the block diagram of Figure 3.15 in series with the approximate process model inverse $G_c(z^{-1})$. The tuning filter can be chosen by the designer to modify the model inverse in order to tune the controller design. When the disturbance prediction filter $F_N(z^{-1})$ and the set point prediction filter $F_{SP}(z^{-1})$ are identical then they can be combined as one filter in the same place as a combined prediction filter next to the tuning filter $F(z^{-1})$ in Figure 3.15 (Bergh (1986)). This duality between disturbance and set point prediction filters and tuning filters has led to some confusion in the literature.

Palmor and Shinnar (1979) note that Minimum Variance controllers designed for Integrated Moving Average ARIMA(0,1,1) disturbances have very good robustness properties, and so recommend this disturbance model

as the basis of good designs. They recommend treating the parameter of the ARIMA(0,1,1) disturbance as a tuning parameter. Similarly Garcia and Morari (1982) recommend a design with a Minimum Variance approximate model inverse and a first order filter for tuning. However, they do not include a disturbance prediction or a set point prediction filter in their design.

Bergh (1986) has done an excellent study comparing the LQOC design method which includes a disturbance prediction filter and a tuning parameter, which can be interpreted as generating a tuning filter, with the IMC design which has only a tuning filter. Bergh shows that the LQOC design has superior discrete output performance to this IMC design at the same level of robustness for autoregressive disturbances typical of process load disturbances.

Zafiriou and Morari (1985) recommend the IMC design using a state deadbeat model inverse and a first order filter for tuning. Morari and Zafiriou (1989) extend the design for inputs other than step inputs.

Bergh (1986) presents a very lucid and clear construction of the elements of controller design which have been reiterated here. Any controller design can be interpreted using these elements. Also, any controller design which is lacking in any of the elements presented here, such as a disturbance prediction filter, can easily be extended using the design method described here to include that element. This idea of taking a controller design concept and expanding it to make it

"complete" will form the basis for most of the developments in this section.

The focus of this section is on tuning. However, because of the duality between tuning and prediction filters it is important to understand both in each controller design. In the case of the LQOC controller design, tuning affects both the tuning and prediction filters Bergh (1986). In order to minimize the effect of the prediction filter a random walk ARIMA(0,1,0) disturbance, equivalent to a step change in the set point, will be used to illustrate the design examples below.

3.2.2 The Linear Quadratic Optimal Controller

The design equations for the LQOC controller were introduced in section 3.1.3. In that section the design with λ=0 was introduced as a model inverse. In this section designs with nonzero λ are compared to the design with λ=0 in order to understand the effect of the tuning parameter. The effect of the tuning parameter λ is interpreted in terms of a tuning filter. In the LQOC design the prediction filter is also affected by the value of λ, but the focus of the discussion here is primarily on the tuning filter.

In equation (3.6) the LQOC controller model inverse is presented. Here we rewrite this equation as

$$G_C(z^{-1})\ F\ (z^{-1}) = \frac{\delta(z^{-1})}{\gamma_0(z^{-1})}\ \frac{\gamma_0(z^{-1})\ \ \gamma(1)}{\gamma(z^{-1})\ \ \gamma_0(1)} \qquad (3.17)$$

where $\gamma_0(z^{-1})$ and $\gamma(z^{-1})$ are obtained from the spectral factorization equation:

$$\gamma(z)\ \gamma(z^{-1}) = w(z)\ w(z^{-1}) + \lambda\ \delta(z)\ \nabla^{d}(z)\ \nabla^{d}(z^{-1})\delta(z^{-1}) \qquad (3.18)$$

with λ equal zero and nonzero respectively. Equation (3.17) shows how to interpret the LQOC model inverse (as presented in section 3.1.3) and the tuning filter in the overall design.

When λ=0 the tuning filter is $F(z^{-1})$=1. When λ=∞ the tuning filter is $F(z^{-1})$=0. For nonzero λ, the tuning filter is of order one higher than the process. Increasing the tuning parameter gradually moves the poles of the tuning filter from the zeros of $\gamma_0(z^{-1})$ to the zeros of $\nabla(z^{-1})\delta(z^{-1})$.

3.2.3 The State Deadbeat Internal Model Controller

The design equations for the state deadbeat IMC design are given in section 3.1.5. In this design (Zafiriou and Morari (1985)) a first order exponential filter is used, where the filter constant is the tuning parameter. The combined model inverse and tuning filter is then:

$$G_C(z^{-1})\ F\ (z^{-1}) = \frac{\delta(z^{-1})}{\sum_i w_i}\ \frac{1 - \lambda}{1 - \lambda\ z^{-1}} \qquad (3.19)$$

When the tuning parameter λ=0 the tuning filter is $F(z^{-1})$=1. When the tuning parameter λ=1 the tuning filter is $F(z^{-1})$=0. As the tuning parameter λ is increased from 0 to 1 the filter is always first order with an increasing time constant. This controller design may require a higher order tuning filter when the process model is of high order as

described by Rivera et al (1986). In fact, the order of the tuning filter should be chosen equal to the order of the process.

The IMC design does not include a disturbance prediction filter or a set point prediction filter. Zafiriou and Morari (1985) mention that the design can be augmented, however, they do not present any insight or design equations as to how this can be done. Alternative derivations of the State Deadbeat controller are given by Jetto (1990) and Lee et al (1991).

3.2.4 The Extended Horizon Controller

The extended horizon controller of Ydstie et al (1985) combines an $n > b$ step ahead prediction filter with a model inverse which cancels the n step ahead prediction. This model inverse is not unique as there are many different sequences of the manipulated input which will cancel the n step ahead prediction. They present three alternative strategies for the controller design: a constant input over the horizon t to t+n-b, an input which changes only once over the horizon, and an input chosen to minimize control effort. They only calculate the control input U_t for the present time and implement it, recalculating the controls at the next control instant.

This design is focused on cancelling the t+n step ahead forecast of the error and does not consider the effect of the control actions on the t+n+1 step ahead forecast. Instead, the nonuniqueness of the possible input pattern over the horizon is resolved by postulating the three strategies described above. In this way a set of inputs U_t, U_{t+1}, ... , U_{t+n-b} is posited to cancel the t+n step ahead forecast of the

error. This design choice does not explicitly cancel the t+n+1 step ahead forecast, so when the control sequence U_{t+1}, U_{t+2}, ... , $U_{t+n+1-b}$ is calculated at time t+1, some of the future control inputs assumed at time t may be modified. As a result, the original t+n forecast of the error will no longer be cancelled! (Note however that if n=b then U_{t+1} has no effect on the process output at t+b, so the minimum variance type controller can always be calculated in this one step at a time manner.)

This deficiency of the Ydstie et al extended controller design is resolved in the next section where a new extended horizon controller design which focuses on cancelling not only the t+n forecast, but also the t+n+1 and other further forecasts of the error.

3.2.5 A New Extended Horizon Controller Design

The design of an extended horizon controller requires both a model inverse a tuning filter and a prediction filter which combine to cancel the forecast error over an extended horizon. The prediction filter for any particular horizon is easily designed as described in section 3.2.1 above. This section creates a new extended horizon controller with very favourable properties by extending the state deadbeat controller design with a new tuning filter design.

The ideas which embody the state deadbeat controller design were stated by Kalman (1954). Although he uses different equations, his explanations can be reformulated in the IMC structure. He notes that:

$$U_t = G_C(z^{-1})\ F(z^{-1})\ \ Y_{sp,t} \qquad (3.20)$$

provided that

$$G_C(1) = 1 / G_M(1) \quad , F(1) = 1 \quad \text{and} \quad G_M(z^{-1}) = G_p(z^{-1}) \qquad (3.21)$$

That is, $G_c(z^{-1})F(z^{-1})$ is the transfer function or the impulse response function relating the control actions to the set point. Kalman's explanation of the state deadbeat response notes that if $G_c(z^{-1})F(z^{-1})$ has no poles and $G_c(z^{-1})$ cancels the process $G_p(z^{-1})$ denominator, then the control action sequence U_t response to a step change in the error is a finite transient. Then, not only does the discrete process response to the control actions have a finite settling time, but so does the response of the underlying continuous process. Thus the elements of the state deadbeat design are that the approximate model inverse cancel the process denominator and both the approximate model inverse and the tuning filter must have no poles. This defines the state deadbeat design.

Kalman explains that with this design the size of the control interval is an implicit tuning parameter, as it is in all discrete controller designs. As the control interval is lengthened the control actions become less severe and the process response becomes slower, as discussed in section 3.1.6 above. In their rejoinder to Kalman's discussion Bergen and Ragazzini (1954) note the need for a modification to this state deadbeat design when the discrete process has a right half plane zero and that there is a need for a method of detuning the design when the control interval is fixed and is not available as a tuning constant. These deficiencies in the state deadbeat design have been addressed by Zafiriou and Morari (1985) where they propose that right

half plane zeros be retained as poles in the approximate model inverse $G_c(z^{-1})$, and that a first order exponential tuning filter $F(z^{-1})$ be used to detune the controller design.

In this section a different tuning filter is proposed based on the concept of an extended horizon controller. The state deadbeat controller is already an extended horizon controller relative to the minimum variance controller: it only fully cancels the process error after r+b control instants have passed, where r is the order of the process model numerator. Once $G_c(z^{-1})$ has been chosen to have no poles and to have a numerator which cancels the process dynamics (the state deadbeat model inverse) then the tuning filter $F(z^{-1})$ can be chosen arbitrarily (so long as it has unity gain), and the process settling time will remain finite so long as the filter has no poles. The selection of the filter $F(z^{-1})$ can be used to arbitrarily extend the process settling time.

This concept for an extended horizon controller using a tuning filter with no poles to increase the settling time of a state deadbeat controller is combined here with the observation from the previous section that the control actions of a state-deadbeat controller become less severe as the size of the control interval is increased. The fundamental concept used here is to choose the tuning filter to make the controller behave as if it had a longer control time.

As an example, consider the first order plus dead time process of equation 3.1 and its discrete equivalent in equation 3.2. The discrete time constant is $\delta_1 = e^{-Tc/Tp}$. If the control interval is

doubled, then the discrete time constant is $\delta_1^2 = e^{-2Tc/Tp}$. So, recalling equation (3.20) we can write:

$$U_t = G_C(z^{-1})\, F(z^{-1})\, E_t = \frac{1 - \delta_1 z^{-1}}{K_P (1 - \delta_1)} \frac{1 + \delta_1 z^{-1}}{(1 + \delta_1)} E_t = \frac{1 - \delta_1^2 z^{-2}}{K_P (1 - \delta_1^2)} E_t \qquad (3.22)$$

which provides a control impulse response to a step change in the error which is identical, in continuous time, to designing the state deadbeat controller with double the control interval, $2T_c$. This is most easily seen by noting that the numerator has no term in z^{-1}. However, this controller executes every T_c and has the disturbance rejection advantages of a short control interval. In unitary feedback form the controller for this example becomes:

$$C(z^{-1}) = \frac{F(z^{-1})\, G_C(z^{-1})}{1 - F(z^{-1})\, G_C(z^{-1})\, G_M(z^{-1})} = \frac{1 - \delta_1^2 z^{-2}}{K_P (1 - \delta_1^2) - z^{-b}(1 - \delta_1)(1 + \delta_1 z^{-1})} \qquad (3.23)$$

The design equation for this extended horizon controller is done for each process pole on a one by one basis. The design procedure requires that the process model be written in factored form:

$$G_p(z^{-1}) = \frac{w_0 + w_1 z^{-1} + \ldots + w_s z^{-s}}{1 + \delta_1 z^{-1} + \ldots + \delta_r z^{-r}} z^{-b} = \frac{w_0 + w_1 z^{-1} + \ldots + w_s z^{-s}}{(1 - D_1 z^{-1}) \ldots (1 - D_r z^{-1})} z^{-b} \qquad (3.24)$$

Then the design equation for the filter for each pole to hold the control action for n control instants is:

$$\frac{(1 - D_i^n z^{-n})}{(1 - D_i^n)} = \frac{(1 - D_i z^{-1})}{(1 - D_i)} F_i(z^{-1}) \tag{3.25}$$

so that

$$F_i(z^{-1}) = \frac{(1 + D_i z^{-1} + D_i^2 z^{-2} + D_i^3 z^{-3} + \ldots + D_i^{n-1} z^{-n+1})}{(1 + D_i^1 + D_i^2 + D_i^3 + \ldots + D_i^{n-1})} \tag{3.26}$$

The complete filter is then the product of the filters for each pole:

$$F(z^{-1}) = F_1(z^{-1}) F_2(z^{-1}) \ldots F_r(z^{-1}) \tag{3.27}$$

The horizon of the overall controller is extended by the order of $F(z^{-1})$. The tuning parameters for this controller are the integer number of control intervals for which the control action generated by each pole is extended. For each pole which has its control action extended by one interval, the overall horizon of the controller is extended by one.

The complete controller with this design is:

$$U_t = G_C(z^{-1}) F(z^{-1}) E_t = \frac{1}{K_P} \frac{(1 - D_1^n z^{-n})}{(1 - D_1^n)} \ldots \frac{(1 - D_r^m z^{-m})}{(1 - D_r^n)} E_t \tag{3.28}$$

where n and m are the horizons for the first and r^{th} poles respectively.

3.2.5.1 Extended Horizon Control With an Infinite Horizon: Closed Loop Controller Design For Open Loop Dynamic Response

A feature of the extended horizon controller is that as the controller horizon is extended the initial control action in response to

an error becomes less and less severe. This can be seen from the controller equation (3.28) which shows that the denominator terms for each pole increases to approach unity as the control horizon is increased (note that $|D_i| < 1$). When the control horizon is extended to infinity the control equation becomes:

$$U_t = G_C(z^{-1})\ F\ (z^{-1})\ \ E_t\ = -\frac{1}{K_p}\ \ E_t \qquad (3.29)$$

So, when the control horizon is infinity with this design the controller response to an error is to move the control action directly to the steady state value which cancels the error and then to hold it there. That is, the controller will eliminate a process error only at steady state. The meaning of cancelling an error at an infinite time in the future is to cancel it at steady state. If it were desired to further detune the extended horizon controller then the filter design would have to be augmented to allow a further reduction in the severity of the initial control action.

This extended horizon controller design with an infinite horizon allows the process to respond to an error or a set point change with its natural dynamics. An alternative derivation which arrives at this same controller design has been presented by Prasad et al (1990). Wong et al (1987) show that in the Dahlin (1968) controller design for a first order system choosing the closed loop time constant equal to the open loop time constant results in a maximum robustness controller design. This suggests a similar controller design, though the Dahlin controller is based on a minimum variance or deadbeat model inverse rather than a state deadbeat model inverse as used here.

3.2.6 The Modified Linear Quadratic Optimal Controller

The design equations for tuning the LQOC controller were introduced in section 3.2.2. In this section a new Modified LQOC design is proposed. The LQOC controller is still given by equation (3.17)

$$G_c(z^{-1})\ F\ (z^{-1}) = \frac{\delta(z^{-1})}{\gamma_0(z^{-1})}\ \frac{\gamma_0(z^{-1})\ \ \gamma(1)}{\gamma(z^{-1})\ \ \gamma_0(1)} \tag{3.17}$$

however $\gamma_0(z^{-1})$ and $\gamma(z^{-1})$ are obtained from the modified spectral factorization equation:

$$\gamma(z)\ \gamma(z^{-1}) = w(z)\ w(z^{-1}) + \lambda\ \delta(z)\ \delta(z^{-1}) \tag{3.18a}$$

with λ equal zero and nonzero respectively. The spectral factorization equation (3.18a) is calculated as if the disturbance were always stationary; that is, d=0 in equation (3.18). This is a different approach from Wilson (1970) and Harris and MacGregor (1987), but it appears to be the same formulation as presented by Kucera (1979). The controller will still have integral action as the gain of the tuning filter in equation (3.20) is still one. However, although this design uses a spectral factorization equation, it does not minimize the objective function of the Linear Quadratic Optimal Controller for nonstationary disturbances.

When λ=0 the tuning filter in this new design is $F(z^{-1})$=1. When λ=∞ the tuning filter is $F(z^{-1})$=$w(z^{-1})\delta(1)/\delta(z^{-1})w(1)$, so that $G_c(z^{-1})F(z^{-1})$=$\delta(1)/w(1)$=$1/K_p$, which is the same as the new extended horizon controller with an infinite horizon. For nonzero λ, the tuning filter is of equal order to the process. Increasing the tuning parameter gradually moves the poles of the tuning filter from the zeros

of $\gamma(z^{-1})$ to the zeros of $\delta(z^{-1})$. If the zeros of $\gamma(z^{-1})$ are near -1 and the zeros of $\delta(z^{-1})$ are in the right half plane, then the poles of the tuning filter will be in the left half plane for small values of λ and will be in the right half plane for large values of λ.

There are two favourable properties of this new Modified LQOC controller. Firstly, if the approximate model inverse in the controller has a left half plane zero then this zero is preserved for small values of the tuning parameter. As seen in the examples of the previous section this can allow for improved performance over the State Deadbeat IMC controller. Secondly, the order of the tuning filter is always matched to the order of the process.

3.2.7 Tuning of Controllers for a FOPDT System

In this section the different controller tuning methods are compared using the example of a first order process with gain K_p=1, time constant T_p=4 and dead-time T_d=0.4 (similar to section 3.1.6 above). The control interval is chosen as T_c=1 so that the process model becomes:

$$G_p(z^{-1}) = \frac{0.1393 + 0.0819\ z^{-1}}{1 - 0.7788\ z^{-1}}\ z^{-1} \qquad (3.30)$$

Where the controller designs require a disturbance model an ARIMA(0,1,0) random walk disturbance is assumed. This disturbance model is also appropriate for a controller designed for step changes in set point.

The LQOC controller for this process is designed using equations (3.17) and (3.18) to give:

$$G_C(z^{-1})\ F\ (z^{-1}) = \frac{1 - 0.7788\ z^{-1}}{0.1393 - 0.0819\ z^{-1}}\ \frac{0.1393 - 0.0819\ z^{-1}}{\gamma(z^{-1})}\ \frac{\gamma(1)}{0.0574} \quad (3.31)$$

where $\gamma(z^{-1})$ is calculated from the spectral factorization:

$$\gamma(z)\ \gamma(z^{-1}) = (.1393 - .0819\ z)\ (.1393 - .0819\ z^{-1}) + \lambda\ (1 - .7788\ z)\ (1 - z)\ (1 - z^{-1})\ (1 - .7788\ z^{-1}) \quad (3.31)$$

where λ is the tuning constant in the design. This equation reveals that $\gamma(z^{-1})$ is of degree two, so the tuning filter in equation (3.31) above will be second order.

The Modified LQOC controller is identical, except that $\gamma(z^{-1})$ is calculated form the spectral factorization:

$$\gamma(z)\ \gamma(z^{-1}) = (.1393 - .0819\ z)\ (.1393 - .0819\ z^{-1}) + \lambda\ (1 - .7788\ z)\ (1 - .7788\ z^{-1}) \quad (3.31)$$

where λ is the tuning constant in the design. $\gamma(z^{-1})$ is then of degree one. For λ near zero the zero of $\gamma(z^{-1})$ is near the zero of $w(z^{-1})$. As λ is increased the zero of $\gamma(z^{-1})$ moves to the right along the real axis towards the zero of $\delta(z^{-1})$. So, for some value of λ the zero of $\gamma(z^{-1})$ is at the origin which means $\gamma(z^{-1})$ is a constant; that is, the design produces a state deadbeat model inverse. Further increases in λ result in a right half plane zero in the filter denominator.

The State Deadbeat Internal Model Controller for this process is designed using equations (3.19) to give:

$$G_C(z^{-1})\ F\ (z^{-1}) = \frac{1 - 0.7788\ z^{-1}}{0.2212}\ \frac{1 - \lambda}{1 - \lambda\ z^{-1}} \quad (3.32)$$

where λ is the tuning constant in the design. This equation shows that the tuning filter is first order. Note that when $\lambda = 0.7788$ then

$G_C(z^{-1})F(z^{-1}) = 1 / K_p = 1$ which is identical to the extended horizon design with an infinite horizon (see equation 3.34 below).

The New Extended Horizon Controller for this process is designed using equation (3.28) to give:

$$G_C(z^{-1})\,F(z^{-1}) = \frac{1 - 0.7788\,z^{-1}}{0.2212} \; \frac{(1 + 0.7788\,z^{-1} + \ldots + 0.7788\,z^{-n})}{(1 - 0.7788^n)} \qquad (3.33)$$

where the integer n is the tuning constant in the design. This equation shows that the tuning filter is of zero order: there is no denominator in the filter, however it has an n^{th} order numerator. When the horizon is infinite (i.e. $n = \infty$) this controller becomes:

$$G_C(z^{-1})\,F(z^{-1}) = 1 \qquad (3.34)$$

Simulation results for a step change in set point from 0 to 1 at time 0 are shown these four controller designs in the following Figures. Figure 3.16 shows the process response and control actions using the LQOC controller with $\lambda = 0$, 0.1, 1.0 and 10. When $\lambda = 0$ the process shows a highly oscillatory response due to the inversion of the process zero (as described in section 3.1 above). However, as the tuning parameter is increased the process response becomes smoother, although it still shows some overshoot over the set point. The manipulated control action is very oscillatory when $\lambda = 0$, but it becomes much smoother as the tuning parameter is increased. However, a feature of the manipulated control action as λ is increased is that the initial control action is not the largest. When $\lambda = 0.1$ the control action peaks with the second control action, when $\lambda = 1.0$ the control action peaks with the fourth control action and when $\lambda = 10$ the control action

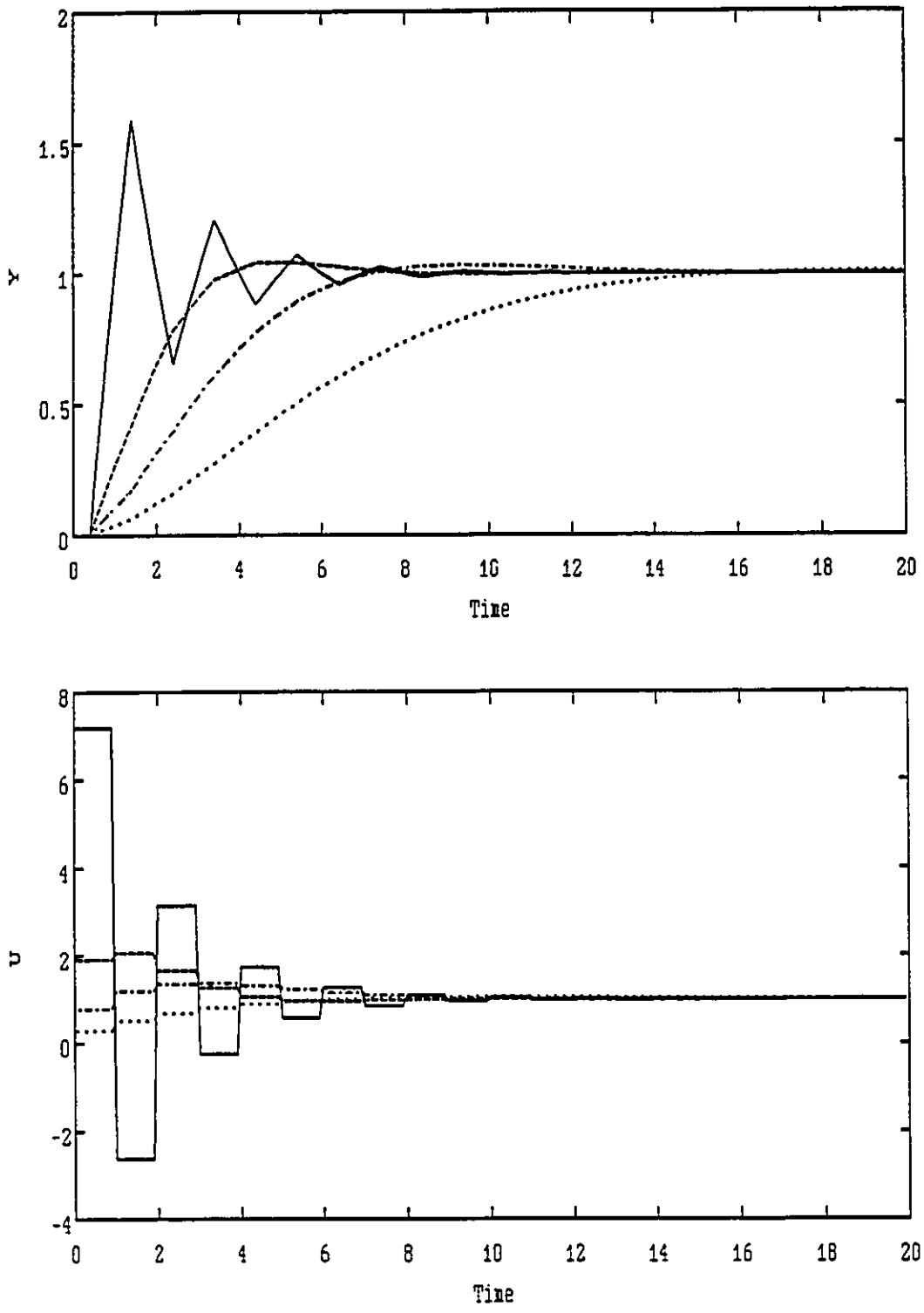


Figure 3.16 FOPDT Step Change in Set Point, LQOC Control, T_c=1
- λ=0, -- λ=0.1, -. λ=1, .. λ=10

peaks with the tenth control action. This increasing initial control action causes the overshoot which is seen in the process response.

Figure 3.17 shows the process response and control actions using the State Deadbeat IMC design with λ = 0, 0.3679, 0.7788 and 0.9048 (i.e. filter time constants of 0, 1, 4 and 10 time intervals). When λ = 0 and 0.3679 the process response rises monotonically to the set point and then levels out. When λ = 0.7788 the filter time constant equals the process time constants and the process responds with its open loop first order dynamics. When λ = 0.9048 the process response is even slower. The manipulated control action when λ = 0 has a large initial control action and then settles out at its steady state value. The oscillatory behaviour has been eliminated through the use of the state deadbeat model inverse. When λ = 0.3679 the initial control action is not so large and decays gradually to its steady state value. When λ = 0.7788 the control action moves immediately to its steady state value and stays there providing the step input which is required for the closed loop process to respond with the open loop process dynamics. When λ = 0.9048 the initial manipulated input is below the final steady state value and it climbs monotonically towards the steady state.

Figure 3.18 shows the process response and control actions using the Extended Horizon design with n = 0, 1, 4 and ∞. When n = 1 the process and manipulated control action responses are identical to the State Deadbeat IMC controller with λ = 0 as they both use the state deadbeat model inverse with no detuning. When n is increased to 1 and 4 the process response takes an additional 1 and 4 control intervals to reach the set point respectively and the initial manipulated control

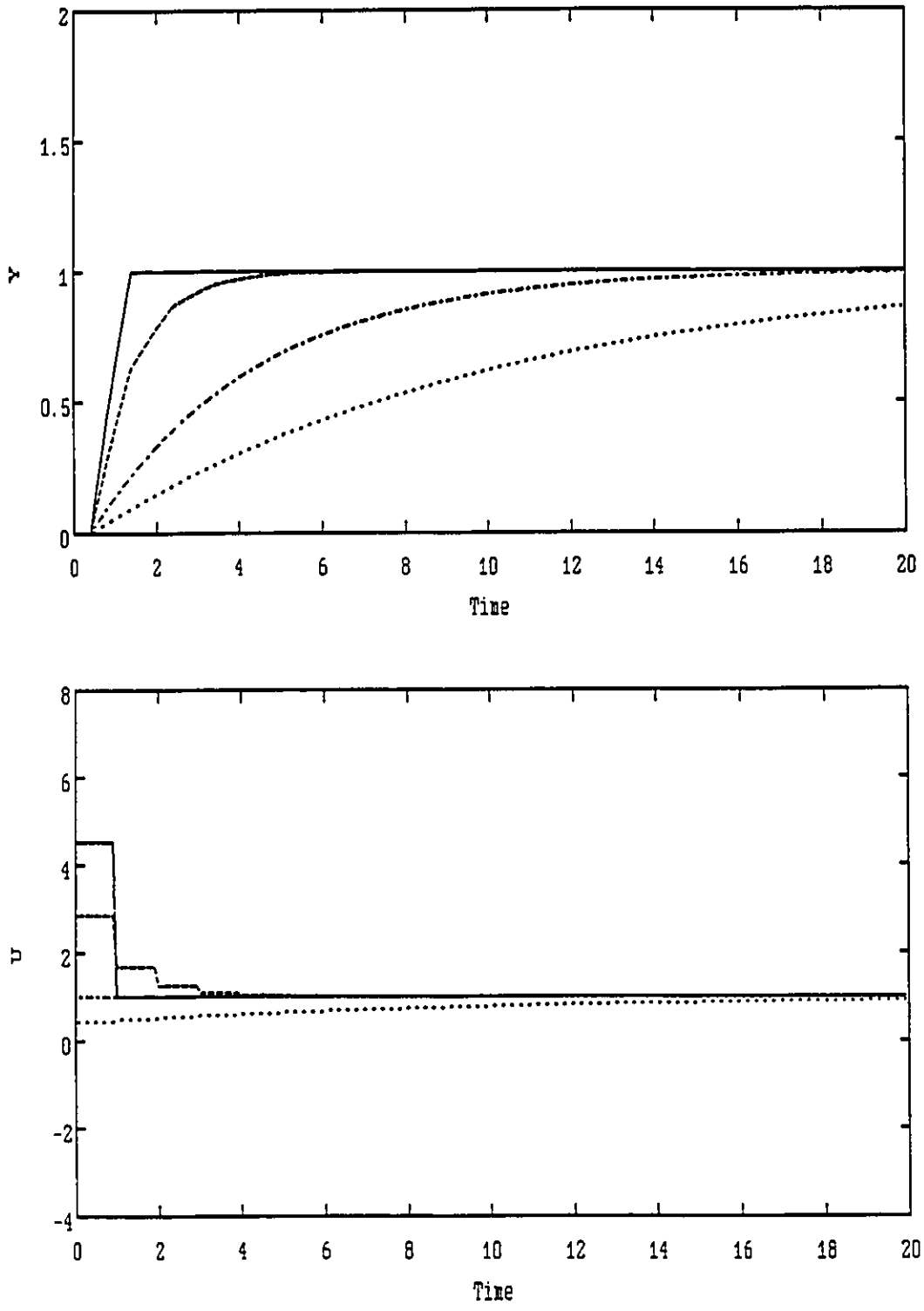


Figure 3.17 FOPDT Step Set Point Change, State Deadbeat IMC, T_c=1
- λ=0, -- λ=0.3679, -. λ=0.7788, .. λ=0.9048

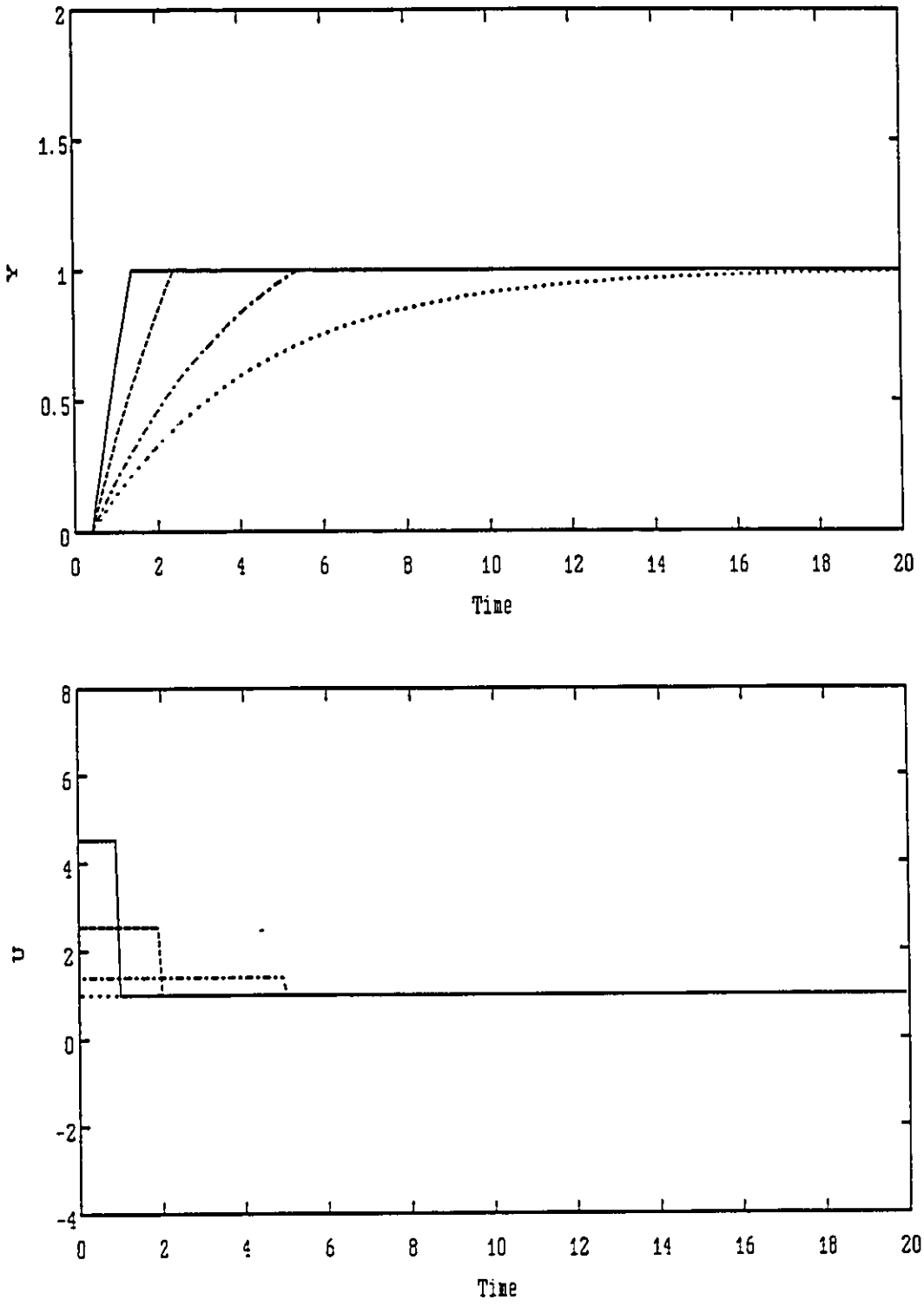


Figure 3.18 FOPDT Step Set Point Change, New Extended Horizon Control
- n=0, -- n=1, -. n=4, .. n=∞

action is reduced and is held for an additional 1 and 4 control time intervals respectively. So, although the control action is calculated every T_c the control response is as if T_c were 2 and 5 respectively. When $n = \infty$ the process responds with its open loop dynamics or as if T_c were ∞. This is the maximum value for the extended horizon tuning parameter and the manipulated input moves directly to its steady state value and holds there providing a step input to the process. This is identical to the State Deadbeat IMC controller with $\lambda = 0.7788$ where the filter time constant is equal to the process time constant. So, the range of tuning provided by the extended horizon controller design spans the same range as the State Deadbeat IMC design for values of λ between 0 and the process time constant. If further detuning is required then the extended horizon design must be augmented with a further tuning filter to reduce the initial control action.

Figure 3.19 shows the process response and control actions using the Modified LQOC design with $\lambda = 0$, 0.1, 1 and 10. When $\lambda = 0$ the process response and the control actions show the same oscillatory behaviour of the LQOC controller. When $\lambda = 0.1$ the process response shows a fast rise time and slight overshoot before settling at the set point. The control action is initially above the final steady state value, then below and then settles quite quickly. This fast process response with slight overshoot is not available for any tuning of the LQOC, the State Deadbeat IMC or the Extended Horizon controller. When $\lambda = 1$ and 10 the process and control action responses are the same as could be obtained with the State Deadbeat IMC controller. This is because the tuning parameter is large enough to pull the pole in the

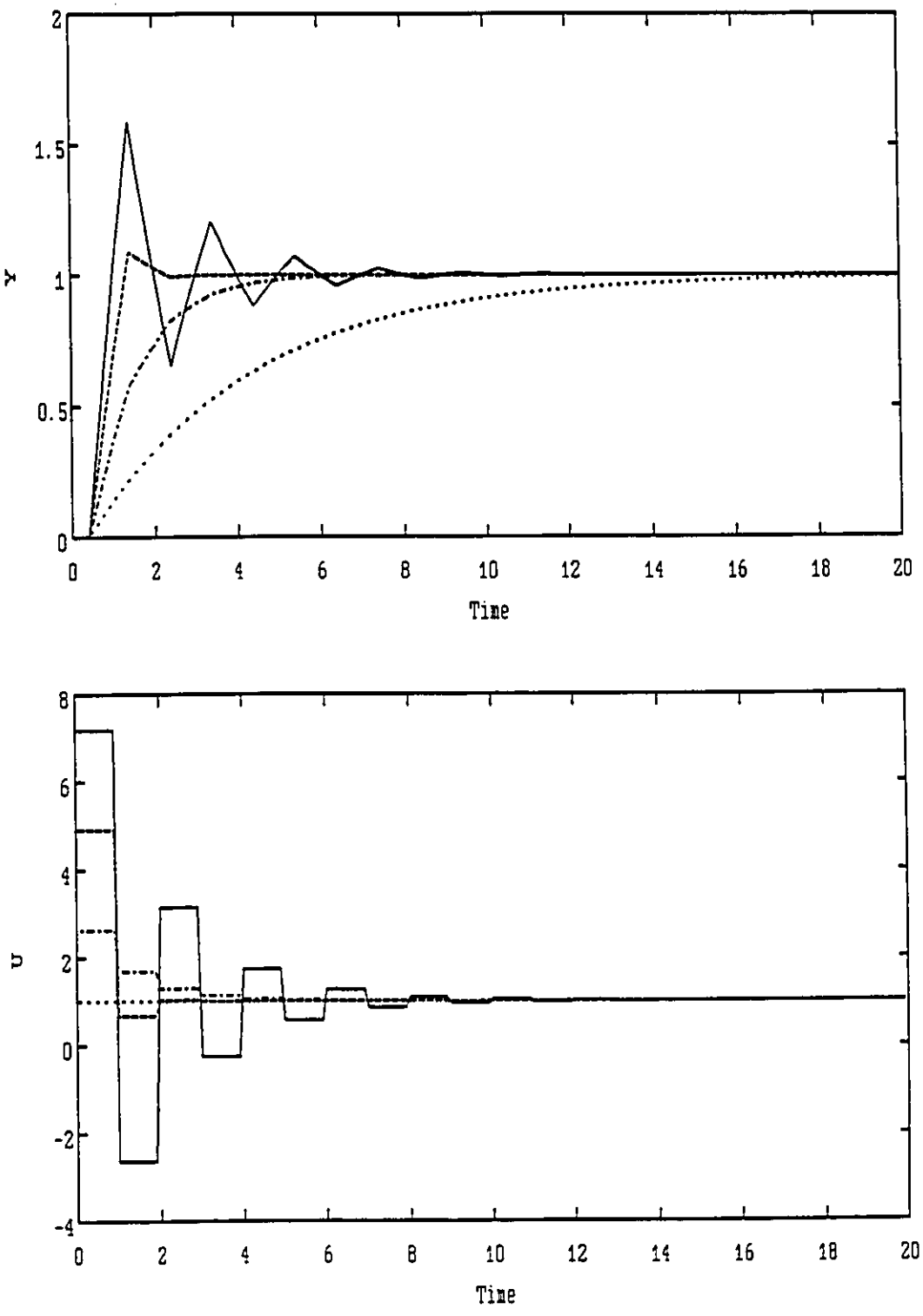


Figure 3.19 FOPDT Step Set Point Change, Modified LQOC Control, T_c=1 - λ=0, -- λ=0.1, -. λ=1, .. λ=10

tuning filter into the right half plane.

3.2.8 Summary

The control actions which are generated by a controller can be moderated by tuning. Different controller designs have different methods of tuning as well as different methods of model inversion. The LQOC controller design uses a model inverse which inverts the model zeros (or a stable reflection of them). Detuning then reduces the harshness of the control actions which can result from the inversion of the model zeros. The tuning filter in the LQOC design is of order one higher than the order of the process. For other than slight detuning, this results in sluggish control actions which approach the final steady state value from below.

The state deadbeat IMC design begins with a model which does not invert the process zeros. The tuning filter then shapes the control actions to reduce their severity and provide a slower process response. The new Extended Horizon controller design presented here also begins with a state deadbeat model inverse, but takes an alternate approach to the tuning filter. The filter is chosen to provide control actions as if the control time were longer. The Extended Horizon design with an infinite horizon provides the same control response as the state deadbeat IMC design when the filter is chosen to have the same poles as the process. In practice the control actions which result from these two different methods of detuning are very similar. The new Extended Horizon design is useful in understanding the relationship between the choice of the size of the control interval and of tuning on the shape of

the control actions. However, the State Deadbeat IMC controller design is much easier to implement.

The Modified LQOC controller uses the same model inverse as the LQOC controller, but the filter is of the same order as the process. In addition the poles of the filter are near the zeros of the process model (or their stable reflection through minus one) for small values of the tuning parameter. When the zeros of the process are in the left half plane this results in a ringing control action. As the tuning parameter is increased the poles of the filter move towards the poles of the process. When the poles of the filter move through the origin into the right half plane the control actions are of the same character as the Sate Deadbeat IMC controller. So, the Modified LQOC controller bridges the gap between LQOC control and State Deadbeat control: for some values of the tuning parameter it preserves a right half plane pole in the tuning filter but reduces the severity so that there is no ringing. This controller can thus take advantage of a left half plane zero in the process model to produce a controller with a faster rise time than the Sate Deadbeat IMC controller.

3.3 The frequency response of a discrete system coincident with a continuous system as the dead-time is varied

The discrete model which has its output coincident with a continuous process where the input is applied through a zero order hold is calculated using modified z transforms (Ragazzini and Franklin (1958), Jury (1958)). In this section, the frequency response of the discrete system is examined as the gain and dead-time of the continuous

process is varied. The parameters of the discrete model are calculated directly from the parameters of the continuous process. The effect of changes in the continuous process gain and dead-time on the discrete frequency response are for the robustness analysis in the next section. In the next section the discrete frequency response is used to evaluate stability of a closed loop system. It is the effect of changes in the continuous process parameters on the frequency response of the discrete model which determines the stability of the closed loop system.

The discrete model which is coincident with a continuous system

$$G_p(s) = \frac{e^{-T_d s}}{A(s)} \tag{3.35}$$

when the input is applied through a zero order hold is:

$$G_p(z^{-1}) = \frac{w(z^{-1})}{\delta(z^{-1})} z^{-b} = (1 - z^{-1}) Z_m \left[\frac{(1 - e^{-T_c s})}{s} \frac{e^{-T_d s}}{A(s)} \right] \tag{3.36}$$

where the coefficients of $w(z^{-1})$ and $\delta(z^{-1})$ are calculated using modified z transforms. Equation (3.35) assumes that the continuous process transfer function has no zero, which is typical of most chemical processes. In this case the order of the numerator of the discrete transfer function polynomial $w(z^{-1})$ is either equal to n-1, where n is the degree of A(s), when the dead-time T_d is an integer multiple of the control interval T_c; or equal to n when T_d is not an integer multiple of T_c.

The gain of the discrete equivalent process model, and the gain of the continuous process model are always equal by definition, so a change in the process gain results in an equivalent change in the gain

of the discrete process model. The effect of the time delay is twofold. Firstly the number of whole periods of delay in the discrete model is calculated from the time delay as

$$b = f + 1 \tag{3.37}$$

where

$$f = \text{int}(T_d / T_c)$$

The effect on the frequency response of adding b whole periods of dead-time is e^{-iwTb}, which contributes pure phase to the process transfer function. As in the continuous case, the amount of phase added is a linear function of the frequency w.

The effect of changes in the continuous process gain and changes in the dead-time which add or subtract integer multiples of the control time have straightforward and independent effects on the gain and phase of the discrete equivalent model respectively. The effect of a change in the process dead-time which is a fraction of a control period T_c is more complex. Through the modified Z transform, a fractional period of dead-time appears in the discrete model as an additional zero. A change in the dead-time which changes the fractional period of delay, moves the location of this zero in the discrete model numerator. The effect of this type of change will be examined separately for first and second order systems.

3.3.1 The effect of a changing fractional period of dead-time on a first order plus dead-time process

This section will show that the modified Z transform of a first order system with a fractional period of delay is a convex combination

of the Z transforms of the same first order plus dead-time process where the dead-time is an integer multiple of the sampling time. The Z transform of the first order system of equation (3.1) where $T_d/T_c = b$, an integer, is from equation (3.2):

$$G_p(z^{-1}) = \frac{K_p(1 - \delta_1)}{1 + \delta_1 z^{-1}} z^{-b} \tag{3.38}$$

The modified Z transform for the same system where the dead-time is a noninteger multiple of the sampling time, $\text{int}(T_d/T_c) + \text{frac}(T_d/T_c) = b + c$, is, by rearranging equation (3.2):

$$G_p(z^{-1}) = \frac{K_p(1 - \delta_1^{1-c})}{1 + \delta_1 z^{-1}} z^{-b} + \frac{K_p(\delta_1^{1-c} - \delta_1)}{1 + \delta_1 z^{-1}} z^{-(b+1)} \tag{3.39}$$

This can be rearranged as:

$$G_p(z\ \ ) = \frac{K_p(1 - \delta_1)\, z^{-b}}{1 + \delta_1 z^{-1}} \quad [\ X\ + (\ 1 - X\ )\ z^{-1}\ ] \tag{3.40}$$

where

$$X = \frac{1 - \delta_1^{1-c}}{1 - \delta_1}$$

Comparing this equation with equation (3.38) shows that as the dead-time is increased continuously from $T_d = bT_c$ to $(b+1)T_c$; that is, the fractional period of delay c is increased from 0 to 1, the modified Z transform transfer function is a convex combination of the Z transform of the process with b whole periods of dead-time and of the Z transform with b+1 periods of delay.

As an example, consider the first order plus dead-time process of equation (3.1), and its discrete equivalent transfer function in equation (3.2), with T_p=1, T_c=1, K_p=1 and the dead-time in the range T_d=0 to 1. Figure 3.20 shows the frequency response of the discrete transfer function when T_d=0, 0.2, 0.4, 0.6, 0.8 and when T_d=1. Also shown are the response at two particular discrete frequencies as the dead-time is increased continuously from 0 to 1. This shows that as the dead-time is continuously varied between two integer multiples of the sampling time, the frequency response connecting the frequency response for the two systems with no fractional dead-time traces a straight line.

In particular, at the discrete frequency $z=\pi$ the process has zero gain for some value of the fractional period of delay. This corresponds to a model inverse with a ringing pole, as described in section 3.1. Observation of the frequency response as the dead-time is varied adds to the understanding of the effect of inverting a zero in the process model when designing a controller. If the nominal process has a fractional period of delay, then the frequency response of the discrete model has a smaller gain at $z=\pi$ than the same process with no fractional period of delay. The discrete model inverse then has a larger gain at $z=\pi$ to offset this reduced gain. If the dead-time is changed in the continuous process for which this controller is designed then not only will the discrete phase change, but the discrete process gain will increase at frequencies near $z=\pi$. The implications of this frequency response behaviour on controller robustness is examined in section 3.4.4.

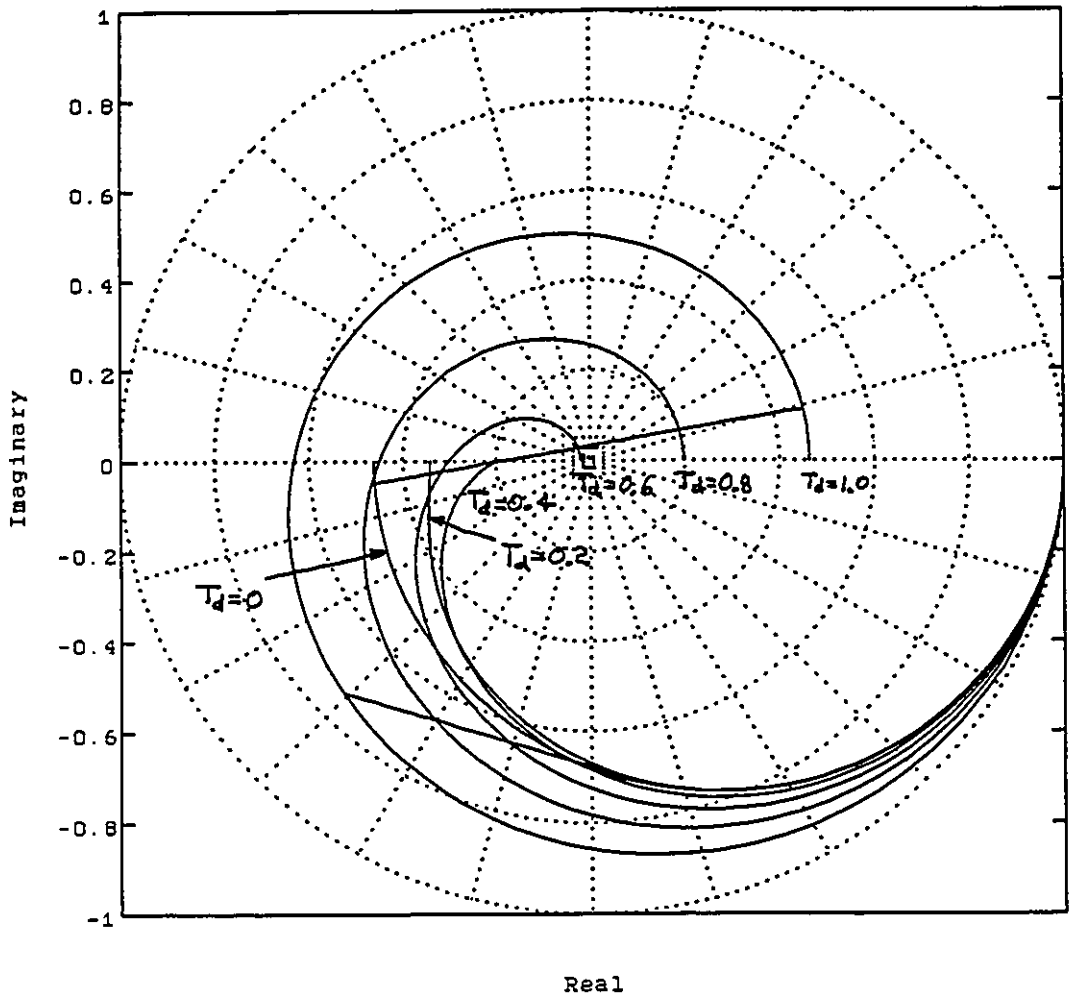


Figure 3.20 FOPDT Process Frequency Response
T_d = 0, 0.2, 0.4, 0.6, 0.8, 1

A continuous change in the dead-time in the continuous process is seen to affect both the gain and the phase of the discrete frequency response. As a result the method of construction for the robustness region presented in the last chapter can not be applied exactly to discrete control of continuous systems. However, it applies as a good approximation when the sampling time is small, because additional whole periods of delay add only pure phase, like dead-time in continuous time. It will also be shown in a later section that although one can not calculate the allowable dead-time variation directly from the phase margin, the approximation provided by the phase margin divided by the discrete frequency is still an easily calculated and useful measure of controller robustness.

3.3.2 The effect of a changing fractional period of dead-time on a second order plus dead-time process

This section shows that the straight line relationship between a changing fractional period of delay in a continuous first order process and the discrete transfer function frequency response does not apply in general. In particular this is demonstrated by using the counter example of a second order system.

The modified Z transform for the second order plus dead-time process of equation (3.3) is given in equation (3.4) above. The example, considered here has T_{p1}=1, T_{p2}=0.5, T_c=1, K_p=1 and the dead-time in the range T_d=0 to 1. Figure 3.21 shows the frequency response of the discrete transfer function when T_d=0, 0.2, 0.4, 0.6, 0.8 and when T_d=1. Also shown are the response at two particular discrete frequencies as the dead-time is increased continuously from 0 to 1.

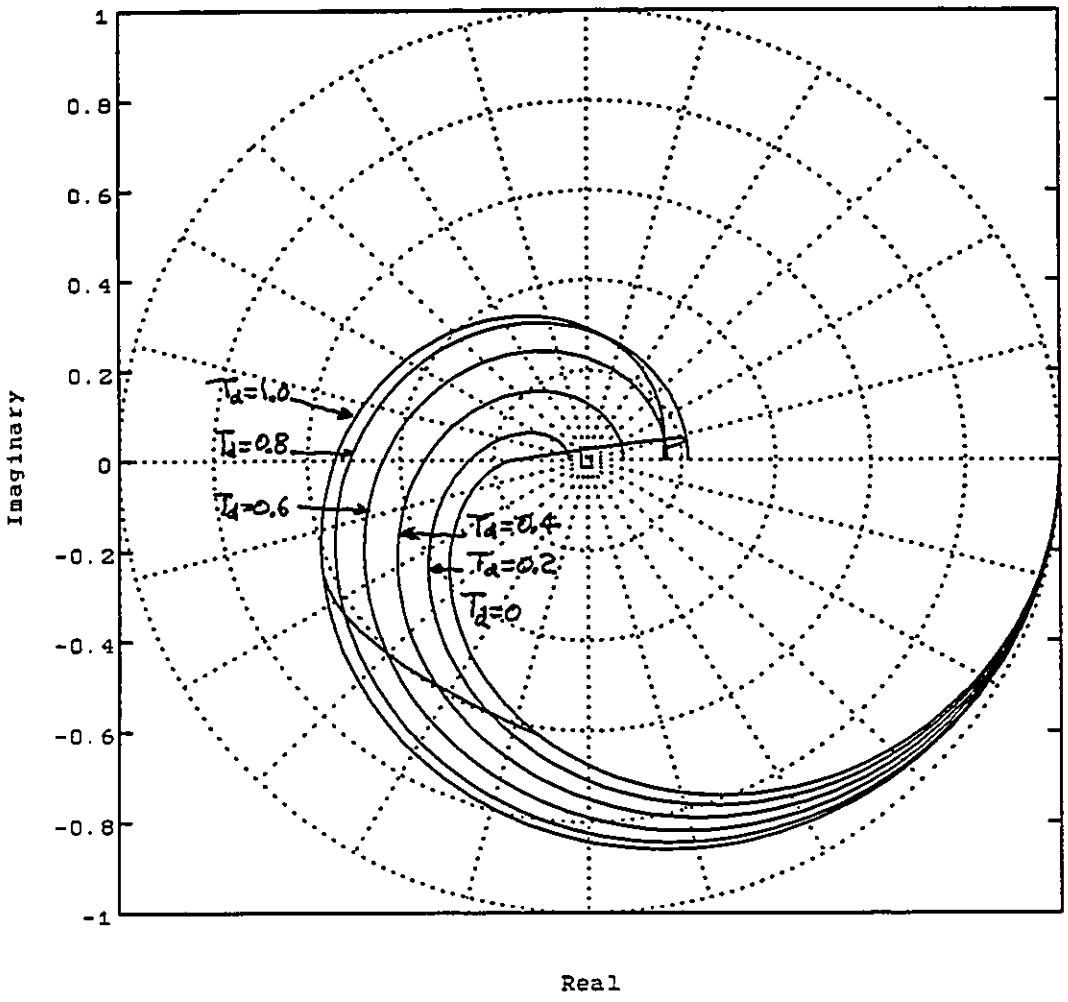


Figure 3.21 SOPDT Process Frequency Response
T_d = 0, 0.2, 0.4, 0.6, 0.8, 1

This shows that as the dead-time is continuously varied between two integer multiples of the sampling time, the frequency response does not move in a straight line between the frequency response for the two systems with no fractional dead-time. In fact, as the discrete frequency approaches π the gain of the frequency response has a maximum for some value of the dead-time between T_d=0.8 and T_d=1.

This observation has important implications for discrete controller design and robustness calculations. Firstly, that the gain of the frequency response may be larger for some fractional dead-time between two integer values of the dead-time needs to be considered when a robustness calculation uses a direct search method to find the range of dead-times for which a system is stable. Searching integer values of the dead-time first in order to prepare for a finer subsequent search may overestimate the system robustness.

3.3.3 Summary

This section has helped clarify the relationship between a continuous process with dead-time and the equivalent discrete process as the dead-time is varied. In particular it was proved that for a first order process when there is a fractional period of dead-time, the frequency response of the discrete process is a convex combination of the discrete process model with no fractional dead-time and of the discrete process model with no fractional dead-time and with one additional whole period of delay. This is believe to be a new result. For more general processes no such simple relationship exists. This was proved with the example of a second order system. In fact, for a second

order process when there is a fractional period of dead-time, the frequency response of the discrete process is not bounded by the frequency response of the discrete process model with no fractional dead-time and of the discrete process model with no fractional dead-time and with one additional whole period of dead-time. This observation is also believed to be new.

3.4 Robustness Analysis for Single-Input Single-Output Discrete Systems

In this section a new definition of robustness for discrete single-input single-output systems is presented. In the previous chapter a new measure of robustness for continuous closed loop systems was defined: the stability region of joint allowable variation in the process gain and the process dead-time. The robustness region was calculated directly from the Nyquist plot of the open loop system frequency response. In this section a similar definition of robustness is advanced for discrete systems: the stability region of joint allowable variation in the process gain and the process phase divided by the frequency. The process phase divided by the frequency is termed the continuous process quasi dead-time, as it approximates the allowable variation in the underlying continuous process dead-time. This robustness region is calculated directly from the Nyquist plot of the discrete open loop system frequency response.

The robustness analysis of this section is consistent with the renewed interest in frequency domain methods for discrete systems (Maciejowski and Samra (1990)). This is by contrast with the

measurement of robustness in terms of uncertainty limits on the system polynomial coefficients (Jury (1990)).

The power and the limitations of this new definition of robustness are shown by application to examples. The new robustness analysis calculation is used to show that the robustness to changes in the process dead-time of Minimum Variance control of a process with no dead-time for random walk disturbances (or step changes in set point) is exactly one control interval. That is, the smaller the size of the control interval the less robust the closed loop system is to changes in the process dead-time. This is believed to be a new result.

The limitation of the method is shown to be that when calculating the robustness of discrete control applied to a process with a fractional period of delay the allowable variation in the quasi dead-time can significantly over-estimate the allowable variation in the true dead-time; however, the relative results when comparing different controllers on the same process are meaningful. As in the continuous time definition of robustness, the quasi-dead time incorporates frequency information in a natural and meaningful way.

3.4.1 Construction of the Region of Joint Allowable Variation in the Process Gain and the Quasi Dead-Time of the Underlying Continuous Process

The discrete Nyquist stability criterion is directly equivalent to that for continuous systems: given a unitary feedback structure consisting of a controller transfer function $C(z^{-1})$ and a process transfer function $G_p(z^{-1})$, the closed loop discrete system is stable if the curve generated by the complex frequency response of the open loop transfer function $C(e^{-is})G_p(e^{-is})$ as s goes from 0 to $+\pi$ (excluding poles of $C(z^{-1})G_p(z^{-1})$ on this path) does not encircle the critical point -1+i0. This somewhat simplified statement of the full Nyquist criterion (Jury (1958)), which applies to stable processes, is sufficient for the development here.

The discrete frequency response can be thought of in a different way. In the description above the substitution $z=e^{is}$ is used. A more common alternative interpretation is to consider the frequency response $C(e^{-iwT_c})G_p(e^{-iwT_c})$ as the discrete frequency variable w goes from 0 to π/T_c (excluding the poles of $C(z^{-1})G_p(z^{-1})$ on this path). In this definition the substitution $z=e^{iwT_c}$ is used. This definition is more helpful as it introduces the concept of a discrete frequency, w. The discrete frequency, w, is limited to values between 0 and π/T_c, the Nyquist frequency. A plot of the discrete frequency response is called a Nyquist plot.

Using the open loop frequency response, two traditional measures of the stability of a process can be calculated: the gain margin and the phase margin. The gain margin is the maximum factor by which the process gain can be increased while maintaining system stability. The

phase margin is the maximum amount of phase which can be added to the process phase while maintaining system stability. The two main reservations with these measures of robustness which were raised in the previous chapter for continuous systems apply equally well to the discrete equivalents. The phase margin does not incorporate any frequency information: a relatively large phase margin translates into only a small allowable dead-time change in the continuous process being controlled if the phase margin occurs at a high frequency. The gain margin assumes that the process model is exact; however, a small change in the process dead-time can cause a large reduction in the gain margin.

In the previous section the relationship between a continuous process and its discrete time transfer function were explored in detail. A change in the continuous process gain results in an identical change in the gain of the discrete transfer function.

However, as seen in the previous section, changes in the continuous process dead-time have a more complex effect on the discrete equivalent transfer function than on the continuous one. Additional dead-time, δT_d, adds $\delta T_d w T_c$ of pure phase to the discrete frequency response (as is the case for the continuous frequency response) only when δT_d is an integer multiple of the control interval T_c. For the special case of a first order process, additional fractional periods of delay modify both the phase and gain so that the frequency response traces a straight line between the frequency response of the nominal process and of the process with an additional whole period of delay. For a general process, the frequency response of the nominal process and of the process with an additional whole period of delay do not even form

a boundary for the process with additional fractional periods of delay. That is, an additional fractional period of delay can add more phase to a discrete frequency response than does an additional whole period of delay.

Although there is no simple general relationship between phase margin and change in dead-time of the underlying continuous process, it is still true that an additional n whole periods of dead-time add pure phase of nwT_c: phase which increases proportionally with the number of whole periods of dead-time. So, dividing the phase margin by the discrete frequency provides a robustness measure. This robustness measure does not correspond exactly to the allowable change in the dead-time of the underlying continuous process but it does capture the fact that a large phase margin at a high frequency means less robustness than the same phase margin at a low frequency. The phase margin divided by the frequency will be called the allowable change in the quasi dead-time of the underlying continuous process.

Consider the open loop transfer function:

$$G(e^{-iwT_c}) = C\,(e^{-iwT_c})\,G_p(e^{-iwT_c}) \qquad (3.41)$$

The development from the previous chapter is repeated here: the allowable gain parameter variation δK_p is defined as the multiplicative change which will bring the closed loop system to the limit of instability. The size of the allowable gain parameter variation is identical to the gain margin: the increase or decrease in the process gain which will lead to an encirclement of the critical point -1+0i by the Nyquist plot. The allowable gain parameter variation is determined

from the Nyquist plot. Equation (3.42) defines the gain at the points where the Nyquist plot crosses the negative real axis.

$$R = \left| C\left(e^{-iwT_c}\right) G_p\left(e^{-iwT_c}\right) \right| \quad \text{whenever} \quad \angle\, C\left(e^{-iwT_c}\right) G_p\left(e^{-iwT_c}\right) = \pm\pi \tag{3.42}$$

Equation (3.43) separates these gain values into those less than 1, indexed by m, and those greater than 1, indexed by n.

$$R = R_m \quad \text{if} \quad R < 1, \qquad R = R_n \quad \text{if} \quad R > 1 \tag{3.43}$$

The positive allowable gain variation δK_p^+, defined in equation (3.43), is the smallest multiplicative gain perturbation larger than 1 which will bring the closed loop system to the limit of stability. The negative allowable gain variation δK_p^- is the largest multiplicative gain perturbation less than 1 which will bring the closed loop system to the limit of stability.

$$\delta K_p^+ = \min R_m^{-1} \qquad \text{and} \qquad \delta K_p^- = \max R_n^{-1} \tag{3.44}$$

The δK_p^+ and δK_p^- are both positive numbers; positive and negative are used here to indicate a gain increase and a gain decrease respectively. If either δK_p^+ or δK_p^- does not exist, then the gain can be increased or decreased indefinitely respectively.

The allowable gain variation δK_p is the minimum gain increase or decrease which will make the closed loop system unstable and does not imply that the system will be unstable for all larger or smaller gains respectively. For robustness analysis it is the first region of instability which is most important: other regions of stability beyond the first onset of instability are not considered here. Very few

systems have a value of δK_p^-, which means they can be made unstable by decreasing the process gain.

The allowable change in the quasi dead-time of the underlying continuous process, δT_{qd}, is calculated by dividing the phase margin by the discrete frequency. The phase margin is defined, as in the previous chapter, as the amount of phase which if added to or subtracted from the Nyquist plot of a closed loop stable system will bring the closed loop system to the verge of instability. Equation (3.45) defines the crossover frequencies w_{co}: the frequencies where the Nyquist Plot crosses the unit circle.

$$\text{Whenever} \quad \left| C\left(e^{-iwT_c}\right) G_p\left(e^{-iwT_c}\right) \right| = 1 \quad \text{then} \quad w_{co} = w \tag{3.45}$$

Equation (3.46) separates the crossovers into two types: positive crossovers where the gain is decreasing, indexed by n, and negative crossovers where the gain is increasing, indexed by m.

$$\text{if} \quad \frac{d\left| C\left(e^{-iwT_c}\right) G_p\left(e^{-iwT_c}\right) \right|}{dw} \begin{cases} < 0 & \text{then } w_{co}^{n+} = w_{co} \\ > 0 & \text{then } w_{co}^{m-} = w_{co} \end{cases} \tag{3.46}$$

Equation (3.47) defines the phase margin θ_{pm}; a positive crossover leads to a positive phase margin, indexed by n, and a negative crossover leads to a negative phase margin, indexed by m.

$$\theta_{pm}^{n+} = \pi + \angle C\left(e^{-iw_{co}^{n+}T_c}\right) G_p\left(e^{-iw_{co}^{n+}T_c}\right) \quad \text{or} \quad \theta_{pm}^{m-} = \angle C\left(e^{-iw_{co}^{m-}T_c}\right) G_p\left(e^{-iw_{co}^{m-}T_c}\right) - \pi \tag{3.47}$$

The phase operator $\angle$ is assumed to produce positive values in the upper half plane and negative values in the lower half plane with absolute

values less than or equal to π. The phase margin may be multiple valued with as many values as there are crossover frequencies.

The allowable quasi dead-time variation, δT_{qd}, is calculated from the phase margin:

$$\delta T_{qd}^{+} = \min_{\text{all } n} \left[\theta_{pm}^{n+} / w_{co}^{n+} \right] \quad \text{and} \quad \delta T_{qd}^{-} = \max_{\text{all } m} \left[\theta_{pm}^{m-} / w_{co}^{m-} \right] \qquad (3.48)$$

The allowable quasi dead-time variation has only two values: one positive and one negative. When there are multiple phase margin values, the smallest value will not necessarily determine δT_{qd} as the calculation depends inversely on the frequency of the crossover. If δT_{qd}^{+} does not exist, then it has the value $+\infty$. If δT_{qd} does not exist, then it has no defined value.

The quasi dead-time variation can not be interpreted as an exact measure of the allowable dead-time change in the continuous process because of the complicated relationship between the dead-time of the underlying continuous process and the frequency response of the equivalent discrete process. However, when the allowable quasi dead-time change has an integer value and the process being studied has no fractional period of dead-time then it does represent an accurate upper bound on the robustness of the system. When the quasi dead-time variation is not an integer value it doesn't have a direct interpretation in terms of the real dead-time of the underlying continuous process. However, despite these limitations, the quasi-dead time does provide a robustness measure which can be used to compare the relative robustness of different systems.

For example, in the previous section it was shown that the frequency response of a first order process with dead-time changes significantly as the amount of fractional dead-time is increased from 0 to 1. But, once a process model is chosen for analysis, the frequency response is fixed. The method described here can then be used to compare the robustness of different controller designs for this process.

The robustness of a feedback control system is now defined as the region of joint allowable gain and quasi dead-time variation. The procedure for constructing the region of joint allowable gain and quasi dead-time variation from the Nyquist plot is described here. The allowable multiplicative gain variation is first calculated from equations (3.42) to (3.44). The gain variation δK_p is stepped incrementally from δK_p^- (or 1 if δK_p^- does not exist) to δK_p^+, the limit beyond which the closed loop system is unstable. The allowable quasi dead-time variation δT_{qd}^- to δT_{qd}^+ will change as the gain is changed. This change occurs since altering of the gain by a factor of δK_p changes the location of the crossover frequencies. The calculations of equations (3.45) through (3.48) are repeated at each value of the gain to determine the allowable dead-time parameter limits δT_{qd}^- to δT_{qd}^+. This region can be plotted as δK_p vs δT_{qd}^- and δKP vs δT_{qd}^+ and represents the robustness of the control system.

3.4.2 Effect of the control interval on the robustness of discrete model based controllers applied to a continuous system with no dead-time

In this section the effect of the size of the control interval on the robustness of a closed loop system with a minimum variance controller design is examined. This analysis complements the examination of the control interval on closed loop system performance in the first section of this chapter. As an example consider a first order process with gain K_p=1 and time constant T_p=1 and no dead-time, as in section 3.1.6 above. For this process the minimum variance controller for an ARIMA(0,1,0) disturbance, the Linear Quadratic Optimal Controller with λ=0 and the state deadbeat IMC controller with a tuning filter $F(z^{-1})$=1 are all identical.

The control intervals which are chosen for this example are T_c=0.01, 0.1, 1.0 and 10.0. The Nyquist plots of the open loop frequency responses is shown in Figure 3.22. The Nyquist plots are identical for all four sampling times! The gain margin is R_m=2 and the phase margin is θ_{pm}=1.0 rad. The effect of the sampling interval is revealed when the frequency responses are shown in a Bode plot. In Figure 3.23 the logarithm of the open loop gain and the phase are plotted against the log of the discrete frequency w. The Bode plot shows that the crossover frequency w_{co} (at which the open loop gain is unity) is $1/T_c$. So, although the plots for the different control intervals all have the same shape, the shorter the control interval the higher the frequency at which a given gain occurs.

The robustness measures for the process are: the allowable gain variation δK_p=R_m=2.0 and the allowable quasi dead-time variation is

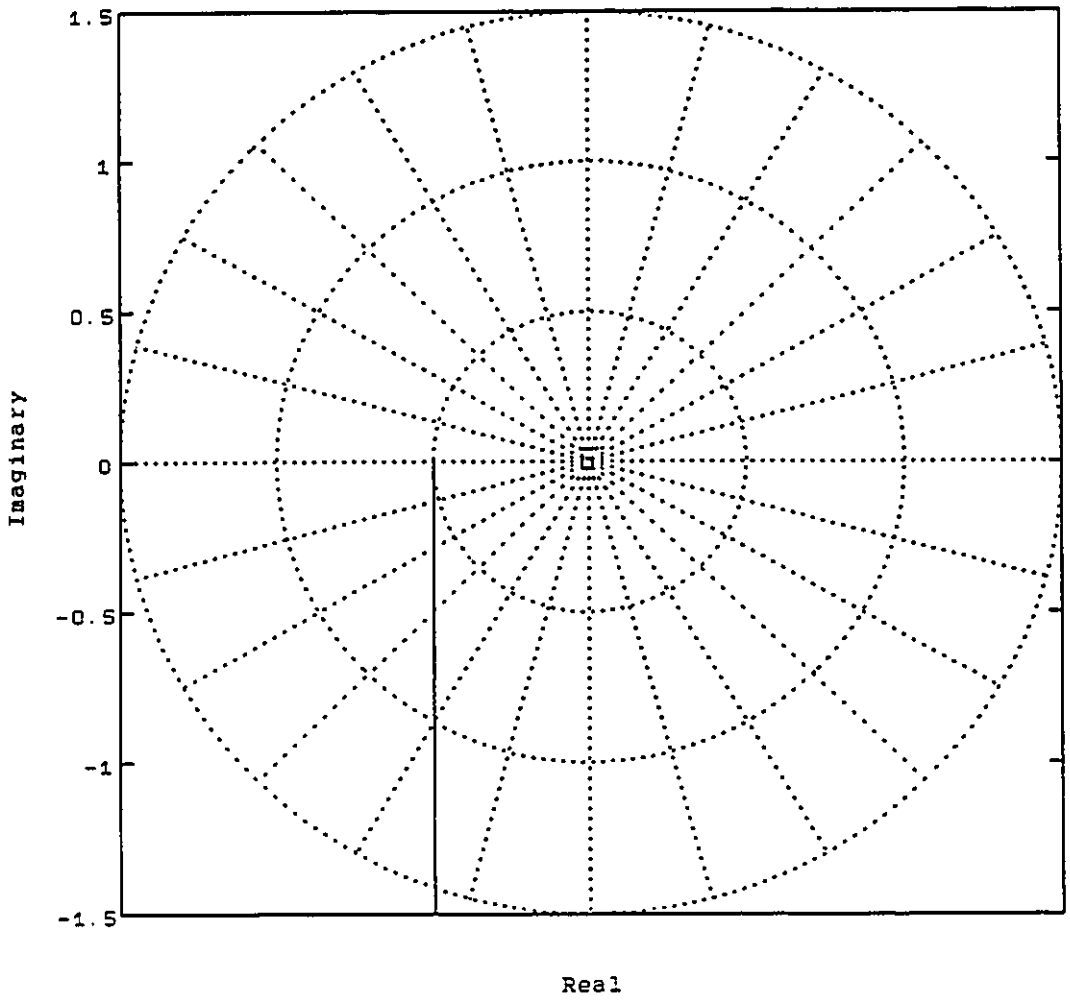


Figure 3.22 Nyquist Plot FO Process with MV Control
-- T_c=0.01, .. T_c=0.1, - T_c=1, -. T_c=10

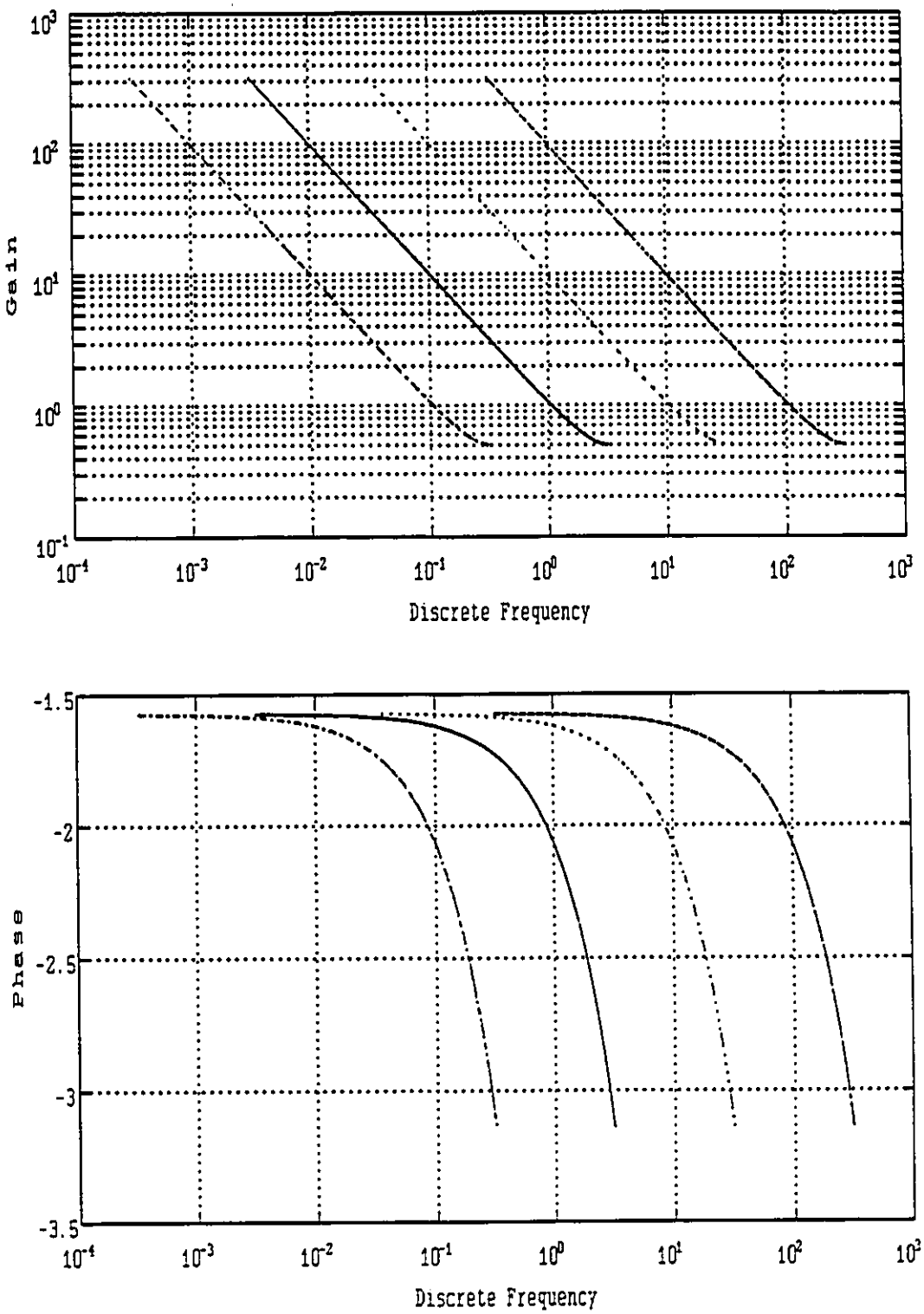


Figure 3.23 Bode Plot FO Process with MV Control
-- T_c=0.01, .. T_c=0.1, - T_c=1, -. T_c=10

$\delta T_{qd}=\theta_{pm}/(1/T_c)-T_c$ (because there is only one crossover the process has only positive values for this measure). Because the allowable quasi dead-time variation is an integer (i.e. 1) number of control intervals it represents an exact amount of dead-time which if added to the underlying continuous process will bring the closed-loop system to the limit of instability. So, the maximum amount of dead-time that can be added to the continuous process is one control interval T_c.

How general is this result? It applies to minimum variance control for ARIMA(0,1,0) disturbances (or step changes in set point) to all invertible processes having no dead-time. For minimum variance control the controller design is equation (3.5) where $\gamma(z^{-1}) = w(z^{-1})$. The controller combined with the process equation (3.4) give the open loop transfer function characteristic equation:

$$1 + C(z^{-1})\ G_p(z^{-1}) \tag{3.49}$$

where minimum variance controller for ARIMA(0,1,0) disturbances is:

$$C(z^{-1}) = \frac{\delta(z^{-1})}{(1 - z^{-f_c-1})\ w(z^{-1})} \tag{3.50}$$

and the process model is:

$$G_p(z^{-1}) = \frac{w(z^{-1})\ z^{-f_p-1}}{\delta(z^{-1})} \tag{3.51}$$

The process dead time in whole periods of delay has been written as f_p in the process model, i.e. the actual process dead-time, and as f_c in the controller. With these substitutions the open loop characteristic equation becomes:

$$1 + C(z^{-1})\, G_p(z^{-1}) = 1 + \frac{\delta(z^{-1})}{(1 - z^{-f_c-1})\, w(z^{-1})} \quad \frac{w(z^{-1})\, z^{-f_p-1}}{\delta(z^{-1})} \tag{3.52}$$

The stability of the open loop characteristic equation is assessed by calculating its roots. To this end the equation can be rearranged to give:

$$1 - z^{-f_c-1} + z^{-f_p-1} = 0 \tag{3.53}$$

So, when f_p=f_c then the characteristic equation has no roots and is unconditionally stable. When the process dead-time is one control time interval T_c larger than the nominal process model used to design the controller $f_p = f_c + 1$. In particular we are considering a process with no dead-time in the model so $f_c = 0$ and f_p -1 so the characteristic equation is:

$$1 - z^{-1} + z^{-2} = 0 \tag{3.54}$$

So, when the process has one control time interval of dead-time and the process model used in the minimum variance controller design has no dead-time, then the roots of the characteristic equation are at $z = 0.5 \pm 0.8667i$, and both of these roots are on the unit circle. That is, adding one additional period of dead-time to the process takes the closed loop system to the limit of stability.

This analysis does not guarantee that adding a fractional period of delay which is less than a whole period of delay to the process model will not make the closed loop system unstable. However, it does guarantee that adding a whole period of delay will take the process to the limit of instability.

This is a very useful result. The robustness of a minimum variance controller for a process with no dead-time is proportional to the size of the control interval. For example, when the sampling time is T_c=0.1 the allowable dead-time variation in the continuous process at most δT_d=0.1 and when the sampling time is T_c=1 then at most δT_d=1.

This result confirms the generally held view that robustness increases as the control interval is lengthened. It allows the designer to use his knowledge about the variability in the process gain and the process dead-time to choose the control interval size for minimum variance control. This example also shows that one need not always construct the entire region of joint allowable variation in the process gain and the process quasi dead-time. The allowable gain variation and the allowable process quasi dead-time variation calculated for the nominal model can be very useful.

3.4.3 Effect of the control interval on the robustness of discrete model based controllers applied to a continuous system with whole periods of dead-time

In this section the effect of the size of the control interval on the robustness of a closed loop system with a minimum variance controller design is examined for the example of a first order plus dead-time process with gain K_p=1 time constant T_p=1 and dead-time T_d=1, as in section 3.1.6 above. The control intervals which are chosen for this example are T_c=0.25, 0.3333, 0.5 and 1.0, so the number of whole periods of delay in the process model are f=4,3,2 and 1 respectively and the process model has no zero due to a fractional period of dead-time.

In this way we can explore the effect of increasing the number of whole periods of delay in different models of the same continuous process.

With these choices the minimum variance controller for ARIMA(0,1,0) disturbances, the Linear Quadratic Optimal Controller with $\lambda=0$ and the state deadbeat IMC controller with a tuning filter $F(z^{-1})=1$ are all identical. The Nyquist plots of the open loop frequency responses are shown in Figure 3.24. The Nyquist plots for all four sampling times overlap so that the differences are not clear from the figure. The gain margin is $R_m=2$ for all four curves. All four curves also have multiple values of the phase margin: $\theta_{pm}= \pm 1.0$ rad.

The Bode plot of the frequency response shows what happens as f increases from 1 to 4. In Figure 3.25 the logarithm of the open loop gain and the phase is plotted against the log of the discrete frequency w. Figure 3.25 shows that when $T_c=1.0$ and f=1 that there are two crossover frequencies and hence two values of the phase margin $\theta_{pm}=+1.0$ rad and $\theta_{pm}=-1.0$ rad. When $T_c=0.5$ and f=2 there are three crossover frequencies and hence three values of the phase margin $\theta_{pm}=+1.0$ and +1.0 rad and $\theta_{pm}=-1.0$ rad. That is, the Nyquist plot wraps an additional half wrap around the origin when f=2 than when f=1. When f=3 there are four crossover frequencies and hence four values of the phase margin $\theta_{pm}=+1.0$ and +1.0 rad and $\theta_{pm}=-1.0$ and -1.0 rad. When f=4 there are five crossover frequencies and hence five values of the phase margin $\theta_{pm}=+1.0$, +1.0 and +1.0 rad and $\theta_{pm}=-1.0$ and -1.0 rad.

So, for each additional whole period of delay the Nyquist plot wraps another half loop around the origin. Also, each successive crossover is at a higher frequency. On the Bode plot of Figure 3.25

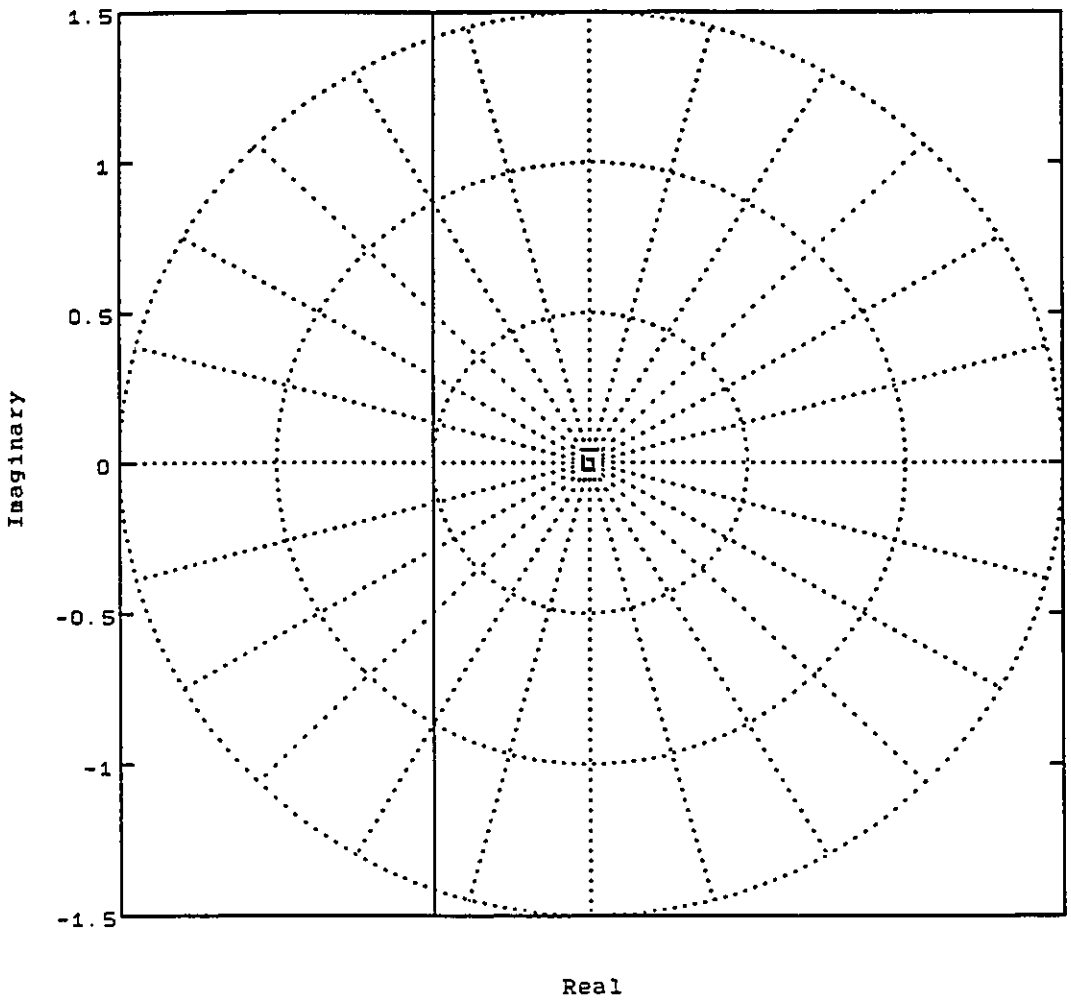
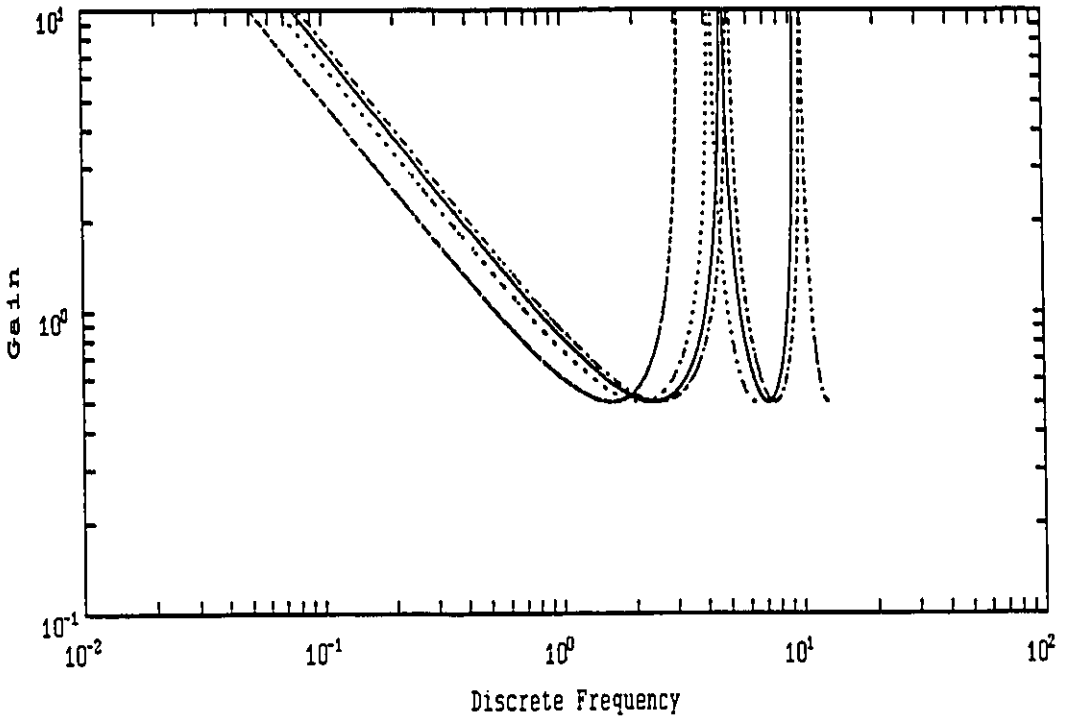


Figure 3.24 FOPDT Process with MV Control
-- T_c=1, .. T_c=0.5, - T_c=0.333, -. T_c=0.25



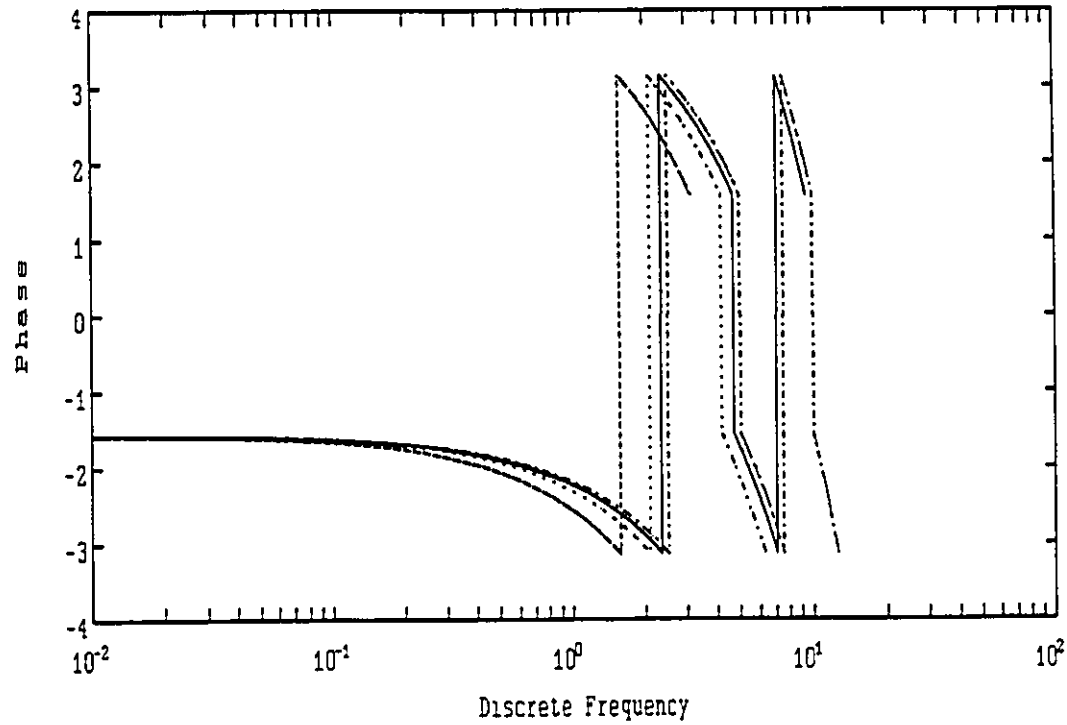


Figure 3.25 FOPDT Process with MV Control
-- T_c=1, .. T_c=0.5, - T_c=0.5, T_c=0.25

when f is odd the plot ends at the Nyquist frequency with a peak. When f is even the Bode plot ends at the Nyquist frequency with a valley. Each additional two periods of delay add an additional peak to the Bode plot.

The quasi dead-time variations which are calculated when f=1,2,3 and 4 are δT_{qd}^{-} = -0.4, -0.3, -0.1212 and -0.1136 respectively and δT_{qd}^{+} = 2.0, 0.2143, 0.1905 and 0.0962 respectively. It is interesting to note that when f is odd $|\delta T_{qd}^{-}| < |\delta T_{qd}^{+}|$; that is, the system stability is more sensitive to a reduction in the quasi dead-time than to an increase. This is because when f is odd the highest frequency crossover is one where the open loop gain is increasing: the Bode gain plot in Figure 3.25 ends at the Nyquist frequency with a peak rather than a valley.

If the control time T_c were reduced towards T_c=0 while always selecting values of T_c such that T_d/T_c is an integer, then the Bode plot would have successively more peaks. In the limit the Nyquist frequency approaches infinity and the Bode plot has an infinite number of peaks, like the Bode plot of continuous LQG control with λ=0 as shown in the previous chapter.

It is interesting to note in this example that the minimum of $|\delta T_{qd}^{-}|$ and $|\delta T_{qd}^{+}|$ is less than T_c. So, minimum variance control of a first order process with dead-time is less robust than minimum variance control of the same process with no dead-time.

3.4.4 Robustness of discrete model based controllers applied to a continuous system with a fractional period of dead-time

In this section the robustness of a closed loop system with a Minimum Variance controller and a State Deadbeat IMC controller is examined for the example of a first order plus dead-time process with gain K_p=1 time constant T_p=1 and dead-time T_d=1, as in section 3.1.6 above. The control interval which is chosen for this example is T_c=1.5, so the number of whole periods of delay is f=0 and the process model is given in equation (3.10). In this way we can compare the robustness of the minimum variance controller for ARIMA(0,1,0) disturbances with the state deadbeat IMC controller with a tuning filter $F(z^{-1})$=1. The process is invertible but it does have a zero so these two controllers are not identical.

The Nyquist plots of the open loop frequency responses are shown in Figure 3.26. The Bode gain and phase plot is shown in Figure 3.27. The Bode gain plot of the system with the state deadbeat controller falls of dramatically as the frequency approaches the Nyquist Frequency. This is because the zero in the process leads to a large gain reduction in the process frequency response near the Nyquist Frequency, as shown in Section 3.3.1 above. The State Deadbeat IMC controller does not compensate for this gain reduction and hence the open loop system gain falls off. The Minimum Variance controller includes the zero of the process as a pole in the approximate model inverse so that the open loop frequency response looks the same as for the process with no dead-time; compare Figures 3.26 and 3.27 with Figures 3.22 and 3.23 in Section 3.4.2 above.

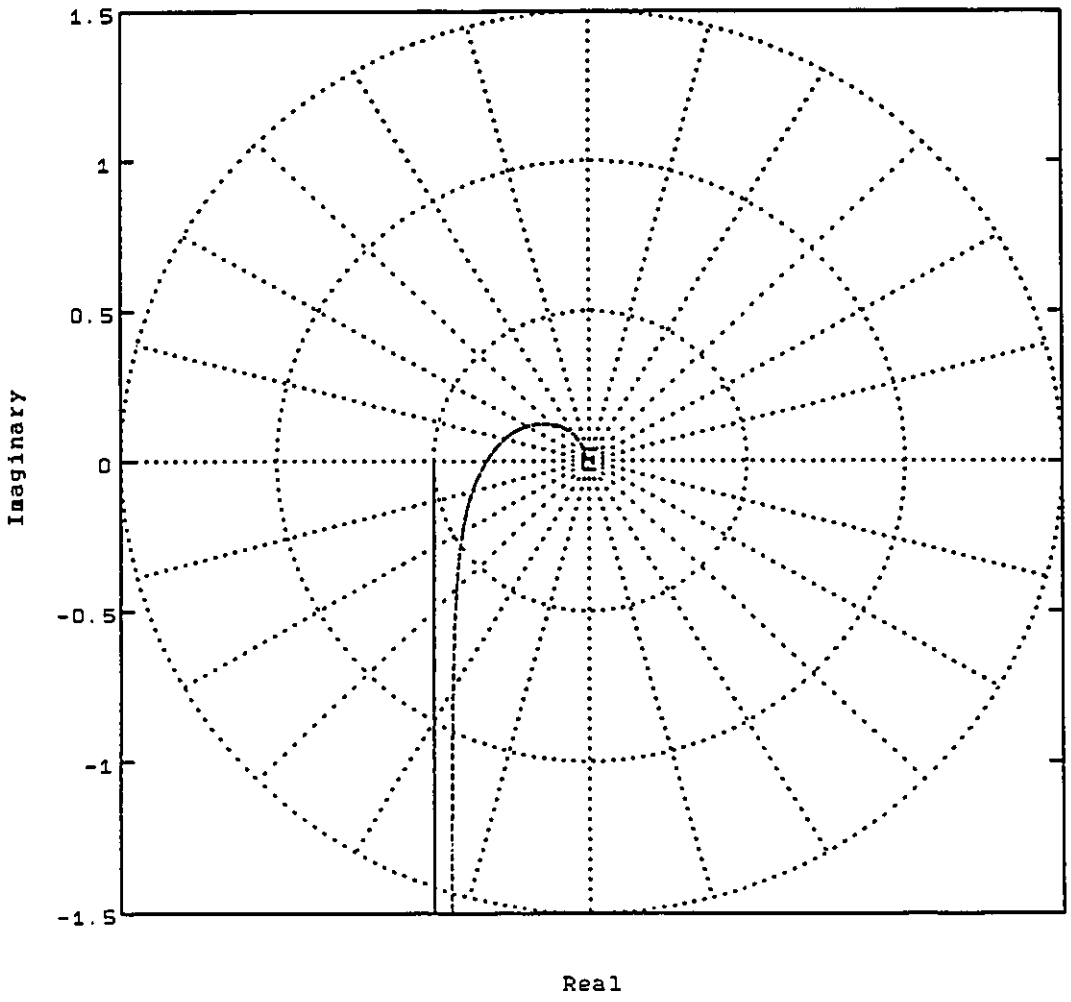


Figure 3.26 Nyquist Plot FO Process with a Fractional Period Dead-Time
- MV, -- State Deadbeat IMC

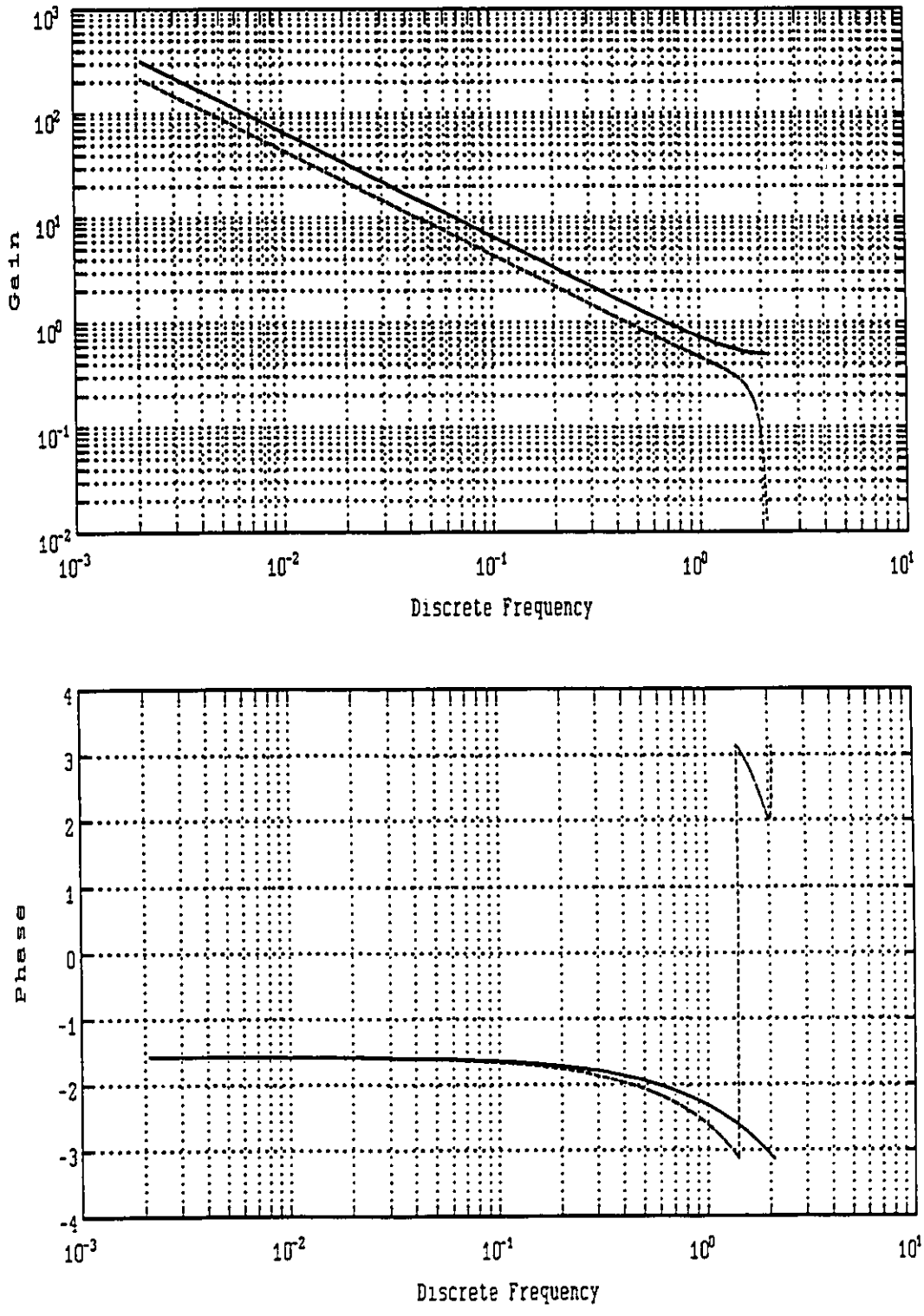


Figure 3.27 Bode Plot FO Process with a Fractional Period of Dead-Time - MV, -- State Deadbeat IMC

The Bode plot shows that the open loop system with the state deadbeat controller has a lower gain at all frequencies than the open loop system with the minimum variance controller which translates into a higher allowable gain variation δK_p=3.02 vs 2.47. The system with the state deadbeat controller has only slightly more phase margin than the minimum variance controller, but the crossover frequency is lower so that this translates into both a higher allowable quasi dead-time variation δT_{qd}=2.4734 vs 1.5. Because the process model has a fractional period of delay this quasi dead-time variation does not translate into an equivalent change in the continuous process dead-time. However, the relative numbers are meaningful: the system with the state deadbeat controller is more robust than the one with the minimum variance controller.

A search method was used to determine the increase in the dead-time of the continuous process which will bring the closed loop system to the limit of instability for this example. The minimum variance controller can tolerate an increase in the process dead-time from 1.0 to 1.5 or δT_d^+=0.5. The state deadbeat controller can tolerate an increase in the process dead-time from 1.0 to 2.49 or δT_d^+=1.5=T_c. This confirms that the state dead-beat controller is more robust. In fact, it has almost as much robustness as the state dead-beat controller for the undelayed process where it was shown that δT_{qd}=T_c (T_c=1.5 in this example).

The minimum variance controller for this process was shown in section 3.1.6 to have a very high ISE because of the ringing nature of the controller output. This large high frequency gain in the controller

is seen here to cause robustness problems as well. Figure 3.20 in Section 3.3.1 shows that the process gain for a first order system with a fractional period of delay can be much lower at the Nyquist frequency than the frequency response of the same system with no fractional period of delay. The minimum variance controller design compensates for this reduced process gain with a large controller gain whereas the state deadbeat controller does not. If the process dead-time is changed from its nominal value then the Minimum Variance controller is very sensitive to the rapidly increasing process gain. The State Deadbeat IMC controller design uses an approximate model inverse which is based on the process with no fractional period of delay. As a result, the state dead-beat controller is more robust than the minimum variance controller for this process with a fractional period of delay.

3.4.5 Summary

This section has presented a new definition for the robustness of a discrete closed loop system: the region of joint allowable variation in the process gain and the quasi dead-time. Given any stable discrete process this method can be used to compare the robustness of different controllers. When the allowable variation in the quasi dead-time is an integer number of control intervals and the process has no fractional period of dead-time then it provides an exact upper limit on the allowable dead-time variation in the underlying continuous process. When the allowable change in the quasi dead-time does not have a direct interpretation in terms of the dead-time of the underlying continuous

process it is nonetheless useful for comparing the robustness of different controllers.

The method was applied to several examples. It was shown that when using minimum variance control for a continuous invertible process with no dead-time and a random walk disturbance then the maximum allowable dead-time increase in the process is one control interval. So, robustness of minimum variance control decreases with the size of the control interval. This example shows that it is not always necessary to plot the entire region of joint allowable variation to get useful robustness results: often the allowable gain variation and allowable quasi-dead time variation alone provide a good measure of robustness.

It was also shown that Minimum Variance control (equivalent to State Deadbeat IMC control) of a first order process with whole periods of dead-time is less robust than control of the same process with no dead-time. The dead-time compensation in the discrete controller produces peaks in the discrete frequency response which adversely affect the system robustness.

Lastly it was shown that for a first order process with a fractional period of delay the use of Minimum Variance control results in a system with less robustness than when the State Deadbeat IMC controller is used. The State Deadbeat IMC controller does not include the process zero as a pole in the approximate model as does the Minimum Variance controller. As a result the State Deadbeat IMC controller has a much lower high frequency gain and is less sensitive to the increase

in the process gain as the dead-time is moved from its fractional value towards an integer value.

This new definition of robustness provides a tool for robustness analysis of discrete controllers applied to the control of continuous processes. Because the robustness region is calculated from the open loop frequency response it allows the robustness implications of features of the frequency response to be interpreted. An understanding of the relationship between changes in the continuous process and the discrete frequency response, as presented in section 3.3, is required to properly interpret the results. Other presentations of discrete controller robustness using frequency response methods, such as Morari and Zafiriou (1989), do not explore any of these relationships. Because of the limitations on interpreting the allowable dead-time change in the continuous process from the allowable change in the quasi dead-time this robustness analysis does not replace parameter search methods such as those used by Bergh (1986).

3.5 Combined performance and robustness analysis for controller tuning

This section examines the tradeoff between robustness, performance, tuning and the size of control interval. In Section 3.1 it was shown that for a State Deadbeat IMC controller, performance (as measured by the ISE of the response to a step change in set point) improves as the control interval is reduced. However, Section 3.4 revealed that this performance is obtained at the price of reduced robustness. In Section 3.2 the relationship between tuning and the size

of the control interval was explained using a new Extended Horizon controller design. The size of the control interval and the tuning parameter in a controller design method both affect the tradeoff between robustness and performance. In addition this section shows that a controller design can have good robustness and performance characteristics and yet require unacceptable control actions. Lennartson (1990) has compared the continuous time variance of several controller design methods for processes without dead-time subject to a stationary disturbance. In that comparison the different controller designs were tuned for equal control action variance. The study here compares different controllers tuned for the same level of robustness.

The examples in this section explore the role of the size of the control interval and controller tuning in the tradeoff between robustness and performance using the Linear Quadratic Optimal Control, Modified LQOC, Internal Model Control and State Deadbeat IMC controller designs. These different controller are applied to examples of a first order and a second order process. The results show that the IMC and State Deadbeat IMC controllers have the best tradeoff between performance and robustness, however they can both produce totally unrealizable control actions. The State Deadbeat IMC controller with a tuning filter order chosen to equal the order of the process and the Modified LQOC controller both produce acceptable control actions and have a similar robustness vs performance tradeoff. The LQOC controller has the poorest tradeoff between performance and robustness among the controller designs which are compared.

3.5.1 Performance and robustness of discrete model based control of a first order process for step changes in set point

In this section the example from Section 3.1.6 of a first order process with gain K_p=1 and time constant T_p=1 with no dead-time is again investigated. Recall that for this example, with an ARIMA(0,1,0) disturbance, the minimum variance, Linear Quadratic Optimal Controller with λ=0, IMC with $F(z^{-1})$=1 and the state deadbeat IMC controller with $F(z^{-1})$=1 are all identical for any choice of the size of the control interval. As shown in the previous section the robustness of minimum variance (and equivalent) control applied to this process decreases with the size of the control interval. This section focuses on the LQOC, Modified LQOC, IMC and State Deadbeat IMC controller designs as they can be detuned.

In the previous section it was shown that for this example the allowable quasi dead-time variation of the system is a good measure of robustness. It was also shown that for the minimum variance controller the allowable quasi dead-time variation is $\delta T_{qd}=T_c$, the size of the control interval. In particular the control interval T_c=1 is chosen as a basis for comparison. So, δT_{qd}=1 will be used for comparison.

In this section the tuning parameter of the LQOC controller, described in Section 3.2.2, and the tuning parameter of the State Deadbeat IMC controller first order filter, described in Section 3.2.3, are used to detune the controllers. The IMC controller and the Modified LQOC controller, described in section 3.2.6 produces the same controller design as the Sate Deadbeat IMC controller for a first order process with no dead-time. As in Section 3.2 both parameters are referred to as λ for simplicity. For a range of control times from T_c=0.01 to 5 these

two controllers are tuned so that they provide the same allowable dead-time variation δT_{qd}=1 as minimum variance control with T_c=1. The appropriate value of the tuning parameter is found numerically using a search method. The ISE performance of the continuous system is then calculated as in Section 3.1.

Figure 3.28 shows the ISE plotted vs the size of the control interval for LQOC and State Deadbeat IMC control where the tuning parameter λ is chosen so that $\delta T_{qd} \geq 1$. For all values of $T_c \geq 1$, λ=0 because the system is more robust than when T_c=1. When T_c=0.8, 0.5, 0.2, 0.1 and 0.01 the values of λ=0.0533, 0.2433, 1.7769, 7.226 and 726.5919 respectively. Figure 3.28 shows that as the size of the interval is reduced below T_c=1 the ISE performance of the closed loop system begins to increase. When the control interval is T_c=0.01 the ISE=0.56, a higher value than when T_c=5.

Figure 3.29 shows the allowable positive quasi dead-time variation, δT_{qd}, plotted vs the size of the control interval, T_c. This figure shows graphically how the robustness of the system decreases linearly with the size of the control interval until $T_c \leq 1$ where the controller is detuned so that δT_{qd}=1.

Figure 3.28 shows the ISE plotted vs the size of the control interval for IMC control where the tuning parameter λ is chosen so that $\delta T_{qd} \geq 1$. For all values of $T_c \geq 1$, λ=0 because the system is more robust than when T_c=1. When T_c=0.8, 0.5, 0.2, 0.1 and 0.01 the values of λ=0.1322, 0.3819, 0.7153, 0.8504 and 0.9844 respectively. (The Modified LQOC controller produces identical results with the values of λ=0.1344, 0.3417, 0.5473, 0.6155 and 0.6745). Figure 3.28 shows that as the size

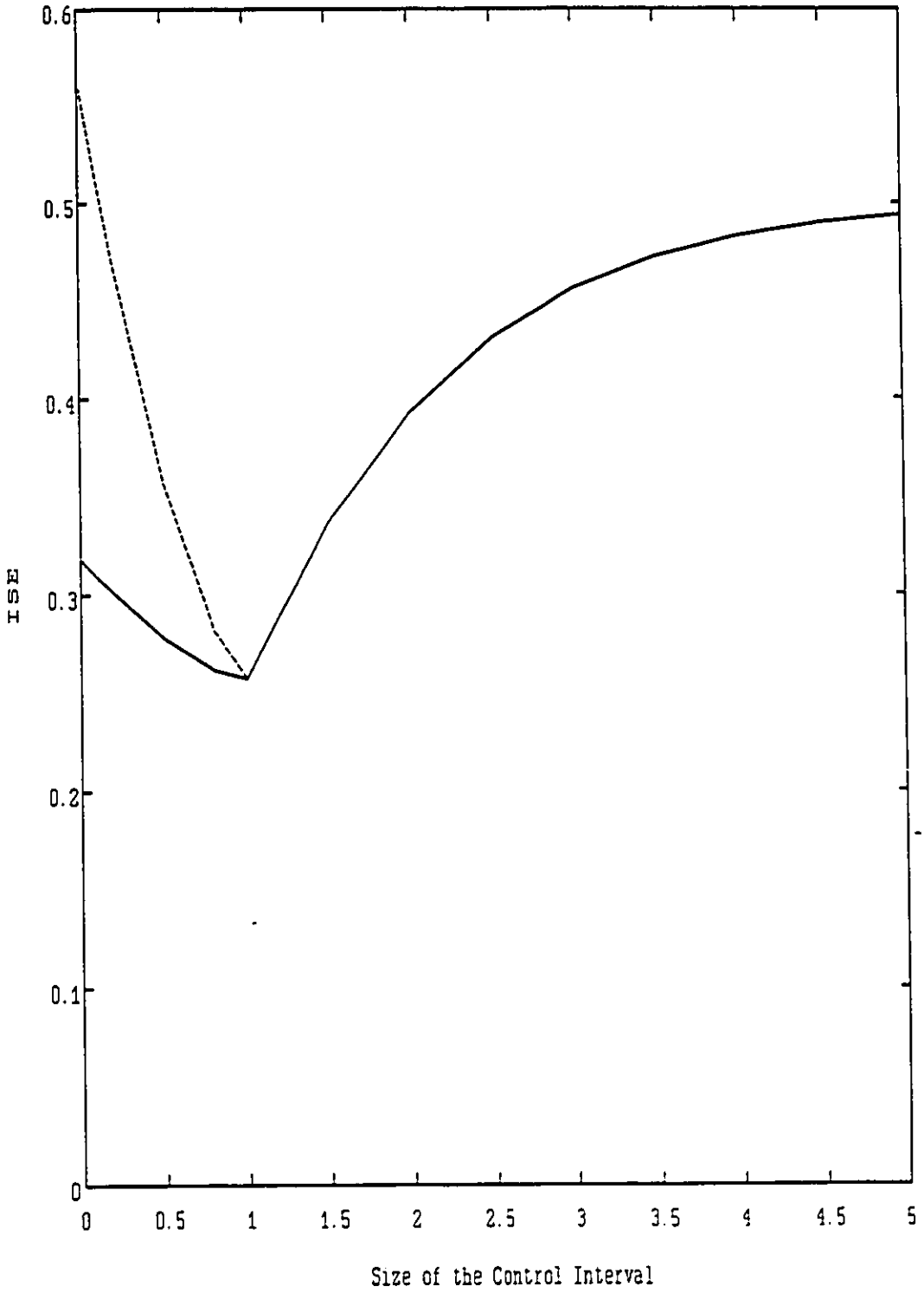


Figure 3.28 FO Process for Step Change in Set Point ISE vs T_c
- State Deadbeat IMC Control, -- LQOC

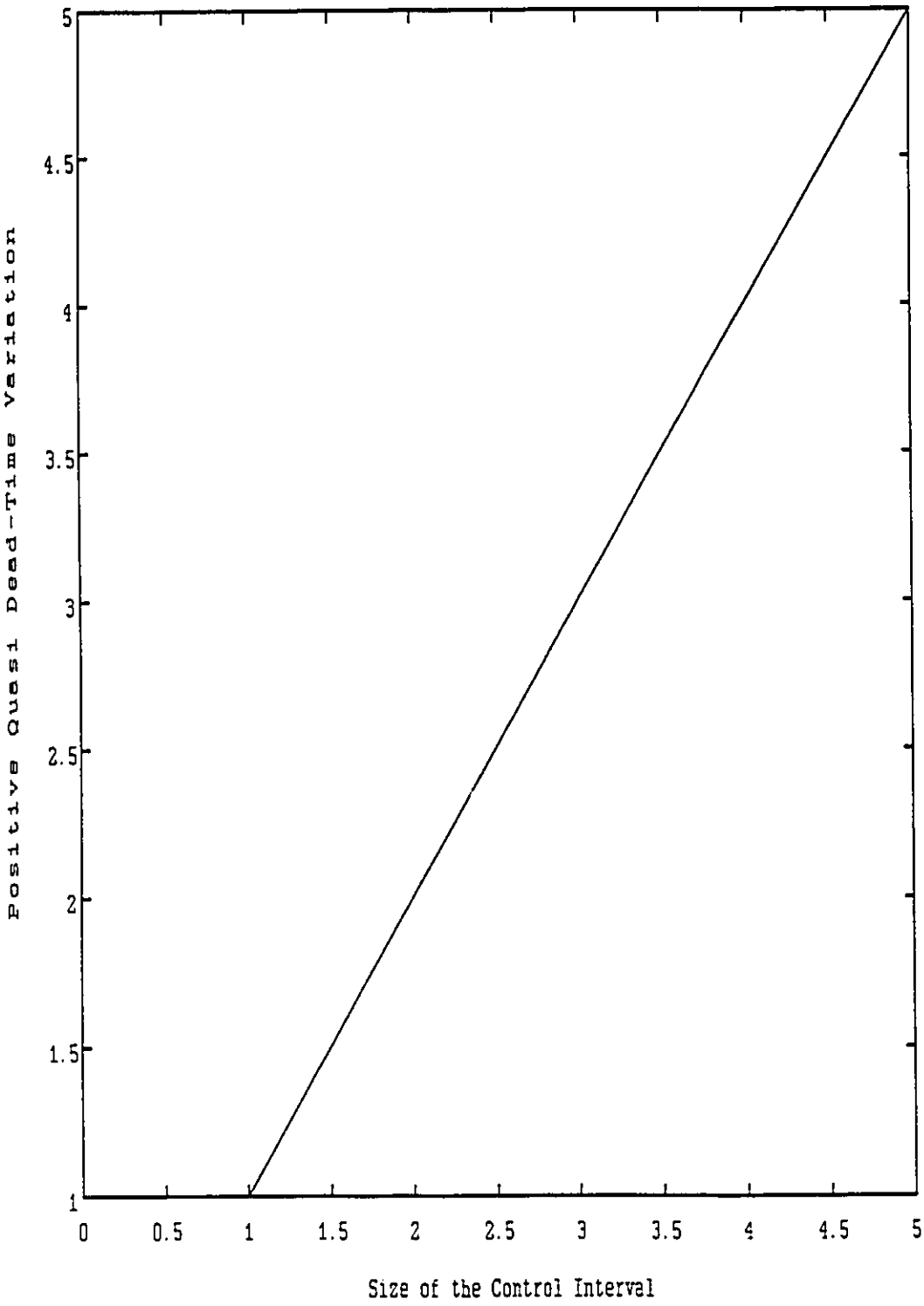


Figure 3.29 Allowable Positive Dead-Time Variation vs Size of the Control Interval

of the interval is reduced below T_c=1 the ISE performance of the closed loop system begins to increase. When the control interval is T_c=0.01 the ISE=0.32, a lower value than when T_c=1.5.

The comparison in Figure 3.28 shows that both the LQOC and State Deadbeat IMC controllers have poorer ISE performance as the control interval is reduced below T_c=1 in this example where the controllers are detuned for robustness. However, the ISE of the State Deadbeat IMC controller is significantly lower than the ISE of the LQOC controller when T_c<1. In order to understand the difference between these two designs the time step set point response of the controllers must be compared. Figure 3.30 shows the process response to a set point change with control interval T_c=0.01 using the LQOC controller with tuning constant λ=726.5919 with the State Deadbeat IMC design with tuning constant λ=0.9844. Comparison of the process responses shows that although the two systems have roughly the same settling time, the system with the IMC design has a much faster initial rise which leads to the lower ISE. Comparison of the controller outputs shows that the LQOC design calculates a control action which initially rises, then overshoots and finally settles out at the steady state value. The IMC design calculates a control action which initially overshoots and then gradually decays to the final steady state value.

The difference between these two controller designs results from their design bases. Because the LQOC controller penalizes the changes in the controller output ∇U_t it results in a controller which takes small steps in the control action. The LQOC controller for this example uses a second order tuning filter as shown in section 3.2.7. Because

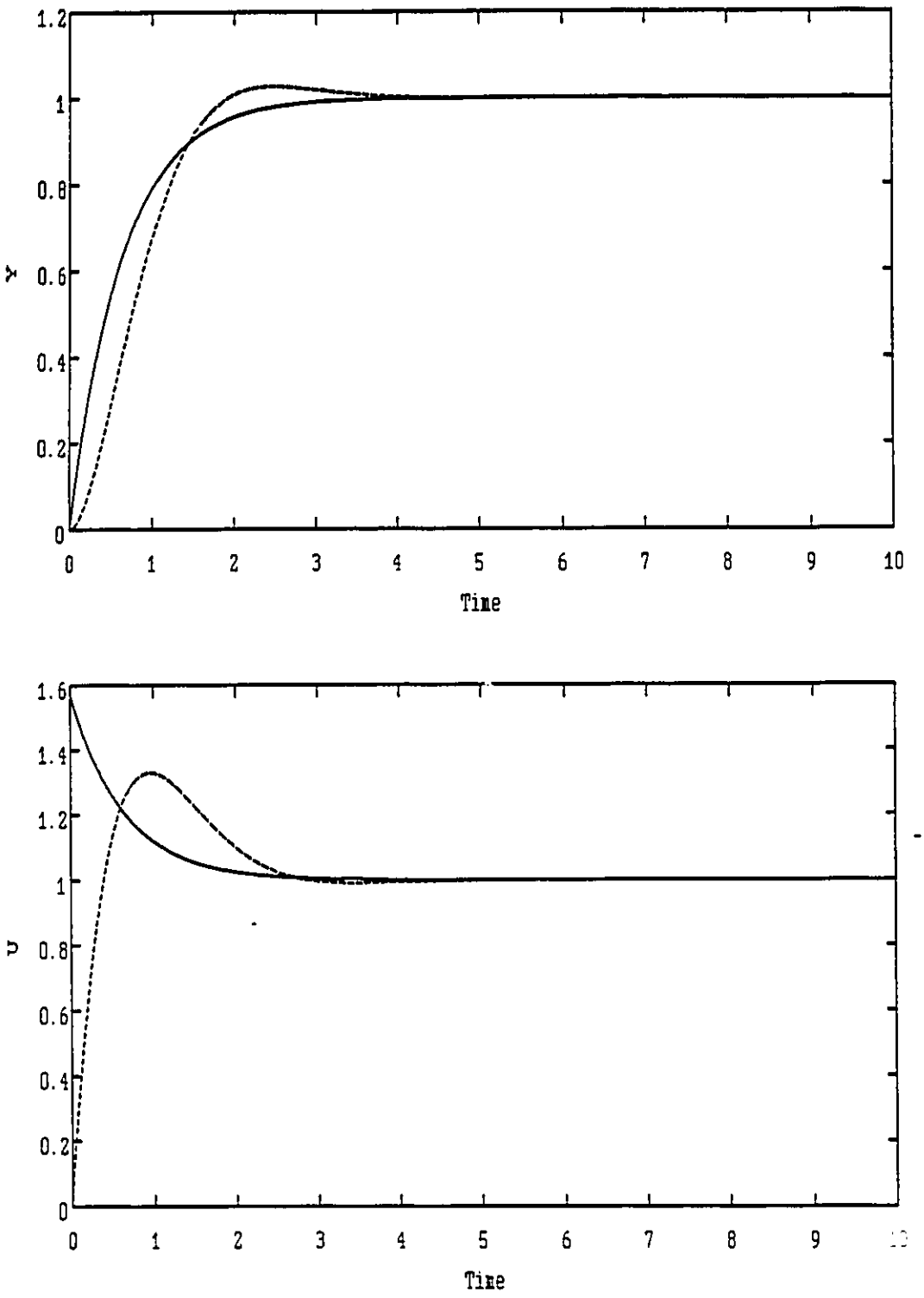


Figure 3.30 FO Process Step Set Point Change, Y and U vs Time - State Deadbeat IMC, -- LQOC

the IMC design applies a first order filter to the minimum variance controller output it results in a controller where the first control action overshoots the final steady state value and then decays towards it in a first order manner.

3.5.2 Performance and robustness of discrete model based control of a second order process for step changes in set point

In this section an example (similar to the one in Section 3.1.7) of a second order process with gain K_p=1 and time constants T_1=1 and T_2=1 with no dead-time is again investigated. For this example, with an ARIMA(0,1,0) disturbance, the Minimum Variance controller, the Linear Quadratic Optimal Controller and modified Linear Quadratic Optimal Controllers with λ=0 and the Internal Model Controller with $F(z^{-1})$=1 are all identical for any choice of the control interval. However, the State Deadbeat IMC controller with $F(z^{-1})$=1 is different in that it replaces the process zero with a gain when calculating the process model inverse. As shown in the previous section the robustness of minimum variance (and equivalent) control applied to this process decreases with the size of the control interval. This section focuses on the LQOC, the Modified LQOC, the IMC and the state deadbeat IMC controller designs as they are detuned.

In this section the tuning parameter of the LQOC controller, described in Section 3.2.2, the tuning parameter of the IMC controller first order filter, described in Section 3.2.3, and the tuning parameter of the State Deadbeat IMC controller first order filter, described in section 3.2.3, are used to detune the controllers. As in Section 3.2 the parameters are all referred to as λ for simplicity. For a range of

control times from T_c=0.01 to 5 these three controllers are tuned so that they provide the same allowable dead-time variation δT_{qd}=1 as minimum variance control with T_c=1. This is done by numerical search. The ISE performance of the continuous system is then calculated as in Section 3.1.

Figure 3.31 shows the ISE plotted vs the size of the control interval for LQOC, Modified LQOC, IMC, State Deadbeat IMC and State Deadbeat IMC with a second order filter where the tuning parameter λ is chosen so that $\delta T_{qd} \geq 1$. For all values of $T_c \geq 1$, λ=0 because the system is more robust than when T_c=1.

For the LQOC controller, when T_c=0.8, 0.5, 0.2, 0.1 and 0.01 the values of λ=0.0054, 0.0273, 0.2166, 0.8941 and 90.4174 respectively. Figure 3.31 shows that as the size of the interval is reduced below T_c=0.8 the ISE performance of the closed loop system begins to increase. When the control interval is T_c=0.01 the ISE=0.74, a lower value than when T_c=2.

For the IMC controller, when T_c=0.8, 0.5, 0.2, 0.1 and 0.01 the values of λ=0.1322, 0.3822, 0.7154, 0.8504 and 0.9844 respectively. Figure 3.31 shows that as the size of the interval is reduced below T_c=1 the ISE performance of the closed loop system continues to decrease until T_c= 0.2 and then it begins to increase. When the control interval is T_c=0.01 the ISE=0.32.

For the State Deadbeat IMC controller, when T_c=0.8, 0.5, 0.2, 0.1 and 0.01 the values of λ=0, 0.2280, 0.6921, 0.8451 and 0.9843 respectively. Figure 3.31 shows that for control intervals T_c=1 and larger that the ISE is higher. This is the same result as observed in

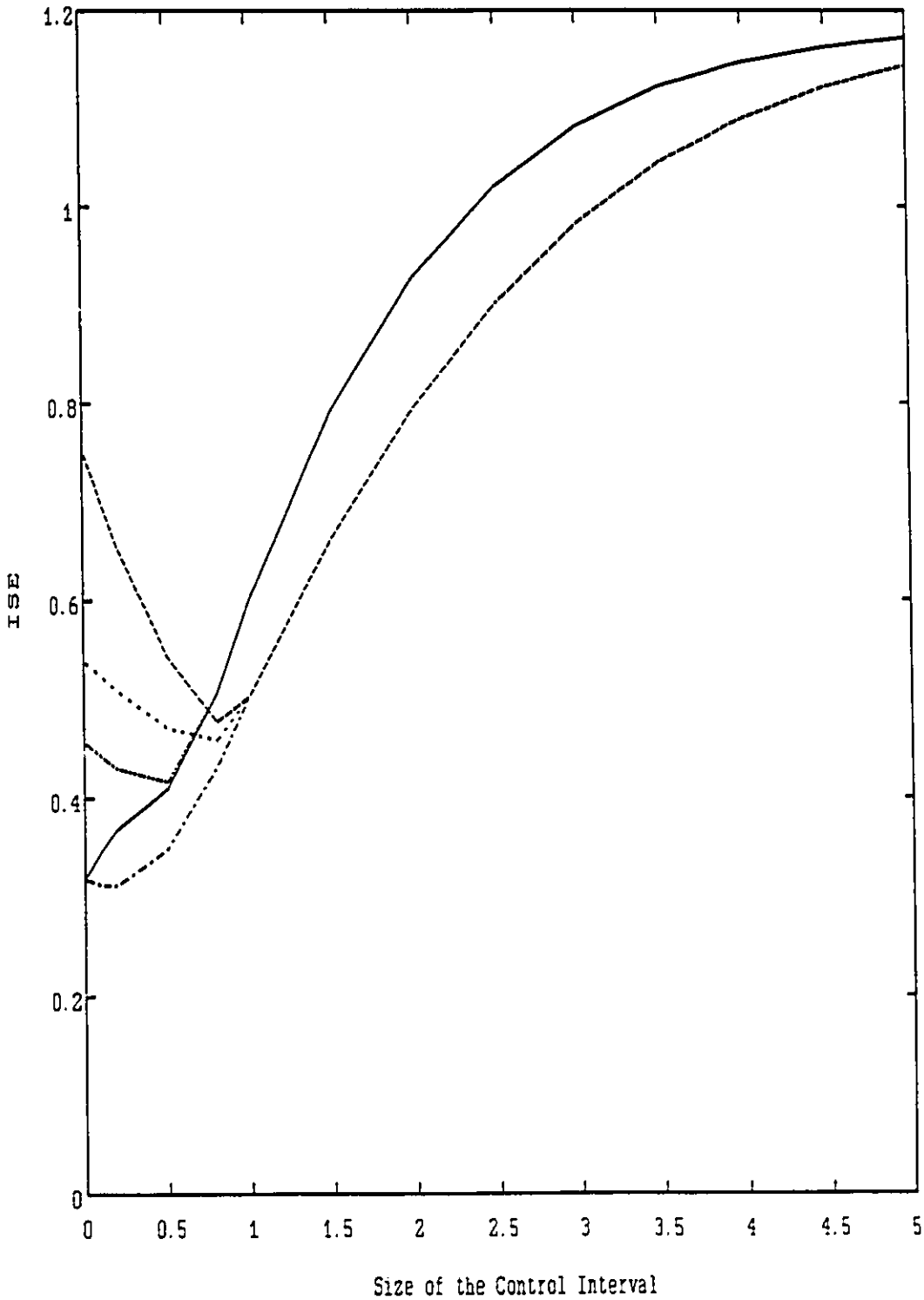


Figure 3.31 SO Process for Step Change in Set Point ISE vs T_c
-. IMC, - State Deadbeat IMC, -.. S.D. IMC 2nd order filter, -- LQOC, .. Modified LQOC

Section 3.1 above and is due to the lack of a pole in the approximate model inverse. As the size of the interval is reduced below T_c=1, the ISE performance of the closed loop system continues to decrease until T_c=0.01. When the control interval is T_c=0.01 the ISE=0.32.

For the Modified LQOC controller, when T_c=0.8, 0.5, 0.2, 0.1 and 0.01 the values of λ=0.0127, 0.0335, 0.0572, 0.0654 and 0.0726 respectively. Figure 3.31 shows that as the size of the interval is reduced below T_c=0.8 the ISE performance of the closed loop system begins to increase. When the control interval is T_c=0.01 the ISE=0.54, a lower value than when T_c=1.5.

For the State Deadbeat IMC controller with a second order filter (two identical poles), when T_c=0.8, 0.5, 0.2, 0.1 and 0.01 the values of λ=0, 0.1315, 0.5488, 0.7521 and 0.9729 respectively. Figure 3.31 shows that as the size of the interval is reduced below T_c=0.8 the ISE performance of the closed loop system begins to increase. When the control interval is T_c=0.01 the ISE=0.45, a lower value than when T_c=0.8.

In order to understand the difference between these designs the time step set point response of the controllers must be compared. Figure 3.32 shows the process response and control actions for a set point change with the LQOC controller designed with control interval T_c=0.01 and tuning constant λ=90.4174. Figure 3.33 shows the process response and control actions for a set point change with the IMC design for the same control interval and tuning constant λ=0.9844. Figure 3.34 shows the process response and control actions for a set point change with the state deadbeat IMC design for the same control interval and

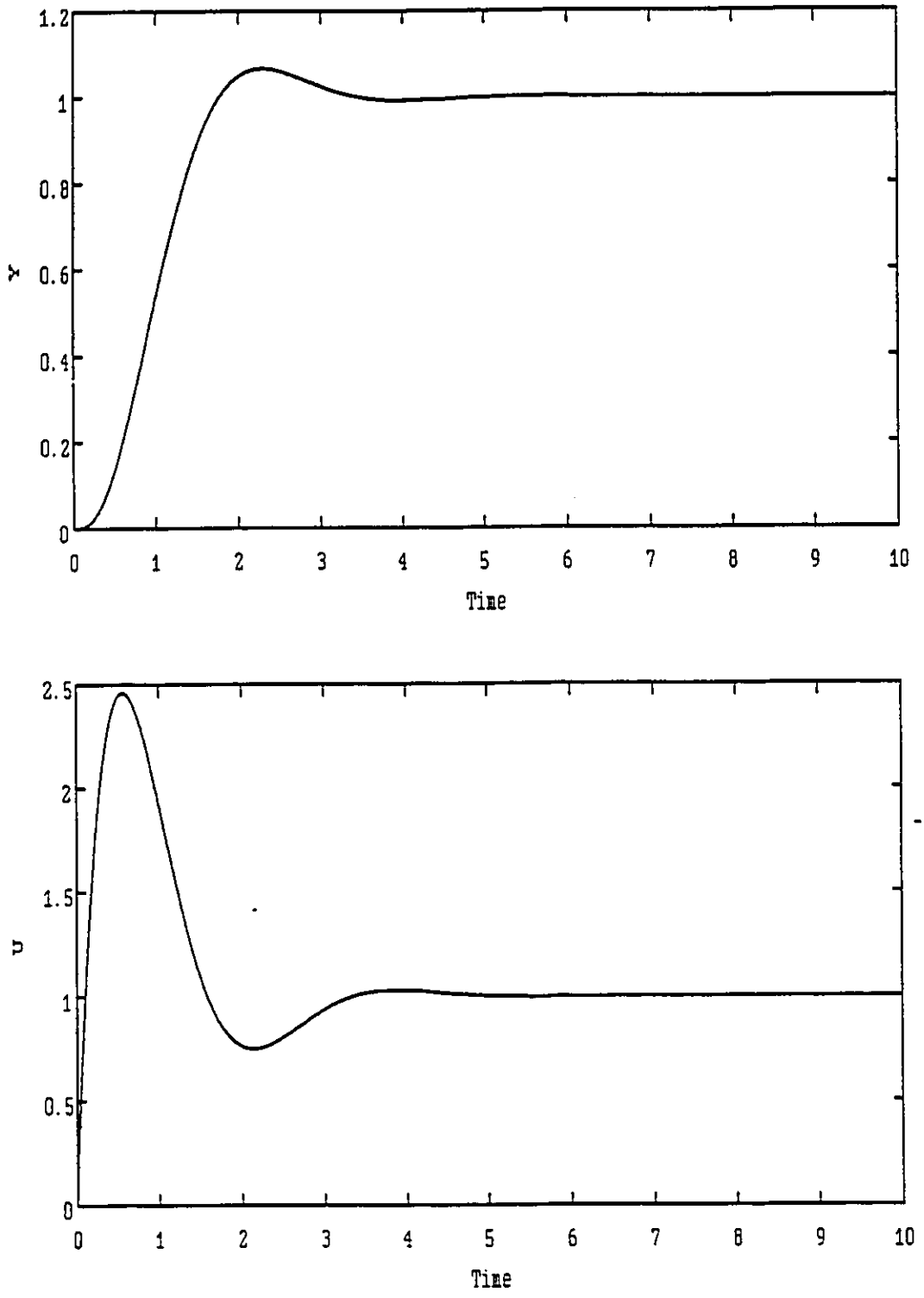


Figure 3.32 SO Process Step Set Point Change with LQOC Control

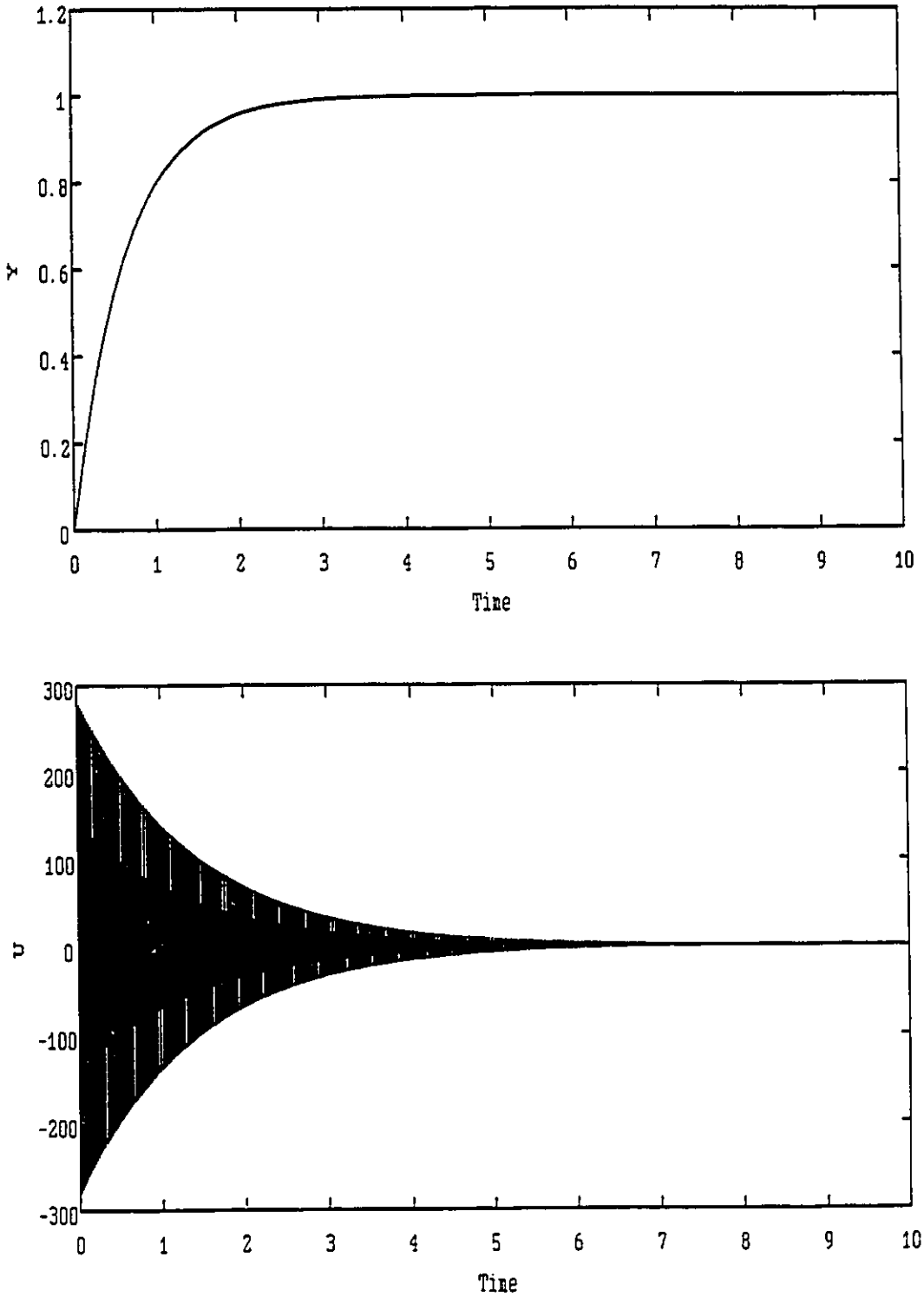


Figure 3.33 SO Process Step Set Point Change with IMC Control

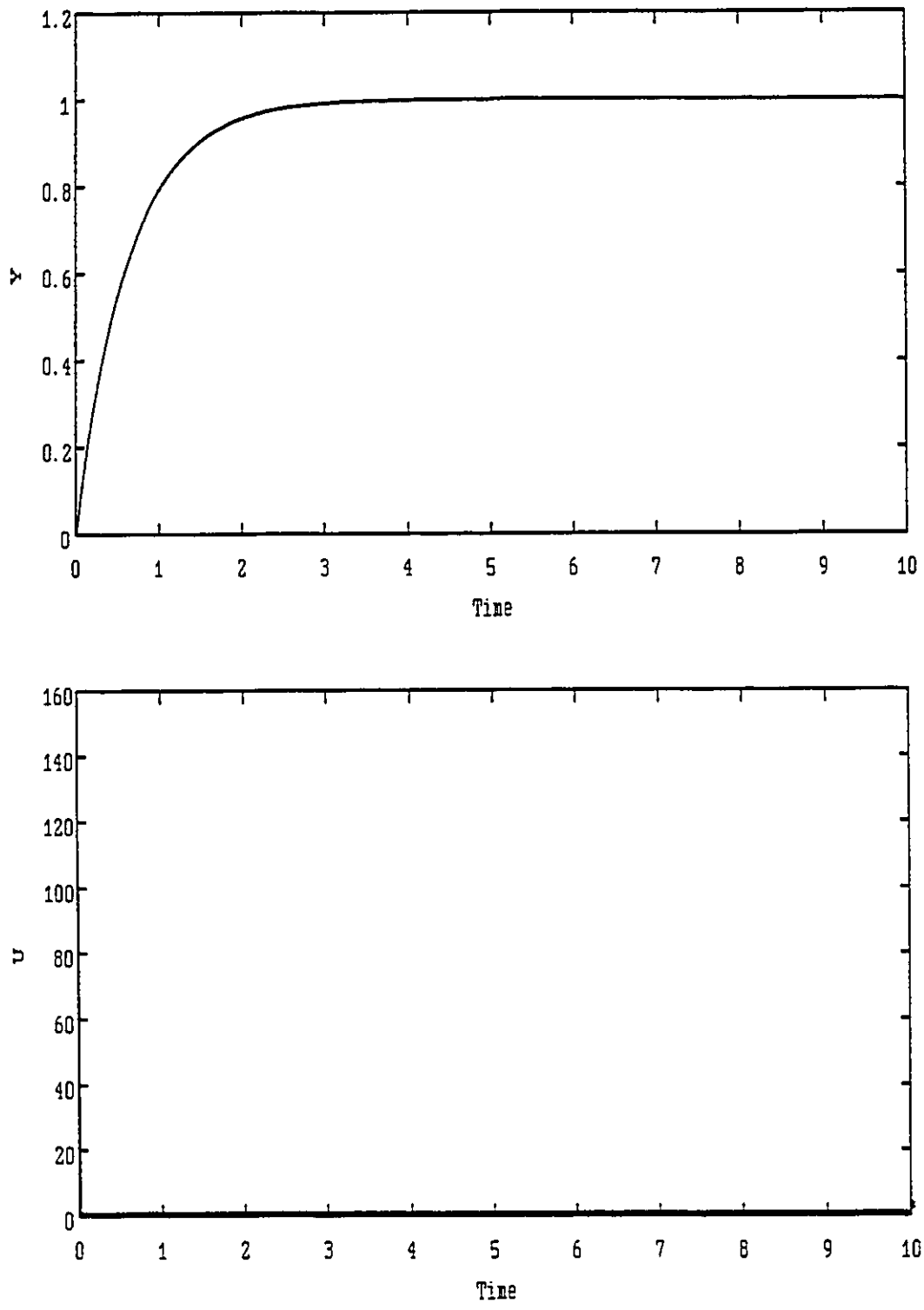


Figure 3.34 SO Process Step Set Point Change with State Deadbeat IMC

tuning constant λ=0.9843.

Comparison of the process responses shows that with the IMC and state deadbeat IMC design the processes have very similar responses with rapid initial rise times. The LQOC design results in a process response with a slow initial rise time and a slight overshoot and oscillation before settling at the new set point. Comparison of the controller outputs shows that the LQOC design calculates a control action which initially rises, then overshoots and finally oscillates and settles out at the steady state value. The IMC design calculates a control action which initially overshoots the final value by a factor of 283 and then rings severely as it decays towards the final steady state value. The state deadbeat IMC design calculates an initial control action which overshoots the final steady state value by a factor of 143 and then settles out at the final steady state value almost immediately.

The difference between these controller designs results from their design bases. Because the LQOC controller penalizes the changes in the controller output ∇U_t it results in a controller which takes small steps in the control action. Because the IMC design applies a first order filter to the minimum variance controller output it results in a severely ringing control action. The State Deadbeat IMC design does not keep the ringing process zero in the controller design so it results in a control action which settles very quickly. However, because the State Deadbeat control applies only a first order filter to a second order approximate model inverse, the control action is clearly unrealizable because of its huge overshoot.

Figure 3.35 shows the process response and control actions for a set point change with the State Deadbeat IMC controller where the tuning filter is 2nd (with two equal poles) to match the process order, with control interval T_c=0.01 and tuning constant λ=0.9729. Figure 3.36 shows the process response and control actions for a set point change with the Modified LQOC design for the same control interval and tuning constant λ=0.0726. Because both designs use a second order tuning filter, the control actions are reasonable. For both of these controllers the first control action is the largest and it is followed by a smooth decay towards the final steady state.

This section shows that performance and robustness analysis are not enough on their own to guarantee an acceptable controller design: the control actions in the resulting design must be realizable. In particular, when designing a controller for a higher order process with a small control interval, as in the example studied here, the tuning filter must be matched to the order of the process to obtain a good design with reasonable control actions. The IMC and State Deadbeat IMC controller designs have too low a filter order and produce unrealizable control actions. The LQOC controller has too high a filter order and produces poor performance for the same level of robustness as the other controller designs. The State Deadbeat IMC controller with a tuning filter matched to the order of the process and the Modified LQOC controller which has a tuning filter naturally matched to the order of the process both produce reasonable control actions as well as good performance and robustness.

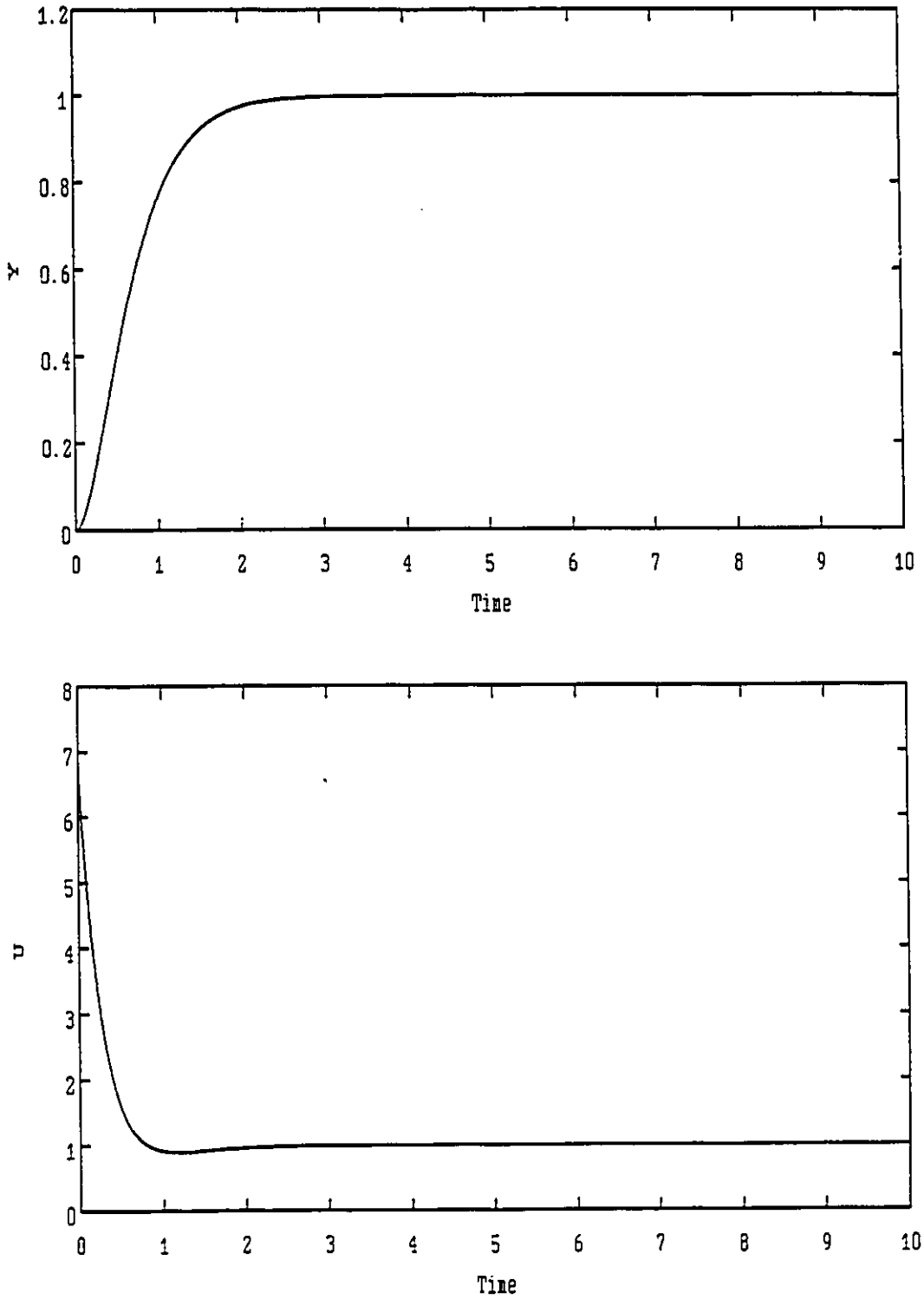


Figure 3.35 SO Process Step Set Point Change with State Deadbeat IMC with 2nd Order Filter

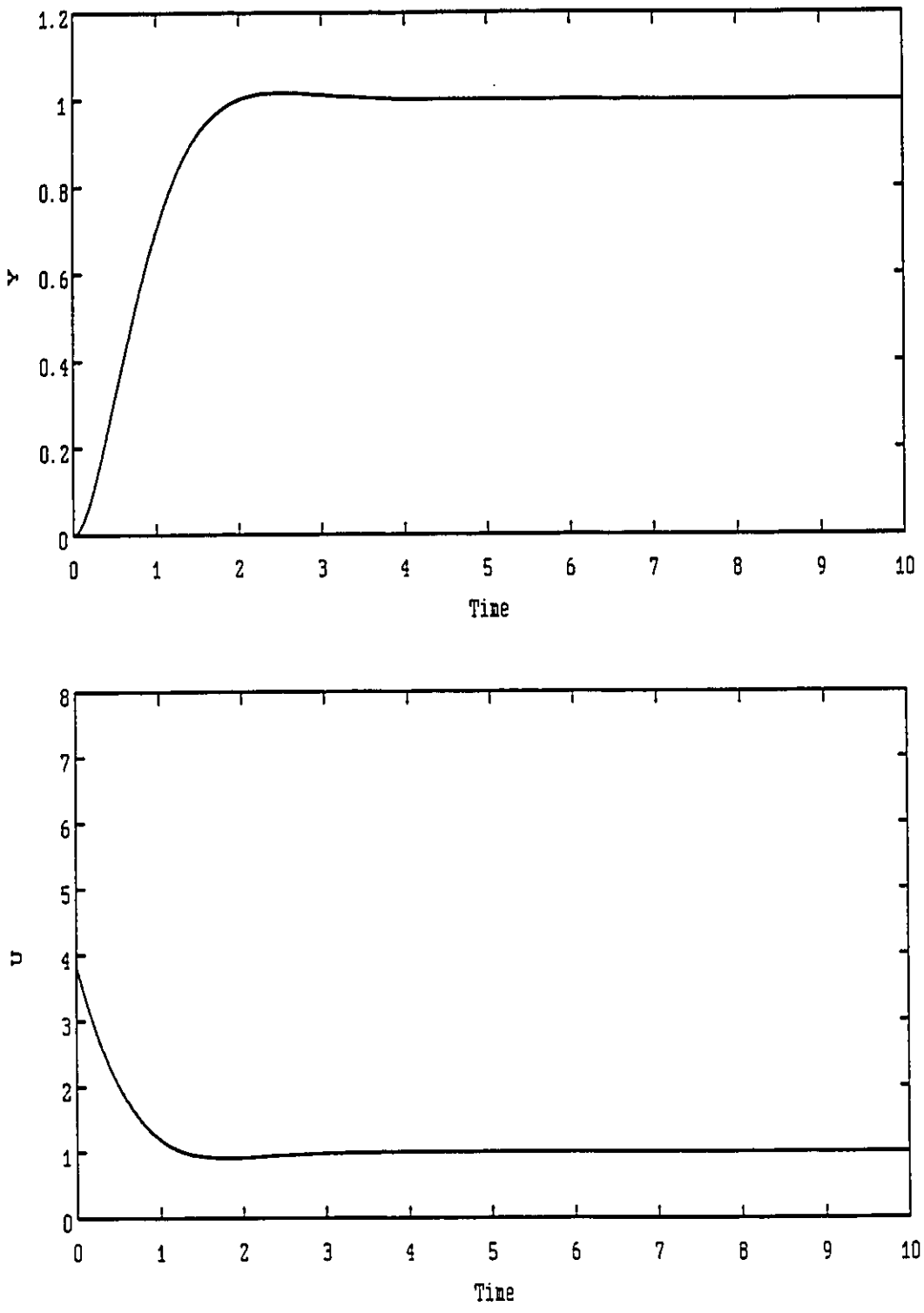


Figure 3.36 SO Process Step Set Point Change with Modified LQOC

3.5.3 Summary

This section has shown that the tuning parameter in a model based controller design can be used to increase the system robustness at the expense of system performance. This tuning counters the adverse effects of a reduction in the control interval on robustness. In addition this section has shown that a controller design can have good robustness and performance characteristics and yet require unacceptable control actions. Examples were used to show that the tuning parameter can be used to provide a desired level of robustness in the LQOC, IMC and state deadbeat IMC and modified LQOC designs. However, the performance vs robustness tradeoffs are not the same for the different controller designs. The results showed that the IMC and State Deadbeat IMC controllers have the best tradeoff between performance and robustness, however they can both produce totally unrealizable control actions. The LQOC controller has the poorest tradeoff between performance and robustness among the controller designs which are compared. The State Deadbeat IMC controller with a tuning filter order chosen to equal the order of the process and the Modified LQOC controller both produce acceptable control actions and have a similar robustness vs performance tradeoff.

3.6 Conclusions and recommendations

This chapter has examined the application of discrete model based controllers to the control of continuous processes. The wide ranging analysis has examined the effects of the size of the control interval and of the controller tuning on the controller performance, robustness and the realizability of the control actions.

That many of the different controller designs which are available do not perform well for some choices of the control interval was demonstrated here by comparing the performance of the Linear Quadratic Optimal Controller (Åström (1970), Box and Jenkins (1970)), the modified Dahlin controller (Dahlin (1968)) and the State Deadbeat Internal Model Control design (Zafiriou and Morari (1985)) with no detuning as the size of the control interval is varied. Comparison of the Integral Square Error and Integral Absolute Error of the continuous process output were used to show under what conditions each of the different controllers perform poorly. The results were explained by using continuous time plots of the process output and manipulated control action response to step changes in the set point. This study confirms many of the desirable properties of the State Deadbeat IMC controller observed by Zafiriou and Morari (1985) and extends their study by showing that LQOC controller outperforms the State Deadbeat IMC controller when the process model numerator has a zero which is not near minus one.

The same discrete controller designs were then compared for the manner in which changes in the tuning parameter effect changes in the

tuning filter. In addition, two new controller tuning methods were proposed: a Modified LQOC controller design and an extended horizon controller design. The Modified LQOC controller calculates a controller for a nonstationary disturbance using the spectral factorization approach as if the disturbance were stationary. This controller has a tuning filter which is of order one less than the LQOC tuning filter. Through the appropriate choice of the tuning filter gain, the resulting controller still has integral action. The Modified LQOC controller was shown to provide a bridge between the LQOC controller of Wilson (1970) and Harris and MacGregor (1987) and the State Deadbeat IMC designs.

The new extended horizon controller design behaves as if it were a state deadbeat controller for the same process but with a longer control interval. This latter controller is an improvement over the extended horizon controller of Ydstie et al (1985) in that if a single set point change is made and no disturbance enters the system then the initial control action sequence will be implemented with no changes. This new controller is less convenient to implement than the State Deadbeat IMC controller to which it is closely related, but it is still instructive as it explains the relationship between tuning and the choice of the size of the control interval.

As a prelude to exploring robustness analysis for discrete control of continuous systems, the effect of a change in the dead-time of a continuous process on the frequency response of the discrete equivalent model was examined. The modified Z transform of a first order process with a fractional period of dead-time was shown to be a convex combination of the Z transforms of the same process with integer

periods of dead-time which bracket the actual dead-time. Through the use of the example of a second order process, it was shown that this result is not general. In fact, for a second order process the frequency response of the process with a fractional period of dead-time can be outside of the boundary formed by the frequency response of the process with integer periods of dead-time which bracket the actual dead-time. This result appears to be new and has implications for studies which determine stability boundaries by a search on the dead-time parameter.

The definition of the robustness region for continuous processes, presented in the previous chapter, was adapted for discrete control of continuous processes as the region of joint allowable variation in the process gain and the quasi dead-time. The quasi dead-time is defined to be the phase margin of the discrete open loop frequency response divided by the discrete frequency. When the allowable variation in the quasi dead-time is an integer number of control intervals and the process has no fractional period of dead-time then it provides an exact upper limit on the allowable dead-time variation in the underlying continuous process. When the allowable change in the quasi dead-time does not have a direct interpretation in terms of the dead-time of the underlying continuous process it is nonetheless useful for comparing the robustness of different controllers. The power and the limitations of this definition of robustness were shown by application to examples.

The method was used to show that a minimum variance controller for a process with no dead-time subject to random walk disturbances

(equivalent to deadbeat control for step changes in set point) can only tolerate an increase in the process dead-time of at most one control interval. That is, the robustness of minimum variance control is proportional to the size of the control interval. This is a new result which provides a control system designer with a useful rule of thumb. It was also shown that robustness of minimum variance control is further reduced when the process has dead-time. The limitation of the method was shown to be that when calculating the robustness of discrete control applied to a process with a fractional period of delay, the allowable variation in the quasi dead-time can significantly over-estimate the allowable variation in the true dead-time. However, the relative results when comparing different controllers on the same process are meaningful. As in the continuous time definition of robustness, the quasi dead-time incorporates frequency information in a natural and meaningful way.

The allowable change in the quasi dead-time measure of robustness and the continuous Integral Square Error measure of performance were used together to compare the robustness vs performance tradeoff of the tuning of LQOC, Modified LQOC, IMC, and State Deadbeat IMC controllers applied to first and second order processes with step changes in set point. The LQOC controller design was shown to have the poorest performance for the same level of robustness. The results were explained using continuous time plots of the process output and the control actions. These plots show additionally that a controller with good performance and good robustness can still require totally impractical control actions. The IMC and State Deadbeat IMC controllers

can require a higher order tuning filter for higher order processes to obtain realizable control actions. The Modified LQOC controller has a filter order which is naturally matched to the order of the process. The poor performance of the LQOC controller is the result of a tuning filter which is of an order one higher than necessary. Bergh and MacGregor (1987) have studied the improved disturbance rejection ability of the LQOC controller over the IMC controller for autoregressive disturbances. Their study, however, focuses on discrete measures of performance for a process example with a fractional period of dead-time and does not make a comparison with a State Deadbeat IMC controller or with the Modified LQOC controller presented here, both of which provide improved robustness for this type of process.

This study has contributed to the understanding of the relationships between approximate model inverses, tuning filters, performance, robustness and realizability of control actions.

CHAPTER 4

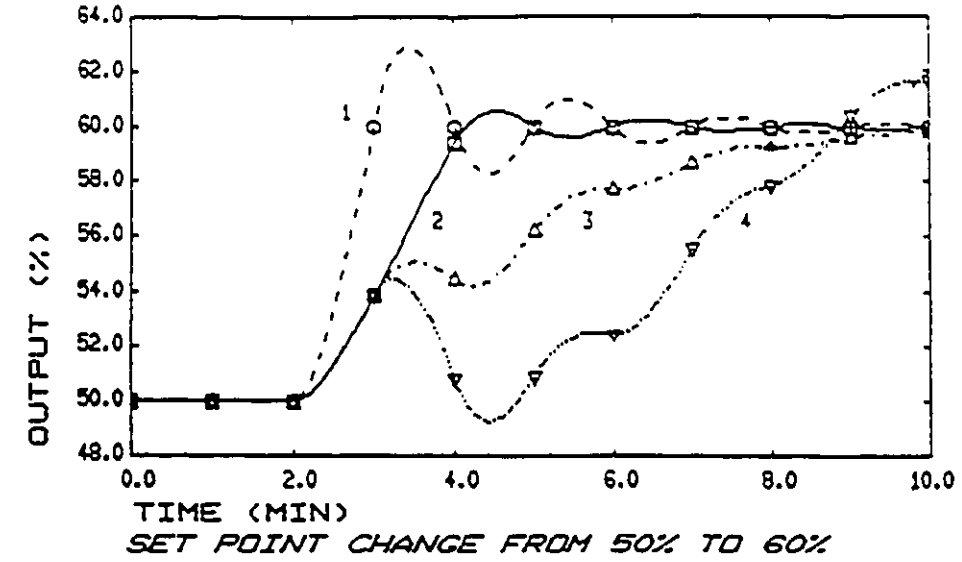
ONE-STEP OPTIMAL CORRECTION FOR INPUT SATURATION IN DISCRETE MODEL BASED CONTROLLERS

One of the most common problems encountered in control of processes is that the control actions calculated by a controller cannot be implemented in full; that is, the control input saturates. Examples include inequality constraints on the input, such as actuators which can only range between fully open and fully closed; inequality constraints on the magnitude of the input changes, such as actuators which should not be opened or closed faster than a certain rate; and general situations where the calculated control action is not applied, such as in advisory control where an operator chooses not to implement the full action calculated by the control equations. Linear controllers only meet their design specifications when the system is in its linearized operating region. If a calculated control input to the system violates a saturation limit then the subsequent control will not in general be satisfactory.

To demonstrate the nature of the problem, a discrete dead-beat or minimum variance controller designed for step changes in setpoint is applied to the continuous time simulation of a second order system with

one period of dead-time. Figure 4.1 shows the process output and the process input responses to a set point change from 50% to 60% applied at time 1. When no limit is placed on the process input, the process output (curve 1 in Figure 4.1) is seen to reach the setpoint (at the control times) one dead-time after the set point change is made. If the process input cannot exceed 100% then the calculated control action will saturate when the set point change is made. If no account is taken of the saturation and the past calculated control actions are used in the controller discrete transfer function to calculate the subsequent control actions, then the resulting performance is very poor (curve 4 in Figure 4.1). If the past actually implemented control actions are used in the controller (a method which has reset windup protection), then one obtains a very slow approach to setpoint (curve 3 in Figure 1). A new algorithm presented in this chapter brings the process output to set point in a manner which is one-step optimal subject to any limit on the input (curve 2 in Figure 1).

Much of the previous research aimed at extending linear control to systems with saturation limits has focused on integrator or reset windup in Proportional-Integral-Derivative controllers. Khandheria and Luyben (1976) present and evaluate several algorithms for reset windup protection. Smith (1972) and Åström and Wittenmark (1984) discuss reset windup protection in their textbooks. Segall and Taylor (1986) (also, Chapter 5 here) and Banyasz et al (1985) have presented new algorithms for reset windup protection. All of these algorithms are advanced on an ad hoc basis and have no particular claims to optimality.



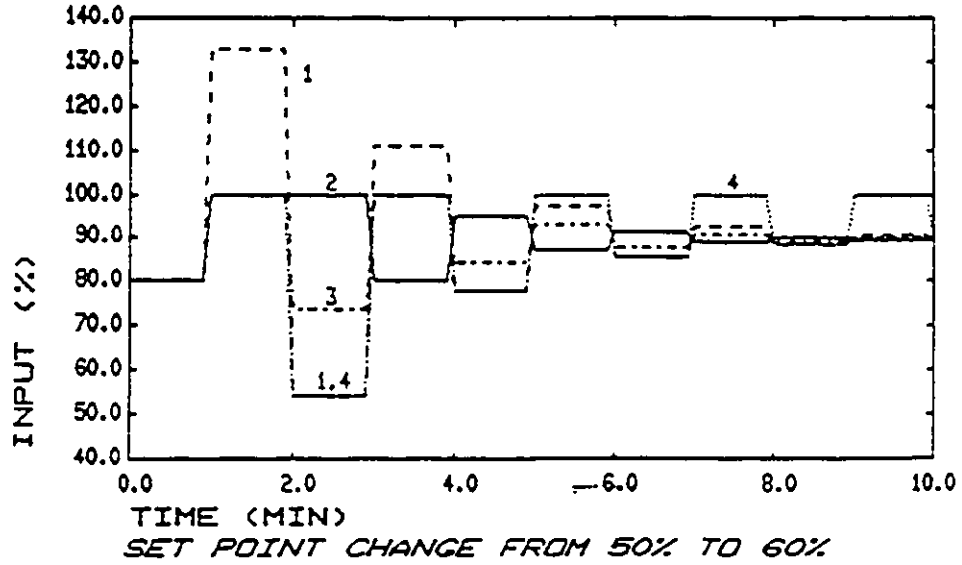


Figure 4.1 Deadbeat or Minimum Variance Control for Second Order System.
Step change in setpoint from 50% to 60% at time 1 Minute.
1. No saturation limit.
2,3,4 Saturation limit on input at 100%.
2. Saturation Correction algorithm used.
3. Reset windup protection only, uses past actually implemented Ut.
4. No Saturation correction, uses past calculated Ut.

Another approach to handling input constraints is through on-line mathematical programming methods (Tabak and Kuo (1971)). Chang and Seborg (1983) and Garcia and Morshedi (1984) have used linear and quadratic programming respectively to handle saturation limits on the inputs and the states of linear multivariable systems. Wong et al (1986) have made both experimental and simulated comparisons between several reset windup protection algorithms and mathematical programming algorithms. Their study shows that the mathematical programming solution using even a one-step horizon performs better than any of the reset windup algorithms. Although it is always possible to use such on-line mathematical programming techniques to calculate the inputs to a process which will be optimal subject to saturation limits, this is rather more power than is normally brought to bear on linear systems with simple constraints on the inputs.

Goodwin (1972) presented a saturation compensation algorithm for single-input single-output one-step optimal LQG controllers. Bezanson (1984) extended this to one-step optimal LQG controllers with more general penalties in the LQG objective. However, these control algorithms involve rational polynomial matrices in z^{-1} which must be implemented either by expanding as infinite polynomials or by introducing auxiliary states. Mäkilä (1982) attempted to implement the result of Goodwin (1972) and to extend it to a multivariable one-step controller. Clarke (1981) presents a univariate saturation correction which involves three equations: a prediction, a control calculation and a revised prediction. Toivonen (1983a,b,c) presented a solution to the infinite step optimal control problem which is based on expectations

over all future time assuming saturation constraints of the form $A \leq U_t \leq B$. This off-line design does not account for the actual history of saturations which has occurred, nor does it allow for time varying constraints. It results essentially in a reduction of the controller gain so that the saturation limits will not be violated. Parrish and Brosilow (1985) presented some empirical rules for saturation compensation for what they refer to as inferential controllers.

This chapter presents a new method for correcting model based discrete transfer function controllers for saturation of the past control inputs which is shown to be one-step optimal. For multivariable systems a further correction is developed which causes other control inputs to compensate simultaneously in a one-step optimal manner for any control inputs which saturate. This material has been published as it was developed in a Technical Report (Segall et al (1987b)), at conferences (Segall et al (1986), (1987a)) and as a journal article (Segall et al (1991)).

4.1 Linear process models

This chapter considers linear multivariable systems which can be modelled by a discrete transfer function with an additive disturbance of the Autoregressive Integrated Moving Average (ARIMA) form (Box and Jenkins (1970), Åström (1970), Kashyap and Rao (1976)). The model is written as:

$$A(z^{-1})\, Y_t = B(z^{-1})\, U_{t-f-1} + C(z^{-1})\, e_t \qquad (4.1)$$

where Y_t is an n-dimensional vector of process output deviations from the setpoint vector $Y_{sp,t}$ or from steady state, U_t is an n-dimensional vector of process input deviations from steady state, f is the number of whole periods of process delay, and e_t is an n-dimensional vector of zero mean white noise processes. The n x n polynomial matrices in the backward difference operator z^{-1} describing the dynamics of the process and the disturbance model are:

$$A(z^{-1}) = I + A_1 z^{-1} + \ldots + A_r z^{-r} ,$$

$$B(z^{-1}) = B_0 + B_1 z^{-1} + \ldots + B_s z^{-s} \quad \text{and} \quad B_0^{-1} \text{ exists},$$

$$C(z^{-1}) = I + C_1 z^{-1} + \ldots + C_q z^{-q} ,$$

and E'E is the variance-covariance matrix of the e_t. The roots of the determinants of the $A(z^{-1})$ and $C(z^{-1})$ polynomial matrices all have their roots on or inside the unit circle in z. The $A(z^{-1})$ and $B(z^{-1})$ matrices may have a common scalar factor ∇^d where ∇ is the backwards difference operator $1-z^{-1}$. This allows for nonstationarity of the disturbance, and d (usually zero or one) is the degree of nonstationarity.

The disturbances may be stochastic in nature in which case the a_t's are usually assumed to be Gaussian, or they may be randomly occurring deterministic disturbances or set point changes such as steps, ramps, exponential changes to a new level, etc., in which case the model structure is still as given in equation (4.1), but the e_t's are highly non-gaussian (MacGregor et al, (1984)).

4.2 A one-step optimal controller with saturation compensation algorithm

This section presents the derivation of a controller which minimizes the one-step objective function:

$$J_t = E\{ Y'_{t+f+1} \, Q \, Y_{t+f+1} + U'_t \nabla^d R \nabla^d U_t \mid Y_t, Y_{t-1}, \ldots, U_{t-1}, U_{t-2}, \ldots \} \qquad (4.2)$$

subject to any set of time-varying constraints on the control vector U_t. Q and R are n x n weighting matrices and Q is positive definite. The control algorithm can easily be generalized to include more general polynomial matrices operating on Y_{t+k+1} and U_t in the objective function equation (4.2). E { } is the expectation operator.

This minimization is subject to the condition that the past inputs U_t may, for any reason whatsoever, have not been implemented as called for. No assumption is made in this derivation about the form of the saturation limits on the past U_t. These may be in the form of time-varying hard limits on the magnitude of the control input, hard limits on the rate of change of the control input, or in general any linear or nonlinear constraints which prevent the calculated control action from being fully implemented.

The derivation of the algorithm which minimizes equation (4.2) subject to hard constraints follows the single-variable derivations of Clarke and Hastings-James (1971), and Clarke and Gawthrop (1975), the multivariable derivations of Åström (1975) for the minimum variance controller, and the MIMO derivations of Koivo (1980) and Mäkilä (1982) which minimize this one-step objective function for the case of no hard constraints.

Substituting the model equation (4.1) into the objective function equation (4.2) gives:

$$\begin{aligned}\underset{U_t}{\text{Min}}\ E\ \{\ & [A^{-1}(z^{-1})B(z^{-1})U_t + A^{-1}(z^{-1})C(z^{-1})e_{t+f+1}]'\ Q \\ & \times [A^{-1}(z^{-1})B(z^{-1})U_t + A^{-1}(z^{-1})C(z^{-1})e_{t+f+1}] \\ & + U_t'\ \nabla^d R\ \nabla^d\ U_t\ \}\end{aligned} \tag{4.3}$$

The effect of the disturbance $A^{-1}(z^{-1})C(z^{-1})e_{t+f+1}$ in equation (4.3) can be separated into forecastable and unforecastable parts by solving the linear Diophantine equation:

$$C(z^{-1}) = A(z^{-1})F(z^{-1}) + z^{-f-1}G(z^{-1}) \tag{4.4}$$

for the least degree solution with respect to $F(z^{-1})$. The degree of $F(z^{-1})$ will be f and $F_0 = I$. This equation can be solved using the methods of Kucera (1979).

Substituting for $A^{-1}(z^{-1})C(z^{-1})$ in equation (4.3) using equation (4.4), and taking conditional expectations gives:

$$\begin{aligned}\underset{U_t}{\text{Min}}\ \{\ & [A^{-1}(z^{-1})B(z^{-1})U_t + A^{-1}(z^{-1})G(z^{-1})e_t\ ]'\ Q \\ & \times [A^{-1}(z^{-1})B(z^{-1})U_t + A^{-1}(z^{-1})G(z^{-1})e_t] \\ & + U_t'\ \nabla^d R\ \nabla^d\ U_t\ + S\ \}\end{aligned} \tag{4.5}$$

where S is the k+1 step ahead forecast error variance for the disturbance:

$$S = E'\ [\ I + F_1'QF_1 + \ldots + F_f'QF_f\ ]\ E \tag{4.6}$$

To minimize the quadratic cost function (4.5) with respect to only the current control action U_t equation (4.5) is differentiated with respect to U_t to give:

$$J_{t,U} = 2B_0'Q\ [A^{-1}(z^{-1})B(z^{-1})U_t + A^{-1}(z^{-1})G(z^{-1})e_t\ ] + 2R\nabla^d U_t \tag{4.7}$$

The random shocks e_t can be expressed in terms of the actual output disturbances as:

$$e_t = C^{-1}(z^{-1}) \; [\; A(z^{-1})Y_t - B(z^{-1})U_{t-f-1}] \tag{4.8}$$

Eliminating e_t from equation (4.7) using equation (4.8) and rearranging gives:

$$\begin{aligned} J_{t,U} = \; & 2B_0'QA^{-1}(z^{-1})G(z^{-1})C^{-1}(z^{-1})A(z^{-1})Y_t \\ & + 2B_0'QF(z^{-1})C^{-1}(z^{-1})B(z^{-1})U_t + 2RV^dU_t \end{aligned} \tag{4.9}$$

Equation (4.9) is now simplified using the relationship (Åström (1978)):

$$C(z^{-1})F^{-1}(z^{-1})A^{-1}(z^{-1})G(z^{-1}) = G(z^{-1})F^{-1}(z^{-1})A^{-1}(z^{-1})C(z^{-1}) \tag{4.10}$$

to give:

$$\begin{aligned} C(z^{-1})F^{-1}(z^{-1})Q^{-1}B_0^{-1'}J_{t,U} = \; & 2G(z^{-1})F^{-1}(z^{-1})Y_t \\ & + 2B(z^{-1})U_t + 2C(z^{-1})F^{-1}(z^{-1})Q^{-1}B_0^{-1'}RV^dU_t \end{aligned} \tag{4.11}$$

The current control action is one-step optimal if it is chosen so that $J_{t,U}$ is zero. Define U_t^* as the control action which would set $J_{t,U}$ to zero in the absence of any saturation limits, and U_t as the control action that is actually implemented, that is:

$$U_t = \text{sat}\{U_t^*\} = \begin{cases} U_t^* \text{ if no limit is encountered} \\ \\ U_{limit} \text{ if a limit is encountered} \end{cases} \tag{4.12}$$

If the calculated optimal control action U_t^* cannot be implemented due to saturation in any form then $J_{t,U}$ will not be zero. In this case it can be seen from equation (4.9) that $J_{t,U}$ will be given by:

$$J_{t,U} = 2[B_0'QB_0 + R]U_t^* \tag{4.13}$$

where U_t^+ is the unimplemented portion of the control action called for at time t, that is:

$$U_t^+ = U_t^* - U_t \tag{4.14}$$

So, when solving equation (4.11) for the optimal control actions which set $J_{t,U}$ to zero, the past values of $J_{t,U}$ may not be zero and are defined by equation (4.13).

To produce a controller which can be implemented express $F^{-1}(z^{-1})$ as the adjoint $F^a(z^{-1})$ divided by the determinant $|F(z^{-1})|$. The determinant is a scalar polynomial and thus it can multiply equation (4.11). Setting $J_{t,U} = 0$ and substituting for past values of $J_{t,U}$ using equation (4.13).

$$G(z^{-1})F^a(z^{-1})Y_t + |F(z^{-1})|B(z^{-1})U_t + C(z^{-1})F^a(z^{-1})Q^{-1}B_0^{-1'}RV^dU_t \tag{4.15}$$
$$+ [I - C(z^{-1})F^a(z^{-1})][B_0 + Q^{-1}B_0^{-1'}R]U_t^+ = 0$$

Equation (4.15) can now be used to calculate the control action U_t^* which would minimize the objective function equation (4.2) if it could be implemented; that is, if no saturation limits are violated it calculates the globally optimal input which may or may not be feasible. The formula is very straightforward involving past values of the process output deviations Y_t, the process inputs that were actually implemented U_t, and the past values of the unimplemented portion of the calculated control action U_t^+. The actually implemented control action can be calculated if the nature of the saturation limits are known, or may be obtained by direct readback from the actuator if this is available.

If the system being controlled is a single-input single-output system then the optimal control input subject to the saturation limits

is either that calculated from equation (4.15) or the limit if a saturation limit is violated. For SISO systems $F^a(z^{-1})=1$ and $|F(z^{-1})|=F(z^{-1})$. If the system being controlled is a multi-input multi-output system then a further correction is required to find the current one-step optimal input subject to the saturation limits. A simultaneous correction for the case where the saturation limits on the current input are of the time varying form $U_{min,t} \leq U_t \leq U_{max,t}$ is derived in section 4.4. The procedure is iterative and converges in at most n steps.

If there is no saturation then U_t^+ will always be zero and equation (4.15) defines the same one-step optimal controller as Koivo (1980) and Mäkilä (1982) which is a generalization of the single-input single-output controller of Clarke and Hastings-James (1971) and Clarke and Gawthrop (1975). Equation (4.11) is a multivariable generalization of the one-step controller with input limits of Goodwin (1972) and Bezanson (1984). It involves rational polynomial matrices in z^{-1} which can be implemented either by expanding as infinite polynomials or by introducing auxiliary states. On the other hand equation (4.15) presented here allows the controller to be implemented easily using a finite number of past vales of Y_t, U_t and U_t^+.

Clarke (1981) presents a univariate saturation correction which involves three equations: a prediction, a control calculation and a revised prediction. Mäkilä's (1982) implementation of the saturation correction of Goodwin (1972) failed to rearrange the rational polynomial matrices correctly as shown in this paper. His algorithm is lacking the correction term involving past U_t^+'s appearing in equation (4.15) here. If the term involving past U_t^+'s in equation (4.15) is ignored then the

controller has reset windup protection since it uses that past implemented control actions in calculating the new control action, but the controller will not return from saturation in an optimal manner.

Although the heuristic anti-reset windup procedure of Banyasz and co-workers (1985) was proposed for SISO PID regulators an extension of it to MIMO model-based controllers can be examined here. The procedure recalculates the deviation from setpoint at saturation using:

$$Y_t(\text{Corrected}) = Y_t + (B_0 + Q^{-1}B_0^{-1'}R)U_t^+$$

Substituting this into equation (4.11) and rearranging, the result is similar to equation (4.15) but the $C(z^{-1})F^{-1}(z^{-1})$ in the last term is replaced by $G(z^{-1})F^{-1}(z^{-1})$. These two terms are only equivalent for step or random walk disturbances.

The controller equation (4.15) can be rewritten in terms of only the past calculated one-step optimal inputs and the past unimplemented part of the control actions:

$$\begin{aligned} G(z^{-1})F^a(z^{-1})Y_t + |F(z^{-1})|B(z^{-1})U_t^* + C(z^{-1})F^a(z^{-1})Q^{-1}B_0^{-1'}R\nabla^d U_t^* \\ - |F(z^{-1})|B(z^{-1})U_t^+ + C(z^{-1})F^a(z^{-1})Q^{-1}B_0^{-1'}R\nabla^d U_t^+ \\ - C(z^{-1})F^a(z^{-1})\,[\,B_0 + Q^{-1}B_0^{-1'}R\,]U_t^+ \qquad = 0 \end{aligned} \tag{4.16}$$

Examination of equation (4.16) shows the different features of this controller. If saturation is ignored completely, that is lines 2 and 3 of equation (4.16) are ignored, then the controller operates only on the past calculated control inputs and it has no reset windup protection. Note that the controller has integral action if $d>0$ as ∇^d is a common factor of the terms multiplying U_t^*. If only line 3 is ignored (equivalent to Mäkilä (1982)), then the controller has reset

windup protection since it uses the past implemented control actions in calculating the new control action, but the controller will not return from saturation in an optimal manner. Also, if line 3 is ignored, it is very important to remove any common left matrix divisors of the equation. This removes any redundant poles and zeroes. If these poles and zeros are not removed they can cause severe oscillations when there is saturation.

There is no guarantee that a one-step optimal controller will be stable, or that the closed loop system using a one-step optimal controller will be stable. For this reason it is necessary to check the poles of the controller and of the closed loop transfer function to ensure stability. The constraint parameter R can be adjusted to obtain stability. The best policy is to ensure both stability and acceptability of the control actions by an appropriate choice of R at the design stage before applying the one-step optimal controller.

The controller equation (4.15) can be interpreted to provide one-step optimal correction for past input saturation for Internal Model Controllers (Garcia and Morari (1982)), Dahlin Controllers (Dahlin (1968)) and pole placement controllers. This will be explained in section (4.6).

The controller with saturation correction derived here provides a closed-form solution to the optimization of the one-step objective function, equation (4.2), taking into account that the past calculated control actions may not have been applied in full. It does not account for whether any future inputs are likely to violate a limit. However, Wong et al (1987) showed by mathematical programming that optimization

with this objective function performed better than any of the anti-reset windup methods which they considered.

4.3 Single-Input Single-Output simulation results

This section presents simulated examples of controller saturation to demonstrate the new saturation correction algorithm. The simulations show both servo and regulatory responses. Because of the duality between controllers designed for setpoint changes or deterministic disturbances with those for regulation of stochastic disturbances (MacGregor et al (1984)), this division is somewhat arbitrary. The setpoint responses show details of the controller action and the process response when a disturbance enters the system. The setpoint examples all use a step change in the setpoint to the controller. The regulatory example treats a nonstationary stochastic disturbance.

The process is an over-damped continuous second order process which is made up of two cascaded first order processes plus dead-time. The gain of the process is 1, the dead-time is 1 minute and the two time constants are 1.44 minutes and 1.09 minutes. The disturbance model is a random walk. With a control interval of 1 minute and a zero order hold between intervals the discrete transfer function is:

$$(1.0 - 0.9\, z^{-1} + 0.2\, z^{-2})\, \nabla\, Y_t = (0.189 + 0.111\, z^{-1})\, \nabla\, U_{t-2} + (1.0 - 0.9\, z^{-1} + 0.2\, z^{-2})\, e_t \qquad (4.17)$$

As defined in equation (4.1), Y_t is a deviation from the set point $Y_{sp,t}$, and U_t is a deviation from the steady state input U_{ss} required to

produce $Y_{sp,t}$. The first controller considered here is the minimum variance controller, which is equivalent to a dead-beat controller for step changes in set point:

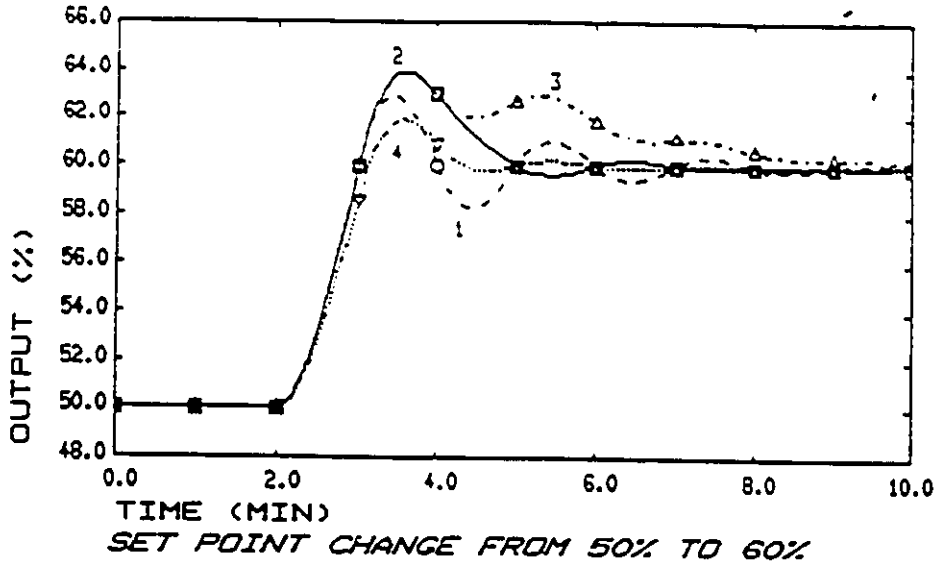
$$(0.189 + 0.3\,z^{-1} + 0.111\,z^{-2})\,\nabla U_t = -(1.0 - 0.9\,z^{-1} + 0.2\,z^{-2})\,Y_t + 0.189\,(-0.9\,z^{-1} + 0.2\,z^{-2})\,U_t^+ \quad (4.18)$$

The response to a step change in the setpoint from 50% to 60% with this controller has already been shown in Figure 4.1 above. In this simulation the initial steady state input is 80%. Figure 4.1 shows the process output and input when there is no saturation limit on the input (labelled 1 in the Figure) to the response when there is a saturation limit at 100% and the algorithm presented here is used (labelled 2 in the figure). In addition, for comparison, the response of the implementation of the controller which simply prevents reset windup by using the past actual values of the U_t in the calculations is shown (labelled 3 in the figure). Lastly, the response of the controller which uses the past calculated U_t^* in the calculations and ignores the saturation altogether, a method with no reset windup protection, is shown (labelled 4 in the figure). The improvement shown by the proposed algorithm over that simply including the usual anti-reset windup features is very obvious as it returns the process to setpoint at the control instants as soon as the input is no longer saturated.

The previous example (with R=0) is not only one-step optimal: it is infinite step optimal. This is because the control input for this simulation hits the saturation limit on the first step. Thus correcting for saturation one step at a time is equivalent to correcting for

saturation for all future steps. In order to understand the difference between what it means to be one-step optimal and infinite step optimal consider the same example except that the steady state input is 10% before the step change is made in the set point. Figure 4.2 shows the process output and the process input responses. The curves are as labelled above, except that the curve labelled 4 shows the true infinite step optimal solution calculated by mathematical programming. The response with no saturation limit on the input (labelled 1) shows that the required input at time 2 is -15.6%. The response using the algorithm presented here (labelled 2) shows that the control input at time 1 is unchanged since the algorithm does not take into account what limits might be encountered in future steps. The saturation on the second step therefore causes the process response to overshoot the set point at time 4. However, thereafter the controller returns the process to set point in an optimal manner. If the standard anti-reset windup protection is used (labelled 3) then the process remains above setpoint for a long time. The infinite step optimal response is shown (labelled 4) for comparison. Because the infinite step controller sees that the saturation limit will be hit at time 2 it chooses a lower control input at time 1. The result is that the process output is slightly below setpoint at time 3, slightly above at time 4, and on set point subsequently. Thus the infinite step controller can improve over the one-step controller by compensating in advance for saturation which will occur in the future.

The constrained one-step optimal controller (with R=0.02) for this same process is:



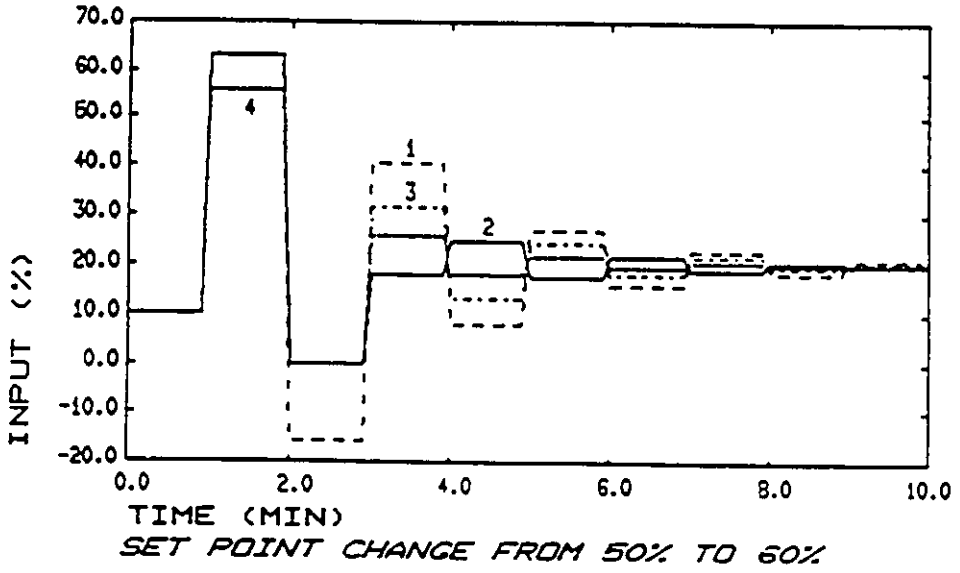


Figure 4.2 Deadbeat of Minimum Variance Control for Second Order System. Step change in setpoint from 50% to 60% at time 1 Minute.
1. No saturation limit.
2,3,4 Saturation limit on input at 0%.
2. Saturation Correction algorithm used.
3. Reset windup protection only, uses past actually implemented Ut.
4. Infinite Step Optimal input.

$$(0.295 + 0.205\ z^{-1} + 0.132\ z^{-2})\ \nabla\ U_t = -\ (1.0 - 0.9\ z^{-1} + 0.2\ z^{-2})\ Y_t + 0.295\ (-\ 0.9\ z^{-1} + 0.2\ z^{-2})\ U_t^+ \qquad (4.19)$$

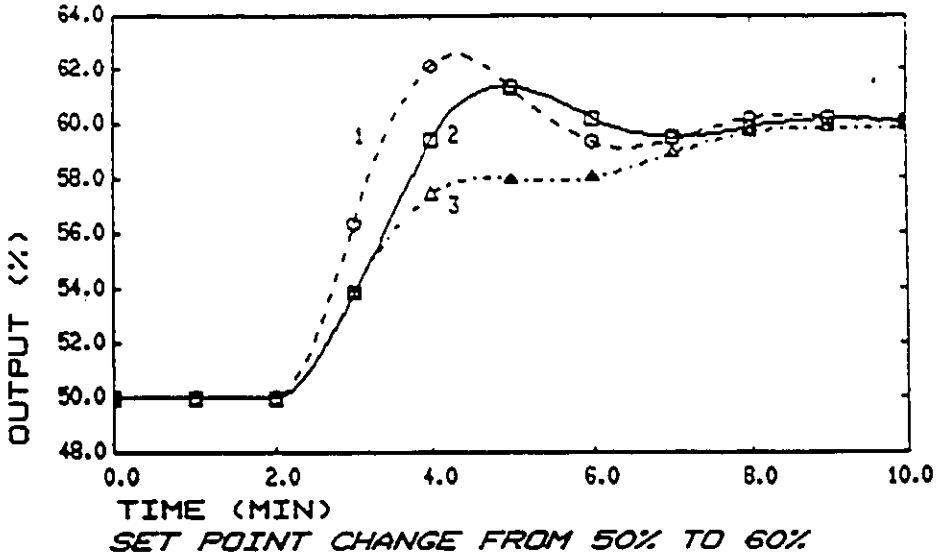
Figure 4.3 shows the process output and the control input using this constrained controller. The curves are labelled as described above. By penalizing the magnitude of the changes in the control actions in the quadratic objective function the initial control action called for is less severe and the saturation effects are also less severe as a result. The process response when there is no saturation (curve 1) does not settle at the set point until near time 10. It is interesting to note that the process response at the control instants with the saturation correction (curve 2) does not match exactly the process response when there is no saturation once the saturation has passed, as was the case for the minimum variance control of Figure 4.1. This is because the penalty on the change in the control actions reduces the effect of zero process output error on the objective function equation (4.2).

Now consider the same process with an ARIMA(0,1,1) disturbance:

$$(1.0 - 0.9\ z^{-1} + 0.2\ z^{-2})\ \nabla\ Y_t = (0.189 + 0.111\ z^{-1})\ \nabla\ U_{t-2} + (1.0 - 0.9\ z^{-1} + 0.2\ z^{-2})\ (1.0 - 0.4\ z^{-1})\ e_t \qquad (4.20)$$

The minimum variance controller for this process, which is equivalent to a Dahlin controller with a specified first order plus dead-time response with a time constant of 1.09 minutes (i.e. a discrete time constant of 0.4), is:

$$(0.189 + 0.224\ z^{-1} + 0.067\ z^{-2})\ \nabla_t U = -\ (0.6 - 0.54\ z^{-1} + 0.12\ z^{-2})\ Y_t + 0.189\ (-\ 1.3\ z^{-1} + 0.52\ z^{-2} - 0.08\ z^{-3})\ U_t^+ \qquad (4.21)$$



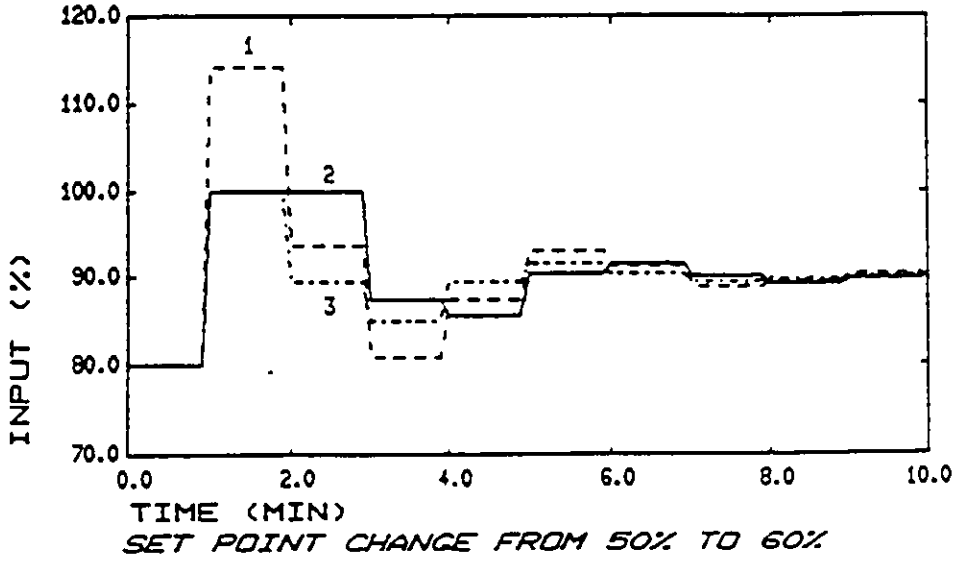
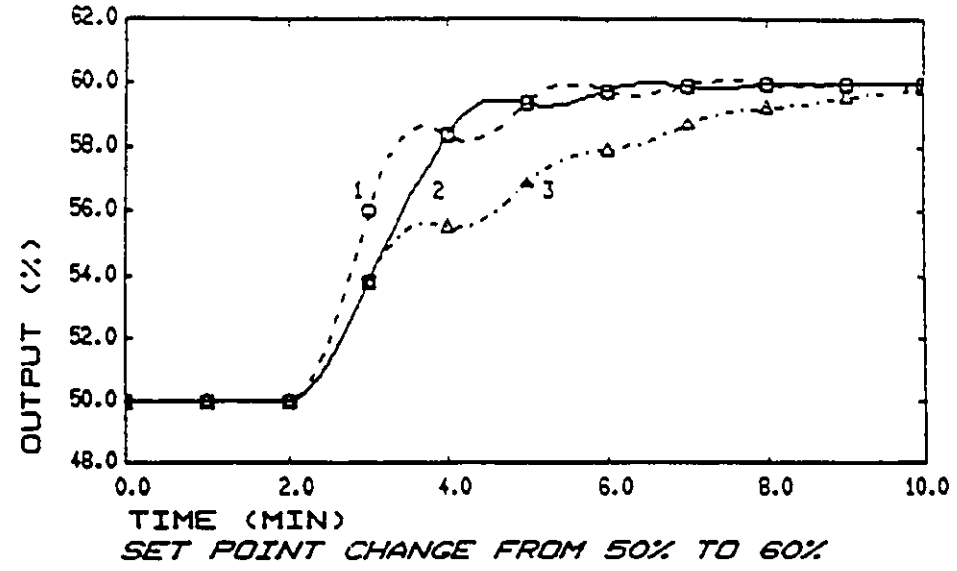


Figure 4.3 One-Step Optimal Control, R-0.02.
Step change in setpoint from 50% to 60% at time 1 Minute.
1. No saturation limit.
2,3 Saturation limit on input at 100%.
2. Saturation Correction algorithm used.
3. Reset windup protection only, uses past actually implemented Ut.

In Figure 4.4 the Dahlin algorithm with no saturation limits (curve 1) can be seen to yield the desired first order response at the control intervals, although there is a substantial amount of intersample ripple. In the presence of an input saturation limit the Dahlin controller is unable to achieve the specified first order response on the first interval. If no saturation correction other than reset windup protection is used, the subsequent response of the Dahlin controller is severely degraded (curve 3). However, using the one-step optimal correction presented here (curve 2) the improvement is very significant. The process output is returned to the desired first order trajectory at the control instants as soon as the input is no longer limited.

This same minimum variance controller is also applied to regulation of the ARIMA(0,1,1) disturbance for which it was designed. The figures show the process output, control input and the cumulative performance index ΣY_t^2 (recall that Y_t has been defined as the deviation from set point $Y_{sp,t}$). The process output setpoint is 70% and the process input steady state is 90%. Figure 4.5 shows the process output with no control (curve 1) compared to the minimum variance control (curve 2). This figure shows the nature of the disturbance and the control inputs which are required to keep the process on set point. Naturally, minimum variance control reduces dramatically the cumulative performance index.

Figure 4.6 shows the response when there is a saturation limit at 100%. The response using the saturation correction presented here (labelled 1) is compared to the response using minimum variance control with reset windup protection only (labelled 2). The cumulative



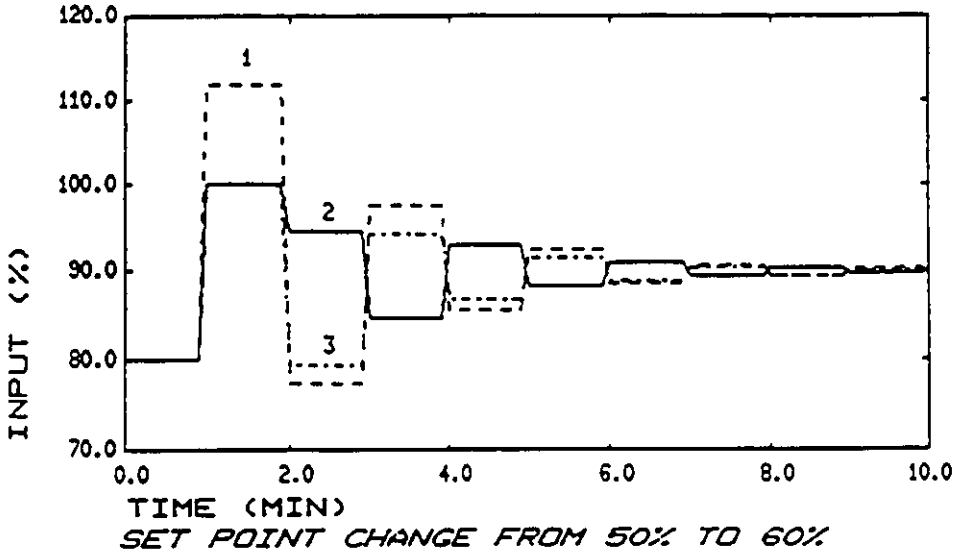
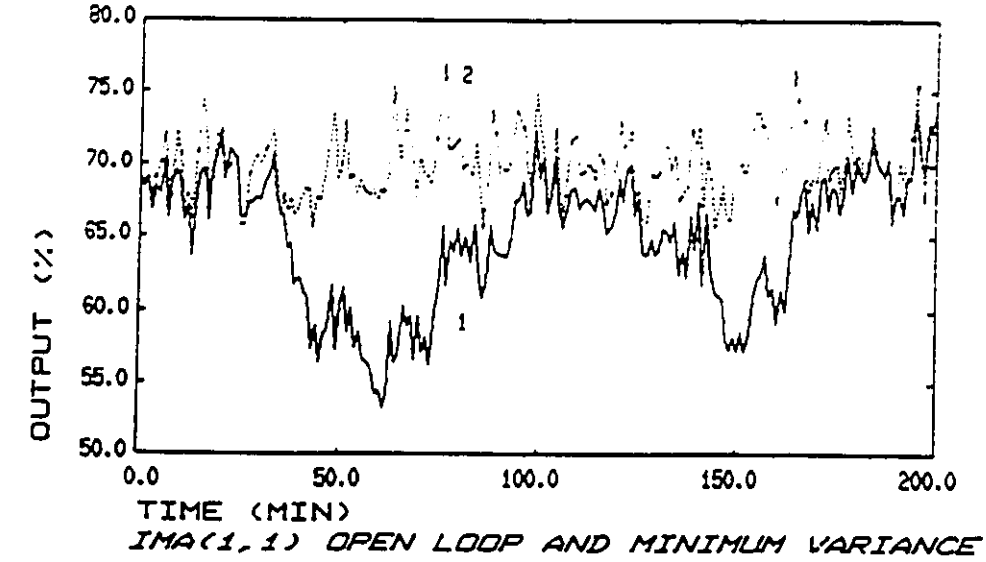
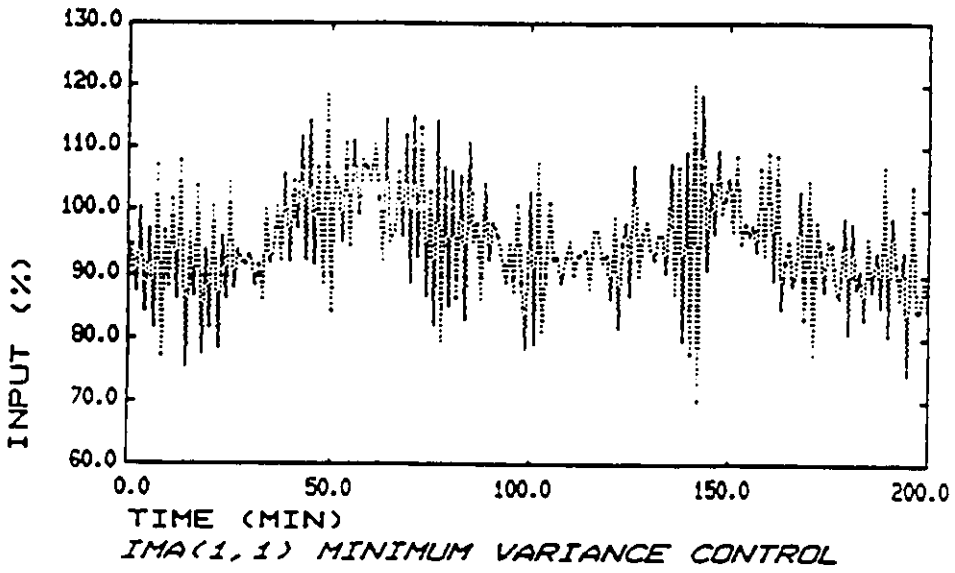


Figure 4.4 Dahlin Controller.
Step change in setpoint from 50% to 60% at time 1 Minute.
1. No saturation limit.
2,3 Saturation limit on input at 100%.
2. Saturation Correction algorithm used.
3. Reset windup protection only, uses past actually implemented Ut.





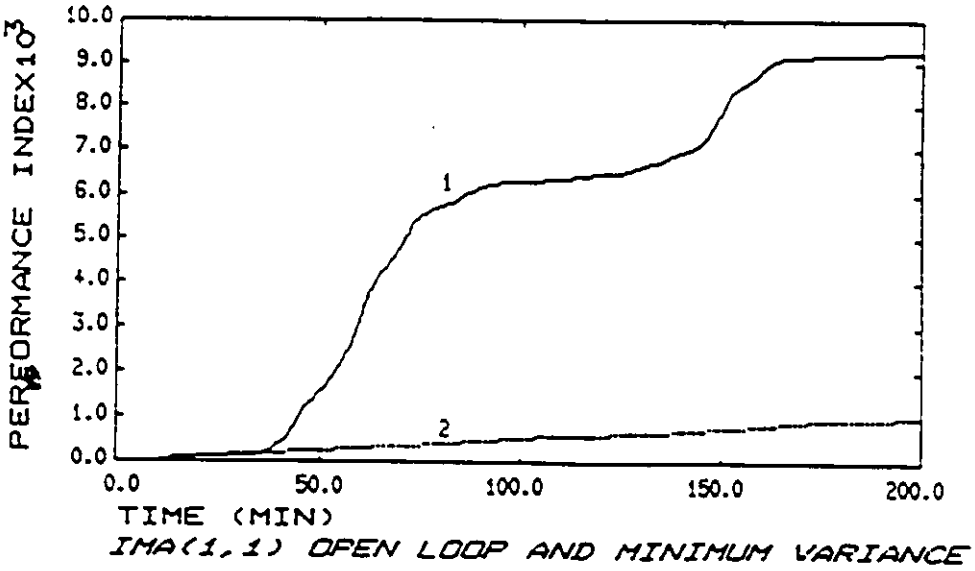
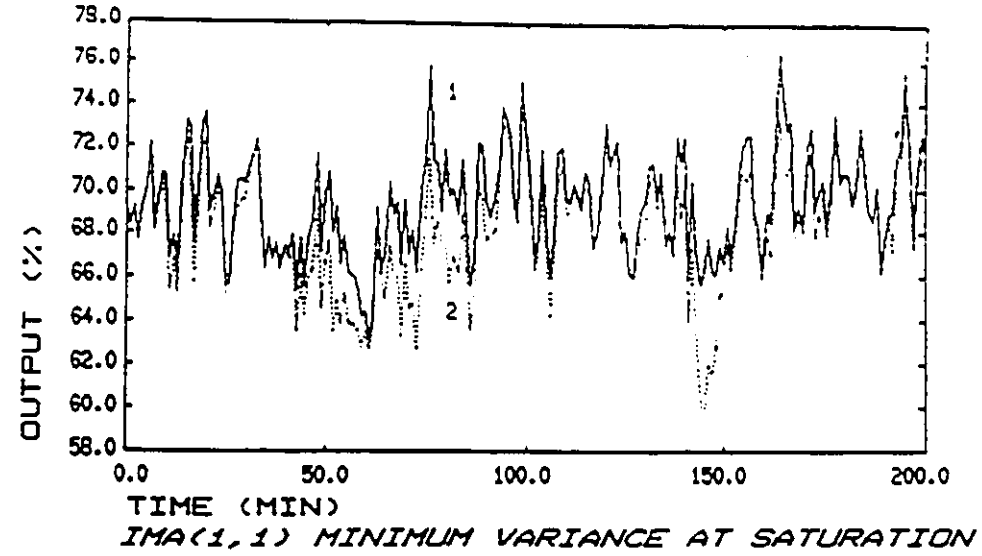
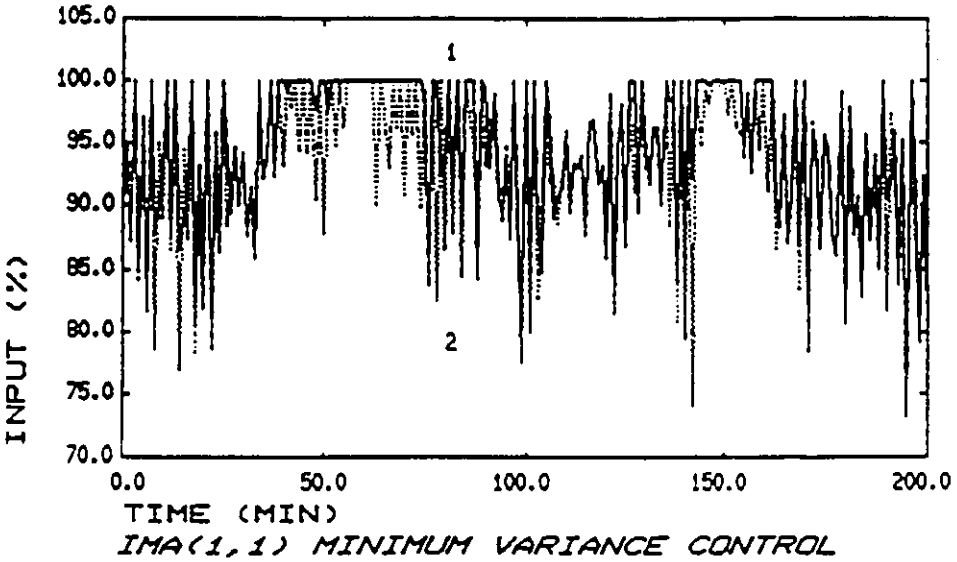


Figure 4.5 Open Loop IMA(1,1) disturbance and Minimum Variance Control
1. Open Loop Disturbance
2. Minimum Variance Control





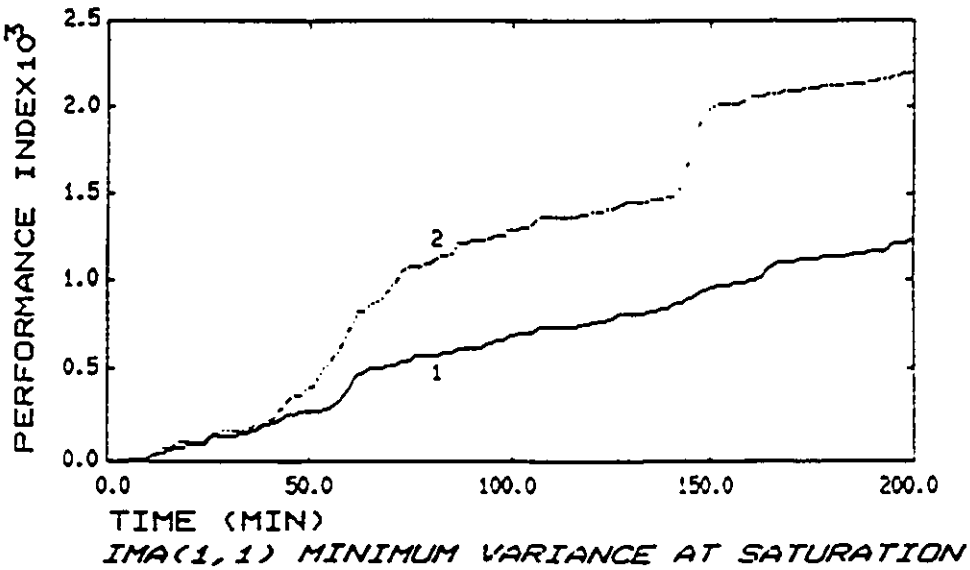


Figure 4.6 Minimum Variance Control of IMA(1,1) disturbance.
Saturation limit at 100%.
1. Saturation Correction algorithm used.
2. Reset windup protection only, uses past actually implemented Ut.

performance index using the one-step optimal saturation protection is only 60% of that using the standard anti-reset windup.

4.4 Simultaneous correction for Multi-Input Multi-Output systems

Equation (4.15) in section 4.2 describes the saturation correction to the controller transfer function using past unimplemented control actions U_t^+. When any inputs in a multi-input multi-output system saturate then a further simultaneous correction is required to calculate the one-step optimal settings of the control inputs subject to saturation limits.

Consider equation (4.11) rearranged as in the original derivation but without setting $J_{t,U}=0$ giving:

$$\begin{aligned} Q^{-1}B_0^{-1'}J_{t,U} = \; & 2G(z^{-1})F^a(z^{-1})Y_t + 2[F(z^{-1})]B(z^{-1})U_t \\ & + 2C(z^{-1})F^a(z^{-1})Q^{-1}B_0^{-1'}R\nabla^d U_t \\ & + 2[I - C(z^{-1})F^a(z^{-1})][B_0 + Q^{-1}B_0^{-1'}R]U_t^+ \end{aligned} \tag{4.22}$$

where $J_{t,U}$ is the gradient of the positive definite quadratic objective function equation (4.2). This equation can be rewritten with obvious substitutions as:

$$J_{t,U} = \alpha(z^{-1})Y_t + \beta(z^{-1})U_t + \gamma(z^{-1})U_t^+ \tag{4.23}$$

The present and past values of Y_t, and all of the past values of U_t and U_t^+ are known so their contribution to the gradient can be lumped together as K_t, giving:

$$J_{t,U} = \beta_0 U_t^* + K_t \tag{4.24}$$

When there is no saturation the optimal control inputs are calculated so as to set the gradient of the objective function to zero:

$$U_t^* = -\beta_0^{-1} K_t \tag{4.25}$$

Because the objective function equation (4.2) is positive definite and quadratic we know that equation (4.25) defines a globally optimal input.

The saturation limits are assumed to be simple, though possibly time varying, bounds:

$$U_{min,t} \leq U_t \leq U_{max,t} \tag{4.26}$$

which define a convex box shape. In order to find the constrained optimum an iterative procedure for determining active constraints is necessary. In order to determine which of the constraints of equation (4.26) violated by the solution of equation (4.25) will be active the gradient must be evaluated using equation (4.23) at the point $U_t = sat\{U_t^*\}$. If the component of the gradient is in the direction of the constraint then the constraint will be active; if the component of the gradient is into the feasible region it will be inactive. This simple active set strategy is due to the constraints being bounds on the inputs (Gill et al (1981)).

If one or more constraints are active then those components of the input are fixed at their limits and the corresponding components of the $J_{t,U}$ will not be zero at the next calculated value of the input U_t. To solve for the remaining unconstrained components of U_t and the nonzero components of $J_{t,U}$ equation (4.24) must be rearranged to collect these terms on the left hand side.

For example, if only the ith input variable will be saturated, we can rearrange equation (4. 4) as:

$$[\beta_0^1|\ldots|e_i|\ldots|\beta_0^n]\begin{bmatrix} U_t^1 \\ \vdots \\ J_{t,U}^i \\ \vdots \\ U_t^n \end{bmatrix} - \begin{bmatrix} K_t^1 \\ \vdots \\ K_t^i \\ \vdots \\ K_t^n \end{bmatrix} - \beta_0^i \, U_{t,Limit}^i \tag{4.27}$$

where β_0 is the ith column of β_0 and U_t^i, $J_{t,U}^i$ and K_t^i are the ith components of U_t, $J_{t,U}$ and K_t respectively, e_t is the ith unit vector and $U_{t,Limit}^i$ is the saturated value of the ith component of the input. Equation (4.27) can be solved for the remaining components of the input U_t and the free component of the gradient $J_{t,U}^i$. When more than one component of the input is saturated the rearrangement shown in equation (4.27) is performed for each component. For computational purposes the inverse of the β_0 matrix can be updated using a rank one update formula for each saturated component of the input. The control inputs calculated from equation (4.27) now become the new iterate for U_t.

The algorithm for finding the constrained optimal input is then:

(i) Calculate the globally optimal control input U_t^* using equation (4.25).

(ii) If any of the components of U_t^* violate their saturation limits set those components to their limiting values and calculate the gradient of the objective function at that point using equation (4.24).

(iii) For each component of the gradient which is in the direction of the constraint, that constraint is active.

(iv) If all of the constraints are active, or if no inactive constraint becomes active and if no active constraint becomes inactive then the constrained optimum has been found.

(v) For each active constraint set the input corresponding to that component to its saturation limit $U^i_{t,Limit}$ and rearrange equation (4.24) as equation (4.27).

(vi) Solve equation (4.27) for the free components of the input and the free components of the gradient.

(vii) Go to step (ii). This algorithm converges in at most n steps.

The Q and B_0 matrices heavily influence the nature of the simultaneous correction. The more interactive the B_0 matrix is, the larger the effect of the simultaneous correction will be. Although the Q matrix has no effect on the Minimum Variance Controller (R=0) of equation (4.15) it can have a very significant effect at saturation when the forecast of the output cannot be cancelled exactly. The weights in the Q matrix determine the relative importance of the various output deviations when calculating the simultaneous correction.

4.5 Multi-Input Multi-Output simulation results

This section presents simulated examples of controller saturation in a multivariable system to demonstrate the corrections for past saturation and the simultaneous saturation correction. The simulations show the servo response to setpoint changes under a variety of saturation conditions.

The system is an experimental two-tank with heater system as shown in Figure 4.7. The state space model for the system is (Hugo (1986)):

$$\begin{bmatrix} \frac{d}{dt} T_T(t) \\ \frac{d}{dt} T_B(t) \end{bmatrix} = \begin{bmatrix} -0.0121 & 0.0527 \\ 0.0208 & -0.0208 \end{bmatrix} \begin{bmatrix} T_T(t) \\ T_B(t) \end{bmatrix} + \begin{bmatrix} 0.0113 & 0.0 \\ 0.0 & 0.00797 \end{bmatrix} \begin{bmatrix} U_T(t) \\ U_B(t) \end{bmatrix} \tag{4.28}$$

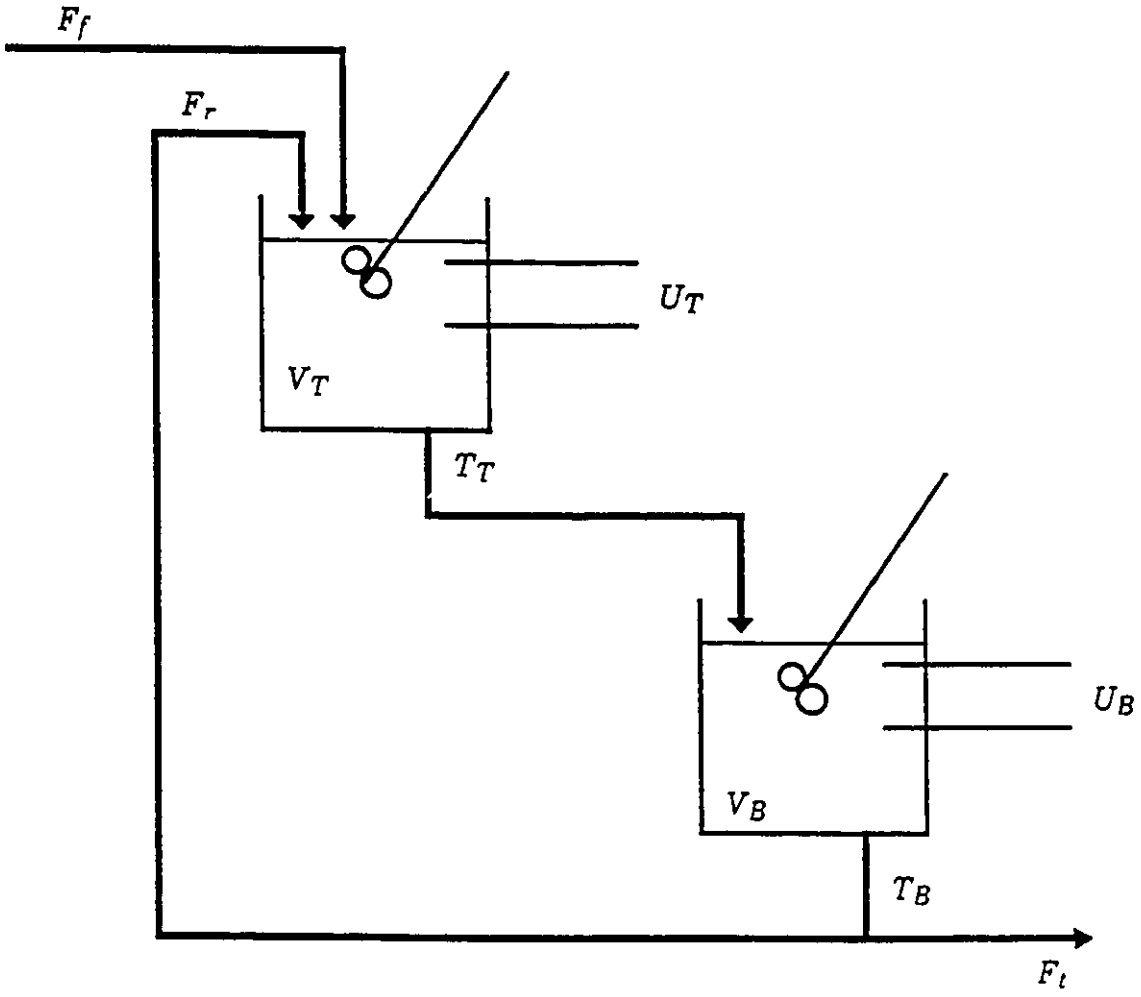
where T represents the temperature in degrees Celsius and U refers to the heat input in Volts at the computer output (range 0-10V) and the time T is in seconds. The subscripts T and B refer to the top and bottom tanks respectively. The steady state values are T_T=30°C, T_B=31°C, U_T=4.5V and U_B=5.5V. The discrete transfer function model based on a 20 second control time is and a random walk disturbance is:

$$A(z^{-1}) = \begin{bmatrix} (1.0 - 1.4764\, z^{-1} + 0.5179\, z^{-2})\,\nabla & 0. \\ 0. & (1.0 - 1.4764\, z^{-1} + 0.5179\, z^{-2})\,\nabla \end{bmatrix} \tag{4.29}$$

$$B(z^{-1}) = \begin{bmatrix} (0.20206 - 0.1335\, z^{-1})\,\nabla & (0.0067432 + 0.0054167\, z^{-1})\,\nabla \\ (0.036021 + 0.030541\, z^{-1})\,\nabla & (0.1303 - 0.1024\, z^{-1})\,\nabla \end{bmatrix}$$

$$C(z^{-1}) = \begin{bmatrix} 1.0 - 1.4764\, z^{-1} + 0.5179\, z^{-2} & 0. \\ 0. & 1.0 - 1.4764\, z^{-1} + 0.5179\, z^{-2} \end{bmatrix} \tag{4.29}$$

The multivariable minimum variance or dead-beat controller of equation (4.15) is, after removing the common left matrix divisor:



$F_f = 6.31 \times 10^{-5}\ m^3/s$

$F_r = 4.90 \times 10^{5}\ m^3/s$

$V_T = 9.29 \times 10^{-3}\ m^3$

$V_B = 5.39 \times 10^{-3}\ m^3$

Figure 4.7 Experimental two tank with heater system.

$$\begin{bmatrix} 1.0 - 1.4764\, z^{-1} + 0.5179\, z^{-2} & 0. \\ 0.0 - 6.2590\, z^{-1} + 4.5700\, z^{-2} & 1. \end{bmatrix} Y_t + \tag{4.30}$$

$$\begin{bmatrix} 0.20206 - 0.1335\, z^{-1} & 0.0067432 + 0.0054167\, z^{-1} \\ 0.038021 - 1.1780\, z^{-1} & 0.1303 - 0.04780\, z^{-1} \end{bmatrix} \bar{v}\, U_t -$$

$$\begin{bmatrix} 0.20201 - 0.2983\, z^{-1} + 0.1046\, z^{-2} & 0.006743 - 0.009956\, z^{-1} + 0.003492\, z^{-2} \\ 0.03802 - 1.2647\, z^{-1} + 0.9234\, z^{-2} & 0.1303 - 0.04221\, z^{-1} + 0.03082\, z^{-2} \end{bmatrix} U_t^+ = 0$$

The controller has real zeros at z = 0.5736 and z = 0.9028 and two real poles at z = 1.0.

The response of the system with this controller to a setpoint change down from 30°C to 28.5°C in the top tank is shown in Figure 4.8. The response with no saturation limit on the heater input is labelled 4. The response using the complete algorithm described here is labelled 1, the response using the correction for past saturation but not the simultaneous correction is labelled 2, and the response using only reset windup protection is labelled 3. The saturation limits on the heater inputs are a low limit of 0V and a high limit of 10V. The complete one-step optimal controller is seen to bring the temperature in the top tank to setpoint in one control time after the saturation has ended while keeping the error equal to zero at the sampling times in the bottom tank. When the simultaneous correction is not used the temperature in the bottom tank drifts only slightly off setpoint. When the saturation correction algorithm is not used both the top and bottom tank temperatures are seen to drift away form set point for a prolonged period of time.

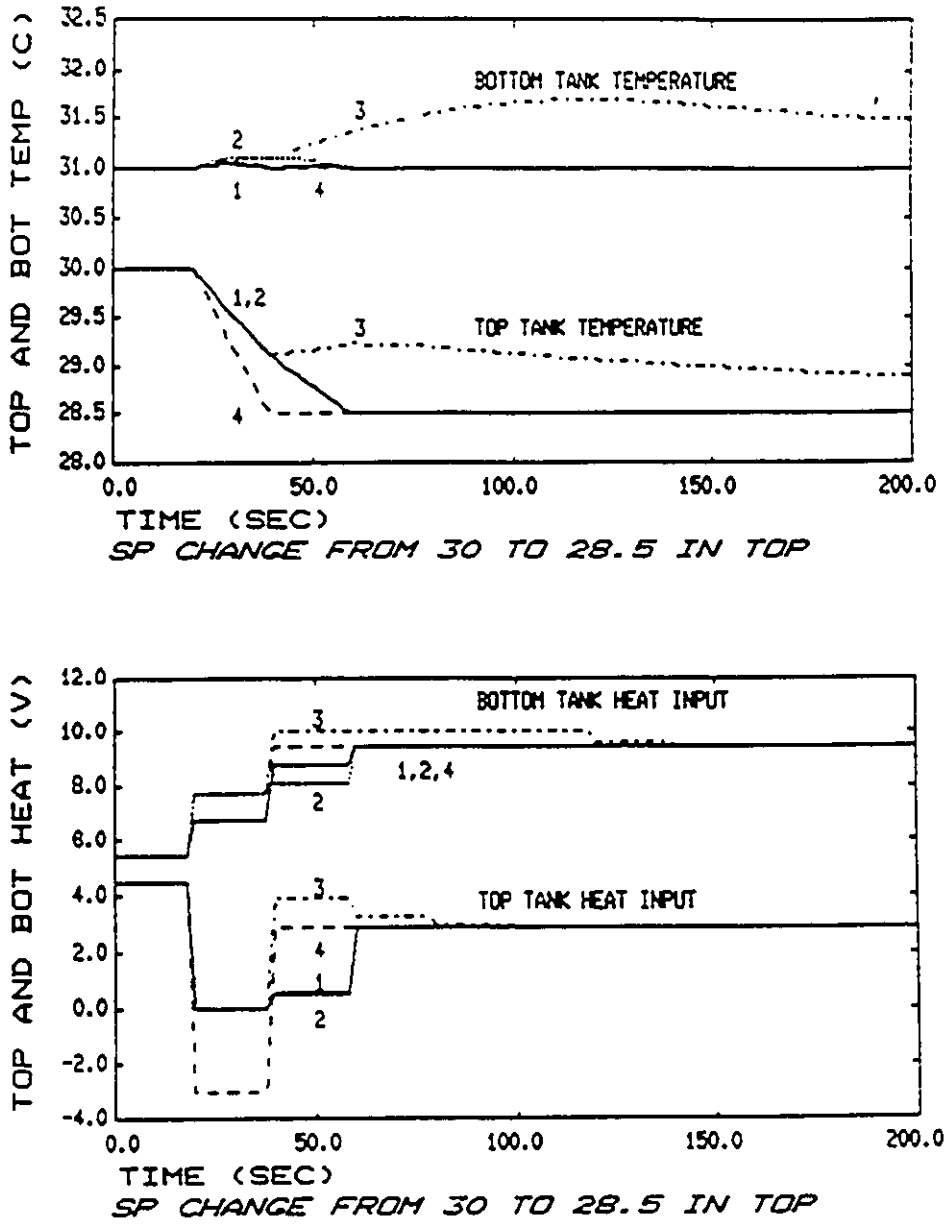
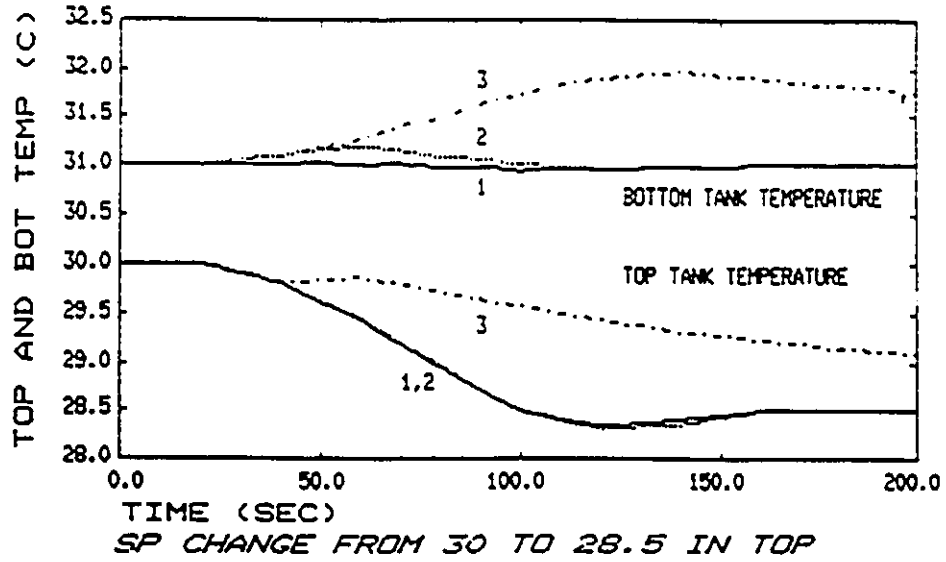


Figure 4.8 Multivariable Deadbeat or Minimum Variance Control for Two Tank with Heater System.
Step change in top tank temperature setpoint from 30°C to 28.5°C at time 20 sec.
1,2,3 Saturation limit on both inputs at 0V and 10V.
1. Saturation Correction algorithm with Simultaneous Correction.
2. Saturation Correction without Simultaneous Correction.
3. Reset windup protection only, uses past actually implemented Ut.
4. No Saturation limits.

Figure 4.9 repeats the previous example but an additional saturation constraint is added: the change in the heater input is restricted to 1.0V per control interval. The one-step optimal controller again gives superior performance and the simultaneous correction is seen to have little effect.

Figure 4.10 shows the response to a prolonged period of saturation. The set point of the bottom tank temperature is changed to the unattainable value of 34°C in the bottom tank. At time 400 seconds, the set point is changed to the attainable value of 32.5°C. The saturation limits are at the high and low values of 10V and 0V. The simultaneous correction is seen to prevent offset in the top tank temperature. The one-step optimal correction without the simultaneous correction allows a slight offset in the top tank temperature for the period of saturation. When reset windup protection only is used, both tank temperatures show poor performance even long after the second set point change.

This last example also shows where an infinite step optimal solution could do better than the one-step optimal solution. Careful consideration of Figure 4.10 shows that if the top tank heat input were increased it would cause some increase in the top tank temperature, but this would eventually bring the bottom tank temperature an equivalent amount closer to set point. However, the one-step optimal solution judges interactions only from the B_0 matrix which shows that the top tank heat input has far less effect on the bottom tank temperature than on the top tank temperature after one step!



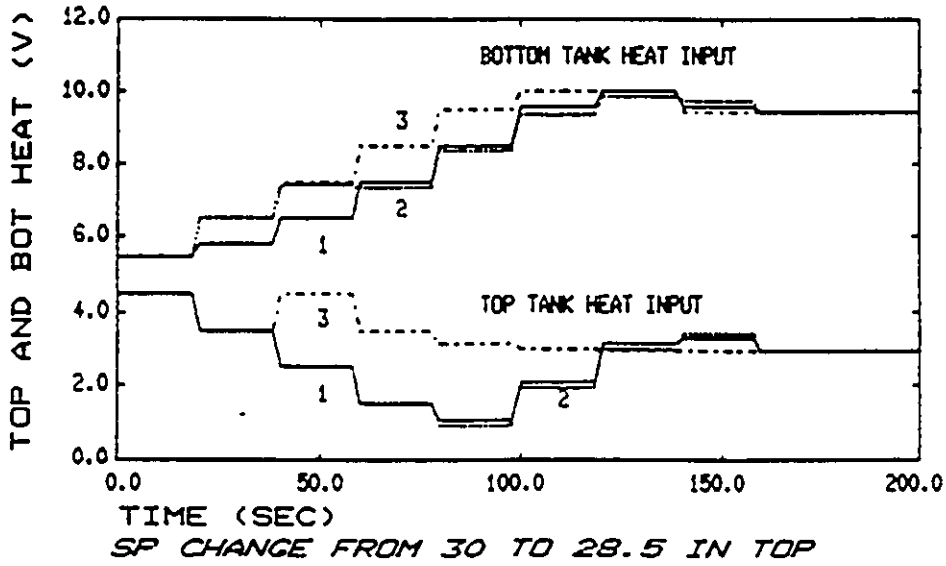


Figure 4.9 Multivariable Deadbeat or Minimum Variance Control for Two Tank with Heater System.
Step change in top tank temperature setpoint from 30°C to 28.5°C at time 20 sec.
1,2,3 Saturation limit on both inputs at 0V and 10V and velocity limits ±1V.
1. Saturation Correction algorithm with Simultaneous Correction.
2. Saturation Correction without Simultaneous Correction.
3. Reset windup protection only, uses past actually implemented Ut.

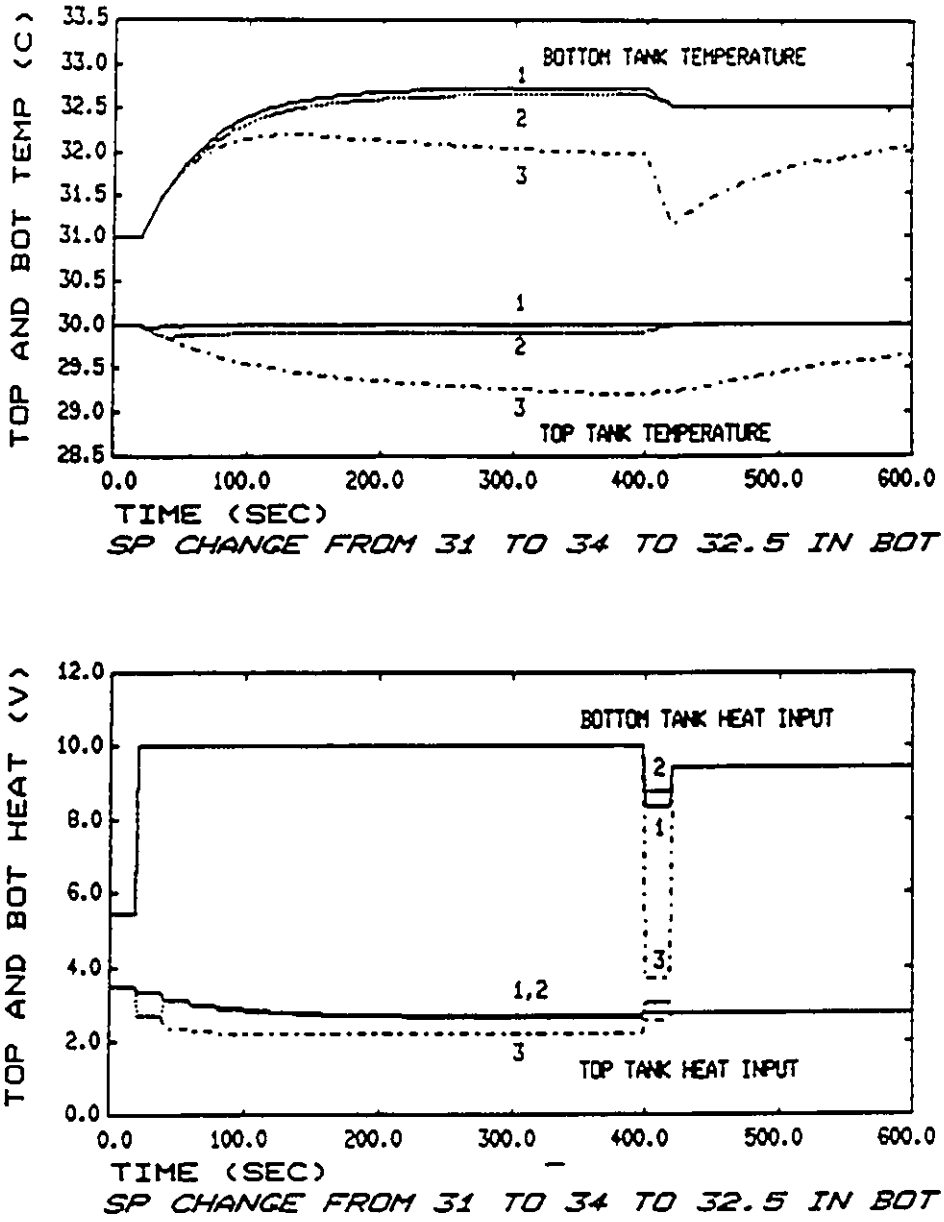


Figure 4.10 Multivariable Deadbeat or Minimum Variance Control for Two Tank with Heater System.
Step change in bottom tank temperature setpoint from 30°C to 28.5°C at time 20 sec.
Step change in bottom tank temperature setpoint from 34°C to 32.5°C at time 20 sec.
1,2,3 Saturation limit on both inputs at 0V and 10V.
1. Saturation Correction algorithm with Simultaneous Correction.
2. Saturation Correction without Simultaneous Correction.
3. Reset windup protection only, uses past actually implemented Ut.

4.6 Extension to other model based controllers

In this section the extension of the one-step optimal saturation correction to other model based controller design techniques is examined. In the examples above it was shown that the one-step optimal saturation correction also applies to dead-beat and Dahlin controllers. This result was expected because of the equivalence between minimum variance controllers for stochastic disturbances and controllers for deterministic disturbances (MacGregor et al (1984)). The examples showed that the correction for past saturation acts to return the forecast of the errors to zero as soon as possible after the saturation stops. That is, the forecast of the errors will be cancelled f+1 control intervals after saturation ends. This is equivalent to saying that the saturation compensation algorithm returns the output behaviour to that defined by the closed loop transfer function of the unrestricted system as soon as possible.

The multivariable minimum variance controller (i.e. R=0) for an invertible process (i.e. the zeroes of $\det(B(z^{-1}))$ are all inside the unit circle) has a closed loop transfer function between the setpoint and the process output given by:

$$G_{CLTF}(z_{-1}) = z^{-f-1}F(z^{-1})C^{-1}(z^{-1})G(z^{-1})F^{-1}(z^{-1}) \qquad (4.31)$$

The closed-loop transfer function depends only on the disturbance model transfer function:

$$G_D(z^{-1}) = A^{-1}(z^{-1})C(z^{-1}) \qquad (4.32)$$

and its forecast function from equation (4.4). This is not surprising as the minimum variance controller calculates the control inputs to set the forecast of the output deviations from setpoint to zero.

If the disturbance model transfer function is considered to be a design tool and is chosen as:

$$G_D(z^{-1}) = \nabla^{-1}C_D(z^{-1}) \tag{4.33}$$

where $C_D(z^{-1}) = \text{diag}(1.0 - C_i(z^{-1}))$ (which is the model of a first order integrated moving average process) then the closed loop transfer function of the resulting minimum variance controller is:

$$G_{CLTF}(z^{-1}) = z^{-f-1}C_D^{-1}(z^{-1})G_D(z^{-1}) \tag{4.34}$$

where $G_D(z^{-1}) = \text{diag}(1.0 - C_i)$ which is a multivariable first order exponential filter with the C_i's as time constants.

So, by choosing $G_D(z^{-1})$ as in Equation (4.33), the minimum variance design yields a controller with a first order exponential filter closed-loop transfer function, that is, a Dahlin controller. Based on this equivalence, equation (4.15) can be used for one-step optimal saturation compensation of a multivariable Dahlin controller.

In the IMC controller design (Garcia and Morari (1982)) the closed loop transfer function is the same first order process as in the Dahlin controller; however, the dead-time compensation is done explicitly as shown in the block diagram Figure 4.11. This controller still requires saturation compensation but the form is slightly different than for the normal minimum variance controller. In the IMC structure the disturbance is calculated as:

$$N_t = Y_t - A^{-1}(z^{-1})B(z^{-1})U_{t-f-1} \tag{4.35}$$

So, we can substitute for e_t in equation (4.7) using:

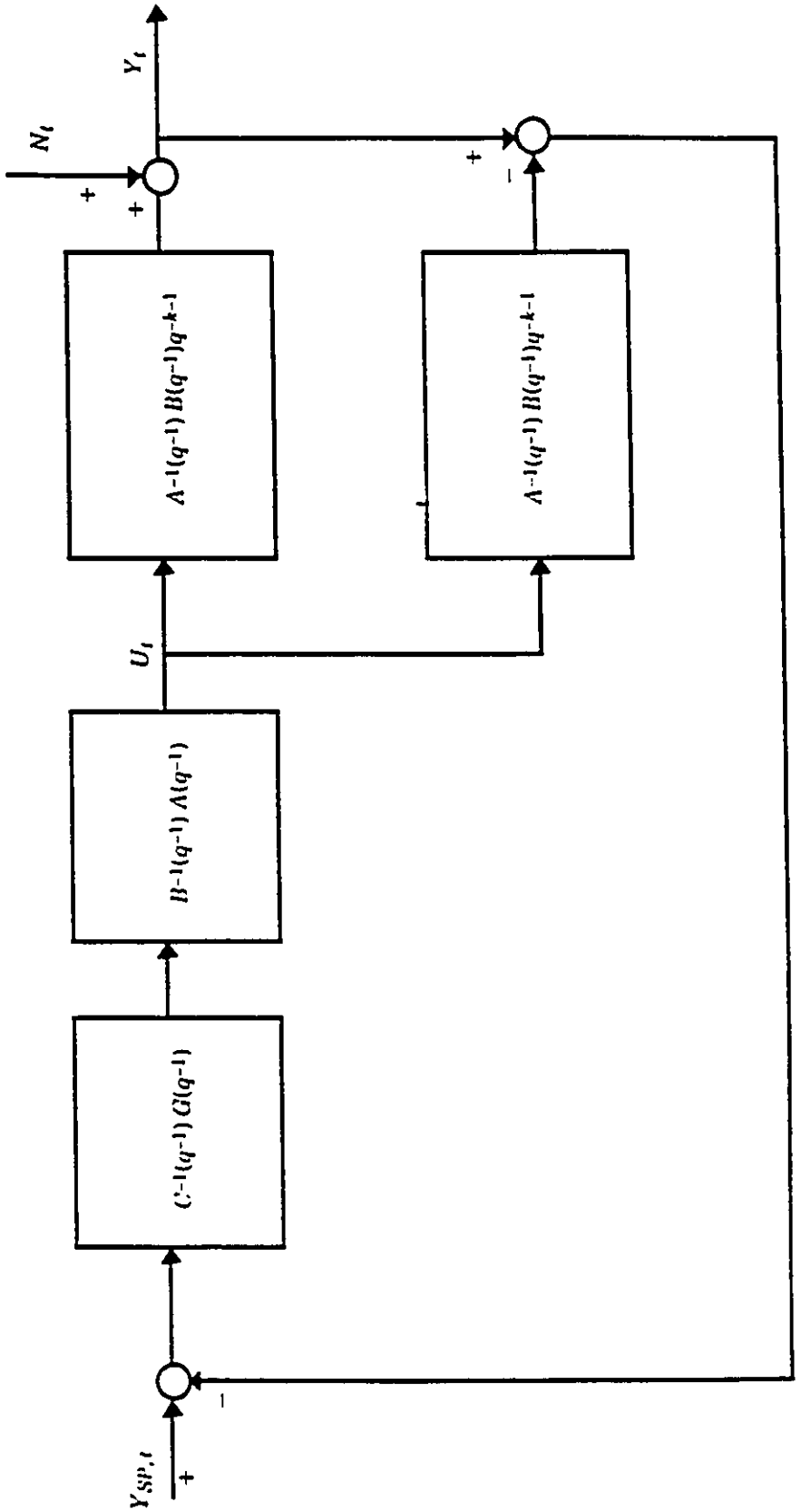


Figure 4.11 Block Diagram of Minimum Variance Controller In Internal Model Control Structure.

$$e_t = C^{-1}(z^{-1})A(z^{-1})N_t \tag{4.36}$$

Also, for the minimum variance design R=0, so equation (4.7) becomes:

$$J_{t,U} = 2B_0'Q\ [A^{-1}(z^{-1})B(z^{-1})U_t + A^{-1}(z^{-1})G(z^{-1})C^{-1}(z^{-1})A(z^{-1})N_t\ ] \tag{4.37}$$

Following the derivation of section 4.2 equation (4.13) still holds. So, the IMC controller equation with compensation for past saturation becomes:

$$A(z^{-1})Q^{-1}B^{-1}{}'J_{t,U} = G(z^{-1})C^{-1}(z^{-1})A(z^{-1})N_t + B(z^{-1})U_t \tag{4.38}$$

which is then implemented using equation (4.13) as:

$$G(z^{-1})C^a(z^{-1})A(z^{-1})N_t + |C(z^{-1})|B(z^{-1})U_t + [I - |C(z^{-1})|A(z^{-1})]B_0U_t^+ = 0 \tag{4.39}$$

Equation (4.38) shows how the one-step optimal correction for past saturation applies to an IMC controller.

The Dynamic Matrix controller of Cutler and Ramaker (1976) uses a particular model form where $A(z^{-1}) = \nabla$ and $C(z^{-1}) = I$ (so $G(z^{-1}) = I$). Equation (4.38) is then implemented as:

$$N_t + A^{-1}(z^{-1})B(z^{-1})U_t = 0 \tag{4.40}$$

and no correction for past saturation is required, so long as the past actually implemented control actions are used in calculating the current control action.

The derivations above show how to correct for past saturation of the control inputs in the Dahlin, IMC and DMC controllers. However, it is also important to note that a simultaneous correction for saturation may be required.

4.7 Summary

This chapter has presented a method for compensating for input saturation in one-step optimal constrained minimum variance controllers. The algorithm is straightforward to implement as a correction operating on the difference between the past calculated control inputs and the inputs which were actually implemented. A further iterative correction for multivariable controllers is presented to simultaneously adjust the remaining control inputs if some inputs saturate. Simulations are presented to show the effect of the saturation correction. Although this correction is derived for minimum variance and one-step optimal constrained minimum variance controllers, the application to Internal Model, Dynamic Matrix and Dahlin controllers is also explained.

CHAPTER 5

SATURATION IN VELOCITY FORM PROPORTIONAL INTEGRAL DERIVATIVE CONTROLLERS

The Proportional Integral Derivative controller is commonly used in the computer control of processes. Controllers sometimes operate near actuator saturation limits out of necessity. The fact that Proportional Integral Derivative controllers written in velocity form can be subject to severe performance degradation when saturation is encountered has only rarely been observed in the literature (Åström and Wittenmark (1984)). This chapter explains the nature of the problem, how it arises from the mathematics of the velocity form and how to fix it. The problems encountered in the tuning of velocity form controllers are also explained.

In addition two solutions to the problem of reset windup protection in PID controllers are presented: one for the position form and one for the velocity form. The position form correction keeps integrating until saturation is due to integral action. The velocity form correction preserves the integral and derivative action while preventing reset windup. These methods of reset windup protection are compared to the method of Smith and Corripio (1985) and the one-step

optimal methods of the previous chapters. Examples are used to show the improvement which is obtained from the proposed solutions. This work has been published as a journal article (Segall and Taylor (1986)).

5.1 Discrete PID Controllers

The discrete approximation to the continuous PID controller is generally written in Position form as (see for example Åström and Wittenmark (1984)):

$$U_t = K\,[Y_t + T_I T_c \sum_{i=1}^{t} Y_i + T_D(Y_t - Y_{t-1})/T_c] + U_{SS} \qquad (5.1)$$

where U_t is the current control action, Y_t is current deviation from the set point $Y_{sp,t}$, K is the controller gain, T_I is the reset rate, T_D is the derivative time, T_c is the control interval, and U_{SS} is the control action when the controller was turned on.

The Velocity form of the controller is often used because it does not need the summation associated with integration. It is obtained by subtracting the Position form, equation (5.1) above, written at time t-1 from that written at time t to give:

$$\nabla U_t = U_t - U_{t-1} = K\,[(Y_t - Y_{t-1}) + T_I T_c Y_t + T_D(Y_t - 2Y_{t-1} + Y_{t-2})/T_c] \qquad (5.2a)$$

and the actuator position is calculated from:

$$U_t = U_{t-1} + \nabla U_t \qquad (5.2b)$$

The equivalence between equation (5.2) and equation (5.1) depends on being able to write (5.2b) as:

$$U_t = U_{SS} + \nabla U_1 + \nabla U_2 + \ldots + \nabla U_{t-1} + \nabla U_t = U_t + \sum_{i=1}^{t} \nabla U_i \qquad (5.3)$$

Substituting for U_t from equation (5.2a) into equation (5.3) produces collapsing sums for the Proportional and Derivative terms and a summation for the Integral term:

$$U_t = U_{SS} + K \sum_{i=1}^{t} (Y_i - Y_{i-1}) + K T_I T_c \sum_{i=1}^{t} Y_i + K T_D \sum_{i=1}^{t} (Y_i - 2Y_{i-1} + Y_{i-2})/T_c \qquad (5.4)$$

where it is assumed that $Y_0=Y_{-1}=0$.

Equation (5.4) will only reduce to equation (5.1) if none of the terms in the collapsing sums are discarded. Therefore, the equivalence of the Velocity form with the Position form via equation (5.4) fails when any of the ∇U_t terms are discarded because of saturation.

Discarding any portion of the control action increment ∇U_t which would cause the output to exceed the saturation limits is common practice in implementing velocity form PID controllers. This is because although it affects the proportional and derivative action, it also limits the summation of the integral action. This is important to prevent reset windup. That is why in the previous chapter, discarding the increments which would cause the output to exceed the saturation limits is referred to as having reset windup protection.

Another discrepancy between the Velocity and Position forms of the PID controller can occur during tuning. Consider proportional only control, so equation (5.2a) reduces to a proportional gain multiplying the difference between the error at time t and time t-1. When steady state is reached there is the inevitable offset, so the error does not go to zero. However, the difference between successive values of the process deviation from set point is zero. Changes in the proportional gain made at this point do not alter the offset because the difference in the errors remains zero. The Position form, equation (5.1), will

respond to gain changes in this situation because the current value of the error is not zero. This anomaly can cause confusion when one is tuning a Velocity form PID controller.

The velocity form can be corrected for tuning changes if the previous values of the tuning parameters K and T_D are retained:

$$\nabla U_t = K_t Y_t - K_{t-1} Y_{t-1} + K_t T_I T_c Y_t + (K_t T_{D,t} Y_t - 2K_{t-1} T_{D,t-1} Y_{t-1} + K_{t-2} T_{d,t-2} Y_{t-2})/T_c \qquad (5.5)$$

5.2 Two Methods of Handling PID Saturation

When a PID controller saturates, the control action calculated by the control algorithm cannot be physically applied to the process. Some maximum (or minimum) value of the control action must be used instead. The way in which the control action is set to the limit must be handled carefully. If the controller has Integral action and the limit is attained due to Integral action, then the integration of the error must be stopped or the controller will suffer reset (integrator) windup. The Proportional and Derivative action must always be applied in full. Proportional and Derivative action can never cause windup since they depend only on the current and immediately previous values of the error. Of course, the control signal sent to the actuator must be limited.

When the controller is written in the Position form as in equation (5.1), the precaution that should be taken is that the integration must be halted when the limit is attained due to Integral action. The algorithm presented by Segall and Taylor (1986) is corrected here to clarify the detection of saturation and the

calculation of the integral action. In FORTRAN 77, this would be done as follows:

```
C
C       CALCULATE THE PROPORTIONAL AND DERIVATIVE ACTION
C
        UPD = K*(Y +TD*(Y-YM1)/TC)
C
C       CALCULATE THE INTEGRAL ACTION
C
        UINT = UIM1 + KP*TI*TC*Y
C
C       CALCULATE THE TOTAL CONTROL ACTION
C
        U = UPD + UINT + USS
C
C       TEST FOR SATURATION
C
        IF (U.GT.UMAX) THEN
C
C            UPPER SATURATION:
C            LIMIT THE INTEGRAL ACTION
C
             IF ((UMAX-USS).LT.UINT) UINT = UMAX - USS
C
C            RECALCULATE THE INTERNAL CONTROL ACTION
C
             U = UPD + UINT + USS
C
        ELSE IF (U.LT.UMIN) THEN
C
C            LOWER SATURATION:
C            LIMIT THE INTEGRAL ACTION
C
             IF ((UMIN-USS).GT.UINT) UINT = UMIN - USS
C
C            RECALCULATE THE INTERNAL CONTROL ACTION
C
             U = UPD + UINT + USS
C
        END IF
C
C       LIMIT THE CONTROL ACTION
C
        IF (U.GT.UMAX) U = UMAX
        IF (U.LT.UMIN) U = UMIN
C
C       UPDATE THE PAST VALUE OF THE PROCESS OUTPUT AND THE INTEGRAL
C
        YM1 = Y
        UIM1 = UINT
```

The reset windup protection described above for the Position form of the PID controller is different from those previously presented in C.L. Smith (1972), Khandheria and Luyben (1976), Åström and Wittenmark (1984). Those algorithms limit the integration as soon as saturation is hit. Rundqwist (1991) presents several reset windup protection algorithms which he calls methods of "conditional integration". However, his list does not include the above algorithm. The algorithm above does not limit the Integral action unless the Integral action would cause saturation. This position form for saturation correction allows the integration to continue until integral action alone would cause saturation. If the saturation is not due to Integral action then the integration is not stopped. If the Integral action would cause saturation then the integral is limited to the value which would bring the control action to the limit. The motivating assumption in this algorithm is that the saturation results from a level shift in the set point or disturbance so that good controller response requires preservation of as much integral action as possible.

The usually prescribed method for reset windup protection for a controller written in Velocity form is to ignore any portion of the calculated ∇U_t which drives the actuator beyond its saturation limit (see for example Åström and Wittenmark (1984) and C.L. Smith (1972)). The Velocity form of the controller in equation (5.2) and the subsequent discussion explains that this course will effect not only the Integral action but also the Proportional and Derivative actions. Åström and Wittenmark (1984) recommend that the Velocity form not be used for P and

PD controllers. However, these problems can be overcome if the controller saturation is handled as shown here.

The way to handle saturation in a Velocity form controller is to limit only that portion of the control action which is due to the integration. The Proportional and Derivative portions of the control action must be applied in full to ensure that the collapsing sums in equation (5.4) reduce to the PID controller of equation (5.1). This is accomplished by calculating an internal control action in the Velocity form of the PID algorithm in which the Proportional and Derivative contributions are not limited. If the controller saturates then the contribution due to the integral action must be limited. The external control action which is applied to the process is always limited. In FORTRAN 77 this would be implemented as follows:

```
C
C       CALCULATE DEL U DUE TO THE PROPORTIONAL AND DERIVATIVE ACTION
C
        DELUPD = K*((Y - YM1) + TD*(Y - 2*YM1 + YM2)/TC)
C
C       ADD THE CONTRIBUTION TO THE OLD INTERNAL CONTROL ACTION
C
        UPD = UPIDM1 + DELUPD
C
C       CALCULATE DEL U DUE TO THE INTEGRAL ACTION
C
        DELUI = K*TI*TC*Y
C
C       ADD THE CONTRIBUTION TO THE PD CONTROL ACTION
C
        UPID = UPD + DELUI
C
C       TEST FOR UPPER SATURATION
C
        IF (UPID.GT.UMAX) THEN
C
C            UPPER SATURATION:
C            LIMIT THE INTEGRAL ACTION INCREMENT
C            FROM THE PREVIOUS EXTERNAL CONTROL ACTION
C
             IF (DELUI.GT.(UMAX-UM1)) DELUI = UMAX - UM1
```

```
C
C            RECALCULATE THE INTERNAL CONTROL ACTION
C
             UPID = UPD + DELUI
C
        END IF
C
C       TEST FOR LOWER SATURATION
C
        IF (UPID.LT.UMIN) THEN
C
C            LOWER SATURATION:
C            LIMIT THE INTEGRAL ACTION INCREMENT
C            FROM THE PREVIOUS EXTERNAL CONTROL ACTION
C
             IF (DELUI.LT.(UMIN-UM1)) DELUI = UMIN - UM1
C
C            RECALCULATE THE INTERNAL CONTROL ACTION
C
             UPID = UPD + DELUI
C
        END IF
C
C       PASS THE INTERNAL CONTROL ACTION
C       TO THE EXTERNAL CONTROL ACTION
C
        U = UPID
C
C       LIMIT THE CONTROL ACTION
C
        IF (U.GT.UMAX) U = UMAX
        IF (U.LT.UMIN) U = UMIN
C
C       UPDATE THE PAST VALUES OF THE PROCESS OUTPUT
C       AND THE IMPLEMENTED CONTROL ACTION AND THE INTERNAL CONTROL
C
        YM2 = YM1
        YM1 = Y
        DELU = U - UM1
        UM1 = U
        UPIDM1 = UPID
```

The algorithm presented by Segall and Taylor (1986) is corrected here to clarify the detection of saturation and the calculation of the integral action. In this algorithm the proportional and derivative contributions to the internally calculated control action, UPID, are never limited. However, when saturation is encountered the integral

contribution is limited to an increment which would take the previously implemented control action, U, to the saturation limit. This is because the neither the integration nor the steady state control action appear explicitly in the velocity form: the previous control action is the only reference point for the integral calculation. So, if the previous control action was at the limit and the current control action is at the limit then there is no integration. The control action, U, that is a applied to the plant is always limited.

This algorithm successfully maintains the equivalence of the Proportional and Derivative action in the Velocity form algorithm with those in the position form algorithm via the collapsing sums of equation (5.4). However, the integration behaves differently at saturation: the velocity form stops integrating whereas the position form continues to integrate until the integral action alone would bring the control action to saturation. Thus the velocity form implementation here is more akin to the methods of C.L. Smith (1972), Khandheria and Luyben (1976), and Åström and Wittenmark (1984).

Although the algorithm uses the velocity form of the controller, the controller output is calculated as a position. The saturation limits against which it is compared are the absolute limits. If an integrating actuator is used, the output increments should be calculated as DELU in the algorithm. Because of the possibility of noise in implementing the actuator increments, the external control action UM1 should be updated by a readback from the actuator. When there is no saturation UPIDM1, the previous internal control action, should be made equal to UM1, the previous external control action obtained by readback.

A similar algorithm was successfully used in a computer data acquisition and control system applied to a pilot-scale packed-bed reactor (Segall (1983)).

The algorithms presented above are not intended to be used during a controller override. If the output of the controller is not being applied to the process, that is an override is in effect, then the controller must be turned off. When the override ends, the controller should be initialized with the current output to the process to ensure a smooth transfer before being restarted.

5.3 Application of the One-Step Optimal Saturation Correction

In the previous chapter a saturation correction was developed for one-step optimal controllers. The key result of that derivation was the controller design equation (4.15):

$$G(z^{-1})F^{a}(z^{-1})Y_t + |F(z^{-1})|B(z^{-1})U_t + C(z^{-1})F^{a}(z^{-1})Q^{-1}B_0^{-1'}R\nabla^{d}U_t \qquad (4.15)$$
$$+ [I - C(z^{-1})F^{a}(z^{-1})][B_0 + Q^{-1}B_0^{-1'}R]U_t^{+} = 0$$

Consider the SISO minimum variance version of this equation (when R is chosen to be R=0):

$$G(z^{-1})\ Y_t + F(z^{-1})B(z^{-1})U_t + [I - C(z^{-1})]B_0U_t^{+} = 0 \qquad (5.6)$$

For a second order discrete process with no zero:

$$(1 + a_1\ z^{-1} + a_2\ z^{-2})\ \nabla\ Y\ = b_0\ \nabla\ U_t\ + (1 + a_1\ z^{-1} + a_2\ z^{-1})\ e_t \qquad (5.7)$$

the controller equation becomes:

$$\nabla\ U = -\ b_0^{-1}\ (1 + a_1\ z^{-1} + a_2\ z^{-2})\ Y_t\ +\ (a_1\ z^{-1} + a_2\ z^{-2})\ U_t^{+} \qquad (5.8)$$

which is a PID controller with saturation correction (note the negative sign results from the different definition of Y in the previous chapter). In equation (5.8), U_t is the current control action being calculated and U_{t-1} is the previous control action as it was actually implemented. The unimplemented portions of the previous two control actions are U^+_{t-1} and U^+_{t-2}.

We can now take a reverse tack: given a PI or PID controller with given tuning constants K_c, T_I and T_D and assume that it is a minimum variance controller for some process (whether we plan to apply it to that process or not is not important). The PID tuning constants can be obtained by rearranging the controller (following MacGregor (1981)) into the form of equation (5.2) to give:

$$K = b_0^{-1}(a_1 + 2a_2), \quad T_I T_c = (1 + a_1 + a_2) \,/\, (a_1 + 2a_2) \text{ and } \quad T_D/T_c = a_2 \,/\, (a_1 + 2a_2)_{-1} \tag{5.9}$$

or

$$b_0^{-1} = K\,(1 + T_I T_c + T_D/T_c), \quad a_1 = b_0(-K - 2KT_D/T_c), \quad a_2 = b_0 K T_D/T_c \tag{5.9a}$$

Then the optimal saturation correction for the PID controller is given in the equation above (5.8). The coefficients can be interpreted using the relationships of equation (5.9) to give:

$$\begin{aligned} \upsilon U_t = {} & K\,[(Y_t - Y_{t-1}) + T_I T_c Y_t + T_D(Y_t - 2Y_{t-1} + Y_{t-2})/T_c] \\ & + 1/(1 + T_I T_c + T_D/T_c)\,[\,(-1 - 2\,T_D/T_c)\,U^+_t + \,T_D/T_c\,U^+_t)\,] \end{aligned} \tag{5.10}$$

Equation (5.10) gives an alternative method of reset windup protection for PID controllers which is based in an ad hoc way on the one-step optimal correction of the previous chapter.

5.4 The Method of Smith and Corripio (1985)

Smith and Corripio (1985) present a method of saturation correction for continuous time PI controllers which they describe as being in common commercial implementation. This method can be adapted to discrete PI controllers. Following their derivation the position form of the PI controller is written:

$$U_t - U_{ss} = K \left[Y_t + T_I T_c \sum_{i=1}^{t} Y_i \right] = K \left[1 + T_I T_c / \nabla \right] Y_t = K Y_t + M_{I,t} \tag{5.11}$$

where ∇ is the backward difference operator and the integral term is written as:

$$M_{I,t} = K T_I T_c / \nabla Y_t \tag{5.12}$$

This latter equation can be rearranged to give:

$$\nabla / (T_I T_c) M_{I,t} = K Y_t \tag{5.13}$$

Similarly equation (5.11) can be rearranged to give:

$$K Y_t = (U_t - U_{ss}) - M_{I,t} \tag{5.14}$$

Now, KY_t can be eliminated between equation (5.13) and (5.14) to give:

$$\nabla / (T_I T_c) M_{I,t} = (U_t - U_{ss}) - M_{I,t} \tag{5.15}$$

This can be rearranged to solve for $M_{I,t}$:

$$M_{I,t} = (U_t - U_{ss}) / (\nabla / T_I T_c + 1) \tag{5.16}$$

which gives the integral term as a first order system driven by the controller output. Smith and Corripio use this equation to calculate the integral action from the control action *after* it is limited. So, because the first order system has unity gain the integral contribution to the control action can never exceed the limit.

In the discrete time implementation the integral action must be calculated from the previous implemented control action in order to use a control action which has been limited:

$$M_{I,t} = T_I T_c/(1 + T_I T_c)\ (U_{t-1} - U_{ss}) + 1\ /(1 + T_I T_c)\ M_{I,t} \qquad (5.17)$$

Three comments can be made about this method of reset windup protection. Firstly, the integral calculation will always lag the proportional action by one control interval. So, performance will be less than with a more normal implementation because the integral action is delayed. In the continuous time implementation of Smith and Corripio this is of course not an issue.

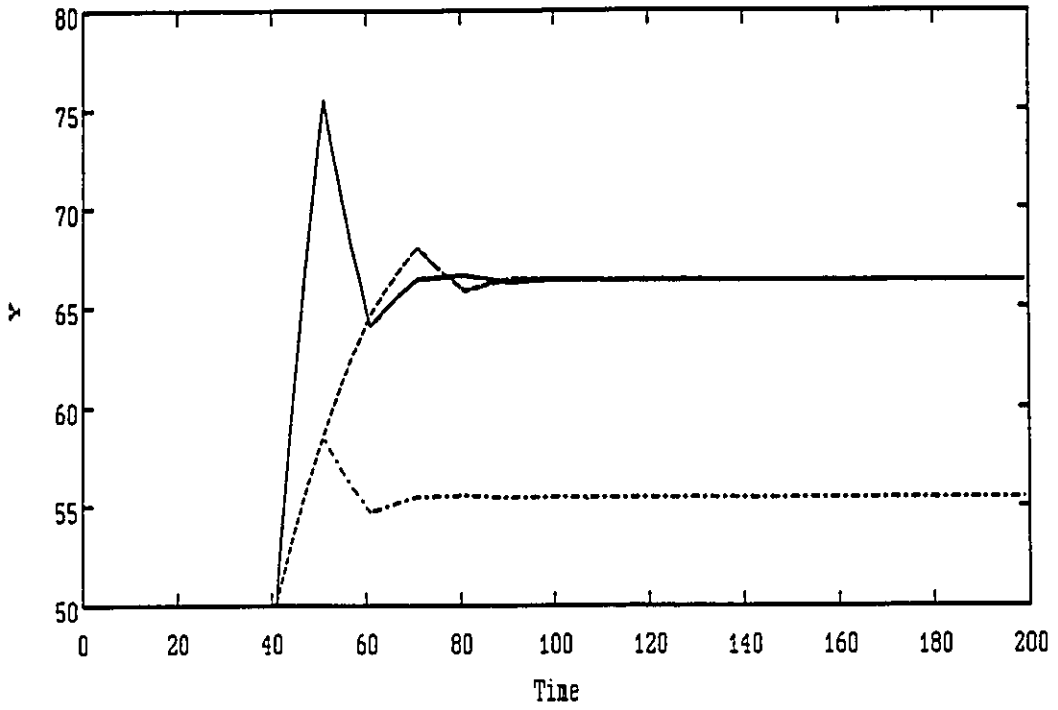
Secondly, once saturation has occurred the integration will happen at a reduced rate compared to the position form algorithm of section 5.2 above. The integration is slowed because equation (5.17) will be driven by the limited control action rather than the previously calculated control action. This would place its performance at saturation between that of the velocity form implementation and the position form implementation of section 5.2 above. In the algorithms of section 5.2 the velocity form algorithm stops integrating once the previous implemented control action is at the saturation limit, while the position from algorithm continues to integrate at the normal rate until the integral contribution alone would bring the control action to the saturation limit.

Thirdly, their method does not appear to generalize readily to controllers which also have Derivative action.

5.5 Simulation Results

Consider a first order plus dead-time system with gain of 1.5 %/%, a time constant of 30 minutes and a dead-time of 1 minute. Figure 5.1 shows the process response, Y, and the control actions, U, for Proportional only control with a controller gain of 3.0 %/% and a control interval of 10 minutes. A set point change from 50% to 70% is made at time 40 minutes. The control action initial steady state is 80%. Figure 5.1 shows that when there is no saturation limit the first control action is 140%. The figure also shows the response when there is an upper saturation limit at 100% and any of the saturation correction methods described in sections 5.2, 5.3 and 5.4 above are used. All of these methods correctly preserve the proportional action during the period of actuator saturation and the process response converges with the process response when there is no saturation.

Figure 5.1 also shows the shows the response of a Velocity form controller where the reset windup problem is handled by simply ignoring all contributions to the output that are above the saturation limit. The first control action at time 40 would be 140%, but this controller discards the extra 40% beyond saturation. The control action increment calculated at the subsequent control interval, time 50, is meant to bring the control action from 140% to 115%; however, because the control action to which it is added is only 100% it erroneously lowers the control action to 75%. The result is that the process response does not converge with the response when there is no saturation and the offset from the new set point is 14.5% rather than the 3.6% obtained when there



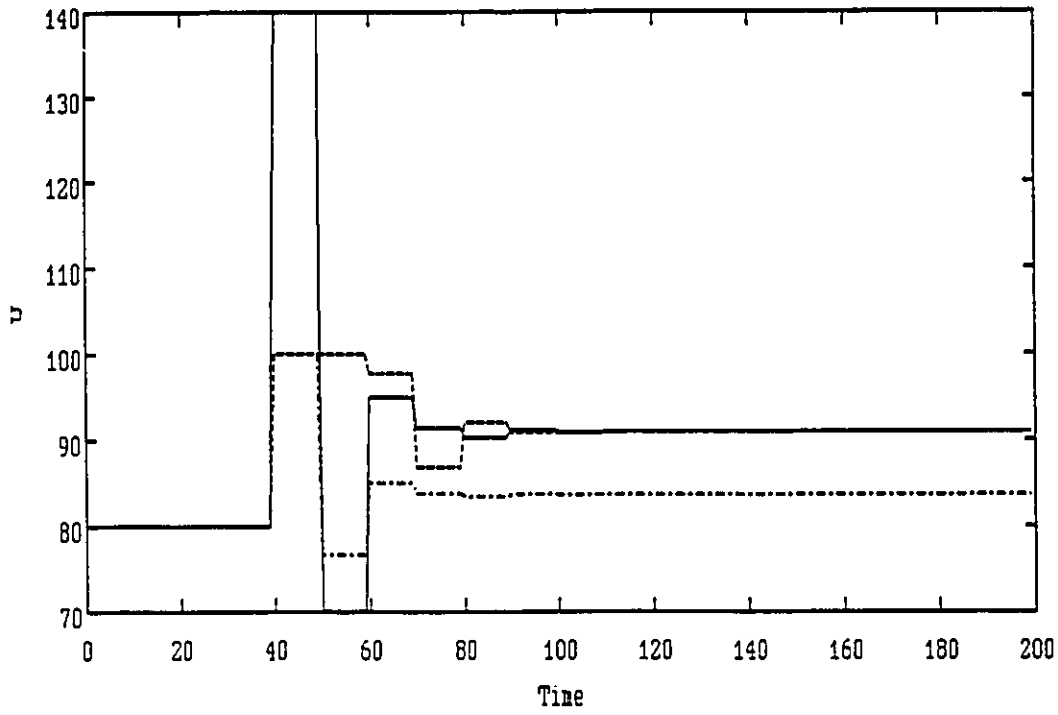


Figure 5.1 Proportional Only Control - No Saturation, -- Correct Saturation Protection, -. Standard Anti-Reset Windup

is no saturation, or when there is saturation but one of the methods described above is used.

For a second example consider a first order process with a gain of 1.5 %/%, a time constant of 15 minutes and no dead-time with a Proportional and Integral action controller with gain of 1.5 %/%, reset rate of 0.0666667 repeats/minute and a control interval of 3 minutes. A set point change from 50% to 70% is made at time 12 minutes and the initial control action steady state is 80%. Figure 5.2 shows the process response and the controller action when there is no saturation limit. Figure 5.2 also shows the responses when the Smith and Corripio implementation of the PI controller described in section 5.4 above is used. The process response using the Smith and Corripio implementation lags behind the response using a normal implementation. With the Smith and Corripio implementation the first control action after the set point change at time 12 minutes is only 110% instead of 116% for the normal implementation. This is because the integral action (being 6%) is not added in until the next time step, following equation (5.17) above.

Figure 5.2 also shows the process response and the control actions when a saturation limit is imposed on the control action at 100% and a Velocity form implementation is used with the normal reset windup protection of discarding the unimplemented portion of the increment. The control only saturates for one control interval and then falls back towards its eventual steady state. This is again because of the effect of discarding a portion of the increment on the proportional action. As a result the process response to the new set point is quite sluggish.

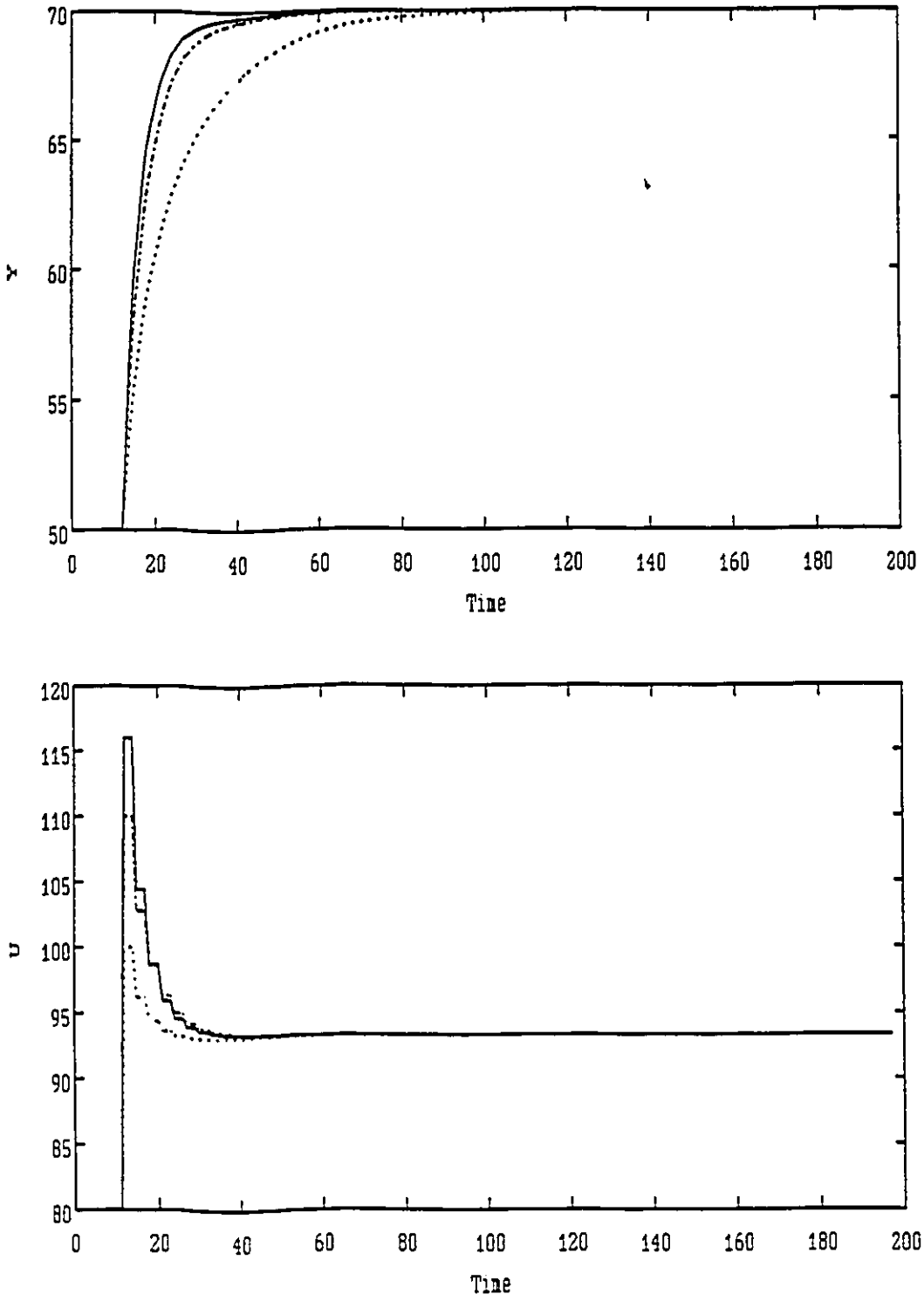


Figure 5.2 PI Control, - No Saturation, -. No Saturation S & C Method
.. Saturation with Normal Anti Reset-Windup Protection

Figure 5.3 shows the results for the same set point change with a saturation limit on the control action at 100%. The Figure shows that for this example the position from algorithm of section 5.2 and the adaptation of the one-step optimal correction algorithm of section 5.3 both produce the same result. Because the saturation is only transient both of these methods continue to integrate and so hold the control action at the saturation limit for 5 control actions. This causes a process response which overshoots the set point before settling. Figure 5.3 also shows the response using the velocity form algorithm of section 5.2 above. Using this method the integral action stops after the first period of saturation so that the control action only stays at the saturation limit for 3 control intervals. This results in a slower process response to set point. Figure 5.3 also shows the response using the method of Smith and Corripio of section 5.4 above. This method does not integrate on the first action but then integrates on the subsequent control actions, but at a reduced rate relative to the position form. The result is that the control action stays at the saturation limit for 3 control intervals, but then the control actions are higher than those of the velocity form control implementation. This gives a process response which is in between those of the position form and velocity form implementations.

Figure 5.4 uses the same process and PI controller tuning as in the previous example. However, two set point changes are made: one from 50% to 70% at time 12 minutes and then from 70% to 50% at time 105 minutes. The initial control action steady state is set at 90% and the saturation limit is at 100%. The set point change to 70% is

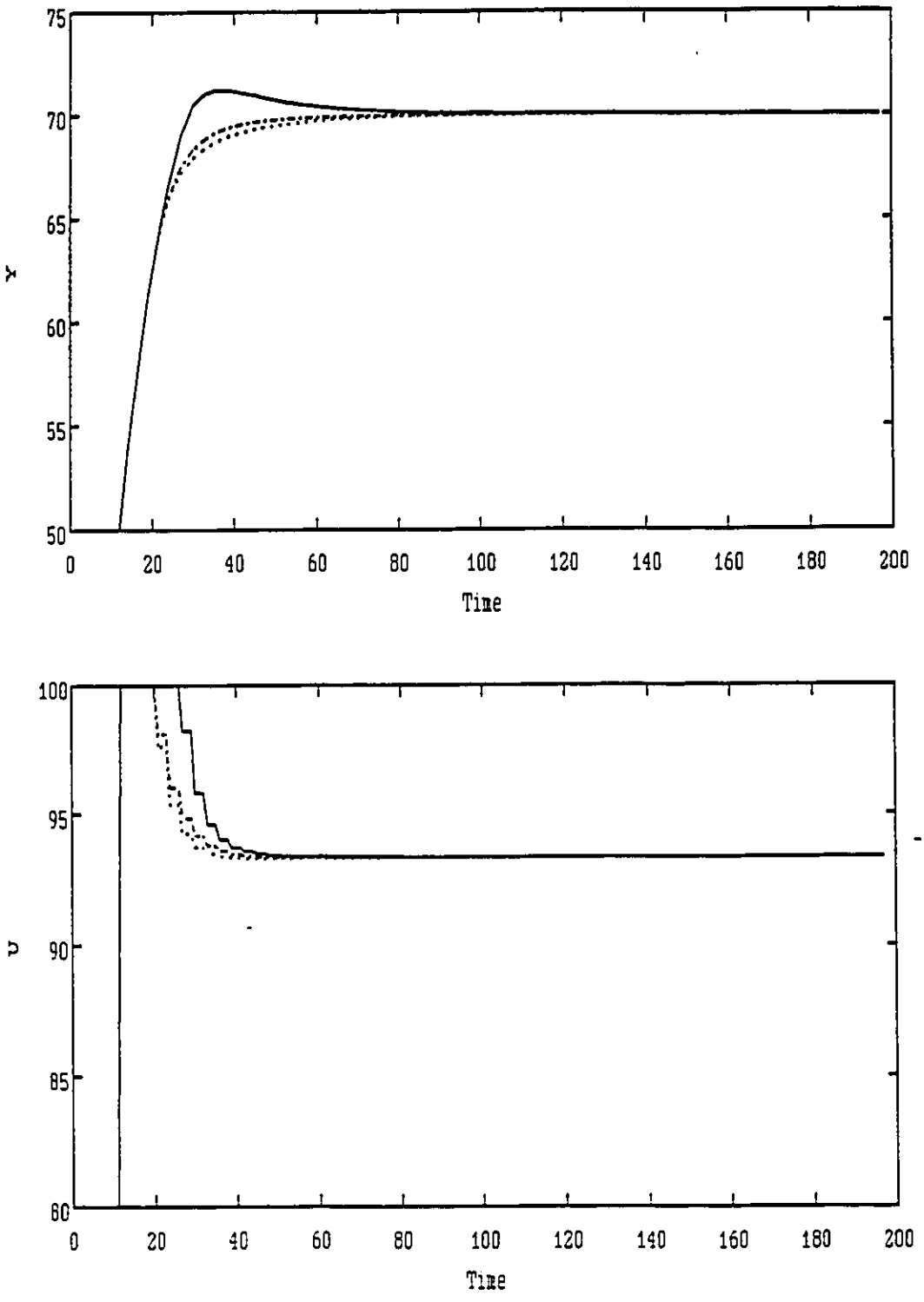


Figure 5.3 PI Control, - Position Form New Algo,
.. Velocity Form New Algo, -. Smith & Corripio Algo

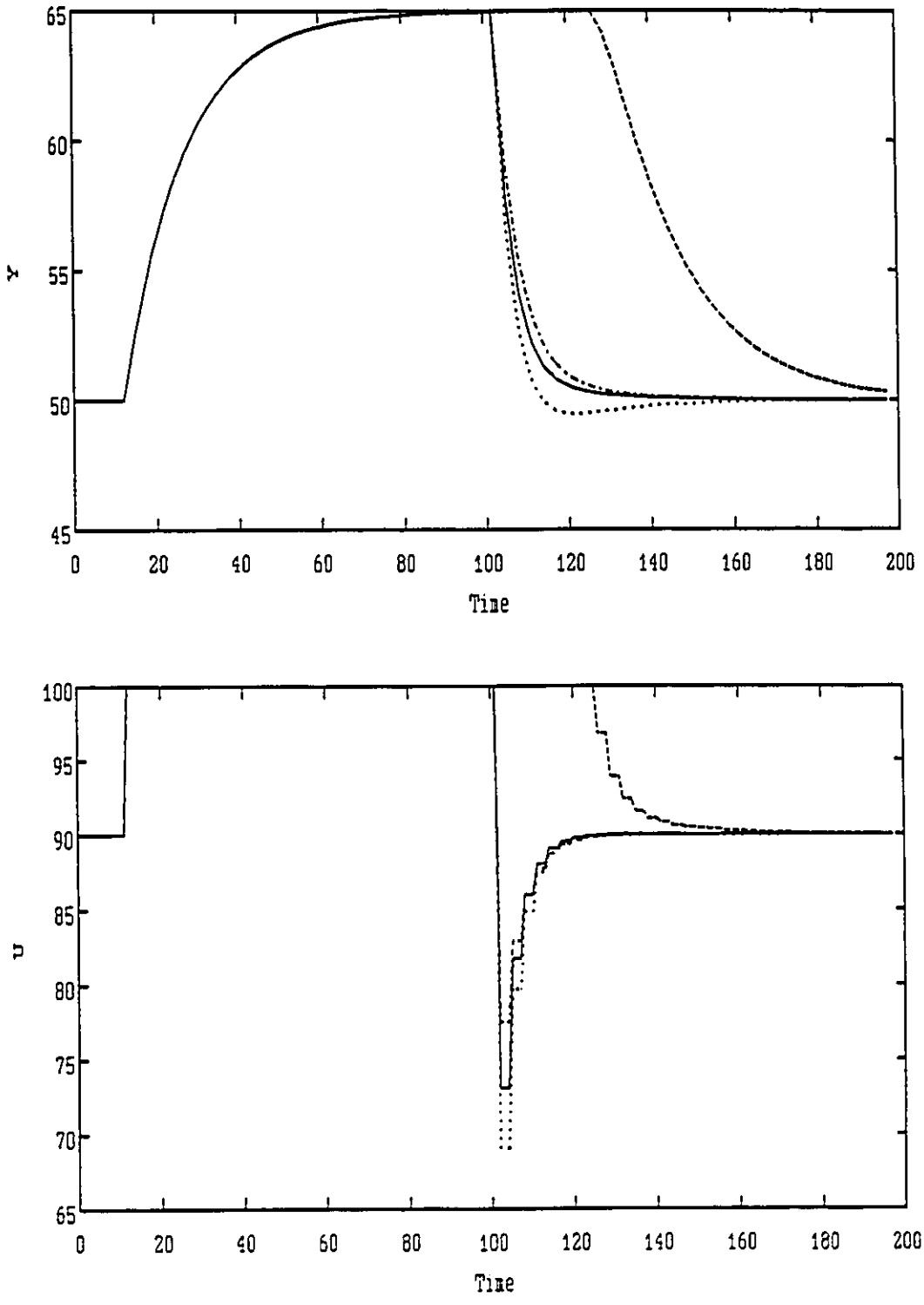


Figure 5.4 PI Control, - New Position Form Algo,
.. New Velocity Form Algo, -- One-step Algo, -. S & C Algo

unattainable and so causes a prolonged period of saturation. All four of the methods saturate on the first control action and then hold the control action there until the second set point change is made. The figure shows the position form implementation using the algorithm of section 5.2. It continues to integrate during the control action saturation until the integral action alone would cause saturation. Then, when the set point change is made downwards the controller response looks very like that of Figure 5.2 (only upside down). Figure 5.4 also shows the Smith and Corripio implementation of section 5.4. It also continues to integrate during the control action saturation until the integral action alone would cause saturation. Again, when the set point change is made downwards the controller response looks very like that of Figure 5.2: the response is slower than that of the position form implementation because of the delay in calculating the integral action. Figure 5.4 shows the response using the Velocity form implementation of section 5.2 above. It stops integrating after the first saturated control action. So, it responds with the largest decrease in the control action when the set point change downward is made. As a result the process response undershoots the set point before settling. The figure shows the response using the adaptation of the one-step optimal correction of section 5.3 above. The control action shows significant windup as it does not return from saturation until long after the second set point change.

Several conclusions can be drawn from these simulation results. Firstly, the adaptation of the one-step optimal correction which was attempted in section 5.3 is not effective. For the example in

Figure 5.4 it suffers from reset windup during prolonged saturation. This is because the method is model based and assumes a model. The calculation of Equation (5.9a) for this example indicates that the assumed process has b_0=.5556, a_1=-0.833 and a_2=0. So the assumed process has a gain of 3.33 and a time constant of 16.5 minutes. This assumed gain is more than twice the actual process gain of 1.5. The problem with the method of section 5.3 is that any given PID controller can arise from many different assumptions of process and disturbance model. The choice of a minimum variance controller for a random walk disturbance was far from correct for this example.

Secondly, the discrete version of the Smith and Corripio implementation of the PI controller does not perform as well as a true PI controller when there is no saturation because of the delay in calculating the integral action. Although their implementation is elegant in the way it handles saturation it performs no better than the position and velocity form corrections presented in section 5.2 above.

Lastly, the new Position and Velocity form implementations of section 5.2 both have different properties which respond differently under different circumstances. In all of these examples the controller with no saturation would cause a monotonic change in the process towards the set point. The position form implementation caused the process response to overshoot the set point when there was transient saturation of the control action. The velocity form implementation caused the process to overshoot the set point when returning from a prolonged period of saturation. For purely Proportional and Derivative control both methods are identical. For purely Integral control both methods

are identical. The differences between the two methods for the ground in between is not that large.

5.6 Summary

Saturation of the control actuator can cause problems for Proportional Integral Derivative controller implementation. The main problems were shown to be preserving the integrity of the Proportional and Derivative action in velocity form controllers and preventing integrator windup. The velocity form implementation of the PID controller requires corrections to allow both online tuning and to preserve the Proportional and Derivative action at saturation. The position form implementation requires only that the integral action not be allowed to cause windup when the controller saturates. This is made easy because the position form calculates the integration explicitly so it can be manipulated. New algorithms to prevent reset windup in both Velocity form and Position form PID controllers were presented. Two other methods of correcting for actuator saturation in PID controllers were found to be less satisfactory: an adaptation of the one-step optimal methods of chapter 4 and an adaptation of the method of Smith and Corripio (1985).

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

The research presented in this thesis has covered topics in the design, analysis and implementation of feedback control systems. The thesis has proposed new discrete controller designs, a new method of robustness analysis for continuous and discrete systems, and has provided solutions to the implementation problem caused by actuator saturation.

The first section of this chapter presents a summary of the conclusions presented in this thesis and recommendations which follow from those conclusions. The subsequent section presents recommendations for further research which might follow from the work in this thesis.

6.1 Conclusions and Recommendations

This thesis has presented a new definition of robustness to process model mismatch for SISO feedback control systems: the region of joint allowable variation in the process gain and dead-time parameters. The method of Tsypkin (1946) for calculating the robustness region of

joint allowable variation in the process gain and dead-time parameters was extended to systems which can be made unstable by a reduction in the dead-time. His analysis only included systems which could be destabilized (or stabilized) by adding dead-time.

This method of robustness analysis provides a clear indication of the robustness of a closed loop feedback system to changes in the process gain and the process dead-time. Frequency domain robustness specifications of Palmor (1980), Doyle and Stein (1981), Morari and Zafiriou (1989) and Lewin (1991) require that the designer specify a priori the magnitude of the process uncertainty as a function of frequency, paying special attention to the "high" frequencies. However, the robustness region method described in this thesis allows a designer to calculate the allowable gain and dead-time change once a controller has been designed. In this way the designer can choose his or her favourite controller design method and then measure the changes in robustness as the controller is tuned. The frequency domain construction used to calculate the robustness region makes it easy to understand the nature of any robustness limitations. The question of what is a "high" frequency is easily answered since the allowable change in the process dead-time incorporates frequency information in a natural and meaningful way.

The limitation on using this definition of robustness is that it is not easily automated. When comparing two different controllers applied to the same process it was shown that the shape of the robustness regions can be very different. The designer must make a judgement on the desired level of robustness. From the perspective of

this study this is not a serious concern. When designing a controller the designer must make decisions about what performance and what control actions are acceptable. A tool which allows a rational and intuitive measurement of robustness is a valuable addition to the control system analysis toolbox.

The robustness region measure of robustness and the Integral Square Error measure of performance were used together to compare the robustness vs performance tradeoff of the tuning of Linear Quadratic Optimal Control, Internal Model Control and Proportional Integral Derivative control for a first order plus dead-time process. The comparison showed that for the same level of robustness PID (Rivera et al (1986)) has better performance than IMC (Rivera et al (1986)) which has better performance than LQOC (Palmor (1982)). This is a surprising result as Rivera et al (1986) claim that IMC gives 10% better performance than PID control, but they do not compare the controllers at the same level of robustness.

That the IMC controller has better performance than the LQOC controller for the same level of robustness should be generalizable to higher order systems, provided that the IMC tuning filter is chosen to match the order of the process. This is because the result that the LQOC controller has an additional filter as compared to the IMC controller, when examined in its Smith Predictor form, is general. This additional filter was shown to be a natural consequence of the LQOC objective function.

However, the result that PID control has better performance than the IMC controller for the same level of robustness likely does not

generalize to higher order systems. This is because the PID controller obtains its performance improvement for a first order plus dead-time process by using derivative action to approximate the dead-time compensation. For a second order plus dead-time process good control requires derivative action simply to compensate for the process dynamics. However, this suggestion is not certain: it may be that the robustness problems caused by dead-time compensation are more severe than the performance improvements obtained by a higher order controller.

The robustness region was shown to be useful for comparing the robustness of these very different controllers. The robustness analysis method was used to analyze the robustness implications of dead-time compensation. The high frequency peaks in the open loop frequency response were shown to cause contraction of the robustness region, and tuning of controllers is shown to reduce these peaks. Although it has previously been observed that high frequency peaks in the frequency response of dead-time compensating controllers cause robustness problems (Palmor (1980), Ioannides et al (1979)), there has been no quantitative analysis of how tuning works to reduce these peaks and improve robustness. The insights from robustness analysis were used in Chapter 2 to explain why the substitution of derivative action for dead-time compensation (Rivera et al (1986)) can provide good robustness and good performance.

The discrete Linear Quadratic Optimal Control (Box and Jenkins (1970), Åström (1970)), the modified Dahlin (Dahlin (1968)) and the State Deadbeat IMC (Zafiriou and Morari (1985)) controllers were compared for ISE performance for step changes in set point when applied

to first order plus dead-time and second order processes as the size of the control interval was varied in Chapter 3. Continuous time simulations of the process output and the control actions were used to explain the results. This study confirmed the results of Zafiriou and Morari (1985) where they use continuous time simulations to show the conditions under which each of these controllers perform more poorly than the State Deadbeat controller. Where the study in this thesis provides an additional contribution is in highlighting that the LQOC controller provides improved performance over the State Deadbeat controller when the process model has a zero which is not near minus one.

Tuning of LQOC, IMC and State Deadbeat IMC controllers was compared and analyzed. The analysis lead to the suggestion of two novel controller designs. The Modified LQOC controller calculates a controller for a nonstationary disturbance using the spectral factorization approach as if the disturbance were stationary. This controller has a tuning filter which is of order one less than the LQOC tuning filter. Through the appropriate choice of the tuning filter gain the resulting controller still has integral action. Also, a new extended horizon controller design is proposed which behaves as if it were a state deadbeat controller for the same process but with a longer control interval. This latter controller is an improvement over the extended horizon controller of Ydstie et al (1985) in that if a single set point change is made and no disturbance enters the system then the initial control action sequence will be implemented with no changes.

The new extended horizon controller is very instructive in understanding the relationship between the selection of the size of the control interval and tuning. It shows that although a designer may choose a short control interval to obtain improved disturbance rejection capability, the controller can be tuned to give the same mild control actions as would be obtained with a longer control interval. However, from a practical implementation point of view the new extended horizon controller design requires a very high order numerator in the tuning filter. Similar mild control actions can be obtained in a more compact form using the State Deadbeat IMC controller or the Modified LQOC controller.

The Modified LQOC controller provides a bridge between the good properties of the LQOC approximate model inverse when the process has a zero far from minus one with the ability to obtain the good properties of the State Deadbeat approximate model inverse when the process has a zero near minus one. When $\lambda=0$ the modified LQOC design is identical to the LQOC design. If the performance is unacceptable increasing the tuning parameter λ moves the poles in the controller from the zeros in the process model towards the poles of the process model, improving performance by reducing the ringing in the control action. This tuning may initially improve the process ISE performance by reducing ringing action in the control action. When $\lambda=\infty$ the Modified LQOC controller provides steady state type control where the process responds according to its open loop dynamics.

The order of the State Deadbeat IMC tuning filter must be chosen by the designer, although a filter order matched to the order of the

process is a good choice. The Modified LQOC controller has a tuning filter which is matched automatically to the order of the process and is of order one less than the LQOC tuning filter. This makes it a very attractive one parameter controller tuning procedure.

As a prelude to introducing robustness analysis for discrete control of continuous systems, the effect of dead-time on the discrete frequency response was examined. The modified Z transform of a first order process with a fractional period of dead-time was shown to be a convex combination of the Z transforms of the same process with integer periods of dead-time which bracket the actual dead-time. This is believed to be a new result. Through the use of the example of a second order process it was shown that this result is not general. In fact, for a second order process the frequency response of the process with a fractional period of dead-time can be outside of the boundary formed by the frequency response of the process with integer periods of dead-time which bracket the actual dead-time. This result appears to be new and has implications for studies which determine stability boundaries by a search on the dead-time parameter.

The definition of the robustness region for continuous processes was adapted for discrete control of continuous processes as the region of joint allowable variation in the process gain and the quasi dead-time. The quasi dead-time is defined to be the phase margin of the discrete open loop frequency response divided by the discrete frequency. When the allowable variation in the quasi dead-time is an integer number of control intervals and the process has no fractional period of dead-

time then it provides an exact upper limit on the allowable dead-time variation in the underlying continuous process.

The power and the limitations of this definition of robustness were shown by application to examples. The method was used to show that a minimum variance controller for a process with no dead-time subject to random walk disturbances (equivalent to deadbeat control for step changes in set point) can only tolerate an increase in the process dead-time of at most one control interval. That is, the robustness of minimum variance control is proportional to the size of the control interval. This is believed to be a new result. In addition it was shown that the robustness of minimum variance control is further reduced when the process has dead-time. The effect of dead-time compensation on discrete minimum variance controllers was shown to be similar to the effect of dead-time compensation on continuous model based controllers: high frequency peaks in the frequency response cause a reduction in the system robustness.

The limitation of the method is shown to be that when calculating the robustness of discrete control applied to a process with a fractional period of dead-time the allowable variation in the quasi dead-time can significantly over-estimate the allowable variation in the true dead-time; however, the relative results when comparing different controllers on the same process are meaningful. Morari and Zafiriou (1989) translate a continuous time frequency response robustness specification to the design of controllers for discrete processes without showing any awareness of the difficulties caused by fractional periods of dead-time on the discrete frequency response. Again, as in

the continuous time definition of robustness, the quasi-dead time incorporates frequency information in a natural and meaningful way. Because of this limitation on the interpretation of the results, this method of robustness analysis does not replace parameter search methods such as used by Bergh (1986). However, the frequency response construction used to calculate the robustness region can help the designer understand the source of robustness problems.

The allowable change in the quasi dead-time measure of robustness and the continuous Integral Square Error measure of performance were combined to compare the robustness vs performance tradeoff of the tuning of discrete LQOC, Modified LQOC, IMC, and State Deadbeat IMC controllers applied to first and second order processes with step changes in set point. The results were explained using continuous time plots of the process output and the control actions. This study showed that a controller with good performance and good robustness can still require totally impractical control actions. The LQOC controller design was shown to have the poorest performance for the same level of robustness. This result is analogous to the result described above that the continuous time LQOC controller had the poorest performance for the same level of robustness. The poor performance of the LQOC controller is because it has a tuning filter which is of order one higher than the process.

The IMC and State Deadbeat IMC controllers with a first order filter applied to a second order process were shown to have good performance and robustness, but can generate unrealizable control actions when the order of the tuning filter is not matched to the order

of the process. The State Deadbeat IMC controller with a filter order matched to the order of the process were shown to have good performance, robustness and realizable control actions. The modified LQOC controller has a filter order which is naturally matched to the order of the process and so has good performance, robustness and realizable control actions. This study contributes to the understanding of the relationships between tuning filters, performance, robustness and realizability of control actions. Bergh and MacGregor (1987) have studied the improved disturbance rejection ability of the LQOC controller over the IMC controller for autoregressive disturbances. Their study, however, focuses on discrete measures of performance for a process example with a fractional period of dead-time and also does not make a comparison with a State Deadbeat IMC controller which provides improved robustness for this type of process.

The general conclusion which follows from Chapter 3 is that the State Deadbeat IMC controller with a tuning filter matched to the order of the process provides a good controller design which gives up stingily on performance and provides reasonable control actions while gaining robustness as it is tuned. The Modified LQOC controller has all of these good properties and in addition is able to take advantage of a zero in the process model to provide improved system performance when this is desired.

One-step optimal controllers, which include minimum variance controllers, are shown in Chapter 4 to suffer severe performance degradation when the control action is subject to a saturation limit. A new saturation correction for one-step optimal Multi-Input Multi-Output

controllers is derived which allows past saturation to be compensated for in a straightforward calculation using the difference between the past calculated and past implemented control actions. This derivation extends the SISO derivations of Goodwin (1972) and Bezanson (1984) to produce a formula which can be implemented directly. In addition it corrects the MIMO derivation of Mäkilä (1982). A further simultaneous correction is derived which is only required for MIMO systems. The iterative correction calculates an immediate correction for control actions which do not saturate for those which do saturate. In addition the application of this method to other discrete model based controllers such as the Dahlin controller, IMC and DMC is explained. This result is very important as the benefits of a model based controller can be lost at saturation due to poor implementation.

That Proportional Integral Derivative controllers written in velocity form can be subject to severe performance degradation when saturation is encountered has only rarely been observed in the literature (Åström and Wittenmark (1984)). Chapter 5 explains the nature of the problem, how it arises from the mathematics of the velocity form and how to fix it. The problems encountered in the tuning of velocity form controllers are also explained. New anti-reset windup protection algorithms are presented for both velocity and position forms of the PID controller. Examples are used to show the improvement which is obtained from the proposed solution compared against standard methods. Again this result is important for obtaining the expected controller performance when a saturation limit is encountered.

6.2 Recommendations for Future Research

There are several topics for further research which are suggested by the studies in this thesis. This section recommends some areas for possible further research.

6.2.1 The Robustness Region for Continuous Controllers

A new definition of robustness was advanced in Chapter 2 for continuous control of continuous systems: the region of joint allowable variation in the process gain and the process dead-time. This region was shown to represent orthogonal components in the frequency domain. The allowable gain variation in the process model is calculated from the allowable gain variation in the open loop frequency response. The allowable dead-time variation in the process model is calculated from the allowable phase variation in the open loop frequency response. The allowable dead-time variation also incorporates frequency information as for a particular value of the dead-time variation the amount of phase added in the frequency domain is proportional to the frequency.

This definition of robustness is very practical for comparing the robustness of different controllers for the same process. The definition is also very useful as it presents robustness in terms with which a control system designer is very familiar: the gain and dead-time of the process model. However, a designer can often describe his or her knowledge of the process uncertainty or process variability in terms of the variability in the process time constant or in variability of the process order in the form of additional time constants. Further

research might make it possible to calculate these other process model uncertainties from the robustness region of the allowable process gain and dead-time variation. This would then aid the designer in better interpreting the robustness region.

6.2.2 The First Order Padé Approximation as a Substitute for Process Dead-Time When Designing Continuous Model Based Controllers

The result in Chapter 2 that a PID controller designed for a first order plus dead-time process using first order Padé approximation (Rivera et al (1986)) provides better performance than IMC and LQOC controllers tuned for the same level of robustness could be generalized to higher order processes with dead-time. For example the IMC controller design for a second order process with dead-time replaced by a first order Padé approximation would result in a controller with 2nd derivative action instead of dead-time compensation. The study of Rivera et al (1986) suggests that the resulting controller would require tuning as a Padé approximation is only an approximation to the dead-time. However, the resulting controller might have favourable robustness properties as it would not include dead-time compensation.

6.2.3 The Robustness Region for Discrete Controllers

The robustness region was extended to discrete control of continuous systems by introducing the concept of the quasi dead-time. The robustness region for discrete control of continuous systems is the region of joint allowable variation in the process gain and the process quasi dead-time. Only when the allowable variation in the quasi dead-time is an integer and the process has no fractional period of dead-time

does it provide an upper bound on the allowable variation in the continuous process dead-time. Despite this limitation, the quasi dead-time was found to be useful for comparing the robustness of different controllers for the same process and for comparing controllers as the size of the control interval is varied. However, when the allowable quasi dead-time variation is not an integer and when the process has a fractional period of dead-time, there are limitations on the absolute accuracy of interpreting the allowable process dead-time variation although relative comparisons can be made. This problem was highlighted for the design of discrete controllers when the control interval is chosen so that the discrete process model has a fractional period of delay.

Further research into this method of robustness analysis might help provide better methods for interpreting the quasi dead-time in terms of the process dead-time. For example, when dealing with a process model with a fractional period of dead-time, it might be possible to analyze a controller design for a different control interval for which the process model would have an integer dead-time and then translate the result to the original control interval under consideration.

6.2.4 The Modified Linear Quadratic Optimal Controller

The modified Linear Quadratic Optimal Controller presented in Chapter 3 was shown to have good robustness and performance properties. However, although it uses spectral factorization to calculate the approximate model inverse it is not clear that it represents the

solution to an optimal control problem. The modified LQOC controller uses the spectral factorization solution which would be optimal for a stationary disturbance but then applies it to the design of a controller for a nonstationary disturbance.

Further research on the nature of the objective function, if any, for which this modified LQOC controller is optimal might be fruitful. This is especially important given the current popularity of controllers implemented through online mathematical programming. In order to obtain the favourable properties of performance and robustness the performance index for which this controller is the optimal solution would be quite valuable.

6.2.5 Saturation Correction

Significant additional work on saturation correction in implementing model based controllers and on reset windup protection in PID controllers has been done by Wong (1992). The research presented here together with Wong's work represent a quite thorough investigation of solutions to the implementation problems caused by the saturation of control actions.

REFERENCES

Åström, K.J. (1970), Introduction to Stochastic Control Theory, Volume 70 in Mathmetics in Science and Engineering, R. Bellman Ed., Academic Press, New York.

Åström, K.J. (1977), "Frequency domain properties of Otto Smith regulators," Int. J. Control, 26, 307-314.

Åström, K.J. (1978), "Stochastic Control Problems" in Mathematical Control Theory Proceedings Canberra Australia 1977, Ed. W.A. Coppel, in Lecture Notes in Mathematics, Ed. A. Dold and B. Eckmann, v.680, Springer-Verlag, New York.

Åström, K.J. and E. Wittenmark (1984), Computer Controlled Systems Theory and Design, Prentice-Hall Information and System Science Series, Ed. T. Kailath, Prentice-Hall, New Jersey.

Banyasz, Cs., J. Jetthessy and L,. Keviczky (1985), "An Adaptive PID Regulator Dedicated for Microprocessor Based Compact Controllers", Proceedings of the 7th IFAC Symposium on Identification and Parameter Estimation, Pergamon, York, England, 1299-1304.

Bergen, A.R. and J.R. Ragazzini (1954), "Sampled-Data Processing Techniques for Feedback Control Systems", AIEE Trans., 73, 236-247.

Bergh, L.G. (1986), Analysis and Design of Discrete Stochastic Controllers and Spatial Modelling, PhD Thesis, McMaster University, Hamilton, Ontario, Canada.

Bergh, L.G. and J.F. MacGregor, "Constrained Minimum Variance Controllers: Internal Model Structure and Robustness Properties," Ind. Eng. Chem. Res., 26, 1558-1564.

Bezanson, L.W. (1984), "Scalar Quadratic Control with Amplitude Constraints", Electronics Letters, 20, 246-247.

Box, G.E.P. and G.M. Jenkins (1970), Time Series Analysis: Forecasting and Control, Holden-Day, London.

Chang, T.S. and D.E. Seborg (1983), "A Linear Programming Approach for Multi-variable Feedback Control with Inequality Constraints", Int. J. Control, 37, 583-597.

Clarke, D.W. and P.J. Gawthrop (1975), "Self-Tuning controller", Proc. Enst. Electr. Eng., 122, 929-934.

Clarke, D.W. and R. Hastings-James (1971), "Design of Digital Controllers for Randomly Disturbed Systems", Proc. Inst. Electr. Eng., 118, 1503-1506.

Coughanowr, D.R. and L.B. Koppel (1965), Process Systems Analysis and Control, McGraw-Hill book Company, New York.

Cutler, C.R. and B.L. Ramaker (April, 1976), "Dynamic Matrix Control - A Computer Control Algorithm", 86th AIChE Meeting

Dahlin, E.B. (1968), "Designing and Choosing Digital Controllers," Instruments and Control Systems, 41, 77-83.

Doyle, J.C. and G. Stein (1981), "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis," IEEE Trans. Automat. Contr., AC-26, 4-17.

Garcia, C.E. and A.M. Morshedi (1984), "Solution of the Dynamic Matrix Control Problem via Quadratic Programming (QDMC)", Proceedings of the Conference of the Canadian Industrial Computing Society, Ottawa, Canada, 13.1-13.3.

Garcia, C.E. and M. Morari (1982), "Internal Model Control. 1. A Unifying Review and Some New Results", Ind. Eng. Chem. Process Des. Dev., 21, 308-323.

Goodwin, G.C.(1972), "Amplitude-Constrained Minimum-Variance Controller", Electronics letters, 8, 187-188.

Harris, T.J. (1985), "A Comparative Study of Model Based Control Strategies", Automatic Control Conference, Boston.

Harris T.J. and B.D. Tyreus (1987), "Comments on 'Internal Model Control. 4. PID Controller Design'," Ind. Eng. Chem. Res., 26, 2161-2162.

Harris, T.J. and J.F. MacGregor (1987), "Design of Multivariable Linear-Quadratic Controllers Using Transfer Functions", AIChE J., 33, 1481-1495.

Harris, T.J., J.F. MacGregor and J.D. Wright (1980), "Self-Tuning and Adaptive Controllers: An Application to Catalytic Reactor Control", Technometrics, 22, 153-164.

Harris, T.J., J.F. MacGregor and J.D. Wright (1982), "An Overview of Discrete Stochastic Controllers: Generalized PID Algorithms with Dead-Time Compensation", Can. J. Chem. Eng., 60, 425-432.

Private communication from A.J. Hugo (1986), Che 761 lab Report, Departmet of Chemical Engineering, McMaster University.

Ioannides, A.C., G.J. Rogers and V. Latham (1979), "Stability limits of a Smith controller in simple systems containing a time delay," Int. J. Control, 29, 557-563.

Jetto, L. (1990), "Deadbeat Controllers With Ripple-Free Requirement for SISO Discrete Systems", IEE Proceedings Pt. D, 137, 323-328.

Jury, E.I. (1958), "Sampled-Data Control Systems", John Wiley and Sons, New York.

Jury, E.I. (1990), "Robustness of a Discrete System", Automation and Remote Control USSR, 51, 571-592.

Kalman R.E. (1954), discussion following A.R. Bergen and J.R. Ragazzini, "Sampled-Data Processing Techniques for Feedback Control Systems", AIEE Trans., 73, 236-247.

Katbab, A. and E.I. Jury (1991), "Generalization and Comparison of Two Recent Frequency-Domain Stability Robustness Results", Int. J. Control, 53, 463-475.

Khandheria, J. and W.L. Luyben (1976), "Experimental Evaluation of Digital Algorithms for Anit-reset Windup", Ind. Eng. Chem. Process Des. Dev., 15, 278.

Koivo, H.N. (1980), "A Multivariable Self-Tuning Controller", Automatica, 16, 352-366.

Kucera, V. (1979), Discrete Linear Control, The Polynomial Equation Approach, John Wiley and Sons, New York.

Kuo, B.C. (1980), Digital Control Systems, Holt, Rheinhart, and Winston, Inc., New York.

Laughlin, D.L., D.E. Rivera and M. Morari (1987), "Smith Predictor Design for Robust Performance", Int. J. Control, 46, 477-504.

Lee, B.K., B.S. Chen and Y.U. Lin, "Model Reference Deadbeat Control", Int. J. Control, 54, 1217-1231.

Lennartson, B. (1990), "On the Choice of Controller and Sampling Period for Linear Stochastic Control", Automatica, 26, 573-578.

Lennartson, B. and T. Söderström (1989), "Investigation of the Intersample Variance in Sampled-Data Control", Int. J. Control, 50, 1587-1602.

Lewin, D.R. (1991), "Robust Performance Specifications for Uncertain Stable SISO Systems", Int. J. Control, 53, 1263-1281.

Maciejowski, J.M. and B.S. Samra (1990), "Gain/Phase Relationships for Discrete-Time Systems", Int. J. Control, 51, 119-137.

MacGregor, J.F. (1973), "Optimal Discrete Stochastic Control Theory for Process Applications", Can. J. Chem. Eng., 51, 468-478.

Private communication from J.F. MacGregor (1984), "Optimal Stochastic and Adaptive Process Control: Theory and Applications", Advanced Process Control Short Course Notes, McMaster University, Hamilton, Ontario.

MacGregor, J.F., T.J. Harris and J.D. Wright (1984), "Duality Between the Control of Processes Subject to Randomly Occuring Deterministic Disturbances and ARIMA Stochastic Disturbances", Technometrics, 26, 389-397.

Mäkilä, P.M. (1982), "Self-Tuning Control With Linear Input Constraints", Opt. Contr. App. and Meth., 3, 337-353.

Morari, M. and E. Zafiriou (1989), Robust Process Control, Prentice Hall, Englewood Cliffs, New Jersey.

Palmor, Z.J. (1980), "Stability properties of Smith dead-time compensator controllers," Int. J. Control, 32, 937-949.

Palmor, Z.J. (1982), "Properties of Optimal Stochastic Control Systems with Dead-Time," Automatica, 18, 107-116.

Palmor, Z.J. and R. Shinnar (1979), "Design of Sampled Data Controllers", Ind. Eng. Chem. Process Des. Dev., 18, 8-30.

Parrish, J.R. and C.B. Brosilow (1985), "Inferential Control Applications", Automatica, 21, 527-538.

Prasad, P.R., V.K. Chawia and P.B. Deshpande (1990), "A New Control Algorithm for Single-Input-Single-Output and Multiple-Input-Multiple-Output Systems: Applications to Multiloop Control Systems", Ind. Eng. Chem. Res., 29, 134-138.

Ragazzini, J.R. and G.F. Franklin (1958), Sampled-Data Control Systems, McGraw-Hill Book Coompany Inc., New York.

Rivera, D.E., M. Morari and S. Skogestad (1986), "Internal Model Control. 4. PID Controller Design", Ind. Eng. Chem. Process Des. Dev., 25, 252-265.

Romagnoli, J.A., M.N. Karim, O.E. Agamennoni and A. Desages (1988), "Controller Designs for Model-Plant Parameter Mismatch", IEE Proceedings Pt. D, 135, 157-164.

Rundqwist, L. (1991), Anti-Reset Windup for PID Controllers, Ph.D. Dissertation, Lund Institute of Technology.

Segall, N.L. and P.A. Taylor (1986), "Saturation of Single-Input-Single-Output Controllers Written in Velocity Form: Reset Windup Protection", Ind. Eng. Chem. Process Des. Dev., 25, 495-498.

Segall, N.L., J.F. MacGregor and J.D. Wright (1986), "One-Step Optimal Correction for Input Saturation in Discrete Model Based Controllers", 36th Canadian Chemical Engineering Conference (Microfiche), Sarnia, Ontario, October.

Segall, N.L., J.F. MacGregor and J.D. Wright (1987a), "One-Step Optimal Correction for Input Saturation in Discrete Model Based Controllers", Proceedings of IFAC 10th World Congress, Munich, July.

Private communication from N.L. Segall, J.F. MacGregor and J.D. Wright (1987b), "One-Step Optimal Correction for Input Saturation in Discrete Model Based Controllers", Report #1013, Department of Chemical Engineering, McMaster University, Hamilton, Ontario.

Segall, N.L., J.F. MacGregor and J.D. Wright (1991), "One-step Optimal Saturation Correction", Automatica, 27, 135-139.

Smith, C.A. and A.B Corripio (1985), Principles and Practice of Automatic Process Control, John Wiley and Sons, New York.

Smith, C.L. (1972), Digital Computer Process Control, Intext Educational Publishers, Scranton, Pennsylvania.

Smith, O.J.M. (1957), "Closer Control of Loops With Dead-Time", Chem. Eng. Prog., 53, 217.

Tabak, D. and B.C. Kuo (1971), "Optimal Control by Mathematical Programming", Prentice-Hall, New Jersey.

Toivonen, H. (1983a), "Suboptimal Control of Linear Discrete Stochastic Systems with Linear Input Constraints", IEEE Trans. Autom. Contr., AC-28, 246-248.

Toivonen, H. (1983b), "Suboptimal Control of Discrete Stochastic Amplitude Constrained Systems", Int. J. Control, 37, 493-502.

Private communication from H. Toivonen (1983c), "Design and Analysis of Discrete Stochastic Amplitude-Constrained Systems-Review of Procedures and Computer Program", University of Trondheim, Norwegian Institute of Technology, Technical report 83-47-W.

Tsypkin, Y.Z. (1946), "Stability of Systems with Delayed Feedback," Avtomat. Telemekh., 7, 107-129, as translated and reproduced in Frequency-Response Methods in Control Systems, A.G.J. MacFarlane ed. (1979), IEEE Press, John Wiley and Sons, New York, N.Y.

Wilson, G.T. (1970), Modelling Linear Systems for Multivariable Control, Ph.D. dissertation, University of Lancaster, U.K.

Wong, J.P.M. (1992), Design and Implementation of Model Based Controller for Systems with Input Saturation and Model Mismatch, Ph.D. Thesis, McMaster University, Hamilton, Ontario, Canada.

Wong, P.M., P.A. Taylor and J.D. Wright (1987), "An Experimental Evaluation of Saturation Algorithms for Advanced Digital Controllers", Ind. Eng. Chem. Process Des. Dev., 26, 1117-1126.

Wong, P.M., P.A. Taylor and J.D. Wright (1987), "Investigation of SISO Digital Controller Robustness in Time Domain", CSChE Conference Proceedings, Montreal

Ydstie, B.E., L.S. Kershenbaum and R.W.H. Sargent (1985), "Theory and Application of an Extended Horizon Self-Tuning Controller", AIChE J., 31, 1771-1780.

Zafiriou E. and M. Morari (1985), "Digital controllers for SISO systems: a review and a new algorithm", Int. J. Control, 42, 855-876.