



Identification of nonlinear interconnected systems

Eleni Pepona

Department of Mathematical Sciences

School of Information Systems, Computing and Mathematics

Brunel University

July 2009

Abstract

In this work we address the problem of identifying a discrete-time nonlinear system composed of a linear dynamical system connected to a static nonlinear component. We use linear fractional representation to provide a unified framework for the identification of two classes of such systems.

The first class consists of discrete-time systems consists of a linear time invariant system connected to a continuous nonlinear static component. The identification problem of estimating the unknown parameters of the linear system and simultaneously fitting a m -th order spline to the nonlinear data is addressed. A simple and tractable algorithm based on the separable least squares method is proposed for estimating the parameters of the linear and the nonlinear components. We also provide a sufficient condition on data for consistency of the identification algorithm. Numerical examples illustrate the performance of the algorithm.

Further, we examine a second class of systems that may involve a nonlinear static element of a more complex structure. The nonlinearity may not be continuous and is approximated by piecewise affine maps defined on different convex polyhedra, which are defined by linear combinations of lagged inputs and outputs. An iterative identification procedure is proposed, which alternates the estimation of the linear and the nonlinear subsystems. Standard identification techniques are applied to the linear subsystem, whereas recently developed piecewise affine system identification techniques are employed for the estimation of the nonlinear component. Numerical examples show that the proposed procedure is able to successfully profit from the knowledge of the interconnection structure, in comparison with a direct black box identification of the piecewise affine system.

As you set out for Ithaka
hope the voyage is a long one
full of adventure, full of discovery.
Laistrygonians and Cyclops,
wild Poseidon, you won't encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.
Keep Ithaka always in your mind.
Arriving there is your ultimate goal.
But do not hurry the journey at all.
Better if it lasts for years;
and to anchor at the island when you are old,
rich with all you have gained on the way,
not expecting that Ithaka will offer you riches.
Ithaka has given you the beautiful journey.
Without her you would not have set out on the road.
She has nothing more to give you.
And if you find her poor, Ithaka has not deceived you.
Wise as you have become, so full of experience,
you must already have understood what Ithakas mean.

Ithaka by Constantine P. Cavafy (1911)

Acknowledgements

On the way to my Ithaca I stumbled across a Marie Curie Early Stage Researcher Training fellowship offered under the NET-ACE project (MEST-CT-2004-6724). For that I am grateful. I was meant to acquire advanced knowledge of engineering mathematics and the skills to become a successful researcher. I got so much more; a deep understanding of myself I would not have gained otherwise.

I would like to thank the people that accompanied me in this journey. Prof. Geoff Rodgers thank you for giving me the opportunity to be a member of the NET-ACE team and for approving all the expenses that occurred in the way; research placements, a handful of books and a couple of conferences! My supervisor, Dr. Paresh Date, I truly thank for the faith he showed in me through all the way and for introducing me to the world of system identification. Your support kept me going for as long as I could honestly take. My apologies for failing your expectations in completing my doctoral studies. Prof. Andrea Garulli and Dr. Simone Paoletti, my dear colleagues from Italy, I thank for the support they provided on my visits to beautiful Siena; our short but fruitful meetings were key points on the advancement of our work. Special thanks to Simone for all the hard work in programming the simulation examples of Chapter 5.

Last but surely not least, I would like to thank my husband, Stathis, for keeping the light on even in the darkest of times.

Contents

1	Introduction	1
1.1	System Identification	1
1.2	Nonlinear Identification Problem	3
1.3	Thesis Outline	9
2	Preliminaries	11
2.1	Hilbert Spaces and Orthogonal Projections	11
2.2	Nonlinear Optimization Techniques	14
2.3	Least Squares Estimation	17
2.3.1	Linear Least Squares	18
2.3.2	Nonlinear Least Squares	20
2.4	Curve Fitting	21
3	Linear Fractional Transformations	24
3.1	LFT Definitions	24
3.2	Applications of LFTs	26
3.2.1	State Space Realizations, Transfer Functions and LFTs	27
3.2.2	Frequency Transformation	27
3.2.3	State Space Parametric Uncertainty	27
3.2.4	Modeling Interconnected Systems	29

4	A Separable Least Squares approach to LFT Identification	31
4.1	The Separable Least Squares Method	32
4.2	Problem Formulation	35
4.2.1	Model Structure	35
4.2.2	Identifiability	37
4.2.3	Identification Criterion	38
4.2.4	Consistency	39
4.3	Parameter estimation algorithm	43
4.3.1	Algorithm Initialization	44
4.4	Numerical Examples	46
4.5	Summary	50
5	Piecewise Affine Identification	52
5.1	PWA identification of nonlinear systems : An overview	53
5.1.1	A bounded-error procedure for PWA identification	54
5.2	Problem Formulation	55
5.2.1	Model Structure	55
5.2.2	Identifiability	58
5.3	Parameter estimation algorithm	61
5.4	Numerical examples	64
5.5	Summary	71
6	Contributions and Future Research	72
6.1	Outline of Contributions	72
6.2	Future Research	73

List of Figures

1.1	The system identification loop	2
1.2	General form of serially interconnected systems	6
1.3	The considered LFT model structure.	8
3.1	(Left) Lower LFT, (Right) Upper LFT	26
3.2	The mass/spring/damper system	28
3.3	The Hammerstein system (on the left) modeled under the LFT framework (on the right).	30
3.4	An interconnected system (on the left) modeled under the LFT framework (on the right).	30
4.1	Approximation of the nonlinear map by a cubic spline in example 4.2.	48
4.2	Samples (\hat{z}_k, \hat{w}_k) at initialization in example 4.3.	49
4.3	(left) \hat{z}_k, \hat{w}_k for updated values of $\hat{\vartheta}_n$ at the last iteration of the L-M algorithm (right) Final fit for updated values of $\hat{\vartheta}_l$ in example 4.3.	49
4.4	Approximation of the nonlinear map by a cubic spline in example 4.4.	51
5.1	First algorithm iteration in Example 5.1: (left) Circles represent the “noisy” signal v_k^1 , while the dashed lines represent a PWA map with two modes that is fitted to the samples (v_k^1, u_k) ; (right) Circles represent the corresponding “smoothed” signal w_k^1	65

5.2	Sequence of estimates of the ARX coefficients $-a_1$, b_1 and g_1 in Example 5.2 (dotted lines are true values).	67
5.3	Example 5.3: (Top left) Samples (v_k^1, \mathbf{r}_k) at the first iteration. (Top right) Number of modes versus the bound δ when a PWA map is fitted to the samples (v_k^1, \mathbf{r}_k) . (Bottom) The same at the third iteration, i.e. for the samples (v_k^3, \mathbf{r}_k)	70
5.4	Sequence of estimates of the FIR coefficients g_1 , g_2 , g_3 , and g_4 in Example 5.3 (dotted lines are true values).	71

List of Tables

4.1	Identification algorithm	45
5.1	Iterative identification algorithm	63
5.2	True and estimated parameters in Example 5.2	66
5.3	Comparison of identification results in Example 5.2	68
5.4	True and estimated parameters in Example 5.3	69

Notation

Symbols

\mathbb{R}	the field of real numbers
\mathbb{R}_+	the field of positive or zero real numbers
\mathbb{C}	the field of complex numbers
\in	belongs to
\equiv	equal by definition
\approx	approximately equal
$A \subseteq B$	A is a subset or equal to B
$x \in \mathbb{R}$	real number x
$x \in \mathbb{R}^n$	n -dimensional real column vector
x_i	the i -th element of vector $x \in \mathbb{R}^n$
0_n	n -dimensional column vector with all elements equal to zero
A^\top	transpose of matrix A
A^\dagger	the pseudo-inverse of matrix A
$diag(a_1, \dots, a_n)$	diagonal matrix $n \times n$ with elements a_1, \dots, a_n in its main diagonal
$rank(A)$	the rank of matrix A
\hat{a}	an estimate of the parameter a
$\inf_{x \in X} f(x)$	infimum of f in X
$Dom(f)$	the domain of function f

$Ker(f)$	kernel of f
f^{-1}	inverse mapping of a bijective mapping f
f'	the derivative of function f
∇f	gradient of a function f
$ x $	absolute value of $x \in \mathbb{R}$
$\ x\ $	Euclidean norm of $x \in \mathbb{R}^n$
X^\perp	orthogonal complement of the vector space X

Acronyms

ARX	Autoregressive Exogeneous
FIR	Finite Impulse Response
LFT	Linear Fractional Transformation
LSE	Least Squares Estimate
LTI	Linear Time Invariant
max, min	point wise maximum or minimum element
arg max, arg min	maximizing or minimizing argument
MIMO	Multi-Input Multi-Output
MISO	Multi-Input Single-Output
MIN PFS	Minimum Partition into Feasible Subsystems
PWA	Piecewise Affine
SISO	Single-Input Single-Output
SLS	Separable Least Squares

Chapter 1

Introduction

System identification is the area of science that deals with the construction of mathematical models for dynamical systems based on observed data from the system. Linear system identification, the area dealing with the identification of linear systems, has been well established for many decades. The inability of the linear system structure to describe the dynamic behaviour of many real systems, however, has driven the research effort in the area of nonlinear system identification. In this chapter we review the aim of system identification and the nonlinear identification problem and we provide a brief outline of this thesis.

1.1 System Identification

Given a set of input-output data, the aim of system identification is to fit a model belonging to a specified model set by optimizing a suitable identification criterion. More analytically, the system identification procedure constitutes of the following four components [55]. The system identification loop can be seen in Figure 1.1.

1. An experiment that will provide the *data record*, the data to be used for the identification of the system. In *experiment design* one can choose when to measure the data, what data to measure and may also choose the input signal. The ultimate goal is thus

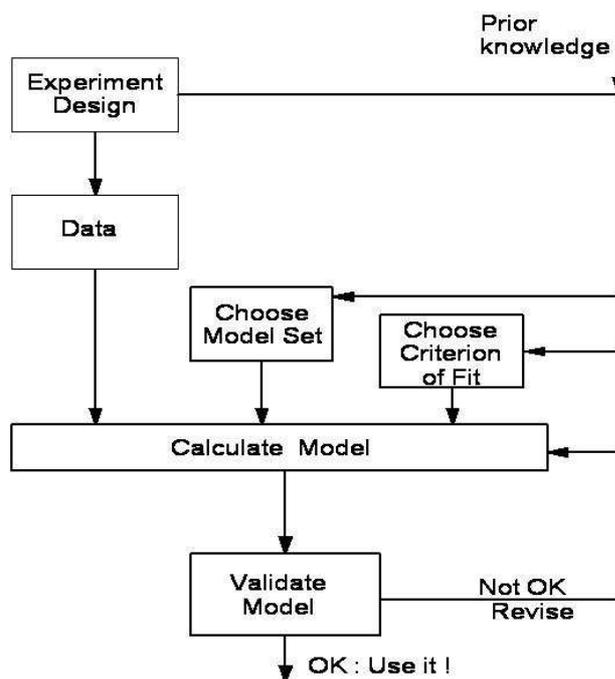


Figure 1.1: The system identification loop

to provide a set of data as informative as possible about the system.

2. A set of candidate models, often called *model structure*. This can be the most challenging part of system identification and most often the choice of the model structure is determined according to *a priori* knowledge of the system structure.
3. The *identification method*, which is the method by which the model structure is fitted to the data. The objective here is to choose the 'best' possible model from the model set according to a specified criterion of fit.
4. The *model validation*. Once a model has been selected from the model set one needs to check whether the model is able to explain any other set of data from the same system.

The problem of identification of linear systems has been exhaustively studied, see [55] for a review. In this thesis, we focus on the problem of identifying nonlinear systems with a specific structure. This problem is discussed in the following section.

1.2 Nonlinear Identification Problem

Up to date most of the nonlinear identification methods proposed in the literature deal with *black-box* input-output models. Black-box identification techniques are applied to dynamical systems for which no *a priory* information as to the structure of the system is available. In such settings there are two measurable sequences available for system identification, the input signal sequence $U_1^N = (u_1, u_2, \dots, u_N)^\top$ and the output signal sequence $Y_1^N = (y_1, y_2, \dots, y_N)^\top$. One then needs to establish a relationship, between the so-called *regressor vector*

$$z_k = z(u_k, u_{k-1}, \dots, u_{k-n_u}, y_{k-1}, y_{k-2}, \dots, y_{k-n_y})^\top \quad (1.1)$$

and the outputs $\{y_k\}_{k=1}^N$, of the form

$$y_k = f(z_k, e_k) \quad (1.2)$$

where f is a nonlinear map which needs to be identified and $e = \{e_k\}_{k=1}^N$ is a sequence of disturbance terms, probably of a known class but not measurable. In this thesis we will study Single Input Single Output (SISO) systems only, that is, u, y and $e \in \mathbb{R}$. An overview of regressor and nonlinear mapping possibilities can be found in [78].

One of the first approaches to be studied was the basis expansions, where the nonlinear mapping f is parameterized as

$$f(z) = \sum_k \alpha_k f_k(z),$$

f_k being *basis functions* of a known structure, thus reducing to a *linear-in-the-parameters* model to be identified. Existence theorems of such expansions are given in [75]. Many alternatives as to the nature of the basis functions have been proposed. The most usual technique is the use of Volterra models [28, 76]. An M^{th} -order Volterra model in discrete

time is given by

$$y_k = \sum_{m=0}^M \sum_{k_1=0}^{\infty} \dots \sum_{k_m=0}^{\infty} h_m(k_1, k_2, \dots, k_m) \prod_{j=1}^m u_{k-k_m},$$

where u is the input, y the output and h_m the m^{th} -order kernel. Recent works employ Laguerre functions for the optimal estimation of the Volterra kernels, see [16, 50] and references therein. Other possibilities include the use of splines or wavelets [22, 82].

According to another approach, a linear (affine) estimator \hat{f} of the nonlinear map f in equation (1.2) is considered to be of the form

$$\hat{f}(z^*) = w_0 + \sum_{k=1}^N w_k z_k,$$

at a given point z^* , where z_k is given by (1.1). The problem is then to estimate the weights w_0 and $\{w_k\}_{k=1}^N$. As mentioned in [73] different methods following this formulation have been proposed. These can be classified into *kernel methods*, *local polynomial modelling*, *least squares support vector machines* and *direct weight optimization*. For a wide review of these methods the interested reader is referred to [73] and references therein.

Recent methods have moved a step further to deal with PieceWise Affine systems (PWA) where the nonlinearity f is considered to be a piecewise affine map. A few notable results are included in the works of [10, 32, 58]. A more comprehensive overview on this topic will follow in Chapter 5.

Further, novel techniques for the estimation of the nonlinearity in equation (1.2) is the use of Neural Networks (NN) [23, 47] and fuzzy models. There is a variety of methods developed in this area in the last decade, from stochastic, dynamic or multilayered NN [49, 67, 80] to Fuzzy NN (FNN) [11, 34], Wavelet NN (WNN) or combinations of these, as in [79].

As mentioned above, all the aforementioned methods are applicable in black-box identification, when no information on the nature of the nonlinearity is available. It is however the case in many applications that the mapping f consists of a linear and a purely nonlinear

component, thus resulting in a nonlinear map overall, that is

$$f(z) = Az + g(z), \tag{1.3}$$

where A a constant matrix of appropriate dimensions and g the purely nonlinear map. Such systems are called *partially linear systems* [29,52]. In such systems it is of great interest to be able to separate the linear from the nonlinear components and identify them separately. Recently a method has been proposed for inferring nonlinear dependence from measured data [52]. Current approaches towards the identification of partially linear systems include the use of support vector machines proposed in [29].

A broader class of systems, which includes those of (1.3), consists of systems formed by interconnected linear and nonlinear components. The most simple example of such an interconnection is the serial one, of Figure 1.2 for example. Depending on the nature of the mappings f_1 , f_2 and f_3 different types of systems are derived. Among these, the Hammerstein system, the Wiener system and their combinations have received great attention over the past few decades.

Let \mathcal{N} denote a static nonlinear map, \mathcal{L} a Linear Time Invariant (LTI) system and I the identity operator. A Hammerstein system consists of a static nonlinear element followed by a linear system, that is $f_1 \equiv \mathcal{N}$, $f_2 \equiv \mathcal{L}$ and $f_3 \equiv I$, whereas the Wiener system is the other way round, a linear dynamical system followed by a static nonlinear map, i.e. $f_1 \equiv \mathcal{L}$, $f_2 \equiv \mathcal{N}$ and $f_3 \equiv I$. Combinations of these have also been studied, the Wiener-Hammerstein model which is a nonlinear element sandwiched by two linear systems ($f_1 \equiv \mathcal{L}_1$, $f_2 \equiv \mathcal{N}$ and $f_3 \equiv \mathcal{L}_2$) and the Hammerstein-Wiener model in which two static nonlinearities surround a linear dynamical system ($f_1 \equiv \mathcal{N}_1$, $f_2 \equiv \mathcal{L}$, $f_3 \equiv \mathcal{N}_2$). The schemes developed for the identification of Hammerstein or Wiener systems usually apply to Wiener-Hammerstein systems as well. This is not the case however for Hammerstein-Wiener systems, mainly because both the input and output signals to the linearity are unobservable. Therefore, Hammerstein-

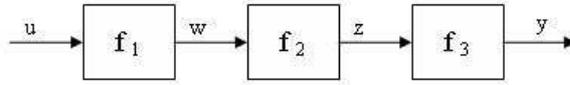


Figure 1.2: General form of serially interconnected systems

Wiener systems are studied separately in the literature. The identification methods proposed for Hammerstein, Wiener or Wiener-Hammerstein systems can be classified into two broad categories, *parametric* and *nonparametric*.

In *parametric methods* the nonlinearity is supposed to be of a known structure, usually polynomial, and the problem consists in estimating the unknown parameters by minimizing some loss function of the data. The path in this direction was set by Narendra and Gallman in their original work [59] where an iterative algorithm was proposed. A noniterative algorithm appeared in [18, 42]. Several parametric schemes have been proposed since. A variation of the [59] algorithm can be found in [70]. A recursive method based on weighted extended least squares was proposed in [12], whereas a different approach using Multilayer Feedforward Neural Networks(MFNN) was discussed in [1]. Recent works include the use of Least Squares Support Vector Machines(LS-SVM) [38] and expansion of the nonlinear map in Fourier series [5]. An ARMAX Hammerstein model was considered in [27] where the regressor vector in (1.1) depends on past values of the disturbance sequence $\{e_k\}_{k=1}^N$ as well, that is

$$z_k = z(u_k, u_{k-1}, \dots, u_{k-n_u}, y_{k-1}, \dots, y_{k-n_y}, e_{k-1}, \dots, e_{k-n_e}).$$

In many studies, identification of Hammerstein or Wiener systems with special structure of the nonlinearity is also considered. In [37] the static nonlinearity is modelled as a piecewise affine map whereas in [19, 20] as piecewise linear. Hard input nonlinearities, like saturation, preload, dead-zone etc. are studied in [6].

In *nonparametric methods* the structure of the nonlinear map is considered unknown

and its estimation is attempted by using approximating smooth functions. Methods for the solution of the identification problem employ kernel regression [41], adaptive orthogonal series [63], best approximation theory concepts [90], polynomial identification techniques [91], wavelet theory concepts [43], linearization techniques [74] or probabilistic methods known as stochastic approximation techniques [40, 83].

Exceptions to the above classification are feasible. For example, a combined parametric-nonparametric algorithm has been proposed in [44]. In another front, identification of Hammerstein systems from a set-membership (worst-case scenario) point of view is studied in [9, 30, 35].

Finally, a few methods for the identification of Hammerstein-Wiener models have been developed. Parametric methods for such systems usually employ the Least Squares technique in order to minimize some loss function [3, 89]. Further, a blind approach is proposed in [4], which allows the recovery of all internal signals based solely on the input and output signal measurements. Nonparametric techniques are also available, for example by representing the system's output in Volterra series as in [81].

A unified framework for representing interconnected systems is based on the Linear Fractional Transformation (LFT) of Figure 1.3. The block \mathcal{L} describes the overall linear dynamics and can be decomposed as

$$\mathcal{L} \sim \begin{bmatrix} \mathcal{L}_y \\ \mathcal{L}_z \end{bmatrix} = \begin{bmatrix} \mathcal{L}_{yu} & \mathcal{L}_{yw} & \mathcal{L}_{ye} \\ \mathcal{L}_{zu} & \mathcal{L}_{zw} & \mathcal{L}_{ze} \end{bmatrix}, \quad (1.4)$$

where \mathcal{L}_{yu} , \mathcal{L}_{yw} and \mathcal{L}_{ye} are the transfer functions from u , w and e to the system output y , respectively. Similarly, \mathcal{L}_{zu} , \mathcal{L}_{zw} , \mathcal{L}_{ze} are the transfer functions from u , w and e to the input of the nonlinear part, z , respectively. The block \mathcal{N} describes the static nonlinear element. A wide class of dynamical systems can be described in this framework. Apart from the Hammerstein and Wiener systems mentioned above, systems which consist of serial connections between multiple linear systems and multiple static nonlinearities fall under this

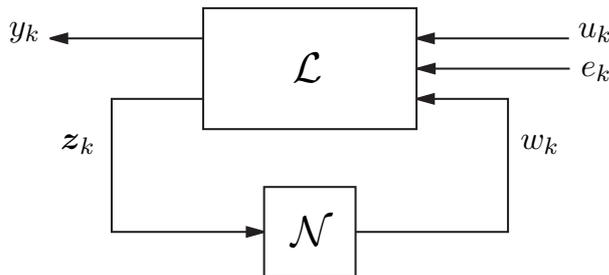


Figure 1.3: The considered LFT model structure.

category, as do partially linear systems described in [29]. The most important representatives of the serially interconnected structures are the Wiener-Hammerstein system considered in [12] and the Hammerstein-Wiener system examined in [89] and [4]. However, more complex interconnected structures are also encountered in practice.

It is apparent that prior structural knowledge and a set of input-output data $\{u_k, y_k\}_{k=1}^N$ do not provide enough information for the identification of both \mathcal{L} and \mathcal{N} . One would need additional information about the internal signals of the LFT system to this end. We can distinguish the following situations:

1. Both z_k and w_k are known.
2. w_k is known, while z_k is unknown.
3. z_k is known, while w_k is unknown.

Identifying the system under the case (1) is trivial. One can employ linear system identification methods for the linear part \mathcal{L} and any nonlinear approximation or curve fitting method to the nonlinear element \mathcal{N} . In case (2), only the linear part $[\mathcal{L}_{yu} \ \mathcal{L}_{ye} \ \mathcal{L}_{yw}]$ of \mathcal{L} can be identified. Thus, we argue that case (3) is the most general situation worth of investigation if one wants to identify both \mathcal{L} and \mathcal{N} .

Although the LFT model structure is a powerful representation of a wide class of nonlinear systems, relatively little work has been done on the identification of the LFT model.

In [68] the authors set up a parameter estimation procedure for LFT models where the forward part is represented by a classical linear regression and the feedback part is given by a nonlinear dynamic map parameterized by a neural network. In [45] the authors propose a nonparametric estimation algorithm for the identification of the static nonlinear map in the interconnected system. Early results on this work are also reported in [24].

All these works assume complete knowledge of the linear component \mathcal{L} . In the work reported in this thesis, we drop this assumption. We assume that the components of the linear dynamical system have a known and specific structure. Thus the identification of the linear part reduces to the estimation of the corresponding parameters. The nonlinear component is thought to be a nonlinear static map, either continuous or piecewise continuous. Two identification techniques for such systems are presented in this work in the subsequent chapters.

1.3 Thesis Outline

In this section we give a brief outline of the thesis.

In Chapter 1 we gave an introduction to the topic of nonlinear system identification and we review current trends and techniques established for the identification of nonlinear systems.

In Chapter 2 we present a selection of preliminary topics, which the reader should be familiar with before proceeding to the study of the work that follows.

Chapter 3 is dedicated to the introduction of the Linear Fractional Transformation. We provide a historic account on the subject, a review of the areas where LFTs are applied and a reasoning for the advantages in using LFTs to model complex interconnected systems.

In Chapter 4 we present a new separable least squares identification technique for a class of nonlinear systems with static nonlinearities, assuming a continuous parametrization of the nonlinear map. To support the main work, we provide an introduction on the subjects

of separable least squares and curve fitting by use of m^{th} order splines.

In Chapter 5 we review briefly the area of piecewise affine identification and propose a piecewise affine approach to identifying a class of systems similar to that of Chapter 4, with the difference that the nonlinear map need not be continuous over polyhedra of lagged inputs and outputs and only needs to be piecewise continuous.

In Chapter 6 we summarize the work presented in this thesis and the major conclusions reached. Open problems and suggestions for future research are also discussed here.

I hereby declare that the work presented in this thesis is a result of my own research apart from the following exceptions:

- The mathematical formulation of the identification algorithm presented in Table 5.1 of Chapter 5 as well as the examples presented in Section 5.4 are a result of joint work with Dr. Simone Paoletti, Assistant Professor, Department of Information Engineering, University of Siena, Siena, Italy.
- The programming of the identification algorithm presented in Table 5.1 of Chapter 5 has been delivered by Dr. Simone Paoletti.

Chapter 2

Preliminaries

In this chapter we present some preliminary theory and results necessary for the understanding of the work presented in the following chapters.

2.1 Hilbert Spaces and Orthogonal Projections

We start with a review of some useful definitions in functional analysis. The material presented here is based on [26] and [36]. The interested reader is referred to the above references for details.

Definition 2.1 Let X be a non-empty set, $X \subseteq \mathbb{R}$. The mapping $d : X \times X \rightarrow \mathbb{R}$, is called a *metric* in X (or a *distance function*) if for any x, y and $z \in X$ the following hold:

1. $d(x, y) \geq 0$ and the equality holds if and only if $x = y$,
2. $d(x, y) = d(y, x)$,
3. $d(x, y) \leq d(x, z) + d(z, y)$.

The pair (X, d) is called a *metric space* and the number $d(x, y)$ is called the *distance* of x from y .

Definition 2.2 Let X be a vector space, $X \subseteq \mathbb{R}^n$. The mapping $\|\cdot\| : X \rightarrow \mathbb{R}$, is called a *norm* in X if for any $x, y \in X$ and $\lambda \in \mathbb{R}$ the following hold:

1. $\|x\| \geq 0$ and the equality holds if and only if $x = 0_n$,
2. $\|\lambda x\| = |\lambda|\|x\|$,
3. $\|x + y\| \leq \|x\| + \|y\|$ (“*triangle inequality*”).

If $\|\cdot\|$ is a norm on the vector space X , then $d(x, y) = \|x - y\|$ is the distance of x from y on X , for all $x, y \in X$. In the following, unless stated otherwise, by $\|x\|$, for any $x \in \mathbb{R}^n$, we will mean the Euclidean norm given by

$$\|x\| = \sqrt{x_1^2 + \dots + x_n^2}. \quad (2.1)$$

A *normed space* is a vector space X with a given norm on X . Such a space is always considered as a metric space for the distance $\|x - y\|$. A metric space M is called *complete* if every Cauchy sequence in M converges in M . A *Banach* space is a normed space which is complete.

Definition 2.3 Let X be a vector space, $X \subseteq \mathbb{R}^n$. The mapping $\langle \cdot, \cdot \rangle : X \times X \rightarrow \mathbb{R}$, is called an *inner product* on X if for any x, y and $z \in X$ and $\lambda, \mu \in \mathbb{R}$ the following hold:

1. $\langle x, x \rangle \geq 0$ and the equality holds if and only if $x = 0_n$,
2. $\langle x, y \rangle = \langle y, x \rangle$,
3. $\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$.

Two vectors $x, y \in X$ are called *orthogonal*, denoted by $x \perp y$, if $\langle x, y \rangle = 0$. Note that the zero element 0_n is orthogonal to every element $x \in X$ and is the only element in X with this property. A vector z is orthogonal to the subspace M of X , if for all $y \in M$ it holds $\langle z, y \rangle = 0$. We write $z \perp M$.

On a vector space X with inner product as above, the mapping $\|\cdot\| : X \rightarrow \mathbb{R}$ given by $\|x\| = \sqrt{\langle x, x \rangle}$ is a norm. This norm is called the norm induced on X by the inner product. A vector space X equipped with the norm induced by the inner product is called a *Hilbert* space if it is complete.

Let H be a Hilbert space, M a closed subspace of H and $x \in H$. There is a unique $y_0 \in M$ such that

$$\|x - y_0\| = d(x, M) = \inf\{\|x - y\| : y \in M\}. \quad (2.2)$$

This unique element $y_0 \in M$ is denoted by $P_M(x)$ and is called the *projection* of x on M . It holds that $x - P_M(x) \perp M$.

Definition 2.4 Let H be a Hilbert space and $A \subseteq H$, $A \neq \emptyset$. The *orthogonal complement* of A , denoted by A^\perp , is defined as

$$A^\perp = \{x \in H : \forall a \in A, \langle x, a \rangle = 0\}. \quad (2.3)$$

A^\perp is a closed subspace of H .

The mapping $x \rightarrow P_M(x)$ of H onto M is linear, continuous and of norm 1 if $M \neq \{0\}$. Its kernel $\text{Ker}(P_M) = P_M^{-1}(0) = M^\perp$ is the subspace orthogonal to M . The linear mapping P_M is called the *orthogonal projection* of H in M and its kernel M^\perp is called the *orthogonal supplement* of M in H . Finally, H is the topological direct sum of M and M^\perp , as stated in the following theorem.

Theorem 2.1 Let H be a Hilbert space and M a closed subspace of H . Then $H = M \oplus M^\perp$, that is, for every $x \in H$ there are unique $x_1 \in M$ and $x_2 \in M^\perp$ such that

$$x = x_1 + x_2.$$

The idea of orthogonal projections will be useful later in Chapter 4.

2.2 Nonlinear Optimization Techniques

In this section we discuss the unconstrained optimization problem [13]

$$\text{minimize } f(x) \tag{2.4}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex and twice differentiable mapping. We assume that there is a solution x^* to the minimization problem (2.4). Since f is differentiable and convex, a point x^* is optimal if and only if

$$\nabla f(x^*) = 0. \tag{2.5}$$

Thus, solving the unconstrained minimization problem (2.4) is equivalent to finding a solution to the equation (2.5), which is an set of n equations with n unknowns, the n variables x_1, x_2, \dots, x_n . The algorithms presented here produce a minimizing sequence $x^j, j = 1, \dots$, given by

$$x^{j+1} = x^j + t^j \Delta x^j, \tag{2.6}$$

where $t^j \in \mathbb{R}$ with $t^j \geq 0$ (the equality holds when x^j is optimal) is called the *step size* at iteration j and $\Delta x \in \mathbb{R}^n$ is called the *step* or *search direction*. The methods presented here are all *descent methods*, meaning that

$$f(x^{j+1}) < f(x^j), \tag{2.7}$$

unless x^j is optimal. To initialize the algorithms we require a suitable starting point $x^0 \in \text{Dom}(f)$. Moreover, we require that the sublevel set S defined below is closed.

$$S = \{x \in \text{Dom}(f) | f(x) \leq f(x^0)\}. \tag{2.8}$$

Since from convexity of f we get that $f(y) \geq f(x^j)$, whenever $\nabla f(x^j)^\top \Delta x^j \geq 0$, the search direction in a descent method must satisfy

$$\nabla f(x^j)^\top \Delta x^j < 0. \quad (2.9)$$

The general descent method algorithm is as follows.

Algorithm 2.1 *General descent method*

Given a starting point $x \in \text{Dom}(f)$.

Repeat

1. Determine a descent direction Δx .
2. *Line search.* Chose a step size $t > 0$.
3. *Update,* $x := x + t\Delta x$.

Until stopping criterion is satisfied.

At the first step of Algorithm 2.1 we determine the descent direction Δx . Next, we perform the *line search*. Here we need to select where along the line $\{x + t\Delta x | t \in \mathbb{R}_+\}$ the next iterate will be. That is, we need to select the step size $t \in \mathbb{R}_+$. This selection can be performed with either the exact line search or an inexact line search.

Following the *exact line search*, t is chosen as the minimizing argument of f along the line $\{x + t\Delta x | t \in \mathbb{R}_+\}$, i.e.

$$t = \arg \min_{s \geq 0} f(x + s\Delta x). \quad (2.10)$$

As noted in [13], this method is used when the cost of the minimization problem with one variable, required in (2.10), is low compared to the cost of computing the search direction itself.

The exact line search is rarely used in practice. Instead *inexact line search* methods are used. In these methods the step length is chosen to *approximately* minimize f along the line $\{x + t\Delta x | t \in \mathbb{R}_+\}$. The method most frequently used in this category is the *backtracking*

line search, see [13].

We have discussed different ways of performing the line search for the second step of Algorithm 2.1. Now we will present the most widely used variations for selecting the descent direction, Δx , in the first step of Algorithm 2.1.

The Gradient Direction. The most simple and natural choice is to choose the so called *gradient direction*

$$\Delta x := -\nabla f(x). \quad (2.11)$$

The Steepest Descent Direction. The *normalized steepest descent direction* is given by

$$\Delta x = \arg \min\{\nabla f(x)^\top u \mid \|u\| = 1\}, \quad (2.12)$$

where, $\nabla f(x)^\top u$ is the directional derivative of f at x in the direction of u and $\|\cdot\|$ is any norm on \mathbb{R}^n . since the first-order Taylor approximation of $f(x+u)$ around x is given by $f(x+u) \approx f(x) + \nabla f(x)^\top u$, the normalized steepest descent is a step of unit norm that gives the largest decrease in the linear approximation of f , [13].

For the Euclidean norm, $\|z\|_2 = (z^\top z)^{1/2}$, the steepest descent direction is $\Delta x = -\nabla f(x)$, which is the gradient descent direction.

For a quadratic norm, $\|z\|_P = (z^\top P z)^{1/2} = \|P^{1/2} z\|_2$, where P is a symmetric, positive definite, $n \times n$ matrix, the normalized steepest descent direction is given by

$$\Delta x = -(\nabla f(x)^\top P^{-1} \nabla f(x))^{-1/2} P^{-1} \nabla f(x).$$

Newton's Direction and Newton's Method. The *Newton step* is given by

$$\Delta x = -\nabla^2 f(x)^{-1} \nabla f(x),$$

for $x \in \text{Dom}(f)$. Newton's method for unconstrained minimization of f is a descent method that uses the Newton step as descent direction and backtracking line search for choosing the step size t . A slight modification to the general descent method of Algorithm 2.1 is that in Newton's method, the stopping criterion is checked after computing the search direction, rather than after the update. Newton's method is summarized in Algorithm 2.2.

Algorithm 2.2 *Newton's method*

Given a starting point $x \in \text{Dom}(f)$ and tolerance $\epsilon > 0$.

Repeat

1. Compute the Newton step $\Delta x = -\nabla^2 f(x)^{-1} \nabla f(x)$ and the decrement $\lambda^2 := \nabla f(x)^\top \nabla^2 f(x)^{-1} \nabla f(x)$.
 2. *Stopping criterion.* **Stop** if $\lambda^2/2 \leq \epsilon$.
 3. *Line search.* Choose step size t by backtracking line search.
 4. *Update,* $x := x + t\Delta x$.
-

We have presented a brief summary of nonlinear minimization techniques for the minimization problem (2.4). In the next section we discuss the least squares method for curve fitting of both linear and non-linear functions.

2.3 Least Squares Estimation

One of the most celebrated techniques in regression analysis is that of the least squares method. Provided that information is available in the form of measurements of some *explaining* variables or *regressors* x_1, x_2, \dots, x_n , regression analysis aims to find an estimate of the *dependent* variable y , connected to \bar{x} 's through the model

$$\hat{y} = g(\bar{x}; \vartheta)$$

such that the difference $\varepsilon = y - \hat{y}$ becomes as small as possible, where $\bar{x} = [x_1 \ \dots \ x_n]^\top \in \mathbb{R}^n$, $y \in \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is a parametric function of the regressors and $\vartheta \in \mathbb{R}^n$ is the parameter vector. In dynamical systems it is often the case that y is the system output, \bar{x} is the system input, g is the mathematical model describing the input-output relationship of the system and ϑ the parameter vector to be estimated for the identification of the system. Given a set of N pairs of measurements (y_k, \bar{x}_k) , $k = 1, \dots, N$, the Least Squares Estimate (LSE) is the minimizing argument of the following quadratic cost function,

$$V_N(\vartheta) = \frac{1}{N} \sum_{k=1}^N (y_k - g(\bar{x}_k))^2 = \frac{1}{N} \sum_{k=1}^N \varepsilon_k^2 = \frac{1}{N} \varepsilon^\top \varepsilon, \quad (2.13)$$

that is,

$$\hat{\vartheta}_N = \arg \min_{\vartheta} V_N(\vartheta). \quad (2.14)$$

As remarked in [55], the parameter $\hat{\vartheta}_N$ is that value of ϑ that gives the best predictor $\hat{y} = g(\bar{x})$, when applied to the data \bar{x}_k , $k = 1, \dots, N$ irrespectively of whether we have imposed a stochastic framework on the problem or not.

2.3.1 Linear Least Squares

The vector function g can take any form. A special case, widely studied in theory and with extensive practical applications, is when we impose a linear parametrization as follows,

$$g(\bar{x}) = \vartheta_1 x_1 + \vartheta_2 x_2 + \dots + \vartheta_n x_n, \quad (2.15a)$$

$$= \vartheta^\top \bar{x}. \quad (2.15b)$$

Thus, we fit y to a linear combination of the regressors x_1, x_2, \dots, x_n .

Remark 2.1 In the following chapters we deal with affine functions. An affine function has

the form

$$g(\bar{x}) = \vartheta_{n+1} + \vartheta^\top \bar{x}. \quad (2.16)$$

Under this framework we can still use the Linear Least Squares formulation by defining the extended parameter vector $\bar{\vartheta} = [\vartheta^\top \ \vartheta_{n+1}]^\top \in \mathbb{R}^{n+1}$ and the extended regression vector $z = [\bar{x} \ 1]^\top \in \mathbb{R}^{n+1}$, thus obtaining

$$g(\bar{x}) = \bar{\vartheta}^\top z. \quad (2.17)$$

The cost function $V_N(\vartheta)$ in (2.13) now takes the form

$$V_N(\vartheta) = \frac{1}{N} \sum_{k=1}^N (y_k - \vartheta^\top \bar{x}_k)^2, \quad (2.18)$$

which is quadratic in ϑ . Thus, it can be minimized analytically (see [56]), yielding a *global* minimum $\hat{\vartheta}_N$ that satisfies the following *normal equations*

$$\left[\frac{1}{N} \sum_{k=1}^N \bar{x}_k \bar{x}_k^\top \right] \hat{\vartheta}_N = \frac{1}{N} \sum_{k=1}^N \bar{x}_k y_k. \quad (2.19)$$

If the matrix on the left-hand side of (2.19) is invertible, the LSE is given by

$$\hat{\vartheta}_N = \left[\frac{1}{N} \sum_{k=1}^N \bar{x}_k \bar{x}_k^\top \right]^{-1} \frac{1}{N} \sum_{k=1}^N \bar{x}_k y_k. \quad (2.20)$$

Given the LSE $\hat{\vartheta}_N$, the *predicted value* of y , \hat{y} is given by

$$\hat{y} = \hat{\vartheta}_N^\top \bar{x}. \quad (2.21)$$

The regression *residuals* $\hat{\varepsilon}$ are the values of the *prediction error* $\varepsilon = y - \vartheta^\top \bar{x}$, associated to the LSE $\hat{\vartheta}_N$, that is

$$\hat{\varepsilon} = y - \hat{\vartheta}_N^\top \bar{x}. \quad (2.22)$$

It can be shown (see [56]) that the sum of squared observations of y can be split into the sum of squared predictions \hat{y} and the sum of squared residuals,

$$\sum_{k=1}^N y_k^2 = \sum_{k=1}^N \hat{y}_k^2 + \sum_{k=1}^N \hat{\epsilon}_k^2. \quad (2.23)$$

We say, that the linear regression is successful when the model (2.15) explains the data y , i.e. the predicted outputs \hat{y} account for the major part of the observed (actual) outputs y in (2.23). Therefore a measure of the model's fit to the data is provided by the *squared correlation coefficient*,

$$R^2 = \frac{\sum_{k=1}^N \hat{y}_k^2}{\sum_{k=1}^N y_k^2} = 1 - \frac{\sum_{k=1}^N \hat{\epsilon}_k^2}{\sum_{k=1}^N \hat{y}_k^2}. \quad (2.24)$$

The closer the value of R^2 to 1 the better the fit of the regression model to the data $\{y_k, \bar{x}_k\}_{k=1}^N$.

2.3.2 Nonlinear Least Squares

Returning to the general form of the minimization problem (2.13)-(2.14), for a general nonlinear function g , we can employ any minimization technique to compute the LSE. According to the general descent method described in Algorithm 2.1 we can use the following iterative minimization scheme:

$$\hat{\vartheta}_N^{j+1} = \hat{\vartheta}_N^j + t_j \Delta \vartheta_j, \quad (2.25)$$

where t_j is the step size and $\Delta \vartheta_j$ is the descent direction at iteration $j = 1, \dots$. The descent direction can be written as

$$\Delta \vartheta^j = R_j^{-1} \nabla V_N(\hat{\vartheta}_N^j), \quad (2.26)$$

where the matrix R_j can be chosen in different ways according to which method one wishes to employ. Usual choices are the following.

- *Gradient Direction*. For $R_j = -I$ we get the gradient direction.

- *Gauss-Newton Direction.* Take $R_j = \frac{1}{N} \varepsilon'^T \varepsilon'$. Then the step direction can be computed as the solution to the following linear quadratic minimization problem

$$\Delta \vartheta_j = \arg \min_{\Delta \vartheta_j} \|\varepsilon' \Delta \vartheta_j - \varepsilon\| = (\varepsilon'^T \varepsilon')^{-1} \varepsilon'^T \varepsilon = \varepsilon'^{\dagger} \varepsilon. \quad (2.27)$$

In the above last equation we have used the notation $z^{\dagger} = (z^T z)^{-1} z^T$ to denote the *pseudo-inverse* of $z \in \mathbb{R}^n$.

- *Levenberg-Marquardt Direction.* Use $R_j = \frac{1}{N} \varepsilon'^T \varepsilon' + \delta_j I$, where δ_j is used instead of the step size t_j . Similarly to the Gauss-Newton direction, the Levenberg-Marquardt direction is computed as the solution to the minimization problem

$$\Delta \vartheta_j = \arg \min_{\Delta \vartheta_j} \|\varepsilon' \Delta \vartheta_j - \varepsilon\| + \delta_j \|\Delta \vartheta_j\| = \begin{bmatrix} \varepsilon' \\ \delta_j I \end{bmatrix}^{\dagger} \begin{bmatrix} \varepsilon^T & 0 & \dots & 0 \end{bmatrix}^T. \quad (2.28)$$

This is the proposed descent direction used later in Chapter 4 for the nonlinear least squares problem encountered in the system identification procedure.

2.4 Curve Fitting

This section is a brief introduction to curve fitting. It is not aimed to be exhaustive and the interested reader is referred to [13] and [51] for a thorough treatment of the topic. The material presented here is based on [13].

In curve fitting problems we are given a set of data and based on certain *a priori* knowledge about the type of function that has generated them we need to choose the member of a finite-dimensional subspace of functions that best fits the data under certain requirements.

Consider a family of functions $f_1, \dots, f_k : \mathbb{R}^n \rightarrow \mathbb{R}$ with common domain $Dom(f_i) = D$, for all $i = 1, \dots, k$. This family of functions is called the *basis functions* of f , where for each

$u \in \mathbb{R}^k$ the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined by

$$f(x) = u_1 f_1(x) + \dots + u_k f_k(x), \quad (2.29)$$

with $Dom(f) = D$. The basis functions generate a subspace F of functions on D . The vector $u \in \mathbb{R}^k$ is called the *coefficient vector* and it is the variable we need to optimize for best fit of the function to the data. According to the level of *a priori* knowledge about the system that generated the data, we might be able to select a member of the family of functions that truly generated the data. In case with little prior knowledge however we have to use more generic families of functions. The families of functions most widely used in curve fitting problems are described below.

Polynomials: A subspace of functions on \mathbb{R} is the set of polynomials of degree less than k . There are many basis functions for this subspace. The simplest one is the family that consists of the power functions $f_i(t) = t^{i-1}$, $i = 1, \dots, k$.

Piecewise-affine functions: First we form a triangulation of the function domain D , i.e. a set grid points $g_1, \dots, g_k \in \mathbb{R}^n$ and a partition of D into a set of simplexes D_1, \dots, D_k such that

$$D = D_1 \cup \dots \cup D_k, \quad int(D_i \cap D_j) = \emptyset \quad \text{for } i \neq j,$$

where $int(A)$ denotes the interior of the set A . Then a piecewise affine function f is constructed by assigning values $f(g_i) = x_i$ to the grid points and then extending the function affinely on each simplex. The basis functions are affine on each simplex and defined by the conditions

$$f_i(g_j) = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Such functions are continuous by construction.

Piecewise polynomials and splines: Piecewise polynomials are defined as polynomials of some maximum degree on each simplex of the triangulation. Piecewise polynomial functions are continuous, that is the polynomials must agree at the boundaries between the simplexes. Spline functions are piecewise polynomial functions that satisfy the further restrictions of having continuous derivatives of up to a certain order, which is the order of the spline function. A choice of basis functions for an m -th order spline is the following

$$\begin{aligned} f_i(x) &= |x - g_i|^m, \quad i = 1, \dots, k - 2 \\ f_{k-1}(x) &= x, \\ f_k(x) &= 1. \end{aligned}$$

Thus, for any given $u \in \mathbb{R}^k$ a spline function of order m is given by

$$f(x) = \sum_{j=1}^{k-2} u_j |x - g_j|^m + u_{k-1}x + u_k.$$

In Chapter 4 we will use cubic splines of this form, i.e. spline functions of third order, to approximate a continuous nonlinear static map.

Chapter 3

Linear Fractional Transformations

In this chapter we present the Linear Fractional Transformation (LFT) and we provide a brief review of its applications in control engineering. Further, we discuss how we can represent complex interconnected structures under the LFT framework and the advantages of such representations. Most of the material presented in Sections 3.1 and 3.2 are based on [48] and [88].

3.1 LFT Definitions

The linear fractional transformation is a term appearing in different branches of physical sciences. In scalar complex analysis we find the following definitions.

Definition 3.1 The one-complex-variable function $F : \mathbb{C} \rightarrow \mathbb{C}$ of the form

$$F(s) = \frac{a + bs}{c + ds} \tag{3.1}$$

with a, b, c and $d \in \mathbb{C}$ is called a *linear fractional transformation*.

If in particular, $c \neq 0$ then $F(s)$ can be written as

$$F(s) = \alpha + \beta s(1 - \gamma s)^{-1},$$

for some α, β and $\gamma \in \mathbb{C}$.

In multivariable complex analysis the linear fractional transformation is defined as follows.

Definition 3.2 Let M be a complex matrix partitioned as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \in \mathbb{C}^{(p_1+p_2) \times (q_1+q_2)},$$

and let $\Delta_l \in \mathbb{C}^{q_2 \times p_2}$ and $\Delta_u \in \mathbb{C}^{q_1 \times p_1}$ be two other complex matrices. The map

$$\mathcal{F}_l(M, \cdot) : \mathbb{C}^{q_2 \times p_2} \rightarrow \mathbb{C}^{q_1 \times p_1}$$

with

$$\mathcal{F}_l(M, \Delta_l) := M_{11} + M_{12}\Delta_l(I - M_{22}\Delta_l)^{-1}M_{21}, \quad (3.2)$$

provided that the inverse $(I - M_{22}\Delta_l)^{-1}$ exists, is called the *lower linear fractional transformation with respect to Δ_l* . The *upper linear fractional transformation with respect to Δ_u* is similarly defined as the map

$$\mathcal{F}_u(M, \cdot) : \mathbb{C}^{q_1 \times p_1} \rightarrow \mathbb{C}^{q_2 \times p_2}$$

with

$$\mathcal{F}_u(M, \Delta_u) := M_{22} + M_{21}\Delta_u(I - M_{11}\Delta_u)^{-1}M_{12}, \quad (3.3)$$

provided that the inverse $(I - M_{11}\Delta_u)^{-1}$ exists.

In Figure 3.1 we see the graphical representation of the lower (left diagram) and the upper (right diagram) LFTs. It is easy to see that the diagram on the left represents the following

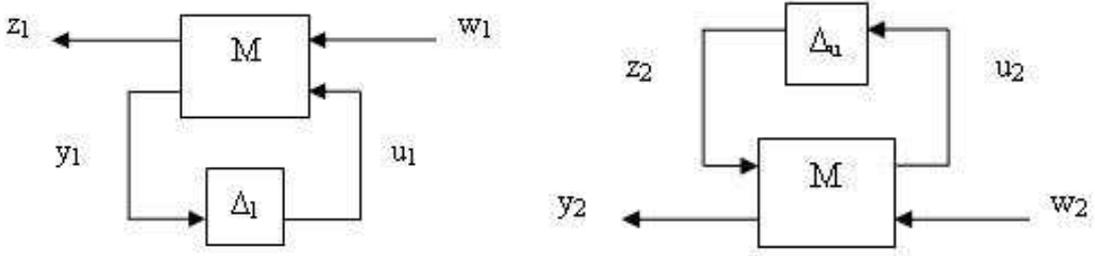


Figure 3.1: (Left) Lower LFT, (Right) Upper LFT

set of equations

$$\begin{bmatrix} z_1 \\ y_1 \end{bmatrix} = M \begin{bmatrix} w_1 \\ u_1 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ u_1 \end{bmatrix}, \quad (3.4a)$$

$$u_1 = \Delta_l y_1 \quad (3.4b)$$

and similarly for the diagram on the right the following hold,

$$\begin{bmatrix} y_2 \\ z_2 \end{bmatrix} = M \begin{bmatrix} u_2 \\ w_2 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} u_2 \\ w_2 \end{bmatrix}, \quad (3.5a)$$

$$u_2 = \Delta_u y_2. \quad (3.5b)$$

3.2 Applications of LFTs

The linear fractional transformations (LFTs) provide a powerful tool for many control engineering applications. Here we review the most important uses of LFTs and we give a brief overview of current research in each field.

3.2.1 State Space Realizations, Transfer Functions and LFTs

The LFT can describe the state space realization of a discrete time system. Let the system be described in state space form by

$$\begin{bmatrix} x_{k+1} \\ y_k \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_k \\ u_k \end{bmatrix} = M \begin{bmatrix} x_k \\ u_k \end{bmatrix}. \quad (3.6)$$

The transfer function is then given by

$$G(z) = D + C(zI - A)^{-1}B = \mathcal{F}_u \left(\begin{bmatrix} A & B \\ C & D \end{bmatrix}, \frac{1}{z}I \right). \quad (3.7)$$

A generalization of the notions of minimality, reachability and observability for parameter dependent multivariable systems modeled by LFTs is presented in [8].

3.2.2 Frequency Transformation

The bilinear transformation between the z -domain and the s -domain is $s = \frac{z+1}{z-1}$, giving

$$\frac{1}{s}I = I - \sqrt{2}Iz^{-1}I(I + z^{-1}I)^{-1}\sqrt{2}I = \mathcal{F}_u(N, z^{-1}I), \quad (3.8)$$

where $N = \begin{bmatrix} I & \sqrt{2}I \\ -\sqrt{2}I & -I \end{bmatrix}$.

3.2.3 State Space Parametric Uncertainty

In many applications there is uncertainty in the parameters of a state space model. The most celebrated example (see [48]) is the idealized mass/spring/damper system shown in Figure 3.2, where F is the applied external force, m is the mass of the block, k is the spring

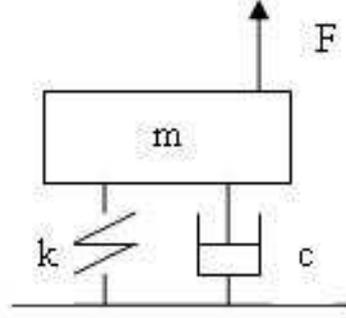


Figure 3.2: The mass/spring/damper system

constant and c is the damping coefficient. The system is described by the equation

$$y'' + \frac{c}{m}y' + \frac{k}{m} = \frac{F}{m}, \quad (3.9)$$

where y is the vertical displacement of the mass. Suppose that the parameters m , c and k are fixed but uncertain, with $m = \tilde{m}(1 + w_m\delta_m)$, $c = \tilde{c}(1 + w_c\delta_c)$ and $k = \tilde{k}(1 + w_k\delta_k)$. Let $x = [x_1 \ x_2]^\top$, where $x_1 = y$ and $x_2 = my'$. Then the above differential equation can be written as

$$\begin{bmatrix} x' \\ y \end{bmatrix} = \mathcal{F}_l(M, \Delta) \begin{bmatrix} x \\ F \end{bmatrix},$$

where,

$$\Delta = \text{diag}(\delta_m, \delta_k, \delta_c) \quad \text{and} \quad M = \begin{bmatrix} 0 & \tilde{m}^{-1} & 0 & -w_m & 0 & 0 \\ -\tilde{k} & -\tilde{c} & 1 & 0 & w_k & w_c \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \tilde{m}^{-1} & 0 & -w_m & 0 & 0 \\ -\tilde{k} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\tilde{c} & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Most of the works found in the literature on LFTs relate to models for uncertain para-

metric dynamical systems. State-order reduction methods for uncertain systems represented by LFTs are discussed in [15] and [33]. Assuming polynomial parametric uncertainty, the authors in [25] propose a method for obtaining low-order LFTs by exploiting the structure of the uncertainty. The authors in [21] examine the problem of model validation for uncertain systems modeled by LFTs in both the cases of structured and un-structured nonlinearities, by using either time- or frequency-domain data. Their results are extended in the more recent work of [87] where the authors discuss new validation approaches under which the simultaneous use of time- and frequency-domain data is possible. In [7] the concept of minimality in uncertain and multidimensional systems modeled by LFTs is related to realization theory results for formal power series. An approach employing parameter-dependent Lyapunov functions for gain-schedule design in LFT systems is discussed in [86].

3.2.4 Modeling Interconnected Systems

The LFT model structure of Figure 1.3 can be used, as discussed in Section 1.1, to model a variety of interconnected systems that consist of linear and nonlinear components.

The Hammerstein system is a serial interconnection of a nonlinear static map, \mathcal{N}_1 , followed by a Linear Time Invariant (LTI) system, \mathcal{L}_1 , as shown on the left diagram of Figure 3.3. It is one of the simplest interconnected systems but with numerous practical applications. The right diagram of Figure 3.3 shows the Hammerstein system in LFT form. The linear system \mathcal{L} has been partitioned as in Equation (1.4), whereas for the static nonlinear map it holds $\mathcal{N} \equiv \mathcal{N}_1$.

More complex interconnected systems can also be modeled under the LFT framework. For example, the system on the left diagram of Figure 3.4, which consists of several linear and nonlinear components connected in serial and parallel configurations, can be modeled under the LFT framework as shown on the right diagram of Figure 3.4. Here, the nonlinear map \mathcal{N} is a block diagonal matrix function, with each nonlinear component on the main diagonal.

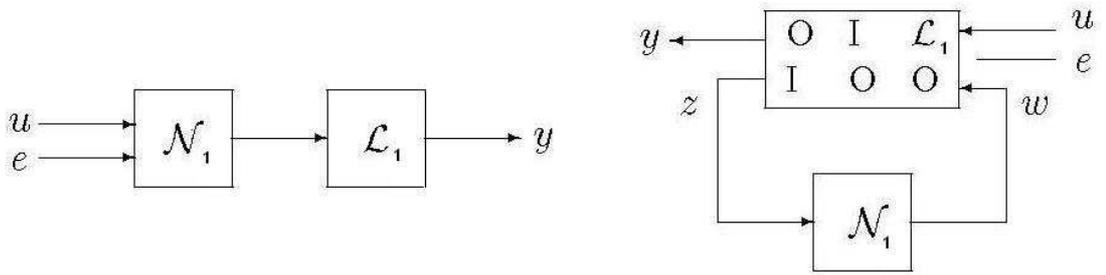


Figure 3.3: The Hammerstein system (on the left) modeled under the LFT framework (on the right).

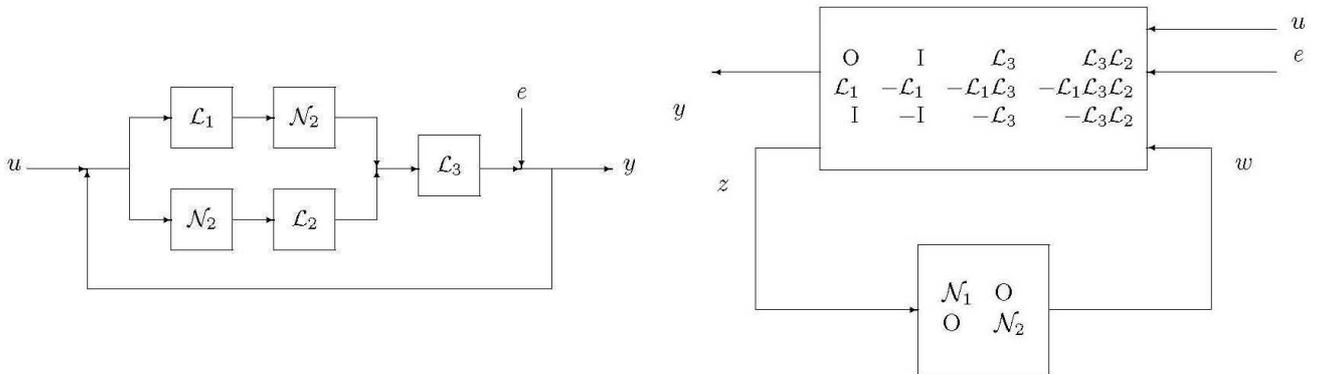


Figure 3.4: An interconnected system (on the left) modeled under the LFT framework (on the right).

Further, the LFT model structure has the property that any serial, parallel or feedback interconnection of two or more LFTs results in a new LFT. This property makes LFTs a versatile tool in modeling complex interconnected structures.

As mentioned in Section 1.1, little work has been done in the area of LFT identification. In Chapters 4 and 5 we discuss the identification problems in the case of LTI linear systems, \mathcal{L} , and continuous or piecewise continuous static nonlinearities \mathcal{N} and we propose two identification techniques. Contributions and future research directions are summarized in Chapter 6.

Chapter 4

A Separable Least Squares approach to LFT Identification

In this chapter, we will consider a subset of LFT systems of Figure 1.3, with \mathcal{L}_y modelled with an AutoRegressive with eXogeneous input (ARX) structure and \mathcal{L}_{zu} modelled with a Finite Impulse Response (FIR) structure. \mathcal{L}_{ze} and \mathcal{L}_{zw} are both assumed to be zero. This subset of systems includes Hammerstein systems, Wiener systems, Wiener-Hammerstein systems and Hammerstein-Wiener systems as special cases. For simplicity of notation, we consider the case where all signals in Figure 1.3 are scalar.

The nonlinear static map will be approximated by a smooth m^{th} -order spline [51]. Splines have long been used for the approximation of static nonlinearities. They are usually considered to provide a much better fit than polynomials as the local action of their coefficients makes them more flexible and adjustable. Approximation of the nonlinear static map by a spline has been proposed in the literature in the following cases. In [46] the authors use splines to approximate the nonlinear part of a Wiener system, whereas approximation using cardinal splines for Hammerstein systems is introduced in [17] and the identification of Hammerstein-Wiener systems under cubic spline representation is discussed in [89]. The framework proposed in this thesis generalises the works of [17], [46] and [89]. In addition,

a major contribution of our approach is that it provides a new and tractable algorithm for identifying Wiener-Hammerstein systems using a spline approximation of the nonlinearity. To the author's knowledge, the literature on spline approximation has not dealt with Wiener-Hammerstein type systems previously.

For the identification of the system model under consideration, we will show that the model is linear in the parameters of the \mathcal{L}_{yu} and \mathcal{L}_{ye} subsystems and the parameters of the spline map, while it is nonlinear in the parameters of the systems \mathcal{L}_{zu} and \mathcal{L}_{yw} . Consequently, a separable least squares approach is taken for the identification of the system under a prediction error (PE) framework. The variable projection method of [39] is employed to reduce the original identification problem to an equivalent problem, where the parameters of \mathcal{L}_{zu} and \mathcal{L}_{yw} are optimized first in a separated prediction error cost minimization and the optimal parameters of \mathcal{L}_{yu} and \mathcal{L}_{ye} are then derived *a posteriori*. For the solution of the equivalent nonlinear parameter minimization problem we use the Levenberg-Marquardt algorithm described in [55], [77] and [84].

The chapter is organized as follows. In Section 4.1 we introduce some preliminary results from the separable least squares optimization theory. The identification problem under consideration is presented in Section 4.2. In Section 4.3 we describe the proposed identification algorithm and propose a scheme for its initialization. Numerical examples illustrating the performance of the algorithm are reported in Section 4.4. Conclusions and possible future research directions are summarized in Section 4.5.

Parts of the work presented in this chapter can be found in [64].

4.1 The Separable Least Squares Method

The Separable Least Squares (SLS) method can be applied to a class of nonlinear minimization problems whose variables separate. Consider a discrete nonlinear system \mathcal{N} , modeled

by

$$\mathcal{N} : y_k = f(\vartheta, \alpha; u_k) = \sum_{j=1}^p \vartheta_j \phi_j(\alpha; u_k), \quad (4.1)$$

where, y_k and u_k , $k \in \mathbb{Z}$, are the system's output and input signals, $\vartheta \in \mathbb{R}^p$, $\alpha \in \mathbb{R}^q$ are parameters that enter the model linearly and non-linearly, respectively, and ϕ_j , $j = 1, \dots, p$ are functions continuously differentiable in α . Note that parameters ϑ and α form two completely disjoint sets. Given a set of data (u_k, y_k) , $k = 1, \dots, N$, the nonlinear least squares optimization method consists in estimating the unknown parameters ϑ and α by minimizing the nonlinear functional

$$r(\vartheta, \alpha) = \|y - f(\vartheta, \alpha; u)\|^2 = \sum_{k=1}^N \left(y_k - \sum_{j=1}^p \vartheta_j \phi_j(\alpha; u_k) \right)^2, \quad (4.2)$$

i.e. solving the nonlinear optimization problem

$$(\hat{\vartheta}, \hat{\alpha}) = \min_{\vartheta, \alpha} r(\vartheta, \alpha) = \min_{\vartheta, \alpha} \|y - f(\vartheta, \alpha; u)\|^2. \quad (4.3)$$

Let $\Phi = \{\Phi\}_{k,j} = \phi_j(\alpha; u_k)$, $k = 1, \dots, N$, $j = 1, \dots, p$. Then, the functional defined in (4.2) can be written as

$$r(\vartheta, \alpha) = \|y - \Phi(\alpha)\vartheta\|^2. \quad (4.4)$$

In order to be able to find a critical point for the functional in (4.4) we need to make the following assumption.

Assumption 4.1 The matrix $\Phi(\alpha)$ has constant rank $\text{rank}(\Phi) \leq \min(N, p)$ for all $\alpha \in \Omega \subseteq \mathbb{R}^q$, with Ω an open set that contains the desired solution.

Let $\Phi^\dagger(\alpha)$ denote the Moore-Penrose generalized inverse of matrix $\Phi(\alpha)$. For each α , the linear operator $P_{\Phi(\alpha)} = \Phi(\alpha)\Phi^\dagger(\alpha)$ is the orthogonal projection on the linear space spanned by the columns of the matrix $\Phi(\alpha)$. Denote by $\Phi^\perp(\alpha) = I - \Phi(\alpha)\Phi^\dagger(\alpha)$, the projector on the orthogonal complement of the column space of $\Phi(\alpha)$. For any given α , the functional in

(4.4) is minimized by the Least Squares Estimate (LSE)

$$\hat{\vartheta}(\alpha) = \min_{\vartheta} \|y - \Phi(\alpha)\vartheta\|^2 \equiv \Phi^\dagger(\alpha)y. \quad (4.5)$$

Thus,

$$\begin{aligned} \min_{\vartheta} r(\vartheta, \alpha) &= r(\alpha) = \|y - \Phi(\alpha)\Phi^\dagger(\alpha)y\|^2 \\ &= \|(I - P_{\Phi(\alpha)})y\|^2 \\ &= \|P_{\Phi(\alpha)}^\perp y\|^2 \end{aligned} \quad (4.6)$$

Hence, instead of solving the minimization problem (4.3) with respect to two parameter vectors, one can minimize first the modified functional $r_2(\alpha)$, given by

$$r_2(\alpha) = \|P_{\Phi(\alpha)}^\perp y\|^2, \quad (4.7)$$

with respect to α , i.e.

$$\hat{\alpha} = \min_{\alpha} r_2(\alpha) = \min_{\alpha} \|P_{\Phi(\alpha)}^\perp y\|^2. \quad (4.8)$$

Once a critical point or a minimizer $\hat{\alpha}$ of (4.7) has been found, $\hat{\vartheta}$ is obtained by replacing $\hat{\alpha}$ in (4.5). This procedure is justified by the following theorem, stated below without proof. The interested reader can refer to [39].

Theorem 4.1 *Let $r(\vartheta, \alpha)$ and $r_2(\alpha)$ be defined as above. We assume that in the open set $\Omega \subset \mathbb{R}^q$, $\Phi(\alpha)$ has constant rank $r_\Phi \leq \min(N, p)$.*

(1) *If $\hat{\alpha}$ is a critical point (or global minimizer in Ω) of $r_2(\alpha)$ and*

$$\hat{\vartheta} = \Phi^\dagger(\hat{\alpha})y, \quad (4.9)$$

then $(\hat{\vartheta}, \hat{\alpha})$ is a critical point (or global minimizer in Ω) of $r(\vartheta, \alpha)$ and $r(\vartheta, \alpha) = r_2(\alpha)$.

(2) *If $(\hat{\vartheta}, \hat{\alpha})$ is a global minimizer of $r(\vartheta, \alpha)$ for $\alpha \in \Omega$, then $\hat{\alpha}$ is a global minimizer of $r_2(\alpha)$ in Ω and $r(\vartheta, \alpha) = r_2(\alpha)$. Furthermore, if there is a unique $\hat{\alpha}$ among the minimizing pairs*

of $r(\vartheta, \alpha)$, then $\hat{\alpha}$ must satisfy (4.9).

One of the advantages of the Separable Least Squares method over traditional nonlinear optimization is summarized in the following theorem of Sjöberg and Viberg, which is stated without proof. The interested reader is referred to [77]. Before stating the theorem we need the following two definitions.

Definition 4.1 The *condition number* of matrix A is defined as $\text{cond}(A) = \frac{\max(\sigma(A))}{\min(\sigma(A))}$ where $\sigma(A)$ are the singular values of the matrix A . If $\text{cond}(A)$ is very large, for example bigger than 10^3 , then the matrix A is said to be *ill-conditioned*.

Definition 4.2 Let $\varepsilon = y - \Phi(\alpha)\vartheta$. An *ill-conditioned minimization problem* is a minimization problem for which ε' is ill-conditioned.

Theorem 4.2 *The separated minimization problem (4.5)-(4.8) is better, or at least as good conditioned as the non-separated minimization problem (4.3).*

However, in identification problems the SLS can be proved a powerful tool. Forming the separated minimization problem usually involves a considerable reduction in the dimension of the problem. This point is illustrated in the identification method presented in Section 4.3.

4.2 Problem Formulation

4.2.1 Model Structure

Consider the discrete-time interconnected dynamical system described by the Linear Fractional Transformation (LFT) of the type shown in Figure 1.3. The system consists of a linear dynamical component interconnected with a static nonlinear element. Although we do not require knowledge of the nonlinear map \mathcal{N} , we do assume that it is smooth enough to be approximated by a m^{th} order spline, i.e. $\mathcal{N}(\cdot) \in C^r$ for $r \geq m$. Signals u_k , y_k and e_k are the

system input, output and the unobservable additive noise at time $k \in \mathbb{Z}$ respectively, while z_k and w_k are internal signals representing the input and the output of the nonlinear map. For simplicity of notation we consider the scalar case, i.e. u_k, y_k, e_k, z_k and $w_k \in \mathbb{R}$.

For the identification of the LFT system in Figure 1.3, the linear system \mathcal{L}_y is modeled as an ARX in (4.10a) below. \mathcal{L}_{zu} in (4.10b) is FIR, while the unknown nonlinear map \mathcal{N} is approximated by a spline of order m as in (4.10c) below.

$$A(q)y_k = B(q)u_k + G(q)w_k + e_k, \quad (4.10a)$$

$$z_k = D(q)u_k, \quad (4.10b)$$

$$w_k = \sum_{j=1}^s \lambda_j |z_k - \kappa_j|^m + \lambda_{s+1} z_k + \lambda_0. \quad (4.10c)$$

In (4.10a), $A(q), B(q), G(q)$ are finite polynomials of known orders n_a, n_b and n_g , respectively, in the delay operator q^{-1} , i.e.

$$A(q) = 1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a}, \quad (4.11a)$$

$$B(q) = b_0 + b_1 q^{-1} + \dots + b_{n_b} q^{-n_b}, \quad (4.11b)$$

$$G(q) = g_0 + g_1 q^{-1} + \dots + g_{n_g} q^{-n_g}. \quad (4.11c)$$

Similarly, $D(q)$ in (4.11b) is a finite polynomial of known order n_d in the delay operator q^{-1} ,

$$D(q) = d_0 + d_1 q^{-1} + \dots + d_{n_d} q^{-n_d}. \quad (4.12)$$

In (4.10c), s is the number of knots (assumed to be known and fixed) used to interpolate the nonlinear input-output data $\{z_k, w_k\}_{k=1}^N$ by an m^{th} order spline and $\kappa_j, j = 1, \dots, s$ are the fixed spline knot points, chosen such that $z_{\min} = \kappa_1 < \dots < \kappa_s = z_{\max}$. We will comment on choosing the number and the location of knots in Section 4.3.

In the next section, we examine the identifiability of the considered system model (4.10).

4.2.2 Identifiability

Let $a = [a_1 \cdots a_{n_a}]^\top$, $b = [b_0 \cdots b_{n_b}]^\top$, $g = [g_0 \ g_1 \ \cdots \ g_{n_g}]^\top$, $d = [d_0 \ \cdots \ d_{n_d}]^\top$ and $\lambda = [\lambda_0 \ \lambda_1 \ \cdots \ \lambda_s \ \lambda_{s+1}]^\top$ and define the parameter vector $\vartheta \in \mathbb{R}^n$, with $n = n_a + n_b + n_g + n_d + s + 5$ given by

$$\vartheta = [a^\top \ b^\top \ g^\top \ d^\top \ \lambda^\top]^\top. \quad (4.13)$$

This parametrization of the system model (4.10) is clearly not unique. For any nonzero constant $\xi \in \mathbb{R}$, no identification experiment can distinguish between the pairs of parameters (g, λ) and $(\xi g, \frac{1}{\xi} \lambda)$. To account for this identifiability issue, we impose a normalizing condition that $g_0 = 1$. This is without loss of generality, since we can always suitably scale the nonlinear map, as it is demonstrated in Example 4.1 below. Similarly, for any nonzero $\beta \in \mathbb{R}$ no identification experiment can distinguish between the products $\lambda_j z_k$ and $(\beta \lambda_j)(\frac{1}{\beta} z_k)$, for $j = 1, \dots, s+1$. To this end, we impose another normalizing condition that $\|d\|_2 = 1$. Finally, since the model (4.10) contains a term $(B(q) + \lambda_{s+1} G(q) D(q)) u_k$, we impose a condition that

$$b_i c_i = 0$$

holds for all $i \in [0, n_l]$, where $c_i = \sum_{j=0}^i g_j d_{i-j}$ and $n_l = \min(n_b + 1, n_g + n_d + 1)$.

Let $\acute{Q} \subseteq \mathbb{R}^n$ be the set of vectors ϑ which satisfy constraints $\|d\|_2 = 1$, $g_0 = 1$ and $b_i c_i = 0$, with c_i defined as above.

Example 4.1 Consider the system

$$y_k = a_1 y_{k-1} + b_1 u_{k-1} + g_1 w_{k-1} + e_k, \quad (4.14a)$$

$$z_k = d_1 u_{k-1} + d_2 u_{k-2}, \quad (4.14b)$$

$$w_k = \sin(z_k). \quad (4.14c)$$

Let $\tilde{w}_k = \sin(z_{k-1})$ and $\tilde{g}_0 = g_1$. Then it holds that $g_1 w_{k-1} = \tilde{g}_0 \tilde{w}_k$. If, further, we define

$\tilde{w}_k = \tilde{g}_0 \tilde{w}_k$, then system (4.14) is equivalent to

$$y_k = a_1 y_{k-1} + b_1 u_{k-1} + \tilde{g}_0 \tilde{w}_k + e_k, \quad (4.15a)$$

$$z_k = d_1 u_{k-1} + d_2 u_{k-2}, \quad (4.15b)$$

$$w_k = \sin(z_k), \quad (4.15c)$$

with $\tilde{g}_0 = 1$.

◇

In the next subsection, we consider an identification criterion with the parameter vector constrained to \hat{Q} .

4.2.3 Identification Criterion

Let

$$\phi_k = [|z_k - \kappa_1|^n \cdots |z_k - \kappa_s|^n \ z_k \ 1]^\top, \quad (4.16)$$

then w_k in (4.10c) can be written as

$$w_k = \phi_k^\top \lambda. \quad (4.17)$$

The predictor \hat{y}_k (which also equals the conditional mean of y_k based on data up to time $k - 1$) can be derived by equation (4.10a) as follows:

$$\hat{y}_k = \left(1 - \hat{A}(q)\right) y_k + \hat{B}(q) u_k + (\hat{G}(q) \hat{\phi}_k)^\top \hat{\lambda}, \quad (4.18)$$

where, by \hat{X} we denote an estimate of X .

The parameter vector ϑ given in (4.13) can be decomposed as follows.

$$\vartheta = [\vartheta_l^\top \ \vartheta_{nl}^\top]^\top, \text{ where} \quad (4.19a)$$

$$\vartheta_l = [a^\top \ b^\top \ \lambda^\top]^\top, \quad (4.19b)$$

$$\vartheta_{nl} = [d^\top \ g^\top]^\top. \quad (4.19c)$$

In the above, $\vartheta_l \in \mathbb{R}^p$ with $p = n_a + n_b + s + 3$ and $\vartheta_{nl} \in \mathbb{R}^q$, with $q = n_g + n_d + 1$ (with $g_0 = 1$). Note that the model (4.10) is linear in ϑ_l and nonlinear in ϑ_{nl} .

The identification criterion is formed as follows : Given a set of N input-output data $\{u_k, y_k\}_{k=1}^N$, find the parameter vector ϑ that minimizes the cost function

$$V_N(\vartheta) = \frac{1}{2N} \sum_{k=1}^N (y_k - \hat{y}_k)^2, \quad (4.20)$$

over \hat{Q} as defined in the last subsection, i.e.

$$\hat{\vartheta} = \arg \min_{\vartheta \in \hat{Q}} V_N(\vartheta). \quad (4.21)$$

We will use the results on asymptotic properties of nonlinear least squares estimator in [85] to establish conditions for data for consistency of the estimate in the next subsection.

4.2.4 Consistency

In order to look at conditions for consistency of the estimator in (4.21), we will denote an estimate of X by \hat{X} , as before, and the true value of a quantity X by \tilde{X} . Accordingly, let $\tilde{\vartheta} \in \hat{Q}$ be the true parameter vector. We say that an estimator $\hat{\vartheta}$ based on data $\{u_k, y_k\}_{k=1}^N$ is consistent if $\hat{\vartheta} \rightarrow \tilde{\vartheta}$ as $N \rightarrow \infty$ with probability 1.

Our main tool to impose conditions on data for consistency of estimator is the following

result from [85] on consistency of nonlinear least squares problem:

Theorem 4.3 *Let $y_k = f(x_k, \tilde{\vartheta}) + \epsilon_k$, where $\tilde{\vartheta} \in \Theta$ is the true parameter vector, x_k is the input that gives rise to observation y_k , ϵ_k are independent, identically distributed errors with a zero mean and a non-zero variance and the parameter space Θ is finite. Assume that the functions $f_k(\vartheta) = f(x_k, \vartheta)$ are continuous in ϑ . Then*

$$\sum_{i=1}^N \left(f(x_i, \tilde{\vartheta}) - f(x_i, \hat{\vartheta}) \right)^2 \rightarrow \infty \text{ as } N \rightarrow \infty, \forall \hat{\vartheta} \neq \tilde{\vartheta} \quad (4.22)$$

implies

$$\hat{\vartheta} \rightarrow \tilde{\vartheta} \text{ w.p. 1, as } N \rightarrow \infty. \quad (4.23)$$

As remarked in [85], the finiteness of the parameter space Θ does not impose any severe restriction since in practice we can only search for the minimum over a finite set, e.g. to the sixth decimal place. To check whether the estimator proposed in the last section is consistent, we make an important assumption that $D(q) = 1$ (which implies $b_0 = 0$ due to the constraints on the parameter set \hat{Q}). Even after this restriction on (4.10), the restricted model set is still quite general and includes Hammerstein systems as special case.

We now seek a *sufficient* condition on data for (4.22) to hold (with $u_k = x_k$ and $f(x_k, \tilde{\vartheta}) = y_k$ as obtained from (4.10a)-(4.10c)); which in turn would be sufficient for the estimator $\hat{\vartheta}$ defined as above to be consistent. For this purpose, we need to define several new quantities related to data as well as to the error in the parameter estimate. Let $p = n_a + n_b + n_g + 1$

and for $t_k > p$, let

$$\begin{aligned}\mathcal{F}_k &= \begin{bmatrix} \mathcal{Y}_k^\top & \mathcal{U}_k^\top & \mathcal{H}_k^\top \end{bmatrix}^\top, \text{ where} \\ \mathcal{Y}_k &= \begin{bmatrix} y_{k-1} & y_{k-2} & \cdots & y_{k-n_a} \end{bmatrix}^\top, \\ \mathcal{U}_k &= \begin{bmatrix} u_{k-1} & u_{k-2} & \cdots & u_{k-n_b} \end{bmatrix}^\top, \\ \mathcal{H}_k &= \begin{bmatrix} \phi_k^\top & \phi_{k-1}^\top & \cdots & \phi_{k-n_g}^\top \end{bmatrix}^\top,\end{aligned}$$

and where ϕ_k is as defined in (4.16). Note that $\mathcal{F}_k \in \mathbb{R}^{n_a+n_b+(n_g+1)(s+2)}$. Next, define two vectors of errors introduced by error in the parameter estimate:

$$\begin{aligned}\Xi &= \begin{bmatrix} (\tilde{a} - \hat{a})^\top & (\tilde{b} - \hat{b})^\top \end{bmatrix}^\top \text{ and} \\ \Psi &= \begin{bmatrix} (\tilde{\lambda} - \hat{\lambda})^\top & (\tilde{g}_1 \tilde{\lambda} - \hat{g}_1 \hat{\lambda})^\top & \cdots & (\tilde{g}_{n_g} \tilde{\lambda} - \hat{g}_{n_g} \hat{\lambda})^\top \end{bmatrix}^\top\end{aligned}\tag{4.24}$$

With this notation, the condition for consistency is given in the following result:

Lemma 4.1 Assume that $D(q) = 1$. Further, assume that the data is persistently exciting in the following sense: there exist an integer $M > p$ and a real constant $\rho > 0$ such that, for every $r \geq p$,

$$\sum_{j=r}^{r+M} (\mathcal{F}_j \mathcal{F}_j^\top) > \rho I\tag{4.25}$$

holds almost surely, where I represents an identity matrix. Then the nonlinear least squares estimate $\hat{\vartheta}$ given in (4.21) is a consistent estimate of the true parameter vector $\tilde{\vartheta}$, i.e.

$$\hat{\vartheta} \rightarrow \tilde{\vartheta} \text{ w.p. 1, as } N \rightarrow \infty.\tag{4.26}$$

Proof: Suppose that $\hat{\vartheta} \neq \tilde{\vartheta}$ and (4.25) holds. Let

$$\left\| \begin{bmatrix} \Xi^\top & \Psi^\top \end{bmatrix}^\top \right\|_2 = \delta,$$

where $\delta > 0$. Simple algebraic manipulation shows that

$$\begin{aligned} \sum_{j=r}^{M+r} (y_j - \hat{y}_j)^2 &= \\ \begin{bmatrix} \Xi^\top & \Psi^\top \end{bmatrix} \left(\sum_{j=r}^{M+r} (\mathcal{F}_j \mathcal{F}_j^\top) \right) \begin{bmatrix} \Xi^\top & \Psi^\top \end{bmatrix}^\top. \end{aligned}$$

Using the definition of the smallest eigenvalue of a symmetric positive semidefinite matrix, this gives

$$\sum_{j=r}^{M+r} (y_j - \hat{y}_j)^2 > \delta \rho \text{ a.s.}$$

Since this lower bound holds for every $r \geq p$,

$$\sum_{r=p}^{p+N} \sum_{j=(r-1)M+r}^{Mr} (y_j - \hat{y}_j)^2 \geq N\delta\rho \text{ a.s.}$$

from which the result follows by letting N tend to infinity. \diamond

Unlike the fairly general consistency results in [55], the result above is for a specific class of systems and under a specific identification criterion, viz., nonlinear least squares. However, the advantage of this result is that it still provides strong consistency condition for a wide and practically relevant class of nonlinear systems (including Hammerstein systems with smooth nonlinearities), without requiring Gaussian noise or even noise with bounded higher order moments. It is difficult to allow for $n_d > 1$ and still yield a meaningful sufficient condition for consistency on data, similar to the rank condition in the previous lemma. However, solving (4.21) using the proposed algorithm with $n_d \neq 1$ still yields satisfactory results, as demonstrated in Section 4.4.

In the following section we describe the algorithm for the parameter estimation of the above system.

4.3 Parameter estimation algorithm

In what follows, let $0_n \in \mathbb{R}^n$ denote a vector with all zero entries. Under the identification criterion (4.21), the parameters of the system (4.10) can be estimated using the separable least squares method [39]. To any *given* vector of nonlinear parameters ϑ_n , there corresponds an optimal vector of linear parameters ϑ_l , given by

$$\hat{\vartheta}_l = J_l^\dagger \mathbf{Y}_N, \quad (4.27)$$

where J_l is the Jacobian matrix of the linear parameter vector with respect to ϑ_l :

$$J_l = \frac{\partial \hat{y}}{\partial \vartheta_l} = \left[\frac{\partial \hat{y}}{\partial a} \quad \frac{\partial \hat{y}}{\partial b} \quad \frac{\partial \hat{y}}{\partial \lambda} \right], \quad (4.28)$$

and \mathbf{Y}_N is an ordered vector of all output measurements. Here, A^\dagger denotes the pseudo-inverse of a matrix A . The cost function $V_N(\vartheta)$ in (4.20) is thus expressed as a function of the nonlinear parameters vector ϑ_n alone, forming the cost function of the separated problem

$$V_{SN}(\vartheta_{nl}) = V_N([\vartheta_{nl}^T \quad \hat{\vartheta}_l^T(\vartheta_{nl})]^T). \quad (4.29)$$

The optimal parameter vector $\hat{\vartheta}_{nl}$ is then given as a solution to the separated minimization problem

$$\hat{\vartheta}_{nl} = \arg \min_{\vartheta_{nl}} V_{SN}(\vartheta_{nl}), \quad (4.30)$$

where the minimum is understood to be a local minimum in general. Note that we have split a non-convex optimization (4.21) in $n_a + n_b + n_g + n_d + s + 4$ variables into two parts: a non-convex optimization (4.30) in $n_g + n_d + 2$ variables (i.e., potentially a much

smaller number of variables) followed by a simple linear least squares problem (4.28) in the remaining $n_a + n_b + s + 2$ variables. This illustrates the clear numerical advantage of using a model structure which allows us to use separable least squares. This structure also allows us to update the parameter vector $\hat{\vartheta}_l$ recursively using standard recursive least squares techniques. In section 4.1 it is shown that if $\hat{\vartheta}_{nl}$ is a minimizing argument of (4.30) and $\hat{\vartheta}_l$ is given by (4.27), then the parameter $\hat{\vartheta} = [\hat{\vartheta}_l^T \hat{\vartheta}_{nl}^T]^T$ is a minimizing argument of (4.21) and vice versa.

To solve problem (4.30) one needs to compute the Jacobian of the separated problem J_s . From [39], this is given by

$$J_s = J_n - J_l J_l^\dagger J_n, \quad (4.31)$$

where, J_l is the Jacobian matrix of the linear parameter vector given in (4.28) and J_n the Jacobian matrix of the nonlinear parameter vector given by

$$J_n = \frac{\partial \hat{y}}{\partial \vartheta_{nl}} = \begin{bmatrix} \frac{\partial \hat{y}}{\partial d} & \frac{\partial \hat{y}}{\partial g} \end{bmatrix}. \quad (4.32)$$

With the Jacobian J_s at hand one can use any step descent iterative algorithm to find the optimal parameter vector $\hat{\vartheta}_{nl}$. In this work, we use the Levenberg-Marquardt algorithm [77], [84]. The proposed identification scheme is summarized in Table 4.1.

Next, we describe a heuristic proposed for initializing this algorithm.

4.3.1 Algorithm Initialization

When no *a priori* knowledge or estimation of the parameter vectors ϑ_l and ϑ_{nl} is at hand, the following initialization process may be employed to obtain the initial estimates needed in the algorithm summarized in Table 4.1.

Table 4.1: Identification algorithm

```

INITIALIZE :  $\vartheta_l^0$  (see Section 4.3.1)
INITIALIZE  $\vartheta_{nl}^0$  (see Section 4.3.1)
Solve problem (4.30) by Levenberg-Marquardt iterations.
  Choose  $\gamma > 1$ ,  $0.95 \leq \rho < 1$  and set  $j = 0$ ,  $\delta = 1$ .
REPEAT
  Set  $j = j + 1$ 
  Compute the next step direction  $\Delta\vartheta_j$  by
    
$$\Delta\vartheta_j = \arg \min_{\Delta\vartheta} \| J_\delta \Delta\vartheta - \epsilon_\delta \|_2$$

  where  $J_\delta = \begin{bmatrix} J_s \\ \sqrt{\delta^j} I_q \end{bmatrix}$  and  $\epsilon_\delta = \begin{bmatrix} \epsilon \\ 0_q \end{bmatrix}$ 
  with  $\epsilon_i = y_i - \hat{y}_i^j$ , for  $i = 1, \dots, N$ .
  Update  $\vartheta_{nl}^{j+1} = \vartheta_{nl}^j + \Delta\vartheta_j$ 
  IF  $V_{SN}(\vartheta_{nl}^{j+1}) > V_{SN}(\vartheta_{nl}^j)$ 
  THEN  $\delta = \gamma\delta$ 
  ELSE  $\delta = \delta/\gamma$ 
  UNTIL  $V_{SN}(\vartheta_{nl}^{j+1}) / V_{SN}(\vartheta_{nl}^j) \leq \rho$ 
RETURN :  $\hat{\vartheta}_{nl}$ 
Estimate  $\vartheta_l$  by use of (4.27).
RETURN :  $\hat{\vartheta}_l$ 

```

1. Solve the linear regression

$$y_k = \mathcal{Y}_k^\top a + \mathcal{U}_k^\top b + \bar{w} + e_k \quad (4.33)$$

in a least squares sense, thus obtaining initial estimates a^0 and b^0 . The constant term \bar{w} accounts for the neglected effect of the nonlinear output signal w in (4.10a).

2. Compute the signal

$$v_k = y_k - \mathcal{Y}_k^\top a^0 - \mathcal{U}_k^\top b^0. \quad (4.34)$$

This is an estimate of the signal w corrupted by the system noise and the model error in the estimation of the a^0 and b^0 parameters.

3. Assuming $\mathcal{N}(\cdot)$ is the identity function, solve in a least squares sense the linear regression

$$v_k = \mathbf{U}_k^\top d + \varepsilon_k, \quad (4.35)$$

where $\mathbf{U}_k = \begin{bmatrix} u_{k-1} & u_{k-2} & \cdots & u_{k-n_d} \end{bmatrix}^\top$. Thus obtain an initial estimate d^0 .

4. Compute an estimate of the internal signal z_k

$$\hat{z}_k = \mathbf{U}_k^\top d^0. \quad (4.36)$$

5. Fit a n^{th} -order spline to the data $(\hat{z}_k, v_k)_{k=1}^N$ of the form

$$v_k = \sum_{j=1}^s \lambda_j |\hat{z}_k - \kappa_j|^n + \lambda_0 + \lambda_{s+1} \hat{z}_k. \quad (4.37)$$

The choice of s , n and κ_i can be made at this stage by observing the graph of \hat{z}_k versus v_k . It was found that using $n = 3$ and using uniformly spaced κ_i over signal range are sufficient for most applications. This gives the first estimate λ^0 of λ .

6. Let $\vartheta_l^0 = \left[(a^0)^\top \quad (b^0)^\top \quad (\lambda^0)^\top \right]^\top$ and $\vartheta_{nl}^0 = \left[(d^0)^\top \quad 0_{n_g}^\top \right]^\top$.

Use the vectors ϑ_n^0 and ϑ_l^0 to initialize the separable least squares algorithm summarized in Table 4.1.

4.4 Numerical Examples

In this section we present some illustrative examples to showcase the performance of the proposed algorithm.

Example 4.2 In this example we consider the following Wiener system.

$$y_k = \frac{z_k}{\sqrt{0.1 + 0.9z_k^2}} + e_k, \quad (4.38a)$$

$$z_k = d_1 u_{k-1} + d_2 u_{k-2} + d_3 u_{k-3}. \quad (4.38b)$$

A data set of $N = 2000$ data points (u_k, y_k) is generated with u_k uniformly distributed in $[-0.7, 2.5]$ and e_k normally distributed with zero mean and 0.1 standard deviation. Half of the data were used for identification purposes and the rest for validating the resulting model. The initialization process described in Section 4.3.1 was used to initialize the separable least squares algorithm. For the approximation of the nonlinear map, a cubic spline, i.e. $n = 3$, was fitted with a fixed *a priori* set of 12 knot points equally spaced in the range of the signal z_k . The true values and the initial and final estimates of the FIR coefficients are given in Table 2.

The performance of the model was measured by means of the Variance Accounted For (VAF%)

$$\text{VAF \%} = \left(1 - \frac{\text{Var}(\hat{y} - y)}{\text{Var}(y)}\right) * 100\% \quad (4.39)$$

The VAF% measures the proportion of the variance in the data $\{y_k\}_{k=1}^N$ that is explained by the fitted values $\{\hat{y}_k\}_{k=1}^N$. Thus, the closer VAF% is to 100% the better the accuracy of the estimated model. In this example, for the validation data set, VAF% = 98.11%. The approximation of the nonlinearity by the cubic spline is shown in Figure 4.2.

Table 4.2. Identification results in Example 4.2

	True	Initial	Final
d_1	-0.8944	-0.9237	-0.8958
d_2	0.2683	0.2157	0.2617
d_3	-0.3578	-0.3236	-0.3439

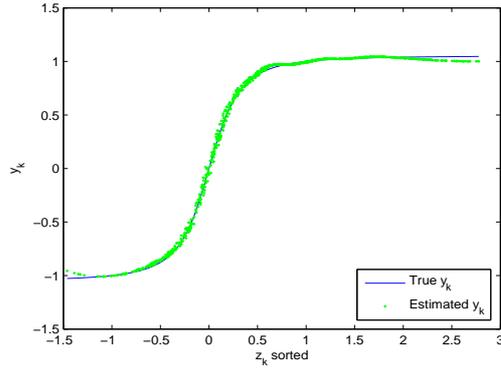


Figure 4.1: Approximation of the nonlinear map by a cubic spline in example 4.2.

Example 4.3 In this example we consider the Wiener-Hammerstein system given in (4.40) below.

$$y_k = a_1 y_{k-1} + w_k + g_1 w_{k-1} + e_k, \quad (4.40a)$$

$$w_k = \tan^{-1}(z_k - 1), \quad (4.40b)$$

$$z_k = d_1 u_{k-1} + d_2 u_{k-2}. \quad (4.40c)$$

A data set of $N = 2000$ data points (u_k, y_k) is generated with u_k uniformly distributed in $[-2.5, 2.5]$ and e_k normally distributed with mean 0 and 0.1 standard deviation. Half of the data were used for identification purposes and the rest for validating the resulting model. The initialization process described in Section 4.3.1 was used to initialize the separable least squares algorithm. For the approximation of the nonlinear map, a cubic spline was fitted with a fixed *a priori* set of 15 knot points equally spaced in the range of the signal z_k . Table 4.3 summarizes the true values and the initial and final estimates of the parameters to be identified.

To provide some insight on the steps of the algorithm, the stages of approximation of the nonlinear map by the cubic spline is shown in Figures 4.3-4.5. Figure 4.3 corresponds to the first approximation at initialization as described in Step (5) of Section 4.3.1. Figure 4.3 (left) shows the slightly enhanced fit achieved once the separated minimization problem (4.30) has

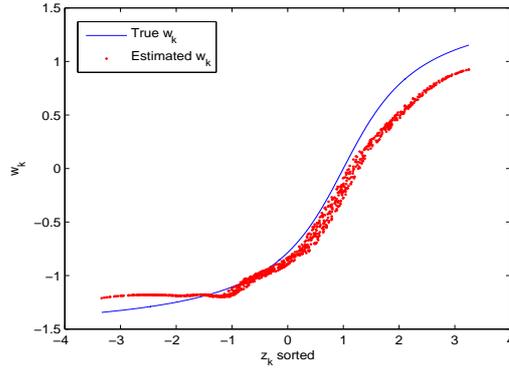


Figure 4.2: Samples (\hat{z}_k, \hat{w}_k) at initialization in example 4.3.

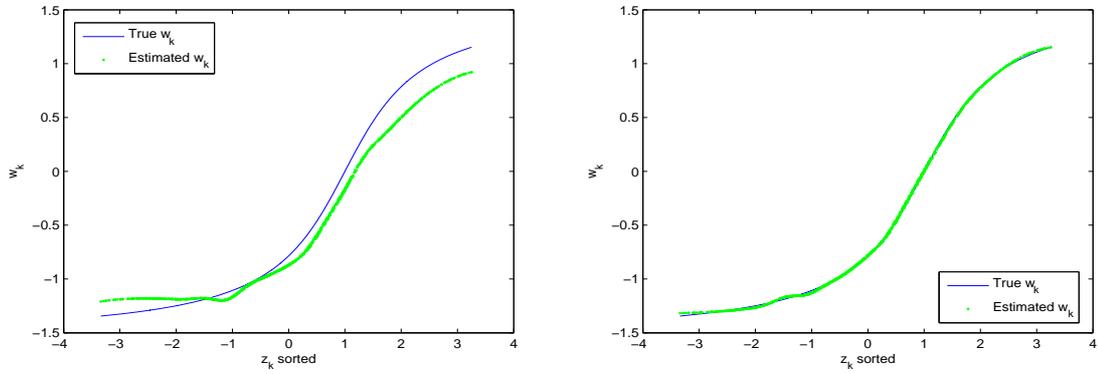


Figure 4.3: (left) \hat{z}_k, \hat{w}_k for updated values of $\hat{\vartheta}_n$ at the last iteration of the L-M algorithm (right) Final fit for updated values of $\hat{\vartheta}_l$ in example 4.3.

been solved. Finally, Figure 4.3 (right) shows the fit achieved once the spline parameters have been updated by use of (4.27). The VAF% in the validation data turned out to be 98.32% .

Table 4.3. Identification results in Example 4.3

	True	Initial	Final
a_1	-0.4	-0.4472	-0.4005
d_1	0.53	0.5663	0.5345
d_2	-0.848	-0.8242	-0.8452
g_1	-0.5	0	-0.5078

Example 4.4 We will close this section with an example of a general LFT model. Consider the following LFT system.

$$y_k = a_1 y_{k-1} + b_1 u_{k-1} + w_k + g_1 w_{k-1} + e_k, \quad (4.41a)$$

$$z_k = y_{k-2}, \quad (4.41b)$$

$$w_k = |z_k - 2| + 0.8 \tan^{-1}(z_k). \quad (4.41c)$$

A data set of $N = 1000$ data points (u_k, y_k) is generated with u_k uniformly distributed in $[-2.5, 2.5]$ and e_k uniformly distributed in $[-0.25, 0.25]$. Half of the data were used for identification purposes and the rest for validating the resulting model. The initialization process described in Section 4.3.1 was used to initialize the separable least squares algorithm. For the approximation of the nonlinear map, a cubic spline was fitted with a fixed *a priori* set of 13 knot points equally spaced in the range of the signal z_k . Table 4.4 summarizes the true values and the initial and final estimates of the parameters to be identified. The approximation of the nonlinearity by the cubic spline is shown in Figure 4.4. The VAF% in the validation data turned out to be 97.85%

Table 4.4. Identification results in Example 4.4

	True	Estimated
a_1	0.4	0.4004
b_1	1.2	1.1989
g_1	0.5	0.5178

4.5 Summary

In this chapter we have proposed a unified framework for a class of interconnected systems with dynamic linear components and static nonlinear components. We considered a subset of LFT systems of Figure 1.3, with \mathcal{L}_y modelled with an AutoRegressive with eXogeneous input (ARX) structure and \mathcal{L}_{zu} modelled with a Finite Impulse Response (FIR) structure. \mathcal{L}_{ze}

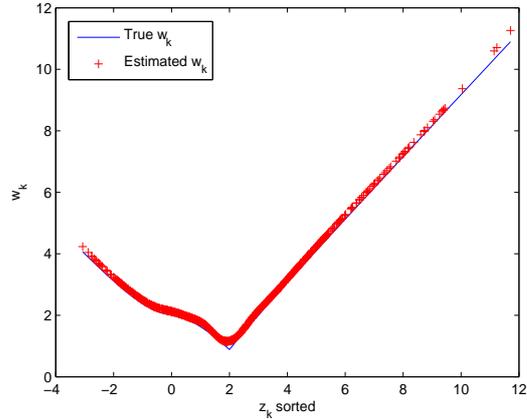


Figure 4.4: Approximation of the nonlinear map by a cubic spline in example 4.4.

and \mathcal{L}_{zw} were both assumed to be zero. Further, the nonlinear static map was approximated by a smooth m^{th} -order spline. We have developed an identification method based on the Separable Least Squares technique. For a restricted model set, we have also provided a sufficient condition on data for the consistency of the parameter estimate. The proposed algorithm is numerically tractable and can be initialized with little prior knowledge. The performance of the algorithm has been demonstrated with numerical examples for identifying a Wiener system as well as a Wiener-Hammerstein system and a general LFT system. Future directions regarding the research work presented here are outlined in Section 6.2.

Chapter 5

Piecewise Affine Identification

Nonlinear system identification has been an active research field in the last few decades, and a number of black-box identification approaches have been proposed as detailed in Chapter 1 and references therein. In particular, PieceWise Affine (PWA) system identification has recently deserved a lot of attention, mainly motivated by the universal approximation properties of PWA maps [14, 53]. Compared to other parametric approaches, PWA black-box identification has a major advantage. PWA maps can model a wide class of *discontinuities*. Therefore, PWA identification is most beneficial when no *a priori* knowledge is available on the continuity of the nonlinearity at hand, or when indeed dealing with a discontinuous nonlinear function.

In the previous chapter we examined the identification problem of a system given in the LFT form of Figure 1.3 under the assumption that the nonlinear static map \mathcal{N} is continuous over the domain of the input signal z_k . In this chapter we relax this assumption. Employing a PWA identification technique for the approximation of the static nonlinear map \mathcal{N} we are able to identify also piecewise continuous maps. We address the identification problem for an LFT interconnection composed by a LTI system and a static nonlinearity. Many nonlinear black-box structures can be used to this aim. In particular, one could identify a PieceWise affine ARX (PWARX) model of the LFT system by using any one of the PWA

system identification techniques proposed in the literature, as will be detailed in Section 5.1. However, the resulting model will typically not reflect the internal structure of the system, and will be more complex than needed in terms of number of parameters to be estimated. Motivated by this, an identification procedure is proposed which exploits the knowledge of the interconnection structure of the system generating the data. Numerical examples show that the proposed iterative scheme is able to exploit the knowledge of the system interconnection structure, thus providing simpler models compared to those obtained when applying black-box PWA identification techniques to the overall system. Parts of the work presented in this chapter can be found in [65], [61] and [66].

In Section 5.1 of this chapter we review the existing techniques in PWA system identification and particularly we present two distinct techniques that we will use in the following. In Section 5.2 we present the considered model structure in SISO form and the associated identifiability requirements, followed by the proposed iterative identification scheme in Section 5.3 and concluded by numerical examples illustrating the performance of the algorithm in Section 5.4. Finally, conclusions and possible future research directions are summarized in Section 5.5.

5.1 PWA identification of nonlinear systems : An overview

A PWA map $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is given by

$$f(x) = \begin{cases} \theta_1^\top \begin{bmatrix} x \\ 1 \end{bmatrix}, & \text{if } x \in \mathcal{X}_1, \\ \vdots \\ \theta_s^\top \begin{bmatrix} x \\ 1 \end{bmatrix}, & \text{if } x \in \mathcal{X}_s, \end{cases} \quad (5.1)$$

where the regressor space $\mathcal{X} \subset \mathbb{R}^n$ is a bounded polyhedron and the set $\{\mathcal{X}_i\}_{i=1}^s$ is a polyhedral partition of \mathcal{X} , i.e. each set \mathcal{X}_i is a convex polyhedron such that $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset, \forall i \neq j$ and $\cup_{i=1}^s \mathcal{X}_i = \mathcal{X}$. The map is parametrised by the parameters $\theta_i \in \mathbb{R}^{n+1}, i = 1, \dots, s$.

Given a set of input-output data $\{u_k, y_k\}_{k=1}^N$ generated by a nonlinear system \mathcal{N} , PWA identification consists in estimating a PWA map f such that

$$y_k = f(x_k) + \varepsilon_k, \quad (5.2)$$

where ε_k are error terms due to the system noise and/or identification error. The regressor vector $x_k \in \mathbb{R}^n$ is constructed by the previous values of the input and output signals. In most of the applications considered in the literature x_k is defined as follows

$$x_k = \begin{bmatrix} y_{k-1} & \dots & y_{k-n_a} & u_{k-1} & \dots & u_{k-n_b} \end{bmatrix}^\top. \quad (5.3)$$

That is, at each region \mathcal{X}_i , an AutoRegressive eXogenous model is identified. The resulting PWA system is thus called PWARX (PieceWise ARX).

The identification process consists in estimating the hyperplanes that define the partition $\{\mathcal{X}_i\}_{i=1}^s$ of the regressors domain \mathcal{X} and the parameters $\{\theta_i\}_{i=1}^s$ of the affine submodels in each region \mathcal{X}_i of the partition. Clearly, the most challenging issue arising here is the classification of the pairs of data points $\{u_k, y_k\}_{k=1}^N$ in their corresponding regions $\mathcal{X}_i, i = 1, \dots, s$. The proposed identification techniques either fix the number of affine submodels a priori [32, 54, 58, 69, 72], or estimate it from data [10, 31, 57]. For an extensive overview on PWA system identification, the interested reader is referred to [71] or [62].

5.1.1 A bounded-error procedure for PWA identification

In this section, the main features of the bounded-error procedure for identification of PWA models presented in [10] are briefly reviewed. Given N data pairs $(w_k, z_k), k = 1, \dots, N$,

where $z \in \mathbb{R}^{n_z}$ and $w \in \mathbb{R}$ are the independent and dependent variables, respectively, the aim of the bounded-error procedure is to fit a PWA map (5.7) such that the bounded-error condition

$$|w_k - f(\mathbf{z}_k)| \leq \delta, \quad \forall k = 1, \dots, N, \quad (5.4)$$

is satisfied for a given quantity $\delta > 0$. The procedure works in three stages. First, a MIN PFS approach [2] is employed to simultaneously classify the data points and estimate a minimum number s of modes so that (5.4) is satisfied (notice that s is not fixed a priori). At second stage, refinement of the data point classification takes place, where misclassified data are reduced and the parameters are estimated afresh. Finally, the polyhedral partition of the domain \mathcal{Z} is estimated using two-class or multiclass linear separation techniques. For further details on the algorithm and its implementation, the interested reader is referred to [10]. It turns out that the bound δ in (5.4) is a tuning parameter of the algorithm that can be used to achieve the desired trade off between model accuracy and complexity. In fact, the smaller δ , the larger is typically the number of modes needed to fit the data (with noisy measurements, this brings on the risk of overfit). Conversely, the larger δ , the worse is the fit, since larger errors are allowed. The choice of a suitable δ can be made by plotting the number of modes versus δ , and selecting δ at the knee of the curve (see also Section 5.4 below, where this idea for the selection of δ is applied).

5.2 Problem Formulation

5.2.1 Model Structure

We consider discrete-time networked dynamical systems defined as the interconnection of linear and nonlinear components, and represented by LFTs of the type shown in Figure 1.3. The block \mathcal{L} is a linear time-invariant dynamical system, while the block \mathcal{N} is assumed to be a static nonlinearity. The scalar signals u_k , y_k and $e_k \in \mathbb{R}$ are the system input, output

and noise at time $k \in \mathbb{Z}$, while $z_k \in \mathbb{R}^{n_z}$ and $w_k \in \mathbb{R}$ are internal signals representing the input and the output of the nonlinear part.

Let Z be a collection of n_y past values of y and the current and n_u past values of the input u , that is $Z = \{y_{k-1}, \dots, y_{k-n_y}, u_k, u_{k-1}, \dots, u_{k-n_u}\}$, where the orders n_y and n_u are assumed to be known. Then z_k is a vector containing some or all of the elements of Z , $z_k \in \mathbb{R}^{n_z}$, where $n_z \leq n_y + n_u + 1$.

For the LFT system in Figure 1.3, the following model class is considered:

$$A(q)y_k = B(q)u_{k-n_k} + G(q)w_k + \varepsilon_k, \quad (5.5a)$$

$$w_k = f(z_k). \quad (5.5b)$$

The linear part \mathcal{L} is described by the ARX model (5.5a), where $\varepsilon_k \in \mathbb{R}$ is the error term, and $A(q), B(q), G(q)$ are finite polynomials of known orders n_a, n_b, n_k and n_g respectively, in the delay operator q^{-1} , namely

$$A(q) = 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a}, \quad (5.6a)$$

$$B(q) = b_0 + b_1q^{-1} + \dots + b_{n_b}q^{-n_b}, \quad (5.6b)$$

$$G(q) = g_0 + g_1q^{-1} + \dots + g_{n_g}q^{-n_g}. \quad (5.6c)$$

The nonlinear part \mathcal{N} is described by the static relation (5.5b), where $f(\cdot)$ is a PWA map of the form

$$f(z) = \begin{cases} \theta_1^\top \phi & \text{if } z \in \mathcal{Z}_1, \\ \vdots & \\ \theta_s^\top \phi & \text{if } z \in \mathcal{Z}_s. \end{cases} \quad (5.7)$$

In (5.7), $\phi = [z^\top \ 1]^\top$, s is the number of modes, $\theta_i \in \mathbb{R}^{n_z+1}$, $i = 1, \dots, s$, are the parameters of each mode, and $\{\mathcal{Z}_i\}_{i=1}^s$ is a complete partition of the domain $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$ where $f(\cdot)$ is defined. Each set \mathcal{Z}_i , $i = 1, \dots, s$, is a convex polyhedron described by $\mathcal{Z}_i = \{z \in \mathbb{R}^{n_z} : H_i \phi \preceq_{[i]} 0\}$,

where $H_i \in \mathbb{R}^{\mu_i \times (n_z+1)}$, $i = 1, \dots, s$, μ_i is the number of linear inequalities defining the i th polyhedral region \mathcal{Z}_i , and the symbol $\preceq_{[i]}$ denotes a μ_i -dimensional vector whose elements can be the symbols \leq and $<$, to avoid that the sets \mathcal{Z}_i overlap over common boundaries.

In the following, it will be useful to write Equation (5.5a) in regression form as follows:

$$y_k = \eta_L^\top \varphi_k + w_k + \varepsilon_k, \quad (5.8)$$

where

$$\eta_L = [\eta_\zeta^\top \ \eta_\omega^\top]^\top, \text{ with} \quad (5.9a)$$

$$\eta_\zeta = [a_1 \ \dots \ a_{n_a} \ b_0 \ b_1 \ \dots \ b_{n_b}]^\top \text{ and} \quad (5.9b)$$

$$\eta_\omega = [g_1 \ \dots \ g_{n_g}]^\top \quad (5.9c)$$

is the unknown parameter vector, $\eta_L \in \mathbb{R}^{n_L}$, $n_L = n_a + n_b + n_g + 2$ and

$$\varphi_k = [\zeta_k^\top \ \omega_k^\top]^\top, \text{ with} \quad (5.10a)$$

$$\zeta_k = [-y_{k-1} \ \dots \ -y_{k-n_a} \ u_k \ u_{k-1} \ \dots \ u_{k-n_b}]^\top \text{ and} \quad (5.10b)$$

$$\omega_k = [w_{k-1} \ \dots \ w_{k-n_g}]^\top, \quad (5.10c)$$

is the (partially unknown) regression vector.

In the following section we discuss the identifiability of the proposed model class (5.5).

5.2.2 Identifiability

In this section we discuss the identifiability of the proposed model.

Let $w_k = f(z_k) = \sum_{i=1}^s (\theta^{(i)})^\top \phi_k \delta_{ik}$, where

$$\delta_{ik} = \begin{cases} 1, & \text{if } z_k \in \mathcal{Z}_i, \\ 0, & \text{otherwise.} \end{cases}$$

Equation (5.5) becomes

$$A(q)y_k = B(q)u_{k-n_k} + \sum_{i=1}^s (\theta^{(i)})^\top \delta_{ik} G(q) \phi_k + \varepsilon_k. \quad (5.11)$$

Using the notation introduced in (5.9) and (5.10), equation (5.11) can be re-written as

$$y_k = \zeta_k^\top \eta_\zeta + \sum_{i=1}^s (\theta^{(i)})^\top \delta_{ik} G(q) \phi_k + \varepsilon_k \quad (5.12)$$

or in more detail, as:

$$\begin{aligned} y_k &= -a_1 y_{k-1} - \dots - a_{n_a} y_{k-n_a} + b_0 u_{k-n_k} + \dots + b_{n_b} u_{k-n_k-n_b} \\ &+ g_0 \underbrace{\sum_{i=1}^s (\theta_1^{(i)} y_{k-1} + \dots + \theta_{n_y+1}^{(i)} u_k + \dots + \theta_{n_z}^{(i)} u_{k-n_u} + \theta_{n_z+1}^{(i)}) \delta_{ik}}_{w_k} \\ &+ g_1 \underbrace{\sum_{i=1}^s (\theta_1^{(i)} y_{k-2} + \dots + \theta_{n_y+1}^{(i)} u_{k-1} + \dots + \theta_{n_z}^{(i)} u_{k-n_u-1} + \theta_{n_z+1}^{(i)}) \delta_{i(k-1)}}_{w_{k-1}} \\ &+ \dots \\ &+ g_{n_g} \underbrace{\sum_{i=1}^s (\theta_1^{(i)} y_{k-n_g} + \dots + \theta_{n_y+1}^{(i)} u_{k-n_g} + \dots + \theta_{n_z}^{(i)} u_{k-n_u-n_g} + \theta_{n_z+1}^{(i)}) \delta_{i(k-n_g)}}_{w_{k-n_g}} \\ &+ \varepsilon_k \end{aligned} \quad (5.13)$$

The model (5.12) is ill-posed. Firstly, for any $\beta \neq 0$, one can choose parameters $\tilde{g}_j = \beta g_j$, $j = 1, \dots, n_g$ and $\tilde{\theta}^{(i)} = \frac{1}{\beta} \theta^{(i)}$, $i = 1, \dots, s$ such that $g_j \theta^{(i)} = \tilde{g}_j \tilde{\theta}^{(i)}$. Without loss of generality, in order to overcome this breach of identifiability we set $g_0 = 1$, since for $g_0 \neq 1$ we can always scale the nonlinear map accordingly.

Remark 5.1 In general ARX LTI systems, requiring that $g_0 = 1$ is a strict constraint, as it implicitly assumes that the corresponding signal w enters the system without delay. However in our case, we can always define the approximating PWA map $f(\cdot)$ such that g_0 is taken non-zero. In support of this argument we provide the following illustrative example. As a result, the choice of $g_0 = 1$ is merely a normalization operation and does not impose any severe restriction on the system structure.

Example 5.1 Consider the following LFT system :

$$y_k = b_1 u_{k-1} + g_1 w_{k-1} + \varepsilon_k, \quad (5.14a)$$

$$w_k = y_{k-1} u_{k-1}^2. \quad (5.14b)$$

To approximate this nonlinearity by a PWA map one could use $\mathbf{z}_k = [y_{k-1} \ u_{k-1}]^T$ and could try to estimate a map $f(\cdot)$ such that $w_k = f(\mathbf{z}_k)$, which corresponds to the system (5.14).

However, it is always possible to define $\tilde{\mathbf{z}}_k = [y_{k-2} \ u_{k-2}]^T$ and estimate a PWA map $\tilde{f}(\cdot)$ such that $\tilde{w}_k = \tilde{f}(\tilde{\mathbf{z}}_k)$. This formulation corresponds to the system

$$y_k = b_1 u_{k-1} + \tilde{g}_0 \tilde{w}_k + \varepsilon_k, \quad (5.15a)$$

$$\tilde{w}_k = y_{k-2} u_{k-2}^2. \quad (5.15b)$$

with $\tilde{g}_0 = g_1$, which is equivalent to system (5.14).

◇

A close inspection of (5.12) reveals another identifiability issue. The possibility of common regressors in y_{k-j} , for some $j = 1, \dots, \max(n_a, n_y + n_g)$ and/or in u_{k-l} for some $l = 1, \dots, \max(n_k + n_b, n_u + n_g)$ in the regression vectors ζ_k of the linear system and ϕ_k of the PWA map, imposes another ill-conditioning on the system. The following lemma provides necessary and sufficient conditions for the identifiability of system (5.12).

Lemma 5.1 The system model of (5.5) with $g_0 = 1$ is identifiable if and only if the following conditions hold :

$$a_{j+l}\theta_j^{(i)} = 0, \quad (5.16)$$

for all $i = 1, \dots, s$, $j = 1, \dots, \min(n_a, n_y)$ and $l = 0, \dots, n_g$ such that $j + l \leq n_a$ and

$$b_{j-n_k+l}\theta_{n_y+1+j}^{(i)} = 0, \quad (5.17)$$

for all $i = 1, \dots, s$, $j = 0, \dots, \min(n_k + n_b, n_u)$ and $l = 0, \dots, n_g$ such that $0 \leq j - n_k + l \leq n_b$.

Conditions (5.16)-(5.17) are equivalent to requiring that there is no common regressor in all pairs of $(\zeta_k, \mathbf{z}_k), (\zeta_k, \mathbf{z}_{k-1}), \dots, (\zeta_k, \mathbf{z}_{k-n_g})$.

Proof.

To prove necessity, suppose for contradiction that there appears one common regressor in one of the pairs $(\zeta_k, \mathbf{z}_k), (\zeta_k, \mathbf{z}_{k-1}), \dots, (\zeta_k, \mathbf{z}_{k-n_g})$. Without loss of generality suppose that the common regressor is y_{k-1} in the pair (ζ_k, \mathbf{z}_k) , i.e. condition $a_1\theta_1^{(i)} = 0$ is violated for some $i = 1, \dots, s$.

Let $\eta_L = \begin{bmatrix} a_1 & \tilde{\eta}_L^T \end{bmatrix}^T$ and $\theta^{(i)} = \begin{bmatrix} \theta_1^{(i)} & (\tilde{\theta}^{(i)})^T \end{bmatrix}^T$ and define $\tilde{\zeta}_k$ and $\tilde{\phi}_k$ accordingly. Then equation (5.11) becomes

$$y_k = (a_1 + \sum_{i=1}^s \theta_1^{(i)} \delta_{ik}) y_{k-1} + \tilde{\zeta}_k^T \tilde{\eta}_L + \sum_{i=1}^s (\tilde{\theta}^{(i)})^T G(q) \tilde{\phi}_k + \varepsilon_k \quad (5.18)$$

But this model is not uniquely identifiable since for any choice of $\alpha \neq 0$ in (5.18) the

transformation

$$\underbrace{\{a_1 + \alpha\}}_{\bar{a}_1} + \sum_{i=1}^s \underbrace{(\theta_1^{(i)} - \alpha)}_{\tilde{\theta}_1^{(i)}} \delta_{ik} \} y_{k-1}$$

will generate the same input/output behaviour. That is, there exist infinitely many parameter vectors η^* , with $\vartheta \neq \vartheta^*$, of the form

$$\begin{aligned} \eta^* &= [(\eta_L^*)^T (\eta_{NL}^*)^T], \\ \eta_L^* &= [\bar{a}_1 \tilde{\eta}_L^T]^T, \\ \eta_{NL}^* &= [(\theta^{(1)})^T \dots [\tilde{\theta}_1^{(i)} \tilde{\theta}^{(i)}]^T \dots (\theta^{(s)})^T]^T, \end{aligned}$$

such that $F_\eta(\mathcal{L}, \mathcal{N}) = F_{\eta^*}(\mathcal{L}, \mathcal{N})$, which contradicts the assumption of global identifiability.

To prove sufficiency, notice that if no common regressors appear in all pairs (ζ_k, \mathbf{z}_k) , $(\zeta_k, \mathbf{z}_{k-1}), \dots, (\zeta_k, \mathbf{z}_{k-n_g})$, i.e. conditions (5.16)-(5.17) hold then equation (5.12) can be written as the linear regression model

$$y_k = [\zeta_k^T w_k \dots w_{k-n_g}] \eta_L + \varepsilon_k \quad (5.19)$$

which is globally identifiable [55].

◇

In the following section we describe an iterative algorithm for the identification of system (5.5) under the identifiability conditions presented above.

5.3 Parameter estimation algorithm

A challenging issue in the considered identification problem is the fact that the internal signal w_k is not measured, and hence it must be estimated along with the polynomials $A(q)$, $B(q)$, and $G(q)$. In this section, we propose an iterative identification procedure which alternates between the estimation of the linear and the nonlinear part. This makes

it possible to recover w_k by means of smoothing techniques, as suggested in [24]. Then, a PWA map $f(\cdot)$ can be fitted to the pairs (w_k, \mathbf{z}_k) . On the other hand, once w_k is known, the identification of the linear part can be easily carried out by means of standard identification techniques [55]. The above discussion suggests the formulation of the iterative identification procedure summarized in Table 5.1, where each iteration consists of two stages, namely the PWA approximation of the nonlinear part and the identification of the linear part. These two stages are described below.

Let $j = 1, 2, \dots$ be the iteration counter, and η_L^{j-1} be the estimate of the ARX coefficients in (5.9) computed at iteration $j - 1$. An estimate η^0 must be provided for initialization. Moreover, let

$$\bar{n} = \max\{n_a, n_b + n_k, n_g + n_y, n_g + n_u\}. \quad (5.20)$$

PWA approximation of the nonlinear part

Given the estimate η_L^{j-1} of the ARX coefficients, standard smoothing techniques are used to recover a suitable signal $\{v_k^j\}_{k=\bar{n}-n_g+1}^N$ according to the estimated linear dynamics

$$y_k = (\eta_\zeta^{j-1})^T \zeta_k + v_k^j + (\eta_\omega^{j-1})^T \begin{bmatrix} v_{k-1}^j \\ \vdots \\ v_{k-n_g}^j \end{bmatrix} + \varepsilon_k, \quad (5.21)$$

with $k = \bar{n} + 1, \dots, N$. Note that (5.21) is obtained by replacing η_L with η_L^{j-1} in (5.12). In general, the role of the error term ε_k depends on the particular smoothing algorithm chosen. A very simple choice for the considered ARX model is to set $\varepsilon_k = 0$ for all k , and then to recover the signal v_k^j as a particular solution of the set of linear equations resulting from (5.21).

The next step is to fit a PWA map $f^j(\cdot)$ to the samples (v_k^j, \mathbf{z}_k) , $k = \bar{n} - n_g + 1, \dots, N$. To this aim we adopt the bounded-error technique proposed in [10]. An attractive feature of this method is that the number of modes of the PWA map is automatically estimated from data. A short description of the adopted procedure is given in Section 5.1.1.

Table 5.1: Iterative identification algorithm

GIVEN: η_L^0, γ
 SET: $j = 0$
 REPEAT
 SET: $j = j + 1$
 % PWA approximation of the nonlinear part
 Estimate $\{v_k^j\}$ according to (5.21)
 Fit a PWA map $f^j(\cdot)$ to the samples (v_k^j, z_k)
 Compute w_k^j as in (5.22)
 % Identification of the linear part
 Compute η_L^j from the linear regression (5.23)
 UNTIL $\|\eta_L^j - \eta_L^{j-1}\|_2 \leq \gamma \|\eta_L^j\|_2$
 RETURN: $\eta_L^j, f^j(\cdot)$

Once a PWA map $f^j(\cdot)$ has been fitted to the data, an estimate of the unknown sequence $\{w_k\}_{k=\bar{n}-n_g+1}^N$ can be obtained as

$$w_k^j = f^j(z_k), \quad k = \bar{n} - n_g + 1, \dots, N. \quad (5.22)$$

It is stressed that, in the proposed scheme, v_k^j is seen as a “noisy” version of w_k^j , where by “noisy” it is meant that v_k^j will typically include the effects of both the system noise and the model error. Hence, the role of the PWA approximation stage is not only to provide an analytic expression of the static nonlinearity, but also to improve the smoothing of the unknown signal w_k . This feature of the proposed procedure is illustrated below in Example 5.1.

Identification of the linear part

Given the estimated signal $\{w_k^j\}_{k=\bar{n}-n_g+1}^N$, one can form the estimated regression vectors $\{\varphi_k^j\}_{k=\bar{n}+1}^N$ by replacing w_k with w_k^j in the definition of ω_k in (5.10c). Then, an estimate η_L^j of the ARX coefficients can be easily computed through standard techniques (for instance,

least squares) from the linear regression

$$y_k - w_k^j = (\eta_L^j)^T \varphi_k^j + \varepsilon_k, \quad k = \bar{n} + 1, \dots, N. \quad (5.23)$$

The procedure terminates when no significant changes occur to the estimated ARX coefficients between two consecutive iterations. This criterion is implemented by checking if

$$\|\eta_L^j - \eta_L^{j-1}\|_2 \leq \gamma \|\vartheta_L^j\|_2, \quad (5.24)$$

where γ is a positive threshold defined by the user, and $\|\cdot\|_2$ denotes the Euclidean norm of a vector.

5.4 Numerical examples

In this section, three numerical examples are presented to illustrate the performance of the proposed identification procedure. Static nonlinearities $\mathcal{N}(\cdot)$ of piecewise affine type are considered. Since in this case the overall interconnected system is still PWA, one can compare directly the models estimated by the proposed iterative procedure and those obtained by identifying a PWARX model of the overall system (i.e., without exploiting the interconnection structure).

Example 5.1 This simple example illustrates the basic features of the proposed identification algorithm. Consider the Hammerstein system with continuous PWA static nonlinearity

$$y_k = -a_1 y_{k-1} + w_k + e_k \quad (5.25a)$$

$$w_k = \begin{cases} \alpha_1 u_k & \text{if } u_k \geq \tau \\ \alpha_2 u_k & \text{if } u_k < \tau, \end{cases} \quad (5.25b)$$

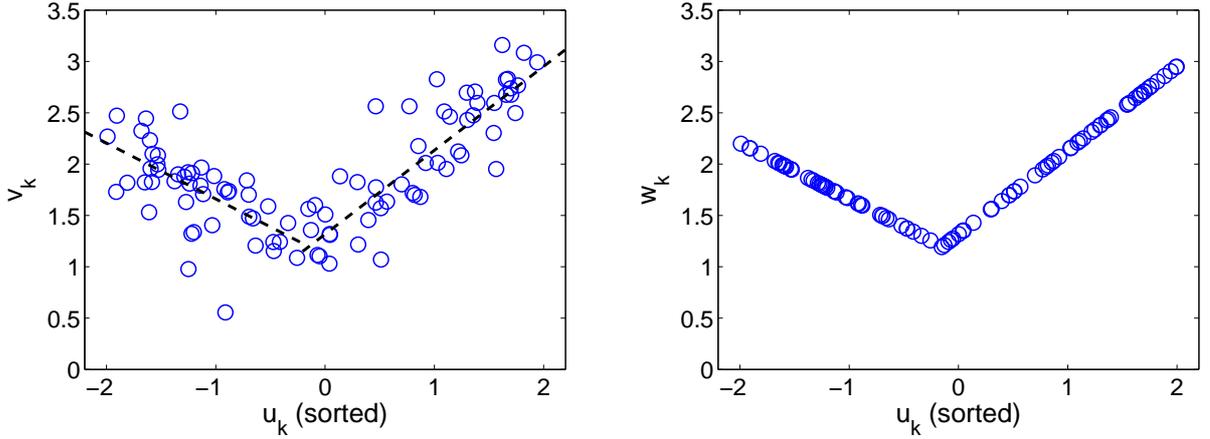


Figure 5.1: First algorithm iteration in Example 5.1: (left) Circles represent the “noisy” signal v_k^1 , while the dashed lines represent a PWA map with two modes that is fitted to the samples (v_k^1, u_k) ; (right) Circles represent the corresponding “smoothed” signal w_k^1 .

where $a_1 = -0.8$, $\alpha_1 = 1$, $\alpha_2 = -0.5$, and $\tau = 0$. A data set of $N = 100$ data points (y_k, u_k) , $k = 1, \dots, N$, is generated with u_k and e_k uniformly distributed in $[-2, 2]$ and $[-0.2, 0.2]$, respectively. The iterative algorithm described in Table 5.1 is then applied with initial estimate $a_1^0 = -0.5$, and $\gamma = 0.001$. Due to the simple linear dynamics (5.25a), at each iteration the signal v_k^j can be readily obtained as $v_k^j = y_k + a_1^{j-1} y_{k-1}$, $k = 2, \dots, N$. It is apparent that v_k^j includes three contributions: the output w_k of the nonlinear part, the noise signal e_k and the model error of the linear part. Figure 5.1 shows the “noisy” signal v_k^1 at the first iteration of the algorithm and the corresponding “smoothed” signal w_k^1 , obtained by fitting a PWA map with two modes to the samples (v_k^1, u_k) , $k = 2, \dots, N$. It can be observed that the shape of the system nonlinearity is already well estimated at the first iteration, while the vertical shift is significant (about 1.2). The iterative algorithm terminates after 5 iterations, providing the estimates $\hat{a}_1 = -0.778$, $\hat{\alpha}_1 = 0.976$, $\hat{\alpha}_2 = -0.497$, $\hat{\tau} = 0.091$, that are all close to the corresponding true values. Quite notably, the vertical shift of the estimated PWA map is reduced to about 0.1, which is clearly acceptable if compared to the noise magnitude.

Example 5.2 Consider the following LFT system:

$$y_k = -a_1 y_{k-1} + b_1 u_{k-1} + w_k + g_1 w_{k-1} + e_k \quad (5.26a)$$

$$w_k = \begin{cases} \theta_1^T \begin{bmatrix} y_{k-2} \\ u_{k-2} \\ 1 \end{bmatrix} & \text{if } h^T \begin{bmatrix} y_{k-2} \\ u_{k-2} \\ 1 \end{bmatrix} \leq 0 \\ \theta_2^T \begin{bmatrix} y_{k-2} \\ u_{k-2} \\ 1 \end{bmatrix} & \text{if } h^T \begin{bmatrix} y_{k-2} \\ u_{k-2} \\ 1 \end{bmatrix} > 0, \end{cases} \quad (5.26b)$$

where $a_1 = -0.4$, $b_1 = 1.2$, $g_1 = -0.5$, and θ_1 , θ_2 , and h are reported in Table 5.2. A data set of $N = 1000$ data points (y_k, u_k) , $k = 1, \dots, N$, is generated with u_k and e_k uniformly distributed in $[-1, 1]$ and $[-0.25, 0.25]$, respectively. To apply the identification algorithm described in Section 5.3, initial estimates $a_1^0 = -0.4645$ and $b_1^0 = 1.2529$ are obtained by fitting the ARX model

$$y_k = -a_1^0 y_{k-1} + b_1^0 u_{k-1} + \bar{w}^0 + \varepsilon_k \quad (5.27)$$

to the data by least squares, while g_1^0 is set equal to 0. The constant term \bar{w}^0 , to be estimated in (5.27), accounts for the effects of the neglected terms depending on the unknown signal w_k (i.e. for the presence of the nonlinearity). Estimation of the static nonlinearity is carried out by applying the bounded-error procedure described in Section 5.1.1, with $\delta = 0.6$ at the first iteration, and $\delta = 0.3$ in the next iterations. Least squares are used for estimation of the ARX coefficients. The iterative algorithm terminates after 13 iterations for $\gamma = 0.001$. The estimates $\hat{a}_1 = -0.3904$, $\hat{b}_1 = 1.2041$ and $\hat{g}_1 = -0.4914$ are finally returned. The sequence of estimates is plotted in Figure 5.2, showing the convergence of the iterative procedure. Concerning the reconstruction of the static nonlinearity (5.26b), two modes are correctly

Table 5.2: True and estimated parameters in Example 5.2

θ_1	$\hat{\theta}_1$	θ_2	$\hat{\theta}_2$	h	\hat{h}
0.5	0.4940	-0.7	-0.6766	1.5	1.4829
1	1.0195	1.5	1.5151	-0.5	-0.4980
-0.25	-0.2826	0	-0.0095	1	1

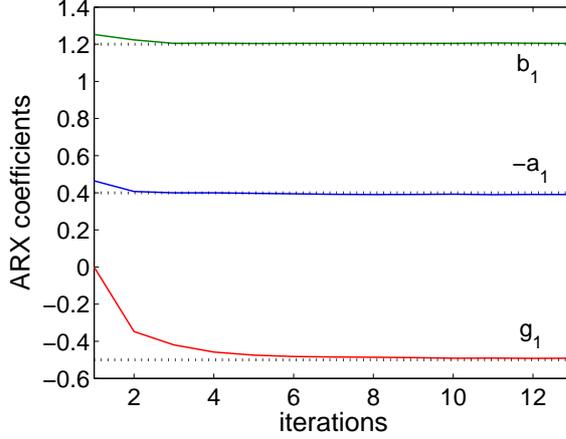


Figure 5.2: Sequence of estimates of the ARX coefficients $-a_1$, b_1 and g_1 in Example 5.2 (dotted lines are true values).

estimated, with parameters $\hat{\theta}_1$, $\hat{\theta}_2$, and \hat{h} reported in Table 5.2. It is apparent that all the estimated parameters of both the ARX dynamics and the static nonlinearity are very close to the corresponding true values.

Identification results obtained by applying the proposed iterative algorithm are compared to those obtained by fitting directly a PWARX model to the data. To this aim, note that the system defined by (5.26) can be rewritten as a PWARX system with four modes, i.e.

$$y_k = \begin{cases} \Theta_{1,1}^T \begin{bmatrix} \mathbf{r}_k \\ 1 \end{bmatrix} & \text{if } \mathbf{r}_k \in \mathcal{R}_{1,1} \\ \Theta_{1,2}^T \begin{bmatrix} \mathbf{r}_k \\ 1 \end{bmatrix} & \text{if } \mathbf{r}_k \in \mathcal{R}_{1,2} \\ \Theta_{2,1}^T \begin{bmatrix} \mathbf{r}_k \\ 1 \end{bmatrix} & \text{if } \mathbf{r}_k \in \mathcal{R}_{2,1} \\ \Theta_{2,2}^T \begin{bmatrix} \mathbf{r}_k \\ 1 \end{bmatrix} & \text{if } \mathbf{r}_k \in \mathcal{R}_{2,2}, \end{cases} \quad (5.28)$$

where $\mathbf{r}_k = [y_{k-1} \ y_{k-2} \ y_{k-3} \ u_{k-1} \ u_{k-2} \ u_{k-3}]^T$, and the mode characterized by subscripts (i, j) , $i, j = 1, 2$, corresponds to the situation when w_k and w_{k-1} are generated by the i th and j th mode of (5.26b), respectively. Note that the LFT representation of the system is more parsimonious than the PWARX representation (5.28), as the former requires 9 parameters (a_1 , b_1 , g_1 , θ_1 and θ_2), while the latter 21 ($\Theta_{1,1}$, $\Theta_{1,2}$, $\Theta_{2,1}$ and $\Theta_{2,2}$). For a given finite data

Table 5.3: Comparison of identification results in Example 5.2

mode	$\frac{\ \tilde{\Theta}_{i,j} - \Theta_{i,j}\ }{\ \Theta_{i,j}\ } \times 100$	$\frac{\ \hat{\Theta}_{i,j} - \Theta_{i,j}\ }{\ \Theta_{i,j}\ } \times 100$
(1, 1)	8.7534	1.7109
(1, 2)	2.7883	2.1367
(2, 1)	9.6381	1.4323
(2, 2)	7.2664	1.5740

set, this is expected to affect the quality of the identification results. In order to directly identify (5.28) from data, the most favorable situation is when the partition of the regressors domain is known. Indeed, in that case the PWA system identification problem reduces to a standard linear identification problem for each mode, and parameter estimates $\tilde{\Theta}_{1,1}$, $\tilde{\Theta}_{1,2}$, $\tilde{\Theta}_{2,1}$ and $\tilde{\Theta}_{2,2}$ can be computed via least squares. For comparison purposes, estimates $\hat{\Theta}_{1,1}$, $\hat{\Theta}_{1,2}$, $\hat{\Theta}_{2,1}$ and $\hat{\Theta}_{2,2}$ are reconstructed from the estimated parameters \hat{a}_1 , \hat{b}_1 , \hat{g}_1 , $\hat{\theta}_1$ and $\hat{\theta}_2$ provided by the iterative procedure. Both $\tilde{\Theta}_{i,j}$ and $\hat{\Theta}_{i,j}$ are compared to the corresponding true values $\Theta_{i,j}$: relative estimation errors are reported in Table 5.3. It turns out that the reconstructed estimates $\hat{\Theta}_{i,j}$ are closer to the true values than the estimates $\tilde{\Theta}_{i,j}$ identified directly. This shows that enhanced identification accuracy can be achieved by exploiting the system LFT structure, rather than identifying a PWARX model of the whole system.

To stress the difficulty of estimating a PWARX model of the whole system in a real situation when both the partition and the number of modes of (5.28) are not known, the bounded-error procedure [10] is also applied to the considered data set. If δ is taken equal to the true noise bound, i.e. $\delta = 0.25$, the procedure fails to reconstruct four linearly separable clusters of points. This is due to the suboptimality of the greedy randomized algorithm employed to partition the data set at initialization, which is not able to single out four distinct dynamics in the considered data set due to the quite high noise level. A valid PWARX model satisfying the given bound is identified by setting $\delta = 0.4$. However, the returned model contains only two modes, representing a quite rough approximation of the

true dynamics (5.28).

Example 5.3 Consider the following LFT system with FIR linear dynamics:

$$y_k = w_k + g_1 w_{k-1} + g_2 w_{k-2} + g_3 w_{k-3} + g_4 w_{k-4} + e_k \quad (5.29a)$$

$$w_k = \begin{cases} \theta_1^T \begin{bmatrix} y_{k-1} \\ u_{k-1} \\ 1 \end{bmatrix} & \text{if } h^T \begin{bmatrix} y_{k-1} \\ u_{k-1} \\ 1 \end{bmatrix} \leq 0 \\ \theta_2^T \begin{bmatrix} y_{k-1} \\ u_{k-1} \\ 1 \end{bmatrix} & \text{if } h^T \begin{bmatrix} y_{k-1} \\ u_{k-1} \\ 1 \end{bmatrix} > 0, \end{cases} \quad (5.29b)$$

where $g_i = \lambda^i$, $i = 1, \dots, 4$, $\lambda = -0.9$, and θ_1 , θ_2 , and h are reported in Table 5.4. This example is challenging, because (5.29a) is a linear regression where the regressors are completely unknown. A data set of $N = 1000$ data points (y_k, u_k) , $k = 1, \dots, N$, is generated with u_k and e_k uniformly distributed in $[-1, 1]$ and $[-0.2, 0.2]$, respectively. To apply the identification algorithm described in Section 5.3, initial estimates g_i^0 , $i = 1, \dots, 4$, are set equal to 0. Least squares are used for estimation of the FIR coefficients, while estimation of the static nonlinearity is carried out by applying the bounded-error procedure described in Section 5.1.1. Figure 5.3 shows how the bound δ is selected at each iteration to fit a suitable PWA map to the samples (v_k^j, z_k) , with $z_k = [y_{k-1} \ u_{k-1}]^T$. At the first iteration, the samples (v_k^1, z_k) do not show a clear partition into distinct affine modes (top left in Figure 5.3), and a large δ (namely, $\delta = 1$) is chosen to cluster the points into two modes (top right). As long as the iterations proceed, the partition of the samples into two distinct affine modes becomes apparent (bottom left), and δ can be reduced. For instance, $\delta = 0.4$ is chosen at

Table 5.4: True and estimated parameters in Example 5.3

θ_1	$\hat{\theta}_1$	θ_2	$\hat{\theta}_2$	h	\hat{h}
0.2	0.2099	-0.5	-0.5014	1	1.0006
-0.8	-0.8316	0.2	0.2106	0	0.0054
-2.2	-2.1680	0	0.0334	1	1

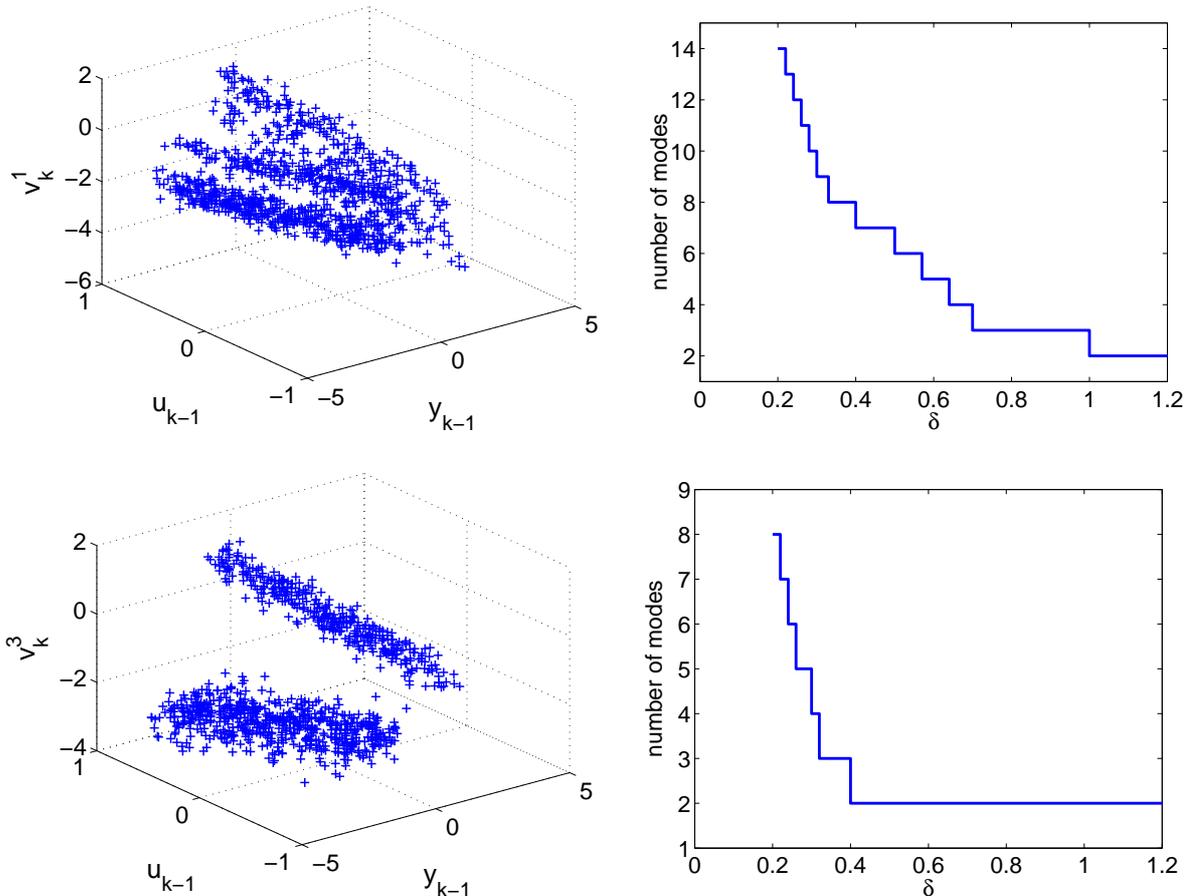


Figure 5.3: Example 5.3: (Top left) Samples (v_k^1, \mathbf{r}_k) at the first iteration. (Top right) Number of modes versus the bound δ when a PWA map is fitted to the samples (v_k^1, \mathbf{r}_k) . (Bottom) The same at the third iteration, i.e. for the samples (v_k^3, \mathbf{r}_k) .

the third iteration (bottom right). The iterative algorithm terminates after 10 iterations for $\gamma = 0.001$. Figure 5.4 shows convergence of the estimates of the FIR coefficients to the corresponding true values. The two modes of the static nonlinearity are also correctly reconstructed, as can be noticed from the estimated parameters $\hat{\theta}_1$, $\hat{\theta}_2$, and \hat{h} reported in Table 5.4.

It is stressed that the equivalent PWARX representation of system (5.29) has up to $2^5 = 32$ modes, corresponding to all possible mode combinations for the 5-tuple (w_k, \dots, w_{k-4}) . In the considered data set, 26 modes out of 32 are visited, and only 8 modes contain a sufficient number of data points to allow for the estimation of the corresponding 11-dimensional parameter vectors. A 8-mode PWARX model is identified by the bounded-error procedure [10]

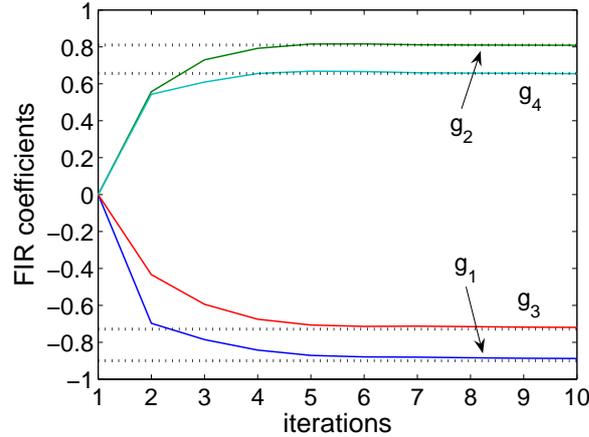


Figure 5.4: Sequence of estimates of the FIR coefficients g_1 , g_2 , g_3 , and g_4 in Example 5.3 (dotted lines are true values).

for $\delta = 0.5$, i.e. for an error significantly larger than the true noise bound.

5.5 Summary

In this chapter an iterative approach for PWA identification of systems described by the LFT interconnection of a linear time-invariant system and a static nonlinearity has been proposed. The nonlinear map needs to be at least piecewise continuous, such that Piecewise Affine(PWA) identification techniques can be applied. By combining the use of classical linear identification tools with recently developed methods for PWA system identification, the proposed approach provides explicit models for both the linear and the nonlinear part of the system. The obtained results are promising, since they show that the identification process can be profitably enhanced by exploiting the knowledge of the interconnection structure of the system. Several open problems are worth to be addressed in the considered identification framework. These are outlined in Section 6.2.

Chapter 6

Contributions and Future Research

In this thesis we have discussed the problem of identifying a discrete time interconnected system represented by a LFT. In the following we summarize our contributions and we outline future research directions and open problems.

6.1 Outline of Contributions

In Chapter 4 we proposed a unified framework for a class of interconnected systems with dynamic linear components and static nonlinear components. For black-box identification, we employed spline approximation for the static nonlinear component. The separable least squares method was used to identify the parameters of the linear dynamical system and to fit a spline to the nonlinear map. For a restricted model set, we provided a sufficient condition on data for the consistency of the parameter estimate. The proposed algorithm is numerically tractable and can be initialized with little prior knowledge. The performance of the algorithm has been demonstrated with numerical examples for identifying a Wiener system, a Wiener-Hammerstein system and a more complex interconnected system. Similar methods have been reported in the literature of separable least squares algorithms for the identification of Wiener, Hammerstein and Hammerstein-Wiener systems (eg. [84], [46]). Our method

provides a generalization of these early works to include more challenging identification tasks, like those of identifying Wiener-Hammerstein systems and general LFT systems.

In Chapter 5 an iterative approach for PWA identification of systems described by the LFT interconnection of a LTI system and a static nonlinearity has been proposed. By combining the use of classical linear identification tools with recently developed methods for PWA system identification, the proposed approach provides explicit models for both the linear and the nonlinear part of the system. We have compared the identification results obtained by exploiting the LFT structure of the system against those obtaining by imposing a black-box framework. The obtained results are promising, since they show that the identification process can be profitably enhanced by exploiting the knowledge of the interconnection structure of the system.

The two methods complement each other. The first method targets to identify systems consisting of rich linear systems interconnected to smooth nonlinear elements. The overall system is thus dominated by the linear part. On the other hand, the second method deals with systems that include richer nonlinear elements, with more complex structure and behaviour. The tradeoff comes in the form of the linear system. Only simple interconnections can be dealt with, like the Hammerstein system. Wiener or Wiener-Hammerstein systems for example can not be considered under this framework as the resulting problem is ill-posed. The two formulations presented in this thesis can be used to model a plethora of systems arising in practical experiments.

6.2 Future Research

There are two theoretically challenging issues identified through the work presented in Chapter 4 for further research. Firstly, the issue of identifiability for nonlinear models of this type needs deeper exploration. Secondly, while the algorithm appears to perform satisfactorily in practice over a large number of simulation examples, the set of inputs which ensure the

consistency of parameter estimates remains to be characterized. From a practical point of view, it will be interesting to consider an application of this methodology in an adaptive identification and control setting, where the computational benefit of separable least squares will be more valuable. Finally, the issues of identifiability and consistency for more general models need deeper exploration and are topics of current research.

Concerning the work presented in Chapter 5, several open problems are worth addressing in the considered identification framework. Relationships with other PWA model structures deserve deeper investigation. The asymptotic properties of the iterative identification algorithm need to be addressed (at least under strong assumptions like the model class matching perfectly the system generating the data). Moreover, the proposed iterative procedure should be extended to the case of multi input-multi output systems and multi-dimensional internal signal w_k . In particular, the latter issue requires the PWA approximation algorithm to deal with vector-valued PWA maps. The considered bounded-error identification procedure can be suitably amended to this aim, as detailed in [60].

Finally, it is worth exploring the relationship between the two system structures presented in this thesis. Intuitively, under proper conditions the system model presented in Chapter 5 forms a subclass of that presented in Chapter 4. However, vigorous mathematical formulation is yet to be achieved.

Bibliography

- [1] H. Al-Duwaish and M. Karim Nazmul. A new method for the identification of Hammerstein model. *Automatica*, 33:1871–1875, 1997.
- [2] E. Amaldi and M. Mattavelli. The MIN PFS problem and piecewise linear model estimation. *Discrete Applied Mathematics*, 118(1-2):115–143, 2002.
- [3] E.W. Bai. An optimal two-stage algorithm for Hammerstein-Wiener 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 Hammerstein models. *IEEE Transactions Automatic Control*, 48:1641–1649, 2003.
- [6] E.W. Bai. Identification of linear systems with hard input nonlinearities of known structure. *Automatica*, 38:853–860, 2004.
- [7] C. Beck. On formal power series representations for uncertain systems. *IEEE Transactions Automatic Control*, 46:314–319, 2001.
- [8] C. Beck and R. D’ Andrea. Noncommuting multidimensional realisation theory: Minimality, reachability and observability. *IEEE Transactions Automatic Control*, 49:1815–1820, 2004.

- [9] G. Belforte and P. Gay. Hammerstein model identification with set-membership errors. In *Proc. of the 38th IEEE Conf. on Decision and Control and the European Control Conference*, pages 592–597, Phoenix, Arizona, USA, December 1999.
- [10] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino. A bounded-error approach to piecewise affine system identification. *IEEE Transactions Automatic Control*, 50(10):1567–1580, 2005.
- [11] C. P. Bottura and G.L. de Oliveira Serra. An algorithm for fuzzy identification of nonlinear discrete-time systems. In *Proc. of the 43rd IEEE Conf. on Decision and Control and the European Control Conference*, pages 5421–5426, Atlantis, Paradise Island, Bahamas, December 2004.
- [12] M. Boutayeb and M. Darouach. Recursive identification method for MISO Wiener-Hammerstein models. *IEEE Transactions Automatic Control*, 40:287–291, 1995.
- [13] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.
- [14] L. Breiman. Hinging hyperplanes for regression, classification, and function approximation. *IEEE Transactions Information Theory*, 39(3):999–1013, 1993.
- [15] J. Doyle C. Beck and K. Glover. Model reduction of multidimensional and uncertain systems. *IEEE Transactions Automatic Control*, 41(10):1466–1477, 1996.
- [16] R. J.G.B. Campello, W.C. do Amaral, and G. Favier. A note on the optimal expansion of Volterra models using laguerre functions. *Automatica*, 42:689–693, 2006.
- [17] K. H. Chan, J. Bao, and W. J. Whiten. Identification of MIMO Hammerstein systems using cardinal spline functions. *Journal Process Control*, 16:659–670, 2006.
- [18] F.H.I. Chang and R. Luus. A non-iterative method for identification of Hammerstein models. *IEEE Transactions Automatic Control*, pages 464–468, 1971.

- [19] H.F. Chen. Strong consistency on recursive identification for Hammerstein systems with discontinuous piece-wise linear memoryless block. *IEEE Transactions Automatic Control*, 50(10):1612–1617, 2005.
- [20] H.F. Chen. Recursive identification for Wiener model with discontinuous piece-wise linear function. *IEEE Transactions Automatic Control*, 51(3):390–400, 2006.
- [21] J. Chen and S. Wang. Validation of linear fractional uncertain models: Solutions via matrix inequalities. *IEEE Transactions Automatic Control*, 41(6):844–849, 1996.
- [22] S. Chen and S.A. Billings. Neural networks for nonlinear dynamic system modelling and identification. *Int. Journal of Control*, 56(2):319–346, 1992.
- [23] S. Chen and S.A. Billings. Neural networks for nonlinear dynamic system modelling and identification. *Int. Journal of Control*, 56(2):319–346, 1992.
- [24] M. S. Claassen. *System identification for structured nonlinear systems*. PhD thesis, Department of Mechanical Engineering, University of California at Berkeley, USA, 2001.
- [25] J. C. Cockburn and B. G. Morton. Linear fractional representations of uncertain systems. *Automatica*, 33(7):1263–1271, 1997.
- [26] J. Dieudonne. *Foundations of Modern Analysis*. Academic Press, New York, 1969.
- [27] F. Ding and T. Chen. Identification of Hammerstein nonlinear ARMAX systems. *Automatica*, 41:1479–1489, 2005.
- [28] F.J. Doyle, R.K. Pearson, and B.A. Ogunnaike. *Identification and Control using Volterra models*. Springer, Berlin, 2002.
- [29] M. Espinoza, J. A. K. Suykens, and B. De Moor. Partially linear models and least squares support vector machines. In *Proc. of the 43rd IEEE Conf. on Decision and Control*, pages 3388–3393, Paradise Island, Bahamas, 2004.

- [30] P. Falugi, L. Giarre, and G. Zappa. Approximation of the feasible parameter set in worst-case identification of Hammerstein models. *Automatica*, 41:1017–1024, 2005.
- [31] G. Ferrari-Trecate and M. Muselli. Single-linkage clustering for optimal classification in piecewise affine regression. In *Proc. of the IFAC Conference on Analysis and Design of Hybrid Systems*, Saint-Malo, France, 2003.
- [32] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari. A clustering technique for the identification of piecewise affine systems. *Automatica*, 39(2):205–217, 2003.
- [33] G. Ferreres. Reduction of dynamic lft systems with lti model uncertainties. *Int. Journal of Robust and Nonlinear Control*, 14:307–323, 2004.
- [34] Y. Gao and M.J. Er. Nonlinear identification and control using a Generalized Fuzzy Neural Network. In *Proc. of the 41st IEEE Conf. on Decision and Control and the European Control Conference*, pages 6418–6423, Las Vegas, USA, December 2002.
- [35] A. Garulli, L. Giarre, and G. Zappa. Identification of approximated Hammerstein models in a worst-case setting. *IEEE Transactions Automatic Control*, 47(12):2046–2050, 2002.
- [36] A. Giannopoulos. *Notes on Functional Analysis*. <http://users.uoa.gr/apgiannop/>, Department of Mathematics, University of Athens, Greece, 2003.
- [37] F. Giri, F.Z. Chaoui, and Y. Rochdi. Parameter identification of a class of Hammerstein plants. *Automatica*, 37:749–756, 2001.
- [38] I. Goethals, K. Pelckmans, J.A.K. Suykens, and B. De Moor. Identification of MIMO Hammerstein models using least squares support vector machines. *Automatica*, 41:1263–1272, 2005.

- [39] G.H. Golub and V. Pereya. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal Numerical Analysis*, 10(2):413–432, 1973.
- [40] W. Greblicki. Stochastic approximation in nonparametric identification of Hammerstein systems. *IEEE Transactions Automatic Control*, 47(11):1800–1810, 2002.
- [41] W. Greblicki and M. Pawlak. Non-parametric identification of Hammerstein systems. *IEEE Transactions Automatic Control*, 35(2):409–418, 1989.
- [42] N.D. Haist, F.H.I. Chang, and R. Luus. Nonlinear identification in the presence of correlated noise using a Hammerstein model. *IEEE Transactions Automatic Control*, pages 552–555, 1973.
- [43] Z. Hasiewicz. Hammerstein system identification by the Haar multiresolution approximation. *Int. Journal of Adaptive Control and Signal Processing*, 13:691–717, 1999.
- [44] Z. Hasiewicz and G. Mzyk. Combined parametric-nonparametric identification of Hammerstein systems. *IEEE Transactions Automatic Control*, 49(8):1370–1375, 2004.
- [45] K. Hsu, T. Vincent, C. Novara, M. Milanese, and K. Poola. Identification of nonlinear maps in interconnected systems. In *Proc. of the 44th IEEE Conf. on Decision and Control and the European Control Conference*, pages 6430–6435, Seville, Spain, 2005.
- [46] M.C. Hughes and D.T. Westwick. Identification of IIR Wiener systems with spline nonlinearities that have variable knots. *IEEE Transactions Automatic Control*, 50(10):1617–1622, 2005.
- [47] G.W. Irwin, K. Warwick, and K.J. Hunt. *Neural Network Applications in Control*. Number 53 in IEE Control Engineering Series. The Institution of Electrical Engineers, London, United Kingdom, 1995.

- [48] A. Packard J. Doyle and K. Zhou. Review of LFTs, LMIs and μ . In *Proc. of the 30th IEEE Conf. on Decision and Control*, pages 1227–1233, 1991.
- [49] S. Jaganatha and F.L. Lewis. Identification of nonlinear dynamical systems using Multilayered Neural Networks. *Automatica*, 32(12):1707–1712, 1995.
- [50] A.H. Kayran and E.M. Eksioğlu. Nonlinear system identification using deterministic multilevel sequences. *Circuits, Systems and Signal Processing*, 24(2):151–181, 2005.
- [51] P. Lancaster and K. Salkauskas. *Curve and Surface Fitting: An Introduction*. UK: Academic, London, 1986.
- [52] D.J. Leith, W.E. Leithead, and R. Murray-Smith. Inference of disjoint linear and non-linear sub-domains of a non-linear mapping. *Automatica*, 42:849–858, 2006.
- [53] J.-N. Lin and R. Unbehauen. Canonical piecewise-linear approximations. *IEEE Transactions Circuits and Systems – I: Fundamental Theory and Applications*, 39(8):697–699, 1992.
- [54] A. Lj. Juloski, S. Wieland, and W. P. M. H. Heemels. A Bayesian approach to identification of hybrid systems. *IEEE Transactions Automatic Control*, 50(10):1520–1533, 2005.
- [55] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Upper Saddle River, NJ, 1999.
- [56] J. Neter M. H. Kutner, C. J. Nachtsheim and W. Li. *Applied Linear Statistical Models*. McGraw-Hill Irwin, New York, fifth edition, 2005.
- [57] E. Münz and V. Krebs. Identification of hybrid systems using apriori knowledge. In *Proc. of the 15th IFAC World Congress*, Barcelona, Spain, 2002.

- [58] H. Nakada, K. Takaba, and T. Katayama. Identification of piecewise affine systems based on statistical clustering technique. *Automatica*, 41(5):905–913, 2005.
- [59] K. S. Narendra and P. G. Gallman. An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Transactions Automatic Control*, 11(3):546–550, 1966.
- [60] S. Paoletti. *Identification of piecewise affine models*. PhD thesis, Department of Information Engineering, University of Siena, Siena, Italy, 2004.
- [61] S. Paoletti, A. Garulli, E. Pepona, and P. Date. Exploiting structure in piecewise affine identification of LFT systems. In *Proc. of the 15th IFAC Symposium on System Identification (To appear)*, 2009.
- [62] S. Paoletti, A. Lj. Juloski, G. Ferrari-Trecate, and R. Vidal. Identification of hybrid systems: A tutorial. *European Journal of Control*, 13(2-3):242–260, 2007.
- [63] M. Pawlak. On the series-expansion approach to the identification of Hammerstein systems. *IEEE Transactions Automatic Control*, 36(6):763–767, 1992.
- [64] E. Pepona and P. Date. Separable least squares identification of a class of nonlinear systems with static nonlinearities. *Int. Journal of Adaptive Control and Signal Processing (Under Review)*, 2009.
- [65] E. Pepona, S. Paoletti, A. Garulli, and P. Date. An iterative procedure for piecewise affine identification of nonlinear interconnected systems. In *Proc. of the 46th IEEE Conf. on Decision and Control*, pages 5098–5103, New Orleans, USA, December 2007.
- [66] E. Pepona, S. Paoletti, A. Garulli, and P. Date. Piecewise affine identification of nonlinear interconnected systems. *IEEE Transactions Control Systems Technology (Under Review)*, 2009.

- [67] A.S. Poznyak and E.N. Sanchez. Nonlinear system identification and trajectory tracking using Dynamic Neural Networks. In *Proc. of the 35th IEEE Conf. on Decision and Control and the European Control Conference*, pages 955–960, December 1996. Kobe, Japan.
- [68] F. Previdi and M. Lovera. Identification of a class of non-linear parametrically varying models. *Int. Journal of Adaptive Control and Signal Processing*, 17(1):33–50, 2003.
- [69] J. Ragot, G. Mourot, and D. Maquin. Parameter estimation of switching piecewise linear systems. In *Proc. of the 42nd IEEE Conf. on Decision and Control*, pages 5783–5788, Maui, HI, 2003.
- [70] S. Rangan, G. Wolodkin, and K. Poola. New results for Hammerstein system identification. In *Proc. of the 34th IEEE Conference on Decision and Control*, pages 697–702, New Orleans, LA, 1995.
- [71] J. Roll. *Local and piecewise affine approaches to system identification*. PhD thesis, Department of Electrical Engineering, Linköping University, Linköping, Sweden, 2003.
- [72] J. Roll, A. Bemporad, and L. Ljung. Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40(1):37–50, 2004.
- [73] J. Roll, A. Nazin, and L. Ljung. Nonlinear system identification via direct weight optimization. *Automatica*, 41:475–490, 2005.
- [74] P. Samuelsson. An integrating linearization method for Hammerstein models. *Automatica*, 41:1825–1828, 2005.
- [75] I.W. Sandberg. Series expansions for nonlinear systems. *Circuits, Systems and Signal Processing*, 2(1):77–87, 1983.
- [76] M. Schetzen. *The Volterra and Wiener theories of nonlinear systems*. Wiley, New York, 1980.

- [77] J. Sjöberg and M. Viberg. Separable non-linear least squares minimization - possible improvements for neural net fitting. *IEEE Workshop Neural Network and Signal Processing*, 7:345–354, 1997.
- [78] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P-Y. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: A unified overview. *Automatica*, 31(12):1691–1724, 1995.
- [79] S. Sristava. Modelling of non-linear systems by FWNNs and their intelligent control. *Int. Journal of Adapted Control and Signal Processing*, 19:505–530, 2005.
- [80] M. Tanaka. Identification of nonlinear systems with missing data using Stochastic Neural Networks. In *Proc. of the 35th IEEE Conf. on Decision and Control and the European Control Conference*, pages 933–934, December 1996. Kobe, Japan.
- [81] G. Vandersteen and J. Schoukens. Measurement and identification of nonlinear systems consisting of linear dynamic blocks and one static nonlinearity. *IEEE Transactions Automatic Control*, 44(6):1266–1231, 1999.
- [82] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, PA, 1990.
- [83] M.T. Wasan. *Stochastic Approximation*. Cambridge University Press, Cambridge, UK, 1969.
- [84] D.T. Westwick and R.E. Kearney. Separable least squares identification of nonlinear Hammerstein models: Application to stretch reflex dynamics. *Annals of Biomedical Engineering*, 29:707–718, 2001.
- [85] C.F. Wu. Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, 9(3):501–513, 1981.
- [86] F. Wu and K. Dong. Gain scheduling control of lft systems using parameter-dependent lyapunov functions. *Automatica*, 42:39–50, 2006.

- [87] D. Xu Z. Ren, G. Gu and J. Chen. LFT uncertain model validation with time- and frequency-domain measurements. *IEEE Transactions Automatic Control*, 44(7):1435–1442, 1999.
- [88] K. Zhou and J.C. Doyle. *Essentials of Robust Control*. Prentice Hall, Upper Saddle River, New Jersey, 1998.
- [89] Y. Zhu. Estimation of an N-L-N Hammerstein-Wiener model. *Automatica*, 38:1607–1614, 2002.
- [90] L. Zi-Qiang. On the identification of controlled plants described by the Hammerstein model. *IEEE Transactions Automatic Control*, 39(3):569–573, 1994.
- [91] L. Zi-Qiang. A nonparametric polynomial identification algorithm for the Hammerstein system. *IEEE Transactions Automatic Control*, 42(10):1435–1441, 1997.