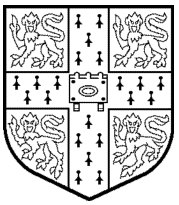
Identification for Control: Deterministic Algorithms and Error Bounds

Paresh Date

Churchill College



Control Group

Department of Engineering

University of Cambridge

A dissertation submitted for

the degree of Doctor of Philosophy

June 2000

Acknowledgements

At the outset, I am pleased to express my gratitude towards my research supervisor Dr. Glenn Vinnicombe. His vision and his patient guidance had a profound impact on this work.

I am indebted to my contemporaries at Control Group, for their contributions towards my understanding of control and for their kinship. I mention in particular Alex Lanzon and Mike Cantoni, who proof-read parts of this thesis and made several valuable suggestions.

On a personal note, I first thank my wife Bhagyashree, for her unfailing love and support. Many friends deserve a mention for making my Cambridge days memorable - special thanks to Meena, Supriya, Satish, Maya and Meenakshi. Continuing support and encouragement of my parents is also gratefully acknowledged.

This work was supported financially by the Cambridge Commonwealth Trust and the Overseas Research Student Award from C.V.C.P., United Kingdom. Travel grants for attending conferences were provided by Cambridge University Engineering Department, Churchill College and the Commonwealth Trust.

As required by the University statute, I hereby declare that this dissertation is not substantially the same as any that I have submitted for a degree or other qualification at any other University. This dissertation is a result of my own work and includes nothing which is the outcome of work done in collaboration.

Abstract

This dissertation deals with frequency domain identification of linear systems in a deterministic set-up. A robustly convergent algorithm for identification of frequency response samples of a linear shift invariant plant is proposed. An explicit bound on the identification error is obtained based on suitable *a priori* assumptions about the plant and the measurement noise. For a finite measurement duration, this algorithm yields (possibly) noisy point frequency response samples of the plant and a worst case error bound. Given such noisy frequency response samples, two different families of worst case identification algorithms are presented. Each of these algorithms yields a model and a bound on the worst case infinity norm of error between the plant and the model, based on *a priori* and (in some cases) *a posteriori* data. One of the families of algorithms is robustly convergent and exhibits a certain optimality for a fixed model order. Both the families of algorithms are shown to be implementable as solutions to certain convex optimisation problems. The ideas and numerical techniques used for implementing these algorithms are further used to propose a method for identification in the $\nu-$gap metric.

Contents

List of Figures

Notation

$\mathbb{R}$	set of real numbers
$\mathbb{Z}$ ($\mathbb{Z}_+$)	set of integers (non-negative integers)
$\mathbb{C}$	set of complex numbers
$\mathbb{D}$	open unit disc in the complex plane
$\partial\mathbb{D}$	unit circle in the complex plane
$\in$	is an element of
$\subset$	is a subset of
$\exists$	there exists
$\forall$	for all
$:=$	the left hand side is defined by the right hand side
l_∞	the space of bounded sequences over $\mathbb{Z}$ or $\mathbb{Z}_+$
$\overline{B}l_\infty(\epsilon)$	$\{\, v \,:\, v \in l_\infty,\, \|v\|_\infty \leq \epsilon \,\}$
$\mathcal{H}_{\infty,\rho}$	$\left\{ f : \mathbb{D}_\rho \mapsto \mathbb{C}^{n\times m},\ f \text{ analytic in } \mathbb{D}_\rho,\ \|f\|_{\infty,\rho} = \sup_{z\in\mathbb{D}_\rho} \overline{\sigma}(f(z)) < \infty \right\}$
$\overline{B}\mathcal{H}_{\infty,\rho}(\gamma)$	$\{\, f : \mathbb{D}_\rho \mapsto \mathbb{C}^{n\times m},\, f \text{ analytic in } \mathbb{D}_\rho,\, \|f\|_{\infty,\rho} := \sup_{z\in\mathbb{D}_\rho} \|f(z)\| \leq \gamma,\, \rho > 1 \}$
$\overline{\delta}$	maximum separation between adjacent, angular measurement frequencies
$[0,\pi)^m$	the set of real $m-$vectors bounded element-wise in $[0,\pi)$
$\overline{\sigma}(\cdot)$	maximum singular value of operator or matrix $(\cdot)$
$\underline{\sigma}(\cdot)$	minimum singular value of operator or matrix $(\cdot)$
$tr(\cdot)$	trace $(\cdot)$
$[a_{ij}]$	matrix with element a_{ij} at position (i,j)

ACRONYMS

LSI	linear shift invariant
SISO	single input single output
MIMO	multiple input multiple output

Chapter 1

Introduction

1.1 Motivation and Background

The main reason to use feedback control is to reduce the effect of uncertainty in the plant description and to reduce the effect of disturbances (e.g. sensor or actuator noise). *Robust* control design aims at designing a controller which guarantees a certain level of the worst case performance over a *deterministic* set of plants. The set of plants is usually defined by a ball of uncertainty in a suitable function space centred at a nominal plant model. To guarantee a level of performance over the entire uncertain plant set, the least we need is an idea about the 'size' of uncertainty (or the radius of the above ball, measured in a suitable metric). Further, we are more interested in modelling how the plant behaves when in closed loop with a particular controller, or preferably, with *any* controller from a class of controllers. Traditional identification methods (e.g. least squares based techniques) do not address these requirements. 'Control Relevant Identification' refers to new identification techniques which focus on obtaining a model of the plant such that

- A deterministic upper bound on the worst case error between the true plant and the identified model is available, in terms of *a priori* information about the true plant and about the measurement noise. The error may be simple additive error or may be measured in more sophisticated, feedback oriented notions of measuring distance. Equivalently, these algorithms give a set of models that a controller must stabilise.

- As the length of data goes to infinity and the noise goes to zero, the identified model converges to the true plant in a suitable metric. If this convergence is independent of *a priori* information, the algorithm to obtain the model is said to be *robustly convergent*.

Most robustly convergent algorithms in literature depend on Fourier Transform based analysis, and hence assume uniformly spaced frequency response samples as *a posteriori* experimental data. In a practical situation, one would like to concentrate on system dynamics in a particular frequency range, and the 'uniform spacing' restriction would be viewed as rather severe. Further, the user has no control of specifying a model set in most methods. Lastly, the assumption on noise is usually simply that the noise is in a ball in l_∞. This inevitably results in rather conservative error bounds, and fails to use a possible *a priori* information that the noise would not reach its infinity norm too often. A suitable framework for constraining the noise further is needed.

In the light of above discussion, we list the prime objectives of this dissertation:

1. To establish a less conservative, rigorous deterministic framework for analysis of measurement noise in frequency response identification;

2. to derive worst case identification algorithms for non-uniform frequency spacing, to identify a model from a pre-defined model set;

3. using the machinery of the gap and the $\nu-$gap metrics, to explore the possibility of deriving feedback oriented identification techniques in a worst case setting.

1.2 Outline of the Dissertation

The dissertation is organised as follows.

Chapter 2: Here we collect together some mathematical preliminaries used in the dissertation. Various signal and system spaces are defined in section 2.1. Section 2.2 defines state-space realisation of a linear system. Section 2.3 introduces some related notions for continuous time systems. In section 2.4, we look at coprime factorisations and graph spaces. Section 2.5 introduces some notions of uncertainty and their relationship with robust control design. In section 2.6 optimisation under LMI constraints is briefly discussed.

Chapter 3: The prevalent methodology of prediction error identification is reviewed in this chapter. Closed loop identification techniques other than direct input-output identification are briefly explored. Methods to quantify uncertainty in identification are outlined.

Chapter 4: This chapter presents a robustly convergent algorithm for worst case identification of point frequency response samples of a linear system. The convergence result does

not require constructing exponential sinusoid inputs. A deterministic characterisation of the allowable noise set is introduced. This characterisation asymptotically models white noise in a particular probabilistic sense. An explicit bound on the identification error is obtained. The method is extended to a robustly convergent algorithm for simultaneous identification of multiple frequency response samples.

Chapter 5: The conventional method for worst case/deterministic identification in $\mathcal{H}_\infty$ is introduced in this chapter. The drawbacks of the conventional algorithms are outlined. A new bounded error type algorithm is suggested. This algorithm is shown to be equivalent to solving an LMI optimisation problem. Modifications are suggested to improve its convergence properties. A simulation example is used to demonstrate an application of the proposed algorithm.

Chapter 6: In this chapter, a stronger notion of robust convergence is introduced. An algorithm for identification with Finite Impulse Response (FIR) models is suggested, which is optimal in a certain sense for a finite model order. *A priori* worst case error bounds for two different cases, corresponding to the plant belonging to two different, specific subsets of $\mathcal{H}_\infty$ are obtained. Various modifications are proposed for obtaining smoother Nyquist plot for the approximation, or incorporating *a priori* knowledge about the poles of the plant. The algorithm is extended to identification of coprime factors. The chapter concludes with a simulation example.

Chapter 7: This chapter reviews the fundamentals of $\nu-$gap metric and proposes a new algorithm for identification in the $\nu-$gap metric. Some of the numerical methods from the previous chapter are used to set up a non-convex, but tractable optimisation problem. An example illustrates the use of the proposed method.

Chapter 8: The last chapter summarises the main contributions of this dissertation and identifies potential directions for future work.

Chapter 2

Preliminaries

In this chapter, we collect together some mathematical definitions useful for subsequent analysis.

2.1 Function Spaces

2.1.1 Signal Spaces

Let $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers respectively. $\mathbb{C}^{m\times n}$ ($\mathbb{R}^{m\times n}$) denotes the space of $m \times n$ complex (real) matrices. $\mathbb{Z}$ represents the set of integers, with $\mathbb{Z}_+ = \{t \in \mathbb{Z}, t \geq 0\}$. l_2^n represents the Hilbert space of square summable sequences over $\mathbb{Z}$ or $\mathbb{Z}_+$, which take values in $\mathbb{C}^n$ or $\mathbb{R}^n$. Inner product in l_2^n is defined by

$$< u, v >= \sum_{t\in\mathbb{Z}} u^*(t)v(t)$$

For complex vector valued signals, $u^*(t)$ represents the conjugate transpose.

For $\rho \geq 1$, let $\mathbb{D}_\rho := \{z \in \mathbb{C} : |z| < \rho\}$. Let $\partial\mathbb{D}_\rho$ and $\overline{\mathbb{D}}_\rho$ denote the boundary and the closure of $\mathbb{D}_\rho$ respectively. $\mathbb{D}_1$, $\partial\mathbb{D}_1$ and $\overline{\mathbb{D}}_1$ are represented as $\mathbb{D}$, $\partial\mathbb{D}$ and $\overline{\mathbb{D}}$ respectively. $l_2^n(\mathbb{Z})$ is isomorphic to the space $\mathcal{L}_2{}^n$ of square integrable functions on $\partial\mathbb{D}$ through Fourier transform:

$$\hat{v}(e^{j\omega}) = \sum_{t\in\mathbb{Z}} v(t)e^{j\omega t}$$

The l_2 norm of a signal is computed by

$$\|v\|_2^2 = \sum_{t\in\mathbb{Z}} v^*(t)v(t) = \frac{1}{2\pi}\int_0^{2\pi} \hat{v}^*(e^{j\omega})\hat{v}(e^{j\omega})d\omega$$

l_2^n may be decomposed into two signal spaces, $l_2^n[0,\infty)$ and $l_2^n(-\infty,0)$. $l_2^n[0,\infty)$ is the space of signals defined for nonnegative time and $l_2^n(-\infty,0)$ is the space of signals defined for negative time. The two spaces are clearly orthogonal. Correspondingly, ${\mathcal{L}_2}^n$ in the frequency domain may be decomposed into two spaces; $\mathcal{H}_2^n$ is the space of Fourier transforms of signals in $l_2^n[0,\infty)$, and $\mathcal{H}_2^{n\perp}$ is the space of Fourier transforms of signals in $l_2^n(-\infty,0)$.

Another signal space of interest is l_∞, the space of bounded sequences over $\mathbb{Z}$ or $\mathbb{Z}_+$. l_∞ norm of a signal is given by

$$\|v\|_\infty = \max_{t\in\mathbb{Z}} |v|$$

A closed ball in l_∞ is defined by

$$\overline{B} l_\infty(\epsilon) := \{\, v \,:\, v \in l_\infty,\, \|v\|_\infty \le \epsilon \,\} \tag{2.1}$$

2.1.2 System Spaces

A system is an operator mapping between signal spaces. The operators we will be dealing with in this thesis are linear shift-invariant operators. A linear operator $P : l_2^m \mapsto l_2^n$ has an infinite matrix representation

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where each $p_{ij} \in \mathbb{C}^{n\times m}$ is for a time index in $\mathbb{Z}_+$. (For $\mathbb{Z}$, a doubly infinite matrix would be used.)

An operator P is said to be *linear shift invariant* (LSI) if $\lambda P = P\lambda$ where λ is a unit delay operator, $(\lambda v)(k) = v(k-1)$. Equivalently, the operator is described by a convolution kernel

$$(Pv)(t) = \sum_{\tau=-\infty}^{\infty} p_\tau v(t-\tau)$$

An LSI operator $P : \mathcal{D}_P \subseteq l_2^m[0,\infty) \mapsto l_2^n[0,\infty)$ has a matrix representation

$$P = \begin{bmatrix} p_0 & 0 & 0 & 0 & \cdots \\ p_1 & p_0 & 0 & 0 & \cdots \\ p_2 & p_1 & p_0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{2.2}$$

where $\mathcal{D}_P = \{u \in l_2^m[0,\infty) : Pu \in l_2^n[0,\infty)\}$ is called the *domain* of P. An LSI operator $P : \mathcal{D}_P \subseteq l_2^m[0,\infty) \mapsto l_2^n[0,\infty)$ is said to be *bounded on* $\mathcal{D}_P$ (or simply, *bounded*) if there exists a finite γ such that $\|Pv\|_2 \leq \gamma\|v\|_2 \,\forall\, v \in \mathcal{D}_P$.

A bounded LSI operator $P : \mathcal{D}_P \subseteq l_2^m[0,\infty) \mapsto l_2^n[0,\infty)$ is equivalent to a multiplication operator $\hat{P} : \hat{\mathcal{D}_P} \subseteq \mathcal{H}_2^m \mapsto \mathcal{H}_2^n$ with a symbol in the *Banach* space of matrix transfer functions

$$\mathcal{L}_\infty := \left\{ f : \partial\mathbb{D} \mapsto \mathbb{C}^{n\times m}, \|f\|_\infty = \operatorname{ess\,sup}_{z\in\partial\mathbb{D}} \overline{\sigma}(f(e^{j\omega})) < \infty \right\}$$

Here, $\hat{\mathcal{D}_P} = \{u \in \mathcal{H}_2^m \,:\, Pu \in \mathcal{H}_2^n\}$. If the domain $\mathcal{D}_P$ of P equals $l_2^m[0,\infty)$, the multiplication operator corresponding to P in $\mathcal{L}_\infty$ has an analytic continuation in $\mathbb{D}$ ([You95], chapter 13). A bounded LSI operator $P : l_2^m[0,\infty) \mapsto l_2^n[0,\infty)$ is equivalent to a multiplication operator $\hat{P} : \mathcal{H}_2^m \mapsto \mathcal{H}_2^n$ with a symbol in *Hardy* space of matrix transfer functions

$$\mathcal{H}_\infty := \left\{ f : \mathbb{D} \mapsto \mathbb{C}^{n\times m},\ f \text{ analytic in } \mathbb{D},\ \|f\|_\infty = \sup_{z\in\mathbb{D}} \overline{\sigma}(f(z)) < \infty \right\}$$

For $f \in \mathcal{H}_\infty$, $\|f\|_\infty = \sup_{z\in\mathbb{D}} \overline{\sigma}(f(z)) = \operatorname{ess\,sup}_{\partial\mathbb{D}} \overline{\sigma}(f(e^{j\omega}))$ [You95]. Dynamical systems represented by multiplication operators with symbols in $\mathcal{H}_\infty$ give an output of bounded energy (i.e. in l_2) for any input of bounded energy, starting from any (finite) initial condition. A system with transfer function in $\mathcal{H}_\infty$ is said to be *stable*.

For a given $\lambda \in \mathbb{D}$, the Taylor series expansion of $P \in \mathcal{H}_\infty$ about the origin is given by[1]

$$P(\lambda) = \sum_{t=0}^{\infty} p_t \lambda^t$$

where p_t are the *impulse response* matrices in (2.2) corresponding to P.

$\mathcal{H}_{\infty,\rho}$ denotes the normed space of functions analytic in $\mathbb{D}_\rho$ and having norm $\|f\|_{\infty,\rho} := \sup_{z\in\mathbb{D}_\rho} |f(z)| < \infty$. A ball in $\mathcal{H}_{\infty,\rho}$ is defined by

$$\overline{B}\mathcal{H}_{\infty,\rho}(\gamma) = \{\, f : f \text{ analytic in } \mathbb{D}_\rho \text{ and } \|f\|_{\infty,\rho} := \sup_{z\in\mathbb{D}_\rho} |f(z)| \leq \gamma,\ \rho > 1\} \tag{2.3}$$

For a SISO plant $P \in \overline{B}\mathcal{H}_{\infty,\rho}(\gamma)$, $P(z) = \sum_{i=0}^{\infty} p_k z^k$, it can be shown that the impulse response parameters p_k obey

$$|p_k| \leq \gamma\rho^{-k} \tag{2.4}$$

A proof may be found in ([Kre88], chapter 13).

[1]Where there is no risk of confusion, we will use same symbol for an LSI operator$P : l_2^m \mapsto l_2^n$ and its multiplication operator symbol in $\mathcal{L}_\infty$.

$\mathcal{R}$ represents the space of all real rational matrix transfer functions. $\mathcal{RL}_\infty$ ($\mathcal{RH}_\infty$) represents the subspace of real rational matrix transfer functions in $\mathcal{L}_\infty$ ($\mathcal{H}_\infty$). A transfer function in $\mathcal{R}$ corresponds to a finite dimensional dynamical system.

For $P \in \mathcal{L}_\infty$, another operator of interest with symbol P is the *Hankel* operator $\mathbf{H}_P$, defined by

$$\mathbf{H}_P : \mathcal{H}_2^\perp \mapsto \mathcal{H}_2 \quad \mathbf{H}_P(u) = \Pi_{\mathcal{H}_2} P u \quad \text{for} \quad u \in \mathcal{H}_2^\perp$$

where $\Pi_{\mathcal{H}_2}$ represents the orthogonal projection onto $\mathcal{H}_2$. More on Hankel operators may be found in [ZDG96].

2.2 State Space Realisation

Consider a discrete time system of equations

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \end{aligned} \tag{2.5}$$

where for any fixed $k \in Z_+$, the *state* $x_k \in \mathbb{R}^p$, the input $u_k \in \mathbb{R}^m$, the output $y_k \in \mathbb{R}^n$ and $A \in \mathbb{R}^{p\times p}$, $B \in \mathbb{R}^{p\times m}$, $C \in \mathbb{R}^{n\times p}$, $D \in \mathbb{R}^{n\times m}$ are bounded, constant matrices. With $x_0 = 0$, this difference equation characterises an LSI operator $P : \mathcal{D}_P \subseteq l_2^m[0,\infty) \mapsto l_2^n[0,\infty)$ which is equivalent to a multiplication operator with symbol in $\mathcal{R}$:

$$y(e^{j\omega}) = P(e^{j\omega})u(e^{j\omega}), \quad P(e^{j\omega}) := C(e^{-j\omega}I - A)^{-1}B + D$$

The system of equations (2.5) is called a *state space realisation* of P. The set of eigen values of A, denoted by spec(A) characterises the stability of P. If spec(A) $\subset \mathbb{D}$, then the multiplication operator corresponding to P has a symbol in $\mathcal{RH}_\infty$. If λ_i is an eigen value of A repeated k times, $\frac{1}{\lambda_i}$ is called a *pole of order k* of P. $\lambda_i = 0$ corresponds to a pole at ∞. If $\lambda_i \in \overline{\mathbb{D}}$ ($\lambda_i \in \mathbb{C} - \overline{\mathbb{D}}$), the pole $\frac{1}{\lambda_i}$ is said to be stable (unstable). We denote the number of unstable poles of P (counting a pole of order k, k-times) by $\eta(P)$.

2.3 Continuous Time Systems

Mostly, the discrete time dynamical systems we will be dealing with are approximations of continuous time systems. In continuous time, l_2 is defined as the Hilbert space of square

integrable signals with inner product

$$< u, v >= \int_{-\infty}^{\infty} u^*(t)v(t)dt \text{ and norm } \|x\|_2 = \sqrt{< x, x >}$$

For bounded, LSI operators in continuous time, the matrix transfer function spaces $\mathcal{L}_\infty$ and $\mathcal{H}_\infty$ are respectively defined as

$$\mathcal{L}_\infty = \left\{ f : j\mathbb{R} \mapsto \mathbb{C}^{m\times n}, \|f\|_\infty := ess \sup_{\omega \in j\mathbb{R}} \overline{\sigma}(f(j\omega)) < \infty \right\}$$
$$\mathcal{H}_\infty = \left\{ f : \mathbb{C} \mapsto \mathbb{C}^{m\times n}, f \text{ analytic in ORHP}, \|f\|_\infty := \sup_{Re(s)>0} \overline{\sigma}(f(s)) < \infty \right\} \quad (2.6)$$

where s is a complex variable, $j\mathbb{R}$ is the imaginary axis and ORHP is the open right half plane. For $f \in \mathcal{H}_\infty$, $\|f\|_\infty = \lim_{\alpha\downarrow 0} \sup_{\alpha>0} \overline{\sigma}(f(\alpha+j\omega)) = \sup_{\omega \in j\mathbb{R}} \overline{\sigma}(P(j\omega))$ (e.g., [GL95], chapter 3).

A continuous time system may be related to an equivalent discrete time system through a bilinear transformation. Suppose, the input and the output of a continuous time system P_c are sampled synchronously at a uniform sampling period T. Then an equivalent discrete time system P_d may be obtained through a bilinear transformation as

$$P_d(z) = P_c\left(\frac{2}{T}\frac{1-z}{1+z}\right)$$

This maps the imaginary axis in the complex plane to $\partial\mathbb{D}$, and ORHP to $\mathbb{D}$. The number of unstable poles and zeroes (in ORHP or in $\mathbb{D}$) is unchanged; as is the supremum of singular value of the frequency response. This is important in the context of this thesis; a *small* difference in the frequency response of a discrete system and its approximation results in a small difference in the frequency responses of the corresponding continuous time systems they represent.

2.4 Coprime Factorisation and Graphs of systems

Two functions M, $N \in \mathcal{H}_\infty$ with same number of columns are said to be *right coprime* if the matrix $\begin{bmatrix} N \\ M \end{bmatrix}$ is left invertible in $\mathcal{H}_\infty$, *i.e.* $\exists Q \in \mathcal{H}_\infty$ s.t. $Q \begin{bmatrix} N \\ M \end{bmatrix} = I$. An ordered pair $\{N, M\}$ is called a *right coprime factorisation* (rcf) of $P \in \mathcal{R}$ (respectively, of $P \in \mathcal{H}_\infty$) if N, M are right coprime, $P = NM^{-1}$ and $M^{-1} \in \mathcal{R}$ (respectively, $M^{-1} \in \mathcal{H}_\infty$). rcf is non-unique; if $\{N, M\}$ is an rcf of P, so is $\{NQ, MQ\}$ for any Q *unimodular* in $\mathcal{H}_\infty$, i.e. Q, $Q^{-1} \in \mathcal{H}_\infty$. For $P \in \mathcal{R}$, rcf always exists, and is unique within multiplication by a

unimodular transfer matrix [Vid85]. For any $P \in \mathcal{H}_\infty^{m\times n}$ corresponding to a system with possibly infinite dimensional state space but finite dimensional input-output spaces, an rcf is given by $\{P, I_{n\times n}\}$.

A particularly useful rcf is the *normalised right coprime factorisation*: An ordered pair $\{N_0, M_0\}$ is called a normalised rcf (nrcf) of P_0 if it is an rcf of P_0 and $\begin{bmatrix} N_0 \\ M_0 \end{bmatrix}$ is *inner*; *i.e.* $N_0^* N_0 + M_0^* M_0 = I$. $G_0 := \begin{bmatrix} N_0 \\ M_0 \end{bmatrix}$ is called the normalised right graph symbol of P_0.

The *Graph* of an operator is the set of all bounded output-input pairs:

$$\mathcal{G}_{P_0} = \left\{ \begin{bmatrix} P_0 u \\ u \end{bmatrix} : u \in \mathcal{D}_{P_0} \right\}$$

$\mathcal{G}_{P_0}$ is related to G_0 by

$$\mathcal{G}_{P_0} = G_0 \mathcal{H}_2$$

Analogous results hold for left coprime factorisation, which will be stated in brief.

- Two functions $\tilde{M}$, $\tilde{N} \in \mathcal{H}_\infty$ with same number of rows are said to be *left coprime* if the matrix $\begin{bmatrix} -\tilde{M} & \tilde{N} \end{bmatrix}$ is right invertible in $\mathcal{H}_\infty$, *i.e.* $\exists Q \in \mathcal{H}_\infty$ s.t. $\begin{bmatrix} -\tilde{M} & \tilde{N} \end{bmatrix} Q = I$.

- An ordered pair $\{\tilde{N}, \tilde{M}\}$ is called a *left coprime factorisation* (lcf) of $P \in \mathcal{R}$ (respectively, of $P \in \mathcal{H}_\infty$) if $\{\tilde{N}, \tilde{M}\}$ are left coprime, $P = \tilde{M}^{-1}\tilde{N}$ and $\tilde{M}^{-1} \in \mathcal{R}$ (respectively, $\tilde{M}^{-1} \in \mathcal{H}_\infty$).

- An ordered pair $\{\tilde{N}_0, \tilde{M}_0\}$ is called a *normalised* lcf (nlcf) of P_0 if it is an lcf of P_0 and $\begin{bmatrix} -\tilde{M}_0 & \tilde{N}_0 \end{bmatrix}$ is *co-inner*; *i.e.* $\tilde{N}_0 \tilde{N}_0^* + \tilde{M}_0 \tilde{M}_0^* = I$. $\tilde{G}_0 = \begin{bmatrix} -\tilde{M}_0 & \tilde{N}_0 \end{bmatrix}$ is called the normalised left graph symbol of P_0.

2.5 Uncertainty Representation

Given input-output data and possibly, some *a priori* information about a dynamical system P, we would like to find a model of the system, $\hat{P}$, generally in $\mathcal{R}$. Our model will never be exact for various reasons:

- No physical system is *exactly* linear and finite dimensional,

- The process of collecting input/output data from the system introduces its own uncertainty (due to actuator/sensor noise, finite word-length of A/D-D/A converters etc),
- Even if the system were linear, a finite number of experiments, each involving a finite data set may fail to characterise it uniquely.

All these causes introduce an error between P and $\hat{P}$. We assume that the true plant lies in a set defined by the 'nominal' model and some quantitative description of uncertainty. In the context of this thesis, we consider two methods for quantifying uncertainty:

- Additive Uncertainty: The true plant lies in a set defined by

$$\left\{ f : f \in \mathcal{RL}_\infty, \|\hat{P} - f\|_\infty \leq \beta, \, \eta(\hat{P}) = \eta(f) \right\} \tag{2.7}$$

  for a given uncertainty level, or error bound, β.

- nrcf Uncertainty: Given nrcf $\{\hat{N}, \, \hat{M}\}$ of the nominal model, the true plant P lies in a set defined by

$$\left\{ f : f = (\hat{N} + \Delta N)(\hat{M} + \Delta M)^{-1}, \, \begin{bmatrix} \Delta N \\ \Delta M \end{bmatrix} \in \mathcal{RH}_\infty, \, \left\| \begin{bmatrix} \Delta N \\ \Delta M \end{bmatrix} \right\|_\infty \leq \beta \right\} \tag{2.8}$$

As shown in [GS90], (2.8) is also equivalent to $\{f : \overrightarrow{\delta_g}(f, \hat{P}) \leq \beta\}$, where the *directed gap* $\overrightarrow{\delta_g}(f, \hat{P})$ between the graph spaces of f and $\hat{P}$ is defined by

$$\overrightarrow{\delta_g}(f, \hat{P}) = \sup_{\begin{bmatrix} y_1 \\ u_1 \end{bmatrix} \in \mathcal{G}_f} \inf_{\begin{bmatrix} y_2 \\ u_2 \end{bmatrix} \in \mathcal{G}_{\hat{P}}} \frac{\left\| \begin{bmatrix} y_1 \\ u_1 \end{bmatrix} - \begin{bmatrix} y_2 \\ u_2 \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} y_1 \\ u_1 \end{bmatrix} \right\|_2} \tag{2.9}$$

One of the main purposes of using feedback control is to reduce the effect of uncertainty about the system dynamics. If a controller C stabilises $\hat{P}$, then it stabilises the plant set described by (2.7) if and only if $\|C(I - \hat{P}C)^{-1}\|_\infty < \beta^{-1}$. Similarly, a controller C stabilises the plant set in (2.8) if and only if it stabilises $\hat{P}$ and

$$\left\| \begin{bmatrix} \hat{P} \\ I \end{bmatrix} (I - C\hat{P})^{-1} \begin{bmatrix} -C & I \end{bmatrix} \right\|_\infty < \beta^{-1} \tag{2.10}$$

If $\hat{P}$ is appropriately frequency weighted, then (2.10) may be linked to nominal performance as well as robust stability [MG92].

Other uncertainty descriptions are described in standard robust control texts [ZDG96], [GL95].

2.6 Linear Matrix Inequalities

Computational methods suggested in this dissertation rely heavily on convex optimisation problems subject to Linear Matrix Inequality (LMI) constraints. LMI constraints have the form

$$F(x) = F_0 + \sum_{i=1}^{m} x_i F_i > 0 \text{ (or } \geq 0) \tag{2.11}$$

where $x \in \mathbb{R}^m$ is variable and the symmetric matrices F_i are given. The inequality symbol indicates positive definiteness (or semi-definiteness, in case of non-strict inequality). Two different LMIs $F^1(x) > 0,\ F^2(x) > 0$ may be combined as a single LMI

$$diag\ \left(F^1(x), F^2(x)\right) > 0$$

(2.11) is a convex constraint in x, *i.e.* the feasible set $\{x \,|\, F(x) > 0\}$ is a convex set. A large variety of linear and quadratic constraints arising in control and identification may be written as LMI constraints. A useful tool for converting quadratic constraints into affine constraints is Schur inequality [BGFB94]:

$$\begin{bmatrix} Q(x) & S(x) \\ S(x)^T & R(x) \end{bmatrix} > 0 \ \Leftrightarrow\ R(x) > 0,\ Q(x) - S(x)R(x)^{-1}S(x)^T > 0 \tag{2.12}$$

where $Q(x) = Q(x)^T$, $R(x) = R(x)^T$ and $S(x)$ are affine functions of the variable x.

The LMI problems we encounter in chapters 5 - 7 are generalised eigen value (gevp) problems. The general form of such problems is

minimise λ subject to $\lambda B(x) - A(x) > 0, \quad B(x) > 0,\ C(x) > 0\,.$

Here $A(x)$, $B(x)$, $C(x)$ are symmetric matrices that depend affinely on x. Efficient numerical methods exist for solving gevp and related optimisation problems, and software packages implementing such optimisation routines (such as LMI Control Toolbox from MATLAB) are commercially available.

Chapter 3

Time Domain Identification

3.1 Introduction

Identification is the determination on the basis of input and output, of a model within a specified set of models, which best approximates the input-output behaviour of the system under test. The quality of approximation is defined in terms of a criterion which is a function of measured inputs and outputs and predicted outputs (and possibly, predicted inputs). The set of systems from which a model is chosen is usually defined by a *model structure*. From a control point of view, model structures of choice are those representing finite order linear difference equations with real coefficients (i.e. real rational transfer function or state space models). Once the structure is chosen, an appropriate excitation is applied to the true plant and input-output data is measured. 'Appropriateness' here refers to the ability to outweigh the effect of disturbances and to excite the system dynamics we wish to model. This measured input-output data is then mapped into a model which best explains this data, in a suitable sense.

Typically, we will be interested in modelling the system accurately over a certain desired bandwidth. There are two ways to carry out this process:

- Apply an excitation 'rich' in desired frequencies and directly identify a parametric model from input-output data. This approach is reviewed in the present chapter.

- Find a non-parametric frequency response of plant by performing one or more experiments with periodic inputs. Then find a parametric model based on this frequency response samples. This approach is discussed in subsequent chapters.

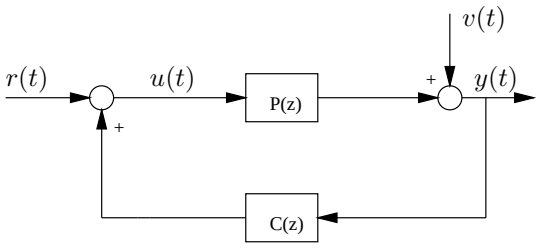


Figure 3.1: Closed Loop Identification

3.2 Prediction Error Identification

3.2.1 Basic Set-up

Consider a closed loop system as shown in fig.(3.1). The following assumptions are made about the system:

1. The reference signal $r(t)$ is quasi-stationary [Lju99], i.e. it is either a stationary stochastic process or a bounded deterministic sequence such that the limits

$$R_r(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} r(t) r^T(t-\tau)$$

   exist for all τ.

2. $v(t) = L(q)e(t)$ where L is monic and inversely stable linear filter and e is a stationary zero mean white process having bounded moments of order $4 + l$, $l > 0$.

$$E e_k \, e_s^T = \Lambda \delta(t-s)$$

   where $\delta(\cdot)$ is Dirac delta function.

3. The plant P is strictly proper, $P(0) = 0$.

4. The controller C is linear and the loop is asymptotically stable.

Note that open loop identification of a stable plant is simply a special case of the above set-up, with $C = 0$.

Assumption 3 can be relaxed; it is sufficient that either C or P contains a delay. However, physical systems do not react instantaneously when subjected to any excitation; hence $P(0) = 0$ is a reasonable assumption. Assumption 4 can also be relaxed so long as the loop is stable; the exact condition the closed loop must satisfy if the controller is nonlinear are given in ([FL98], section 3).

3.2.2 The Method

An approach considered here disregards presence of feedback and identifies a model *directly* from its input-output data.

We first need some formal definitions. Define a set of monic and inversely stable filters:

$$\mathcal{L} = \{\, L \,|\, L \in \mathcal{R},\, L^{-1} \in \mathcal{RH}_\infty\,,\, L(0) = I\}$$

Next, define a set of *ordered pairs* (L, P) of transfer matrices :

$$\mathcal{LP} = \{\, (L, P) \,|\, L \in \mathcal{L},\, P \in \mathcal{R},\, P(0) = 0,\, L^{-1}P \in \mathcal{RH}_\infty\}$$

Then a *model structure* is a differentiable mapping from a connected and open subset of n dimensional parameter space $\mathbb{R}^n$ to a subset $\mathcal{LP}_A$ of $\mathcal{LP}$,

$$\mu : \mathbb{R}^n \supseteq D_\mu \rightarrow \mathcal{LP}_A \subseteq \mathcal{LP},\; \mu(\theta) = (\, L(q,\theta),\; P(q,\theta)\,)$$

such that the gradient $\frac{\partial}{\partial\theta}\begin{bmatrix} L(q,\theta) & P(q,\theta)\end{bmatrix}$ is stable. This model structure is used to describe the relationship between the measurable input $u(t)$ and measured output $y(t)$ as

$$y(t) = P(q,\theta)u(t) + L(q,\theta)e(t) \tag{3.1}$$

where $e(t)$ is a zero mean white process. q is a backward shift operator.

Define one-step ahead prediction error for a model structure μ at a parameter vector (θ^*) and at sample time t by

$$\epsilon(t, \mu(\theta^*)) = y(t) - \hat{y}(t|\mu(\theta^*)) \tag{3.2}$$

where $\hat{y}(t|\mu(\theta^*))$ is a one-step ahead prediction of the output based on the data up to sample time $t-1$. To find an optimal one-step ahead prediction, (3.1) is rewritten as

$$y(t) = [I - L^{-1}(q,\theta^*)]y(t) + L^{-1}(q,\theta^*)P(q,\theta^*)u(t) + e(t) \tag{3.3}$$

For a sufficiently large t (i.e. barring the starting up of the IIR filters from zero initial conditions), all the terms on the right hand side except the disturbance $e(t)$ are determined from the past data up to $y(t-1)$. It may be shown that the following choice of one-step ahead prediction minimises the variance of the prediction error $\epsilon(t, \mu(\theta^*))$ [Lju99] :

$$\hat{y}(t|\mu(\theta^*)) = [I - L^{-1}(q,\theta^*)]\, y(t) + L^{-1}(q,\theta^*)P(q,\theta^*)u(t) \tag{3.4}$$

Restricting (L, P) to $\mathcal{LP}$ ensures that the above one-step ahead predictor is stable. Let

$$Z^N = [y(1)\, u(1)\, y(2)\, u(2)\, \cdots y(N)\, u(N)] \tag{3.5}$$

For this data set and a model structure $\mu(\theta)$, the parameter estimate $\hat{\theta}_N$ is defined by

$$V_N(\mu(\hat{\theta}_N), Z^N) = \min_{\theta \in D_\mu} V_N(\mu(\theta), Z^N) \tag{3.6}$$

where

$$V_N(\mu(\theta), Z^N) = l\left(\frac{1}{N}\sum_{t=1}^{N} \epsilon(t, \mu(\theta))\epsilon^T(t, \mu(\theta))\right) \tag{3.7}$$

and $l(\cdot)$ is a scalar, positive valued, twice differentiable function such that $l(Q + \Delta Q) \geq l(Q)$ whenever $\Delta Q \geq 0$, with equality achieved only at $\Delta Q = 0$. A candidate cost function considered here $l(Q) := tr(Q)$. In ([SS89], chapter 7), it is shown that the choice $l(Q) = det(Q)$ corresponds to maximum likelihood estimation in the case of gaussian noise. Different parameterisations of L and P are discussed in [Lju99]. The optimisation problem (3.6) is usually solved by some modified variant of Gauss-Newton method [Fle87].

3.2.3 Convergence

In the discussion that follows, $\mathcal{P}(x > \tau)$ is used to denote the probability of an event $\{x > \tau\}$ for a random variable x. Let x_N be an indexed sequence of random variables and x^* be a random variable. Then $x_N \to x^*$ with probability (w.p.) 1 as $N \to \infty$ if $\lim_{N\to\infty} \mathcal{P}(x_N \to x^*) = 1$ ([Pap91], chapter 8, [Lju99], Appendix 1). Here, $\{x_N \to x^*\}$ represents an event consisting of all outcomes ζ such that the sequence $x_N(\zeta)$ converges to $x^*(\zeta)$.
Define

$$\overline{V}(\mu(\theta)) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E\left(\epsilon(t, \mu(\theta))^T \epsilon(t, \mu(\theta))\right)$$

where E denotes the expectation operator. Suppose the true system generating data is given by

$$y(t) = P_0\, u(t) + L_0\, e(t)$$

where L_0, P_0, e and the reference r which generates the input u (possibly, along-with feedback) satisfy the relevant assumptions in section 3.2.1. The following lemma summarises the main results on the asymptotic behaviour of prediction error estimate:

Lemma 3.1 *([FL98])*

1. *Suppose* $\exists$ *a unique minimiser* $\tilde{\theta} \in D_\mu$ *s.t.* $\overline{V}(\mu(\tilde{\theta})) = \min_{\theta \in D_\mu} \overline{V}(\mu(\theta))$.

$$\begin{aligned} \textit{Then} \qquad & \tilde{\theta} = \arg\min_{\theta \in D_\mu} \int_{-\pi}^{\pi} tr(\phi_\epsilon)\, d\omega \quad \textit{and} \\ & \hat{\theta}_N \to \tilde{\theta} \ \textit{ w.p. } 1 \quad \textit{as } N \to \infty \end{aligned} \tag{3.8}$$

where $\phi_\epsilon = L_\theta^{-1} \begin{bmatrix} (P_0 - P_\theta) & (L_0 - L_\theta) \end{bmatrix} Q \begin{bmatrix} (P_0 - P_\theta)^* \\ (L_0 - L_\theta)^* \end{bmatrix} L_\theta^{-*}$, $Q = \begin{bmatrix} \phi_u & \phi_{ue} \\ \phi_{eu} & \Lambda \end{bmatrix}$ *and* ϕ_{ue} *is the cross-spectrum between* u *and* e. Λ *is as defined in section 3.2.1. For brevity,* $P(e^{j\omega}, \theta)$ *is written as* P_θ, $P_0(e^{j\omega})$ *as* P_0, *and the same holds for* L_0, L_θ.

2. *Define* $T(q, \theta) := \begin{bmatrix} P(q, \theta) \\ L(q, \theta) \end{bmatrix}$. *Then, as model order* $n \to \infty$ *and data length* $N \to \infty$ *such that* $\lim_{n,\, N \to \infty} \frac{n^2}{N}$ *is finite* [1],

$$Cov\, vec\,(\hat{T}_N(\omega)) \approx \frac{n}{N} \begin{bmatrix} \phi_u(\omega) & \phi_{ue}(\omega) \\ \phi_{eu}(\omega) & \Lambda \end{bmatrix}^{-T} \otimes L_0(e^{j\omega})\, \Lambda\, L_0(e^{j\omega})^* \tag{3.9}$$

where $\otimes$ *denotes Kroneker product.*

We make a number of observations regarding these two convergence results-

- If the input is a summation of n periodic signals, then the matrix Q in the definition of ϕ_ϵ is positive definite at least at n frequencies. If the true plant P and the model P_θ are both of order $n-1$ or less and $Q(e^{j\omega_i}) > 0 \,\forall\, i \in [1, n]$ then

$$\begin{aligned} & tr\left((P - P_\theta)\phi_u(P - P_\theta)^*\right)(e^{j\omega_i}) = 0 \,\forall\, i \in [1, n] \quad \text{if and only if} \\ & P(e^{j\omega_i}) \equiv P_\theta(e^{j\omega_i}) \,\forall\, i \in [1, n] \end{aligned}$$

  In other words, the error spectrum between the model and the plant is guaranteed to be nonzero at some frequency $e^{j\omega_i}$ so long as the model doesn't converge to the true plant. This condition is referred to as *persistent excitation of order n* ([Lju99], chapter 13). If the matrix Q is positive definite at all frequencies, the input is said to be persistently exciting.

- The input spectrum ϕ_u consists of contributions from noise and reference signal:

$$\phi_u = S_i\, \phi_r\, (S_i)^* + S_i C L_0 \Lambda L_0^* (S_i C)^*$$

[1] The exact technical conditions on n, N and on the data $\{y, u\}$ under which (3.9) holds are somewhat more involved; see ([FL98], Appendix A.2). However, only the implications of (3.9) are of interest here.

and the cross-spectrum ϕ_{ue} is given by

$$\phi_{ue} = S_i C L_0 \Lambda$$

Here S_i is the sensitivity function, $S_i = (I - CP)^{-1}$. Consider a SISO system for simplicity, with $\Lambda = \lambda_0$. From standard linear algebra ([ZDG96], chapter 2)

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & \times \\ \times & \times \end{bmatrix}$$

Using this result, the upper left term in (3.9) is given by

$$\begin{aligned} cov(\hat{P}_N(q,\theta)) &= \frac{n}{N}\frac{1}{\phi_u}\left(1 + \frac{|\phi_{ue}|^2}{\phi_u}\frac{1}{\lambda_0 - \frac{|\phi_{ue}|^2}{\phi_u}}\right)\phi_v \\ &= \frac{n}{N}\frac{\phi_v}{|S_i|^2\phi_r} \end{aligned}$$

This shows that the asymptotic covariance of transfer matrix at any frequency depends on spectral ratio of noise to the part of input signal injected from reference at that frequency. For a given (or constrained) spectral distribution of input power ϕ_r, feedback will improve asymptotic noise to signal ratio over frequency range where $|S_i|$ is large, and will deteriorate it where $|S_i|$ is small.

- As [FL98] shows, ϕ_ϵ may also be written as

$$\phi_\epsilon = L_\theta^{-1}\begin{bmatrix}(P_0 + B_\theta - P_\theta) & (L_0 - L_\theta)\end{bmatrix}\tilde{Q}\begin{bmatrix}(P_0 + B_\theta - P_\theta)^* \\ (L_0 - L_\theta)^*\end{bmatrix}L_\theta^{-*}$$

where $\tilde{Q} = \begin{bmatrix}\phi_u & 0 \\ 0 & \Lambda - \phi_{eu}\phi_u^{-1}\phi_{ue}\end{bmatrix}$ and $B_\theta = (L_0 - L_\theta)\phi_{eu}\phi_u^{-1}$. This 'bias term' B_θ is small if the term $\phi_{eu}\phi_u^{-1}$ is small and/or noise model is flexible enough. If $\phi_{eu} \approx 0$ (typically for an open loop system), we may choose *not to* parameterise L and use instead a fixed data pre-filter, $\hat{L}$. Then the expression (3.8) becomes

$$\hat{\theta}_N \to D_c := \arg\min_\theta \int_{-\pi}^{\pi} tr(\hat{L}^{-1}(P_0 - P_\theta)\phi_u(P_0 - P_\theta)^*\hat{L}^{-*})d\omega \tag{3.10}$$

This is seen as a best 2− norm approximation of $P_0(e^{j\omega})$, weighted by ϕ_u and filtered by $\hat{L}$.

3.2.4 Other Closed Loop Identification Methods

The direct method described so far is not the only method for closed loop identification. To see an alternate scenario, consider a SISO plant P_0 in closed loop with controller C as in fig. (3.1). Suppose, our control objective is to minimise some norm of $J(P,C) := \frac{P}{1+PC}$. One may choose a model structure,

$$y(t) = \frac{P_\theta}{1 + P_\theta C}\, r(t) + e(t) \tag{3.11}$$

Crucially, the disturbance term and the 'input' term are such that $\phi_{re} = 0$. As in (3.10), one may now use a fixed noise filter to shape the bias error in identification and still get good estimates. To pose a numerically more tractable problem, one may further approximate (3.11) through iterations. i^{th} iteration proceeds as:
Perform an identification experiment in closed loop with the controller C_i in place. Then

$$\hat{P}_{i+1} = \arg\min_{P_\theta} \left\| J(P_0, C_i) - \frac{P_\theta}{1 + P_{i-1}C_i} \right\| \tag{3.12}$$

$$\text{and} \quad C_{i+1} = \arg\min_{C} \|J(\hat{P}_{i+1}, C)\| \tag{3.13}$$

where minima are taken over appropriate sets of model/controllers. $\|J(P_0, C_i) - J(P, C_i)\|$ is usually approximated by a least squares problem over a finite time or frequency domain data using a suitable choice of fixed noise filter and a suitable excitation r. Iterative strategies such as these are widely discussed in literature. [vdHS95] provides a comprehensive review. There are a variety of iterative algorithms depending on the control objective $J(P,C)$; there are iterative identification and control design algorithms intended for use with LQ control [ZBG95], with $\mathcal{H}_\infty$ optimisation [BM98] and with model predictive control [RPG92]. The main advantage of these methods over direct prediction error appears to be their ability to shape the bias errors over the desired bandwidth. Their main weakness is a lack of performance guarantees - the achieved performance $\|J(P_0, C_i)\|$ is not guaranteed to be non-increasing with successive experiments.

3.2.5 Difference in Closed Loop Behaviour

The asymptotic value of cost function in (3.8) can be used to bound the mean squared difference in closed loop performance of P_0 in feedback with any controller C and P_θ in feedback with the same controller C. Given a nominal controller C that stabilises the true

plant P_0, the closed loop transfer function $H(P_0, C)$ is defined by

$$H(P_0, C) := \begin{bmatrix} P_0 \\ I \end{bmatrix} (I - CP_0)^{-1} \begin{bmatrix} -C & I \end{bmatrix} \tag{3.14}$$

It is desired that the difference in the behaviour of closed loops $H(P_0, C)$ and $H(P_\theta, C)$ is small in some sense. The *pointwise* difference in the closed loop performance of nominal plant P_0 and a model P_θ for the same controller C can be bounded from above as [Vin93]

$$\overline{\sigma}(H(P_0, C) - H(P_\theta, C))(e^{j\omega}) \leq \kappa(P_0, P_\theta)(e^{j\omega})\,\overline{\sigma}(H(P_0, C))(e^{j\omega})\,\overline{\sigma}(H(P_\theta, C))(e^{j\omega}) \tag{3.15}$$

where $\kappa(P_0, P_\theta)$ is the pointwise *chordal distance*,

$$\kappa(P_0, P_\theta)(e^{j\omega}) := \overline{\sigma}\left((I + P_0 P_0^*)^{-\frac{1}{2}}(P_0 - P_\theta)(I + P_\theta^* P_\theta)^{-\frac{1}{2}}\right)(e^{j\omega})$$

The upper bound makes sense only if C stabilises both P_0 and P_θ.

Consider prediction error identification as described in section 3.2.2 for a SISO system, with $\phi_{ue} \approx 0$ and a fixed noise filter $\hat{L} = 1$. From (3.10),

$$\hat{\theta}_N \to D_c := \arg\min_\theta \int_{-\pi}^{\pi} \phi_\epsilon \, d\omega$$

where $\phi_\epsilon = |P_0 - P_\theta|^2 \phi_u$. From the definition of $\kappa(P_0, P_\theta)$, clearly
$\kappa^2(P_0, P_\theta)(e^{j\omega})\phi_u(e^{j\omega}) \leq \phi_\epsilon(e^{j\omega})$. Using this fact and (3.15), it follows that

$$\begin{aligned} \overline{\sigma}^2(H(P_0, C) - H(P_\theta, C))(e^{j\omega})\phi_u(e^{j\omega}) &\leq \|H(P_0, C)\|_\infty^2 \, \|H(P_\theta, C)\|_\infty^2 \, \kappa^2(P_0, P_\theta)(e^{j\omega})\phi_u(e^{j\omega}) \\ &\leq \|H(P_0, C)\|_\infty^2 \, \|H(P_\theta, C)\|_\infty^2 \, \phi_\epsilon(e^{j\omega}) \end{aligned}$$

$$\begin{aligned} \text{so that } \frac{1}{2\pi} \int_{-\pi}^{\pi} &\overline{\sigma}^2(H(P_0, C) - H(P_\theta, C))(e^{j\omega})\phi_u(e^{j\omega}) d\omega \\ &\leq \|H(P_0, C)\|_\infty^2 \, \|H(P_\theta, C)\|_\infty^2 \, \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_\epsilon(e^{j\omega}) \, d\omega \end{aligned}$$

Suppose $\phi_u(e^{j\omega}) = k^2$, i.e. u is a white process with intensity k^2. For any controller C with $\|H(P_0, C)\|_\infty$, $\|H(P_\theta, C)\|_\infty$ not too large, the above result shows that a good asymptotic prediction error fit ensures a small *mean squared* difference in closed loop behaviour. We will re-visit (3.15) in chapter 7.

3.3 Quantifying Uncertainty

Besides finding a model of system, one is interested in knowing the extent of uncertainty in the model that a controller designed for the model must cope with. This would require some *a*

priori knowledge - or assumptions - about the system and the measurement noise. Depending on the *a priori* knowledge, there are two ways of obtaining uncertainty information.

If the model $\hat{P}$ is obtained from frequency domain data (i.e. using periodic inputs) using a worst case identification algorithm ([HJN91], [MP92], [GK92]), appropriate, deterministic *a priori* assumptions about the noise and the true system give us worst case error bound α of the form $\|\hat{P} - P_0\|_\infty \leq \alpha$. Usually, the assumptions relate to a lower bound on the rate of decay of impulse response of the true system and an upper bound on the maximum amplitude of noise. More will be said about *a priori* assumptions in deterministic frequency domain identification in later chapters.

On the other hand, with probabilistic assumptions, one gets a point-wise error bound of the form

$$\mathcal{P}(|P(e^{j\omega_i}) - \hat{P}(e^{j\omega_i})| \leq \alpha) > \beta_i \tag{3.16}$$

In [dVvdH95], certain deterministic assumptions about the impulse response of the true system and stochastic assumptions about the measurement noise are used to derive point-wise error bounds (3.16) for a finite number of frequencies, using periodogram averaging technique. The stochastic part of these bounds is valid only as the length of data tends to infinity. In chapter 4, we obtain deterministic as well as non-asymptotic probabilistic point-wise error bounds like (3.16) for bounded white measurement noise.

For n frequency response samples and $\beta_i = \beta$, (3.16) gives

$$\mathcal{P}\left(\bigcup_{i=1}^{n} \left\{|P(e^{j\omega_i}) - \hat{P}(e^{j\omega_i})| \leq \alpha\right\}\right) > \beta^n$$

Clearly, one needs a confidence level β *very* close to unity (typically, $> 99\%$) for these results to be of practical use.

If the error due to under-modelling is ignored, the prediction error method gives a measure of uncertainty in the form of a consistent estimate of covariance matrix $Cov\,(\hat{\theta}_N) := \Lambda_\theta$. [BGS99] shows how to map this into a more tractable uncertainty description in frequency domain. Asymptotically, the scalar $(\hat{\theta}_N - \theta_0)^T \Lambda_\theta^{-1} (\hat{\theta}_N - \theta_0)$ has a chi-squared distribution. From χ^2 tables and for a given confidence level η, one may obtain χ such that

$$\mathcal{P}\left((\hat{\theta}_N - \theta_0)^T \Lambda_\theta^{-1} (\hat{\theta}_N - \theta_0) < \chi^2\right) > 1 - \eta \tag{3.17}$$

In [BGS99], the following problem is solved for transfer function models for a SISO plant: Given a model $\hat{P}_\theta$ corresponding to a parameter vector $\hat{\theta}_N$, a frequency ω and a confidence

ellipsoid parameter χ^2, minimise γ such that

$$(\hat{\theta}_N - \theta_0)^T \Lambda_\theta^{-1} (\hat{\theta}_N - \theta_0) < \chi^2$$
$$\text{and} \quad \kappa(P_0, P_\theta)(e^{j\omega}) \leq \gamma$$

[BGS99] shows that this may be formulated as an LMI optimisation. The worst case chordal distances computed at different frequencies may be used, for example, to design a controller using extended loop shaping procedure [Vin].

3.4 The Frequency Domain Approach

By 'frequency domain identification' we mean the following:

- identification of (possibly noisy) frequency response samples of the true plant and
- obtaining a transfer function that best describes these response samples in a suitable sense.

The frequency response samples can also be obtained by non-parametric methods from time domain data [Lju99] or by using periodic inputs. The later approach typically requires multiple experiments to collect the frequency response samples; hence is more time consuming and potentially expensive as compared to the prediction error approach. However, it has some important advantages:

- Many systems have different small signal and large signal behaviour. If we are interested in large signal behaviour, it is not possible to excite the system dynamics in large signal mode over a band of frequencies, due to actuator constraints. It may be possible, however, to concentrate all the allowable input energy at *one* frequency and identify large signal response of the system at that frequency.
- Spectral noise to signal ratio is a fundamental limitation on the accuracy of time domain identification. An extreme way to improve it is to apply a periodic input signal at a particular frequency. If noise does not contain a periodic component, one expects a very low noise to signal ratio and hence very good non-parametric frequency response estimates.

In the rest of the dissertation, we will be concerned solely with frequency domain identification.

Chapter 4

Point frequency Response Identification

4.1 Introduction

This work suggests a robustly convergent algorithm for worst case identification of point frequency response samples of LSI, stable, discrete time system. A deterministic characterisation of the allowable noise set is introduced, which captures the 'low correlation' property of a white noise realisation.

In what follows, given a real vector $v = [v_0\, v_1\, \ldots\, v_{N-1}]^T$, the circular autocorrelation of v is given by

$$R_v(\tau) = \sum_{t=0}^{N-1} v_{t+\tau}\, v_t^* \tag{4.1}$$

where the index $t + \tau$ is taken as a modulo-N summation.

4.2 Point Frequency Response Identification

4.2.1 Problem Formulation

Assumptions :

- The plant P, for which frequency response point samples are to be estimated is a stable, LSI SISO plant. Further, $P \in \overline{B}\mathcal{H}_{\infty,\rho}(\gamma)$.

- Let $0 < \eta \leq 1$ and $0 < \sigma_N < \infty$ be known constants. The additive noise v corrupting the output measurement belongs to *either* of the following two sets,

$$V_{\sigma_N,\eta,N} = \left\{ v \in \mathbb{R}^N : \frac{1}{N} R_v(0) \leq \sigma_N^2, \max_{\tau \in [1,N-1]} \frac{1}{N} |R_v(\tau)| \leq \eta\, \sigma_N^2 \right\} \tag{4.2}$$

$$V^{\omega}_{\sigma_N,\eta,N} = \left\{ v \in \mathbb{R}^N : \frac{1}{N} R_v(0) \leq \sigma_N^2, \sup_{\omega \in [0,\pi)} \frac{1}{N} \left| \sum_{k=0}^{N-1} v_k\, e^{j\omega k} \right| \leq \eta\, \sigma_N \right\} \tag{4.3}$$

  Here $N \in \mathbb{Z}_+$ is the experiment duration in number of samples and $R_v(\tau)$ is circular correlation defined as in (4.1).

***A posteriori* information** :

A vector of noisy output of the plant for a sinusoid input:

$$Y = \begin{bmatrix} y_0 & y_1 & \dots & y_{N-1} \end{bmatrix}^T$$

where $y_k = \sum_{m=0}^{k} p_m u_{k-m} + v_k$. p_m are the impulse response parameters of P and the sinusoid input $\{u_k\}$ is given by

$$u_k = \alpha \cos \omega_0 k,\ \alpha \in \mathbb{R},\ \omega_0 \in [0, \pi),\ k \geq 0 \tag{4.4}$$

The frequency ω_0 satisfies

$$\omega_0 = \frac{m\pi}{n} \text{ for some integers } \quad m \geq 0,\ n > 0 \tag{4.5}$$

Find :

An algorithm $A_N : \mathbb{R}^N \times [0, \pi) \times \mathbb{R} \mapsto \mathbb{C}$ which maps the experimental data into a point frequency response estimate P_{ω_0} such that the worst case errors defined by

$$e_1(A_N : \omega_0, \gamma, \rho, \eta, \sigma_N, \alpha, N) = \sup_{\substack{P \in \overline{B}\mathcal{H}_{\infty,\rho}(\gamma) \\ v \in V_{\sigma_N,\eta,N}}} |P_{\omega_0} - P(e^{j\omega_0})|$$

$$e_2(A_N : \omega_0, \gamma, \rho, \eta, \sigma_N, \alpha, N) = \sup_{\substack{P \in \overline{B}\mathcal{H}_{\infty,\rho}(\gamma) \\ v \in V^{\omega}_{\sigma_N,\eta,N}}} |P_{\omega_0} - P(e^{j\omega_0})|$$

converge as

$$\begin{aligned} \lim_{\sigma_N \to 0} \lim_{N \to \infty} e_i(A_N : \omega_0, \gamma, \rho, \eta, \sigma_N, \alpha, N) &= 0 \quad i = 1, 2 \\ \lim_{\eta \to 0} \lim_{N \to \infty} e_i(A_N : \omega_0, \gamma, \rho, \eta, \sigma_N, \alpha, N) &= 0 \quad i = 1, 2 \end{aligned} \tag{4.6}$$

Here, $P_{\omega_0} = A_N(Y, \omega_0, \alpha)$. Further, find an upper bound on $e_i(A_N : \omega_0, \gamma, \rho, \eta, \sigma_N, \alpha, N)$ as a function of ω_0, γ, ρ, η, σ_N, α and N.

4.2.2 The Noise Sets

Based on the definition of the noise sets, we list some significant properties of the two noise sets.

- In both the noise sets, η may be interpreted as a bound on a particular, deterministic correlation coefficient. To see this, let $v = \begin{bmatrix} v_0 & v_1 & \dots & v_{N-1} \end{bmatrix}^T \in \mathbb{R}^N$ be a vector such that

$$\frac{1}{N} R_v(0) \leq \sigma_N^2 \tag{4.7}$$

Let λ_τ represent a circular $\tau-$ shift on vector v,

$$(\lambda_\tau)(v) = \begin{bmatrix} v_\tau & v_{\tau+1} & \dots & v_{\tau+N-1} \end{bmatrix}^T$$

where $\tau + k$ for any k is a modulo-N summation. Note that $< v, v >=< \lambda_\tau v, \lambda_\tau v >= R_v(0) \leq N\sigma_N^2$, where $< x, y >:= y^* x$ is the scalar product. Then $v \in \mathbb{R}^N$ satisfying (4.7) is in $V_{\sigma_N,\eta,N}$ if

$$\sup_{\tau \in [0,N-1]} \frac{| < v, \lambda_\tau v > |}{\sqrt{< v, v >< \lambda_\tau v, \lambda_\tau v >}} \leq \eta$$

Next, define a complex frequency vector

$$S_\omega := \begin{bmatrix} 1 & e^{j\omega} & e^{2j\omega} & \dots & e^{(N-1)j\omega} \end{bmatrix}^T$$

Note that $< S_\omega, S_\omega >= N$. Any $v \in \mathbb{R}_N$ satisfying (4.7) is in $V^\omega_{\sigma_N,\eta,N}$ if

$$\sup_{\omega \in [0,\pi)} \frac{| < v, S_\omega > |}{\sqrt{< v, v >< S_\omega, S_\omega >}} \leq \eta$$

- A vector $v \in \mathbb{R}^N$ may be considered as an l_2 signal, allowed to be non-zero only on $[0, N-1]$. From the definition of $R_v(0)$, clearly $R_v(0) = \|v\|_2^2$. Let $\hat{v}(e^{j\omega}) = \sum_{t \in \mathbb{Z}} v(t) e^{j\omega t}$. Using the frequency domain definition of $2-$ norm,

$$\|v\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} \hat{v}^*(e^{j\omega}) \hat{v}(e^{j\omega}) d\omega \leq \sup_{\omega \in [0,\pi)} |\hat{v}(e^{j\omega})|^2 \tag{4.8}$$

$$\text{If } \sup_{\omega \in [0,\pi)} |\hat{v}(e^{j\omega})|^2 = \sup_{\omega \in [0,\pi)} \left| \sum_{k=0}^{N-1} v_k e^{j\omega k} \right|^2 \leq N^2 \eta^2 \sigma_N^2, \tag{4.9}$$

(4.8) implies that

$$R_v(0) = \|v\|_2^2 \leq (N\eta^2) N \sigma_N^2 \tag{4.10}$$

By definition (4.3), any $v \in \mathbb{R}^N$ is in $V^{\omega}_{\sigma_N,\eta,N}$ if and only if it satisfies (4.9) and (4.7). If $\eta \leq \frac{1}{\sqrt{N}}$, (4.9) $\Rightarrow$ (4.7). The interesting case for us is $\frac{1}{\sqrt{N}} \leq \eta \leq 1$, when r.m.s. value of noise vector v and the magnitude of its largest periodic component are independently constrained.

- A ball in l_∞ of size σ_N also includes all stochastic noise signals bounded by σ_N. However, it also includes pathological noise signals which would hardly occur, e.g. a periodic signal of amplitude σ_N. By choosing an appropriate $\eta < 1$, we can exclude these signals from our noise set. Both $V_{\sigma_N,\eta,N}$ and $V^{\omega}_{\sigma_N,\eta,N}$ asymptotically include a 'typical' bounded white noise sequence, as the next result shows.

Theorem 4.1 *For each N, let $v = [\, v_0\, v_1\, \ldots\, v_{N-1}\,]^T$ be a vector of independent, zero mean random variables v_i bounded in $[-K, K]$ and having an identical, finite central moment $E|v_i|^2 = \lambda^2$. Let $0 < \eta \leq 1$ be a constant and let $\sigma : \mathbb{Z}_+ \mapsto \mathbb{R}$ be any bounded, positive valued function such that, for some $N_0 \in \mathbb{Z}_+$,*

$$\sigma_N^2 > \lambda^2 \quad \forall\, N \,>\, N_0 \tag{4.11}$$

1. *Let $V_{\sigma_N,\eta,N}$ be as defined in (4.2). Then*

$$\mathcal{P}(v \in V_{\sigma_N,\eta,N}) \overset{N\to\infty}{\longrightarrow} 1$$

2. *Let $V^{\omega}_{\sigma_N,\eta,N}$ be as defined in (4.3). Then*

$$\mathcal{P}(v \in V^{\omega}_{\sigma_N,\eta,N}) \overset{N\to\infty}{\longrightarrow} 1$$

Proof: See Appendix A.

Certain physical processes (e.g. thermal noise) can be adequately modelled as white noise. Many others, however, are best modelled as either band limited or filtered white noise (e.g. random temperature variations in a chemical reactor may be better modelled as a low frequency disturbance than as a 'white' disturbance). To see how well the proposed noise sets deal with filtered white noise, a simple simulation experiment is performed. Consider a vector of independent, normally distributed random variables, each with zero mean and unit variance,

$$v = \begin{bmatrix} v_0 & v_1 & \ldots & v_N \end{bmatrix}^T$$

with $N = 2000$ and let

$$F_\alpha(z) = \frac{(1-\alpha)z}{1-\alpha z}$$

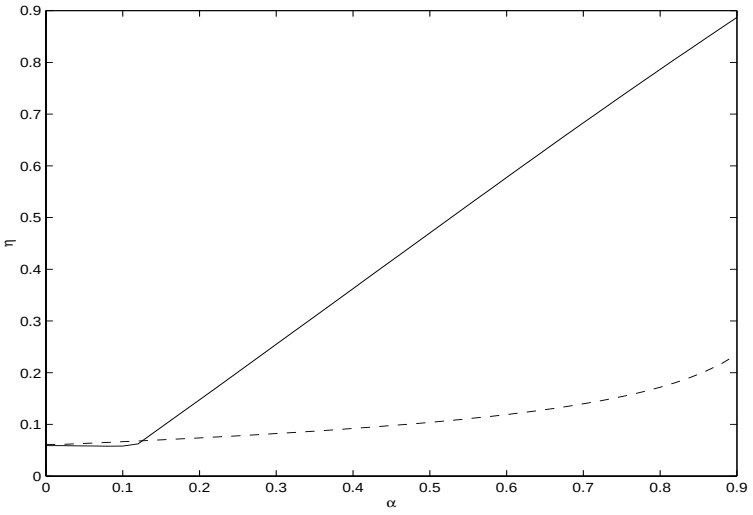


Figure 4.1: Variation of allowable η for filtered white noise

where z is a unit delay. Define

$$
\begin{aligned}
v_\alpha &= F_\alpha v \\
\eta_1(\alpha) &= \max_{\tau\in[1,N]} \frac{|R_{v_\alpha}(\tau)|}{R_{v_\alpha}(0)} \\
\eta_2(\alpha) &= \frac{\sup_{\omega\in[0,\pi)} \frac{1}{N}\left|\sum_{k=0}^{N-1}(v_\alpha)_k\, e^{j\omega k}\right|}{\sqrt{\frac{1}{N}R_{v_\alpha}(0)}}
\end{aligned}
$$

Then $\eta_1(\alpha)$, $\eta_2(\alpha)$ indicate the minimum permissible values of η for $v_\alpha \in V_{\sigma_N,\eta,N}$ and for $v_\alpha \in V^{\omega}_{\sigma_N,\eta,N}$ respectively (with $\sigma_N^2 = \frac{1}{N}R_{v_\alpha}(0)$). Fig. (4.1) shows plots of $\eta_1(\alpha)$ (solid) and $\eta_2(\alpha)$ (dashed) as α is varied from 0 to 0.9, for a typical zero mean, unit variance random vector v generated by *randn* command in MATLAB. It is seen that $V^{\omega}_{\sigma_N,\eta,N}$ still includes v_α for a small value of η, but $V_{\sigma_N,\eta,N}$ requires rather large values of η even for a modest 'colouring' of white noise.

Noise sets with low correlation properties are also investigated in relation to robust $\mathcal{H}_2$ analysis in [Pag95] and in relation to affinely parameterised approximation problems in [VD], [VD97], [TL95] and [FS99].

4.2.3 Identification Algorithm

The estimate P_{ω_0} is given by

$$\begin{aligned} P_{\omega_0} &= \frac{1}{\alpha N}\sum_{k=0}^{N-1} Y_k \qquad \text{if } \omega_0 = 0 \\ &= \frac{2}{\alpha N}\sum_{k=0}^{N-1} Y_k\, e^{j\omega_0 k} \quad \text{otherwise} \end{aligned} \tag{4.12}$$

Remark 4.1 *The implementation of this algorithm suggested is actually identical to the 'classical' method available in literature ([Lju99], ch. 6). The difference is that, here the algorithm is justified purely from a worst case perspective, rather than from a statistical perspective.*

4.2.4 Upper Bounds on Identification Errors

Theorem 4.2 *For the algorithm stated above,*

$$e_1(A_N : \omega_0, \gamma, \rho, \eta, \sigma_N, \alpha, N) \le L(\gamma, \rho, N, \omega_0) + \beta\frac{\sigma_N}{\alpha}\sqrt{\eta + \frac{1-\eta}{N} + \left(1 - \frac{1}{N}\right)\left|\sin\frac{\omega_0 N}{2}\right|} \tag{4.13}$$

$$e_2(A_N : \omega_0, \gamma, \rho, \eta, \sigma_N, \alpha, N) \le L(\gamma, \rho, N, \omega_0) + \beta\,\eta\frac{\sigma_N}{\alpha} \tag{4.14}$$

where

$$L(\gamma, \rho, N, \omega_0) = \frac{\gamma\rho(1-\rho^{-N})}{N(\rho-1)^2} \qquad \textit{if } \omega_0 = 0 \tag{4.15}$$

$$= \frac{\gamma\rho(1-\rho^{-N})}{N(\rho-1)}\left(\frac{1}{\rho-1} + \frac{1}{\sin\omega_0}\right) \qquad \textit{otherwise} \tag{4.16}$$

(4.17)

$$\textit{and} \quad \beta = 1 \qquad \textit{if } \omega_0 = 0$$

$$= 2 \qquad \textit{otherwise}$$

Proof : See Appendix A.

Remark 4.2 *For* a priori *information $v \in V^{\omega}_{\sigma_N,\eta,N}$, robust convergence follows since γ, ρ, η, σ_N, α are arbitrary. For* a priori *information $v \in V_{\sigma_N,\eta,N}$, robust convergence with respect to η requires an additional condition on N:*

$$N = \frac{2\pi}{\omega_0}m, \textit{ for some integer } \quad m \tag{4.18}$$

that is, the data consists of an integer number of cycles of input. This is always possible due to condition (4.5) on ω_0.

The factor $(N \sin \omega_0)$ in $L(\gamma, \rho, N, \omega_0)$ may appear surprising at first. We show here that this factor basically relates to the integer number of sinusoid input cycles completed by data. Let $0 < \omega_0 = \frac{2\pi}{r} < \frac{\pi}{2}$ where r is rational. Let $N = mr + q$, where $q < r$ and m is an integer. Using standard trigonometric inequalities, it can be seen that

$$\begin{aligned} N(\omega_0 - \tfrac{\omega_0^3}{6}) \quad &\leq N \sin \omega_0 \quad \leq N\omega_0 \\ \Leftrightarrow 2\pi(m + \tfrac{q}{r})(1 - \tfrac{2\pi^2}{3r^2}) \quad &\leq N \sin \omega_0 \quad \leq 2\pi(m + \tfrac{q}{r}) \\ \text{Since } \tfrac{q}{r} < 1, \quad N \sin \omega_0 \to \infty &\Longleftrightarrow m \to \infty \end{aligned}$$

Remark 4.3 *If the input were* $u(k) = \alpha \sin \omega_0 k$, $k \geq 0$, *it can be shown that*

$$\begin{aligned} P_{\omega_0} &= \frac{1}{\alpha N} \sum_{k=0}^{N-1} Y_k \qquad \text{if } \omega_0 = 0 \\ &= \frac{-2j}{\alpha N} \sum_{k=0}^{N-1} Y_k \, e^{j\omega_0 k} \quad \text{otherwise} \end{aligned} \tag{4.19}$$

is a robustly convergent algorithm under additional condition (4.18). The development is identical to the one above.

4.3 Identification of Multiple Frequency Response Samples

Instead of applying one sinusoid at a time, a single experiment may be performed by applying the sum of sinusoids of different frequencies and correlating the output with each frequency. Under the same *a priori* assumptions as before, such an experiment also leads to a robustly convergent algorithm.

4.3.1 Problem Formulation

Assumptions : Same as in section 4.2.1. For notational convenience, define a set of $m-$ vectors bounded element-wise in $[0, \pi)$ as:

$$[0, \pi)^m := \{x : x \in \mathbb{R}^m, \, x_i \in [0, \pi) \, \forall \, i \in [1, m]\} \tag{4.20}$$

***A posteriori* information** :

A vector of noisy output of plant for a multiple sinusoid input:

$$Y = \begin{bmatrix} y_0 & y_1 & \dots & y_{N-1} \end{bmatrix}^T$$

where $y_k = \sum_{q=0}^{k} p_q u_{k-q} + v_k$. p_q are impulse response parameters of P. The input u is given by a sum of sinusoids,

$$u_k = \sum_{i=1}^{m} \alpha_i \cos \omega_i k,\ \alpha_i \in \mathbb{R},\ \omega_i \in [0, \pi),\ k \geq 0 \tag{4.21}$$

Define a corresponding vector of (angular) frequencies

$$W := [\omega_1\ \omega_2\ \ldots\ \omega_m]^T$$

and a vector of corresponding amplitudes

$$\alpha := [\alpha_1\ \alpha_2\ \ldots\ \alpha_m]^T$$

Here each ω_i satisfies $\omega_i = \frac{2\pi l_i}{r_i}$ with $r_i > 0$ and $l_i \geq 0$.

Find :

An algorithm $A_N^m : \mathbb{R}^N \times [0, \pi)^m \times \mathbb{R}^m \mapsto \mathbb{C}^m$ which maps the experimental data into a $m-$vector of point frequency response estimates

$$P_\omega = [\, P_{\omega_1},\, P_{\omega_2},\, \ldots,\, P_{\omega_m}]^T,\ P_{\omega_i} := (A_N^m(Y, W, \alpha))\,(i)$$

such that the worst case errors defined by

$$e_1(A_N^m : W, \gamma, \rho, \eta, \sigma_N, \alpha, N) = \sup_{i \in [1,m]} \sup_{\substack{P \in \overline{B}\mathcal{H}_{\infty,\rho(\gamma)} \\ v \in V_{\sigma_N,\eta,N}}} |P_{\omega_i} - P(e^{j\omega_i})|$$

$$e_2(A_N^m : W, \gamma, \rho, \eta, \sigma_N, \alpha, N) = \sup_{i \in [1,m]} \sup_{\substack{P \in \overline{B}\mathcal{H}_{\infty,\rho(\gamma)} \\ v \in V^{\omega}_{\sigma_N,\eta,N}}} |P_{\omega_i} - P(e^{j\omega_i})|$$

converge as

$$\begin{aligned} \lim_{\sigma_N \to 0} \lim_{N \to \infty} e_p(A_N^m : W, \gamma, \rho, \eta, \sigma_N, \alpha, N) &= 0 \quad p = 1, 2 \\ \lim_{\eta \to 0} \lim_{N \to \infty} e_p(A_N^m : W, \gamma, \rho, \eta, \sigma_N, \alpha, N) &= 0 \quad p = 1, 2 \end{aligned} \tag{4.22}$$

Further, find an upper bound on $e_p(A_N^m : W, \gamma, \rho, \eta, \sigma_N, \alpha, N)$ as a function of γ, ρ, η, σ_N, α, N and W.

4.3.2 Identification Algorithm

The estimates P_{ω_i} are given by

$$\begin{aligned} P_{\omega_i} &= \frac{1}{\alpha_i N} \sum_{0}^{N-1} Y_k \quad \text{if } \omega_i = 0 \\ &= \frac{2}{\alpha_i N} \sum_{0}^{N-1} Y_k e^{j\omega_i k} \quad \text{otherwise} \end{aligned} \tag{4.23}$$

4.3.3 Bound on Worst Case Error

Theorem 4.3 *For the algorithm stated above,*

$$e_1(A_N^m : W, \gamma, \rho, \eta, \sigma_N, \alpha, N) \leq L(\gamma, \rho, N, \alpha, W) + \frac{2\sigma_N}{\alpha_{min}} \sqrt{\eta + \frac{1-\eta}{N} + \left(1 - \frac{1}{N}\right) \max_{i \in [1,m]} \left| \sin \frac{\omega_i N}{2} \right|} \tag{4.24}$$

$$e_2(A_N^m : W, \gamma, \rho, \eta, \sigma_N, \alpha, N) \leq L(\gamma, \rho, N, \alpha, W) + 2\frac{\eta \sigma_N}{\alpha_{min}} \tag{4.25}$$

$$\textit{Here} \quad L(\gamma, \rho, N, \alpha, W) = \frac{\gamma\rho(1-\rho^{-N})}{N(\rho-1)} \left(\frac{1}{\rho - 1} + (m-1)\psi_\alpha \phi_W \right)$$

where $\alpha_{min} = \min_{i \in [1,m]} \alpha_i$ *and* $\psi : \mathbb{R}^m \mapsto \mathbb{R}$, $\phi : \mathbb{R}^m(\pi) \mapsto \mathbb{R}$ *are functions defined by*

$$\psi_\alpha := \max_{i,j \in [1,m]} \frac{\alpha_i}{\alpha_j} \tag{4.26}$$

$$\phi_W := \max_{\substack{i,j \in [1,m] \\ \omega_i \geq \omega_j}} \left(\frac{1}{sin\left(\frac{\omega_i + \omega_j}{2}\right)} + \frac{1}{sin\left(\frac{\omega_i - \omega_j}{2}\right)} \right) \tag{4.27}$$

Proof : See Appendix A.

Intuitively, the task of separating information about the system response at different frequencies from a single experiment seems a difficult task if the frequencies are closely spaced. The definition of ϕ_W captures this fact; the error bound deteriorates as the difference $\max\{\omega_i - \omega_j, (\pi - \omega_i) - \omega_j\}$ becomes smaller.

4.4 Probabilistic Bounds in Bounded White noise

Instead of using deterministic noise sets, one may also use probabilistic noise description, along-with deterministic assumptions on the true plant. This leads to probabilistic bounds, or confidence intervals. The following result gives one such confidence interval:

Lemma 4.1 *Suppose the true plant* $P \in \overline{B}\mathcal{H}_{\infty,\rho}(\gamma)$ *and let the measurement noise be zero mean, white stochastic process bounded in* $[-K, K]$. *Given the same a posteriori information as in section 4.2.1 and given any confidence level* $0 < \zeta < 1$, *the estimate* P_{ω_0} *as defined in (4.12) satisfies*

$$\mathcal{P}\left(|P_{\omega_0} - P(e^{j\omega_0})| \leq L(\gamma, \rho, N, \omega_0) + \frac{4K}{\alpha\sqrt{N}} \left\{ \log\left(\frac{4}{1-\zeta}\right) \right\}^{\frac{1}{2}} \right) > \zeta \tag{4.28}$$

Here $L(\gamma, \rho, N, \omega_0)$ *is as defined in the equation (4.15).*

Proof : See Appendix A.

If $P_{\omega_0} := A_N(Y, \omega_0, \alpha)$ is considered as a sequence of random variables indexed by N, it follows from the proof of (4.28) and from the definition of $L(\gamma, \rho, N, \omega_0)$ that

$$\mathcal{P}\left(|P_{\omega_0} - P(e^{j\omega_0})| > \epsilon\right) \overset{N\to\infty}{\longrightarrow} 0, \ \forall\, \epsilon$$

In other words, P_{ω_0} tends to $P(e^{j\omega_0})$ *in probability* ([Pap91], chapter 8). Besides this nice convergence result and its non-asymptotic nature, (4.28) is still somewhat pessimistic from a practical point of view. To see this, suppose, we want a confidence level $\zeta = 0.99$ and want to restrict noise term in the error bound to 0.1. Further, suppose $\frac{K}{\alpha} = 1$. The constraint $\frac{4}{\sqrt{N}} \left\{\log\left(\frac{4}{1-0.99}\right)\right\}^{\frac{1}{2}} \leq 0.1$ yields $N \geq 9587$ as a lower bound on the number of samples required to achieve this. Experiment durations as long as this may be unacceptable in many practical cases.

Chapter 5

New Untuned Algorithms for Worst Case Identification

5.1 Introduction

Worst case identification has attracted a lot of attention since its definitive formulation in [HJN91]. There are two objectives in this problem formulation -

- To map the experimental frequency response data into a stable plant transfer function via a 'convergent' algorithm. An algorithm is said to be convergent if the worst case identification error (with respect to noise and the unknown system) tends to zero as the number of data points tends to infinity and the (deterministic) noise tends to zero.
- To provide bounds on the worst case identification error.

This problem can be approached in two different ways, leading to *untuned* and *tuned* algorithms. Both approaches assume that the true, unknown plant belongs to a certain subset Ψ of the set of stable systems and that the additive noise corrupting the measurement belongs to a certain norm bounded set Ξ. Suitable choices of Ψ and Ξ enable us to derive bounds on the worst case identification error.

The identification algorithm is said to be untuned if it is independent of the definitions of Ψ and Ξ; and is said to be tuned otherwise. Examples of algorithms of the the former type may be found in [HJN91], [GK92], [M$\ddot{\text{o}}$91] and [MP92]; while the later type is investigated in [CNF95], [GLZ96]. An extension of the tuned algorithms to the multivariable case is studied in [CFN96]. Tuned algorithms may give a better result if the plant and the noise

conform to the assumptions about them, but may fail to give any sensible result at all if the assumptions are wrong.

Conventionally, the worst case identification problem is formulated as follows.

Assumptions :

- The plant $P(z)$ whose transfer function is to be identified is a stable, LSI, SISO discrete time system. Further, $P(z) \in \overline{B}\mathcal{H}_{\infty,\rho_1}(\gamma_1)$.
- Noise corrupting measurement belongs to $\overline{B}l_\infty(\epsilon)$.

Here, the noise bound $\epsilon \in [0, \infty)$, the lower bound on stability margin $\rho_1 \in (1, \infty)$ and the upper bound on gain $\gamma_1 \in [0, \infty)$ are assumed to be known constants.

***A posteriori* Information :** A vector of uniformly spaced noisy frequency response samples

$$P_\omega = \begin{bmatrix} P_{\omega_1} & P_{\omega_2} & \dots & P_{\omega_m} \end{bmatrix}^T, \quad P_{\omega_i} = P(e^{j\omega_i}) + v_i$$

where $\omega_i = \frac{2\pi(i-1)}{m}$, $i = 1, 2, \dots, m$.

Find : An algorithm $A_m : \mathbb{C}^m \mapsto \mathcal{H}_\infty$ such that the worst case error defined by

$$e(A_m : \rho_1, \gamma_1, \epsilon, m) = \sup_{\substack{P \in \overline{B}\mathcal{H}_{\infty,\rho_1}(\gamma_1) \\ v \in \overline{B}l_\infty(\epsilon)}} \|A_m(P_\omega) - P\|_\infty$$

converges as

$$\lim_{m\to\infty, \epsilon\to 0} e(A_m : \rho_1, \gamma_1, \epsilon, m) = 0 \tag{5.1}$$

In addition, derive explicit bounds on the above error. [1]

Existing untuned algorithms rely on a two step procedure:

- Let
$$\tilde{P}_m(k) = \frac{1}{m} \sum_{i=0}^{m-1} P_{\omega_i} e^{-jk(\frac{2i\pi}{m})} \; k = 1, 2, \dots, m$$
be the inverse Discrete Fourier Transform Coefficients. Define an $\mathcal{L}_\infty$ approximation by
$$\hat{P}_1 = \sum_{k=-m}^{m} w_{m,k} \tilde{P}_m(k) z^k$$
Here $w_{m,k}$ is a smoothing function.

[1] Note that $A_m(P_\omega)$ depends on the true plant P and noise v through the relation $P_{\omega_i} = P(e^{j\omega_i}) + v_i$.

- Solve Nehari's problem to determine a stable plant nearest to $\hat{P}_1$:

$$\hat{P} = \arg \min_{f \in \mathcal{H}_\infty} \|f - \hat{P}_1\|_\infty$$

Different untuned algorithms differ in the smoothing or window functions $w_{m,k}$ used in the first step. An example of smoothing function is the triangular window

$$w_{m,k} = \begin{array}{ll} 1 - \frac{|k|}{m} & -m \leq k \leq m \\ 0 & \text{otherwise} \end{array}$$

The Nehari approximation step makes the overall scheme nonlinear in *a posteriori* data. The convergence property of an untuned algorithm is said to be *robust* if (5.1) holds for every triplet $(\rho_1, \gamma_1, \epsilon) \in [1, \infty) \times [0, \infty) \times [0, \infty)$. Partington [Par92] has shown that it is impossible to construct a linear, robustly convergent algorithm. Since $\|\hat{P} - P\|_\infty \leq \|\hat{P}_1 - P\|_\infty + \|\hat{P} - \hat{P}_1\|_\infty \leq 2\|\hat{P}_1 - P\|_\infty$, robust convergence can be ensured by a proper choice of window function in the first step. [GK92] gives sufficient conditions on a window function for the algorithm to be robustly convergent.

Existing untuned, convergent algorithms seem to suffer from following disadvantages -

- The class of transfer function models to which the identified model will belong can not be selected *a priori.*

- Uniform frequency spacing is necessary. If nonuniform spacing is used, it needs to be interpolated into uniformly spaced samples to use existing results [AGK92].

- It is not clear how to obtain the *a priori* information (γ, ρ) necessary to compute the worst case error bound. Moreover, this bound may be conservative, and may not give a true indication of how good - or how poor - the model is. Provided the Nyquist plot of the model is sufficiently smooth and under reasonable (qualitative) assumptions that

  – the Nyquist plot of the true plant is also smooth and

  – the measurement noise during point frequency response identification does not have a periodic component,

  the worst case fit $\max_i |P_{\omega_i} - A_m(P_\omega)(i)|$ over a sufficiently dense grid of measurement frequencies $\{\omega_i\}$ is, at least, a good indication of the 'true' distance between the plant and the model, rather than the bound on the worst case error. However, most of the existing algorithms do not offer any direct control over the worst case fit $\max_i |P_{\omega_i} - A_m(P_\omega)(i)|$.

A notable exception which overcomes most of the above anomalies is the finite impulse response approximation algorithm by Mäkilä and Partington [MP92]. More will be said about the algorithm in [MP92] in chapter 6. Here, an alternative untuned algorithm to obtain rational models from a specific subset of $\mathcal{H}_\infty$ is suggested. It will be motivated by illustrating the shortcomings of direct interpolation of the given frequency response samples using Pick's theorem.

5.2 Identification using Pick interpolation

First we state Pick's interpolation theorem [KT85].

Pick's Theorem : Let $a_i \in \mathbb{D}$ and $b_i \in \overline{\mathbb{D}},\ i = 1, 2, \ldots, n$ be two sets of complex numbers. $\exists$ an analytic function

$$\hat{f} \in \mathcal{H}_\infty,\ \hat{f} : \mathbb{D} \mapsto \overline{\mathbb{D}} \tag{5.2}$$

such that

$$\hat{f}(a_i) = b_i,\ i = 1, 2, \ldots, m \tag{5.3}$$

if and only if the Pick matrix defined by

$$P = \left[\frac{1 - b_i b_j^*}{1 - a_i a_j^*}\right],\ i, j = 1, 2, \ldots, m$$

is nonnegative definite. Further, if the Pick matrix is nonnegative definite, $\exists$ a rational function $\hat{f}$ of degree at most m which satisfies conditions (5.2) and (5.3).

To apply Pick's theorem to our present problem, the following corollary is used:

Corollary 5.1 *Let P_{ω_i} and ω_i $i = 1, 2, \ldots, m$ be two sets of complex numbers with $\omega_i \in \partial\mathbb{D}$. Let $\rho_2 \in \mathbb{R},\ \rho_2 > 1$ be a given constant. Then there exists a function $\hat{P}(z)$ analytic and bounded in $\mathbb{D}_{\rho_2}$ such that $\hat{P}(e^{j\omega_i}) = P_{\omega_i}$.*

Proof : If $P_{\omega_i} = 0\ \forall\ i \in [1, m]$, then $\hat{P}(z) = 0$ is the required function. Suppose instead, at least one P_{ω_i} is nonzero. Let $z_i = \frac{e^{j\omega_i}}{\rho_2}$. Consider matrix

$$\hat{Q} = E - \alpha^2 F,\ \alpha \in \mathbb{C} \tag{5.4}$$

where

$$E = \left[\frac{1}{1 - z_i z_j^*}\right], \quad F = \left[\frac{P_{\omega_i} P_{\omega_j}^*}{1 - z_i z_j^*}\right]\ i, j = 1, 2, \ldots, m$$

Let λ_{max} be the largest eigen value of $E^{-1}F$. As shown in [KT85], for any $\alpha \leq \alpha_{max} = \frac{1}{\sqrt{\lambda_{max}}}$, there exists a rational analytic function $\hat{P}_\alpha : \mathbb{D} \mapsto \overline{\mathbb{D}}$ such that $\hat{P}_\alpha(z_i) = \alpha P_{\omega_i}$. A solution to the original interpolation problem is then given by

$$\hat{P}(z) = \frac{1}{\alpha}\hat{P}_\alpha(\frac{z}{\rho_2}) \in \mathcal{H}_\infty \qquad \blacksquare \tag{5.5}$$

Implementation details of Pick interpolation procedure may be found in ([DFT92], chapter 9). As $m \to \infty$, however, α_{max} may become arbitrarily small, making $\|\hat{P}\|_{\infty,\rho_2}$ arbitrarily large. It seems a better way is to find points $q_i \in \mathbb{C}$ such that

- $|q_i - P_{\omega_i}|$ is 'small';
- An analytic function belonging to a *norm bounded* subset of $\mathcal{H}_\infty$ interpolates the q_i's.

This idea motivates an alternate problem formulation for worst case identification.

5.3 Alternate Problem Formulation

Assumptions : Identical to those in section 5.1.

A posteriori Information : A vector of (not necessarily uniformly spaced) noisy frequency response samples

$$P_\omega = [P_{\omega_1}\ P_{\omega_2}\ \ldots\ P_{\omega_m}]^T,\ P_{\omega_i} = P(e^{j\omega_i}) + v_i \tag{5.6}$$

where $\omega_i \in [0, \pi)$, $i = 1, 2, \ldots, m$.

Also given is a corresponding vector of (angular) frequencies

$$W = [\omega_1\, \omega_2\ \ldots\ \omega_m]^T \tag{5.7}$$

$\overline{\delta} = \max_i |\omega_{i+1} - \omega_i|$ is the maximum separation between adjacent angular frequencies. Let $[0, \pi)^m$ be as defined in (4.20).

Given user chosen parameters $\rho_2 > 1$, $\gamma_2 > 0$,

Find an algorithm $A_{\overline{\delta}_1} : C^m \times [0, \pi)^m \mapsto \overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)$ such that

$$\max_i |\hat{P}(e^{j\omega_i}) - P_{\omega_i}|$$

is minimised, where $\hat{P}(z) = A_{\overline{\delta}_1}(P_\omega, W)$. Further, find an upper bound on the worst case error defined by

$$e\,(A_{\overline{\delta}_1} : \rho_1, \gamma_1, \rho_2, \gamma_2, \epsilon, \overline{\delta}) := \sup_{P \in \overline{B}\mathcal{H}_{\infty,\rho_1}(\gamma_1),\, v \in \overline{B}l_\infty(\epsilon)} \| A_{\overline{\delta}_1}(P_\omega, W) - P \|_\infty \tag{5.8}$$

Remark 5.1 *Actually, it is possible - and indeed, more meaningful - to use noise sets defined in chapter 4 in place of* $\overline{B}l_\infty(\epsilon)$*. This will complicate notation, however, as the error definition and convergence will involve both the length of each experiment and the number of experiments (i.e. the number of frequency response samples). Hence we keep using* $\overline{B}l_\infty(\epsilon)$ *as the noise set; keeping in mind that* ϵ *may actually be replaced by (4.13) or (4.14).*

5.3.1 New Identification Algorithm

As stated above, our aim is to solve

$$\min_{f\in\overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)} \max_i |f(e^{j\omega_i}) - P_{\omega_i}| \tag{5.9}$$

The steps to solve the above problem are enumerated below.

1. Find $\hat{x}, \hat{w}_1, \hat{w}_2, \dots, \hat{w}_n$ which solve the following minimisation problem

$$\min_{x\in\mathbb{R},\, w_1,w_2,\dots,w_m\in\mathbb{C}} \quad \lambda \tag{5.10}$$

subject to LMI constraints

$$\text{diag}\left(\begin{bmatrix} \lambda x & w_i^* \\ w_i & \lambda x \end{bmatrix}\right) \geq 0 \tag{5.11}$$

$$\text{diag}\left(\begin{bmatrix} 1 & xP_{\omega_i}^* + w_i^* \\ xP_{\omega_i} + w_i & 1 \end{bmatrix}\right) \geq 0 \tag{5.12}$$

$$\begin{bmatrix} E^{-1} & \text{diag}\,(xP_{\omega_i}^* + w_i^*) \\ \text{diag}(xP_{\omega_i} + w_i) & E \end{bmatrix} \geq 0 \tag{5.13}$$

and

$$x \geq \gamma_2^{-1} \tag{5.14}$$

Here

$$E = \left[\frac{1}{1 - \frac{e^{j(\omega_i-\omega_j)}}{\rho_2^2}}\right] \quad i,j = 1,2,\dots,n$$

and diag (x_i) denotes a diagonal matrix with $x_i, i = 1,2,\dots,m$ along its diagonal.

2. Find a rational analytic function $f : \mathbb{D} \mapsto \overline{\mathbb{D}}$ of degree $\leq 2m$ such that

$$f\left(\frac{e^{j\omega_i}}{\rho_2}\right) = \hat{x}P_{\omega_i} + \hat{w}_i,\ f\left(\frac{e^{-j\omega_i}}{\rho_2}\right) = \hat{x}P^*_{\omega_i} + \hat{w}^*_i,\ i = 1, 2, \ldots, m \tag{5.15}$$

As will be seen in the proof of theorem 5.1, the existence of the interpolant f is ensured by (5.11) - (5.14). Complex conjugate interpolation data is required to obtain a real rational transfer function. If $f(z)$ which satisfies (5.15) is such that $f(z) = f_1(z)+jf_2(z)$ where f_1, f_2 are real rational functions, $f_1(z)$ also obeys $f_1\left(\frac{e^{j\omega_i}}{\rho_2}\right) = \hat{x}P_{\omega_i} + \hat{w}_i,\ i = 1, 2, \ldots, m$ [DFT92].

3. The identified transfer function model is then given by

$$\hat{P}(z) = \hat{x}^{-1} f_1(\frac{z}{\rho_2}) \tag{5.16}$$

Note that solving the above problem requires no assumptions as to how (ρ_1, γ_1) and (ρ_2, γ_2) are related, i. e. the *a priori* information is *not* in-built in the algorithm. The user is free to choose (ρ_2, γ_2) so as to obtain a desired compromise between a good worst case error fit and an acceptable smoothness for $\hat{P}(z)$ (since $\|\hat{P}'\|_\infty \leq \frac{\gamma_2}{\rho_2 - 1}$).

We now prove that the above procedure indeed gives the required transfer function.

Theorem 5.1 *When $\hat{P}(z)$ is obtained as outlined above,*

$$\max_{i\in[1,m]} |\hat{P}(e^{j\omega_i}) - P_{\omega_i}| = \min_{f \in \overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)} \max_{i\in[1,m]} |f(e^{j\omega_i}) - P_{\omega_i}| \tag{5.17}$$

Proof : Let $\tilde{P} \in \overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)$ such that

$$\sup_{z\,\in\,\mathbb{D}_{\rho_2}} |\tilde{P}(z)| \leq x^{-1} \tag{5.18}$$

for some $x \in \mathbb{R}$.

$$\text{Let} \quad \tilde{P}(e^{j\omega_i}) = P_{\omega_i} + \tilde{w}_i,\ i = 1,\ 2,\ \ldots,\ m \tag{5.19}$$

Then we seek to minimise $\max_i |\tilde{w}_i|$.

$$\text{Define} \quad \hat{f}(z) = x\tilde{P}(\rho_2 z) \tag{5.20}$$

and $w_i = x\,\tilde{w}_i$. Then equations (5.20), (5.18) $\Leftrightarrow \sup_{z\in\mathbb{D}} |\hat{f}(z)| \leq 1$. Also, $\tilde{P}$ analytic in $\mathbb{D}_{\rho_2} \Leftrightarrow \hat{f}(z)$ is analytic in $\mathbb{D}$.

Multiplying both sides of (5.19) by x and using (5.18),

$$xP_{\omega_i} + w_i \in \overline{\mathbb{D}} \tag{5.21}$$

Also, using definition of $\hat{f}(z)$,

$$\hat{f}\left(\frac{e^{j\omega_i}}{\rho_2}\right) = xP_{\omega_i} + w_i \tag{5.22}$$

It follows that the Pick matrix corresponding to the interpolation pairs $\left(\frac{e^{j\omega_i}}{\rho_2},\ xP_{\omega_i} + w_i\right)$ must be nonnegative definite, i. e.

$$Q = E - \text{diag}(xP_{\omega_i} + w_i)\ E\ \text{diag}(xP^*_{\omega_i} + w^*_i) \geq 0 \tag{5.23}$$

Inequality (5.12) is equivalent to (5.21), which can be verified using (2.12). Inequality (5.13) is equivalent to (5.23). Inequality (5.14) follows from (5.18) and from the fact that $\tilde{P} \in \overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)$. Finally,

$$\max_{i\in[1,m]} |\tilde{w}_i| = \max_{i\in[1,m]} x^{-1}|w_i|$$

Since $x > 0$, the right hand side can be minimised by solving

$$\min \lambda \text{ subject to } \max_{i\in[1,m]} (x^2\lambda^2 - |w_i|^2) \geq 0 \tag{5.24}$$

This is achieved by minimising λ subject to (5.11), the equivalence again follows from (2.12).

■

Thus we have formulated our identification problem as a convex optimisation problem. For any allowable choice of $(\gamma_2,\ \rho_2)$, consider a choice of decision variables $x = \gamma_2^{-1} + q,\ q \in (0, \infty), \quad w_i = -xP_{\omega_i}\ \forall\, i \in [1, m]$. This corresponds to a solution $\hat{P}(z) = 0$.
We deduce two important properties of our problem formulation from this solution:

- The above solution is feasible for any choice of pair $(\rho_2,\ \gamma_2)$ such that $\rho_2 \in (1,\ \infty),\ \gamma_2 \in (0,\ \infty)$. Thus the optimisation problem has a nonempty feasible solution set for all allowable $\gamma_2,\ \rho_2$.

- This solution remains feasible if the inequalities in (5.11) - (5.14) are replaced by *strict* inequalities, i.e. the problem is *strictly feasible* [BGFB94].[2] For a strictly feasible problem, the infimum obtained by solving the problem with non-strict inequalities is same as the one obtained by solving the problem with strict inequalities [BGFB94]. The strict inequalities problem is tractable by easily available and efficient commercial software, without the added complexity of reducing non-strict LMIs to strict LMIs.

[2]The only nontrivial step in proving this is to prove that E is positive definite. This is proved using Pick's theorem in [CNF95].

5.3.2 Worst Case Error Bound

We use a result from [HJN91] which gives a bound on infinity norm on the worst case slope of an exponentially stable system:

Fact 5.1 *Let* $f \in \overline{B}\mathcal{H}_{\infty,\rho}(\gamma)$. *Then*

$$\| f^{(k)}\|_\infty \leq \frac{\gamma k!}{(\rho-1)^k}$$

where $f^{(k)} = \frac{d^k f(z)}{dz^k}$.

Theorem 5.2 *Let* $\hat{P}(z) = A_{\overline{\delta}1}(P_\omega, W)$. *Then*

$$e(A_{\overline{\delta}1} : \rho_1, \gamma_1, \rho_2, \gamma_2, \epsilon, \overline{\delta}) \leq \left(\frac{\gamma_1}{\rho_1 - 1} + \frac{\gamma_2}{\rho_2 - 1}\right)\frac{\overline{\delta}}{2} + \lambda + \epsilon \tag{5.25}$$

$$\textit{where} \quad \lambda = \max_{i\in[1,m]} |\hat{P}(e^{j\omega_i}) - P_{\omega_i}|$$

Proof : Consider any point $\omega^* \in [\omega_i, \omega_{i+1})$. Applying triangle inequality to the error at this frequency

$$\begin{aligned} |\hat{P}(e^{j\omega^*}) - P(e^{j\omega^*})| \;&\leq\; |\hat{P}(e^{j\omega^*}) - \hat{P}(e^{j\omega_i})| + |\hat{P}(e^{j\omega_i}) - P_{\omega_i}| \\ &\quad + |P_{\omega_i} - P(e^{j\omega_i})| + |P(e^{j\omega_i}) - P(e^{j\omega^*})| \\ &\leq\; \|\hat{P}'\|_\infty \frac{\overline{\delta}}{2} \;+\; \lambda \;+\; \epsilon \;+\; \|P'\|_\infty \frac{\overline{\delta}}{2} \end{aligned} \tag{5.26}$$

where $f'(z) := \frac{df}{dz}$. From fact 5.1,

$$\|P'\|_\infty \leq \frac{\|P\|_{\infty,\rho_1}}{\rho_1 - 1} \leq \frac{\gamma_1}{\rho_1 - 1}$$

Similarly, $\|\hat{P}'\|_\infty \leq \frac{\gamma_2}{\rho_2 - 1}$. Substituting these values in (5.26) gives the result. ∎

5.3.3 Asymptotic Behaviour of Worst Case Error

Let $\underline{\delta} = \min_i |w_{i+1} - w_i|$ be the minimum angular frequency separation. Define $e_1^{inf} := \lim_{\overline{\delta}\to 0} \lim_{\epsilon\to 0} e(A_{\overline{\delta}1} : \rho_1, \gamma_1, \rho_2, \gamma_2, \epsilon, \overline{\delta})$ where $\lim_{\overline{\delta}\to 0}$ is taken such that the ratio $\frac{\overline{\delta}}{\underline{\delta}}$ is always bounded.

Lemma 5.1

$$\begin{aligned} e_1^{inf} \;&\leq\; \gamma_1 \textit{ always} && (5.27) \\ &=\; 0 \textit{ if } \overline{B}\mathcal{H}_{\infty,\rho_1}(\gamma_1) \subseteq \overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2) && (5.28) \end{aligned}$$

Proof : First, let $\overline{B}\mathcal{H}_{\infty,\rho_1}(\gamma_1) \subseteq \overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)$. Since $P \in \overline{B}\mathcal{H}_{\infty,\rho_1}(\gamma_1) \subseteq \overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)$,

$$\lambda = \min_{F \in \overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)} \max_i |F(e^{j\omega_i}) - P_{\omega_i}| \leq \max_i |P(e^{j\omega_i}) - P_{\omega_i}| \leq \epsilon \tag{5.29}$$

Substituting in (5.25) and taking appropriate limits leads to (5.28). If $\overline{B}\mathcal{H}_{\infty,\rho_1}(\gamma_1) \not\subseteq \overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)$, then the trivial solution $\hat{P}(z) = 0$ gives

$$e_1^{inf} \leq \lim_{\overline{\delta}\to 0} \lim_{\epsilon \to 0} \max_i |P_{\omega_i}| \leq \gamma_1 \qquad \blacksquare$$

Suppose $\overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)$ reflects our knowledge about the true plant set $\overline{B}\mathcal{H}_{\infty,\rho_1}(\gamma_1)$. Then the algorithm converges if our *a priori* information is correct. This resembles convergence of tuned algorithms. Even if the *a priori* information is incorrect, however, the proposed algorithm retains its bounded error property.

5.4 Achieving Robust Convergence

A part of the bound on the worst case error is fixed by *a priori* information, γ_1, ρ_1 and ϵ. The remaining part depending on *a posteriori* information is

$$e_r(A_{\overline{\delta}1} : \rho_2, \gamma_2, \overline{\delta}) = \|\hat{P}'\|_{\infty,\rho_2} \frac{\overline{\delta}}{2} + \lambda \tag{5.30}$$

where $\lambda = \max_i |\hat{P}(e^{j\omega_i}) - P_{\omega_i}|$, $\hat{P}(z) = A_{\overline{\delta}1}(P_\omega, W)$ and $\|\hat{P}'\|_\infty \leq \left(\frac{\|\hat{P}\|_{\infty,\rho_2}}{\rho_2 - 1}\right)$ from [HJN91].
Instead of minimising worst case absolute distance, consider the following problem:
Given the same *a priori* and *a posteriori* information as before and a user chosen lower bound on stability margin ρ_2,

Find an algorithm $A_{\overline{\delta}2} : C^m \times [0,\pi)^m \mapsto \mathcal{H}_{\infty,\rho_2}$ such that

$$\left(\frac{\|\hat{P}\|_{\infty,\rho_2}}{\rho_2 - 1} \frac{\overline{\delta}}{2} + \max_i |\hat{P}(e^{j\omega_i}) - P_{\omega_i}| \right) \tag{5.31}$$

is minimised, where $\hat{P}(z) = A_{\overline{\delta}2}(P_\omega, W)$. The procedure to solve the above problem is given below.

1. Solve

$$\min_{x \in \mathbb{R},\, w_1, w_2, \ldots, w_n \in \mathbb{C}} \quad \lambda \tag{5.32}$$

subject to LMI constraints

$$\text{diag}\left(\begin{bmatrix} \lambda x - k & w_i^* \\ w_i & \lambda x - k \end{bmatrix} \right) \geq 0 \tag{5.33}$$

and constraints given by (5.12), (5.13). Here $k = \left(\frac{1}{\rho_2 - 1}\right) \frac{\overline{\delta}}{2}$.

Steps 2 and 3 are identical to those in in section 5.3.1.

By using arguments nearly identical to those used in case of $A_{\overline{\delta}1}$, it can be shown that the above procedure solves (5.31). Further, define the worst case error as before

$$e_2\left(A_{\overline{\delta}2} : \rho_1, \gamma_1, \rho_2, \epsilon, \overline{\delta}\right) := \sup_{P \in \overline{B}\mathcal{H}_{\infty,\rho_1}(\gamma_1),\, v \in \overline{B}l_\infty(\epsilon)} \| A_{\overline{\delta}2}(P_\omega, W) - P \|_\infty \tag{5.34}$$

with the asymptotic value defined as $e_2^{inf} := \lim_{\overline{\delta}\to 0} \lim_{\epsilon\to 0} e(A_{\overline{\delta}2} : \rho_1, \gamma_1, \rho_2, \epsilon, \overline{\delta})$. We have the following result:

Lemma 5.2

$$\begin{aligned} e_2^{inf} &\leq \gamma_1 \textit{ always} && (5.35)\\ &= 0 \textit{ if } \rho_2 \leq \rho_1 && (5.36) \end{aligned}$$

Proof: similar to the proof of lemma 5.1.

Since $\frac{\|\hat{P}\|_{\infty,\rho_2}}{\rho_2 - 1}$ is an upper bound on $\|\hat{P}'\|_\infty$, the above lemma introduces the idea of trade-off between the worst case fit achieved and the worst case slope of approximation (which controls the behaviour of the model in-between the measurement samples). This idea will be one of the central themes in chapters 6 and 7.

Note that lemma 5.2 is a stronger statement of convergence than lemma 5.1. A natural extension is to go a step further and pose a problem:
Given the same *a priori* and *a posteriori* information as before,
Find an algorithm $A_{\overline{\delta}3} : C^m \times [0,\pi)^m \mapsto \mathcal{H}_\infty$ such that $A_{\overline{\delta}3}$ solves

$$\min_{\rho\in(1,\infty)} \min_{\hat{P}\in H_{\infty,\rho}} \left(\frac{\|\hat{P}\|_{\infty,\rho}}{\rho - 1}\frac{\overline{\delta}}{2} + \max_i |\hat{P}(e^{j\omega_i}) - P_{\omega_i}| \right) \tag{5.37}$$

where $\hat{P}(z) = A_{\overline{\delta}3}(P_\omega, W)$.

Lemma 5.3 *$A_{\overline{\delta}3}$ is robustly convergent algorithm.*

Proof: Defining the worst case error and its asymptotic value as before, the above result follows from lemma 5.2.

The outermost minimisation is non-convex in ρ. An approximate solution may be sought for finite data size by solving (5.31) for different values of ρ.

5.5 Simulation Example

We illustrate the use of the first algorithm proposed here by identifying a discrete transfer function $P(z) = \frac{N(z)}{D(z)}$, with $N(z)$, $D(z)$ given by

$$N(z) = 2(0.141z^6 + 0.2618z^5 + 0.5229z^4 + 0.1691z^3 + 0.2736z^2 + 0.2674z + 0.042)$$
$$D(z) = 0.2897z^7 + 0.1111z^6 + 0.0776z^5 + 0.3671z^4 + 0.4575z^3 + 0.2973z^2 + 0.3557z + 1$$

This transfer function is the same as that used in [GXZ93] to demonstrate a tuned algorithm. It is scaled up by 2 to avoid (5.12) from becoming trivially true. 32 uniformly spaced noisy point frequency response samples were identified by the algorithm described in chapter 4. For each frequency, length of data used for (4.12) is $N = 500$. Gaussian white noise of variance approximately 10% of 2-norm of 'clean' output data is added. Matlab's LMI control toolbox was used for solving the optimisation problem (5.9) for these frequency response samples. First, the parameters for $\overline{B}\mathcal{H}_{\infty,\rho_2}(\gamma_2)$ were $\gamma_2 = 10 \max_i |P_{\omega_i}|$, $\rho_2 = 1.05$. In figures (5.1), the solid line represents the true transfer function and the dashed line represents the frequency response of the identified model. '$*$' denotes the frequency response samples P_{ω_i}. The effect of spurious poles and zeroes is clearly evident in the model plots. Changing the value of γ_2 to upto $\gamma_2 = 1.1 \max_i |P_{\omega_i}|$ does not yield any significant improvement in the 'smoothness' of Nyquist plot (or worst case slope).

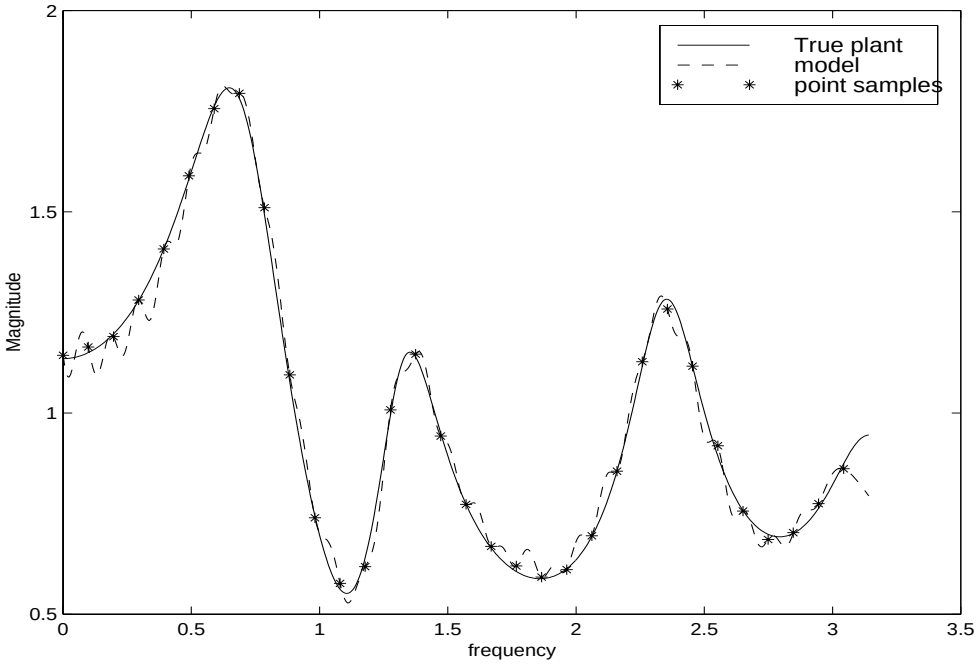


Figure 5.1: Magnitude plots of true plant and identified model

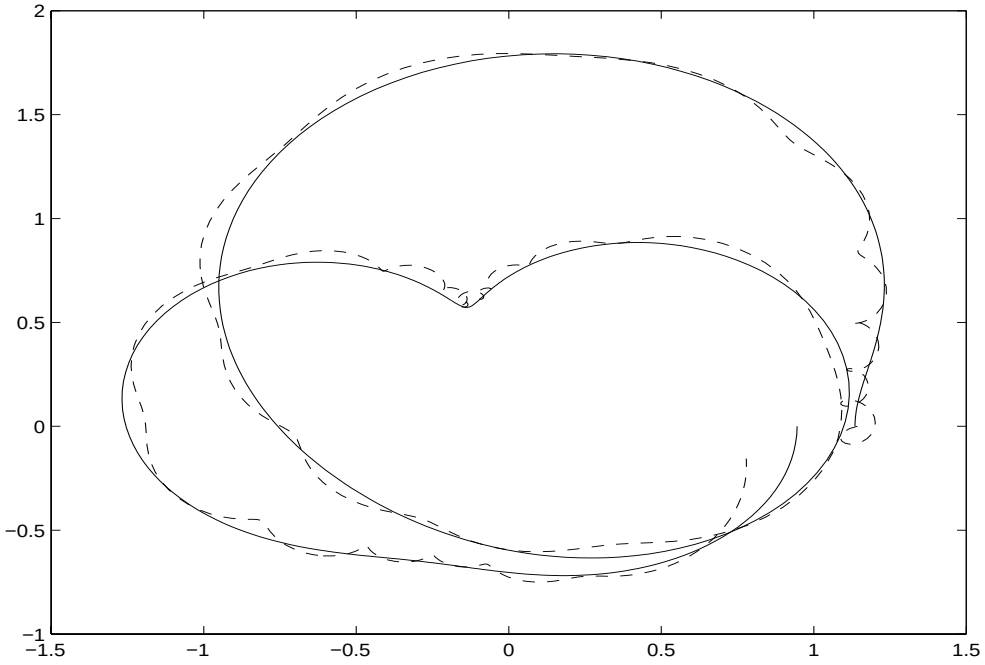


Figure 5.2: Nyquist plots of true plant and identified model

Chapter 6

Worst Case Identification using FIR Models

6.1 Introduction

In section 5.4, the idea of obtaining a model that gives the best worst case fit subject to a constraint on its slope (in terms of γ_2, ρ_2) is introduced. The algorithms in chapter 5 still share one major disadvantage with other worst case methods - the model order is a function of length of data. This restriction is removed in this chapter, obtaining a robustly convergent algorithm which is optimal, in a certain sense, for a finite model order. Further motivation of this work comes from the analysis of the optimal achievable worst case error by Zames et al in [ZLW94] and from a subsequent tuned FIR approximation scheme suggested by Glaum, Lin and Zames [GLZ96].

The Kolmogorov n-width of a subset $\mathcal{P} \subset \mathcal{H}_\infty$, in $\mathcal{H}_\infty$ is defined by [Pin85]

$$d_n(\mathcal{P}) = \inf_{X_n} \sup_{x \in \mathcal{P}} \inf_{y \in X_n} \|x - y\|_\infty \tag{6.1}$$

where the left-most infimum is taken over all n-dimensional subspaces of $\mathcal{H}_\infty$. Thus $d_n(\mathcal{P})$ represents the optimal worst case error one can achieve by approximating a system in $\mathcal{P}$ by a linear combination of n basis functions in $\mathcal{H}_\infty$. A subspace which actually achieves this error is called an optimal subspace for $\mathcal{P}$. We are interested in two plant sets:

$$\begin{aligned} \mathcal{P}_1(\gamma, \rho) &= \overline{B}\mathcal{H}_{\infty,\rho}(\gamma) \quad \text{and} \\ \mathcal{P}_2(M) &= \{f : f \in \mathcal{H}_\infty,\ \|f'\|_\infty \leq M\} \end{aligned} \tag{6.2}$$

where $f'(z) := \frac{df}{dz}$. For both $\mathcal{P}_1(\gamma,\, \rho)$ and $\mathcal{P}_2(M)$, an optimal subspace is

$$S_n = span\{1, z, \ldots, z^{n-1}\} \tag{6.3}$$

where *span* is over the real field. Further, the exact values of n-widths in these cases are [Pin85]

$$d_n(\mathcal{P}_1(\gamma,\, \rho)) = \gamma\rho^{-n},\ d_n(\mathcal{P}_2(M)) = \frac{M}{n} \tag{6.4}$$

Clearly, $d_n(\mathcal{P}_i) \overset{n\to\infty}{\longrightarrow} 0,\ i = 1,\, 2$. Contrary to intuition however, an optimal approximation is *not* always given by the first n impulse response parameters.

6.2 Problem Formulation

Assumption : The plant transfer function $P(z)$ to be identified is a stable, linear, shift invariant, SISO discrete time system.

***A priori* Information :**

- $P(z)$ belongs to *either* of the plant sets $\mathcal{P}_i$, $i = 1,\, 2$.
- Noise v corrupting measurement belongs to $\overline{B}l_\infty(\epsilon)$.

Here, $\epsilon \in [0, \infty)$, either $M \in [0, \infty)$ or $\rho \in (1, \infty)$ and $\gamma \in [\,0, \infty)$ are assumed to be known constants.

***A posteriori* Information :** A vector of (not necessarily uniformly spaced) noisy frequency response samples

$$P_\omega = [P_{\omega_1}\ P_{\omega_2}\ \ldots\ P_{\omega_m}]^T,\ P_{\omega_i} = P(e^{j\omega_i}) + v_i \tag{6.5}$$

where $\omega_i \in (0, \pi)$, $i = 1,\, 2,\, \ldots,\, m$.

Also given is a corresponding vector of (angular) frequencies

$$W = [\omega_1\, \omega_2\ \ldots\ \omega_m]^T \tag{6.6}$$

$\overline{\delta} = \max_i |\omega_{i+1} - \omega_i|$ and $\underline{\delta} = \min_i |\omega_{i+1} - \omega_i|$ are the maximum and the minimum separation between adjacent angular frequencies respectively. Let $[0, \pi)^m$ be as defined in (4.20).

Find : An algorithm $A_{\overline{\delta}} : \mathbb{C}^m \times [0, \pi)^m \mapsto S_n$ such that the worst case errors defined by

$$e_i(A_{\overline{\delta}} : x_i, \epsilon, n, \overline{\delta}) = \sup_{P \in \mathcal{P}_i(x_i), v \in \overline{B}l_\infty(\epsilon)} \| A_{\overline{\delta}}(P_\omega, W) - P \|_\infty \tag{6.7}$$

converge as follows

$$\begin{aligned} \lim_{\overline{\delta}\to 0} e_i(A_{\overline{\delta}} : x_i, \epsilon, n, \overline{\delta}) &\leq d_n(\mathcal{P}_i) + \beta\epsilon \\ \lim_{n\to\infty} e_i(A_{\overline{\delta}} : x_i, \epsilon, n, \overline{\delta}) &\leq g_i(x_i, \overline{\delta}, \epsilon) \end{aligned} \tag{6.8}$$

where $i = 1, 2$, $x_1 = (\gamma, \rho)$, $x_2 = M$, and $\overline{\delta} \to 0$ in such a way that $\frac{\overline{\delta}}{\underline{\delta}}$ remains bounded. g_i are some functions bounded for fixed x_i, monotonically increasing in both $\overline{\delta}$ and ϵ such that $\lim_{\epsilon\to 0} \lim_{\overline{\delta}\to 0} g(\alpha, \overline{\delta}, \epsilon) = 0$. In addition, derive explicit bounds on the above error.

An important feature of this formulation is the separation of the error convergence with respect to the increasing model order and that with respect to the increasing number of data points. Existing robustly convergent algorithms rely on model order being some *increasing* function of the number of data points for the error to go to zero. The proposed algorithm removes this constraint, giving a significantly stronger convergence property.

6.3 Identification Algorithm

$A_{\overline{\delta}}(P_\omega, W) \in S_n$ is a solution of the following optimisation problem:

$$\text{minimise } \lambda \text{ subject to } f \in S_n \text{ and}$$

$$\max \left\{ \max_{i\in[1,m]} |\, f(e^{j\omega_i}) - P_{\omega_i}|,\ \tilde{k}_1 \| f' \|_\infty \right\} < \lambda \tag{6.9}$$

where $\tilde{k}_1 = k_1 \left(\overline{\delta}\right)^r$, $0 < k1 < \infty$ and $0 < r < 1$ are constants. For finite data, user only needs to specify a single constant $\tilde{k}_1$; the decomposition of $\tilde{k}_1$ into a constant and $\overline{\delta}$-dependent part will be used later in section 6.4 to prove robust convergence.

Recalling the definition of S_n in (6.3), $A_{\overline{\delta}}(P_\omega, W)$ must be of the form

$$A_{\overline{\delta}}(P_\omega, W) = \sum_{k=0}^{n-1} a_k z^k, \quad a_k \in \mathbb{R}$$

The steps to implement (6.9) are enumerated below.

- Define

$$A = \begin{bmatrix} 0_{1\times n-3} & 0 \\ I_{n-3} & 0 \end{bmatrix},\ B = \begin{bmatrix} 1 \\ 0_{n-3\times 1} \end{bmatrix} \tag{6.10}$$

$$C = \begin{bmatrix} 2a_2 & 3a_3 & \dots & (n-1)a_{n-1} \end{bmatrix},\ D = a_1 \tag{6.11}$$

$$\phi_i = \begin{bmatrix} 1 & e^{j\omega_i} & \dots & e^{(n-1)j\omega_i} \end{bmatrix},\ i = 1, 2, \dots, m$$

$$\theta = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \end{bmatrix}^T$$

$$\begin{aligned} L_1 &= \begin{bmatrix} 0_{n-2\times n-2} & 0_{n-2\times 1} & C^T \\ 0_{1\times n-2} & 0 & D \\ C & D & 0 \end{bmatrix} \\ L_2 &= \operatorname{diag}\left(\begin{bmatrix} 0 & \theta^T\phi_i^* - P_{\omega_i}^* \\ \phi_i\theta - P_{\omega_i} & 0 \end{bmatrix} \right) \end{aligned} \tag{6.12}$$

$$\begin{aligned} R_1 &= -\begin{bmatrix} A^TXA - X & A^TXB & 0_{n-2\times 1} \\ B^TXA & B^TXB - 1 & 0 \\ 0_{1\times n-2} & 0 & -1 \end{bmatrix} \\ R_2 &= \tilde{k}_1\, I_{2m} \end{aligned} \tag{6.13}$$

where $X \in \mathbb{R}^{n-2\times n-2}$ is an arbitrary symmetric matrix.

- $$\text{Solve} \qquad \inf_{a_0, a_1, \dots, a_{n-1}} \lambda \tag{6.14}$$

$$\text{subject to} \qquad X > 0 \tag{6.15}$$

$$\text{and} \qquad L_i < \lambda\, R_i,\ i = 1, 2 \tag{6.16}$$

- Let $\hat{a}_k$, $k = 0, 1, \dots, n-1$ be the values of decision variables which minimise λ. Then

$$A_{\overline{\delta}}(P_\omega, W) = \sum_{k=0}^{n-1} \hat{a}_k z^k$$

It remains to be shown that the procedure outlined above actually amounts to solving (6.9).

Theorem 6.1 *Let $\hat{P}(z) = \sum_{k=0}^{n-1} \hat{a}_k z^k$, with $\hat{a}_k$ obtained as above. Then*

$$\max\left\{ \max_i |\hat{P}(e^{j\omega_i}) - P_{\omega_i}|,\ \tilde{k}_1 \|\hat{P}'\|_\infty \right\} = \inf_{f \in S_n} \max\left\{ \max_i |\, f(e^{j\omega_i}) - P_{\omega_i}|,\ \tilde{k}_1 \|\, f'\|_\infty \right\} \tag{6.17}$$

Proof : Let $f(z) = \sum_{k=0}^{n-1} a_k z^k$. Note that minimising $\max\{|\, f(e^{j\omega_i}) - P_{\omega_i}|, \tilde{k}_1 \|\, f'\|_\infty\}$ is same as minimising $\max\{\tilde{k}_1^{-1}|\, f(e^{j\omega_i}) - P_{\omega_i}|, \|\, f'\|_\infty\}$. From [GA94] and from definitions (6.10), (6.11), the conditions

$$\begin{bmatrix} A^TXA - X & A^TXB & \lambda^{-1}C^T \\ B^TXA & B^TXB - 1 & \lambda^{-1}D \\ \lambda^{-1}C & \lambda^{-1}D & -1 \end{bmatrix} < 0 \tag{6.18}$$

and $X > 0$ correspond to $\|f'\|_\infty < \lambda$. Also, $\tilde{k}_1^{-1}\,|\, f(e^{j\omega_i}) - P_{\omega_i}| < \lambda$ is enforced by

$$\text{diag}\left(\begin{bmatrix} \tilde{k}_1\,\lambda & P^*_{\omega_i} - \theta^T\phi_i^* \\ P_{\omega_i} - \phi_i\theta & \tilde{k}_1\,\lambda \end{bmatrix}\right) > 0 \tag{6.19}$$

Rearranging these LMIs gives (6.16). ∎

Remark 6.1 *In [MP92], a closely related problem is solved :*

$$\begin{aligned} &\textit{minimise} \quad \lambda \textit{ subject to} \quad f \in S_n \textit{ and} \\ &\max\left\{\max_{i\in[1,m]} |\, f(e^{j\omega_i}) - P_{\omega_i}|,\ k_1\|\, f'\|_1\right\} < \lambda \end{aligned} \tag{6.20}$$

where $\|\, f'\|_1 = \sum_{i=0}^{n-1} k|f_k|$, f_k *being parameters of* f. *Solution to this problem also yields a robustly convergent algorithm. The algorithm presented here has a significantly stronger convergence property. It also generalises nicely to the coprime factor identification case, as will be seen in section 6.6.*

6.4 Convergence and *a priori* Error Bounds

Let $\omega^* \in [0, \pi)$. For convenience of notation, denote $\hat{P}(z) = \left(A_{\overline{\delta}}(P_\omega, W)\right)(z)$. Then by triangle inequality,

$$\begin{aligned} &|\hat{P}(e^{j\omega^*}) - P(e^{j\omega^*})| \le |\hat{P}(e^{j\omega^*}) - \hat{P}(e^{j\omega_i})| + \\ &|\hat{P}(e^{j\omega_i}) - P_{\omega_i}| + |P_{\omega_i} - P(e^{j\omega_i})| + |P(e^{j\omega_i}) - P(e^{j\omega^*})| \\ &\le \|\hat{P}'\|_\infty \tfrac{\overline{\delta}}{2} + \lambda + \epsilon + \|P'\|_\infty \tfrac{\overline{\delta}}{2} \end{aligned} \tag{6.21}$$

where $\lambda = |P_{\omega_i} - \hat{P}(e^{j\omega_i})|$ and $f'(z) = \frac{df}{dz}$. For any P in $\mathcal{P}_1$ or in $\mathcal{P}_2$, $P' \in \mathcal{H}_\infty$ and $\|P'\|_\infty \frac{\overline{\delta}}{2} \xrightarrow{\overline{\delta}\to 0} 0$. Hence the convergence result (6.8) holds for $A_{\overline{\delta}}$ if, for $P \in \mathcal{P}_k$, $k = 1, 2$, the

FIR approximation $\hat{P}$ satisfies the following properties:

$$\lim_{n\to\infty} \max\left(\|\hat{P}'\|_\infty \frac{\overline{\delta}}{2}, \max_i |\hat{P}(e^{j\omega_i}) - P_{\omega_i}|\right) \leq g(x_i, \overline{\delta}, \epsilon) < \infty \tag{6.22}$$

$$\lim_{\overline{\delta}\to 0} \max\left(\|\hat{P}'\|_\infty \frac{\overline{\delta}}{2}, \max_i |\hat{P}(e^{j\omega_i}) - P_{\omega_i}|\right) \leq \beta\,\epsilon + d_n(\mathcal{P}_1)\,\forall\, i \in [1,\, m] \tag{6.23}$$

where $\lim_{\overline{\delta}\to 0} \lim_{\epsilon\to 0} g(x_i, \overline{\delta}, \epsilon) \longrightarrow 0$ for fixed x_i and β is independent of both data and *a priori* information. The error bound is derived and the convergence of the algorithm is proved by demonstrating the existence of feasible solution to the optimisation problem (6.9) which have the required convergence properties (6.22) and (6.23).

For $P \in \mathcal{P}_2$, a feasible solution with the required properties is given by the following result of Glaum et al ([GLZ96, lemma 3.1]):

Lemma 6.1 *([GLZ96]) For a priori information $v \in \overline{B}l_\infty(\epsilon)$, $P \in \mathcal{P}_1$ and a posteriori information as in section 6.2, there exists a $\hat{P}(z) \in S_n$ which satisfies the following constraints:*

$$|\hat{P}(e^{j\omega_i}) - P_{\omega_i}| \leq \epsilon + \frac{M}{n} + \frac{M\delta}{4} \tag{6.24}$$

$$\|\hat{P}'\|_\infty < M + \frac{2M}{n\delta} \tag{6.25}$$

where $\delta > 0$ is arbitrary.

Using an argument identical to that used for proving the above lemma in [GLZ96] and employing fact 5.1, one can prove the following result for $P \in \mathcal{P}_1$:

Lemma 6.2 *For a priori information $v \in \overline{B}l_\infty(\epsilon)$, $P \in \mathcal{P}_1$ and a posteriori information as in section 6.2, there exists a $\hat{P}(z) \in S_n$ which satisfies the following constraints :*

$$|\hat{P}(e^{j\omega_i}) - P_{\omega_i}| \leq \epsilon + \gamma\rho^{-n} + \frac{\gamma}{\rho - 1}\frac{\delta}{4} \tag{6.26}$$

$$\|\hat{P}'\|_\infty < \frac{\gamma}{\rho - 1} + \frac{2\gamma\rho^{-n}}{\delta} \tag{6.27}$$

where $\delta > 0$ is arbitrary.

Theorem 6.2

$$\lim_{n\to\infty} e_i(A_{\overline{\delta}} : x_i,\, \epsilon,\, n, \overline{\delta}) \leq g(x_i, \overline{\delta}, \epsilon)$$

$$\lim_{\overline{\delta}\to 0} e_i(A_{\overline{\delta}} : x_i,\, \epsilon,\, n, \overline{\delta}) \leq 2\,\epsilon + d_n(\mathcal{P}_i) \tag{6.28}$$

where $i = 1, 2$, $x_1 = \{\gamma, \rho\}$, $x_2 = M$ and $g(\cdot)$ is such that $\lim_{\overline{\delta}\to 0} \lim_{\epsilon\to 0} g(x_i, \overline{\delta}, \epsilon) = 0$ for fixed x_i.

Proof: Let $\tilde{k}_1 = k_1 \left(\overline{\delta}\right)^r$ where $0 < k_1 < \infty$ and $0 < r < 1$ are constants. For $P \in \mathcal{P}_2$, it is clear from lemma 6.1 that

$$\begin{aligned}
&\min_{f \in S_n} \max \left\{ \max_{i \in [1,m]} |\, f(e^{j\omega_i}) - P_{\omega_i}|,\ k_1 \| f' \|_\infty \left(\overline{\delta}\right)^r \right\} \\
&< \max \left(\epsilon + \frac{M}{n} + \frac{M\delta}{4},\ k_1 \left\{ M + \frac{2M}{n\delta} \right\} \left(\overline{\delta}\right)^r \right)
\end{aligned} \tag{6.29}$$

for an arbitrary $\delta > 0$. Choose $\delta = \left(\frac{\overline{\delta}}{n}\right)^\zeta$ for an arbitrary ζ, $0 < \zeta < r$ and let

$$\begin{aligned}
\lambda_{opt} &:= \max \left(\epsilon + \frac{M}{n} + \frac{M\delta}{4}, k_1 \left\{ M + \frac{2M}{n\delta} \right\} \left(\overline{\delta}\right)^r \right) \\
&= \max \left(\epsilon + \frac{M}{n} + \frac{M}{4} \left(\frac{\overline{\delta}}{n}\right)^\zeta, \left\{ k_1 M \left(\overline{\delta}\right)^r + \frac{2k_1 M \left(\overline{\delta}\right)^{r-\zeta}}{n^{1-\zeta}} \right\} \right)
\end{aligned} \tag{6.30}$$

Using (6.21), (6.29) and (6.30),

$$e(A_{\overline{\delta}} : M,\ \epsilon,\ n, \overline{\delta}) \leq \left(\frac{1}{k_1} \frac{\left(\overline{\delta}\right)^{1-r}}{2} + 1 \right) \lambda_{opt} + M\frac{\overline{\delta}}{2} + \epsilon \tag{6.31}$$

$$\begin{aligned}
\text{Also,} \quad \lim_{n \to \infty} \lambda_{opt} &= \max \left(\epsilon,\ k_1\, M\, \left(\overline{\delta}\right)^r \right) \\
\text{and} \quad \lim_{\overline{\delta} \to 0} \lambda_{opt} &= \epsilon + \frac{M}{n}
\end{aligned} \tag{6.32}$$

The result follows from (6.31) and (6.32). Proof for $P \in \mathcal{P}_1$ follows similarly using lemma 6.2.

■

6.5 Modifications and Extensions

6.5.1 Obtaining Smooth Approximations

For a fixed n and P_ω, $\tilde{k}_1$ in (6.9) serves as a trade off between smoothness of the approximation (in terms of $\|\hat{P}'\|_\infty$) and the achieved fit to the data. A tighter control over the smoothness of $\hat{P}(z)$ may be obtained by using $\|\hat{P}'\|_2$ or $\|\hat{P}''\|_\infty$. The following result lends justification to the use of $\|\hat{P}''\|_\infty$:

Lemma 6.3 *Let $\hat{P} \in \mathcal{H}_\infty$. Let ω_i and ω^* be two angular frequencies such that $|\omega_i - \omega^*| \leq \frac{\overline{\delta}}{2}$. Then*

$$|\hat{P}(e^{j\omega_i}) - \hat{P}(e^{j\omega^*})| \ \leq\ |\hat{P}'(e^{j\omega_i})| \frac{\overline{\delta}}{2} + \|\hat{P}''\|_\infty \frac{\overline{\delta}^2}{4} \tag{6.33}$$

Proof : Define $f : [0,\, \pi] \mapsto \mathbb{C}$, such that

$$f(\omega) := \hat{P}(e^{j\omega})$$

From ([Rud76], chapter 5),

$$|\, f(\omega_i) - f(\omega^*)| \leq \sup_{\omega \,\in\, [\omega_i, \omega^*]} |\, f'(\omega)||\omega_i - \omega^*| \tag{6.34}$$

where $f'(\omega) = \frac{df}{d\omega}$. Let the supremum of $f'(\omega)$ be attained at a frequency $\tilde{\omega}$. Then using (6.34),

$$\begin{aligned}
|\, f(\omega_i) - f(\omega^*)| &\leq |\, f'(\tilde{\omega})||\omega_i - \omega^*| \\
&= |\, f'(\tilde{\omega}) + f'(\omega_i) - f'(\omega_i)||\omega_i - \omega^*| \\
&\leq |\, f'(\tilde{\omega}) - f'(\omega_i)||\omega_i - \omega^*| + |\, f'(\omega_i)||\omega_i - \omega^*| \qquad (6.35) \\
&\leq \left(\sup_{\omega \,\in\, [\omega_i, \tilde{\omega}]} |\, f''(\omega)||\omega_i - \tilde{\omega}| \right) \frac{\overline{\delta}}{2} + |\, f'(\omega_i)| \frac{\overline{\delta}}{2} \qquad (6.36) \\
&\leq \sup_{\omega \,\in\, [0,\, \pi]} |\, f''(\omega)| \, \frac{\overline{\delta}^2}{4} + |\, f'(\omega_i)| \, \frac{\overline{\delta}}{2} \qquad (6.37)
\end{aligned}$$

Note that (6.36) follows from (6.35) by employing (6.34) for $f'(\omega)$. Now, using chain rule for differentiation gives

$$\left|\, f'(\omega)\right| = \left| \frac{df}{d\omega} \right| = \left| \frac{d\hat{P}}{d(e^{j\omega})} \right| \left| \frac{d(e^{j\omega})}{d\omega} \right| = \left| \frac{d\hat{P}}{d(e^{j\omega})} \right| \tag{6.38}$$

Similarly, $|\, f''(\omega)| = \left| \frac{d^2 \hat{P}}{d(e^{j\omega})^2} \right|$. As $\hat{P}(z) \in \mathcal{H}_\infty$, $\hat{P}'(z) \in \mathcal{H}_\infty$, $\hat{P}''(z) \in \mathcal{H}_\infty$. Therefore $|\, f'(\omega_i)| = |\hat{P}'(e^{j\omega_i})|$ and $\sup_{[0,\pi]} |\, f''(\omega)| = \sup_{[0,\pi]} |\hat{P}''(e^{j\omega})| = \|\hat{P}''\|_\infty$. This completes the proof. ∎

Let $\hat{P}(z) = A_{\overline{\delta}}(P_\omega, W) = \sum_{k=0}^{n-1} \hat{a}_k z^k$. Then the condition

$$\|\hat{P}''\|_\infty < \lambda$$

is affine in the parameters $\hat{a}_k$ of $\hat{P}(z)$. It is hence possible to modify (6.9) and pose an LMI problem

$$\text{minimise } \lambda \text{ subject to}$$

$$\max \left(\max_i |\hat{P}(e^{j\omega_i}) - P(e^{j\omega_i})|,\, \tilde{k}_1 \, \|\hat{P}'\|_\infty,\, \tilde{k}_2 \, \|\hat{P}''\|_\infty \right) < \lambda \tag{6.39}$$

where $\tilde{k}_1$, $\tilde{k}_2$ are user chosen nonzero weights. One may choose to replace $\tilde{k}_1 \|\hat{P}'\|_\infty$ by $\max_i \tilde{k}_1 |\hat{P}'(e^{j\omega_i})|$. As $\tilde{k}_1$ is a free parameter, both representations are equally good. The former leads to a higher number of decision variables, the latter, in general, to a higher number of LMI constraints.

Lemma 6.4 *For a priori information $v \in \overline{B}l_\infty(\epsilon)$, $P \in \mathcal{P}_1(\gamma, \rho)$ and a posteriori information as in section 6.2, there exists $\hat{P}(z) \in S_n$ such that*

$$\begin{aligned} &\max \left(|\hat{P}(e^{j\omega_i}) - P_{\omega_i}|, \tilde{k}_1 \|\hat{P}'\|_\infty, \tilde{k}_2 \|\hat{P}''\|_\infty \right) \\ &\leq \max \left(\epsilon + \alpha d_n(\mathcal{P}_1), \tilde{k}_1 \frac{\gamma}{(\rho-1)}, \tilde{k}_2 \frac{2\gamma}{(\rho-1)^2} \right) \end{aligned}$$

where $\alpha = \frac{\rho}{\rho-1}$.

Proof : Let $P(z) = \sum_{i=0}^{\infty} p_k z^k$ and choose $\hat{P}(z) = \sum_{i=0}^{n-1} p_k z^k$. The result follows using fact 5.1 and using

$$|\hat{P}(e^{j\omega_i}) - P_{\omega_i}| \leq \epsilon + |\hat{P}(e^{j\omega_i}) - P(e^{j\omega_i})| \qquad \blacksquare \tag{6.40}$$

Using (6.22)- (6.23) and the above lemma, it is seen that (6.39) only slightly weakens the convergence properties (6.8) for $P \in \mathcal{P}_1$ (i.e. the worst case error is within a constant factor α).

Similarly, $\|\hat{P}'\|_2$ may also be incorporated in the min-max problem (6.39) as an additional smoothness constraint. With $\hat{P}(z) = \sum_{k=0}^{n-1} a_k z^k$ and C, D as defined in (6.11),

$$\begin{bmatrix} \lambda & 0 & C^T \\ 0 & \lambda & D \\ C & D & \lambda \end{bmatrix} > 0 \Leftrightarrow \|\hat{P}'\|_2 < \lambda$$

6.5.2 Using fixed pole structures

Let $S(z) = \sum_{k=0}^{l-1} s_k z^k$ be a user chosen polynomial of degree $r-1$ such that $S^{-1} \in \mathcal{H}_\infty$. $S(z)$ may represent the user's knowledge about pole locations, based on elementary tests like step response. Define a function $\hat{d}_n$ by

$$\begin{aligned} \hat{d}_n &= d_n(\mathcal{P}_1) \quad n > 0 \\ &= \gamma \qquad\quad n \leq 0 \end{aligned} \tag{6.41}$$

Consider the following problem :

Given *a priori* information $v \in \overline{B}l_\infty(\epsilon)$, $P \in \mathcal{P}_1(\gamma, \rho)$, the same *a posteriori* information as in

section 6.2 and additionally a user chosen polynomial $S(z)$ of degree $l-1$,

Find : An algorithm $A_{\overline{\delta}} : \mathbb{C}^m \times [0,\pi)^m \mapsto S_n$ such that, for IIR transfer function $\hat{P}(z) = \frac{A_{\overline{\delta}}(P_\omega, W)}{S(z)}$, the worst case error defined by

$$e_1(A_{\overline{\delta}} : \gamma, \rho, \epsilon, n, \overline{\delta}) = \sup_{P \in \mathcal{P}_1(\gamma,\rho), v \in \overline{B}l_\infty(\epsilon)} \|\hat{P} - P\|_\infty, \tag{6.42}$$

converges as

$$\lim_{\overline{\delta}\to 0} e_1(A_{\overline{\delta}} : \gamma, \rho, \epsilon, n, \overline{\delta}) \;\leq\; \alpha \hat{d}_{n-l} + \beta\,\epsilon \tag{6.43}$$

for some constants α, β independent of n and $\overline{\delta}$.

Solution : Let $Q(z) = \sum_{k=0}^{n-1} (A_{\overline{\delta}}(P_\omega, W))(k) z^k$. Then

$$\hat{P}'(z) = \left(\frac{SQ' - QS'}{S^2}\right)(z) \tag{6.44}$$

The above expression is affine in the parameters of $Q(z)$. Consequently, the following problem may be written as an affine problem in the parameters of $Q(z)$,

$$\text{minimise } \lambda \text{ subject to}$$

$$\max\left\{ \max_i |\hat{P}(e^{j\omega_i}) - P_{\omega_i}|,\; \tilde{k}_1 \|\hat{P}'\|_\infty \right\} < \lambda \tag{6.45}$$

(6.45) can be solved by a procedure identical to the one in section 6.3, with suitable redefinitions of matrices. The following lemma ensures convergence property.

Lemma 6.5 *Let $S(z) = \sum_{k=0}^{l-1} s_k z^k$. For a priori information $v \in \overline{B}l_\infty(\epsilon)$, $P \in \mathcal{P}_1(\gamma,\rho)$ and a posteriori information as in section 6.2, there exists $Q(z) \in S_n$ such that*

$$\left\| \left(\frac{Q}{S}\right)'(z) \right\|_\infty \leq \frac{\gamma}{(\rho - 1)}$$

and

$$\left| \left(\frac{Q}{S}\right)(e^{j\omega_i}) - P_{\omega_i} \right| \leq \epsilon + \frac{1}{\rho - 1} \hat{d}_{n-l}$$

Proof : If $n < l$, $\hat{d}_{n-l} = \gamma$ and $Q(z) = 0$ conforms to the requirements. For $n \geq l$, let $P(z) = \sum_{i=0}^{\infty} p_k z^k$ and let $f(z) = \sum_{i=0}^{n-l} p_k z^k$, $Q(z) = f(z)S(z)$. Clearly, $Q(z) \in S_n$ and

$$\hat{P}(z) = \frac{Q(z)}{S(z)} = f(z) \in \mathcal{P}_1(\gamma,\rho)$$

The result then follows using fact 5.1 to bound $\|\hat{P}'\|_\infty$ and using (6.40). Convergence property (6.43) follow by using (6.22)- (6.23) and taking appropriate limits. ∎

Using a fixed pole structure weakens the asymptotic property; in the worst case, the assumed fixed poles will only worsen our approximation. However, a good starting estimate of fixed poles may reduce the number of FIR parameters needed for a good approximation.

Numerical Implementation

Implementing $\|\{\frac{Q}{S}\}'\|_\infty < \lambda$ where Q depends on the decision variables and S is fixed is possible, but rather cumbersome. Alternatively, since

$$\left\|\frac{Q}{S}\right\|'_\infty \leq \left\|\begin{bmatrix} Q' & Q \end{bmatrix}\right\|_\infty \left\|\begin{bmatrix} \frac{1}{S} \\ \frac{-S'}{S^2} \end{bmatrix}\right\|_\infty$$

$\left\|\begin{bmatrix} Q' & Q \end{bmatrix}\right\|_\infty$ may be used in place of $\|\hat{P}'\|_\infty$ in (6.45). This gives a much simpler implementation.

6.5.3 Use of Other Norms

Define an error vector $e := \begin{bmatrix} f(e^{j\omega_1}) - P_{\omega_1} & f(e^{j\omega_2}) - P_{\omega_2} & \dots & f(e^{j\omega_m}) - P_{\omega_m} \end{bmatrix}^T$. Then (6.9) can be re-written as

$$\text{minimise } \lambda \text{ subject to } f \in S_n \text{ and}$$

$$\max \left\{ |e|_\infty,\ \tilde{k}_1 \|f'\|_\infty \right\} < \lambda \tag{6.46}$$

where $|\cdot|_\infty$ denotes the infinity norm of $m-$ vector. $\infty-$norm is very sensitive to effect of *outliers* i.e. measurements for which the corresponding disturbance is exceptionally high. The effect of outliers may be reduced by choosing other norms for the error vector e. One obvious alternative would be

$$\text{minimise } \lambda \text{ subject to } f \in S_n \text{ and}$$

$$\max \left\{ |e|_2,\ \tilde{k}_1 \|f'\|_2 \right\} < \lambda \tag{6.47}$$

This is a much simpler optimisation problem; with only two constraints in n variables as opposed to $m+1$ constraints in $\mathcal{O}(n^2)$ variables in (6.9). There is no guarantee of worst case convergence, however.

For a cost of increase in computation, a further refinement in the problem formulation is possible. For $v \in \mathbb{C}^m$ and $1 \leq K \leq m$, define the $K-norm$:

$$|v|_K = K^{-1} \sup_{\mathcal{S} \subset \{1,2,\dots,m\}, |\mathcal{S}|=K} \sum_{i \in \mathcal{S}} |v_i|$$

Here $|\mathcal{S}|$ is the cardinality of set $\mathcal{S}$. The cases $K = 1$, $K = m$ respectively represent l_∞ norm and normalised l_1 norm. In general, it represents an average of magnitude of K largest elements of the vector. Consider a problem:

$$\text{minimise } \lambda \text{ subject to } f \in S_n \text{ and}$$

$$\max \left\{ |e|_{\mathcal{K}},\ \tilde{k}_1 \|f'\|_\infty \right\} < \lambda \tag{6.48}$$

For a fixed K, this clearly gives a robustly convergent algorithm, which may be proven using the same feasible solutions as in lemma 6.1 and lemma 6.2. Use of K-norms and other measures for rejection of outliers are discussed in details in [PM99].

6.6 Identification of Coprime Factors

To address the problem of identification of coprime factors of a plant using an FIR based model set, define the following sets,

$$\mathcal{P}_1^g = \{G : G = [N\ D]^T;\ N, D\ \in \mathcal{P}_1(\gamma, \rho)\}$$

$$\mathcal{P}_2^g = \{G : G = [N\ D]^T;\ N, D\ \in \mathcal{P}_2(M)\}$$

As a candidate model set, define a set of functions

$$S_{1,2} = \{f : f = [f_1\ f_2]^T, f_1 \in S_{n_1},\ f_2 \in S_{n_2}\} \tag{6.49}$$

Suppose point frequency response samples of coprime factors of a plant are given. The plant coprime factor samples may be estimated by using closed loop measurements for a known stabilising, and stable controller. The problem is to map these samples into a model in $S_{1,2}$ via a robustly convergent algorithm.

6.6.1 Problem Formulation

A priori Information :

- The (not necessarily normalised) right graph $G = [N\ D]^T$ of the SISO plant P belongs to either of the sets $\mathcal{P}_i^g,\ i = 1, 2$.
- Noise v corrupting measurement is in $\overline{B}l_\infty(\epsilon)$.

A posteriori Information :

A matrix of (not necessarily uniformly spaced) noisy frequency response samples

$$\begin{aligned} G_\omega &= [G_{\omega_1}\ G_{\omega_2}\ \ldots\ G_{\omega_m}] \\ G_{\omega_i} &= G(e^{j\omega_i}) + v_i \\ &= [N(e^{j\omega_i})\ D(e^{j\omega_i})]^T + [v_1(i)\ v_2(i)]^T \end{aligned} \tag{6.50}$$

where $\omega_i \in [0, \pi)$, $i = 1, 2, \ldots, m$ and $v_1(i)$, $v_2(i) \in \overline{B}l_\infty(\epsilon)$.

W, $\overline{\delta}$, and $\underline{\delta}$ are as defined in section 6.2.

Find :

An algorithm $A_{\overline{\delta}} : \mathbb{C}^{2\times m} \times [0, \pi)^m \mapsto S_{1,2}$ such that the worst case errors defined by

$$e_i(A_{\overline{\delta}} : x_i, \epsilon, n_1, n_2, \overline{\delta}, \epsilon) = \sup_{G \in \mathcal{P}_i^g, v \in \overline{B}l_\infty(\epsilon)} \| A_{\overline{\delta}}(G_\omega, W) - G \|_\infty \tag{6.51}$$

converge as follows

$$\begin{aligned} \lim_{\overline{\delta}\to 0} e_i(A_{\overline{\delta}} : x_i, n_1, n_2, \overline{\delta}, \epsilon) &\le d_{n_1}(\mathcal{P}_i) + d_{n_2}(\mathcal{P}_i) + \beta\epsilon \\ \lim_{n_1, n_2 \to \infty} e_i(A_{\overline{\delta}} : x_i, \epsilon, n_1, n_2, \overline{\delta}, \epsilon) &\le g_i(x_i, \overline{\delta}, \epsilon) \end{aligned} \tag{6.52}$$

where $i = 1, 2$, $x_1 = (\gamma, \rho)$, $x_2 = M$, and $\overline{\delta} \to 0$ in such a way that $\frac{\overline{\delta}}{\underline{\delta}}$ remains bounded. g_i are functions monotonically non-decreasing in both $\overline{\delta}$ and ϵ such thatlim$_{\epsilon\to 0}$ lim$_{\overline{\delta}\to 0}$ $g_i(x_i, \overline{\delta}, \epsilon) = 0$. β is a constant independent of both data and *a priori* information. In addition, derive explicit bounds on the above error.

6.6.2 Identification Algorithm and Error Bounds

$A_{\overline{\delta}}(G_\omega, W)$ is a solution to the following optimisation problem :

$$\text{minimise } \lambda \text{ subject to } f \in S_{1,2} \text{ and}$$

$$\max \left\{ \max_{i\in[1,m]} \|f(e^{j\omega_i}) - G_{\omega_i}\|_2,\ \tilde{k}_1 \|f'\|_\infty \right\} < \lambda \tag{6.53}$$

where $\tilde{k}_1$ is a user specified weight and $S_{1,2}$ is as defined in (6.49). This may be written as an LMI optimisation using (6.18), in a form similar to (6.14)- (6.16).

Theorem 6.3

$$\begin{aligned} \lim_{n\to\infty} e_i(A_{\overline{\delta}} : x_i,\ \epsilon,\ n, \overline{\delta}) &\le g(x_i, \overline{\delta}, \epsilon) \\ \lim_{\overline{\delta}\to 0} e_i(A_{\overline{\delta}} : x_i,\ \epsilon,\ n, \overline{\delta}) &\le (2 + \sqrt{2})\epsilon + d_{n_1}(\mathcal{P}_i) + d_{n_2}(\mathcal{P}_i) \end{aligned} \tag{6.54}$$

where $i = 1, 2$, $x_1 = \{\gamma, \rho\}$, $x_2 = M$ *and* $g(\cdot)$ *is such that* lim$_{\overline{\delta}\to 0}$ lim$_{\epsilon\to 0}$ $g(x_i, \overline{\delta}, \epsilon) = 0$ *for fixed* x_i.

Proof : Using the facts

$$\left\| \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \right\|_\infty \leq \|f_1\|_\infty + \|f_2\|_\infty$$
$$\left\| \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \right\|_\infty \leq \sqrt{2}\max\{\|f_1\|_\infty, \|f_2\|_\infty\}$$

for any SISO $f_1,\ f_2 \in \mathcal{RH}_\infty$, and

$$\left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 \leq |a| + |b|$$
$$\left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 \leq \sqrt{2}\max\{|a|, |b|\}$$

for any complex scalars a, b, the above bounds can be easily derived from the equations (6.21) and (6.24) - (6.27). ∎

The convergence property (6.52) follows by taking appropriate limits and using (6.4).

6.7 Simulation Example

To illustrate the use of these algorithms, a discrete transfer function $\frac{N(z)}{D(z)}$ is considered, with $N(z)$, $D(z)$ given by

$$N(z) = 2(0.141z^6 + 0.2618z^5 + 0.5229z^4 + 0.1691z^3 + 0.2736z^2 + 0.2674z + 0.042)$$
$$D(z) = 0.2897z^7 + 0.1111z^6 + 0.0776z^5 + 0.3671z^4 + 0.4575z^3 + 0.2973z^2 + 0.3557z + 1$$

The same *a posteriori* data as used for the simulation example in section 5.5 is used here. With $n = 23$, the optimisation (6.39) was carried out for two different combinations of $\tilde{k}_1, \tilde{k}_2$ using Matlab's LMI control toolbox [Mat95]. In fig. (6.1), the solid line and the dashed line respectively indicate the true and the estimated frequency responses for $\tilde{k}_1 = 0.0050$, $\tilde{k}_2 = 0$. The achieved value of λ is 0.0644. For the same data, using $\tilde{k}_1 = 0.0050$, $\tilde{k}_2 = 0.0008$ gives a smoother FIR approximation, with only a small increase in the achieved cost ($\lambda = 0.0907$). The response of this second model is compared with that of the plant in fig. (6.2). Both FIR approximations are seen to be significantly smoother than the approximation in fig. (5.2).

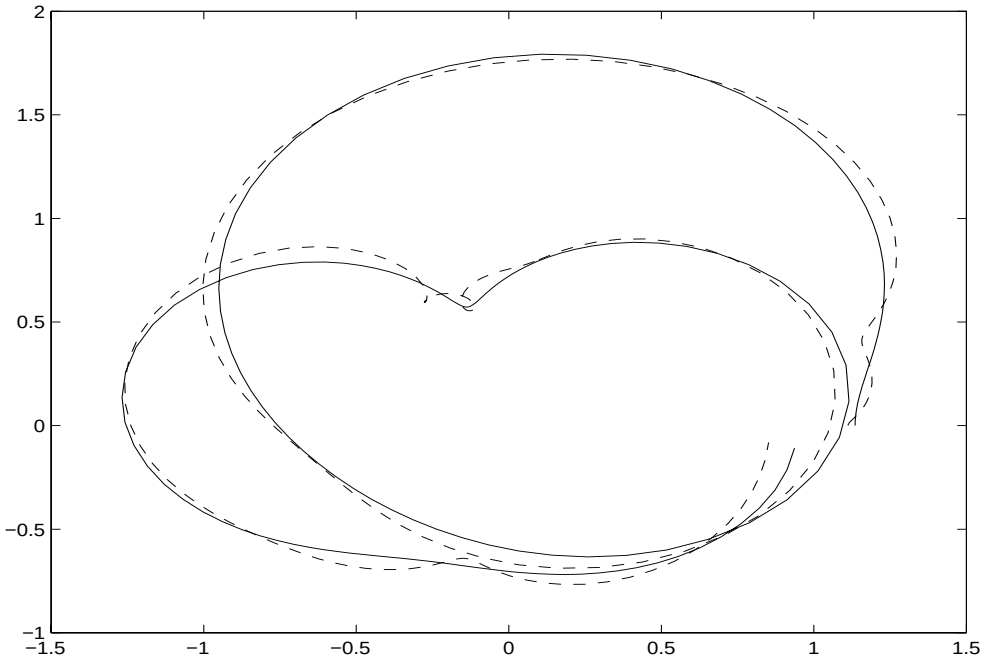


Figure 6.1: Nyquist Plots: $\tilde{k}_1 = .005, \tilde{k}_2 = 0$

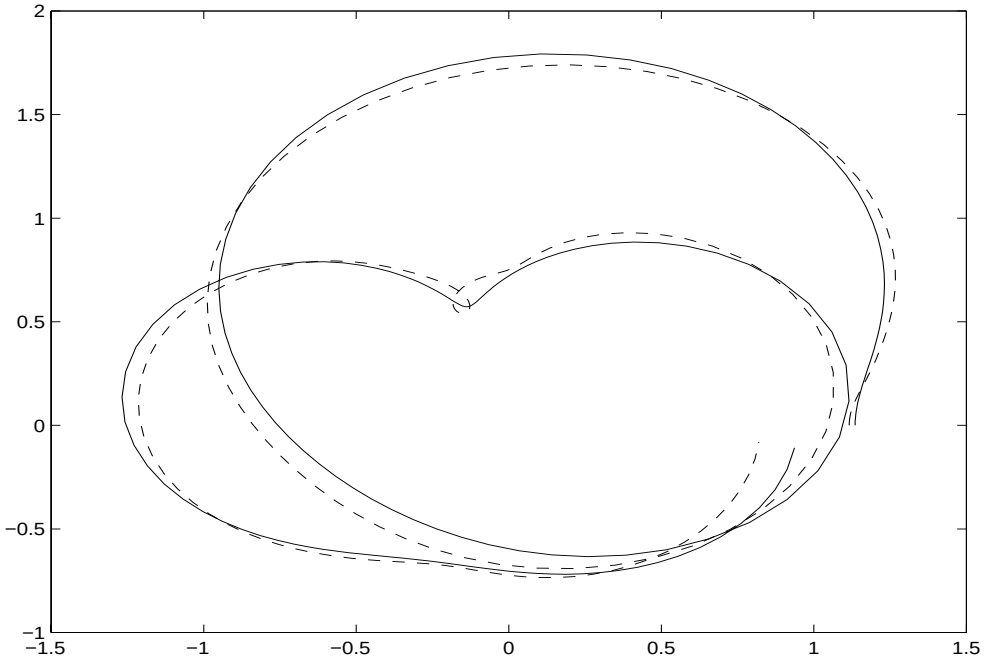


Figure 6.2: Nyquist Plots: $\tilde{k}_1 = .005, \tilde{k}_2 = .0008$

Chapter 7

Identification in the $\nu-$gap metric

7.1 The $\nu-$gap metric

For a plant P_i, let G_i and $\tilde{G}_i$ represent normalised right and normalised left graph symbols respectively. The $\nu-$gap between two plants P_0 and P_1 is defined as [Vin93]

$$\begin{aligned}\delta_\nu(P_0, P_1) &= \inf_{Q, Q^{-1} \in \mathcal{L}_\infty} \| G_0 - G_1 Q \|_\infty \text{ if } I(P_0, P_1) = 0 \\ &= 1 \quad \text{otherwise} \end{aligned} \tag{7.1}$$

where $I(P_0, P_1) :=$ wno $\det(G_1^* G_0) =$ wno $\det(\tilde{G}_0 \tilde{G}_1^*)$ and wno (g) denotes the winding number of $g(z)$ evaluated on the standard Nyquist contour indented around any poles on $\partial\mathbb{D}$. For a real rational transfer matrix X such that $X, X^{-1} \in \mathcal{RL}_\infty$, the winding number wno det (X) is the excess of zeros of X in $\mathbb{D}$ over the poles of X in $\mathbb{D}$. Using the fact that $\begin{bmatrix} G_1^* \\ \tilde{G}_1 \end{bmatrix}$ is unitary, we have

$$\overline{\sigma}(G_0 - G_1 Q)(e^{j\omega}) = \overline{\sigma}\left(\begin{bmatrix} G_1^* \\ \tilde{G}_1 \end{bmatrix} (G_0 - G_1 Q) \right)(e^{j\omega}) \tag{7.2}$$

$$= \overline{\sigma}\left(\begin{bmatrix} G_1^* G_0 - Q \\ \tilde{G}_1 G_0 \end{bmatrix} \right)(e^{j\omega}) \tag{7.3}$$

since $\tilde{G}_1 G_1 = \tilde{M}_1 (P_1 - P_1) M_1 = 0$. Setting $Q = G_1^* G_0$ whenever $\det(G_1^* G_0) \neq 0$ gives an alternate definition for the $\nu-$gap metric

$$\begin{aligned}\delta_\nu(P_0, P_1) &:= \|\tilde{G}_1 G_0\|_\infty \qquad \text{if } I(P_0, P_1) = 0 \\ &= 1 \qquad \text{otherwise} \end{aligned} \tag{7.4}$$

When the winding number condition is satisfied, $\delta_\nu(P_0, P_1)$ equals the $\mathcal{L}_2$ *-gap*, defined by

$$\delta_{\mathcal{L}_2}(P_0, P_1) := \|\tilde{G_1} G_0\|_\infty = \sup_\omega \kappa(P_0, P_1)(e^{j\omega}) \tag{7.5}$$

where $\kappa(P_0, P_1)(e^{j\omega})$ is the pointwise *chordal* distance defined as in section 3.2.5,

$$\kappa(P_0, P_1)(e^{j\omega}) := \overline{\sigma}\left((I + P_1 P_1^*)^{-\frac{1}{2}}(P_0 - P_1)(I + P_0^* P_0)^{-\frac{1}{2}}\right)(e^{j\omega})$$

The equivalence follows as $\tilde{G_1} G_0 = \tilde{M_1}(P_1 - P_0) M_0$ and $\tilde{M_1}^* \tilde{M_1} = (I + P_1 P_1^*)^{-1}$, $M_0 M_0^* = (I + P_0^* P_0)^{-1}$.

$\delta_\nu(P_0, P_1)$ is a measure of difference in closed loop performance of P_0 in feedback with a controller C and P_1 in feedback with the same controller C. Given a nominal controller C that stabilises a (possibly frequency weighted) plant P_0, a useful closed loop performance measure is

$$b_{P_0,C} = \|H(P_0, C)\|_\infty^{-1} \tag{7.6}$$

where the closed loop transfer function $H(P_0, C)$ is defined as in (3.14),

$$H(P_0, C) = \begin{bmatrix} P_0 \\ I \end{bmatrix} (I - CP_0)^{-1} \begin{bmatrix} -C & I \end{bmatrix}$$

It is known that [Vin93] any controller stabilising P_0 and achieving $b_{P_0,C} > \alpha$ stabilises the plant set

$$\{P_1 : \delta_\nu(P_0, P_1) \leq \alpha\}$$

More importantly, the *pointwise* difference in the closed loop performance of nominal plant P_0 and a perturbed plant P_1 for the same controller C can be quantified in terms of $\kappa(P_0, P_1)$ as [Vin93]:

$$\begin{aligned} \kappa(P_0, P_1)(e^{j\omega}) &\leq \overline{\sigma}(H(P_0, C) - H(P_1, C))(e^{j\omega}) \\ &\leq \kappa(P_0, P_1)(e^{j\omega})\, \overline{\sigma}(H(P_0, C))(e^{j\omega})\, \overline{\sigma}(H(P_1, C))(e^{j\omega}) \end{aligned} \tag{7.7}$$

The upper bound has been mentioned previously in section 3.2.5. Clearly, the upper bound in (7.7) makes sense only if C stabilises both P_0 and P_1.

Let η $(H(P, C))$ represent the number of closed loop unstable poles of $H(P, C)$ and let $b_{\mathcal{L}_2}(P, C) = \|H(P, C)\|_{\mathcal{L}_\infty}^{-1}$, regardless of stability of $H(P, C)$. $b_{\mathcal{L}_2}(P, C) = 0$ if $H(P, C)$ has a pole on $\partial\mathbb{D}$. Suppose a controller C is designed to stabilise a model P_0 of the true plant P_1. Then from ([Vin], section 3.8), a 'good' model (small $\delta_{\mathcal{L}_2}(P_0, P_1)$) and a 'good' controller C

(with large $b_{\mathcal{L}_2}(P_0, C)$) ensures that C stabilises P_1, provided the winding number condition is satisfied:

Lemma 7.1 *([Vin]) Given P_0, P_1, C satisfying $b_{\mathcal{L}_2}(P_0, C) > \delta_{\mathcal{L}_2}(P_0, P_1)$,*

$$\eta\left(H(P_1, C)\right) = \eta\left(H(P_0, C)\right) + I(P_1, P_0) \tag{7.8}$$

From (7.4), (7.7) and (7.8), it follows that any controller C that stabilises P_0 with a good $b_{P_0,C}$ also stabilises P_1, *without any significant deterioration in performance*, provided $\delta_\nu(P_0, P_1)$ is small.

To emphasise the difference between measuring the size of identification error by $\nu-$gap and by any other standard metrics, consider a pair of continuous time plants,

$$P_{\epsilon_+} = \frac{1}{s+\epsilon},\ P_{\epsilon_-} = \frac{1}{s-\epsilon},\ 0 < \epsilon \ll 1$$

The difference between P_{ϵ_+} and P_{ϵ_-} is unbounded in $\mathcal{H}_2$ or $\mathcal{H}_\infty$ norms. However, for a simple controller $C = -\delta,\ \delta > 1$, a little algebra reveals that

$$H(P_{\epsilon_+}, C) - H(P_{\epsilon_-}, C) = \frac{2\epsilon}{(s+\delta+\epsilon)(s+\delta-\epsilon)} \begin{bmatrix} -\delta & -1 \\ \delta^2 & \delta \end{bmatrix} \tag{7.9}$$

$$\text{and} \quad \|H(P_{\epsilon_+}, C) - H(P_{\epsilon_-}, C)\|_\infty = \frac{2\epsilon(1+\delta^2)}{(\delta+\epsilon)(\delta-\epsilon)}$$

so that, from a feedback point of view, the two systems P_{ϵ_+} and P_{ϵ_+} should be judged close for a small ϵ. The $\nu-$gap metric captures this fact correctly, as

$$G_{\epsilon_-}^* G_{\epsilon_+} = \frac{(s+\epsilon+1)(s+\epsilon-1)}{(s+\sqrt{1+\epsilon^2})(s-\sqrt{1+\epsilon^2})}$$

so that $\text{wno}\,\det\left(G_{\epsilon_-}^* G_{\epsilon_+}\right) = 0$ and $\delta_\nu(P_{\epsilon_+}, P_{\epsilon_-})$ may be shown to be $\frac{2\epsilon}{1+\epsilon^2}$.

The $\nu-$gap metric is closely related to *gap* metric [GS90] defined by

$$\delta_g(P_0, P_1) := \max\left\{\overrightarrow{\delta_g}(P_0, P_1), \overrightarrow{\delta_g}(P_1, P_0)\right\} \tag{7.10}$$

$$= \inf_{Q, Q^{-1} \in \mathcal{H}_\infty} \|G_0 - G_1 Q\|_\infty \tag{7.11}$$

where $\overrightarrow{\delta_g}(P_0, P_1)$ is the directed gap as defined in (2.9). The equivalence of (7.10) and(7.11) is proved in ([Vin], chapter 7). Comparing (7.1) and (7.10), it follows that $\delta_g(P_0, P_1)$ bounds $\delta_\nu(P_0, P_1)$ from above. Also important is the fact that gap metric is just a number, whereas $\nu-$gap metric is a supremum over frequency of a pointwise distance function; the value of

this function at all frequencies provides useful information about the difference between the two plants.

From the above discussion, a logical aim for control oriented identification is to minimise $\nu-$gap between the true plant and the model. This would, however, require infinite data. To facilitate the formulation of an identification problem formally on a finite set of frequencies, some relevant quantities are now defined. As *a posteriori* information in the identification process, suppose that a vector of (not necessarily uniformly spaced) frequency response samples of the normalised right graph symbol of the true plant $P_0(z)$ is given:

$$G_\omega := [G_0(e^{j\omega_1})\ G_0(e^{j\omega_2})\ \dots\ G_0(e^{j\omega_m})]^T \tag{7.12}$$

where $\omega_i \in (0, \pi)$, $i = 1, 2, \dots, m$.

Define a function $\delta_W : \mathcal{R} \times \mathcal{R} \times [0, \pi)^m \to [0, 1]$, such that

$$\begin{aligned}\delta_W(P_0, P_1) &:= \inf_{Q, Q^{-1} \in \mathcal{L}_\infty} \sup_{i \in [1,m]} \overline{\sigma}(\, G_0 - G_1 Q)(e^{j\omega_i}) \text{ if } I(P_0, P_1) = 0 \\ &= 1 \ otherwise\end{aligned} \tag{7.13}$$

Next, given any controller C, define a performance measure over finite frequency set as

$$b^w_{P_0,C} := \left\{ \sup_{i \in [1,m]} \overline{\sigma}\left(H(P_0, C)\right)(e^{j\omega_i}) \right\}^{-1} \tag{7.14}$$

Let

$$\Delta H_{01} := \sup_{i \in [1,m]} \left\{ \overline{\sigma}(H(P_0, C) - H(P_1, C))(e^{j\omega_i}) \right\}$$

be the performance degradation measure. Then

$$\delta_W(P_0, P_1) \ \leq \ \Delta H_{01} \leq \frac{\delta_W(P_0, P_1)}{b^w_{P_0,C} b^w_{P_1,C}} \tag{7.15}$$

follows immediately from (7.7).

7.2 An 'ideal' Algorithm

In view of the discussion above, a reasonable problem for control oriented identification with a finite data would be

$$\textbf{A1:} \qquad \min_{P_1 \in \mathcal{S}} \delta_W(P_0, P_1)$$

for given frequency response samples $G_0(e^{j\omega_i})$, where $\mathcal{S}$ is a model set yet to be determined. A possible choice would be an affinely parameterised set $S_{1,2}$, defined in (6.49). For $f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$, $f_1 \in S_{n1}$, $f_2 \in S_{n2}$, let

$$Fac(f) := f_1 f_2^{-1}$$

In addition to minimising $\delta_W(P_0, P_1)$, we may wish to control the rate of change of pointwise graph of model between two measurement frequencies on $\partial\mathbb{D}$. An appropriate measure for this purpose is $\mathcal{V}_i = \|\tilde{G}_i G_i'\|_\infty$. As shown in [Vin96], $\kappa(P_i(e^{j\omega_1}), P_i(e^{j\omega_2})) < \mathcal{V}_i|\omega_1 - \omega_2|$. In this case **A1** may be modified to

$$\textbf{A2:} \qquad \min_{f \in \mathcal{S}_{1,2}} \max(\delta_W(P_0, P_1), k\mathcal{V}_1)$$

where k is a user chosen weight and $P_1 = Fac(f)$. Note that this problem is similar to (6.9); both problems give a model that trades off worst case fit achieved and the complexity of model. The measures of both the worst case fit and complexity differ, however, which makes **A2** more control relevant than (6.9) and also numerically more demanding.

If the true plant belongs to a set $\mathcal{S}_\alpha := \{P_0 : \mathcal{V}_0 < \alpha\}$, then the worst case error bound

$$\sup_{P_0 \in \mathcal{S}_\alpha} \delta_\nu(P_0, P_1) \leq (\alpha + \mathcal{V}_1)\frac{\overline{\delta}}{2} + \delta_W(P_0, P_1)$$

follows by triangle inequality. Here $\overline{\delta}$ is the maximum separation between adjacent angular frequencies, as before. Additionally, if the plant $P_0 \in \mathcal{R}$, one may guarantee convergence of the worst case error in **A2** as the length of data goes to infinity and the model order goes to infinity (since $P_1 = P_0$, $\mathcal{V}_1 = \alpha$ is a feasible solution).

Having set out what we would like to do, let us see what we *can* do. Even with finite data, minimising $\delta_W(P_0, P_1)$ is a very difficult problem, since $\kappa(P_0, P_1)$ is non-convex in parameters of P_1 and implementing $I(P_0, P_1) = 0$ would require more information about P_0. Here, a three-step, approximate procedure is used to obtain a model.

- Given $P_0(e^{j\omega_i})$, obtain $G_0(e^{j\omega_i})$ within multiplication by a unitary matrix.
- Solve

  $$\min_{f \in S_{1,2}} \max \left\{ \max_i \kappa(P_0, Fac(f))(e^{j\omega_i}), k\|f'\|_\infty \right\}$$

  where k is a user chosen constant. Note that $\|f'\|_\infty$ is used as an approximation to $\mathcal{V}_{Fac(f)}$. Let P_1 be a solution.

- Given P_1, $P_0(e^{j\omega_i})$ and a controller stabilising P_0, solve

$$\min_{P\in\mathcal{R},\, I(P,P_0)=0} \delta_{\mathcal{L}_2}(P_1, P)$$

Let P_2 be the solution. Then by triangle inequality, $\delta_W(P_1, P_2) \leq \max_i \kappa(P_0, P)(e^{j\omega_i}) + \delta_{\mathcal{L}_2}(P_1, P_2)$.

These three steps are discussed in detail in sections 7.3, 7.4 and 7.5 respectively. [1]

7.3 Finding Frequency Response of Normalised Right Graph Symbol

Consider a multivariable plant P_0. The point frequency response samples (possibly matrices) $P_0(e^{j\omega})$ may be obtained by sinusoidal testing in open loop for a stable plant, or in closed loop with a linear, shift-invariant controller for an unstable plant. From $P_0(e^{j\omega})$, we want to find $G_0(e^{j\omega})$ where G_0 is a normalised right graph of P_0. An equivalent approximation problem may be formulated.

Lemma 7.2 *Given* $P_0, P_1 \in \mathcal{RL}_\infty$, $Q, Q^{-1} \in \mathcal{L}_\infty$, $\exists\ \hat{Q}, F$ *such that* $\hat{Q}$, $\hat{Q}^{-1}$, $F \in \mathcal{L}_\infty$,

$$\overline{\sigma}\left(G_0 - G_1 Q\right)(e^{j\omega}) = \overline{\sigma}\left(F - G_1\hat{Q}\right)(e^{j\omega})\ \forall\, \omega$$

and at any ω, *the complex matrix* $F(e^{j\omega})$ *can be written as a function of the point frequency response matrix* $P_0(e^{j\omega})$.

Proof : At any point $e^{j\omega}$ on $\partial\mathbb{D}$, let $P_0(e^{j\omega}) = (U\Sigma V^*)$ be the singular value decomposition of $P_0(e^{j\omega})$, with $\Sigma = diag\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$. Define

$$Y := \left(V(I + \Sigma^T\Sigma)^{\frac{1}{2}}V^*\right) \tag{7.16}$$

Then $(I + P_0^*P_0)(e^{j\omega}) = V(I + \Sigma^T\Sigma)V^* = Y^*Y$. Let X be the outer spectral factor of $(I + P_0^*P_0)$, *i.e.* $X^*X = (I + P_0^*P_0)$, $X, X^{-1} \in \mathcal{H}_\infty$. Then $G_0 := \begin{bmatrix} P_0X^{-1} \\ X^{-1} \end{bmatrix}$ represents the normalised right graph symbol of P_0, within multiplication by a constant unitary matrix. Since $X(e^{j\omega})^*X(e^{j\omega}) = Y^*Y$, and both $X(e^{j\omega})$ and Y are square complex matrices, $X(e^{j\omega}) = \phi_\omega Y$ where ϕ_ω is a square matrix such that $\phi_\omega^*\phi_\omega = I$ ([ZDG96], lemma 2.14). As ϕ_ω at any

[1] The last two steps in this procedure for identification in the $\nu-$gap metric are analogous to the two-step procedure used in untuned worst case methods for identification in $\mathcal{H}_\infty$ norm [HJN91].

point $e^{j\omega}$ on $\partial\mathbb{D}$ is bounded, there exists a function $\psi \in \mathcal{L}_\infty$ such that $\psi(e^{j\omega}) = \phi_\omega$. Further, since $\phi_\omega^*\phi_\omega = I$, ψ must be such that $\psi^{-1} \in \mathcal{L}_\infty$. For a given $Q, Q^{-1} \in \mathcal{L}_\infty$, define $\hat{Q} := Q\psi$ and $F := \begin{bmatrix} P_0X^{-1}\psi \\ X^{-1}\psi \end{bmatrix}$. Then $F, \hat{Q}, \hat{Q}^{-1} \in \mathcal{L}_\infty$ and

$$F(e^{j\omega}) = [(P_0(e^{j\omega})Y^{-1} \;\; Y^{-1}]^T \tag{7.17}$$

is determined solely from $P_0(e^{j\omega})$, with $Y(e^{j\omega})$ as in (7.16). Also,

$$\begin{aligned} \overline{\sigma}\,(F - G_1\hat{Q})(e^{j\omega}) \; &= \; \overline{\sigma}\,((G_0 - G_1Q)\psi)(e^{j\omega}) \\ &= \; \overline{\sigma}\,(G_0 - G_1Q)(e^{j\omega}) \qquad \blacksquare \end{aligned}$$

Given plant frequency response samples $P_0(e^{j\omega_i})$, $F(e^{j\omega_i})$ defined as above may be used to find a model P_1 and a $\hat{Q}$ which minimise $\max_i \overline{\sigma}(F - G_1\hat{Q})(e^{j\omega_i})$ subject to some smoothness constraint on G_1. For a stable SISO plant P_0, note that $F(e^{j\omega_i}) = \left[\frac{P_0(e^{j\omega_i})}{\sqrt{1+|P_0(e^{j\omega_i})|^2}} \;\; \frac{1}{\sqrt{1+|P_0(e^{j\omega_i})|^2}}\right]^T$. Further, G_1 need not be normalised, as the normalising transfer function can be absorbed into $\hat{Q}$. To see this, let $f \in S_{1,2}$, $P_1 = Fac(f)$ and $f^*f > 0$. Let X be an outer spectral factor of f^*f. Then fX^{-1} is a normalised right graph symbol of $Fac(f)$, unique within multiplication by a unitary matrix. For $\hat{Q}$, $\hat{Q}^{-1} \in \mathcal{L}_\infty$ and F as above, $\overline{\sigma}(F - G_1\hat{Q})(e^{j\omega_i}) = \overline{\sigma}(F - fX^{-1}\hat{Q})(e^{j\omega_i}) = \overline{\sigma}(F - f\tilde{Q})(e^{j\omega_i})$, with $\tilde{Q} = X^{-1}\hat{Q}$, $\tilde{Q}, \tilde{Q}^{-1} \in \mathcal{L}_\infty$.

If $f^*f(e^{j\omega}) = 0$ at some $e^{j\omega}$, one can perturb the solution to $f + \delta f \in S_{1,2}$, such that $(f + \delta f)^*(f + \delta f)(e^{j\omega}) > 0$, $\forall e^{j\omega} \in \partial\mathbb{D}$. Since the cost function is continuous in f, a small δf will cause only a small increase in cost.

7.4 Approximation in the $\mathcal{L}_2$-gap

An iterative procedure for approximation in the $\mathcal{L}_2$-gap is given below. SISO case is discussed here for ease of notation. Let S_n and $S_{1,2}$ be defined as in (6.3) and (6.49) respectively.

Given : A vector of frequency response samples P_ω defined as in (7.12).

Initialisation : Set $k = 1$, $\hat{Q}_0 = 1$.

Step A : Solve

$$\min_{f_k \in S_{1,2}} \left(\max_i \overline{\sigma}\,(F - f_k\hat{Q}_{k-1})(e^{j\omega_i}),\, \alpha\|f_k'\|_\infty \right) \tag{7.18}$$

where α is a user chosen constant, $F(e^{j\omega_i})$ is obtained from $P_0(e^{j\omega_i})$ using (7.17) and $f_k' = \frac{df_k}{dz}$. This problem is similar to (6.53). Both the constraints $\overline{\sigma}\,(F - f_k\hat{Q}_{k-1})(e^{j\omega_i}) \leq \lambda$

and $\alpha\|f_k'\|_\infty < \lambda$ may be written as affine constraints in the parameters of f_k, the latter using (6.18) and suitable definitions of matrices A, B, C, D. Thus, (7.18) may be written as an LMI optimisation problem. Let $\hat{f}_k$ be the solution to (7.18).

Step B : Solve

$$\min_{Q_k(e^{j\omega_i})} \max_{i\in[1,m]} \overline{\sigma}\,(F - \hat{f}_k Q_k)(e^{j\omega_i}) \tag{7.19}$$

This step is an unconstrained linear least squares problem, and has a closed form solution:

$$\hat{Q}_k(e^{j\omega_i}) = \frac{\hat{f}_k^*(e^{j\omega_i})F(e^{j\omega_i})}{\hat{f}_k^*(e^{j\omega_i})\hat{f}_k(e^{j\omega_i})} \tag{7.20}$$

If $\max_i \overline{\sigma}(F - \hat{f}_k\hat{Q}_k)(e^{j\omega_i})$ is less than a specified tolerance, stop; otherwise set $k := k+1$ and go back to step A.

If $\hat{f}_k^*\hat{f}_k(e^{j\omega_i}) = 0$ for some $i \in [1,m]$, we can replace $\hat{f}$ in (7.20) by $\hat{f}_k + \delta\hat{f}_k$ for a small perturbation $\delta\hat{f}_k$, such that $(\hat{f}_k + \delta\hat{f}_k)^*(\hat{f}_k + \delta\hat{f}_k) \neq 0$. As the cost function is continuous in $\hat{f}_k$, a small perturbation in $\hat{f}_k$ results in a small increase in cost. This, of course, is of theoretical interest; yielding $\hat{f}_k^*\hat{f}_k(e^{j\omega_i}) = 0$ *exactly* is a situation rather unlikely to occur in practice.

Suppose, α is sufficiently small so that the solution of (7.18) is the same as the solution of $\min_{f_k\in S_{1,2}} \max_i \overline{\sigma}\,(F - f_k\hat{Q}_{k-1})(e^{j\omega_i})$. [2] Define $v_{k,l} := \max_i(F - \hat{f}_k\hat{Q}_l)(e^{j\omega_i})$. Thus $v_{k,k-1}$ and $v_{k,k}$ give the achieved cost at steps A and B respectively at k^{th} iteration. A particularly attractive property of the above procedure is that the achieved cost in our approximation improves with each iteration.

Lemma 7.3 *For $k \geq 1$,*

$$v_{k,k} \leq v_{k,k-1} \leq v_{k-1,k-1}$$

Proof : Consider step A of k^{th} iteration. Since $\hat{f}_{k-1}$ is a feasible solution for (7.18), we have $v_{k,k-1} \leq v_{k-1,k-1}$. Similarly, as $\hat{Q}_{k-1}$ is a feasible solution for (7.19), $v_{k,k} \leq v_{k,k-1}$. ∎

Further, a local optimality-like property may be proven for this procedure. Define $w_{k,j} := (F - \hat{f}_k\hat{Q}_j)(e^{j\omega_i})$. Let $\hat{f}_k = \hat{f}_{k-1} + \lambda\Delta$, where $\lambda \in \mathbb{R}$ and $\Delta \in S_{1,2}$, $\|\Delta\|_\infty = 1$. The following result is based on a similar result in [VM97].

Lemma 7.4 *At any point $e^{j\omega_i}$,*

$$\frac{d}{d\lambda}\{w_{k,k}^* w_{k,k}\}\Big|_{\lambda=0} = \frac{d}{d\lambda}\{w_{k,k-1}^* w_{k,k-1}\}\Big|_{\lambda=0} \tag{7.21}$$

[2] In a number of simulation experiments, this assumption is seen to hold for any $\alpha < 1$, with a sensible choice of n_1 and n_2..

Proof : From (7.20), $w^*_{k,k}\hat{f}_k = 0$. This gives

$$\frac{d}{d\lambda}\{w^*_{k,k}w_{k,k}\} = w^*_{k,k}(-\Delta\hat{Q}_k) + (-\Delta\hat{Q}_k)^* w_{k,k} \tag{7.22}$$

$$\frac{d}{d\lambda}\{w^*_{k,k-1}w_{k,k-1}\} = w^*_{k,k-1}(-\Delta\hat{Q}_{k-1}) + (-\Delta\hat{Q}_{k-1})^* w_{k,k-1} \tag{7.23}$$

The right hand sides of (7.22), (7.23) are equal at $\lambda = 0$. ∎

This means that if there is a descent direction for minimising $\overline{\sigma}(F - f_k Q_k)$ simultaneously over (f_k, Q_k) at a point $(\hat{f}_{k-1}, \hat{Q}_{k-1})$, then usually, there also exists a descent direction at $(\hat{f}_{k-1}, \hat{Q}_{k-1})$ for minimising $\overline{\sigma}(F - f_k\hat{Q}_{k-1})$ over f_k alone.

7.5 Satisfying the Winding Number Constraint

Let P_0 represent the true plant. Let $P_1 = f_{1,k}f_{2,k}^{-1}$, where $\hat{f}_k = [f_{1,k}\, f_{2,k}]^T$ is a solution to (7.18) at k^{th} iteration. Then the above procedure ensures that $\max_i\, \kappa(P_1, P_0)(e^{j\omega_i}) < v_{k,k}$, but it doesn't guarantee that $I(P_1, P_0) = 0$. This means that P_1 may not be stabilised by a controller C, given that C stabilises P_0.

Suppose, a controller C which stabilises the true plant P_0 results in $\eta\,(H(P_1, C)) = k$ and $b_{\mathcal{L}_2}(P_1, C) = \alpha$ for a model P_1. From lemma 7.1, if another model P_2 satisfies $I(P_2, P_1) \leq -k$ and $\delta_{\mathcal{L}_2}(P_1, P_2) < \alpha$, then $\eta\,(H(P_2, C)) = 0$. A method for finding P_2 which guarantees $I(P_2, P_1) \leq -k$ is outlined in the next theorem.

Theorem 7.1 *Given P_1 and $\gamma < 1$, the following are equivalent.*

(i). $\exists\ P_2 : I(P_2, P_1) \leq -k$ *and* $\delta_{\mathcal{L}_2}(P_1, P_2) < \gamma$

(ii). $\gamma > \sqrt{1 - \sigma_k^2(\mathbf{H}_{G_1})}$

where $\sigma_k(\mathbf{H}_G)$ is the k^{th} largest singular value of Hankel operator $\mathbf{H}_G$.

When the condition (ii) is satisfied, such a P_2 may be constructed as $P_2 = \{Fac(X)\}^*$ where $X = [x_1\ x_2]^T$ satisfies $x_2^{-1} \in \mathcal{R}$, $X \in \mathcal{H}_{\infty(-),n-k}$ and $\| G_3 - X \|_\infty < \gamma$. Here G_3 is the normalised right graph symbol for $P_3 := P_1^*$, deg$(P_1) = n$ and $\mathcal{H}_{\infty(-),n-k}$ denotes the space of functions analytic outside the unit disk except for at most $n - k$ stable poles.

Proof : See Appendix B.

As illustrated in the proof of this result, one can construct a P_2 which solves

$$\inf_{\substack{P_2 \in \mathcal{R} \\ I(P_2, P_1) \leq -k}} \delta_{\mathcal{L}_2}(P_1, P_2) \tag{7.24}$$

Then by triangle inequality,

$$\max_i \kappa(P_0, P_2)(e^{j\omega_i}) \leq v_{k,k} + \delta_{\mathcal{L}_2}(P_1, P_2) \tag{7.25}$$

Also, from (7.8), $\eta\left(H(P_2, C)\right) = 0$ provided $b_{\mathcal{L}_2}(P_1, C) > \delta_{\mathcal{L}_2}(P_1, P_2)$. Since C stabilises P_0, again applying (7.8) for P_2, P_0, C gives

$$I(P_2, P_0) = 0 \tag{7.26}$$

provided $b_{\mathcal{L}_2}(P_2, C) > \delta_{\mathcal{L}_2}(P_0, P_2)$. $\delta_{\mathcal{L}_2}(P_0,\, P_2)$, in turn, may be approximated by $\max_i\, \kappa(P_0, P_2)(e^{j\omega_i})$. If a 'good' chordal distance fit is obtained over a sensible grid of frequencies, this condition is easy to satisfy; as the simulation example in section 7.6 demonstrates. Finally (7.25), (7.26) combine to give

$$\delta_W(P_0, P_2) \leq v_{k,k} + \delta_{\mathcal{L}_2}(P_1, P_2)$$

If both P_0 and P_1 are stable (*i.e.* stabilised by $C = 0$), $I(P_1, P_0) = 0$ provided $b_{\mathcal{L}_2}(P_1, 0) > \delta_{\mathcal{L}_2}(P_1, P_0)$. This is a very simple condition to check approximately, and does not require knowledge of any other controller. If the plant is unstable, a stabilising controller is needed to build a model P_2. This is a reasonable requirement; a controller would be needed to collect any operational data anyway.

A new, related algorithm for identification of normalised coprime factors appears in [Gu99]. This algorithm is based on discrete Fourier Analysis and is robustly convergent. On the other hand, it relies on uniform frequency spacing and does not cater for the important winding number correction.

7.6 Simulation Example

Consider a continuous time plant

$$P(s) = \frac{1.875s^2 + 18.75s + 48.75}{s^4 + 6.1s^3 + 16.4845s^2 + 38.431s + 60.197}$$

This plant has two pairs of lightly damped resonant poles and a pair of resonant zeroes. It was mapped to z-plane by Tustin transformation, with a sampling period 0.5 sec. 20 frequency response samples are used; 16 uniformly spaced between 0 and $\frac{\pi}{2}$ and 4 uniformly spaced between $\{\frac{\pi}{2}, \frac{7\pi}{8}\}$.

This data is used for $\mathcal{L}_2$ -gap approximation, with $n_1 = 5$, $n_2 = 5$ and $\alpha = 0.001$ in (7.18). After 12 iterations, the approximation P_1 obtained gives $b_{\mathcal{L}_2}(P_1, 0) = 0.58$,

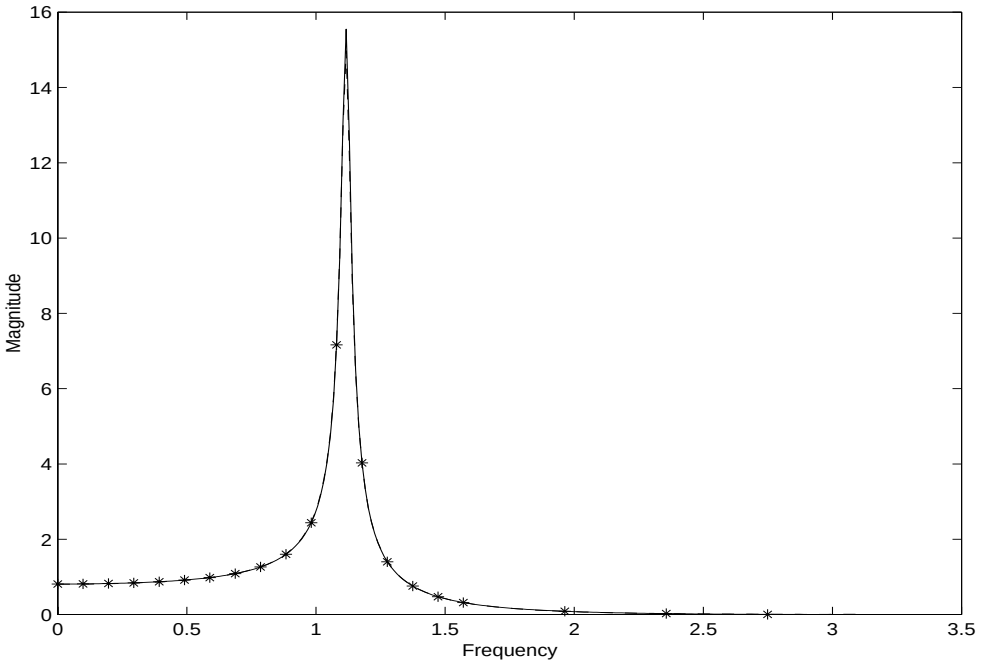


Figure 7.1: Magnitude Plots: True plant and $\nu-$gap Approximation

$\max_i \kappa(P_1, P_0)(e^{j\omega_i}) = 0.0031$ and P_1 is stable. Hence the Hankel norm approximation step as in theorem 7.1 is deemed unnecessary and $\delta_W(P_1, P_0) \approx 0.0031$. True $\delta_\nu(P_1, P_0)$ in this case is 0.008. In fig. (7.1), the solid line shows the true plant response, '$\star$' indicating the frequency response samples. The dashed line indicates the response of the $\nu-$gap approximation.

As another example, the same data is used for $\nu-$gap approximation with a lower order plant, with $n_1 = 3$, $n_2 = 3$ and $\alpha = 0.001$. After 12 iterations, the worst case chordal distance obtained is 0.0606; with true $\nu-$gap distance 0.0607. The magnitude of true plant (solid) and the model (dashed) frequency response is shown in fig. (7.2). This very good result for a lower order model is not entirely unexpected. Using the results for optimal $\nu-$gap approximation from ([Vin], chapter 8), it may be shown that $0.0603 \leq \inf_{\hat{P}:deg(\hat{P})\leq 2} \delta_\nu(P, \hat{P}) \leq 0.0616$. Our simulation results based on finite data conform to these bounds.

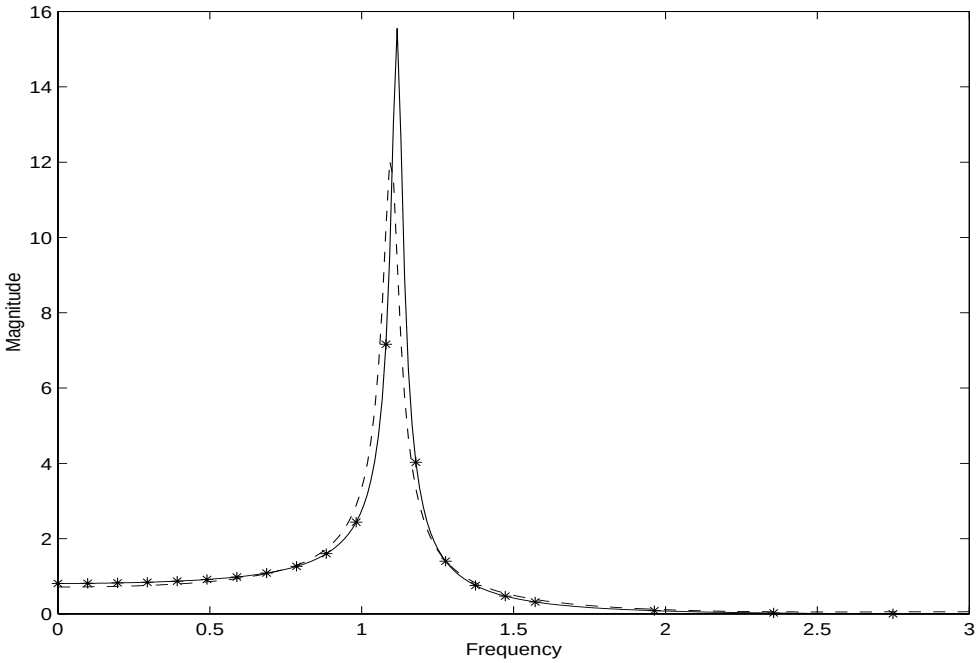


Figure 7.2: Magnitude Plots: True plant and lower order $\nu-$gap Approximation

Chapter 8

Concluding Remarks

8.1 Main Contributions

Here we summarise the main contributions of this dissertation.

- A new robustly convergent algorithm is suggested for point frequency response identification in chapter 4. A deterministic characterisation of white noise is introduced, which is potentially less conservative than a 'ball-in-l_∞' type description and is shown to capture rigorously the notion of uncorrelated nature of noise in a deterministic sense. The identification algorithm is also extended to identification of multiple point frequency response samples.

- In chapter 5, a family of new worst case identification algorithms is proposed which maps a given set of (possibly noisy) point frequency response samples into a rational model from a user specified model set. These algorithms introduce the idea of a trade-off between the worst case fit obtained and the worst case slope of the frequency response of the model.

- The idea of this trade-off is further developed in chapter 6 to introduce a robustly convergent algorithm for identification with FIR models. This algorithm is shown to be optimal, in a particular sense, for a finite model order. Various further modifications are suggested to improve the quality of the obtained model (in terms of its slope) and to incorporate the knowledge of poles of the plant.

- A small $\nu-$gap between the plant and the model guarantees that a controller achieving a high robustness margin $b_{P,C}$ for the model will stabilise the true plant without a signif-

icant degradation in performance. Chapter 7 uses the numerical techniques developed in the previous chapter to formulate and solve an important problem for identification in the $\nu-$gap metric. A rational model is obtained along-with an approximate bound on the $\nu-$gap error.

8.2 Recommendations for Future Research

- The algorithm for identification in the $\nu-$gap metric proposed in chapter 7 is not robustly convergent. As seen in section 7.2, it is better to constrain $\mathcal{V}_i$ rather than constraining $\|f'\|_\infty$ to control the worst case rate of change of chordal distance. The problem of constraining the complexity measure $\mathcal{V}_i$ as defined in section 7.2 is still open.

- Chapter 3 reviewed classical time domain identification approach. Classical methods obtain a model that closely approximates the *output* of plant for the *same* input. In the light of the definition (2.9) of directed gap and its relation to the $\nu-$gap metric, such a model may not be the best model from a feedback perspective. The problem of rigorous assessment (and possibly, of a modification) of classical time domain methods from a deterministic, feedback point of view yet remains to be addressed.

- Another useful direction for research will be to develop a deterministic characterisation of noise, similar to the one introduced in chapter 4, for identification in the $\nu-$gap metric. Such deterministic characterisations may also be useful to study finite sample properties in time domain identification.

Appendix A

Appendix to chapter 4

To prove theorem 4.1, two technical results are needed. The first is used to bound the sample mean of bounded independent random variables [Hoe63]:

Theorem A.1 (Hoeffding's theorem) *Let $z_0, z_1, \ldots, z_{N-1}$ be independent random variables bounded by $a \leq z_i \leq b$ and having finite first and second central moments. Define $\overline{z} = \frac{1}{N}\sum_{N=0}^{N-1} z_N$ and $\mu = E(\overline{z})$. Then for $\delta > 0$,*

$$P(\overline{z} - \mu > \delta) \leq e^{\frac{-2N\delta^2}{(b-a)^2}}$$

To apply Hoeffding's inequality to $R_v(\tau)$, we need to separate $R_v(\tau)$ into sums of independent random variables. The following result, by Paganini [Pag95], allows us to do this:

Lemma A.1 *(***Paganini)** *Let $N \geq 3$, and $v_0, v_1, \ldots, v_{N-1}$ be independent random variables. Fix $1 \leq \tau < N$. Then $R_v(\tau)$ may be expressed as $R_v(\tau) = \Sigma_1 + \Sigma_2 + \Sigma_3$, where each Σ_i is a sum of N_i independent random variables, and $N_i \geq \frac{N}{5}$.*

To illustrate what this result means, let $N = 6$ and let $v_0, v_1, \ldots, v_5$ be independent random variables. Consider $R_v(\tau)$ for $\tau = 1$:

$$\begin{aligned} R_v(1) &= v_0v_1 + v_1v_2 + v_2v_3 + v_3v_4 + v_4v_5 + v_5v_1 \\ &= (v_0v_1 + v_2v_3 + v_4v_5) + (v_1v_2 + v_3v_4) + (v_1v_5) \end{aligned}$$

The term in each bracket is a sum of independent random variables. The above lemma says that such a grouping of $R_v(\tau)$ into three summations of independent random variables is possible for any $\tau \in [1, N-1]$ and for any N. Further, the number of terms in each of the three summations is no smaller than $\frac{N}{5}$.

A.1 Proof of theorem 4.1

(i). Argument in this proof is along the lines of the proof of theorem 4.3 in [Pag95]. The basic idea is as follows:

- Given a function σ_N satisfying (4.11), express the complement of $V_{\sigma_N,\eta,N}$ as a subset of a specific, countable union of sets.
- Show that the probability of a random vector v belonging to this union of sets goes to zero, as $N \to \infty$.

The complement of $V_{\sigma_N,\eta,N}$ can be written as

$$V^c_{\sigma_N,\eta,N} = \{v \in \mathbb{R}^N : \frac{1}{N} R_v(0) > \sigma_N^2\} \cup \{v \in \mathbb{R}^N : \max_{1 \le \tau \le N-1} \frac{1}{N} |R_v(\tau)| > \eta\, \sigma_N^2\} \quad \text{(A.1)}$$

Now we show that the probability of each of the two sets (whose union includes $V^c_{\sigma_N,\eta,N}$) goes to zero as $N \to \infty$.

For $N > N_0$, let $\sigma_N^2 = \lambda^2 + \alpha_N^2$, $\alpha_N > 0$ from assumption (4.11). Now $|v_i|^2$ are independent random variables with identical mean λ^2 and bounded in $[0, K^2]$. Hence

$$\begin{aligned} \mathcal{P}\left(\frac{1}{N} R_v(0) > \sigma_N^2\right) &= \mathcal{P}\left(\frac{1}{N}\sum_{i=0}^{N-1} |v_i|^2 > \sigma_N^2\right) \\ &= \mathcal{P}\left(\frac{1}{N}\sum_{i=0}^{N-1} |v_i|^2 - \lambda^2 > \alpha_N^2\right) \end{aligned} \quad \text{(A.2)}$$

This probability may be bounded using Hoeffding's inequality, so that

$$\mathcal{P}\left(\frac{1}{N} R_v(0) > \sigma_N^2\right) \le e^{\left(\frac{-2N\alpha_N^4}{K^4}\right)} \overset{N\to\infty}{\longrightarrow} 0 \quad \text{(A.3)}$$

Next, using Paganini's lemma, $R_v(\tau) = \Sigma_1 + \Sigma_2 + \Sigma_3$ where each Σ_i is a sum of N_i independent random variables bounded in $[-K^2, K^2]$ and having mean 0 (As v_i, v_k are independent and therefore uncorrelated, $E\,(\,v_i v_k\,) = E\,(\,v_i)\;E\;(v_k\,)\,$). Hence

$$\mathcal{P}\left(\frac{R_v(\tau)}{N} > \eta\, \sigma_N^2\right) \le \sum_{i=1}^{3} \mathcal{P}\left(\frac{\Sigma_i}{N_i} > \eta\, \sigma_N^2\right) \quad \text{(A.4)}$$

Again using Hoeffding's inequality for $N > N_0$ and using $N_i \ge \frac{N}{5}$,

$$\mathcal{P}\left(\frac{R_v(\tau)}{N} > \eta\, \sigma_N^2\right) \le \sum_{i=0}^{3} e^{\frac{-N_i \eta^2 \sigma_N^4}{2K^4}} \le 3e^{\frac{-N\eta^2\sigma_N^4}{10K^4}} \quad \text{(A.5)}$$

Now $-R_v(\tau)$ can also be split into three summations of zero mean, independent random variables bounded in $[-K^2, K^2]$ in a similar fashion. Hence,
$\mathcal{P}\left(-\frac{R_v(\tau)}{N} > \eta\,\sigma_N^2\right) \leq 3e^{\frac{-N\eta^2\sigma_N^4}{10K^4}}$ also holds. Since

$$\mathcal{P}\left(\frac{|R_v(\tau)|}{N} > \eta\,\sigma_N^2\right) = \mathcal{P}\left(\frac{R_v(\tau)}{N} > \eta\,\sigma_N^2\right) + \mathcal{P}\left(-\frac{R_v(\tau)}{N} > \eta\,\sigma_N^2\right)$$

it follows that

$$\mathcal{P}\left(\frac{|R_v(\tau)|}{N} > \eta\,\sigma_N^2\right) \leq 6e^{\frac{-N\eta^2\sigma_N^4}{10K^4}} \tag{A.6}$$

Bounding each $\mathcal{P}\left(\frac{|R_v(\tau)|}{N} > \eta\,\sigma_N^2\right)$ in this fashion and using the fact that

$$\left\{v \in \mathbb{R}^N : \max_{1\leq\tau\leq N-1} \frac{1}{N}|R_v(\tau)| > \eta\sigma_N^2\right\} \subseteq \bigcup_{\tau=1}^{N-1} \left\{v \in \mathbb{R}^N : \frac{1}{N}|R_v(\tau)| > \eta\,\sigma_N^2\right\}$$

we get

$$\mathcal{P}\left(\max_{1\leq\tau\leq N-1} \frac{|R_v(\tau)|}{N} > \eta\,\sigma_N^2\right) \leq 6Ne^{\frac{-N\eta^2\sigma_N^4}{10K^4}} = 6e^{\left(logN - \frac{N\eta^2\sigma_N^4}{10K^4}\right)} \overset{N\to\infty}{\longrightarrow} 0 \tag{A.7}$$

From (A.1), (A.3) and (A.7), it follows that $\mathcal{P}(v_N \in V^c_{\sigma_N,\eta,N}) \overset{N\to\infty}{\longrightarrow} 0$.

(ii). The proof of this part proceeds in two steps. Let $r = N^\alpha$ where $\alpha > 1$ is an arbitrary integer. For a given r, define a set of uniformly spaced angular frequencies

$$\Omega = \left\{0, \frac{\pi}{r}, \frac{2\pi}{r}, \ldots, \frac{(r-1)\pi}{r}\right\}$$

Next, define a set $\tilde{V}^\omega_{\sigma_N,\eta,N}$ by

$$\tilde{V}^\omega_{\sigma_N,\eta,N} := \left\{v \in \mathbb{R}^N : \frac{1}{N}R_v(0) \leq \sigma_N^2,\ \sup_{\omega\in\Omega} \frac{1}{N}\left|\sum_{k=0}^{N-1} v_k\, e^{j\omega k}\right| \leq \eta\,\sigma_N\right\} \tag{A.8}$$

First, we establish a bound on $\mathcal{P}\left(v \in V^\omega_{\sigma_N,\eta,N}\right)$ in terms of $\mathcal{P}\left(v \in \tilde{V}^\omega_{\sigma_N,\eta,N}\right)$. Then $\mathcal{P}\left(v \in \left\{\tilde{V}^\omega_{\sigma_N,\eta,N}\right\}^c\right)$ is bounded using Hoeffding's inequality.

- The first step in this part is along the lines of a similar proof in [VD]. Note that

$$\sup_{\omega\in[0,\pi)} \left|\frac{d}{d\omega} \sum_{k=0}^{N-1} v_k\, e^{j\omega k}\right| \leq N \sup_{\omega\in[0,\pi)} \left|\sum_{k=0}^{N-1} v_k\, e^{j\omega k}\right| \tag{A.9}$$

  A proof may be found in ([Pin85], chapter 8). Suppose, $\sup_{\omega\in[0,\pi)} \left|\sum_{k=0}^{N-1} v_k\, e^{j\omega k}\right|$ is achieved at $\omega = \omega_0$. Using mean value theorem ([Rud76], chapter 5) , we get

$$\left|\sum_{k=0}^{N-1} v_k\, e^{j\omega_0 k}\right| \leq \left|\sum_{k=0}^{N-1} v_k\, e^{j\omega k}\right| + |\omega - \omega_0|N \left|\sum_{k=0}^{N-1} v_k\, e^{j\omega_0 k}\right| \quad \forall\omega \in [0,\pi) \tag{A.10}$$

Choose $\omega \in \Omega$, $|\omega - \omega_0| \leq \frac{\pi}{r}$ and note that

$$\sup_{\omega \in [0,\pi)} \left| \sum_{k=0}^{N-1} v_k \, e^{j\omega_0 k} \right| \leq \sup_{\omega \in \Omega} \left| \sum_{k=0}^{N-1} w_k \, e^{j\omega k} \right| \tag{A.11}$$

where w is defined by

$$w = \frac{1}{1 - \frac{N\pi}{r}} v \tag{A.12}$$

Now w_k are independent, zero mean random variables bounded in $\left[-\frac{K}{1-\frac{N\pi}{r}}, \frac{K}{1-\frac{N\pi}{r}} \right]$ and having a variance $E(|w_k|^2) = \frac{\lambda^2}{(1-\frac{N\pi}{r})^2}$. To avoid defining yet another symbol, suppose for some $N_0 \in \mathbb{Z}_+$,

$$\sigma_N^2 > E(|w_k|^2) \quad \forall \, N > N_0 \tag{A.13}$$

Using the definitions of $\tilde{V}^{\omega}_{\sigma_N,\eta,N}$ and $V^{\omega}_{\sigma_N,\eta,N}$, it follows from (A.11) that

$$\mathcal{P}\left(v \in V^{\omega}_{\sigma_N,\eta,N}\right) \geq \mathcal{P}\left(w \in \tilde{V}^{\omega}_{\sigma_N,\eta,N}\right) \tag{A.14}$$

- This step follows on similar lines as in part (i) of the proof. The aim is to show that

$$\mathcal{P}\left(w \in \{\tilde{V}^{\omega}_{\sigma_N,\eta,N}\}^c\right) \overset{N\to\infty}{\longrightarrow} 0$$

with w defined as in (A.12). The complement of $\tilde{V}^{\omega}_{\sigma_N,\eta,N}$ can be written as

$$\{\tilde{V}^{\omega}_{\sigma_N,\eta,N}\}^c = \{v \in \mathbb{R}^N : \frac{1}{N} R_v(0) > \sigma_N^2\} \cup \left\{ v \in \mathbb{R}^N : \sup_{\omega \in \Omega} \frac{1}{N} \left| \sum_{k=0}^{N-1} v_k \, e^{j\omega k} \right| > \eta \, \sigma_N \right\} \tag{A.15}$$

For convenience of notation, let

$$\begin{aligned} X_{cos} &= \left\{ v \in \mathbb{R}^N : \sup_{\omega \in \Omega} \left(\frac{1}{N} | \sum_{k=0}^{N-1} v_k \, \cos \omega k \, | > \frac{\eta \, \sigma_N}{\sqrt{2}} \right) \right\} \\ X_{sin} &= \left\{ v \in \mathbb{R}^N : \sup_{\omega \in \Omega} \left(\frac{1}{N} | \sum_{k=0}^{N-1} v_k \, \sin \omega k \, | > \frac{\eta \, \sigma_N}{\sqrt{2}} \right) \right\} \end{aligned} \tag{A.16}$$

Then clearly,

$$\{\tilde{V}^{\omega}_{\sigma_N,\eta,N}\}^c \subseteq \{v \in \mathbb{R}^N : \frac{1}{N} R_v(0) > \sigma_N^2\} \cup X_{cos} \cup X_{sin} \tag{A.17}$$

In a manner similar to (A.3), the probability $\mathcal{P}\left(w \in \{v \in \mathbb{R}^N : \frac{1}{N} R_v(0) > \sigma_N^2\}\right)$ in (A.17) can be bounded using (A.13) and can be shown to go to 0 asymptotically

as $N \to \infty$. To bound the probability $\mathcal{P}(w \in X_{cos})$ note that $w_k cos\, \omega_i k$ are independent random variables and $E\left(\frac{1}{N}\sum_{k=0}^{N-1} w_k \cos \omega_i k\right) = 0$ due to linearity of expectation operator. Now

$$\begin{aligned}\mathcal{P}\left(\frac{1}{N}\left|\sum_{k=0}^{N-1} w_k \cos \omega_i k\right| > \frac{\eta\sigma_N}{\sqrt{2}}\right) &= \mathcal{P}\left(\frac{1}{N}\sum_{k=0}^{N-1} w_k \cos \omega_i k > \frac{\eta\sigma_N}{\sqrt{2}}\right) \\ &\quad + \mathcal{P}\left(-\frac{1}{N}\sum_{k=0}^{N-1} w_k \cos \omega_i k > \frac{\eta\sigma_N}{\sqrt{2}}\right) \\ &\leq 2e^{-\frac{N\eta^2\sigma_N^2}{4K^2}(1-\frac{N\pi}{r})^2} \end{aligned} \tag{A.18}$$

where the last step follows from Hoeffding's theorem. The probability $\mathcal{P}\left(\frac{1}{N}\left|\sum_{k=0}^{N-1} w_k \sin \omega_i k\right| > \frac{\eta\sigma_N}{\sqrt{2}}\right)$ can be similarly bounded. Adding up these probabilities over r frequencies, we have

$$\mathcal{P}(w \in X_{cos} \cup X_{sin}) \leq 4re^{-\frac{N\eta^2\sigma_N^2}{4K^2}(1-\frac{N\pi}{r})^2} \leq e^{-\left(\frac{N\eta^2\sigma_N^2}{4K^2}(1-\frac{N\pi}{r})^2 - \log(4r)\right)} \tag{A.19}$$

Since $r = N^\alpha$, it follows that

$$\mathcal{P}(w \in X_{cos} \cup X_{sin}) \overset{N\to\infty}{\longrightarrow} 0 \tag{A.20}$$

From (A.14), (A.17), (A.3) and (A.20), the result

$$\mathcal{P}\left(v \in \left\{V^{\omega}_{\sigma_N,\eta,N}\right\}^c\right) \overset{N\to\infty}{\longrightarrow} 0$$

follows. ∎

A.2 Proof of theorem 4.2

Suppose $\omega_0 \neq 0$. The output at time k may be written as

$$y_k = \alpha \sum_{q=0}^{k} p_q \cos \omega_0 (k-q) + v_k \tag{A.21}$$

Thus

$$\begin{aligned} P_{\omega_0} &= \frac{2}{N}\sum_{k=0}^{N-1}\left(\sum_{q=0}^{k} p_q \cos\omega_0(k-q)\right)e^{j\omega_0 k} + \frac{2}{\alpha N}\sum_{k=0}^{N-1} v_k e^{j\omega_0 k} \\ &= \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{q=0}^{k} p_q e^{j\omega_0 q}\right) + \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{q=0}^{k} p_q e^{j(2\omega_0 k - \omega_0 q)}\right) + \frac{2}{\alpha N}\sum_{k=0}^{N-1} v_k e^{j\omega_0 k} \end{aligned}$$

(using standard trigonometric identities to expand $\cos\omega_0(k-q)e^{j\omega_0 k}$)

$$
\begin{aligned}
&= \sum_{k=0}^{N-1} p_k e^{j\omega_0 k} - \frac{1}{N}\sum_{k=0}^{N-1} k p_k e^{j\omega_0 k} + \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{q=0}^{k} p_q e^{j(2\omega_0 k-\omega_0 q)}\right) \\
&+ \frac{2}{\alpha N}\sum_{k=0}^{N-1} v_k e^{j\omega_0 k} \\
&= P(e^{j\omega_0}) - \sum_{k=N}^{\infty} p_k e^{j\omega_0 k} - \frac{1}{N}\sum_{k=0}^{N-1} k p_k e^{j\omega_0 k} \\
&+ \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{q=0}^{k} p_q e^{j(2\omega_0 k-\omega_0 q)}\right) + \frac{2}{\alpha N}\sum_{k=0}^{N-1} v_k e^{j\omega_0 k}
\end{aligned}
$$

which gives

$$
\begin{aligned}
|P_{\omega_0} - P(e^{j\omega_0})| \le \sum_{k=N}^{\infty}|p_k| + \frac{1}{N}\sum_{k=0}^{N-1} k|p_k| + \frac{1}{N}\sum_{k=0}^{N-1}\left|\left(\sum_{q=0}^{k} p_q e^{j(2\omega_0 k-\omega_0 q)}\right)\right| \\
+ \frac{2}{\alpha N}\left|\sum_{k=0}^{N-1} v_k e^{j\omega_0 k}\right|
\end{aligned} \tag{A.22}
$$

Using (2.4), the first term may be bounded as

$$
\sum_{k=N}^{\infty}|p_k| \le \sum_{k=N}^{\infty}\gamma\rho^{-k} = \gamma\frac{\rho^{-N}}{1-\rho^{-1}} \tag{A.23}
$$

To bound the second term, note that for $|x| < 1$,

$$
\begin{aligned}
\sum_{k=0}^{N-1} k x^k &= x\sum_{k=0}^{N-1} k x^{k-1} = x\sum_{k=0}^{N-1}\frac{d}{dx}x^k = x\frac{d}{dx}\sum_{k=0}^{N-1} x^k \\
&= x\frac{d}{dx}\left(\frac{1-x^N}{1-x}\right) = \frac{(1-x^N)x-(1-x)Nx^N}{(1-x)^2}
\end{aligned} \tag{A.24}
$$

Substituting $x = \rho^{-1}$ gives

$$
\frac{1}{N}\sum_{k=0}^{N-1} k|p_k| \le \frac{\gamma(1-\rho^{-N})\rho^{-1}}{N(1-\rho^{-1})^2} - \frac{\gamma\rho^{-N}}{1-\rho^{-1}} \tag{A.25}
$$

Next, rewrite the third term as

$$
\frac{1}{N}\left|\sum_{k=0}^{N-1} e^{j2\omega_0 k}\left(\sum_{q=0}^{k} p_q e^{j\omega_0 q}\right)\right| = \frac{1}{N}\left|\sum_{k=0}^{N-1} p_k e^{j\omega_0 k}\left(\sum_{q=k}^{N-1} e^{j(2\omega_0 q)}\right)\right| \tag{A.26}
$$

The following identity is useful to bound the trigonometric summation [Jef94]:

Fact A.1 *For $x \neq 2q\pi$,*

$$\begin{aligned}\sum_{k=1}^{N}\cos(kx) &= \frac{\sin\frac{Nx}{2}}{\sin\frac{x}{2}}\cos\left(\frac{N+1}{2}x\right)\\ \sum_{k=1}^{N}\sin(kx) &= \frac{\sin\frac{Nx}{2}}{\sin\frac{x}{2}}\sin\left(\frac{N+1}{2}x\right)\end{aligned} \tag{A.27}$$

Here,

$$\begin{aligned}\frac{1}{N}\left|\sum_{k=0}^{N-1}p_k e^{j\omega_0 k}\left(\sum_{q=k}^{N-1}e^{j(2\omega_0 q)}\right)\right| &= \frac{1}{N}\left|\sum_{k=0}^{N-1}p_k e^{j\omega_0 k}\left(\sum_{q=1}^{N-k}e^{j2\omega_0(q+k-1)}\right)\right|\\ &= \frac{1}{N}\left|\sum_{k=0}^{N-1}p_k e^{j\omega_0 k}e^{j2\omega_0(k-1)}\left(\sum_{q=1}^{N-k}e^{j2\omega_0 q}\right)\right|\\ &= \frac{1}{N}\left|\sum_{k=0}^{N-1}p_k e^{j\omega_0(3k-2)}\left(\frac{\sin(N-k)\omega_0}{\sin\omega_0}e^{j\omega_0(N-k+1)}\right)\right| \qquad \text{(A.28)}\\ &\leq \frac{1}{N\sin\omega_0}\sum_{k=0}^{N-1}|p_k| \leq \frac{\gamma}{N\sin\omega_0}\frac{1-\rho^{-N}}{1-\rho^{-1}} \qquad \text{(A.29)}\end{aligned}$$

To bound the fourth term when $v \in V_{\sigma_N,\eta,N}$,

$$\left|\sum_{k=0}^{N-1}v_k e^{j\omega_0 k}\right| = \left(\sum_{k=0}^{N-1}\sum_{t=0}^{N-1}v_k\, v_t\, e^{j\omega_0(k-t)}\right)^{\frac{1}{2}} \tag{A.30}$$

Now

$$\begin{aligned}\sum_{k=0}^{N-1}\sum_{t=0}^{N-1}v_k\, v_t\, e^{j\omega_0(k-t)} &= \sum_{k=0}^{N-1}v_k^2 + \sum_{k=0}^{N-2}e^{j\omega_0}\, v_k\, v_{k+1} + \sum_{k=0}^{N-3}e^{2j\omega_0}\, v_k\, v_{k+2} + \cdots +\\ &\quad e^{j\omega_0(N-1)}\, v_0\, v_{N-1} + \sum_{k=0}^{N-2}e^{-j\omega_0}\, v_k\, v_{k+1} + \sum_{k=0}^{N-3}e^{-2j\omega_0}\, v_k\, v_{k+2} +\\ &\quad \cdots + e^{-j\omega_0(N-1)}\, v_0\, v_{N-1}\\ &= R_v(0) + \sum_{p=1}^{N-1}e^{j\omega_0 p}\left\{\sum_{k=0}^{N-1-p}v_k v_{k+p} + e^{-j\omega_0 N}\sum_{k=N-p}^{N-1}v_k v_{k+p}\right\}\\ &= \sum_{p=0}^{N-1}e^{j\omega_0 p}R_v(p) + (e^{-j\omega_0 N}-1)\sum_{p=1}^{N-1}e^{j\omega_0 p}\sum_{k=N-p}^{N-1}v_k v_{k+p} \qquad \text{(A.31)}\end{aligned}$$

where $k+p$ is a modulo-N summation. Using (A.31) and the fact that

$$\sum_{k=N-p}^{N-1}v_k\, v_{k+p} \leq \sum_{k=N-p}^{N-1}\frac{v_k^2+v_{k+p}^2}{2} \quad \text{we get}$$

$$\frac{1}{N}\left|\sum_{k=0}^{N-1} v_k e^{j\omega_0 k}\right| \leq \frac{1}{N}\left\{\sum_{p=0}^{N-1}|R_v(p)| + \frac{|e^{-j\omega_0 N}-1|}{2}\sum_{p=1}^{N-1} R_v(0)\right\}^{\frac{1}{2}} \tag{A.32}$$

$$\leq \frac{1}{N}\left\{N\sigma_N^2 + (N-1)\eta N\sigma_N^2 + \left|\sin\frac{\omega_0 N}{2}\right|(N-1)N\sigma_N^2\right\}^{\frac{1}{2}}$$

$$\leq \sigma_N\sqrt{\eta + \frac{1-\eta}{N} + \left(1-\frac{1}{N}\right)\left|\sin\frac{\omega_0 N}{2}\right|} \tag{A.33}$$

When $v \in V^{\omega}_{\sigma_N,\eta,N}$, the fourth term in (A.22) is bounded directly from the definition of $V^{\omega}_{\sigma_N,\eta,N}$ as

$$\frac{2}{\alpha N}\left|\sum_{k=0}^{N-1} v_k e^{j\omega_0 k}\right| \leq 2\frac{\eta\,\sigma_N}{\alpha} \tag{A.34}$$

Substituting from (A.23), (A.25), (A.29), (A.33) and (A.34) in (A.22) yields (4.13) - (4.14).

For $\omega_0 = 0$, we have

$$P_{\omega_0} = \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{q=0}^{k} p_q\right) + \frac{1}{\alpha N}\sum_{k=0}^{N-1} v_k$$

$$= \sum_{k=0}^{N-1} p_k - \frac{1}{N}\sum_{k=0}^{N-1} k p_k + \frac{1}{\alpha N}\sum_{k=0}^{N-1} v_k$$

$$= P(e^{j\omega_0}) - \sum_{k=N}^{\infty} p_k - \frac{1}{N}\sum_{k=0}^{N-1} k p_k + \frac{1}{\alpha N}\sum_{k=0}^{N-1} v_k$$

$$\text{which gives} \quad |P_{\omega_0} - P(e^{j\omega_0})| \leq \sum_{k=N}^{\infty}|p_k| + \frac{1}{N}\sum_{k=0}^{N-1} k|p_k| + \frac{1}{\alpha N}\left|\sum_{k=0}^{N-1} v_k\right|$$

Bounding the three terms individually as before yields the necessary error bound. ∎

A.3 Proof of theorem 4.3

The output at time k may be written as

$$y_k = \alpha_i \sum_{q=0}^{k} p_q \cos\omega_i(k-q) + \sum_{\substack{r=1\\ r\neq i}}^{m}\left(\sum_{q=0}^{k}\alpha_r p_q \cos\omega_r(k-q)\right) + v_k \tag{A.35}$$

Following the same analysis as in the proof of theorem 4.2 results in

$$|P_{\omega_i} - P(e^{j\omega_i})| \leq \sum_{k=N}^{\infty}|p_k| + \frac{1}{N}\sum_{k=0}^{N-1} k|p_k| + \frac{1}{N}\sum_{k=0}^{N-1}\left|\left(\sum_{q=0}^{k} p_q e^{j(2\omega_i k - \omega_i q)}\right)\right|$$

$$+ \frac{2}{\alpha_i N}\left|\sum_{k=0}^{N-1} v_k e^{j\omega_i k}\right| + \frac{2}{\alpha_i N}\left|\sum_{\substack{r=1\\ r\neq i}}^{m}\sum_{k=0}^{N-1}\left(\sum_{q=0}^{k}\alpha_r p_q \cos\omega_r(k-q)\right)e^{j\omega_i k}\right| \tag{A.36}$$

The first four terms are bounded in a similar fashion as in the earlier case. To bound the last term, consider any ω_s, $s \neq i$.

$$\begin{aligned}
&\frac{2}{\alpha_i N}\sum_{k=0}^{N-1}\left(\sum_{q=0}^{k}\alpha_s p_q \cos\omega_s(k-q)\right)e^{j\omega_i k} = \frac{1}{\alpha_i N}\sum_{k=0}^{N-1}e^{j(\omega_i+\omega_s)k}\left(\sum_{q=0}^{k}\alpha_s p_q e^{-j\omega_s q}\right)\\
&+\frac{1}{\alpha_i N}\sum_{k=0}^{N-1}e^{j(\omega_i-\omega_s)k}\left(\sum_{q=0}^{k}\alpha_s p_q e^{j\omega_s q}\right)\\
&=\frac{1}{\alpha_i N}\sum_{k=0}^{N-1}\alpha_s p_k e^{-j\omega_s k}\left(\sum_{q=1}^{N-k}e^{j(\omega_i+\omega_s)(q+k-1)}\right)\\
&+\frac{1}{\alpha_i N}\sum_{k=0}^{N-1}\alpha_s p_k e^{j\omega_s k}\left(\sum_{q=1}^{N-k}e^{j(\omega_i-\omega_s)(q+k-1)}\right) \qquad \text{(A.37)}
\end{aligned}$$

Bounding this expression using (A.27) and the definitions (4.26), (4.27) of ψ_α, ϕ_W and adding over the remaining $(m-1)$ frequencies gives the result. ∎

A.4 Proof of lemma 4.1

From the proof of theorem 4.2,

$$|P_{\omega_0} - P(e^{j\omega_0})| \;\leq\; L(\gamma,\rho,N,\omega_0) + \frac{2}{\alpha N}\left|\sum_{k=0}^{N-1} v_k e^{j\omega_0 k}\right| \qquad \text{(A.38)}$$

For a given $\delta \leq K$, it follows that

$$\mathcal{P}\big(|P_{\omega_0} - P(e^{j\omega_0})| \;\leq\; L(\gamma,\rho,N,\omega_0)+\delta\big) > \mathcal{P}\left(\frac{2}{\alpha N}\left|\sum_{k=0}^{N-1} v_k e^{j\omega_0 k}\right| \leq \delta\right) \qquad \text{(A.39)}$$

For a given ω_0, define sets

$$\begin{aligned}
\hat{X}_{cos} &= \left\{v\in\mathbb{R}^N :\; \left(\frac{2}{\alpha N}|\sum_{k=0}^{N-1} v_k\, \cos\omega k\,| \;>\; \frac{\delta}{\sqrt{2}}\right)\right\}\\
\hat{X}_{sin} &= \left\{v\in\mathbb{R}^N :\; \left(\frac{2}{\alpha N}|\sum_{k=0}^{N-1} v_k\, \sin\omega k\,| \;>\; \frac{\delta}{\sqrt{2}}\right)\right\} \qquad \text{(A.40)}
\end{aligned}$$

Then

$$\left\{v\in\mathbb{R}^N : \frac{1}{N}\left|\sum_{k=0}^{N-1} v_k e^{j\omega_0 k}\right| > \delta\right\} \subseteq \hat{X}_{cos}\cup\hat{X}_{sin} \qquad \text{(A.41)}$$

Since $E(v_k)=0$, $E\left(\frac{2}{\alpha N}\sum_{k=0}^{N-1} v_k\cos(\omega_0 k)\right)=0$, $E\left(\frac{2}{\alpha N}\sum_{k=0}^{N-1} v_k\sin(\omega_0 k)\right)=0$, due to linearity of expectation operator; and $\frac{2}{\alpha}v_k\cos(\omega_0 k)$ are independent random variables bounded

in $[-\frac{2K}{\alpha}, \frac{2K}{\alpha}]$. Now

$$\begin{aligned}
&\mathcal{P}\left(\frac{2}{\alpha N}\left|\sum_{k=0}^{N-1} v_k \cos(\omega_0 k)\right| > \frac{\delta}{\sqrt{2}}\right) = \\
&\mathcal{P}\left(\frac{2}{\alpha N}\sum_{k=0}^{N-1} v_k \cos(\omega_0 k) > \frac{\delta}{\sqrt{2}}\right) + \mathcal{P}\left(-\frac{2}{\alpha N}\sum_{k=0}^{N-1} v_k \cos(\omega_0 k) > \frac{\delta}{\sqrt{2}}\right) \\
&\leq 2e^{-\frac{N\delta^2\alpha^2}{16K^2}}
\end{aligned} \tag{A.42}$$

where the last inequality follows from Hoeffding's theorem. Bounding the probability of $\mathcal{P}\left(\frac{1}{N}\left|\sum_{k=0}^{N-1} v_k \sin(\omega_0 k)\right| > \frac{\delta}{\sqrt{2}}\right)$ in a similar fashion and then using (A.41),

$$\mathcal{P}\left(\frac{1}{N}\left|\sum_{k=0}^{N-1} v_k e^{j\omega_0 k}\right| > \delta\right) \leq 4e^{-\frac{N\delta^2\alpha^2}{16K^2}} \tag{A.43}$$

Combining (A.39) and (A.43) gives

$$\mathcal{P}\left(|P_{\omega_0} - P(e^{j\omega_0})| \leq L(\gamma, \rho, N, \omega_0) + \delta\right) > 1 - 4e^{-\frac{N\delta^2\alpha^2}{16K^2}} \tag{A.44}$$

Equating the right hand side with ζ and then solving for δ in terms of ζ gives the required result. ∎

Appendix B

Appendix to chapter 7

In the two proofs that follow, some properties of the winding number of determinant of a square transfer matrix will be used repeatedly:

$$\begin{aligned} \text{wno det}\,(AB) &= \text{wno det}\,(A) + \text{wno det}\,(B) \\ \text{wno det}\,(A^*) &= \text{wno det}\,(A^{-1}) = -\,\text{wno det}\,(A) \quad \text{and} \\ X \in \mathcal{L}_\infty,\, \|X\|_\infty < 1 &\Rightarrow \text{wno det}\,(I+X) = 0 \end{aligned}$$

Proofs can be found in ([Vin], chapter 1).
Now, a technical result critical to the proof of theorem 7.1 is given.

Lemma B.1

$$I(P_1, P_2) = I(P_2^*, P_1^*) + deg(P_1) - deg(P_2)$$

Proof : The proof is based on the following result from ([Vin], chapter 8):

Lemma B.2 *([Vin]) Given $\{N, M\}$ a normalised right coprime factorisation of P, there exists an Ω with the following properties*

(i). $\Omega \in \mathcal{H}_\infty$

(ii). $wno\ det\Omega = deg(P)$

(iii). $\Omega^*\Omega = I$

(iv). $\{\Omega N^*, \Omega M^*\}$ *is a normalised left coprime factorisation of* P^*.

Here, deg(P) represents the Mcmillan degree of P. Let Ω_1, Ω_2 be such transfer functions corresponding to G_1 and G_2 respectively, defined as in the above lemma. Then

$$\begin{aligned} I(P_1, P_2) &= \text{wno det}\,(G_2^* G_1) \\ &= \text{wno det}\,\{(\Omega_2^*\Omega_2)(G_2^* G_1)(\Omega_1^*\Omega_1)\} \\ &= \text{wno det}\,(\Omega_2^*) + I(P_2^*, P_1^*) + \text{wno det}\,(\Omega_1) \\ &= I(P_2^*, P_1^*) + deg(P_1) - deg(P_2) \qquad \blacksquare \end{aligned} \tag{B.1}$$

B.1 Proof of theorem 7.1

(ii) $\Rightarrow$ (i):
Let $P_3 = P_1^*$ and note that $\sqrt{1-\sigma_k^2(\mathbf{H}_{G_1})} = \sigma_{n-k+1}(\mathbf{H}_{G_3})$ ([Vin], chapter 8). From ([ZDG96], theorem 8.7),

$$\inf_{X \in \mathcal{H}_{\infty(-),n-k}} \|G_3 - X\|_\infty = \sigma_{n-k+1}(\mathbf{H}_{G_3}) \tag{B.2}$$

and the infimum is achieved. Hence for $\gamma > \sigma_{n-k+1}(G_3)$, $\exists\ X_1 \in H_{\infty(-),n-k}$ such that $\|G_3 - X_1\|_\infty < \gamma$. For such an X_1, $\eta(X_1) \geq deg(X_1) - (n-k)$, from the definition of $H_{\infty(-),n-k}$.

Let $X_1 = FQ^{-1}$ be a right coprime factorisation of X_1. Then wno det $(Q) = \eta(X_1)$. Put $F = [F_1\ F_2]^T$. If F_2^{-1} exists and if F_1, F_2 are right coprime, put $H = F$. Otherwise put $H = [F_1\ F_2 + \epsilon I]^T$ where ϵ is sufficiently small so that $(F_2 + \epsilon I)^{-1}$ exists, F_1, $F_2 + \epsilon I$ are right coprime and

$$\|G_3 - HQ^{-1}\|_\infty < \gamma \tag{B.3}$$

is satisfied. Let $P_2 = \{Fac(H)\}^*$. Now

$$\|G_3 - X_1\|_\infty = \left\| \begin{bmatrix} G_3^* \\ \tilde{G_3} \end{bmatrix} (G_3 - X_1) \right\|_\infty \tag{B.4}$$

(since $\begin{bmatrix} G_3^* \\ \tilde{G_3} \end{bmatrix}$ is inner and doesn't affect the norm.)

$$= \left\| \begin{bmatrix} I - G_3^* X_1 \\ -\tilde{G_3} X_1 \end{bmatrix} \right\|_\infty = \sigma_{n-k+1}(\mathbf{H}_{G_3}) < 1 \tag{B.5}$$

Hence we have

$$\begin{aligned} \text{wno det}\,(G_3^* X_1) &= \text{wno det}\,(G_3^* H Q^{-1}) = 0 \\ \Rightarrow \text{wno det}\,(G_3^* H) &= \text{wno det}\,(Q) \\ \Rightarrow I(P_2^*, P_1^*) &= \eta(X_1) \geq deg(X_1) - (n-k) \\ \Rightarrow I(P_1^*, P_2^*) &< (n-k) - deg(X_1) \qquad \text{(B.6)} \\ \Rightarrow I(P_2, P_1) &< (n-k) - deg(X_1) \\ &\quad + deg(P_2) - deg(P_1) \qquad \text{(B.7)} \end{aligned}$$

where (B.7) follows from (B.6) using lemma B.1. Since $deg(P_1) = n$ and $deg(P_2) \leq deg(H) \leq deg(X_1)$, (B.7) gives

$$I(P_2, P_1) < -k$$

as desired. Further, using (7.5) and (B.3), it follows that $\delta_{\mathcal{L}_2}(P_1, P_2) = \delta_{\mathcal{L}_2}(P_1^*, P_2^*) < \gamma$.

(i) $\Rightarrow$ (ii):

The proof is along the same lines as the proof of Theorem 8.6 in [Vin]. Let P_2 be such that $I(P_2, P_1) \leq -k$ and $\delta_{\mathcal{L}_2}(P_2, P_1) = \epsilon < \gamma$. Then we need to show that $\epsilon \geq \sigma_{n-k+1}(\mathbf{H}_{G_3})$, where $P_3 = P_1^*$ as before.

Let $P_4 = P_2^*$ and define $X := G_4(G_3^* G_4)^{-1}(1-\epsilon^2)$. Note that

$$\frac{X^*}{1-\epsilon^2} = (G_4^* G_3)^{-1} G_4^* = (\Omega_4 G_4^* G_3)^{-1} \Omega_4 G_4^*$$

where Ω_4 is as defined in lemma B.2. $(\Omega_4 G_4^* G_3)$ and $\Omega_4 G_4^*$ are $\mathcal{RH}_\infty$ functions, and

$$\begin{aligned} \text{wno det}\,(\Omega_4 G_4^* G_3) &= deg(P_2) + I(P_1^*, P_2^*) \\ &= deg(P_2) + I(P_2, P_1) \\ &\quad + deg(P_1) - deg(P_2) \\ &\leq n - k \qquad \text{(B.8)} \end{aligned}$$

so that X^* has no more than $n-k$ unstable poles, which implies $X \in \mathcal{H}_{\infty(-),n-k}$. Since $\begin{bmatrix} G_3^* \\ \tilde{G}_3 \end{bmatrix}$ is inner,

$$\|G_3 - X\|_\infty = \left\| \begin{bmatrix} G_3^* \\ \tilde{G}_3 \end{bmatrix} (G_3 - X) \right\|_\infty = \left\| \begin{bmatrix} \epsilon^2 I \\ (\tilde{G}_3 G_4)(G_3^* G_4)^{-1}(1-\epsilon^2) \end{bmatrix} \right\|_\infty \qquad \text{(B.9)}$$

From ([Vin], lemma 2.2), $\|\tilde{G}_3 G_4\|_\infty = \epsilon$ and $(\tilde{G}_3 G_4)^*(\tilde{G}_3 G_4) + (G_3^* G_4)^*(G_3^* G_4) = I$ imply that $\|(\tilde{G}_3 G_4)(G_3^* G_4)^{-1}\|_\infty = \frac{\epsilon}{\sqrt{1-\epsilon^2}}$. From (B.9),

$$\|G_3 - X\|_\infty = \sqrt{\epsilon^4 + (1-\epsilon^2)^2 \frac{\epsilon^2}{1-\epsilon^2}} = \epsilon \tag{B.10}$$

The proof is completed by noting that $\epsilon < \sigma_{n-k+1}$ contradicts (B.2). ∎

Bibliography

[AGK92] H. Akcay, G. Gu, and P.P. Khargonekar. Identification in $\mathcal{H}_\infty$ with nonuniformly spaced frequency response measurements. In *Proc. of Amer. Contr. Conf.*, pages 246–250, 1992.

[BGFB94] S. Boyd, L.El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* SIAM, 1994.

[BGS99] X. Bombois, M. Gevers, and G. Scorletti. Controller validation for a validated model set. In *Proc. of European Contr. Conf.*, Karlsruhe, Germany, 1999.

[BM98] J. M. Böling and P. M. Mäkilä. On control relevant criteria in $\mathcal{H}_\infty$ identification. *IEEE Trans. Automat. Contr.*, 43:694–698, 1998.

[CFN96] J. Chen, J.A. Farell, and C.N. Nett. $\mathcal{H}_\infty$ identification of multivariable systems by tangential interpolation methods. *IEEE Trans. Automat. Contr.*, 41:1822–1828, 1996.

[CNF95] J. Chen, C.N. Nett, and M.K.H. Fan. Worst case system identification in $\mathcal{H}_\infty$: Validation of a priori information, essentially optimal algorithms, and error bounds. *IEEE Trans. Automat. Contr.*, 40:1260–1265, 1995.

[DFT92] J.C. Doyle, B.A. Francis, and A.R. Tannenbaum. *Feedback Control Theory.* Macmillan, 1992.

[dVvdH95] D. K. de Vries and P. M. J. van den Hof. Quantification of uncertainty in transfer function estimation: A mixed probalistic - worst case approach. *Automatica*, 31:543–557, 1995.

[FL98] U. Forsell and L. Ljung. Closed loop identification revisited. Technical report, Linköping University, 1998.

[Fle87] R. Fletcher. *Practical Methods of Optimisation*. John Wiley and Sons, 1987.

[FS99] H. Fukushima and T. Sugie. Identification of unfalsified plant model sets based on low-correlated bounded noise model. In *Proc. of 38th Conf. on Dec. and Contr.*, pages 5088–5093, 1999.

[GA94] P. Gahinet and P. Apkarian. A linear matrix inequality approach to $\mathcal{H}_\infty$ control. *Int. Journal of Robust and Nonlin. Contr.*, 4:421–448, 1994.

[GK92] G. Gu and P.P. Khargonekar. A class of algorithms for identification in $\mathcal{H}_\infty$. *Automatica*, 28:299–312, 1992.

[GL95] M. Green and D. J. N. Limebeer. *Linear Robust Control*. Prentice Hall, 1995.

[GLZ96] M. Glaum, L. Lin, and G. Zames. Optimal $\mathcal{H}_\infty$ approximation by systems of prescribed order using frequency response data. In *Proc. 35th IEEE Conf. Des. Contr.*, pages 2318–2321, 1996.

[GS90] T. T. Georgiou and M. C. Smith. Optimal robustness in the gap metric. *IEEE Trans. Automat. Contr.*, 35:673–687, 1990.

[Gu99] G. Gu. Modelling of normalised coprime factors with ν-metric uncertainty. *IEEE Trans. Automat. Contr.*, 44:1498–1511, 1999.

[GXZ93] G. Gu, D. Xiong, and K. Zhou. Identification in $\mathcal{H}_\infty$ using pick's interpolation. *Syst. and Contr. Lett.*, 20:263–272, 1993.

[HJN91] A.J. Helmicki, C.A. Jacobson, and C.N. Nett. Control oriented system identification: A worst case/deterministic approach in $\mathcal{H}_\infty$. *IEEE Trans. Automat. Contr.*, 36:1163–1176, 1991.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Amer. Statistical Asso. Journal*, (3):13–30, 1963.

[Jef94] A. Jeffrey, editor. *Table of Integrals, Series, and Products*. Academic Press, 1994.

[Kre88] E. Kreyzig. *Advanced Engineering Mathematics*. John Wiley and Sons, 1988.

[KT85] P.P. Khargonekar and A. Tannenbaum. Non-euclidian metrics and the robust stabilization of systems with parameter uncertainty. *IEEE Trans. Automat. Contr.*, 30:1005–1013, 1985.

[Lju99] L. Ljung. *System Identification: Theory For The User*. Prentice Hall, 1999.

[MÖ91] P. M. Mäkilä. Robust identification and Galois sequences. *Int. J. Control*, 54:1189–1200, 1991.

[Mat95] Mathworks Inc. *LMI control toolbox manual*, 1995.

[MG92] D. C. McFarlane and K. Glover. A loop shaping design procedure using $\mathcal{H}_\infty$ synthesis. *IEEE Trans. Automat. Contr.*, 37:759–769, 1992.

[MP92] P. M. Mäkilä and J. R. Partington. Robust identification of strongly stabilizable systems. *IEEE Trans. Automat. Contr.*, 37:1709–1716, 1992.

[Pag95] F. Paganini. *Sets and Constraints in the Analysis of Uncertain Systems*. PhD thesis, Cal. Inst. of Tech., 1995.

[Pap91] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, Inc., 1991.

[Par92] J. R. Partington. Robust identification in $\mathcal{H}_\infty$. *Journal of Math. Anal. Appl.*, 166:428–441, 1992.

[Pin85] A. Pinkus. *n-widths in Approximation Theory*. Springer-Verlag, 1985.

[PM99] J. R. Partington and P. M. Mäkilä. Robust identification and the rejection of outliers. 1999.

[RPG92] D. E. Rivera, J.F. Pollard, and C. E. Garcia. Control relevant prefiltering: A systematic design approach and case study. *IEEE Trans. Automat. Contr.*, 37:964–974, 1992.

[Rud76] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.

[SS89] T. Södderström and P. Stoica. *System Identification*. Prentice Hall, 1989.

[TL95] A. Tikku and L. Ljung. Worst case identification in l_1 for f.i.r. linear systems. In *Proc. of 34th Conf. on Dec. and Contr.*, pages 2998–3003, 1995.

[VD] S. R. Venkatesh and M. A. Dahleh. On system identification of complex systems with finite data. pre-print.

[VD97] S. R. Venkatesh and M. A. Dahleh. Identification in the presence of classes of unmodeled dynamics and noise. *IEEE Trans. Automat. Contr.*, 42:1620–1635, 1997.

[vdHS95] P. M. J. van den Hof and R. J. P. Schrama. Identification and control - closed loop issues. *Automatica*, 31:1751–1770, 1995.

[Vid85] M. Vidyasagar. *Control Systems Synthesis: A Facorisation Approach*. MIT Press, 1985.

[Vin] G. Vinnicombe. *Uncertainty and Feedback: $\mathcal{H}_\infty$ loop-shaping and the $\nu-$gap metric*. Imperial College Press. In Press.

[Vin93] G. Vinnicombe. Frequency domain uncertainty and the graph topology. *IEEE Trans. Automat. Contr.*, 38:1371–1383, 1993.

[Vin96] G. Vinnicombe. The robustness of feedback systems with bounded complexity controllers. *IEEE Trans. Automat. Contr.*, 41:795–803, 1996.

[VM97] G. Vinnicombe and S. Miyamoto. On reduced order $\mathcal{H}_\infty$ loop-shaping controllers: A design method and examples of local optima. In *Proc. of 4th European Contr.Conf.*, 1997.

[You95] N. Young. *An Introduction to Hilbert Space*. Cambridge University Press, 1995.

[ZBG95] Z. Zang, R. R. Bitmead, and M. Gevers. Iterative weighted least squares identification and weighted LQG control. *Automatica*, 31:1577–1594, 1995.

[ZDG96] K. Zhou, J.C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, 1996.

[ZLW94] G. Zames, L. Lin, and L.Y. Wang. Fast identification $n-$widths and uncertainty principles for LTI and slowly varying systems. *IEEE Trans. Automat. Contr.*, 39:1827–1838, 1994.