The Pennsylvania State University

The Graduate School

Department of Electrical Engineering

SEMI-BLIND ROBUST IDENTIFICATION AND MODEL (IN)VALIDATION

A Thesis in

Electrical Engineering

by

Wenjing Ma

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2007

The thesis of Wenjing Ma was reviewed and approved* by the following:

Mario Sznaier
Professor of Electrical Engineering
Thesis Advisor, Co-Chair of Committee

Constantino M. Lagoa
Associate Professor of Electrical Engineering
Co-Chair of Committee

Octavia I. Camps
Associate Professor of Electrical Engineering

W. Kenneth Jenkins
Professor of Electrical Engineering
Department Head of Electrical Engineering

Qian Wang
Assistant Professor of Mechanic Engineering

*Signatures are on file in the Graduate School.

Abstract

In this thesis, we study a so-called semi-blind robust identification motivated from the fact that sometimes for system Identification only partial input data is exactly known. Derived from a time-domain algorithm for robust identification, this semi-blind robust identification is stated as a non convex problem. We develop a convex relaxation, by combining two variables into a new variable, to reduce it to an LMI optimization problem. Applying this convex relaxation, a macro-economy modelling problem can be solved. For future work of application on Intrusion Detection, a sampling algorithm for blind identification is also briefly presented.

Accordingly, we consider the problem of semi-blind (in)validation which is shown to be non convex. Two different relaxations — a deterministic and a risk-adjusted convex relaxation — are explored to solve this non convex problem. We demonstrate an application of the semi-blind (in)validation on the problem of detecting and isolating faults from noisy input-output measurements. The results of this application using both two relaxations are presented through an experimental example.

Furthermore, the problem of identification of Wiener Systems, a special type of nonlinear systems, is analyzed from a set-membership standpoint. We propose an algorithm for time-domain based identification by pursuing a risk-adjusted approach to reduce it to a convex optimization problem. An arising non-trivial problem in computer vision, tracking a human in a sequence of frames, can be solved by modelling the plant as Wiener system using the proposed identification method.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I will start out by thanking my advisor, Dr. Mario Sznaier, for his guidance and leadership throughout my doctoral program. You have provided me with a challenging, yet nurturing and invaluable experience and for that I am grateful. You have been a great mentor. I would also like to thank my co-advisor, Dr. Constantino Lagoa, for his endless patience and support throughout this process. I have the highest respect for both of you and will continue to reflect upon the knowledge that I have learned from you throughout my career. Thanks also goes to my committee members: Dr. Octavia Camps, Dr. Kenneth Jenkins, and Dr. Qian Wang, for their valuable time and fruitful suggestions.

Additional thanks to my colleagues in the Robust System Lab, Dr. Cecilia Mazzaro, Dr. Muhittin Yilmaz, Dr. Tao Ding and Dr. Hwasup Lim, for their assistance on the projects and their friendship.

Finally, I would like to thank my parents, Linxia Ma and Tongkuang Ma, for believing in me. Without your constant encouragement, I never would have gotten to this point. Your unwavering support got me through the toughest times in graduate school.

Chapter 1

Motivation and Contribution

1.1 Motivation

The goal of this dissertation is to study several open problems that have not been noticed while are very important to the control community. It focuses on Robust or Control–Oriented Identification and Model (In)Validation. An additional motivation for studying these problems is the variety applications of the operator theoretic tools, such as in Economy, Fault Detection and Isolation (FDI) and Computer Vision.

The purpose of using system identification in the context of the well established robust control design leads to the origin of robust identification around year 1990 (see [33, 34, 35, 36]). It is required that the robust identification method yields a bounded, uncertain model set description, in the form of a nominal model with an explicit, worst-case identification error bounds on the plant. Unlike traditional identification, one advantage of robust identification is that it does not requires any information on the statistical properties of the noise or the model structure of the system to be identified.

A lot of research has been done on the problem of developing robust identification procedures in the past decade. ℓ_1 identification (see [59, 115, 14, 41, 43, 52, 55, 54]) with time-domain experimental data defines an ℓ_1 error bound, which can be used in

ℓ_1 optimal control. It is useful in approximate verification of the stability, time invariance, and linearity of the systems ([84]), etc., before a frequency response experiment[1] is performed. $\mathcal{H}_\infty$-based identification (see [36, 37, 16, 31, 30, 73, 51, 53]), using frequency–domain experimental data, measures the uncertainty bound in terms of the $\mathcal{H}_\infty$–norm, which is adequate for $\mathcal{H}_\infty$ optimal control or μ–synthesis procedures. Furthermore, mixed $\mathcal{H}_\infty/\ell_1$ identification (see [13, 85, 70]) produces a smaller consistency set of indistinguishable models and a consequent smaller worst–case identification error through combining both time–domain and frequency–domain experimental data. Especially for Linear Time Invariant (LTI) systems, it is by now relatively mature and efficient algorithms (see [12, 86], etc.) are available to obtain both nominal models and worst-case identification error bounds. One special case called blind robust identification[2], which deal with the case of unknown input sequence, is addressed in [58].

However, all of these approaches assume that the input–output experimental data are generated exactly from initial time $t_0 = 0$ by ignoring the effects of inputs prior to the initial time, which would lead to artificially high identification errors. To overcome the high identification error due to this fact, a so–called Semi–Blind robust identification is stated and one contribution of this dissertation is to develop an interpolatory algorithm for it.

Following the developments in the robust identification, model (in)validation, known as determining whether or not the experimental corrupted input-output data

[1] Although all experiments are performed in the time–domain, a sinusoidal sweep at the input of the plant to obtain an output magnitude and phase response is defined here as a frequency response experiment.

[2] for blind identification, [2] study it in a deterministic context by assuming the statistical properties of the input

is consistent with the given uncertain model, has been acknowledged as an momentous component in control system modeling. If the experimental data is consistent with the given model, i.e. the data can be reproduce to form the model, the model is said to be *not invalidated*. Otherwise, it is said to be *invalidated*. Note that we can not really validate a model since the model can only be tested by a finite data length. Since a set of models obtained through a robust identification algorithm will always be not invalidated by the experimental data used to compute it with the identical framework used for the experiment, there are only two reasons by which the identified set of models may be invalidated: *(a)* a *different* setup structure, *(b)* a *new* set of experimental data.

Model (in)validation techniques for LTI systems have been extensively studied in the past few years. It has been initiated in [93]. For general Linear Fractional Transformation (LFT) uncertain models (see [57, 11, 17, 93]), the problem is partially solved. In the case of LTI, causal, unstructured uncertainty, the problem can be reduced to a convex optimization problem that can be efficiently solved, by applying norm constrained interpolation theory. In the case of slow linear time-varying (SLTV) structured uncertainty, the problem has been completely solved by obtaining tractable, convex necessary and sufficient conditions for it (see [57]). Whereas, in the case of structured LTI uncertainty, the problem still remains open and was shown to be an NP-hard problem (see in [99, 91]). Note that all these approaches ignore the effects of inputs prior to the initial time. Therefore, a so–called Semi–Blind model (in)validation is motivated for the same reason as for Semi–Blind robust identification.

The techniques addressed above are also of practical interest. The performance of economical and financial systems can be optimized by using robust control tools (see

[87, 100, 63]). The benefits of applying robust identification and model (in)validation to obtain and validate models of economic subsystems are multiple. For example, it indicates when the model is no longer compatible with the measured data and thus provide a mechanism to answer Lucas' critique (see in [50]) questioning the usefulness of using the identified models for macro-economic forecasting. Furthermore, the similarity between the setup structure of FDI problem and semi–blind model (in)validation allows us to apply semi–blind model (in)validation framework on FDI and offers us a well-performed estimation for the faults.

Motivated by the two facts that *(a)* most robust identification algorithms just deal with linear systems and *(b)* there is a special type of nonlinear systems, consists of the cascade of a LTI model and a memoryless, static non–linearity, arising in many practical situations, we turn our attention to the identification of this special type of nonlinear systems, called Wiener system. One of the useful applications of wiener system identification is the problem of human motion modeling and tracking. This application can override the difficulty of searching very high dimensional spaces due to the fact that even small size of images require processing hundreds of pixels.

1.2 Contribution and Organization of the dissertation

This dissertation shows that the so–called semi–blind robust identification and semi–blind model (in)validation are both non-convex and it explores some relaxations to reduce them to solvable convex optimization problems. The contributions of this dissertation can be valued not only theoretically, by providing answers to the unnoticed problem of ignoring effects of priori inputs. It can also be valued practically through

illustrating how dynamic nominal models and uncertainties can combined to solve the problem of macro-economic modeling and how model (in)validation framework can be applied to solve the problem of Fault Detection and Isolation. One more contribution of this dissertation is the development of the wiener system identification algorithm from a set–membership standpoint. This technique is utilized to settle a non-trivial problem in computer vision: human motion modeling and tracking.

This thesis is organized as follows.

Chapter 2. gives all the notations that has been used throughout the thesis and briefly stated some mathematical background theorems.

Chapter 3. first formulates the time-domain robust identification problem and then presents an interpolatory identification algorithm. Afterward, a convex relaxation for semi-blind identification is developed and an application on macro-economic subsystem modeling is illustrated. At the end of this chapter, blind robust identification is explained and a sampling algorithm is shown.

Chapter 4 addresses model (in)validation with time-domain measurements. We explore two different convex relaxations, one deterministic and one stochastic, for the semi-blind model (in)validation. We seek a minimum bound of the uncertainty for a new set of experimental data of the identified model of the macro-economy subsystem in Chapter 2 with a suitable noise level. At last, a FDI problem is interpreted in a semi-blind model (in)validation framework.

Chapter 5 analyzes the problem of Wiener System identification from a set-membership standpoint. We reduce the problem to a convex LMI optimization form

by pursuing a risk-adjusted approach. The results are illustrated with an application on a non-trivial problem in computer vision: tracking a human in a sequence of frames.

Finally, Chapter 6, the last chapter, summarizes all the work in the thesis.

Chapter 2

Preliminaries

This chapter provides the notation we use throughout the whole proposal and some background theorems for the problems under consideration.

2.1 Notation

$\mathcal{R}$	set of real numbers.
$\mathbf{x}$	column vector.
A^H	conjugate transpose of matrix A.
A^{-1}	inverse of square matrix A.
$\lambda(A)$	eigenvalue of A.
$\sigma(A)$	singular value of A.
$\overline{\sigma}(A)$	maximum singular value of A.
$A > (\geq) 0$	$A = A^H$ is positive(semi) definite.
$\mathcal{R}^n$	n–dimensional real–number plan.
$\mathbf{I}, \mathbf{0}$	the identity and zero matrices of compatible dimensions (when omitted).
$\mathcal{C}^{m \times n}$	Complex plan of dimension $m \times n$.
$(\mathcal{X}, m)$	metric space of elements in $\mathcal{X}$ equipped with the metric $m(x_1, x_2)$.

$\|\mathbf{x}\|_p$ — p–norm of vector $\mathbf{x}$, defined as:

$$\|\mathbf{x}\|_p \doteq \left(\sum_{i=0}^{\infty} |x_i|^p\right)^{\frac{1}{p}},$$

$p \in [1, \infty)$ and $\|\mathbf{x}\|_\infty \doteq \max_i |x_i|$.

$\|A\|_p$ — induced p–norm of matrix A, defined as:

$$\|A\|_p \doteq max_{x \neq 0} \left(\frac{\|Ax\|_p}{\|x\|_p}\right).$$

$\mathcal{B}\mathcal{X}(\gamma)$ — open γ-ball in a normed space $\mathcal{X}$: $\mathcal{B}\mathcal{X}(\gamma) = \{x \in \mathcal{X} \colon \|x\|_{\mathcal{X}} < \gamma\}$.

$\overline{\mathcal{B}}\mathcal{X}(\gamma)$ — closure of $\mathcal{B}\mathcal{X}(\gamma)$.

$r(\mathcal{A})$ — radius of set $\mathcal{A} \subseteq \mathcal{X}$: $r(\mathcal{A}) \doteq \inf_{x \in \mathcal{X}} \sup_{a \in \mathcal{A}} m(x, a)$.

$d(\mathcal{A})$ — diameter of set $\mathcal{A} \subseteq \mathcal{X}$: $d(\mathcal{A}) \doteq \sup_{x, a \in \mathcal{A}} m(x, a)$.

$\mathcal{X}^{m \times n}$ — normed space of dimension $m \times n$ with each entry in $\mathcal{X}$

ℓ^p — extended Banach space of vector valued real sequences equipped with the norm: $\|x\|_p \doteq \left(\sum_{i=0}^{\infty} \|\mathbf{x}_i\|_p^p\right)^{\frac{1}{p}} < \infty$, for $p \in [1, \infty]$ and $\|x\|_\infty \doteq \sup_i \|\mathbf{x}_i\|_\infty$.

$\mathcal{L}_\infty$ — Lebesgue space of complex-valued matrix functions essentially bounded on the unit circle, equipped with the norm: $\|G\|_\infty \doteq ess \sup_{|z|=1} \overline{\sigma}(G(z))$.

$\mathcal{H}_\infty$ Space of functions with bounded analytic continuation inside the unit disk, equipped with the norm: $\|G\|_\infty \doteq ess\sup_{|z|<1} \overline{\sigma}(G(z))$.

$\mathcal{H}_\infty^{m\times n}$ Space of matrix functions of dimension $m\times n$ with each entry in $\mathcal{H}_\infty$.

$\mathcal{H}_{\infty,\rho}$ space of transfer functions analytic in $|z| \le \rho$,equipped with the norm $\|G\|_{\infty,\rho} \doteq ess\sup_{|z|<\rho} \overline{\sigma}(G(z))$.

$\mathcal{BH}_\infty$ open unit ball in the normed space: $\mathcal{BH}_\infty$: $\{h \in \mathcal{H}_\infty, \|h\|_\infty < 1\}$.

$\overline{\mathcal{B}}\mathcal{H}_\infty$ closed unit ball in the normed space $\mathcal{H}_\infty$.

$\mathcal{BH}_\infty^N$ set of $(N-1)^{\text{th}}$ order FIR transfer matrices that can be completed to belong to $\mathcal{BH}_\infty$, i.e. $\mathcal{BH}_\infty^N \doteq \{H(z) = \mathbf{h}_0 + \mathbf{h}_1 z + \ldots + \mathbf{h}_{N-1} z^{N-1} : H(z) + z^N G(z) \in \mathcal{BH}_\infty, \text{ for some } G(z) \in \mathcal{H}_\infty\}$.

$\mathbf{T}_x$ lower triangular block Toeplitz matrix associates with any finite sequence $\{x_k, k = 0, 1, \cdots, n-1\}$, or any column vector $\mathbf{x} = [x_0, x_1, \cdots, x_{n-1}]^T$:

$$\mathbf{T}_x = \begin{bmatrix} x_0 & 0 & \ldots & 0 \\ x_1 & x_0 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ x_{n-1} & x_{n-2} & \ldots & x_0 \end{bmatrix}$$

.

In the case of an LTI system, from an input–output viewpoint, any operator of interest G will be represented either by a (rational) complex–valued transfer function:

$$G(z) \doteq \sum_{0}^{\infty} g_k z^k$$

or a minimal state–space realization:

$$G \equiv \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$$

We denote by $\mathbf{T}_G$ and $\mathbf{\Gamma}_G$ the Toeplitz and Hankel operators associated with an ℓ_∞ stable system.

For an input sequence applied in the interval $[0, \infty]$:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} g_0 & 0 & 0 & \cdots & \cdots \\ g_1 & g_0 & 0 & 0 & \cdots \\ g_2 & g_1 & g_0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{bmatrix}}_{\mathbf{T}_G} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \end{bmatrix}$$

where u_k and y_k denote the inputs and outputs sequences to the system.

For an input sequence applied in $(-\infty, -1]$:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} g_1 & g_2 & g_3 & \cdots \\ g_2 & g_3 & \cdots & \cdots \\ g_3 & \vdots & \ddots & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{\mathbf{\Gamma}_G} \begin{bmatrix} u_{-1} \\ u_{-2} \\ u_{-3} \\ \vdots \end{bmatrix}$$

When dealing with finite sequences of length N, we will represent these operator by the finite matrices $\mathbf{T}_G^N$ and $\mathbf{\Gamma}_G^N$ the finite $N \times N$ upper left sub-matrices of $\mathbf{T}_G$ and $\mathbf{\Gamma}_G$ respectively.

2.2 Mathematical Background Theorem

For preparation, this section introduces some repeatedly applied theorems and properties in the dissertation.

First, let's introduce the well-known Carathéodory-Fejér Interpolation Problem:

Problem 1. *(see [12])Given complex numbers $c_i, i = 0, 1, \cdots, n-1$, find a function $h \in \overline{\mathcal{B}}\mathcal{H}_\infty$ such that*

$$h(\lambda) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1} + \lambda^n \hat{g}(\lambda), \tag{2.1}$$

where $\hat{g} \in \overline{\mathcal{B}}\mathcal{H}_\infty$

In other words, the first n Taylor series coefficients of h are required to match the given complex numbers, i.e.,

$$\frac{h^{(k)}(0)}{k!} = c_k, \qquad k = 0, 1, \cdots, n-1.$$

Of immediate interest Carathéodory-Fejér Interpolation Theorem, which plays an important role in time–domain robust identification interpolatory algorithms , is presented as following,

Theorem 1. *(see [12]) Given complex numbers $c_i, i = 0, 1, \cdots, n-1$,there exists a function $h \in \overline{\mathcal{B}}\mathcal{H}_\infty$ such that it satisfies* (2.1) *if and only if*

$$\mathbf{I} - \mathbf{T}_c^H \mathbf{T}_c \geq 0 \tag{2.2}$$

where T_c is an $n \times n$ lower triangular Toeplitz matrix associated with sequence $c = [c_0, c_1, \cdots, c_{n-1}]$. The function is non–unique if and only if $\mathbf{I} - \mathbf{T}_c^H \mathbf{T}_c$ is positive definite, and unique if and only if $\mathbf{I} - \mathbf{T}_c^H \mathbf{T}_c$ is rank deficient. All the interpolating functions can be found following these steps:

1. *Determine the rank of $\mathbf{I} - \mathbf{T}_c^H \mathbf{T}_c$. Let it be m, and denote the leading principal minor of T_c as T_m. If $m < n$, then a unique rational interpolating function of degree m can be found as*

$$h(\lambda) = \frac{\bar{h}_{m-1} + \bar{h}_{m-2}\lambda + \cdots + \bar{h}_0 \lambda^{m-1}}{\hat{h}_0 + \hat{h}_1 \lambda + \cdots + \hat{h}_{m-1}\lambda^{m-1}}, \tag{2.3}$$

where $\hat{h}^T \doteq [\hat{h}_0, \hat{h}_1, \cdots, \hat{h}_{m-1}]$ *is the eigenvector of* $\mathbf{I} - \mathbf{T}_m^H \mathbf{T}_m$ *corresponding to the largest eigenvalue.*

2. *For* $m = n$*, form the matrix function*

$$G(\lambda) = \begin{bmatrix} g_{11}(\lambda) & g_{12}(\lambda) \\ g_{21}(\lambda) & g_{22}(\lambda) \end{bmatrix}$$

such that

$$G(\lambda) = \mathbf{I} + (\lambda - 1) C_n (\lambda \mathbf{I} - R)^{-1} (\mathbf{I} - \mathbf{T}_c^H \mathbf{T}_c)^{-1} (\mathbf{I} - R^H)^{-1} C_n^H \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

where

$$R \doteq \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix},$$

and

$$C_n = \begin{bmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ 1 & 0 & \cdots & 0 \end{bmatrix}.$$

The required interpolating function $h(\lambda)$ *is given as*

$$h(\lambda) = \frac{g_{11}(\lambda) f(\lambda) + g_{12}(\lambda)}{g_{21}(\lambda) f(\lambda) + g_{22}(\lambda)},$$

where $f(\lambda) \in \overline{\mathcal{B}\mathcal{H}}_{\infty}$ *is arbitrary*

One popular choice for $h(\lambda)$ is the central interpolant, corresponds to the choice of $f(\lambda) = 0$.

The extension of the above Carathéodory-Fejér Interpolation problem to matrix analytic functions leads to the so called Tangential Carathéodory-Fejér Interpolation:

Problem 2. *Given matrices* $\mathbf{U}_i \in \mathcal{C}^{l\times j}$ *and* $\mathbf{V}_i \in \mathcal{C}^{k\times j}, i = 0, 1, \cdots, n-1$, *find a matrix function* $\hat{\mathbf{F}} \in \overline{\mathcal{B}\mathcal{H}}_{\infty}^{k\times l}$ *such that*

$$\hat{\mathbf{F}}(\lambda) = \mathbf{F}_0 + \mathbf{F}_1\lambda + \cdots + \mathbf{F}_{n-1}\lambda^{n-1} + \lambda^n G(\lambda)$$

for some $G \in \overline{\mathcal{B}\mathcal{H}}_{\infty}^{k\times l}$ *and*

$$\mathbf{T}_F[\mathbf{U}_0^T, \cdots, \mathbf{U}_{n-1}^T]^T = [\mathbf{V}_0^T, \cdots, \mathbf{V}_{n-1}^T]^T \tag{2.4}$$

where $\mathbf{T}_F$ *is the lower triangular block Toeplitz matrix associated with the sequence* $F_i, i = 0, 1, \ldots, n-1$

$$\mathbf{T}_F = \begin{bmatrix} F_0 & 0 & \cdots & 0 \\ F_1 & F_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ F_{n-1} & F_{n-2} & \cdots & F_0 \end{bmatrix}$$

By converting this tangential Carathéodory-Fejérproblem into a Nudelman problem (see [12]), it leads to the following theorem:

Theorem 2. *(see [12]) There exists a matrix function* $\hat{\mathbf{F}} \in \overline{\mathcal{B}}\mathcal{H}_{\infty}^{k\times l}$ *such that* (2.4) *holds if and only if*

$$\mathbf{T}_U^H \mathbf{T}_U - \mathbf{T}_V^H \mathbf{T}_V \geq 0 \tag{2.5}$$

The problem admits a unique solution if and only if $\mathbf{T}_U^H \mathbf{T}_U - \mathbf{T}_V^H \mathbf{T}_V \geq 0$ *and* $\mathbf{T}_U^H \mathbf{T}_U - \mathbf{T}_V^H \mathbf{T}_V$ *is rank deficient, and the solution is non–unique if and only if* $\mathbf{T}_U^H \mathbf{T}_U - \mathbf{T}_V^H \mathbf{T}_V > 0$.

Chapter 3

Robust Identification for LTI System

This chapter summarizes a Time-domain Identification Algorithm for LTI systems, develops a method in time-domain for the so-called Semi-Blind Robust Identification, applies the Semi-Blind Robust Identification on Macro-Economic Modeling, and describes the Blind Robust Identification at the end.

3.1 Background of a time-domain Robust Identification

System identification is a procedure using mechanisms and algorithms that process finite, partial, and corrupted information to yield abstract mathematical descriptions for real world physical system. The available information generally consists of any *a priori* knowledge about the physical plant and *a posteriori* experimental data measured from the system's response in operation.

Classical identification approaches ([48, 107, 94, 47]) assume the system has a prescribed model structure and a stochastic property that the data is corrupted by a process with known statistical properties. Most of these identification procedures are based on least squares methods that estimate the parameters of the hypothesized models from the corrupted measurements. In these approaches, as long as the prescribed model is an accurate representation of the physical system, the only source of uncertainty is the noise in the measurements.

Noise description is only one of the factors affecting the quality of an identified model. A more important factor is the unrealistic presumption that a fixed model structure may fully represent the system to be identified. In practice, only partial information of the physical system is available, model parameters might change due to different operation conditions, and real systems are often too complex to be accurately modeled from first principles. Therefore, in the situations that the measurements are known within an accuracy range while their statistical information might be questionable or the available fixed model structure assumption is unrealistic, a deterministic bounded noise descriptions are a practical and sound alternative to stochastic ones. Using this deterministic identification error bounds, the problem of system identification can be formulated as a set–membership approach which finds the sets of parameter values that are consistent with the known noise bounds. A survey of the set–membership formulation can be found in [60].

In the classical identification community, it is well understood that there is a trade-off between the bias error, the one induced by the fact that the true plant is not within the assumed model structure, and the variance error, the one due to the noise in the data (see [49, 102, 106]). For a given experimental record, there is an optimum model order that balances both sources of error and gives the smallest total identification error. Invoked by this trade–off, the deterministic description gains an advantage over the stochastic description of the to be identified dynamics (see [62]).

Moreover, the fact that robust control analysis and design tools (see [86, 113]) are based on a deterministic worst–case approach, with no previous assumption of a fixed

model structure, leads to the reconcilement of robust control theory and the system identification that use set–membership approach. For this reason, the attempt to get suitable linear models for already existing powerful robust control tools, such as $\mathcal{H}_\infty$,μ-synthesis, is called Control–Oriented or Robust Identification theory. Intense research was initiated in the 1990s on deterministic identification procedures, which can be used as a first step in a robust control design orientation. It departs from traditional approaches by using a deterministic worst–case approach with no prior assumption about the order of the system. Instead, robust identification procedures are based on *a priori* assumptions on the class of systems and noise (see for example [12, 74, 86, 101])and on the *a posteriori* experimental data. Using this information, robust identification seeks to get a bounded model set description of the physical plant, in the form of a nominal model and a deterministic, worst–case bound on the identification error. It considers model uncertainty from tow different sources: measurement noise and lack of knowledge of the system itself due to limited information supplied by the experiment. Therefore, this procedure depends on not only on the *a posteriori* experimental data but also on the *a priori* assumptions on the class of systems to be identified.

3.1.1 Problem Formulation

We denote the class of systems which are discrete-time,causal,linear and stable as $H(z) = H_d(\frac{1}{z})$, with $z \in \mathcal{C}$, and $H_d(z)$ being the standard z-transform. Therefore, causal stable systems $H(z)$ is analytic inside the unit circle, with the standard z-transform of the plant impulse response evaluated at $\frac{1}{z}$:

$$H(z) = \sum_{k=0}^{\infty} h(k) z^{-k} \tag{3.1}$$

For simplicity, in the following, we discuss single-input/single-output (**SISO**) models, although all results can be applied to multi-input/multi-output (**MIMO**) systems.

In a robust identification setting, a linear system H is identified from the *a posteriori* experimental data (a finite set of measurements of the input-output data obtained in time–domain[1] with the output data corrupted by noise η) on the *a priori* information (assumed to consists of some set description of the admissible set of plants $\mathcal{S}$, assumed to include the actual system H, and noise $\mathcal{N}$) . In the case of discrete–time, SISO, LTI plants, one typical choice for S is the set of exponentially stable systems whose impulse response can be bounded in magnitude by $K\rho^{-k}$,$\forall k \geq 0$, with $K > 0$, $\rho > 1$, i.e. $H \in \mathcal{S} = \overline{\mathcal{B}}\mathcal{H}_{\infty,\rho}(K)$. The noise that corrupt the experimental output data is usually assumed to be additive and bounded at each sample time by a positive constant ϵ.

Given the actual linear system $H \in \mathcal{S}$ and a set of finite experimental input/output data with additive noise (collected in vectors $\mathbf{u}, \mathbf{y}, \eta$ respectively), the experiment can be denoted in terms of an operator as follows:

$$\mathbf{y} = E\big(H, \eta\big) \tag{3.2}$$

Note that the same output $\mathbf{y}$ may be produced by different combinations of model and noise due to the fact that the information provided by $\mathbf{y}$ is incomplete and corrupted.

[1] Time–domain robust identification collects the input–output data generated from time–domain; Frequency–domain robust identification collects the input–output data generated from frequency–domain

For a given $\mathbf{y}$, there exists a smallest set of plants that are indistinguishable from the corrupted experimental data, known as the *consistency* set $\mathcal{T}$:

$$\mathcal{T}(\mathbf{y}) = \{H \in \mathcal{S} \colon \mathbf{y} = E\big(H, \eta\big), \eta \in \mathcal{N}\} \tag{3.3}$$

This even may also happen in the noise free case, since the experimental evidence is incomplete[2]. Therefore, the operator is not invertible and no direct operation over $\mathbf{y}$ can provide the model H.

The identification framework works only when the *a priori* information is consistent with the *a posteriori* experimental data. Since there is no guarantee that the *a priori* information will always be coherent with the *a posteriori* experimental data, robust identification procedures must always first test the *consistency* of the two information.

Definition 1. *The* a priori *information* $\big(\mathcal{S}, \mathcal{N}\big)$ *is consistent with the* a posteriori *experimental data,* $\mathbf{y} = [y_0, y_1, \cdots, y_{N-1}]^T$ *and* $\mathbf{u} = [u_0, u_1, \cdots, u_{N-1}]^T$*, if and only if the set*

$$\mathcal{T}(\mathbf{y}) \doteq \{h \in \mathcal{S} | y_k = (h * u)_k + \eta_k, \text{ for some sequence } \eta_k \in \mathcal{N}, k = 0, 1, \ldots, N-1\} \tag{3.4}$$

is nonempty.

If $\mathcal{T} = \emptyset$, then the experimental data $\{\mathbf{y}, \mathbf{u}\}$ invalidates the *a priori* assumptions about the class of models and noise, that is, the experimental data cannot be explained

[2] e.g. consider the problem of finding an arbitrary LTI discrete-time model whose first n impulse response coefficients are fixed.

by models in these sets. In this case, it is necessary to improve the information on the class of models and measurement noise by finding a more accurate candidate set $\mathcal{S}$ or by tightening the bounds of the measurement noise.

After defining the consistency set, robust identification problem can be precisely stated as follows:

Problem 3. *Given the* a priori *information, sets* $(\mathcal{S},\mathcal{N})$*, and the* a posteriori *experimental data* $\mathbf{y}$,$\mathbf{u}$*, determine:*

1. *whether the* a priori *and* a posteriori *information are consistent, i.e., the consistency set* $\mathcal{T}(\mathbf{y})$ *is nonempty;*

2. *a nominal model which belongs to the consistency set* $\mathcal{T}(\mathbf{y})$ *if it is nonempty.*

However, even $(\mathcal{S},\mathcal{N})$ and $(\mathbf{u},\mathbf{y})$ are consistent, it could happen that the candidate set $\mathcal{S}$ does not include the real plant g_{real}, caused by the falseness of the *a priori* information. For example, when a very noisy measurement is used for identification, there may exists $g_1 \in \mathcal{T}(\mathbf{y})$ such that

$$E(g_1,\eta_1) = E(g_{real},\eta_{real}) = \mathbf{y}$$

with $\eta_1,\eta_{real} \in \mathcal{N}$, while g_{real} is not included in $\mathcal{T}(\mathbf{y})$. From a practical point of view, we can reduce the possibilities of having the real plant g_{real} excluded from candidate set $\mathcal{S}$ by performing as many types of experiments[3] as possible. Furthermore, a final model

[3] such as different time, frequency, and/or sensors

validation, using a new set of experimental data to test the consistency set, is necessary before the identified model can be used by engineers[4].

3.1.2 Time-Domain Interpolatory Identification Algorithm

Many papers addressed nonparametric identification of models with a worst–case global bound (see [36, 72, 15], etc.), however, in many cases, part of the model has a clear parametric structure. To avoid conservative results caused by disregarding this information, in this dissertation, we study the case of parametric identification.

Consider systems of the form $\mathcal{S}$:

$$\mathcal{S} \doteq \{G(z) = H(z) + P(z)\}. \tag{3.5}$$

$H \in \overline{\mathcal{BH}}_{\infty,\rho}(K)$ for some two positive constants $K > 0, \rho > 1$, represents the nonparametric components of the operator. $P(z)$ represents the parametric components, and belong to the class $\mathcal{P}$ of affine operators:

$$\mathcal{P} \doteq \{P(z) = \mathbf{p}^T G_p(z), \mathbf{p} \in \mathbb{R}^{N_p}\}, \tag{3.6}$$

where the N_p components $G_{p_i}(z)$ of vector $G_p(z)$ are known, linearly independent, rational transfer functions.

The following *Lemma* provides a necessary and sufficient condition for the consistency of the *a priori* and the *a posteriori* information.

[4]Model (In)Validation procedure is introduced in Chapter 4

Lemma 1. *(see [71]) Given $K > 0$,$\rho > 1$ and two vector sequences $(\mathbf{u}, \mathbf{y})$, there exists an operator $S \in \mathcal{S}$ such that the consistency set $\mathcal{T}(\mathbf{y})$ is nonempty, if and only if there exists a vector $\mathbf{h}$ satisfying:*

$$\mathbf{M}(\mathbf{h}) \doteq \begin{bmatrix} KR^{-2} & (\mathbf{T}_h)^T \\ \mathbf{T}_h & KR^2 \end{bmatrix} \geq 0 \tag{3.7}$$

$$\mathbf{y} - \mathbf{T}_u \mathbf{P}\mathbf{p} - \mathbf{T}_u \mathbf{h} \in \mathcal{N}$$

where

$$\mathbf{P} \doteq \begin{bmatrix} g_0^1 & g_0^2 & \cdots & g_0^{N_p} \\ g_1^1 & g_1^2 & \cdots & g_1^{N_p} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N-1}^1 & g_{N-1}^2 & \cdots & g_{N-1}^{N_p} \end{bmatrix} \tag{3.8}$$

g_k^i *denotes the k-th Markov parameter of the i-th transfer function $G_{p_i}(z)$, h_k is the k-th Markov parameter of the nonparametric component $H(z)$, $\mathbf{T}_u$ and $\mathbf{T}_h$ are the lower Toepliz matrix associated with the finite sequence* $\mathbf{u}$ *and column vector* $\mathbf{h}$ *respectively, and $R = diag[1, \rho, \rho^2, \cdots, \rho^{N-1}]$*

Proof. From theorem 2.3.6 in [12], it follows that, given a finite sequence $\{\hat{\mathbf{h}}_i\}_{i=0}^{N-1}$ there exists $\hat{H} \in \overline{\mathcal{B}\mathcal{H}}_\infty$ such that $\hat{H} = \hat{\mathbf{h}}_o + \ldots + \hat{\mathbf{h}}_{N-1} z^{N-1} + \ldots..$ if and only if $\left(\mathbf{T}_{\hat{h}}\right)^T \mathbf{T}_{\hat{h}} \leq \mathbf{I}$. Combining the fact that

$$H(z) \in \overline{\mathcal{B}\mathcal{H}}_{\infty,\rho}(K) \Longleftrightarrow \hat{H}(z) = \frac{1}{K} H(\rho z) \in \overline{\mathcal{B}\mathcal{H}}_\infty$$

with coefficients $\hat{h}_k = \rho^k h_k/K$, after some algebra, the existence of $H(z) \in \overline{\mathcal{BH}}_{\infty,\rho}(K)$ is equivalent to

$$KR^{-2} - \frac{1}{K}\left(\mathbf{T}_h\right)^T R^{-2}\left(\mathbf{T}_h\right) \geq 0 \tag{3.9}$$

Now let's recall the widely used **Schur Complement** and its property:

Definition 2. *In linear algebra and the theory of matrices, the Schur complement (named after Issai Schur) of a block of a matrix within the larger matrix is defined as follows. Suppose A, B, C, D are respectively $p \times p$, $p \times q$, $q \times p$ and $q \times q$ matrices, and D is invertible. Let*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

*so that M is a $(p+q) \times (p+q)$ matrix. Then the **Schur complement** of the block D of the matrix M is the $p \times p$ matrix*

$$A - BD^{-1}C$$

One of its property says: If M is a positive definite symmetric matrix, then so is the Schur Complement in D of M,

i.e.,

$$\begin{bmatrix} A & B \\ B^H & D \end{bmatrix} \geq 0 \Longleftrightarrow D > 0 \quad \text{and} \quad A - BD^{-1}B^H \geq 0. \tag{3.10}$$

Then, the first condition in (3.7) can be easily obtained by applying Schur Complement on inequality (3.9).

The second condition is a restatement of the conditions

$$\begin{aligned} s &= h + \mathbf{p}^T g_p, \\ y_k &= (s * u)_k + \eta_k \text{ for some } \eta_k \in \mathcal{N} \end{aligned}$$

□

Once the consistency is established, a nominal model can be obtained through:

1. Find a pair of feasible solutions **p**,**h** satisfying the LMIs (3.7);

2. Given the feasible pair $(\mathbf{p}, \mathbf{h})$, the identified system can be computed as:

$$G_{id}(z) = H(z) + \mathbf{p}^T G_p(z) \tag{3.11}$$

where one admissible model of $H(z)$ can be chosen as the central[5] interpolant with the following state-space realization (see [6] and [79] for details):

$$\begin{aligned} \mathbf{A}_H &= \rho\big[\mathbf{A} - \mathbf{X}^{-1}\mathbf{C}_-^T\mathbf{C}_-(\mathbf{A}-\mathbf{I})\big]^{-1} \\ \mathbf{B}_H &= \rho\big[\mathbf{C}_-^T\mathbf{C}_-(\mathbf{A}^T - \mathbf{A} - \mathbf{I}) - (\mathbf{A}^T - \mathbf{I})\mathbf{M}_R\mathbf{A}\big]^{-1}\mathbf{C}_-^T \\ \mathbf{C}_H &= KC_+\{\mathbf{I} - \big[\mathbf{A} - \mathbf{X}^{-1}\mathbf{C}_-^T\mathbf{C}_-(\mathbf{A}-\mathbf{I})\big]\} \\ \mathbf{D}_H &= K\mathbf{C}_+\big[\mathbf{X}\mathbf{A} - \mathbf{C}_-^T\mathbf{C}_-(\mathbf{A}-\mathbf{I})\big]^{-1}\mathbf{C}_-^T \end{aligned} \tag{3.12}$$

[5]Recall that all the solutions for the nonparametric component can be parameterized as a lower LFT in terms of a free parameter $Q(z) \in \mathcal{BH}_{\infty,\rho}$. The central interpolant corresponds to the choice of $Q(z) = 0$

with

$$\mathbf{X} = \mathbf{C}_{-}^{T}\mathbf{C}_{-} + \left(\mathbf{A}^{T} - \mathbf{I}\right)\mathbf{M}_{R},$$

and

$$\mathbf{A} = \begin{pmatrix} 0 & \mathbf{I}_{(N-1)\times(N-1)} \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{C}_{-} = [1, 0, \cdots, 0]_{N}, \qquad \mathbf{C}_{+} = \frac{\mathbf{h}^{T} R\rho}{K}$$

3.2 Semi-Blind Robust Identification

In the previous section, we identify a system G by assuming that the experimental data are generated exactly from initial time $t_0 = 0$. Ignoring the effects of inputs prior to the initial time t_0 would lead to artificially high identification errors when the situation of only partially input-output experimental data (the data after time t_0) are exactly known happens. This motivates the following so-called *semi-blind* identification.

3.2.1 Information Consistency

Encapsulating the effect of inputs prior to time t_0 in some unknown, non-zero initial conditions $\mathbf{x}_0$, *semi-blind* robust identification problem is stated as:

Problem 4. *Given an unknown plant,* a priori *sets of candidate models and noise* $\left(\mathcal{S}, \mathcal{N}\right)$ *and a finite set of samples of the input* $\mathbf{u}$ *to the plant and its corresponding output* $\mathbf{y}$ *corrupted by additive measurement noise* η*, find a model* g *compatible with both*

the a priori *information and the* a posteriori *experimental data, that is* $g \in \mathbf{T}(\mathbf{y})$*, with*

$$\mathbf{T}(\mathbf{y}) \doteq \{g \in \mathcal{S}\colon y_k - (\sum_{i=0}^{k} g_i u_{k-i} + C_g * A_g^{k-1}\mathbf{x}_0) \in \mathcal{N}, k = 0, 1, \cdots, N-1, \quad \textit{for some } \mathbf{x}_0\} \tag{3.13}$$

where

$$g \equiv \left(\begin{array}{c|c} A_g & B_g \\ \hline C_g & D_g \end{array}\right) \qquad g_0 = D_g \quad g_i = C_g A_g^{i-1} B_g$$

Assuming our system is controllable, with some admissible past input sequence $\{u_k, k = -1, -2, \cdots\} \in \mathcal{U}^-$ acting on $(-\infty, -1]$, any physically meaningful initial condition $\mathbf{x}_0$ can be achieved by driving the system from the very beginning $x(-\infty) = 0$. That is, any output sequence $\{y_k, k = 0, 1, \cdots, N-1\}$ generated from the input sequence $u = \{u_k, k = 0, 1, \cdots, N-1\}$ with an initial condition of $\mathbf{x}_0$,

$$y_k = \sum_{i=0}^{k} g_i u_{k-i} - C_g * A_g^{k-1}\mathbf{x}_0, k = 0, 1, \cdots, N-1 \tag{3.14}$$

can be replaced, shown as follows, by using the effect of the unknown past input sequence $u^- \doteq \{u_k, k = -1, -2, \cdots\}$

$$y_k = \left(\mathbf{T}_g \mathbf{u}^-\right)_k + \left(\mathbf{\Gamma}_g \mathbf{u}\right)_k, k = 0, 1, \cdots, N-1 \tag{3.15}$$

where $\mathbf{T}_g$ and $\mathbf{\Gamma}$ reperesent the Toeplitz and Hankel operators associated with the system g respectively. Therefore, the semi-blind identification problem can be restated as below:

Problem 5. *Given:*

1. *an unknown plant* $G(z) = H(z)+P(z)$*, where* $H(z)$ *and* $P(z)$ *are the nonparametric portion and parametric portion of plant* $G(z)$*, respectively, as stated in equation* (3.5)*;*

2. a priori *sets of candidate models and noise* $(\mathcal{S}, \mathcal{N})$*;*

3. *a characterization of the set of the past input sequence* $\mathbf{u}^- \doteq \{u_k, k = -1, -2, \cdots, -N^-\} \in \mathcal{U}^-$ *applied prior to time* t_0*,*

4. *a finite set of samples of the input* $\mathbf{u} = \{u_0, u_1, \cdots, u_{N-1}\}$ *and output data* $\mathbf{y} = \{y_0, y_1, \cdots, y_{N-1}\}$*, corrupted by additive measurement noise* η

Find whether the set $\mathcal{T}(\mathbf{y})$ *is non empty, where*

$$\mathcal{T}(\mathbf{y}) = \{G \in \mathcal{S} \colon \mathbf{y} = \mathbf{T}_G^N \mathbf{u} + \mathbf{\Gamma}_G^{N^-} \mathbf{u}^- + \eta \text{ for some sequence } \eta \in \mathcal{N}, \text{and } \mathbf{u}^- \in \mathcal{U}^-\} \tag{3.16}$$

If $\mathcal{T}(\mathbf{y})$ *is non empty, then find a model* $G \in \mathcal{T}(\mathbf{y})$*.*

The set $\mathcal{T}(\mathbf{y})$ defined in (3.16) is the set of all systems G compatible with the *a priori* information that could have generated the observed experimental data. The first term in (3.16) gives the response of G to the input $\mathbf{u}$ and the second represent the responses of G to the input $\mathbf{u}^-$. Consequently, (3.16) is just a restatement of the fact that the *a priori* information and the *a posteriori* experimental data are consistent.

3.2.2 A convex relaxation

By invoking Lemma 1, we have

$$\mathbf{y} - \mathbf{T}_u \mathbf{P}\mathbf{p} - \mathbf{T}_u \mathbf{h} - \mathbf{\Gamma}_G^{N^-} \mathbf{u}^- \in \mathcal{N}, \tag{3.17}$$

which leads to a BMIs in h_k, the k–th Markov parameter of nonparametric component $H(z)$, and $\mathbf{u}^-$, due to the multiplication of $\mathbf{\Gamma}_G^N \mathbf{u}^-$. However, BMIs generically lead to non-convex, NP-hard optimization problems. To avoid this difficulty, we propose a convex relaxation of Problem 5 in the sequel.

Assuming that the set of admissible past inputs has the form

$$\mathcal{U}^- = \mathcal{B}\ell^p(K_{u^-}), \quad \text{i.e.} \quad \|\mathbf{u}^-\|_{\ell_q} \leq K_{u^-};$$

and that a bound $\|\mathbf{\Gamma}_G^{N^-}\|_{\ell_q \to \ell_\infty} \leq K_G$ is available as part of the *a priori* information, the following convex relaxation can be obtained for Problem 5 by replacing $\mathbf{\Gamma}_G^{N^-} \mathbf{u}^-$ by a new variable $\mathbf{x}$:

Problem 6. *Given an unknown plant $G(z)$ as stated in equation* (3.5)*,* a priori *sets of candidate models and noise $(\mathcal{S}, \mathcal{N})$, a bound K_{u^-} on the norm of sequence $\mathbf{u}^-$ and a finite set of samples of the input-output data $(\mathbf{u}, \mathbf{y})$. Find whether*

$$\mathcal{T}(\mathbf{y}) = \{G \in \mathcal{S} \colon \mathbf{y} = (\mathbf{T}_H^N + \mathbf{T}_P^N)\mathbf{u} + \mathbf{x} + \eta \text{ for some } \|\mathbf{x}\|_{\ell_q} \leq K_G K_{u^-} \text{ and some sequence } \eta \in \mathcal{N}\} \tag{3.18}$$

is empty. If $\mathcal{T}(\mathbf{y})$ *is non empty, then find a model* $G \in \mathcal{T}(\mathbf{y})$.

Straightforward application of Lemma 1 leads now to the following results:

Proposition 1. *Problem 6 has a solution if and only if the following set of LMIs in* $\mathbf{h}$ *and* $\mathbf{x}$ *is feasible:*

$$\begin{aligned} \mathbf{M}(\mathbf{h}) &\doteq \begin{bmatrix} KR^{-2} & \left(\mathbf{T}_h\right)^T \\ \mathbf{T}_h & KR^2 \end{bmatrix} \geq 0 \\ \mathbf{y} - \mathbf{T}_u\mathbf{P}\mathbf{p} - \mathbf{T}_u\mathbf{h} - \mathbf{x} &\in \mathcal{N} \\ \|\mathbf{x}\|_{\ell_q} &\leq K_G K_u - \end{aligned} \tag{3.19}$$

where the last two inequalities should be interpreted in a component-wise sense.

3.3 Application on Macro-Economic Modeling

3.3.1 Motivation

Central to the robust control tools is the modeling of the system under consideration as a Linear system subject to norm bound uncertainty. Robust optimization techniques can then be used to optimize the worst-case performance of the economical and financial system models([87, 100, 88, 63]).

Success of these techniques hinges upon the ability to obtain both a nominal model and an uncertainty description suitable to be used in a robust optimization context. Classical Bayesian analysis has been used extensively in economical systems modeling. In this approach, a parametric model, based on first principle is postulated and *a priori* probability distributions are assigned to each parameter. However, obtaining these

parametric models and the associated probability distributions poses difficult practical problems([81]). In addition, these probability distributions should be validated, as more experimental data becomes available. [109] proposes to accomplish this by incorporating learning to update posteriors. However, since this approach is stochastic in nature, it can neither provide the uncertainty bounds required by the robust control tools nor conclusively invalidate the *a priori* assumptions.

To avoid these difficulties, we apply robust identification and model (in)validation tools to obtain and validate models of economic sub–systems, together with a worst–case uncertainty description. The benefits of this approach are multiple:

1. It allows for addressing economic modeling problems from an input–output point of view. Thus, it does not require knowledge of a state space realization of the system, or even its order.

2. It allows to naturally incorporate knowledge of the system dynamics whenever it is available.

3. It provides mechanisms to invalidate *a priori* assumptions about the dynamics of the system and to detect when these assumptions or the identified model are no longer valid.[6]

4. It provides worst-case estimates of the identification error that can be used to determine for how long the predictions of the model will be valid, in the absence of additional data.

[6] addressed in Chapter 4.

The potential of the application of Robust Identification technique[7] to macro-economic modeling is illustrated by establishing a model of the subsystem of the US economy relating FED funds rates to inflation. The resulting model is able to correctly forecast the inflation rate over 149 quarters of real data, outperforming currently existing comparable linear models.

3.3.2 The *a priori* Information and *a posteriori* experimental data

The experimental data, obtained from the websites *www.marketvector.com* and *www.federalreserve.gov*, consists of historical values of the quarterly U.S. FED funds rate in percent at an annual rate, i_k, and the corresponding measured values of the quarterly inflation in the Gross-Domestic Product(GDP) chain–weighted price index in percent at an annual rate, y_k. In order to account for the difficulty in exactly measuring inflation, we assume that the measured inflation is given as

$$y_k = \pi_k + \eta_k,$$

where y_k, π_k, η_k denote the measured inflation, the actual inflation and additive noise, respectively. We postulate a simplified scenario where the inflation π_k is assumed to be the output of an unknown, stable LTI system in response to the input sequence i_k. Note that only a partial (in this case post first quarter of 1961, 1961.1) experimental data record is available which lead to the semi-blind identification problem.

[7] the application of model (in)validation is shown in Chapter 4.

In order to apply the framework introduced in section 3.2, we need a characterization of the set of past inputs $\mathcal{U}^-$. Since the FED funds rate has changed considerably over the period under consideration, the quarterly *change* in the FED funds rate is substantially lower and more uniform across the period of interest than the FED funds rate itself. Thus, identify a system G mapping the *change* in the FED funds rate,

$$u_k = i_k - i_{k-1},$$

to inflation π_k rather than mapping the FED funds rate i_k to inflation π_k leads to a much tighter bound on the past inputs. This system then should include an integrator in its parametric portion. Next, from a Fourier analysis of the input/output data, an upper bound of ρ is determined as $\rho \leq 1.02$. With these assumptions, the *a priori* information is shown as below:

$$\begin{aligned}
\mathcal{S} &= \left\{ H(z) = p_1 \frac{z}{z-1} + G_{np}(z) \,|\, G_{np}(z) \in \mathcal{BH}_{\infty,\rho}(K) \right\} \\
\mathcal{N} &= \{\eta \in \ell_\infty : |\eta_k| \leq \epsilon\}, \epsilon = 0.2 \\
\mathcal{U}^- &= \left\{ u \in \ell_\infty : |u_k| \leq u_{max} \right\}, u_{max} = 0.5 \\
x_k &= p \cdot i_k + x_k^{Np}, \; |x_k^{Np}| \leq K \frac{\rho^{-N}}{\rho - 1} 0.5
\end{aligned} \tag{3.20}$$

The last equation comes from the facts that:

$$
\begin{aligned}
x_k &= p\sum_{j=-\infty}^{k}(i_j - i_{j-1}) + \mathbf{\Gamma}_H \mathbf{u}^- = pi_k + \mathbf{\Gamma}_H \mathbf{u}^- \\
|\left(\mathbf{\Gamma}_H \mathbf{u}^-\right)_k| &= |\sum_{j=-\infty}^{-1} h_{k-j} u_j| \leq |u_{max}| \cdot \sum_{k+1}^{\infty} |h_j| \\
&\leq |u_{max}| \cdot K \cdot \sum_{k+1}^{\infty} \rho^{-k} = |u_{max}| K \frac{\rho^{-k}}{\rho - 1}
\end{aligned}
\tag{3.21}
$$

where the last inequality stems from:

$$
\mathbf{H} \in \overline{\mathcal{B}}\mathcal{H}_{\infty,\rho}(K) \Longrightarrow |h_k| \leq K\rho^{-k}
$$

3.3.3 Identification Results

Running the semi-blind robust identification algorithm on the *a priori* information (3.20) only using $N = 21$ historical values of inflation and FED funds rates(from 1961.1 to 1966.1), a 4-th order model[8] is obtained after eliminating unobservable/uncontrollable states:

$$
\begin{aligned}
\pi_{k+4} &= 0.526\pi_{k+3} + 0.074\pi_{k+2} - 0.13\pi_{k+1} + 0.506\pi_k \\
&+ 0.522u_{k+4} + 0.208u_{k+3} + 0.184u_{k+2} + 0.267u_{k+1} - 0.024u_k
\end{aligned}
\tag{3.22}
$$

[8] Here, we select the central model

For comparison, let's introduce an empirical, simplified model of this subsystem, the Rudebusch and Svensson(RS) model, which is already available in the literature [82]:

$$\begin{aligned} \pi_{t+1} &= 0.7\pi_t - 0.1\pi_{t-1} + 0.28\pi_{t-2} + 0.12\pi_{t-3} + 0.14y_t + \epsilon_{t+1} \\ y_{t+1} &= 1.16y_t - 0.25y_{t-1} - 0.1(\bar{i_t} - \bar{\pi_t}) + \eta_{t+1} \end{aligned} \tag{3.23}$$

where $\bar{\pi} = \frac{1}{4}\sum_{k=0}^{3}(\pi_{t-k})$ is the 4-quarter average of the inflation, $y_t = 100(\frac{q_t - q_t^*}{q_t^*})$ is the percentage gap between actual real(q_t) and potential GDP(q_t^*), $\bar{i_t} = \frac{1}{4}\sum_{k=0}^{3}(i_{t-k})$ is the 4-quarter average of the FED funds rate and ϵ_{t+1},η_{t+1} are i.i.d zero mean disturbances. This model was obtained by least squares fitting of historical data from the first quarter of 1961 through the second quarter of 1996. As shown in Figure 3.2, the RS model fits the training data, however, that

1. this technique requires a large amount of data points, while at the same time not exploiting *a priori* information available about the subsystem in question, and,

2. no conclusions can be drawn about modeling and future prediction errors.

These difficulties are all avoided in model (3.22) by utilizing the Robust Identification. Moreover, the forecasting power of the model (3.22) is illustrated in Figure 3.1, where it was used to estimate the inflation for the entire 1961.1-2003.2 quarters, using as inputs past historical data for inflation and changes in FED funds rates. As shown there, the model (3.22) yields a worst-case error of 3.2 comparing the RS model (3.23) has a worst-case error of 3.75, (roughly 11% higher) even though it was obtained considering 142 data points.

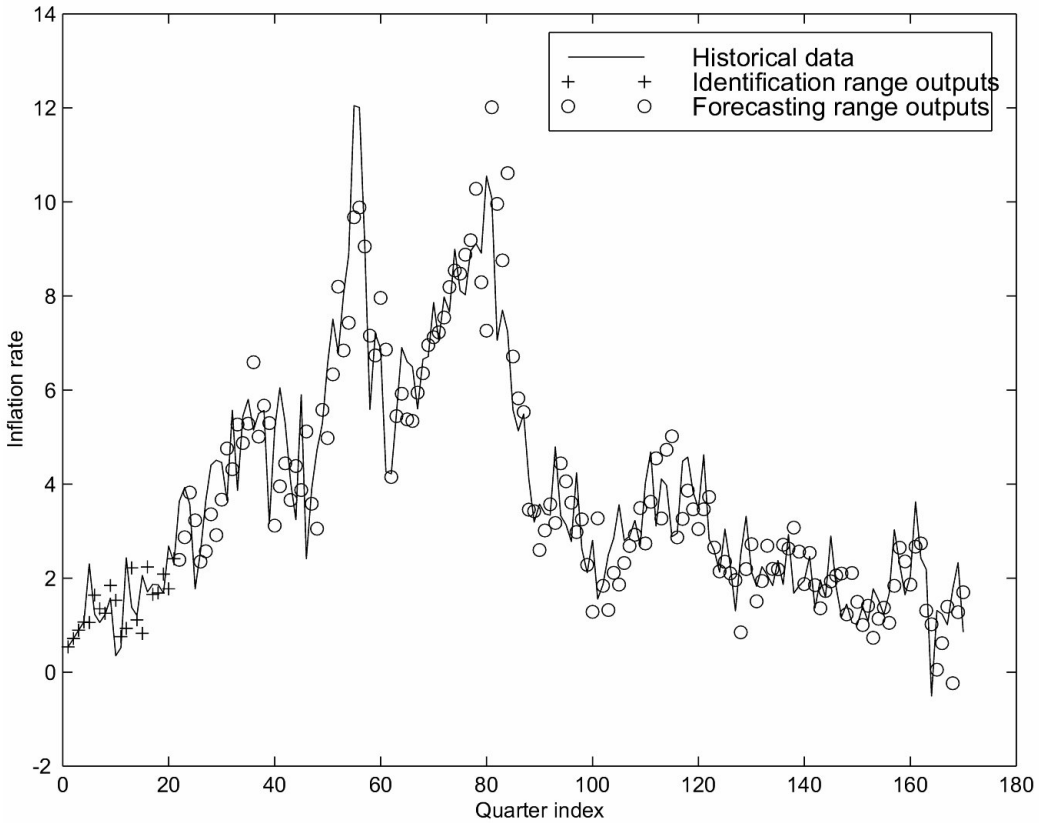

Fig. 3.1. Historical Inflation Data and the identification of the macro-economy subsystem with ℓ_∞–norm bounded observation noise

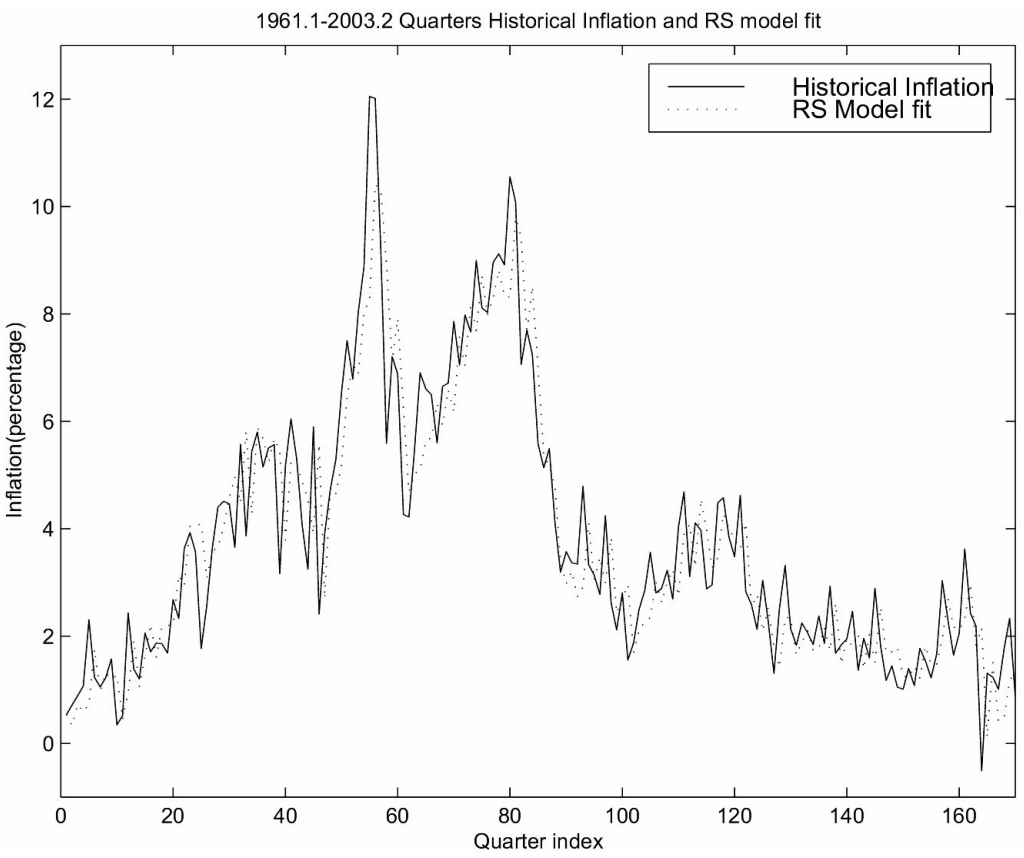

Fig. 3.2. Historical Inflation Data and RS model fit of the macro-economy subsystem

3.3.4 Worst-Case prediction error bounds

One of the advantages of Robust Identification is its ability to provide worst-case bounds on the prediction error. Begin by noting that the algorithm is interpolatory, that is it produces a model inside the consistency set $\mathcal{T}(\mathbf{y})$. Thus, since the "true" system must also belong to the consistency set[9], it follows that, given the first N measurements y_i, $y = 0, \ldots, N-1$ a bound on the worst-case prediction error at $t = N$ is given by:

$$\begin{aligned} |e_N| &\leq \sup_{g_1, g_2 \in \mathcal{T}(\mathbf{y})} |\left[(\mathbf{T}_{g_1} - \mathbf{T}_{g_2})\mathbf{u} + \mathbf{\Gamma}_{g_1} \mathbf{u}^-_1 - \mathbf{\Gamma}_{g_2} \mathbf{u}^-_2\right]_N| \\ &= d\left[\mathcal{T}(\mathbf{y})\right] \leq \sup_{\mathbf{y}} d\left[\mathcal{T}(\mathbf{y})\right] = \mathcal{D}(\mathcal{I}) \end{aligned} \tag{3.24}$$

where $d(\cdot)$ and $\mathcal{D}(\mathcal{I})$ denote the diameter of the set $\mathcal{T}(\mathbf{y}$, in the sup-metric, and the diameter of information, respectively. Moreover, since the *a priori* sets $\mathcal{S}$ and the noise set $\mathcal{N}$ are both convex and symmetric, with points of symmetry $g_s = 0$ and $\eta_s = 0$ respectively, it can be shown (Refer to Lemma 10.3 in [86]) that:

$$\mathcal{D}(\mathcal{I}) \leq 2 \sup_{g \in \mathcal{T}(\mathbf{0})} | \, p \cdot i_N + \sum_{j=0}^{N} h_{N-j} \cdot u_j + K \cdot K_u - \frac{\rho^{-N}}{\rho^{-1}} \, | \tag{3.25}$$

where $\mathcal{T}(\mathbf{0})$ indicate the set of operators compatible with the zero outcome: $y_k = 0$, $k = 0, 1, \cdots, N-1$. This leads to the following convex optimization problem:

$$\max |p \cdot i_N + \sum_{j=0}^{N} h_{N-j} u_j + K \cdot K_u - \frac{\rho^{-N}}{\rho^{-1}}| \tag{3.26}$$

[9] As long as the *a priori* information is indeed correct

subject to:

$$\begin{aligned} \mathbf{M}(\mathbf{h}) &\doteq \begin{bmatrix} KR^{-2} & \left(\mathbf{T}^{N}_{h}\right)^{T} \\ \mathbf{T}^{N}_{h} & KR^{2} \end{bmatrix} \geq 0 \\ \|\mathbf{T}^{N}_{u}\mathbf{P}\mathbf{p} + \mathbf{T}^{N}_{u}\mathbf{h} + \mathbf{x}\|_{\ell_2} &\leq \epsilon \\ \|\mathbf{x}\|_{\ell_q} &\leq K_G K_u - \end{aligned} \tag{3.27}$$

Next, we illustrate the use of this bound by obtaining an estimate of the prediction error at $N = 22$, the first historical data point not used in identification. Solving the problem above with $u_{max} = 0.6$ and $i_{max} = 6$, yields $|e_{22}| \leq 2.37$ (the actual error, computed a posteriori, is 1.05). The bounds corresponding to $N = 25$ and $N = 30$ are 1.99 and 1.95, showing that, as expected, the error gets smaller as more points are used in the identification.

3.3.5 A White–noise–like Observation Noise

As of now, we assume the observation noise η to be ℓ_∞–norm bounded, but otherwise unrestricted. However, to be more in the spirit of economic literature (see [63, 64, 1, 66, 20]), in this subsection, we propose an extension of using a different characterization of observation noise. We make the assumption of a ℓ_2–norm bounded noise and furthermore allow the noise to resemble a stochastic process, white–noise–like sequence. Comparing with the previous the *a priori* information and the *a posteriori* experimental data in subsection 3.3.2, the only change is the noise set $\mathcal{N}$, shown as

following:

$$\mathcal{N} = \left\{ \eta \in \mathcal{B}\ell_2(\epsilon) \colon \eta \text{ resembles a white-noise-like sequence} \right\}, \epsilon = 0.7 \tag{3.28}$$

where $\epsilon = 0.7$ indicates that the noise is ℓ_2-norm bounded by 10% of the ℓ_2 norm of the historical output, 0.7. This noise set can be interpret by two constraints:

1. noise sequence η is energy bounded:

$$\eta^T \eta \leq \epsilon^2,$$

which is trivial and can be reduced to an LMI:

$$\begin{bmatrix} \epsilon^2 & \eta^T \\ \eta & 1 \end{bmatrix} \geq 0, \tag{3.29}$$

according to Schur complement argument.

2. noise sequence η resembles a white-noise-like(uncorrelated) sequence. According to [69], this constraint can be interpreted as (for detail, see in Appendix C):

$$|\mathbf{R}_\eta(\tau)| \leq \gamma \cdot \mathbf{R}_\eta(0), \tau = 1, 2, \cdots, \tau_0 - 1 \tag{3.30}$$

where $\mathbf{R}_\eta(\tau)$ represents the circular autocorrelation of signal sequence η with time gap τ, γ represents the decay rate which indicates the accuracy of the white-noise-like, and τ_0 represents the maximum possible time constant of $G_{np}(z)$, respectively.

In the following, the second constrain of the noise set $\mathcal{N}$ in 3.30 is shown to be non-convex and we must reduce it to a convex presentation, in order to apply the identification technique, owning to the attribute of the Semi–Blind Robust Identification procedure. Two different approaches are explored later in this subsection.

Combining

$$\begin{aligned}\mathbf{R}_{\eta}(\tau) =& E\{\eta(k+\tau)\eta(k)\}\\ =& \eta(\tau)\eta(0) + \eta(\tau+1)\eta(1) + \cdots + \eta(N-1)\eta(N-1-\tau)\\ =& \eta^{T}(\tau+1:N)\cdot\eta(1:N-\tau)\end{aligned}$$

with

$$\begin{aligned}\mathbf{R}_{\eta}(0) =& E\{\eta(k)\eta(k)\}\\ =& \eta^{T}\cdot\eta \leq \epsilon^{2},\end{aligned}$$

inequality (3.30) can be rewritten as:

$$|\eta^{T}(\tau+1:N)\cdot\eta(1:N-\tau)| \leq \gamma\cdot\epsilon^{2} \tag{3.31}$$

Notice that as noise sequence η is an unknown variable, inequality (3.31) is not jointly convex on account of the non–symmetric multiplication of $\eta^{T}(\tau+1:N)\cdot\eta(1:N-\tau)$. Thus, we seek some relaxation to reduce this non-convex problem to a set of LMIs.

Noting

$$\eta(\tau+1:N) = \eta_{n} \quad \text{and} \quad \eta(1:N-\tau) = \eta_{1},$$

then (3.31) can be separated as two parts:

$$\gamma \cdot \epsilon^2 - \eta_n^T \eta_1 \geq 0 \tag{3.32}$$

$$\gamma \cdot \epsilon^2 + \eta_n^T \eta_1 \geq 0 \tag{3.33}$$

Rewrite (3.32) as:

$$2\gamma\epsilon^2 - (\eta_n^T + \eta_1^T)(\eta_n + \eta_1) + \eta_1^T \eta_1 + \eta_n^T \eta_n \geq 0$$

Following the assumption of

$$0 \leq \eta_n^T \eta_n \leq \epsilon^2 \quad \text{and} \quad 0 \leq \eta_1^T \eta_1 \leq \epsilon^2,$$

we relax (3.32) to:

$$2\gamma\epsilon^2 + 2\epsilon^2 - (\eta_n^T + \eta_1^T)(\eta_n + \eta_1) \geq 0$$

$\Longleftrightarrow$

$$\begin{bmatrix} 2(\gamma+1)\epsilon^2 & \eta_n^T + \eta_1^T \\ \eta_n + \eta_1 & 1 \end{bmatrix} \geq 0 \tag{3.34}$$

The same as (3.32), inequality (3.33) can be rewritten as:

$$2\gamma\epsilon^2 - (\eta_n^T - \eta_1^T)(\eta_n - \eta_1) + \eta_1^T \eta_1 + \eta_n^T \eta_n \geq 0$$

and relaxed to:

$$2\gamma\epsilon^2 + 2\epsilon^2 - (\eta_n^T - \eta_1^T)(\eta_n - \eta_1) \geq 0$$

$\Longleftrightarrow$

$$\begin{bmatrix} 2(\gamma+1)\epsilon^2 & \eta_n^T - \eta_1^T \\ \eta_n - \eta_1 & 1 \end{bmatrix} \geq 0 \tag{3.35}$$

Summarily, after some relaxation, the non–convex noise set in 3.28 is reduced to a set of LMIs as following:

$$\begin{bmatrix} \epsilon^2 & \eta^T \\ \eta & 1 \end{bmatrix} \geq 0,$$

$$\begin{bmatrix} 2(\gamma+1)\epsilon^2 & \eta_n^T + \eta_1^T \\ \eta_n + \eta_1 & 1 \end{bmatrix} \geq 0, \tag{3.36}$$

$$\text{and} \quad \begin{bmatrix} 2(\gamma+1)\epsilon^2 & \eta_n^T - \eta_1^T \\ \eta_n - \eta_1 & 1 \end{bmatrix} \geq 0$$

Next, let's introduce an alternative relaxation. Consider the Hankel matrix of the noise sequence

$$\mathbf{\Gamma} = \begin{bmatrix} \eta(1) & \eta(2) & \eta(3) & \cdots & \cdots & \eta(N) \\ \eta(2) & \eta(3) & \cdots & \cdots & \eta(N) & \eta(1) \\ \cdots & \cdots & \cdots & \eta(N) & \eta(1) & \eta(2) \\ \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \eta(N) & \eta(1) & \cdots & \cdots & \eta(N-2) & \eta(N-1) \end{bmatrix}$$

where N is the total length of the noise sequence, i.e., the length of the input–output data sequence. It is easy to see that:

$$\mathbf{\Gamma}^T\mathbf{\Gamma} = \begin{bmatrix} \mathbf{R}_\eta(0) & \mathbf{R}_\eta(1)+\mathbf{R}_\eta(N-1) & \cdots & \mathbf{R}_\eta(N-1)+\mathbf{R}_\eta(1) \\ \mathbf{R}_\eta(1)+\mathbf{R}_\eta(N-1) & \mathbf{R}_\eta(0) & \cdots & \cdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{R}_\eta(N-1)+\mathbf{R}_\eta(1) & \cdots & \cdots & \mathbf{R}_\eta(0) \end{bmatrix} \tag{3.37}$$

Further,

$$\begin{aligned} \lambda_i(\mathbf{\Gamma}^T\mathbf{\Gamma}) &= \sigma_i(\mathbf{\Gamma}^T\mathbf{\Gamma}) \\ &= \left[\mathbf{F}^H(\omega_i)\mathbf{F}(\omega_i)\right]^{\frac{1}{2}} \end{aligned} \tag{3.38}$$

where λ_i, σ_i are the eigenvalues and singular values of matrix $\mathbf{\Gamma}^T\mathbf{\Gamma}$, respectively. $\mathbf{F}(\omega_i)$ is the discrete fourier transform of the matrix $\mathbf{\Gamma}^T\mathbf{\Gamma}$ (see in [96]).

The following can be derived by simple algebra from equations (3.37) and (3.38):

$$\begin{aligned}
\overline{\sigma}\left(\mathbf{\Gamma}^T\mathbf{\Gamma}\right) =& \sigma_{i=0}(\mathbf{\Gamma}^T\mathbf{\Gamma}) \\
=& \mathbf{R}_\eta(0) + 2\big[\mathbf{R}_\eta(1) + \mathbf{R}_\eta(N-1)\big] + 2\big[\mathbf{R}_\eta(2) + \mathbf{R}_\eta(N-2)\big] \\
&+ \cdots + 2\big[\mathbf{R}_\eta((N-1)/2) + \mathbf{R}_\eta((N+1)/2)\big] \qquad \text{for } N = \text{odd} \\
\\
\overline{\sigma}\left(\mathbf{\Gamma}^T\mathbf{\Gamma}\right) =& \mathbf{R}_\eta(0) + 2\big[\mathbf{R}_\eta(1) + \mathbf{R}_\eta(N-1)\big] + 2\big[\mathbf{R}_\eta(2) + \mathbf{R}_\eta(N-2)\big] \\
&+ \cdots + 2\big[\mathbf{R}_\eta(N/2-1) + \mathbf{R}_\eta(N/2+1)\big] + 2\big[\mathbf{R}_\eta(N/2)\big] \qquad \text{for } N = \text{even}
\end{aligned} \tag{3.39}$$

Associate with inequality (3.30) and

$$\mathbf{R}_\eta(0) = E\{\eta(k)\eta(k)\} = \eta^T \cdot \eta \leq \epsilon^2,$$

we have:

$$\begin{aligned}
\overline{\sigma}\left(\mathbf{\Gamma}^T\mathbf{\Gamma}\right) &\leq \epsilon^2 + 2(N-1)\gamma\epsilon^2 \qquad \text{for } N = \text{odd} \\
\overline{\sigma}\left(\mathbf{\Gamma}^T\mathbf{\Gamma}\right) &\leq \epsilon^2 + 2N\gamma\epsilon^2 \qquad \text{for } N = \text{even}
\end{aligned} \tag{3.40}$$

which can be easily reduced to:

$$\begin{bmatrix} [2(N-1)\gamma+1]\epsilon^2 & \mathbf{\Gamma}^T \\ \mathbf{\Gamma} & \mathbf{I} \end{bmatrix} \geq 0 \qquad \text{for } N = \text{odd}$$

$$\begin{bmatrix} [2N\gamma+1]\epsilon^2 & \mathbf{\Gamma}^T \\ \mathbf{\Gamma} & \mathbf{I} \end{bmatrix} \geq 0 \qquad \text{for } N = \text{even} \tag{3.41}$$

Therefore, alternatively, when we use odd number of input–output experimental data to do the identification, noise set $\mathcal{N}$ in (3.28) can be relaxed as

$$\begin{bmatrix} [2(N-1)\gamma+1]\epsilon^2 & \mathbf{\Gamma}^T \\ \mathbf{\Gamma} & \mathbf{I} \end{bmatrix} \geq 0, \tag{3.42}$$

and when we use even number of input–output experimental data to do the identification, noise set $\mathcal{N}$ in (3.28) can be relaxed as

$$\begin{bmatrix} [2N\gamma+1]\epsilon^2 & \mathbf{\Gamma}^T \\ \mathbf{\Gamma} & \mathbf{I} \end{bmatrix} \tag{3.43}$$

For simplicity, here we only run the semi-blind robust identification algorithm outlined in section 3.2 with the relaxation (3.36) for noise set. Only using $N = 21$ historical values of inflation and FED funds rates(from 1961.1 to 1966.1), a 5-th order

model[10] is obtained after eliminating unobservable/uncontrollable states:

$$\begin{aligned}\pi_{n+5} &= 0.73\pi_{n+4} - 0.19\pi_{n+3} + 0.03\pi_{n+2} - 0.02\pi_{n+1} + 0.42\pi_n \\ &+ 0.53u_{n+5} + 0.10u_{n+4} + 0.23u_{n+3} + 0.21u_{n+2} + 0.18u_{n+1} + 0.03\end{aligned} \tag{3.44}$$

This time the forecasting power of the model (3.44) is illustrated in Figure 3.3, where it was used to estimate the inflation for the entire 1961.1-2003.2 quarters, using as inputs past historical data for inflation and changes in FED funds rates. As shown there, the model (3.44) yields a worst-case error[11] of 24% in terms of the percentage of ℓ_2–norm of the historical output.

3.4 Blind Robust Identification

Following the two previous sections 3.1 and 3.2, now we address the problem of Blind Robust Identification[12] (see [58]): Given a finite set of samples of noisy output data in the time domain and the *a priori* information, find an admissible *model and input pair* that make the *a priori* information consist with the *a posteriori* experimental data. This problem is close to Problem 3, except that now not only the plant and measurement noise but also the samples of input are unknown and are assumed to lie in some bounded set, $\mathcal{U}$. From Lemma 1, with $\mathbf{h}$ and $\mathbf{u}$ both unknown, the feasibility problem (3.7) turns to be non-convex in this case.

[10] Central model is selected

[11] Since in the *a priori* information , noise set is ell_2–norm bounded, we also scale the identification error in terms of ell_2–norm.

[12] The corresponding blind model (in)validation is studied in [56].

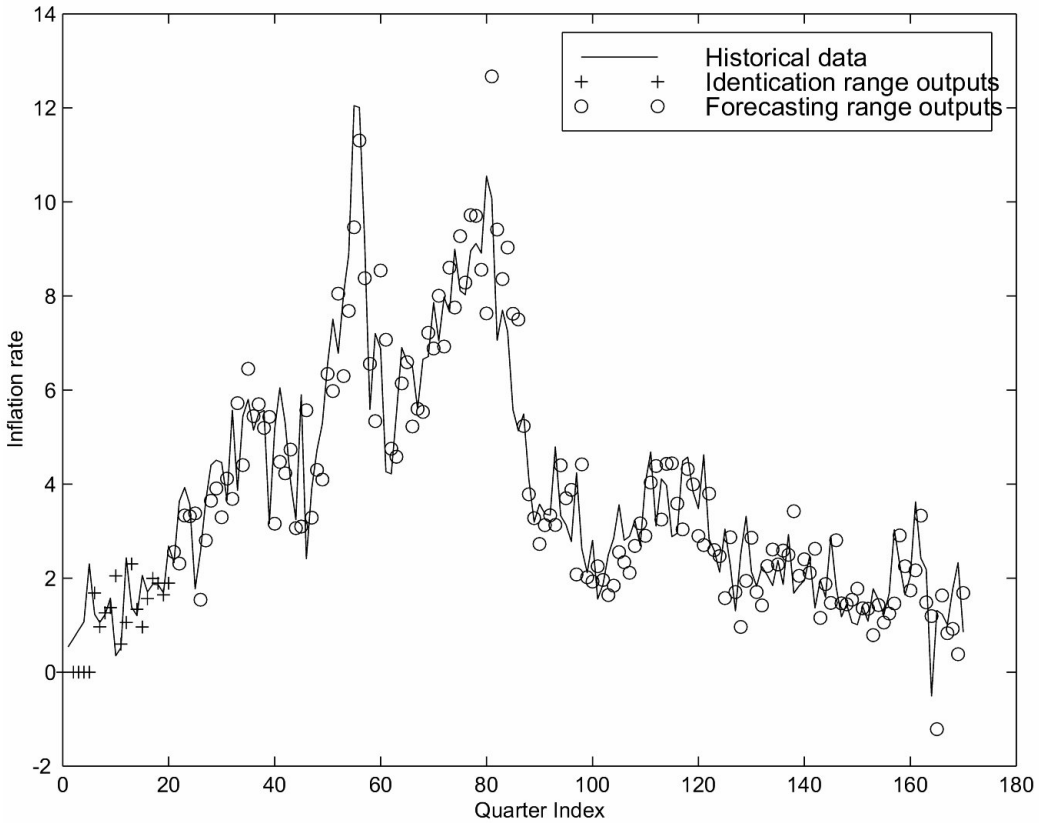

Fig. 3.3. Historical Inflation Data and the identification of the macro-economy subsystem with white–noise–like observation noise

With one suitable choice for the set of admissible inputs $\mathcal{U} = \mathcal{BH}_{\infty}^{N}$, let's begin by recalling an algorithm, developed in [46], that generates uniformly distributed finite samples over a bounded set. It will be used to obtain a convex, computationally tractable stochastic relaxation of the blind robust identification problem in the case of unstructured uncertainty and noisy measurements.

Algorithm 1.

Let $k = 0$. Generate N_1 samples uniformly distributed over the set

$$\{\mathbf{H}_0 : \overline{\sigma}(\mathbf{H}_0) \leq 1\}. \tag{3.45}$$

1. *Let $k := k + 1$. For every generated sample $(\mathbf{H}_0^i, \mathbf{H}_1^i, \ldots, \mathbf{H}_{k-1}^i)$, consider the partition*

$$\begin{bmatrix} \mathbf{H}_k^i & \cdots & \mathbf{H}_1^i & \mathbf{H}_0^i \\ \mathbf{H}_{k-1}^i & \cdots & \mathbf{H}_0^i & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{H}_0^i & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_k^i & \mathbf{B} \\ \mathbf{C} & \mathbf{A} \end{bmatrix} \tag{3.46}$$

and let the matrices $\mathbf{Y}$ and $\mathbf{Z}$ be a solution of the linear equations

$$\begin{aligned} \mathbf{B} &= \mathbf{Y}(\mathbf{I} - \mathbf{A}^T\mathbf{A})^{\frac{1}{2}}; \\ \mathbf{C} &= (\mathbf{I} - \mathbf{A}\mathbf{A}^T)^{\frac{1}{2}}\mathbf{Z}, \end{aligned} \tag{3.47}$$

2. *Let* $\mathbf{J}(\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{k-1}) \doteq |(\mathbf{I} - \mathbf{Y}\mathbf{Y}^T)^{\frac{1}{2}}|^m |(\mathbf{I} - \mathbf{Z}^T\mathbf{Z})^{\frac{1}{2}}|^s$. *Generate*

$$\left\lfloor N_s \mathbf{J}(\mathbf{H}_0^i, \mathbf{H}_1^i, \ldots, \mathbf{H}_{k-1}^i) \right\rfloor, \tag{3.48}$$

samples uniformly over the set $\{\mathbf{W} : \overline{\sigma}(\mathbf{W}) \leq 1\}$ *and for each of those samples* $\mathbf{W}^i$, *compute*

$$\mathbf{H}_k^i = -\mathbf{Y}\mathbf{A}^t\mathbf{Z} + (\mathbf{I} - \mathbf{Y}\mathbf{Y}^T)^{\frac{1}{2}}\mathbf{W}^i(\mathbf{I} - \mathbf{Z}^T\mathbf{Z})^{\frac{1}{2}}. \tag{3.49}$$

3. *If* $k \leq N$ *go to step 1. Otherwise, stop.*

Next, with N_s samples of the input, generated by Algorithm 1, we may attempt to *approximately* check consistency between the *a priori* information and the *a posteriori* experimental data. Here is the *Identification Algorithm*:

Algorithm 2. *Given the experimental data* $\mathbf{y}$, *the* a priori *information, choose* N_1 *and generate* N_s *samples of the input* $\{\mathbf{u}^i\}_{i=1}^{i=N_s}$ *from the set* $\mathcal{BH}_\infty^N$.

1. *For each* $\mathbf{u}^i$, *solve the convex problem* (3.7) *in* $\mathbf{h}$.

2. *If there exists at least one feasible* $\mathbf{h}$, *stop. With the feasible pair* $\mathbf{h}_{feas}$ *and* $\mathbf{u}_{feas} = \mathbf{u}^i$, *one admissible model will be computed as the central interpolant shown in equations* (3.12).

   Otherwise, consider next sample $\mathbf{u}^{i+1}$ *and go back to step 1.*

The algorithm 2 finishes either by finding one admissible pair of $(\mathbf{h}, \mathbf{u})$ or after N_s steps. This lead to one admissible model obtained, but not necessarily well suited for

application purposes. We may reformulate this algorithm by repeating Step 1 N_s times and by selecting $(\mathbf{h_{id}}, \mathbf{u_{id}})$ as the feasible pair that achieves the minimum value of K.

Remark 1. *In the case* $\mathcal{U} = \mathcal{BH}_{\infty,\rho}$ *for some* $\rho \geq 1$*, [46] gives an algorithm to generate uniformly distributed samples of the input,* $(u_0^i, u_1^i, \cdots, u_{N-1}^i), i = 1, \cdots, N_s$.

Remark 2. *If the number of samples is large enough, the risk of missing a consistent pair* $(\mathbf{h}, \mathbf{u})$ *can be made arbitrarily small.*

Chapter 4

Time-domain Model (In)Validation

This chapter summarizes Model (In)Validation for causal, LTI, discrete-time system with time-domain measurements, proposes two relaxations for semi-blind (in)validation, and presents the application of semi-blind (in)validation framework on the problem of FDI.

4.1 Model (In)Validation for LTI system

Derived from robust identification procedure, a description of the to be identified system, consists of a nominal model and an appropriate uncertainty description associate with it, is suitable to be used by control synthesis algorithms. However, due to the facts that sometimes the identification bound we obtained from identification procedure may be too conservative to be useful in practice or the *a priori* assumptions may be difficult to verify, from a practical standpoint, before using this description to synthesize controller, it should be validated using *new* experimental data, which has not been used in the identification process, to test its ability to represent the actual system's behavior.

4.1.1 Problem Formulation

In order to suite the robust control setting, a widely known description, Linear Fractional Transformation (LFT) form[1] is used for model (in)validation, shown in Figure 4.1.

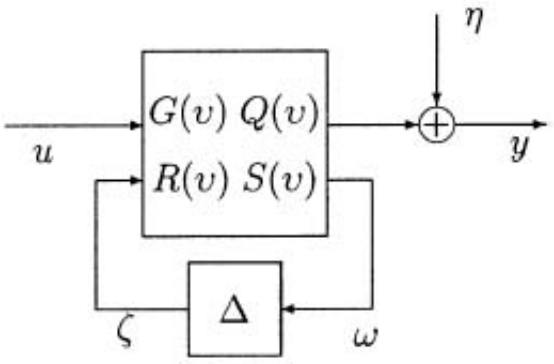

Fig. 4.1. The linear fractional uncertainty model

Consider a general class of uncertain system models described by a LFT, We can present the model by the input-output mapping:

$$y = \left[G + Q\Delta\left(\mathbf{I} - S\Delta\right)^{-1}R\right] * u + \eta \tag{4.1}$$

where signal u and y represent a test input and its corresponding output, respectively, and $*$ indicates convolution. η represents an additive measurement noise at the plant

[1] Two special cases of additive uncertainty and multiplicative uncertainty are addressed in the literature (see [45, 115] and [76, 92]).

output and is assumed to belong to a certain convex set $\mathcal{N}$, and Δ is the modeling uncertainty, assumed to belong to $\mathcal{BH}_\infty(\gamma)$. The quadruple $\big(G, Q, R, S\big)$ represents the nominal model of the plant G, stable and obtained through robust identification or given, and a description of how uncertainty enters the model, given by the blocks Q, R, S.

The model validation problem is to determine whether or not the corrupted experimental data of the input u, output y are consistent with the known model $\big(G, Q, R, S\big)$ and the the given noise set $\mathcal{N}$ and the uncertainty $\Delta \in \mathbf{\Delta}$, i.e.:

Problem 7. *Given the model* $\big(G, Q, R, S\big)$*, the* a priori *sets* $\mathcal{N}$ *and* $\mathbf{\Delta} \doteq \mathcal{BH}_\infty(\gamma)$*, and the experiments data* $\mathbf{u} \doteq [u_0, u_1, \cdots, u_{N-1}]^T$*,* $\mathbf{y} \doteq [y_0, y_1, \cdots, y_{N-1}]^T$*, determine if there exists a feasible pair* $\big(\Delta, \eta\big) \in \big(\mathcal{BH}_\infty(\gamma), \mathcal{N}\big)$ *such that the experimental data can be reproduced from the given model, i.e. equation* (4.1) *holds.*

If the answer is negative, then the given model does not provide a correct description of the physical system, which lead to the revision of the identification step.

4.1.2 (In)Validation with Time-domain measurements

In this subsection, we recast the (in)validation problem stated above in Problem 7 into LMI feasibility form.

To start, we recall a theorem addressed in [45]:

Theorem 3. *Given sequences* $u = \{u_0, u_1, \cdots, u_{l-1}\} \in \mathcal{R}^m$ *and* $y = \{y_0, y_1, \cdots, y_{l-1}\} \in \mathcal{R}^p$*, there exists a stable causal, linear, time-invariant operator* Δ *with*

$$\|\Delta\|_\infty \leq \gamma$$

and such that

$$\Delta(u_0, u_1, \cdots, u_{l-1}, *, *, \cdots) = (y_0, y_1, \cdots, y_{l-1}, *, *, \cdots)$$

if and only if

$$\mathbf{T}_y^T \mathbf{T}_y \leq \gamma^2 \mathbf{T}_u^T \mathbf{T}_u \tag{4.2}$$

where $\mathbf{T}_u$ *and* $\mathbf{T}_y$ *are the associated lower Toeplitz matrices formed form the sequences* u *and* y*, respectively.*

This theorem gives a necessary and sufficient condition for the existence of a causal, LTI operator mapping an input sequence u to an output sequence y.

Applying Theorem 3 and according to the structure of the system in Fig 4.1, we reduce the following LMIs:

Theorem 4. *Given time-domain measurements of the input* $\mathbf{u} = [u_0, u_1, \cdots, u_{n-1}]^T$ *and the output* $\mathbf{y} = [y_0, y_1, \cdots, y_{n-1}]^T$*, the LTI model* $\big(G, Q, R, S\big)$ *is not invalidated by this experimental data if and only if there exists a vector* $\zeta = [\zeta_0, \zeta_1, \cdots, \zeta_{n-1}]^T$*, such that*

$$\mathbf{M}(\zeta) \doteq \begin{bmatrix} \mathbf{X}_\zeta & \mathbf{T}_\zeta^T \\ \mathbf{T}_\zeta & \mathbf{Y}_\zeta \end{bmatrix} \geq 0 \tag{4.3}$$

$$\mathbf{y} - \mathbf{T}_G^n \mathbf{u} - \mathbf{T}_Q^n \zeta = \eta \in \mathcal{N}$$

where

$$\mathbf{X}_{\zeta} = (\mathbf{T}^{n}_{R}\mathbf{T}_{u})^{T}\mathbf{T}^{n}_{R}\mathbf{T}_{u} + (\mathbf{T}^{n}_{R}\mathbf{T}_{u})^{T}\mathbf{T}^{n}_{S}\mathbf{T}_{\zeta} + (\mathbf{T}^{n}_{S}\mathbf{T}_{\zeta})^{T}\mathbf{T}^{n}_{R}\mathbf{T}_{u}$$

$$\mathbf{Y}_{\zeta} = (\frac{1}{\gamma^2} - (\mathbf{T}^{n}_{S})^{T}\mathbf{T}^{n}_{S})^{-1}$$

hold

Proof. From Theorem 3 we have :

$$(\mathbf{T}_{\zeta})^{T}\mathbf{T}_{\zeta} \leq \gamma^{2}(\mathbf{T}_{\omega})^{T}\mathbf{T}_{\omega} \tag{4.4}$$

Structure of the model gives:

$$\begin{aligned} \mathbf{y} &= \mathbf{T}^{n}_{G}\mathbf{u} + \mathbf{T}^{n}_{Q}\zeta + \eta \\ \omega &= \mathbf{T}^{n}_{R}\mathbf{u} + \mathbf{T}^{n}_{S}\zeta \end{aligned} \tag{4.5}$$

Replace ω in the right hand side of (4.4) by the expression in (4.5), and reorder terms, the following yields:

$$\mathbf{T}^{T}_{\zeta}[\frac{1}{\gamma^2}\mathbf{I} - (\mathbf{T}^{n}_{S})^{T}\mathbf{T}^{n}_{S}]\mathbf{T}_{\zeta} \leq (\mathbf{T}^{n}_{R}\mathbf{T}_{u})^{T}\mathbf{T}^{n}_{R}\mathbf{T}_{u} + (\mathbf{T}^{n}_{S}\mathbf{T}_{\zeta})^{T}\mathbf{T}^{n}_{R}\mathbf{T}_{u} + (\mathbf{T}^{n}_{R}\mathbf{T}_{u})^{T}\mathbf{T}^{n}_{S}\mathbf{T}_{\zeta} \tag{4.6}$$

Finally, useing Schur complement, LMI (4.3) can be directly obtained. □

4.2 Semi-Blind Model (In)Validation

Motivated by the same reason of Semi-Blind Identification, in this section we study the problem of Semi-Blind (In)Validation. The situation here is also only the input-output experimental data applied after some initial time t_0 are known exactly.

4.2.1 Problem Formulation

As mentioned in section 3.2, the effect of the inputs priori to time t_0 can be encapsulate in some unknown, non-zero initial conditions x_0. Then, the *semi–blind* model (in)validation problem is stated as:

Problem 8. *Given experimental data* $\mathbf{y} = [y_0, y_1, \cdots, y_{n-1}]^T$,$\mathbf{u} = [u_0, u_1, \cdots, u_{n-1}]^T$, *an uncertain model described by an LFT as in Figure 4.1 and set descriptions* $\mathcal{N}, \mathbf{\Delta}$ *of admissible noise and uncertainty, determine if there exists at least one pair* $\big(\eta, \Delta\big) \in \big(\mathcal{N}, \mathbf{\Delta}\big)$ *that can reproduce*

$$\mathbf{y} = \mathbf{T}^{n}_{\hat{G}}\mathbf{u} + \mathbf{T}^{ic}_{\hat{G}}\mathbf{x_0} \tag{4.7}$$

for some unknown initial condition $\mathbf{x_0}$, *where*

$$\hat{G} = [G + Q\Delta(\mathbf{I} - S\Delta)^{-1}R] \tag{4.8}$$

presents the given uncertain system, $\mathbf{T}^{n}_{\hat{G}}$ *and* $\mathbf{T}^{ic}_{\hat{G}}$ *denote the operators that map the input and initial conditions of the uncertain system* $\hat{G}$ *to its output, respectively.*

Again, the same as in section 3.2, by driving the system from the very beginning $x(-\infty) = 0$ with some admissible past sequence $\{u_k, k = -1, -2, \cdots, -N^{-}\}$ to achieve

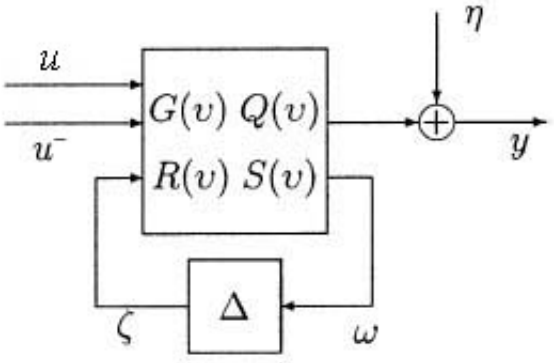

Fig. 4.2. System structure for semi-blind model (in)validation

any physically meaningful initial condition $\mathbf{x_0}$, the semi-blind (in)validation problem can be restated as following:

Problem 9. *Given:*

1. *system structure as in Figure 4.2 and the nominal plant* $\big(G, Q, R, S\big)$*;*

2. *Experimental data* $\mathbf{u} = [u_0, \cdots, u_{n-1}]^T$*, and* $\mathbf{y} = [y_0, \cdots, y_{n-1}]^T$*;*

3. *descriptions of noise set* $\mathcal{N}$*, and uncertainty set* $\mathbf{\Delta}$*:*

$$\mathbf{\Delta} \doteq \mathcal{BH}_\infty(\gamma);$$

4. *set description of the sequence*

$$\mathbf{u}^- \doteq \{u_k, k = -1, \cdots, -N^-\} \in \mathcal{U}^-$$

applied prior to time t_0*,*

determine if there exists at least one triple $\left(\Delta, \eta, \mathbf{u}^{-}\right) \in \left(\mathbf{\Delta}, \mathcal{N}, \mathcal{U}^{-}\right)$ *that can reproduce the experimental data:*

$$\mathbf{y} = \mathbf{T}^{n}_{\hat{G}}\mathbf{u} + \mathbf{\Gamma}^{N^{-}}_{\hat{G}}\mathbf{u}^{-} + \eta \tag{4.9}$$

where $\hat{G}$ *is the given uncertain system, shown in* (4.8)*, and* $\mathbf{\Gamma}_{\hat{G}}$ *denote the Hankel operator associate with the uncertain system* $\hat{G}$*.*

Straightforward application of theorem 3 shows that the model is not invalidated by the experimental data if and only if the following inequalities are feasible:

$$\mathbf{T}^{T}_{\zeta}\mathbf{T}_{\zeta} \leq \gamma^{2}\mathbf{T}^{T}_{\omega}\mathbf{T}_{\omega} \tag{4.10}$$

with some $\left(\Delta, \eta, \mathbf{u}^{-}\right) \in \left(\mathbf{\Delta}, \mathcal{N}, \mathcal{U}^{-}\right)$. Combing with

$$\omega = \mathbf{T}^{n}_{R}\mathbf{u} + \mathbf{\Gamma}^{N^{-}}_{R}\mathbf{u}^{-} + \mathbf{T}^{n}_{S}\zeta, \tag{4.11}$$

obtained from the structure in Figure 4.2, problem 9 is reduced to the feasibility of the inequality

$$\mathbf{T}^{T}_{\zeta}\mathbf{T}_{\zeta} \leq \gamma^{2}\left(\mathbf{T}^{n}_{R}\mathbf{T}_{u} + \mathbf{\Gamma}^{N^{-}}_{R}\mathbf{T}_{u^{-}} + \mathbf{T}^{n}_{S}\mathbf{T}_{\zeta}\right)^{T}\left(\mathbf{T}^{n}_{R}\mathbf{T}_{u} + \mathbf{\Gamma}^{N^{-}}_{R}\mathbf{T}_{u^{-}} + \mathbf{T}^{n}_{S}\mathbf{T}_{\zeta}\right) \tag{4.12}$$

Unfortunately, (4.12) is not jointly convex in all the variables involved, due to the cross-term of

$$\left(\mathbf{\Gamma}^{N^{-}}_{R}\mathbf{T}_{u^{-}}\right)^{T}\left(\mathbf{\Gamma}^{N^{-}}_{R}\mathbf{T}_{u^{-}}\right) \quad \text{and} \quad \left(\mathbf{\Gamma}^{N^{-}}_{R}\mathbf{T}_{u^{-}}\right)^{T}\left(\mathbf{T}^{n}_{S}\mathbf{T}_{\zeta}\right)$$

on the right hand side of (4.12).

To avoid the difficulties associated with solving non-convex problem, in the following two subsections, we propose two different convex relaxations.

4.2.2 Deterministic Convex Relaxation

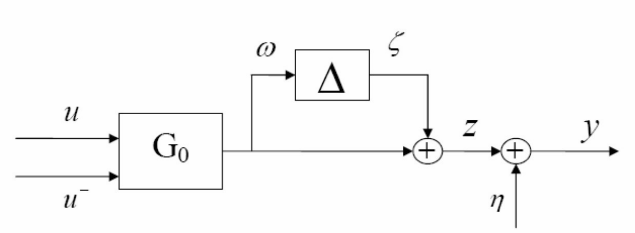

Fig. 4.3. structure of system with multiplicative uncertainty

Consider a special case of Problem 9 shown in Figure 4.3, where the nominal dynamic is subject to *multiplicative, unstructured* uncertainty, and with the assumption that the noise set $\mathcal{N} = \mathcal{B}\ell_2(\epsilon)$ and the uncertainty set $\mathbf{\Delta} \doteq \mathcal{B}\mathcal{H}_\infty(\gamma)$.

While the problem is not jointly in all the variables involved, a convex relaxation can be obtained by considering the alternative setup shown in Figure 4.4, where the measurement noise is moved to a place it is also affected by the uncertainty Δ:

$$\mathbf{y} = \big(\mathbf{I} + \Delta\big)\big(\mathbf{T}^{n}_{G_0}\,\mathbf{u} + \mathbf{\Gamma}^{N^-}_{G_0}\,\mathbf{u}^- + \tilde{\eta}\big) \tag{4.13}$$

When compared to the original setup in Figure 4.3, it can be easily seen that the only difference is in the measurement noise, $\eta \doteq \big(\mathbf{I} + \Delta\big)\tilde{\eta}$. Thereupon, the existence of a

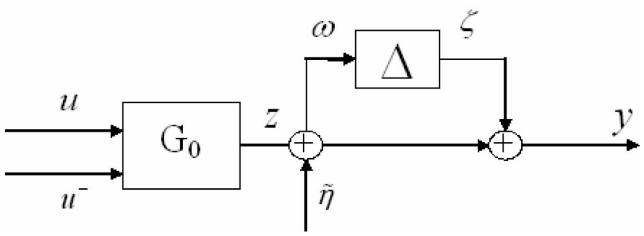

Fig. 4.4. An alternative system structure for Semi-Blind model (in)validation

triple $\left(\Delta, \tilde{\eta}, \mathbf{u}^{-}\right) \in \left(\mathbf{\Delta}, \tilde{\mathcal{N}}, \mathcal{U}^{-}\right)$ with

$$\tilde{\eta} \in \tilde{\mathcal{N}} \doteq \mathcal{B}\ell_2(\tilde{\epsilon}) = \mathcal{B}\ell_2\left(\frac{\epsilon}{1+\gamma}\right)$$

satisfying (4.13) leads to the satisfaction of:

$$\mathbf{y} = \left(\mathbf{I} + \Delta\right)\left(\mathbf{T}_{G_0}^{n}\,\mathbf{u} + \mathbf{\Gamma}_{G_0}^{N^{-}}\mathbf{u}^{-}\right) + \eta \tag{4.14}$$

with the triple $\left(\Delta, \eta, \mathbf{u}^{-}\right) \in \left(\mathbf{\Delta}, \mathcal{N}, \mathcal{U}^{-}\right)$. i.e., Problem 9 can be solved through searching a solution to the model (in)validation problem in Figure 4.4 with noise level as $\tilde{\epsilon}$. Note, if $\|\Delta\|_{\infty} \ll 1$ then this approximate relaxation is not too conservative. Furthermore, in the sequel we'll present that this relaxes to a convex optimization problem.

Theorem 5. *There exists a feasible triple* $\left(\Delta, \tilde{\eta}, \mathbf{u}^{-}\right) \in \left(\mathbf{\Delta}, \tilde{\mathcal{N}}, \mathcal{U}^{-}\right)$ *that satisfies* (4.13) *if and only if there exists at least one pair of finite sequences*

$$\mathbf{u}^{-} = \{u_{-1}, u_{-2}, \cdots, u_{-N^{-}}\} \in \mathcal{U}^{-}$$

and

$$\tilde{\eta} = \{\tilde{\eta}_0, \tilde{\eta}_0, \cdots, \tilde{\eta}_{n-1}\} \in \tilde{\mathcal{N}}$$

such that the following LMI holds with some $\gamma < 1$*:*

$$\mathbf{A} \doteq \begin{bmatrix} \mathbf{X} & (\mathbf{T}_\omega)^T \\ \mathbf{T}_\omega & -(\gamma^2 - 1)^{-1} \end{bmatrix} \geq 0 \tag{4.15}$$

where:

$$\mathbf{X} \doteq -\left[(\mathbf{T}_y)^T \mathbf{T}_y - (\mathbf{T}_y)^T \mathbf{T}_\omega - (\mathbf{T}_\omega)^T \mathbf{T}_y\right]$$

$$\mathbf{T}_\omega \doteq \underset{G_0}{\mathbf{T}^n} \mathbf{T}_u + \underset{G_0}{\mathbf{\Gamma}^{N^-}} \mathbf{T}_{u^-} + \mathbf{T}_{\tilde{\eta}}$$

Proof. From the setup Figure 4.4, we have the following relations between signals:

$$\begin{aligned} \zeta &= \mathbf{y} - \omega \\ \omega &= \underset{G_0}{\mathbf{T}^n} \mathbf{u} + \underset{G_0}{\mathbf{\Gamma}^{N^-}} \mathbf{u}^- + \tilde{\eta} \end{aligned} \tag{4.16}$$

Apply Theorem 3 on signal ω, ζ and replace ζ with $\zeta = y - \omega$,

$$(\mathbf{T}_y - \mathbf{T}_\omega)^T (\mathbf{T}_y - \mathbf{T}_\omega) \leq \gamma^2 \underset{\omega}{\mathbf{T}^T} \mathbf{T}_\omega \tag{4.17}$$

yields.

For the case of $\gamma < 1$, the following can be obtained by reordering (4.17):

$$\begin{aligned} \mathbf{T}_y^T \mathbf{T}_y - \mathbf{T}_y^T \mathbf{T}_\omega - \mathbf{T}_\omega^T \mathbf{T}_y &\leq (\gamma^2 - 1)\mathbf{T}_\omega^T \mathbf{T}_\omega \\ \iff -(\mathbf{T}_y^T \mathbf{T}_y - \mathbf{T}_y^T \mathbf{T}_\omega - \mathbf{T}_\omega^T \mathbf{T}_y) - \mathbf{T}_\omega^T [-(\gamma^2 - 1)] \mathbf{T}_\omega &\geq 0 \end{aligned} \tag{4.18}$$

Then, (4.15) can be achieved directly through applying Schur Complement on (4.18). □

This deterministic relaxation reduces the problem to a conventional LMI optimization, but it is limited to the case of unstructured multiplicative uncertainty. Moreover, as we noted before, if the condition $\|\Delta\|_\infty \ll 1$ does not hold, LMIs (4.15) may be infeasible when the original problem 9 has a solution. Hence, to handle completely general uncertainty structures, we offer a risk–adjusted relaxation at the expense of computational time in the following subsection.

4.2.3 Risk-Adjusted Convex Relaxation

The risk–adjusted relaxation proposed here can handle arbitrary uncertainty structures in return for an arbitrarily small probability of missing a feasible solution. In addition, it has polynomial, rather than exponential, computational complexity growth with the problem data ([97]).

First, let's consider the situation that Δ is known. If Δ is fixed and given, then the original problem 9 can be directly reduced to a convex LMI optimization problem presented as below:

Lemma 2. *Given a fixed $\Delta = \Delta^i$, Problem 9 has a solution if and only if there exists at least one feasible sequence $\mathbf{u}^- \in \mathcal{U}^-$ such that*

$$\mathbf{y} - \left(\mathbf{T}^n_{\hat{G}}\mathbf{u} + \mathbf{\Gamma}^{N^-}_{\hat{G}}\mathbf{u}^-\right) = \eta \in \mathcal{N} \tag{4.19}$$

where

$$\hat{G} \doteq [G + Q\Delta(\mathbf{I} - S\Delta)^{-1}R] \tag{4.20}$$

is the given uncertain system.

Proof. With Δ given, the interpolation constraint (4.10) which causes the problem to be non-convex has been removed. Then, the remaining requirements are only that the past sequence $\mathbf{u}^-$ stays in $\mathcal{U}^-$ and the noise has to belong to $\mathcal{N}$. □

This obviously lead to a convex problem since there is only one unknown variable $\mathbf{u}^-$ in (4.19).

Next, stimulated by Lemma 2, the main idea of this method is to uniformly sample the set of admissible uncertainty $\mathbf{\Delta}$, in an attempt to find at least one element $\Delta^i \in \mathbf{\Delta}$ that can make the experimental data be reproduced by (4.9) with $\eta \in \mathcal{N}$ and $\mathbf{u}^- \in \mathcal{U}^-$. This yields another difficulty that the set $\mathbf{\Delta}$ is *infinite dimensional*. However, since Δ is a causal operator only its first n Markov parameters affect the output $\mathbf{y}$. Thus, instead of generating samples over $\mathcal{BH}_\infty(\gamma)$, we only need to (i) sample the set $\mathcal{BH}^n_\infty(\gamma)$, using Algorithm 1 in section 3.4, and (ii) combine the samples[2].

[2] In the case of structured uncertainty, the same construction can be used block-wise.

Associate the above two ideas, the following algorithm is derived to reduce the semi-blind model (in)validation problem to an non–convex LMI optimization problem:

Algorithm 3. *Given the experimental data* $(\mathbf{u}, \mathbf{y})$*, the uncertain model shown in Figure 4.2 and descriptions of sets* $(\mathbf{\Delta}, \mathcal{N}, \mathcal{U}^-)$*, Problem 9 can be solved through:*

1. *Choose* N_1 *and generate* N_s *samples of* $\{\Delta^i\}$ *on set* $\mathbf{\Delta} \doteq \mathcal{BH}_\infty^n(\gamma)$ *using Algorithm 1.*

2. *For each* Δ^i*, find a feasible* $\mathbf{u}^-$ *such that*

$$\mathbf{y} - \left(\mathbf{T}_{\hat{G}}^n \mathbf{u} + \mathbf{\Gamma}_{\hat{G}}^{N^-} \mathbf{u}^-\right) = \eta \in \mathcal{N} \tag{4.21}$$

   where $\hat{G}$ *is the given uncertain system, shown in* (4.20)*, and*

$$\mathbf{u}^- \in \mathcal{U}^- \tag{4.22}$$

   hold.

3. *If there exists at least one feasible* $\mathbf{u}^-$*, stop. Otherwise, try a next sample* Δ^{i+1} *and go back to step 1.*

The algorithm finishes either by finding one feasible solution of $\mathbf{u}^-$ or after N_s samples.

Remark 3. *Let* (ν, δ) *be two positive constants in* $(0, 1)$*. If* N_1 *in Algorithm 1is chosen to satisfy*

$$N_1 \geq \frac{\ln(1/\delta)}{\ln(1/(1-\nu))},$$

then, with probability greater than $1 - \delta$*, the probability of not finding a feasible pair* $\Delta^i, \mathbf{u}^-\}$ *when one exists is smaller than* ν*.*

Thus, by introducing an (arbitrarily small) risk of missing a feasible solution, we can substantially alleviate the computational complexity entailed in robustly testing the validity of the nominal plant subject to structured uncertainty and measurement noise considering the effects of the past unknown inputs.

4.2.4 Model Validation for the Macro-Economic Subsystem

After the model is identified, it should be validated with *new* experimental data before practical use. The (in)validation procedure outlined above is applied to the macroeconomic system introduced in section 3.3, with $\mathbf{u}^-$ and $\mathbf{u}$ represent the past and current FED fund rates, $\mathbf{y}$ represents the annual inflation, and the identified nominal model shown in (3.22). The procedure seeks a minimum size of the uncertainty that will not (in)validate the experimental data for a suitable given measurement error level. As the experimental data to be tested here is of length 166, which causes risk-adjusted relaxation take too much time to run the Algorithm, we only use the deterministic relaxation. For comparison purpose, the same process is performed on the empirical RS model. Table 4.1 shows the results and the comparison.

In all cases the initial conditions at $t = t_0$ were assumed to have been reached by a sequence of past inputs

$$\{u_k, k = -4, -3, -2, -1\}, \text{ with } |u_k| \leq 0.5.$$

As shown there, the modeling error (measured in terms of the norm of the uncertainty $\Delta(z)$ incurred for the semi–blind robust identification method is between 30% to 50% lower than that of the RS model.

In addition, we as well apply the (in)validation procedure to the model (3.44) identified with the white–noise–like observation noise sequence, and hence a minimum size of the uncertainty that will not (in)validate the experimental data for a suitable given measurement error level is found. Considering the length of the experimental data to be validated, we also only use the deterministic relaxation. Table 4.2 shows the results[3] and clearly denotes when a system is no longer valid or compatible for the underlying dynamics during the optimization. Note that the noise level is in percentage of ℓ_2–norm of the historical output.

For both situations, as we expected, the uncertainty (measured in terms of its $\mathcal{H}_\infty$–norm) decreases when we relax the noise level, which witness the facts that both noise and uncertainty contributes to the identification error and there is a trade off between these two contributions.

[3] The same as in Table 4.1, in all cases, assume the past sequence $\mathbf{u}^- = \{u_k, k = -4, \cdots, -1\}$, $|u_k| \leq 0.5$.

Table 4.1. Model Invalidation Results of the macro-economy subsystem with ℓ_∞–norm bounded observation noise

Data range	Noise Level	Identified Model $\|\Delta\|_\infty$	RS Model $\|\Delta\|_\infty$
5-170	1.72	0.5	0.63
5-170	2.00	0.3	0.4575
5-170	3.37	0.0009	0.0370

Table 4.2. Model Invalidation Results of the macro-economy subsystem with whiter–noise–like observation noise

Data range	Noise Level	Identified Model $\|\Delta\|_\infty$
6-170	24%	0.18
6-170	20%	0.59
6-170	15%	infeasible

4.3 Application on FDI (Fault detection and Isolation)

This subsection addresses the problem of detecting and isolating faults in noisy MIMO uncertain systems, subject to unstructured dynamic uncertainty. It demonstrates that this problem can be efficiently solved using a semi-blind model (in)validation framework.

4.3.1 Introduction

The problem of Fault Detection and Isolation (FDI) in control systems has been the subject of considerable attention during the past two decades. This research has resulted in a variety of methods (see for instance [24, 26] and references therein). Among these methods, those based on analytical or *model–based* approaches are specially appealing, since do not require additional hardware. However, a drawback of these approaches is their (potential) *fragility*: a mismatch between the actual plant and the model used in the FDI algorithm can result in false alarms. To avoid this difficulty, the algorithm must be robust both against modeling errors and exogenous disturbances. Robust FDI methods have been well studied (see for instance [18, 23, 24, 38, 42, 83, 95, 112] and references therein). A potential disadvantage of these methods is the difficulty in isolating the exact location of the fault and in detecting simultaneous faults. Motivated by [89], we propose to address these issues by recasting the problem into a robust model (in)validation form. A similar approach was pursued in [39], where, for frequency domain data, the problem of fault detection was recast as a μ analysis one and reduce to an LMI optimization using the well known μ upper bound. However, when in addition

to establishing that a fault has occurred it is desired to isolate its location, the resulting problem is no longer convex[4], and thus computationally hard to solve. In order to obtain tractable solutions, we use the two convex relaxations developed for semi-blind model (in)validation: one deterministic and one stochastic.

The proposed new FDI framework has the following advantages over existing methods: (a) It can handle *arbitrary* fault dynamics and uncertainty structures, (b) It provides an estimate of which fault has occurred, and (c) Its computational complexity grows polynomially with the dimension of the plant.

4.3.2 Application of Semi-Blind Validation Framework on FDI

Now let's demonstrate the application of the semi-blind (in)validation framework on the problem of detecting and isolating faults from noisy input-output measurements of a MIMO uncertainty system.

We consider the FDI problem for the systems of the following structure:

$$y = \left[G_0(z,\Delta_0) + \sum_{i=1}^{N_f} f_i G_i(z,\Delta_i)\right] * u + d \tag{4.23}$$

where the transfer matrices $G_0(z,\Delta_0)$, $G_i(z,\Delta_i), i = 1,\cdots,N_f$ represent the zero-faults plant and dynamic faults models respectively, and $\Delta_i \in \mathbf{\Delta}_i \subseteq \mathcal{BH}_\infty$ represent the (structured)dynamic model uncertainty. d represents a measurement noise with the only information of $\|d\|_2 \leq \epsilon$. The scalars $f_i \in [0,1]$ are the fault indicators, with $f_i = 0$

[4]This problem is equivalent to a model (in)validation problem with structured uncertainty ([89]), which is generically NP–hard ([98]).

indicates no faults involved and $f_i = 1$ indicates the extreme case of total faults. Note that this formulation allows for the uncertainty to enter the dynamics in an arbitrary way.

Then, in this context, the FDI problem can be stated as:

Problem 10. *Given a model of the plant under normal conditions* $G_0(\lambda, \Delta_0)$*, failure dynamics* $G_i(\lambda, \Delta_i)$*, a bound* ϵ *on the measurement noise, uncertainty sets* $\mathbf{\Delta}_i$*, and input/output experimental measurements, determine: (i) whether a fault has occurred, and (ii) in that case, isolate it and determine its strength.*

Comparing (4.23) with (4.9) in semi-blind model validation, $G_0(z, \Delta_0) * u$ corresponds with $\mathbf{T}^n_{\hat{G}}\mathbf{u}$, $\left[\sum_{i=1}^{N_f} f_i G_i(z, \Delta_i)\right] * u$ corresponds with $\mathbf{\Gamma}^{\mathbf{N}^-}_{\hat{\mathbf{G}}}\mathbf{u}^-$, and d corresponds with η. Problem 10 leads to a non-convex problem due to the non-convexity of $\left[\sum_{i=1}^{N_f} f_i G_i(z, \Delta_i)\right] * u$. Applying semi-blind model (in)validation framework, it can be restated as:

Problem 11. *Given the* a priori *information:* $G_i(\lambda)$*, uncertainty sets* $\mathbf{\Delta_i}, \Delta_i \in \mathbf{\Delta_i}$*, measurement noise set*

$$\mathcal{N} = \{d\colon \|d\|_2 \leq \epsilon\},$$

find one feasible set $\{\Delta_i, d, f\}$ *that can reproduce the input/output experimental measurements data* $\mathbf{u}, \mathbf{y}$ *as in* (4.23)*. If* $f = 0$ *then no fault occurs, otherwise, the fault location and value is identified by the elements of* f*.*

This problem can be reduced to an LMI optimization problem using the two relaxations developed for semi-blind model (in)validation: one deterministic and one stochastic.

Note that in general, due to the presence of uncertainty and noise, there may exist more than one set $\left\{\Delta_i, d, f\right\}$ that explains the experimental input–output data. In that case, to avoid ambiguities, we will select, among all possible solutions, the one corresponding to the minimum value of $\|f\|_2$. This choice minimizes the number of false alarms, since it tries to explain, whenever possible, the experimental data as being produced by the normal (non–failure) dynamics, possibly affected by dynamic uncertainty and measurement noise.

Hence, from the analysis above it follows that finding minimum $\|\mathbf{f}\|$ such that Problem 11 has at least one feasible solution reduces to a convex LMI minimization problem. Furthermore, when we use the stochastic relaxation, according to Remark 3, if N_1 in Algorithm 1 is chosen to satisfy,

$$N_1 \geq \frac{\ln(1/\delta)}{\ln(1/(1-\nu))},$$

then

$$\mathbf{Prob}\{\mathbf{Prob}[\|\mathbf{f}_{min}\|_2 < \|\mathbf{f}_{min}^{N_1}\|_2] \leq \nu\} \geq (1-\delta),$$

where $\mathbf{f}_{min}$ and $\mathbf{f}_{min}^{N_1}$ denotes the real faults and the estimated faults found by stochastic relaxation, respectively. Roughly speaking, with confidence $1-\delta$, the stochastic relaxation will find, with probability $1-\nu$, the estimation. In addition, it can be argued that a purely deterministic approach to FDI could be potentially overly optimistic, since the system will be deemed to be operating under no–fault conditions even if there exist a *single* combination of uncertainty and noise such that the corresponding $\|\mathbf{f}\|_2 = 0$. On

the other hand, in such cases the approach proposed here will indicate, (with probability close to 1) the existence of a fault[5].

4.3.3 Illustrative Examples

This subsection gives several examples to illustrate the potential of solving FDI problem through the semi-blind model (in)validation framework.

Example 1: Consider the following system subject to multiplicative uncertainty:

$$y = (I + \Delta)(G_0 + \sum_{i=1}^{3} f_i G_i)u + d \tag{4.24}$$

where

$$\begin{aligned}
G_0 &= \frac{s^4 - 4.75s^3 - 2.48s^2 - 1.19s - 0.56}{s^4 + 1.92s^3 + 1.61s^2 + 0.83s + 0.16}, \\
G_1 &= \frac{5.07s^4 + 3.91s^2 + 0.94}{s^4 + 2.55s^3 + 3.76s^2 + 4.16s + 3.18}, \\
G_2 &= \frac{31.75s^3 + 1.8s}{s^4 + 2.55s^3 + 3.76s^2 + 4.16s + 3.18}, \\
G_3 &= \frac{75.75s^2 + 65}{s^4 + 2.55s^3 + 3.76s^2 + 4.16s + 3.18}
\end{aligned}$$

and assume that the available *a priori* information is $\|\Delta\|_\infty \leq 0.3$ and $\|d\|_2 \leq 0.94$[6]. Here, to facilitate comparisons with existing approaches, we have formulated the problem in the continuous time domain. Thus, in order to apply the techniques proposed in this paper, a discrete–time model of the system above was obtained by using samplers and zero order holds with a sampling time of 0.1 seconds. Further, since the closest existing

[5] see also [114] for a similar argument used in the context of model (in)validation

[6] this noise level corresponds to 10% of the energy of the step response of the plant.

technique to the one proposed here ([21]) requires a–priori knowledge of the McMillan degree of the uncertainty, we will assume that the model uncertainty Δ is just a constant.

The second and third columns in Table 4.3 show the results of applying the proposed risk–adjusted and deterministic methods to several faults. The experimental data consisted of $N_t = 20$ samples of the step response, corrupted by noise d, with $\|d\|_2 = 0.85$, of the model (4.24) corresponding to $\Delta = 0.28$. For the risk–adjusted relaxation $N_1 = 250$ samples of the uncertainty where used, which guarantees, with confidence 0.99, a probability of 0.98 of finding the minimum $\|f\|_2$ that explains the experimental data, resulting on a computation time of 42 seconds using Matlab's LMI toolbox on a 1.7 Ghz Xeo processor. For comparison, the deterministic relaxation required 4 seconds. Both methods successfully identified and isolated faults in all cases, with the risk–adjusted relaxation slightly outperforming the deterministic one, as expected due to the moderately large uncertainty level.

Real Fault			risk adjusted ($N_1 = 250$, $prob = 98\%$)			deterministic			observer-based ($FAR = 2\%$)
0.0	0.0	0.0	10^{-5}*[0.10	0.11	0.11]	10^{-5}*[0.11	0.12	0.12]	No Fault
1.0	0.0	0.0	0.95	0.00	0.00	0.92	0.00	0.00	Fault
0.0	1.0	0.0	0.04	0.95	0.00	0.05	0.94	0.00	Fault
0.0	0.0	1.0	0.00	0.00	0.97	0.00	0.00	0.96	Fault
0.45	0.21	0.56	0.36	0.21	0.54	0.33	0.21	0.53	Fault
0.20	0.10	0.05	0.12	0.10	0.04	0.10	0.09	0.03	No Fault
0.10	0.05	0.1	0.03	0.04	0.08	0.02	0.04	0.07	No Fault

Table 4.3. Example 1: Fault estimates corresponding to the case $\|\Delta\|_\infty \leq 0.3$ and 10% noise level.

Next we compare them against the probabilistic residual based method proposed in [21]. To this effect, consider state space realizations of the nominal and fault dynamics,

$$G_i = \left(\begin{array}{c|c} A_i & B_i \\ \hline C_i & D_i \end{array}\right) \qquad i = 0, \ldots 3,$$

and note that for the case of constant Δ, the model (4.24) can be written in the form:

$$\begin{aligned} \dot{x} &= (A + \Delta A + \Delta A_F)x + (B + \Delta B + \Delta B_F)u \\ y &= (C + \Delta C + \Delta C_F)x + (D + \Delta D + \Delta D_F)u + d \end{aligned} \tag{4.25}$$

where

$$
\begin{aligned}
A &= \begin{bmatrix} A_o & 0 \\ 0 & 0 \end{bmatrix},\ B = \begin{bmatrix} B_o^T & 0 \end{bmatrix}^T, \\
C &= \begin{bmatrix} C_o & 0 \end{bmatrix},\ D = D_o \\
\Delta A &= 0;\ \Delta B = 0, \\
\Delta C &= \delta \cdot C,\ \Delta D = \delta \cdot D, \\
|\delta| &\leq 0.1, \\
\Delta A_F &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & A_1 & & \\ \vdots & & \ddots & \\ 0 & & & A_3 \end{bmatrix}, \\
\Delta B_F &= \begin{bmatrix} 0 & B_1^T \dots B_3^T \end{bmatrix}^T \\
\Delta C_F &= (1+\delta) \begin{bmatrix} 0 & f_1 C_1 & \cdots & f_3 C_3 \end{bmatrix}, \\
\Delta D_F &= (1+\delta) \sum f_i D_i
\end{aligned}
\tag{4.26}
$$

In its simplest form, the approach proposed by [21] uses an observer based on the nominal dynamics to generate a residual r:

$$
\begin{aligned}
\dot{\hat{x}} &= (A - LC)\hat{x} + (B - LD)u \\
r &= y - C\hat{x} - Du
\end{aligned}
\tag{4.27}
$$

A fault is deemed to have occurred whenever $\|r\|_2 > J_{th}$, where the threshold J_{th} is selected so that the probability of false alarms is lower than a pre–set level, that is

$$P\left(\|r|_2 > J_{th}| \text{ no fault is present }\right) \leq \alpha,$$

where α is given. In this case, using a gain $L = 0$ (corresponding to both the optimal $\mathcal{H}_2$ and $\mathcal{H}_\infty$ filters) and setting the False Alarm Rate (FAR) to 2%, yields a threshold $J_{th} = 17.45$. As shown shown in the last column of Table 4.3, while for larger values of f all methods have comparable performance, both invalidation based approaches outperform the residual based one in the case of smaller f. Note also that the former approaches, in addition to establishing the existence of a fault, were able to accurately estimate its components, albeit at the cost of increased on–line computational time. This increase is small for the case deterministic relaxation, but more substantial for the risk–adjusted one. This suggests that those cases where either the residual or the deterministic approaches are applicable may benefit from using a mixed strategy where these methods are used first to establish the existence of a fault, followed by use of the risk–adjusted method to establish its location.

Example 2: Consider now the problem of FDI in a simplified model of the yaw damper system of a jet transport. The transfer function from the rudder and aileron deflections to the yaw rate and bank angle can be represented by a model of the form (4.24) with (see [89] for details)

$$G_i = \frac{N_i}{D_i},\ i = 0, \ldots, 3,$$

$$N_{i}(z) = \begin{pmatrix} N^{i}_{11} & N^{i}_{12} \\ N^{i}_{21} & N^{i}_{22} \end{pmatrix},$$

where G_{0} denotes the nominal plant, and where

$$\begin{aligned}
D_{o}(z) &= z^4 - 3.81z^3 + 5.45z^2 - 3.46z + 0.83 \\
D_{i}(z) &= z^8 - 7.7z^7 + 26.2z^6 - 50.8z^5 + 61.4z^4 - 47.6z^3 + 23.1z^2 - 6.4z + 0.78,\ i = 1,2,3 \\
N^{o}_{11} &= -0.44z^3 + 1.3z^2 - 1.3z + 0.42,\ N^{o}_{12} = 0.11z^3 - 0.34z^2 + 0.33z - 0.11 \\
N^{o}_{21} &= 10^{-3} * (5z^3 - 7z^2 - 4.3z + 5.3),\ N^{o}_{22} = 10^{-2} * (5.3z^3 - 4.6z^2 - 5.1z + 4.5) \\
N^{1}_{11} &= -0.28z^7 - 1.42z^6 + 2.64z^5 - 1.73z^4 - 0.93z^3 + 2.16z^2 - 1.26z + 0.26 \\
N^{1}_{12} &= -0.07z^7 + 0.37z^6 - 0.68z^5 + 0.45z^4 + 0.24z^3 - 0.56z^2 + 0.32z - 0.07 \\
N^{1}_{21} &= 10^{-5} * (7.1z^7 + 46.6z^6 - 181.8z^5 + 138.2z^4 + 106.6z^3 - 169.6z^2 + 46.9z + 6.1) \\
N^{1}_{22} &= -10^{-4} * (0.2z^7 + 1.4z^6 - 5.8z^5 + 4.4z^4 + 3.4z^3 - 5.4z^2 + 1.5z + 0.2) \\
N^{2}_{11} &= 10^{-2} * (z^7 - 1.2z^6 - 7.8z^5 + 22.9z^4 - 23.5z^3 + 8.9z^2 + 0.6z - 0.9 \\
N^{2}_{12} &= -10^{-3} * (1.9z^7 - 2.4z^6 - 14.9z^5 + 43.7z^4 - 44.8z^3 + 16.9z^2 + 1.2z - 1.7) \\
N^{2}_{21} &= -10^{-3} * (2.3z^7 - 2.9z^6 - 18z^5 + 53z^4 - 54z^3 + 21z^2 + 1.5z + 2) \\
N^{2}_{22} &= 10^{-4} * (6z^7 - 7z^6 - 47z^5 + 137z^4 - 141z^3 + 53z^2 + 4z - 5) \\
N^{3}_{11} &= 10^{-6} * (2.4z^7 + 53z^6 - 27z^5 - 206z^4 + 215z^3 + 13z^2 - 49z - 2) \\
N^{3}_{12} &= -10^{-6} * (z^7 + 24z^6 - 12z^5 - 93z^4 + 97z^3 + 6z^2 - 22z - 0.9) \\
N^{3}_{21} &= 10^{-5} * (1.5z^7 + 32z^6 - 16z^5 - 126z^4 + 132z^3 + 8.2z^2 - 30z - 1.2) \\
N^{3}_{22} &= -10^{-5} * (0.36z^7 + 7.8z^6 - 4.0z^5 - 30z^4 + 32z^3 + 2.0z^2 - 7.2z - 0.3)
\end{aligned}$$

The only a priori information available is that the nominal and failure dynamics are subject to multiplicative model uncertainty with $\|\Delta\|_\infty \leq 0.3$, and that $\|d\|_2 \leq 2.6$[7]. The experimental data consist of 20 samples of the response of $(I + \tilde{\Delta})G_f$[8], where $\tilde{\Delta}$ is given by[9]:

$$\tilde{\Delta} = \frac{0.018}{D_\Delta} \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}, \tag{4.28}$$

$$D_\Delta = z^4 + 1.87z^3 + 1.27z^2 + 0.37z + 0.04$$

$$\Delta_{11} = 1.9z^4 + 2.5z^3 - 0.24z^2 - 1.04z - 0.25,$$

$$\Delta_{12} = 0.5z^4 + 0.8z^3 + 0.25z^2 - 0.09z - 0.03,$$

$$\Delta_{21} = 2.9z^4 + 3.0z^3 - 1.66z^2 - 2.3z - 0.51,$$

$$\Delta_{22} = 3z^4 + 3.5z^3 - 1.12z^2 - 2.1z - 0.47$$

As shown in Table 4.4, both relaxations were able to provide good estimates of the fault indicators. Similar results,shown in Table 4.5, were also obtained with lower uncertainty and noise levels. For the results in Table 4.5, we tighten the uncertainty norm bound and the noise level to 0.2 and 10%, respectively. As expected, the performance of both methods become closer with decreasing uncertainty and noise levels.

[7]This noise level corresponds to 20% of the energy of the response of plant

[8]Here G_f denotes the transfer function of the failure mode under consideration.

[9]This corresponds to a randomly generated uncertainty with $\|\tilde{\Delta}\|_\infty \leq 0.297$.

Real Fault			Risk Adjusted Estimate ($N_1 = 250$)			Deterministic Relaxation		
0.0	0.0	0.0	10^{-6}*[0.7814	0.8356	0.8715]	10^{-3}*[0.4572	0.4592	0.4514]
1.0	0.0	0.0	0.7313	0.0194	0.0222	0.6336	0.0797	0.0194
0.0	1.0	0.0	0.0400	0.5868	0.0261	0.0755	0.5118	0.0406
0.0	0.0	1.0	0.0215	0.0000	0.6567	0.0447	0.0000	0.5674
0.85	0.78	0.001	0.6694	0.4318	0.0500	0.6113	0.4183	0.0402
0.45	0.21	0.56	0.3248	0.1749	0.3424	0.2966	0.0876	0.3058

Table 4.4. Risk adjusted versus deterministic relaxation for Example 2 with $\|\Delta\|_\infty \leq 0.3$ and 20% noise level

We apply the semi-blind model (in)validation framework on the problem of robust fault detection and isolation for systems described by a parameterized fault model, subject to arbitrary dynamic uncertainty. In general this setup leads to non–convex, NP hard problems. To remove this limitation, we use the two convex relaxations proposed for semi-blind model (in)validation. The deterministic relaxation reduces the problem to a conventional LMI optimization, but is limited to the case of unstructured multiplicative uncertainty, on the other hand, the risk–adjusted relaxation can, in return for an (arbitrarily small) probability of a false alarm, handle completely general uncertainty structures. Further, for risk–adjusted relaxation, since the number of samples needed for reliable fault estimations is relatively small, it is feasible to generate and store these samples off–line, leading to further reduction of the computational complexity of the problem that needs to be solved on–line. Both relaxations have comparable performance for relatively low uncertainty levels, with the stochastic relaxation outperforming the convex one as the uncertainty level increases.

Real Fault			Risk Adjusted Estimate ($N_1 = 250$)			Deterministic Relaxation		
0.0	0.0	0.0	10^{-6}*[0.9122	0.8285	0.8516]	10^{-3}*[0.2452	0.2667	0.2446]
1.0	0.0	0.0	0.8251	0.0107	0.0038	0.7730	0.0779	0.0090
0.0	1.0	0.0	0.0376	0.7523	0.0000	0.0564	0.6999	0.0177
0.0	0.0	1.0	0.0168	0.0074	0.7744	0.0267	0.0000	0.7221
0.85	0.78	0.001	0.7586	0.5444	0.0089	0.7143	0.5302	0.0405
0.45	0.21	0.56	0.3919	0.1746	0.3946	0.3694	0.1206	0.3634

Table 4.5. Risk adjusted versus deterministic relaxation for Example 2 with $\|\Delta\|_\infty \leq 0.2$ and 10% noise level

Chapter 5

Wiener System Identification

This chapter addresses the problem of identification of Wiener systems from a set-membership standpoint and the power of this approach is illustrated by a non-trivial problem arising in computer vision: tracking a human in a sequence of frames.

5.1 Introduction

Wiener systems are a special type of nonlinear systems consisting of the cascade of a Linear Time Invariant (LTI) plant and a memoryless, static non–linearity. The problem of identifying such models from experimental data has received considerable attention, since these systems arise in many practical situations in a wide range of applications, including control (see [111, 117]), communications (see [27, 19]), and biology (see [8, 10]). On the basis of the *a priori* information and input/output measurements of the whole system, the identification process is to construct mathematical descriptions of the nonlinear subsystem as well as the dynamic part. One source of the difficulties of Wiener System identification, the unmeasurable signal interconnecting subsystems, lead to the fact that neither subsystem can be entirely identified.

Several techniques have been proposed in the past to solve this problem. [111] proposes a recursive identification algorithm, derived from a parametric wiener model

without applying inversion of the nonlinearity, in the context of a stochastic approximation framework. Without parametric restriction on the nonlinear characteristic of the memoryless subsystem, [28] develop a nonparametric algorithm recovering the feature of the whole system and [29] applies kernel regression estimate to obtain the invertible part of the nonlinearity. [108, 32] use a so–called subspace methods (see [105, 103, 104]), which based on the analysis of structured subspaces, to separately identify the linear part of the system. Unfortunately, a potential difficulty with these approaches is that they are stochastic in nature, relying on the use of white Gaussian inputs.

Some research has been done in literature to handle this difficulty. [75] addresses this limitation by explicitly modeling the inverse of the nonlinearity, coupled with the use of subspace based approaches. [3] presents an optimal two-stage identification algorithm combining recursive least squares and singular value decomposition and [4] uses a blind approach to recover all the internal variables solely based on the output measurement. However, all of these approaches require additional assumptions on the nonlinearity (either invertibility or a special structure), which do not hold in several problems of practical interest. Finally, [5] develops a frequency domain based algorithm by exploring the fundamental frequency and harmonics generated by the nonlinearity. However, at the present time this approach is restricted only to frequency domain data (e.g. steady state outputs generated by sinusoidal inputs).

In this chapter, we propose an algorithm for time–domain based identification that avoids these difficulties by pursuing a risk–adjusted approach. Here, in return for an (arbitrarily) small risk of not being able to establish consistency of the data, the problem is reduced to a convex optimization problem. In the second part of this

chapter we illustrate these results with a non–trivial problem arising in computer vision: tracking a human in a sequence of frames. The challenge here arises from the changes in appearance undergone by the target and the large number of pixels to be tracked. By using the proposed identification method, we show that the problem can be solved by modeling the plant as a Wiener system. This formalizes some recent conjectures (see [61]) where it has been argued that this motion can be explained by considering linear dynamics in a low dimensional manifold, accounting for the physics of the motion, followed by a static non-linearity that accounts for appearance changes in the target.

5.2 Problem Statement

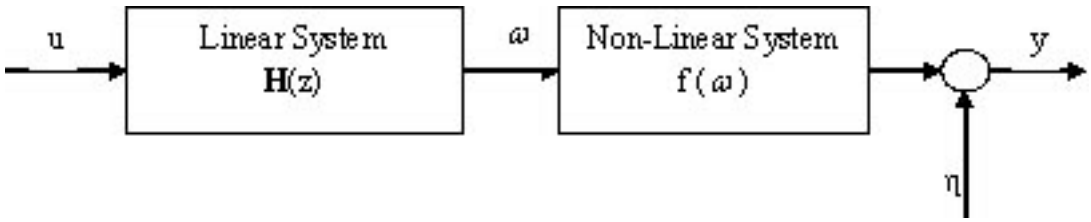

Fig. 5.1. Wiener System Structure

Consider the Wiener system shown in Figure 5.1 consisting of the interconnection of a LTI system $H(z)$ and a memoryless nonlinearity $f(.)$. The corresponding equations

are given by:

$$\begin{aligned} y_k &= f(\omega_k) + \eta_k \\ \omega_k &= (h * u)_k \end{aligned} \tag{5.1}$$

where $*$ denotes convolution and the signals $u \in \mathcal{R}^{n_u}$ and $y \in \mathcal{R}^{n_y}$ represent the experimental data: a known finite input sequence and its corresponding output sequence, corrupted by unknown but norm–bounded measurement noise η. Actually, there is no standard formulation on how the noise corrupt the wiener system. An assumption , which appears to be the usual one, is that the noise enters as an additive output measurement noise as shown in Figure 5.1 (see in [108, 110]). An alternative assumption is to assume that the noise enters between the LTI subsystem and the static nonlinear block (see in [28, 116]). Note that the intermediate signal $\omega \in R^{n_\omega}$ (the output of the LTI system) is not measurable. Our goal is to, given the experimental data $\mathbf{u} = \left[u_0, u_1, \cdots, u_{N-1}\right]$ and $\mathbf{y} = \left[y_0, y_1, \cdots, y_{N-1}\right]$, and some *a priori* information about the plant, establish whether they are consistent, and if so, find a model that interpolates the experimental data within the measurement error level.

In the sequel, we will make the following standard assumptions about the *a priori* information:

A1.- The linear portion of the plant belongs to the set

$$\mathcal{S} \doteq \overline{\mathcal{B}}\mathcal{H}_{\infty,\rho}(K),$$

that is, we consider exponentially stable plants with a stability margin of $(\rho - 1)$ and a peak response to complex exponential inputs bounded by some known K.

A2.- The static nonlinearity can be expanded in terms of some known basis functions:

$$f_i(.) = \sum_{j=1}^{n_f} b_{i,j}\psi_j(.), i = 1, 2, \cdots, n_y; \qquad \psi_j(.) \text{ known} \tag{5.2}$$

A3.- The measurement noise satisfies:

$$\eta \in \mathcal{N} \doteq \{\eta \colon \|\eta_k\|_\infty \leq \epsilon\}$$

With these assumptions, the problem under consideration can be precisely stated as:

Problem 12. *Given the* a priori *information* $\mathcal{S}, \mathcal{N}, \psi_{i,j}(.)$ *and the* a posteriori *experimental data* $\{\mathbf{y}, \mathbf{u}\}$*, determine:*

1. *if the* a priori *and* a posteriori *information are consistent, i.e., the consistency set*

$$\mathcal{T}(\mathbf{y}) \doteq \{H \in \overline{\mathcal{B}}\mathcal{H}_{\infty,\rho}(K) \colon\ y_k = \mathbf{B}\mathbf{\Psi}\left[(h * u)_k\right] + \eta_k, \quad \textit{for some } \mathbf{B} \in R^{n_y \times n_f} \textit{ and some sequence} \quad \eta_k \in \mathcal{N}, k = 0, 1, \ldots, N-1\} \tag{5.3}$$

is nonempty, where

$$\mathbf{\Psi} = \left[\psi_1, \psi_2, \cdots, \psi_{n_f}\right]^T$$

2. If $\mathcal{T} \neq \emptyset$, find a nominal model $\{H, f(.)\}$ that interpolates the experimental data.

5.3 Risk-adjusted set-membership Identification of Wiener Systems

In this section we present a solution to Problem 12 based on the use of risk adjusted ideas. We begin by recasting the problem into an equivalent, albeit non–convex, optimization form.

5.3.1 Establishing Consistency

Lemma 3. *Given*

1. the a priori *information:*

$$K > 0, \rho > 1,$$

and a nonlinear vector function

$$\mathbf{\Psi}(\cdot) = \left[\psi_1, \psi_2, \cdots, \psi_{n_f}\right]^T,$$

2. two vector sequences of experimental data

$$\mathbf{y} = \left[y_0, y_1, \cdots, y_{N-1}\right]^T,$$

$$\mathbf{u} = \left[u_0, u_1, \cdots, u_{N-1}\right]^T,$$

there exist a linear operator

$$\mathbf{H}(z) \in \overline{\mathcal{B}}\mathcal{H}_{\infty,\rho}(K)$$

and a nonlinear mapping

$$f(\cdot) = \mathbf{B}\boldsymbol{\Psi}(\cdot)$$

such that the consistency set $\mathcal{T}(\mathbf{y})$ *is nonempty, if and only if there exist a vector* $\mathbf{h}$ *and a matrix* $\mathbf{B}$ *satisfying:*

$$\mathbf{M}(\mathbf{h}) \doteq \begin{bmatrix} KR^{-2} & \left(\mathbf{T}_h\right)^T \\ \mathbf{T}_h & KR^2 \end{bmatrix} \geq 0 \tag{5.4}$$

$$\mathbf{y} - \mathbf{B} \cdot \boldsymbol{\Psi}(\mathbf{T}_h \mathbf{u}) \in \mathcal{N}$$

where $\mathbf{T}_u$ *and* $\mathbf{T}_h$ *are the lower Toeplitz matrix associated with the sequences* $\mathbf{u}$ *and vector* $\mathbf{h}$ *respectively, and* $R = diag[1, \rho, \rho^2, \cdots, \rho^{N-1}]$.

Proof. The first condition in (5.4) follows now from the fact that

$$H(z) \in \overline{\mathcal{B}}\mathcal{H}_{\infty,\rho}(K) \Longleftrightarrow \frac{1}{K} H(\rho z) \in \overline{\mathcal{B}}\mathcal{H}_{\infty}$$

followed by a Schur complement argument. The second condition is a restatement of the conditions

$$\begin{aligned} \omega_k &= (h * u)_k, \\ y_k &= f(\omega_k) + \eta_k \text{ for some } \eta_k \in \mathcal{N} \end{aligned}$$

□

Note that the second constraint in (5.4) leads to a computationally hard to solve, non–convex optimization problem, even in the case where $\boldsymbol{\Psi}(.)$ is affine, on account of

the bilinearity of $\mathbf{B} \cdot \mathbf{\Psi}(\mathbf{T}_h \mathbf{u})$ in variables $\mathbf{B}$ and $\mathbf{h}$. In the next subsection we propose to circumvent this difficulty by pursuing a risk–adjusted approach.

5.3.2 A Risk-adjusted relaxation

In this subsection we exploit some recently introduced results on sampling systems in $\overline{\mathcal{B}\mathcal{H}}_\infty$ in an attempt to find a system such that, when cascaded with the nonlinearity $\mathbf{B\Psi}$, explains the observer input/output data. Note that instead of sampling $\overline{\mathcal{B}\mathcal{H}}_\infty$, we generate samples over $\overline{\mathcal{B}\mathcal{H}}_\infty^N$ to avoid the difficulty caused by the infinite dimension of $\overline{\mathcal{B}\mathcal{H}}_\infty$. The basis of doing so is that only the first N Markov parameters affect the output y for a causal operator. This renders the second constraint in (5.4) convex, allowing for recasting the identification problem into a convex LMI optimization form. This idea is formalized in the following algorithm:

Algorithm 4.

1.- Choose N_1 *and generate* N_s *samples* $\{\hat{\mathbf{h}}\}_{i=1}^{N_s}$ *of the set* $\overline{\mathcal{B}\mathcal{H}}_{\infty,\rho}^N(K)$ *by using the algorithm 1 to sample* $\overline{\mathcal{B}\mathcal{H}}_\infty^N$*, followed by the transformation*

$$\mathbf{h}^i = K \cdot [1, \rho^{-1}, \rho^{-2}, \cdots, \rho^{-(N-1)}] \cdot \hat{\mathbf{h}}^i.$$

Set $i = 1$.

2.- Solve the following optimization problem:

$$\mu(\mathbf{h}_i) = \min_{\mathbf{B}} \|\mathbf{y} - \mathbf{B\Psi}(\mathbf{T}_{\mathbf{h}^i}^N \mathbf{u})\|_\infty \tag{5.5}$$

3.- If $\mu(\mathbf{h}^i) \leq \epsilon$ *or* $i = N_s$*, stop. Otherwise, set* $i = i + 1$ *and go back to step 2.*

The algorithm finishes either by finding one feasible pair $\{\mathbf{h}, \mathbf{B}\}$ or after N_s steps, in which case the *a posteriori* experimental data is deemed to invalidate the *a priori* information. As we show next, if the number of samples N_s is large enough, the risk of incorrectly concluding that $\mathcal{T}(\mathbf{y}) = \emptyset$ can be made arbitrarily small, possibly at the expense of increased computational time.

Lemma 4. *Let* (ν, δ) *be two positive constants in* $(0, 1)$*. If* N_s *is chosen such that*

$$N_s \geq \frac{\ln(1/\delta)}{\ln(1/(1-\nu))},$$

then, with probability greater than $1 - \delta$*, the probability of not finding a feasible pair* $\{\mathbf{h}, \mathbf{B}\}$ *when one exists is smaller that* ν*.*

Proof. Note that $\mathcal{T}(\mathbf{y}) \neq \emptyset$ if there exists at least one $\mathbf{h} \in \overline{\mathcal{B}}\mathcal{H}^N_{\infty,\rho}(K)$ such that $\mu(\mathbf{h}) \leq \epsilon$. Direct application of the results in [97] shows that if the number of samples is at least N_s then

$$\begin{aligned} \mathbf{Prob}\big\{\mathbf{Prob}\{\exists \mathbf{h} \in \overline{\mathcal{B}}\mathcal{H}^N_{\infty,\rho}(K) \colon \mu(\mathbf{h}) \leq \epsilon \colon \\ \{\mu(\mathbf{h}_i)\}_{i=1}^{N_s} > \epsilon\}\big\} \leq \nu\big\} \geq (1-\delta), \end{aligned} \tag{5.6}$$

which yields the desired result. □

5.4 Application on Human Motion Modeling and Tracking

5.4.1 Motivation

The problems of modeling and tracking human motion using as input images from a sequence of video frames has been the subject of extensive research in the computer vision community, see for instance [65, 40, 25, 44, 77, 78, 90] and references therein. A difficulty with these approaches stems from the need to search very high dimensional spaces in order to find the correct values of the parameters of the underlying model. This is due to the fact that even using small size images and considering only human silhouettes still requires processing hundreds of pixels from each frame. In [61], nonlinear dimensionality reduction methods are used to map the high dimensional images into low dimensional manifolds that preserve neighborhood properties of the original data. It has been conjectured that this problem can be partitioned into two decoupled problems: (a) a tracking problem in a low dimensional manifold, accounting for the *dynamics* of the motion, and (b) a nonlinear, static mapping that accounts for the changes in appearance of the target. Moreover, the results there strongly suggest that the dynamics that account for the motion in the low dimensional manifold are linear, although no formal proof of the fact is given. [9] establish a linear operator by taking the dynamical appearance and motion of the target as the output driven by a stochastic signal.

5.4.2 The *a priori* Information and *a posteriori* experimental data

Motivated by the above results, in this subsection we apply the proposed Wiener systems identification framework to the problem of human motion modeling and tracking.

The starting point is to *postulate* that such motion can be modeled as the impulse response of a Wiener system whose output is the observed images. Furthermore, the inverse mapping of the dimensionality reduction gives the nonlinear subsystem $f(\cdot)$ and the dynamics in the low dimensional manifold devotes the linear subsystem $H(z)$. The proposed framework can then be used to substantiate this hypothesis by establishing consistency of the *a priori* assumptions and the *a posteriori* experimental data, and to find a suitable model.

A widely used nonlinear dimensionality reduction method, Local Linear Embedding (LLE) (see [80]), is considered here. Since the output human silhouettes are very high-dimensional, learning the mapping from the output image space, $\mathcal{R}^d$ (this dimension is decided by the number of pixels in each frame of the output silhouettes) to the embedding space, low-dimension space, requires very large number of samples, in order to be able to interpolate. This fact leads to the infeasibility of learning the dimensionality reduction mapping, i.e., the nonlinear subsystem $f(\cdot)$ is not invertible. We use Generalized Radial Basis Function (GRBF) interpolation framework to learn the nonlinear mapping from embedding space to the visual output space (see [22, 61]).

As shown in Figure 5.1, the set of output silhouettes and their corresponding points in the embedding space are denoted as

$$\mathbf{y} = y_k \in \mathcal{R}^d, k = 0, \cdots, N-1$$

and

$$\omega = \omega_k \in \mathcal{R}^e, k = 0, \cdots, N-1,$$

respectively. Note, e is the dimensionality of the embedding space. Let

$$c_j \in \mathcal{R}^e, j = 1, \cdots, N_c$$

be a set of N_c centers in the embedding space where such centers can be obtained using k-means clustering (see [80]). Then, we can learn the nonlinear mappings from the embedding space to each individual pixel in the output image space by solving for multiple interpolants

$$f^i : \mathcal{R}^e \longrightarrow \mathcal{R}$$

where i is i-th dimension(pixel) in the output image space and f^i is a radial basis function interpolant. Of particular interest are functions of the form:

$$f^i(\omega) = p^i(\omega) + \sum_{j=1}^{N_c} w_j^i \phi(|\omega - t_j|) \tag{5.7}$$

where $\phi(\cdot)$ is a real-valued basic function, w_j are real coefficients, $|\cdot|$ is the norm on $\mathcal{R}^e$. Typical choices[1] for the basis function includes,

$$\text{thin-plate spline:} \quad \phi(u) = u^2 \log(u),$$

$$\text{Gaussian:} \quad \phi(u) = \exp(-cu^2),$$

$$\text{biharmonic:} \quad \phi(u) = u,$$

$$\text{and triharmonic:} \quad \phi(u) = u^3.$$

$p^i(x)$ is a linear polynomial with coefficients q^i, $q^i \in \mathcal{R}^{(e+1)}$. Therefore, the whole mapping can be written in a matrix form as:

$$f(\omega) = \mathbf{B}\mathbf{\Psi}(\omega) \tag{5.8}$$

, where $\mathbf{B}$ is a $d \times (N_c + e + 1)$ dimensional matrix with the i-th row $[w^i_1, \cdots, w^i_{N_c}, q^i]$ and the vector function $\mathbf{\Psi}(\omega)$ is

$$\left[\phi(|\omega - t_1|), \cdots, \phi(|\omega - t_{N_c}|), 1, \omega^T\right]^T$$

The matrix $\mathbf{B}$ represents the coefficients for d different nonlinear mappings, each from a low-dimension embedding space into real numbers.

[1] All these different $\phi(\cdot)$ give almost the same result for the learning mapping

Motivated by this procedure and the physics of the practical problem of human motion modeling and tracking, the following assumptions concerning the *a priori* information can be made to accomplish wiener system identification:

1.- The output of the LTI part, ω, evolves in a 3-dimensional space[2] (this hypothesis is motivated by the physics of the problem, where ω is related to the coordinates of the centroid of the target).

2.- Choosing the basis function as Gaussian, the static nonlinearity $f(\omega)$ is given:

$$\begin{aligned} f(\omega) &= \mathbf{B}\boldsymbol{\Psi}(\omega) \\ &= \mathbf{B}\left[\phi(|\omega - t_1|), \phi(|\omega - t_2|), 1, \omega^T\right]^T \end{aligned}$$

with $\phi(x) = \exp(-0.8x^2)$ and

$$\begin{bmatrix} t_1 & t_2 \end{bmatrix} = \begin{bmatrix} 0.6833 & -0.7552 \\ -0.4521 & 0.4997 \\ -0.0033 & 0.0036 \end{bmatrix}$$

3.- Measurements of the pixels values are corrupted by measurement noise of up to 10% of their peak value. (This accounts for both actual measurement noise and errors in establishing pixel correspondences across frames).

The experimental data, shown in Figures 5.2 and 5.4, consists of the first 20 frames of a human walking on a treadmill, each having 1728 pixels, taken from the CMU

[2] i.e. $e = 3$

MOBO database. Then, The measurements vector $\mathbf{y}$, of dimension 1728 ($d = 1728$), was constructed by row-wise stacking the grey-scale values of the pixels in each frame.

5.4.3 Identification Results

Applying Algorithm 1 with $N_s = 470$ samples, which yields a probability of 99% of establishing consistency with confidence 99% led to a feasible pair $\{\mathbf{h}, \mathbf{B}\}$ and hence a model explaining the experimental data. The central interpolant[3] corresponding to the linear portion of the model is given as the following 5^{th} order system (after a model reduction step):

$$H(z) = \begin{bmatrix} \dfrac{-0.03z^5 + 0.19z^4 - 0.30z^3 + 0.22z^2 - 0.08z + 0.02}{z^5 - 1.84z^4 + 1.13z^3 + 0.92z^2 - 1.57z + 0.83} \\ \\ \dfrac{-0.08z^5 + 0.09z^4 - 0.02z^3 - 0.08z^2 + 0.07z - 0.03}{z^5 - 1.84z^4 + 1.13z^3 + 0.92z^2 - 1.57z + 0.83} \\ \\ \dfrac{0.14z^5 - 0.35z^4 + 0.27z^3 + 0.12z^2 - 0.30z + 0.16}{z^5 - 1.84z^4 + 1.13z^3 + 0.92z^2 - 1.57z + 0.83} \end{bmatrix} \tag{5.9}$$

The corresponding static output nonlinearity is given by $\mathbf{B\Psi}$, where a surface plot of the matrix $\mathbf{B}$ is shown in Figure 5.6. Here the x, y axes correspond to the index of the matrix and the z axis to the matrix value. Note that in most cases, the rows of $\mathbf{B}$ are

[3] All solutions in the consistency set can be parameterized in terms of a free parameter $Q(z) \in \overline{\mathcal{BH}}_{\infty,\rho}$. The central solution corresponds to $Q(z) = 0$.

sharply peaked around one or two values, indicating that these pixels can be explained using fewer elements of the bases $\Psi_{i,j}$.

The impulse response of the identified system is shown in Figures 5.3 and 5.5. As illustrated there, the system is able to correctly predict the appearance of the target in the next two frames (21 and 22). For comparison purposes, the actual output, not used for training, is shown in the last 2 frames of Figure 5.4. The worst case prediction error in this interval, found by comparing the actual and predicted sequences for the 1728 pixels, is 21.73% of the ℓ_∞ norm of the actual output. The mean square error is 0.1393 in terms of the ℓ_∞ norm of the error signal.

Fig. 5.2. First 11 frames of a walking person video sequence

Fig. 5.3. First 11 frames of the output of the identified Wiener System

Fig. 5.4. Frames 12 to 22 of the actual walking sequence (Frames 21 and 22 were not used in the identification)

Fig. 5.5. Frames 12 to 22 predicted by the identified system

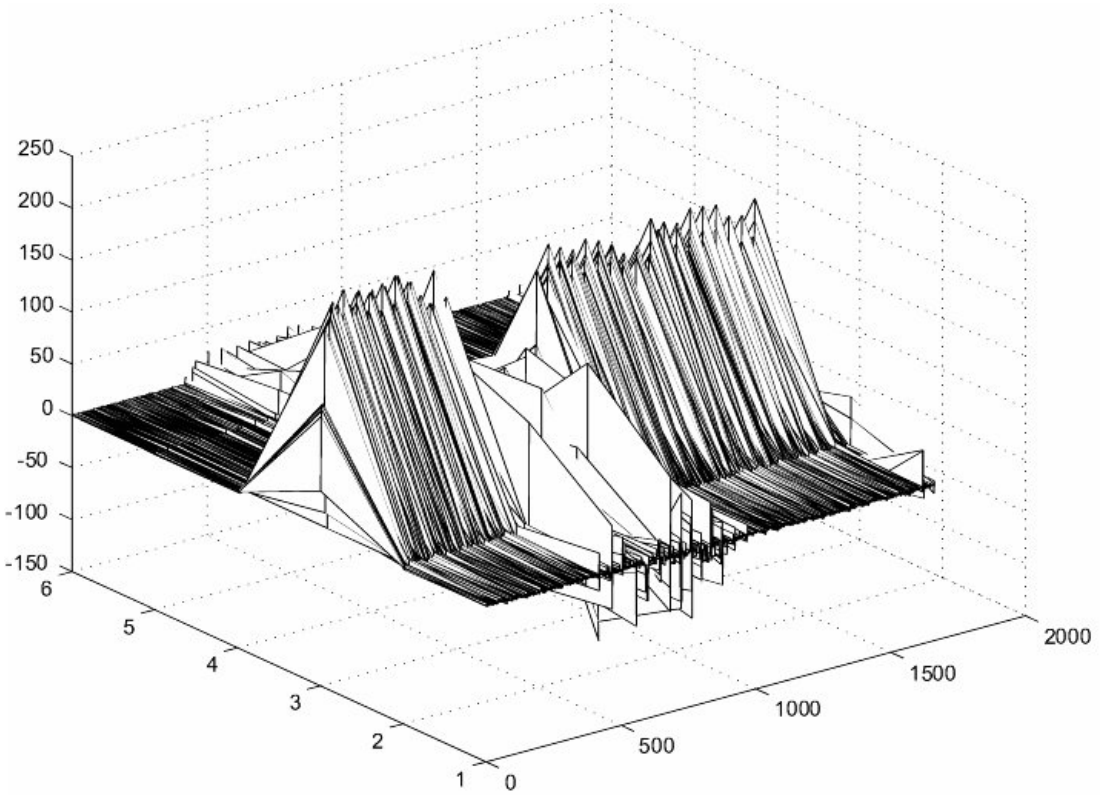

Fig. 5.6. Surface of **B** for the nonlinear part of the model

Chapter 6

Conclusion

This dissertation addresses several open problems which have not been noticed but play important roles to the robust identification area in the control community. It takes care of the effect of the system inputs prior to an initial time which leads to an artificial high identification error. Applications of the so–called semi–blind robust identification and semi–blind model (in)validation on Economy and Fault Detection and Isolation (FDI) are illustrated respectively. Moreover, motivated by the problem of human motion modeling and tracking, a set-membership standpoint identification method is developed for the wiener system identification.

Chapter 3 studies the so–called semi–blind robust identification problem. Due to the fact that most of the research on robust identification assumes that the input–output experimental data are generated exactly from an initial time $t = t_0$, which is not true for most practical situations, the robust identification results in a high identification error. To improve this disadvantage, we states the semi–blind robust identification problem as: identify a system with the situation where only the input–output data applied after some initial time $t = t_0$ are known exactly. While the only information available about inputs prior to time t_0 is the characteristic of its norm–bounded set. By encapsulating the effects of inputs prior to time t_0 in some unknown, non–zero initial conditions x_0 and assuming our system controllable, this problem is shown as an non-convex, NP–hard,

BMI problem utilizing the time–domain interpolatory identification algorithm. Through relaxing the bound of the partial output due to the past input sequence, this BMI problem can be reduced to a solvable convex optimization problem and from thereupon a suitable mathematical representation for the physical system is obtained.

Notice that robust control tools have the potential to optimize the performance of economical and financial systems, provided that a suitable model of the system under consideration, as well as bounds of the identification error, are available. In chapter 3, we illustrate the potential of the application of semi–blind robust identification technique on economy and finance by identifying and validating[1] a subsystem of the macro–economy relating inflation and interest–rates. This application benefits on several aspects, such as it provides worst–case estimates of the identification error that can be used to determine for how long the predictions of the model will be valid, in the absence of additional data. The resulting model obtained using semi–blind robust identification is able to correctly forecast the inflation rate over 149 quarters of real data, outperforming a widely used linear model, even though the latter was obtained using data registers 7 times longer. Research currently under way seeks to extend these techniques to piecewise linear models, which allows for addressing situations where the systems dynamics change over time.

For future works use, blind robust identification procedure is also introduced in chapter 3. Utilizing an algorithm[2] that generates uniformly distributed finite samples over bounded set (see [46]), an identification algorithm (see [58]), which finishes either by

[1] contained in chapter 4

[2] this algorithm is also used for semi–blind model (in)validation and wiener system identification in chapter 4 and chapter 5, respectively

finding a feasible solution for the *a priori* information and the *a posteriori* experimental data or by running out of the samples, is presented.

Chapter 4 focuses on the problem of semi–blind model (in)validation. Motivated by the same facts as for semi–blind robust identification, the goal of semi–blind model (in)validation is to test the validity of the identified system, consists of a nominal plant and an appropriate uncertainty description, with a set of *new* experimental data in the situation where only the input–output data after initial time $t = t_0$ are known exactly. It is shown that, in general, this problem is not directly solvable due to the non-joint convexness in all variables. To remove this limitation, we propose two convex relaxations: one deterministic and one stochastic. Considering a special case that the nominal dynamic is subject to *multiplicative* unstructured uncertainty, the deterministic relaxation moves the measurement noise to a place where it is also affected by the uncertainty Δ and modifies the noise set bound correspondingly. It reduces the problem to a conventional LMI optimization, however, it is not only limited to the special case of multiplicative uncertainty but also too conservative when the condition $\|\Delta\|_\infty \leq 1$ can not be hold. The stochastic relaxation overrides these difficulties by uniformly sampling the admissible uncertainty set $\mathbf{\Delta}$, in return for an (arbitrarily small) probability of missing a feasible solution.

Noticing the similarity between the setup structure of semi–blind model (in)validation and the structure of FDI problem, in chapter 4, we states FDI problem in a semi–blind model (in)validation framework. Two illustrative examples, one SISO and one MIMO, are given to demonstrate the potential of this application. Comparing the performance of these (in)validation based approaches with the observer–based approach (see [21]) in

the first example, the following results are noticed: for larger values of $\|f\|$ all methods have comparable performance, while in the case of smaller $\|f\|$ both invalidation based approaches outperform the observer–based one. Furthermore, the former approaches are able to accurately estimate the components of a fault, in addition to establishing its existence, at the expense of on–line computational time. This suggests that a mixed strategy of using observer–based or deterministic approaches first to establish the existence of a fault, followed by use of the stochastic method to establish its location, may give some advantage. In the second example, two sets of estimations are obtained with two different levels of uncertainty bound and noise bound. As we expected, the comparison of them indicates that the performance of both relaxations becomes better with decreasing uncertainty and noise bounds. In addition, for relatively low uncertainty levels, both relaxations have comparable performance, while the stochastic relaxation outperforming the convex one as the uncertainty level increases.

Chapter 5 addresses the problem of identification for wiener system, consists of the cascade of a LTI model and a memoryless, static non–linearity. Motivated by the fact that most robust identification algorithms just handle linear systems and wiener system is widely applied in many engineering problems, the identification of wiener model attracts our attention. Actually it has been an active research area for many years. In this chapter, we propose an algorithm, which avoids the requirement of either special statistic characteristic for input signal or special system structure, from a deterministic set–membership standpoint for wiener systems using time–domain data. As shown, in principle, this formulation leads to a non–convex, computationally hard to solve optimization problem. However, by exploiting recently introduced results on sampling of

transfer functions, the problem can be relaxed to a convex optimization, at the price of an arbitrarily small probability of mis-identifying the plant.

One meaningful application on a problem that has been the object of considerably attention in the computer vision community: modeling the evolution of human motion in a sequence of two-dimensional images is illustrated in chapter 5. Applying wiener system identification framework, the difficulty of searching very high dimensional spaces due to the huge numbers of pixels of each frame is overcome. By modeling this evolution as the impulse response of a Wiener system, we were able to establish that the problem can be indeed decoupled into two simpler subproblems: (i) tracking the trajectory of an LTI system in a low dimensional subspace and (ii) finding a nonlinear static mapping that accounts for appearance changes. By decoupling the intrinsic dynamics of the target from changes in its appearance, this decomposition is the first step towards designing faster, more robust trackers. Efforts are currently underway to generalize the techniques proposed in this chapter to cases where both the linear dynamics and the nonlinearity are slowly time varying.

6.1 Related Publications

A list of publications related with the results of this dissertation, either in part or as a whole, is given below:

- M. Sznaier, W. Ma, M. Yilmaz, and C. Lagoa, "Semi-Blind Robust Identification/Model (In)Validation with Applications to Macro-Economic Modeling.", International Federation of Automatic Control, July 2005.

- M. Sznaier, W. Ma and C. Lagoa., "A Risk Adjusted Approach to Robust Simultaneous Fault Detection and Isolation.", International Federation of Automatic Control, July 2005.

- W. Ma, H. Lim, M. Sznaier, and O. Camps, "Risk-adjusted Identification of Wiener System", Proceedings of the IEEE Conference on Decision and Control, December 2006.

- M. Sznaier, W. Ma, and C. Lagoa, "A Risk-Adjusted Approach to Robust Simultaneous Fault Detection and Isolation", Automatica, to be published.

- M. Yilmaz, W. Ma, M. Sznaier, and A. Yavash, "A Robust Identification/Model (In)Validation Approach to Macro-Economic Modeling", Journal of Economic Dynamics and Control, under review.

Appendix A

Identification Error Analysis

[1] Robust Identification provides an identification error that is valid no matter which model $g \in \mathcal{S}$ is the real plant leading to a worst–case approach, due to its deterministic feature. In addition, the identification error should "cover" all models $g \in \mathcal{S}$ that, combined with all possible noise vectors $\eta \in \mathcal{N}$, are consistent with the experiments. To this effect, given an identification algorithm $\mathcal{A}$ which maps both *a priori* and *a posteriori* information to a candidate nominal model, it's worst-case error can be defined as:

$$e\big(\mathcal{A}\big) \doteq \sup_{\eta \in \mathcal{N}, g \in \mathcal{S}} m\{g, \mathcal{A}\big[E(g,\eta), \mathcal{S}, \mathcal{N}\big]\} \tag{A.1}$$

where $E(\cdot,\cdot)$ is the experiment operator and $m(\cdot, cdot)$ is a specific metric. This identification error is a *global* one in terms of the experiments since it considers all possible experimental outcomes. A *local* error related with a specific experiment $\mathbf{y}$ can be defined as:

$$e\big(\mathcal{A}, \mathbf{y}\big) \doteq \sup_{g \in \mathcal{T}(\mathbf{y})} m\big[g, \mathcal{A}(\mathbf{y}, \mathcal{S}, \mathcal{N})\big] \tag{A.2}$$

According to the definition of the set of all possible experimental data,

$$\mathbf{Y} \doteq \{E(g,\eta) | g \in \mathcal{S}, \eta \in \mathcal{N}\},$$

[1] refer to [86]

the set $\{g \in \mathcal{T}(\mathbf{y}), \mathbf{y} \in \mathbf{Y}\}$ is equivalent to $\mathcal{S}$. Then, the following is obtained by replacing $\mathbf{y}$ by $E(g,\eta)$ in (A.2):

$$\sup_{g\in\mathcal{S}}\sup_{\eta\in\mathcal{N}} m\{g, \mathcal{A}\big[E(g,\eta),\mathcal{S},\mathcal{N}\big]\} = \sup_{\mathbf{y}\in\mathbf{Y}}\sup_{g\in\mathcal{T}(\mathbf{y})} m\{g, \mathcal{A}\big[\mathbf{y},\mathcal{S},\mathcal{N}\big]\}$$

i.e.,

$$e\big(\mathcal{A}\big)= \sup_{\mathbf{y}\in\mathbf{Y}} e\big(\mathcal{A},\mathbf{y}\big) \tag{A.3}$$

In order to obtain mathematically tractable bound for these errors, let's introduce several definitions first(see [86]).

Definition 3. *The radius and diameter of a subset $\mathcal{A}$ of a metric space $\big(\mathcal{X}, m\big)$ are*

$$r\big(\mathcal{A}\big)= \inf_{x\in\mathcal{X}}\sup_{a\in\mathcal{A}} m(x,a) \tag{A.4}$$

$$d\big(\mathcal{A}\big)= \sup_{x,a\in\mathcal{A}} m(x,a) \tag{A.5}$$

When considering the set $\mathcal{A}$ as represented by a single "central" point (which might not belong to $\mathcal{A}$), the radius of the set can be interpreted as the maximum error that is measured in the metric $m(\cdot)$. The diameter is the maximum distance between any tow points in the set.

Recall that the set $\mathcal{T}(\mathbf{y}) \subset \mathcal{S}$ is the largest set of indistinguishable models. Therefore, in some sense, the following definition represents the "size" in a certain metric, of the largest set of indistinguishable models according to all possible *a priori* and *a posteriori* information.

Definition 4. *The radius and diameter of information are defined as*

$$\mathcal{R}(\mathcal{I}) \doteq \sup_{\mathbf{y}\in\mathbf{Y}} r\big[\mathcal{T}(\mathbf{y})\big] \tag{A.6}$$

$$\mathcal{D}(\mathcal{I}) \doteq \sup_{\vec{y}\in\mathbf{Y}} d\big[\mathcal{T}(\mathbf{y})\big] \tag{A.7}$$

This "size" cannot be decreased by any identification algorithm, unless extra information is added to the problem. In fact, it depends only on the *a priori* information because it does not correspond to a particular experiment, but to all possible experiments.

Based on the concepts of radius and diameter of a set and of information, the following Lemmas(see [86]) offer a lower bound for the worst-case identification error of any identification algorithm :

Lemma 5. *The worst-case identification error defined in* (A.1) *satisfies the following inequality:*

$$e(\mathcal{A}) \geq \mathcal{R}(\mathcal{I}) \geq \frac{1}{2}\mathcal{D}(\mathcal{I}) \tag{A.8}$$

for any algorithm $\mathcal{A}$*.*

Of special interest in this dissertation, the interpolation algorithm, denoted by $\mathcal{A}_I$, is the algorithm that interpolates the given experimental samples so that a nominal model g_{id} in the consistency set $\mathcal{T}(\mathbf{y})$ can be generated. For this special class of algorithms, there is one further lemma that offers an upper bound for worst–case error.

Lemma 6. *For any interpolation algorithm $\mathcal{A}_I$, the following upper bound holds:*

$$\mathcal{D}(\mathcal{I}) \geq e(\mathcal{A}_I) \tag{A.9}$$

The consequence of the above two lemmas makes interpolation algorithms very appealing in the sense of *nearly* optimal, i.e. it can obtain the minimal error bound $\mathcal{R}(\mathcal{I})$ up to a factor of 2, that is,

$$\mathcal{D}(\mathcal{I}) \geq e(\mathcal{A}_I) \geq \frac{1}{2}\mathcal{D}(\mathcal{I}) \tag{A.10}$$

In the view of (A.7) and (A.10), the question whether there exists a worst–case experiment $\mathbf{y}_0$ in the sense of achieving the upper bounds for $r\big(\mathcal{T}(\mathbf{y})\big)$ and $r\big(\mathcal{T}(\mathbf{y})\big)$ evokes our attention. The following Lemma gives an answer by stating that the experimental output, achieved by the centers of the *a priori* sets, $\mathcal{S}$ and $\mathcal{N}$, provides the least amount of information (largest diameter of indistinguishable set) if both of the sets are symmetric and convex.

Lemma 7. *If the* a priori *sets $\mathcal{S}$ and $\mathcal{N}$ are symmetric and convex with respect to symmetry points $c_{\mathcal{S}}$ and $c_{\mathcal{N}}$, respectively, and the experimental mapping is linear, then, the diameter of information satisfies:*

$$\mathcal{D}(\mathcal{I}) = \sup_{\mathbf{y}\in\mathbf{Y}} d\big[\mathcal{T}(\mathbf{y})\big] = d\big[\mathcal{T}(\mathbf{y}_0)\big], \tag{A.11}$$

$$\mathbf{y}_0 = E(c_{\mathcal{S}}, c_{\mathcal{N}})$$

Furthermore, it holds that

$$d\big[\mathcal{T}(\mathbf{y}_0)\big] = 2 \sup_{g \in \mathcal{T}(\mathbf{y}_0)} m(g, 0) \tag{A.12}$$

The bound on the worst-case identification error is given by twice the maximum distance from any element in $\mathcal{T}(\mathbf{y}_0)$ to the center of symmetry of $\mathcal{S}$ and $\mathcal{N}$.

Appendix B

Convergence

[1] As we mentioned in Chapter 3, robust identification considers model uncertainty from two different sources: measurement noise and lack of knowledge of the system itself due to limited information supplied by the experiment. These two sources, viewed as "corruption" and "partialness" respectively, initiates the concern of convergence. An identification algorithm $\mathcal{A}$ is said to be convergent when the worst–case *global* identification error $e(\mathcal{A})$ in (A.1) goes to zero as both the *a priori* and *a posteriori* information tends to be "complete", in the sense of zero "partialness" and zero "corruption". When both zero "partialness", either the set $\mathcal{S}$ tends to the real plant or the experiment operator provides all necessary data, and zero "corruption", the noise set $\mathcal{N}$ reduces to a single element (usually zero noise), the diameter of information tends to zero. It is convenient to perceive that when the consistency set $\mathcal{T}(\mathbf{y})$ goes to only one element: the real system, i.e., $g_{id} \rightarrow g^*$ (where g_{id} is the identified model and g^* is the real plant), the *a priori* and *a posteriori* information is completed and there is no uncertainty.

Thus, the definition of convergence can be stated as follows:

Definition 5. *An identification algorithm $\mathcal{A}$ converges if and only if*

$$\lim_{\mathcal{D}(\mathcal{I})\rightarrow 0} e(\mathcal{A}) = 0 \tag{B.1}$$

[1] refer to [86]

Straightforward from Lemma 5, all interpolation algorithms are convergent.

Note that the worst–case error defined in (A.1) depends on the classes of models and measurement noise. According to the definition of convergence, it is reasonable to expect the sets $\mathcal{S}$ and $\mathcal{N}$ to meet some necessary conditions so that the existence of convergent algorithms can be guaranteed. The following lemma provides a condition on $\mathcal{N}$ that precludes the existence of these types of algorithms (see [**?**]):

Lemma 8. *If the* a priori *class $\mathcal{S}$, which belongs to metric space (μ, m), contains a ball of size γ, say,*

$$\mathcal{B}_{\gamma}(g_0) = \{g \in \mu, m(g, g_0) \leq \gamma\} \in \mathcal{S},$$

then,

$$\mathcal{R}(\mathcal{I}) \geq \gamma$$

for any number of measurements and any noise level.

At the very least, the class of models $\mathcal{S}$ must be uniformly bounded in order to have a bounded consistency set, and furthermore it should be compact and include only stable and proper models in order to guarantee convergent algorithms.

Appendix C

White Noise Descriptions in the Time Domain

[1] In order to decide whether an discrete–time experimental signal of finite length N is a sample of white noise or not, we need to a perform a hypothetic test in terms of some statistic. A common choice (see [7, 47]) is the sample circular autocorrelation sequence (correlogram), defined as

$$\mathbf{R}_{\eta}(\tau) = E\{\eta(k+\tau)\eta(k)\}, \tau = 0, 1, \cdots, N-1 \tag{C.1}$$

It should approximate a delta function, which represents the expected correlation of white noise. In other words, a discrete–time signal $\eta(k)$ is considered as white if its circular autocorrelation $\mathbf{R}_{\eta}(\tau)$, for τ in a certain range (e.g. $\tau = 1, 2, \cdots, T$), is relatively small with respect to

$$\mathbf{R}_{\eta}(0) = E\{\eta(k)\eta(k)\} = \|\eta\|_2^2.$$

For example, as shown in Figure C.1, we can specify that the correlogram, which is normalized to $\mathbf{R}_{\eta}(0) = 1$, falls inside a band of width γ around zero.

Thus, the following definition is motivated:

[1] refer to [69]

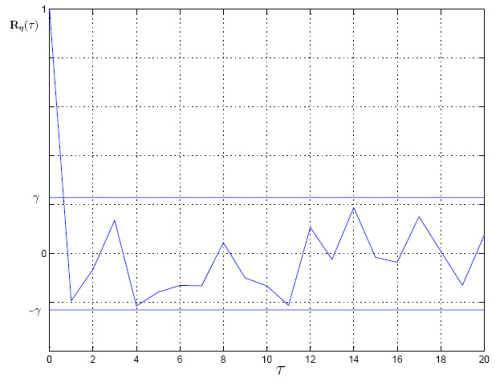

Fig. C.1. Correlogram of a random sequence

Definition 6. *The signal sequence*

$$\eta = \big[\eta(0), \eta(1), \cdots, \eta(n-1)\big]$$

of length n is categorized as white with accuracy γ, up to lag T, in the time domain sense, if the following condition is satisfied:

$$\{\eta \in \mathcal{R}^N : |\mathbf{R}_\eta(\tau)| \leq \gamma \mathbf{R}_\eta(0), \quad \textit{for} \quad \tau = 1, 2, \cdots, T\} \tag{C.2}$$

From the classical statistical point of view, the choice of γ is associated to a level of significance of the test. Obviously, the choice $\gamma = 0$ would give a clean white noise, however, unfortunately, realistic white noise signals will not have exactly zero autocorrelations in a finite horizon setting. Setting an appropriate decay rate for γ,

$$\text{e.g.} \qquad \gamma = \frac{1}{N^\alpha}, \quad \text{with} \quad \alpha \leq \frac{1}{2},$$

gives a high probability to the signal set.

There are other statistical tests that can also be applied to the correlogram (see [68, 67]). In this dissertation, we prefer the one in Definition 6, not only because it involves quadratic constraints on signal η, which lead to more tractable worst–case analysis, but also because it provides a very tight characterization of stochastic noise.

Appendix D

Locally Linear Embedding

[1] Locally Linear Embedding (LLE) is an unsupervised learning algorithm that computes low dimensional, neighborhood preserving embeddings of high dimensional data. In the sequel, we introduce this algorithm in detail.

Providing sufficient data consisting of N real–valued vectors,

$$\mathbf{x}_i \in \mathcal{R}^d, i = 1, 2, \cdots, N,$$

we expect each data point and its neighbors to lie on or close to a locally linear patch of the manifold. In the simplest formulation of LLE, K nearest neighbors per data point are identified in the sense of Euclidean distance. Alternatively, neighbors can be identified by choosing all points within a ball of fixed radius or some more complicated rules.

Reconstruction errors are measured as following:

$$\varepsilon(W) = \sum_{i=1}^{N} |\mathbf{x}_i - \sum_j W_{ij}\mathbf{x}_j|^2, \tag{D.1}$$

where W_{ij} is the reconstruction weight that presents the contribution of the jth data point to the ith reconstruction, specially, $W_{ij} = 0$ for the data points that does not belong to the set of the neighbors of $\mathbf{x}_i$. $\sum_{i=1}^{N}$ represents the summation of the squared distances

[1] refer to [80]

between all the data points and their reconstructions. Weights W_{ij} are computed by minimizing $\varepsilon(W)$ in (D.1) with constrain:

$$\sum_{j=1}^{K} W_{ij} = 1, \tag{D.2}$$

called "sum-to-one" constrain, leading the best reconstruction for each data $\mathbf{x}_i$.

Suppose the data lie on or near a smooth nonlinear manifold of dimension $e \ll d$. Then, there exists a linear mapping that maps the high dimensional coordinates of each neighborhood to global internal coordinates on the manifold. The reconstruction weights W_{ij} characterize intrinsic geometric properties of each neighborhood of the data that are invariant to exactly such transformations, due to the fact that they obey an important symmetry: for any particular data point, they are invariant to rotations, rescalings, and translations of that data point and its neighbors. Therefore, the embedded manifold coordinates in e dimensions can be reconstructed by the same weights W_{ij} that reconstruct the ith data point in d dimensions.

Based on the above idea, in the final step of the algorithm, each high dimensional data points $\mathbf{x}_i$ is mapped to a low dimensional point $\mathbf{y}_i$ representing global internal coordinates on the manifold. Points $\mathbf{y}_i$ computed through minimizing the embedding cost function:

$$\Phi(\mathbf{y}) = \sum_{i=1}^{N} |\mathbf{y}_i - \sum_{j} W_{ij}\mathbf{y}_j|^2, \tag{D.3}$$

gives the best reconstruction. Note, this time, the weights W_{ij} are fixed and obtained through the previous steps while optimizing the coordinates $\mathbf{y}_i$.

To recap, we summarize the LLE algorithm in the following:

Algorithm 5 (LLE Algorithm)**.**

Step 1. Compute the neighbors for each data point $\mathbf{x}_i$

Step 2. Compute the weights W_{ij} *by minimizing* $\varepsilon(W)$ *in* (D.1) *with constrain* (D.2)

Step 3. Compute the low dimensional data points $\mathbf{y}_i$ *by minimizing* (D.3) *with fixed weights* W_{ij} *obtained from step 2.*

Implementation of the algorithm is fairly straightforward. The algorithm only has one free parameter: the number of neighbors per data point, K. Once the neighbors of each data point are chosen, the optimal weights W_{ij} and coordinates $\mathbf{y}_i$ can be computed by standard methods in linear algebra.

References

[1] E. Anderson, L. P. Hansen, and T. Sargent. Robustness, detection and the price of risk. *mimeo*, 2000.

[2] E. Bai and M. Fu. Blind system identification and channel equalization of iir systems without statistical information. *IEEE Transactions on Signal Processing*, 47(7):1910–1921, 1999.

[3] E. W. Bai. An optimal two-stage identification algorithm for hammerstein-weiner nonlinear systems. *Automatica*, 34:333–338, 1998.

[4] E. W. Bai. A blind approach to the hammerstein-wiener model identification. *Automatica*, 38:967–979, 2002.

[5] E. W. Bai. Frequency domain identification of wiener models. *Automatica*, 39:1521–1530, 2003.

[6] J. A. Ball, I. Gohberg, and L. Rodman. Interpolation of rational matrix functions.

[7] M. S. Bartlett. *An Introduction to Stochastic Processes*. Cambridge: Cambridge University. Press., 1955.

[8] A. C. Brinker. A comparison of results from parameter estimation of impulse responses of the transient visual systems. *Biological Cybern*, 61:139–151, 1989.

[9] O. I. Camps, H. Lim, C. Mazzaro, and M. Sznaier. A caratheodory-fejer approach to robust multiframe tracking. *IEEE International Conference on Computer Vision*, pages 1048–1055, 2003.

[10] P. Celka and P. Colditz. Nonlinear nonstationary wiener model of infant eeg seizures. *IEEE Transactions on Biomedical Engineering*, 49:556–564, 2002.

[11] J. Chen. Frequency domain tests for validation of linear fractional uncertain models. *IEEE Transactions on Automatic Control*, 42(6):748–760, 1997.

[12] J. Chen and G. Gu. *Control-oriented system identification: An $\mathcal{H}_\infty$ Approach*. John Wiley & Sons, Inc., New York, 2000.

[13] J. Chen and C. N. Nett. The carathéodory-fejér problem and $\mathcal{H}_\infty/\ell_1$ identification: A time domain approach. *IEEE Transactions on Automatic Control*, 40(4):729–735, 1995.

[14] J. Chen, C. N. Nett, and M. K. H. Fan. Optimal nonparametric system identification from arbitrary corrupt finite time series: A control oriented approach. *Proceedings of 1992 American Control Conference, Chicago, IL, USA*, pages 279–285.

[15] J. Chen, C. N. Nett, and M. K. H. Fan. Worst-case system identification in $\mathcal{H}_\infty$: Essentially optimal algorithms, error bounds, and validation of a priori information. *Proceedings of American Control Conference*, pages 251–257, 1992.

[16] 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 Transactions on Automatic Control*, 40(7):1260–1265, 1995.

[17] J. Chen and S. Wang. Validation of linear fractional uncertain models: Solutions via matrix inequalities. *IEEE Transactions on Automatic Control*, 41(6):844–849, 1996.

[18] E. Collins and T. Song. Robust $\mathcal{H}_\infty$. *Journal of Guidance, Control and Dynamics*, (5):857–864.

[19] S. C. Cripps. *RF Power Amplifiers for Wireless Communications*. Artech House, Norwood, MA, 1999.

[20] M. A. Dahleh and I. J. Diaz-Bobillo. *Control of Uncertain Systems: A Linear Programming Approach*. Prentice Hall, Englewood Cliffs, NJ, 1996.

[21] S. X. Ding, P. Zhang, and P. M. Frank. Application of probabilistic robustness technique to the fault detection system design. *42nd IEEE Conference on Decision and Control.*, pages 972–977, 2003.

[22] A. Elgammal and C.S. Lee. Inferring 3d body pose from silhouettes using activity manifold learning. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 681–688, 2004.

[23] A. Emaimi-Naeimi, M. M. Akhter, and S. M. Rock. Effect of model uncertainty on fault detection: The threshold selector. *IEEE Transactions on Automatic Control*, 33(12):1106–1115, 1998.

[24] P. Frank and X. Ding. Survey of robust residual generation and evaluation in observer based fault detection systems. *Journal of Process Control*, 7(6):403–429, 1997.

[25] D. Gavrila and L. Davis. 3-d model-based tracking of humans in action: a multi-view approach. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 73–80, 1996.

[26] J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, New York, 1998.

[27] G. Giunta, G. Jacoviti, and A. Neri. Bandpass nonlinear system identification by higher cross correlation. *IEEE Transactions on Signal Processing*, 39:2092–2095, 1991.

[28] W. Greblicki. Nonparametric identification of wiener system. *IEEE Transactions on Information Theory*, 38:1487–1493, 1992.

[29] W. Greblicki. Nonparametric approach to wiener system identification. *IEEE Transactions on Circuits and Systems*, 44:538–545, 1997.

[30] G. Gu and P. Khargonekar. Linear and nonlinear algorithms for identification in $\mathcal{H}_\infty$ with error bounds. *IEEE Transactions on Automatic Control*, 37(7), 1992.

[31] G. Gu and P. P. Khargonekar. A class of algorithms for identification in $\mathcal{H}_\infty$. *Automatica*, 28(2):299–312, 1992.

[32] B.R.J. Haverkamp. Subspace identification of continuous-time wiener models. *IEEE Conference of Decision and Control*, 2:1846–1847, 1998.

[33] A. J. Helmicki, C. Jacobson, and C. Nett. $\mathcal{H}_\infty$ identification of stable lsi systems: a scheme with direct application to controller design. *Proceedings of American Control Conference, Pittsburgh*, pages 1428–1434, 1989.

[34] A. J. Helmicki, C. Jacobson, and C. Nett. Identification in $\mathcal{H}_\infty$: a robust convergent, nonlinear algorithm. *Proceedings of American Control Conference, San Diego*, 1990.

[35] A. J. Helmicki, C. A. Jacobson, and C. N. Nett. Identification in $\mathcal{H}_\infty$: linear algorithms. *Proceedings of American Control Conference, San Diego*, 1990.

[36] A. J. Helmicki, C. A. Jacobson, and C. N. Nett. Control oriented system identification: A worst case/deterministic approach in $\mathcal{H}_\infty$. *IEEE Transactions on Automatic Control*, 36(10):1163–1176, 1991.

[37] A. J. Helmicki, C. A. Jacobson, and C. N. Nett. Worst–case/deterministic identification in $\mathcal{H}_\infty$: the continuous–time case. *IEEE Transactions on Automatic Control*, 37(5), 1992.

[38] D. Henry and A. Zolghadri. Design and analysis of robust residual generators for sysems under feedback control. *Automatica*, 41:251–264, 2005.

[39] D. Henry, A. Zolghadri, M. Monsion, and S. Ygorra. Off–line robust fault diagnosis using the generalized structured sigular value. *Automatica*, 38:1347–1358, 2002.

[40] D. Hogg. Model-based vision: a program to see a walking person. *Image and Vision Computing*, 1:5–20, 1983.

[41] C. A. Jacobson, C. N. Nett, and J. R. Partington. Worst case system identification in ℓ_1: Optimal algorithms and error bounds. *Systems and Control Letters*, 19(6):419–424, 1992.

[42] B. Jiang, J. Wang, and Y. Soh. An adaptive technique for robust diagnosis of faults with independent effects on system outputs. *International Journal of Control*, 75(11):792–802, 2002.

[43] B. Kacewicz and M. Milanese. Optimal finite-sample experiment design in the worst-case ℓ_1 system identification. *Decision and Control, 1992., Proceedings of the 31st IEEE Conference on*, 1:56–61, 1992.

[44] I.A. Kakadiaris and D. Metaxas. Model-based estimation of 3d human motion with occlusion based on active multi-viewpoint selection. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 18–20, 1996.

[45] P. P. Khargonekar, A. Tikku, J. Krause, K. Poolla, and K. M. Nagpal. A time–domain approach to model validation. *IEEE Transactions on Automatic Control*, 39:951–959, 1994.

[46] C. M. Lagoa, X. Li, M. C. Mazzaro, and M. Sznaier. Sampling random transfer functions. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*. Springer–Verlag, London, 2004.

[47] L. Ljung. *System Identification Theory for the User*. Prentince–Hall, Englewood Cliffs, NJ, 1987.

[48] L. Ljung. Development of system identification. *Proceedings IFAC World Congress*, G:141–146, 1996.

[49] L. Ljung and Z. Yuan. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *IEEE Transactions on Automatic Control*, AC–30(6):514–530, 1991.

[50] Jr. Lucas, R. E. Econometric policy evaluation: A critique. *Carnegie Rochester Series on Public Policy*, 1:19–46, 1976.

[51] P. Mäkilä. Identification of stabilizable systems: closed–loop approximation. *International Journal of Control*, 54(3), 1991.

[52] P. Mäkilä. Robust identification and galois sequences. *International Journal of Control*, 54(5), 1991.

[53] P. Mäkilä and J. Partington. Robust identification of strongly stabilizable systems. *IEEE Transactions on Automatic Control*, 37(11), 1992.

[54] P. M. Mäkilä. Worst–case input–output identification. *International Journal of Control*, 56(3), 1992.

[55] P. M. Mäkilä, J. R. Partington, and T. K. Gustafsson. Robust identification. *Proc. IFAC Symp. System Identification*, 1:45–63, 1994.

[56] C. Mazzaro, M. Sznaier, and O. Camps. A model (in)validation approach to gait classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11):1820–1825, 2005.

[57] C. M. Mazzaro and M. Sznaie. Convex necessary and sufficient conditions for frequency domain model (in)validation under sltv structured uncertainty. *IEEE Transactions on Automatic Control*, 49(10):1683–1692, 2004.

[58] C. M. Mazzaro and M. Sznaie. A set-membership approach to blind identification. *43rd IEEE Conference on Decision and Control*, 5:5176–5181, 2004.

[59] M. Milanese. Worst case ℓ_1 identification. In M. Milanese, J. Norton, H. Piet-Lahanier, and E. Walter, editors, *Bounding Approaches to System Identification*. Springer, Politecnico di Torino, Italy, 1994.

[60] M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6):997–1009, 1991.

[61] V. Morariu and O. Camps. Modelling correspondences for multi camera tracking using nonlinear manifold learning and target dynamics. *Proceedings of 2006 IEEE Computer Vision and Pattern Recognition*, pages 545–552.

[62] B. Ninness and G. C. Goodwin. Estimation of model quality. *Automatica*, 31(12):1771–1797, 1995.

[63] A. Onatski and J. Stock. Robust monetary policy under model uncertainty in a small model of the us econnomy. *Macroeconomic Dynamics*, 6:85–110, 2002.

[64] A. Onatski and N. Williams. Modeling model uncertainty. *Journal of the European Economic Association, MIT Press*, 1(5), 2003.

[65] J. O'Rourke and N. I. Badlet. Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:522–536, 1980.

[66] A. Orphanides. Monetary policy evaluation with noisy information. *manuscript, Board of Governors of the Federal Reserve Systems*, 1998.

[67] F. Paganini. Set descriptions of white noise and worst case induced norms. *Proceedings of Conference on Decision and Control*, 1993.

[68] F. Paganini. A set based approach for white noise modeling. *IEEE Transactions on Automatic Control*, 41(10), 1996.

[69] F. Paganini. *Sets and Constraints in the Analysis of Uncertain Systems*. Ph.D Thesis at California Institute of Technology, Pasadena, CA., 1996.

[70] P. Parrilo, Sznaier M., and R. Sánchez-Peña. Mixed time/frequency based robust identification. *Preceedings of Conference on Decision and Control, Kobe, Japan*, 1996.

[71] P. Parrilo, R.f Sánchez Peña, and M. Sznaier. A parametric extension of mixed time/frequency robust identification. *IEEE Transactions on Automatic Control*, 44(2):364–369, 1999.

[72] J. R. Partington. Robust identification and interpolation in $\mathcal{H}_\infty$. *International Journal of Control*, 54:1281–1290, 1991.

[73] J. R. Partington. Robust identification in $\mathcal{H}_\infty$. *Journal of Mathematical Analysis and Applications*, 166(22):428–441, 1992.

[74] J. R. Partington. Interpolation identification, and sampling. In *London Mathematical Society Monographs, New Series 17*. Oxford University Press, Inc., 1997.

[75] R. Raich, G. T. Zhou, and M. Viberg. Subspace based approaches for wiener system identification. *IEEE Transactions on Automatic Control*, 50:1629–1634, 2005.

[76] S. Rangan and K. Poolla. Time–domain validation for sample–data uncertainty models. *IEEE Transactions on Automatic Control*, 41:980–991, 1996.

[77] J.M. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. *International Conference on Computer Vision*, pages 612–617, 1995.

[78] K. Rohr. Towards model-based recognition of human movements in image sequence. *Computer Vision and Graphic Image Processing Conference*, 59:94–115, 1994.

[79] H. Rotstein. A nevanlinna–pick approach to time–domain constrained $\mathcal{H}_\infty$ control. *Journal of Control and Optim.*, 34(4), 1996.

[80] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[81] G. Rudebusch. Is the fed too timid? monetary policy in an uncertain world. *Reviews of Economics and Statistics*, 83:203–217, 2001.

[82] G. Rudebusch and Svensson L. Policy rules for inflation targeting. In J. B. Taylor, editor, *Monetary Policy Rules*, pages 203–246. University of Chicago Press, Chicago, 1999.

[83] A. Saberi, A. A. Stoorvogel, P. Sannuti, and H. Niemann. Fundamental problems in fault detection and identification. *International Journal of Robust and Nonlinear Control*, 10(14):1209–1236, 2000.

[84] R. Sánchez-Peña and C. Galarza. Mixed time/frequency based robust identification. *IEEE Transactions on Control Systems Technology*, 2(1), 1994.

[85] R. Sánchez-Peña and Sznaier M. Robust identification with mixed time/frequency experiments: consistency and interpolation algorithms. *Preceedings of Conference on Decision and Control, New Orleans*, 1996.

[86] R. Sánchez-Peña and M. Sznaier. *Robust Systems Theory and Applications*. John Wiley & Sons, Inc., 1998.

[87] J. T. Sargent. Comment. In J. B. Taylor, editor, *Monetary Policy Rules*, pages 144–154. University of Chicago Press, Chicago, 1999.

[88] J. T. Sargent and L. Hansen. Wanting robustness in macroeconomics robust control and filtering of forward looking models. *Unpublished manuscript*, 2000.

[89] D. K. Shim and M. Sznaier. A caratheodory-fejer approach to simultaneous fault detection and isolation. *Proceedings of 2003 American Control Conference*, pages 2979–2984, 2003.

[90] H. Sidenbladh, M.J. Black, and D.J. Fleet. Towards model-based recognition of human movements in image sequence. *European Conference on Computer Vision*, pages 702–718, 2000.

[91] R. Smith and J. C. Doyle. Towards a methodology for robust parameter identification. *Proceedings of 1990 American Control Conference, Seattle, Washington, USA*, pages 2394–2399, 1990.

[92] R. Smith and G. Dullerud. Continuous–time control model validation using finite experimental data. *IEEE Transactions on Automatic Control*, 41:1094–1105, 1996.

[93] R. S. Smith and J. C. Doyle. Model validation: A connection between robust control and identification. *IEEE Transactions on Automatic Control*, 37(7):942–952, 1992.

[94] T. Söderström and P. Stoica. *System Identification*. Prentince–Hall, Englewood Cliffs, NJ, 1989.

[95] J. Stoustrup and H. Niemann. Optimal threshold functions for fault detection and isolation. *Proceedings of 2003 American Control Conference*, pages 1782–1787, 2003.

[96] M. Sznaier and O. Camps. Robust identification of periodic systems with applications to texture inpainting. *44th IEEE Conference on Decision and Control, Seville, Spain.*, pages 609–614, 2005.

[97] R. Tempo, E. W. Bai, and F. Dabbene. Probabilistic robustness analysis: Explicit bounds for the minimum number of sampling points. *Proceedings of IEEE Conference on Decision and Control*, pages 3418–3423, 1996.

[98] O. Toker and J. Chen. Time domain validation of structured uncertainty model sets. *35th IEEE Conference on Decision and Control. Kobe, Japan*, pages 255–260, 1996.

[99] O. Toker and J. Chen. On computational complexity of invalidating structured uncertainty models. *Systems and Control Letters*, 33(3):199–207, 1998.

[100] A. Tornell. Robust-$\mathcal{H}_\infty$. *Online Papers from UCLA Department of Economics, NBER 237*, pages 3418–3423.

[101] J. Traub, G. Wasilkowski, and H. Woźniakowski. *Information–Based Complexity.* Academic Press, Inc., 1988.

[102] P. Van Overschee and B. De Moor. Subspace algorithms for the stochastic identification problem. *Automatica*, 29(3):649–660, 1993.

[103] M. Verhaegen. Identification of the deterministic part of mimo state space models given in innovation form from input–output data. *Special Issue on Statistical Signal Processing and Control of Automatica*, 30(1):61–74, 1993.

[104] M. Verhaegen. Subspace model identification. part iii: Analysis of the ordinary output–error state space model identificaiton algorithm. *International Journal of Control*, 58(3):555–586, 1993.

[105] M. Verhaegen. Application of a subspace model identification technique to identify lti sysstems operating in closed loop. *Automatica*, 29(4):1027–1040, 1994.

[106] B. Wahlberg. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *IEEE Transactions on Automatic Control*, AC–36(5):551–562, 1991.

[107] A Weinmann. *Uncertain Models and Robust Control*. Springer–Verlag, Wien, 1991.

[108] D. Westwick and M. Verhaegen. Identifying mimo wiener systems using subspace model identification methods. *Signal Proces*, 52:235–258, 1996.

[109] V. Wieland. Monetary policy, parameter uncertainty and optimal learning. *Journal of Monetary Economics*, 46:199–228, 2000.

[110] T. Wigren. Recursive prediction error identification methods. *Automatica*, 29(4):1011–1025, 1993.

[111] T. Wigren. Convergence analysis of recursive identification algorithms based on the nonlinear wiener model. *IEEE Transactions on Automatic Control.*, 39(11):2191–2206, 1994.

[112] M. Zhong, S. X. Ding, J. Lam, and H. Wang. An lmi approach to design robust fault detection filters for uncertain lti systems. *Automatica*, 39:543–550, 2003.

[113] K. Zhou, J. Doyle, and K. Glover. *Robust and Optimal Control*. Prentince–Hall, Englewood Cliffs, NJ, 1996.

[114] T. Zhou. Unfalsified probability estimation for a model set based on frequency domain data. *International Journal of Control*, 73(5), 2000.

[115] T. Zhou and H. Kimura. Time domain identification for robust control. *Systems and Control Letters*, 20(3):167–178, 1993.

[116] Y. C. Zhu. Parametric wiener model identification for control. *Proceedings of 14th IFAC World Congress*, 1999.

[117] L. Zi-Qiang. Controller design oriented model identification method for hammerstein system. *Automatica*, 29:767–771, 1993.

Vita

Wenjing Ma was born in Tongling, Anhui Province, P.R. China, in July 1982. She received the B.S. degree with honors in Automation from Northeastern University, in Shenyang, Liaoning Province, China in July 2002. In August 2002 she was enrolled in the Ph.D program in Electrical Engineering at the Pennsylvania State University. Since August 2003, she has worked as a Research Assistant in the Robust Systems Lab under the direction of Dr. Mario Sznaier. Her research interests include semi-blind robust identification and model (in)validation and the applications on Macro-Economic Modeling and Forecast, Fault Detection and Isolation, and Human Motion Modeling and Tracking.