Controller Design by Frequency Domain Approximation

Lilja, Mats

1989

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

Controller Design by Frequency Domain Approximation

Mats Lilja

Lund 1989

Controller Design by Frequency Domain Approximation

av

Mats Lilja
Civ. ing., Hk

Akademisk avhandling som för avläggande av teknisk doktorsexamen vid Tekniska fakulteten vid Universitetet i Lund kommer att offentligen försvaras i sal M:B, Maskinteknikhuset, Lunds Tekniska Högskola, fredagen den 27 oktober 1989, kl 10.15.

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

Document name
DOCTORAL DISSERTATION

Date of issue
August 1989

Document Number
CODEN: LUTFD2/(TFRT-1031)/1-107/(1989)

Author(s)
Mats Lilja

Supervisor
Karl Johan Åström

Sponsoring organisation
STU Contract 89-00403P

Title and subtitle
Controller Design by Frequency Domain Approximation

Abstract

Approximation is important in many areas of applied mathematics. It enters in a natural way in control systems design where a controller of a specified order is computed such that a certain closed loop performance or robustness is achieved. The least squares approximation method is emphasized because of its simplicity. In many cases only a few points on the process Nyquist curve are required in order to obtain good design results. Least squares approximation can also be used to shape the loop in such a way that open loop variations give acceptable variations in the closed loop. Different design trade-offs such as closed loop bandwidth vs. conroller complexity can conveniently be studied.

Key words
Approximation, Feedback control, Pole placement, Frequency domain, Least squares

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title

ISBN

Language
English

Number of pages
107

Recipient's notes

Security classification

To Kerstin and Susanna

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

Published 1989
Printed in Sweden
Studentlitteratur

Contents

1

Introduction

Control system design is complicated because it is a compromise between many different factors. Examples of factors that are important are

- Command signal following
- Load disturbances
- Measurement noise
- Model uncertainty
- Constraints
- Regulator complexity

There are few design methods that consider all these factors simultaneously. Design methods will typically focus on one or two issues. It is then necessary for the designer to check the other factors by analysis or simulation.

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

In the classical approach the main concern was to design a feedback compensator such that a certain stability margin was achieved. The emphasis was then on model uncertainties. Feedback was used to decrease sensitivity to disturbances and model errors. The compensator design was done mainly by graphical methods evolving from the Nyquist stability criterion. There was no a priori specification of the controller complexity and the compensator consisted of a number of cascaded lead-lag filters, successively determined from a Bode diagram of the loop frequency response. The graphical procedure gave good engineering insight but it was hard to computerize. The method also supplied little information

about the transient properties. The development of analytical design methods in the fifties made it possible to give specifications on the transient performance. At the same time, less attention was given to robustness and sensitivity issues. In the last two decades analytical methods has been developed in which robustness has regained its importance. In the analytical design methods a process model together with a closed loop specification is given. The design method then gives a controller, typically of order at least as high as the order of the process model. Choice of a suitable model complexity therefore becomes a key design issue when using analytical design methods.

In this thesis we have considered design problems where a process model in terms of a frequency response is available. There are good experimental methods to measure frequency responses (see [Åström, 1975]). In the last decade an increasing interest has been given to the problem on how to reduce the controller order while at the same time maintaining a good closed loop performance. In Fig. 1.1 the different paths leading to a low order controller are shown.

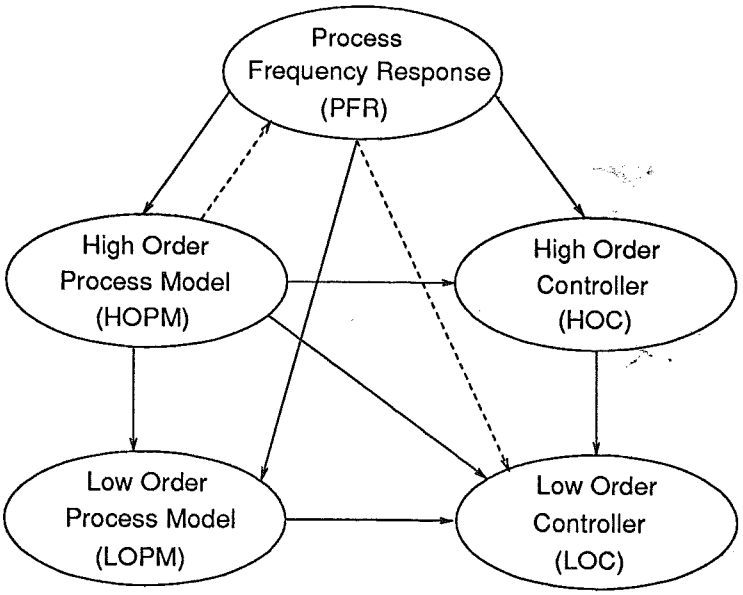


Figure 1.1 Approaches to control system design.

The most important of these are

A (PFR) ⇒ (HOPM) ⇒ (LOPM) ⇒ (LOC)
B (PFR) ⇒ (HOPM) ⇒ (HOC) ⇒ (LOC)
C (PFR) ⇒ (LOPM) ⇒ (LOC)
D (PFR) ⇒ (LOC)

In method A the process model is reduced before the controller design is done while the model reduction is performed after the controller design step in method B. In methods C and D no intermediate high order models are used. Method

D corresponds to a design where the low order controller is computed directly from the given process frequency response, like in the classical design methods. Assuming that instead of having the process represented by a frequency response, a high order process model has been given. This gives the two modified design approaches

C′ (HOPM) ⇒ (PFR) ⇒ (LOPM) ⇒ (LOC)
D′ (HOPM) ⇒ (PFR) ⇒ (LOC)

Many of these steps clearly involve approximation. In Chapter 2 some different approximation methods are reviewed. The emphasis is put on the discrete least squares method and the optimal Hankel norm method (weighted and unweighted). The Hankel norm method is used in the approximation step (HOPM) ⇒ (LOPM) while the least squares method operates on a frequency response, (PFR) ⇒ (LOPM). These approximation methods are applied to some different high order process models. A new way to compute the optimal Hankel norm approximation of a rational SISO transfer function is proposed in Chapter 2. A new algorithm for the classical problem of approximating a transfer function by a rational transfer function with ime delay, is also presented. Chapter 3 contains a description of the pole placement design method. Some of the weaknesses of this method are pinpointed by some design examples. Combining some of the features of pole placement with least squares approximation in the frequency domain gives a new design method which is presented in Chapter 4. This method follows the scheme D′ indicated by dashed arrows in Fig. 1.1. Some of its properties are:

- It resembles pole placement in the respect that a desired closed loop model is specified.
- The method uses only the frequency response at a finite number of frequencies. These frequencies have to be chosen by the user.
- The order of the controller is relatively independent of the order of the desired closed loop model.
- The approximation can be done at any self-conjugated set in the complex plane, which means that there is no principal difference between discrete time design (approximation on the unit circle) and continuous time design (approximation on the imaginary axis).
- There are no problems with common factors in the polynomials $B(s)$ and $A(s)$ of the transfer function $G(s) = B(s)/A(s)$ since only the values $G(i\omega)$ are used in the design calculations.

Practically all design methods have design variables that the user chooses to influence the trade-offs in a design. For example, in LQG the design variables are the weighting matrices characterizing the loss function. In our case the major design variables are given by

- A set of approximation frequencies Ω.
- A desired closed loop model $G_m(s)$.
- An observer polynomial $A_o(s)$.
- The order of the feedback compensator

The fact that frequencies are chosen establishes a strong link to the basic ideas of classical control. In Chapter 5 least squares fitting is used to shape the loop frequency response in a desirable way. This is inspired by the approach of Bode. It is demonstrated that open loop variations described by one parameter need not be eliminated by high gain feedback. Instead they can be turned into closed loop variations of an acceptable nature such as invariance of the damping in the step response. If an invariant response is desired it can be obtained using a two degree of freedom structure. This will, however, often lead to slower responses to command signals. Chapter 6 contains some final conclusions and suggestions for future research.

Preface

Most of the results in this thesis are of empirical nature and has been developed in the matrix language PRO-MATLAB. The work is motivated by the need of control synthesis methods which give insight into the relationship between process model order, closed loop specifications and controller complexity. One goal has been to find a design methodology which avoids large and demanding optimization procedures and which operates directly on frequency responses. This leads naturally to the use of least squares approximation. A collection of MATLAB functions has been written and used as a toolbox for design of low order controllers.

The thesis contains a review of some different approximation methods and of the pole placement method. The reader is assumed to have basic knowledge of automatic control and complex analysis.

Acknowledgements

First of all, I would like to thank my supervisor Professor Karl Johan Åström who suggested the thesis subject. He has been a persistent source of encouragement and enthusiasm. I would also like to thank all my colleagues at the Department. Special thanks go to Kjell Gustafsson who has given many valuable comments on the manuscript and with whom I have had many interesting discussions. I am also indebted to Per Hagander who has proof-read Chapter 2. The excellent computer resources at the Department are supported by Leif Andersson. He has written the TEXmacros which made the type-setting of this thesis much smoother than it would have been else. The thesis has been partially supported by the Swedish Board of Technical Development under contract 89-00403P.

2

Approximation of Transfer Functions

2.1 Introduction

In this chapter approximation of transfer functions will be discussed. Given a transfer function $G(s)$ the problem is to find a transfer function $\hat{G}(s)$ which in some sense is close to $G(s)$. The approximation $\hat{G}(s)$ is in most cases taken to be a rational function. The reason for this is that much analytic control theory deals with systems described by rational transfer functions.

In section 2.2 a short overview of different approximation methods is presented. Section 2.3 is dealing with least squares approximation in the equation error formulation [Lévy, 1959] and section 2.4 contains some facts about Hankel norm approximation together with a new way of computing the Hankel norm approximation of a given degree for a rational function. A non-linear in the error least squares method is presented in section 2.5.

2.2 Preliminaries

A transfer function $G(s)$ is said to be stable if all the poles of $G(s)$ are located in the left half plane. A function $F(s)$ is said to be rational of order (m, n) if it can be written as

$$F(s) = \frac{x_0 s^m + x_1 s^{m-1} + \cdots + x_m}{s^n + y_1 s^{n-1} + \cdots + y_n}$$

Let $G(s)$ be a given stable transfer function for which an approximation $\hat{G}(s)$ is to be found. There are several ways of measuring how good the approximation $\hat{G}$ is. A common type of distance function is the metric d induced by a norm $\|\cdot\|$:

$$d(G, \hat{G}) = \|G - \hat{G}\|$$

An often used norm is the L^∞ norm (Tchebycheff norm):

$$\|G - \hat{G}\|_\infty = \sup_{\omega>0} |G(i\omega) - \hat{G}(i\omega)|$$

Finding a fixed order rational approximation in this norm leads to optimization problems which require iterative techniques. The Hankel norm is much easier to calculate. It is related to the Tchebycheff norm by

$$\|G\|_H = \inf \|G - F\|_\infty$$

where G is a stable rational function and the infimum is taken over all functions F bounded on the imaginary axis and having all its poles in the right half plane. This characterization of the Hankel norm immediately yields the relationship

$$\|G\|_H \leq \|G\|_\infty$$

for all stable rational functions G. Another popular norm is the L^2 norm (on the imaginary axis) defined by

$$\|G\|_2^2 = \int_0^\infty |G(i\omega)|^2 d\omega$$

All norms have weighted versions defined by

$$\|G\|_W = \|WG\|$$

for any norm $\|\cdot\|$ and weighting function $W(i\omega)$. Using the L^2 norm formally with the special weighting

$$W(i\omega) = \sum_{k=1}^{N} \delta(\omega - \omega_k)$$

for a finite set of frequencies $\{\omega_1, \ldots, \omega_N\}$ gives a discrete seminorm

$$\|G\|^2 := \sum_{k=1}^{N} |G(i\omega_k)|^2$$

A seminorm $\|\cdot\|$ differs from a norm in that there exist nonzero functions f such that $\|f\| = 0$. All approximants $\hat{G}$ fulfilling the interpolation condition

$\hat{G}(i\omega_k) = G(i\omega_k)$, $k = 1, \ldots, N$ are those for which $\|G - \hat{G}\| = 0$. Minimizing the error $G - \hat{G}$ in this seminorm corresponds to least squares fitting of the parameters

$$\theta = (\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n, \hat{b}_1, \ldots, \hat{b}_n)^T \tag{2.1}$$

in a rational approximant

$$\hat{G}(s) = \frac{\hat{B}(s)}{\hat{A}(s)} = \frac{\hat{b}_1 s^{n-1} + \hat{b}_2 s^{n-2} + \cdots \hat{b}_n}{s^n + \hat{a}_1 s^{n-1} + \hat{a}_2 s^{n-2} + \cdots + \hat{a}_n} \tag{2.2}$$

One advantage of using the least squares method in approximation is that the computation effort is independent of what type of function G is (non-rational etc) since only the function values at the N points $\{i\omega_1, \ldots, i\omega_N\}$ have to be known. The approximation problem can be written as

$$\min_\theta J_1(\theta) = \min_\theta \sum_{k=1}^{N} |G(i\omega_k) - \hat{G}(i\omega_k)|^2 = \sum_{k=1}^{N} \left| G(i\omega_k) - \frac{\hat{B}(i\omega_k)}{\hat{A}(i\omega_k)} \right|^2 \tag{2.3}$$

where θ is given by Eq. 2.1. This gives, however, a least squares criterion that is non-linear in the error. By using the equation error $\hat{B} - G\hat{A}$, we get the loss function [Levy, 1959]:

$$J(\theta) = \sum_{k=1}^{N} W_k^2 |\hat{B}(i\omega_k) - \hat{A}(i\omega_k) G(i\omega_k)|^2 \tag{2.4}$$

instead. This formulation leads to a drastic simplification because the error then is linear in θ. The weightings W_k^2 are used to de-emphasize high frequencies and to emphasize certain frequencies.

2.3 Least squares fitting in frequency domain

Model without time delay

Assume that the frequency response for the process is given at some frequencies

$$G(i\omega_k), \quad k = 1, 2, \ldots, N$$

The problem is to find a rational approximation $\hat{G}(s)$ of the form (2.2) where $2n \leq N$ which minimizes a certain loss function J. As mentioned in Section 2.2, a particularly convenient method is to use the equation error (or "linear-in-the-error") least squares criterion (2.4). In Eq. 2.2 the approximant is of relative degree one but in general, the degrees of the polynomials $\hat{B}$ and $\hat{A}$ may, of course,

be chosen independently. A common restriction is, however, that $\deg \hat{B} \leq \deg \hat{A}$ (i.e. $\hat{G}$ is proper). The minimization of J has an explicit solution. This gives a simple optimization problem as opposed to the minimization (2.3). which has to be solved by optimization methods requiring substantial computation effort. The weightings W_k^2 are often chosen as $|W(i\omega_k)|^2$ for some rational function $W(s)$. This function should decrease with frequency in order to avoid large weighting of high frequencies. The fitting will else deteriorate at lower frequencies at the expense of getting a good fitting at higher frequencies. Another possibility is to use an iterative scheme where the weight function W is updated after each iteration:

ALGORITHM 2.1—[Sanathan and Koerner, 1963]

1 Make a least squares fitting using some initial weighting $W^{[1]}(s)$ giving the rational transfer function

$$\frac{\hat{B}^{[1]}(s)}{\hat{A}^{[1]}(s)}$$

2 Update the weight function according to:

$$W^{[k+1]}(s) = 1/\hat{A}^{[k]}(s)$$

and compute the next least squares approximation

$$\frac{\hat{B}^{[k+1]}(s)}{\hat{A}^{[k+1]}(s)}$$

Other approaches using this technique, but with different updating, are found in [Lawrence and Rogers, 1979] and [Stahl, 1984]. The initial weight function could sometimes be chosen as $W^{[1]} \equiv 1$ but this may give a too bad $\hat{A}^{[1]}(s)$. A better choice of initial weighting is some a priori estimate of $1/\hat{A}(s)$. Presently, there are no convergence proofs available for any of the iterative methods in [Sanathan and Koerner, 1963], [Lawrence and Rogers, 1979] and [Stahl, 1984].

Let $\varepsilon(s)$ denote the transfer function error $G(s) - \hat{G}(s)$. To get an error amplitude $|\varepsilon(i\omega)|$ which is almost constant with respect to ω a modified version of the above iterative method can be used.

1' Iterate according to Sanathan-Koerner until (if possible) $\hat{G}$ converges (pointwise at the chosen frequencies).

2' Make one more iteration but this last time with the updating $W^{[k+1]}(s) = \varepsilon^{[k]}(s)/\hat{A}^{[k]}(s)$.

This heuristic method will be applied in one of the examples later in this chapter.

Continuing with the analysis, we introduce the following notation

$$\phi = \begin{pmatrix} (i\omega_1)^{n-1} & (i\omega_1)^{n-2} & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ (i\omega_N)^{n-1} & (i\omega_N)^{n-2} & \cdots & 1 \end{pmatrix}$$

and

$$\psi = \Gamma \begin{pmatrix} (i\omega_1)^n \\ \vdots \\ (i\omega_N)^n \end{pmatrix} \qquad \Phi = \begin{pmatrix} -\Gamma\phi & \phi \end{pmatrix}$$

where $\Gamma = \mathrm{diag}(\{G(i\omega_k)\}_{k=1}^N)$. The loss function in Eq. 2.4 can now be written as

$$J(\theta) = (\Phi\theta - \psi)^* Q(\Phi\theta - \psi)$$

where $*$ denotes conjugate transpose and $Q = \mathrm{diag}\{|W(i\omega_k)|^2\}_{k=1}^N$ is the weighting matrix. The explicit solution to this problem is given by

$$\hat{\theta} = (\Phi^* Q \Phi)^{-1} \Phi^* Q \psi \tag{2.5}$$

Numerically, however, it is more preferable to use QR factorization or singular value decomposition [Golub and van Loan, 1987].

The points of approximation need not necessarily be located on the imaginary axis. An arbitrary point set $\mathcal{Z}$ in the complex plane could be chosen as the approximation set. Since all transfer functions $F(s)$ considered here, are assumed to have the property

$$\overline{F}(s) = F(\overline{s}) \tag{2.6}$$

(equivalent to $F(s)$ having a real inverse Laplace transform), the set $\mathcal{Z}$ must be closed under conjugation. This also implies that the coefficients in the approximating transfer function, will be real as the following lemma states.

LEMMA 2.1
If the approximation set $\mathcal{Z}$ is self conjugate and the weight function W has the property $|W(z)| = |W(\overline{z})|$ then the solution $\hat{\theta}$ in Eq. 2.6 is real.

Proof: Let

$$\Phi = \begin{pmatrix} -G(z_1)z_1^{n-1} & \cdots & -G(z_1)z_1 & -G(z_1) & z_1^{n-1} & \cdots z_1 & 1 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ -G(z_N)z_N^{n-1} & \cdots & -G(z_N)z_N & -G(z_N) & z_N^{n-1} & \cdots z_N & 1 \end{pmatrix}$$

$$\psi = \begin{pmatrix} -G(z_1)z_1^n & \cdots & -G(z_N)z_N^n \end{pmatrix}^T$$

where $\mathcal{Z} = \{z_1, \ldots, z_N\}$. Assuming that $\mathcal{Z}$ is closed under conjugation we have that $\overline{\Phi} = \Pi\Phi$ and $\overline{\psi} = \Pi\psi$ for some $N \times N$ permutation matrix Π. The assumptions $Q = \text{diag}\{|W(z)|^2\}_{k=1}^N$ in 2.5 and $|W(z)| = |W(\overline{z})|$ imply that $\Pi Q = Q\Pi$. Furthermore, since Π is a permutation matrix, it is orthogonal. Using these facts together with 2.5 gives

$$\overline{\hat{\theta}} = (\overline{\Phi}^* Q \overline{\Phi})^{-1} \overline{\Phi}^* Q \overline{\psi} = (\Phi^* \Pi^T Q \Pi \Phi)^{-1} \Phi^* \Pi^T Q \Pi \psi = (\Phi^* Q \Phi)^{-1} \Phi^* Q \psi = \hat{\theta}$$

which is the desired result. ∎

Remark. The result in Lemma 2.1 can be viewed in the following way. Since both G and $\hat{G}$ have the property (2.6) the error $\varepsilon = G - \hat{G}$ will also have that property. If both z and $\overline{z}$ belong to the approximation set $\mathcal{Z}$ the corresponding pair of errors can be rewritten according to

$$\begin{cases} \varepsilon(z) = G(z) - \hat{G}(z) \\ \varepsilon(\overline{z}) = G(\overline{z}) - \hat{G}(\overline{z}) \end{cases} \iff \begin{cases} \text{Re}\,\varepsilon(z) = \text{Re}(G(z) - \hat{G}(z)) \\ \text{Im}\,\varepsilon(z) = \text{Im}(G(z) - \hat{G}(z)) \end{cases}$$

which shows that the complex least squares problem can be solved as a real least squares problem. □

The most common choice of approximation set $\mathcal{Z}$ is points on the imaginary axis. One reason for this is that it is easy to measure the frequency response on the imaginary axis. Another motivation is that the stability boundary is often chosen to coincide with the imaginary axis. In the following sections, $\Omega = \{\omega_1, \ldots, \omega_m\}$, $\omega_j > 0$ will refer to the approximation set $\mathcal{Z} = \{i\omega_1, \ldots, i\omega_m, -i\omega_1, \ldots, -i\omega_m\}$.

Connections with Padé approximations

A Padé approximation at a point ζ could be thought of as a rational function analogue for Taylor polynomials [Cheney, 1966]. Given the power series of a function $G(s)$ at a point s_1

$$G(s) = \sum_{j=0}^{\infty} g_j (s - s_1)^j$$

it is possible to find a rational function $\hat{G}(s) = \hat{B}(s)/\hat{A}(s)$ of fixed order such that $|G(s) - \hat{G}(s)| \leq M|s - s_1|^m$ for some positive integers m and M and in some disc $|s - s_1| < \rho$ with $\rho > 0$. One method to compute the coefficients in the rational function is to multiply both sides of the equality $G(s) = \hat{G}(s) + \text{terms of order} \geq m$ by $\hat{A}(s)$ and identify coefficients of terms of order less than $m-1$ on both sides. Another method is to use continued fraction expansions, that is, expansions of

type

$$a_0 + \cfrac{1}{a_1 + \cfrac{s}{a_2 + \cfrac{s}{a_3 + \cfrac{s}{a_4}}}}$$

The latter method may have some computational advantages but will not be treated here.

There is also a generalization to multi-point Padé approximations [Hwang and Chen, 1987] i.e. the function to be approximated is given as its power series expansions at different points s_i, $i = 1, 2, \ldots, n$:

$$G(s) = \sum_{j=0}^{\infty} g_{i,j}(s - s_i)^j, \qquad i = 1, 2, \ldots, n$$

and the rational approximant fulfills the condition

$$\hat{G}(s) = \sum_{j=0}^{m_i - 1} g_{i,j}(s - s_i)^j + O((s - s_i)_i^m), \qquad i = 1, 2, \ldots, n \tag{2.7}$$

The special case of $m_i = 1, \quad i = 1, 2, \ldots, n$ corresponds to interpolation of the function $G(s)$ with a rational function $\hat{G}(s)$ at the points $s_i, \quad i = 1, 2, \ldots, n$. The general multi-point Padé approximation problem ($m_i \geq 1$) can be approximated by the ordinary interpolation problem ($m_i = 1$) by instead of using the derivatives $G^{(k)}(s), \quad k = 1, \ldots, m_i$ taking $m_i + 1$ points in a neighborhood of s_i but the smaller neighborhood, the more badly conditioned problem is obtained. The use of QR factorization together with pivoting can, however, yield good approximations in this way.

EXAMPLE 2.1—[Hwang and Chen, 1987]
Consider the problem of approximating the non-rational function

$$G(s) = e^{-\sqrt{s}}$$

with a multipoint Pad'e approximation of order $(1, 4)$ at $s_i = i - 1$ with $m_i = i$ for $i = 1, 2, 3$. The corresponding interpolation problem is to choose $\mathcal{Z} = \{0, 1, 1 + \epsilon, 2 - \epsilon, 2, 2 + \epsilon\}$ with a small value of $\epsilon > 0$. With $\epsilon = 0.001$ the result is

$$\hat{G}(s) = \frac{139.474s + 26.11}{s^4 + 6.15312s^3 + 164.882s^2 + 251.959s + 26.11}$$

to be compared with

$$\hat{G}^{\text{Pade}'}(s) = \frac{139.437s + 26.09751}{s^4 + 6.14955s^3 + 164.84275s^2 + 251.87964s + 26.09751}$$

The multi-point Padé approximation can be further generalized by solving a least squares problem, namely minimization of the function

$$J = \sum_{i=1}^{n} \sum_{j=1}^{m_i} |\hat{G}^{(j)}(s_i) - G^{(j)}(s_i)|^2 \tag{2.8}$$

over all rational functions of some order (k, ℓ). This establishes a connection between the multi point Padé approximation and the least squares approximation (Eq. 2.3) in that they corresponds to the two special cases $\sum_i m_i = k + \ell$ and $m_i = 1$, $i = 1, 2, \ldots, n$ respectively. □

Model with time delay

For a process model including a time delay

$$\hat{G}(s) = \frac{\hat{B}(s)}{\hat{A}(s)} e^{-\hat{\tau}s} = \frac{\hat{b}_1 s^{n-1} + \hat{b}_2 s^{n-2} + \cdots \hat{b}_n}{s^n + \hat{a}_1 s^{n-1} + \hat{a}_2 s^{n-2} + \cdots + \hat{a}_n} e^{-\hat{\tau}s}$$

the Φ-matrix in will be dependent of the time delay $\hat{\tau}$. Let $D(\tau)$ denote the matrix

$$D(\tau) = \operatorname{diag}\big(\{\exp(-i\omega_k \tau)\}_{k=1}^{N}\big)$$

The modified Φ-matrix can then be written as

$$\Phi(\tau) = \begin{pmatrix} -\Gamma\phi & D(\tau)\phi \end{pmatrix} \tag{2.9}$$

and the corresponding loss function with unity weighting to simplify notation.

$$J(\theta, \tau) = \sum_{k=1}^{N} |W(i\omega_k)|^2 |\hat{A}(i\omega_k) G(i\omega_k) - \hat{B}(i\omega_k)|^2 = \big|\Phi(\tau)\theta - \psi\big|^2$$

This gives a non-quadratic minimization problem in $2n + 1$ variables θ and τ. It turns out, however, that the problem can be separated into a quadratic minimization in θ and a non-quadratic minimization in the time delay τ. The scalar minimization in τ is carried out first and then θ is found by the procedure described in previous section.

THEOREM 2.1
Let $\Phi^*\Phi$ be non-singular and let P denote the matrix valued function

$$P(\tau) = I - \Phi(\tau)\big(\Phi(\tau)^*\Phi(\tau)\big)^{-1}\Phi(\tau)^*$$

Every local minimum of J with respect to τ and θ is given by

$$\hat{\theta} = \Theta(\hat{\tau})$$

where $\hat{\tau}$ is a local minimum of the function f defined by

$$f(\tau) := J\big(\Theta(\tau), \tau\big) = \psi^* P(\tau)\psi$$

where

$$\begin{aligned}\Theta(\tau) &:= \min_{\theta} J(\theta, \tau) \\ &= \big(\Phi^*(\tau)\Phi(\tau)\big)^{-1}\Phi(\tau)^*\psi\end{aligned}$$

Proof: A necessary condition for J to have a local minimum at $\tau = \hat{\tau}$, $\theta = \hat{\theta}$ is

$$\left(\frac{\partial J}{\partial \theta} \quad \frac{\partial J}{\partial \tau} \right)_{\theta=\hat{\theta},\tau=\hat{\tau}} = 0$$

For each fixed τ, there exists a unique solution to the quadratic minimization problem in θ, given by $\Theta(\tau)$. Computing the first derivative of f, with respect to τ gives

$$\begin{aligned}\frac{df}{d\tau} &= \left.\frac{\partial J}{\partial \tau}\right|_{\theta=\Theta(\tau)} + \left.\frac{\partial J}{\partial \theta}\right|_{\theta=\Theta(\tau)} \frac{d\Theta}{d\tau} \\ &= \left.\frac{\partial J}{\partial \tau}\right|_{\theta=\Theta(\tau)}\end{aligned}$$

where the last step follows from the fact that $\frac{\partial J}{\partial \theta} \equiv 0$ on the curve $(\theta, \tau) = (\Theta(\tau), \tau)$. This shows that any stationary point of J, with respect to θ and τ, corresponds to a stationary point of f with respect to τ. If a stationary point $(\hat{\theta}, \hat{\tau})$ to $J(\theta, \tau)$ in fact is a local minimum, then J has a local minimum on every curve $(\theta(v), \tau(v))$, where v is a scalar parameter, passing through $(\hat{\theta}, \hat{\tau})$. One such curve is $(\theta, \tau) = (\Theta(v), v)$ which implies that f has a local minimum at $(\hat{\theta}, \hat{\tau})$. ∎

An interesting observation is that the scalar function $f(\tau)$, and thus $J(\theta, \tau)$ will actually have several local minima. In case $G(s)$ has no poles in the closed right half plane the interesting minimum occurs for the smallest τ larger than zero. If the choice of frequencies and weighting is reasonable this minimum will give a stable transfer function. Local minima at larger values of τ corresponds often to unstable transfer functions.

The minimization of f with respect to τ can be done for example by the following modified Newton-Raphson algorithm:

ALGORITHM 2.2
Update the time delay τ according to

$$\tau_{n+1} = \tau_n - \frac{\frac{df}{d\tau}(\tau_n)}{\alpha\left|\frac{d^2f}{d\tau^2}(\tau_n)\right| + (1-\alpha)\frac{d^2f}{d\tau^2}(\tau_n)}$$

where $0.5 < \alpha < 1$. The modification is done in order to make local maxima repulsive and local minima attractive. Values of α less than one are chosen to get a stronger "repulsion" from a maximum. However, if the initial value of τ is chosen exactly at a maximum, the algorithm will of course get stuck. Since ψ is independent of τ, the first and second derivatives of f are given by

$$\frac{df}{d\tau} = \psi^* \frac{dP}{d\tau}\psi$$
$$\frac{d^2f}{d\tau^2} = \psi^* \frac{d^2P}{d\tau^2}\psi$$

Recalling the definitions of P,Φ and D,

$$D(\tau) = \operatorname{diag}\big(\{\exp(-i\omega_k\tau)\}_{k=1}^N\big)$$
$$\Phi(\tau) = \begin{pmatrix} -\Gamma\phi & D(\tau)\phi \end{pmatrix}$$
$$P(\tau) = I - \Phi(\tau)\big(\Phi(\tau)^*\Phi(\tau)\big)^{-1}\Phi(\tau)^*$$

the first derivative of P is computed according to

$$\frac{dD}{d\tau} = \operatorname{diag}\big(\{-i\omega_k \exp(-i\omega_k\tau)\}_{k=1}^N\big)$$
$$\frac{d\Phi}{d\tau} = \begin{pmatrix} 0 & \frac{dD}{d\tau}\phi \end{pmatrix}$$
$$\frac{dP}{d\tau} = Q + Q^* \qquad Q := -P\frac{d\Phi}{d\tau}\big(\Phi^*\Phi\big)^{-1}\Phi^*$$

and the second derivative of P is given by

$$\frac{d^2D}{d\tau^2} = \operatorname{diag}\big(\{-\omega_k^2 \exp(-i\omega_k\tau)\}_{k=1}^N\big)$$
$$\frac{d^2\Phi}{d\tau^2} = \begin{pmatrix} 0 & \frac{d^2D}{d\tau^2}\phi \end{pmatrix}$$
$$P_1 := \frac{d\Phi}{d\tau}\big(\Phi^*\Phi\big)^{-1}\Phi^*$$
$$P_2 := \frac{d^2\Phi}{d\tau^2}\big(\Phi^*\Phi\big)^{-1}\Phi^*$$
$$\frac{d^2P}{d\tau^2} = \frac{dQ}{d\tau} + \frac{dQ^*}{d\tau} \qquad \frac{dQ}{d\tau} = P(2P_1^2 - P_2) + Q^*Q + QQ^*$$

The implementation of the algorithm has been done in the matrix manipulation package PRO-MATLAB.

EXAMPLE 2.2—Eight cascaded low pass filters
Many finite dimensional system of high order are well described by a low order system followed by a time delay. One such system is

$$G(s) = \frac{1}{(s+1)^8}$$

A first order transfer function with time delay is found by using Algorithm 2.2 with a initial value $\tau = 0$ and with $\Omega = \{0.01, 0.1, 0.2, 0.4\}$. The iteration is stopped when the absolute difference between two consecutive values of τ is less than a specified value $\delta_{\min}$. In this case the algorithm converges after four iterations when $\delta_{\min} = 0.01$. The result is

$$\hat{G}(s) = \frac{0.2838}{s + 0.2824} e^{-5.105s}$$

The Nyquist curves of $\hat{G}(s)$, and $G(s)$ are shown in Fig. 2.1.

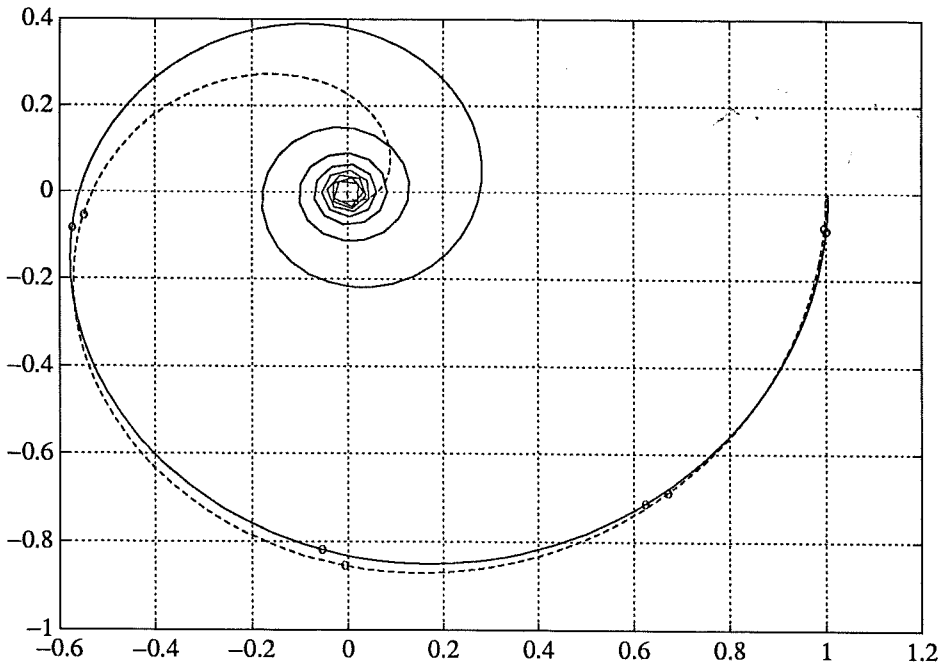


Figure 2.1 Nyquist curves of $G(s) = (s+1)^{-8}$ (dashed curve) and a time delayed first order approximation $\hat{G}(s)$ (solid curve). The approximation frequencies $0.01, 0.1, 0.2$ and 0.4 are marked.

A fairly good approximation is obtained below 0.4 rad/s. To get a better approximation without time delay a third order model is required. □

2.4 Hankel norm approximation

The Hankel Operator and the Hankel Norm

Consider the linear, time-invariant system

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$$

with transfer function $G(s) = C(sI - A)^{-1}B$. Assuming that the eigenvalues of A are in the open left half plane we can define the Hankel operator corresponding to this system by the operator $\Gamma_G : L_2(0,\infty) \to L_2(0,\infty)$ given by

$$(\Gamma_G v)(t) = \int_0^\infty Ce^{A(t+\tau)}Bv(\tau)d\tau$$

One interpretation of this operator is that if the input $u(t) = v(-t)$ for $t < 0$ and $u(t) = 0$ for $t > 0$ is applied to the system then the output for $t > 0$ will be $(\Gamma_G v)(t)$. The adjoint operator Γ_G^* is given by

$$(\Gamma_G^* y)(t) = \int_0^\infty B^* e^{A^*(t+\tau)}C^* y(\tau)d\tau$$

The Hankel singular values are defined as the singular values of the Hankel operator Γ_G, i.e. the eigenvalues of the operator $\Gamma_G^*\Gamma_G$. By introducing the controllability gramian P and the observability gramian Q defined by

$$P := \int_0^\infty e^{At}BB^*e^{A^*t}dt$$

$$Q := \int_0^\infty e^{A^*t}C^*Ce^{At}dt$$

it can be shown [Glover, 1984] that the Hankel singular values can be written as

$$\sigma_i(\Gamma_G) = \sqrt{\lambda_i(PQ)}, \quad i = 1, 2, \ldots, n$$

As a convention the σ_i's are sorted according to decreasing magnitude, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. The gramians can be computed by solving the Lyapunov equations

$$AP + PA^* + BB^* = 0$$
$$A^*Q + QA + C^*C = 0$$

The Hankel norm $\|\cdot\|_H$ is defined as the largest Hankel singular value. In [Glover, 1984] it is shown that the Hankel norm of a function G with all poles in the open left half plane, can be characterized by

$$\|G\|_H = \inf \|G - F\|_\infty$$

where the infimum is taken over all functions F having all poles in the right half plane. This implies that

$$\|G\|_H \leq \|G\|_\infty \tag{2.10}$$

giving a lower bound on the Tchebycheff norm of G.

Remark. In discrete time, the L_2 norm of a transfer function $H(z)$ is

$$\|H\|_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |H(e^{i\omega})|^2 d\omega$$

For causal functions H the inequality in Eq. 2.10is extended to [Kung and Lin, 1981]

$$\|H\|_2 \leq \|H\|_H \leq \|H\|_\infty$$

Approximation in the Hankel Norm

Let R_n denote the set of all proper real-rational functions of order n, i.e. functions of the form $r = p/q$ where p and q are polynomials with $\deg p \leq \deg q \leq n$ and with real coefficients. Introduce the Lesbesgue space L_∞ of all complex-valued functions bounded on the imaginary axis. Furthermore, introduce the Hardy space H_∞ of all functions F analytic in the right half plane with

$$\sup_{\mathrm{Re}\, s>0} |F(s)| < \infty$$

Similarly, let $\hat{H}_\infty$ denote all functions F analytic in the left half plane with

$$\sup_{\mathrm{Re}\, s<0} |F(s)| < \infty$$

The stable projection of a function $G \in L_\infty$ is defined as the stable part $G_- \in H_\infty$ of the decomposition

$$G(s) = G_-(s) + G_+(s)$$

and will be denoted by $G_- = \Pi_{H_\infty} G$. Given a stable, proper, rational transfer function $G(s)$, the goal is to find a stable approximant $\hat{G}(s) \in R_n$ such that the Hankel norm of the error $\|G - \hat{G}\|_H$ is minimized. To find the solution we need the following

THEOREM 2.2—[Adamjan, Arov, and Krein, 1971]
Let $\hat{H}_\infty^{[n]}$ denote all functions, bounded on the imaginary axis, which can be written as $H(s) = \hat{G}(s) + F(s)$, where $\hat{G} \in R_k \cap H_\infty$ and $F \in \hat{H}_\infty$. Given a function $G \in L_\infty$, the infimum

$$\inf_{H \in H_\infty^{[k]}} \|G - H\|_\infty$$

is attained by the function H given by the unique solution to

$$G(s) - H(s) = \sigma_k(\Gamma_G) E_k(s) \tag{2.11}$$

where $|E_k(i\omega)| = 1$, $\omega \in \mathbb{R}$. In more detail, $E_k(i\omega) = \overline{\hat{\xi}}(i\omega)/\hat{\xi}(i\omega)$, where $\hat{\xi}(s)$ is the Laplace transform of $\xi(t) \in L_2(0,\infty)$ and $\xi(t)$ satisfies the equation $\Gamma_G \xi = \sigma_k(\Gamma_G)\overline{\xi}$. The pair of functions $(\xi, \overline{\xi})$ is called a Schmidt pair.

LEMMA 2.2— Let the polynomials A, B, X and Y be defined by

$$\begin{aligned} A(s) &= s^n + a_1 s^{n-1} + \cdots + a_n \\ B(s) &= b_0 s^n + b_1 s^{n-1} + \cdots + b_n \\ X(s) &= s^{n-1} + x_1 s^{n-2} + \cdots + x_{n-1} \\ Y(s) &= y_0 s^{n-1} + y_2 s^{n-2} + \cdots + y_{n-1} \end{aligned}$$

Consider the polynomial equation

$$A(s)Y(s) - B(s)X(s) = \lambda A^*(s)X^*(s) \tag{2.12}$$

where the notation $P^*(s) = P(-s)$ is used and the polynomials X and Y are unknown. The solutions to the Eq. 2.12 are found via the eigenvectors $\{v_1, \ldots, v_n\}$ corresponding to the finite eigenvalues of the generalized eigenvalue problem

$$(\Phi - \lambda\Psi)v = 0 \tag{2.13}$$

where the matrices Φ and Ψ are defined as

$$\Phi = \begin{pmatrix} 1 & 0 & \cdots & 0 & -b_0 & 0 & \cdots & 0 \\ a_1 & 1 & & 0 & -b_1 & -b_0 & & 0 \\ a_2 & a_1 & & 0 & -b_2 & -b_1 & & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_n & a_{n-1} & \cdots & a_1 & -b_n & -b_{n-1} & \cdots & -b_1 \\ 0 & a_n & \cdots & a_2 & 0 & -b_n & \cdots & -b_2 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_n & 0 & 0 & \cdots & -b_n \end{pmatrix}$$

and

$$\Psi = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & & 0 & -a_1 & -1 & & 0 \\ 0 & 0 & & 0 & a_2 & a_1 & & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & (-1)^n a_n & (-1)^n a_{n-1} & \cdots & (-1)^n a_1 \\ 0 & 0 & \cdots & 0 & 0 & a_n & \cdots & a_2 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & (-1)^{2n-1} a_n \end{pmatrix}$$

For each eigenvector v_j the coefficients of the polynomials X and Y are defined by

$$v_j = (y_0, y_1, \ldots, y_{n-1}, 1, x_1, \ldots, x_{n-1})^T$$

.

Proof: Identifying coefficients in both members of Eq. 2.8 gives the result. ∎

Remark. This generalized eigenvalue problem can always be rewritten as an ordinary eigenvalue problem. The matrix Φ is always non-singular due to the fact that $B(s)$ and $A(s)$ do not have any common factors. Therefore we can rewrite the problem as $(\lambda^{-1} I - \Phi^{-1}\Psi))v = 0$. □

Combining Theorem (2.2) and Lemma (2.2) we arrive at the new result

THEOREM 2.3— Given a rational function $G(s) \in R_N$,
the stable function $\hat{G}(s) \in R_n$, $n < N$, minimizing the Hankel norm of the error $G - \hat{G}$, is given by the stable projection of the function

$$H(s) = \frac{y_0 s^{N-1} + y_1 s^{N-2} + \cdots y_{N-1}}{s^{N-1} + x_1 s^{N-2} + \cdots x_{N-1}}$$

where $(y_0, y_1, \ldots, y_{N-1}, 1, x_1, \ldots, x_{N-1})^T$ is the eigenvector corresponding to the eigenvalue, with the n:th largest finite magnitude, of the generalized eigenvalue problem Eq. (2.13)

Proof: By definition of H, minimizing $\|G - \hat{G}\|_H$ is equivalent to minimizing

$$\inf_{F \in \hat{H}_\infty} \|G - \hat{G} - F\|_\infty$$

Theorem 2.2 gives the solution

$$\begin{cases} G(s) - H(s) = \sigma_k(\Gamma_G) E_k(s) \\ \hat{G}(s) = \Pi_{H_\infty} H(s) \end{cases}$$

Since G and H both are rational we must have that $E_k(s)$ is rational. Let $G(s) = B(s)/A(s)$ and $H(s) = Y(s)/X(s)$ with A, B, X and Y polynomials. By considering the fact that $|E_k(i\omega)| = 1$, $\omega \in \mathbb{R}$, we can write Equation (2.11) as

$$\frac{B(s)}{A(s)} - \frac{Y(s)}{X(s)} = \sigma_n(\Gamma_G)\frac{A^*(s)X^*(s)}{A(s)X(s)}$$

This is precisely Equation (2.12) in Lemma 2.2 with $\lambda = -\sigma_n(\Gamma_G)$. Since the generalized eigenvalue problem (2.13) has precisely n finite eigenvalues, we can identify all the Hankel singular values $\sigma_i(\Gamma_G)$, $i = 1, 2, \ldots, n$ as the absolute values of these eigenvalues. ■

Remark. A computational advantage with this method is that no realizations are needed, since the coefficients of the transfer function are used directly. Furthermore, the Hankel norm approximations of all orders $0, \ldots, n-1$ can be computed from the eigenvectors $v_1, \ldots, v_n$. This means that only *one* generalized eigenvalue problem has to be solved to get *all* Hankel norm approximations. These features are shared with a similar method [Harshavardhana, Jonckheere, and Silverman, 1984]. The method presented in that paper has, however, a slightly more complicated proof. □

Since the optimal Hankel norm approximation is independent of the direct term of the transfer function, it is not unique. If $\hat{G}_s(s)$ denotes the unique *strictly* proper, fixed order Hankel norm approximation of $G(s)$ then every other Hankel norm approximation (of the same order) is found from $\hat{G}_s(s)+\hat{D}$ where $\hat{D}$ is a constant. The extra degree of freedom obtained by the direct term in the approximation can be used to reduce the Tchebycheff norm of the error. A reasonable, heuristic choice of the direct term is given by $\hat{D}$ such that $|G(0) - \hat{G}_s(0) - \hat{D}| = |G(\infty) - \hat{G}_s(\infty) - \hat{D}|$.

Weighted Hankel Norm Approximation

Typically the error $G(i\omega) - \hat{G}(i\omega)$ has almost constant magnitude when unweighted Hankel norm approximation is used. The error magnitude curve can be shaped by using a weighting function in the Hankel norm approximation [Latham and Anderson, 1985], [Anderson, 1986], [Hung and Glover, 1986]. The minimization problem then becomes

$$\min \|W(s)(G(s) - \hat{G}(s))\|_H$$

where the minimization is taken over all stable $\hat{G}(s)$ of a fixed degree k. The computation is done in the following steps:

1. Find the stable and unstable projections of $W(s)G(s)$:

$$W(s)G(s) = G_+(s) + G_-(s)$$

where $G_-(s)$ is the stable part.

2. Find the (unweighted) order k Hankel norm approximation $\hat{G}_-(s)$ of $G_-(s)$.
3. Find $\hat{G}(s)$ from

$$W(s)\hat{G}(s) = \hat{G}_+(s) + \hat{G}_-(s)$$

where $\hat{G}_+(s)$ is unstable. The function $\hat{G}(s)$ then minimizes the Hankel norm of the weighted error $W(s)(G(s) - \hat{G}(s))$.

As in the unweighted case the approximation is not unique. A heuristic choice of direct term is

$$\hat{D} = \frac{W(\infty)G(\infty) + W(0)(G(0) - \hat{G}(0))}{W(\infty) + W(0)}$$

This implies that

$$\left|\frac{G(\infty) - \hat{G}(\infty)}{G(0) - \hat{G}(0)}\right| = \left|\frac{W(0)}{W(\infty)}\right|$$

i.e. the ratio between the error magnitudes at $\omega = \infty$ and $\omega = 0$ will be the inverse ratio of the corresponding magnitudes of the weighting function. A motivation for this choice is that the error is expected to be larger at frequencies for which the weighting function magnitude is small and vice versa. Ideally, with this point of view, the error magnitude should essentially look like the weighting function magnitude turned upside down. An important fact is, however, that the Hankel norm depends not only on the *magnitude* of the weighting function. Let $A_w(s)$ and $B_w(s)$ be polynomials with all roots strictly in the left half plane. The four weighting functions

$$W_1(s) = \frac{B_w(s)}{A_w(s)}, \qquad W_2(s) = \frac{B_w(s)}{A_w(-s)}$$
$$W_3(s) = \frac{B_w(-s)}{A_w(s)}, \qquad W_4(s) = \frac{B_w(-s)}{A_w(-s)}$$

all have the same magnitude on the imaginary axis. In [Latham and Anderson, 1985] the case W_4 is analyzed and in [Hung and Glover, 1986] the cases W_3 and W_4 are considered when both $A_w(s)$ and $B_w(s)$ are of first order.

EXAMPLE 2.3—Eight cascaded low pass filters
Recalling the eighth order system in Section 1.1 we will in this section illustrate the different approximation methods reviewed in 2.2. The transfer function

$$G(s) = \frac{1}{(s+1)^8}$$

will thus be approximated by rational transfer functions of different orders using the following methods:

- Unweighted Hankel norm approximation
- Weighted Hankel norm approximation
- Least squares approximation

Unweighted Hankel norm approximation

The unweighted Hankel norm approximations of orders one to seven are computed. The Nyquist curves and step responses for the system together with the approximations are shown in Figure 2.2. One observation is that all the approximating transfer functions has some zeros in the right half plane. This illustrates the fact that to be able to match the phase decrease at higher frequencies the low order approximations must have non-minimum phase zeros. The third order unweighted Hankel norm approximation

$$\hat{G}(s) = \frac{-0.0274s^3 + 0.0679s^2 - 0.0864s + 0.0483}{s^3 + 0.6628s^2 + 0.3155s + 0.0470}$$

has zeros in 1.072 and $0.7044 \pm i1.0726$ and poles in -0.2142 and $-0.2243 \pm i0.4114$. The heuristic choice of direct term as described in Section 2.4 is used here. More generally the following three properties have that in common that they model large phase decrease:

- Large relative degree
- Non-minimum phase zeros
- Time delay

When designing controllers of a fixed low order, the non-minimum phase zeros of the process transfer function approximations put limits on the achievable bandwidth. The magnitudes of the approximation errors $G(i\omega) - \hat{G}_k(i\omega)$, $k = 1, \ldots, 7$ are plotted in Figure 2.3. Notice that the error magnitudes are almost constant with respect to frequency.

Weighted Hankel norm approximation

Sometimes greater accuracy is required in a limited frequency range. This is achieved by using weighted Hankel norm approximation as described in Section 2.4. As an illustration the following four types of weighting functions are considered:

$$W_1(s) = \frac{B_w(s)}{A_w(s)}, \qquad W_2(s) = \frac{B_w(s)}{A_w(-s)}$$

$$W_3(s) = \frac{B_w(-s)}{A_w(s)}, \qquad W_4(s) = \frac{B_w(-s)}{A_w(-s)}$$

where $A_w(s) = s^2 + \omega_0 s + \omega_0^2$ and $B_w(s) = s^2 + 2\omega_0 s + \omega_0^2$. The error magnitudes for the different weightings are shown in Figure 2.4 together with corresponding

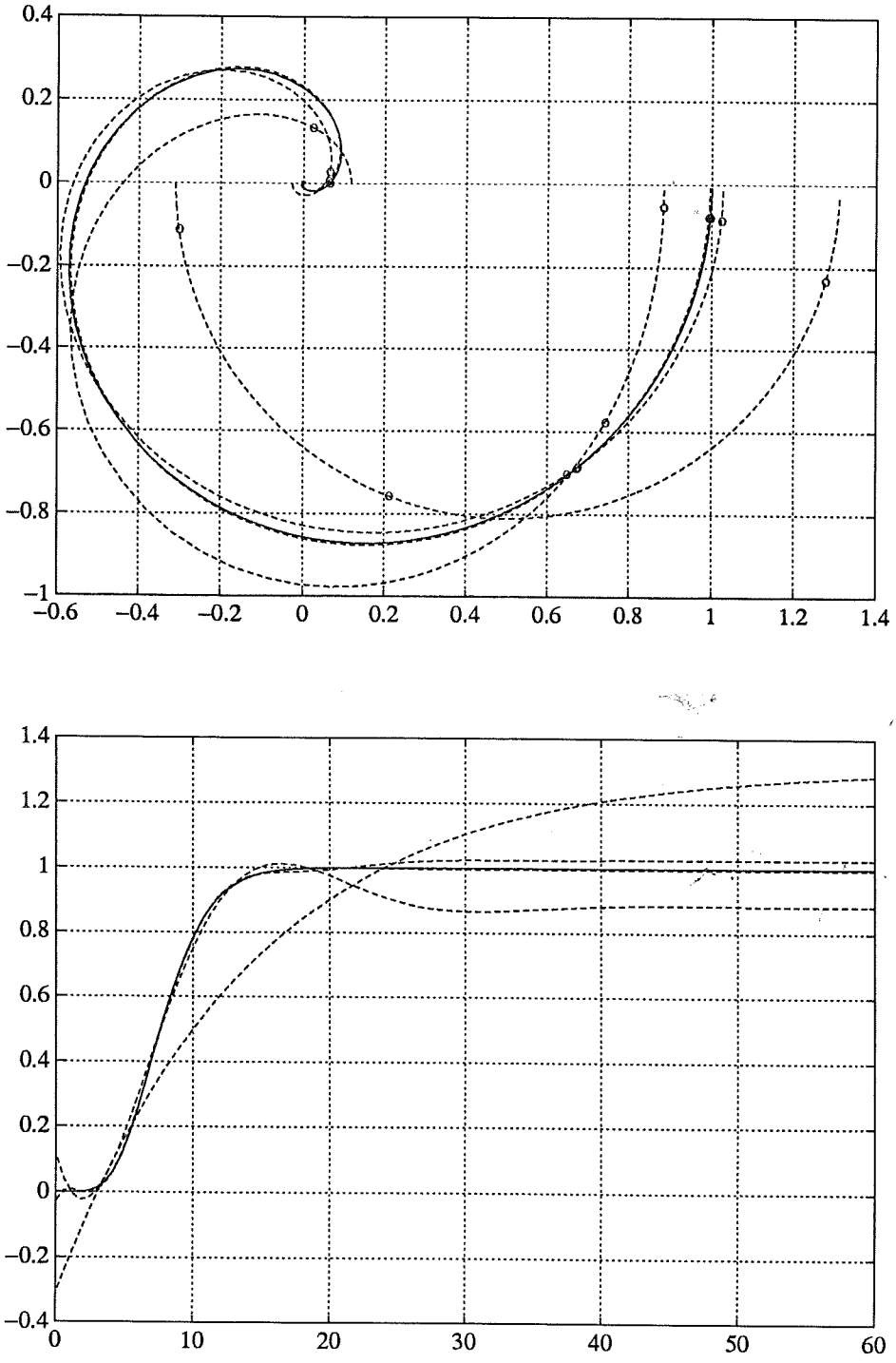


Figure 2.2 Nyquist curves (top) and step responses (bottom) of $G(s) = (s+1)^{-8}$ (solid curves) and unweighted Hankel norm approximations of orders 1 to 4 (dashed curves). Markings are done for $\omega = 0.01$, 0.1 and 1.

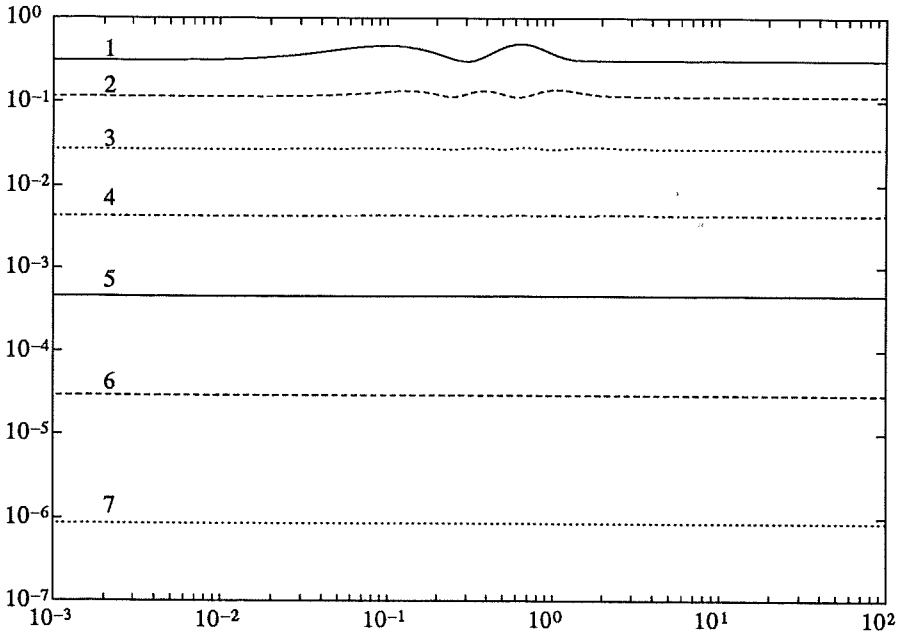


Figure 2.3 Magnitude of approximation errors for unweighted Hankel norm approximations of orders 1 to 7 for $G(s) = (s+1)^{-8}$

weighting function for the values $\omega_0 = 0.2$ and $\omega_0 = 1$. The smallest L^∞-error is obviously obtained by the weighting function with all poles and all zeros in the right half plane ($W_4(s)$). The weighting function $W_4(s)$ results in an error magnitude that is shaped like a mirror image (upside down) of the corresponding weighting function.

Least squares approximation

If the exact shape of the error magnitude curve is of less importance, it is sufficient to use the least squares method as described in Section 2.3. Furthermore, this method requires only a finite number of values of the full order transfer function. Let the approximation set be $\Omega = \{0.01, 0.1, 0.2, 0.4\}$ and the weighting function be $W(s) \equiv 1$.

The Nyquist curves and error magnitudes of the rational approximations of orders $(3,3)$, $(2,3)$ and $(2,2)$ are shown in Figure 2.5. In this case the weighting function $W(s) \equiv 1$ was used since the transfer function is relatively easy to approximate with lower order rational functions. The approximation of order $(3,3)$

$$\hat{G}(s) = \frac{-0.1875s^3 + 0.3039s^2 - 0.2602s + 0.1076}{s^3 + 1.2358s^2 + 0.6009s + 0.1076}$$

has zeros in 0.7842 and $0.4183 \pm i0.7464$ and poles in -0.4108 and $-0.4125 \pm i0.3031$. The "notches" in the error magnitudes are typical for pointwise approx-

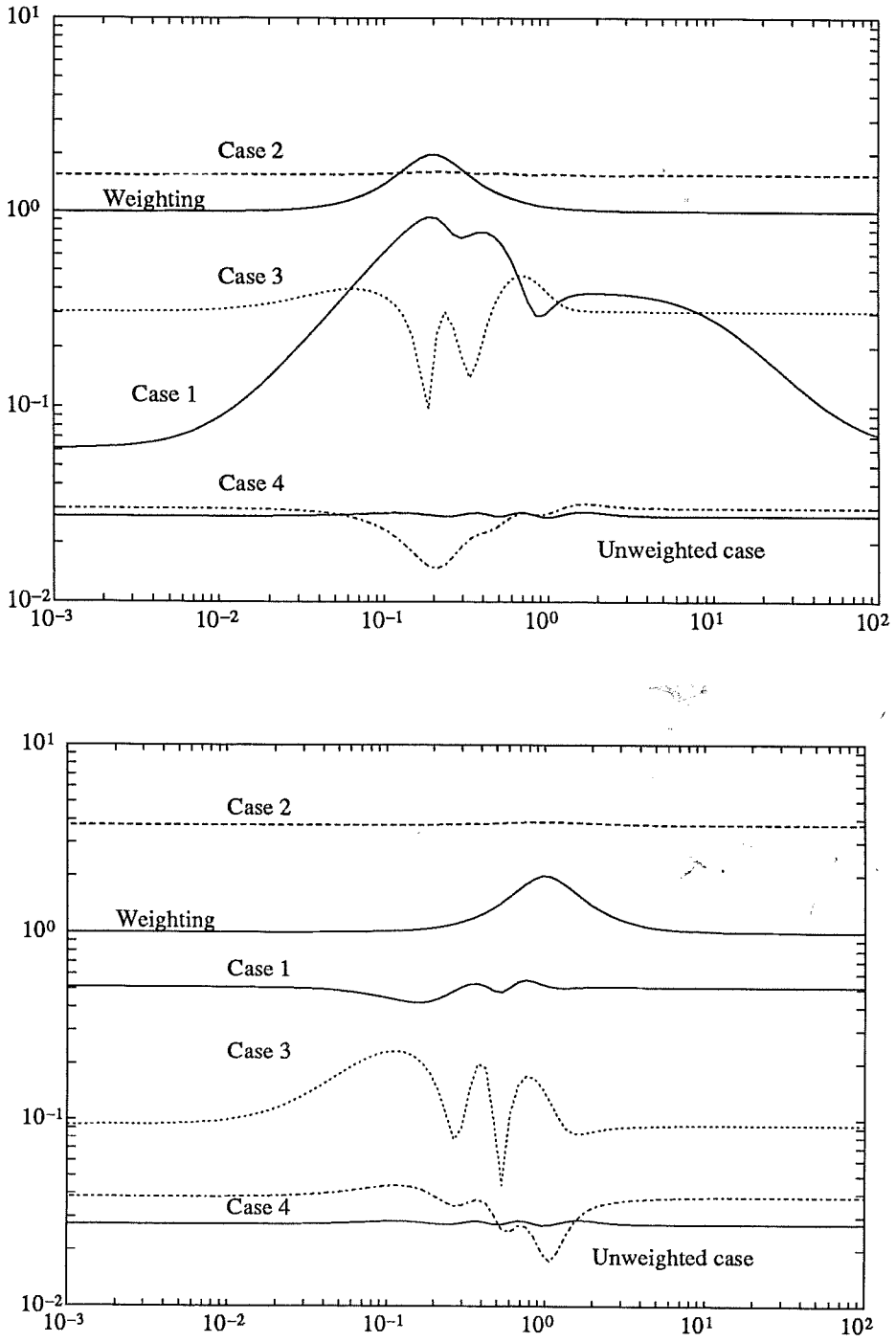


Figure 2.4 Magnitude of approximation errors for weighted Hankel norm approximations of order 3 of $G(s) = (s+1)^{-8}$. $\omega_0 = 0.2$ (top) and $\omega_0 = 1$ (bottom).

imation with few points. In the interpolation case this comes from the fact that

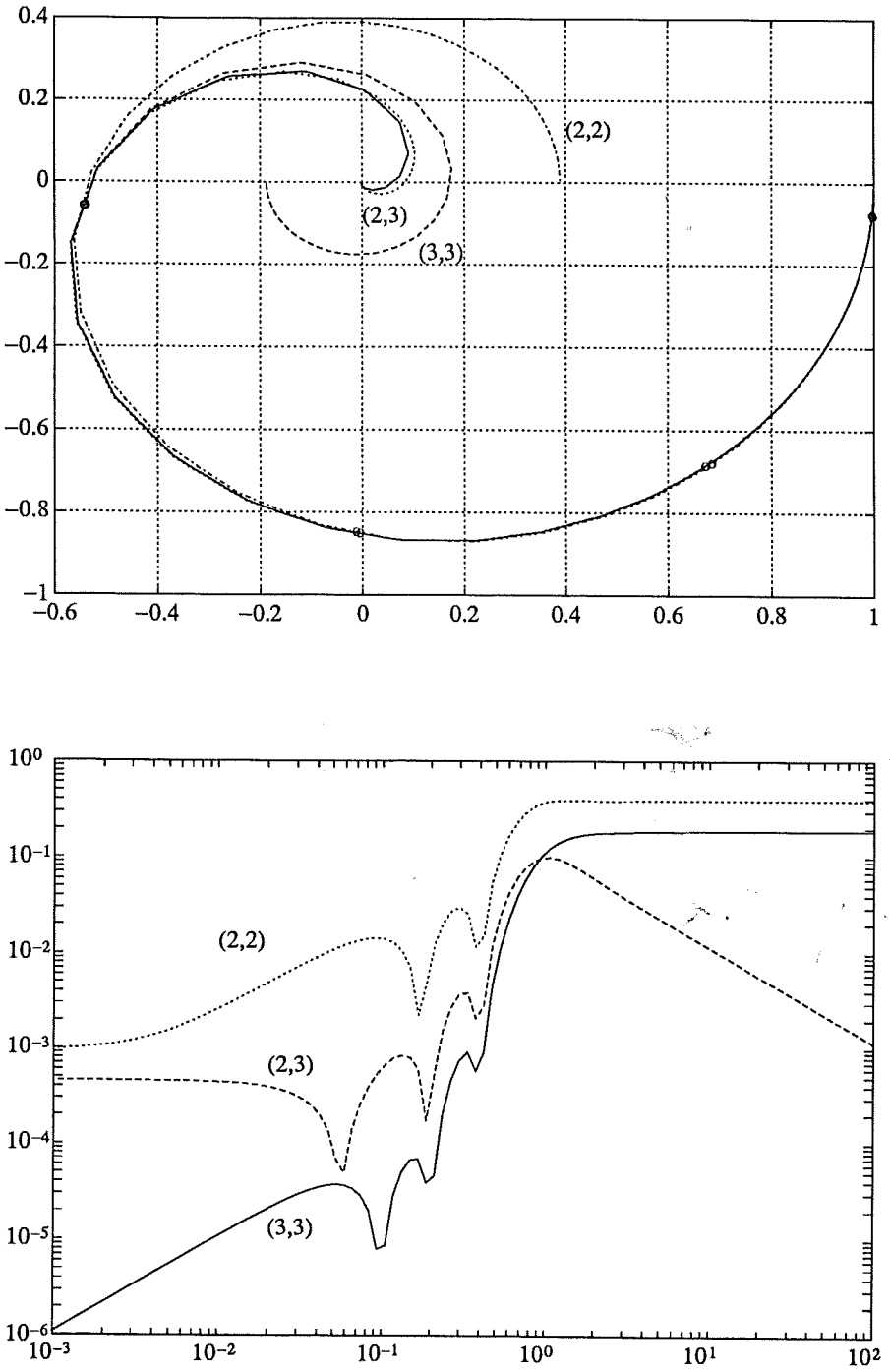


Figure 2.5 Nyquist curves of $G(s) = (s+1)^{-8}$ and least squares rational approximations of orders $(3,3)$, $(2,3)$ and $(2,2)$ (top) and the corresponding error magnitudes (bottom). The approximation frequencies $\Omega = \{0.01, 0.1, 0.2, 0.4\}$ are marked.

the error is zero at the approximation points which clearly is seen as notches in a log-log diagram. By using a larger number of approximation points the error magnitude can be considerably smoothened.

When incorporating higher approximation frequencies the weighting becomes increasingly important. This is due to the fact that the polynomial $\hat{A}(s)$ becomes large at large values of s. As mentioned in Section 2.3, the weighting function $w(s)$ should include some approximation of $1/\hat{A}(s)$. The heuristic Algorithm (2.1) in Section 2.3 was applied to the transfer function $G(s) = 1/(s+1)^8$. A rational approximation of order $(3,3)$ was computed using 20 logarithmically equally spaced frequencies between 0.01 and 100 rad/s. Five iterations were done in the algorithm. The initial weighting used was $W^{[1]}(s) \equiv 1$. This choice of weighting often results in bad conditioning for high order LS-approximations. In these cases it is recommendable to choose $W(s) = 1/p(s)$ for some polynomial $p(s)$ not necessarily of the same degree as $\hat{A}(s)$. Figure 2.6 shows the error magnitude for the iterated LS-approximation together with the error magnitude of the third order unweighted Hankel norm approximation.

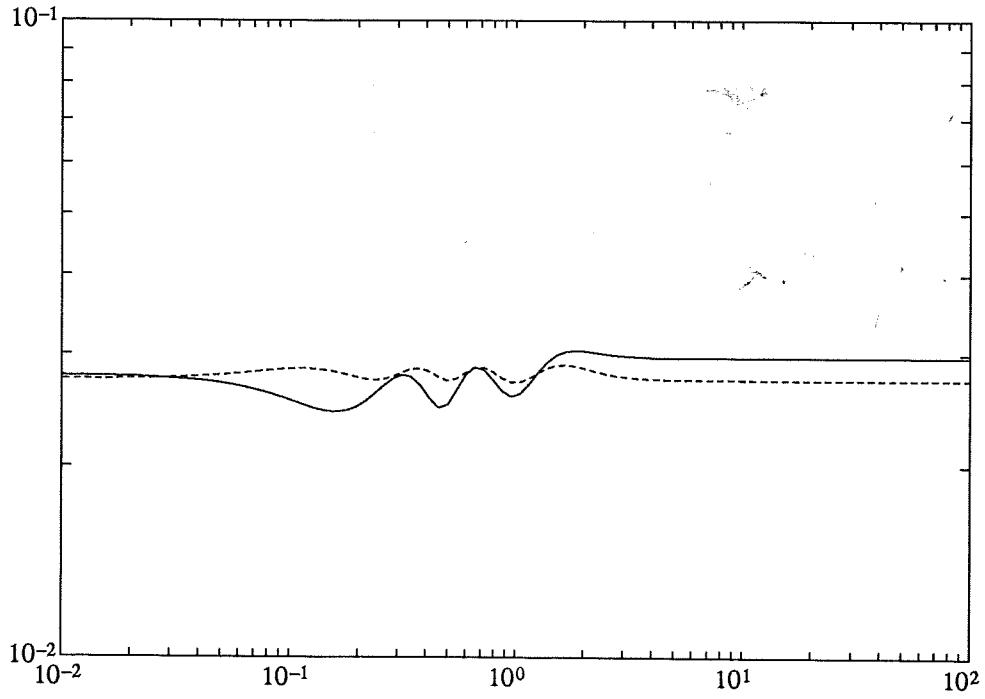


Figure 2.6 Error magnitudes of two different rational approximations of $G(s) = (s+1)^{-8}$. The solid curve shows a rational $(3,3)$ LS-approximation with iteration in the weights and the dashed curve shows a third order unweighted Hankel norm approximation.

The advantage with this method is that approximations $\hat{G}(s)$ for which the

error magnitude is almost constant (as for unweighted Hankel norm approximations) can be obtained for non-rational transfer functions $G(s)$. □

Remark. From a control theory point of view, it is sometimes desirable that the reduced order process transfer function $\hat{G}(s) = \hat{B}(s)/\hat{A}(s)$ has a relative degree ($= \deg \hat{A} - \deg \hat{B}$) which is at least one. This is the case when the controller contains a direct term. If both the process and the controller contains a direct transmission then the closed loop system can be undefined or non-physical for certain controller gains. Even if the loop has zero gain at $s = 0$ there can be problems as the following example [Willems, 1971] shows. The loop transfer function is $1 - e^{-\tau s}$ and the feedback is positive. This gives a closed loop transfer function which is $e^{\tau s}$ which is not feasible. The problem lies in the assumption that the loop is infinitely fast. By assuming that the total loop transfer function has a relative degree of at least one, such problems are avoided. □

EXAMPLE 2.4—Sixth order doubly resonant system
The transfer function

$$G(s) = \frac{-s+1}{s^6 + 3s^5 + 5s^4 + 7s^3 + 5s^2 + 3s + 1}$$

has two poorly damped pole pairs. The poles are given by -1.7549, -0.5698, $-0.2151 \pm i1.3071$ and $-0.1226 \pm i0.7449$ and the zero is located at 1.

Hankel norm approximation

The Hankel norm approximations of order 3 and 4 yields the error magnitude curves shown in Figure 2.7 with weighting functions $W(s) \equiv 1$ and $W(s) = \frac{s^2-s+1}{s^2-2s+1}$ respectively. The error magnitudes clearly reflects the shape of the weighting function. The third order models are given by

$$\hat{G}(s) = \frac{-0.32932s^3 + 0.64888s^2 - 0.86209s + 0.38705}{s^3 + 0.68151s^2 + 0.75731s + 0.29117}$$

$$\hat{G}_W(s) = \frac{-0.48339s^3 + 0.76875s^2 - 1.0336s + 0.45638}{s^3 + 0.73622s^2 + 0.78391s + 0.30766}$$

The poles of the transfer functions are -0.4463 and $-0.1176 \pm i0.7991$ in the unweighted case and -0.4674 and $-0.1344 \pm i0.8001$ in the weighted case. It is interesting to note that the resonance frequencies of the approximants are slightly higher than the lower resonance frequency of the original model. A plausible explanation is that this is an influence by the upper resonance frequency in $G(s)$. The fourth order models are given by

$$\hat{G}(s) = \frac{0.14742s^4 - 0.55034s^3 + 1.2176s^2 - 1.4938s + 0.72253}{s^4 + 1.4318s^3 + 2.4269s^2 + 1.1135s + 0.84746}$$

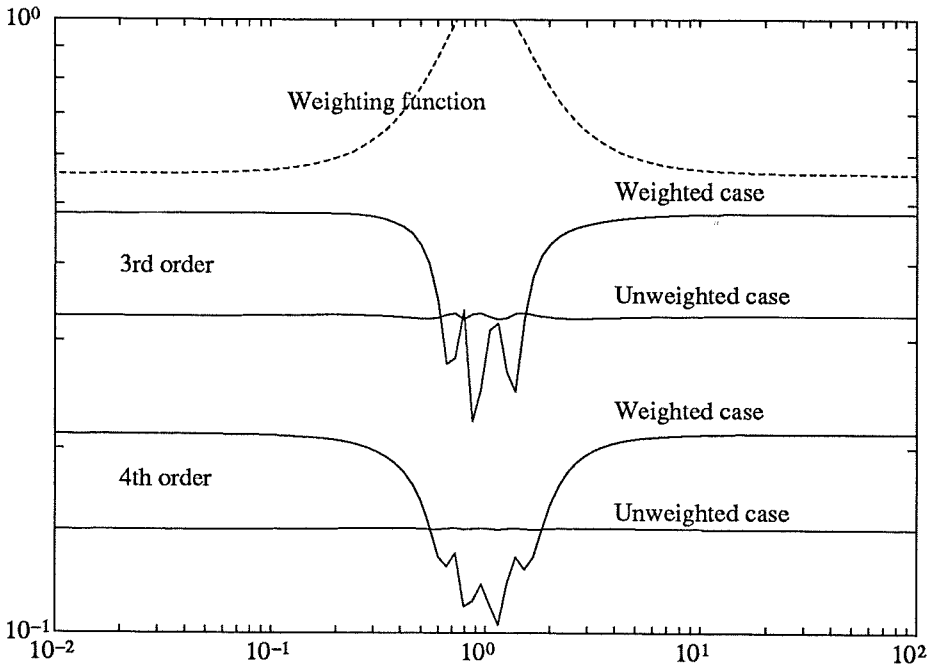


Figure 2.7 Error magnitudes for third and fourth order unweighted and weighted Hankel norm approximations (solid curves) of the transfer function $G(s) = (-s + 1)/(s^6 + 3s^5 + 5s^4 + 7s^3 + 5s^2 + 3s + 1)$. The magnitude of the weighting function is also shown (dashed curve). To get good scalings of the axes in the plot, the weighting function has been multiplied by a constant factor.

and

$$\hat{G}_W(s) = \frac{0.21126s^4 - 0.61495s^3 + 1.3291s^2 - 1.5352s + 0.71603}{s^4 + 1.3321s^3 + 2.5147s^2 + 1.0645s + 0.90782}$$

The poles of the transfer functions are given by $-0.1630 \pm i0.7332$ and $-0.5530 \pm i1.0939$ in the unweighted case and $-0.1588 \pm i0.7250$ and $-0.5152 \pm i1.1790$ in the weighted case. In both cases the second pair of complex conjugated poles is better damped than the corresponding pole pair in G. This could be explained by the attempt of the approximations to capture the influence of the two real poles of G. For the same reason, the higher resonance frequency has been decreased compared to the corresponding resonance frequency in the original model.

Least squares approximation

The least squares method gives rational approximants of orders $(2,3)$, $(3,3)$ and $(3,4)$ which have the error magnitudes shown in Figure 2.8. The approximation set used in this case was $\Omega = \{0.01, 0.1, 0.3, 0.7, 1\}$ and the weighting function $W(s) \equiv 1$. The $(3,4)$ approximant is unstable. This is mainly due to the fact that the fourth order model "lacks" information at frequencies near the second

resonance. Instead an unstable pole-zero pair appears in the approximation. The second poorly damped pole pair has imaginary parts ± 1.3. By extending the approximation set by the frequency $\omega = 1.4$ (for example), the $(3,4)$ approximant becomes stable with poles at $-0.1847 \pm i0.7502$ and $-0.6617 \pm i1.0347$ and zeros at 0.5302 and $0.3772 \pm i1.2495$. As in the 4th order Hankel norm approximations, the second complex pole pair has a larger damping than the corresponding pole pair in the original system. Another observation is that the stable $(3,4)$ approximation has a larger error magnitude at low frequencies than the $(2,3)$ approximation. Instead the error magnitude is decreased in the frequency range above $\omega = 1$. □

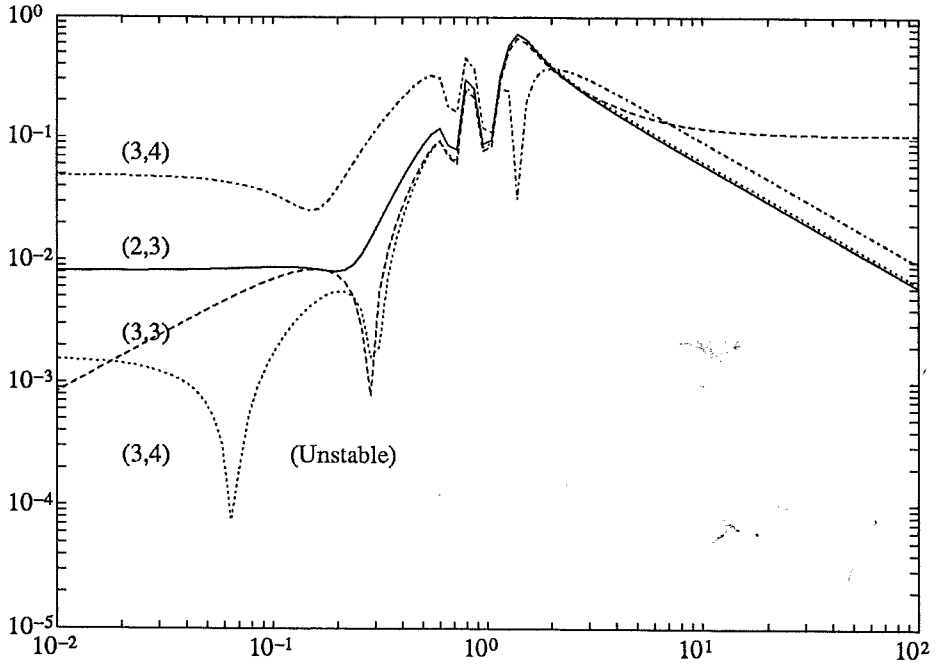


Figure 2.8 Error magnitudes for third and fourth order least squares approximations. One of the fourth order approximations is unstable. The approximation frequencies are $\Omega = \{0.01, 0.1, 0.3, 0.7, 1\}$ for all cases except the stable $(3,4)$ approximation in which case $\omega = 1.4$ is added.

EXAMPLE 2.5—Third order system with time delay
Consider the system with transfer function

$$G(s) = \frac{1}{s^3 + 2s^2 + 2s + 1} e^{-2s}$$

The least squares approximations of orders $(2,3)$, $(3,3)$, $(3,4)$, $(4,4)$ and $(4,5)$ was computed on the approximation set $\Omega = \{0.01, 0.1, 0.2, 0.4, 0.8\}$ with unity weighting. Figure 2.9 shows the Nyquist curves and the step responses for the

system and its approximation. Computation of Hankel norm approximations of non-rational functions will not be considered here. An approximation very close to the Hankel norm approximation of a certain order can, however, be found by first computing a least squares approximation of higher order and then finding a Hankel norm approximation to this LS-approximation. As an example of this procedure, a least squares approximation of order $(7,7)$ was computed using the approximation set $\Omega = \{\omega_1, \ldots, \omega_{10}\}$, with ω_k logarithmically spaced between 0.1 rad/s and 10 rad/s. The LS-approximation was computed with initial weight $W^{[0]}(s) = s^{-3}$ and iterated one step using $W^{[1]}(s) = 1/\hat{A}^{[0]}(s)$. A fifth order Hankel approximation of the seventh order LS-approximation was then computed. The resulting error magnitude curve is shown in Figure 2.10. The error magnitude of a LS-approximation of order $(5,5)$, using the same approximation frequencies, is included in the plot for comparison. This LS-approximation was not iterated and the weighting $W(s) = W^{[0]}(s)$ was used. The error magnitude for the "combined approximation" deviates not very much from a constant which is a property shared with Hankel norm approximations. □

EXAMPLE 2.6—Heat conduction
One type of transfer function appearing in connection with diffusion and heat conduction problems is

$$G(s) = e^{-\sqrt{s}}$$

Using the approximation frequencies $\Omega = \{0.01, 0.4, 1, 2, 4\}$ rational approximations of orders $(2,3)$, $(3,4)$ and $(4,5)$ were computed with 2 iterations in Algorithm 2.1 (except in the $(4,5)$ case since this corresponds to interpolation). The Nyquist curves of $G(s)$ and the three approximations are shown in Figure 2.11. Notice the "bubbles" on the Nyquist curves of the approximations. This phenomena is probably due to the fact that the slope of the magnitude curve of $G(i\omega)$ varies continuously with ω, while magnitude curves of rational transfer functions has constant slope between the cut off frequencies. A requirement on the rational approximations would then be to have a large "density" of pole-zero pairs on the negative real axis in order to achieve a continuously varying slope of the magnitude curve. This phenomena is actually seen in the approximants. The transfer function $\hat{G}_{45}(s)$ has, for instance, poles in -0.01460, -0.1013, -0.3937, -1.192, -3.367 and -18.45 while the zeros are -0.01602, -0.1199, -0.5145, -1.871 and 27.084. □

2.5 Conclusions

Approximation of functions is a fundamental tool in a many areas of applied mathematics. In control theory the main interest lies in approximation by rational functions. Two of the available methods of rational approximation have been investigated in this chapter, namely optimal Hankel norm approximation and

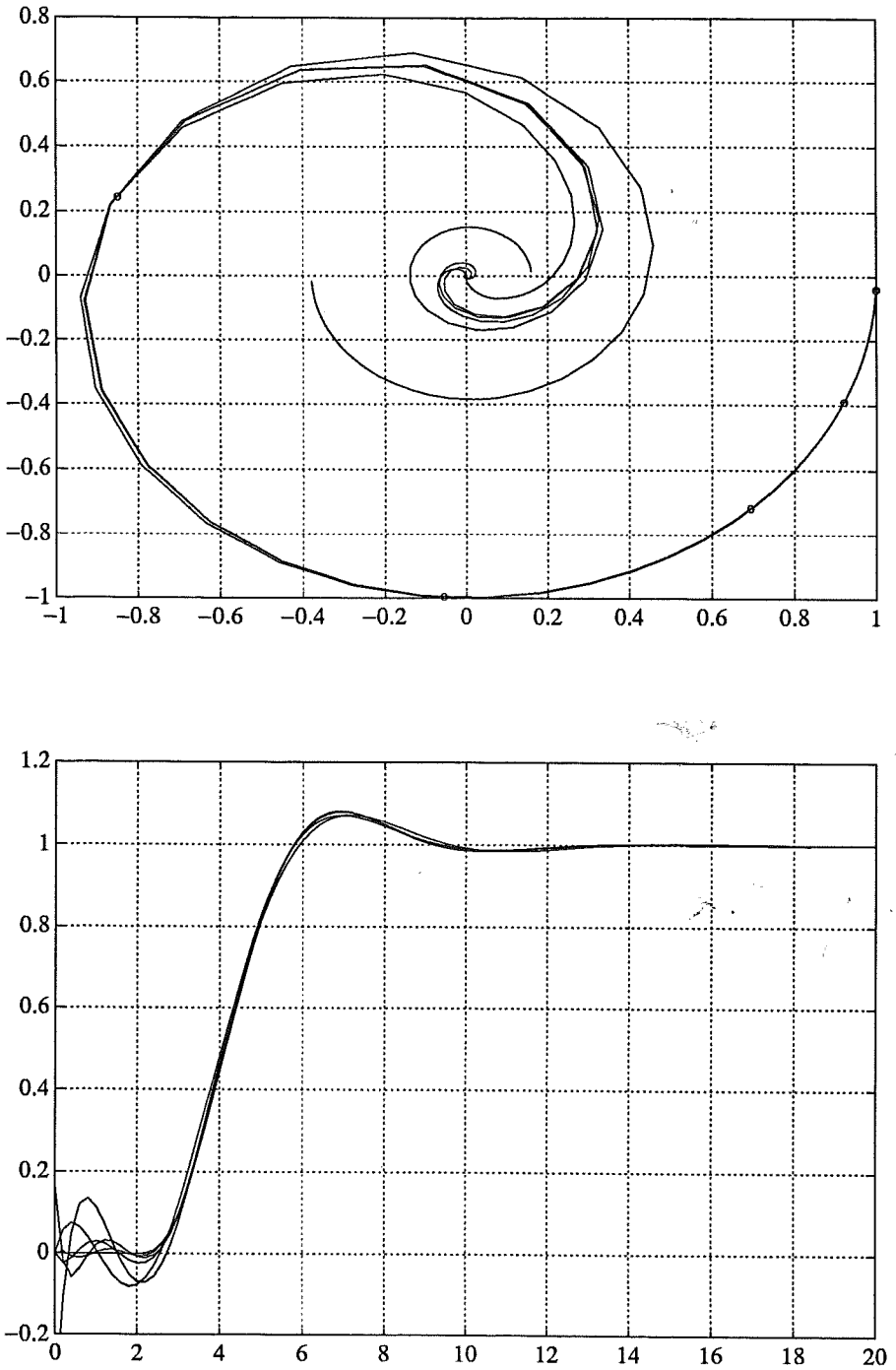


Figure 2.9 Nyquist curves (top) and step responses (bottom) of $G(s) = e^{-2s}/(s^3 + 2s^2 + 2s + 1)$ and its LS-approximations of orders $(2,3)$, $(3,3)$, $(3,4)$, $(4,4)$ and $(4,5)$. The approximation frequencies are $\Omega = \{0.01, 0.1, 0.2, 0.4, 0.8\}$.

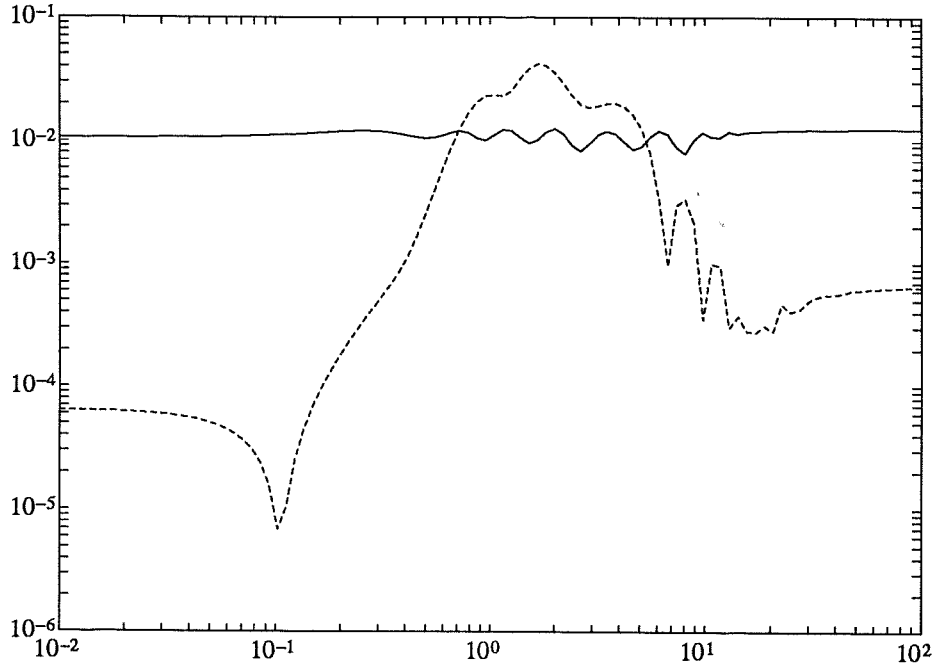


Figure 2.10 Error magnitude for a 5th order Hankel approximation of a 7th order LS-approximation (solid curve) of the transfer function $G(s) = e^{-2s}/(s^3+2s^2+2s+1)$. The error magnitude of a LS-approximation of order $(5,5)$ is also shown (dashed curve).

discrete least squares approximation. The former method has been developed in the last two decades beginning with the papers [Adamjan, Arov, and Krein, 1971] and [Glover, 1984].

In the Hankel norm approximation case there are L_∞-error bounds available which make this method attractive. By introducing weighting it is possible to have good control of the shape of the error magnitude curve. Another advantage is that stable rational functions always have stable Hankel norm approximants. One drawback is that it is not possible to directly obtain a Hankel norm approximation from frequency response data.

The discrete least squares method is easy to apply to frequency response data. Thus, there are no difficulties in approximating non-rational transfer functions. However, there are no currently known bounds on the L_∞-error for any class of transfer functions and distribution of approximation frequencies, except in trivial cases. One such case is when the transfer function $G(s)$ to be approximated is allowed to have poles arbitrarily close to the imaginary axis. It is then easily seen that there exist no bounds on the L_∞-norm of the approximation error since a G with poles on the imaginary axis may even be interpolated by a transfer function $\hat{G}$ at the approximation frequencies ω_k assuming that no pole of G coincides with $\pm i\omega_k$. The stability of the resulting approximants is

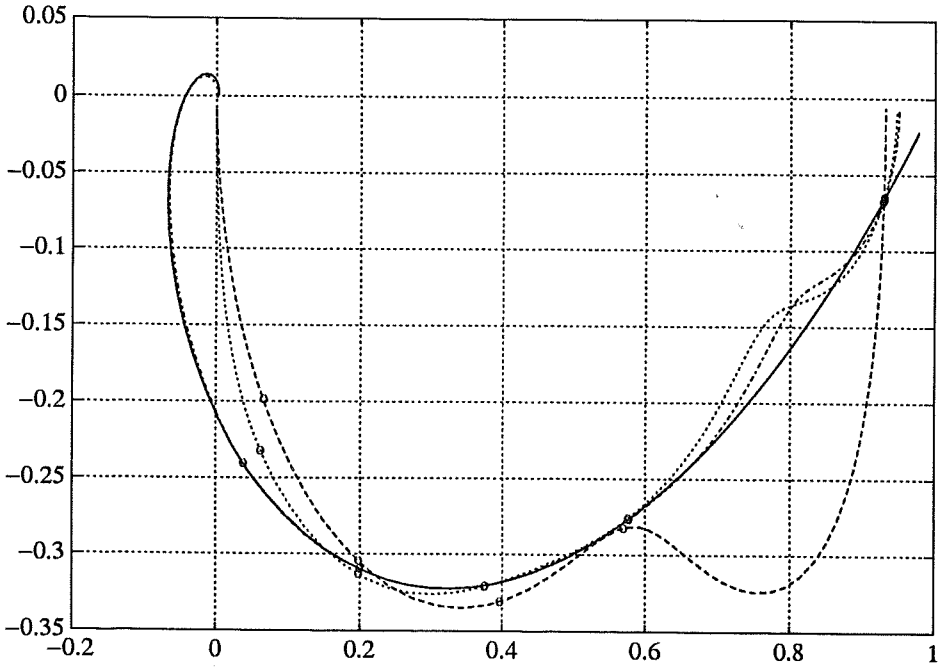


Figure 2.11 Nyquist curves of $G(s) = e^{-\sqrt{s}}$ (solid curve) and LS-approximations of orders $(2,3)$ (dashed curve), $(3,4)$ (dotted curve) and $(4,5)$ (dash-dotted curve). The approximation frequencies are $\Omega = \{0.01, 0.2, 0.5, 1, 2\}$.

not guaranteed either. In some examples it has been shown that the choice of approximation frequencies is in fact crucial to get stable approximants.

Least squares and Hankel norm techniques can be combined by first obtaining a high order model from a measured frequency response by least squares fitting and then applying optimal Hankel norm model reduction to the high order model. A similar suggestion is found in [Wahlberg, 1987, Part I] concerning the combination of identification methods and model reduction methods.

In this chapter Hankel norm approximation and least squares approximation have been applied to some different transfer functions. The examples give insight into some of the properties of the methods. A new method to compute Hankel norm approximations for rational SISO transfer functions has been proposed. This method uses the coefficients of the transfer functions directly, without any transformation to state space. A new non-linear least squares method is also presented, in which a set of points on a given Nyquist curve is fitted to a rational transfer function with a time delay.

3

Pole Placement Design

3.1 Introduction

Pole placement is one of the simpler direct design procedures. The key idea is to find a feedback law such that the closed-loop poles have the desired locations. Let the process to be controlled be described by

$$y = \frac{B}{A}(u + d) \tag{3.1}$$

where u is the control variable, y the process output and d a load disturbance. The symbols A and B denote polynomials in the differential operator $\frac{d}{dt}$ for continuous-time systems or the forward shift operator q for discrete-time systems. It is assumed that A and B are relatively prime, i.e. that they do not have any common factors. Further, it is assumed that A is monic, i.e. that the coefficient of the highest power in A is unity. The relative degree is defined as $\deg A - \deg B$. Let the desired response from the reference signal r to the output be described by the dynamics

$$y_m = \frac{B_m}{A_m} r \tag{3.2}$$

A two degree of freedom linear regulator can be described by

$$Ru = Tr - S(y + n) \tag{3.3}$$

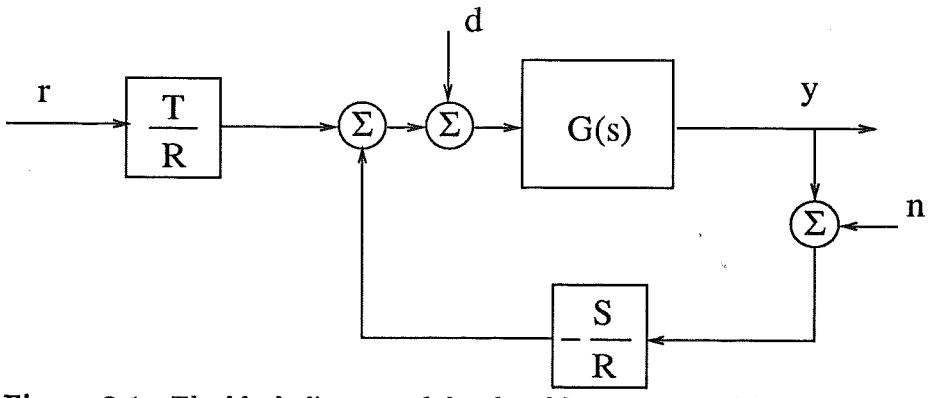


Figure 3.1 The block diagram of the closed loop system with disturbances

where n represents measurement noise. The block diagram of the closed loop system is shown in Figure 3.1. In the rest of the chapter it will be assumed that the process is strictly proper ($\deg A - \deg B \geq 1$) and that the controller is proper ($\deg R \geq \deg S$ and $\deg R \geq \deg T$). The properness of the controller implies that the condition

$$\deg A_m - \deg B_m \geq \deg A - \deg B \tag{3.4}$$

must hold. Elimination of u between Eqs. (3.1) and (3.4) gives

$$\begin{aligned} y &= \frac{BT}{AR+BS}r + \frac{BR}{AR+BS}d - \frac{BS}{AR+BS}n \\ u &= \frac{AT}{AR+BS}r - \frac{BS}{AR+BS}d - \frac{AS}{AR+BS}n \end{aligned} \tag{3.5}$$

To achieve the desired input-output response, the following condition must be true

$$\frac{BT}{AR+BS} = \frac{B_m}{A_m} \tag{3.6}$$

The denominator $AR+BS$ is the closed-loop characteristic polynomial. Eq. (3.6) implies that A_m must divide this polynomial. To carry out the design, the polynomial B is factored as

$$B = B^+B^-$$

where B^+ is a monic polynomial whose zeros are stable and sufficiently well damped to be canceled by the regulator. When $B^+ = 1$, there is no cancellation of any zeros. Since B^+ is canceled, it also divides the closed-loop characteristic polynomial. The other factors of this polynomial are A_m and A_o where A_o is referred to as the observer polynomial. This gives the Diophantine-Aryabhatta-Bezout (DAB) equation

$$AR + BS = A_oA_mB^+$$

It follows from this equation that B^+ divides R. Hence

$$R = XB^+$$

$$AX + B^-S = A_oA_m \tag{3.7}$$

The solution of the DAB equation (Eq. 3.7) is equivalent to solving a set of linear equations (see below).

Equation (3.7) has a unique solution if A and B^- are relatively prime. Furthermore, it follows from Eq. (3.6) that B^- must divide B_m and that

$$T = A_oB_m/B^- \tag{3.8}$$

From Eq. (3.5) it is seen that the polynomial A_o is canceled only in the transfer functions from the reference input r. This means that the dynamics corresponding to A_o is not controllable from r. The polynomial A_o will thus influence the response to the disturbances d and n but not the response to the reference signal r.

The DAB equation viewed as a matrix equation

In the following we put $B^- = B$ and $B_m = b_0B$, where b_0 is a constant, for simplicity. Introduce for brevity $P = A_mA_o$. The DAB-equation

$$AR + BS = P \tag{3.9}$$

where R and S are the unknown polynomials, can be written as a linear matrix equation. For convenient notation, the case $\deg A - \deg B = 1$ will be considered. Let

$$\begin{aligned}
A(s) &= s^n + a_1s^{n-1} + \cdots + a_n \\
B(s) &= b_1s^{n-1} + b_2s^{n-2} + \cdots + b_n \\
R(s) &= s^k + r_1s^{k-1} + \cdots + r_k \\
S(s) &= s_0s^\ell + s_1s^{\ell-1} \cdots + s_\ell \\
P(s) &= s^m + p_1s^{m-1} + \cdots + p_m
\end{aligned}$$

and introduce $\theta = (1, r_1, \ldots, r_k, s_0, s_1, \ldots, s_\ell)^T$ and $\phi = (1, p_1, \ldots, p_m)^T$, where $m = n + k$. Furthermore, let $M = \mathrm{Sylv}(A, B, k+1, \ell+1)$ denote the $(n + k +$

$1) \times (k+\ell+2)$ Sylvester matrix of the polynomials A and B given by

$$M = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ a_1 & 1 & \cdots & 0 & b_1 & 0 & \cdots & 0 \\ a_2 & a_1 & \cdots & 0 & b_2 & b_1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_n & a_{n-1} & \cdots & a_1 & b_n & b_{n-1} & \cdots & b_1 \\ 0 & a_n & \cdots & a_2 & 0 & b_n & \cdots & b_2 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_n & 0 & 0 & \cdots & b_n \end{pmatrix}$$

The DAB-equation, Eq. (3.9) can then be written as

$$M\theta = \phi \tag{3.10}$$

The matrix M is non-singular precisely when A and B do not have any common factor. It follows that Eq. 3.10 has a unique solution precisely when M is of full rank and $\ell = n-1$, i.e.

$$\deg S = \deg A - 1$$

Pre-specified controller dynamics

In many situations it is of interest to give some specifications on the controller. To eliminate a step disturbance in d, for example, it is necessary to increase the loop gain at low frequencies. This can be done by specifying that the factor s is to be included in $R(s)$ thus introducing integration in the controller. In the general case the polynomial $R(s)$ is factorized as $R_1(s)R_2(s)$ with $R_1(s)$ being the pre-specified part of $R(s)$. In a similar manner pre-specified factors can be included in the polynomial $S(s)$ so that $S(s) = S_1(s)S_2(s)$. This is often used to introduce notch filters into the controller by letting S_1 contain poorly damped zeros located near the resonances to be damped out.

The DAB-equation 3.9 is then modified to

$$AR_1R_2 + BS_1S_2 = P \tag{3.11}$$

with the unknown polynomials being R_2 and S_2. To get a unique solution to Eq. (3.11) viewed as a system of linear equations, the condition

$$\deg A + \deg R = \deg P = \deg R_2 + \deg S_2 + 1$$

must be true. Furthermore, assuming that $\deg A_m = \deg A$ yields

$$\deg R = \deg A_o$$

and

$$\deg S = \deg A_m + \deg R_1 + \deg S_1 - 1$$

Robustness of the Design

Consider a pole placement design based on an approximate model

$$\hat{G}(s) = \frac{\hat{B}(s)}{\hat{A}(s)}$$

Let $G(s)$ be the transfer function of the system to be controlled. Assume that $\hat{G}$ and G have the same number of poles outside the stability region and that $G_m = B_m/A_m$ is stable. A sufficient condition for stability of the closed loop system is obtained by using the following (generalized) version of Rouchés theorem [Åström and Wittenmark 1984]:

LEMMA 3.1
Assume that the functions $L(s)$ and $\hat{L}(s)$ have the same number of poles outside the stability region $\mathcal{D}$ and that the function $1 + L(s)$ has no zeros outside $\mathcal{D}$. Then the function $1 + \hat{L}(s)$ has no zeros outside $\mathcal{D}$.

Proof: Apply Rouchés theorem to the functions $1 + L(s)$ and $1 + \hat{L}(s)$ which are analytic outside $\mathcal{D}$. ∎

We then obtain

THEOREM 3.1—Robustness
The actual closed loop system is stable if

$$|G(s) - \hat{G}(s)| < \left|\hat{G}(s) + \frac{R(s)}{S(s)}\right| = \left|\frac{\hat{G}(s)T(s)}{G_m(s)S(s)}\right|$$

for $s = i\omega$, $\omega \in \mathbb{R}$.

Proof: By putting $L(s) = 1 + G(s)S(s)/R(s)$ and $\hat{L}(s) = \hat{G}(s)S(s)/R(s)$ in the lemma and using the identity

$$\frac{\hat{G}(s)T(s)}{R(s) + \hat{G}(s)S(s)} = G_m(s) \tag{3.12}$$

the result follows. ∎

This shows that the process has to be more accurate at frequencies where $|G_m(s)|$ is large compared to $|\hat{G}(s)|$. Another way to see this is by considering the sensitivity function

$$G_{\text{sens}}(s) = \frac{dG_{\text{cl}}}{dG}(s)\frac{G(s)}{G_{\text{cl}}(s)}$$

where

$$G_{\text{cl}}(s) := \frac{G(s)T(s)}{R(s) + G(s)S(s)} \tag{3.13}$$

is the actual closed loop transfer function. The magnitude of $G_{\text{sens}}(s)$ gives the gain from the relative error in the process transfer function G to the relative error in the closed loop transfer function $G_{\text{cl}}(s)$. Some calculations give that

$$G_{\text{sens}}(s) = \frac{R(s)}{R(s) + G(s)S(s)} \tag{3.14}$$

From Eqs. (3.13) and (3.14) it follows that

$$G_{\text{sens}}(s) = \frac{G_{\text{cl}}(s)}{G(s)} \frac{R(s)}{T(s)} \tag{3.15}$$

Assuming that $G_{\text{cl}} \approx G_m$ the sensitivity function is large at frequencies where $|G_m/G|$ is large. Equation (3.15) also reveals that the sensitivity is smaller near the controller poles. This is natural since the loop gain is large at those frequencies.

Solutions to the DAB equation

The solutions to the DAB equation (Eq. 3.9) can be parametrized by a polynomial Q. Assume that one solution is given by $R = R_0$, $S = S_0$. Then all solutions to Eq. 3.9 are given by

$$\begin{cases} R = R_0 + BQ \\ S = S_0 - AQ \end{cases} \tag{3.16}$$

This gives a free parameter Q to adjust while keeping the closed loop characteristic polynomial P fixed. The freedom in Q can be used to "shape" the loop transfer function

$$L(s) = \frac{B(s)S(s)}{A(s)R(s)}$$

to get better stability margins or to decrease the sensitivity to disturbances and uncertainties in the process model. The inclusion of a pre-specified factor R_1 in R as in Eq. (3.11) is done by letting $R_0 = R_1 R_{20}$ and $Q = R_1 Q_2$ with Q_2 being the free parameter. The degree of Q must then be chosen such that the controller is proper. Notice that if the coefficients in Q becomes very large the feedback compensation S/R approximates $-A/B$, i.e. the loop transfer function is close to -1. This gives a very sensitive design.

A generalization of Eq. (3.16) is the Youla parametrization, where the polynomials are replaced by certain rings of rational functions [Vidyasagar, 1986].

3.2 Unstable controllers

Some unstable processes have the property that they can be stabilized by a stable controller. This property is often called strong stabilizability. The process must then satisfy the "parity interlacing property" (P.I.P.), i.e. it has an even number of poles between each pair of right half plane zeros (see [Youla, Bongiorno, and Lu 1974]). The result is generalized to arbitrary domains of stability in [Vidyasagar, 1986]. In pole placement it is often the case that the controller is unstable. A typical situation is when the specified bandwidth is high compared to the bandwidth of the process. Unstable controllers can also appear when the desired bandwidth is significantly lower than the open process bandwidth. It is mostly undesirable to use unstable controllers. One reason is that the stability margin is often very small when the controller is unstable (see Section 3.3). Another drawback is the appearance of limit cycles when using unstable controllers in servo mechanisms containing non-linearities [Wallenborg, 1987]. This section treats some special cases where the stability of the controller can be related to the location of the closed loop poles.

Stability of the controller for special pole placements

The DAB equation (Eq. 3.9) can be extended to a parametrized DAB equation

$$A(s)R(s,p) + B(s)S(s,p) = P(s,p) \tag{3.17}$$

equivalent to the parametrized linear equation (compare 3.10)

$$M\theta(p) = \phi(p)$$

Equation 3.17 is solved for each fixed value of p. This constitutes an implicit parametrization of the polynomials R and S. Each coefficient in the polynomials R and S will thus be functions of p. A special one-parameter family of polynomials P is given by

$$P(s,p) = p^m P_1(s/p), \quad m = \deg P_1 \tag{3.18}$$

where p is a real, scalar parameter and P_1 is some fixed polynomial. Notice that the parameter p can be interpreted as a scaling parameter for the magnitude of the roots of P leaving the pattern of the roots unchanged.

LEMMA 3.2
With the parametrization 3.18, the coefficients in the polynomials R and S will be polynomials in the parameter p.

Proof: Assume that $P_1(s) = s^m + c_1 s^{m-1} + \cdots + c_m$. It then follows that $P(s,p) = s^m + c_1 p s^{m-1} + \cdots + c_m p^m$. Assuming that M is invertible, the

controller parameters can be expressed by

$$\begin{pmatrix} 1 \\ r_1(p) \\ \vdots \\ r_{n-1}(p) \\ s_0(p) \\ \vdots \\ s_{n-1}(p) \end{pmatrix} = M^{-1} \begin{pmatrix} 1 \\ c_1 p \\ \vdots \\ c_m p^m \end{pmatrix} \tag{3.19}$$

which shows that $r_1(p), \ldots, s_{n-1}(p)$ are polynomials in the scaling parameter p. ■

A simple extension of the parametrization in Eq. 3.18 is to have one fixed part $P_2(s)$ of $P(s)$ according to

$$P(s,p) = p^m P_1(s/p) P_2(s), \quad m = \deg P_1$$

The formula in Eq. 3.19 is then modified to

$$\begin{pmatrix} 1 \\ r_1(p) \\ \vdots \\ r_{n-1}(p) \\ s_0(p) \\ \vdots \\ s_{n-1}(p) \end{pmatrix} = M^{-1} N \begin{pmatrix} 1 \\ c_1 p \\ \vdots \\ c_{m_1} p^{m_1} \end{pmatrix} \tag{3.20}$$

where $P_1(s) = s^{m_1} + c_1 s^{m_1-1} + \cdots + c_{m_1}$ and $P_2(s) = s^{m_2} + d_1 s^{m_2-1} + \cdots + d_{m_2}$ and the matrix N is given by the $(k+n) \times m_2$ matrix

$$N = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ d_1 & 1 & \cdots & 0 \\ d_2 & d_1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ d_{m_2} & d_{n-1} & \cdots & d_1 \\ 0 & d_{m_2} & \cdots & d_2 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_{m_2} \end{pmatrix}$$

The values of p for which the controller is unstable can be found by checking the zeros of the Hurwitz determinants of appropriate orders considered as polynomials in p. A necessary condition for $R(s)$ to be Hurwitz is that the coefficients $r_1, \ldots, r_{n-1}$ are positive. Lemma 3.2 then gives

THEOREM 3.2
A necessary condition for stability of the controller for a certain value of the scaling parameter p is that the polynomials given by the first n elements of $M^{-1}\phi(p)$ are positive for that value of p.

Proof: The theorem is a trivial consequence of Equation 3.17 and Lemma 3.2. ∎

Stability of the controller for special process models

A necessary condition for a polynomial to be Hurwitz is that all its coefficients are larger than zero. This can be used to prove the following simple result

LEMMA 3.3
A Hurwitz polynomial $X(s)$ must satisfy the condition

$$X^{(k)}(s) \geq 0, \quad s \in \mathbb{R}^+$$

for all $k \geq 0$.

Proof: Since X is Hurwitz all its coefficients must be positive. This is true also for the derivatives of X. Then $X^{(k)}(s)$ is a sum of positive numbers if s is real and positive which proves the statement in the lemma. ∎

This lemma gives a simple check of stability of the controller when applying pole placement to some special process models. A special case, which is easy to analyze, is when the process has a zero of multiplicity 2 on the positive real axis.

THEOREM 3.3
Consider the DAB equation, Eq. 3.9 with $B(s) = (s-\beta)^2$, $\beta > 0$. Then a sufficient condition for the controller to be unstable is

$$P'(\beta)A(\beta) - P(\beta)A'(\beta) < 0 \tag{3.21}$$

Proof: Expressing the left hand side in Eq. 3.21 using Eq. 3.9 gives the result since Lemma 3.3 implies $R'(\beta) < 0$. ∎

EXAMPLE 3.1
Consider the process model

$$G(s) = \frac{(s-1)^2}{(s+1)^3}$$

Let $A_m(s)A_o(S) = P(s) = (s+p)^5$ where $p > 0$. Then Theorem 3.3 gives that the controller is unstable if $p > 7/3 \approx 2.333$. Since the controller is only of second order the exact interval of p for which the controller is stable can be calculated from Theorem 3.2. The polynomials $r_1(p)$ and $r_2(p)$ are given by

$$\begin{cases} r_1(p) = (-3p^5 - 5p^4 + 10p^3 + 30p^2 + 25p - 25)/16 \\ r_2(p) = (5p^5 + 15p^4 + 10p^3 - 10p^2 - 15p + 11)/16 \end{cases}$$

The polynomial $r_2(p)$ is strictly positive for $p > 0$ which implies that the controller can never have a single pole in the right half plane. Checking the zeros of $r_1(p)$ gives that the values of p for which the controller is stable are $p_- \leq p \leq p_+$ where $p_- \approx 0.5678$ and $p_+ = \sqrt{5} \approx 2.236$ which is to be compared with the approximate value $p_+ = 7/3 \approx 2.667$ above. An interesting observation is that unstable controllers appear both when the closed loop is made "fast" ($p > p_+$) and "slow" ($p < p_-$). □

EXAMPLE 3.2—Eight cascaded low pass filters
The freedom of choosing arbitrary poles in pole placement may be deceptive. The following example illustrates this assertion. A controller with integration is computed for the system

$$G(s) = \frac{1}{(s+1)^8}$$

Specifying the desired closed loop model as $G_m(s) = 256/(s+2)^8$ and the observer polynomial as $A_o(s) = (s+2)^8$ gives the controller

$$\begin{aligned} R(s) &= s^8 + 24s^7 + 260s^6 + 1672s^5 + 7050s^4 + 20264s^3 + 39796s^2 + 51480s \\ S(s) &= 39203s^8 + 325064s^7 + 1.1858 \cdot 10^6 s^6 + 2.4880 \cdot 10^6 s^5 + 3.2884 \cdot 10^6 s^4 + \\ &\quad + 2.8075 \cdot 10^6 s^3 + 1.5144 \cdot 10^6 s^2 + 472808s + 65536 \\ T(s) &= 256s^8 + 4096s^7 + 28672s^6 + 114688s^5 + 286720s^4 + 458752s^3 + \\ &\quad + 458752s^2 + 262144s + 65536 \end{aligned}$$

As can be seen, the coefficients in the polynomials are very large, which can cause problems in the implementation of the controller. Moreover, the Nyquist curve of the loop transfer function $B(s)S(s)/A(s)R(s)$ (Fig. 3.2, top) shows that the stability margin is rather poor. The Bode plot of the feedback compensation $S(s)/R(s)$ (Fig. 3.2, bottom) gives some indication on the minimal degree of a stable controller (with the exception of a pole in $s = 0$) achieving approximately the desired closed loop frequency response. The phase increases from $-90°$ to a maximum of $315°$. Requiring the controller to be proper, stable and causal implies that the minimal order of the controller is larger than $405/90 \approx 4.5$. The controller must therefore be at least of fifth order to match the phase increase. A compensator $\hat{S}/\hat{R}$ of fifth order satisfying these conditions was computed using

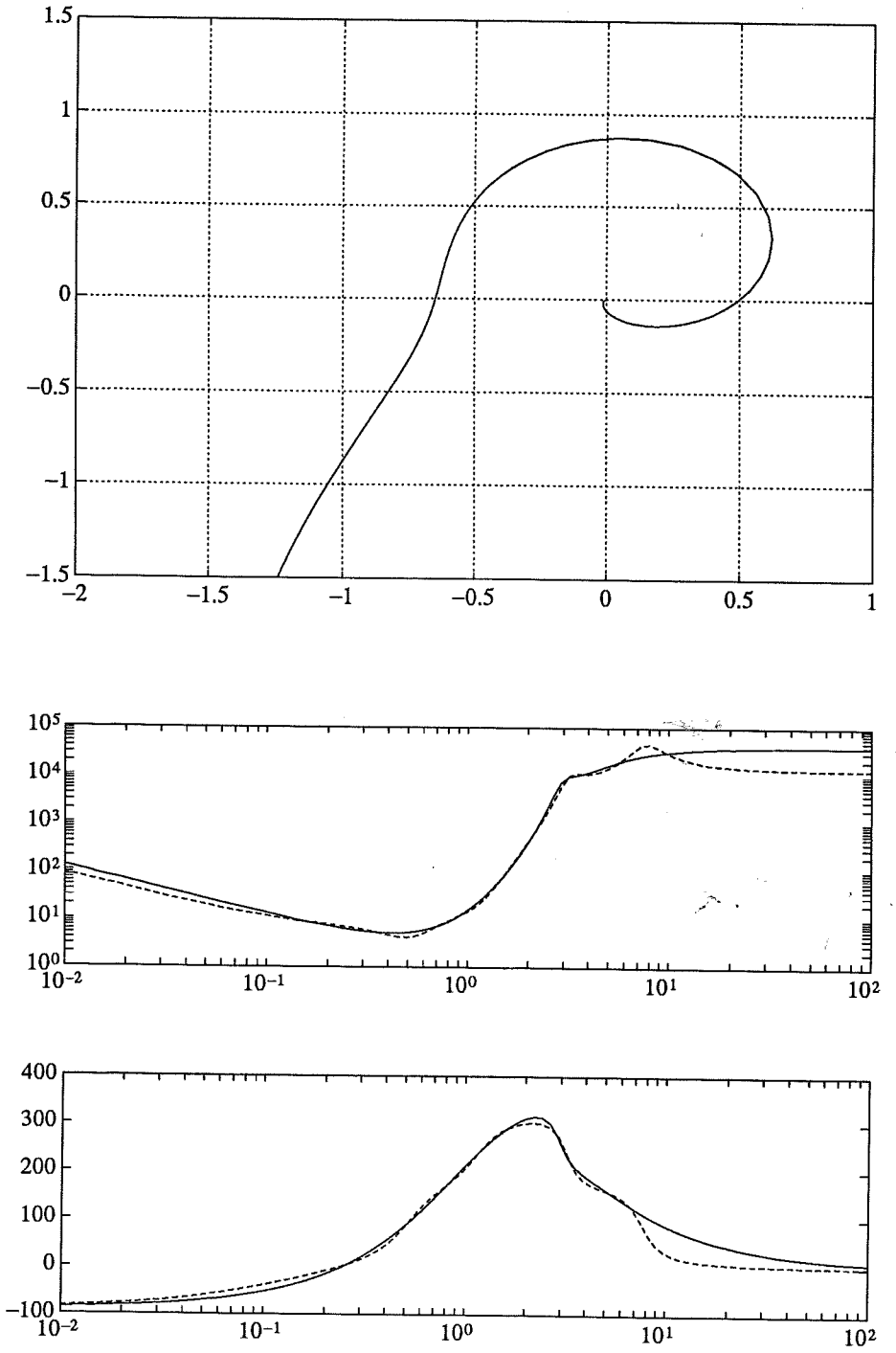


Figure 3.2 The Nyquist curve of the loop using a controller with integration and with $A_m(s) = A_o(s) = (s+2)^8$ (top). The Bode plot of the compensator $S(s)/R(s)$ shows a phase increase of over 400° (bottom). A fifth order least squares approximation is also shown (dashed curves).

weighted least squares approximation with the approximation frequencies in the interval [0.1, 10]. The resulting polynomials $\hat{R}$ and $\hat{S}$ are given by

$$\hat{R}(s) = s^5 + 3.35689s^4 + 70.7206s^3 + 71.8036s^2 + 600.499s$$
$$\hat{S}(s) = 13038s^5 + 14062s^4 + 25070s^3 + 10430s^2 + 5572.0s + 528.55$$

The fifth order controller is stable and gives a stable closed loop system. □

EXAMPLE 3.3—Unstable controllers
One reason to avoid controllers with poles in the right half plane is the problem of modeling errors. Consider a third order least squares approximation $\hat{G}(s)$ to $G(s) = (s+1)^{-8}$ at the frequencies $\omega = 0.01$, 0.1, 0.2 and 0.4:

$$\hat{G}(s) = \frac{0.11539s^2 - 0.14238s + 0.072869}{s^3 + 1.0052s^2 + 0.44057s + 0.072836}$$

Using this stable approximation for an RST-controller design with $A_m(s) = A_o(s) = (s+1)^3$ gives:

$$R(s) = s^3 - 12.909s^2 + 29.64s$$
$$S(s) = 155.15s^3 + 173.21s^2 + 79.528s + 13.723$$
$$T(s) = 13.723s^3 + 41.17s^2 + 41.17s + 13.723$$

The controller has two poles in the open right half plane and yields obviously a stable closed loop system for the approximate third order process model. The actual closed loop system is, however, unstable. This is not too surprising, since the approximation frequencies are low, compared to the bandwidth of the closed loop system. A clearer view of this can be obtained by checking the Nyquist curves of the loop transfer functions of the nominal system (the approximation) and the original system (Figure 3.3, top). Notice the poor stability margins for the nominal loop. Applying the principle of variation of the argument gives that since the controller has two right half plane poles, the loop Nyquist curve must encircle the point -1 one time. The actual loop Nyquist curve, however, does not encircle -1 which means that the closed loop system must be unstable in this case. The two Nyquist curves start diverging at about $\omega = 0.5$ which is above the approximation frequencies. The large high frequency roll-off and phase shift of $G(i\omega)$ above $\omega = 0.7$ makes it impossible for the actual loop Nyquist curve to approach -1. Attempts to match $\hat{G}(s)$ to $G(s)$ at higher frequencies will, on the other hand, give larger errors.

Another motivation for preferring stable controllers is robustness considerations. Using $A_m(s) = A_o(s) = (s+4)^8$ yields an unstable controller (two poles in RHP). The Nyquist curve of the loop (Figure 3.3, bottom) encircles -1 which is necessary to get a stable closed loop system since the controller is unstable. Note

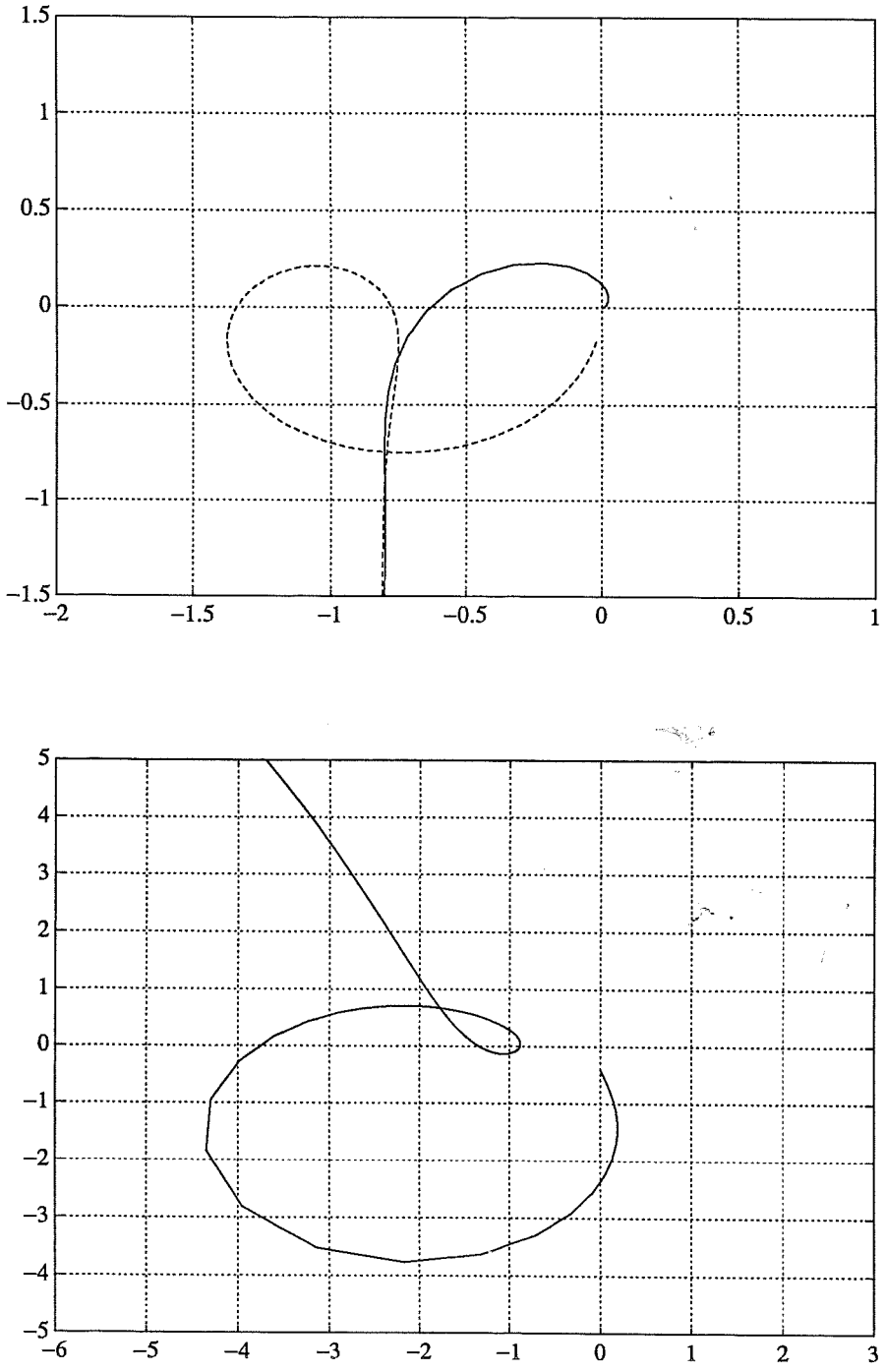


Figure 3.3 The Nyquist curves of the actual loop (solid curve) and the nominal loop (dashed curve) with a controller design based on a third order approximation of $G(s) = 1/(s+1)^8$ (top). The Nyquist curve of the loop using an eighth order controller with integration and with $A_m(s) = A_o(s) = (s+4)^8$ is shown at the bottom.

that even small perturbations in the process model will cause the encirclement around -1 to disappear, i.e., the closed loop system becomes unstable. From the plot it is clear that the stability margin (minimal distance to -1) is ≈ 0.1. □

EXAMPLE 3.4—Sixth order doubly resonant system
The system

$$G(s) = \frac{-s+1}{s^6+3s^5+5s^4+7s^3+5s^2+3s+1}$$

has two resonances and has already been encountered in Chapter 2. A sixth order controller with integration will be designed for the system. The specified closed loop model poles (roots of $A_m(s)$) are evenly spread out on a circular arc with a minimum relative damping of $\zeta = 1/\sqrt{2}$ and for different distances ω_m to the origin (Fig. 3.4). Introduce the shorthand notation

$$A_m(s) = \text{Butt}(n, \omega_m, \alpha) \tag{3.22}$$

with $n = 6$ and $\alpha = 45°$ in this case. The special modification of the Butterworth pole pattern is motivated by the fact that it does not yield the poorly damped step responses for large n as in the ordinary Butterworth pattern (in which case $\alpha = (1 - 1/n)90°$ in Eq. 3.22). We could just as well have chosen e.g. a Bessel pole pattern which also gives better damping. The observer polynomial $A_o(s)$ is chosen identical to $A_m(s)$.

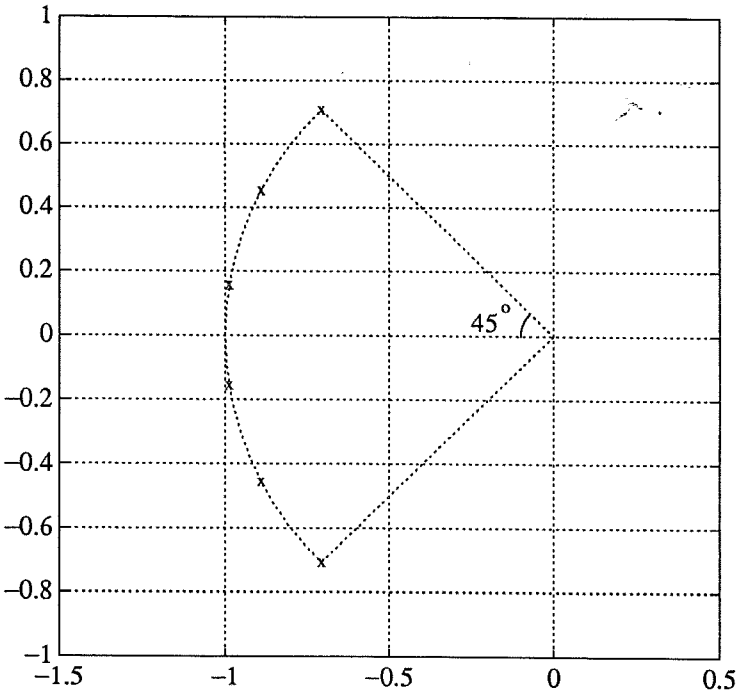


Figure 3.4 The pole pattern used in Example 2 with radius $\omega_m = 1$.

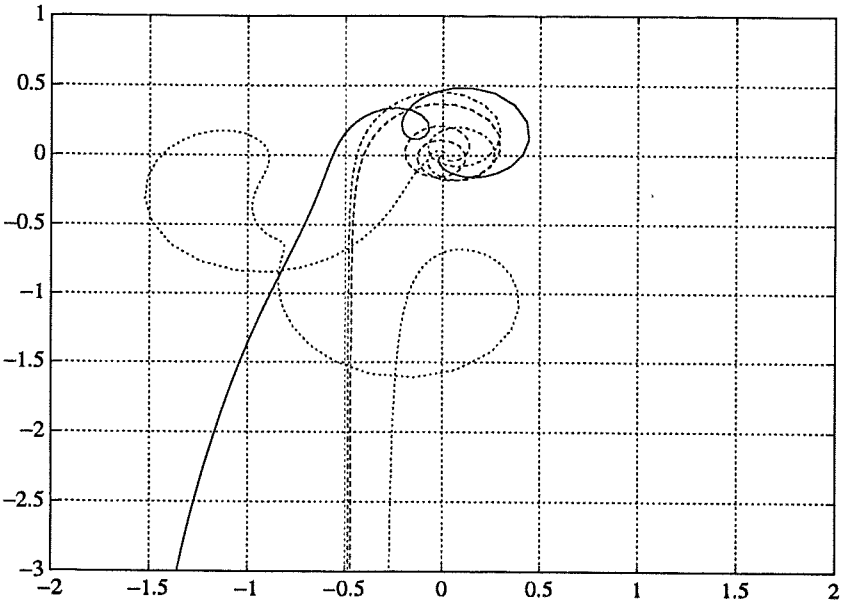


Figure 3.5 Open loop compensated Nyquist curves in four design cases for the process $G(s) = (-s+1)/(s^6+3s^5+5s^4+7s^3+5s^2+3s+1$. The three first cases are sixth order controllers where the magnitudes of the model and observer poles are $\omega_m = 1$ (solid curve), 1.4 (dashed curve) and 2 (dotted curve). The dash-dotted curve show a seventh order design with first order roll-off and $\omega_m = \omega_o = 1.4$.

The resulting open loop Nyquist curves are shown in Fig. 3.5 (top) for $\omega_m = 1$, 1.4 and 2. Notice the case $\omega_m = 2$, for which the controller has two poles in the right half plane. This forces the Nyquist curve of the open loop to encircle the point -1 in order to get a stable closed loop system. This gives a poor stability margin and a decrease of gain will result in instability. For $\omega_m = 1$ the Nyquist curve passes rather close to -1, while $\omega_m = 1.4$ yields a better stability margin. The case of an unstable controller results in quite large control actions as can be seen from the plot. The Bode plots of the feedback compensation S/R (Fig. 3.6) also shows the increase of compensator high frequency gain as the specified bandwidth increases. Included in both Fig. 3.5 and Fig. 3.6 is a feedback compensation of relative degree one which means that the degree of observer polynomial A_o must be increased from 6 to 7. The polynomials A_m and A_o are chosen as Butt$(6, 1.4, 45°)$ and Butt$(7, 1.4, 45°)$ respectively. The high frequency roll-off can be motivated by the attenuation of high frequency uncertainties in the process transfer function. By choosing even higher degree on the observer polynomial, a steeper high frequency roll-off is obtained. □

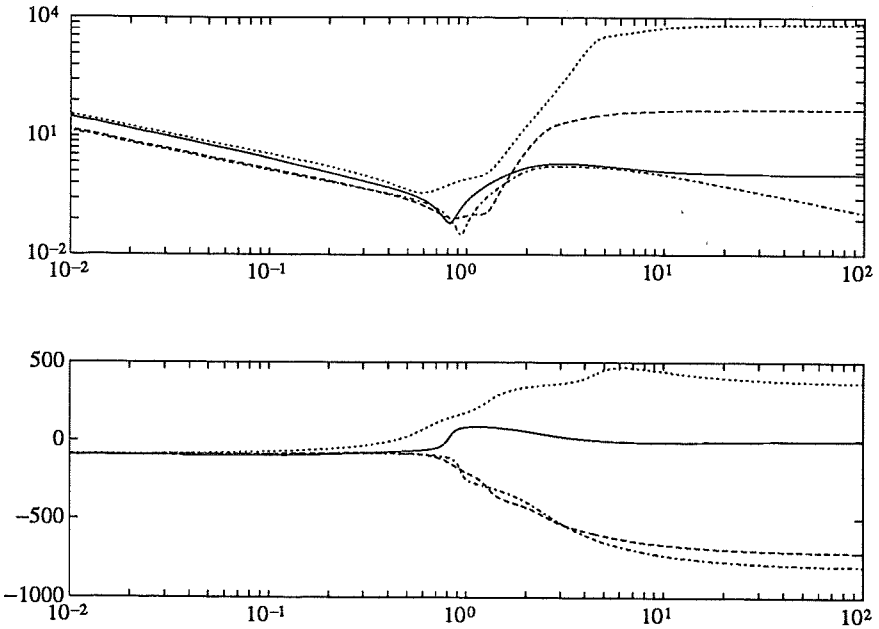


Figure 3.6 Bode plots of the feedback compensator in four design cases for the process $G(s) = (-s+1)/(s^6+3s^5+5s^4+7s^3+5s^2+3s+1$. The three first cases are sixth order controllers where the magnitudes of the model and observer poles are $\omega_m = 1$ (solid curve), 1.4 (dashed curve) and 2 (dotted curve). The dash-dotted curve shows a seventh order design with first order roll-off and $\omega_m = \omega_o = 1.4$.

3.3 Conclusions

The pole placement design method is popular because of its simplicity. Given the process dynamics and a specification of all closed loop poles there are straight forward procedures to calculate a controller that satisfies the specifications. There are, however, several situations where the method leads to a poor control design which has given the method a somewhat bad reputation among practitioners. The pole placement method has been reviewed in this chapter. It has been observed that the flexibility in the choice of pole locations is in fact a weakness, in the sense that it hides some of the important design trade-offs. Firstly, a high closed loop bandwidth requires a large controller gain. The controller may even be unstable. This is sometimes necessary in order to stabilize the plant but in connection with high closed loop bandwidths it may give very poor robustness as has been demonstrated. Secondly, the order of the controller is determined by the order of the plant. The controller order may therefore be high which may cause problems in the implementation of the controller.

Another drawback is that the order of the specified closed loop model must be of high order and often equal to the order of the process model. This means that it may be difficult to choose pole locations such that a desired closed loop performance is obtained.

4

Frequency Domain Least Squares Design

4.1 Introduction

In Chapter 3 some of the properties of pole placement design were discussed. A constraint appearing in pole placement design is that the order of the regulator is directly determined by the order of the process model, the order of the desired closed loop model and the order of the observer polynomial. One way to relax this constraint is to combine pole placement ideas with frequency domain methods. Many robustness properties are more transparent in the frequency domain. The idea is to replace the DAB equation Eq. (3.8) with the equation

$$\frac{G(s)T(s)}{R(s)+G(s)S(s)} = G_m(s) \tag{4.1}$$

for $s \in \mathcal{Z}$, where $\mathcal{Z}$ is some finite point set in the complex plane. Here, $G(s) = B(s)/A(s)$ and $G_m(s) = B_m(s)/A_m(s)$ are the open loop process transfer function and the desired closed loop transfer function respectively. However, in Eq. (4.1) there is no reason for restricting $G(s)$ to be rational, since only the values of G at a finite number of points has to be specified.

Another important observation is, as mentioned above, that the polynomials R, S and T and the order of the desired closed loop transfer function $G_m(s)$ may be chosen relatively independently in contrast to the pole placement design. This means, for example, that PI and PID controllers can be designed for a large class of process models.

A simple calculation gives that

$$E(s) = -1 + F_R(s)\frac{R(s)}{t_0} + F_S(s)S(s)\frac{S(s)}{t_0} \tag{4.4}$$

where

$$F_R(s) = \frac{G_m(s)}{G(s)A_o(s)}R(s)$$
$$F_S(s) = \frac{G_m(s)}{A_o(s)}S(s)$$

From Eq. 4.4 it follows that the controller parametrization

$$\theta = \frac{1}{t_0}(1, r_1, \ldots, r_{n_R}, s_0, \ldots, s_{n_S})^T$$

gives linearity in $E(s)$. Choose a set of complex numbers $\mathcal{Z} = \{z_i\}_{i=1}^N$ and define the loss function

$$J = \sum_{i=1}^{N} |E(z_i)|^2 \tag{4.5}$$

Let $n_R = \deg R$ and $n_S = \deg S$ and introduce the notation

$$\phi_R(s) = F_R(s)\begin{pmatrix} s^{n_R} & s^{n_R-1} & \ldots & 1 \end{pmatrix}$$
$$\phi_S(s) = F_S(s)\begin{pmatrix} s^{n_S} & s^{n_S-1} & \ldots & 1 \end{pmatrix}$$
$$\phi(s) = \begin{pmatrix} \phi_R(s) & \phi_S(s) \end{pmatrix}$$

$$\Phi = \begin{pmatrix} w_1\phi(z_1) \\ w_2\phi(z_2) \\ \vdots \\ w_N\phi(z_N) \end{pmatrix}, \quad \psi = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix}$$

The loss function can then be written as

$$J(\theta) = (\Phi\theta - \psi)^*(\Phi\theta - \psi)$$

where $*$ denotes conjugate transpose. Since J is a non-negative definite and quadratic it has a unique global minimum precisely when the matrix Φ has full column rank. Column rank deficiency occurs e.g. if

$$N < n_R + n_S + 2 \tag{4.6}$$

Eq. 4.1is solved in a least squares sense as described in Section 4.2. The number of approximation points may therefore be larger than the number of parameters.

The design method described in this chapter corresponds to method D' in Chapter 1. As an alternative, an intermediate low order process model $\hat{G}(s) = \hat{B}(s)/\hat{A}(s)$ can be used (method C' in Chapter 1). The DAB equation

$$\hat{A}(s)R(s) + \hat{B}(s)S(s) = A_m(s)A_o(s)$$

yields a low order controller which gives approximately the desired performance when applied to the model $G(s)$. The approximation $\hat{G}$ must then in particular be good at the frequency range where the approximate loop transfer function $\hat{G}S/R$ is near the point -1 (cf. Theorem 3.1). The method C' imposes some degree constraints on the polynomials R and S in contrast to the method D'. One advantage of using a low order process model $\hat{G}$ is that the Q parametrization described in Chapter 3 can be used to shape the loop transfer function.

4.2 Least squares fitting of controller parameters

The process model considered in this chapter will be slightly more general than in Eq. (3.1), namely

$$y = G(\frac{d}{dt})(u + d) \tag{4.2}$$

where G is a meromorph function. The same control structure as in Chapter 3 will be used namely

$$R(\frac{d}{dt})u = -S(\frac{d}{dt})y + T(\frac{d}{dt})r \tag{4.3}$$

where the polynomials R, S and T are given by

$$\begin{aligned} R(s) &= s^{n_R} + r_1 s^{n_R-1} + \cdots + r_{n_R} \\ S(s) &= s_0 s^{n_S} + s_1 s^{n_S-1} + \cdots + s_{n_S} \\ T(s) &= t_0 A_o(s) \end{aligned}$$

The polynomial $A_o(s)$,which can be interpreted as an observer polynomial (compare section 3.1), is specified by the user. Let the desired closed loop transfer function be given by $G_m(s)$. The actual closed loop transfer function is given by

$$G_{cl}(s) = \frac{G(s)T(s)}{R(s) + G(s)S(s)}$$

Introduce the relative closed loop model error

$$E(s) = \frac{G_{cl}(s) - G_m(s)}{G_{cl}(s)}$$

This can be avoided by choosing sufficiently many complex frequencies z_j. Another case, for which Φ lacks full column rank, is when there are solutions of lower order.

The approximation set $\mathcal{Z}$ is a finite point set in the complex plane. By imposing the condition that $\mathcal{Z}$ be self-conjugated it is guaranteed that the resulting polynomial coefficients will be real (see Lemma 2.1). Natural choices for $\mathcal{Z}$ are points on the imaginary axis (continuous time) or points on the unit circle (discrete time). This corresponds to fitting the ordinary Nyquist curve of the closed loop system to the Nyquist curve of the desired closed loop system at certain frequencies. Note that the special case $N = \deg R + \deg S + 2$,i.e. the number of approximation points equals the number of controller parameters, gives a controller where the closed loop transfer function interpolates the desired closed loop transfer function at the specified points.

By a simple modification of the functions $F_R(s)$ and $F_S(s)$ it is possible to incorporate pre-specified factors in the polynomials $R(s)$ and $S(s)$ by introducing $R(s) = R_1(s)R_2(s)$ and $S(s) = S_1(s)S_2(s)$ with R_1 and S_1 being the fixed parts of the respective polynomials. The condition Eq. 4.6 is modified to $N < n_{R_2} + n_{S_2} + 2$. For example, if integration is to be included in the controller, the pre-specified part of R is chosen as $R_1(s) = s$ and the function F_R is modified to $R_1(s)F_R(s) = sF_R(s)$.

Other transfer functions, such as

$$\frac{AS}{AR + BS}$$

(from noise n to output y) can be chosen as the functions to be approximated. This gives the possibility to "shape" different loops in a simple way.

4.3 Some design considerations

There are several design objectives that has to be specified. Some of the choices to be done are

- The approximation set $\mathcal{Z}$.
- The desired closed loop model G_m.
- The observer polynomial A_o.
- The order of the controller polynomials R, S and T.

The approximation set $\mathcal{Z}$

As mentioned above the approximation points are often chosen on the imaginary axis (continuous time) or on the unit circle (discrete time). Another alternative is to choose approximation points near the poles of the specified closed loop model and the observer poles. The motivation for this is to get a closed loop system with poles near the specified poles. This will of course be achieved at the cost of the approximation quality on the imaginary axis (on the unit circle in the discrete time case). In the following, the notation used in Chapter 2, $\Omega = \{\omega_1, \ldots, \omega_m\}$ for $\mathcal{Z} = \{i\omega_1, \ldots, i\omega_m, -i\omega_1, \ldots, -i\omega_m\}$, will be used and ω_k, $k = 1, \ldots, m$ will be referred to as the approximation frequencies.

The choice of approximation frequencies strongly affects the stability of both the controller and the resulting closed loop system. Too high approximation frequencies often results in an unstable controller depending on the order of the desired model G_m and the order of the controller (see below). Intuitively, this occurs when the model G_m is "too hard to follow" at high frequencies due to the required amount of phase lead. On the other hand, too low approximation frequencies may give a controller with poles in the right half plane corresponding to frequencies well above the approximation frequencies.

EXAMPLE 4.1—Flexible servo
Consider the system (from [Wallenborg, 1987])

$$G(s) = \frac{113.6(s^2 + 0.07s + 16)}{(s + 0.1)(s^2 + 0.4s + 125)} \tag{4.7}$$

to be controlled by a second order controller with integration with specifications given by $A_m(s) = \text{Butt}(3, 4, 45°)$ and $A_o(s) = \text{Butt}(2, 8, 45°)$ (for notation see Ex. 3.4). The approximation frequencies are chosen as $\Omega = \{0, 2, 4, 8\}$. This gives the unstable compensator

$$\frac{S(s)}{R(s)} = \frac{-0.30592s^2 - 3.1225s - 4.7401}{s(s - 9.9835)}$$

which yields an unstable closed loop system. By decreasing the approximation frequencies to $\Omega = \{0, 0.5, 1, 2\}$ the compensator becomes stable:

$$\frac{S(s)}{R(s)} = \frac{1.9406s^2 + 23.304s + 30.627}{s(s + 64.01)}$$

This compensator gives a stable closed loop system which matches the specification well as Fig. 4.1shows, where the step response of the closed loop system (solid curve) is shown together with the step response of the desired closed loop system (dotted curve). □

The desired closed loop model G_m

From the discussion on robustness in Chapter 3 it is understood that the desired closed loop model G_m should not deviate too much from the process model G, especially when controllers of low order are used. Let R, S and T satisfy Eq. (4.1) for some choice of approximation set $\mathcal{Z}$ (interpolation). The feedback compensation can then be written as

$$\frac{S(s)}{R(s)} = \frac{G_m(s)}{G(s)} \frac{1}{T(s)/S(s) - G_m(s)}, \quad s \in \mathcal{Z} \tag{4.8}$$

Equation (4.8) shows that the feedback gain must be large at points in $\mathcal{Z}$ where G_m/G is large which is the case when the bandwidth of the closed loop system is large compared to the open loop system. This also requires large positive phase in the controller. The order of the controller limits the amount of phase lead in the loop. For closed loop models of low order compared to the process model some non-minimum phase must therefore be included. One method is to use approximations of low order obtained from the methods in Chapter 2. Let

$$\hat{G}(s) = \frac{\hat{B}(s)}{\hat{A}(s)}$$

be such an approximation. As observed in Chapter 2, the polynomial $\hat{B}(s)$ in many cases has zeros in the right half plane. A method for choosing G_m then is

$$G_m(s) = \frac{b_0 \hat{B}(s)}{A_m(s)}$$

i.e. the zeros of the model are held fixed thus only varying the poles and the constant b_0. In the traditional pole placement this corresponds to keeping all non-minimum phase zeros of the process. However, since the zeros of $\hat{B}$ are not the "real" process zeros it is possible to "move" them in spite of the fact that they are located in the right half plane. This is another way of "speeding up the system" instead of moving the poles. It should be noted that no real cancellation of zeros takes place, but if the "virtual zeros" in the right half plane are moved too far, the resulting controller will typically be unstable. As pointed out in Ex. 3.2, this will often result in an unstable closed loop system.

EXAMPLE 4.2—Flexible servo
Continuing with Ex. 4.1 the closed loop specification is changed to $A_m(s) = \text{Butt}(3, 6, 45°)$ to make the closed loop response faster. Still using $A_o(s) = \text{Butt}(2, 8, 45°)$ and $\Omega = \{0, 0.5, 1, 2\}$ gives the compensator

$$\frac{S(s)}{R(s)} = \frac{-0.10872s^2 - 2.8008s - 4.9363}{s(s - 4.8939)}$$

which is unstable. Attempts to find a stable compensator by adjustment of the frequencies Ω fails. The closed loop system is also unstable. This means that in this case the order of the compensator must be increased to three, which ends up with the ordinary pole placement procedure, since both G and G_m are of third order. The third order controller is, however, unstable which may explain why all the computed second order compensators were unstable.

Returning to the second order design with $\Omega = \{0, 0.5, 1, 2\}$, $A_m(s) = \text{Butt}(3, 4, 45°)$ and $A_o(s) = \text{Butt}(2, 8, 45°)$ another modification will be done. The polynomial B_m is modified as $B_m(s) = b_0 B(s/p)$ where p is a constant such that $G_m(0) = 1$. Putting $p = 1.4$ gives the compensator

$$\frac{S(s)}{R(s)} = \frac{0.62505s^2 + 2.3987s + 3.113}{s(s + 1.6429)}$$

The step responses are seen in Fig. 4.1.

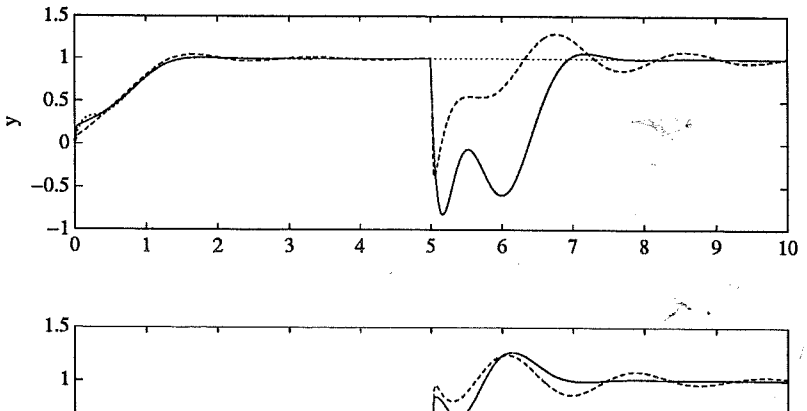


Figure 4.1 Nyquist curves of the flexible servo with second order compensation with integration for two cases. The specifications are in both cases given by $\Omega = \{0, 0.1, 0.2, 0.4\}$, $A_m(s) = \text{Butt}(3, 4, 45°)$ and $A_o(s) = \text{Butt}(3, 8, 45°)$. Two cases are shown, one for which $B_m(s) = b_0 B(s)$ (solid curve) and one with $B_m(s) = b_0 B(s/1.4)$ (dashed curve). The step response of the desired closed loop model is shown for the case $B_m(s) = b_0 B(s)$ (dotted curve).

The response to the load disturbance is faster but more oscillative in the case $p = 1.4$. □

The observer polynomial A_o

In conventional pole placement design, the observer polynomial A_o corresponds to the closed loop dynamics which is canceled from the reference input r to y and u. This is not the case for the design method described here since the zeros of A_o do not, in general, appear among poles of the actual closed loop system. The rules for choosing A_o is similar to when choosing A_m, namely not to move the roots too far away from the original poles since this might cause instability of the controller.

The order of the polynomials R, S and T

The degree of the polynomial T is given by $\deg T = \deg A_o$ since it is defined as a constant multiple of A_o. If $\deg A_o$ is chosen too low compared to $\deg R$ and $\deg S$ there might be solutions of lower order. As already mentioned, this results in loss of column rank in Φ. In certain cases, a too low order of A_o gives an unstable controller due to the fact that there may exist an exact solution which corresponds to a non-proper compensator S/R. In this case $\deg A_o$ must be increased. The polynomials R and S could also have almost common factors due to the fact that the closed loop specification $G_m(s)$ almost admits solutions of lower order. The orders of R and S should then be decreased, especially if the almost common zeros of R and S are located in the right half plane.

Another situation is that $G_m(s)$ might require more phase lead than is possible to achieve with the specified order of the controller. A diagnostic of this is that R possesses unstable zeros without any counterparts in S. By increasing the orders of R and S the controller will in most cases become stable.

EXAMPLE 4.3—Flexible servo
Keeping $\Omega = \{0, 0.5, 1, 2\}$ and $A_m(s) = \text{Butt}(3, 4, 45°)$ from the design in Ex. 4.1 while increasing ω_o in $A_o(s) = \text{Butt}(2, \omega_o, 45°)$ from 8 to 10 gives the unstable compensator

$$\frac{S(s)}{R(s)} = \frac{-0.71795s^2 - 12.421s - 17.105}{s(s - 34.561)}$$

Stability of the compensator is not achieved by modifying the approximation frequencies. The zeros of $S(s)$ are in the left half plane so there is no "near cancellation" of poles and zeros in the compensator. To get a stable controller the controller order must therefore be increased. □

Modification of the polynomial T

In some cases the closed loop response to a step in the reference input results in a large initial peak in the control signal. For a PID controller this corresponds to too much derivative action. By modifying the polynomial T while keeping R and S, this can be eliminated to some extent. The idea is to replace T by T_f, where $T_f(s)$ is an approximation of $T(s)F(s)$ for some low pass filter $F(s)$ and

$\deg T_f = \deg T$. A simple method is to use a least squares approximation in some frequency range. As an alternative to this, a pre-filter may be included after the reference input. This increases, however, the order of the controller.

Robust Performance

High gain control is a standard way to achieve robust performance. This leads, however, to sensitivity to measurement noise. Another drawback is that saturation of the control signal will typically occur, which leads to poor performance. The technique developed in this section can be employed to enhance the performance robustness without having to use high gain feedback.

Consider a set of process frequency responses given as transfer functions $\{G_1(s), G_2(s), \ldots, G_M(s)\}$ evaluated at certain frequencies $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$. The corresponding closed loop transfer functions are given by

$$G_{cl,j}(s) = \frac{G_j(s)T(s)}{R(s) + G_j(s)S(s)}, \quad j = 1, \ldots, M$$

and the relative closed loop model errors

$$E_j(s) = \frac{G_m(s) - G_{cl,j}(s)}{G_{cl,j}(s)}$$

Consider the loss function

$$J(\theta) = \sum_{j=1}^{M} \rho_j^2 \sum_{k=1}^{N} |E_j(i\omega_k)|^2 \tag{4.9}$$

where θ contains the controller parameters. Minimizing J can be interpreted as a compromise between M different designs since putting one weighting $\rho_j \neq 0$ and the rest of the weightings $= 0$ gives the loss function in Eq. 4.5 for the single process $G(s) = G_j(s)$.

EXAMPLE 4.4
The process models

$$G_j(s) = \frac{1}{s^3 + 3s^2 + a_j s + 2}, \quad j = 1, 2$$

are given, where $a_1 = 4$ and $a_2 = 3$ and the frequency responses are given at the frequencies $\Omega = \{0.02, 0.4, 0.8, 2\}$. The desired closed loop transfer function is chosen as

$$G_m(s) = \frac{1}{s^3 + 2.414s^2 + 2.414s + 1}$$

and the observer polynomial $A_o(s) = s^3 + 2.414s^2 + 2.414s + 1$ which gives a rather small bandwidth of the closed loop system so that the loop gain is kept at

a moderate level. The controller is chosen to be of third order with integration. Consider the minimization of the criterion Eq.4.9. Introduce the normalization $\rho_2 = 1$. Two limiting cases are recognized, namely $\rho_1 \to \infty$ and $\rho_1 \to 0$. As $\rho_1 \to \infty$ the polynomials R and S approaches the solution obtained by the conventional pole placement method using the process model G_1 and similarly $\rho_1 \to 0$ corresponds to using G_2 as the nominal process model. The step responses for the closed loop systems together with the desired closed loop step response are shown in Fig.4.2. for $\rho_1 = 10^{-5}$, 0.1, 1 and 10000. □

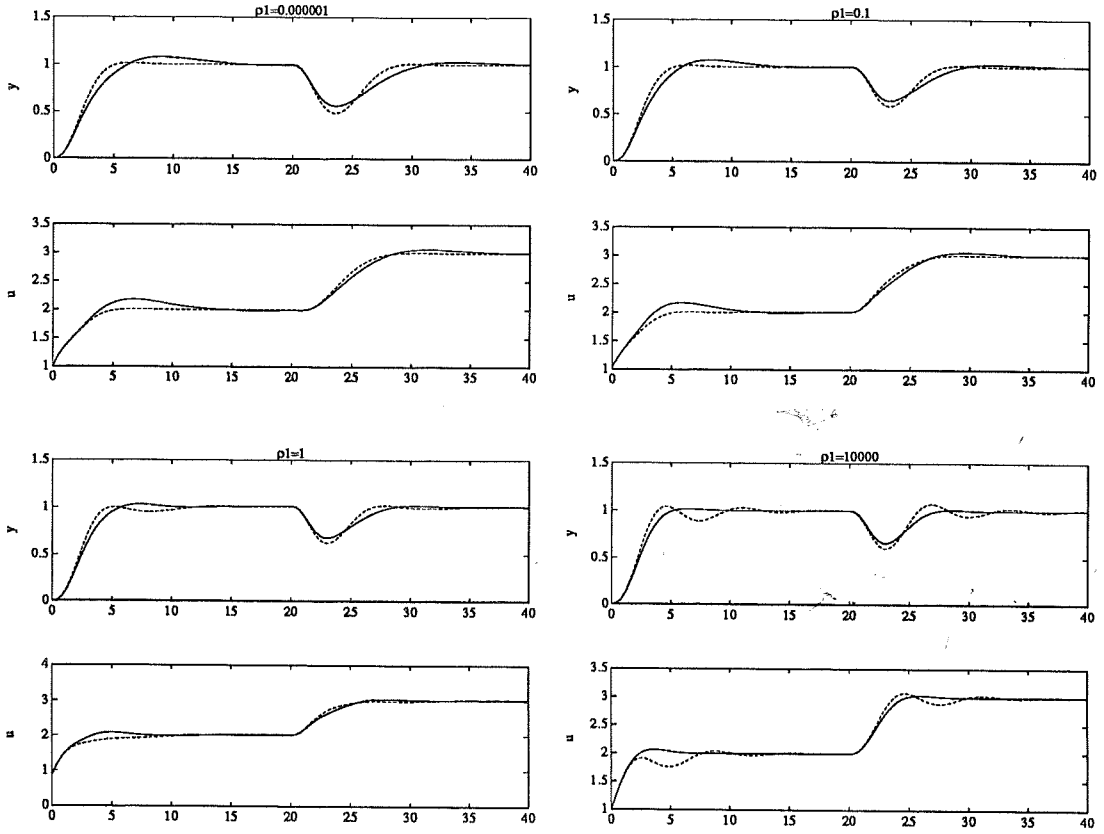


Figure 4.2 Closed loop step responses for 4 different weightings ($\rho_1 = 10^{-5}$, 0.1, 1 and 10000) for the 2-process example (solid and dashed curves). The approximation frequencies were chosen as $\Omega = \{0.02, 0.2, 0.4, 2\}$.

4.4 Some examples

To illustrate the method described in Section 4.2 some controller designs of low order will be done in this section. Integration is included and each example starts with the lowest possible order (i.e. one).

EXAMPLE 4.5—Eight cascaded low pass filters
The system

$$G(s) = \frac{1}{(s+1)^8}$$

Starting with the choice of approximation frequencies, it is important to know for which frequencies the Nyquist curve intersects the real and imaginary axes. These frequencies will be denoted by ω_{-90°, ω_{-180°, ... with obvious notation. For this specific case we have that $\omega_{-n\cdot 90^\circ} = \tan n\frac{\pi}{16}$, $n = 1, 2, \ldots$. The numerical values for the first frequencies are $\omega_{-90^\circ} \approx 0.199$, $\omega_{-180^\circ} \approx 0.414$ and $\omega_{-270^\circ} \approx 0.668$.

First order controller

To begin with, a first order controller with integration (a PI-controller) will be designed. The phase of the controller varies between -90° and 0°. This means that the bandwidth of the closed loop system should not exceed the bandwidth of the open loop non-compensated system. Taking the values of the "axis frequencies" into account the approximation frequencies are chosen as $\Omega = \{0.01, 0.1, 0.2\}$. The desired closed loop model is somewhat arbitrarily chosen to be of third order with relative degree one. Before specifying this, a third order process model is fitted to $G(s)$ at the chosen frequencies. This gives

$$\hat{G}(s) = \frac{\hat{B}(s)}{\hat{A}(s)} = \frac{0.1229s^2 - 0.1607s + 0.0807}{s^3 + 1.0955s^2 + 0.4847s + 0.0807}$$

The Nyquist curves for $G(s)$ and $\hat{G}(s)$ fits very well for frequencies below 0.2 rad/s as Figure 4.3 indicates. The zeros of $\hat{B}(s)$, $0.653 \pm i0.479$, are kept in the model $G_m(s)$. The polynomials A_m and A_o are both chosen to Butt$(3, 0.3, 45^\circ)$. A first order controller with integration obtained by fitting at the frequencies $\Omega = \{0, 0.01, 0.1, 0.2\}$ results in the closed loop Nyquist curves shown in Figure 4.4. The zero frequency is included in order to get a small error in the closed loop static gain. By choosing a sufficiently large weighting for this frequency the error can be made arbitrarily small. In general there may be two sets of frequencies – Ω_1 at which very small errors are required (nearly interpolation) and Ω_2 with larger tolerance. This can be achieved by choosing weightings of Ω_1 much larger than the weightings of Ω_2. The controller output u and the process output y in closed loop are shown in Figure 4.5 for the case when the reference signal r is a unit step at $t = 0$ and the load disturbance d is a unit step at $t = 50$. The step response of the desired model from r to y is also included in the plot. The step responses follows very close. If the desired closed loop poles are moved further out from the origin the fitting will deteriorate and when the polynomials A_m and A_o are chosen as Butt$(3, \omega_m, 45^\circ)$ with $\omega_m = 1.1$ the actual closed loop system is unstable.

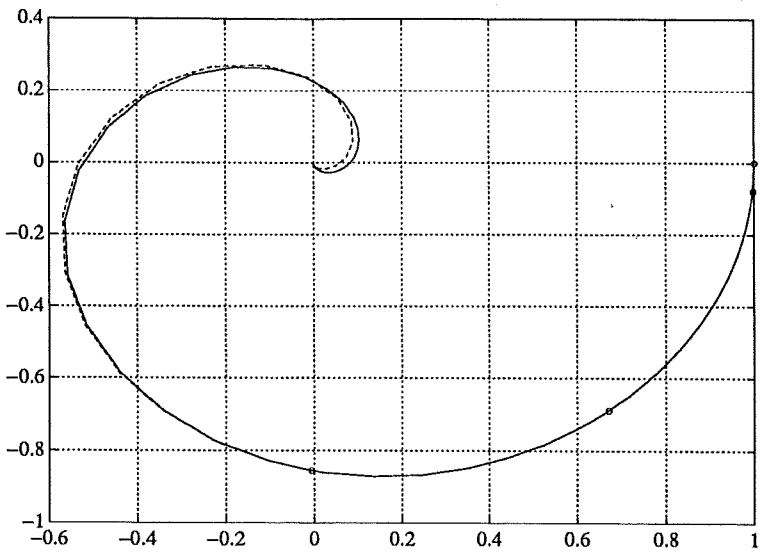


Figure 4.3 Nyquist curves of $G(s) = 1/(s+1)^8$ (dashed curve) and a third order approximation $\hat{G}(s)$ (solid curve). The approximation frequencies $\Omega = \{0.01, 0.1, 0.2\}$ are marked with o's.

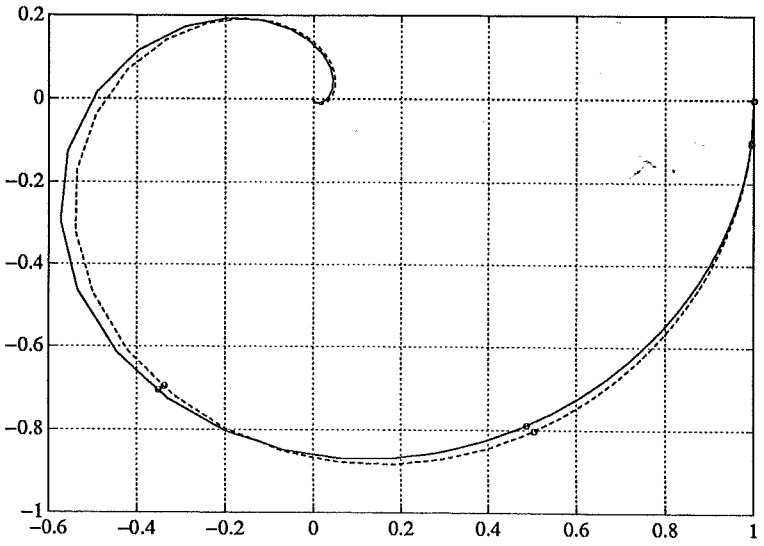


Figure 4.4 The plant $G(s) = (s+1)^{-8}$ controlled by a first order controller with integration. Nyquist curves of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies are given by $\Omega = \{0, 0.01, 0.1, 0.2\}$.

Second order controller

By increasing the order of the controller, it is possible to get more phase lead in the loop which means that the bandwidth can be further increased. A second

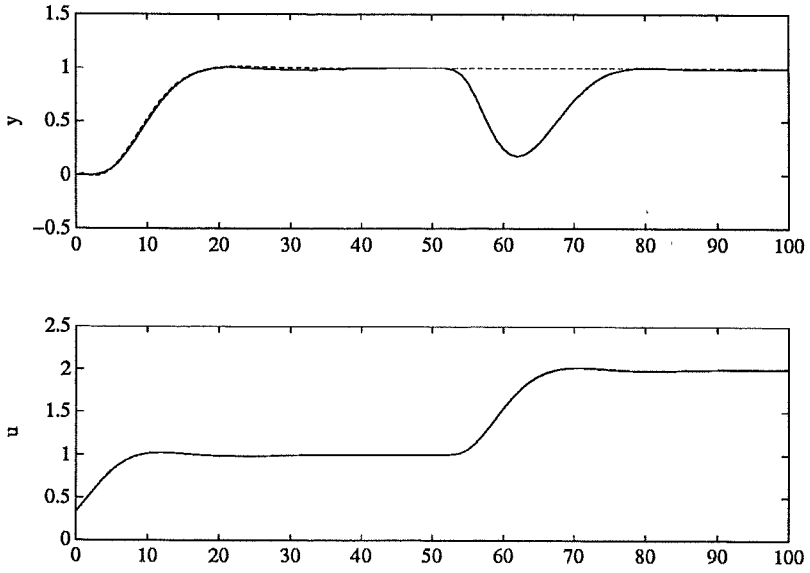


Figure 4.5 The plant $G(s) = (s+1)^{-8}$ controlled by a first order controller with integration. Step responses of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies are given by $\Omega = \{0, 0.01, 0.1, 0.2\}$.

order controller with integration (PID-controller) is computed using the approximation frequencies $\Omega = \{0, 0.01, 0.1, 0.2, 0.4\}$. The desired closed loop poles and the observer poles are moved to the radius $\omega_m = 0.4$ yielding the Nyquist curves in Figure 4.6. The step response of the closed loop system is well in accordance with the specified model as Figure 4.7 shows.

When the specified poles are moved further out from the origin, the design procedure gives an unstable controller. The reason for this is intuitively that more phase lead is required (see Example 3.2). The only way a fixed order controller can give such a phase advance is to make it unstable.

The resulting closed loop system will also be unstable. This is due to the fact that the instability of the controller implies that the Nyquist curve should encircle the point -1. This is not the case, since the approximation frequencies are much lower than the frequencies for which the loop Nyquist curve is close to -1. In fact, the instability occurs for much smaller values of ω_m (≈ 0.47) than in the first order case. The reason for this is that in the first order case the single controller pole is fixed to zero, so that there is no possibility to have any controller poles in the open right half plane. This clearly indicates that unstable controllers are not desirable.

Third order controller

The bandwidth of the closed loop system can be made larger if the controller order is increased. In the third order case the values $\omega_m = 0.5$ and $\omega_o = 0.5$ were chosen and the approximation frequencies were the same as in the second

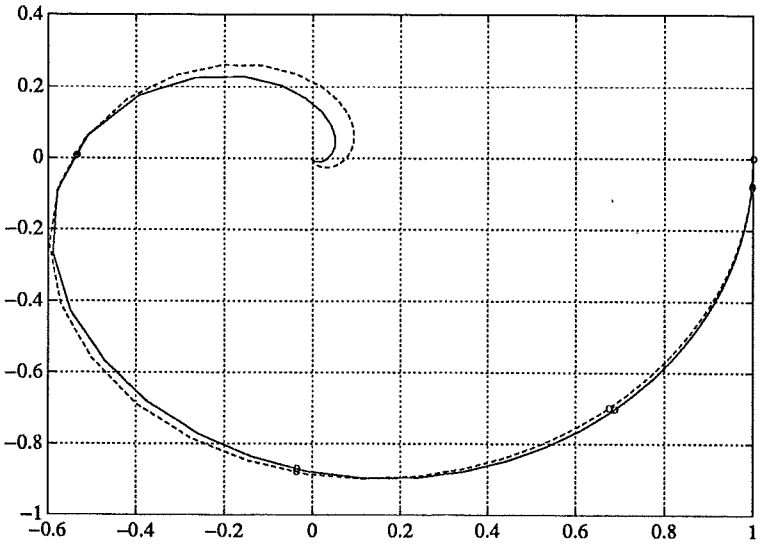


Figure 4.6 The plant $G(s) = (s+1)^{-8}$ controlled by a second order controller with integration. Nyquist curves of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies are given by $\Omega = \{0, 0.01, 0.1, 0.2, 0.4\}$ and are marked by o's.

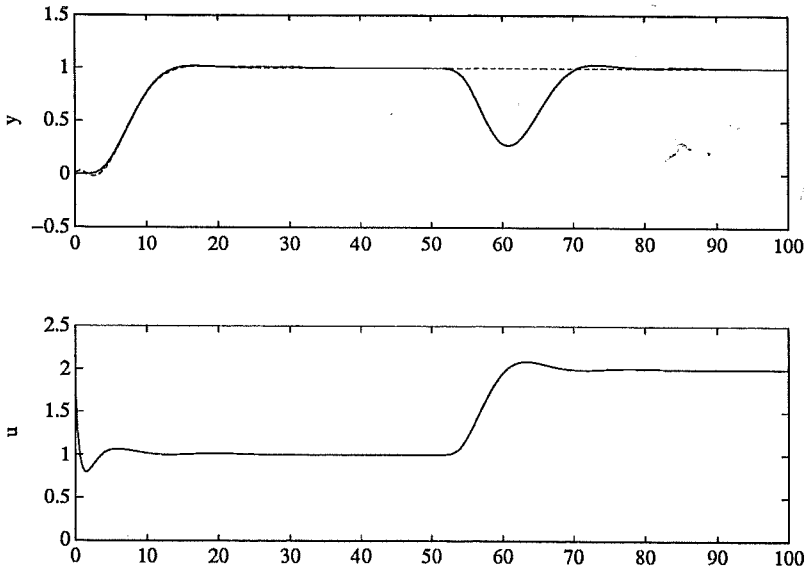


Figure 4.7 The plant $G(s) = (s+1)^{-8}$ controlled by a second order controller with integration. Step responses of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies are given by $\Omega = \{0, 0.01, 0.1, 0.2, 0.4\}$.

order case, namely $\Omega = \{0, 0.01, 0.1, 0.2, 0.4\}$. The Nyquist curves of the closed loop system and the reference model fits well together below $\omega = 0.4$ (Fig.4.8).

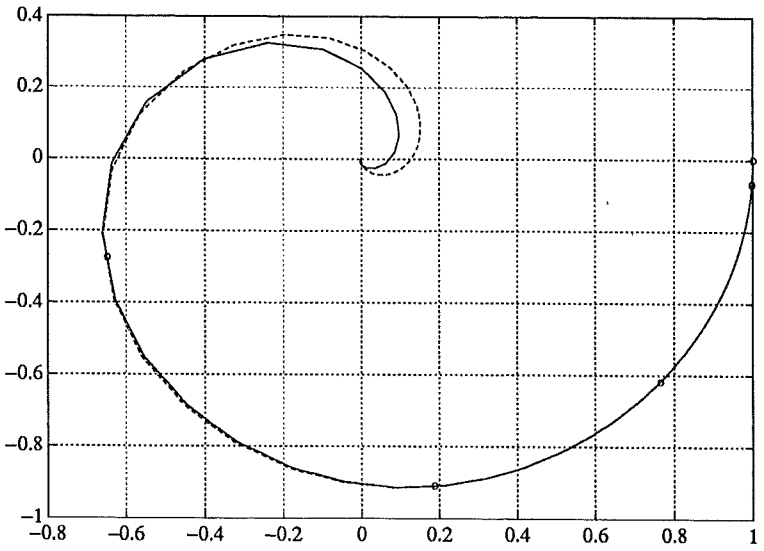


Figure 4.8 The plant $G(s) = (s+1)^{-8}$ controlled by a second order controller with integration. Nyquist curves of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies are given by $\Omega = \{0, 0.01, 0.1, 0.2, 0.4\}$ and are marked by o's.

In this case the controller poles are relatively well damped but the controller will eventually become unstable as the desired bandwidth is increased. The actual and desired closed loop step responses are shown in Figure 4.9.

Higher order controllers

As the order of the controller is increased, some caution has to be taken. The number of approximation points (frequencies) must be sufficiently large. Also, the order of the desired closed loop model may have to be increased to get a stable controller. Another possibility is to force the controller to be stable by keeping the polynomial R fixed to a stable polynomial R_1. The order of the controller may then have to be increased to get a good Nyquist curve fitting. As mentioned in Section 3.2 the freedom to choose R_1 can be utilized to adjust the loop Nyquist curve without changing the closed loop Nyquist curve too much. □

EXAMPLE 4.6—Sixth order doubly resonant system
Designing PI or PID compensators for the system

$$G(s) = \frac{-s+1}{s^6 + 3s^5 + 5s^4 + 7s^3 + 5s^2 + 3s + 1}$$

is a non-trivial task mainly due to the fact that the phase decreases rather in a relatively short frequency interval. This is caused by the two resonances, which are rather close together, and the right half plane zero. Computing the first intersec-

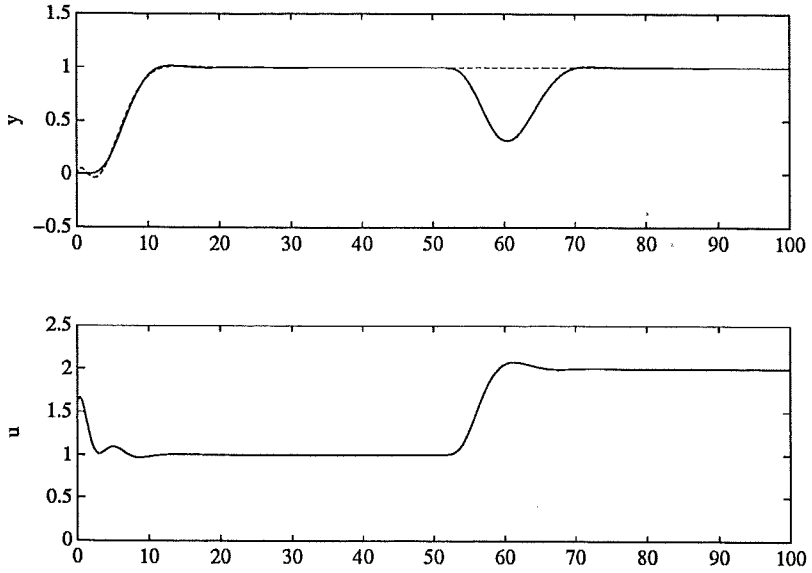


Figure 4.9 The plant $G(s) = (s+1)^{-8}$ controlled by a third order controller with integration. Step responses of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies are given by $\Omega = \{0, 0.01, 0.1, 0.2, 0.4\}$.

tions between the Nyquist curve and the axes gives the frequencies $\omega_{-90^\circ} \approx 0.40$, $\omega_{-180^\circ} \approx 0.69$ and $\omega_{-270^\circ} \approx 0.86$.

First order controller

Considering the first "axis frequency" ω_{-90° a reasonable set of approximation frequencies is $\Omega = \{0, 0.1, 0.2, 0.4\}$. A third order unweighted LS-approximation of the process at these frequencies is given by

$$\hat{G}(s) = \frac{\hat{B}(s)}{\hat{A}(s)} = \frac{0.2784s^2 - 0.5647s + 0.3107}{s^3 + 0.816s^2 + 0.678s + 0.3107}$$

As Figure 4.10 shows, the approximation is satisfactorily in the range of Ω. The complex conjugated pole pair of the approximate model is rather close to the slower resonant pole pair of the process. The polynomial $\hat{B}$ is used in a third order desired closed loop model. Using a first order controller (PI-controller) means limited opportunities to damp the resonant modes. The lower resonance frequency of the process is approximately 0.745. A realistic choice of bandwidth is therefore below this frequency. A design using the polynomial $A_m(s) = \text{Butt}(3, 0.4, 45^\circ)$ and the observer polynomial $A_o(s) = s + 0.6$ gives the Nyquist curves in Figure 4.11. This choice of A_m gives a desired closed loop bandwidth of $\omega_B \approx 0.32$ which is less than half the lower resonance frequency of the process model. Moving the observer pole -1 farther into the left half plane gives a more oscillatory response to the load disturbance. The weighting for $\omega = 0$ was increased in order to obtain

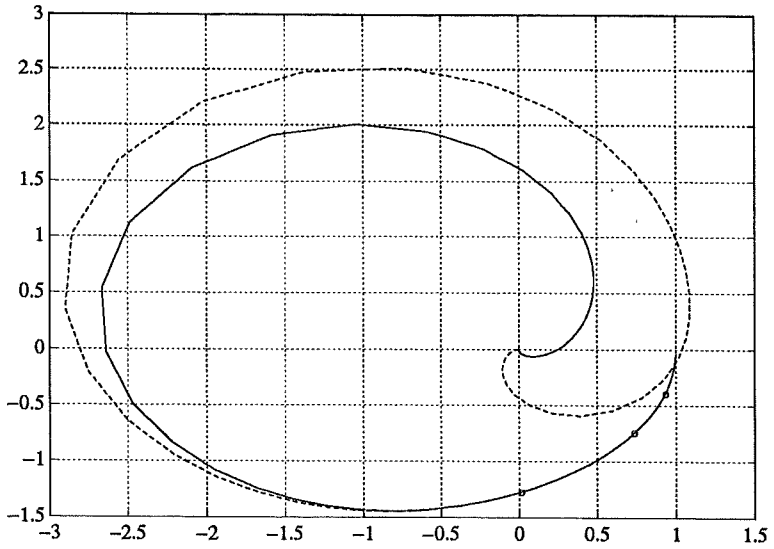


Figure 4.10 Nyquist curves for $G(s) = (1-s)/(s^6+3s^5+5s^4+7s^3+5s^2+3s+1)$ (dashed curve) and a $(2,3)$ LS-approximation $\hat{G}(s)$ (solid curve).

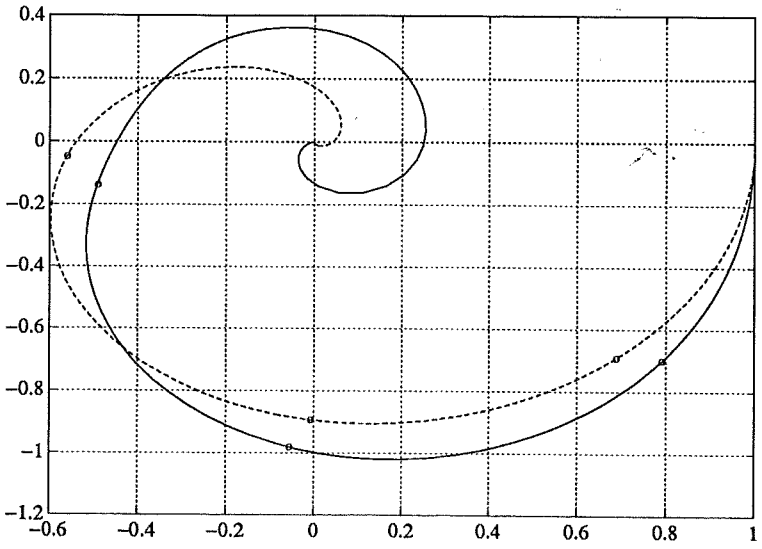


Figure 4.11 The plant $G(s) = (1-s)/(s^6+3s^5+5s^4+7s^3+5s^2+3s+1)$ controlled by a first order controller with integration. Nyquist curves of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies $\Omega = \{0, 0.1, 0.2, 0.4\}$ are marked with o's. The polynomials A_m and A_o were chosen as $A_m(s) = \text{Butt}(3, 0.4, 45°)$ and $A_o(s) = s + 0.6$.

the correct stationary gain. The desired and actual closed loop step responses

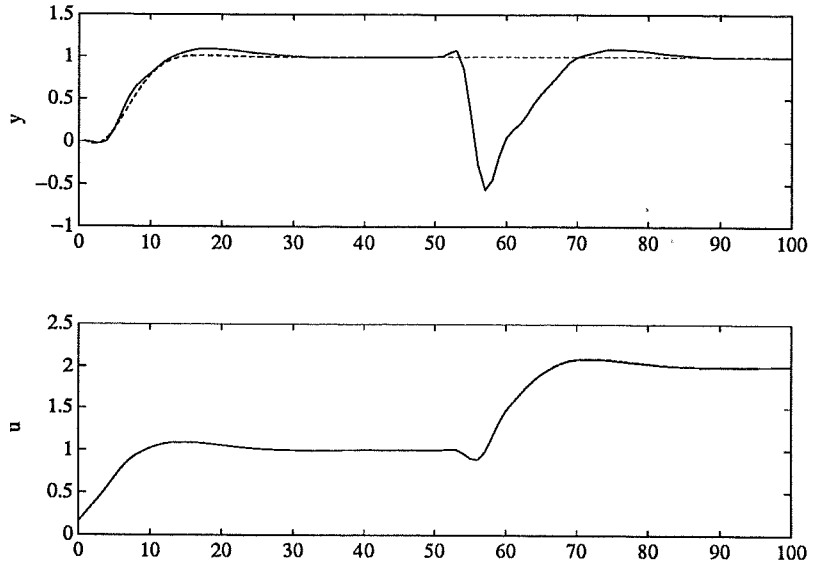


Figure 4.12 The plant $G(s) = (1-s)/(s^6+3s^5+5s^4+7s^3+5s^2+3s+1)$ controlled by a first order controller with integration. Step responses of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies are given by $\Omega = \{0, 0.1, 0.2, 0.4\}$. The polynomials A_m and A_o were chosen as $A_m(s) = \text{Butt}(3, 0.4, 45°)$ and $A_o(s) = s + 0.6$.

are seen in Fig. 4.12. The resulting feedback compensation is

$$\frac{S(s)}{R(s)} = \frac{-0.1181s + 0.1067}{s}$$

The polynomial S has a zero in the right half plane. The intuition behind this is that in order to damp the resonances the compensation must have just enough phase lag in order to "rotate" the loop Nyquist curve clockwise so that the large magnitude at the lower resonance "points" into the right half plane.

Second order controller

Using $\Omega = \{0, 0.1, 0.2, 0.4\}$, $\omega_m = 0.4$ and $\omega_o = 0.6$ as in the first order case, gives the closed loop Nyquist curve shown in Figure 4.13. The resonant modes are, as expected, much better damped than in the first order case. The actual closed loop system also gives a significantly better approximation of the desired closed loop system. This is also visible in the closed loop step responses (see Fig. 4.14). In this case the feedback compensation is given by

$$\frac{S(s)}{R(s)} = \frac{-0.0992s^2 + 0.1232s + 0.04725}{s^2 + 0.3648s}$$

This corresponds to a PI-compensator similar to the one in the previous case, cascaded with a lead compensation. The lead compensation has maximum phase at $\omega \approx 0.33$, which is approximately the bandwidth of the closed loop system.

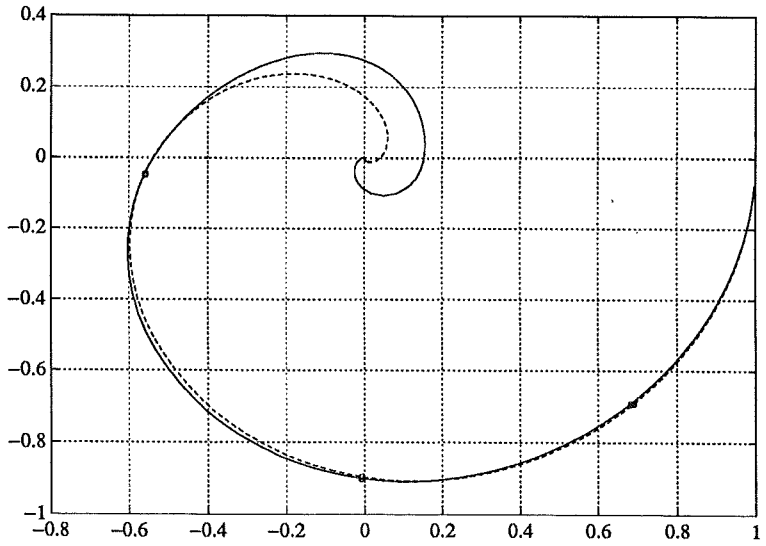


Figure 4.13 The plant $G(s) = (1-s)/(s^6+3s^5+5s^4+7s^3+5s^2+3s+1)$ controlled by a second order controller with integration. Nyquist curves of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies $\Omega = \{0, 0.1, 0.2, 0.4\}$ are marked with o's. The polynomials A_m and A_o were chosen as $A_m(s) = \text{Butt}(3, 0.4, 45°)$ and $A_o(s) = \text{Butt}(2, 0.6, 45°)$.

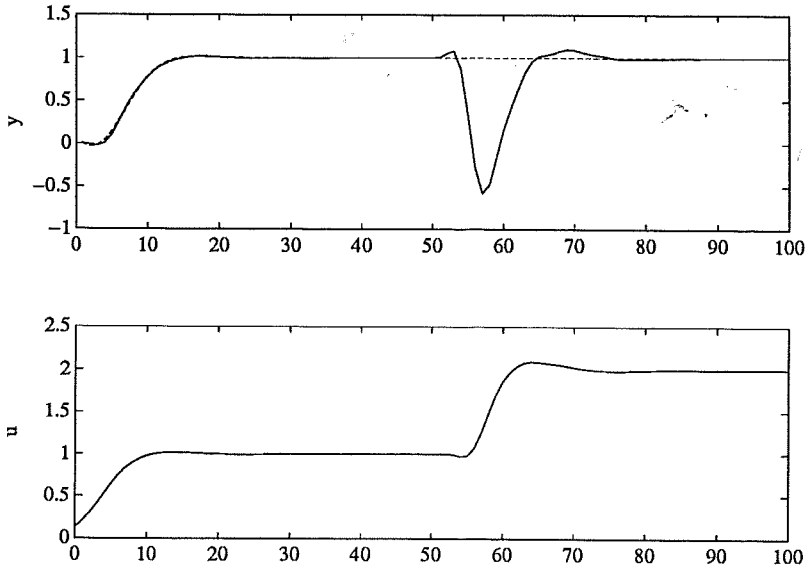


Figure 4.14 The plant $G(s) = (1-s)/(s^6+3s^5+5s^4+7s^3+5s^2+3s+1)$ controlled by a second order controller with integration. Step responses of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies are given by $\Omega = \{0, 0.1, 0.2, 0.4\}$. The polynomials A_m and A_o were chosen as $A_m(s) = \text{Butt}(3, 0.4, 45°)$ and $A_o(s) = \text{Butt}(2, 0.6, 45°)$.

Increasing ω_m and ω_o will give a poorer damping of the resonances. The non-zero pole of the controller will eventually move into the right half plane. This can not be remedied by changing the approximation frequencies.

Third order controller

Increasing the controller order to three gives more phase lead capability. The bandwidth can be increased a little higher, compared to the second order case, before oscillations appear in the closed loop step response. A third order controller with integration is computed using $\Omega = \{0, 0.1, 0.2, 0.4\}$, $\omega_m = 0.5$ and $\omega_o = 0.7$. The closed loop Nyquist curve approximates the desired one rather well at frequencies in the range of Ω (see Fig. 4.15) but above $\omega = 1$ the approximation is not quite as good.

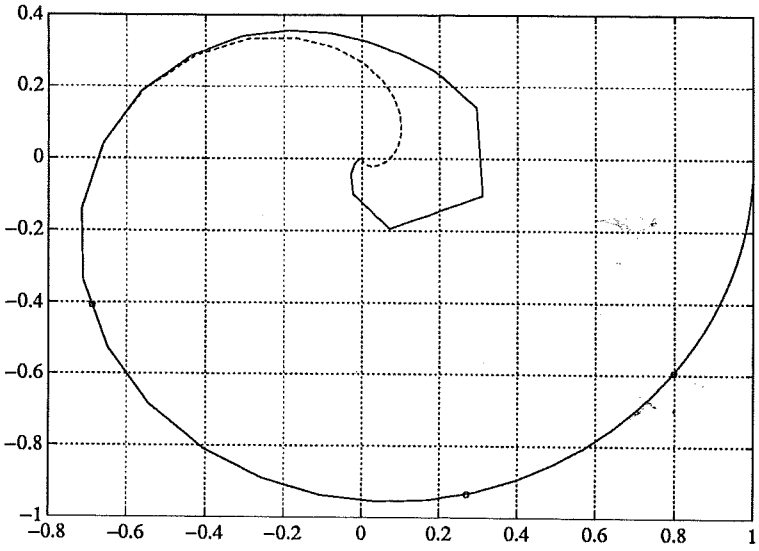


Figure 4.15 The plant $G(s) = (1-s)/(s^6+3s^5+5s^4+7s^3+5s^2+3s+1)$ controlled by a third order controller with integration. Nyquist curves of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies $\Omega = \{0, 0.1, 0.2, 0.4\}$ are marked with o's. The polynomials A_m and A_o were chosen as $A_m(s) = \text{Butt}(3, 0.5, 45°)$ and $A_o(s) = \text{Butt}(3, 0.7, 45°)$.

The approximation gets slightly better when increasing the approximation frequencies. It is important to update the approximate process model $\hat{G}$ when changing approximation frequencies. Incorporating frequencies above the higher resonance gives an unstable controller. The step responses are shown in Fig. 4.16. □

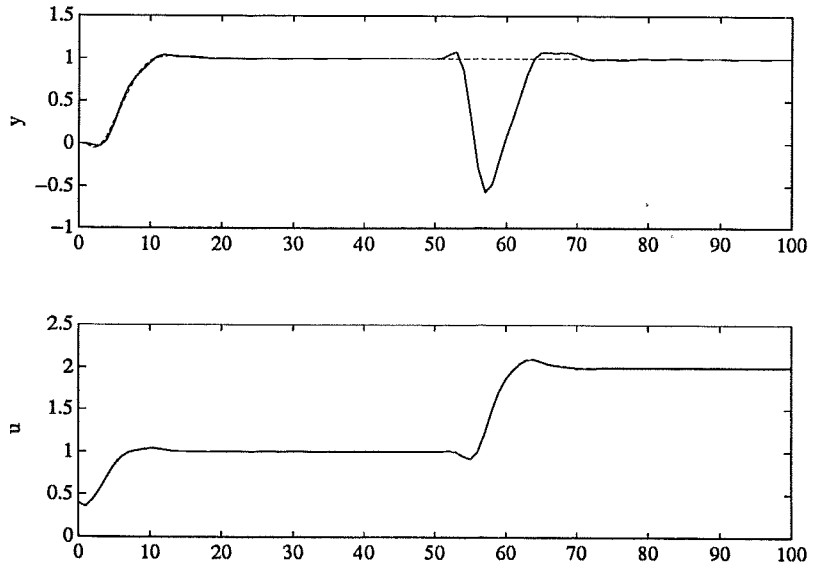


Figure 4.16 The plant $G(s) = (1-s)/(s^6+3s^5+5s^4+7s^3+5s^2+3s+1)$ controlled by a third order controller with integration. Step responses of desired closed loop system (dashed curve) and actual closed loop system (solid curve). The fitting frequencies are given by $\Omega = \{0, 0.1, 0.2, 0.4\}$. The polynomials A_m and A_o were chosen as $A_m(s) = \text{Butt}(3, 0.5, 45°)$ and $A_o(s) = \text{Butt}(3, 0.7, 45°)$.

4.5 Conclusions

A new method for design of low order controllers for stable SISO systems has been proposed. The method uses a least squares method to find a linear controller of specified order from a number of points on the Nyquist curve of the process. Since the design is made in the frequency domain the process model may be of infinite order. By repeating the optimization for controllers having different complexity it is possible to find the benefits of increasing the order of the controller. The design variables consist of the approximation frequencies Ω (or more generally $\mathcal{Z}$), the desired closed loop transfer function $G_m(s) = B_m(s)/A_m(s)$, the observer polynomial $A_o(s)$, and the orders of the controller polynomials R and S.

The method combines good features of classical frequency domain design methods and pole placement. The advantage with pole placement design is that it is possible to understand how the design variables influences the transient response properties of the closed loop system. The drawback is that it is very difficult to see the influences on the controller gain. The classical frequency domain design methods have the opposite properties. In this case it is difficult to see how the design variables affect the transient response while it is easy to see the influence on the controller gain. The new method suggested in this chapter combines the attractive properties from each of the two design techniques.

The proposed method is believed to be a good starting point for design of PI and PID controllers from measured plant frequency responses thereby continuing

and refining the well-known method of Ziegler and Nichols. The present method is illustrated by some examples where the order of the controllers is considerably lower than the order of the process. The controllers thus obtained are shown to give good accuracy with respect to the desired closed loop step response.

5

Robust Control Design

5.1 Introduction

Ever since the work in [Bode, 1945], frequency domain methods have been used in the design of control systems. One motivation for Bodes work was to design feedback systems that are insensitive to process variations. This viewpoint was further emphasized and developed in [Horowitz, 1963]. The development of state space methods in the fifties tended, however, to de-emphasize the "classical" frequency domain techniques. At the same time less interest was given to the problem of plant variation. A renewed interest in this problem has occurred in the last decade through the LQG-LTR approach [Doyle and Stein, 1981], [Stein and Athans, 1987] and the H_∞-methods [Zames, 1981], [Francis, Helton and Zames, 1984], [Glover and Doyle, 1988].

In the Horowitz method, a structure with two degrees of freedom is used (Fig. 5.1).

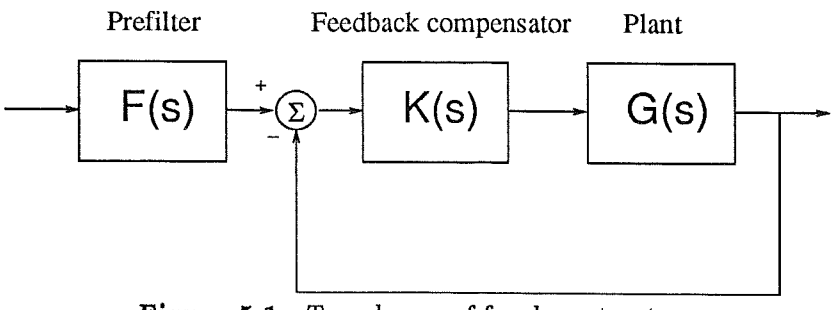


Figure 5.1 Two degree of freedom structure

The feedback compensator $K(s)$ is designed with respect to the uncertainties in the process. This is done using graphical methods to account for variations in the process frequency response. The uncertainties are given as "templates" (see next section) for a number of frequencies. The idea is to keep the loop gain just high enough to meet the requirements of robustness against the plant variations specified by the templates. The actual computation of the compensator $K(s)$ is done by using a "patching technique" where a number of lead and lag compensators are successively cascaded. This often results in a compensator of excessively high order. The pre-filter $F(s)$ is used to shape the closed loop response and is typically chosen as to decrease the bandwidth of the closed loop system.

The LQG-LTR procedure consists of an LQG design where the the robustness properties of the loop transfer function from the full state LQ design (infinite gain margin and at least 60° phase margin) are recovered in the design of the Kalman-Bucy filter. This is done by introducing fictitious noise with special properties. The resulting loop transfer function converges pointwise on the imaginary axis to the LQ loop transfer function as a certain parameter q goes to infinity. This means that the LQG-LTR loop Nyquist curve will approach $K/i\omega$ at large frequencies as $q \to \infty$. The sensitivity to noise and high frequency unmodeled dynamics is thereby increased. There is also a dual version of this method starting with the Kalman filter design and then computing a sequence of LQ-regulators which recover the loop transfer function of the pure filter case.

The methods involving H_∞-theory are variations on the theme: Minimize the H_∞-norm of the weighted sensitivity function

$$\|W_1(I+GK)^{-1}W_2\|_\infty$$

over all controllers $K(s)$ achieving internal stability, i.e. for which the transfer function

$$\begin{pmatrix} (I+GK)^{-1} & (I+GK)^{-1}G \\ K(I+GK)^{-1} & K(I+GK)^{-1}G \end{pmatrix}$$

is stable. In the single input single output case, this corresponds to maximizing the minimal distance between the open loop Nyquist curve and the point -1. The uncertainty is thus described by a circular disc at each frequency. This gives a rather conservative robustness.

The method presented in this chapter is most closely related to the Horowitz method. The novelty of this approach is that the computation of the compensator is viewed as an approximation problem. Using a least squares criterion a compensator of specified order is computed to match the desired open loop frequency response at certain frequencies.

5.2 Description of plant uncertainty

To describe model uncertainty in the frequency domain it is convenient to introduce sets called *templates* (compare [Horowitz, 1982]). A template for a certain frequency ω is defined as a subset of the complex plane given by $G(i\omega, \mathcal{P}) = \{G(i\omega, p) : p \in \mathcal{P}\}$ where $\mathcal{P}$ is a possibly infinite index set.

Plant uncertainty can be classified according to

- Structured uncertainty
  - Parametric
  - Non-parametric
- Unstructured uncertainty
  - Additive
  - Multiplicative
  - Coprime factorization

Parametric structured uncertainty

A plant model with parametric structured uncertainty is given by a family of transfer functions, e.g.

$$G(s,p) = \frac{b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n}{s^n + a_1 s^{n-1} + \cdots + a_n} e^{-\tau s}$$

where p is a parameter vector containing the coefficients b_j and a_j together with the time delay τ. The parameter vector belongs to a subset $\mathcal{P}$ of $\mathbb{R}^{2n+1}$. Often, $\mathcal{P}$ is chosen as a bounded hypercube in $\mathbb{R}^{2n+1}$, i.e. a cartesian product of bounded intervals.

Non-parametric structured uncertainty

Another way to represent structured plant uncertainty is by a number of measured frequency responses. In a way, this could be classified as a parametric representation, namely

$$G(i\omega, p) = \{G_j(i\omega)\}_{j=1}^{M}$$

where M is the number of frequency responses and $p = \{1, 2, \ldots, M\}$. Nevertheless, it seems more natural to regard this as a non-parametric way to describe uncertainty. A typical situation is when frequency responses are measured for different operating conditions. The templates for each frequency consists in this case of M points in the complex plane.

Additive unstructured uncertainty

Additive unstructured uncertainties are defined by

$$G(s) = G_0(s) + \Delta_G(s)$$

where $G_0(s)$ is the nominal transfer function and $\Delta_G(s)$ is an asymptotically stable perturbation which is bounded on the imaginary axis according to

$$|\Delta_G(i\omega)| \leq M(\omega)$$

for some bounded function M. In this case, each template is a circular disc with radius equal to $M(\omega)$.

Multiplicative unstructured uncertainty

Multiplicative unstructured perturbations can be written as

$$G(s) = G_0(s)(1 + \delta_G(s))$$

where δ_G is asymptotically stable with

$$|\delta_G(i\omega)| \leq m(\omega)$$

for some bounded function $m(\omega)$. This relative error formulation is advantageous when working in Bode or Nichols diagrams. Each template is a circular disc with radius $|G_0(i\omega)|m(\omega)$.

Coprime factorization unstructured uncertainty

The third and latest approach to unstructured uncertainties is the stable coprime factorization formulation:

$$G(s) = N(s)D^{-1}(s) = (N_0(s) + \Delta_N(s))(D_0(s) + \Delta_D(s))^{-1}$$

where N and D are coprime elements of H_∞. A clear advantage with this approach is that the perturbed model need not have the same number of poles in the right half plane as the unperturbed model. Also, there is attractive solution methods available for the corresponding H_∞ sensitivity minimization problem. There is no simple characterization of the templates for this approach since the perturbations appear both in the numerator and in the denominator of the transfer function.

In the following, all uncertainties will be of the structured type described by parameter variations in a transfer function. Instead of using high gain control to eliminate variations in the open loop transfer function the idea is to shape the transfer function in such a way that the closed loop behavior varies in a

"pleasant" way. A typical example is to allow the response speed to change while the damping is kept unchanged as some parameter in the process changes. The shaping will be done using the weighted least squares method presented in Section 5.3. Different types of parametric structured plant variations are exemplified in Section 5.4.

5.3 Robust control as an approximation problem

One of the key ideas in Bodes' work is to shape the loop in such a way that the closed loop system is insensitive to gain variations in the loop. The ideal loop transfer function is, according to [Bode, 1945, p. 455]

$$L(s) = \frac{k_0}{\left(\sqrt{1 + s^2/\omega_0^2} + s/\omega_0\right)^\alpha} \tag{5.1}$$

For $\omega < \omega_0$ the argument of $L(i\omega)$ drops from 0 to $-\alpha\pi/2$ and for $\omega \geq \omega_0$ the argument is constant $= -\alpha\pi/2$. To get a more realistic loop transfer function high frequency roll-off must be introduced as Bode remarks on p. 471. A related work is [Oustaloup and Bergeon, 1987] where loop transfer functions of non integral order are considered.

The ideal loop transfer function can be approximated by using least squares approximation at some frequencies near the cross-over frequency to compute a compensator of a specified order. The compensator order is related to the length of the frequency range in which the phase is to be kept constant. By changing the angle ϕ_c the phase margin can be adjusted. The Bode ideas can be reformulated as an approximation problem which avoids the graphical constructions done in the Horowitz method.

In general a desired loop transfer function

$$L(s) = \frac{B_l(s)}{A_l(s)}$$

is specified at certain frequencies. Given the nominal process transfer function $G_{\text{nom}}(s)$, a controller

$$R(\frac{d}{dt})u(t) = -S(\frac{d}{dt})y(t)$$

where R and S are polynomials, can be computed such that

$$G_{\text{nom}}(s)\frac{S(s)}{R(s)} \approx L(s)$$

Using the least squares method the loss function can be written as

$$J = \sum_{k=1}^{N} |W(i\omega_k)|^2 \left| R(i\omega_k) - S(i\omega_k)\frac{G_{\text{nom}}(i\omega_k)}{L(i\omega_k)} \right|^2 \tag{5.2}$$

which is minimized with respect to the coefficients in $R(s)$ and $S(s)$. Introducing a reference input r the control law is modified to

$$R(\frac{d}{dt})u(t) = S(\frac{d}{dt})(y(t) - r(t))$$

The closed loop transfer function from r to y is then given by

$$G_{\rm cl}(s) = \frac{L(s)}{1 + L(s)} \tag{5.3}$$

which is equivalent to

$$L(s) = \frac{G_{\rm cl}(s)}{1 - G_{\rm cl}(s)} \tag{5.4}$$

The loss function in Eq. 5.2 could therefore be rewritten as

$$J_{\rm cl} = \sum_{k=1}^{N} |W(i\omega_k)|^2 \left| R(i\omega_k) - S(i\omega_k)G_{\rm nom}(i\omega_k)\frac{1 - G_{\rm cl}(i\omega_k)}{G_{\rm cl}(i\omega_k)} \right|^2$$

for some desired closed loop transfer function $G_{\rm cl}(s)$.

5.4 Robustness against parametric plant variations

A rather general description of parametric (and thus structured) plant variations is given by

$$G(s,p) = \frac{B(s,p_B)}{A(s,p_A)}e^{-\tau s}$$

where the parameter vector $p = (p_A, p_B, \tau)$ belongs to some subset $\mathcal{P} \subset \mathbb{R}^{n_p}$ for some integer n_p. For simplicity, $\mathcal{P}$ is assumed to be a closed bounded hypercube. Each template $\mathcal{G}(i\omega) = G(i\omega, \mathcal{P})$ consists of a subset of the complex plane. In the special case $n_p = 1$ the templates are curve segments in $\mathbb{C}$. For $n_p > 1$ the templates are given by closed subsets in $\mathbb{C}$. If $n_p = 2$ the boundary of the template consists of four curve segments. A compensator S/R will alter the magnitude and phase of the loop transfer function $L = GS/R$. In a Nichols diagram the logarithm of the magnitude is plotted against the phase. This means that each "compensated" template $L(i\omega, \mathcal{P})$ is just a translation of the "original" template $G(i\omega, \mathcal{P})$ when viewed in a Nichols diagram, since the multiplication by $S(i\omega)/R(i\omega)$ corresponds to addition by a vector in the diagram. To get a closed loop system insensitive to changes in the parameters p, the templates $L(i\omega, \mathcal{P})$ which are close -1 must be small. If the templates are small at high frequencies, a high gain feedback might satisfy the requirements. A system with a time delay fails to fulfill that condition, since the phase uncertainty becomes larger with increasing frequency.

Assume that $n_p = 1$. Typical examples are when p is a pure gain or when p is a time delay. The templates $L(i\omega, \mathcal{P})$, which consists of curve segments, are mapped under the closed loop mapping, Eq. 5.3 to "closed loop" templates $G_{\text{cl}}(i\omega, \mathcal{P})$ also consisting of curve segments. Robustness means that the closed loop response varies in an acceptable way as p varies. This can be formalized as

$$G_{\text{cl}}(s, p) = G_{\text{nice}}(s, \rho(p)) \tag{5.5}$$

where $G_{\text{nice}}(s, \rho)$ is a family of transfer functions parametrized by ρ which corresponds to "nice" step responses. "Nice" could mean, for example, that the variations in the closed loop step responses are very small. If the high frequency uncertainty of the plant is negligible and the plant is minimum phase then high gain feed back could be used to accomplish this. However, if high gain feedback is unfeasible then some variations in the response of the closed loop system must be tolerated. The problem is then to decide which closed loop properties that could be allowed to vary.

One alternative is to allow the speed (bandwidth) of the closed loop response to vary while keeping a constant damping. This can be expressed by

$$G_{\text{cl}}(s, p) = G_{\text{nice}}\left(\frac{s}{\rho(p)}\right) \tag{5.6}$$

where $\rho(p)$ is a positive real-valued function. Equation 5.6 can be interpreted as a frequency scaling where $\rho(p)$ measures the response speed relative to the nominal speed $\rho(p_{\text{nom}}) = 1$. This "damping invariance" also implies that any two Nyquist curves $G_{\text{cl}}(i\omega, p_1)$ and $G_{\text{cl}}(i\omega, p_2)$ consist of the same point set in $\mathbb{C}$ but are equipped with different parametrizations in the frequency ω, i.e. the frequencies are "slided" along the Nyquist curves. The open loop transfer function is given by

$$L(s) = \frac{G_{\text{nice}}(s/\rho(p))}{1 - G_{\text{nice}}(s/\rho(p))} =: L_{\text{nice}}\left(\frac{s}{\rho(p)}\right) \tag{5.7}$$

The ideal loop transfer function in Eq. 5.7 is approximated in some frequency range, depending on the variations in p.

The converse is to keep the closed loop bandwidth constant but to allow the damping to change. There is no general parametrization as in Eq. 5.6. A simplified version is to keep the natural frequency ω_0 constant in the second order transfer function family

$$G_{\text{nice}}(s, \zeta) = \frac{\omega_0}{s^2 + 2\zeta\omega_0 s + \omega_0^2} \tag{5.8}$$

which gives roughly the same bandwidth for different relative dampings ζ. The corresponding open loop transfer function is computed from Eq. 5.4 as

$$L(s) = \frac{\omega_0^2}{s(s + 2\zeta\omega_0)}$$

EXAMPLE 5.1
An embarrassingly simple but illustrative example is

$$G(s,a) = \frac{1}{s+a}$$

where a is uncertain. The compensator

$$K(s) = \frac{\omega_0}{s}$$

gives the desired closed loop parametrization

$$G_{\text{cl}}(s,a) = G_{\text{nice}}(s,\zeta)$$

with $\zeta(a) = \frac{a}{2\omega_0}$. □
Many other parametrizations of acceptable closed loop variation could be given in the one parameter case $n_p = 1$ but the main trade-off is between speed and damping. In the multi parameter case $n_p > 1$ the closed loop variations must also be parametrized by n_p parameters. The function ρ in Eq. 5.5 is then a mapping from $\mathcal{P} \subset real R^{n_p}$ into $\mathbb{R}^{n_p}$ for a given G_{nice}. For the case $n_p = 2$ one possibility would be to combine the two one-parameter cases "constant damping" and "constant bandwidth". □

The most useful of the one parameter cases is the "constant damping" case Eq. 5.6 since it guarantees a certain stability margin. The rest of the chapter will therefore be devoted to this case.

Variations in the process gain

As a special case of parametric plant variation the gain variation problem will be considered. The process transfer function is thus assumed to be of the form

$$G(s) = \frac{bB_0(s)}{A_0(s)} e^{-\tau_0 s}$$

where the scalar parameter b belongs to some interval, $[b_-, b_+]$ and $B_0(s)$ and $A_0(s)$ are fixed polynomials and the time delay τ_0 is a fixed scalar. Assume that the design of the loop is done according to the "constant damping" method in Eq. 5.6. Ideally, this would mean that the phase margin is kept constant over a certain frequency range. The transfer function

$$L(s) = \frac{k}{s^\alpha} \tag{5.9}$$

where k is some constant, has the property that the phase is constant, $\arg L(i\omega) = -\alpha\pi/2$ for all frequencies ω. This is a simplified version of the ideal loop

transfer function of Bode in Eq. 5.1. The transfer function L in Eq. 5.9 is invariant to gain variations in the sense that

$$k_1 L(s) = L(s/\rho(k_1))$$

where

$$\rho(k_1) = k_1^{1/\alpha}$$

The phase margin of the loop transfer function in Eq. 5.9 is given by

$$\varphi_m = (1-\alpha)\pi/2 \tag{5.10}$$

This means that in this ideal case the damping of the closed loop will be independent of the gain and only the speed of the response will vary. The exponent α is in the Bode case chosen to 1.5 which corresponds to a phase margin of 45° but any phase margin φ_m can be achieved by choosing α according to Eq. 5.10.

The actual loop transfer function should thus approximate the ideal loop transfer function given by Eq. 5.9 in the frequency range determined by the gain variation of the process. This can be achieved with a controller computed by the least squares method with loss function given by Eq. 5.2 and with approximation frequencies chosen from this frequency range.

A general loop transfer function with uncertainty in the gain only can be written as

$$L(s) = kL_0(s) \tag{5.11}$$

where k varies. The constant k in Eq. 5.11 is chosen with respect to the desired closed loop bandwidth ω_b.

By definition, the closed loop bandwidth ω_b is given by

$$\left| \frac{L(i\omega_b)}{1+L(i\omega_b)} \right| = \frac{1}{\sqrt{2}}$$

which is equivalent to

$$|L(i\omega_b) - 1| = \sqrt{2} \tag{5.12}$$

Inserting Eq. 5.11 into Eq. 5.12 gives after some simple calculations

$$k = g + \sqrt{1+g^2} \tag{5.13}$$

where $g = \operatorname{Re} L_0(i\omega_b)/|L_0(i\omega_b)|^2$ and only positive values of k are considered. For the special case in Eq. 5.9 the formula for k is

$$k = \omega_b^\alpha \left(\cos\frac{\alpha\pi}{2} + \sqrt{1+\cos^2\frac{\alpha\pi}{2}} \right) \tag{5.14}$$

It is reasonable to choose approximation frequencies in an interval $[\omega_-, \omega_+]$ around the nominal closed loop bandwidth. The length of the interval must

be chosen with respect to the gain variation of the process. If $b \in [k_-, k_+]$ and the nominal value is $k = k_{\text{nom}}$ then a possible choice of frequency interval for L as in Eq. 5.9 is

$$\begin{aligned} \omega_- &= (b_{\text{nom}}/b_-)^{1/\alpha}\omega_b \\ \omega_+ &= (b_{\text{nom}}/b_+)^{1/\alpha}\omega_b \end{aligned} \tag{5.15}$$

which are the closed loop bandwidths for the lower and upper values of b.

EXAMPLE 5.2
Consider the uncertain system

$$G(s) = \frac{b}{s(s+2)}$$

where the gain $b \in [1, 10]$ and the nominal value $b_{\text{nom}} = 2$. The frequency range in which the phase is to be kept constant is determined by the desired bandwidth ω_b of the nominal closed loop system. Let $\omega_b = 1$. The approximation frequencies were chosen as $\Omega = \{0.63, 1, 1.59, 2.92\}$ where the values are logarithmically spread between ω_- and ω_+ according to Eq. 5.15. An attempt to fit a second order compensator with integration resulted in an unstable compensator. This motivated a compensator order of three which turned out to give a stable compensator. The resulting compensated Nyquist curve is rather similar to a straight line in the interesting frequency range (see Fig. 5.2).

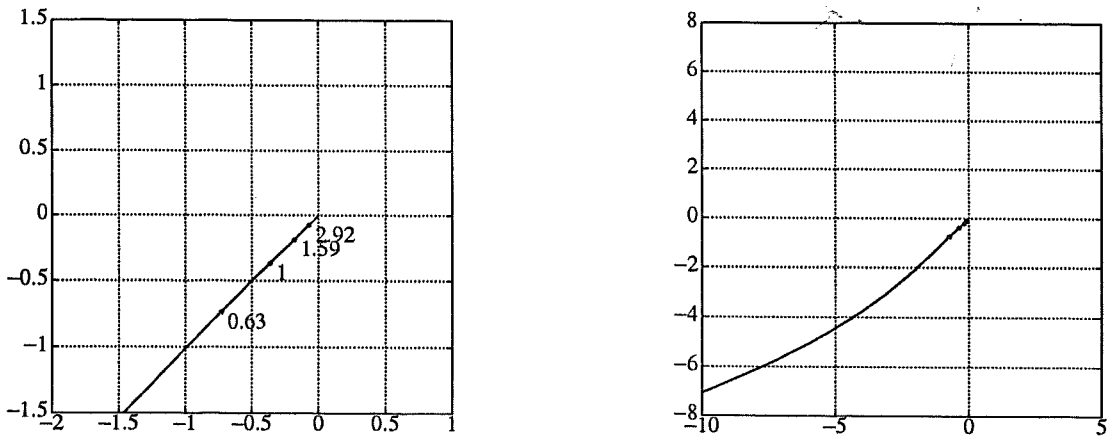


Figure 5.2 The Nyquist curve of the process $G(s) = b/s(s+2)$ with a third order compensator such that the phase margin is kept constant as b varies between 1 and 10. The approximation frequencies are marked with 'o':s. The Nyquist curve is shown in two different scales.

To verify the robustness of the closed loop damping with respect to gain variations the time response of the closed loop system is plotted in Fig. 5.3 for the

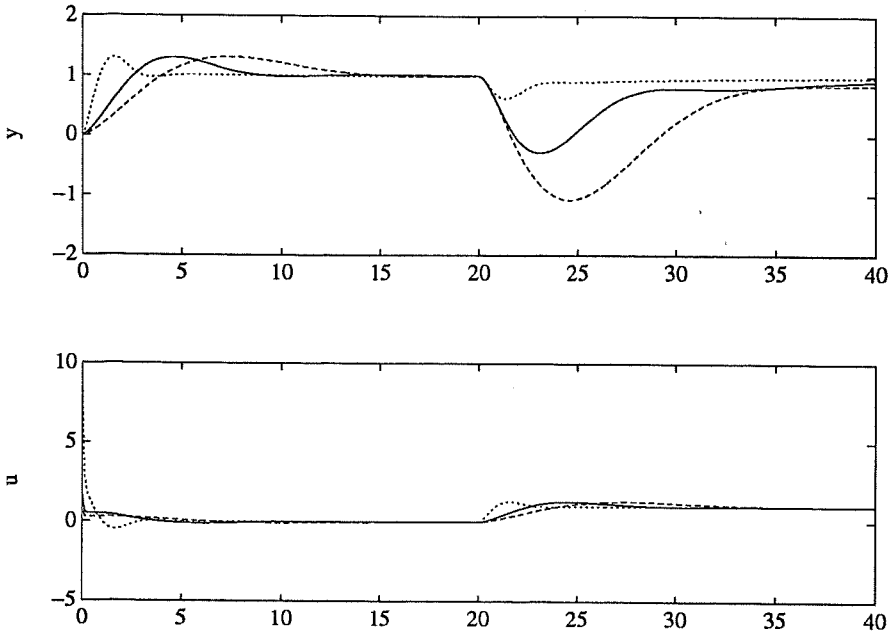


Figure 5.3 Step responses of the process $G(s) = b/s(s+2)$ with a third order compensator such that the phase margin is kept constant as b varies between 1 and 10. The values of b are 2 (solid curve), 1 (dashed curve) and 10 (dotted curve). The desired bandwidth for the nominal case $b = 2$ is $\omega_b = 1$

values $b = 1, 2$ (nominal) and 10. Clearly, the step responses share approximately the same relative damping. The slow mode appearing in the load disturbance response can be partly explained by the relatively low bandwidth, $\omega_b = 1$. When increasing ω_b to 4 in the nominal case $b = 2$ the step responses in Fig. 5.4 are obtained instead. Notice the large initial peak in the control signal. This can be removed by using an RST structure with $T \neq S$. This may, however, destroy the independence of the relative damping with respect to the gain variation. □

EXAMPLE 5.3—Double integrator
A double integrator with uncertain gain

$$G(s) = \frac{k}{s^2}, \qquad k \in [1, 5]$$

is to be controlled by a third order controller with integration. Using the desired loop transfer function in Eq. 5.9 gives a compensator such that the loop Nyquist curve starts with a phase of $-90°$ instead of $-270°$ which should be the case since the process contains two integrators and the compensator has one integrator. This implies an unstable closed loop system (even in the nominal case). The desired loop transfer function in Eq. 5.9 is therefore modified to

$$L(s) = \frac{k_1}{s^{\alpha_1}} + \frac{k_2}{s^{\alpha_2}}$$

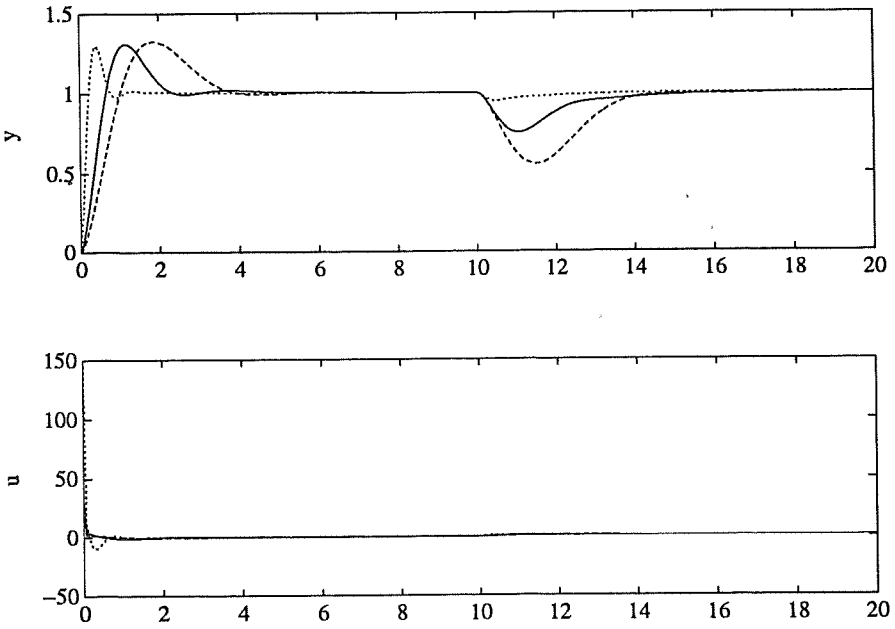


Figure 5.4 Step responses of the process $G(s) = b/s(s+2)$ with a third order compensator such that the phase margin is kept constant as b varies between 1 and 10. The values of b are 2 (solid curve), 1 (dashed curve) and 10 (dotted curve). The desired bandwidth for the nominal case $b = 2$ is $\omega_b = 4$.

with $\alpha_1 = 1.5$ and $\alpha_2 = 3$. At low frequencies $L(i\omega) \approx k_2(i\omega)^{-3}$ and at high frequencies $L(i\omega) \approx k_1(i\omega)^{-1.5}$. This gives the required low frequency phase asymptote $-270°$ while at the same time the "constant phase" frequency range is maintained. The bandwidth of the closed loop system depends in this case on two constants, k_1 and k_2 which means that the constants are not uniquely defined. Letting $k_1 = 4$ and $k_2 = 2$, which corresponds to a bandwidth of $\omega_b \approx 4$, gives the loop transfer Nyquist curve shown in Fig. 5.5. The approximation frequencies were chosen as $\Omega = \{1, 2, 4, 8\}$. The corresponding step responses are shown in Fig. 5.6. Notice that the damping varies slightly as the gain varies since the loop transfer function has been modified. □

EXAMPLE 5.4—Eight cascaded low pass filters

The Bode "constant phase" method will now be demonstrated for the system with transfer function

$$G(s) = \frac{b}{(s+1)^8}$$

A controller with integration will be designed such that the phase margin of the compensated loop is approximately constant as b varies in the interval $[0.5, 2]$ and with the closed loop bandwidth $\omega_b = 0.1$ at the nominal gain $b_{\text{nom}} = 1$. Since the phase of this system drops rather rapidly in the frequency range of the open

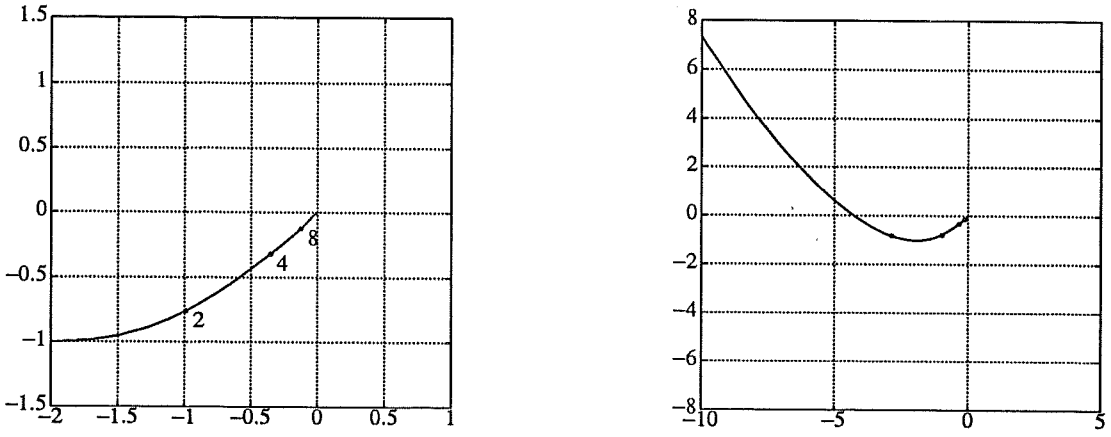


Figure 5.5 The Nyquist curve of the the process $G(s) = k/s^2$ with a third order compensator such that the phase margin is kept approximately constant as k varies between 1 and 5. The desired bandwidth for the nominal case $k = 1$ is $\omega_b \approx 4$ The approximation frequencies are marked with 'o':s. The Nyquist curve is shown in two different scales.

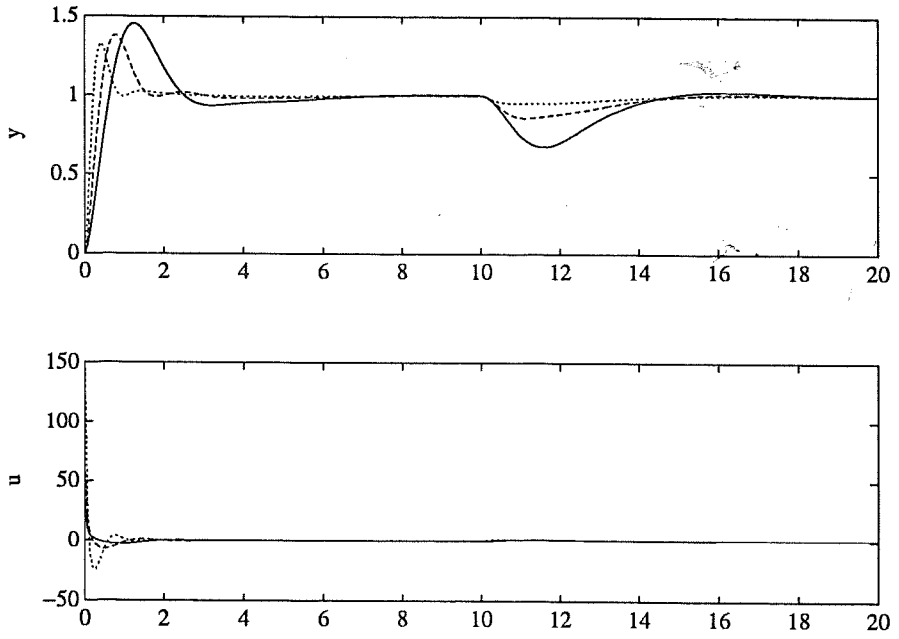


Figure 5.6 Step responses of the process $G(s) = k/s^2$ with a third order compensator such that the phase margin is kept approximately constant as k varies between 1 and 5. The desired bandwidth for the nominal case $k = 1$ is $\omega_b \approx 4$ (solid curve). The dashed and dotted curves corresponds to $k = 2$ and $k = 5$ respectively.

loop bandwidth, the approximation frequencies Ω have to be chosen rather low in order to get a stable compensator S/R. With $\Omega = \{0.05, 0.1, 0.15, 0.2\}$ a third order controller was computed. The desired open loop transfer function had to

be modified with high frequency roll off:

$$L(s) = \frac{B_r(s)}{A_r(s)} s^{-\alpha}$$

where $B_r = 0.05$, $A_r(s) = s^3 + 3s^2 + 3s + 2$ and $\alpha = 1.5$. The resulting open loop transfer function is shown in Fig. 5.7.

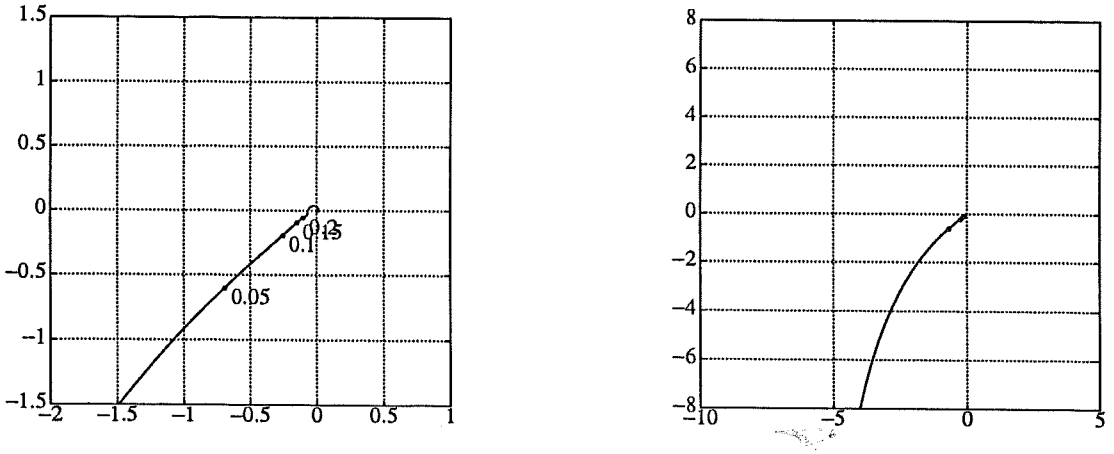


Figure 5.7 The system $G(s) = (s+1)^{-8}$ with a third order compensator giving almost constant phase over the range of approximation frequencies $\Omega = \{0.05, 0.1, 0.15, 0.2\}$. The approximation frequencies are marked with 'o':s. The Nyquist curve is shown in two different scales.

Figure 5.8 shows that the closed loop is considerable slower than the uncompensated open loop system and it seems not worthwhile to have "damping invariance" with respect to the process gain in this case.

Variations in the time delay of the process

Another type of plant variation is a varying time delay. The process model can then be written as

$$G(s) = \frac{B_0(s)}{A_0(s)} e^{-\tau s}$$

where $\tau \; in [\tau_-, \tau_+]$ and the polynomials B_0 and A_0 are fixed. Using the "constant damping" method as in Eq. 5.6 an ideal loop transfer function would be

$$L(s) = ke^{-\tau_L s} \tag{5.16}$$

since this has the invariance property

$$L(s)e^{-\tau_1 s} = L(s/\rho(\tau_1))$$

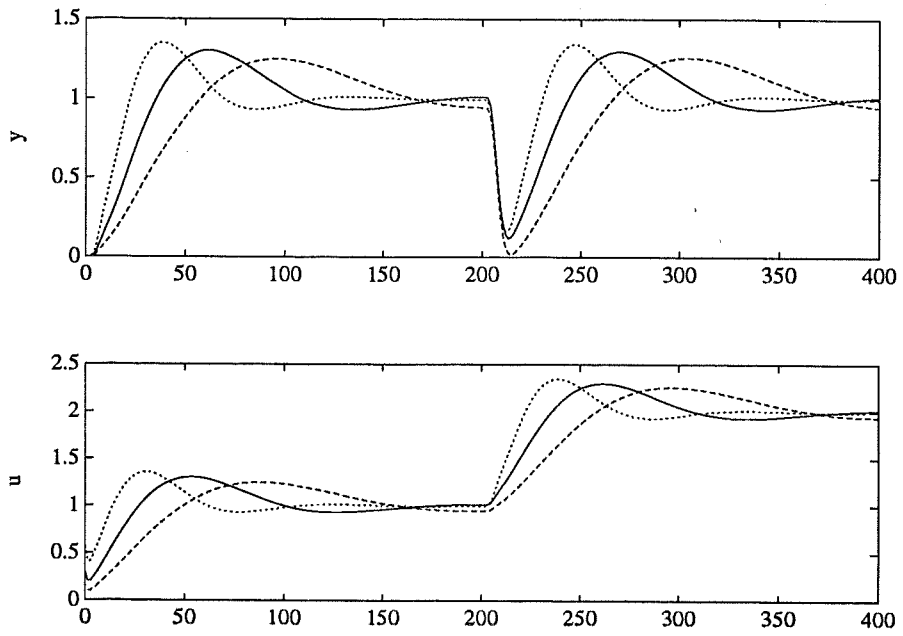


Figure 5.8 Step responses of the process $G(s) = b/(s+1)^8$ with a third order compensator such that the phase margin is kept constant as b varies between 0.5 and 2. The values of b are 1 (solid curve), 0.5 (dashed curve) and 2 (dotted curve). The desired bandwidth for the nominal case $b = 2$ is $\omega_b = 0.1$

where

$$\rho(\tau_1) = \frac{\tau_L}{\tau_1 + \tau_L}$$

The gain margin for this choice of loop transfer function is given by

$$g_m = \frac{1}{k}$$

A change in the time delay of the loop will merely change the speed of the closed loop response while the damping is kept constant. As in the case of gain variation the actual loop transfer function approximates the ideal loop transfer function in a certain frequency interval. In this case, the frequency interval depends on the interval $[\tau_-, \tau_+]$.

EXAMPLE 5.5—First order system with time delay
The process model

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

has a varying time delay $\tau \in [0.5, 2]$ which has the nominal value $\tau_{\text{nom}} = 1$. A second order controller without integration will be designed which gives approximate damping invariance with respect to the variation in the time delay. The ideal loop transfer function for the nominal case $\tau = 1$ was chosen as

$$L(s) = 0.3e^{-2s}$$

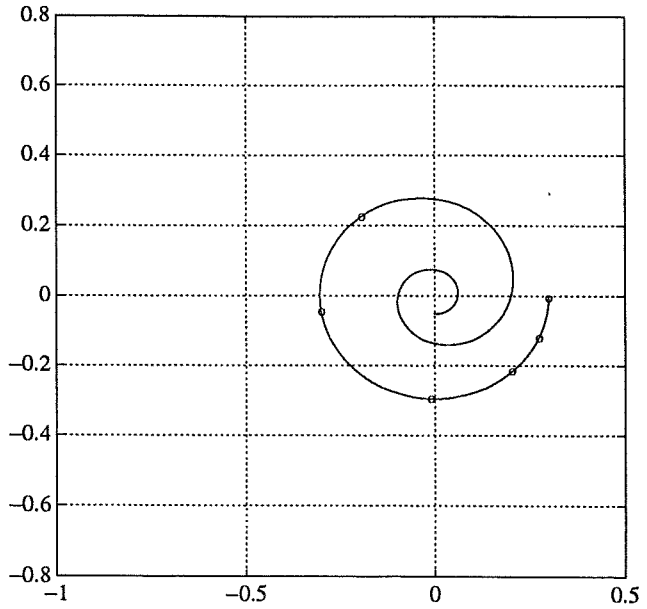


Figure 5.9 The system $G(s) = e^{-s}/(s+1)$ with a second order compensator (without integration) giving almost constant magnitude over the range of approximation frequencies $\Omega = \{0.01, 0.2, 0.4, 0.8, 1.5, 2\}$.

with the approximation frequencies $\Omega = \{0.01, 0.2, 0.4, 0.8, 1.5, 2\}$. The approximating loop transfer function is shown in Fig. 5.9. The resulting compensator $K(s)$ is computed as

$$K(s) = \frac{S(s)}{R(s)} = \frac{-0.47807s^2 + 0.73557s + 1.7069}{s^2 + 2.807s + 5.6816}$$

The reference input r is introduced in the controller in such a way that the transfer function G_{cl} from r to y has the stationary gain $G_{\text{cl}}(0) = 1$. The step response of the closed loop system is shown in Fig. 5.10for $\tau = 0.5$, 1 and 2. The simulations were carried out using a fifth order rational least squares approximation of $G(s) = e^{-\tau s}/(s+1)$ for each value of τ. The step responses have approximately the same damping as the time delay τ varies.

The reason for not introducing integration in the controller in Ex. 5.5 is that the loop transfer function in Eq. 5.16 lacks the required large low frequency gain. One way to circumvent this is to introduce the "trade-off"

$$L(s) = k_1 e^{-\tau_L s} + k_2 \frac{1}{s} \tag{5.17}$$

Unfortunately, this will not give the desired property due to a slow mode which appears in the closed loop system. This slow mode will destroy the "damping invariance" property since the second term in Eq. 5.17 will dominate at the

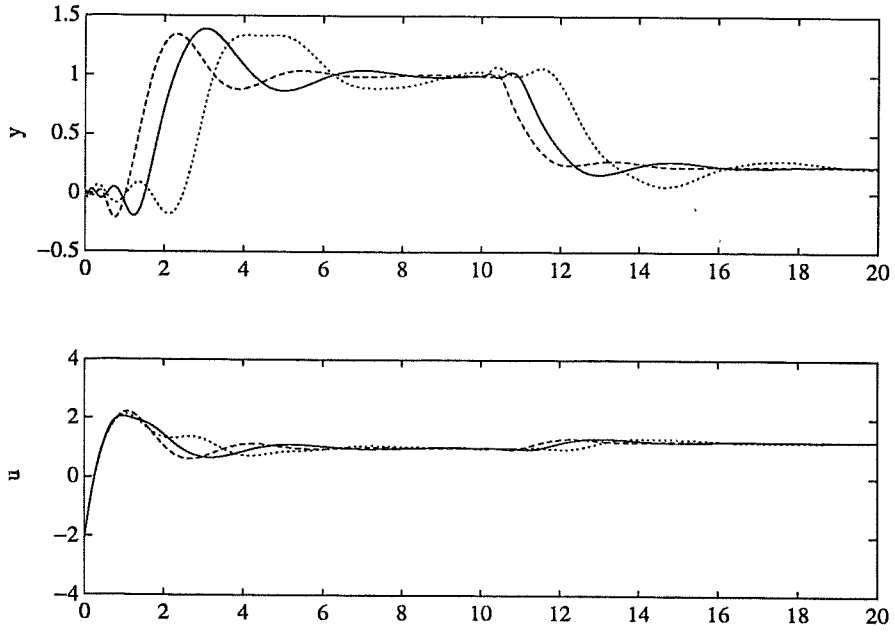


Figure 5.10 Step responses of the process $G(s) = e^{-\tau s}/(s+1)$ with a second order compensator such that the amplitude margin is kept constant as τ varies between 0.5 and 2. The values of τ are 1 (solid curve), 0.5 (dashed curve) and 2 (dotted curve).

corresponding low frequencies, no matter what values of k_1 and k_2 that are chosen. □

5.5 Conclusions

Robust control can be viewed as transforming variations in the plant by feedback to variations in the closed loop system of benign nature. As an alternative to high gain feedback, the approach found in [Bode, 1945] has been considered. Some examples have illustrated that plant variations of different types can be turned into closed loop variations in such a way that the bandwidth varies without change of damping. The controller accomplishing this is computed by least squares curve fitting at a number of frequencies. The advantage with the method is that a certain acceptable closed loop variation can be obtained to any degree of accuracy. The accuracy is determined by the order of the compensator. It is therefore natural to start with a compensator of low order and then increasing the order until the desired accuracy is achieved. The method is similar to the Horowitz method but the new ingredient is the application of least squares curve fitting. This gives an easy way to compute the controller and may also give a controller of lower order compared to the Horowitz way of successively cascading lead-lag compensators.

6

Conclusions

A common way to obtain a process model is to measure the frequency response of the process. At this stage it is not necessary to make assumptions about the order of the process. It has therefore been of interest to find design methods which operate directly on frequency response data. Such methods are the classical frequency domain methods developed by Nyquist, Black, and Bode during the thirties and the forties. In these approaches the main concern was to achieve robustness against variations in the process characteristics. This was done by introducing lead-lag compensation in the loop. It was not straightforward to see how the closed loop transient properties were influenced by this compensation. In the pole placement method the situation is otherwise. The transient performance is specified while the robustness properties have to be checked after the design.

In Chapter 4 a design method has been proposed which shows similarities with both the classical frequency domain approach and the pole placement method. The parameters in a two degree of freedom controller are computed by least squares fitting to a desired closed loop transfer function at certain frequencies. The design specification consists of the set of approximation frequencies Ω, the desired closed loop model $G_m(s)$, the observer polynomial $A_o(s)$, and the orders of the controller polynomials R and S. The weighting of the least squares method could also be regarded as one of the design specifications. Both the order of the regulator and the order of the desired closed loop model can be chosen much lower than the order of the process. Thus, only the "essential dynamics" has to be specified (cf. the dominant pole design in [Åström and Hägglund, 1985]). Examples have been given where the closed loop step responses are well in accordance with the specifications. The trade-off between the performance and the regulator complexity has also been demonstrated.

Loop shaping has been treated in Chapter 5 where the least squares method is used to compute a compensator such that the loop approximates some "ideal

loop characteristic" in a certain frequency range. The goal is to turn open loop variations into "harmless" closed loop variations. These ideas originates from the work [Bode, 1945] where process ideal loop gain in the case of process gain variations was considered. Some examples are given where variations in the open loop gain gives closed loop step responses which differs only by a change in time scale.

The design methods described in Chapter 4 and 5 are not straightforward to extend to the multivariable case. The major importance of the methods is that much insight is gained about different trade offs in SISO control design. Design of low order controllers such as PI and PID controllers is conveniently done with the method in Chapter 4. The design methods are also believed to be instructive to use in education.

References

ADAMJAN, AROV and KREIN (1971): "Analytical properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem," *Math. USSR Sbornik*, 15.

ANDERSON, B. D. O. (1986): "Weighted Hankel-norm approximation: Calculation of bounds," *Systems Control Lett.*, 7.

ÅSTRÖM and WITTENMARK (1984): *Computer Controlled Systems*, Prentice-Hall.

ÅSTRÖM and HÄGGLUND (1985): "Dominant pole design," Technical report TRFT-7282, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

ÅSTRÖM (1975): "Lectures on System Identification: Frequency Response Analysis," Technical report TRFT-7504, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

ÅSTRÖM (1988): "Dominant pole placement design of PI regulators," Technical report TRFT-7381, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

BODE, H. W. (1945): *Network Analysis* and *Feedback Amplifier Design*, D. Van Nostrand, New Jersey.

CHENEY, E. W. (1966): *Approximation Theory*, McGraw-Hill Book Company, pp. 173 – 189.

DOYLE, J. and STEIN, G. (1981): "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis," *IEEE Trans. of Autom. Control*, 26.

FRANCIS, B, and ZAMES, G (1984): "On optimal sensitivity theory for SISO feedback systems," *IEEE Trans. Automatic Control*, 29.

GLOVER, K. (1984): "All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds," *Int. J. Control*, 39.

GLOVER, K. and DOYLE, J. (1988): "State-space formulae for all stabilizing controllers that satisfy an H_∞ bound and relations to risk sensitivity," *Systems and Control Letters*, **11**, 167 – 172.

GOLUB, G. H. and VAN LOAN, C. F. (1987): *Matrix Computations*, John Hopkins University Press, Baltimore, Maryland, pp. 162 – 169.

GOODWIN, G. and SALGADO (1989): "Quantification of uncertainty using an embedding principle," Proc. American Control Conference, June 1989, Pittsburg USA.

HARSHAVARDHANA, JONCKHEERE and SILVERMAN (1984): "Eigenvalue and generalized eigenvalue formulations for Hankel norm reduction directly from frequency response data," Proc. IEEE Conf. on Decision and Control, Las Vegas, NV, Dec. 1984.

HWANG and CHEN (1987): "Solution of General Padé Fitting Problem Via Continued Fraction Expansion," *IEEE Trans. of Autom. Control*, 32.

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

HOROWITZ, I. (1982): "Quantitative feedback theory," *IEE Proc.*, **129-D**, 6.

KUNG, S-Y. and LIN, D. W. (1981): "Optimal Hankel-Norm Model Reductions: Multivariable Systems," *IEEE Trans. of Autom. Control*, 26.

LATHAM, G. A. and ANDERSON, B. D. O. (1985): "Frequency-weighted optimal Hankel-norm approximation of stable functions," *Systems and Control Letters*, 5.

LAWRENCE, P. J. and ROGERS, G. J. (1979): *Proc. Inst. Electr. Engrs.*, 126, p. 104.

LEVY, E.C. (1959): "Complex curve fitting," *I.R.E. Trans. of Autom. Control*, 4, p. 37.

OUSTALOUP, A. and BERGEON, B. (1987): "Frequency space synthesis of a robust dynamic command," Proc. IFAC 10th world congress on Aut. Control, Munich.

PRO-MATLAB (1987): *PRO-MATLAB User's guide*, The MathWorks, Inc., Sherborn, Mass., USA.

SANATHANAN, C.K. and KOERNER, J. (1963): "Transfer function synthesis as a ratio of two complex polynomials," *IEEE Trans. of Autom. Control*, 8, p. 56.

STAHL, H (1984): "Transfer function synthesis using frequency response data," *Int. J. Control*, 39, p. 541.

STEIN, G. and ATHANS, M. (1987): "The LQG/LTR Procedure for Multivariable Feedback Control Design," *IEEE Trans. of Autom. Control*, 32.

VIDYASAGAR, M. (1986): *Control System Synthesis – A Factorization Approach*, MIT Press, Cambridge, Massachusetts, USA, pp. 53 – 59.

WAHLBERG, B. (1987): "On the identification and approximation of linear systems," Ph.D. Thesis, Department of Electrical Engineering, Linköping University, Linköping, Sweden.

WALLENBORG, A. (1987): "Control of Flexible Servo Systems," Technical report TRFT-3188 (Licentiate thesis), Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

WILLEMS, J. C. (1971): *The Analysis of Feedback Systems*, The M.I.T. Press, Camridge, Massachusetts, p. 95.

YOULA, D. C., BONGIORNO, J. J., and LU, C. N. (1974): "Single-Loop Feedback-Stabilization of Linear Multivariable Dynamical Plants," *Automatica*, 10.

ZAMES, G. (1981): "Feedback and optimal sensitivity: Model reference transformation, multiplicative seminorms, and approximate inverses," *IEEE trans. Automatic Control*, 26, 301 – 320.

A

Some MATLAB functions

Most of the computations in the thesis have been done in PRO-MATLAB. During the work it has been found convenient to include some new MATLAB functions listed in this section, each with a short description.

Least squares approximation

abtau	Least squares fitting to a rational function times an exponential
abz	Least squares fitting to a rational function
rsz	Least squares fitting of a one-degree-of freedom controller
rstz	Least squares fitting of a two-degree-of freedom controller

Hankel norm approximation

hankelu	Unweighted Hankel norm approximation
hankelw	Weighted Hankel norm approximation

Polynomials

besselpoly	Bessel polynomial
butterpoly	Butterworth polynomial (with generalization)
dab	Solution of the DAB (Diophantine-Aryabhatta-Bezout) equation
mconv	Product of more than two polynomials
polyadd	Sum of two polynomials
polyder	Derivative of a polynomial
polystr	Formatting of a polynomial as a TEXstring
rloc	Root locus plot with special choice of step length
rstpoly	Pole placement using the DAB equation
tmodify	Modification of the polynomial T

Transfer functions and frequency responses

ampl	Amplitude plot of a transfer function
bandwidth	Approximate bandwidth from a frequency response
bpl	Bode plot in a fresh diagram
bopl	Bode plot in an old diagram
gadd	Sum of two transfer functions
gval	Computation of frequency response of $e^{-\tau s}B(s)/A(s)$
nipl	Nichols plot of a transfer function
nypl	Nyquist plot of a transfer function
phpl	Phase plot of a transfer function
arg	Argument of complex vector (in degrees)

Time responses

apparent	Computes apparent time delay and apparent time constant
stepresp	Step response of $e^{-\tau s}B(s)/A(s)$
stepplot	Plots a step response computed with **stepresp**
pcrit	Frequency scaling test of controller stability
yucl	Step response of closed loop system
yupl	Plots a step response computed with **yucl**
yup	Same as **yupl** but in an old diagram

A convenient way of using MATLAB is to use the text macro facility. This consists of command execution by evaluation of strings. As an illustration a start up file with text macros is listed below.

```
i=sqrt(-1);
lgwlow=-2;lgwhigh=2;nww=300;
dozz='ww=logspace(lgwlow,lgwhigh,nww)'';zz=i*ww;';
eval(dozz);
doam = 'am=conv(amfix,butterpoly(nam-length(amfix)+1,wm,phim));';
doao = 'ao=conv(aofix,butterpoly(nao-length(aofix)+1,wo,phio));';
phim = 45;
phio = 45;
aofix = 1;
amfix = 1;
doz='z=i*w;';
dog='g=gval(b,a,tau,z);';
dof='f=abs(gval(bf,af,0,z));';
bf = 1;
af = 1;
abh='[ah,bh,eabh]=abz(z,g,f,na,nb,ahfix,bhfix);tauh=0;';
ahfix = 1;
bhfix = 1;
dorstz='[r,s,t,erst]=rstz(z,g,f,am,bm,ao,r1,nr,ns);';
dorst='[r,s,t]=rstpoly(a,1,b,am,1,ao,r1);';
dorsth='[r,s,t]=rstpoly(ah,1,bh,am,1,ao,r1);';
dogm='eval(doam);eval(dobm);';
dobm = 'bn=real(poly(kbh*roots(bh)));';
dobm = [dobm 'bm=bn/bn(max(size(bn)))*am(max(size(am)));'];
```

```
kbh = 1;
do1 = 'eval([doz dog dof abh]);';
do2 = 'eval([doz dog dof dogm doao dorstz]);';
do3 = 'eval([dogm doao dorst]);';
do5 = 'eval([dogm doao dorsth]);';
clp='clpo=roots(polyadd(conv(a,r),conv(b,s))),';
clph='clpoh=roots(polyadd(conv(ah,r),conv(bh,s))),';
clpn='clop=roots(conv(am,ao)),';
dogg='gg=gval(b,a,tau,zz);';
doggh='ggh=gval(bh,ah,tauh,zz);';
doggm='ggm=gval(bm,am,0,zz);';
pl1='eval([dozz doggh dogg dowmark]);';
pl1 = [pl1 'nypl([ggh gg],'' '','''o'',wmark,ww);'];
pl2='eval([dozz doggm dogcl dowmark]);';
pl2 = [pl2 'nypl([gcl ggm],'' '','''o'',wmark,ww);'];
np = 300;
docl='cl=yucl(a,b,r,s,t,tid,np);';
doclh='clh=yucl(ah,bh,r,s,t,tid,np);';
```

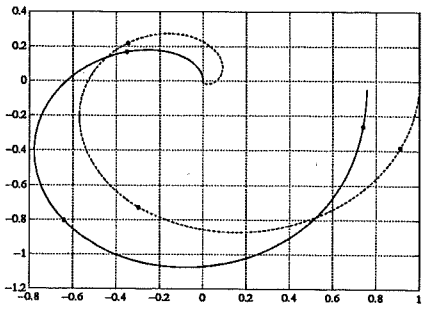
B

A sample session

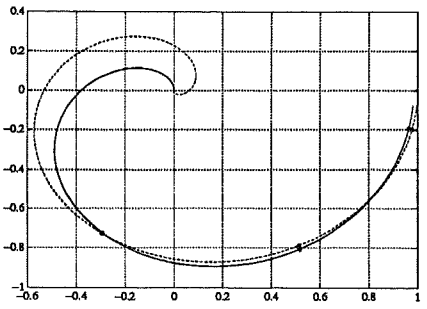
The MATLAB functions and text macros in Appendix A are used in the following interactive session. The design method in Chapter 4 is illustrated by a PID-design for the system $G(s) = 1/(s+1)^8$ (cf. Ex. 4.5). The designer starts with an indirect method (Method C′ in Chapter 1) where a second order approximant $\hat{G}(s)$ is used in a second order pole placement design. After experimenting with different choices of approximation frequencies Ω and observer polynomial $A_o(s)$ the designer decides to increase the order of the closed loop model $G_m(s)$. He then uses the direct method D′ and finally arrives at a PID-controller with which he is satisfied.

```
>> % Process model: G(s) = 1/(s+1)^8
>> b=1;a=poly(-ones(1,8));tau=0;
>> % Find LS approximation of order 2 at Omega={0.1,0.5,1}
>> w=[0.1 0.5 1];
>> na = 2; nb = 1;
>> eval(do1);ah

felet =

    0.7500    0.3281    0.0787


ah =

    1.0000   -0.0035    0.0541

>> % Unstable model. Maybe the approximation frequencies are too high.
>> % Check axis intersections of the Nyquist curve of G(s)
>> waxes = table1([arg(gg) ww],-[90 180 270 360])

waxes =
```

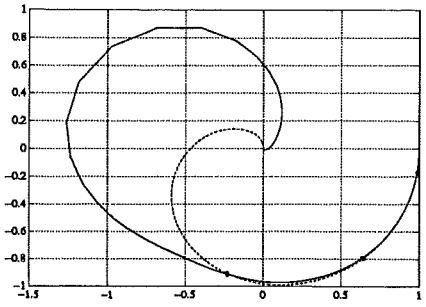
```
    0.1989    0.4142    0.6684    1.0011

>> % Either decrease approximation frequencies or increase model order
>> % Let's try the first alternative
>> w=w/2

w =

    0.0500    0.2500    0.5000

>> eval(do1);ah

felet =

    0.2103    0.3526    0.0520


ah =

    1.0000    0.2288    0.0615

>> % The approximant is stable! What about the quality of it?
>> eval(pl1)
```



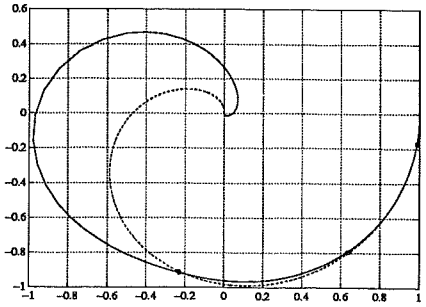

```
>> % Bad curve fitting! Decrease approximation frequencies further
>> w=w/2

w =

    0.0250    0.1250    0.2500

>> eval([do1 pl1]);ah

felet =

    0.0140    0.0186    0.0042


ah =

    1.0000    0.4295    0.0752
```




```
>> % This is better. Now, let's try pole placement design
>> % based on this model
>> % Reasonable pole locations: A_m=Butt(2,0.3,45) and A_o=Butt(2,0.6,45)
>> nam=2;nao=2;r1=[1 0];
>> wm=0.3;phim=45; wo=0.6;phio=45;
>> eval(do5)
>> r,s,t

r =

    1.0000    1.6962         0


s =

    5.2315    2.3362    0.4378


t =

    1.2161    1.0319    0.4378

>> % Check stability of closed loop system
>> eval(clp)

clpo =

  -2.3029 + 0.4621i
  -2.3029 - 0.4621i
  -1.6230 + 1.1276i
  -1.6230 - 1.1276i
  -0.6827 + 1.1397i
  -0.6827 - 1.1397i
  -0.0573 + 0.4384i
  -0.0573 - 0.4384i
  -0.1821 + 0.1604i
  -0.1821 - 0.1604i

>> % One of the complex pole pairs is very poorly damped
>> % Let's check the closed loop Nyquists
>> eval(pl2)
>> grid
```

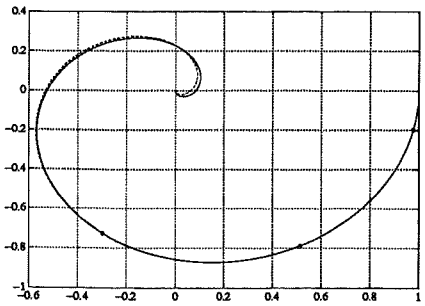


```
>> % Hmmmm... This doesn't look nice at all
>> % Maybe we should have increased the model order after all.
>> na = 3; nb = 2;
>> % No, wait -- Let's decrease wo instead
>> na=2;nb=1;
>> wo = 0.4;
>> eval([do2 clp pl2])

felet =

    0.0048    0.0015    0.0003


clpo =

  -2.1201 + 0.4368i
  -2.1201 - 0.4368i
  -1.5118 + 1.0568i
  -1.5118 - 1.0568i
  -0.6586 + 1.0626i
  -0.6586 - 1.0626i
  -0.1023 + 0.4043i
  -0.1023 - 0.4043i
  -0.1679 + 0.1549i
  -0.1679 - 0.1549i
```



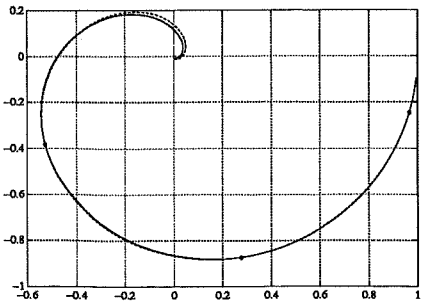
```
>> % Slightly better but not enough. Try inceasing process model order
>> na=3;nb=2;
>> eval([do1 pl1])
```

```
felet =

   1.0e-15 *

    0.2238    0.1110    0.1110
```

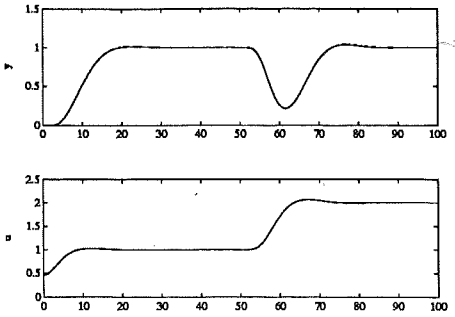



```
>> % Not bad ...
>> % Now we have to use the Chapter 4 method (controller still of 2nd order)
>> % Increase order of desired closed loop model
>> nam=3;
>> eval(do2)

felet =

    0.0013    0.0004    0.0000

>> r,s,t

r =

    1.0000    0.5883         0


s =

    0.7459    0.4157    0.0743


t =

    0.4643    0.2627    0.0743

>> eval([clp pl2])

clpo =

  -1.8977 + 0.3690i
  -1.8977 - 0.3690i
  -1.3778 + 0.8905i
  -1.3778 - 0.8905i
  -0.6464 + 0.8874i
```

```
  -0.6464 - 0.8874i
  -0.1732 + 0.2879i
  -0.1732 - 0.2879i
  -0.1991 + 0.1211i
  -0.1991 - 0.1211i

>> grid
```




```
>> % Quite acceptable curve matching
>> % Let's have a look at the step responses
>> tid=100; np=300;
>> eval(docl);yup(cl)
```




```
>> % This is good enough
>> exit
```