MODELING REQUIREMENTS FOR PROCESS CONTROL

Thesis by

Daniel Eduardo Rivera

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1987

(Submitted November 26, 1986)

Thus says **God the Lord,**

    who created the heavens ...

I will lead the blind by a way they do not know,

In paths they do not know I will guide them.

I will make darkness into light before them

And rugged places into plains.

These are the things I will do,

    and I will not leave them undone.

Isaiah 42:5a,16 NAS

Con mucho cariño dedico esta tésis a mis padres,

Eduardo y Carmen Lina.

ACKNOWLEDGMENTS

I am thankful for the many individuals who contributed positively to my academic and personal life while at Caltech. In particular, I am indebted to the following persons:

- Professor Manfred Morari, for his leadership, motivation, high standards of excellence, and assistance in all stages of this work.

- Professor John Doyle, who eagerly and unselfishly gave advice on a number of aspects of this thesis.

- To my colleagues Evanghelos Zafiriou, Tony Skjellum, and Jorge Mandler who provided their advice and friendship. Evanghelos and Tony deserve special mention for making 134 Spalding Lab a group forum for discussing civil liberties issues, religion, politics, and sexual attitudes (all at the expense of fellow office mates Chris Webb and Mike Creed ; thanks guys, for putting up with us).

I am also grateful to the members of the Iglesia Misionera Hispanoamericana for contributing to my life in ways Caltech could not. I'd like to thank Jose and Kathy Endara, Steve and Doris Klein, Obed and Cindy Alvarez, Ramiro and Esperanza del Muro, Luis and Mercedes Perez, Esthela Gonzalez, and Alfredo Ortega for their friendship.

The housekeeping staff at Keck Dormitory have been special friends. Many thanks to Carmen de la Fuente, Hector Diaz, and Ana Rosa Garcia.

Jan Owen deserves special thanks for typing this thesis and for performing many secretarial duties on my behalf. Jan and sidekick Betty Benjamin have also been good friends. I also do not want to forget George Griffith, who always "walked the extra mile" for me.

Finally, I wish to thank the members of the Caltech Christian Fellowship for their friendship and support. I will fondly remember the moments of prayer and fellowship shared with John McGill, Jim Cline, Eric Aslakson, and others.

The generous support of a National Science Foundation Graduate Fellowship is gratefully acknowledged.

MODELING REQUIREMENTS FOR PROCESS CONTROL

by

Daniel Eduardo Rivera

ABSTRACT

Modeling and control system design have traditionally been viewed as distinct, independent problems. Not all model characteristics, however, are relevant to the control system design problem. One can expect, then, that parsimonious, more effective controllers are possible if control considerations are incorporated in the modeling stage.

The synergism of dynamic modeling and process control, as pertaining to the fields of low-order controller design, model reduction, and model identification, is investigated in this thesis. The guiding theoretical framework is the robust control paradigm using the Structured Singular Value, which addresses controller design in the presence of model uncertainty.

The main contribution of this thesis is the development of a control-relevant model reduction methodology. The effectiveness of reduction is increased by incorporating the closed-loop performance/robustness specifications, plant uncertainties, and setpoint/disturbance characteristics explicitly as weights in the reduction procedure. The efficient computation of the control-relevant reduction problem is indicated and illustrated with examples taken from the control of a methanation reactor and a binary distillation column.

ABSTRACT (Continued)

A low-order controller design methodology for single-input, single-output plants is also presented. The basis for this methodology is the combination of the control-relevant reduction problem with the Internal Model Control (IMC) design procedure. The relationship between low-order IMC controllers and classical feeback compensators is examined. It is shown that for many models common to the process industries, the controllers obtained from the low-order compensator design technique are of the PID type.

Finally, a model identification methodology is established using spectral time series analysis to obtain plant transfer function and uncertainty estimates directly from experiments. The control-relevant model reduction procedure can then be used to fit the "full-order" frequency response to a "reduced-order" parametric model. Model validation for control purposes is achieved by insuring that the robustness condition is satisfied.

TABLE OF CONTENTS

CHAPTER I: INTRODUCTION

1. Motivation

Today's chemical industry is faced with new and significant challenges to its once glorious past. While many segments of the industry experienced substantial growth during the 60's and early 70's, American chemical and refining companies during the late 70's and 80's have witnessed drastic changes in the economic climate. Among the factors that have affected chemical manufacturers are detrimental exchange and interest rates, production overcapacity, heightened competition, and fluctuating feedstock prices (e.g., petroleum). Many companies have had to abandon the heady expectations of the past and learn to "ride the tiger" (Hirsig and Schlanger, 1984), that is, learn to cope with up and down business cycles.

The need for improved process control has emerged as one of the crucial requirements for success in an uncertain economic environment. As indicated by Shell Development's Garcia and Prett (1986), modifying process designs to achieve maximum profitability requires that market conditions remain unchanged for a substantial period of time (usually in excess of five years) in order to achieve a return on investment. A much more cost-effective alternative is to apply advanced control and optimization strategies on existing equipment in order to identify and maintain optimal process conditions. Shell is not the only major corporation that has recognized the merits of advanced control. At Du Pont, for example, a recent company-wide survey estimates that savings amounting to $ 500,000,000.00 per year can be obtained through improved control (Smith, 1986).

The benefits of advanced control extend to both the specialty and bulk chemical businesses. One example is Amoco Chemicals, where the manufacture of a wide variety of specialty acids is accomplished using the same process equipment (Ali, 1986). Advanced control strategies are needed to provide the flexibility necessary for achieving optimum performance for each reaction system.

Faced with such a vast challenge, the control engineer must be equipped with adequate tools that meet the fundamental needs of industrial practice. In particular, these tools must generate controllers that are robust to plant uncertainty and input constraints, are easily adjustable on-line, and are parsimonious, that is, they are expressed as compactly as possible. Furthermore, these tools must be simple enough (either computationally or through user-friendly, possibly "expert" computer-aided design software) such that plant personnel can confidently perform the designs.

A good model is the first requirement for a successful design strategy that meets the objectives outlined in the previous paragraph. In chemical engineering systems, dynamic models vary greatly in terms of size and complexity. While the temperature response of a stirred tank, for example, can be described by simple first-order lag with deadtime model, the concentration response of a staged separation process, such as a distillation column, is fundamentally modeled by a large system of differential equations. Distributed parameter systems, such as tubular reactors, also give rise to complex modeling problems

(Mandler et. al, 1986).

High-order, complex process models present difficulties for both control system design and implementation. The need for simple, practical controllers has spawned a number of research areas in control system engineering. These are:

Low-Order Controller Design. Efforts most commonly associated with this field are simple tuning rules for PID controllers, such as those provided by Ziegler and Nichols (1942), which require a only a minimal amount of information regarding the plant (e.g., the frequency response at crossover). Other more general procedures, involving nonlinear programming or the reduction of a high-order compensator, also characterize this area.

Model Reduction. The traditional model reduction problem consists of reducing the number of states of a high-order linear model either by matching the predominant features of the full model (statistical moments, dominant poles, etc.) or by minimizing a functional objective, such as the integral square error between full and reduced models. Based on the reduced-order model, control system design is more readily achieved. Designing controllers from reduced-order plants is also a common low-order controller design technique.

A good summary of model reduction literature is found in Bosley and Lees (1972).

Model Identification. This field bypasses the need for first-principles modeling by indicating how a suitable process model can be generated from experiments. A good survey paper on the subject is that by Astrom and Eykhoff (1971).

The problems of low-order controller design, model reduction, and model identification have received extensive treatment in the literature. What motivation is there, then, for any further study of these topics ? The main issue is that the relationship between the modeling and control problems has not been examined to a satisfactory level. In many cases, mathematical convenience has given second place to a fundamental understanding of what constitutes an adequate model for control purposes.

The interrelationship between control and modeling is particularly significant in the the fields of model reduction and identification, where the tendency is to view the control and reduction/identification problems as mutually exclusive. Most work in this area incorrectly assumes that a "separation principle" holds between both problems. The effectiveness of these techniques, however, can be improved if control considerations are incorporated in the problem statement.

The desire to study the aforementioned problems from the viewpoint of robust control (Doyle, 1984) is also a motivating factor for this study. In particular, we wish to define reduction, identification, and low-order controller design methodologies which are derived directly from the robust control paradigm.

Finally, we do not wish to lose sight of the needs of the practicing engineer. It is intended that the proposed reduction, identification, and low-order design methodologies be simple to code and computationally efficient. Physically meaningful quantities should be requested from the users of this software, and the resulting controller designs should be robust, parsimonious, and easily adjustable on-line.

2. Thesis Overview

The main focus of this thesis centers on developing a control-relevant model reduction and identification methodology, based on the robust control paradigm. The extension to the low-order controller design arises naturally by applying the Internal Model Control (IMC) design procedure (Morari et al., 1987) to a control-relevant reduced model.

The thesis is organized as follows. **Chapter II** introduces the IMC design procedure and examines its relation to classical feedback. It is seen that for most models common to the process industries, the "optimal" controller is of the PID type. **Chapter III** defines control-relevant plant and controller reduction problems for single-input, single-output (SISO) systems and outlines the efficient solution of these problems using quadratic and linear programming. In **Chapter IV**, the control-relevant reduction procedure is extended to multivariable systems. **Chapter V** combines Chapters II and III, resulting in a comprehensive low-order controller tuning methodology for SISO systems. **Chapter VI** represents ground-breaking efforts on the control-relevant model identification problem, outlining the use of spectral analysis to obtain the plant frequency response and an associated uncertainty description directly from experiments. Control-relevant reduction serves a dual purpose as a parameter estimation procedure for control purposes. Conclusions and suggestions for future research are given in **Chapter VII.**

References

Ali, M.H. Amoco Chemical Research and Development, personal communication (1986).

Astrom, K.J. and P. Eykhoff. Automatica, 7, 123 (1971).

Bosley, M.J. and F.P. Lees. Automatica, 8, 765 (1972).

Doyle, J., Notes from the Honeywell/ONR Workshop on Advances in Multivariable Control, Minneapolis (1984).

Garcia, C.E. and D.M. Prett, "Advances in Model Predictive Control," Proceedings of the Third International Conference on Chemical Process Control, Asilomar, California, (1986).

Hirsig, A.R. and M.O. Schlanger. Chemical Eng. Progress, 80, 24 (1984).

Morari, M., E. Zafiriou, and C.G. Economou. Robust Process Control, Springer-Verlag, to appear (1987).

Mandler, J.A., M. Morari, and J.H. Seinfeld, Chemical Eng. Sci., 41, 1577 (1986).

Smith, W.D. Du Pont Co., personal communication (1986).

Ziegler, J.G. and N.B. Nichols. Trans. ASME, 64, 759 (1942).

CHAPTER II:

INTERNAL MODEL CONTROL: PID CONTROLLER DESIGN

(Published in Ind. and Eng. Chem. Proc. Des. and Dev., 25, 252 (1986))

INTERNAL MODEL CONTROL: PID CONTROLLER DESIGN

Daniel E. Rivera

Manfred Morari

Sigurd Skogestad

Chemical Engineering, 206-41

California Institute of Technology

Pasadena, California 91125

Abstract

For a large number of single input-single output (SISO) models typically used in the process industries, the Internal Model Control (IMC) design procedure is shown to lead to PID controllers, occasionally augmented with a first-order lag. These PID controllers have as their only tuning parameter the closed-loop time constant or equivalently, the closed-loop bandwidth. On-line adjustments are therefore much simpler than for general PID controllers. As a special case, PI and PID tuning rules for systems modeled by a first-order lag with deadtime are derived analytically. The superiority of these rules in terms of both closed-loop performance and robustness is demonstrated.

1. Introduction

Synthesis and tuning of control structures for SISO systems comprises the bulk of process control problems. In the past, hardware considerations dictated the use of the PID controller, but through the use of computers controllers have now advanced to the stage where virtually any conceivable control policy can be implemented. Despite these advances, the most widely used controller is still of the PID-type. Finding design methods which lead to the optimal operation of PID controllers is therefore of significant interest.

For controller tuning, simplicity, as well as optimality, is important. The three modes of the ordinary PID controller, k_c, τ_I, and τ_D, do not readily translate into the desired performance and robustness characteristics which the control system designer has in mind. The presence of simple rules which relate model parameters and/or experimental data to controller parameters serves to simplify the task of the designer.

The literature contains a number of these "tuning rules"; possibly the best known are the Ziegler-Nichols rules proposed in 1942. Given the wide use of the first-order lag/deadtime model for chemical processes, tuning rules for PID control of this structure have received wide attention in the literature. Most common are the rules proposed by Cohen and Coon (1953). Smith (1972) contains a good summary of efforts in this area.

Our intention is to present a clearer and more logical framework for PID-controller design which is simple to understand and implement while possessing a sound fundamental basis. Instead of fixing a control structure and then attempting to "extract" optimality from this controller (as is usually the case with classical methods), our approach will be to postulate a model, state desirable control objectives, and from these proceed in a straightforward manner to obtain both the appropriate controller structure and parameters.

The Internal Model Control (IMC) structure provides a suitable framework for satisfying these objectives. IMC was introduced by Garcia and Morari (1982) but a similar concept has been used previously and independently by a number of other researchers. Using the IMC design procedure, controller complexity depends exclusively on two factors: the complexity of the model and the performance requirements stated by the designer. The goal of this article is to show that for the objectives and simple models common to chemical process control, the IMC design procedure leads naturally to PID-type controllers, occasionally augmented by a first-order lag. Furthermore, the proposed procedure provides valuable insight regarding controller tuning effects on both performance and robustness.

2. Performance and Robustness Measures

Probably the best indicator of performance is the sensitivity function

$$S = \frac{1}{1+gc} = \frac{e}{y_s-d} \tag{1}$$

(The nomenclature should be apparent from Fig. 1). It is desirable to keep the sensitivity function small over as wide a frequency range as possible. For any proper system, $|S|$ will approach unity as the frequency becomes large. Instead of the sensitivity function, the closed loop bandwidth can be used as a simple performance measure; it is the frequency ω_b at which $|S|$ first reaches $1/\sqrt{2}$

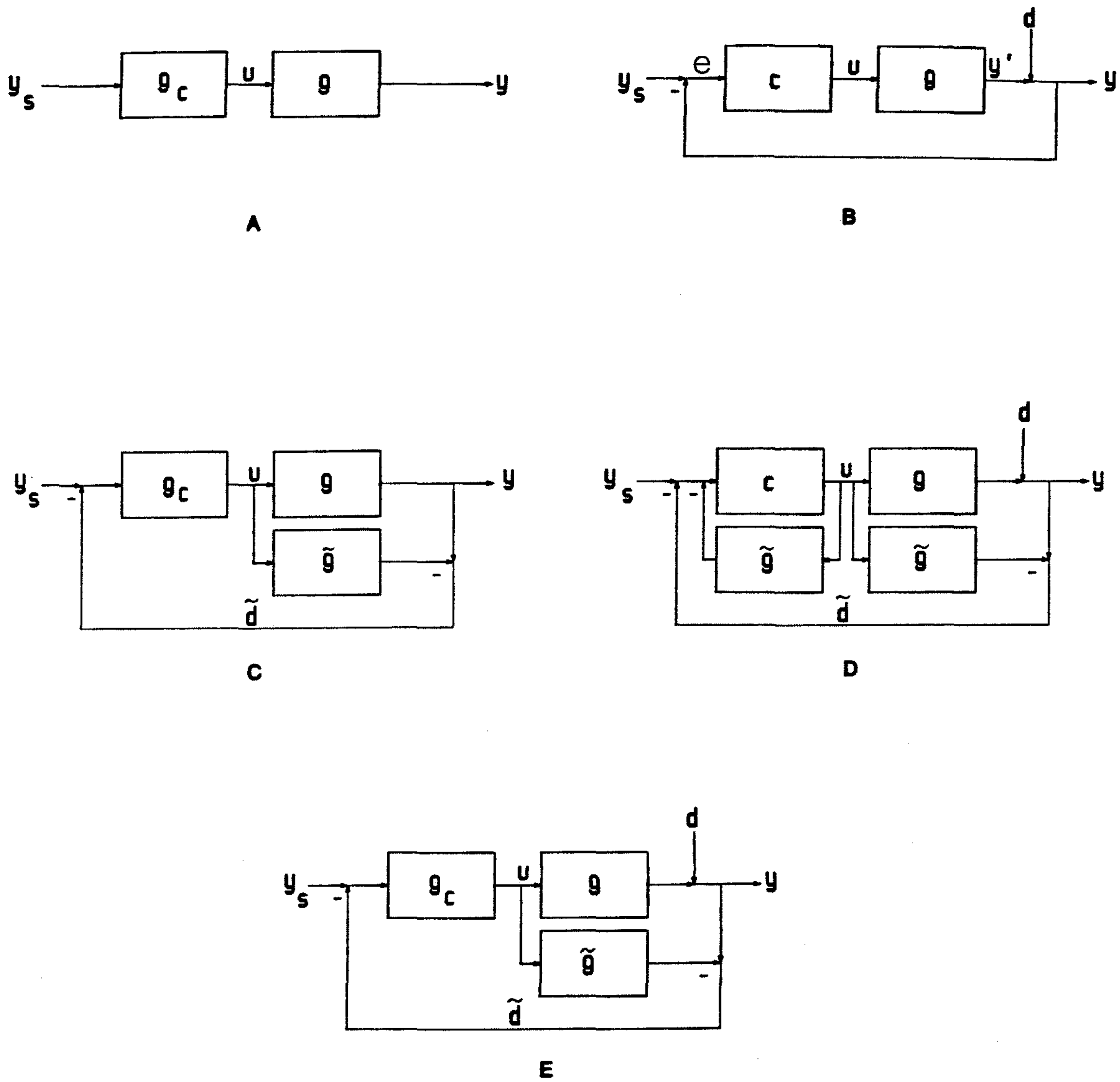


Figure 1. Evolution of the IMC structure.

$$|S| < \frac{1}{\sqrt{2}} \qquad \forall\ \omega < \omega_b \tag{2}$$

Increasing the bandwidth implies less attenuation of the reference signal, more effective disturbance rejection, and a faster response. For a phase margin (PM) less than or equal to $\frac{\pi}{2}$ (the most common situation) the bandwidth is less than or equal to the (gain) crossover frequency ω_c, defined as the frequency at which the open-loop gain first drops to unity

$$|gc| > 1 \qquad \forall\ \omega < \omega_c \tag{3}$$

Occasionally, we will also refer to the Integral Square Error (ISE) and to the Integral Absolute Error (IAE) for a specified setpoint or disturbance change to compare the performance of different controllers:

$$J = ISE = \int_0^\infty (y-y_s)^2 dt \tag{4}$$

$$J' = IAE = \int_0^\infty |y-y_s| dt \tag{5}$$

It is crucial in control system design to insure the stability and performance of the closed-loop system in the presence of plant/model mismatch, i.e.,to guarantee robustness. We will use a superscript (˜) to distinguish the (known) model and its properties from the (generally unknown) "real" plant. Plant/model mismatch can be caused, for example, by model reduction, (the representation of a high-order system by a low-order approximate model) or by system parameters which depend on the operating conditions. Though we do not know the real plant g it is often reasonable to assume it to be a member of a family Π of linear plants defined by a norm bounded multiplicative error e_m

$$\Pi = \{g: |e_m| \leq \ell_m\} \tag{6}$$

where

$$e_m = \frac{g-\tilde{g}}{\tilde{g}} \tag{7}$$

Usually $|e_m|$ approaches a value equal to or greater than 1 for high frequencies.

We will also establish in the following that the <u>complementary sensitivity function</u> $\tilde{H}$

$$\tilde{H} = \frac{\tilde{g}c}{1+\tilde{g}c} \tag{8}$$

is a good robustness measure. The name "complementary sensitivity" follows from the equality

$$\tilde{H} + \tilde{S} = 1 \tag{9}$$

Let us assume that g, $\tilde{g}$ and c have no poles in the open right-half plane (RHP) and that the closed-loop system with the "nominal" plant $\tilde{g}$ and the controller c is stable. Then Doyle and Stein (1981) have shown that the closed-loop system is stable for all plants in the family Π if and only if

$$|\tilde{H}| < \frac{1}{\ell_m} \qquad \forall\omega \tag{10}$$

Because ℓ_m increases with frequency and eventually exceeds 1, $|\tilde{H}|$ has to drop below 1 at some frequency. Because of (9), $|\tilde{S}|$ has to be close to 1 in this frequency range. Thus the achievable closed-loop bandwidth is limited by the bandwidth over which the process model is good. The smallest uncertainty $\ell_m(\omega)$ is allowed at the frequency where $|\tilde{H}(j\omega)|$ has its maximum peak. As a consequence the M-value defined by (11) (e.g.,Rosenbrock, 1974)

$$M = \max_{\omega} |\tilde{H}| \tag{11}$$

is a suitable robustness indicator.

M is convenient and widely accepted as more useful than gain margins (GM) or phase margins (PM). Gain and phase margins only measure robustness with respect to model uncertainties which are independent of ω, and thus tend to be overly optimistic. The following relationships indicate how M establishes lower bounds on GM and PM:

$$GM \geq 1 + \frac{1}{M} \tag{12}$$

$$PM \geq 2 \sin^{-1} \left(\frac{1}{2M}\right) \cong \frac{1}{M} \tag{13}$$

For the special case of M=1, (12) and (13) become

$$GM \geq 2 \tag{14}$$

$$PM \geq 60^{\circ} \tag{15}$$

One must note that M by itself yields only a qualitative indication of robustness. The allowable uncertainty in specific model parameters can be deduced from M only when the bandwidth ω_b is known. Consider, for example, an analysis of the allowable deadtime error in a closed-loop system (the plant deadtime exceeds that of the model by the quantity δ):

$$g = \tilde{g}e^{-s\delta} \tag{16}$$

Because the deadtime error introduces a phase lag of $\omega\delta$ at frequency ω, the system will remain stable for the deadtime error δ if

$$\delta < \frac{PM}{\omega_c} \tag{17}$$

Substituting (13), this exact condition can be replaced by the more conservative condition

$$\delta < \frac{1}{\omega_c M} \tag{18}$$

For PM = 90°, $\omega_b = \omega_c$ and (18) becomes

$$\delta < \frac{1}{\omega_b M} \tag{19}$$

Equation (19) clearly illustrates the trade-off between performance and robustness. Good performance (high ω_b) is obtained only at the expense of robustness (small allowed deadtime error).

Our study is aimed at systems of "type-1" and "type-2" (Wiberg, 1971):

$$\text{Type 1:} \quad \lim_{s\to 0} \; sgc \neq 0 \tag{20}$$

$$\text{Type 2:} \quad \lim_{s\to 0} \; s^2gc \neq 0 \tag{21}$$

Type 1 and Type 2 systems exhibit no offset to step and ramp changes on (y_s-d) respectively. Furthermore, the following limits hold:

$$\lim_{s\to 0} \; H(s) = 1 \tag{22}$$

$$\lim_{s\to 0} \; S(s) = 0 \tag{23}$$

3. Internal Model Control (IMC)

3. 1. Fundamentals

The goal of control system design is fast and accurate setpoint tracking

$$y \cong y_s \quad \forall \; t, \quad \forall \; d \tag{24}$$

This implies that the effect of external disturbances should be corrected as efficiently as possible (good regulatory behavior)

$$y' \cong y_s - d \quad \forall \; t, \quad \forall \; d \tag{25}$$

Furthermore, the control system designer wishes to obtain (24) and (25) while also being assured of insensitivity to modeling error.

It is well known that an open-loop (feedforward) arrangement (Fig. 1A) represents the optimal way to satisfy (24). For the open-loop scheme, the stability question is trivial (the system is stable when both the controller and the system are stable); also the controller is easy to design ($g_c=\tilde{g}^{-1}$). The disadvantages are the sensitivity of the performance to plant/model mismatch and the inability to cope with unmeasured disturbances. With the feedback arrangement (Fig. 1B) the situation is reversed. Plant/model mismatch and unmeasured disturbances can be dealt with effectively, but tuning is complicated by the closed-loop stability problem.

We can now augment the open-loop and closed-loop systems as indicated in Fig. 1C and 1D without affecting performance: In Fig. 1C, $\tilde{d} = 0$, and therefore the system is still open loop, in Fig. 1D the two blocks $\tilde{g}$ cancel each other. Relating Fig. 1C and 1D through the definitions

$$g_c = \frac{c}{1+c\tilde{g}} \tag{26}$$

$$c = \frac{g_c}{1-\tilde{g}g_c} \tag{27}$$

we arrive at the general structure in Fig. 1E which has the advantages of both the open-loop and closed-loop structures: When the model of the plant is perfect ($g=\tilde{g}$) and there are no disturbances ($d=0$), feedback is not needed and structure E behaves identically to structure A. Because the plant model $\tilde{g}$ appears explicitly in E, this structure is referred to as the Internal Model Control (IMC) structure. As a simplification we can say that the controller in E can be designed with the ease of an open-loop controller while retaining the benefits of a feedback system. It is our goal to describe, in detail, such a design procedure.

From the block diagram for the IMC structure (Fig. 1E) follow the relationships

$$u = \frac{g_c}{1+g_c(g-\tilde{g})}\,(y_s-d) \tag{28}$$

$$y = \frac{gg_c}{1+g_c(g-\tilde{g})}(y_s-d) + d = H(y_s-d) + d \tag{29}$$

$$e = y_s - d = \frac{1-g_c\tilde{g}}{1+g_c(g-\tilde{g})}(y_s-d) = S(y_s-d) \tag{30}$$

Four properties can be shown which suggest the advantages of this structure:

P1: Dual Stability. Assume $g = \tilde{g}$. Then the system is effectively open-loop and "closed-loop stability" is implied by the stability of g and g_c:

$$y = gg_c(y_s-d) + d \tag{31}$$

While for the classical structure (Fig. 1B) it is not at all clear what type of controller c and what parameter choices lead to closed-loop stable systems, the IMC structure guarantees closed-loop stability for all stable controllers g_c.

P2: Perfect control. Assume that the controller is equal to the model inverse ($g_c=\tilde{g}^{-1}$) and that the closed-loop system in Fig. 1E is stable. Then $y = y_s$ for all $t > 0$ and all disturbances d(t).

P3: Type 1 system. Assume that the controller steady-state gain is equal to the inverse of the model gain

$$g_c(0) = \tilde{g}(0)^{-1} \tag{32}$$

and that the closed-loop system in Fig. 1E is stable. Then the system is of type 1 and the control error vanishes asymptotically for all asymptotically constant inputs y_s and d. This property implies no offset at steady-state, and follows from (30) via the Final Value Theorem.

P4: Type 2 system. Select g_c to satisfy P3 and

$$\frac{d}{ds}(g_c\tilde{g})\Big|_{s=0} = 0 \tag{33}$$

Then the system is of type 2 and the control error vanishes asymptotically for all asymptotically ramp shaped inputs y_s and d (Brosilow, 1983). (P4 also follows from (30) via the Final Value Theorem).

P1 simply expresses the fact that in the absence of plant/model mismatch the stability issue is trivial, as long as the open-loop system is stable. P2 asserts that the ideal open-loop controller leads to perfect closed-loop performance when the IMC structure is employed. P3 and P4 state that inherent integral action can be achieved without the need for introducing additional tuning parameters. P2, however, represents an idealized situation. We know intuitively that P2 requires an infinite controller gain; this is confirmed by substituting $g_c = \tilde{g}^{-1}$ in (27). By setting $g_c(0) = \tilde{g}(0)^{-1}$ as postulated for P3 we find $c(0) = \infty$, which implies integral control action, as expected.

There are several reasons why the "perfect controller" implied by P2 cannot be realized in practice.

1. Right half plane (RHP) zeros: If the model has a RHP zero, the controller $g_c = \tilde{g}^{-1}$ has a RHP pole and if $\tilde{g} = g$ the closed-loop system will be unstable according to P1.
2. Time delay: If the model contains a time delay, the controller $g_c = \tilde{g}^{-1}$ is predictive and cannot be realized.
3. Constraints on the manipulated variables: If the model is strictly proper, then the perfect controller $g_c = \tilde{g}^{-1}$ is improper, which implies $\lim_{\omega \to \infty} |g_c| = \infty$. Thus infinitely small high-frequency disturbances would give rise to infinitely large excursions of the manipulative variables which are physically unrealizable.

4. Modeling error: If $g \neq \tilde{g}$, P1 does not hold and the closed-loop system will generally be unstable for the controller $g_c = \tilde{g}^{-1}$.

In resolving these four issues, the ideal of perfect control must be abandoned. The IMC design procedure handles this in two steps; first, performance is addressed with no regards to robustness or input constraints. Next, a filter is introduced and designed for properness (input constraints) and robustness without looking at how this affects the performance. Though there obviously does not exist any separation principle which makes this approach "optimal", the design procedure is very simple and direct. Also, there seem to be very few cases where other more complicated and indirect procedures (e.g.,LQG) give better results. The freedom which the designer is given to choose the filter makes it possible to take into account considerations which may be difficult to pin down mathematically.

Step 1: Factor the model

$$\tilde{g} = \tilde{g}_+ \tilde{g}_- \tag{34}$$

such that $\tilde{g}_+$ contains all the time delays and RHP zeros; consequently $\tilde{g}_-^{-1}$ is stable and does not involve predictors.

Step 2: Define the IMC controller by

$$g_c = \tilde{g}_-^{-1} f \tag{35}$$

where f, a low-pass filter, must be selected such that g_c is proper or, if "derivative" action is allowed (as in the ideal PID controller), such that g_c has a zero excess of at most 1. By definition of the factorization in (34), g_c is realizable.

Having introduced these definitions, the closed-loop relationships (29) and (30) become

$$y = \frac{\tilde{g}_+ f(1+e_m)}{1+\tilde{g}_+ f e_m}(y_s+d) + d = \frac{\tilde{H}(1+e_m)}{1+\tilde{H}e_m}(y_s-d) + d \tag{36}$$

$$e = y_s - y = \frac{1-\tilde{g}_+ f}{1+\tilde{g}_+ f e_m}(y_s-d) = \frac{1-\tilde{H}}{1+\tilde{H}e_m}(y_s-d) \tag{37}$$

For the special case of a perfect model (e_m=0) (36) and (37) reduce to

$$y = \tilde{g}_+ f(y_s-d) + d = \tilde{H}(y_s-d) + d \tag{38}$$

$$e = (1-\tilde{g}_+ f)(y_s-d) = \tilde{S}(y_s-d) \tag{39}$$

Eqs. (38) and (39) demonstrate clearly that for the case of no plant/model mismatch, the nominal closed-loop transfer function $\tilde{H}$ = $\tilde{g}_+ f$ is at the designer's discretion except that 1) $\tilde{g}_+$ must contain all the delays and RHP zeros and 2) f must be of sufficiently high order to avoid physically unrealizable control action. Thus the closed-loop transfer function can be designed directly and not ambiguously via c as in the classic controller design procedure (Fig. 1B).

Our treatment is not complete without indicating how to select $\tilde{g}_+$ and f.

3.2. Factorization of $\tilde{g}$

Assume $\tilde{g}$ = g. For step inputs in y_s and d, selecting $\tilde{g}_+$ and f such that $|\tilde{g}_+ f| = 1 \quad \forall\omega$ minimizes the ISE (Holt and Morari, 1984, 1985). This implies that f must be unity and that $\tilde{g}_+$ has the form of an allpass

$$\tilde{g}_+ = e^{-\theta s} \prod_i \frac{-\beta_i s+1}{\beta_i s+1} \qquad \mathrm{Re}(\beta_i) > 0 \tag{40}$$

where β_i^{-1} are all the RHP zeros and θ is the time delay present in $\tilde{g}$. As a consequence of this factorization, poles corresponding to the LHP image of the RHP zeros have been added to the closed-loop response.

For step inputs in y_s and d selecting f to be unity and $\tilde{g}_+$ as

$$\tilde{g}_+ = e^{-\theta s} \prod_i (-\beta_i s+1) \qquad \mathrm{Re}(\beta_i) > 0 \tag{41}$$

minimizes the IAE (Holt and Morari, 1984, 1985).

When $\tilde{g}$ is a minimum phase model, $\tilde{g}_+ = 1$.

3.3. Filter Selection

In order to satisfy P3 (zero offset to step inputs) we adopt the following convention for $\tilde{g}_+(s)$ and $f(s) = p(s)/q(s)$

$$\tilde{g}_+(0) = p(0) = q(0) = 1 \tag{42}$$

The simplest filter f satisfying (42) is of the form

$$f(s) = \frac{1}{(\varepsilon s+1)^r} \tag{43}$$

where r is sufficiently large to guarantee that the IMC controller g_c is proper. If $g = \tilde{g}$ and $\tilde{g}_+ = 1$ (i.e., the model is minimum phase) then $y/y_s = \tilde{H} = f$. The parameter ε , which can be adjusted by the operator, determines the speed of response. For a minimum phase system, the bandwidth is proportional to $1/\varepsilon$

$$\omega_b = \omega_c = 1/\varepsilon \quad \text{for } r = 1 \tag{44}$$

$$1/\varepsilon > \omega_b > 1/r\varepsilon > \omega_c \quad \text{for } r > 1 \tag{45}$$

For nonminimum phase systems, the achievable bandwidth is inherently limited by the plant. For example, consider the following representative factorizations with $\varepsilon = 0$

$$\tilde{g}_+ = e^{-s\theta}: \qquad \omega_c = \frac{\pi}{3\theta} \qquad \omega_b = \frac{0.724}{\theta} = 0.69\omega_c \tag{46}$$

$$\tilde{g}_+ = \frac{-\beta s+1}{\beta s+1}: \qquad \omega_c = \frac{1}{\beta\sqrt{3}} \qquad \omega_b = \frac{1}{\beta\sqrt{7}} = 0.65\omega_c \tag{47}$$

$$\tilde{g}_+ = -\beta s+1: \qquad \omega_c = \infty \qquad \omega_b = \frac{1}{\beta\sqrt{2}} \tag{48}$$

For $\varepsilon > 0$, ω_c and ω_b decrease from the bounds established through (46)-(48). For $r = 1$, exact formulas for the bandwidth and crossover are included in Appendix A; these are effectively approximated by

$$\tilde{g}_+ f = \frac{e^{-s\theta}}{\varepsilon s+1} : \qquad \omega_c \cong \omega_b \cong \frac{1}{\theta+\varepsilon} \tag{49}$$

$$\tilde{g}_+ f = \frac{-\beta s+1}{\beta s+1}\,\frac{1}{(\varepsilon s+1)} : \qquad \omega_c \cong \omega_b \cong \frac{1}{2\beta+\varepsilon} \tag{50}$$

$$\tilde{g}_+ f = \frac{-\beta s+1}{\varepsilon s+1} : \qquad \omega_c \cong \omega_b \cong \frac{1}{\beta+\varepsilon} \tag{51}$$

One notices from these expressions that until $1/\varepsilon$ is of an order of magnitude comparable to θ or β respectively, ω_c and ω_b are virtually unaffected by the presence of the filter. Thus making ε very small for non-minimum phase systems has little effect on the bandwidth and performance but is very detrimental to the robustness, as we will see later. For ε large compared to θ or β approximately the same proportionality holds as for MP systems (Eq. 44).

ε is directly related to important closed-loop characteristics, unlike the parameters available in the general lead/lag network c of the classical structure (e.g., PID controllers). The larger ε is, the slower the response and the smaller the actions of the manipulated variable. With (43), the maximum peak for $|f|$ is 1, i.e. , the robustness characteristics are good.

For $r > 1$, filter forms other than (43) can lead to faster response. For example, for $r = 2$ the filter

$$f = \frac{1}{\varepsilon^2 s^2+2\zeta\varepsilon s+1} \tag{52}$$

with damping factor $\zeta = 0.5$ minimizes the ISE (Frank, 1974); however, with this filter $|f|_{max} = 1.15$, thus performance improvement occurs at the expense of a reduced robustness margin. In practice, choosing filters with structures more general than (43) is usually not worthwhile.

Additional conditions on f are necessary in order to satisfy P4 (zero offset to ramp inputs). With the adopted conventions (42), (33) becomes

$$\tilde{g}_+'(0) = q'(0) - p'(0) \tag{53}$$

where the prime denotes differentiation with respect to s. An example of a filter satisfying (53) is

$$f = \frac{(2\varepsilon - \tilde{g}_+'(0))s+1}{(\varepsilon s+1)^2} \tag{54}$$

where, as before, the adjustable parameter ε is, for minimum phase systems the closed-loop time constant and $1/\varepsilon$ is proportional to the closed-loop bandwidth. Values of $\tilde{g}_+'(0)$ for representative factorizations are

$$\frac{d}{ds}\,(e^{-s\theta})\Big|_{s=0} = -\theta \tag{55}$$

$$\frac{d}{ds}\,(-\beta s+1)\Big|_{s=0} = -\beta \tag{56}$$

$$\frac{d}{ds}\,\left(\frac{-\beta s+1}{\beta s+1}\right)\Big|_{s=0} = -2\beta \tag{57}$$

Because in general $\tilde{g}_+'(0) < 0$, one obtains

$$M = \max_{\omega} |\tilde{H}| > 1\ , \tag{58}$$

i.e., M is strictly greater than unity for all filters satisfying (53). Again, the tighter performance specification (no offset for ramps) is paid for with decreased robustness margins.

3.4. Accounting for Modeling Error

Thus far, all the discussion on filter selection has assumed a perfect model, in which case ε can be selected freely; this is not the case in practice where plant/model mismatch exists. It follows from (10) that for robust stability of the closed-loop system

$$\left|\tilde{g}_+ f\right| < \frac{1}{\ell_m} \qquad \forall\ \omega \tag{59}$$

Assuming for simplicity $\left|\tilde{g}_+\right| = 1$ it becomes clear that the filter magnitude $|f|$ must be small wherever the plant/model mismatch e_m is large. Because ℓ_m approaches or exceeds 1 for high frequencies in all practical situations, we find again that <u>the allowable range for ε is limited by the degree of plant/model mismatch</u>. As stated previously, the closed-loop bandwidth can never be larger than the bandwidth over which the process model is valid. The models used in process control are usually good enough to set $1/\varepsilon$ at least equal to the open-loop bandwidth.

In the presence of plant/model mismatch, the structure of f fails to automatically guarantee the shape of the response. However, for the suggested IMC design procedure ($\tilde{g}_+(0)f(0) = 1$ and $f = (\varepsilon s+1)^{-r}$) and using the Triangle Inequality

$$|e| \le \frac{\left|1-\tilde{g}_+ f\right|}{1-\left|\tilde{g}_+ f e_m\right|} \left|y_s-d\right| \tag{60}$$

one can discern general frequency intervals for which (60) and the ideal error function

$$|e| = \left|1-\tilde{g}_+ f\right|\left|y_s-d\right| = \left|\tilde{S}\right|\left|y_s-d\right| \tag{61}$$

are very similar. At low frequency, ($\omega \ll 1/\varepsilon$), $\tilde{g}_+ f \cong 1$ and $e \cong 0$. For $\omega \gg 1/\varepsilon$, $|f|$ is exceedingly small, $\left|\tilde{g}_+ f e_m\right| \cong 0$ and (60) and (61) become close to each other. For $\omega \cong 1/\varepsilon$ the situation is uncertain. We conclude that for ε sufficiently large the closed-loop response to high frequency or low frequency inputs (e.g., steps) will become similar to the response of the nominal system $\tilde{g}_+ f$.

In summary, the key advantage of the IMC design procedure is that all controller parameters are related in a unique, straightforward manner to the model parameters. There is only one adjustable parameter ε which has intuitive appeal because it determines the speed of response of the system. Furthermore, ε is approximately proportional to the closed-loop bandwidth which must always be smaller than the bandwidth over which the process model is valid. This leads to a good initial estimate of ε, which can be adjusted on-line if necessary.

4. IMC in the Context of Classical Control

For linear systems the IMC controller g_c represents an alternate parametrization of the classic controller c, albeit with very useful properties. Through the transformation

$$c = \frac{g_c}{1-g_c\tilde{g}} = \frac{\tilde{g}_-^{-1}}{f^{-1}-\tilde{g}_+} \tag{62}$$

Fig. 1B and 1E become equivalent. If there is no delay in g, c is rational and can be implemented as a lead/lag network. Indeed, for minimum phase systems ($\tilde{g}_+=1$) and a first-order filter ($f=(\varepsilon s+1)^{-1}$), c becomes

$$c = \frac{1}{\varepsilon}\,\frac{\tilde{g}^{-1}}{s} \tag{63}$$

4.1. IMC Implemented as a PID Controller (Table 1)

Naturally, one would expect that for certain process models, the lead/lag network c obtained from (62) via the IMC design procedure is a PID controller. Indeed, we find that IMC leads to PID controllers for virtually all models common in industrial practice (Table 1). Note that Table 1 includes systems with pure integrators and RHP zeros. Occasionally, the PID controllers are augmented by a first-order lag with time constant τ_F. A few remarks regarding Table 1 are

Table 1: IMC-based PID controller parameters

Controller form:

$$c = \frac{1}{(\tau_F s+1)} k_c \left(1+ \frac{1}{\tau_I s} + \tau_D s\right)$$

ε is the only adjustable parameter; for most cases ε is equivalent to the closed-loop time constant, $1/\varepsilon$ is approximately the closed-loop bandwidth.

In all cases there exists no offset for step setpoint/disturbance changes.

Comments:

1. ISE optimal for step setpoint changes when $\varepsilon = 0$.
2. IAE optimal for step setpoint changes when $\varepsilon = 0$.
3. ISE optimal for step setpoint changes when $\varepsilon = \beta$.
4. Filter/factorization option 1. (64) Practical recommendation $\varepsilon > \beta/2$.
5. Filter/factorization option 2. (66) Practical recommendation $\varepsilon > \beta$.
6. No offset for ramp setpoint/disturbance changes.

	Model	$\frac{y}{y_s} = \tilde{g}_+ f$	Controller	$k_c k$	τ_I	τ_D	τ_F
A	$\frac{k}{\tau s+1}$	$\frac{1}{\epsilon s+1}$	$\frac{1}{k}\frac{\tau s+1}{\epsilon s}$	$\frac{\tau}{\epsilon}$	τ	-	
B	$\frac{k}{(\tau_1 s+1)(\tau_2 s+1)}$	$\frac{1}{\epsilon s+1}$	$\frac{(\tau_1 s+1)(\tau_2 s+1)}{k\epsilon s}$	$\frac{\tau_1+\tau_2}{\epsilon}$	$\tau_1+\tau_2$	$\frac{\tau_1\tau_2}{\tau_1+\tau_2}$	-
C	$\frac{k}{\tau^2 s^2+2\zeta\tau s+1}$	$\frac{1}{\epsilon s+1}$	$\frac{\tau^2 s^2+2\zeta\tau s+1}{k\epsilon s}$	$\frac{2\zeta\tau}{\epsilon}$	$2\zeta\tau$	$\frac{\tau}{2\zeta}$	-
D	$k\frac{-\beta s+1}{\tau s+1}$	(2,3,5) $\frac{-\beta s+1}{\epsilon s+1}$	$\frac{\tau s+1}{k(\beta+\epsilon)s}$	$\frac{\tau}{\beta+\epsilon}$	τ	-	-
E	$k\frac{-\beta s+1}{\tau s+1}$	(1,4) $\frac{-\beta s+1}{(\beta s+1)(\epsilon s+1)}$	$\frac{\tau s+1}{ks(\beta\epsilon s+2\beta+\epsilon)}$	$\frac{\tau}{2\beta+\epsilon}$	τ	-	$\frac{\beta\epsilon}{2\beta+\epsilon}$
F	$k\frac{-\beta s+1}{\tau^2 s^2+2\zeta\tau s+1}$	(2,3,5) $\frac{-\beta s+1}{\epsilon s+1}$	$\frac{\tau^2 s^2+2\zeta\tau s+1}{k(\beta+\epsilon)s}$	$\frac{2\zeta\tau}{\beta+\epsilon}$	$2\zeta\tau$	$\frac{\tau}{2\zeta}$	-
G	$k\frac{-\beta s+1}{\tau^2 s^2+2\zeta\tau s+1}$	(1,4) $\frac{-\beta s+1}{(\beta s+1)(\epsilon s+1)}$	$\frac{\tau^2 s^2+2\zeta\tau s+1}{k(\beta\epsilon s+2\beta+\epsilon)}$	$\frac{2\zeta\tau}{2\beta+\epsilon}$	$2\zeta\tau$	$\frac{\tau}{2\zeta}$	$\frac{\beta\epsilon}{2\beta+\epsilon}$
H	$\frac{k}{s}$	$\frac{1}{\epsilon s+1}$	$\frac{1}{k\epsilon}$	$\frac{1}{\epsilon}$	-	-	-
I	$\frac{k}{s}$	(6) $\frac{2\epsilon s+1}{(\epsilon s+1)^2}$	$\frac{2\epsilon s+1}{k\epsilon^2 s}$	$\frac{2}{\epsilon}$	2ϵ	-	-
J	$\frac{k}{s(\tau s+1)}$	$\frac{1}{\epsilon s+1}$	$\frac{\tau s+1}{k\epsilon}$	$\frac{1}{\epsilon}$	-	τ	-
K	$\frac{k}{s(\tau s+1)}$	(6) $\frac{2\epsilon s+1}{(\epsilon s+1)^2}$	$\frac{(\tau s+1)(2\epsilon s+1)}{k\epsilon^2 s}$	$\frac{2\epsilon+\tau}{\epsilon^2}$	$2\epsilon+\tau$	$\frac{2\epsilon\tau}{2\epsilon+\tau}$	-

Table 1 (Continued)

L	$k\frac{-\beta s+1}{s}$	(2,3,5) $\frac{-\beta s+1}{\varepsilon s+1}$	$\frac{1}{k(\beta+\varepsilon)}$	$\frac{1}{\beta+\varepsilon}$	-	-	-
M	$k\frac{-\beta s+1}{s}$	(1,4) $\frac{-\beta s+1}{(\beta s+1)(\varepsilon s+1)}$	$\frac{1}{k(\beta\varepsilon s+2\beta+\varepsilon)}$	$\frac{1}{2\beta+\varepsilon}$	-	-	$\frac{\beta\varepsilon}{2\beta+\varepsilon}$
N	$k\frac{-\beta s+1}{s}$	(5,6) $\frac{(-\beta s+1)((\beta+2\varepsilon)s+1)}{(\varepsilon s+1)^2}$	$\frac{(\beta+2\varepsilon)s+1}{ks(\beta+\varepsilon)^2}$	$\frac{\beta+2\varepsilon}{(\beta+\varepsilon)^2}$	$\beta+2\varepsilon$	-	-
O	$k\frac{-\beta s+1}{s}$	(4,6) $\frac{(-\beta s+1)(2(\beta+\varepsilon)s+1)}{(\beta s+1)(\varepsilon s+1)^2}$	$\frac{2(\beta+\varepsilon)s+1}{ks(\beta\varepsilon^2 s+\varepsilon^2+4\beta\varepsilon+2\beta^2)}$	$\frac{2(\beta+\varepsilon)}{2\beta^2+4\beta\varepsilon+\varepsilon^2}$	$2(\beta+\varepsilon)$	-	$\frac{\beta\varepsilon^2}{2\beta^2+4\beta\varepsilon+\varepsilon^2}$
P	$k\frac{-\beta s+1}{s(\tau s+1)}$	(2,3,5) $\frac{-\beta s+1}{\varepsilon s+1}$	$\frac{\tau s+1}{k(\beta+\varepsilon)}$	$\frac{1}{\beta+\varepsilon}$	-	τ	-
Q	$k\frac{-\beta s+1}{s(\tau s+1)}$	(1,4) $\frac{-\beta s+1}{(\beta s+1)(\varepsilon s+1)}$	$\frac{\tau s+1}{k(\beta\varepsilon s+2\beta+\varepsilon)}$	$\frac{1}{2\beta+\varepsilon}$	-	τ	$\frac{\beta\varepsilon}{2\beta+\varepsilon}$
R	$k\frac{-\beta s+1}{s(\tau s+1)}$	(5,6) $\frac{(-\beta s+1)((\beta+2\varepsilon)s+1)}{(\varepsilon s+1)^2}$	$\frac{(\tau s+1)((\beta+2\varepsilon)s+1)}{ks(\beta+\varepsilon)^2}$	$\frac{\beta+2\varepsilon+\tau}{(\beta+\varepsilon)^2}$	$\beta+2\varepsilon+\tau$	$\frac{\tau(\beta+2\varepsilon)}{\beta+2\varepsilon+\tau}$	-
S	$k\frac{-\beta s+1}{s(\tau s+1)}$	(4,6) $\frac{(-\beta s+1)(2(\beta+\varepsilon)s+1)}{(\beta s+1)(\varepsilon s+1)^2}$	$\frac{(\tau s+1)(2(\beta+\varepsilon)s+1)}{ks(\beta\varepsilon^2 s+\varepsilon^2+4\beta\varepsilon+2\beta^2)}$	$\frac{2(\beta+\varepsilon)+\tau}{2\beta^2+4\beta\varepsilon+\varepsilon^2}$	$2(\beta+\varepsilon)+\tau$	$\frac{2\tau(\beta+\varepsilon)}{2(\beta+\varepsilon)+\tau}$	$\frac{\beta\varepsilon^2}{2\beta^2+4\beta\varepsilon+\varepsilon^2}$

Table 1 (Continued)

appropriate:

Remark 1: When the PID controller of the specified form is applied to the model $\tilde{g}$, the closed-loop system is stable for all values of $\varepsilon > 0$.

Remark 2: For about one-third of the cases studied ε appears only in the expression for the controller gain k_c. For cases A,B,C,H, and J, the controller gain is inversely proportional to ε, thus demonstrating that on-line PID controller adjustment is effectively achieved by simply manipulating k_c. These are minimum phase models, for which $\omega_b = \omega_c = 1/\varepsilon$ and the model itself imposes no limitations on the bandwidth. For cases D,F,L, and P, the controller gain k_c is the only parameter dependent on ε, but because of the presence of a RHP zero, there is a maximum gain which cannot be surpassed no matter how small ε is. Cases D,F,L, and P correspond to systems factored according to (41), in which case ω_b is limited by approximately $\frac{1}{\beta+\varepsilon}$ (recall eq. (51)); the IMC design procedure recognizes naturally that increasing the gain beyond a certain value leads to performance deterioration and eventually stability.

For a significant number of the considered models ε appears in all the parameters of the classic feedback controller, e.g.,K and R. It is not surprising then that for such processes trial and error tuning of PID controllers is notoriously difficult. However, the IMC parametrization shows how all the controller parameters may be adjusted simultaneously in an effective manner.

Remark 3: In all cases there is no offset for setpoint and/or disturbance step changes. If the process has an integrator, a step disturbance entering through the integrator becomes a ramp, thus requiring that there should be no offset for ramp changes. This performance specification is met in cases I,K,N,O,R, and S by selecting the filter f to be of the form shown in (54).

Remark 4: For systems with RHP zeros, two options for choosing $\tilde{g}_+f$ are available.

Option 1: $\tilde{g}_+f$ follows (40):

$$\tilde{g}_+f = \left(\frac{-\beta s+1}{\beta s+1}\right) \frac{1}{\varepsilon s+1} \tag{64}$$

(64) is optimal in the ISE sense to step changes when $\varepsilon = 0$ (Holt and Morari, 1985). For no offset to ramps f has to be selected in accordance with (54):

$$\tilde{g}_+f = \frac{-\beta s+1}{\beta s+1} \frac{(2(\beta+\varepsilon)s+1)}{(\varepsilon s+1)^2} \tag{65}$$

(64) and (65) require augmenting the PID controller with a lag term $(\tau_F s+1)$. The filter constant ε may, in principle, be chosen freely. However, as already pointed out in Eq. (47), a single RHP zero factored according to (40) limits the bandwidth to $1/\beta\sqrt{7}$. Therefore selecting $\varepsilon << \beta\sqrt{7}$ has very little effect on the response. We recommend that to improve robustness, $\varepsilon > \beta/2$.

Option 2: $\tilde{g}_+f$ follows (41)

$$\tilde{g}_+ = \frac{-\beta s+1}{\varepsilon s+1} \tag{66}$$

For step inputs (66) is IAE optimal when $\varepsilon = 0$ and ISE optimal when $\varepsilon = \beta$ (Holt and Morari, 1985). For no offset to ramps:

$$\tilde{g}_+f = \frac{(-\beta s+1)((\beta+2\varepsilon)s+1)}{(\varepsilon s+1)^2} \tag{67}$$

Option 2 gives a simpler controller and is favorable for situations where $\varepsilon > \beta$ is acceptable. It results in a PID controller without the need for an additional lag (as shown in cases D,F,L,N,P and R). However, noting that the closed-loop transfer function is not strictly proper, one must require that

$$\lim_{\omega\to\infty} \left|\tilde{g}_{+}f\right| < 1 \tag{68}$$

or (59) will be violated for high frequencies (where $\ell_m \geq 1$) and instability is bound to occur in all practical situations. This explains why $\varepsilon > \beta$ is required for D,F,L,N,P, and R. The effect of this practical recommendation is that the RHP zero is pushed outside the bandwidth of the closed-loop system.

In practice, there exists no ideal PID controller as required in Option 2. An additional lag is always present in the controller to provide roll-off at high frequencies. Option 1 suggests a "practical" PID controller with an "optimal" roll-off element $(\tau_F s+1)^{-1}$.

Remark 5: No systems with LHP zeros are listed in Table 1. As seen from (62), LHP zeros translate into lags in the feedback controller structure when the IMC design procedure is used. Therefore, for models with LHP zeros, the PID-controller from Table 1 should be augmented with the corresponding lags.

Remark 6: Controller complexity, as stated in the introduction, depends on the model and the control system objectives. Consider the cases H (a pure integrator), A (a first-order model) and B (a second-order noninteracting model), for which the desired closed-loop response is that of a first-order lag. Only a proportional controller is necessary for H, a PI-controller must be used for A, while a PID controller is needed for B. Likewise, consider cases P-S, where the process model is the same: as the demands on the control system increase (as in requiring no offset to ramp changes), so does the complexity of the controller.

Remark 7: Table 1 can also be used for systems with delays by approximating the deadtime with a Padé element; the entry for the rational approximate model then provides the controller parameters. This procedure is illustrated with two examples.

Example 1

$$g(s) = \frac{1-k_1 e^{-\theta s}}{s} \tag{69}$$

Using a first-order Padé approximation

$$\tilde{g}(s) = \frac{1-k_1+\frac{\theta}{2}(1+k_1)s}{s(\frac{\theta}{2}s+1)} \tag{70}$$

If $k_1 > 1$ then (70) has a RHP zero and a controller from entries P-S can be selected. If $k_1 < 1$ the resulting LHP zero should be removed by a simple lag, as explained in Remark 5. PID parameters can then be obtained from entries J or K.

Example 2:

$$g(s) = \frac{ke^{-\theta s}}{\tau s+1} \tag{71}$$

A "zeroth-order" Padé approximation ($e^{-\theta s} \cong 1$) yields

$$\tilde{g}(s) = \frac{k}{\tau s+1} \tag{72}$$

Entry A in Table 1 provides a PI controller for this structure. The "zeroth-order" Padé approximation is equivalent to designing a controller with no information on the deadtime.

A first-order Padé approximation yields

$$\tilde{g}(s) = \frac{k(-\frac{\theta}{2}s+1)}{(\tau s+1)(\frac{\theta}{2}s+1)} \tag{73}$$

Entries F and G (PID controller, PID controller with first-order lag) are applicable to this problem. This problem is discussed in more detail in Section 5.

4.2. Effects of the Padé Approximation

Examples 1 and 2 are interesting because they indicate circumstances under which three term lead-lag controllers can be used to control processes with deadtime. The use of the Padé approximation, however, introduces modeling error, which consequently limits the achievable bandwidth ω_b and the minimum value for ε. From (59), one obtains a good guess on the smallest value of ε which still maintains a stable control system.

$|e_m|$ for the zeroth and first-order Padé approximations are shown in Fig. 2. For the first-order Padé approximation, $|e_m| = 1$ at $\omega \cong 3/\theta$ and thus a sufficient condition for stability is to choose $\varepsilon > \theta/3$. For the zeroth-order approximation $|e_m| = 1$ at $\omega \cong 1/\theta$ and therefore $\varepsilon > \theta$ is required. Because ω_b is inherently limited by $0.724/\theta$ (recall (46)), one can expect that using the first-order Padé approximation will yield designs very close to optimal (i.e., if no approximation were present). The zeroth-order approximation will be adequate, however, when small bandwidths and low-frequency inputs are involved.

5. IMC Based PID Control For a First-Order Lag With Deadtime

The important role of the first-order lag/deadtime model (71) in process control mandates a more detailed discussion of Ex. 2. Our attention is directed to a further understanding of the PI and PID rules generated by cases A and F; the advantages of the augmented PID controller (case G) are also indicated.

5.1. Tuning Procedures

The IMC-based controllers obtained using first and zeroth-order Padé approximations for the time delay are (cases A and F in Table 1)

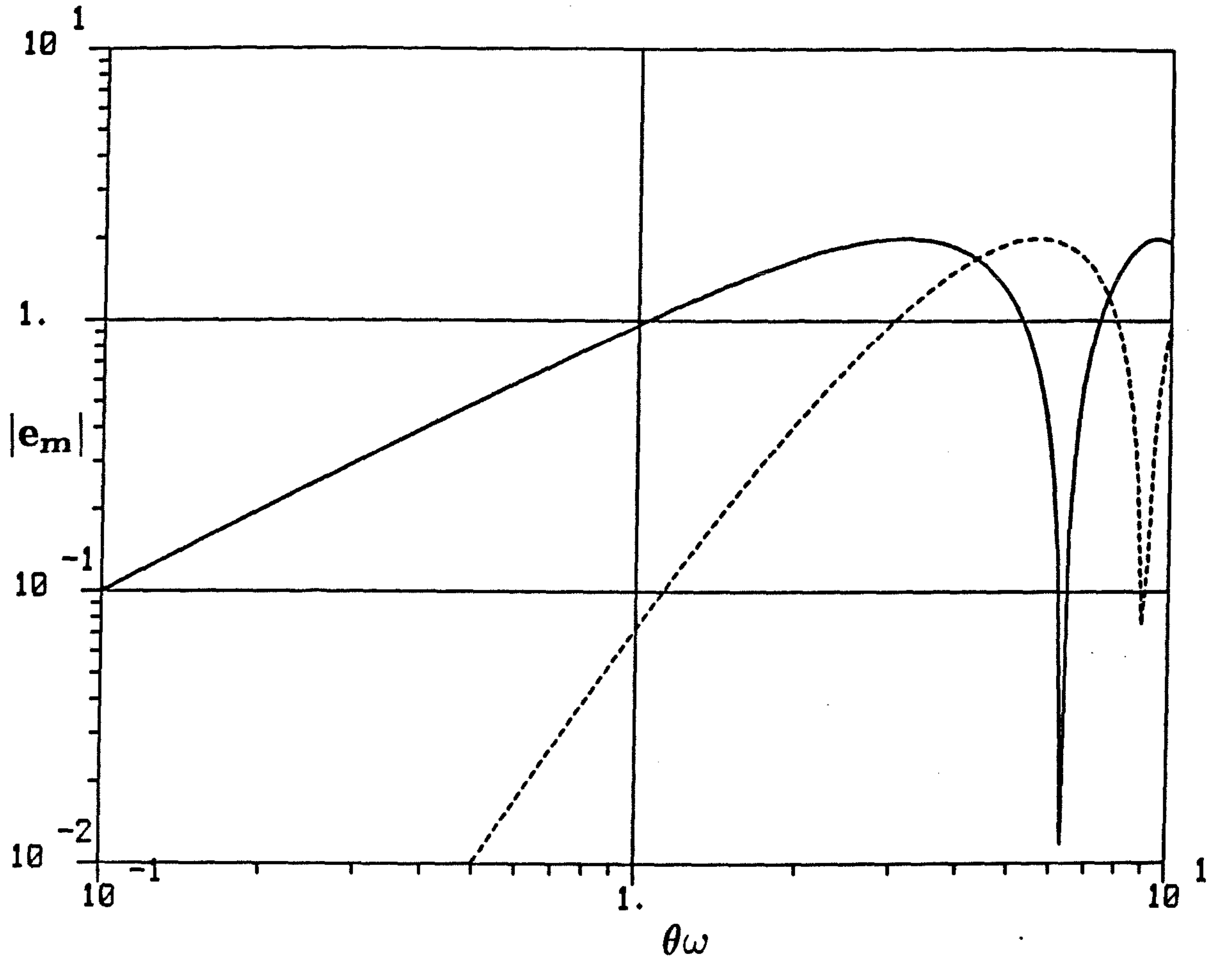


Figure 2

Multiplicative uncertainty $|e_m|$ for the zeroth (——) and first-order (····) Padé approximation.

Zeroth-order: $|e_m| = |1-e^{-\theta s}|$

First-order: $|e_m| = \left| \dfrac{1-\frac{\theta}{2}s}{1+\frac{\theta}{2}s} - e^{-\theta s} \right|$

$|e_m|$ for the zeroth-order approximation is equal to $|S|$ for $H = e^{-\theta s}$.

$$\underline{\text{PID}}: \quad c = \frac{(1+\tau s)(1+\frac{\theta}{2}\ s)}{k(\frac{\theta}{2} + \varepsilon)s} \tag{74}$$

$$\underline{\text{PI}}: \quad c = \frac{(1+\tau s)}{k\varepsilon s} \tag{75}$$

Option 2 (Eq. 66) was chosen for the filter for the first-order Padé approximation in order to get a PID controller without an additional lag term. These controllers are represented compactly in Table 2. The closed-loop transfer functions for system (71) with these controllers indicate a number of advantages:

$$\text{PID:} \quad y = \frac{\dfrac{e^{-\theta s}}{(\frac{\varepsilon}{\theta} + \frac{1}{2})\theta s}}{1+ \frac{1}{2}\ \theta s} + e^{-\theta s}\ (y_s-d) + d \tag{76}$$

$$\text{PI:} \quad y = \frac{e^{-\theta s}}{(\frac{\varepsilon}{\theta})\theta s + e^{-\theta s}}\ (y_s-d) + d \tag{77}$$

- The closed-loop response is independent of the system time constant τ (the process lag $(1+\tau s)$ is cancelled by the controller).
- Time is scaled by θ.
- The shape of the response depends on ε/θ only.

In other words, specifying one value of ε/θ for any first-order lag with deadtime model results in an identical response when time is scaled by θ, regardless of k, θ, and τ. For instance, if the deadtime in system I is twice as long as the deadtime in system II, then for a specific ε/θ, the response characteristics will be identical except that it will take the response of system I exactly twice as long to reach the same point as system II. The choice of the "best" ratio ε/θ must be based on performance and robustness considerations.

Controller	kk_c	τ_I	τ_D	Recommended ε/θ ($> \frac{0.1\tau}{\theta}$ always)
PID	$\frac{2\tau+\theta}{2\varepsilon+\theta}$	$\tau + \frac{\theta}{2}$	$\frac{\tau\theta}{2\tau+\theta}$	>0.8
PI	$\frac{\tau}{\varepsilon}$	τ	-	>1.7
Improved PI	$\frac{2\tau+\theta}{2\varepsilon}$	$\tau + \frac{\theta}{2}$	-	>1.7

Table 2: IMC-based PID parameters for

$$g(s) = \frac{ke^{-\theta s}}{\tau s+1}$$

and practical recommendations for ε/θ.

For the PID controller Figure 3 demonstrates the dependence of the step response on ε/θ. $\varepsilon/\theta = 0.4$ is fairly close to the value where instability occurs (ε/θ=0.145), and the large overshoot and poorly damped oscillations are therefore not surprising. Note that $\varepsilon/\theta = 0.5$ is the lower value recommended in Table 1 for models with a RHP zero factored according to (66). For $\varepsilon/\theta = 0.8$ the response looks very good: the rise time is about 1.5θ and the settling time is 4.5θ; the overshoot is about 10% and the decay ratio is quite good. For ε/θ = 2.5 the response becomes highly overdamped and almost identical to that of a first-order system with time constant ε/θ and delay θ.

The scaled form of the closed-loop transfer functions (76) and (77), allows convenient design plots to be made (Figs. 4 and 5). The performance measure J, the integral square error to a step disturbance/setpoint change, and the robustness measure M have been plotted as a function of ε/θ. In Figs. 4 and 5, J is normalized by J_{opt}, the error corresponding to the optimum response $y/y_s = e^{-\theta s}$. In theory, a Smith Predictor with infinite gain (k_c=∞) accomplishes this response.

For PID control (Fig. 4), J/J_{opt} reaches a minimum of 1.092 for $\varepsilon/\theta = 0.68$. At this point, M = 1.3. For practical purposes a better compromise between performance and robustness is attained for ε/θ = 0.8; here the ISE is almost minimum but M has dropped to 1. For PI control (Fig. 5), $\varepsilon/\theta = 1.4$ results in the minimum J/J_{opt} value 1.55. M for this setting is approximately 1.3. M = 1 first occurs at $\varepsilon/\theta \cong 2$, where $J/J_{opt} \cong 1.7$.

Figure 4 also confirms that the first-order Padé approximation leads to relatively little performance deterioration. For $\varepsilon/\theta = 0.8$

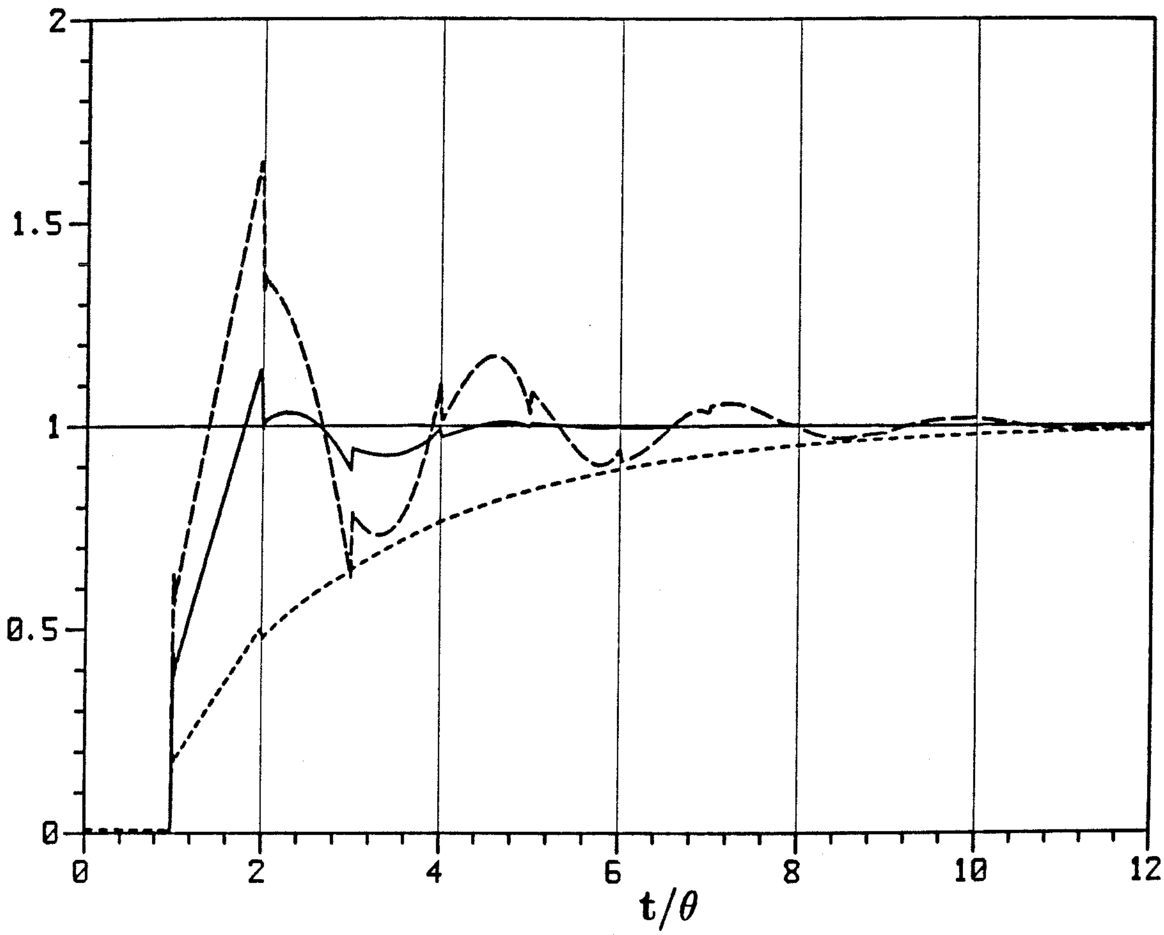


Figure 3

IMC-PID tuning rule. Effect of ε/θ on the closed-loop response to a unit step setpoint change. $g(s) = ke^{-\theta s}/(\tau s+1)$. —— : $\varepsilon/\theta = 0.8$; ---- : $\varepsilon/\theta = 0.4$; ···· : $\varepsilon/\theta = 2.5$.

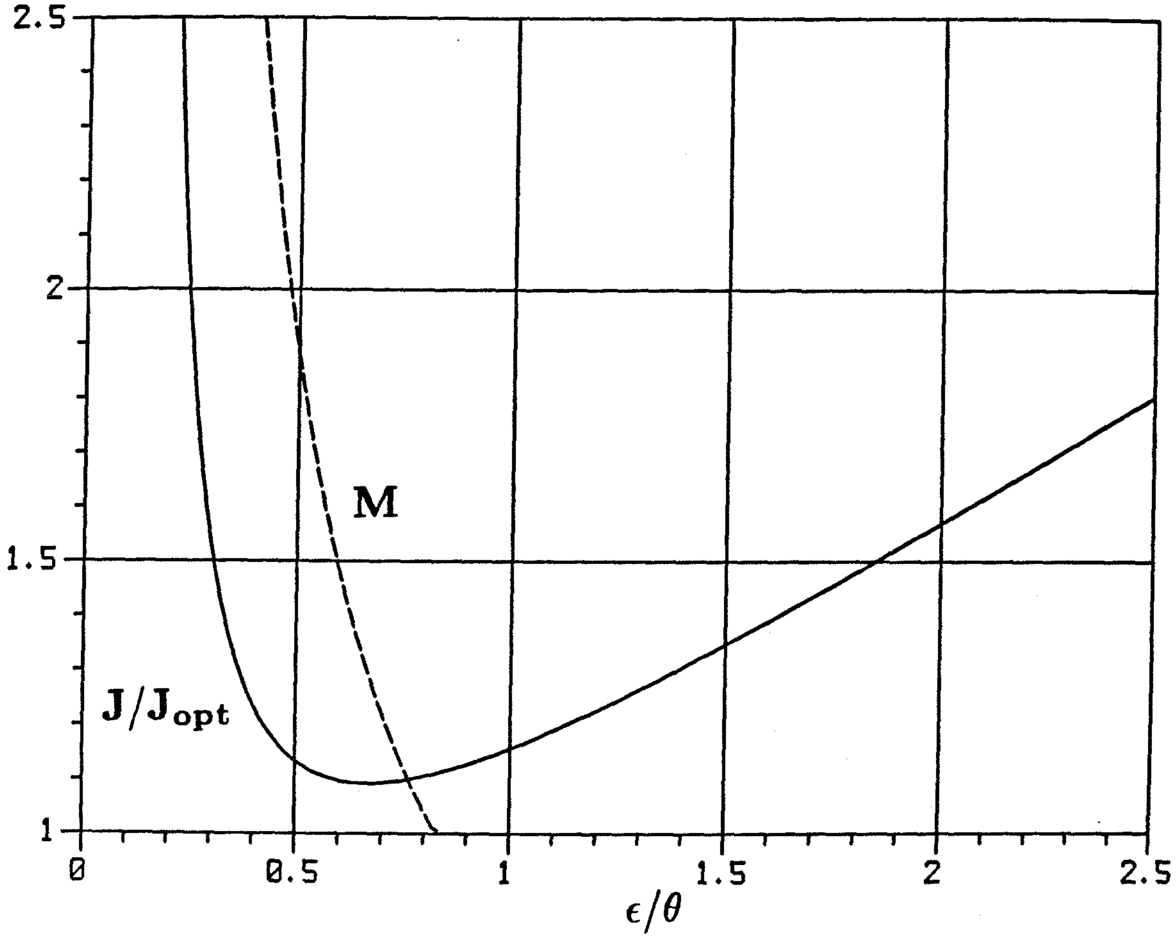


Figure 4

IMC-PID tuning rule (74). Effect of ε/θ on M and ISE (J) for step changes. $g(s) = ke^{-\theta s}/(\tau s+1)$.

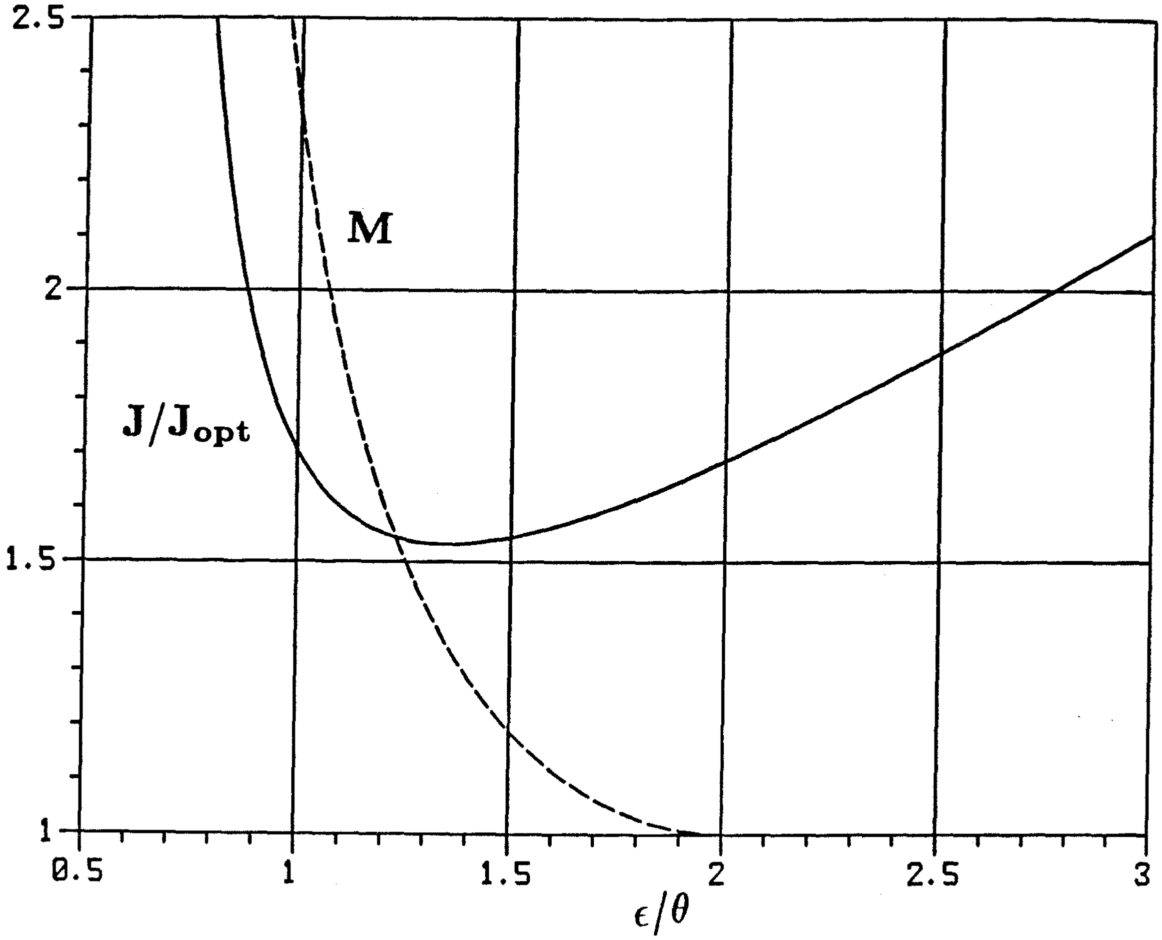


Figure 5

IMC-PI tuning rule (75). Effect of ε/θ on M and ISE (J) for step changes. $g(s) = ke^{-\theta s}/(\tau s+1)$.

the result is a PID controller that performs with only 10% greater ISE than the optimal Smith Predictor, while retaining favorable robustness characteristics. Compared to the PI controller, however, the Smith Predictor provides significant performance improvement; one must realize that the PI rule originates from a reduced model with no dependence on the process deadtime; an alternate rule is described in Section 5.3 which takes into account this deficiency.

Figures 4 and 5 have been obtained under the assumption of no plant uncertainty; only the model error induced by the Padé approximation is considered. Significant plant uncertainty within the bandwidth of the controller will require the designer to select a larger value of ε. This consideration is of particular concern when $\varepsilon/\theta << 1$. Because for the process industries the closed-loop bandwidth can rarely exceed ten times the open-loop bandwidth ($10/\tau$), a practical requirement is to always select $\varepsilon > \tau/10$. For the IMC-PID parameters this inequality is dominant for $\theta/\tau < 1/7$; for the PI parameters it will become important for $\theta/\tau < 1/14$.

5.2. Comparisons with Other Methods

Next we compare the IMC-PID parameters with the classic Ziegler-Nichols and Cohen-Coon tuning rules (Figs. 6a & b). The first notable difference between these rules and those from IMC is that J and M depend strongly on θ/τ, while for the IMC rules the performance and robustness measures are independent of this ratio. The Cohen-Coon rules give reasonable performance ($J/J_{opt} < 1.3$) for $0.6 < \theta/\tau < 4.5$. In this range M varies between 2.7 and 1.0, i.e., robustness is quite poor, especially for small ratios of θ/τ. The performance obtained with the closed-loop Ziegler-Nichols parameters is good for the range

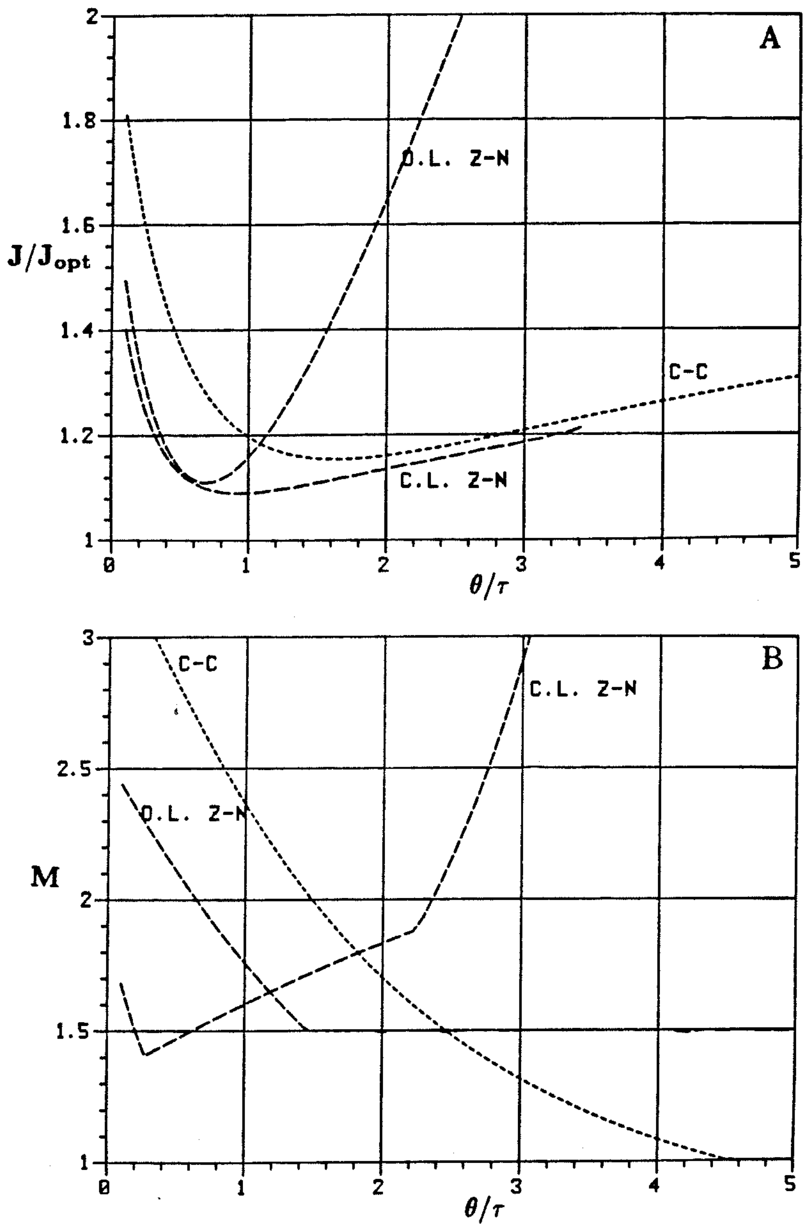


Figure 6

PID controllers for $g(s) = e^{-\theta s}/(\tau s+1)$. Performance (Fig. 6A) and robustness (Fig. 6B) properties of the Cohen-Coon, open-loop Ziegler-Nichols, and closed-loop Ziegler-Nichols tuning rules.

$0.2 < \theta/\tau < 3.5$ but again the robustness is poor except for $\theta/\tau \cong 0.3$. Indeed, for $\theta/\tau > 4$ the closed-loop system is unstable with the c-l Ziegler-Nichols parameters. In terms of performance the open-loop Ziegler-Nichols parameters are only useful in the range $0.2 < \theta/\tau < 1.4$. The advantages of the IMC tuning rules are further demonstrated through simulations (Fig . 7).

It should be emphasized, however, that by themselves the higher M values for the Ziegler-Nichols and Cohen-Coon settings do not imply that these control systems can tolerate less plant/model mismatch than IMC before becoming unstable. As was explained in Section II, model error tolerance depends on both M and the closed-loop bandwidth ω_b. Thus only for a particular bandwidth/performance specification is IMC more robust than Ziegler-Nichols and Cohen-Coon. Comparing Ziegler-Nichols and Cohen-Coon with a small bandwidth/poor performance and IMC with a larger bandwidth/better performance can demonstrate a larger robustness of the former despite larger M values.

In Figs. 7a-c, the presence of pure derivative action (which is physically unrealizable) leads to a somewhat jerky response and to even more violent moves in the manipulated variable. If the proper IMC controller implied by entry G is used, the consequences are an increase in the ISE and a slower speed of response (Fig. 8); the response, however, is smooth and looks more attractive (Fig. 9). Here $\varepsilon/\theta = 0.45$ was chosen to obtain good robustness characteristics (M=1).

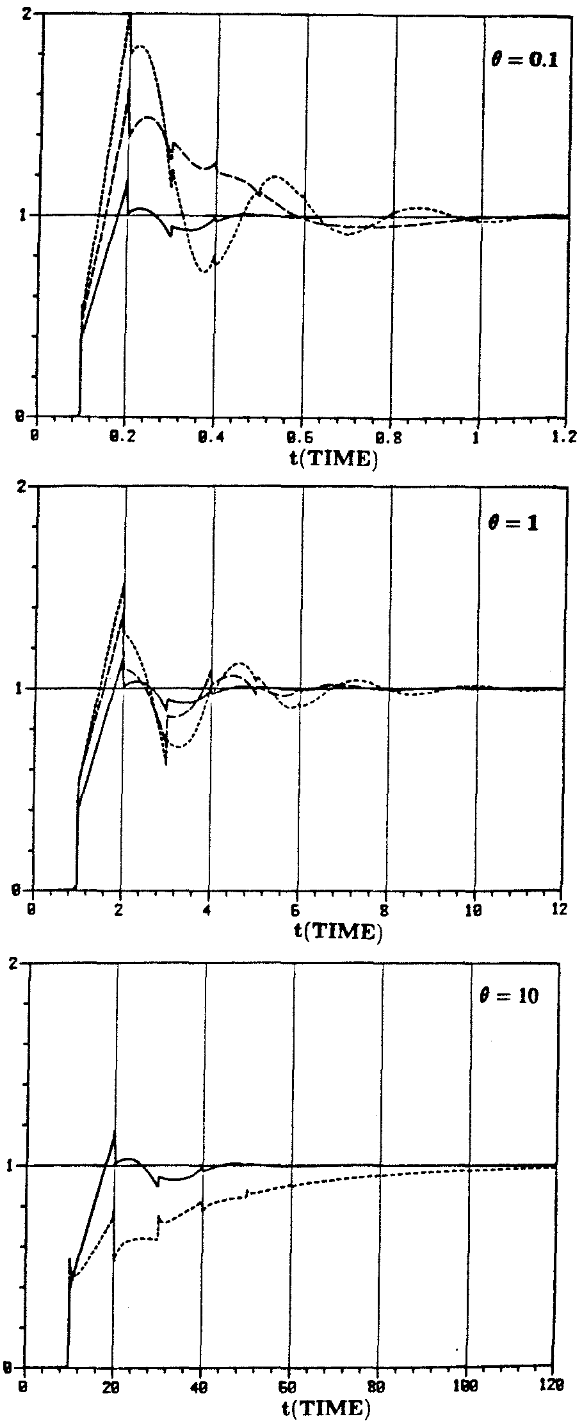


Figure 7

PID tuning rules for process $e^{-\theta s}/(\tau s+1)$. Closed-loop responses to a unit step setpoint change for $\theta = 0.1$, 1, and 10. ——— : IMC ($\varepsilon/\theta = 0.8$); ---- : closed-loop Ziegler-Nichols (unstable for $\theta = 10$); ···· : Cohen-Coon.

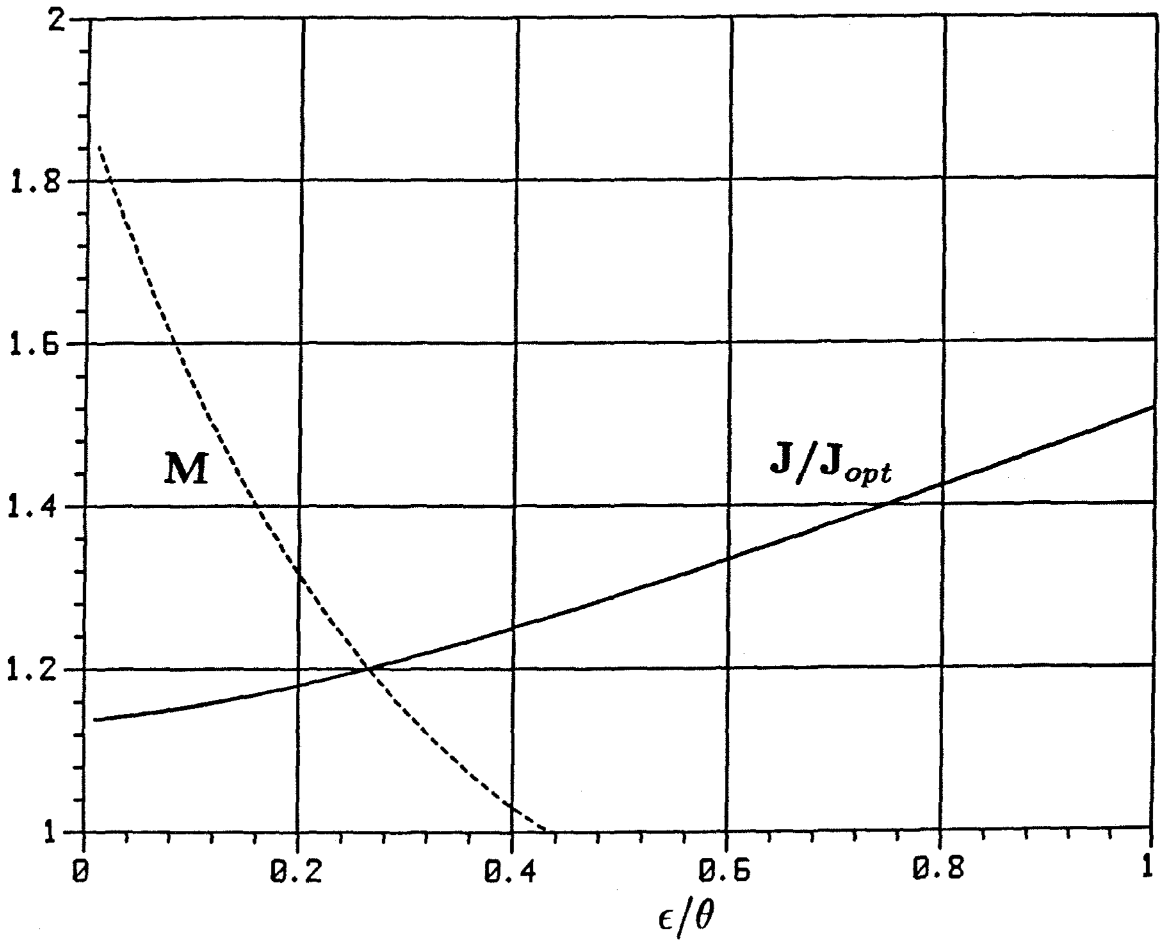


Figure 8

IMC-PID controller with exponential filter (Table 1, entry G). Effect of ε/θ on M and the ISE (J) for step changes. $g(s) = ke^{-\theta s}/(\tau s+1)$

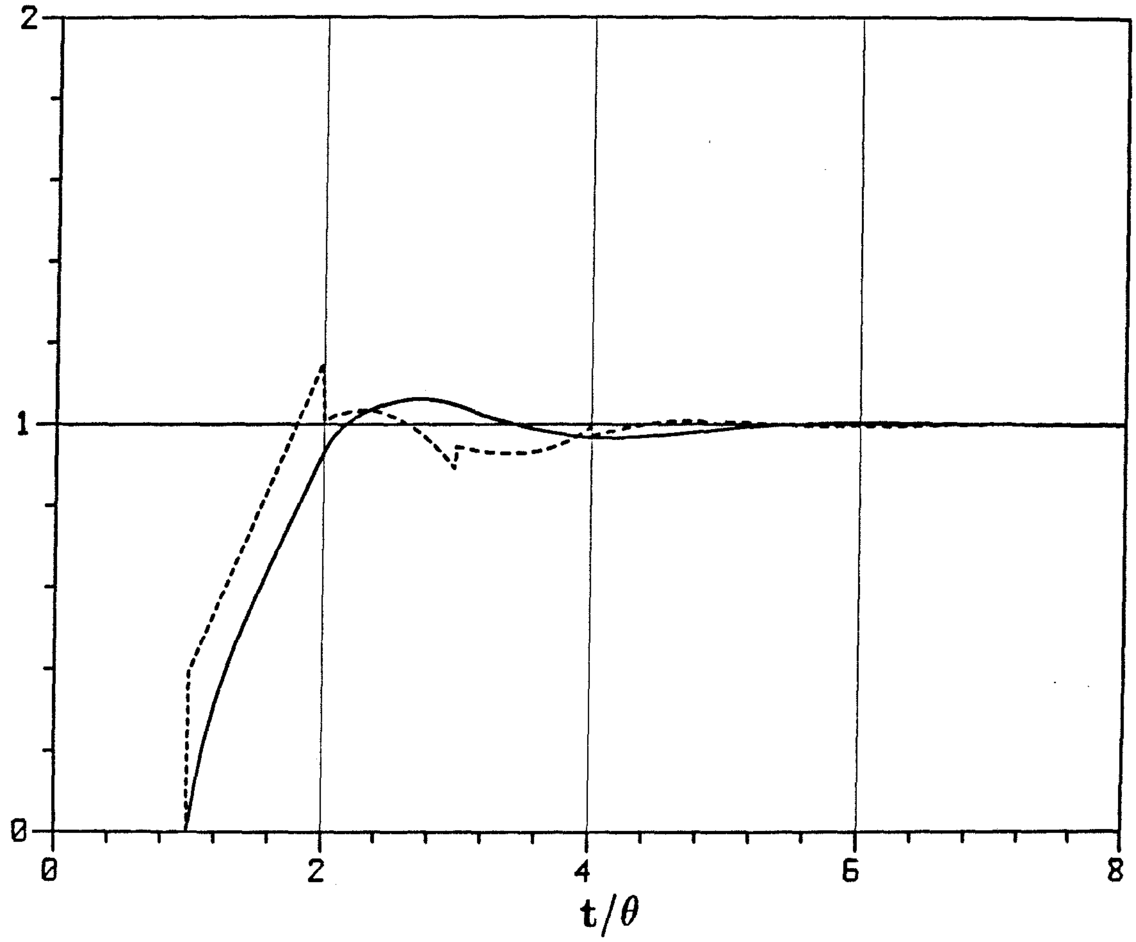


Figure 9

IMC-PID controllers. Closed-loop responses to a unit step setpoint change. $g(s) = ke^{-\theta s}/(\tau s+1)$.

—— : IMC-PID with exponential filter (Table 1, entry G), $\varepsilon/\theta = 0.45$.

···· : IMC-PID (Table 1, entry F), $\varepsilon/\theta = 0.8$.

5. 3. Development of an Improved PI Rule

The IMC PI rule (from entry A, Table 1), despite its compactness and simplicity (as evidenced by Table 2 and Fig. 5) has at best 55% greater performance cost than the optimal Smith Predictor and is not overall superior to the Z-N and C-C expressions. This is a consequence of the zeroth-order Padé approximation, and can be remedied by incorporating the deadtime in the internal model through other means.

The IMC design procedure prescribes that first a process model $\tilde{g}$ should be established which closely approximates the real process; the controller structure and parameters follow directly from $\tilde{g}$. As a second step, the filter parameters are adjusted to compensate for the plant/model mismatch. In the context of the present example it is clear that a zeroth-order Padé approximation is inadequate. If, in order to obtain a PI controller, a first-order lag is used to approximate a first-order lag with deadtime, it appears reasonable to increase the model's lag over that of the process in order to account for the presence of the delay. Thus we postulate the model

$$\tilde{g} = \frac{k}{(\tau\lambda)s+1} \tag{78}$$

where λ depends on the process time delay. λ must be chosen such that "best" approximates the first-order lag/delay process. Rivera (1984) has established

$$\lambda = 1 + 0.5(\theta/\tau) \tag{79}$$

as suitable.

The PI rules obtained using (79) appear in Table 2. Comparing the "improved" PI controller

$$c = \frac{1+(\tau+\theta/2)s}{k\varepsilon s} \tag{80}$$

with the PID controller (74) and the PI-controller (75) based on the zeroth-order Padé approximation, the following becomes clear

- For small time delays ($\tau/\theta >> 1$) the improved PI rule and the original one are the same.
- For very large time delays ($\tau/\theta << 1$) the term $(1+\tau s)$ in (74) will be outside the closed-loop bandwidth, and the improved PI controller and PID controller are equivalent when the following relationship is used:

$$(\frac{\varepsilon}{\theta})_{PI} = (\frac{\varepsilon}{\theta})_{PID} + 0.5 \tag{81}$$

Thus for very large time delays the PID controller approaches the PI controller with some gain correction according to (70), i.e., derivative action becomes ineffective.

So far, no rules have been given on how to select ε for the improved IMC-PI controller. To provide an idea of the performance and robustness properties Fig. 10 was constructed. This plot indicates the maximum values of J/J_{opt} and M (over the entire θ/τ range) as a function of ε/θ. From Figs. 4, 5, and 10 one finds that $\varepsilon/\theta = 1.7$ provides a reasonable compromise between performance and robustness ($(J/J_{opt}) = 1.58$, $M = 1.15$). Note that this is slightly higher than $\varepsilon/\theta = 1.3$ which is suggested from (81) based on the PID rule ($(\frac{\varepsilon}{\theta})_{PID}=0.8$). Fig. 11 demonstrates that with this choice one obtains performance and robustness properties equal or superior to those of the Ziegler-Nichols and Cohen-Coon PI rules. Simulation results (Figs. 12a-c) confirm this.

Not only do the improved IMC-PI parameters lead to better performance and robustness than the traditional methods, the IMC design procedure also makes the search for the appropriate parameters

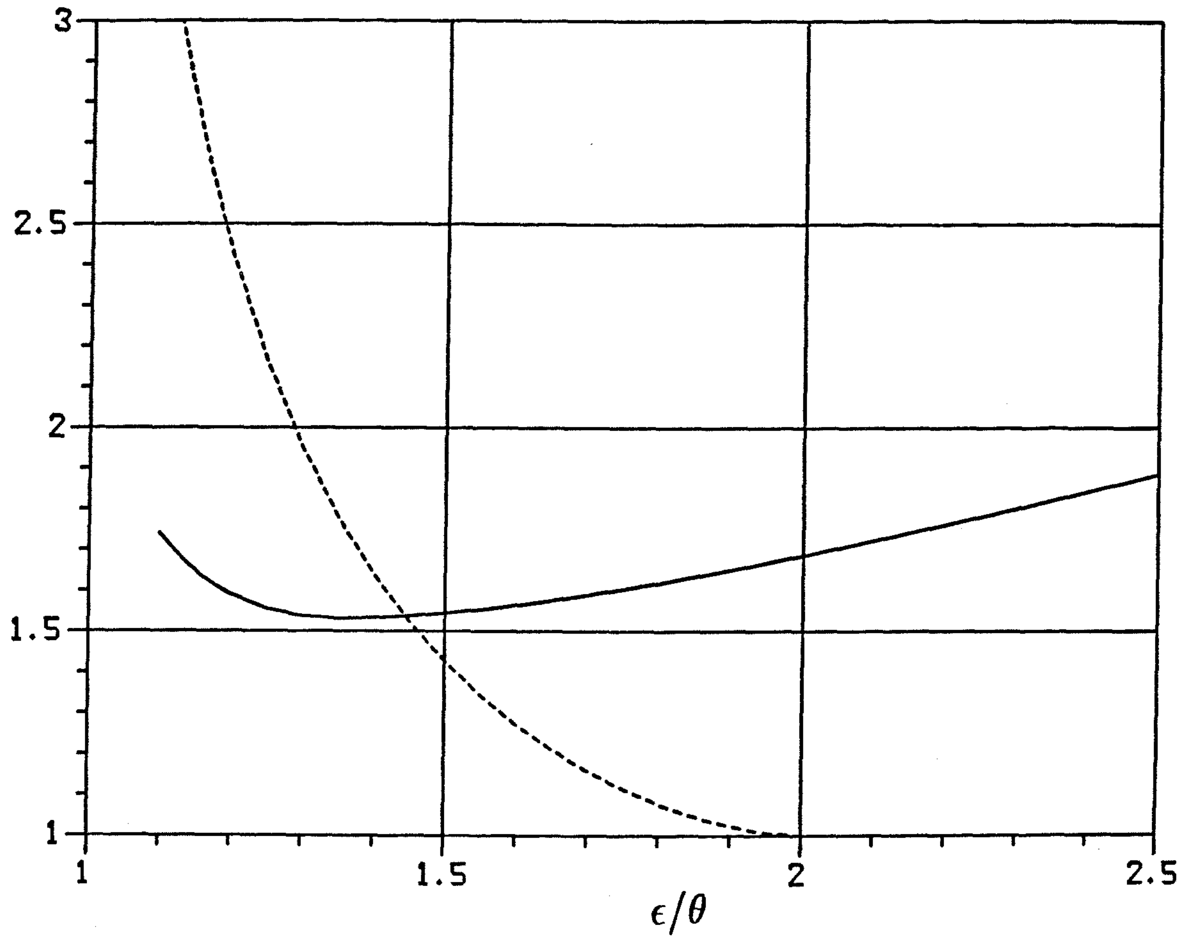


Figure 10

Improved IMC-PI tuning rule. Lower bound on performance and robustness for all θ/τ. $g(s) = ke^{-\theta s}/(\tau s+1)$. ——: $(J/J_{opt})_{max}$; ···· : M_{max}.

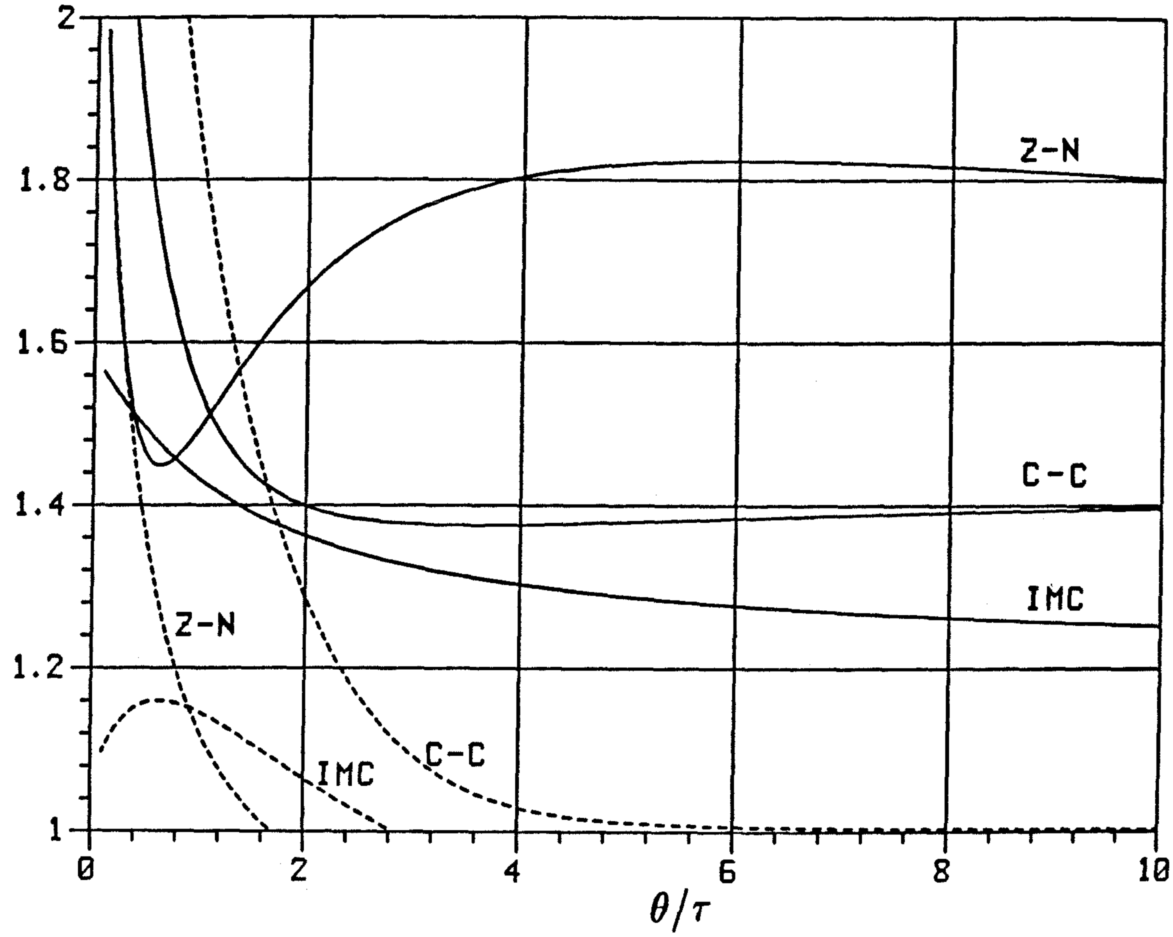


Figure 11

PI tuning rules for $g(s) = ke^{-\theta s}/(\tau s+1)$. Performance and robustness properties for the improved IMC-PI tuning rule (ε/θ = 1.7), the closed-loop Ziegler-Nichols rule, and the Cohen-Coon rule. ——— : J/J_{opt}; ···· : M.

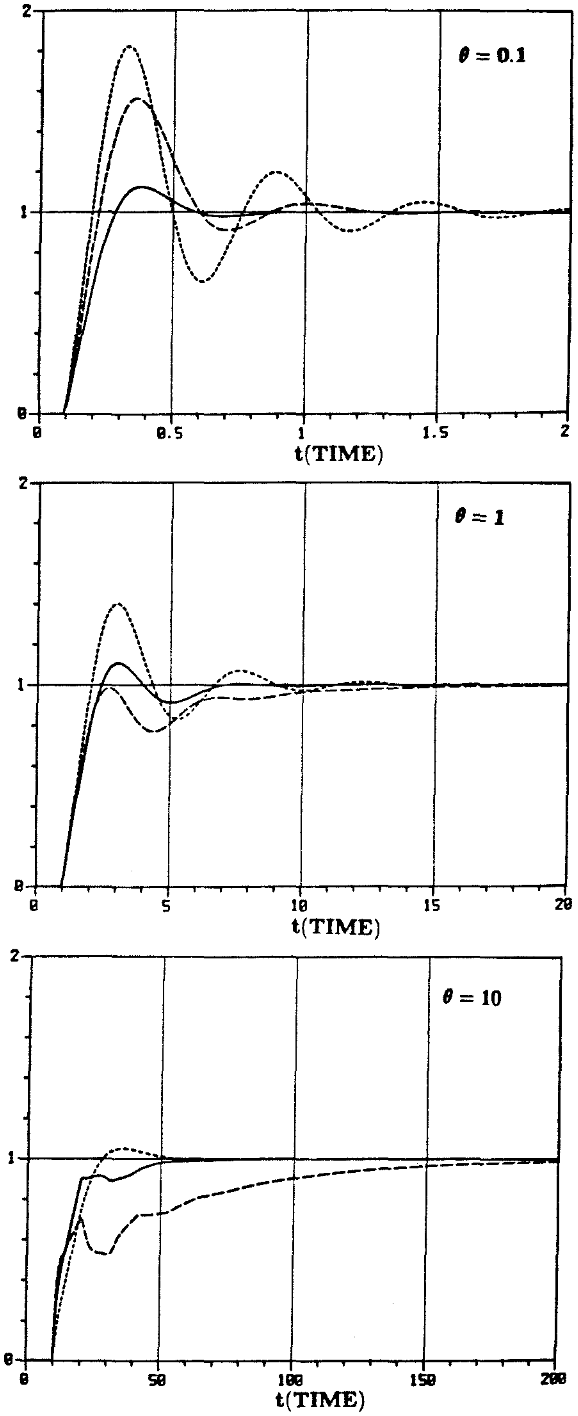


Figure 12

PI tuning rules for $g(s) = e^{-\theta s}/(\tau s+1)$. Closed-loop responses to a unit step setpoint change for $\theta = 0.1$, 1, and 10. ——— : IMC-improved ($\varepsilon/\theta = 1.7$); ---- : closed-loop Ziegler-Nichols; ···· : Cohen-Coon.

simpler. In the IMC context, the PI controller is reparametrized with the parameters λ and ε. While in general it is necessary to search over k_c and τ_I simultaneously, IMC allows to search first for λ to obtain a good model fit and then for ε to obtain good performance and robustness.

5.4. Robustness to Deadtime Errors

The following criterion which is sufficient for stability in the face of deadtime error was derived in Section 2

$$\delta < \frac{1}{\omega_c M} \tag{82}$$

For the IMC-PID controller and $\varepsilon/\theta \geq 0.8$ ($M = 1$), the following crossover frequency approximation holds (see Appendix B)

$$\frac{1}{\omega_c} \cong \frac{\theta}{2} + \varepsilon \tag{83}$$

Likewise, for the improved PI controller ($\varepsilon/\theta \geq = 1.7$, M varies with τ, as shown in Fig. 11), one can approximate ω_c as

$$\frac{1}{\omega_c} \cong \varepsilon \tag{84}$$

(83) and (84) lead to convenient expressions for the allowable deadtime error in terms of the design parameters ε, θ, and M

$$\text{PID:} \qquad \delta < \varepsilon + \frac{\theta}{2} \tag{85}$$

$$\text{Improved PI:} \qquad \delta < \frac{\varepsilon}{M} \tag{86}$$

Table 3 confirms that (85) and (86) provide extremely accurate predictions for the allowable deadtime error. Furthermore, it shows that the suggested PI and PID settings provide deatime error robustness in excess of 100%.

5.5 Tuning Based on Crossover Information

Controller	Deadtime Range	Allowable deadtime error δ/θ	
		Approximate Bound	Exact Bound
PID $\frac{\varepsilon}{\theta} = 0.8$	$0 < \frac{\theta}{\tau} < \infty$	1.30	1.36
Improved PI $\frac{\varepsilon}{\theta} = 1.7$	$\frac{\theta}{\tau} = 0.1$	1.54	1.55
	$\frac{\theta}{\tau} = 0.6$	1.46	1.46
	$\frac{\theta}{\tau} = 1.0$	1.50	1.52
	$\frac{\theta}{\tau} = 10.0$	1.70	1.97

Table 3: Allowable deadtime error δ/θ for IMC-based PID and PI parameters. The exact bound is obtained from (17) while the approximate bounds are obtained from (85) and (86) for the PID and improved-PI rules, respectively.

The classic Ziegler-Nichols identification procedure (1942) and its modern counterpart by Aström and coworkers (1983) do not provide a parametric model directly but only the ultimate gain K_u and the ultimate period P_u. P_u is related to the phase crossover frequency ω_ϕ by

$$P_u = \frac{2\pi}{\omega_\phi} \tag{87}$$

K_u is the inverse of the process gain at ω_ϕ

$$K_u = |g(\omega_\phi)|^{-1} \tag{88}$$

Generally a good estimate of the process time delay θ is also available from step tests. As Shinskey (1979) argues, it is reasonable to model processes with

$$2\theta < P_u < 4\theta \tag{89}$$

as a first-order lag with deadtime as denoted by (71). If θ is known, k and τ can be estimated from K_u and P_u through the formulas

$$\tau = \frac{P_u \tan\left(\frac{\pi(1-\ (2\theta))}{P_u}\right)}{2\pi} \tag{90}$$

$$k = \frac{1}{K_u}\sqrt{1+\left[\frac{2\pi\tau}{P_u}\right]^2} \tag{91}$$

With the aid of (90) and (91) the PI and PID tuning rules from Table 2 can be applied when K_u, P_u and θ are available and a model of the form (71) is assumed.

6. Summary

IMC derived PID controller parameters are listed in Table 1 for most models commonly used in process control. In some cases, the IMC design procedure leads naturally to the need for a first-order lag to augment the PID controller structure. The single adjustable parameter ε is directly related to the speed of response, with $1/\varepsilon$ approximately equal to the bandwidth of the closed-loop system. When the PID controller with the specified structure is applied to the model the closed-loop system is stable for all $\varepsilon > 0$. In practice $1/\varepsilon$ should be chosen to be smaller than the bandwidth over which the process model is valid. A good generally conservative initial guess is to set ε equal to the dominant time constant of the open-loop system.

If the model includes LHP zeros, these zeros should be cancelled first by an appropriate lag and then the entries from Table 1 can be used. If the system includes deadtime, Padé approximations may be used to simplify the model; the result is a simple, lead-lag type controller. The zeroth-order approximation requires $\varepsilon > \theta$, for the first-order approximation $\varepsilon > \theta/3$.

When a 1st-order Padé approximation is used for a first-order lag with deadtime, the IMC design technique yields the PID parameters listed in Table 2. The ISE to a step change is minimal for $\varepsilon/\theta = 0.68$. A better tradeoff between performance and robustness is reached for $\varepsilon/\theta = 0.8$. For small deadtimes ($\theta/\tau < 1$), robustness considerations (unmodeled dynamics) will dictate a larger ε.

By approximating a first-order lag with deadtime model by a first-order lag without deadtime, the IMC procedure leads to the "improved" PI parameters in Table 2. For a choice of $\varepsilon/\theta = 1.7$ the lower bound on performance, over the entire θ/τ range, is $J/J_{opt} = 1.58$ and $M = 1.15$.

Furthermore, simple expressions for the allowable deadtime error δ are available in terms of the parameter ε. For the IMC-PID rule this expression is

$$\delta < (\varepsilon + \frac{\theta}{2}) \tag{85}$$

and for the improved PI rule

$$\delta < \frac{\varepsilon}{M} \tag{86}$$

7. Conclusions

We have shown that for most of the models used to describe the dynamics of chemical process systems the PID controller is the natural choice. In the absence of nonlinearities, constraints, or multivariate interactions it is infeasible to improve the performance with more complex controllers unless higher order, more accurate process models are available.

Furthermore, by substituting Padé approximations, these PID rules have been extended to models with deadtime. For the particular case of a first order lag with deadtime process. the improvement of the ISE for a step setpoint/disturbance by the Smith Predictor over a PID controller is at most 10% <u>regardless</u> of θ/τ. For small values of θ/τ this 10% improvement is generally not attainable because of model uncertainties. For large values of θ/τ some improvement is possible if the process model is valid over a large enough bandwidth.

Although we show that PID-type controllers are adequate for most common process models, we find that the classical feedback structure is inadequate for a clear understanding of control system design. IMC formed the basis of all the rules in Tables 1 and 2. If one were to use IMC directly and not insist on the traditional PID parameters, no rules and no involved tables would be needed. The IMC design procedure is generally applicable regardless of the system involved. No special provisions are required to deal with every single type of system. The complexity of the rules in Tables 1 and 2 demonstrates that the PID parameters k_C, τ_I, and τ_D are the consequences of a long hardware tradition rather than because they represent the most practical tuning tools. The unfortunate parametrization of the PID controller might also explain why some modern control methods (possessing structures that fall under that of IMC) have claimed improvements in control quality over PID for simple systems where a properly tuned PID controller would have yielded an equally good result. The results presented here also clearly point out the limitations of PID controllers. The practical occurrence of systems where no nonlinearities, constraints, or multivariate interactions are present are very rare. In all other situations the PID controller must be "patched up" with anti-reset windup, deadtime compensators, and decouplers, while the IMC technique allows a unified treatment of all cases.

Finally, we must acknowledge (Lau and Balhoff, 1984) that the discrete form of IMC (Garcia and Morari, 1982), because of the increased number of tuning parameters and the added flexibility allowed by the discrete representation in formulating control objectives, can lead to performance and robustness improvements not possible with the PID parameters suggested here.

Acknowledgment

The idea for a paper of this type arose from discussion with J.P. Shunta and J. Richards from DuPont Corp., Wilmington, DE. We are thankful for criticism from C. Economou and K. Levien. Financial support from the National Science Foundation (CPE-8115022) and the Department of Energy (DOE contract DE-AC02-80ER10645) is gratefully acknowledged.

References

Aström, K.J. and T. Hägglund. "Automatic Tuning of Simple Regulators for Phase and Amplitude Margin Specifications." IFAC Workshop on Adaptive Control, San Francisco (1983).

Brosilow, C. Case Western Reserve University, Cleveland, OH, personal communication (1983).

Cohen, G. H. and G. A. Coon. Trans. ASME, 75, 827 (1953).

Doyle, J. and G. Stein. IEEE Trans. Automatic Control, AC-26, 4 (1981).

Frank, P. M. Entwurf von Regelkreisen mit vorgeschriebenem Verhalten, G. Braun Verlag, Karlsruhe (1974).

Garcia, C. E. and M. Morari. Ind. Eng. Chem. Process Des. Dev., 21, 308 (1982).

Holt, B. R. and M. Morari. "The Effect of Dead Time on Dynamic Resilience", Chem. Eng. Sci., 40, 1229 (1985).

Holt, B. R. and M. Morari. Chem. Eng. Sci., 40, 59 (1985).

Lau, H. K. and R. A. Balhoff. Shell Development, Houston, TX, personal communication (1984).

Rivera, D. "Process Control Using Low-Order Models," M.S. Thesis, University of Wisconsin, Madison, WI (1984).

Rosenbrock, H. H. Computer-Aided Control System Design, Academic Press (1974).

Shinskey, F. G. Process Control Systems, McGraw-Hill, New York (1979).

Smith, C. L. Digital Computer Process Control, Intext Educational Publishers, Scranton, PA (1972).

Wiberg, D. M. State Space and Linear Systems, Schaum, McGraw-Hill, New York (1971).

Ziegler, J. G. and N. B. Nichols. Trans. ASME, 64, 759 (1942).

Appendix A

Analytical forms for ω_b and ω_c can be found for the following:

Option 1: $\tilde{g}_+ f = \frac{-\beta s+1}{\beta s+1} \frac{1}{(\varepsilon s+1)}$ (64)

The sensitivity operator is:

$$S = 1-\tilde{g}_+ f = (\frac{(2\beta+\varepsilon)s(\frac{\varepsilon\beta}{(2\beta+\varepsilon)} s+1)}{(\beta s+1)(\varepsilon s+1)} \tag{A.1}$$

which yields

$$\omega_b = \frac{\sqrt{-(7\beta^2+8\beta\varepsilon+\varepsilon^2)+\sqrt{(7\beta^2+8\beta\varepsilon+\varepsilon^2)^2+4(\beta\varepsilon)^2}}}{\sqrt{2}\ \beta\varepsilon} \tag{A.2}$$

Using the asymptote approximation for the amplitude of (A.1) one obtains

$$\omega_b \cong \frac{1}{2\beta+\varepsilon}$$

To obtain the cross-over frequency, we have cg:

$$cg = \frac{\tilde{g}_+}{f^{-1}-\tilde{g}_+} = \frac{-\beta s+1}{\varepsilon\beta s^2 + (\varepsilon+2\beta)s} \tag{A.3}$$

from which one obtains:

$$\omega_c = \frac{\sqrt{-(\varepsilon^2+4\beta\varepsilon+3\beta^2)+\sqrt{(\varepsilon^2+4\beta\varepsilon+3\beta^2)^2+4(\beta\varepsilon)^2}}}{\sqrt{2}\ \varepsilon\beta} \tag{A.4}$$

Again, using the asymptote approximation for the amplitude of (A.3) one obtains the simpler expression

$$\omega_c \cong \frac{1}{2\beta+\varepsilon}$$

Option 2: $\tilde{g}_+ f = \frac{-\beta s+1}{\varepsilon s+1}$ (66)

The sensitivity operator is

$$S = \frac{(\varepsilon+\beta)s}{(\varepsilon s+1)} \tag{A.5}$$

which leads to

$$\omega_b = \frac{1}{\sqrt{2(\varepsilon+\beta)^2-\varepsilon^2}} \tag{A.6}$$

or, from the asymptote amplitude approximation to (A.5),

$$\omega_b \cong \frac{1}{\varepsilon+\beta}$$

For the cross-over frequency the expressions are

$$cg = \frac{-\beta s+1}{(\varepsilon+\beta)s} \tag{A.7}$$

from which one obtains

$$\omega_c = \frac{1}{\sqrt{(\varepsilon+\beta)^2-\beta^2}} \tag{A.8}$$

and from the asymptote amplitude ratio

$$\omega_c \cong \frac{1}{\varepsilon+\beta}$$

Option 3: $\tilde{g}_+ f = \dfrac{e^{-\theta s}}{\varepsilon s+1}$

For this case it is not possible to write the bandwidth or cross-over expressions in explicit form. One can obtain approximate expression by representing the deadtime as a Padé approximation

$$\tilde{g}_+ f \cong \frac{1-\frac{\theta}{2}s}{1+\frac{\theta}{2}s}\,\frac{1}{(\varepsilon s+1)}$$

and then proceeding according to Option 1. The resulting expressions are

$$\omega_b \cong \omega_c \cong \frac{1}{\theta+\varepsilon}$$

Appendix B: Cross-over Approximation for PI and PID Rules

The expression for cg arising from the use of the IMC-PID rule to a first-order lag with deadtime process is

$$cg = \frac{(1+ \frac{\theta}{2}\ s)e^{-\theta s}}{(\frac{\theta}{2} + \varepsilon)s} \tag{B.1}$$

One can solve explicitly for the crossover frequency to obtain

$$\omega_c = \frac{1}{\theta\sqrt{(\frac{\varepsilon}{\theta} + \frac{1}{2})^2-(\frac{1}{2})^2}} \tag{B.2}$$

For e/θ > 0.8, it is reasonable to neglect the latter $(\frac{1}{2})^2$ term and thus approximate ω_c as

$$\omega_c \cong \frac{1}{\theta\sqrt{(\frac{\varepsilon}{\theta} + \frac{1}{2})^2}}$$

$$= \frac{1}{\varepsilon + \frac{\theta}{2}}$$

For the IMC-improved PI rule, cg is

$$cg = \frac{(1+(\tau+ \frac{\theta}{2}\ )s)e^{-\theta s}}{\varepsilon s(1+\tau s)} \tag{B.3}$$

As $\theta \to 0$, it is clear that the crossover frequency reaches the value

$$\omega_c \to \frac{1}{\varepsilon} \tag{B.4}$$

which implies

$$\omega_c \cong \frac{1}{\varepsilon} \tag{B.5}$$

Assuming, however, that $\tau = 0$ (the worst case), we see that (B.5) is still a good approximation. Consider that for $\tau = 0$, the crossover frequency is determined explicitly by

$$\omega_c = \frac{1}{\theta\sqrt{(\frac{\varepsilon}{\theta})^2 - (\frac{1}{2})^2}}$$

For $\varepsilon/\theta > 1.7$, $(\varepsilon/\theta)^2 \gg (1/2)^2$ and therefore

$$\omega_C \cong \frac{1}{\varepsilon}$$

is a suitable approximation for the crossover frequency.

CHAPTER III:

CONTROL-RELEVANT MODEL REDUCTION PROBLEMS FOR SISO H_2 , H_∞, AND μ-CONTROLLER SYNTHESIS

(Accepted for publication, International Journal of Control)

Control-Relevant Model Reduction Problems for SISO H_2, H_∞, and μ-Controller Synthesis

Daniel E. Rivera

Manfred Morari

Chemical Engineering, 206-41

California Institute of Technology

Pasadena, CA 91125

Abstract

The problem of model reduction in the context of control system design is investigated. Starting from closed-loop objectives (H_2, H_∞, and μ), equivalent weighted "open-loop" plant and controller reduction problems are developed. The control-relevant weight function incorporates explicitly all the important characteristics of the control problem, such as the setpoint/disturbance spectrum and the designer requirements for the sensitivity/complementary sensitivity functions. Furthermore, these control-relevant reduction problems are complimented with validation procedures that indicate rigorously the effects of the reduction problem on the desired performance objectives. A simple algorithm that uses standard regression routines is presented to solve these problems.

1. Introduction

Model reduction techniques seek to derive from a "full" order model (that is, a complete process model derived from fundamental principles or identified experimentally) a "reduced" order model (i.e., one containing fewer states) which best approximates the behavior of the complete model for a given application. In a general mathematical sense, the model reduction problem (MRP) is an optimization problem which requires minimizing a function of the weighted error between full and reduced models:

$$\mathrm{MRP} \equiv \min_{\mathrm{model}} \; f(\mathrm{Weight}, \mathrm{Error}) \tag{1}$$

Reduced-order modeling is a significant tool for simplifying the design and implementation of control systems; for this reason reduction methods have received a considerable amount of attention in the past 20 years.

Historically, the selection of functionals, weights, and error description for model reduction has been done with greatest emphasis on computational convenience. The traditional methods attempt to match certain predominant features of the full model (statistical moments, dominant poles, etc.) or minimize a functional objective such as the integral square error, the maximum error, or the integral absolute error between full and reduced models. Recent advances, most particularly model reduction through balanced realization, have revolutionized the computational aspect of the reduction problem while leaving a basic question unanswered: is the reduction objective relevant to the closed-loop control problem? Implicitly stated in these efforts is the belief that a good "open-loop" matching serves well for all applications in dynamic simulation and control system design. The review papers of Bosley and Lees (1972) and Decoster and van Cauwenberghe (1976a,b) provide a good summary of these "open-loop" procedures.

In contrast to the fervent activity on "open-loop" model reduction, the link between the reduction objective and closed-loop control (henceforth referred to as control-relevant model reduction) has received comparatively little attention in the systems literature, although its importance has begun to be emphasized in recent years (Enns, 1984, Hyland and Bernstein, 1984, Skelton and Owens, 1986). In the absence of a "separation principle" guiding the control-relevant reduction problem (that is, reduction followed by control system design, or vice versa, will always be suboptimal compared to the integrated process of reduction and design) there is much to be gained from understanding this link, as it can lead to obtaining smaller parsimonious models and, consequently, result in simpler yet more effective control systems. The most common approach to this technique has been an indirect one; usually a proposed reduction technique will account for control applications by incorporating a user-defined weighting function. The problem remains, however, of finding a relevant weighting function systematically rather than on an ad hoc basis.

Recently, Enns (1984) and Anderson (1985) have proposed weights and computational techniques which consider the stability issue of model reduction, that is, assuring that the reduced plant or controller model will result in a stable control system despite reduction error. Performance degradation effects have been studied by a number of investigators (Mitra, 1969; Hyland and Bernstein, 1984), but are of limited usefulness because they usually imply a specific controller design paradigm and the conditions presented for "control-optimal" reduction are complex and numerically involved. The efforts in this paper, while also centered on preserving performance in the face of model reduction error, attempt to be as independent as possible of a control system design procedure and address a variety of control objectives, including robust performance (Doyle, 1984, 1985) which has been a subject of recent theoretical and practical interest. Furthermore, the resulting control-relevant reduction problems are expressed in terms of physically meaningful quantities. The framework for this analysis is the Q-parametrization/Internal Model Control structure (Fig. 1b), which relates reduced models directly to feedback compensators. As a consequence, it is possible to define model reduction problems directly from the control objectives; this in turn will lead to defining weighted "open-loop" reduction problems which are easy to solve through standard regression algorithms.

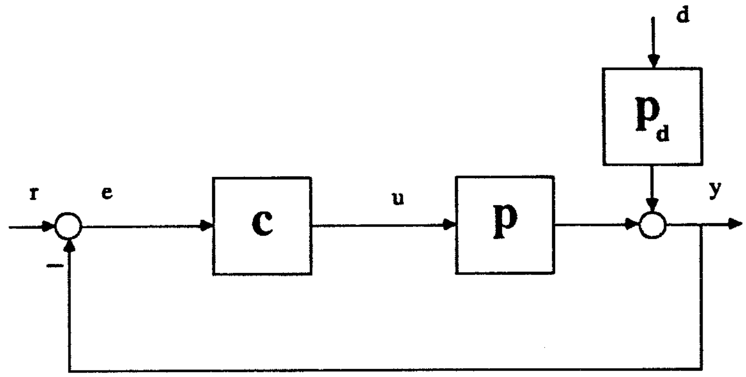
2. Preliminary: Control Objectives Considered

This work will focus on single-input, single-output systems and will assume that the plant p with u as its input and y as its output is open-loop stable. $\tilde{p}$ represents the full-order model, while $\hat{p}$ refers to the reduced-order model.

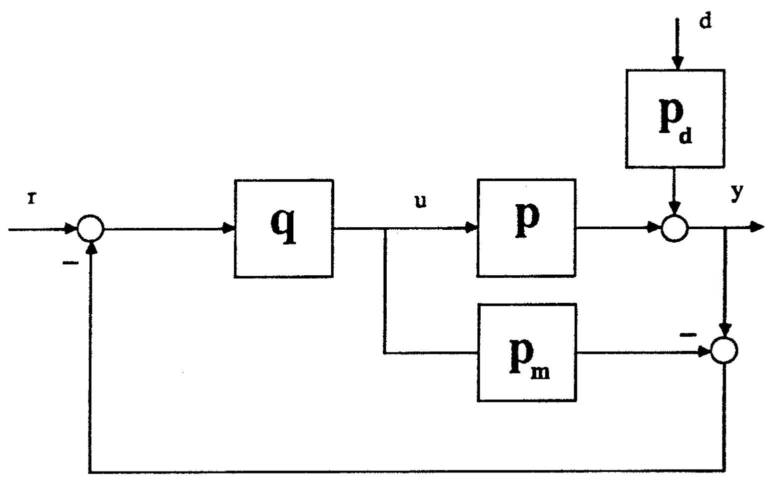
A number of control objectives will be of interest in this study. The first we will consider is the 2-norm (H_2), evaluated in the frequency domain through Parseval's Theorem:

$$||e||_2=\left(\frac{1}{\pi}\int_0^\infty |e(j\omega)|^2 d\omega\right)^{1/2} \tag{2}$$

$e(j\omega)$ is the error signal for the classical feedback controller (Fig.



(A)



(B)

Figure 1

Classical feedback structure (A) and the Internal Model Control structure (B).

1a), represented mathematically as the product of the sensitivity function ε and the setpoint/disturbance characteristics of the problem

$$e(s) = \varepsilon(s)(r(s) - p_d(s)d(s)) \tag{3}$$

The other objective of interest is the infinity norm on the weighted sensitivity function (H_∞) (Zames, 1981)

$$||w_p\varepsilon||_\infty = \sup_\omega |w_p\varepsilon(j\omega)| \tag{4}$$

From (4) one finds that the following property applies to the infinity norm:$1/|w_p|$ represents an upper bound on the sensitivity function, that is

$$|\varepsilon(j\omega)| \leq 1/|w_p(j\omega)| \qquad \forall\omega \tag{5}$$

if and only if

$$||w_p\varepsilon||_\infty \leq 1 \tag{6}$$

The previous two control objectives assume implicitly the absence of plant uncertainty ($p=\tilde{p}$). If uncertainty plays a significant effect in the control design problem, a more meaningful objective is that obtained from the μ-synthesis theory developed by Doyle and co-workers (1984). The robust performance theorem presented by these investigators is the following:

Theorem 1: Condition (5) is satisfied for all members of the family of plants π

$$\pi: \{p: |(p-\tilde{p})\tilde{p}^{-1}| \leq \ell_m\} \tag{7}$$

if and only if

$$\mu = \sup_\omega (|\tilde{\eta}\ell_m| + |w_p\tilde{\varepsilon}|) \leq 1 \tag{8}$$

$\tilde{\eta}$ is the complementary sensitivity function, related to $\tilde{\varepsilon}$ according to $\tilde{\eta} + \tilde{\varepsilon} = 1$. Both $\tilde{\eta}$ and $\tilde{\varepsilon}$ are based on the full-order model. The uncertainty description (7) is generic for SISO systems in the sense that all other norm-bounded uncertainty descriptions (additive, input or output multiplicative uncertainty) can be lumped into a single ℓ_m description without conservativeness; furthermore, comparing (4) and (8) it is clear that for the case of no uncertainty, μ and the infinity norm are equal. We will therefore consider μ in the rest of the analysis of this paper and assume that the reader will understand that it incorporates the H_∞ problem as specified in (4).

3. Internal Model Control -- Relating Reduced Models to Feedback Compensators

The lack of a convenient framework linking modeling to control system design has been a continuing obstacle towards understanding the consequences of reduced-order modeling in the control context. Given the diversity of controller design methodologies available -- LQG, pole placement, and loopshaping, to name a few, -- one would generally expect that the control implications of model reduction are greatly dependent on the design technique chosen and therefore difficult to assess in a general sense. These difficulties are avoided if, as a basis for control-relevant reduction analysis, the classical feedback structure is reparametrized into a form that allows for more insight, particularly, one that allows that the control problem be addressed with a minimal amount of information regarding the controller design procedure. Specifically, this parametrization should be able to express the reduction problem in terms of designer-specified performance requirements, the plant uncertainty description, and the setpoint/disturbance characteristics of the problem.

Such benefits are obtained by examining the reduction problem through the Internal Model Control structure. The IMC structure, introduced by Garcia and Morari in 1982 (it is equivalent to the Q-parametrization (Zames 1981) for stable systems) is an alternate albeit equivalent representation of the classical feedback structure. The relationship is denoted by the equations

$$q = c(1+cp_m)^{-1} \tag{9}$$

$$c = q(1-p_m q)^{-1} \tag{10}$$

and illustrated in Figures 1a and 1b. q represents the IMC controller.

We first consider the case of no model uncertainty ($\ell_m = 0$, $p = \tilde{p}$, i.e., the full-order model describes the plant perfectly) and use the full-order model to design the control system ($p_m = \tilde{p}$, i.e., no model reduction is performed). This case is important because it allows one to define the plant-inherent limitations to control which in turn will influence the choice of the reduced model. From (10) it becomes clear that incorporating the plant inverse in the IMC controller as $q = \tilde{p}^{-1}\tilde{\eta}$ leads to a closed-loop system possessing the desired transfer functions:

$$e = \tilde{\varepsilon}(r-p_d d) = (1-\tilde{\eta})(r-p_d d) \tag{11}$$

$$y = \tilde{\eta}(r-p_d d) + p_d d \tag{12}$$

The system is effectively open-loop and closed-loop stability is implied by the stability of $\tilde{p}$ and q. $\tilde{\eta}$, the complementary sensitivity function, is mostly at the designer's discretion, subject to the following limitations:

<u>Limitation 1</u>. Model-Inherent Limitations: Nonminimum phase elements in the plant will limit the best achievable performance for any system. Deadtime in $\tilde{p}$ could cause q to be non-causal, while Right-Half Plane zeros could result in unstable poles in q. This is avoided by requiring that the plant be factored according to

$$\tilde{p} = \tilde{p}_+\tilde{p}_- \quad (13)$$

with $\tilde{p}_+$ containing all the nonminimum phase behavior. Setting $\tilde{\eta}$ as

$$\tilde{\eta} = \tilde{p}_+ \quad (14)$$

results in a causal and internally stable controller.

Limitation 2. Input magnitude/robustness limitations: The control action must be bounded, which implies that q must be proper. If $\tilde{p}$ is proper, $\tilde{\eta}$ must be proper with a sufficiently steep roll-off. The definition of $\tilde{\eta}$ must thus be complemented with a low-pass filter f such that

$$\tilde{\eta} = \tilde{p}_+ f \quad (15)$$

Detailed discussion of procedures for computing $\tilde{\eta}$ from plant models can be found in Morari, Zafiriou and Economou (1987).

When $\hat{p}$, the reduced-order model, is used to design the control system (see Fig. 2), the limitations on control system stability and performance created by the full/reduced model mismatch must be considered. Representing this mismatch in terms of the multiplicative error e_m

$$e_m = (\tilde{p}-\hat{p})\hat{p}^{-1} \quad (16)$$

and determining q on the basis of the reduced-order model ($q = \hat{p}^{-1}\hat{\eta}$, with $\hat{\eta}$ obtained from $\hat{p}$) results in expressions for the sensitivity and complementary sensitivity functions that incorporate the reduction error:

$$\tilde{\varepsilon} = (1-\hat{\eta})(1+e_m\hat{\eta})^{-1} \quad (17)$$

$$\tilde{\eta} = (1+e_m)\hat{\eta}(1+e_m\hat{\eta})^{-1} \quad (18)$$

the full/reduced model mismatch forces the designer to reconsider the following:

1. Stability. A stable system is no longer assured if $\hat{p}$ and q are

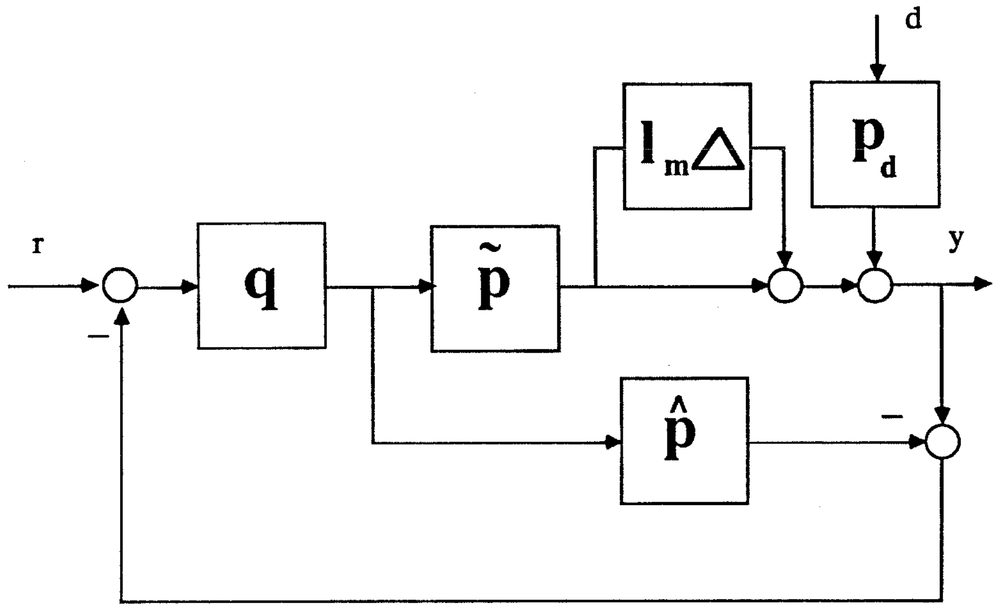


Figure 2

IMC structure for control-relevant model reduction.

stable. It follows from (17) and (18) that for closed-loop stability the Nyquist plot of $e_m\hat{\eta}$ should not encircle (-1,0). A sufficient condition (obtained through the Small Gain Theorem) establishes bounds on the magnitude of $|e_m\hat{\eta}|$

$$|e_m\hat{\eta}| < 1 \qquad \forall\omega \tag{19}$$

2. <u>Performance</u>. e_m corrupts the performance expected from $\hat{\eta}$. This effect can be assessed quantitatively by evaluating the performance objectives (2) and (8) using (17) and (18):

$$||e||_2 = \left(\frac{1}{\pi}\int_0^\infty \left|(1-\hat{\eta})(1+e_m\hat{\eta})^{-1}(r-p_d d)\right|^2 d\omega\right)^{1/2} \tag{20}$$

$$\mu = \sup_\omega \left(\left|(1+e_m)\hat{\eta}(1+e_m\hat{\eta})^{-1}\right|\ell_m + \left|w_p(1-\hat{\eta})(1+e_m\hat{\eta})^{-1}\right|\right) \tag{21}$$

The previous two equations indicate the usefulness of the IMC parametrization as a reduced model assessment tool; from a specified $\hat{\eta}$ (which reflects desired/attainable performance based on the reduced model), e_m (the model reduction error), ℓ_m (the multiplicative plant uncertainty description), w_p (the sensitivity function bound), and $r-p_d d$ (the setpoint/disturbance characteristics), an objective function of choice can be readily evaluated and compared with the objective evaluated on the basis of $\tilde{\eta}$ obtained from full-order model characteristics. Furthermore, for the robust performance objective defined by (21), Theorem 1 can be used to rigorously determine that the performance specification has been met, despite reduction error.

The expressions denoted by (20) and (21) agree with the definition of the generalized MRP defined in (1). The functionals to be minimized are the integral and the maximum of the weighted sensitivity/complementary sensitivity functions; most importantly, the <u>multiplicative</u> error e_m, rather than the commonly utilized <u>additive</u>

error e_a ($e_a = \tilde{p} - \hat{p}$), is the relevant reduction error measure for control applications. It is also clear from these expressions that the relationship between the reduction and control error is not a linear one; because of this arrangement, the solution to the model reduction problems posed by (20) and (21) is numerically difficult and requires the use of nonlinear programming routines for computation. Generalized statements regarding convergence and existence of solutions are difficult, if not impossible to state in general. Rivera (1984) has implemented a programming solution for the problem defined by (20) and found it to be practical only when a limited number of model parameters need to be obtained.

4. Obtaining Open-Loop Reduction Criteria from Closed-Loop Control Objectives

The numerical difficulties associated with a rigorous solution to the control-relevant model reduction problem make it desirable to obtain computationally convenient formulations in the spirit of the "open-loop" methods which have predominated the model reduction literature. Our purpose in this section will be to indicate the analysis that leads to stating the MRP's defined by (20) and (21) in the following form:

$$\text{MRP} = \min_{\text{model}} \quad f(\text{Weight } |e_m|^k) \qquad (22)$$

The first step in deriving near-equivalent "open-loop" reduction problems is to conveniently approximate the functions $\tilde{\eta}$ and $\tilde{\varepsilon}$. Assuming that condition (19) holds (which, as discussed in the previous section, implies stability of the closed-loop system despite reduction error), then (17) and (18) can be expanded into geometric series

$$\tilde{\varepsilon} = (1-\hat{\eta})(1-e_m\hat{\eta} + (e_m\hat{\eta})^2 - (e_m\hat{\eta})^3 + \ldots) \tag{23}$$

$$\tilde{\eta} = (1+e_m)\hat{\eta}(1-e_m\hat{\eta} + (e_m\hat{\eta})^2 - (e_m\hat{\eta})^3 + \ldots) \tag{24}$$

Further simplification is achieved by neglecting second-order terms and higher; this approximation is valid if $|e_m\hat{\eta}| << 1$ over the bandwidths of $(1-\hat{\eta})$ and $\hat{\eta}$, respectively

$$\tilde{\varepsilon} \mathrel{\hat{=}} (1-\hat{\eta})(1-e_m\hat{\eta}) \tag{25}$$

$$\tilde{\eta} \mathrel{\tilde{=}} \hat{\eta}(1+(1-\hat{\eta})e_m) \tag{26}$$

Having obtained the simplified expressions (25) and (26), one is now able to relate the multiplicative error in a linear fashion to the functions $\tilde{\eta}$ and $\tilde{\varepsilon}$; the first term of (25) and (26) incorporates the reduced-model inherent limitations to control, while the second term denotes the degradation occurring from full/reduced model mismatch. The control-relevant reduction problem requires obtaining a model such that the degradation resulting from the multiplicative error is minimized.

Substituting (25) and (26) into the control objectives defined by (2) and (8) one obtains

$$||e||_2 \mathrel{\tilde{=}} ||e||_{2,approx} = \left(\frac{1}{\pi}\int_0^\infty \left(|1-\hat{\eta}|^2|1-e_m\hat{\eta}|^2|r-p_d d|^2\, d\omega\right)^{1/2}\right. \tag{27}$$

$$\mu \mathrel{\tilde{=}} \mu_{,approx} = \sup_\omega \left(|\tilde{\eta}(1+(1-\hat{\eta})e_m)|\ell_m + |w_p(1-\hat{\eta})(1-e_m\hat{\eta})|\right) \tag{28}$$

From the property

$$||a||_\alpha - ||b||_\alpha \leq ||a+b||_\alpha \leq ||a||_\alpha + ||b||_\alpha,\ \alpha = 1, \ldots, p, \ldots, \infty \tag{29}$$

it is possible to separate the effect of e_m on the 2-norm/μ objectives:

$$||(1-\hat{\eta})(r-p_d d)||_2 - ||(1-\hat{\eta})(r-p_d d)\hat{\eta}e_m||_2 \leq ||e||_{2,approx} \leq$$
$$||(1-\hat{\eta})(r-p_d d)||_2 + ||(1-\hat{\eta})(r-p_d d)\hat{\eta}e_m||_2 \tag{30}$$

$$|| \; |\hat{\eta}\ell_m| + |w_p(1-\hat{\eta})| \; ||_\infty - ||(|1-\hat{\eta}|\ell_m + |w_p(1-\hat{\eta})|)\hat{\eta}e_m||_\infty \leq \mu,\text{approx} \leq$$
$$|| \; |\hat{\eta}\ell_m| + |w_p(1-\hat{\eta})| \; ||_\infty + ||(|1-\hat{\eta}|\ell_m + |w_p(1-\hat{\eta})|)\hat{\eta}e_m||_\infty \qquad (31)$$

Control-relevant model reduction problems of the form (22) are obtained by minimizing this degradation term:

$$\text{P1:} \quad \min_{\hat{p}} \left(\int_0^\infty \left|(1-\hat{\eta})^2(r-p_d d)^2\right| \left|\hat{\eta}e_m\right|^2 d\omega = \min_{\hat{p}} \; J_1 \qquad (32)$$

$$\text{P2:} \quad \inf_{\hat{p}} \; \sup_{\omega} \; (|(1-\hat{\eta})\ell_m| + |w_p(1-\hat{\eta})|)|\hat{\eta}e_m| = \inf_{\hat{p}} \; J_2 \qquad (33)$$

The weight functions of these MRP's incorporate the following aspects of the control system design problem:

1. The nominal sensitivity function $(1-\hat{\eta})$: this portion of the weight attenuates the low frequencies.
2. The nominal complementary sensitivity $\hat{\eta}$: this function attenuates the high frequencies and provides a roll-off point for the multiplicative error.
3. The setpoint/disturbance characteristics of the problem $(r-p_d d)$: In most cases, the setpoint/disturbance spectrum will attenuate the error at high frequencies. Certain inputs, such as step changes, will place greater emphasis on the low-frequency component of the error.
4. The sensitivity function bound $1/|w_p|$: this function attenuates the low frequencies.
5. The uncertainty description ℓ_m. Depending on the nature of the problem, the effect of uncertainty could extend to all frequencies. In general, ℓ_m tends to be small at low frequencies, increasing towards infinity for unstructured cases.

The associated problem of controller reduction can also be addressed in a fashion similar to that of the plant reduction problem. This is achieved by re-writing the sensitivity function as:

$$(1+\tilde{p}\hat{c})^{-1} = \tilde{\varepsilon}(1-\tilde{\varepsilon}\tilde{p}(\tilde{c}-\hat{c}))^{-1} \tag{34}$$

with $\tilde{\varepsilon}$ representing the sensitivity function obtained from the full-order controller.

Requiring that

$$\left|\tilde{\varepsilon}\tilde{p}(\tilde{c}-\hat{c})\right| < 1 \qquad \forall\omega \tag{35}$$

and expanding (34) into a geometric series results in the following expressions for the sensitivity and complementary sensitivity functions

$$(1+\tilde{p}\hat{c})^{-1} \cong (1-\tilde{\eta}) + (\tilde{c}-\hat{c})\tilde{p}(1-\tilde{\eta})^2 \tag{36}$$

$$\tilde{p}\hat{c}(1+\tilde{p}\hat{c})^{-1} \cong \tilde{\eta} - (\tilde{c}-\hat{c})\tilde{p}(1-\tilde{\eta})^2 \tag{37}$$

and to the controller reduction problems

$$\text{P3:} \quad \min_{\hat{c}} \int_0^\infty \left|(1-\tilde{\eta})^2\tilde{p}(r-p_d d)\right|^2\left|\tilde{c}-\hat{c}\right|^2 d\omega = \min_{\hat{c}} J_3 \tag{38}$$

$$\text{P4:} \quad \inf_{\hat{c}} \sup_{\omega} (\ell_m+|w_p|)\left|1-\tilde{\eta}\right|^2\left|\tilde{p}\right|\left|\tilde{c}-\hat{c}\right| = \inf_{\hat{c}} J_4 \tag{39}$$

As before, the weights in (38) and (39) directly incorporate the sensitivity/complementary sensitivity functions, the setpoint/disturbance spectrum, the sensitivity function bound, and the uncertainty description. Contrary to the plant reduction problem, however, the additive error $(\tilde{c}-\hat{c})$ between full and reduced controllers is the suitable error description.

It is of interest to compare the proposed model reduction problems with those presented by Enns (1984). Enns' control-relevant criterion is to assure stability in the face of reduction error in an environment free from plant uncertainty. The relevant functional is thus the infinity norm, and the specific MRP's are

Plant-order reduction:

$$\inf_{\hat{p}} \sup_{\omega} \left|\hat{\eta} e_m\right| \tag{40}$$

Controller-order reduction:

$$\inf_{\hat{c}} \sup_{\omega} \left|(1-\tilde{\eta})\tilde{p}\right|\left|\tilde{c}-\hat{c}\right| \tag{41}$$

The problems defined by (40) and (41) correspond closely to problems (33) and (39) defined for the μ objective; the extra factor

$$(\ell_m + \left|w_p\right|)\ \left|1-\eta\right| \tag{42}$$

$\eta = \hat{\eta}$: plant reduction $\quad\quad \eta = \tilde{\eta}$: controller reduction

is included, indicating the frequency range over which $\left|\hat{\eta} e_m\right|$ or $\left|(1-\tilde{\eta})\tilde{p}(\tilde{c}-\hat{c})\right|$ must be kept much less than one.

An issue that remains to be addressed is the computational aspect of the problem. Minimizing the functionals that appear in the proposed problems has been discussed previously in the literature; for the 2-norm (ISE) in the frequency domain, one can refer to the work of Noldus and Decoster (1976); with regards to the μ problem, the method of frequency-weighted balanced realizations (Enns, 1984) can be used. Two obstacles, however, are encountered in the case of plant reduction. The first is that most literature approaches consider the additive error e_a, while the criterion of interest is the multiplicative error e_m. The second is the presence of $\hat{\eta}$ in the control-relevant weight; the nonminimum phase characteristics of the reduced model (represented in $\hat{p}_+$) are not known a priori. Both of these obstacles are recognized by Enns, who indicates that in such cases the additive error problem can be stated as a fixed point problem (the reduced model inverse $\hat{p}^{-1}$ now forms part of the weight) with the method of successive approximation used to obtain a solution. The reduced model resulting from each iteration is factored (for the purposes of updating $\hat{\eta}$) and introduced in the weight; assuming existence of a solution and that the functional of the weighted error is a contraction, the method will converge.

5. Synthesis Methods for Control-Relevant Model Reduction

Despite the availability of computational methods for solving the problems developed in the previous section, there are still motivating reasons for exploring alternate computational schemes. First of all, the problems presented in the previous section are still nonlinear programming problems and require significant computational effort. Furthermore, when using the balanced realization technique it is required that the weights and models be in state-space form, thus forbidding the use of models with deadtime. Systems with poles on the $j\omega$-axis are also excluded. As an added drawback, error bounds for the frequency-weighted case are currently not available. Our goal, then, is to present simpler, regression-style computational schemes for solving the proposed problems which are easy to implement using widely accessible software packages.

The computational procedure presented in this section is patterned after the generalized framework for frequency response matching proposed by Stahl (1984). The philosophy behind this method is that the nonlinear programming problem can be reformulated as an iterative linear one.

The first step towards developing this alternate computational scheme is to consider the full-order model $\tilde{p}$ or $\tilde{c}$ nonparametrically in terms of its frequency response (z triplets consisting of the full-order plant or controller evaluated at a frequency with its associated real (R) and imaginary (I) parts).

$$((\omega_1,R_1,I_1),\ldots,(\omega_i,R_i,I_i),\ldots,(\omega_z,R_z,I_z)) \tag{43}$$

Because of this full-order plant representation, the previously stated MRP's must thus be defined in terms of discrete, rather than continuous operators. For plant reduction we have

$$\text{P1:} \quad J_1 \tilde{=} J_1' = \sum_{i=1}^{z} \left|(1-\hat{\eta})\hat{\eta}(r-p_d d)\right|^2_{s=j\omega_i} \; |e_m(j\omega_i)^2 \Delta\omega_i \tag{44}$$

$$\Delta\omega_i = \omega_i - w_{i-1}$$

$$\text{P2:} \quad J_2 \tilde{=} J_2' = \max_{\omega_i}((\ell_m(\omega_i)+\left|w_p(j\omega_i)\right|)\left|(1-\hat{\eta})\hat{\eta}\right|_{s=j\omega_i} \; \left|e_m(j\omega_i)\right|) \tag{45}$$

and for controller reduction the discretized objectives are

$$\text{P3:} \quad J_3 \tilde{=} J_3' = \sum_{i=1}^{z} \left|(1-\tilde{\eta})^2\tilde{p}(r-p_d d)\right|^2_{s=j\omega_i} \; \left|e_c(j\omega_i)\right|^2 \Delta\omega_i \tag{46}$$

$$\text{P4:} \quad J_4 \tilde{=} J_4' = \max_{\omega_i} \; ((\ell_m(\omega_i)+\left|w_p(j\omega_i)\right|)\left|(1-\tilde{\eta})^2\tilde{p}\right|_{s=j\omega_i} \; \left|e_c(j\omega_i)\right| \tag{47}$$

$$e_c(j\omega_i) = \tilde{c}(j\omega_i) - \hat{c}(j\omega_i)$$

It should be clear to the reader that as z increases, (44)-(47) more closely match their continuous counterparts, (32), (33), (38), and (39). We will consider reduced models of the form

$$\hat{p} \text{ or } \hat{c} = \frac{a_0+a_1 s+\ldots+a_n s^n}{b_0+b_1 s+\ldots+b_m s^m} e^{-\theta s} = \frac{N(s)}{D(s)} e^{-\theta s} \tag{48}$$

where $b_0 = 1$ for a model without integrator and $b_0 = 0$, $b_1 = 1$ for a model with a single integrator. Combining the definitions (43) and (48) leads to representing e_m and e_c as follows:

$$e_m(j\omega_i) = \left.\frac{D(s)(R_i+jI_i)-N(s)e^{-\theta s}}{N(s)e^{-\theta s}}\right|_{s=j\omega_i} = \left.\frac{e_L}{N(s)e^{-\theta s}}\right|_{s=j\omega_i} \tag{49}$$

$$e_c(j\omega_i) = \left.\frac{D(s)(R_i+jI_i)-N(s)e^{-\theta s}}{D(s)}\right|_{s=j\omega_i} = \left.\frac{e_L}{D(s)}\right|_{s=j\omega_i} \tag{50}$$

where

$$e_L(j\omega_i) = (b_0+b_1s+\ldots+b_ms^m)(R_i+jI_i) - (a_0+a_0s+\ldots+a_ns^n)e^{-\theta s}\big|_{s=j\omega_i} \quad (51)$$

e_L is a linear function of the parameters $[b_0,b_1,\ldots,b_m]$ and$[a_0,a_1,\ldots,a_n]$, and can therefore be represented in matrix form as follows:

$$\mathbf{e}_L^T = [\mathrm{Re}(e_L(j\omega_1)),\ldots,\mathrm{Re}(e_L(j\omega_i)),\ldots,\mathrm{Re}(e_L(j\omega_z)),$$
$$\mathrm{Im}(e_L(j\omega_1)),\ldots,\mathrm{Im}(e_L(j\omega_i)),\ldots,\mathrm{Im}(e_L(j\omega_z))] \quad (52)$$

$$\mathbf{e_L} = \mathbf{Mp} - \mathbf{v},\ \mathbf{M} = \begin{bmatrix}\mathbf{M_R}\\ \hline \mathbf{M_I}\end{bmatrix},\ \mathbf{v} = \begin{bmatrix}\mathbf{v_R}\\ \hline \mathbf{V_I}\end{bmatrix} \quad (53)$$

For a system without integrator ($b_0 = 1$), the form for $\mathbf{p}$, $\mathbf{M_R}$, $\mathbf{M_I}$, $\mathbf{v_R}$, and $\mathbf{v_i}$ is the following:

$$\mathbf{p}^T = [a_0,\ldots,a_n,b_1,\ldots,b_m] \quad (54)$$

$$\mathbf{M_R} = \begin{bmatrix} \cos(\theta\omega_1) \ldots \mathrm{Re}((j\omega_1)^n\ e^{-\theta j\omega_1}) & I_1\omega_1 \ldots \mathrm{Re}((-j\omega_1)^m\ (R_1+jI_1)) \\ \cos(\theta\omega_i) \ldots \mathrm{Re}((j\omega_i)^n\ e^{-\theta j\omega_i}) & I_i\omega_i \ldots \mathrm{Re}((-j\omega_i)^m\ (R_i+jI_i)) \\ \cos(\theta\omega_z) \ldots \mathrm{Re}((j\omega_z)^n\ e^{-\theta j\omega_z}) & I_z\omega_z \ldots \mathrm{Re}((-j\omega_z)^m\ (R_z+jI_z)) \end{bmatrix}$$

$$\mathbf{M_I} = \begin{bmatrix} -\sin(\theta\omega_1) \ldots \mathrm{Im}((j\omega_1)^n\ e^{-\theta j\omega_1}) & -R_1\omega_1 \ldots \mathrm{Im}((-j\omega_1)^m\ (R_1+jI_1)) \\ -\sin(\theta\omega_i) \ldots \mathrm{Im}((j\omega_i)^n\ e^{-\theta j\omega_i}) & -R_i\omega_i \ldots \mathrm{Im}((-j\omega_i)^m\ (R_i+jI_i)) \\ -\sin(\theta\omega_z) \ldots \mathrm{Im}((j\omega_z)^n\ e^{-\theta j\omega_z}) & -R_z\omega_z \ldots \mathrm{Im}((-j\omega_z)^m\ (R_z+jI_z)) \end{bmatrix}$$

$$\mathbf{v}_R^T = [R_1,\ \ldots,\ R_i,\ \ldots,\ R_z] \qquad \mathbf{v}_I^T = [I_1,\ \ldots,\ I_i,\ \ldots,\ I_z]$$

For a system with integrator ($b_0=0$, $b_1=1$), we have

$$\mathbf{p}^T = [a_0,\ \ldots,\ a_n,\ b_2,\ \ldots,\ b_m] \quad (55)$$

$$\mathbf{M_R} = \begin{bmatrix} \cos(\theta\omega_1) \ldots \mathrm{Re}((j\omega_1)^n\ e^{-\theta j\omega_1}) & R_1\omega_1^2 \ldots \mathrm{Re}((-j\omega_1)^m\ (R_1+jI_1)) \\ \cos(\theta\omega_i) \ldots \mathrm{Re}((j\omega_i)^n\ e^{-\theta j\omega_i}) & R_i\omega_i^2 \ldots \mathrm{Re}((-j\omega_i)^m\ (R_i+jI_i)) \\ \cos(\theta\omega_z) \ldots \mathrm{Re}((j\omega_z)^n\ e^{-\theta j\omega_z}) & R_z\omega_z^2 \ldots \mathrm{Re}((-j\omega_z)^m\ (R_z+jI_z)) \end{bmatrix}$$

$$\mathbf{M_I} = \begin{bmatrix} -\sin(\theta\omega_1) \ \dots \ \mathrm{Im}((j\omega_1)^n \ e^{-\theta j\omega_1}) & I_1\omega_1^2 \ \dots \ \mathrm{Im}((-j\omega_1)^m \ (R_1+jI_1)) \\ -\sin(\theta\omega_i) \ \dots \ \mathrm{Im}((j\omega_i)^n \ e^{-\theta j\omega_i}) & I_1\omega_i^2 \ \dots \ \mathrm{Im}((-j\omega_i)^m \ (R_i+jI_i)) \\ -\sin(-\theta\omega_Z) \ \dots \ \mathrm{Im}((j\omega_Z)^n \ e^{-\theta j\omega_Z}) & I_1\omega_Z^2 \ \dots \ \mathrm{Im}((-j\omega_Z)^m \ (R_Z+jI_Z)) \end{bmatrix}$$

$$\mathbf{v_R^T} = [-I_1\omega_1, \ \dots, \ -I_i\omega_i, \ \dots, \ -I_Z\omega_Z] \qquad \mathbf{v_I^T} = [R_1\omega_1, \ \dots, \ R_i\omega_i, \ \dots, \ R_Z\omega_Z]$$

The matrix representation for e_L can be extended to preassign values of specific reduced model parameters and thus account for previous knowledge of the process. If $\mathbf{p_F}$ is the portion of $\mathbf{p}$ to be preassigned to a desired value and $\mathbf{p_{id}}$ is the portion that remains to be identified, $\mathbf{M}$ must be partitioned according to

$$\mathbf{e_L} = [\mathbf{M_{id}} \mid \mathbf{M_F}] \left[\frac{\mathbf{p_{id}}}{\mathbf{p_f}}\right] - \mathbf{v} \tag{56}$$

where $\mathbf{M_{id}}$ and $\mathbf{M_f}$ contain the column vectors corresponding to the coefficients in $\mathbf{p_{id}}$ and $\mathbf{p_F}$, respectively. Rearranging, (56) becomes

$$\mathbf{e_{L,id}} = \mathbf{M_{id}} \ \mathbf{p_{id}} - \mathbf{v_{id}} \tag{57}$$

with

$$\mathbf{v_{id}} = \mathbf{v} - \mathbf{M_F p_F} \tag{58}$$

Our objective is to obtain expressions for $|e_m|$ and $|e_c|$ based on $\mathbf{e_L}$; wishing to conserve a real-valued matrix representation, we introduce a factor which preserves the magnitude but alters the phase

$$e_m'(j\omega_i) = \left.\frac{N(s)e^{-\theta s}}{|N(s)e^{-\theta s}|}\right|_{s=j\omega_i} \quad e_m(j\omega_i) = \left.\frac{e_L(j\omega_i)}{|N(s)e^{-\theta s}|}\right|_{s=j\omega_i} \tag{59}$$

$$e_c'(j\omega_i) = \left.\frac{D(s)}{|D(s)|}\right|_{s=j\omega_i} \quad e_c(j\omega_i) = \left.\frac{e_L(j\omega_i)}{|D(s)|}\right|_{s=j\omega_i} \tag{60}$$

To incorporate e_m' and e_c' into the matrix representation of e_L we define

$$\mathbf{e_m'^T} = [\mathrm{Re}(e_m'(j\omega_1)), \ \dots, \ \mathrm{Re}(e_m'(j\omega_i)), \ \dots, \ \mathrm{Re}(e_m'(j\omega_Z)),$$
$$\mathrm{Im}(e_m'(j\omega_1)), \ \dots, \ \mathrm{Im}(e_m'(j\omega_i)), \ \dots, \ \mathrm{Im}(e_m'(j\omega_Z))] \tag{61}$$

$$\mathbf{e_m'} = \mathbf{W_m(Mp-v)} \tag{62}$$

$$\mathbf{e_c'T} = [\mathrm{Re}(e_c'(j\omega_1)), \ldots, \mathrm{Re}(e_c'(j\omega_i)), \ldots, \mathrm{Re}(e_c'(j\omega_Z)),$$
$$\mathrm{Im}(e_c'(j\omega_1)), \ldots, \mathrm{Im}(e_c'(j\omega_i)), \ldots, \mathrm{Im}(e_c'(j\omega_Z))] \tag{63}$$

$$\mathbf{e_c'} = \mathbf{W_c(Mp-v)} \tag{64}$$

In all cases the weight matrix has the general form

$$\mathbf{W_m} = \begin{bmatrix} \mathbf{W_m^*} & \mathbf{0} \\ 0 & \mathbf{W_m^*} \end{bmatrix} \quad ; \quad \mathbf{W_c} = \begin{bmatrix} \mathbf{W_c^*} & \mathbf{0} \\ 0 & \mathbf{W_c^*} \end{bmatrix} \tag{65}$$

The appropriate weight submatrices are

$$\mathbf{W_m^*} = \mathrm{diag}(1/\left|N(j\omega)\ e^{-\theta j\omega_1}\right|, \ldots, 1/\left|N(j\omega_i)\ e^{-\theta j\omega_i}\right|, \ldots, 1/\left|N(j\omega_Z)\ e^{-\theta j\omega_Z}\right| \tag{66}$$

$$\mathbf{W_c^*} = \mathrm{diag}(1/\left|D(j\omega_1)\right|, \ldots, 1/\left|D(j\omega_i)\right|, \ldots, 1/\left|D(j\omega_Z)\right|) \tag{67}$$

To obtain a weighted error description compatible with that derived in the previous section, we must define a second weight matrix which contains the control-relevant weight. This weight is defined as:

$$\mathbf{W_\ell} = \begin{bmatrix} \mathbf{W_\ell^*} & \mathbf{0} \\ 0 & \mathbf{W_\ell^*} \end{bmatrix} \qquad \ell = 1,2,3,4 \tag{68}$$

with submatrices defined according to

$$\begin{aligned} \mathbf{W_1^*} = \mathrm{diag}(&\left|(1-\hat{\eta})\hat{\eta}(r-p_d d)\right|_{s=j\omega_1} \sqrt{\Delta\omega_1}\ , \ldots, \\ &\left|(1-\hat{\eta})\hat{\eta}(r-p_d d)\right|_{s=j\omega_i} \sqrt{\Delta\omega_i}\ , \ldots, \\ &\left|(1-\hat{\eta})\hat{\eta}(r-p_d d)\right|_{s=j\omega_Z} \sqrt{\Delta\omega_Z} \end{aligned} \tag{69}$$

$$\begin{aligned} \mathbf{W_2^*} = \mathrm{diag}((&\ell_m(\omega_1) + \left|w_p(j\omega_1)\right|)\left|(1-\hat{\eta})\hat{\eta}\right|_{s=j\omega_1}, \ldots, \\ (&\ell_m(\omega_i) + \left|w_p(j\omega_i)\right|)\left|(1-\hat{\eta})\hat{\eta}\right|_{s=j\omega_i}, \ldots, \\ (&\ell_m(\omega_Z) + \left|w_p(j\omega_Z)\right|)\left|(1-\hat{\eta})\hat{\eta}\right|_{s=j\omega_Z}) \end{aligned} \tag{70}$$

$$\begin{aligned} \mathbf{W_3^*} = \mathrm{diag}(&\left|(1-\tilde{\eta})^2\tilde{p}(r-p_d d)\right|_{s=j\omega_1} \sqrt{\Delta\omega_1}\ , \ldots, \\ &\left|(1-\tilde{\eta})^2\tilde{p}(r-p_d d)\right|_{s=j\omega_i} \sqrt{\Delta\omega i}\ , \ldots, \end{aligned} \tag{71}$$

$$\left|(1-\tilde{\eta})^2\tilde{p}(r-p_d d)\right|_{s=j\omega_z} \sqrt{\Delta\omega_z}$$

$$\begin{aligned} \mathbf{W}_4^* = \operatorname{diag}((\ell_m(\omega_1) + \left|w_p(j\omega_1)\right|)\left|(1-\tilde{\eta})^2\tilde{p}\right|_{s=j\omega_1}, \ldots, \\ (\ell_m(\omega_i) + \left|w_p(j\omega_i)\right|)\left|(1-\tilde{\eta})^2\tilde{p}\right|_{s=j\omega_i}, \ldots, \\ (\ell_m(\omega_z) + \left|w_p(j\omega_z)\right|)\left|(1-\tilde{\eta})^2\tilde{p}\right|_{s=j\omega_z}) \end{aligned} \tag{72}$$

The relationship between the weighted error and the discrete objectives (44)-(47) are

$$J_1' = \left|\left|\mathbf{W}_1\ \mathbf{e}_\mathbf{m}'\right|\right|_2^2 \tag{73}$$

$$\left|\left|\mathbf{W}_2\ \mathbf{e}_\mathbf{c}'\right|\right|_\infty \leq J_2' \leq \sqrt{2}\ \left|\left|\mathbf{W}_2\ \mathbf{e}_\mathbf{c}'\right|\right|_\infty \tag{74}$$

$$J_3' = \left|\left|\mathbf{W}_3\ \mathbf{e}_\mathbf{c}'\right|\right|_2^2 \tag{75}$$

$$\left|\left|\mathbf{W}_4\ \mathbf{e}_\mathbf{c}'\right|\right|_\infty \leq J_4' \leq \sqrt{2}\ \left|\left|\mathbf{W}_4\ \mathbf{e}_\mathbf{c}'\right|\right|_\infty \tag{76}$$

$$||\cdot|| = \text{p-vector norm}$$

The generalized problem solved by using this representation is

$$\min_{\theta,\mathbf{p}} \ \left|\left|\mathbf{W}_\ell \mathbf{W}_\mathbf{h}(\mathbf{Mp}-\mathbf{v})\right|\right|_\alpha \quad \alpha = 2,\infty; \quad \ell = 1,2,3,4; \ \mathbf{h} = \mathbf{m} \text{ or } \mathbf{c} \tag{77}$$

From (77) one realizes the benefit of the defined weighted error representation. The original problem, nonlinear in nature, can now be decomposed into a single parameter nonlinear search for the deadtime and an iterative quadratic (for the 2-norm case) or linear (for the μ case) programming problem for the parameters of the rational portion of the model, as illustrated in Figure 3.

Figure 3 charts the proposed computational procedure. The programming problem is embedded in the nonlinear search routine for deadtime. The input sequence of the program requires that the user specify the problem at hand (plant or controller reduction), the objective of interest (H_2 of μ), the desired closed-loop performance ($\hat{\eta}$ and w_p), the uncertainty description ℓ_m, the setpoint/disturbance

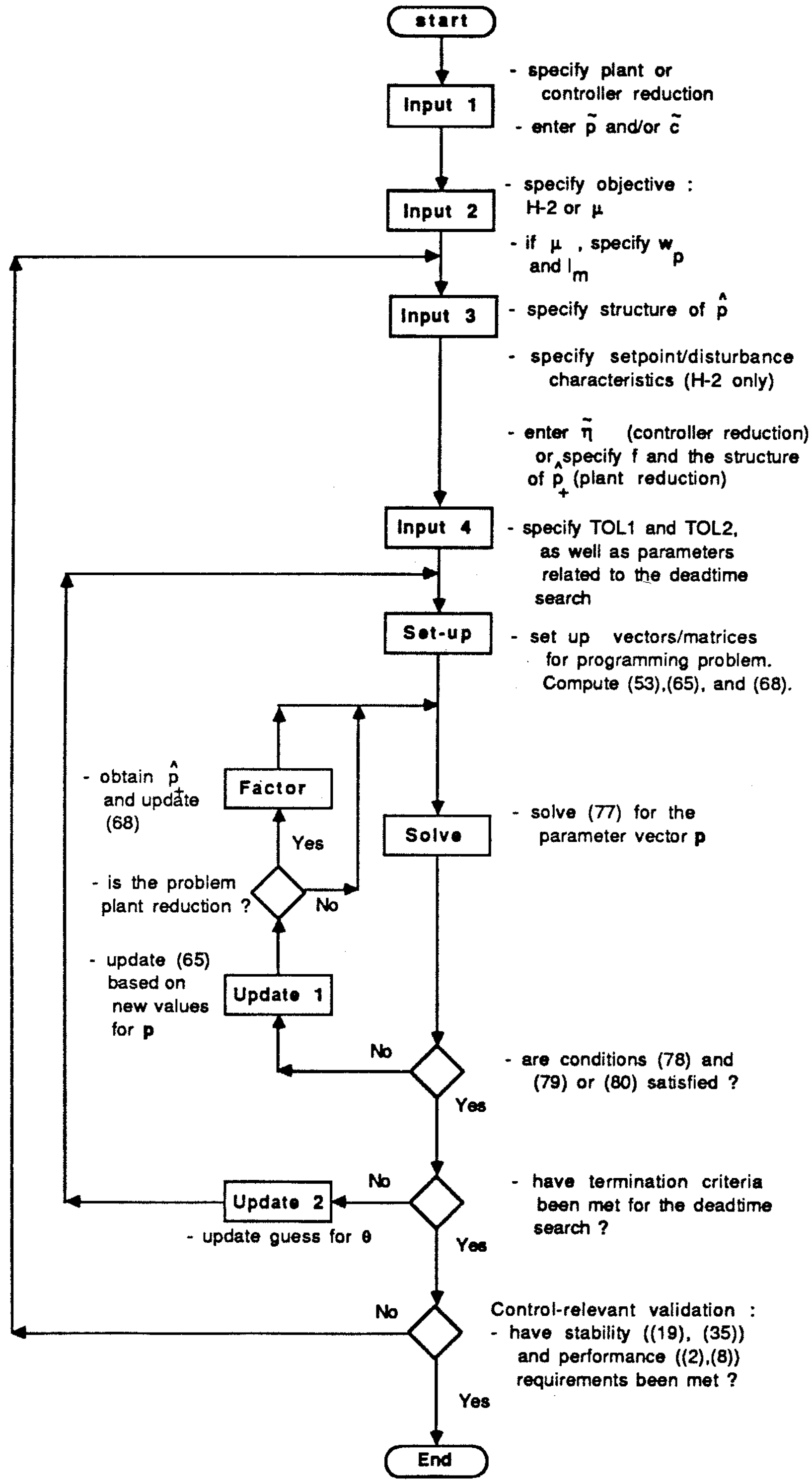


Figure 3

Flowchart for the control-relevant model reduction algorithm.

characteristics ($r-p_d d$), the structure of the reduced-order model, a guess for the deadtime (if desired), and various tolerances which indicate termination of the computational procedure. This information is sufficient to define the weight $\mathbf{W}_\ell$. Initially, the weight functions $\mathbf{W}_c$ and $\mathbf{W}_m$ are defined on the basis of user-supplied values for N(s) and D(s) (our experience has shown that setting N(s) or D(s) equal to 1 is adequate). For a given value of θ, the problem defined by (77) is readily solved as a least-squares problem (for the H_2 criterion), or by the simplex method (for the μ criterion). The authors were successful in utilizing the IMSL routines LLSQF and RLLMV for this purpose. Having obtained an initial solution vector $\mathbf{p}_1$, the results are then used to update the weight matrices $\mathbf{W}_c$ or $\mathbf{W}_m$; termination is determined when the elements of $\mathbf{p}$ do not vary appreciably with added iterations. In our application, we require that two termination criteria be satisfied; one is the maximum relative change in the parameters

$$\left|\left| \frac{p_{k+1}-p_k}{p_{k+1}} \right|\right|_\infty \leq \text{TOL1} \tag{78}$$

The other is the relative error in the residuals

$$\left| \frac{||\mathbf{W}_i^{(k+1)} e_m'||_\alpha - ||\mathbf{W}_i^{(k)} e_m'||_\alpha}{||\mathbf{W}_i^{(k+1)} e_m'||_\alpha} \right| \leq \text{TOL2} \tag{79}$$

for controller reduction

$$\left| \frac{||\mathbf{W}_i^{(k+1)} e_c'||_\alpha - ||\mathbf{W}_i^{(k)} e_c'||_\alpha}{||\mathbf{W}_i^{(k+1)} e_c'||_\alpha} \right| \leq \text{TOL2} \tag{80}$$

The reader will note from Figure 3 and the previous discussion that the plant reduction problem carries an intermediate step. Because an a priori knowledge of the reduced model is not available, initially the user can only define the low pass filter portion f in $\hat{\eta}$.

Internally, the program must compute the factor $\hat{p}_+$ based on a previously defined formula. The most convenient one is the allpass, defined as

$$\hat{p}_+ = \prod_{i}^{N} \frac{-\beta_i s+1}{\beta_i s+1} \quad \mathrm{Re}(\beta_i) \geq 0 \tag{81}$$

The final step in the control-relevant reduction procedure is the validation of the resulting reduced model. This involves two steps:

1. <u>Stability Validation</u>. One must insure that the reduced-order controller (obtained directly or implied by $\hat{p}$ and the choice of $\hat{\eta}$) leads to a stable closed-loop system when applied to the full-order plant (i.e., nominal stability). This is confirmed in the plant reduction problem by verifying that the Small Gain Theorem, condition (19), be satisfied. Its equivalent for the controller reduction case is (35).
2. <u>Performance Validation</u>. Performance validation requires being able to ascertain the degradation caused by reduction. For the μ-synthesis problem, performance degradation is rigorously assessed by confirming that Theorem 1 is satisfied. For the 2-norm case, performance degradation can be measured by computing the H_2 objective (20).

The effectiveness of the proposed techniques hinges on satisfying the previously specified diagnostics. In the case that an obtained model fails the validation criteria, the designer can either decrease the bandwidth and/or increase the allowable resonance peak implied by w_p^{-1} and $\hat{\eta}$ (i.e., make the response more sluggish or tolerate more oscillation and overshoot) or increase the number of parameters in the model by making new choices for n and m (i.e., decrease the mismatch).

6. Examples

The following examples demonstrate the usefulness of the proposed methods.

Example 1: Mandler et al. (1986) have considered the problem of outlet temperature control of a 49th-order methanation reactor, whose impulse response to inlet gas temperature is presented in Figure 4. The process is subject to step decreases in the entering carbon monoxide concentration; this effect is adequately approximated by the transfer function

$$p_d d = \frac{-0.333}{(51s+1)s} \tag{82}$$

It was desired to obtain a fourth-order semiproper model to describe the reactor

$$\hat{p} = \frac{a_0+a_1s+a_2s^2+a_3s^3+a_4s^4}{1+b_1s+b_2s^2+b_3s^3+b_4s^4} \tag{83}$$

The H_∞ model reduction problem was chosen for this case study. A reasonable weight function for this problem is

$$w_{p_1}^{-1} = \left(\frac{120s}{60s+1}\right)(s+1)(51s+1) \tag{84}$$

(84) incorporates two main features which affect the nature of the control problem. The first term, 120s/(60s+1), denotes the inherent limitation caused by the 60 second deadtime ont he bandwidth of the sensitivity function. The two lead terms, (s+1) and (51s+1), incorporate our knowledge of the process disturbance and recognize that at frequencies higher than 0.02 rad/sec, the sensitivity function will become significantly attenuated by the disturbance.

The details of the control-relevant reduction procedure are as follows: The full-order model was represented by 150 points in the frequency interval [0.0001,1]. The complementary sensitivity function $\hat{\eta}$

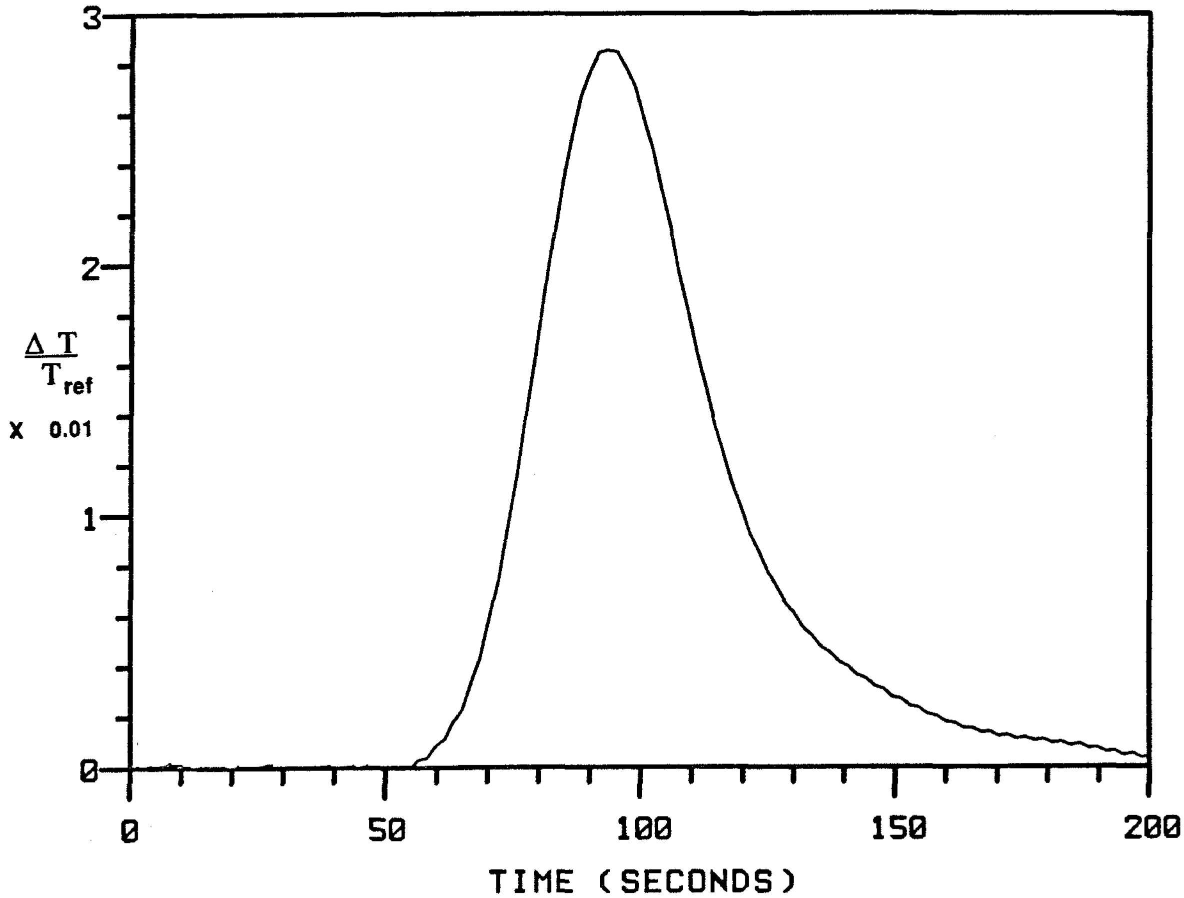


Figure 4

Impulse response for full-order reactor model, Example 1.

was chosen according to formula

$$\hat{\eta} = \prod_{i}^{N} \left(\frac{-\beta_i s+1}{\beta_i s+1}\right) \left(\frac{1}{\lambda s+1}\right) \tag{85}$$

The value $\lambda = 20$ was chosen because it corresponds to three times the closed-loop bandwidth implied by the 60-second deadtime. The feedback controller corresponding to a model according to (83) and $\hat{\eta}$ from (85) is readily obtained from (10).

The final model parameters were obtained after 6 linear program iterations. The validation criteria (represented in Table 1) indicate that both stability and performance requirements have been satisfied.

For comparison purposes, unweighted frequency response matching according to the objective

$$J_{OL} = \left[\frac{1}{\pi}\int_{0}^{\infty} \left|\tilde{p}(i\omega)-\hat{p}(\omega)\right|^2 d\omega\right]^{1/2} \tag{86}$$

and the reduction problem proposed by Enns (Eq. 40) were examined. (86) represents the integral of the squared impulse response energy which traditionally has been considered as an adequate objective for both model reduction and identification. The solution technique used to solve (86) is that proposed by Sanathanan and Koerner (1962), which also falls under the generalized framework presented by Stahl (1984). The solution to (40) was achieved by using the method presented in Section 5, with the weight submatrices of (68) defined according to

$$\mathbf{W}_s^* = \mathrm{diag}\left(\left|\hat{\eta}\right|_{s=j\omega_1}, \ldots, \left|\hat{\eta}\right|_{s=j\omega_i}, \ldots, \left|\hat{\eta}\right|_{s=j\omega_Z}\right) \tag{87}$$

In the process of obtaining a solution, limit cycle problems were encountered. As recommended by Enns, all plants composing the limit set were examined, and the one resulting in the lowest residual was chosen.

Table 1: Summary of Results, Example 1.

Method	Model	$\max_\omega \lvert\hat{\eta} e_m\rvert$	$\max_\omega \lvert w_{p_1}\tilde{\varepsilon}\rvert$	$\max_\omega (\lvert\tilde{\eta}\rvert \ell_m + \lvert w_{p_2}\tilde{\varepsilon}\rvert)$	$\lVert e\rVert_2$ Bound
1	$\frac{0.8142-31.65s+448.63s^2-3929.4s^3+5534s^4}{1+40.38s+1627s^2+13720s^3+0.228x10^6s^4}$	0.762	1.035	1.387	584.4
2	$\frac{1.2244-29.84s+799.4s^2-4337.6s^3+60507s^4}{1+45.63s+1599s^2+11153s^3+0.1895x10^6s^4}$	0.429	0.922	1.274	85.03
3	$\frac{1.2434-47.94s+977.4s^2-6967s^3+68332s^4}{1+59.78s+1654.3s^2+15635s^3+0.1817x10^6s^4}$	0.443	0.841	-	-
4	$\frac{1.0679-38.312s+835s^2-4790s^3+59927s^4}{1+43.53+1687.8s^2+9123s^3+0.1840x10^6s^4}$	0.527	-	1.1650	44.70

Key:
1 Impulse-response matching (open-loop)
2 Enns' stability-preserving weight
3 Control-relevant H_∞, w_{p_1}-weight
4 Control-relevant μ; w_{p_2}-weight

The reduced models, along with the corresponding values for the H_∞ norm, are shown in Table 1; closed-loop responses are shown in Figure 5. The most notable difference is that the response obtained from the control-relevant reduced model rapidly settles to the steady-state, while the others lag significantly longer.

Figures 6 and 7 provide clues towards understanding the benefits of the control-relevant weight. In Figure 6, the open-loop frequency responses of the full and reduced models (from (33) and (86)) are compared. Because of the control-relevant weight, the model obtained from (33) has a good fit at the low frequencies, which are most relevant to the control problem; the model obtained from unweighted error minimization is unsatisfactory because it attempts to match the plant at all frequencies, even those that have no bearing on the control problem. Based on examining the open-loop frequency responses, it would appear as if a higher order reduced model is necessary; a higher-order controller is thus implied. In fact, the H_∞ norm for the control system obtained from (86) exceeds one, requiring that either the model order be increased or the weight specification relaxed.

Greater insights are obtained by examining the multiplicative error (Figure 7), which, as stated previously, represents the adequate error criterion for control-relevant model reduction. The control-relevant model has low error at the important frequency interval. The unweighted problem has significant low-frequency error, thus causing deterioration of the response at final time.

Control-relevant plant reduction of the reactor takes on an added dimension when model uncertainty is considered. Figure 8 shows the Bode plot of the multiplicative uncertainty ℓ_m arising from model

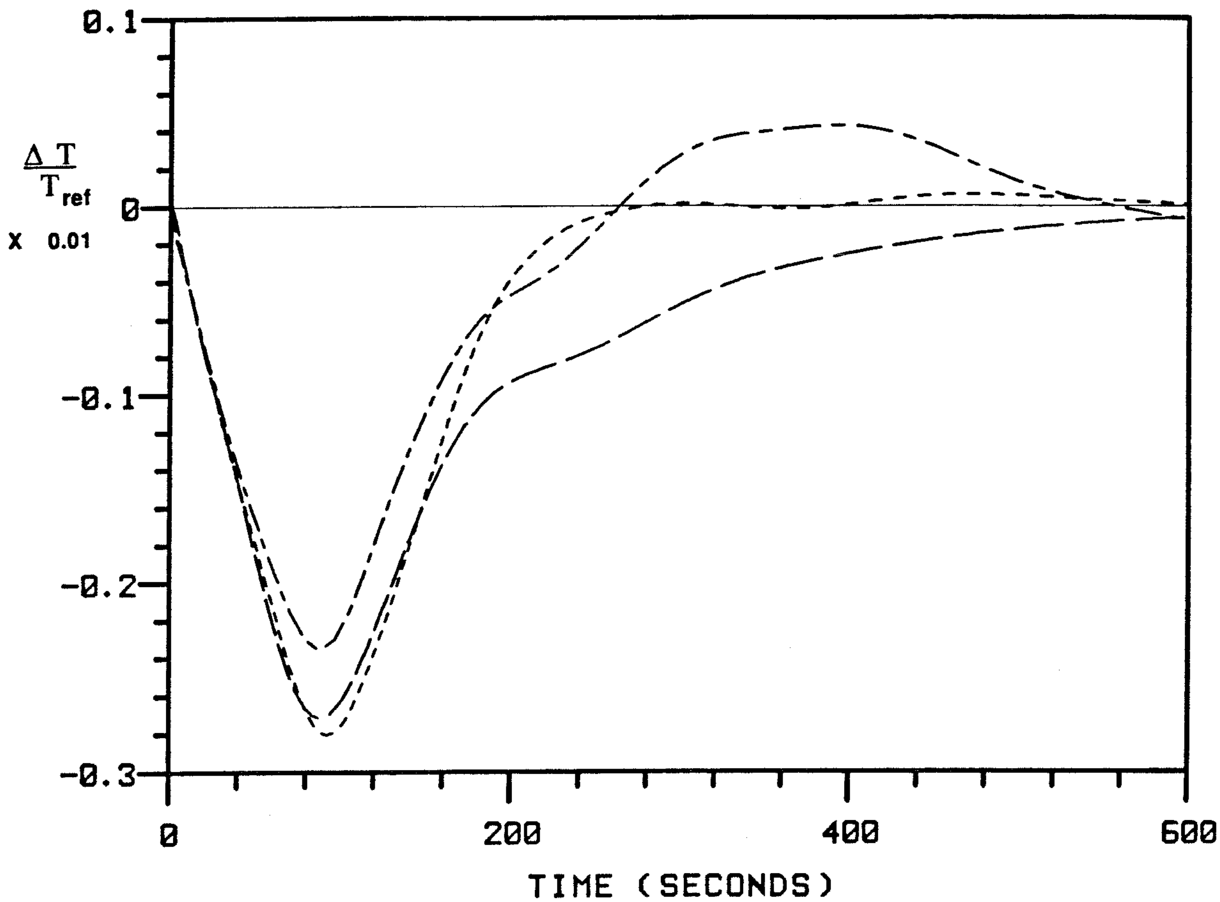


Figure 5

Closed-loop responses for control systems generated from various reduced-order models, Example 1. (— — —) unweighted; (- - - -) H_∞ control-relevant; (— - - —) stability-weighted (Enns').

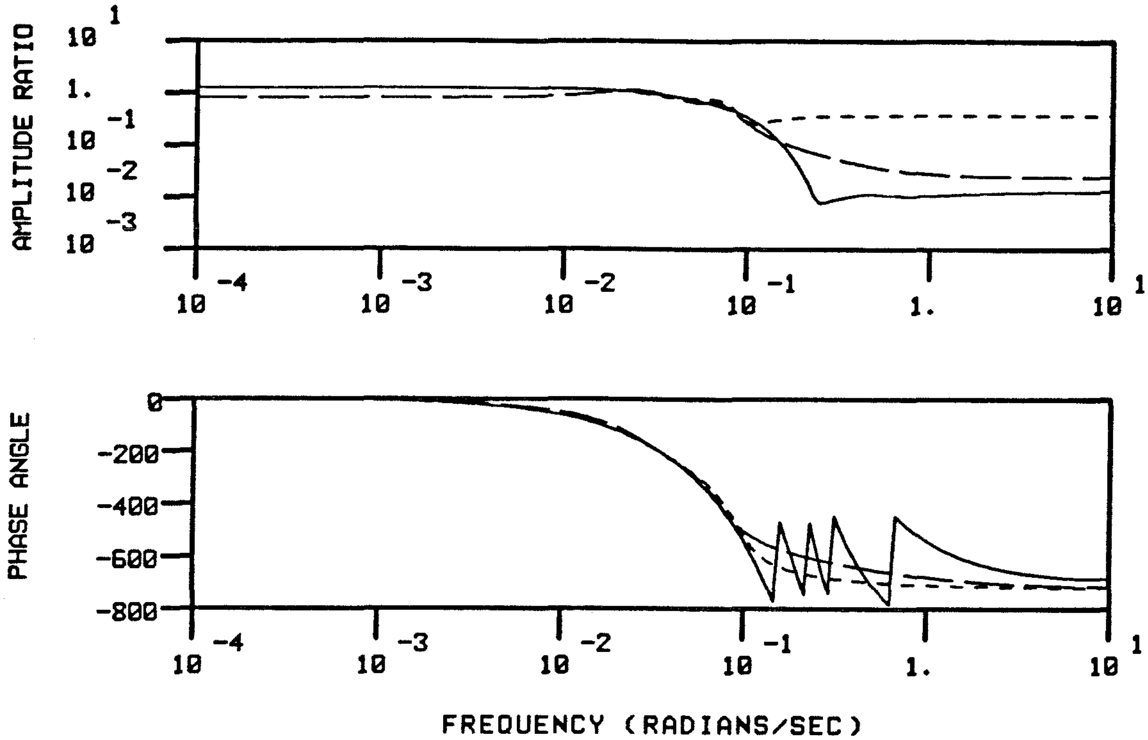


Figure 6

Open-loop Bode plots for selected models from Example 1. (——) full-order model; (– – –) unweighted; (----) H_∞ control-relevant.

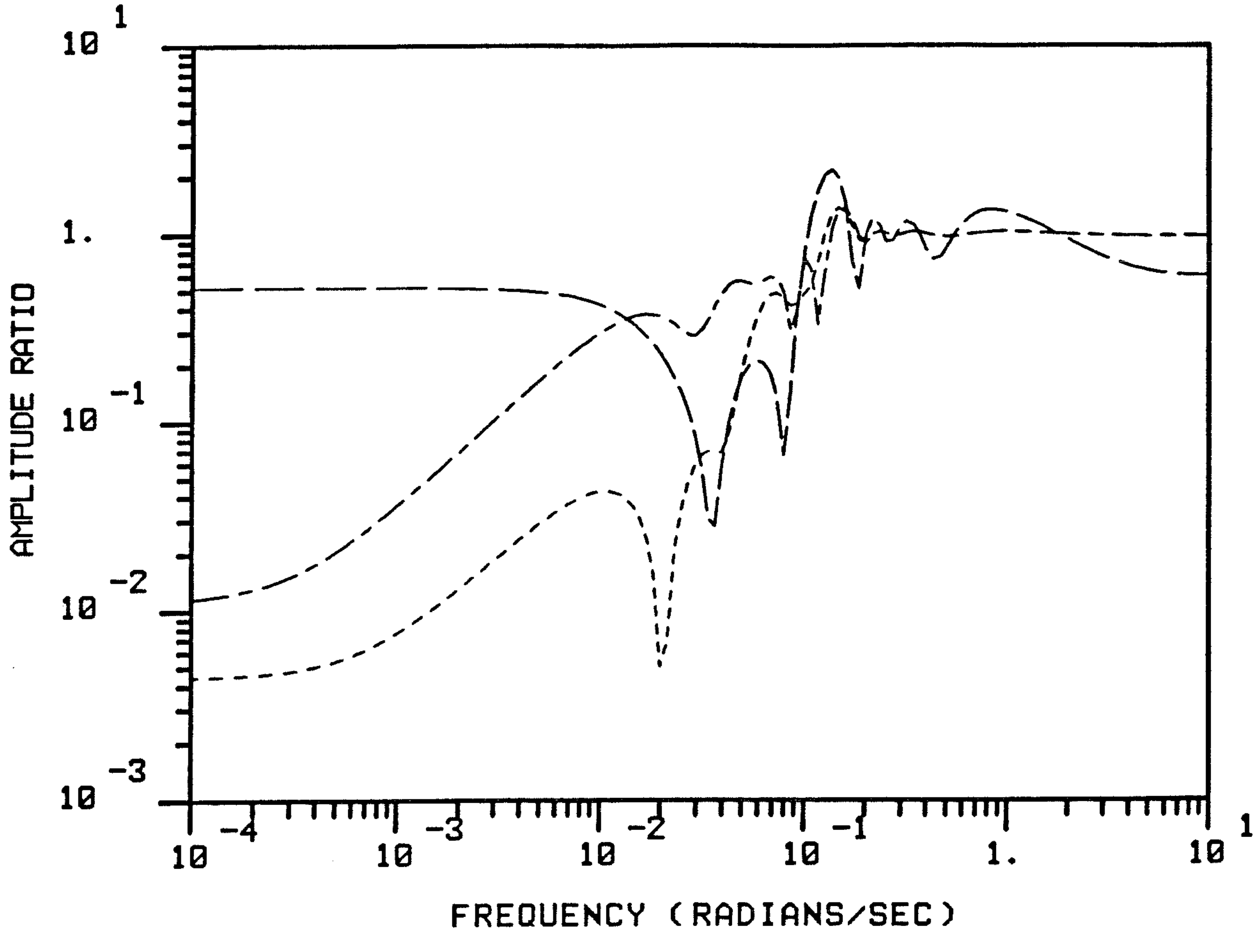


Figure 7

$|e_m|$ for models from Example 1.

(——) unweighted; (--∺-) H_∞ control-relevant; (—--—) stability-weighted (Enns').

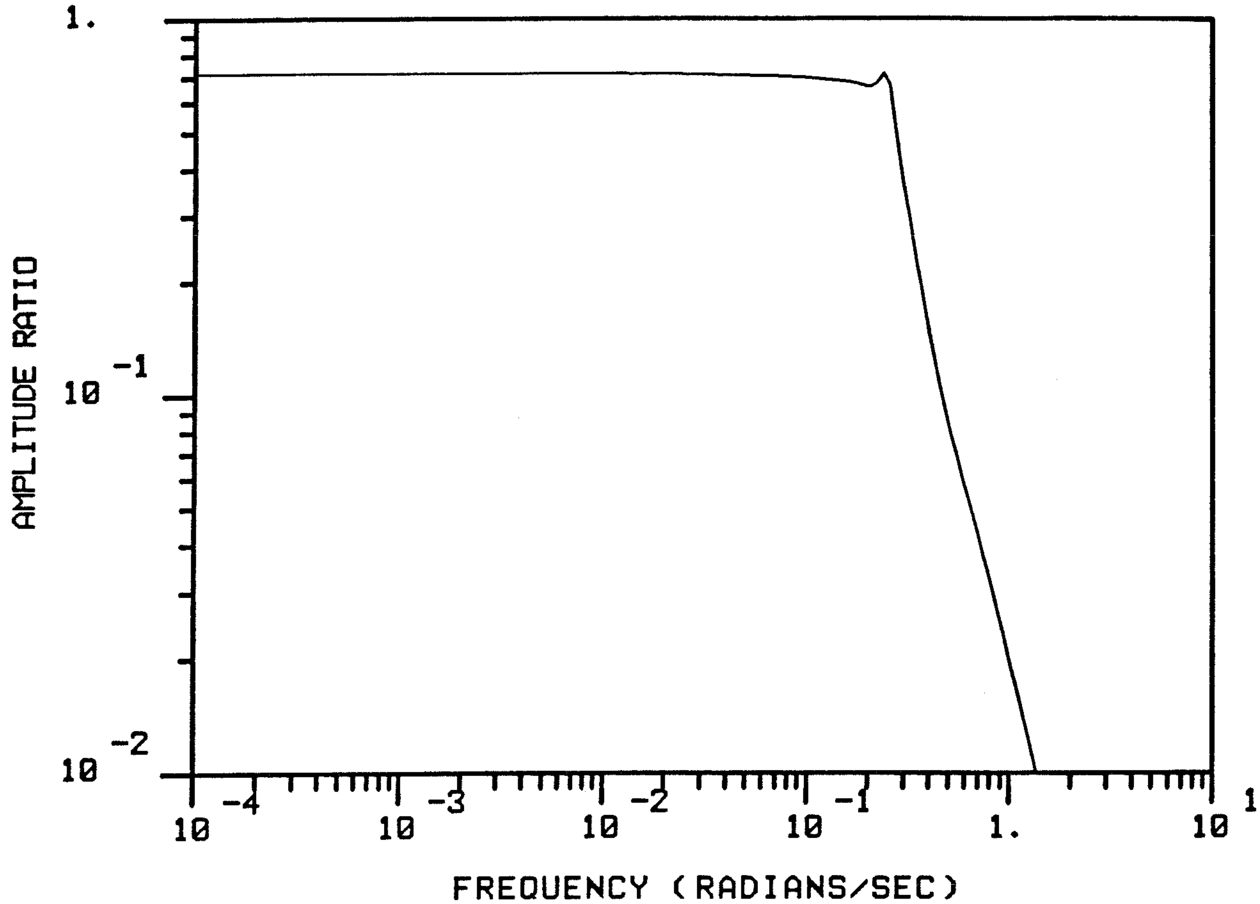


Figure 8

ℓ_m from model nonlinearity, Example 1.

nonlinearity. From the fact that the sensitivity function is bounded for a set of plants according to

$$|\varepsilon| \leq \frac{|\tilde{\varepsilon}|}{1-|\tilde{\eta}|\ell_m} \; ; \; |\tilde{\eta}|\ell_m < 1 \qquad \forall\omega \tag{88}$$

one can redefine w_{p_1} as follows

$$w_{p_2}^{-1} = \frac{\left|\left(\frac{120s}{60s+1}\right)(51s+1)(s+1)\right|}{1-\frac{\ell_m}{|20s+1|}} \tag{89}$$

leading to a new model, as described in Table 1. Although this model results in lower μ than that obtained from the competing methods, one notes that it exceeds one. In order to satisfy Theorem 1, the problem must be re-computed using a higher order for the reduced model or the restrictions posed by w_{p_2} must be relaxed. In our case, we wish to ascertain the performance attainable from the models in Table 1. Substituting (17) into (88) to obtain the formula

$$w_p^{-1} = \frac{|1-\hat{\eta}|}{|1+\hat{\eta}e_m|+|\hat{\eta}(1+e_m)\ell_m|} \tag{90}$$

yields the weight function required to make $\mu = 1$. The two-norm bound for the entire set of plants ℓ_m can be computed according to

$$||e||_2 \leq \left[\frac{1}{\pi}\int_0^{\infty} \left|w_p^{-1}(r-p_d d)\right|^2 d\omega\right]^{1/2} \tag{91}$$

From Table 1, one observes that this bound is significantly reduced by the reduced model resulting from the proposed control-relevant procedure.

<u>Example 2</u>. Controller reduction -- four-disk example.

Enns (1984) presents the example of the control of a four disk system, that consists of "four disks (unity inertia), connected by a flexible wire (unity spring constant) with a motor for applying torques to the third disk and a sensor for measuring angular displacement of the first disk". The model describing this system is

$$\tilde{p} = \frac{(s^2+0.04s+1)(0.207s+1)(0.03132s^2-0.1416s+1)}{4s^2(1.709s^2+0.0523s+1)(0.503s^2+0.0283s+1)(0.292s^2+0.0216s+1)} \quad (92)$$

A full-order compensator designed using LQG loop shaping is

$$\tilde{c} = \frac{0.02916(19.88s+1)(1.709s^2+0.0523s+1)(0.503s^2+0.0283s+1)}{(0.124s^2+0.418s+1)(0.118s^2+0.134s+1)(s^2+0.04s+1)}$$

$$\times \frac{0.292s^2+0.0216s+1}{(10.133s^2+0.693s+1)(1.919s+1)}) \quad (93)$$

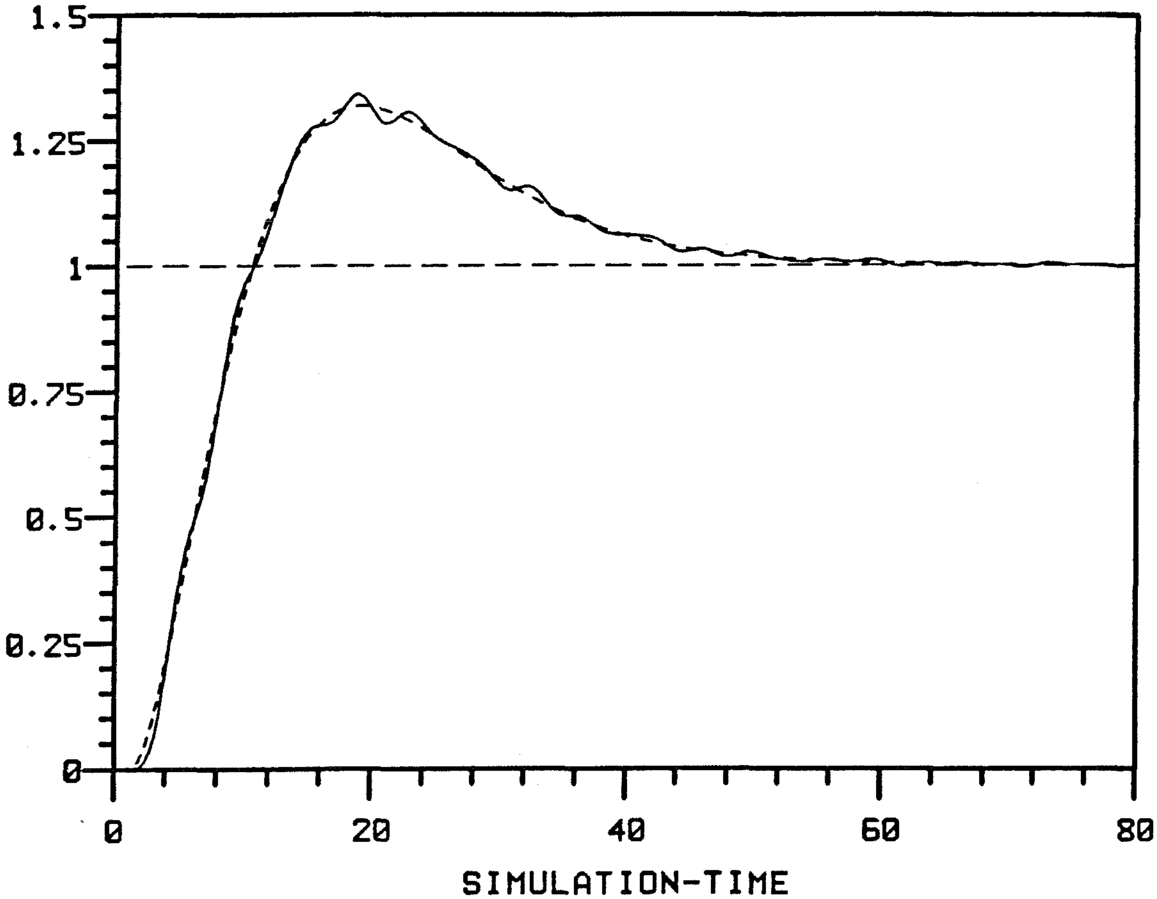
The challenge of this problem is the fact that any reduced-order controller obtained from an internally balanced realization (or any unweighted reduction method, for that matter) results in an unstable closed-loop system. Solving the problem posed by (41), Enns obtains the sixth-order controller

$$\hat{c} = 0.0232 \frac{(19.417s+1)(1.708s^2+0.07843s+1)(0.463s^2+0.0408s+1)}{(s^2+0.04s+1)(3.11s^2+3s+1)(0.0924s^2+0.0121s+1)} \quad (94)$$

Wishing to obtain an even more parsimonious description for this controller, we have used the H_2- optimal controller reduction problem defined by (38) which results in the following 4th order controller (step setpoint changes considered)

$$\hat{c} = \frac{0.0290+0.5764s+0.0723s^2+0.93541s^3}{1+3.1705s+1+3.5229s^2+3.2390s^3+2.3928s^4} \quad (95)$$

Figure 9 compares the closed-loop responses from (93) and (95), indicating that a fourth-order controller obtained from the control-relevant analysis performs adequately. In contrast, attempts to solve for a fourth-order controller using the method of Section 5 and Enns' weight failed to provide a stable controller. Figure 10 compares the additive error ($|\tilde{c}(i\omega) - \hat{c}(i\omega)|$) resulting from models (94) and (95). The added parsimony of the 4th order controller is obtained by the fact that the error is still kept small (less than 0.0001 in magnitude) over



<u>Figure 9</u>

Closed-loop responses, Example 2. (——) H_2 relevant, 4th-order model; (----) full 9th-order controller.

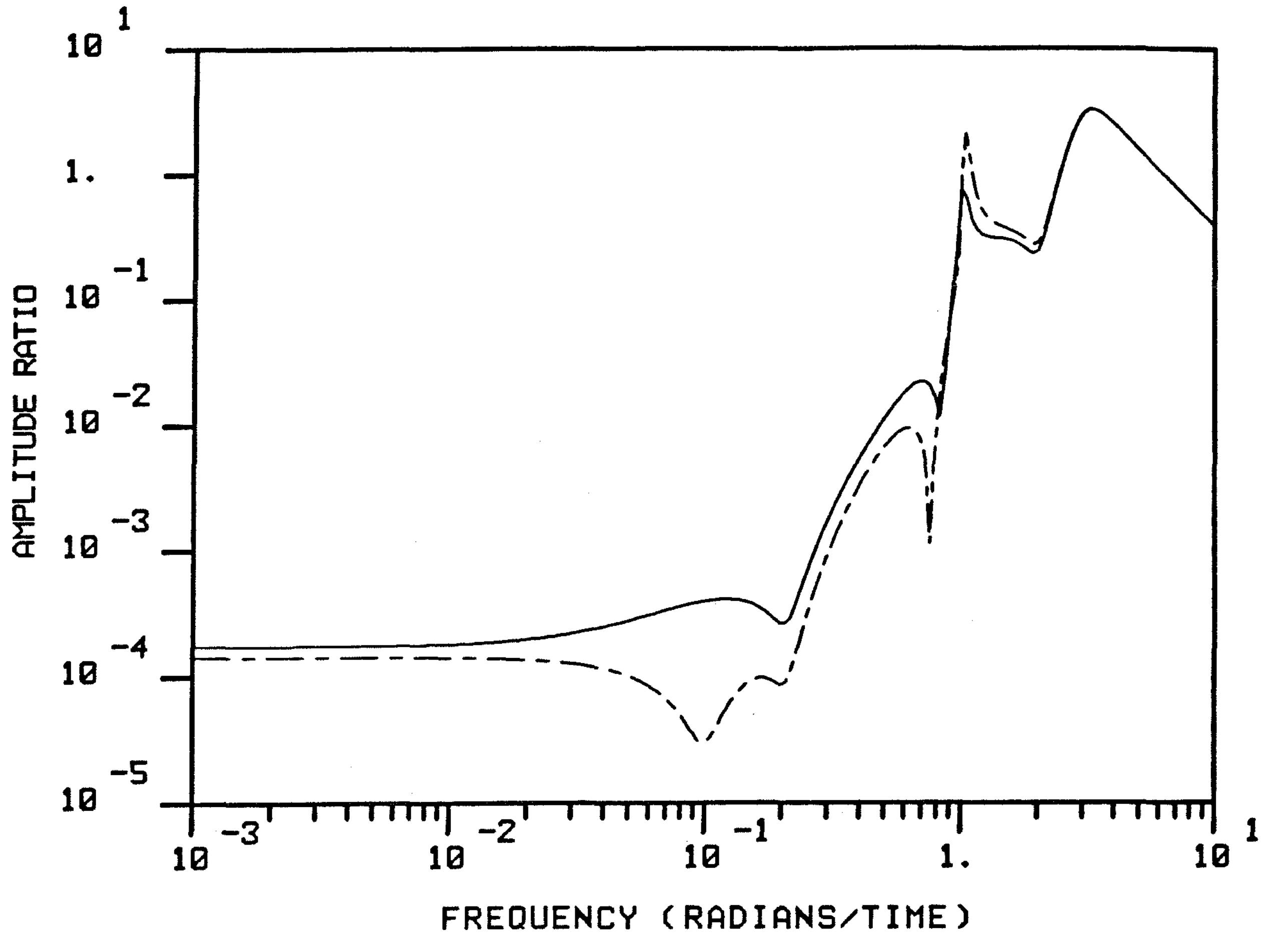


Figure 10

$|e_c|$ for controllers of Example 2. (——) H_2 relevant, 4th-order model; (—-—) stability-weighted 6th-order model (Enns').

the relevant frequency interval.

7. Summary and Conclusions

Despite the vast amount of interest in model reduction during the last two decades, the control and systems literature lacks a clear, simple framework for understanding the effectiveness of reduced models in the control context. The IMC structure provides an answer to this dilemma by relating the controller parameters in a unique, straightforward manner to those of the reduced model. These features lead naturally to using IMC as a control-relevant reduced model assessment procedure. A reduced-order model is readily incorporated into a feedback compensator by a simple algebraic relationship (10); furthermore, the designer can directly specify a desired achievable closed-loop response and bounds $(\hat{\eta}, w_p)$, the setpoint/disturbance characteristics, and the uncertainty description ℓ_m and readily evaluate how the control system performance (H_2, H_∞, or μ) is affected by the mismatch between full and reduced models (e_m).

The IMC parametrization also provides the necessary insight for defining a control-relevant weight for model reduction. This weight in turn allows developing model reduction problems which are near optimal with respect to the closed-loop performance. For the case of plant reduction, these problems require minimizing a weighted multiplicative error; for controller reduction, the weighted additive error between full and reduced-order models is minimized. The presented weights incorporate explicitly all the aforementioned characteristics of the closed-loop control problem. This philosophy contrasts that of most traditional methods for which the error criterion and the weighting functions are selected in an ad hoc manner. Furthermore, the presented problems can be formulated as simple-to-compute programming problems, and can be implemented readily through widely available software. Two examples were presented which attest to the improvements possible by using the control-relevant weights.

The implications of this work extend beyond the problem of model reduction. IMC demonstrates that any feedback compensator can be re-written in terms of a model and a filter; therefore, whenever a high-order plant is controlled by a low-order compensator, a model reduction problem is involved. The presented methods are thus low-order controller design tools as well as model simplification tools. Furthermore, because the full-order plant is represented by a set of frequency dependent real and imaginary parts, the presented methods can serve for parameter estimation in process identification as well; this is the subject of current work (Rivera, 1987).

References

Anderson, B. D. O., "Linear Controller Approximation: A Method with Bounds," IFAC Workshop on Model Errors Concept, Boston (1985).

Bosley and Lees, Automatica, 8, 765 (1972).

Decoster, M. and A. R. van Cauwenberghe, Journal A., 17, 68 (1976a); 125 (1976b).

Doyle, J., Notes from the Honeywell/ONR Workshop on Advances in Multivariable Control, Minneapolis (1984).

Doyle, J. "Structured Uncertainty in Control System Design," Proceedings of the 1985 Control and Decision Conference, Ft. Lauderdale, FL.

Enns, D., "Model Reduction for Control System Design," Ph.D. Thesis, Stanford University (1984).

Garcia, C. and M. Morari, I&EC Proc. Des. Dev., 21, 308 (1982).

Hyland, D. C. and D. S. Bernstein, Proc. 23rd Conference on Decision and Control, Las Vegas (1984).

Mitra, D., Proc. IEE, 116, 1439 (1969).

Mandler, J. A., M. Morari and J. H. Seinfeld, Chemical Engineering Science, 41, 1577 (1986).

Morari, M., E. Zafiriou and C. E. Economou, Robust Process Control, Springer-Verlag, to appear (1987).

Noldus, E. and M. Decoster, Regelungstechnik, 24, 122 (1976).

Rivera, D. E., "Process Control Using Low-Order Models," M.S. Thesis, University of Wisconsin (1984).

Rivera, D. E., "Modeling Requirements for Process Control," Ph.D. Thesis, California Institute of Technology (1987).

Sanathanan, C. K. and J. Koerner, IEEE Trans. Auto. Control, AC-8, 56 (1963).

Skelton, R. E. and D. H. Owens, IFAC Newsletter, May 1986.

Stahl, H., Int. J. Control, 39, 541 (1984).

Zames, G., IEEE Trans. Auto. Control, AC-26, 301 (1981).

CHAPTER IV:

PLANT AND CONTROLLER REDUCTION PROBLEMS FOR ROBUST PERFORMANCE

(Submitted for publication to the

IEEE Trans. Autom. Cntrl., November, 1986)

PLANT AND CONTROLLER REDUCTION PROBLEMS FOR ROBUST PERFORMANCE

Daniel E. Rivera

Manfred Morari

Chemical Engineering, 206-41

California Institute of Technology

Pasadena, California 91125

Abstract

Model reduction problems which consider preserving closed-loop performance in the presence of uncertainty are presented. These are formulated as weighted multiplicative error problems (for plant reduction) and weighted additive error problems (for controller reduction), with the weight function incorporating explicitly such control information as the desired sensitivity operator bound, the setpoint/disturbance spectrum, and the process uncertainties. The solution to these problems using currently available methodologies, such as frequency-weighted balanced realization, is considered. Finally, the benefit of the proposed problems is illustrated with examples taken from the control of a binary distillation column.

1. Introduction

The problem of model reduction is of significant practical importance in control system design, and has been a subject of continuing study since the early 60's. The vast number of references cited by Genesio and Milanese (1976) attests to the fervent activity conducted on this problem.

Sorting through the multitude of suggested approaches, one can identify two recent significant contributions: the first is the simplified computation of the reduction problem by means of balanced realizations and Hankel-norm approximations (Moore, 1981; Glover, 1984) while the second is the development of control-relevant weight functions for plant and controller reduction which consider preserving nominal stability of the control system in spite of reduction error (Enns, 1984; Anderson, 1985). Given the importance of the field of robust control to the design of practical controllers, the focus of this paper is on developing weighted reduction problems which maintain desired levels of closed-loop control performance in the presence of both plant uncertainty and reduction error. The basis for our analysis is the structured singular value μ (Doyle, 1982) which allows the assessment of robust stability and performance in the presence of structured uncertainty (Doyle, 1985 a&b). The computational aspect of the proposed problems is addressed, particularly the use of the frequency-weighted balanced realization algorithm described by Enns (1984).

2. Theoretical Development

The methodology presented applies to linear, time-invariant, and stable multivariable plants subject to norm-bounded perturbations, such as those described by Doyle, Wall, and Stein (1982) and summarized in Table 1. Any combination of these perturbations in the classical feedback structure can in turn be represented as a block-diagonal perturbation problem according to the $G\Delta$ interconnection structure (Figure 1) for which the μ analysis theorem applies:

Theorem 1 (Doyle, 1982): Robust Stability, Structured.

The generalized plant G is stable for all perturbations described by the set

$$X_\infty = \bigcup_{j=1}^{\infty} X_j \tag{1}$$

$$X_\delta = \mathrm{diag}(\overbrace{\Delta_1,\Delta_1,\ldots,\Delta_1}^{m_1},\ \overbrace{\Delta_2,\Delta_2,\ldots,\Delta_2}^{m_2},\Delta_3,\Delta_3,\ldots,\Delta_{n-1},\ \overbrace{\Delta_n,\Delta_n,\ldots,\Delta_n}^{m_n})$$

$$\Delta_j \in C^{k_j \times k_j} \quad \bar{\sigma}(\Delta_j) \leq \delta,\ \delta \in [0,\infty)\ ,\ \textit{for each } j = 1,2,\ldots,n$$

iff

$$||G||_\mu \leq 1 \tag{2}$$

where

$$||G||_\mu \triangleq \sup_\omega \mu\,[G(j\omega)] \tag{3}$$

with μ defined as

$$\frac{1}{\mu(M)} = \min_{\Delta\varepsilon X_\infty} \{\bar{\sigma}(\Delta)\,|\det(I-M\Delta) = 0\}\ ,\ M = G(j\omega) \tag{4}$$

The results of Theorem 1 must be qualified. First of all, (4) is not a convenient formula for computing μ, as the implied optimization problem is cumbersome and not unimodal. An upper bound for μ which has significantly better computational properties is (Doyle, 1982)

$$||G||_\mu \leq ||DMD^{-1}||_\infty\ ,\ ||\cdot||_\infty \triangleq \sup_\omega \bar{\sigma}(\cdot) \tag{5}$$

where D is the set of real diagonal matrices defined as

$$\mathcal{D} = \{\mathrm{diag}(d_1 I_{k_1}, d_2 I_{k_1},\ldots,d_{m_1} I_{k_1}, d_{m_1+1} I_{k_2},\ldots,d_m I_{k_n})\,|\,d_i \varepsilon R^+ = (0,\infty)\} \tag{6}$$

Description	Representative Types of Uncertainty Characterized	Representative Types of Performance Specs
$P' = (I+\Delta)P$	- output (sensor) error - neglected HF dynamics - changing numbers of rhp zeros	- sensor noise attenuation - output response to output commands
$P' = P(I+\Delta)$	- input (acuator) errors - neglected HF dynamics - changing numbers of rhp zeros	- input response to input commands
$P' = P + \Delta$	- additive plant errors - uncertain rhp zeros	- input response to output commands
$P' = (I+\Delta)^{-1}P$	- LF plant parameter errors - changing numbers of rhp poles	- output sensitivity - output errors to output commands and disturbances
$P' = P(I+\Delta)^{-1}$	- LF plant parameter errors - changing numbers of rhp poles	- input sensitivity - input errors to input commands and disturbances
$P' = (P^{-1}+\Delta)^{-1}$	- uncertain rhp poles	- output errors to input commands and disturbances

Table 1. Representative robustness/performance conditions.

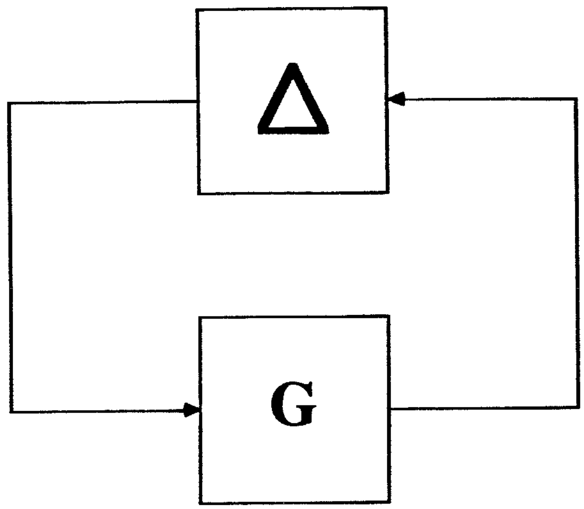


Figure 1. G–Δ interconnection structure.

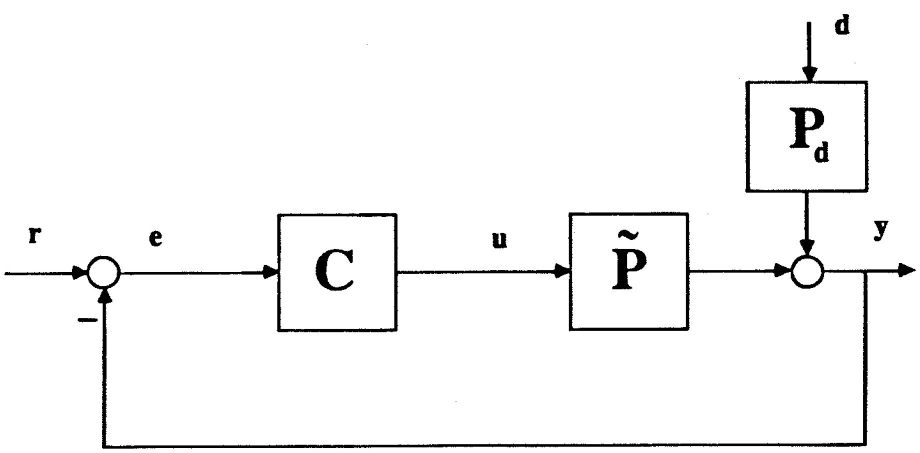


Figure 2. Classical Feedback Structure.

$$m = \sum_{j=1}^{n} m_j \qquad D \rightarrow \inf_{D \varepsilon \mathcal{D}} \bar{\sigma}(DMD^{-1})$$

(5) is an exact bound when Δ has three or fewer blocks.

Secondly, it must be understood that Theorem 1 also incorporates the robust performance problem. The performance conditions are incorporated in the $G\Delta$ structure in the form of additional perturbations (Doyle, 1985 a&b). The $G\Delta$ structure thus allows for the performance and uncertainty aspects of the feedback problem to be captured in a unified manner.

We now consider the full-order model $\tilde{P}$ subject to control using the compensator C, as illustrated in Figure 2. The output sensitivity/complementary sensitivity operators are, respectively,

$$\tilde{S} = (I+\tilde{P}C)^{-1} \tag{7}$$

$$\tilde{H} = \tilde{P}C(I+\tilde{P}C)^{-1} = I - \tilde{S} \tag{8}$$

(7) and (8) represent the nominal performance characteristics of the control system.

The first step towards developing the control-relevant reduction problems is to perform an affine parametrization of M in terms of $\tilde{S}$

$$M = N_{11} + N_{12}\tilde{S}N_{21} \tag{9}$$

N_{11}, N_{12}, and N_{21} are independent of the feedback controller C. For the feedback structure subject to the perturbations listed in Table 1, obtaining the parametrization (9) can be done by inspection without need for detailed computation, as shown by Skogestad and Morari (1986c). The link with the reduction problem is made by relating the controller designed on the basis of the reduced-model $\hat{P}$ to the nominal performance represented by $\tilde{S}$, and consequently, as indicated by (9) and Theorem 1, to the closed-loop robust performance. Representing the full-order model in terms of the reduced-order model as

$$\tilde{P} = (I+E_m)\hat{P} \tag{10}$$

and substituting (10) into (7) and (8) results in the following expressions for the complementary sensitivity/sensitivity operators:

$$\tilde{S} = (I-\hat{H})(I+E_m\hat{H})^{-1} \tag{11}$$

$$\tilde{H} = (I+E_m)\hat{H}(I+E_m\hat{H})^{-1} \tag{12}$$

E_m denotes the multiplicative error between the full and reduced-order models

$$E_m = (\tilde{P}-\hat{P})\hat{P}^{-1} \tag{13}$$

and is the suitable error measure for control-relevant plant reduction. Its presence in (11) and (12) indicates the degradation occurring in the closed-loop system as a result of full/reduced model mismatch.

Substituting (11) in expression (9), one obtains

$$M = N_{11} + N_{12}(I-\hat{H})(I+E_m\hat{H})^{-1}N_{21} \tag{14}$$

M is now expressed in terms of the reduction error E_m, and substituting this into (5), a control-relevant reduction problem is obtained. Solving the associated optimization problem, however, remains a complex task. Further analysis is thus needed to simplify the problem into a form amenable for computation.

One can show from (11) and (12) that

$$\bar{\sigma}(E_m\hat{H}) < 1 \qquad \forall\omega \tag{15}$$

is a sufficient condition for nominal stability of the reduced-order control system if $\hat{P}$ and C lead to an internally stable control system. Assuming (15) holds, (14) can be expanded into a Neuman series according to:

$$M = N_{11} + N_{12}(I-\hat{H})(I-E_m\hat{H}+(E_m\hat{H})^2-(E_m\hat{H})^3+\ldots)N_{21} \tag{16}$$

The matrices N_{21}, $(I-\hat{H})$, and N_{21} define a bandwidth, or "control-relevant frequency interval", over which the reduction should be performed. Assuming that $\bar{\sigma}(E_m\hat{H}) << 1$ over this frequency interval, it is then safe to neglect the higher-order terms to obtain an approximate M

$$M \cong M_{approx} = N_{11} + N_{12}(I-\hat{H})N_{21} - N_{12}(I-\hat{H})E_m\hat{H}N_{21} \tag{17}$$

$$M_{approx} = \hat{M} + M_{error} \tag{18}$$

The benefit of an expression of the form (18) is that the problem is now decomposed into a linear combination of the M obtained on the basis of the reduced model $\hat{P}$ and an error/degradation term, consequence of reduction. Substituting (18) into the simplified formula for computing μ stated in (5), and using the properties of function norms one observes that

$$||D\hat{M}D^{-1}||_\infty - ||DM_{error}D^{-1}||_\infty \leq ||DM_{approx}D^{-1}||_\infty \leq ||D\hat{M}D^{-1}||_\infty + ||DM_{error}D^{-1}||_\infty \tag{19}$$

One can then formulate a control-relevant plant reduction problem of the desired form by minimizing the term that incorporates the degrading effects of the multiplicative error E_m

$$\inf_{\hat{P}} \quad ||DN_{12}(I-\hat{H})E_m\hat{H}N_{21}D^{-1}||_\infty \tag{20}$$

(20) is a frequency-weighted problem whose weight functions incorporate explicitly control system design information; this represents a refreshing change from previous efforts which suggest weights based on ad hoc considerations.

The previous analysis extends itself in a straightforward manner for the case of controller reduction. This is achieved by first re-writing (11) as

$$(I+\tilde{P}\hat{C})^{-1} = (I-(I-\tilde{H})\tilde{P}(\tilde{C}-\hat{C}))^{-1}(I-\tilde{H}) \tag{21}$$

As before, requiring that the Small Gain Theorem holds

$$||(I-\tilde{H})\tilde{P}(\tilde{C}-\hat{C})||_\infty \leq 1 \tag{22}$$

leads to a Neuman series, which, truncated after the first term

$$M \tilde{=} M_{approx} = N_{11} + N_{12}(I+(I-\tilde{H})\tilde{P}(\tilde{C}-\hat{C}))(I-\tilde{H})N_{21} \tag{23}$$

produces the controller reduction problem

$$\inf_{\hat{C}} \; ||DN_{12}(I-\tilde{H})\tilde{P}(\tilde{C}-\hat{C})(I-\tilde{H})N_{21}D^{-1}||_\infty \tag{24}$$

Example 1:

For purposes of example consider the control problem represented in Fig. 3. This is a robust performance problem, in which the performance specifications are represented in terms of weights on the output signal, and the plant model is subject to output, input, and additive uncertainties. The form of Δ and G are

$$\Delta = \mathrm{diag}(\Delta_I,\Delta_A,\Delta_O,\Delta_P) \tag{25}$$

$$G = W_1 G^I W_2 \tag{26}$$

where

$$W_1 = \mathrm{diag}(W_{1I},W_{1A},W_{1O},W_{1P}) \tag{27}$$

$$W_2 = \mathrm{diag}(W_{2I},W_{2A},W_{2O},W_{2P}) \tag{28}$$

$$G^I = \begin{bmatrix} -\tilde{P}^{-1}\tilde{H}\tilde{P} & -\tilde{P}^{-1}\tilde{H} & -\tilde{P}^{-1}\tilde{H} & -\tilde{P}^{-1}\tilde{H} \\ \tilde{P}^{-1}\tilde{S}\tilde{P} & -\tilde{P}^{-1}\tilde{H} & -\tilde{P}^{-1}\tilde{H} & -\tilde{P}^{-1}\tilde{H} \\ \tilde{S}\tilde{P} & \tilde{S} & -\tilde{H} & -\tilde{H} \\ \tilde{S}\tilde{P} & \tilde{S} & \tilde{S} & \tilde{S} \end{bmatrix} \tag{29}$$

W_1 and W_2 are frequency-dependent functions which are used to determine the specific character of the perturbation deltas. The parametrization (9) of M in terms of $\tilde{S}$ is

$$N_{11} = W_1 \begin{bmatrix} -I & -\tilde{P}^{-1} & -\tilde{P}^{-1} & -\tilde{P}^{-1} \\ 0 & -\tilde{P}^{-1} & -\tilde{P}^{-1} & -\tilde{P}^{-1} \\ 0 & 0 & -I & -I \\ 0 & 0 & 0 & 0 \end{bmatrix} W_2 \tag{30}$$

$$N_{12}^T = (\tilde{P}^{-1}\tilde{P}^{-1} \; I \; I)W_1^T \tag{31}$$

$$N_{21} = (\tilde{P} \; I \; I \; I)W_2 \tag{32}$$

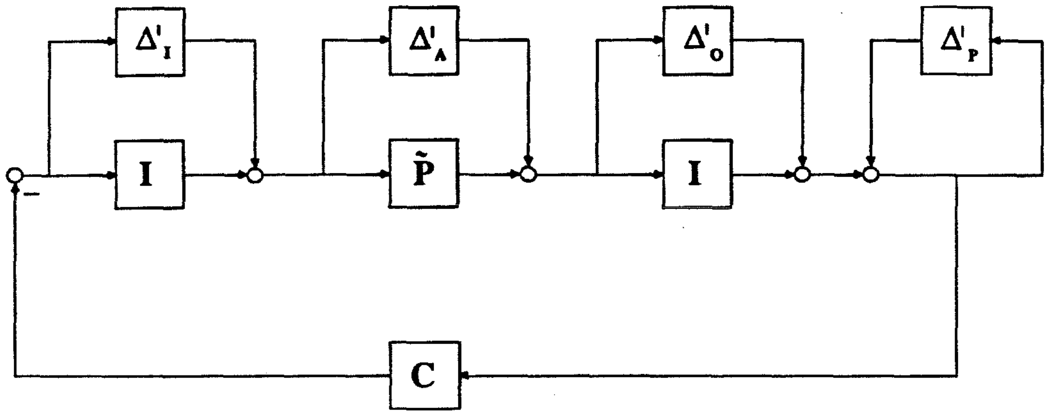


Figure 3. Feedback structure, Example 1.

$$\Delta'_I = W_{2I}\Delta_I W_{1I}$$

$$\Delta'_A = W_{2A}\Delta_A W_{1A}$$

$$\Delta'_O = W_{2O}\Delta_O W_{1O}$$

$$\Delta'_P = W_{2P}\Delta_P W_{1P}$$

The control-relevant plant reduction problem is therefore

$$\inf_{\hat{P}} \quad \left|\left| DW_1 \begin{bmatrix} \tilde{P}^{-1} \\ \tilde{P}^{-1} \\ I \\ I \end{bmatrix} (I-\hat{H})E_m\hat{H}(\tilde{P} \;\; I \;\; I \;\; I)W_2D^{-1} \right|\right|_\infty \tag{33}$$

while the equivalent controller reduction problem is

$$\inf_{\hat{C}} \quad \left|\left| DW_1 \begin{bmatrix} \tilde{P}^{-1} \\ \tilde{P}^{-1} \\ I \\ I \end{bmatrix} (I-\tilde{H})\tilde{P}(\tilde{C}-\hat{C})(I-\tilde{H})(\tilde{P} \;\; I \;\; I \;\; I)W_2D^{-1} \right|\right|_\infty \tag{34}$$

At this point it is worthwhile to compare problems (20) and (24) with those proposed by Enns (1984) and Anderson (1985). As stated previously, these investigators consider maintaining nominal stability of the closed-loop system despite reduction error; this is achieved by insuring that conditions (15) and (22) are met:

Plant reduction:

$$\inf_{\hat{P}} ||E_m\hat{H}||_\infty \tag{35}$$

Controller reduction:

$$\inf_{\hat{C}} \quad ||(I-\tilde{H})\tilde{P}(\tilde{C}-\hat{C})||_\infty \tag{36}$$

One notices that problems (20) and (24) incorporate problems (35)-(36) with some extra factors. This can be readily explained by recalling the derivation of the performance-relevant reduction problems. While problems (35) and (36) require that the Small Gain condition be kept uniformly small over all frequency, the performance-relevant problem only demands that the Small Gain condition be satisfied (i.e., be less than one) with the performance weights defining the bandwidth over which the condition should be minimized. Clearly, the performance problem is superior, as over this relevant bandwidth an even lower-order model (compared to that obtained from (35), (36), or any of the wide variety of unweighted reduction procedures) might be suitable, resulting in added parsimony without compromising the closed-loop properties of the system.

A special note regarding the H_2-optimal control problem:

Convenient control-relevant reduction problems can also be formulated for the H_2-optimal control problem, defined by minimizing the objective:

$$J = ||W_o\tilde{S}W_i||_2 \triangleq \left[\frac{1}{\pi}\int_0^\infty tr[(W_o\tilde{S}W_i)^*(W_o\tilde{S}W_i)]d\omega\right]^{1/2} \qquad (37)$$

W_o and W_i represent matrix valued frequency-dependent weights on the sensitivity operator; * denotes the complex conjugate transpose. Assuming, as done previously, that (15) and (22) hold and are much less than 1 over the relevant frequency bandwidth, the derivation of control-relevant plant and controller problems for the H_2 objective follows through exactly as before, resulting in the following problem definitions:

Plant reduction:

$$\inf_{\hat{P}} \; ||W_o(I-\hat{H})E_m\hat{H}W_i||_2 \qquad (38)$$

Controller reduction:

$$\inf_{\hat{C}} \; ||W_o(I-\tilde{H})\tilde{P}(\tilde{C}-\hat{C})(I-\tilde{H})W_i||_2 \qquad (39)$$

As with the μ-synthesis problems (20) and (24), (38) and (39) incorporate explicitly all information required for the closed-loop design.

3. Computational Issues

The subject of frequency-weighted model reduction computation remains one of the foremost research problems in the field of reduced-order modeling. While the focus of our study centers on formulating control-relevant problems, we wish to indicate how some currently available techniques can be utilized to solve the problems presented in Section 2.

Moore (1981) first proposed the method of internally balanced realizations, which, despite being suboptimal with respect to the H_2 and H_∞ objectives, is computationally convenient and provides information (second-order modes) which are readily translated into useful error bounds. Enns (1984) has extended the method of balanced realizations to include input and output weights

$$\inf_{\text{model}} \quad \| W_{bro} E_a W_{bri} \|_{BR} \tag{40}$$

$$E_a = \tilde{P} - \hat{P} \text{ (plant reduction); } \tilde{C} - \hat{C} \text{ (controller reduction)}$$

Difficulties with (40) arise from the fact that, unlike the internally (unweighted) balanced realization problem, the reduced models obtained may not be stable if both input and output weights are non-unity. Furthermore, no simple error bounds based on the "second-order modes" of the balanced grammians have been proposed. The computational intensity of the problem also increases with increasing order of the weight functions.

An alternate computational technique is obtained by representing the plant $\tilde{P}$ in terms of its frequency response, using the properties of equivalent norms (Stone, 1962), such as

$$n^{-3/2} \sum_{i,j} |a_{ij}| \leq \bar{\sigma}(A) \leq \sum_{i,j} |a_{ij}| \text{ ; } n = \text{order of } A \tag{41}$$

to rewrite the control-relevant reduction problems in terms of the programming problem (Stahl, 1984)

$$\inf_{\mathbf{p}} \quad ||\mathbf{W}(\mathbf{M}\mathbf{p}-\mathbf{v})||_\alpha \tag{42}$$

$||\cdot||_\alpha$ - α-vector norm; $\alpha = 2, \infty$

p represents a vector of reduced-model parameters, while **W, M,** and **v** incorporate the control-relevant weight and plant information. A detailed description of the development of problem (42) for the scalar case is provided by Rivera and Morari (1986); for multivariable systems, however, a number of restrictions apply, such as maintaining the same poles for all elements in the reduced model transfer function matrix.

Several considerations must be addressed prior to solving problems (20), (24), (38) and (39) with the methods denoted by (40) and (42). First we examine the inherent difficulties associated with the plant reduction problem. Because of the lack of previous knowledge on the reduced model $\hat{P}$, $\hat{H}$, and hence the control-relevant weights, cannot be fully specified a priori. Using $\hat{H}$ based on the full-model is out of the question, as that essentially involves computing the full-order controller. To complicate matters, the multiplicative error E_m, rather than the additive error E_a (the error criterion used in (40)) must be minimized, for which few efficient methods exist.

Enns (1984) recognizes these difficulties, and suggests that the problem be formulated as a fixed-point problem and the method of successive approximation used to obtain a solution. The reduced model obtained at each iteration is used to update the weight; assuming existence of a solution and that the functional of the weighted error is a contraction, the method will converge. Applying these recommendations to our problem results in the following procedure:

1. For the initial iteration step i = 1, specify $\hat{H}$ as $\hat{H}$ = F, where F

is a matrix of low-pass filter elements that determine the bandwidth of the nominal reduced-order closed-loop response.

2. Set $\hat{P}_1 = \tilde{P}$, and compute μ based on (5) with the D scales obtained through Osborne's method (1960). (This step is not necessary for the H_2 plant reduction problem (38)).

3. Solve problem (20) or (38) either using the frequency-weighted balanced realization technique to obtain the model $\hat{P}_{i+1}$

$$\min_{\hat{P}_{i+1}} \quad \| W_{bro} E_a W_{bri} \|_{BR} \tag{43}$$

$$\text{H-2 problem:} \quad W_{bro} = W_o(I-\hat{H}) \qquad W_{bri} = \hat{P}_i^{-1}\,\hat{H}W_i$$

$$\mu \text{ problem:} \quad W_{bro} = DN_{12}(I-\hat{H}) \qquad W_{bri} = \hat{P}_i\hat{H}N_{21}D^{-1}$$

or the iterative programming problem denoted by (42).

4. Recompute $\hat{H}$ to take into account any non-minimum phase elements in $\hat{P}$. The factor $\hat{P}_+$, which contains all nonminimum phase behavior, can be obtained as described by Zafiriou and Morari (1985), or through spectral or inner/outer co-prime factorizations (Anderson, 1967; Enns, 1984). The factor $\tilde{P}_+$ is used to update the function $\hat{H}$ according to

$$\hat{H} = \hat{P}_+F \tag{44}$$

5. Obtain E_m according to (13). From E_m update $\tilde{H}$ (12); re-compute μ and D-scales. If

$$|\mu_{i+1}-u_i| \leq \text{TOL} \tag{45}$$

or

$$|J_{i+1}-J_i| \leq \text{TOL} \tag{46}$$

holds, where TOL is a user-specified tolerance, terminate the procedure. Otherwise, increment i by 1 and return to step 3.

Following successful completion of the reduction procedure, one must verify that 1) the reduced model does not contain RHP poles and 2) the Small Gain condition (15) has been satisfied. Failure to satisfy these criteria usually implies that control specifications need to be relaxed, or the reduced plant order needs to be increased.

Solving the controller reduction problem, on the other hand, is significantly less complicated (in making this statement we ignore the amount of effort that might have been required to obtain the full-order controller). $\hat{H}$ is know a priori, the D scales need only be computed once, and the mismatch between full and reduced controllers is adequately expressed as an additive error. The weight function is thus completely well-defined at the beginning of the problem, eliminating the need for any successive substitution.

The benefits of the proposed problems are further illustrated through case studies.

4. Case Studies

Case Study 1. H_2-relevant plant reduction.

We consider the H_2-optimal control of a 40-tray, high-purity binary distillation column, as described by Skogestad and Morari (1986 a&b). Of interest is the regulation of top and bottom composition (y_D and x_B, respectively) using the reflux (L) and boilup (V) rates. The resulting linearized plant model presented by Skogestad and Morari is a 41-order model,

$$\begin{bmatrix} dy_D \\ dx_B \end{bmatrix} = G_{LV} \begin{bmatrix} dL \\ dV \end{bmatrix} \tag{47}$$

with the distinguishing feature that G_{LV} is ill-conditioned (ratio of maximum to minimum singular values = 141); this is a consequence of the high composition sensitivity to product flowrates.

Two-state models using internally, stability-weighted (35), and performance-weighted (38) balanced realizations were obtained. Initially, F was set to a bandwidth of 0.05 rad/sec

$$F = \frac{1}{20s+1} I_2 \tag{48}$$

The step setpoint changes considered were represented in W_i as

$$W_i = \begin{bmatrix} 1.0/s & 0 \\ 0 & -0.5/s \end{bmatrix} \tag{49}$$

The output weight (W_O) was set to identity. For an identity W_O and the setpoint change indicated by (49), the form of the H_2-optimal factorization $\hat{P}_+$ needed to update $\hat{H}$ is indicated by Morari et al. (1987) as

$$\hat{P}_+ = \hat{P}_A \hat{P}_A^{-1}(0) \tag{50}$$

$\hat{P}_A(s)$ is an all-pass matrix obtained directly from spectral or inner-outer factorization.

The first benefit of using the performance-relevant weights was noted while conducting the reduction procedure. The stability-weighted problem required six iterations before the objective function value varied by less that 1 percent, while the performance-weighted problem took only three iterations to reach the same level of convergence. This represents tremendous time savings, as each iteration involves the balancing of the full-order model with weights, an extremely time consuming procedure. Having computed the reduced-order models, the resulting closed-loop responses can be obtained from (12), (44), and (50).

Figures 4a&b compare the closed-loop responses obtained from controllers designed on the basis of the Q-parametrization/Internal

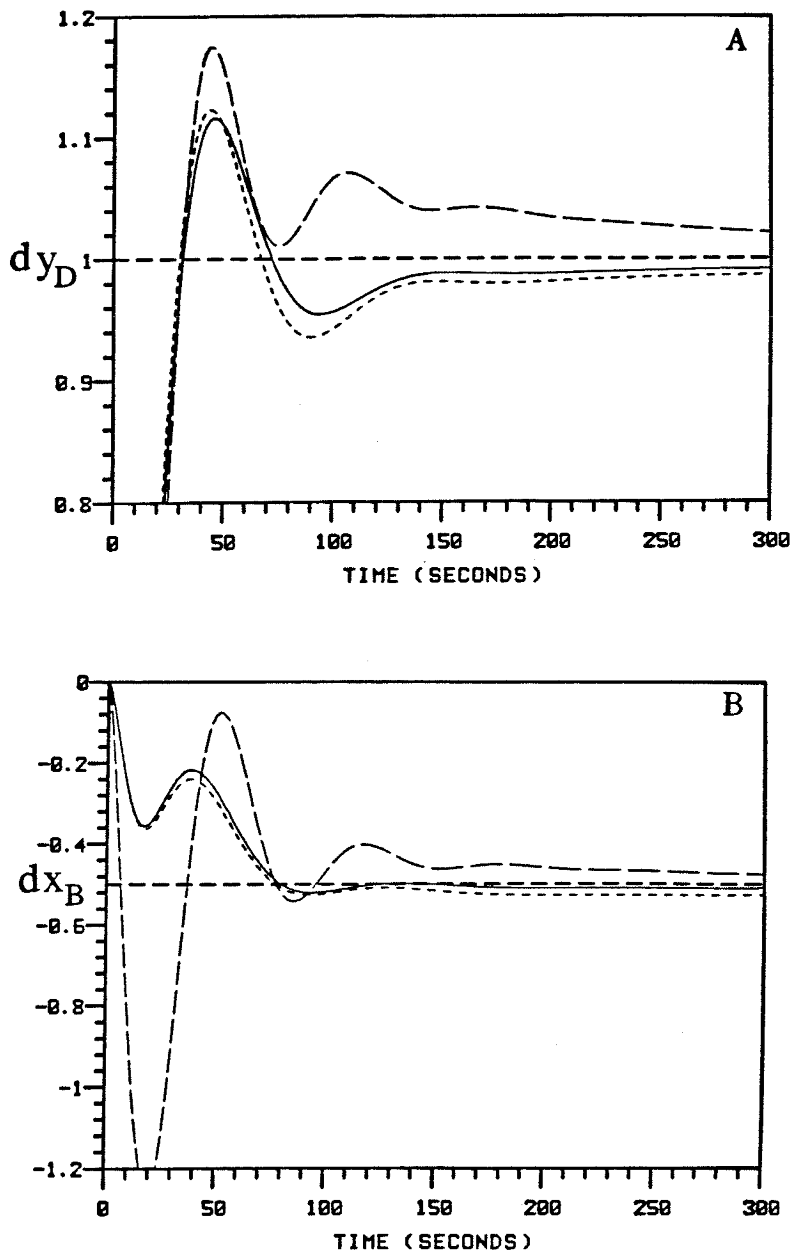


Figure 4

Closed-loop responses for Case Study 1. Controllers obtained from: (— —) internally-balanced model; (----) stability-weighted model; (——) performance-weighted model. (A) Distillate composition. (B) Bottoms composition.

Model Control procedure (Morari et al., 1987)

$$C = \hat{P}_M^{-1}\ \hat{P}_A^{-1}(0)F(I-\hat{P}_A\hat{P}_A^{-1}(0)F)^{-1} \tag{51}$$

$\hat{P}_M$ denotes the minimum-phase portion of $\hat{P}$, obtained from spectral factorization. The control resulting from the internally balanced model is unsatisfactory, as its responses are clearly inferior, particularly that of the bottoms composition. Control based on the stability-weighted model is significantly better, while the performance-weighted model provides additional improvements, particularly in the middle and final segments of the response.

These results are readily justified when one examines the variation in $\bar{\sigma}(E_m\hat{H})$, as noted by Figure 5. The inadequacy of the unweighted internal balancing procedure is evident, as $\bar{\sigma}(E_m\hat{H})$ exceeds 1 at low frequencies. Unweighted balancing considers all frequencies equally important, and hence most effort is devoted to fitting over the high-frequency range which has little influence on the control problem. Based on an unweighted criterion, it would appear as if a higher-order model is necessary for successful control system design. The stability weight improves the low-frequency fit, but not as significantly as the performance weight hence the faster integral action in the performance-weighted model's response. Comparing (35) and (38) one notes that for the problem specifications, the performance weight is essentially the stability weight with an extra degree of "roll-off"; this narrows the relevant frequency interval and is most likely the cause for the improved numerical properties of the performance-weighted problem.

<u>Case Study 2</u>. μ-relevant controller reduction

Skogestad and Morari (1986a) present the μ-optimal controller synthesis of the distillation column described in Case Study 1. The

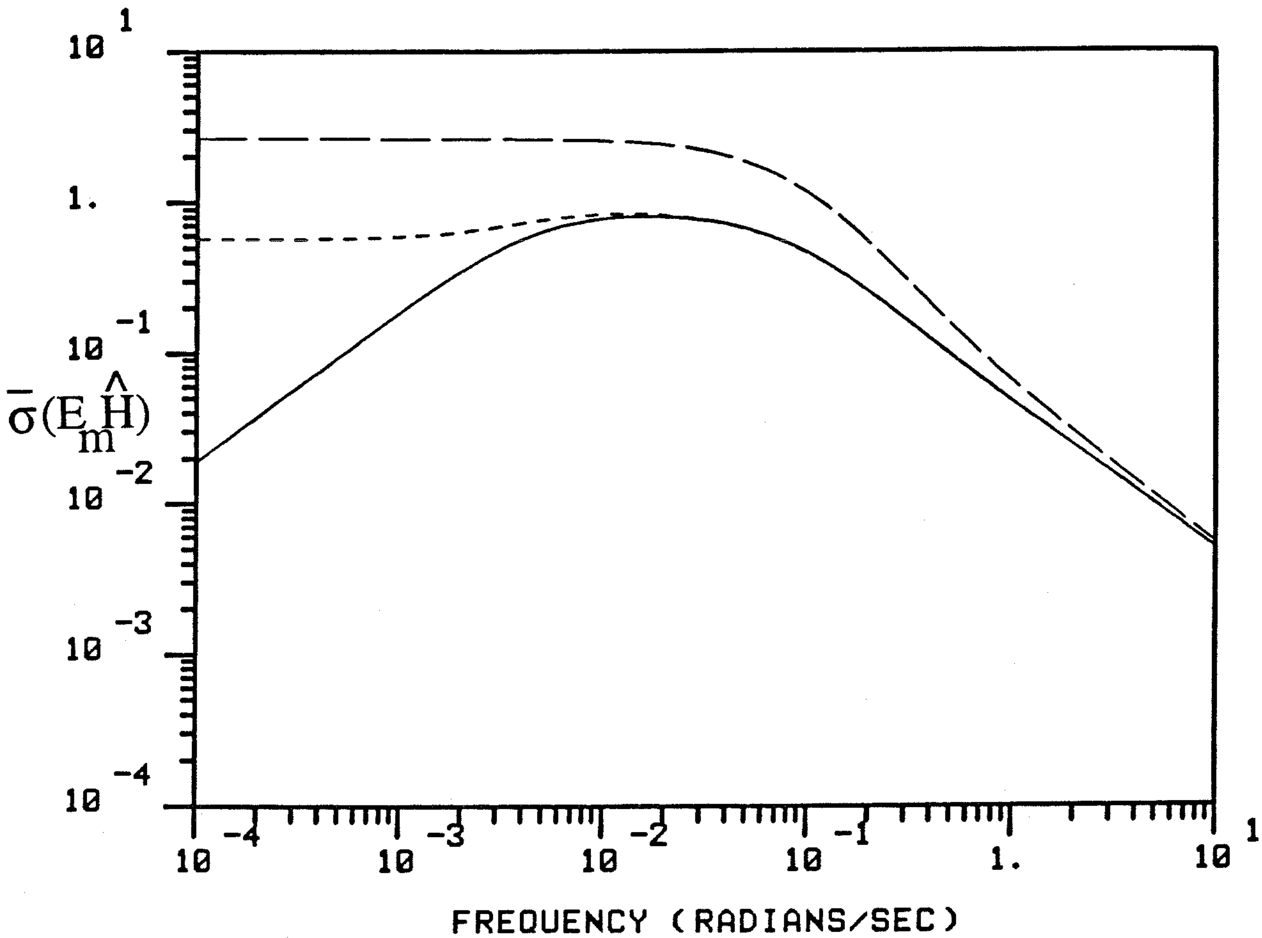


Figure 5

Small Gain condition for models in Case Study 1. (— —) internally-balanced model; (----) stability-weighted model; (——) performance-weighted model.

control problem calls for satisfying performance despite multiplicative input uncertainty. This robustness problem is a special case of Example 1, with the plant model

$$\tilde{P} = \frac{1}{1+75s}\begin{bmatrix} 0.878 & -0.864 \\ 1.082 & -1.096 \end{bmatrix} \tag{52}$$

the input uncertainty description

$$W_{1I} = 0.2\,\frac{5s+1}{0.5s+1}\,I_2 \;;\; W_{2I} = I_2 \tag{53}$$

and the performance weight

$$W_{1P} = 0.5\,\frac{10s+1}{10s}\,I_2 \;;\; W_{2P} = I_2 \tag{54}$$

On the basis of (52)-(54), Skogestad and Morari obtain a seventh-order μ-optimal controller (See Appendix A). Table 2 and Figures 6 and 7 indicate the performance degradation resulting from controller reduction employing the internally-balanced, stability- and performance-weighted balanced realization techniques. The performance-weighted problem leads to clearly superior controllers, with no degradation experienced while going from a 7th to a 6th order controller. An interesting phenomenon occurs when using the stability-preserving weights: μ resulting from the 5th and 4th-order controllers is greater than that obtained from unweighted balancing. Unlike Case Study 1 where preserving nominal stability led to improvements in the nominal performance, the robust performance problem presented here requires performance-oriented weights to obtain an improved result.

5. Comparison with Modal Reduction Techniques

Besides the method of balanced realization, system diagonalization followed by state residualization (more generally known as aggregation or modal reduction) was also attempted in reducing the controller model of Case Study 2. It is worthwhile to comment on the relative

Reduced-Order	Internal Balancing μ	% increase	Stability-Weighted μ	% increase	Performance-Weighted μ	% increase
6	1.62	44.6	1.25	11.6	1.12	0
5	1.62	44.6	1.90	69.6	1.31	17
4	2.71	142	7.50	569.6	2.00	78.6

Table 2. μ-optimal controller reduction, Example 3. μ for full-order, seventh-order controller: 1.12. % increase refers to percent increase over the full-order optimal μ.

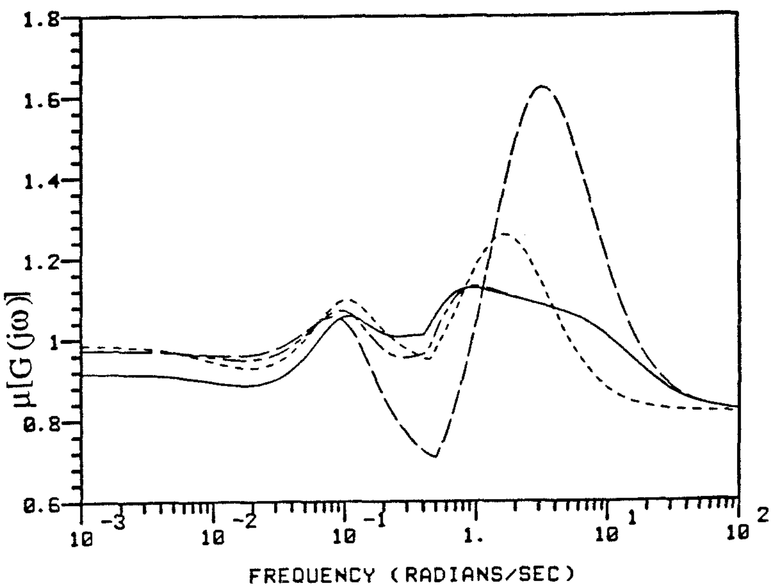


Figure 6

μ [G(jω)] for sixth-order controllers, Case Study 2. (— —) internally-balanced model; (----) stability-weighted model; (——) performance-weighted model; (— - —) full-order model.

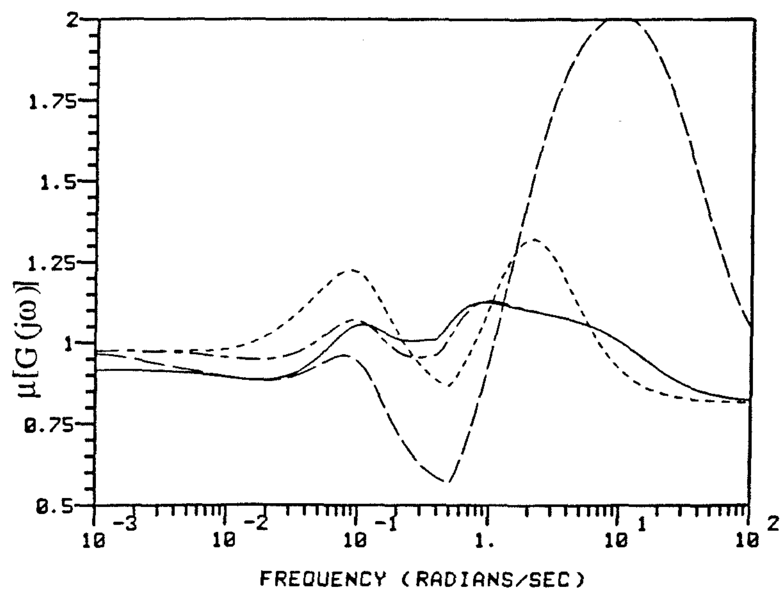


Figure 7

μ [G(jω)] for performance-weighted models, Case Study 2. (— —) 4th-order model; (----) 5th-order model; (——) 6th-order model; (— - —) full-order model merits and disadvantages of both techniques.

The basis for aggregation techniques is that if plant eigenvalues are far apart, those that are most negative contribute the least to the open-loop response of the system and hence can be safely neglected. The steady-state is conserved in the process. Balanced realization, on the other hand, consists of equating the system observability and controllability grammians. These "balanced" grammians are diagonal matrices which are related to the energy transmitted from states to outputs and from inputs to states, respectively. The balancing method consists of solving two Lyapunov equations, followed by two symmetric eigenvalue problems for obtaining the balancing transformation; asymptotic stability of all reduced-order models is assured for the unweighted and one-side weighted cases. H_2 and H_∞ error bounds are also readily generated from the balanced grammian elements, but steady-state agreement is not necessarily maintained.

Both methods are useful because of their relative computational simplicity; however, both are suboptimal because what is really desired is to minimize the weighted H_2 and μ objective functions presented in earlier sections of this paper. As noted by Sinha and Lastman (1985), certain limiting situations can arise where these methods fail: for the aggregation method, if all system eigenvalues are closely-spaced, the choice of retained dominant modes is not clear. For balanced realization, the removal of weakly observable or controllable modes does not necessarily remove the non-dominant system poles. As a consequence, a common ad hoc reduction procedure is obtain a balanced reduction model, then perform aggregation to remove the high-frequency poles that were "skipped" by the balancing procedure.

The phenomenon of high-frequency pole retention was observed in Case Study 2, as one can notice by examining the controllers presented in Appendix A. We decided to confirm the importance, from a controllability/observability standpoint, of the high-frequency poles retained. Our test consisted of aggregating the full-order controller to 6th, 5th, and 4th order, followed then by computing the balanced grammians for each of these controllers. The results are summarized in Appendix B. One notices from the resulting grammians that there did not exist a one-to-one correspondence between the high-frequency poles and the weakly controllable observable states. It is interesting to note that the second example in Lastman and Sinha (1985) also exhibits this behavior.

Clearly, while the balanced realization is a computationally convenient method, it is not a panacea for all model reduction. Additional investigation into improved computational techniques is still warranted. A promising method, which would deal directly with the H_2 and H_∞ objectives, is the multivariable extension of the quadratic and linear programming methods described by Rivera and Morari (1986). Such a technique would be useful for the following reasons:

1. Because the full-order model is represented nonparametrically in terms of its frequency response, speed of computation would be dependent mostly on the desired reduction order. With balanced realization/aggregation techniques, the dimensions of the full-order model are the most limiting factor.
2. Error bounds are readily obtainable from the residual of the quadratic/linear program.
3. The theory and practice of large-order quadratic and linear

programming is fairly well understood. Standard software packages could be used without modification to solve the reduction problems.

6. Concluding Remarks

In this paper we have developed a variety of control-relevant model reduction problems derived directly from the closed-loop H_2 and μ objectives. The proposed problems can be solved through existing techniques, such as frequency-weighted balanced realizations, and resulted in improved performance with less or equivalent computational effort than comparable methods. However, the development of more effective weighted reduction schemes (with non-conservative error bounds, assurances of reduced-model stability, and the possible use of a multiplicative as well as additive error criteria) would be a welcome addition to the field and would greatly increase the usefulness of the proposed problems.

Acknowledgments

We are thankful for the assistance provided by S. Skogestad and Prof. John Doyle.

References

Anderson, B. D. O. IEEE Trans. Autom. Control, AC-12, 410 (1967).

Anderson, B. D. O. IFAC Workshop on Model Error Concepts, Boston (1985).

Doyle, J. IEE Proc., 129, 242 (1982).

Doyle, J. IFAC Workshop on Model Error Concepts, Boston (1985a).

Doyle, J. IEEE Control and Decision Conference, Ft. Lauderdale (1985b).

Doyle, J., J. Wall, and G. Stein. IEEE Control and Decision Conference (1982).

Enns, D. Ph.D. Thesis, Dept. of Aeronautics and Astronautics, Stanford University (1984).

Genesio, R. and M. Milanese, IEEE Trans. Autom. Control, 21, 118 (1976).

Glover, K. Int. J. Control, 39, 1115 (1984).

Lastman, G. J. and N. K. Sinha. IEEE Trans. on Autom. Cntrl., AC-30, 301 (1985).

Morari, M., E. Zafiriou, and C. G. Economou. Robust Process Control, Springer-Verlag, to appear (1987).

Moore, B. C. IEEE Trans. Autom. Control, AC-26, 17 (1981).

Osborne, E. E. I. Assoc. Comput. Mach., 7, 338 (1960).

Rivera, D. E. and M. Morari. "Control-Relevant Reduction Problems for SISO H_2, H_∞, and μ-Controller Synthesis", in press, International Journal of Control (1986).

Skogestad, S. and M. Morari. "Ill-Conditioned Plants: High-Purity Distillation", presented at the 1986 AIChE Annual Meeting, Miami Beach (1986a).

Skogestad S. and M. Morari. "Effect of Disturbance Directions on

Closed-Loop Performance", submitted to I&EC Proc. Des. and Dev. (1986b).

Skogestad, S. and M. Morari. "Deriving Simple Bounds from μ Conditions", in preparation (1986c).

Stahl, H. Int. J. Control, 39, 541 (1984).

Stone, B. J. Numerische Mathematik, 4, 114-116 (1962).

Zafiriou, E. and M. Morari. "Robust Digital Controller Design for Multivariable Systems," presented at the AIChE Annual Meeting, Chicago (1985).

Appendix A

Full-order controller state-space description:

$$\dot{x} = Ax + Bu$$

$$y = Cx + Du$$

$$A = \mathrm{diag}(-1.0\text{x}10^{-7}, -1.0\text{x}10^{-7}, -0.1064, -0.151, -8.993, -583.8, -586.7)$$

$$B^T = \begin{bmatrix} 3.245 & -1.762 & 2.245 & 5.478 & -20.33 & 1867 & 900.6 \\ -2.635 & 3.408 & 2.804 & -4.380 & -25.42 & -1493 & 1126 \end{bmatrix}$$

$$C = \begin{bmatrix} 1.774 & 1.102 & 2.478 & 4.962 & 22.42 & -1692 & -232.3 \\ -4.686\text{x}10^{-2} & -1.069 & -2.483 & 4.956 & -22.45 & -1689 & 232.6 \end{bmatrix}$$

$$D = \begin{bmatrix} 5866 & -3816 \\ 5002 & -4878 \end{bmatrix}$$

Reduced-order controllers from performance-weighted balanced realization:

6th-order controller:

$$A = \mathrm{diag}(-1.002\text{x}10^{-7}, -3.272\text{x}10^{-6}, -0.1510, -9.032, -583.8, -586.8)$$

$$B^T = \begin{bmatrix} -65.13 & 72.24 & 5.492 & -90.86 & 1867 & 67.22 \\ -90.09 & 90.31 & -4.394 & -113.6 & -1494 & 84.03 \end{bmatrix}$$

$$C = \begin{bmatrix} 0.6564 & 0.7171 & 4.949 & 5.033 & -1691 & -311.2 \\ 0.6555 & 0.5425 & 4.941 & -5.040 & -1689 & 311.6 \end{bmatrix}$$

$$D = \begin{bmatrix} 5866 & -3816 \\ 5002 & -4878 \end{bmatrix}$$

5th-order controller:

$$A = \mathrm{diag}(-1.0\text{x}10^{-7}, -1.0\text{x}10^{-7}, -0.1510, -372.3, -583.8)$$

$$B^T = \begin{bmatrix} 3.245 & -1.762 & -5.505 & 632 & 1867 \\ -2.635 & 3.408 & 4.404 & 790.2 & -1484 \end{bmatrix}$$

$$C = \begin{bmatrix} 1.774 & 1.102 & -4.938 & -233.2 & -1691 \\ -4.686\text{x}10^{-2} & -1.069 & -4.930 & 233.6 & -1689 \end{bmatrix}$$

$$D = \begin{bmatrix} 5866 & -3816 \\ 5002 & -4878 \end{bmatrix}$$

4th-order controller:

$$A = \mathrm{diag}(-1.0\mathrm{x}10^{-7},\ -1.0\mathrm{x}10^{-7},\ -0.1510,\ -583.8)$$

$$B^T = \begin{bmatrix} 3.245 & -1.762 & 5.509 & 1867 \\ -2.635 & 3.408 & -4.399 & -1494 \end{bmatrix}$$

$$C = \begin{bmatrix} 1.774 & 1.102 & 4.926 & -1691 \\ -4.686\mathrm{x}10^{-2} & -1.069 & 4.942 & -1698 \end{bmatrix}$$

$$D = \begin{bmatrix} 5866 & -3816 \\ 5002 & -4878 \end{bmatrix}$$

Appendix B

Balanced grammians for:

1. Full-order model:

$$W_C = W_O = \mathrm{diag}(1.49\text{x}10^8,\ 1.22\text{x}10^7,\ 4.9\text{x}10^3,\ 1.77\text{x}10^3,\ 1.63\text{x}10^2,\ 6.1\text{x}10,\ 5.7\text{x}10)$$

2. Residualized model with $\lambda = -586.7$ omitted.

$$W_C = W_O = \mathrm{diag}(1.49\text{x}10^8,\ 1.22\text{x}10^7,\ 4.9\text{x}10^3,\ 1.63\text{x}10^2,\ 6.1\text{x}10,\ 5.9\text{x}10)$$

3. Residualized model with $\lambda = -583.8$ and $\lambda = -586.7$ omitted.

$$W_C = W_O = \mathrm{diag}(1.49\text{x}10^8,\ 1.22\text{x}10^7,\ 1.63\text{x}10^2,\ 6.1\text{x}10,\ 5.9\text{x}10)$$

4. Residualized model with $\lambda = -8.993$, -583.8, and -586.7 omitted.

$$W_C = W_O = \mathrm{diag}(1.49\text{x}10^8,\ 1.22\text{x}10^7,\ 1.63\text{x}10^2,\ 6.1\text{x}10)$$

CHAPTER V:

LOW-ORDER SISO CONTROLLER TUNING METHODS FOR THE H_2, H_∞, AND μ OBJECTIVE FUNCTIONS

(To be submitted to Automatica)

LOW-ORDER SISO CONTROLLER TUNING METHODS FOR THE H_2, H_∞, AND μ OBJECTIVE FUNCTIONS

Daniel E. Rivera

Manfred Morari

Chemical Engineering, 206-41

California Institute of Technology

Pasadena, California 91125

Abstract

A methodology for synthesizing low-order compensators that addresses the need for closed-loop robustness, allows for on-line controller adjustment, and is computationally simple is outlined. The method consists of applying the Internal Model Control (IMC) design procedure to a <u>control-relevant</u> reduced-order model, that is, a model obtained by incorporating features of the closed-loop problem as weights in the reduction procedure. As a consequence, lower-order yet better performing controllers, as compared to those resulting from equivalent methods, are obtained. The computational algorithm is outlined, and it is shown that the model reduction problem can be solved efficiently through standard linear or quadratic programming algorithms, while only the IMC filter parameters and process deadtime need to be obtained through more elaborate search techniques. The benefits of the proposed algorithm are shown through examples.

1. Introduction

Process control problems are characterized by models of widely varying complexity. While the temperature response of a stirred tank, for example, might be fully described by a first-order lag with deadtime model, staged separation processes, such as a distillation column, are fundamentally modeled by large systems of differential equations. For low-order models, the optimal controller is of low-order and easy to design (Rivera et al., 1986), but for high-order models, the corresponding optimal compensator can be numerically difficult to obtain and implement. In these cases, the problem of low-order controller tuning is of significant practical value, allowing the possibility of ease of implementation without substantial degradation from optimal performance.

Historically, The low-order controller tuning problem has been attacked using a wide variety of approaches. The literature on this subject can be classified into three broad categories, which are:

1. Direct low-order controller tuning. Featured in this category are the vast array of PID tuning rules based on minimal knowledge regarding the plant, such as its step response (Cohen and Coon, 1953) or its frequency response at the cross-over frequency (Ziegler and Nichols, 1942; Astrom and Hagglund, 1984). Also included are more general techniques which are either closely linked to a particular design methodology, such as pole placement (Davison, 1972; Rao and Lamba, 1975) or which formulate the low-order controller tuning problem as an optimization problem (Harris and Mellichamp, 1985), this usually requiring the solution of a nonlinear programming problem to obtain the controller parameters.

2. Controller order reduction. This approach requires first designing the optimal, high-order controller and then reducing it to a desired order. Controller order reduction is discussed by Wilson, Seborg, and Fisher (1974), Rivera and Morari (1986), and Enns (1984).

3. Plant order reduction. This represents the most common approach to low-order controller tuning. In this two-step procedure, the plant model is first reduced, then, on the basis of the reduced model, a low-order controller is obtained. The literature on plant-order reduction is vast; the reader can refer to a number of surveys (Genesio and Milanese, 1975 ; Bosley and Lees, 1972; Decoster and van Cauwenberghe, 1976) which present the diversity of the field.

With so many approaches postulated for this problem, why then, another low-order controller tuning methodology? The reason is that a number of significant issues required for satisfactory control have not been suitably addressed by previous efforts. Particularly, the low-order controller be robust to model uncertainty; it must be suitable for on-line adjustment and furthermore, it must be parsimonious with respect to the control problem at hand, that is, it must allow for the lowest compensator order possible by capturing the important features of the control problem and no more. Furthermore, the methodology must be flexible to deal with controller structures beyond the PID controller, and must not constitute a sophisticated nonlinear programming exercise.

The Internal Model Control (IMC) design procedure (Morari et. al., 1987) offers a convenient framework for addressing these issues. The basis for the IMC design procedure is the transformation of the classical feedback structure to an alternate yet equivalent structure that replaces the controller c (Figure 1a) with the simpler-to-design q (Figure 1b). Because the design of q incorporates the plant model directly, IMC results in an unambiguous relationship between a plant model and an appropriate feedback compensator.

When IMC is applied to a high-order model, a high-order q results. The extension to low-order controller tuning results from designing q based on a reduced-order description of the plant model. However, the blind application of any of the diverse techniques proposed for model reduction can be very inefficient. Because there exists no "separation principle" between the reduction and control design problems, the process of reduction followed by controller design (or vice versa) can be highly suboptimal. Without an understanding of the link between both problems, higher order than necessary reduced models may be required to obtain satisfactory control. Secondly, tests that indicate the attainable closed-loop stability and performance from a reduced-order model must be available. Rivera and Morari (1986) have developed a "control-relevant" reduction procedure which incorporates these considerations, enabling the designer to obtain the most parsimonious description possible for a particular set of closed-loop specifications.

This paper integrates the control-relevant model reduction procedure developed by Rivera and Morari (1986) with the IMC design procedure to obtain a low-order controller tuning methodology for the H_2, H_∞, and μ objective functions. The efficient solution of the problems posed in this paper is presented, along with examples demonstrating the usefulness and flexiblitity of the proposed methods.

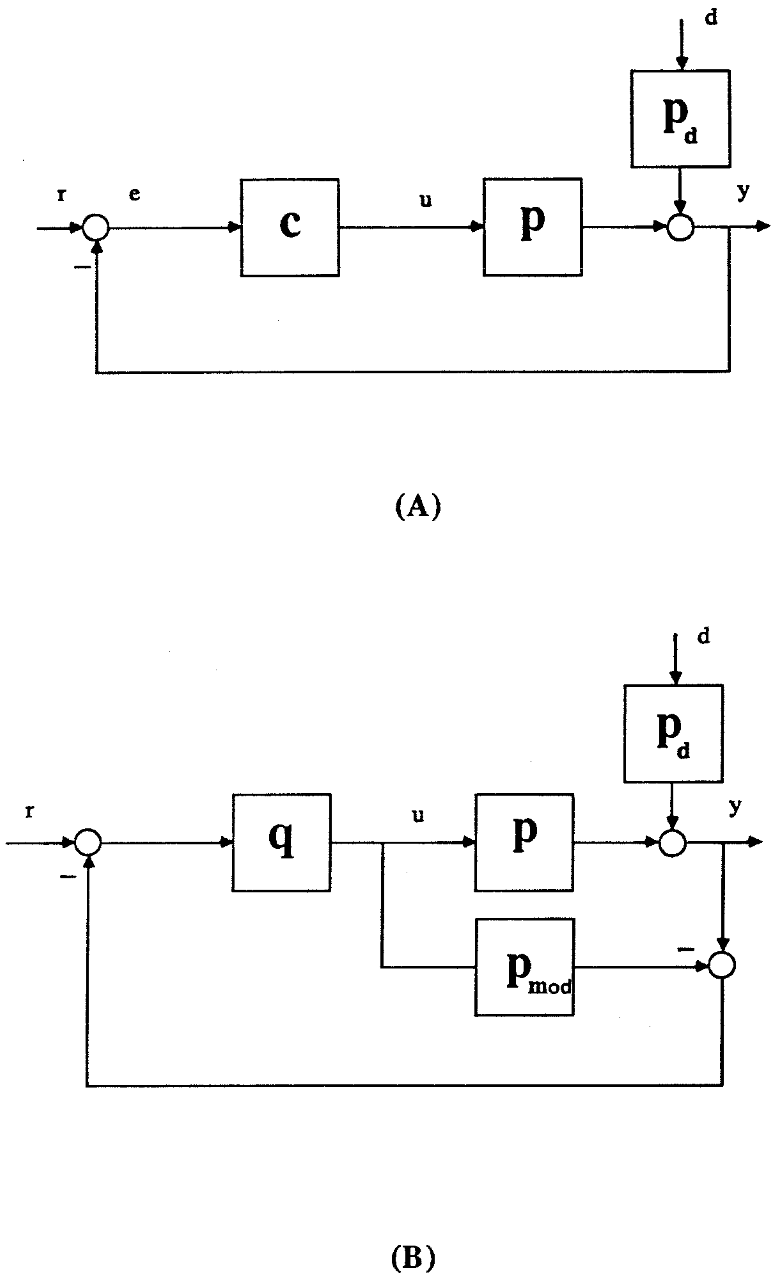


Figure 1

Classical feedback structure (A) and the Internal Model Control structure (B).

2. Performance and Robustness Objectives

This work will focus on single-input, single-output systems and will assume that the plant p with u as its input and y as its output is open-loop stable. $\tilde{p}$ represents the full-order model, while $\hat{p}$ refers to the reduced-order model.

A number of control objectives will be of interest in this study. The first is the H_2 (2-norm), evaluated in the frequency domain through Parseval's Theorem:

$$J = ||e||_2 = \left(\frac{1}{\pi}\int_0^\infty |e(j\omega)|^2\, d\omega\right)^{1/2} \tag{1}$$

$e(j\omega)$ is the error signal for the classical feedback controller (Fig. 1a), represented mathematically as the product of the sensitivity function ε and the setpoint/disturbance characteristics of the problem

$$e = \varepsilon(r-p_d d) = (1+pc)^{-1}(r-p_d d) \tag{2}$$

The second objective of interest is the infinity norm on the weighted sensitivity function (Zames, 1981)

$$||w_p\varepsilon||_\infty = \sup_\omega |w_p\varepsilon(j\omega)| \tag{3}$$

From (3) one finds that $1/|w|$ represents an upper bound on the sensitivity function, that is

$$|\varepsilon(j\omega)| \leq 1\,|w_p(j\omega)| \qquad \forall\omega \tag{4}$$

if and only if

$$||w_p\varepsilon||_\infty \leq 1 \tag{5}$$

The problem formulated by (3) allows the control system designer much flexibility in specifying performance. For instance, selecting w_p as a constant

$$w_p = \alpha \tag{6}$$

is equivalent to requiring that the maximum amplification in the sensitivity function not exceed $1/\alpha$. Another more versatile, yet simple expression for w_p is

2

$$w_p = \alpha\left[\frac{\gamma s+1}{\gamma s}\right] \tag{7}$$

in which the user specifies two parameters: $1/\gamma$ denotes the closed-loop system bandwidth, while $1/\alpha$ sets a limit on the maximum amplification in the sensitivity function. Furthermore, the integrator in (7) indicates that the sensitivity function must have one-pole roll-off.

The H_2 and H_∞ control objectives assume implicitly the absence of plant uncertainty ($p = \tilde{p}$). If uncertainty plays a significant effect in the control design problem, a more meaningful objective is that obtained from the μ synthesis theory developed by Doyle and co-workers (1984). The robust performance theorem presented by these investigators is the following:

<u>Theorem 1</u>: Condition (4) is satisfied for all members of the family of plants Π

$$\Pi:\ \{p: \left|(p-\tilde{p})\tilde{p}^{-1}\right| \leq \ell_m\} \tag{8}$$

if and only if

$$\mu = \sup_{\omega} \left(\left|\tilde{\eta}\ell_m\right| + \left|w_p\tilde{\varepsilon}\right|\right) \leq 1 \tag{9}$$

$\tilde{\eta}$ is the complementary sensitivity function, related to according to $\tilde{\varepsilon}\ \tilde{\eta} + \tilde{\varepsilon} = 1$. Both $\tilde{\eta}$ and $\tilde{\varepsilon}$ are obtained on the basis of the full-order model:

$$\tilde{\varepsilon} = (1+\tilde{p}c)^{-1} \tag{10}$$

$$\tilde{\eta} = \tilde{p}c(1+\tilde{p}c)^{-1} \tag{11}$$

The uncertainty description (8) is generic for SISO systems in the sense that all other norm-bounded uncertainty descriptions (additive, input, or output multiplicative uncertainties) can be lumped into a single lm description without conservativeness; furthermore, comparing (3) and (9) it is clear that for the case of no uncertainty, μ and the infinity norm are equal. We will therefore consider only the 2 and μ norm objectives in the rest of the analysis of this paper, assuming the reader understands that the μ analysis incorporates the infinity norm problem as well.

One should also note that in the case of no performance specification ($w_p = 0$), Theorem 1 reduces to

$$\left|\tilde{\eta}\ell_m\right| \leq 1 \qquad \forall\omega \tag{12}$$

This is the robust stability condition (Small Gain Theorem) and places an important practical restriction on the closed-loop system bandwidth.

3. The Internal Model Control Design Procedure

The Internal Model Control design procedure, introduced by Garcia and Morari (1982) and further elucidated by Morari et al., (1987) forms the basis for this study. The Internal Model Control structure, shown in Figure 1b, is equivalent to the Q-parametrization structure used by Zames (1981) and is related to the classical feedback structure (Fig. 1a) according to the equations

$$q = c(1+cp_{mod})^{-1} \tag{13}$$

$$c = q(1-p_{mod}q)^{-1} \tag{14}$$

p_{mod} represents the nominal plant model. For a full-order controller design, $p_{mod} = \tilde{p}$; to obtain a reduced-order controller, $p_{mod} = \hat{p}$.

The IMC design procedure consists of two steps; the first considers H_2 performance without regard for robustness or constraints, while the second involves detuning the optimal controller such that robustness and input constraint limitations are met. Though, as in the case of the control-relevant reduction problem, there is no separation principle which makes this approach "optimal," the design procedure is very simple and direct. Also, there appear to be very few cases where other more complicated and indirect procedures (e.g., LQG) give better results.

This section summarizes the two-step procedure which is more adequately explained in Morari et al. (1987). Section 4 develops the control-relevant reduction problem, which is then more fully integrated with the IMC design procedure in Section 5. The steps entailing the IMC design procedure are as follows:

STEP 1: Nominal performance

$\tilde{q}$ is designed to yield a "good" system response for the input(s) of interest, without regard for constraints or model uncertainty. We thus choose $\tilde{q}$ such that it is H_2 optimal, which entails the following sub-steps:

A. Factor the plant and input models according to

$$\tilde{p} = \tilde{p}_A \tilde{p}_M \tag{15}$$

$$d' = d_A d_M \quad d' \equiv p_d d \quad \text{or} \quad r(s) \tag{16}$$

p_A and d_A are all-pass functions containing all the process nonminimum phase behavior (deadtime and RHP zeros)

$$p_A,\ d_A \equiv \prod_{i=1}^{N} \frac{-\beta_i s+1}{\bar{\beta}_i s+1} \quad \mathrm{Re}(\beta_i) > 0 \tag{17}$$

As a consequence of the factorization (16), p_M and d_M are minimum-phase functions.

B. Synthesize $\tilde{q}$ according to

$$\tilde{q} = p_M^{-1} d_M^{-1} \{p_A^{-1} d_M\}_* \tag{18}$$

$\{\ \}_*$ is the operator which denotes that after a partial fraction expansion of the operand, all terms involving the poles of p_A^{-1} are omitted. Note that $\tilde{q}$ parametrized according to (18) assures nominal stability. Table 1 summarizes the form of (18) for some common input forms.

STEP 2: Robust Stability and Performance

The presence of uncertainty and input constraints will require that $\tilde{q}$ be rolled off. Therefore $\tilde{q}$ is augmented by a low-pass filter f

$$q = \tilde{q}f \ , \ f = \frac{n_f(r)}{d_f(s)} \tag{19}$$

The order of f is determined by the requirement that q be (strictly) proper and the degree of roll-off required to satisfy the robust performance and stability constraints (Eqns. 9 and 12). The structure of f is determined by the asymptotic properties of d'; for asymptotically step setpoints/disturbances, the filter must satisfy

$$n_f(0) = d_f(0) = f(0) = 1 \tag{20}$$

A one-parameter filter achieving this is

$$f = \frac{1}{(\lambda s+1)^n} \tag{21}$$

If the problem calls for no offset to ramp setpoint/disturbances, the filter requirements are, in addition to (20),

$$n_f'(0) = d_f'(0) \tag{22}$$

where the prime denotes differentiation with respect to s. A one parameter filter structure satisfying (22) is

$$f = \frac{(n\lambda s+1)}{(\lambda s+1)^n} \tag{23}$$

The parameters of f must be determined such that (9) and (12) are satisfied. While this selection might be clear for the cases when l_m = 0 or when only robust stability (12) needs to be satisfied, the

Form of d'	q
$\frac{1}{s}$	p_M^{-1}
$\frac{1}{\tau s+1}$	$\frac{p_M^{-1}}{p_A(-\frac{1}{\tau})}$
$\frac{1}{s(\tau s+1)}$	$p_M^{-1}\left[1+(1-p_A^{-1}\ (-\frac{1}{\tau}))\ \tau s\right]$
$\frac{1}{s^2}$	$p_M^{-1}(1-p_A'(0)s)$
$\frac{1}{s^2(\tau s+1)}$	$p_M^{-1}\left[\tau^2 s^2(p_A^{-1}(\frac{1}{\tau}) - \frac{p_A'(0)}{\tau} - 1) - sp_A'(0) + 1\right]$

Table 1. H_2-optimal factorizations for common input forms. ' denotes differentiation with respect to s.

selection for robust performance is not obvious. In this case, a gradient search (as discussed by Zafiriou and Morari (1986)), might be useful.

Having completed the two-step design, the *nominal complementary sensitivity function* is defined from $\tilde{\eta} = \tilde{p}q$. The equivalent classical compensator implied by this design is expressed as

$$c = \tilde{\eta}\tilde{p}^{-1} (1-\tilde{\eta})^{-1} \tag{24}$$

For many process models common to industrial practice (e.g., first and second-order lags), the structure of (24) is that of a PID-type controller. Full details on this equivalence are presented by Rivera et al. (1986).

4. Defining the Control-Relevant Plant Reduction Problem

The previous discussion indicated how an IMC controller is obtained from a full-order linear plant model; to design a low-order controller for a high-order system, the IMC design procedure must be performed on a reduced-order model $\hat{p}$. Model reduction, however, places additional restrictions on control system stability and performance. Our goal, then, is to formulate model reduction problems that 1) take advantage of the fact that the ultimate goal of the reduction procedure is closed-loop design, and 2) are computationally convenient.

The first step is to represent the mismatch between full and reduced models in some convenient form. A suitable description is the multiplicative error e_m

$$e_m = (\tilde{p}-\hat{p})\hat{p}^{-1} \tag{25}$$

which along with q determined on the basis of the reduced model ($q = \hat{q}f$, with $\hat{q}$ obtained from $\hat{p}$) results in expressions for the sensitivity and complementary sensitivity functions that incorporate the reduction error:

$$\tilde{\varepsilon} = (1-\hat{\eta})(1+e_m\hat{\eta})^{-1} \tag{26}$$

$$\tilde{\eta} = (1+e_m)\hat{\eta}(1+e_m\hat{\eta})^{-1} \tag{27}$$

e_m affects the IMC design procedure in the following manner:

1. Nominal Stability. A stable system is not assured by the parametrization of q (Equation 18). It follows from (26) and (27) that for closed-loop stability the Nyquist plot of $\hat{\eta}e_m$ should not encircle (-1,0). A sufficient condition (obtained through the Small Gain Theorem) establishes bounds on the magnitude of e_m

$$\left|\hat{\eta}e_m\right| < 1 \qquad \forall\omega \tag{28}$$

2. Performance. e_m corrupts the performance expected from $\hat{\eta}$. This effect can be assessed quantitatively by evaluating the performance objectives (1) and (9) using (26) and (27):

$$||e||_2 = \left(\frac{1}{\pi}\int_0^\infty \left|(1-\hat{\eta})(1+e_m\hat{\eta})^{-1}(r-p_d d)\right|^2 d\omega\right)^{1/2} \tag{29}$$

$$\mu = \sup_\omega \left(\left|(1+e_m)\hat{\eta}(1+e_m\hat{\eta})^{-1}\right| \ell_m + \left|w_p(1-\hat{\eta})(1+e_m\hat{\eta})^{-1}\right|\right) \tag{30}$$

(29) and (30) indicate the usefulness of the IMC structure for assessing the "control-adequacy" of a reduced models; low-order controller tuning; from a specified $\hat{\eta}$ (which reflects desired/attainable nominal performance based on the reduced model), e_m (the model reduction error), and $r-p_d d$ (the setpoint/disturbance characteristics), an objective function of choice can be readily evaluated and compared with that obtained on the basis of $\tilde{\eta}$ obtained from the full model. Furthermore, for the robust performance objective defined by (9), Theorem 1 can be applied to rigorously determine that the performance specification has been met, despite reduction error.

While (29) and (30) are rigorous statements of the control-relevant reduction problem, the nonlinear relationship between the control error e and the model reduction error e_m makes their solution numerically difficult and requires the use of nonlinear programming routines for computation. Generalized statements regarding convergence and existence of solutions are difficult, if not impossible to state in general. Rivera (1984) has implemented a programming solution for the problem defined by (29) and found it to be practical only when a limited number of parameters need to be obtained.

A computational convenient problem formulation is the form

$$\min_{\hat{p}} \; ||(\text{Weight}) \times e_m||_\alpha \;, \; \alpha = 2,\infty \tag{31}$$

A model reduction problem written according to (31) can be solved using the methodology presented by Stahl (1984); this is discussed in Section 5.

The first step towards rewriting (29) and (30) in the form of (31) is the need for convenient approximations of $\tilde{\eta}$ and $\tilde{\varepsilon}$. Assuming that condition (28) holds (which, as discussed in the previous section, implies nominal stability of the closed-loop system despite reduction error), then (26) and (27) can be approximated by the truncated geometric series

$$\tilde{\varepsilon} \;\tilde{=}\; (1-\hat{\eta})(1-e_m\hat{\eta}) \tag{32}$$

$$\tilde{\eta} \;\tilde{=}\; \hat{\eta}(1+(1-\hat{\eta})e_m) \tag{33}$$

This approximation implies that $\left|\hat{\eta}e_m\right| << 1$ over the bandwidths of $(1-\hat{\eta})$ and $\hat{\eta}$, respectively.

Substituting (32) and (33) into the objective expressions (1) and (9) one obtains

$$||e||_2 \cong ||e||_{2,\text{approx}} = (\frac{1}{\pi}\int_0^\infty |(1-\hat{\eta})(1-e_m\hat{\eta})(r-p_d d)|^2\, d\omega)^{1/2} \tag{34}$$

$$\mu \cong \mu_{\text{approx}} = \sup_\omega \; (|\hat{\eta}(1+(1-\hat{\eta})e_m)|\; \ell_m + |w_p(1-\hat{\eta})(1-e_m\hat{\eta})| \tag{35}$$

From the property

$$||a||_\alpha - ||b||_\alpha \le ||a+b||_\alpha \le ||a||_\alpha + ||b||_\alpha \;,\; \alpha = 1, \ldots, p, \ldots, \infty \tag{36}$$

it is now possible to separate the degradation caused by e_m on the 2-norm/μ objectives

$$||(1-\hat{\eta})(r-p_d d)||_2 - ||(1-\hat{\eta})(r-p_d d)\hat{\eta}e_m||_2 \le ||e||_{2,\text{approx}} \le$$
$$|(1-\hat{\eta})(r-p_d d)||_2 + ||(1-\hat{\eta})(r-p_d d)\hat{\eta}e_m||_2 \tag{37}$$

$$||\; |\hat{\eta}\ell_m| + |w_p(1-\hat{\eta})|\; ||_\infty - ||\; (|(1-\hat{\eta})\ell_m| + |w_p(1-\hat{\eta})|)\; \hat{\eta}e_m||_\infty \le \mu_{\text{approx}} \le$$
$$||\; |\hat{\eta}\ell_m| + |w_p(1-\hat{\eta})|\; ||_\infty + ||\; (|(1-\hat{\eta})|\ell_m + |w_p(1-\hat{\eta})|)\hat{\eta}e_m||_\infty \tag{38}$$

Control-relevant plant reduction problems of the form (31) are obtained by minimizing this degradation term:

$$\text{P1:} \quad \min_{\hat{p}} (\int_0^\infty |(1-\hat{\eta})(r-p_d d)\hat{\eta}e_m|^2\, d\omega) = \min_{\hat{p}}\; J_1 \tag{39}$$

$$\text{P2:} \quad \inf_{\hat{p}} \; \sup_\omega \; (|(1-\hat{\eta})\ell_m| + |w_p(1-\hat{\eta})|)\; |\hat{\eta}e_m| = \inf_{\hat{p}}\; J_2 \tag{40}$$

(39) and (40) represent frequency-weighted reduction problems, with weights that incorporate specifically the nominal sensitivity/complementary sensitivity functions, the setpoint/disturbance characteristics, the sensitivity function bound, and the plant uncertainty description.

5. Synthesis Methods for Low-Order Controller Design

In this section, we integrate the IMC design procedure and the control-relevant model reduction problem to develop a comprehensive low-order controller design technique. The first step requires that the user select the model and filter structures. Among the considerations involved in selecting a suitable structure for $\hat{p}$ include the limitations of the control hardware and the features of the full-order model (whether integrators or significant deadtime is present, for example). Selecting the appropriate filter structure is influenced by the model structure (properness must be maintained) and on the setpoint/disturbance characteristics (no offset for exponentially ramp or setpoint changes). For the case that a user wishes to tune a PID-type controller, the paper by Rivera et al. (1986) indicates what model/filter combinations are appropriate. Following the choice of structure, the tuning problem ensues. This consists of 1) a search for the control-relevant reduced model parameters and 2) a search for the optimal filter parameters.

5.1 Solving the Control-Relevant Model Reduction Problem

The computational procedure for finding the model parameters is patterned after the generalized framework for frequency-response matching proposed by Stahl (1984). The philosophy behind this method is that nonlinear problems such as (39) and (40) can be reformulated as iterative quadratic or linear programs. The application of Stahl's methodology to this problem is exhaustively discussed in Rivera and Morari (1986). This section highlights some of the major points.

The first step in developing this alternate computational scheme is to consider the full-order model $\tilde{p}$ nonparametrically in terms of its frequency response (z triplets consisting of the full-order plant or controller evaluated at a frequency with its associated real (R) and imaginary (I) parts).

$$((\omega_1, R_1, I_1), \ldots, (\omega_i, R_i, I_i), \ldots, (\omega_Z, R_Z, I_Z)) \tag{41}$$

Because of this full-order plant representation, problems (39) and (40) must be defined in terms of discrete, rather than continous operators

$$\text{P1:} \quad J_1 \cong J_1' = \sum_{i=1}^{z} \left|(1-\hat{\eta})\hat{\eta}(r-p_d d)\right|^2_{s=j\omega_i} \left|e_m(j\omega_i)\right|^2 \tag{42}$$

$$\Delta\omega_i = \omega_i - \omega_{i-1}$$

$$\text{P2:} \quad J_2 \cong J_2' = \max_{\omega_i} \left((\ell_m(\omega_i) + \left|w_p(j\omega_i)\right|)\left|(1-\hat{\eta})\hat{\eta}\right|_{s=j\omega_i} \left|e_m(j\omega_i)\right|\right) \tag{43}$$

Assuming some uniform spacing with respect to frequency, increasing z results in (42) and (43) becoming a closer match to their continuous counterparts.

The reduced-order model structure considered is

$$\hat{p} = \frac{a_0+a_1s+\ldots+a_ns^n}{b_0+b_1s+\ldots+b_ms^n} e^{-\theta s} = \frac{N(s)}{D(s)} e^{-\theta s} \tag{44}$$

where $b_0 = 1$ for a model without integrator and $b_0 = 0$, $b_1 = 1$ for a model with a single integrator. Combining (41) and (44) leads to representing e_m as follows:

$$e_m(j\omega_i) = \left.\frac{D(s)(R_i+jI_i)-N(s)e^{-\theta s}}{N(s)\ e^{-\theta s}}\right|_{s=j\omega_i} = \left.\frac{e_L}{N(s)e^{-\theta s}}\right|_{s=j\omega_i} \tag{45}$$

where

$$e_L(j\omega_i) = (b_0+b_1a+\ldots+b_ms^m)(R_i+jI_i)-(a_0+a_1s+\ldots+a_ns^n)\ e^{-\theta s}\big|_{s=j\omega_i} \tag{46}$$

eL is a linear function of the parameters $(b_0,b_1,\ldots,b_m)$ and $(a_0,a_1,\ldots,a_n)$, and can therefore be represented in matrix form as follows:

$$\mathbf{e}_L^T = [\mathrm{Re}(e_L(j\omega_1)),\ \ldots,\ \mathrm{Re}(e_L(j\omega_i)),\ \ldots,\ \mathrm{Re}(e_L(j\omega_z)),$$
$$\mathrm{Im}(e_L(j\omega_1),\ \ldots,\ \mathrm{Im}(e_L(j\omega_i)),\ \ldots,\ \mathrm{Im}(e_L(j\omega_z))] \tag{47}$$

$$\mathbf{e_L} = \mathbf{Mp-v}\ ;\ \mathbf{M} = \begin{bmatrix}\mathbf{M_R}\\ \hline \mathbf{M_I}\end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix}\mathbf{v_R}\\ \hline \mathbf{v_I}\end{bmatrix} \tag{48}$$

The essence of the computational technique is that problems (42) and (43) can be written according to the general form

$$\min_{\theta,\mathbf{p}} ||\mathbf{W}_{\ell}\mathbf{W}_{\mathbf{m}}(\mathbf{Mp}-\mathbf{v})||_{\alpha} \quad \alpha = 2, \infty \; ; \; \ell = 1,2 \tag{49}$$

$\mathbf{W_m}$ is a weight matrix that contains $1/N(s)e^{-\theta s}$, the nonlinear factor in e_m. $\mathbf{W}_{\ell}$ is the control-relevant weight. Solving for $\mathbf{p}$ in (49) is achieved by iteratively solving a linear least squares problem (for the H_2 objective) or a min max problem (for the μ objective). At every iteration, $\mathbf{W_m}$ is updated with the parameters from previous problem solution. Because a previous knowledge of the reduced-model's nonminimum phase behavior cannot be obtained, initially $\hat{\eta}$ in $\mathbf{W}_{\ell}$ is set to f, and then updated according to Table 1 at every iteration. Convergence is determined after both the residual and parameter estimates error between successive iterations reaches a user-specified tolerance.

The deadtime parameter, however, must still be computed by a direct nonlinear search. Our experience has been that this search need not be very elaborate, and can be effectively accomplished through the use of a Golden Section Search.

5.2. Solving for the Optimal Filter Parameters

As with the deadtime parameter, the search for filter parameters does not call for a sophisticated algorithm. Because in most practical situations a single filter parameter would be most convenient, a simple search procedure such as the Golden Section search would be all that is required. For the general case, however, a gradient search would be useful. Recalling that

$$\frac{d|a|^2}{d\lambda_i} = 2|a| \frac{d|a|}{d\lambda_i} a + \frac{da^*}{d\lambda_i} a + a^* \frac{da}{d\lambda_i} = 2\mathrm{Re}\left[a^* \frac{da}{d\lambda_i}\right] \tag{50}$$

* stands for the complex conjugate. One can obtain corresponding gradient expressions for the objectives (1) and (9)

$$\frac{dJ}{d\lambda_i} = \frac{1}{J\pi} \int_0^\infty \mathrm{Re}\left[\gamma_2(1-\tilde{\eta})^* \frac{df}{d\lambda_i}\right] |r-p_d d|^2 \, d\omega \tag{51}$$

$$\frac{d\mu}{d\lambda_i} = \frac{\ell_m}{|\tilde{\eta}|} \mathrm{Re}\left[\tilde{\eta}^* \gamma_1 \frac{df}{d\lambda_i}\right] + \frac{w_p}{|1-\tilde{\eta}|} \mathrm{Re}[(1-\tilde{\eta})\gamma^*{}_2 \frac{df}{d\lambda_i}]\Big|_{\omega = \omega_5}$$

$$\omega_s = \text{frequency at which sup is evaluated} \tag{52}$$

where

$$\gamma_1 = \frac{\hat{p}\hat{q}(1+e_m)}{(1+\hat{\eta}e_m)^2} \frac{df}{d\lambda_i} \tag{53}$$

$$\gamma_2 = -\gamma_1 \tag{54}$$

$\frac{df}{d\lambda}$ for the filter (21) is

$$\frac{df}{d\lambda} = \frac{-ns}{(\lambda s+1)^{n+1}} \tag{55}$$

and for (23)

$$\frac{df}{d\lambda} = \frac{-n(n-1)\lambda s^2}{(\lambda s+1)^{n+1}} \tag{56}$$

5.3. Mode of Operation for the Model and Filter Parameter Searches

Combining the algorithms discussed in 5.1 and 5.2 can be very flexible, and will vary greatly according to problem requirements. For example, in the control-relevant reduction problem, the user may elect to fix the deadtime to a particular value, hence eliminating the need for the Golden Section search. Because of the close relationship between the filter parameters and the closed-loop response, the user could also avoid the filter parameter search and get adequate controllers simply from solving the control-relevant reduction problem.

Two general operating schemes are proposed in Figure 2. The first considers a gradient search for f with the model reduction problem (MRP) embedded in the nonlinear programming problem. In this "integrated" approach, the top-level optimization is the filter parameter search, with the control-relevant reduction problem solved at every evaluation of the filter parameters. The second scheme, the "decomposed" approach, first solves for the control-relevant model, then for the filter parameters. The decomposed scheme, while not optimal, can be executed much faster than the integrated approach and, provided good initial filter parameter estimates are given, without substantial performance degradation.

The final step in the design procedure is the validation of the resulting low-order control system. This involves two steps:

1. Stability Validation. The loop gain transfer function $\tilde{p}\hat{c}$ must satisfy the Nyquist Stability criterion.
2. Performance Validation. Performance validation requires being able to ascertain the degradation caused by reduction. For the μ synthesis problem, performance degradation is rigorously assessed by confirming that Theorem 1 is satisfied. For the 2-norm case, performance degradation can be measured by computing (1).

The effectiveness of the proposed techniques hinges on satisfying the previously specified diagnostics. In the case that the designed controller fails the validation criteria, the designer can either decrease the bandwidth and/or increase the allowable resonance peak implied by $\hat{\eta}$ and/or w_p (the performance specs) (i.e., make the response more sluggish or tolerate more oscillation and overshoot) or increase

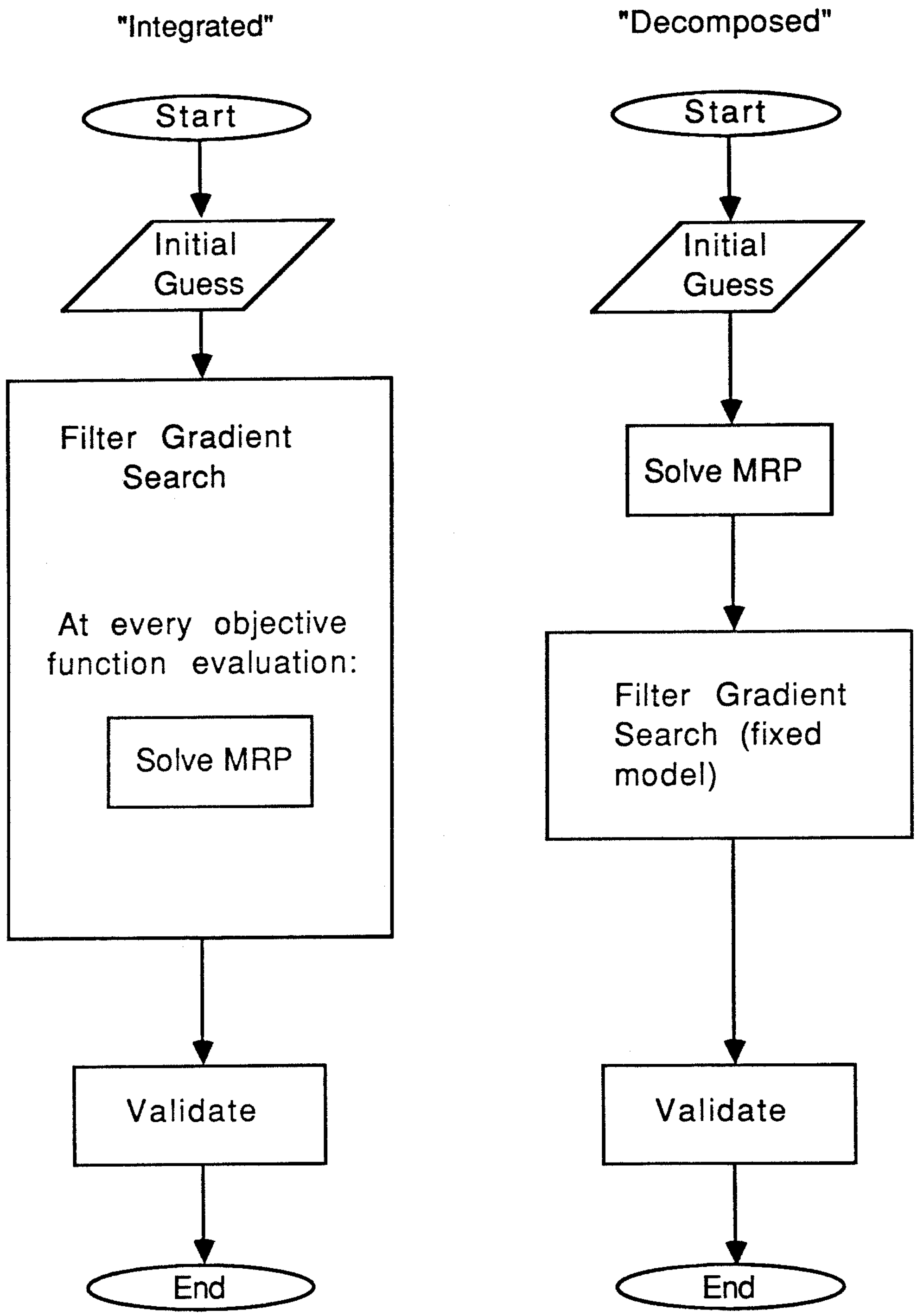


Figure 2. **Schemes for conducting the low-order controller design procedure. MRP stands for "model reduction problem"**

the number of parameters in the reduced model by making new choices for n and m (i.e., decrease the mismatch).

6. Examples

We will compare the proposed low-order controller tuning methodology with the results of direct low-order controller tuning (Example 1), and with the IMC design procedure applied to models obtained from unweighted reduction (Example 2).

Example 1. Consider the control of the following fourth-order model, proposed by Aström and Hägglund (1984):

$$\tilde{p} = \frac{1}{(s+1)(0.2s+1)(0.05s+1)(0.01s+1)} \tag{57}$$

(57) contains four widely spaced poles, at -1,-5,-20, and -100. Aström and Hägglund compare the Ziegler-Nichols settings with those derived from their "dominant pole" design procedure for the "bumpless" PID controller structure

$$u = K(e_p + \frac{1}{\tau_I}\int e(t)dt + \tau_D \frac{de_d}{dt}) \tag{58}$$

Note that $e = r-y$, $e_p = \beta r - y$ with $0 < \beta < 1$, and $e_a = -y$. Both the Aström-Hägglund and Ziegler-Nichols design utilize plant frequency response information close to or at the crossover frequency as the basis for their design. The Ziegler-Nichols settings are

$$K = 15.15, \quad \tau_I = 0.314, \quad \tau_D = 0.0785, \quad \beta = 1 \tag{59}$$

while those of the Aström-Hägglund controller are

$$K = 14.17, \quad \tau_I = 0.407, \quad \tau_D = 0.1018, \quad \beta = 0.17 \tag{60}$$

Corresponding closed-loop output and manipulated variable responses to a unit step setpoint change are shown in Figures 3 a & b. The Ziegler-Nichols controller produces large oscillations in both the output and input responses, while the dominant pole design results in a marked

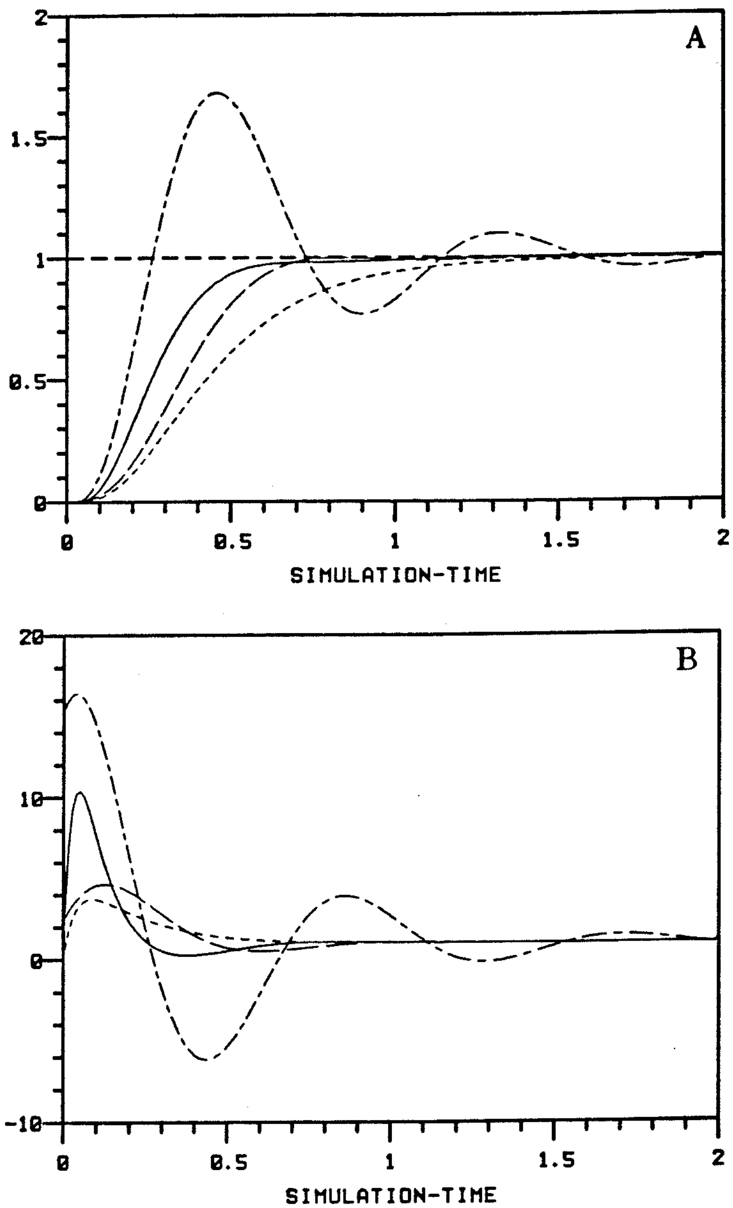


Figure 3. Output responses (A) and input policies (B) for Example 1. (——) IMC, λ = 0.1; (---) IMC, λ = 0.2; (— —) Aström/Hägglund; (— ·) Ziegler-Nichols settings.

performance improvement.

An equivalent low-order H_2- optimal IMC design entails different considerations than those of the Aström-Hägglund and Ziegler-Nichols procedures. First, we select a second-order model structure:

$$\hat{p} = \frac{a_0+a_1s}{1+b_1s+b_2s^2} \tag{61}$$

The need for a smooth input policy starting from the origin leads to the requirement that the filter f result in q with a pole excess greater than 1; a suitable choice is a second-order filter with a single adjustable parameter λ

$$f = \frac{1}{(\lambda s+1)^2} \tag{62}$$

From Rivera et. al. (1986), it is clear that, for step inputs, the model (61) and the filter (62) lead to a PID controller with two additional lags.

As stated earlier, λ is an indicator of the closed-loop time constant and therefore intelligent choices can be made based on process characteristics. The filter parameter will be fixed in this example, requiring only the search for the four rational model parameters, all accomplished by the iterative least-squares procedure discussed in Section 5.1.

IMC designs were performed for λ = 0.1 and 0.2. The resulting reduced models and feedback controllers are

$\lambda = 0.1$:

$$\hat{p}_1 = \frac{0.96636-0.0410s}{0.23019s^2+1.1622s+1} \tag{63}$$

$$c_1 = \frac{0.23019s^2+1.1622s+1}{4.11\text{x}10^{-4}s^3+1.79\text{x}10^{-2}s^2+0.275s} \tag{64}$$

$\lambda = 0.2$:

$$\hat{p}_2 = \frac{0.9928-0.0445}{0.22395s+1.2034s+1} \tag{65}$$

$$c_2 = \frac{0.22395s^2+1.2034s+1}{1.76x10^{-3}s^3+5.73x10^{-2}s^2+0.485s} \tag{66}$$

Closed-loop output and input responses are shown in Figure 3. By manipulating λ, the speed of response, as well as the input magnitude can be set at will. Once implemented, on-line adjustment can be readily accomplished by re-specifying λ and re-computing the feedback controller law.

Example 2. Mandler et al. (1986) have considered the problem of outlet temperature control of a 49th-order methanation reactor, whose impulse response to inlet gas flowrate is presented in Figure 4. We will design H_∞ optimal controllers for unit step setpoint changes, applying the IMC design procedure with optimal filter search to fourth-order models of the form

$$\hat{p} = \frac{a_0+a_1s+a_2s^2+a_3s^3+a_4s^4}{1+b_1s+b_2s^2+b_3s^3+b_4s^4} \tag{67}$$

A reasonable performance weight function is the form suggested by (7), augmented with an extra term:

$$w_{p_1} = \frac{0.5(100s+1)}{100s(s+1)} \tag{68}$$

The bandwidth requirement is 0.01 rad/sec, the maximum amplification is 2, and the extra lag factor reflects (s+1) the fact that the step setpoint change attenuates the sensitivity function after ω > 1; beyond ω > 1, the amplification condition is relaxed.

A first-order filter structure according to (21) was chosen, with the "decomposed" approach applied twice for obtaining the optimal λ. Initially, the filter guess was set to λ = 20, then repeated with the optimal λ from the first iteration, λ = 34.1.

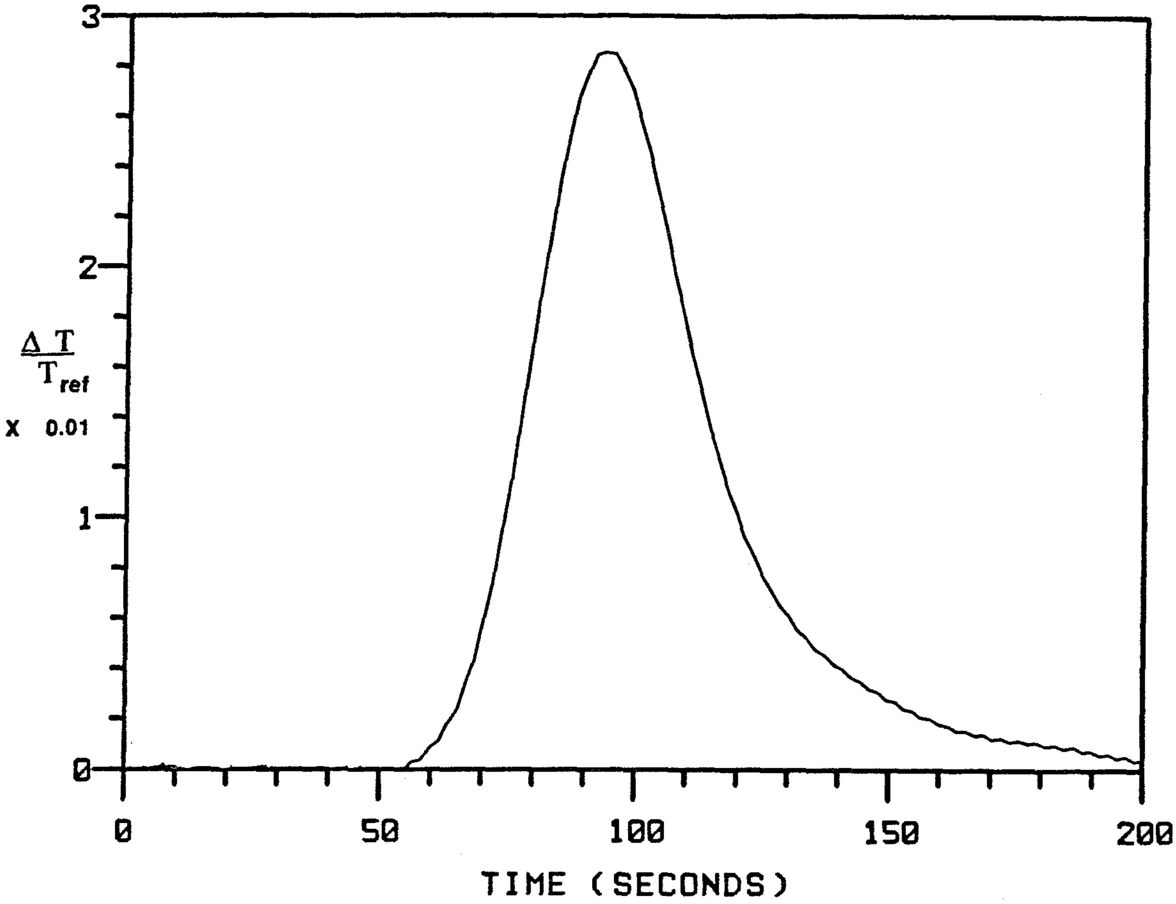


Figure 4

Impulse response for full-order reactor model, Example 2.

For comparison purposes, unweighted frequency response matching according to the objective

$$J_{OL} = \left[\frac{1}{\pi} \int_0^{\infty} \left| \tilde{p}(i\omega) - \hat{p}(i\omega) \right|^2 d\omega \right]^{1/2} \tag{69}$$

was examined. (69) represents the integral of the squared impulse response energy which traditionally has been considered as an adequate objective for both model reduction and identification. The solution technique used to solve (69) is that of Sanathanan and Koerner (1962), which also falls under the generalized framework presented by Stahl (1984). The model obtained by solving (69) was then "fed" to the optimal filter search.

These reduced models, along with the corresponding values for the H_∞ norm, are shown in Table 2. Closed-loop responses are shown in Figure 5. The improvement caused by the use of control-relevant reduction is obvious, with the performance obtained from the control-relevant reduced-order model showing significantly less overshoot and a faster settling time.

Examining the multiplicative error, which, as stated earlier, is the suitable measure for determining a model's effectiveness for control applications, provides the clue that explains the benefits of the control-relevant reduction procedure. As noted in Figure 6, the control-relevant model has low multiplicative error at the important frequency interval for the control problem, defined by the bandwidth 0.02 rad/sec. The unweighted model, on the other hand, has significant low-frequency error, which is the "compromise" that the unweighted reduction procedure must make for an improved fit at the high frequencies, which have little effect on the control problem. Based on

Table 2. Summary of results, Example 2.

Method	Model	λ_{opt}	$\max_{\omega} \lvert w_{p_1}\tilde{\varepsilon}\rvert$	$\max_{\omega} (\lvert\tilde{\eta}\rvert \ell_m + \lvert w_{p_2}\tilde{\varepsilon}\rvert)$	$\lVert e\rVert_2$ bound
1	$\frac{0.8142-31.65s+448.63s^2-3929.4s^3+5534s^4}{1+40.38s+1627s^2+13720s^3+0.228\times10^6s^4}$	15.86	1.10	1.05	32.51
2	$\frac{1.2195-41.891s+893.52s^2-5854.2s^3+66432s^4}{1+53.95s+1674.2s^2+12041s^3+2.034\times10^5s^4}$	68.01	0.901	-	-
3	$\frac{1.2463-40.32s+915.57s^2-5956.1s^3+66679s^4}{1+55.4s+1743.1s^2+13438s^3+2.21\times10^5s^4}$	94.93	-	0.953	28.2

Key:

1. Impulse-response matching (open-loop).
2. Control-relevant H_∞, w_{p_1}-weight.
3. Control-relevant μ, w_{p_2}-weight.

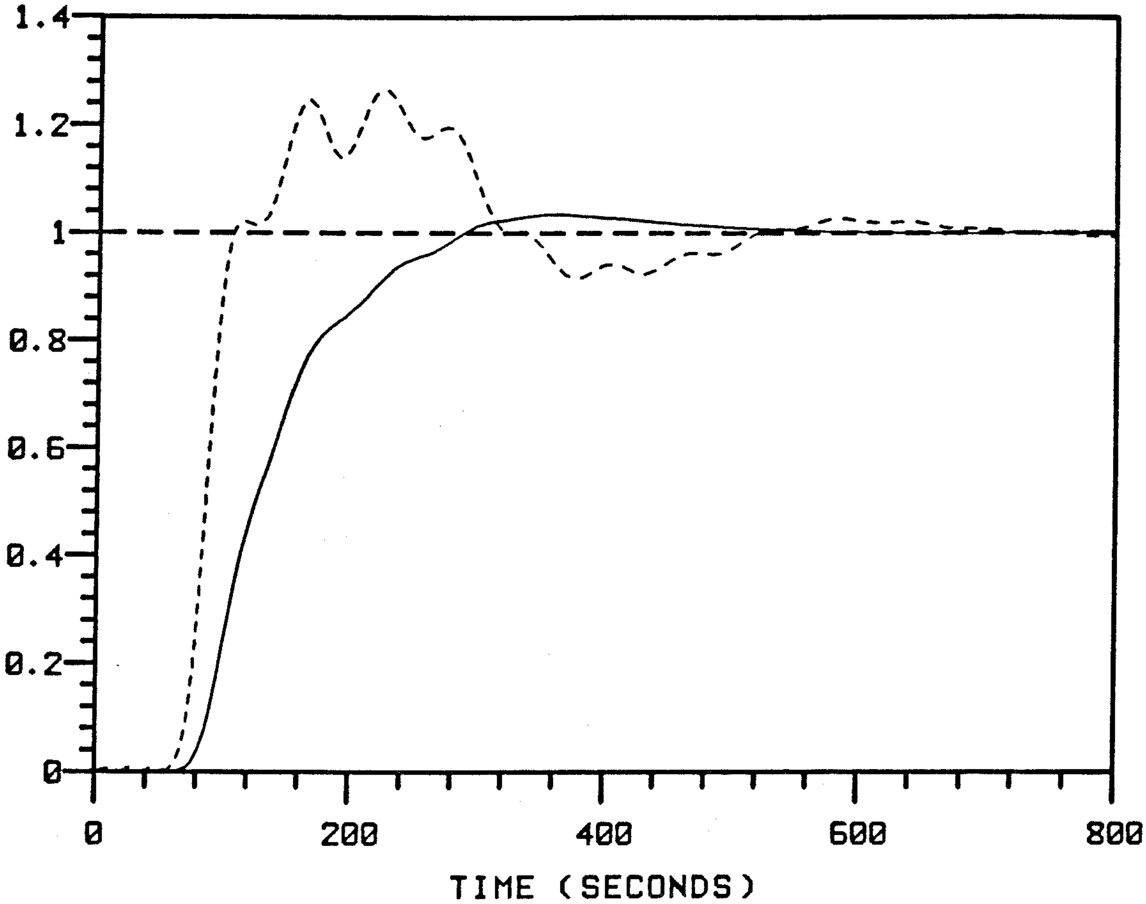


Figure 5. Closed-loop responses to a unit step setpoint change, H_∞ design, Example 2. (——) from control-relevant model; (---) from unweighted model.

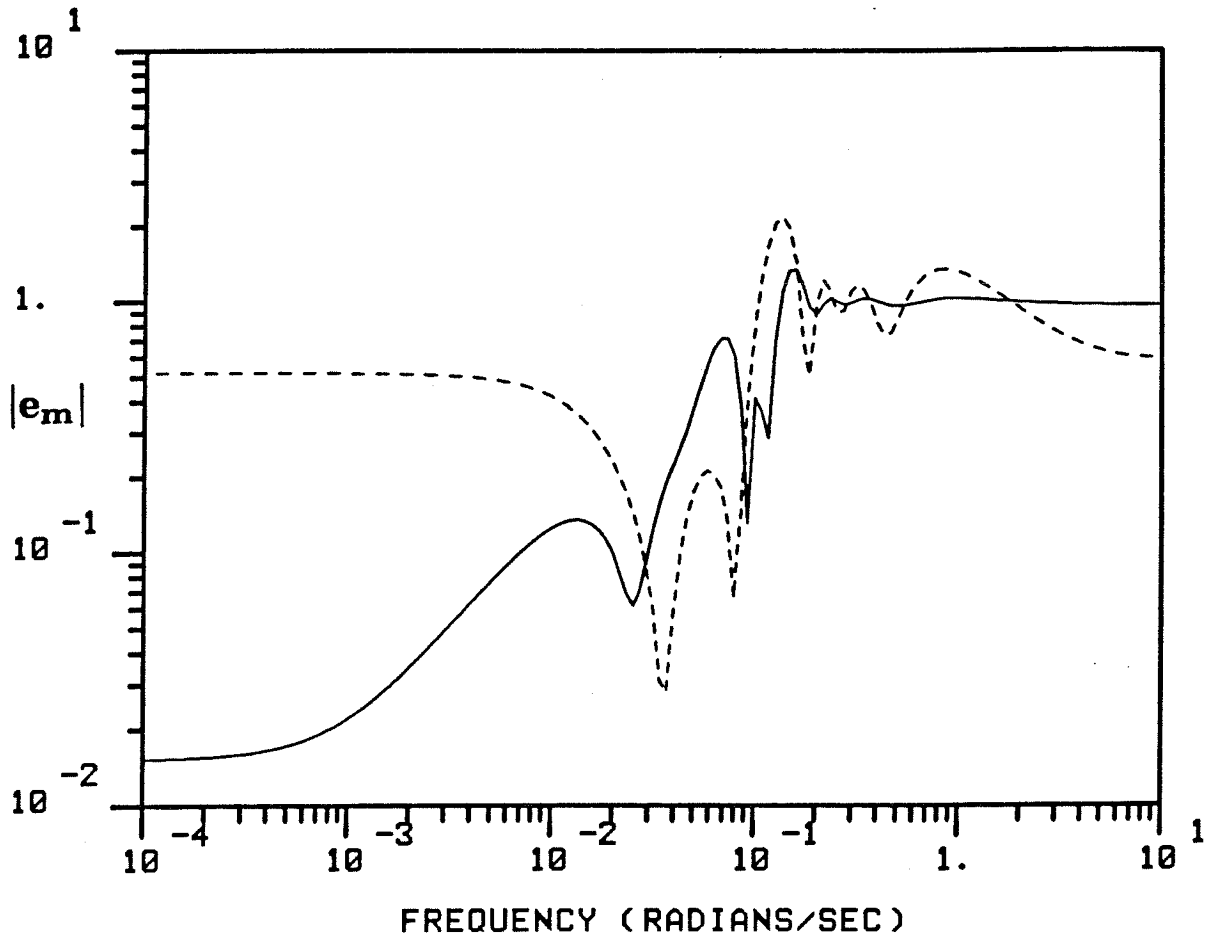


Figure 6. Multiplicative error between full and reduced models, Example 2. (——) control-relevant model; (---) unweighted model.

an unweighted criterion, it would seem as if a higher order model is necessary for satisfactory control, but such is not the case.

The design of a reduced-order control system was also attempted for the case that uncertainty is present. A plot of the uncertainty description l_m is shown in Figure 7; as discussed by Mandler et al. (1986), it accounts for system nonlinearity. The first step is to redefine the performance weight (68) to take into account the added limitations imposed by uncertainty; from the fact that the sensitivity function is bounded for a set of plants according to

$$|\varepsilon| \leq \frac{|\tilde{\varepsilon}|}{1-|\eta \ell_m|} \quad ; \quad |\tilde{\eta}|\ell_m < 1 \tag{70}$$

one can re-specify (68) as follows

$$w_{p_2} = \frac{0.5(100s+1)}{100s(s+1)} \left(1- \left| \frac{1}{68s+1} \right| \ell_m\right\} \tag{71}$$

The choice of $\tilde{\eta}$ comes from the filter constant obtained in the H_∞ case. A new model, using the "decomposed" procedure, is presented in Table 2, and is compared with the optimal filter parameter from the unweighted model. As with the H_∞ case, Theorem 1 is satisfied for the control system derived from the control-relevant model, while the unweighted model fails to meet specifications. Figure 8 compares how μ varies with frequency for both designs. Again we see that the unweighted model results in increased μ at low frequencies, while μ at high frequencies is unnecessarily low. The control-relevant model possesses a better fit over the frequency interval of importance to the control problem, and therefore yields an improved result. Table 2 also compares the 2-norm error bounds for both designs. This error bound is obtained by evaluating (70) using (26), then substituting into the two-norm expression (1). A 13% decrease in the error bound is achieved,

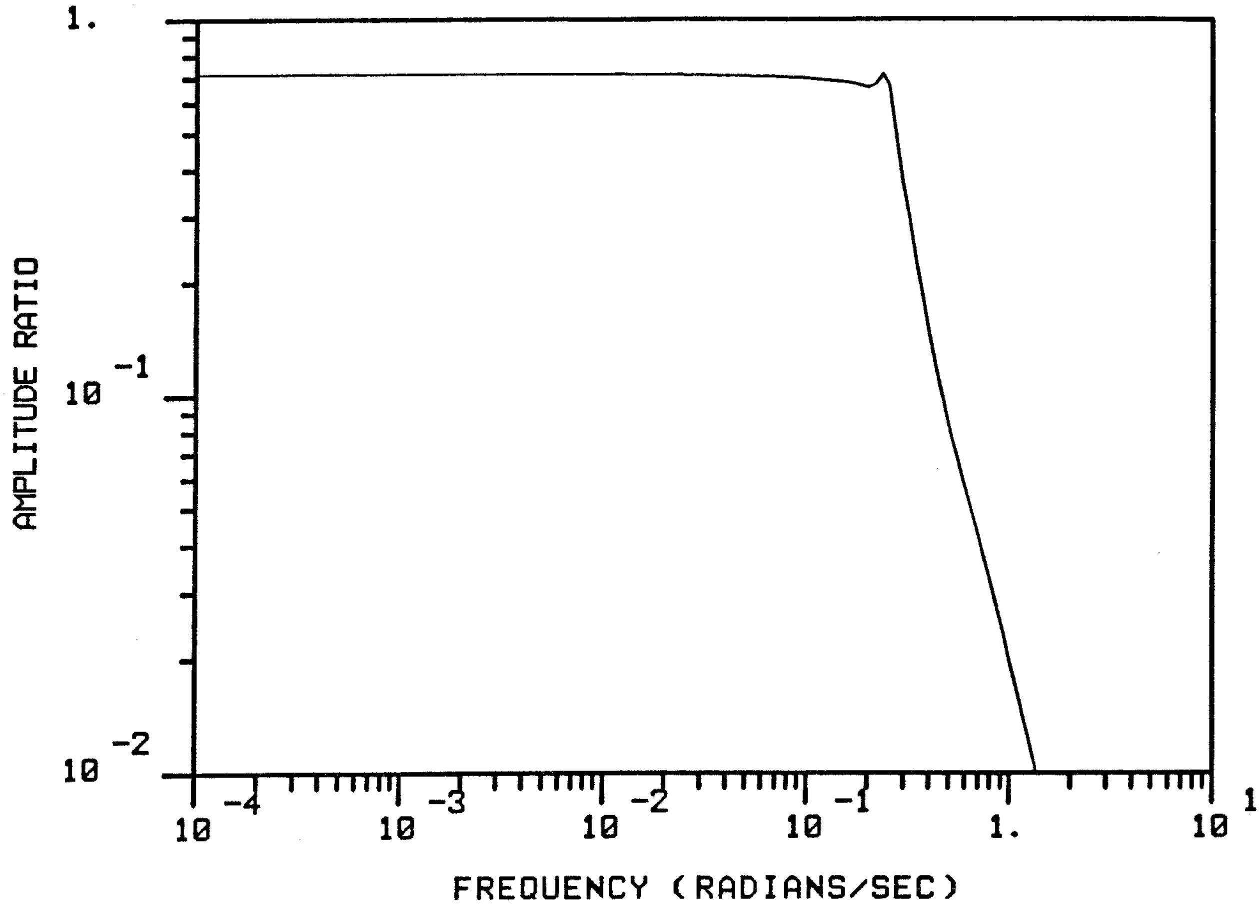


Figure 7

ℓ_m from model nonlinearity, Example 2

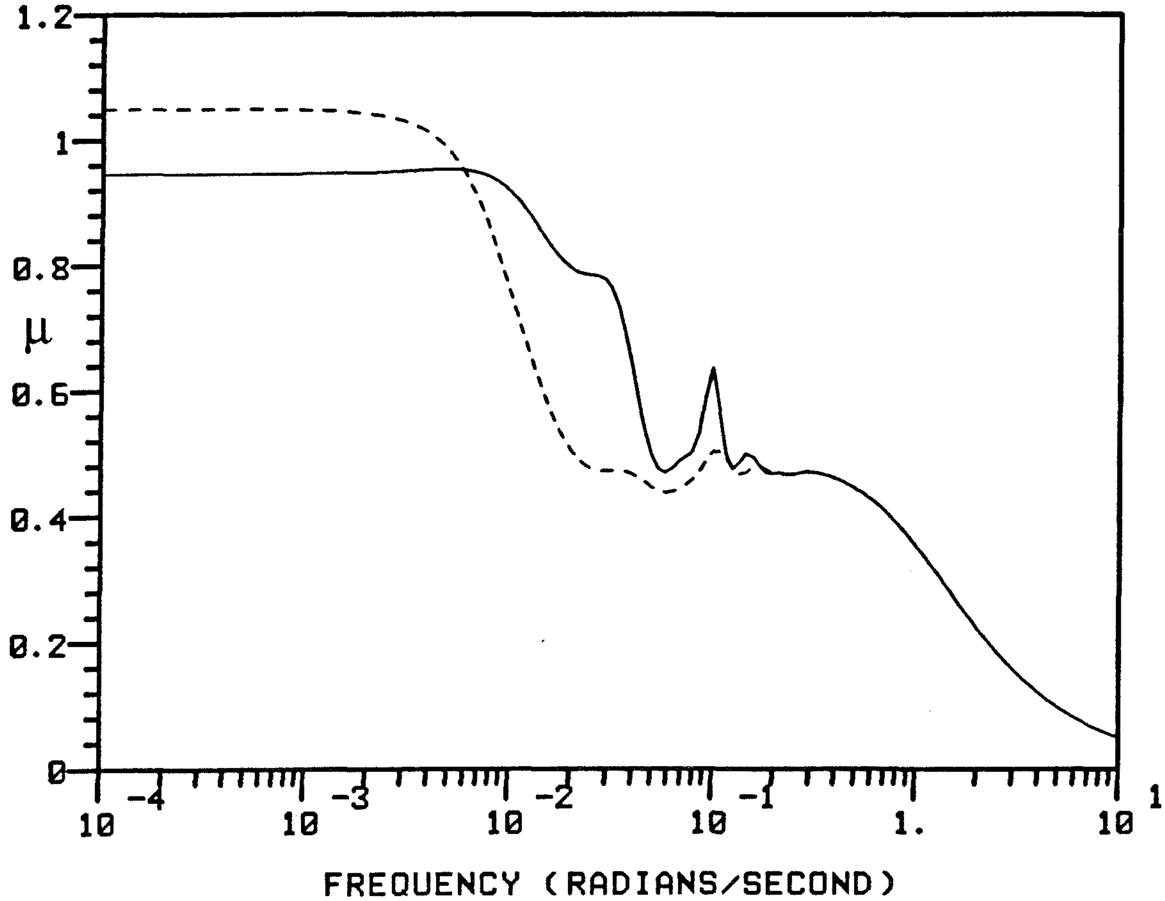


Figure 8. μ with respect to frequency, Example 2.

thanks to the control-relevant weight.

7. Summary and Conclusions

A low-order controller tuning methodology, based on the IMC design procedure and satisfying the H_2, H_∞, and μ control objectives has been presented. The essence of the method is the use of a control-relevant model reduction procedure, which incorporates closed-loop considerations to obtain more accurate, parsimonious model representations for design purposes. The IMC design procedure, consisting of two simple steps, relates the reduced model and closed-loop performance requirements unambiguously to a feedback compensator.

The usefulness of the proposed methodology consists in its great flexibility and ease of computation. While the rational model parameters are computed through quadratic and linear programming algorithms, only the model deadtime and filter parameters need to be computed through Golden Section or gradient search techniques.

Because the computational methods presented require that the full-order model be described by its frequency response, the low-order tuning tools discussed in this paper can be extended to the problem of autotuning as well. Through the use of spectral time series analysis (Jenkins and Watts, 1969), the process frequency response and uncertainty description can be obtained from experiments; this problem is the subject of current investigation (Rivera, 1987).

References

Aström, K.J. and T. Hagglund, "A Frequency Domain Method for Automatic Tuning of Simple Feedback Loops," Proceedings of the 23rd Conference on Decision and Control, Las Vegas, (1984).

Bosley, M.J. and Lees, F.P. Automatica, 8, 765, 1972.

Cohen, G.H. and G.A. Coon. Trans. ASME, 75, 827 (1953).

Davison, E.J. IEEE Trans. Auto. Control, AC-17, 621, (1972).

Decoster, M. and A.R. van Cauwenberghe. Journal A, 17, 68, (1972).

Doyle, J. Notes from the Honeywell/ONR Workshop on Advances in Multivariable Control, Minneapolis (1984).

Enns, D. "Model Reduction for Control System Design," Ph.D. Thesis, Dept. of Aeronautics, Stanford Univ. (1984).

Garcia, C. and M. Morari, I & EC Proc. Des. Dev., 21, 308 (1982).

Genesio and Milanese, IEEE Trans. Autom. Cntrl., 21, 118, (1976).

Harris, S.L. and D.A. Mellichamp. AIChE Journal, 31, 484, (1985).

Jenkins G. and D. Watts. Spectral Analysis and Its Applications, Holden-Day, (1968).

Mandler, J. A., M. Morari, and J. H. Seinfeld. Chemical Engineering Science, 41, 1577 (1986).

Morari, M., E. Zafiriou and C. E. Economou, Robust Process Control, Springer-Verlag, to appear (1986).

Rao, S.V. and S.S. Lamba. Proc. IEE, 122 , 197 (1975)

Rivera, D.E. "Modeling Requirements for Process Control," Ph.D dissertation, California Institute of Technology, (1987).

Rivera, D.E. and M.Morari. "Control-relevant Model Reduction for SISO H_2, H_∞, and μ Controller Synthesis," Int. Journal of Control, in press (1986).

Rivera, D.E., M.Morari, and S.Skogestad. I&EC Proc. Des. and Dev., 25 252 (1986).

Sanathanan C. K. and J. Koerner, IEEE Trans. Auto. Cntrl., AC-8, 56 (1963).

Stahl, H. Int. J. Control, 39, 541 (1984).

Wilson, R.G., D.G. Fisher, and D.E. Seborg. AIChE Journal, 20, 1131, 1974.

Zafiriou. E. and M. Morari. "Synthesis of the IMC Filter by Using the Structured Singular Value Approach," 1986 American Control Conference, Seattle.

Zames, G. IEEE Trans. Auto. Cntrl., AC-26, 301 (1981).

Ziegler, J.G. and N.B. Nichols. Trans. ASME, 64, 759 (1942).

CHAPTER VI:

A CONTROL-RELEVANT IDENTIFICATION METHODOLOGY

A CONTROL-RELEVANT IDENTIFICATION METHODOLOGY

Daniel E. Rivera

Chemical Engineering, 206-41

California Institute of Technology

Pasadena, California 91125

Abstract

Dynamic model identification when the purpose is robust control system design is considered. Through the use of spectral time series analysis, a nominal plant frequency response and an uncertainty description suited for robustness analysis are obtained. For a user-specified model structure, parameter estimation is achieved through a control-relevant procedure that incorporates the uncertainty description and designer requirements for the closed-loop system to obtain a parsimonious fit of the frequency response to an s-domain model. The robustness theory further serves as validation tool for determining the control worthiness of the parametric model.

1. Introduction

The subject of dynamic model identification is an important one in process control, particularly when designing control systems for processes that are difficult to model from fundamental principles. Classical approaches to identification consist of fitting experimental data (from a pulse, step, or pseudo-random binary input) to a linear parametric model through an optimization procedure (e.g., least-squares, maximum likelihood); the absence of correlation in the residuals is the usual measure for determining the adequacy of the fitted model. Classical techniques are "open-loop" procedures in which control system design considerations are usually ignored. A good survey of classical identification methods is found in Aström and Eykhoff (1971).

The classical identification approach is inherently deficient when the purpose of the identification is control-system design. This occurs because the identification objective is unrelated to the control objective; in the optimization step, an attempt is made to fit the data over all regions of time and frequency, some of which are meaningless to the control system design problem. As a result, classical identification techniques produce models of unnecessarily high order. Furthermore, most plants are inherently nonlinear, hence a linear description is adequate only over a narrow operating region. The plant is better represented by a family of linear plants, or equivalently, by a nominal linear plant with an associated uncertainty description. The control system design of uncertain plants is the focus of robust control paradigm developed by Doyle (1985), which establishes the conditions for maintaining both stability and performance for a variety of uncertainty descriptions.

The purpose of this paper is to present an identification methodology that is control-relevant because it originates from the robust control paradigm. The initial portion of the paper introduces the Structured Singular Value (SSV) as a suitable measure for robustness. The identification procedure is then formulated with the robustness condition in mind. The proposed methodology involves three aspects: 1) the use of spectral time series analysis to obtain the nominal plant frequency response and an uncertainty description, 2) the use of frequency-weighted parameter estimation for fitting the nominal plant frequency response to a designer-specified model structure, and 3) validation procedures that evaluate the control-adequacy (i.e., establish achievable closed-loop performance) arising from the identified model. The result of the control-relevant approach is that lower-order, yet better performing controllers, are possible as compared to those designed on the basis of classically-identified models. Furthermore, the control-relevant procedure is simple and can be implemented using widely available signal processing software.

2. Robust Stability and Performance Measures

Our concern in this paper is fitting experimental data to multivariable linear time-invariant models. A linear model, in most cases, does not provide a complete description of the plant dynamics. The effects arising from nonlinearity, parameter variations, or unmodeled dynamics, which we will refer to generically as "uncertainty," can be captured in terms of norm-bounded perturbations on the linear plant model. The issue of robustness, then, centers on being able to ascertain that control-loop stability and performance are maintained despite uncertainty. The structured singular value μ

represents a useful tool for robustness assessment. As shown by Doyle (1985), any linear interconnection of inputs, outputs, transfer functions and perturbations can be rearranged to fit the $G\Delta$ structure (Figure 1), for which the robustness theorem applies:

Theorem 1. Robust Stability, Structured.

Consider the set of perturbations defined by

$$\mathbf{\Delta} = \{\mathrm{diag}(\Delta_1, \Delta_2, \ldots, \Delta_n) \mid \Delta_i \in \mathbb{C}^{k_j \times k_j}\} \tag{1}$$

and its bounded subset

$$B\mathbf{\Delta} = \{\Delta \in \mathbf{\Delta} \mid \bar{\sigma}(\Delta) < 1\} \tag{2}$$

The generalized plant G is stable for all perturbations described by (2) iff

$$\|G\|_\mu \leq 1 \tag{3}$$

where

$$\|G\|_\mu \triangleq \sup_\omega \mu[G(j\omega)] \tag{4}$$

with μ, the Structured Singular Value, defined as

$$\frac{1}{\mu(M)} = \min_{\Delta \in \mathbf{\Delta}} \{\bar{\sigma}(\Delta) \mid \det(I - M\Delta) = 0\}, \; M = G(j\omega) \tag{5}$$

It must be understood that Theorem 1 also incorporates the robust performance problem. The performance conditions are incorporated in the $G\Delta$ structure in the form of additional perturbations. The $G\Delta$ structure thus allows for the performance and uncertainty characteristics of the feedback problem to be captured in a unified manner.

3. Spectral Time Series Analysis

The field of spectral time series analysis is concerned with obtaining frequency-domain information from records of time-domain data. Exhaustive treatment of this topic is presented in the texts by

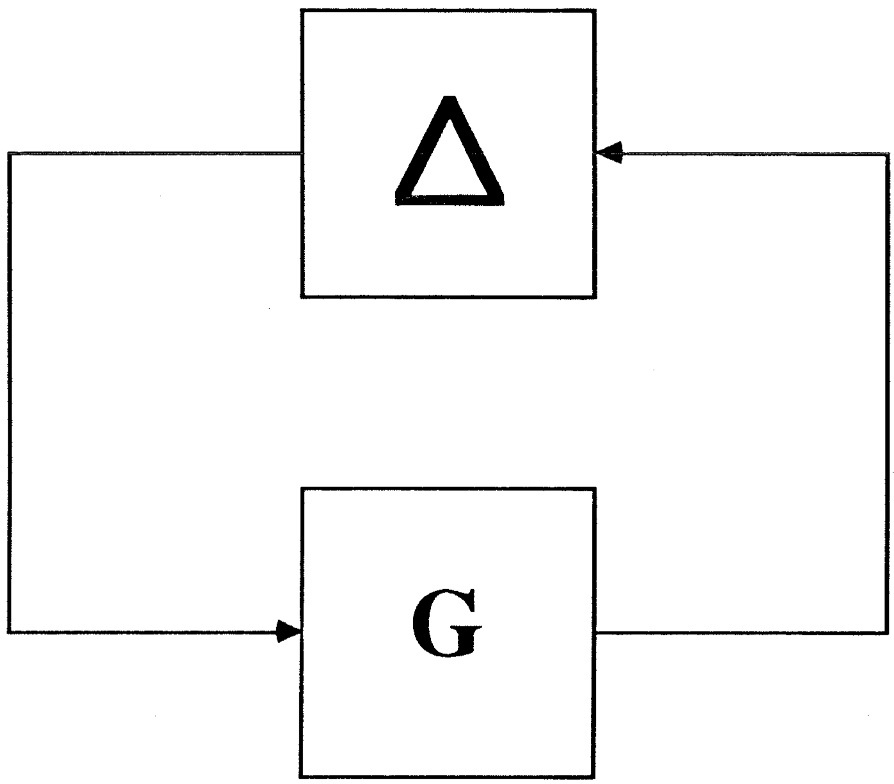


Figure 1. G-Delta Interconnection Structure

Jenkins and Watts (1969) and Koopmans (1974). This section highlights the main aspects of the theory, and we encourage the interested reader to review the aforementioned texts for more information on the statistical details involved.

We consider two vector processes, X_t and Y_t

$$X_t^T = (x_{1t}, \ldots, x_{jt}, \ldots, x_{qt}) \tag{6}$$

$$Y_t^T = (y_{1t}, \ldots, y_{it}, \ldots, y_{rt}) \tag{7}$$

X_t is the vector of plant inputs, while Y_t is the vector of outputs. $\{x_{jt}\}$ and $\{y_{it}\}$ are records (time series) of data assumed to be stationary, (i.e., each series has a constant mean), of equal length, and obtained by sampling at an interval of δ seconds. Furthermore, it is assumed that no external disturbances are corrupting the output time series. The steps involved in obtaining the transfer function estimate $\tilde{P}$ for the plant P from X_t and Y_t are as follows:

1. Compute the auto- and cross-covariances for the augmented vector process:

$$A_t = \left[\frac{X_t}{Y_t}\right] \tag{8}$$

$$\{a_{jt}\} = \text{element in } A_t$$

using the formula

$$c_{ij}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (a_{it}-\bar{a}_i)(a_{j(t+k)}-\bar{a}_j) \qquad 0 \leq k \leq L-1 \tag{9}$$

N is the record length, L are the number of covariance lags (usually chosen between 20 to 30% of N), and the overbar represents the mean value of each element in A_t. The resulting covariance matrix has $(q+r)x(q+r)$ dimensions and is represented as

$$C(k) = \begin{bmatrix} C_{XX}(k) & C_{XY}(k) \\ C_{YX}(k) & C_{YY}(k) \end{bmatrix} \quad (10)$$

2. Take the smooth Fourier transform of the covariance matrix to obtain the (q+r)x(q+r) spectral matrix (containing the auto- and cross-spectral estimates)

$$F(n) = \begin{bmatrix} F_{XX} & F_{XY} \\ F_{YX} & F_{YY} \end{bmatrix} \quad 0 \leq n \leq N_f \quad (11)$$

N_f indicates the number of linearly-spaced frequencies (recommended number is 2 to 3 times L). $\{f_{ij}\}$, the elements of F, are complex quantities

$$f_{ij}(n) = R_{ij} + jI_{ij} \quad (12)$$

with real and imaginary parts estimated by the formulas

$$R_{ij} = 2\delta\{\ell_{ij}(0) + 2 \sum_{k=1}^{L-1} \ell_{ij}(k)w(k) \cos \frac{n\pi k}{N_f}\} \quad 0 \leq n \leq N_f \quad (13)$$

$$I_{ij} = 4\delta \sum_{k=1}^{L-1} q_{ij}(k)w(k) \sin \frac{n\pi k}{N_f} \quad 1 \leq n \leq N_f - 1 \quad (14)$$

$$I_{ij}(0) = I_{ij}(N_f) = 0$$

where

$$\ell_{ij}(k) = \frac{1}{2} \{c_{ij}(k+S) + c_{ji}(k-S)\}$$

$$q_{ij}(k) = \frac{1}{2} \{c_{ij}(k+S) - c_{ji}(k-S)\}$$

The spectral estimates are spaced $1/(2 N_f\delta)$ frequencies apart, ranging from frequencies 0 to $1/2\delta$.

Step 2 requires that w(k), the lag "window" and S, the number of "alignment" lags be specified. A summary of the properties of common lag windows is found in Table 1. Having selected a window type, the parameter M, called the truncation point, must be specified. While

Description	Lag Window	ν, Degrees of Freedom
Rectangular	$w(u) = \begin{cases} 1, & \lvert u\rvert \le M \\ 0, & \lvert u\rvert > M \end{cases}$	$\frac{N}{M}$
Bartlett	$w(u) = \begin{cases} \frac{1-\lvert u\rvert}{M}, & \lvert u\rvert \le M \\ 0, & \lvert u\rvert > M \end{cases}$	$\frac{3N}{M}$
Tukey	$w(u) = \begin{cases} \frac{1}{2}\left(1+\cos\left[\frac{\pi u}{M}\right]\right), & \lvert u\rvert \le M \\ 0, & \lvert u\rvert > M \end{cases}$	$2.667\ \frac{N}{M}$
Parzen	$w(u) = \begin{cases} 1-6\left[\frac{u}{M}\right]^2 + 6\left[\frac{\lvert u\rvert}{M}\right]^3, & \lvert u\rvert \le \frac{M}{2} \\ 2\left[1-\frac{\lvert u\rvert}{M}\right]^3, & \frac{M}{2} < \lvert u\rvert \le M \\ 0, & \lvert u\rvert > M \end{cases}$	$3.71\ \frac{N}{M}$

Table 1. Properties of Common Lag Window. M: Truncation point for window; N: record length.

increasing M decreases the bias in the estimate, the variance associated with the estimate increases. Choosing the appropriate lag window and truncation parameter requires a trade-off between estimate "fidelity" (bias) and "stability" (variance). Practical guidelines are provided by Jenkins and Watts in Chapter 7.

The selection of alignment lags is important for systems that exhibit pronounced phase lag (i.e., systems with deadtime). By selecting S corresponding to the process time delay (deadtime = $S\delta$), the smoothing process is significantly improved.

3. Obtain the transfer function estimate from

$$\tilde{P} = F_{YX}F_{XX}^{-1} \tag{15}$$

and the residual spectrum from

$$F_{ZZ} = F_{YY} - \tilde{P}F_{XY} \tag{16}$$

where

$$Z = Y(j\omega) - \tilde{P}X(j\omega)$$

The residual spectrum is a consequence of smoothing ; it is zero if no smoothing (i.e., $w(k) = 1$ for $M=N$) is applied.

From the literature, one notices that the residual spectrum is utilized in two distinct ways. The most common is to label the residual spectrum as the power spectral density of an external disturbance. The second is to define "confidence intervals" for the transfer function estimate; this latter choice considers the residual as an indicator of plant uncertainty. The most general case presented by Jenkins and Watts (1969) is the multi-input, single-output (MISO) case, for which the joint confidence distribution applies

$$F_{z_i z_i}\,(v_i(P-\tilde{P})F_{XX}(P-\tilde{P})^* v_i^T) \leq \frac{2q}{\nu-2q}\, f_{2q,\nu-2q}\,(1-\alpha) \tag{17}$$

$f_{2q,\nu-2q}$ $\equiv$ Fisher statistic for 2q and ν-2q degrees of freedom, evaluated at a $(1-\alpha)\%$ confidence level.

$$v_1 = [1, 0, \ldots, 0] \qquad v_r = [0, \ldots, 0, 1]$$

The vector v_i specifies the ith row of the error transfer function $P-\tilde{P}$. (17), however, is not written in a form suitable for Theorem 1; a singular value bound is needed. Noting that F_{XX} is a Hermitian matrix, it can be decomposed according to

$$F_{XX} = U\Lambda^{1/2}\,\Lambda^{1/2}\,U^H \tag{18}$$

$$\Lambda = \mathrm{diag}(\lambda_{max}, \ldots, \lambda_{min})$$

U is a unitary matrix, while Λ is the matrix of eigenvalues. As a result of (18), one can perform the following manipulations

$$v_i(P-\tilde{P})F_{XX}(P-\tilde{P})^* v_i^T = \|\, v_i(P-\tilde{P})U\Lambda^{1/2}\,\|_E^2 = \bar{\sigma}^2(v_i(P-\tilde{P})U\Lambda^{1/2}) \tag{19}$$

(recall that the Euclidean norm and the maximum singular value are identical for vectors). A singular value bound of suitable form is thus obtained

$$\bar{\sigma}(v_i(P-\tilde{P})U\Lambda^{\,1/2}\,F_{z_i z_i}^{-1/2}\,f_{2q,\nu-2q}^{-1/2}\,(1-\alpha)) \leq 1 \tag{20}$$

which is assembled into a weighted Δ perturbation problem compatible with Theorem 1 through the following definition for the matrices L and E

$$L^T = [L_{a_1}, \ldots, L_{ai}, \ldots, L_{ar}] \tag{21}$$

$$L_{ai} = \Lambda^{-1/2}\,U^H\,F_{z_i z_i}^{1/2} f_{2q,\nu-2q}^{1/2}\,(1-\alpha)$$

$$\Delta_a = \mathrm{diag}(\Delta_1, \ldots, \Delta_i, \ldots, \Delta_r)$$

$$E = \mathrm{diag}\,(\overbrace{u, \ldots, u}^{r})\ ,\ u = (1, 0, \ldots, 0) : 1 \times q \text{ vector}$$

Figure 2 presents the block diagram for this uncertainty description.

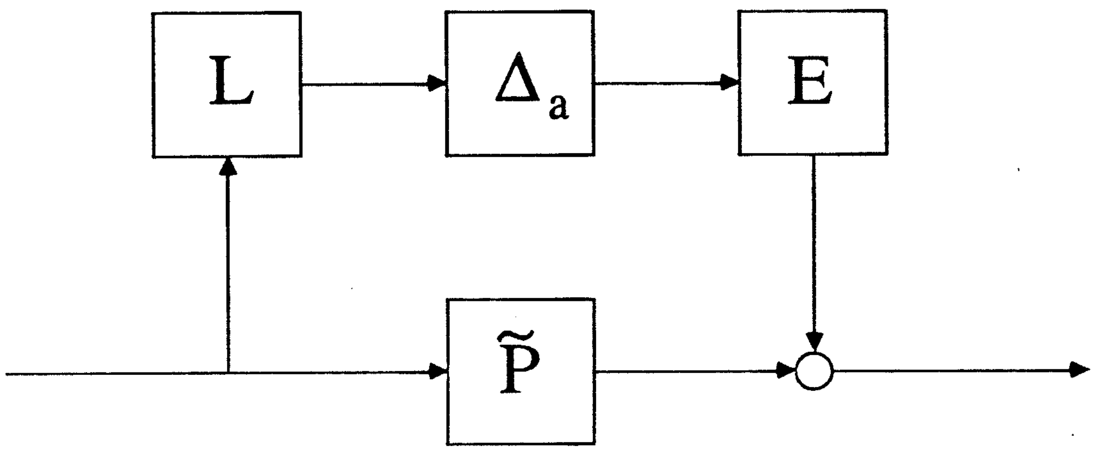


Figure 2. Block diagram description for the plant P, in terms of $\tilde{P}$, the nonparametric model, and L and E, the uncertainty description.

Equation (20) generates some helpful physical insights regarding the control-relevant identification problem:

1. <u>The Noise-to-Input Signal (N/IS) Ratio as an Indicator of the Extent of Plant Uncertainty</u>.

From the expression

$$\bar{\sigma}(v_i(P-\tilde{P}))\ \underline{\sigma}\ (U\Lambda^{1/2}\ F_{z_i z_i}^{-1/2} f^{-1/2}) \leq \bar{\sigma}(v_i(P-\tilde{P})U\Lambda^{1/2}\ F_{z_i z_i}^{-1/2} f^{-1/2}) \qquad (22)$$

one obtains the singular value bound

$$\bar{\sigma}(v_i(P-\tilde{P})) \leq \left[\frac{F_{z_i z_i}}{\lambda_{min}}\right]^{1/2} f^{1/2} \qquad (23)$$

The quantity $F_{z_i z_i}/\lambda_{min}$ is, in effect, the noise-to-input signal ratio evaluated at a particular frequency, and serves as a simple indicator of the significance of uncertainty for the chosen set of experimental conditions. For most physical systems, the N/IS ratio is likely to be a function of the input characteristics. A reasonable conjecture appears to be that if, for a large magnitude input signal, the N/IS ratio remains small over the frequency range of interest, uncertainty effects can be neglected and control-system designs that do not incorporate robustness explictly (such as standard H_2- optimal control) can be expected to perform adequately.

The importance of the N/IS ratio has also been noted by Ljung (1984), although his treatment is restricted to SISO systems.

2. <u>Usefulness of a Multivariable Normal Input Signal.</u>

The derivation of (23) entailed the conservative step denoted by (22). If a multivariable-normal input signal is used, such as the multivariable pseudo-random input signal (Briggs and Godfrey, 1966), the input spectrum F_{XX} is conveniently expressed as a scalar times identity.

$$F_{XX} = \Upsilon(jn)I_q \tag{24}$$

Using (24), expressions (20) and (23) are equivalent. Other considerations, however, must be incorporated in the choice of a suitable input signal. One example is whether a multilevel pseudo-random random sequence, as opposed to a binary one, will be more effective for control purposes.

The interpretation given to the residual spectrum has serious consequences on the control problem, and therefore this decision must be based on a careful physical understanding of the process. To label the residual exclusively as an external disturbance when significant uncertainty is present will lead to an unstable closed-loop system. Alternatively, using the residual spectrum strictly to define plant uncertainty when significant external disturbances are present can lead to unpredictable results, depending on the character of the disturbance.

A comprehensive identification methodology for control purposes must recognize the presence of both uncertainty and external disturbances. The details of such a methodology are, unfortunately, beyond the scope of this paper; more investigation on this problem is definitely warranted. One reasonable approach is to perform more than one experiment on the plant. In the first experiment, the plant input is held fixed, and the resulting output spectrum is treated as the disturbance power spectral density. In the second, the plant input is manipulated with a sufficiently wide magnitude signal such that the contribution of the disturbance to the plant output is negligible. As stated previously, one can conjecture that a larger magnitude will draw out more of the nonlinear character of the process, and hence a better uncertainty description will be obtained.

4. Control-Relevant Parameter Estimation

The parameter estimation problem consists of curvefitting the frequency response $\tilde{P}$ obtained from spectral analysis to a user-defined model structure. While very elementary approaches can be used for this purpose (such as estimating the model parameters directly from a Bode plot), a more elegant and useful approach is to incorporate the description of the control problem to define what features of the frequency response need to be conserved. In this manner, the most parsimonious model description (with respect to the control problem at hand) is obtained.

The problem of control-relevant parameter estimation can be viewed as equivalent to the control-relevant model reduction problem (Rivera and Morari, 1986a). The frequency-response $\tilde{P}$ is represents the "full-order" model, while $\hat{P}$, the parametric model, is the "reduced-order" model. The uncertainty description from Section 3, as well as designer requirements for the nominal and robust performance of the closed-loop system, are integrated to define a "control-relevant" weight.

The particular case considered in this paper is the robust performance of a closed-loop system with the plant uncertainty description (21) subject to output sensitivity requirements, represented by the weight functions W_{p_1} and W_{p_2}. The corresponding block diagram is shown in Figure 3. Theorem 2 is the re-statement of Theorem 1 for this problem:

Theorem 2: Robust Performance, Structured

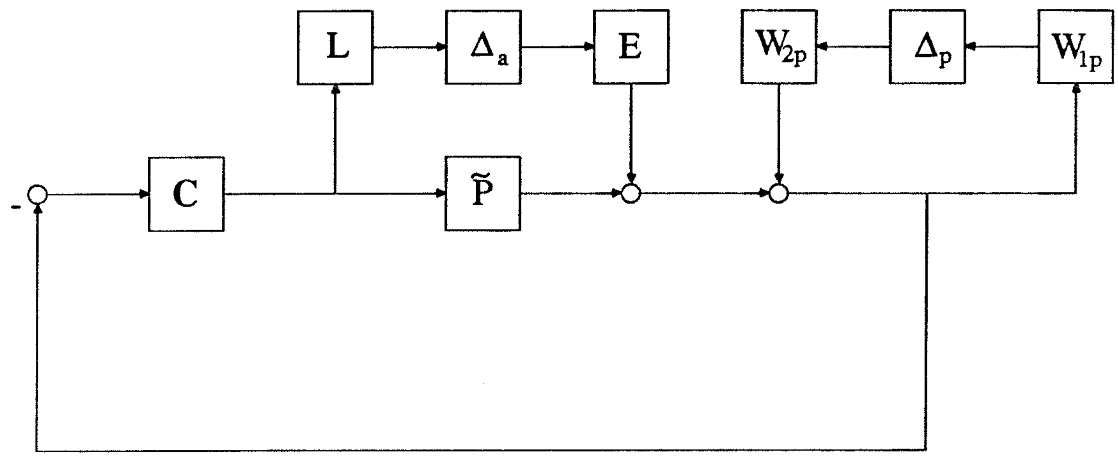


Figure 3 . Feedback structure for control-relevant identification.

For the set of perturbations defined by

$$\Delta = \mathrm{diag}(\Delta_a \ \ \Delta_p) \tag{25}$$

where Δ_p represents the performance Δ. The generalized plant G exhibits robust performance iff

$$||G||_\mu \leq 1 \tag{26}$$

where

$$G = \begin{bmatrix} -L\tilde{P}^{-1}(I+E_m\hat{H}(I+E_m\hat{H})^{-1}E) & -L\tilde{P}^{-1}(I+E_m)\hat{H}(I+E_m\hat{H})^{-1}W_{p_2} \\ W_{p_1}(I-\hat{H})(I+E_m\hat{H})^{-1}E & W_{p_1}(I-\hat{H})(I+E_m\hat{H})^{-1}W_{p_2} \end{bmatrix} \tag{27}$$

$\hat{H}$ and $I-\hat{H}$ are the nominal complementary sensitivity and sensitivity operators, respectively

$$I - \hat{H} = (I+\hat{P}C)^{-1} \tag{28}$$

$$\hat{H} = \hat{P}C(I+\hat{P}C)^{-1} \tag{29}$$

while E_m denotes the multiplicative error

$$E_m = (\tilde{P}-\hat{P})\hat{P}^{-1} \tag{30}$$

The model reduction problem corresponding to Theorem 2 is shown by Rivera and Morari (1986a) to be

$$\inf_{\hat{P}} \ ||D\begin{pmatrix} L\tilde{P}^{-1} \\ W_{p_1} \end{pmatrix} (I-\hat{H}) \ E_m\hat{H}[E \quad W_{p_2}] \ D^{-1}||_\infty \tag{31}$$

with D defined according to

$$\mathcal{D} = \mathrm{diag}(d_1I_q, d_2I_q, \ldots, d_rI_q, d_{r+1}I_q | d_i \in R^+ = (0,\infty)\} \tag{32}$$

$$D \ni \inf_{D\varepsilon\mathcal{D}} \ \bar{\sigma}(DMD^{-1}) \ , \ M = G(j\omega)$$

An iterative algorithm for solving (31) for a single-input, single-output plant is discussed by Rivera and Morari (1986b). For the general multivariable case, the solution to (31) with $\tilde{P}$ described by its frequency response remains a research issue. In such a case, one possible option is to obtain a high-order $\hat{P}$ using the general frequency-response matching methodologies discussed by Noldus and Decoster (1976) or Stahl (1984), followed then by the use of frequency-weighted balanced realization, as discussed by Rivera and Morari (1986a), to obtain a low-order $\hat{P}$.

5. Control-Relevant Model Validation

Control-relevant model validation involves ascertaining that $\hat{P}$ (obtained from whatever means) can be used to design a control system satisfying the user's robust performance requirements. From the discussion in Section 4, it becomes evident that μ is also useful as a model validation measure. For example, model validation for the problem described by Figure 3 is achieved by insuring that Theorem 2 is satisfied. Note that a convenient way of defining the complementary sensitivity operator $\hat{H}$ is through the formula

$$\hat{H} = \hat{P}_+ F \tag{33}$$

$\hat{P}_+$ contains all the nonminimum-phase behavior in $\hat{P}$, while F is a matrix of low-pass filter elements. Computing $\hat{H}$ according to (33) is discussed in Morari et al. (1987).

6. Summary and Conclusions

A control-relevant identification procedure, based on the robust control paradigm, has been formulated. The field of spectral time series analysis (Jenkins and Watts, 1969) and the control-relevant model reduction problem (Rivera and Morari, 1986a,b) have been utilized to obtain plant and uncertainty descriptions compatible with the structured singular value framework of Doyle (1984). The end result of this methodology is that parsimonious control structures satisfying desired performance levels can be achieved.

The treatment provided in this paper is by no means complete. Additional investigation on experimental design, uncertainty descriptions, and computational issues related to the parameter estimation problem is still warranted.

7. Acknowledgment

We are indebted to Prof. John Doyle for elucidating some of the issues discussed in this paper.

References

Aström, K. J. and P. Eykhoff. Automatica, 7, 123 (1971).

Briggs, P. A. N. and K. R. Godfrey. Proc. IEE, 113, 1259 (1966).

Doyle, J., "Structured Uncertainty in Control System Design". presented at the 24th Conference on Decision and Control, Ft. Lauderdale, Florida (1985).

Jenkins, G. and D. Watts. Spectral Analysis and its Applications, Holden-Day (1969).

Koopmans, L. H. The Spectral Analysis of Time Series, Academic Press (1974).

Ljung, L. "Frequency-Domain Properties of Identified Transfer Functions", Report LiTH-ISY-I-0668, Department of Electrical Engineering, Linkoping University, Sweden (1984).

Morari, M., E. Zafiriou and C. G. Economou. Robust Process Control, Springer-Verlag, to appear (1987).

Noldus, E. and M. Decoster. Regelungstechnik, 24, 122 (1976).

Rivera, D. E., M. Morari and S. Skogestad. Ind. Eng. Chem. Proc. Des. and Dev., 25, 252 (1986).

Rivera, D. E. and M. Morari, "Plant and Controller Reduction Problems for Robust Performance," submitted to the IEEE Trans. on Autom. Cntrol., November (1986a).

Rivera, D. E. and M. Morari, "Control-Relevant Model Reduction Problems for SISO H_2, H_∞, and μ-Controller Synthesis," International Journal of Control, in press (1986b).

Stahl, H. Int. J. Control, 39, 541 (1984).

CHAPTER VII: CONCLUSIONS

1. Summary

The synergism of modeling and control, as pertaining to the fields of low-order control design, model reduction, and model identification has been investigated. A summary of the main contributions of this thesis follows:

Model Reduction: Starting from the H_2 , H_∞ , and μ control objectives, control-relevant plant and controller reduction problems have been formulated. These are weighted optimization problems with weights that incorporate explicitly the control problem description and designer requirements for closed-loop performance. The efficient solution of these problems, using the method of balanced realizations as well as linear and quadratic programming, is outlined. Examples demonstrating the effectiveness of the control-relevant analysis include the control of a *methanation reactor and a* binary distillation column.

Low-Order Controller Design: The Internal Model Control design procedure was used to relate process models common to industrial practice to the PID controller. The IMC parametrization offers significant insights regarding the on-line adjustment of these controllers through the use of a single adjustable parameter which has a direct effect on the closed-loop bandwidth.

PID tuning rules for the first-order lag with deadtime model were developed. These were shown to be, for all practical purposes, equally performing to the optimal Smith Predictor.

The control-relevant model reduction problem was integrated with the IMC design procedure to provide a more general low-order controller design methodology. PID control of a high-order system, for example, can be achieved by using control-relevant reduction to obtain a model belonging to Table 1, Chapter II. The selection of the filter parameter is accomplished either through a simple search or from physical arguments.

Model Identification: Introductory efforts on a comprehensive, control-relevant identification methodology were presented. It was shown how the results of spectral time series analysis can be translated to meet the requirements of the robust control theory. The control-relevant reduction problem serves a dual purpose as a control-relevant parameter estimation problem, with the Structured Singular Value acting as a model validation measure for control purposes.

2. Applications of the Thesis

This work was partly motivated by the need to provide analysis and synthesis tools useful to the plant-level engineer. Along these lines, a number of computer programs have been written that implement the results of this thesis. The programs REDUCE, TUNE, REAL (BALANCE option), and SPECTRAL all form part, or will be forming part, of the CONSYD software package (Holt, et al., 1986). REDUCE implements the control-relevant reduction methodology outlined in Chapter III. The BALANCE subroutine in REAL allows the user to compute weighted balanced realizations and thus carry out the results of Chapter IV. TUNE allows the user to execute the low-order controller tuning strategy developed in Chapter V. The mu-optimal design option in TUNE has been incorporated as the main controller design routine for the prototype version of the expert system ROBEX (Heersink, 1986). A preliminary version of the program SPECTRAL, which performs spectral time series analysis, has been drafted.

The PID tuning rules, developed in Chapter II, have had a significant impact. At Shell Oil, these rules have become a favorite design tool among plant engineers (Garcia, 1986). Others have used them successfully in both simulated and actual plant environments (O'shima, 1984 ; Brambilla, 1984 ; Levien and Morari, 1985).

3. Suggestions for Future Research

It seems that an immutable characteristic of research is that new questions always arise to replace the old ones that were answered. As indicated in the conclusion segment of Chapter IV, such is the case with the model reduction problem. Improving the effectiveness of the computational techniques would be a welcome addition to the literature. One reasonable approach would be to develop a less restrictive multivariable generalization of the programming approach of Stahl (1984). The resulting methodology would also be useful for solving the control-relevant parameter estimation problem of Chapter VI, which was not addressed in this thesis.

The development of a comprehensive, control-relevant identification methodology merits a complete thesis. The control problem considered in Chapter VI addressed robust weighted sensitivity to additive uncertainty perturbations (under the assumption that the uncertainty and disturbance descriptions have been accurately obtained from separate experiments); the real problem at the plant, however, is more complicated: the uncertainty is more structured and not limited to the process but to the plant sensors and actuators as well. The multi-input, single-output uncertainty description presented should be generalized for multiple outputs. Furthermore, more research is required on the design of experiments, keeping in mind the restrictions imposed by industrial practice. It should also be established for what types of processes will identification be useful. Some systems, like ill-conditioned plants, might be inherently difficult to identify. Case studies performed on simulated and actual plants will also be necessary to test the effectiveness of the theory.

References

Brambilla, A. "Analysis of Control Systems for Atmospheric Crude Columns," seminar presented at the Department of Chemical Engineering, University of Wisconsin-Madison, May 1984.

Garcia, C.E. Shell Development Co., personal communication, (1986).

Heersink, R., "ROBEX: Expert System Technology In Process Control Design Systems," Master's Research Report, California Institute of Technology, (1986).

Holt, B.R., M. Morari, and W.H. Ray, "CONSYD - Integrated Software for Computer-Aided Control System Design and Analysis," Comp. and Chem. Eng., in press (1986).

Levien, K.L. and M. Morari, "Internal Model Control of a Coupled Distillation Column," Proceedings of the 1985 American Control Conference, Boston.

O'shima M. "Control Simulation of a Packed Bed Reactor," Summary Report, University of Wisconsin-Madison, (1984).

Stahl, H. Int. J. Cont., 39, 541 (1984).