UNIVERSITY OF CALIFORNIA, SAN DIEGO

# Block-oriented Nonlinear System Identification Using Semidefinite Programming

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Engineering Sciences (Mechanical Engineering)

by

Younghee Han

Committee in charge:

Professor Raymond de Callafon, Chair
Professor Robert Bitmead
Professor Mauricio de Oliveira
Professor Philip Gill
Professor Michael Todd

2012

The dissertation of Younghee Han is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2012

DEDICATION

To my parents and my husband

EPIGRAPH

*Damn straight! Today, the mad scientist can't get a doomsday device, tomorrow it's the mad grad student. Where will it end?*

—Professor Hubert J. Farnsworth

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my thanks to my advisor Prof. Raymond de Callafon for his guidance, advice, and patience throughout my graduate school career. Next I wish to thank my Doctoral Committee: Prof. Robert Bitmead, Prof. Mauricio de Oliveira, Prof. Philip Gill and Prof. Michael Todd for their advice, suggestions, and review of this thesis. I would like to thank the faculty, staff, secretaries, and graduate students of the Mechanical Engineering Department. I also want to express my thanks to my colleagues in the System Identification and Control Laboratory for their help and support. I wish to thank my parents for all their support and encouragement throughout my graduate school career. Finally, I would like to thank my husband, Gregory Bishop for all of his encouragement, support, and love.

Most of the material in this dissertation has been published or accepted for publication. The results in Chapter 3 have been accepted for publication in

Y. Han and R. de Callafon, Hammerstein system identification using nuclear norm minimization, Automatica, to appear 2012.

The results in Chapter 4 have been accepted for publication in

Y. Han and R. de Callafon, 2012, Identification of a Wiener System via Semidefinite Programming, 16th IFAC Symposium on System Identification 2012, Brussels, Belgium.

The results in Chapter 5 have been accepted for publication in

Y. Han and R. de Callafon, Identification of Wiener-Hammerstein Benchmark Model using Convex Optimization, Control Engineering Practice Special Issue, to appear 2012.

The results in Chapter 6 have been published in

Y. Han and R. de Callafon, Output Error Identification of Closed-loop Hammerstein Systems, IEEE Conference on Decision and Control 2011, Orlando, US, and

Y. Han and R. de Callafon, Closed-loop Identification of Hammerstein Systems Using Iterative Instrumental Variables, IFAC World Congress 2011, Milano, Italy.

## VITA

| | |
|---|---|
| 2001 | Bachelor of Science in Mechanical Engineering, Ajou University, South Korea |
| 2005 | Master of Science in Mechanical Engineering, University of Kentucky |
| 2006 | Graduate Teaching Assistant, University of Kentucky |
| 2008 | Summer Intern, Samsung Electronics |
| 2010 | NREIP Intern, Space and Naval Warfare Systems Center, San Diego |
| 2011 | NREIP Intern, Space and Naval Warfare Systems Center, San Diego |
| 2008-20012 | Graduate Student Researcher/Teaching Assistant, University of California, San Diego |
| 2012 | Doctor of Philosophy in Engineering Sciences (Mechanical Engineering), University of California, San Diego |

## PUBLICATIONS

Y. Han and R. de Callafon, Identification of Wiener-Hammerstein Benchmark Model using Convex Optimization, *Control Engineering Practice Special Issue*, to appear 2012.

Y. Han and R. de Callafon, Hammerstein system identification using nuclear norm minimization, *Automatica*, to appear 2012.

Y. Han and R. de Callafon, Evaluating Track-Following Servo Performance of High-Density Hard Disk Drives Using Patterned Media, *IEEE Transactions on Magnetics*, Vol. 45, Issue 12, 5352 5359, 2009.

J.G. DeHaven, Y. Han, H.S. Tzou, Transition of Membrane/Bending Neural Signals on Transforming Adaptive Shells, *Journal of Vibration and Control*, Vol. 13, No. 7, 1007-1029, 2007.

W.K. Chai, Y. Han, K. Higuchi, H.S. Tzou, Micro-actuation characteristic of rocket conical shell sections, *Journal of Sound and Vibration*, Vol. 293, 1-2, 286-298, 2006.

Y. Han and R. de Callafon, 2012, Identification of a Wiener System via Semidefinite Programming, *16th IFAC Symposium on System Identification* 2012, Brussels, Belgium.

Y. Han and R. de Callafon, Output Error Identification of Closed-loop Hammerstein Systems, *IEEE Conference on Decision and Control* 2011, Orlando, US.

Y. Han and R. de Callafon, Closed-loop Identification of Hammerstein Systems Using Iterative Instrumental Variables, *IFAC World Congress* 2011, Milano, Italy.

Y. Han and R. de Callafon, J. Jaffe, Dynamic Modeling and Pneumatic Switching Control of a Submersible Drogue, *Int. Conference on Informatics in Control, Automation and Robotics* 2010, Madeira, Portugal.

Y. Han and R. de Callafon, Evaluation of Track Following Performance for Patterned Servo Sectors in Hard Disk Drives, *IEEE Conference on Decision and Control* 2009, Shanghai, China.

Y. Han and R. de Callafon, Evaluation of Track Following Servo Performance of High Density Hard Disk Drives Using Bit-Patterned Media, *The Magnetic Recording Conference* 2008, Singapore.

Y. Han, H.S. Tzou, Transition of neural signals on adaptive cylindrical paraboloidal shells, *ASME 2006 International Mechanical Engineering Congress and Exposition*, Chicago, Illinois, US.

Y. Han and H.S. Tzou, Nano- and micro-actuations of hard disks head positioning system: Actuator Design and Performance Evaluation, *ASME 2006 International Mechanical Engineering Congress and Exposition*, Chicago, Illinois, US.

Y. Han and H.S. Tzou, Nano- and micro-actuations of hard disks head positioning system: Precision Control and Evaluation, *ASME 2006 International Mechanical Engineering Congress and Exposition*, Chicago, Illinois, US.

J.G. DeHaven, Y. Han and H.S. Tzou, Transition of Neural Signal on Cylindrical Shells with Various Curvatures, *ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Long Beach, California, US.

ABSTRACT OF THE DISSERTATION

**Block-oriented Nonlinear System Identification Using Semidefinite Programming**

by

Younghee Han

Doctor of Philosophy in Engineering Sciences (Mechanical Engineering)

University of California, San Diego, 2012

Professor Raymond de Callafon, Chair

Identification of block-oriented nonlinear systems has been an active research area for the last several decades. A block-oriented nonlinear system represents a nonlinear dynamical system as a combination of linear dynamic systems and static nonlinear blocks. In block-oriented nonlinear systems, each block (linear dynamic systems and static nonlinearity) can be connected in many different ways (series, parallel, feedback) and this flexibility provides the block-oriented modeling approach with an ability to capture a large class of nonlinear systems. However, intermediate signals in such block-oriented systems are not measurable and the inaccessibility of such measurements is the main difficulty in block-oriented nonlinear system identification.

Recently a system identification method using rank minimization has been introduced for linear system identification. Finding the simplest model within a feasible model set restricted by convex constraints can often be formulated as a rank minimization problem. In this research, the rank minimization approach is extended to block-oriented nonlinear system identification. The system parameter estimation problem is formulated as a rank minimization problem or the combination of prediction error and rank minimization problems by constraining a finite dimensional time dependency of a linear dynamic system and by using the monotonicity of static nonlinearity. This allows us to reconstruct non-measurable intermediate signals and once the intermediate signals have been reconstructed, the identification of each block can be solved with the standard Prediction Error method or Least Squares method.

The research work presented in this dissertation proposes a new approach for block-oriented system identification by tackling the inaccessibility of measurement of intermediate signals in block-oriented nonlinear systems via rank minimization. Since the rank minimization problem is non-convex, the rank minimization problem is relaxed to a semidefinite programming problem by minimizing the nuclear norm instead of the rank. The research contributes to advances in block-oriented nonlinear system identification.

# 1

# Introduction to block-oriented nonlinear system identification

## 1.1 Block-oriented nonlinear system identification

System identification has become increasingly essential in all branches of engineering and in other disciplines. When conducting system identification, linear models are usually considered first because solutions for identification and modeling of linear dynamical systems are well known. A linear model can be adequate if the nonlinearity of the system is not severe or the real system is operated on a limited operating range so that the linearity assumption is satisfied. However, only limited types of systems can be approximated by linear systems. If the nonlinearity of the system is severe or a wide operating range is considered, the linear assumption may not be valid and a nonlinear model becomes necessary to capture the nonlinearity of the system. Block-oriented models, a specific class of nonlinear systems, have simple structures yet provide much better approximations than linear models for nonlinear dynamic systems with specific purposes. This dissertation focuses on identification of block-oriented nonlinear systems that consist of linear dynamic systems and static nonlinear blocks. Only Single Input and Single Output (SISO) systems will be considered in this dissertation.

The output of a dynamic system depends on the past input and past output. On the other hand, the output of a static nonlinearity depends only on the current input. Common static nonlinear functions are saturation, deadzone, etc. Static nonlinearity is input dependent. Only the range that is excited by an input can be identified using system identification. Therefore, an input has to be chosen wisely for a specific identification purpose during experimental design. In block-oriented nonlinear systems, each block (linear dynamic systems and static nonlinearity) can be connected in many different ways (series, parallel, feedback) and this flexibility provides the block-oriented modeling approach with an ability to capture a large class of nonlinear systems. However, intermediate signals in such block-oriented systems are not measurable and the inaccessibility of such measurements is the main difficulty in block-oriented nonlinear system identification [29].

### 1.1.1 Hammerstein system identification

The simplest block-oriented nonlinear systems are Hammerstein and Wiener systems. In these systems, only a single static nonlinear block and a single linear dynamic block are used. A Hammerstein system is a series connection of a static nonlinearity followed by a linear dynamic system, as shown in Figure 1.1.

$$
\begin{aligned}
y(t) &= G(q)x(t) + v(t) \\
x(t) &= f(u(t)).
\end{aligned}
$$

A Hammerstein structure is a good approach to modeling systems with actuator nonlinearity or other nonlinear effects that can be brought to the system input [29]. Hammerstein models have been successfully used to represent the nonlin-

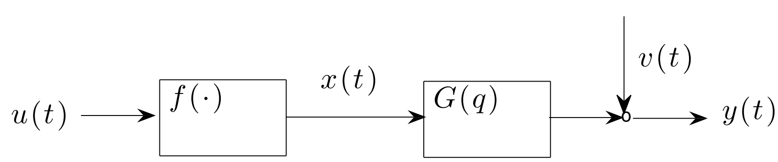



**Figure 1.1**: Hammerstein system consists of a static nonlinear block followed by a linear dynamic block.

ear dynamics of various systems. For example, the nonlinear dynamics of various chemical processes, electrical systems, biological systems, thermal systems, etc. have been modeled with the Hammerstein model [5] [21] [23] [29] [38] [42] [43] [44] [58] [66]. Usually, Hammerstein systems are parametrized linear in both the input static nonlinearity and the linear dynamic system, so the parameter estimation is reduced to an ordinary least squares (LS) technique or any of its improved versions [1]. Early research on the identification of Hammerstein systems can be found in [53] where an iterative identification method was proposed for Hammerstein systems utilizing the alternate adjustment of the parameters of the linear and nonlinear parts of the systems. Many authors have tackled the identification of Hammerstein systems. A non-iterative method to estimate the parameters by minimizing the equation error was proposed in [14]. Due to the equation error minimization, this method is sensitive to noise and may result in biased estimation for colored noise. An iterative technique for the estimation of parameters in a Hammerstein model was developed in [36] to deal with colored noise. Modified formulations of the generalized least squares (GLS) estimation algorithm for system parameter identification was presented in [37] to deal with the biased estimation produced by the least squares (LS) method, but there is no guarantee that the procedure converges to the optimal solution. [32] proposed an algorithm using a nonparametric kernel estimate of regression functions calculated from dependent data. A two-step identification method of the LS parameter estimation based on correlation functions was introduced in [34]. The authors in [3] discussed discrete Hammerstein model identification using a blind system identification approach. An optimal two-stage identification for Hammerstein-Wiener systems was presented in [4] and the authors in [66] revisited an optimality result established in [4] showing that the two-stage algorithm (TSA) provides the optimal estimation of a bilinearly parameterized Hammerstein system in the sense of a weighted nonlinear LS criterion formulated with a special weighting matrix. However, these methods suffer from an over-parametrization problem, which requires parameter separation via a singular value decomposition (SVD).

### 1.1.2 Wiener system identification

On the other hand, a Wiener system has a block oriented structure in which a linear dynamical system is followed by a static output nonlinearity, as shown in Figure 1.2.

$$x(t) = G(q)u(t)$$
$$y(t) = f(x(t)) + v(t).$$

A Wiener structure is a good approach to modeling systems with sensor nonlin-

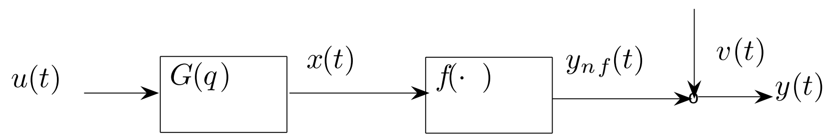



**Figure 1.2**: Wiener system consists of a linear dynamic block followed by a static nonlinear block.

earity or other nonlinear effects that can be brought to the system output [25] [29] [35] [67]. It has been shown that Wiener models can be used to effectively capture various nonlinear dynamics, such as chemical processes and biological systems [10] [29] [33]. The identification of Wiener systems involves estimating the parameters describing the linear dynamical and the output static nonlinear blocks from the measured input and output data. The most common assumptions used in Wiener system identification are the Gaussian assumption of the input signal and the invertibility of the static nonlinearity [7] [8] [22] [31]. These assumptions are popular because, if the input signal is Gaussian noise, the identification of the linear dynamical block can be separated from the identification of the static nonlinear function based on separability assumption and parameterization of the output static nonlinearity is possible for the inverse of the given static nonlinearity. However, the Gaussian input assumption is too restrictive for practical application and the invertibility of the static nonlinearity assumption excludes hard nonlinearities, such as saturation, common in control systems.

### 1.1.3 Wiener-Hammerstein system identification

More complicated block-oriented systems can be created by using more than a single linear dynamic and static nonlinear blocks. A Wiener-Hammerstein system is a dynamical system characterized by a series connection of three parts: a linear dynamical system, a static nonlinearity, and another linear dynamical system, as shown in Figure 1.3. Wiener-Hammerstein models can be used to represent sensor systems, electro-mechanical systems in robotics, mechatronics, biological, and chemical systems [29] [39].

$$
\begin{aligned}
x_1(t) &= G_1(q)u(t) \\
x_2(t) &= f(x_1(t)) \\
y(t) &= G_2(q)x_2(t) + v(t).
\end{aligned}
$$

Early works on Wiener-Hammerstein system identification can be found in [7]

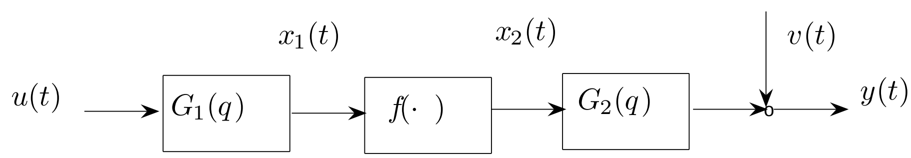



**Figure 1.3**: Wiener-Hammerstein system consisting of the cascade of a linear dynamic block $G_1(q)$, a static non-linear block $f(\cdot)$ and a linear dynamic block $G_2(q)$.

[8]. In this early research, the correlation analysis-based identification method under Gaussian excitation has been proposed. The authors in [15] introduced a time-domain identification method based on the Maximum Likelihood principle. The authors in [12] presented a simple technique for recursive identification of the Wiener-Hammerstein model with extension to the multi-input single-output (MISO) case. More recent work can be found in [2] [17] [22] [52] [54] [57] [65]. The authors in [54] proposed an identification method using the polynomial non-linear state space (PNLSS) approach. The authors in [2] presented a method iteratively identifying the linear system and the Hammerstein system by minimizing the square norm of output prediction error and by using the orthogonal decomposition subspace method (ORT).

### 1.1.4 Hammerstein-Wiener system identification

On the other hand, a Hammerstein-Wiener system is characterized by a series connection of a static nonlinear block, a linear dynamic block, and another static nonlinear block, as shown in Figure 1.3. A HammersteinWiener model is an efficient way of modeling nonlinear dynamic systems in which both actuator and sensor nonlinearities are present. Many authors have studied the identification of Hammerstein-Wiener systems [4] [18] [64]. Hammerstein-Wiener models have been successfully applied to modeling of electro-mechanical systems, chemical processes, radio frequency components, and audio and speech processing [20] [40] [48]. In [4], Bai proposed an optimal two stage identification algorithm using the recursive least squares and the singular value decomposition.

$$
\begin{aligned}
x_1(t) &= f_1(u(t)) \\
x_2(t) &= G(q)x_1(t) \\
y(t) &= f_2(x_2(t)) + v(t).
\end{aligned}
$$

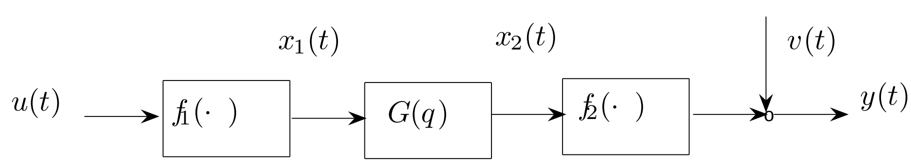



**Figure 1.4**: Hammerstein-Wiener system consisting of the cascade of a static non-linear block $f_1(\cdot)$, a linear dynamic block $G_1(q)$, and a static non-linear block $f_2(\cdot)$.

### 1.1.5 Closed-loop Hammerstein system identification

A feedback connection of a block-oriented system and a controller can be used to model a closed-loop nonlinear dynamic system. Closed-loop nonlinear dynamic system identification techniques are useful for control relevant identification. From the control design point of view, the use of data gathered from closed-loop experiments provides advantages for designing control systems to satisfy typical control performance requirements, such as stability. For example, a feedback connection of a Hammerstein system and a controller provides a good approach to modeling systems with static actuator nonlinearity or input saturation during closed-loop experiments as shown in Figure 1.5.

$$\begin{aligned} u(t) &= r(t) - C(q)y(t) \\ x(t) &= f(u(t)) \\ y(t) &= G(q)x(t) + v(t). \end{aligned}$$

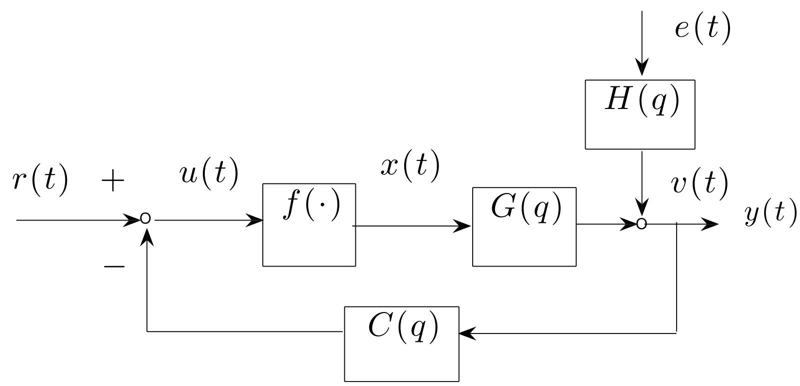



**Figure 1.5**: Closed-loop Hammerstein system.

One of the early works dealing with closed-loop Hammerstein system identification can be found in [6]. In this work, Beyer et al. proposed a closed-loop identification method for Hammerstein systems using the LS method, the GLS method and the maximum likelihood method. In addition, Linard et al. [46] extended closed-loop identification methods (a two-stage method and using right coprime factorizations) for linear dynamic systems to nonlinear dynamic systems and De Bruyne et al. [19] presented gradient expressions for a closed-loop parametric identification scheme. However, these methods are based on linearization of a nonlinear map between time domain signals. Recently, van Wingerden and Verhaegen [62] presented an algorithm to identify MIMO Hammerstein systems under open and closed-loop conditions. They formulated an optimized predictor based subspace identification algorithm in the dual space. Laurain et al. [45] presented an IV method dedicated to closed-loop Hammerstein systems.

## 1.2 Motivation for a new rank minimization approach and problem statement

### 1.2.1 Introduction to the rank minimization approach

Intermediate signals in block-oriented systems are not measurable and the inaccessibility of such measurements is the main difficulty in block-oriented nonlinear system identification [29]. The research work presented in this dissertation proposes a new approach for block-oriented system identification by tackling the inaccessibility of measurement of intermediate signals in block-oriented nonlinear systems via rank minimization. In system identification, the prediction error minimization method is the most commonly used approach to formulate a system parameter estimation problem. However, the prediction error minimization approach commonly results in nonlinear cost functions when it is used in block-oriented system identification due to the non-convexity of the error. As an alternative, the rank minimization approach is proposed in this study. Based on the complexity of a system, the system parameter estimation problem can be formulated as a rank minimization problem or as the combination of prediction error and rank minimization problems by constraining a finite dimensional time dependency between signals. This allows us to reconstruct non-measurable intermediate signals and, once the intermediate signals have been reconstructed, the identification of each block can be solved with either the standard Prediction Error method or Least Squares method. Since the rank minimization problem is non-convex, the rank minimization problem is relaxed to a convex problem by minimizing the nuclear norm instead of the rank. Through this process, a parameter estimation problem for a block-oriented model can be formulated as a Semidefinite Programming (SDP) problem.

### 1.2.2 Problem statement

This study focuses on identification of nonlinear dynamic systems using block-oriented models. In this dissertation, only linear dynamic systems and static nonlinear functions will be considered as subsystems of a block-oriented representation of a nonlinear dynamic system. In order to define a system identification problem using a block-oriented structure, the finite time dependency of a linear dynamic system and the static relationship of a nonlinear function will be considered as follows:

I. The static nonlinear function has no memory:
The current output from the static nonlinear function only
depends on the current input to the static nonlinear function.

II. The linear dynamical system has a finite, but unknown,
McMillan degree $n$:
$y(t) = \phi^T(t)\theta$, where
$\phi^T(t) = [u(t) \ \cdots \ u(t-n_b) \ y(t-1) \ \cdots \ y(t-n_a)]$,
$\theta$ is the linear system parameter,
and $n \leq max(n_b - 1, n_a)$.

The order of a finite dimensional model can be expressed as the rank of a matrix that is filled with input and output measurement. Finding the simplest model within a feasible model set restricted by convex constraints can often be formulated as a rank minimization problem [27]. Based on this idea, the rank minimization approach is used to formulate a convex parameter estimation problem via nuclear norm relaxation, where the nuclear norm of $H$ is defined as the summation of its singular values as $||H||_* = \sum_{i=1}^{r} \sigma_i(H)$.

## 1.3 Contributions

The main objective of this dissertation is to introduce a new approach to the modeling of block-oriented nonlinear systems. The new approach for block-oriented system identification involves tackling the inaccessibility of measurement of intermediate signals in block-oriented nonlinear systems via rank minimization. Using nuclear norm relaxation, the rank minimization problem is converted to a semi-definite programming problem.

Most of the material in this dissertation has been published or accepted for publication. The results in Chapter 3 have been accepted for publication in

Y. Han and R. de Callafon, Hammerstein system identification using nuclear norm minimization, Automatica, to appear 2012.

The results in Chapter 4 have been accepted for publication in

Y. Han and R. de Callafon, 2012, Identification of a Wiener System via Semidefinite Programming, 16th IFAC Symposium on System Identification 2012, Brussels, Belgium.

The results in Chapter 5 have been accepted for publication in

Y. Han and R. de Callafon, Identification of Wiener-Hammerstein Benchmark Model using Convex Optimization, Control Engineering Practice Special Issue, to appear 2012.

The results in Chapter 6 have been published in

Y. Han and R. de Callafon, Output Error Identification of Closed-loop Hammerstein Systems, IEEE Conference on Decision and Control 2011, Orlando, US,

and

Y. Han and R. de Callafon, Closed-loop Identification of Hammerstein Systems Using Iterative Instrumental Variables, IFAC World Congress 2011, Milano, Italy.

## 1.4 Outline of this work

The dissertation is organized as follows. Chapter 1 contains a literature review of block-oriented nonlinear system identification and introduces a new rank minimization approach to solving block-oriented nonlinear system identification. Chapter 2 introduces preliminary concepts regarding a new approach to block-oriented nonlinear system identification. Chapter 3 deals with open-loop identification of Hammerstein systems. The parameter estimation problem is formulated as a rank minimization problem by constraining a finite dimensional time dependency between signals and by using subspace interpretation and the LQ decomposition of the data matrix. Then, the rank minimization is relaxed to a convex optimization problem using a nuclear norm. Chapter 4 deals with open-loop identification of Wiener Systems. A new method is proposed for identifying Wiener systems with monotonically non-decreasing nonlinearity. The identification problem is formulated as a convex semidefinite programming (SDP) problem by constraining a finite dimensional time dependency between signals. The proposed method is robust to output noise and neither the Gaussian assumption of the input signal nor the invertibility of the static nonlinearity is necessary. Chapter 5 deals with open-loop identification of Wiener-Hammerstein systems. The identification of Wiener-Hammerstein systems is formulated as a non-convex rank minimization problem by using the over-parameterization technique. The non-convex rank minimization problem is then reformulated as a convex optimization problem using trace minimization. An iterative approach is proposed to update two unmeasurable intermediate signals. Chapter 6 deals with closed-loop identification of Hammerstein systems. An iterative nuclear norm minimization algorithm is proposed for the OE minimization problem of closed-loop Hammerstein systems. The basic idea is to express the nonlinear parameter estimation as an iterative nuclear norm minimization problem using gradient-based updates. Chapter 7 summarizes the conclusions of this dissertation.

# 2

# Preliminaries

Preliminary material about linear dynamic systems, static nonlinearity and semidefinite programming is presented in this chapter. A modeling approach to linear dynamic systems and static nonlinearity is presented. The rank minimization approach and nuclear norm relaxation are explained.

## 2.1 Linear dynamic systems

Let $g_k,\ k = 0, 1, \cdots$ be a causal sequence of impulse responses of $G(q)$. The relationship between the input $u(t)$ and the output $y(t)$ can be described by the convolution as

$$y(t) = \sum_{k=0}^{\infty} g_k u(t-k). \tag{2.1}$$

In this study, we will approximate the relationship in (2.1) by a finite dimensional system.

$$\hat{y}(t) = \sum_{k=0}^{L-1} \hat{g}_k u(t-k) \tag{2.2}$$

where the finite order sequence of $\hat{g}_k,\ k = 0, 1, \cdots, N-1$ is the impulse response of the approximated system. This finite impulse response (FIR) approximation of $G(q)$ will be used to formulate a convex optimization problem to estimate system parameters. Based on (2.1) and (2.2), the error is defined by

$$\begin{aligned} e(t) &= y(t) - \hat{y}(t) \\ &= y(t) - \sum_{k=0}^{L-1} \hat{g}_k u(t-k). \end{aligned}$$

Thus,

$$\begin{aligned} ||e(t)||_2^2 &= \sum_{t=1}^{N} \left[ y(t) - \sum_{k=0}^{L-1} \hat{g}_k u(t-k) \right]^2 \\ &= \sum_{t=1}^{N} y^2(t) - 2 \sum_{k=0}^{L-1} \hat{g}_k R_{yu}(k) \\ &\quad + \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} \hat{g}_k \hat{g}_l R_{uu}(k-l) \end{aligned}$$

where

$$\begin{aligned} R_{yu}(k) &= \sum_{t=1}^{N} y(t)u(t-k) \\ R_{uu}(k) &= \sum_{t=1-k-l}^{N-k-l} u(t)u(t+k). \end{aligned}$$

If $L$ tends toward infinity, $\hat{g}_k$ will satisfy $|\hat{g}_k| << 1$ for $k \geq L$, resulting in $R_{x_1 u}(k) \rightarrow \sum_{l=0}^{L-1} \hat{g}_l R_{uu}(l-k)$. Then,

$$\lim_{N \rightarrow \infty, L \rightarrow \infty} ||e(t)||_2^2 = 0$$

As a result, the estimate $\hat{y}(t)$ will converge to $y(t)$ provided that $N \rightarrow \infty$ and $L \rightarrow \infty$.

## 2.2 Static nonlinearity

In a static nonlinear function, the current output only depends on the current input. It is well known that the static nonlinear function can be approximated as a linear combination of a finite set of basis functions as

$$f(u(t)) \approx \hat{f}(u(t)) = \sum_{m=1}^{M} \lambda_m \xi_m(u(t)) \tag{2.3}$$

where $\lambda_m$ are weighting parameters to be estimated and $\xi_m(\cdot)$ are basis functions. Modeling nonlinearity linearly in the parameters as in (2.3) results in a linear least squares problem for the static nonlinear system parameter estimation. Users can choose any basis functions suitable for their purpose, such as polynomial and radial basis functions. In this dissertation, polynomial approximation and piecewise linear approximation using triangle basis functions are used.

### 2.2.1 Polynomial approximation

A polynomial function is often used to approximate a static nonlinear function due to its simplicity. With the polynomial basis functions, a $n_f^{th}$ order polynomial function $\hat{f}(u(t))$ is defined by

$$\hat{f}(u(t)) = \lambda_0 + \lambda_1 u(t) + \lambda_2 u^2(t) + \cdots + \lambda_{n_f} u^{n_f}(t). \tag{2.4}$$

Weierstrass's Theorem below guarantees that the polynomial approximation $\hat{f}(u(t))$ in (2.4) will converge to $f(u(t))$ as $n_f$ tends toward to infinity for an arbitrary interval.

**Weierstrass's Theorem** If $f(u(t))$ is a given continuous function for an arbitrary interval $i \leq u(t) \leq j$, and $\epsilon$ is a small magnitude positive constant, there is a polynomial $\hat{f}(u(t))$ such that

$$|f(u(t)) - \hat{f}(u(t))| < \epsilon \ \forall x \in [i, j].$$

A disadvantage of polynomial approximation is that high-degree polynomials have oscillatory behavior and parameter estimation is often numerically ill-conditioned [41] [69].

### 2.2.2 Piecewise linear approximation

A piecewise linear approximation of the static nonlinearity $f(\cdot)$ using piecewise triangle functions is shown in Figure 2.1. Using triangle basis functions $\xi_m(\cdot)$, the static nonlinearity $f(\cdot)$ is assumed to satisfy the following condition

$$\sup_{u(t) \in [u_{min}, u_{max}]} \lim_{M \to \infty} \sum_{m=1}^{M} |\lambda_m \xi_m(u(t)) - f(u(t))| = 0 \tag{2.5}$$

where the center location vector $m = [m_1 \ \cdots \ m_M]^T$, specifying the center locations of triangle basis functions, spans the amplitude of the input vector $u = [u(1) \ \cdots \ u(N)]^T$ and the amplitude vector $\lambda = [\lambda_1 \ \cdots \ \lambda_M]^T$, specifying the amplitudes of triangle basis functions at the center $m$, is to be estimated. The condition in (2.5) indicates that the static nonlinearity $f(\cdot)$ can be approximated arbitrarily well with a dense grid of triangular basis functions. In order to define a piecewise

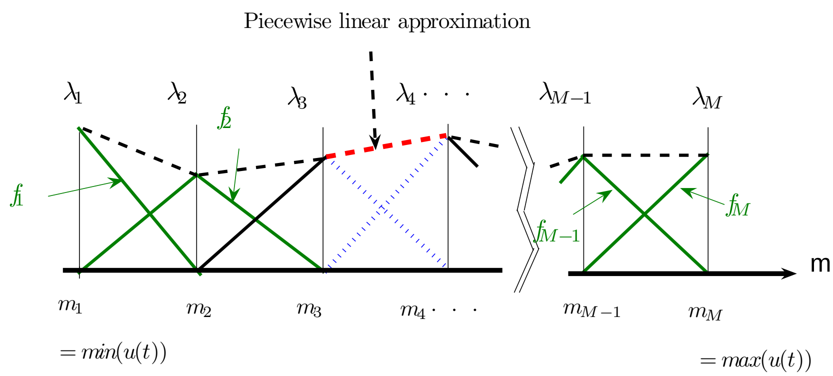



**Figure 2.1**: Triangle basis functions.

linear approximation of the static nonlinearity $f(\cdot)$, a finite value $M$ in (2.5) can be chosen, whereas the points $m_1, \cdots, m_M$ of a grid over $[u_{min}, u_{max}]$ can be chosen linearly spaced or at strategic locations. Each triangle function $\xi_m(u(t))$ in (2.5) has nonzero values through two segments and zeros elsewhere except for the first and the last intervals of the grid.

$$\xi_m(u(t)) = \begin{cases} \dfrac{u(t) - m_{l-1}}{m_l - m_{l-1}} & \text{for } m_{l-1} \le u(t) < m_l \\ \dfrac{m_{l+1} - u(t)}{m_{l+1} - m_l} & \text{for } m_l \le u(t) < m_{l+1} \\ 0 & \text{Otherwise} \end{cases}$$

$$\xi_1(u(t)) = \begin{cases} \dfrac{m_2 - u(t)}{m_2 - m_1} & \text{for } m_1 \le u(t) < m_2 \\ 0 & \text{Otherwise} \end{cases}$$

$$\xi_M(u(t)) = \begin{cases} \dfrac{u(t) - m_{M-1}}{m_M - m_{M-1}} & \text{for } m_{M-1} < u(t) \le m_M \\ 0 & \text{Otherwise} \end{cases}$$

In each segment of the $m$-axis in Figure 2.1, the resulting linear function as indicated by the (red/shaded) dashed line is defined by two overlapping triangle functions in the segment, as indicated by the two (blue) dotted lines. In feedback control systems, non-smooth static nonlinearity, such as saturation, is common. A piecewise linear approximation is an excellent way to estimate such nonlinearity for feedback control systems since we can achieve good approximation with only a small number of parameters.

## 2.3 Semidefinite programming

Semidefinite programming (SDP) deals with minimization of a linear cost function subject to positive semidefinite symmetric matrix variables with an affine space. A typical semidefinite programming problem can be written as

$$\text{minimize } c^T x$$
$$\text{subject to } F(x) \geq 0,$$
$$\text{where}$$
$$F(x) = F_0 + \sum_{i=1}^{m} x_i F_i$$

where $x \in \mathbb{R}^m$, $c \in \mathbb{R}^m$, and $F_i \in \mathbb{R}^{n \times m}$ are symmetric matrices. Since such constraints are convex, SDP is a special case of convex optimization and includes special cases of linear programming, quadratic programming, etc. Semidefinite programming is one of the largest classes of optimization problems that can be solved with reasonable efficiency. SDP has many applications in engineering, such as robust control and combinatorial optimization, and can be applied to system identification [13].

## 2.4 Rank minimization and nuclear norm relaxation

The order of a finite dimensional model can be expressed as the rank of a matrix that is filled with input and output measurement. Finding the simplest model within a feasible model set restricted by convex constraints can often be formulated as a rank minimization problem. From the control point of view, identifying the simplest model that can capture the most important dynamic characteristic of a system is important. Unfortunately, the rank minimization is not convex. Minimizing the nuclear norm instead of the rank of the matrix is a convex relaxation of the rank minimization problem where the nuclear norm of $Z$ is defined as the summation of its singular values as [26] [27]:

$$||Z||_* = \sum_{i=1}^{r} \sigma_i(Z).$$

The motivation for this nuclear norm relaxation is that over the set

$$\{Z \mid ||Z|| \leq 1\},$$

$||Z||_*$ is the convex envelope of the function $Rank(Z)$. While the original rank minimization problem is difficult to solve, the nuclear norm problem is a convex optimization problem which is easy to solve. Let $h : C \rightarrow \Re$, where $C \subseteq \Re^n$. The convex envelope of $h$ (on $C$) is defined as the largest convex function $g$ such that $g(x) \leq h(x)$ for all $x \in C$ [26].

**Lemma 1** *[26] The convex envelope of the function $\phi(X) = Rank(X)$, on $C = \{X \in \Re^{m\times n} | \ ||X|| \leq 1\}$, is $\phi_{env}(X) = ||X||_*$.*

**Proof** : [26] To prove the theorem we use conjugate functions. Recall that the conjugate $h^*$ of a function $h : C \rightarrow \Re$, where $C \subseteq \Re^n$, is defined as

$$h^*(y) = \sup\{y^T x - h(x) | x \in C\}.$$

A basic result of convex analysis is that $h^{**}$, i.e., the conjugate of the conjugate, is the convex envelope of the function $h$, provided some technical conditions, which are valid here.

Part 1 . Computing $\phi^*$: The conjugate of the rank function $\phi$, on the set of matrices with (spectral) norm less than or equal to one, is

$$\phi^*(Y) = \sup_{||X||\leq 1} (trY^T X - \phi(X)). \tag{2.6}$$

Let $q = min\{m, n\}$, and note that by Von Neumann's trace theorem we have

$$trY^TX \leq \sum_{i=1}^{q} \sigma_i(Y)\sigma_i(X),$$

where $\sigma_i(\cdot)$ denotes the $i$th largest singular value. Let $X = U_X Z_X V_X^T$ and $Y = U_y \Sigma_Y V_Y^T$ be the singular value decompositions (SVDs) of X and Y. Since the term $\phi(X)$ in (2.6) is independent of $U_X$ and $V_X$, we pick $U_X = U_Y$ and $V_X = V_Y$ to maximize the first term in (2.6). It follows that

$$\phi^*(Y) = \sup_{||X|| \leq 1} \left( \sum_{i=1}^{q} \sigma_i(Y)\sigma_i(X) - Rank(X) \right).$$

If $X = 0$, we have $\phi^*(Y) = 0$ for all $Y$, and if $Rank(X) = r, 1 \leq r \leq q$, then $\phi^*(Y) = \sum_{i=1}^{r} \sigma_i(Y) - r$. So $\phi^*(Y)$ can be expressed as:

$$\phi^*(Y) = max\{0, \sigma_1(Y) - 1, \cdots, \sum_{i=1}^{r} \sigma_i(Y) - q\}.$$

The largest term in this set is the one that sums all positive $(\sigma_i(Y) - 1)$ terms. We conclude that

$$\phi^*(Y) = \sum_{i=1}^{q} (\sigma_i(Y) - 1)_+$$

where $a_+$ denotes the positive part of $a$, i.e., $a_+ = max\{0, a\}$.

Part 2. Computing $\phi^{**}$: We will now find the conjugate of $\phi^*$, defined as

$$\phi^{**}(Z) = \sup(trZ^TY - \phi^*(Y))$$

for all $Z \in C^{m \times n}$. As before, we choose $U_Y$ and $V_Y$ such that $U_z^T U_Y = I$ and $V_Y^T V_Z = I$ to get

$$\phi^{**}(Z) = \sup(\sum_{i=1}^{q} \sigma_i(Z)\sigma_i(Y) - \phi^*(Y)).$$

We will consider two cases, $||Z|| > 1$ and $||Z|| \leq 1$: If $||Z|| > 1$, we can choose $\sigma_1(Y)$ large enough so that $\phi^{**}(Z) \to \infty$. To see this, note that in

$$\phi^{**}(Z) = \sup_Y \left( \sum_{i=1}^{q} \sigma_i(Z)\sigma_i(Y) - (\sum_{i=1}^{q} \sigma_i(Y) - r) \right),$$

the coefficient of $\sigma_1(Y)$ is $(\sigma_1(Z) - 1)$ which is positive. Now let $||Z|| \leq 1$. If $||Y|| > 1$, then $\phi^*(Y) = 0$ and the supremum is achieved for $\sigma_i(Y) = 1, i = 1, \cdots, q$, yielding

$$\phi^{**}(Z) = \sum_{i=1}^{q} \sigma_i(Z) = ||Z||_*.$$

We will now show that if $||Y|| > 1$, $\phi^{**}(Z)$ is always smaller than the value given above. We have

$$\phi^{**}(Z) = \sup_{||Y||>1} \left( \sum_{i=1}^{q} \sigma_i(Z)\sigma_i(Y) - (\sum_{i=1}^{q} \sigma_i(Y) - 1) \right).$$

Consider the expression inside the sup. By adding and subtracting the term $\sum_{i=1}^{q} \sigma_i(Z)$ and rearranging the terms, we get

$$\begin{aligned} &= \sum_{i=1}^{r} (\sigma_i(Y) - 1)(\sigma_i(Z) - 1) + \sum_{i=r+1}^{q} (\sigma_i(Y) - 1)\sigma_i(Z) \\ &< \sum_{i=1}^{q} \sigma_i(Z) \end{aligned}$$

where the last inequality holds since the first two sums on the second line always have a negative value. In summary, we have shown

$$\phi^{**}(Z) = ||Z||_*$$

over the set $\{Z| \ ||Z|| \leq 1\}$. Thus, over this set, $||Z||_*$ is the convex envelope of the function $Rank(Z)$. □

Lemma 1 has the following implications. Suppose the feasible set is bounded by $Q$, i.e., for all $X \in C$, we have $||X|| \leq Q$. The convex envelop of $Rank X$ on $\{X| \ ||X|| \leq Q\}$ is given by $\frac{1}{Q}||X||_*$. In particular, for all $X \in C$, we have $Rank X \geq \frac{1}{Q}||X||_*$. Thus, by solving the nuclear norm minimization problem, we obtain a lower bound on the optimal value of the original rank minimization problem [26][27].

Nuclear norm minimization coupled with linear constraints lead to a Semidefinite Programming (SDP) problem that converts a non-convex problem to a convex optimization problem by defining a feasible convex set. This is easier to solve and the solution is close to the solution of the original non-convex problem [26] [27].

# 3

# Open-loop Identification of Hammerstein Systems

## 3.1 Introduction

A Hammerstein system has a block oriented structure where a static input nonlinearity and a linear dynamic system are separated, as shown in Figure 3.7. Hammerstein structure is very efficiency in modeling systems with actuator nonlinearity or other nonlinear effects that can be brought to the system input [29]. It has been shown that such a model structure can effectively represent and approximate many industrial processes. For example, the nonlinear dynamics of various chemical processes, electrical systems, biological systems, thermal system, etc. have been modeled with the Hammerstein model [5] [21] [23] [29] [38] [42] [43] [44] [58] [66]. Early research on the identification of Hammerstein systems can be found in [53]. Many authors have tackled the identification of Hammerstein systems and a short overview is included here as a reference. A non-iterative method to estimate the parameters by minimizing the equation error was proposed in [14]. Due to the equation error minimization, this method is sensitive to noise and may result in biased estimation for colored noise. An iterative technique for the estimation of parameters in a Hammerstein model was developed in [36] to deal with colored noise. Modified formulations of the generalized least squares (GLS) estimation algorithm for system parameter identification was presented in [37] to deal with the biased estimation produced by the least squares (LS) method, but there is no guarantee that the procedure converges to the optimal solution. A two-step identification method of the LS parameter estimation based on correlation functions was introduced in [34]. The authors in [3] discussed discrete Hammerstein model identification using a blind system identification approach. The authors in [66] revisited an optimality result established in [4] showing that the two-stage algorithm (TSA) provides the optimal estimation of a bilinearly parameterized Hammerstein system in the sense of a weighted nonlinear LS criterion formulated with a special weighting matrix. However, these methods suffer from an over-parametrization problem, which requires parameter separation via a singular value decomposition (SVD). A more comprehensive overview of block-oriented nonlinear system identification can be found in [29].

In this chapter, we propose a method that extends the rank minimization approach to Hammerstein system identification and does not need a bilinear parametrization and singular value decomposition (SVD), which are commonly used in two-step approaches for Hammerstein system identification. Regarding Figure 3.7, the objective of the study in this chapter is to formulate a procedure that allows the characterization and identification of the nonlinear static function $f(\cdot)$ and the linear dynamic system $G(q)$ individually based on the input $u(t)$ and the output $y(t)$ observation. This is done in a novel way by the reconstruction of the intermediate signal $x(t)$ with conditions on the finite dimensional dynamic representation of the linear systems $G(q)$ and the memoryless static nonlinearity $f(\cdot)$. A similar method of finding a feasible model consistent with the input and output data under certain constraints was also considered in [68].

As mentioned in Chapter 1, finding the simplest model within a feasible model set restricted by convex constraints can often be formulated as a rank minimization problem [27]. Based on this idea, in this study the rank minimization problem is used to formulate a convex optimization problem via nuclear norm relaxation, where the nuclear norm is defined as the summation of its singular values. The use of nuclear norm approximation with application to system identification can be found in [47] and with application to Hammerstein systems in [24].

## 3.2 Problem description

### 3.2.1 Hammerstein system

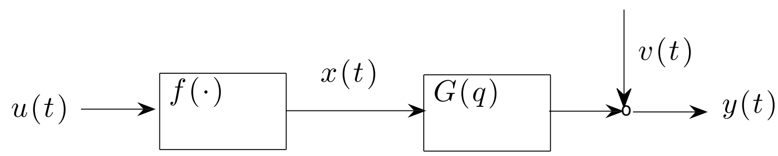



**Figure 3.1**: Hammerstein system consists of a static nonlinear block followed by a linear dynamic block.

The system to be modeled is shown in Figure 3.1. We propose a method to identify the unknown linear dynamical systems $G(q)$ and a static nonlinear function $f(\cdot)$ from a finite number of observations of the data $u(t)$ and $y(t)$ by reconstructing the unmeasurable intermediate signal $x(t)$ via rank constrained Semidefinite Programming (SDP). The SDP problem will be formulated in such a way that $x(t)$ and $u(t)$ are related via a memoryless static nonlinearity and that $x(t)$ and $y(t)$ are related via a linear dynamical system with the smallest McMillan degree. Once $x(t)$ has been reconstructed, the identification of $G(q)$ from $x(t)$ and $y(t)$ can be solved with a standard Prediction Error (PE) identification method [49]. The static nonlinearity and the system parameters will be estimated by finding a feasible model consistent with the input and output data, and satisfying the following basic properties of the Hammerstein system:

**Condition 1**

*I. The static nonlinear function (the relation between $u(t)$ and $x(t)$) has no memory.*

*II. The linear dynamical system has a finite, but unknown McMillan degree $n$, relating a finite number of the past input samples to the past output samples.*

The properties in Condition 1 are used to formulate a procedure to reconstruct the unmeasurable intermediate signal $x(t)$ based on rank minimization.

### 3.2.2 Modeling of static nonlinearity

In this section, the input static nonlinearity is modeled as a piecewise linear function using piecewise triangle functions as shown in Section 2.2. Let $\hat{x}(t) = f(u(t), \lambda)$ be the approximation of $x(t)$ and $\lambda$ is the amplitude parameter

$$\lambda = \begin{bmatrix} \lambda_1 & \cdots & \lambda_M \end{bmatrix}^T \tag{3.1}$$

on $m = [m_1 \ \cdots \ m_M]^T$. In each segment of the $m$-axis, the resulting linear function is defined by two overlapping triangle functions in the segment. Thus, $\hat{x}(t)$ can be written as

$$\hat{x}(t) = \rho(u(t))\lambda \tag{3.2}$$

where $\rho(u(t))$ is defined as

$$\rho(u(t)) = \left[\cdots \ 0 \ \frac{m_{k+1} - u(t)}{m_{k+1} - m_k} \ \frac{u(t) - m_k}{m_{k+1} - m_k} \ 0 \ \cdots\right] \tag{3.3}$$
$$\text{for } \ m_k \leq u(t) < m_{k+1}$$

where $m_k$ and $m_{k+1}$ are the center locations of the triangle basis functions. There could be many possible combinations of a static nonlinear block and a Finite Impulse Response (FIR) linear block that satisfy Condition 1 and (3.2). In order to limit the number of possible selections for a linear block and a static nonlinear block, a monotonically non-decreasing static nonlinearity with the maximum slope of 1 (fixing the maximum slope of the static nonlinear function in one interval, likely at the origin, equals 1) is considered as follows:

**Condition 2**

*I. The static nonlinear function is monotonically non-decreasing with the maximum slope of* 1*:*
$(\hat{x}(i) - \hat{x}(j))(\hat{x}(i) - \hat{x}(j) - u(i) + u(j)) \leq 0$
$\forall i > j.$

Without loss of generalization, this monotonicity assumption on the unknown static nonlinearity with the maximum slope 1 guarantees a solution for an FIR linear system and serves as a normalization condition on the static nonlinearity, so that the static gain of the Hammerstein system is modeled by the static gain of the linear system $G(q)$. The maximum slope of the static nonlinear function can be chosen by a user. If there exists a static nonlinear function that satisfies Condition 2 for sampled data, Condition 2 will be satisfied by any two values chosen from the sample range by the mean values theorem. A Hammerstein system with a monotonically non-decreasing static nonlinear function can be used to model many control, mechanical, electrical, chemical, and biological systems with various static nonlinear functions, such as saturation, deadzone, quantization, etc. The examples can be found in [5][21][23][38][60].

### 3.2.3 Input-output map of the dynamical system

Let $g(i),\ i = 0, 1, \cdots$ be the causal sequence of unit impulse responses for $G(q)$. The relationship between the intermediate signal $x(t)$ and the output $y(t)$ can be described by the convolution as

$$y(t) = \sum_{i=0}^{\infty} g(i)x(t-i) + v(t)$$

where $v(t)$ is noise. Due to Condition 1 (finite McMillan degree), for a finite data sequence of $N = n_1 + n_2$ data points and a zero initial condition, the relationship between the intermediate signal $x(t)$ and the output $y(t)$ can be described by

$$Y = HX_p + TX_f + V \tag{3.4}$$

where $Y$ is the data matrix, including the future output $y(t)$, defined by

$$Y = \begin{bmatrix} y(1) & \cdots & y(n_2) \\ y(2) & \cdots & y(n_2+1) \\ \vdots & \ddots & \vdots \\ y(n_1) & \cdots & \cdots y(n_1+n_2-1) \end{bmatrix}, \tag{3.5}$$

$X_p$ is the data matrix, including the past intermediate signal, defined by

$$X_p = \begin{bmatrix} x(0) & x(1) & \cdots & x(n_2-1) \\ 0 & x(0) & \cdots & x(n_2-2) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots x(0) \end{bmatrix},$$

$X_f$ is the data matrix, including the future intermediate signal, defined by

$$X_f = \begin{bmatrix} x(1) & x(2) & \cdots & x(n_2) \\ x(2) & x(3) & \cdots & x(n_2+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(n_1+1) & x(n_1+2) & \cdots & x(n_1+n_2) \end{bmatrix}, \tag{3.6}$$

$H$ is the Hankel matrix defined by

$$H = \begin{bmatrix} g(1) & \cdots & g(n_2) \\ g(2) & \cdots & g(n_2+1) \\ \vdots & \ddots & \vdots \\ g(n_1) & \cdots & \cdots g(n_1+n_2-1) \end{bmatrix},$$

$T$ is the Toeplitz matrix defined by

$$T = \begin{bmatrix} g(0) & 0 & \cdots & 0 & 0 \\ g(1) & g(0) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ g(n_1-1) & g(n_1-2) & \cdots & g(0) & 0 \end{bmatrix},$$

and $V$ is the matrix, including noise data, defined by

$$V = \begin{bmatrix} v(1) & \cdots & v(n_2) \\ v(2) & \cdots & v(n_2+1) \\ \vdots & \ddots & \vdots \\ v(n_1) & \cdots & \cdots v(n_1+n_2-1) \end{bmatrix}.$$

The order of the linear dynamical system is determined by the $rank(H)$ as $H$ is simply the product of the extended observability and controllability matrices [30]. A lower order model, consistent with the input and output signals can be estimated by minimizing the rank of $H$.

## 3.3 System identification

### 3.3.1 Problem formulation

The effect of $X_f$ to $Y$ in (3.4) can be removed by the orthogonal projection of $Y$ onto the null space of $X_f$. With the projection matrix $X_f^{\perp} = I - X_f^T(X_f X_f^T)^{\dagger} X_f$, the effect of $X_f$ is removed [49]. Then,

$$YX_f^{\perp} = HX_pX_f^{\perp} + VX_f^{\perp}.$$

In order to remove the effect of noise, the projection $Y$ can be subsequently weighted by matrices $W_1$ and $W_2$ such that

$$W_1YX_f^{\perp}W_2 = W_1HX_PX_f^{\perp}W_2 + W_1VX_f^{\perp}W_2 \tag{3.7}$$

in which $W_1$ and $W_2$ are chosen to be rank-preserving and such that $W_1VX_f^{\perp}W_2 \rightarrow 0$ as the number of samples $N \rightarrow \infty$. Details of the role and choice of weighting matrices $W_1$ and $W_2$ can be found in [51] [61]. The result in (3.7) indicates that the rank minimization problem for $H$ can be rewritten as the rank minimization problem for $YX_f^{\perp}$. Unfortunately, the rank minimization problem for $YX_f^{\perp}$ cannot be solved directly since $X$ is unknown. In this section, the rank minimization problem of $YX_f^{\perp}$ is reformulated using $LQ$ decomposition of data matrix $\begin{bmatrix} X_f \\ Y \end{bmatrix}$ so that the rank minimization problem can be solved without knowing $X_f^{\perp}$.
Let the LQ decomposition (the transpose of the QR decomposition) of the data matrix $\begin{bmatrix} X_f \\ Y \end{bmatrix}$ be given by

$$\begin{bmatrix} X_f \\ Y \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} \tag{3.8}$$

where $L_{11}, L_{22}$ are lower triangular and $Q_1, Q_2$ are orthogonal. From (3.8), $Y$ can be written as

$$Y = L_{21}L_{11}^{-1}X_f + L_{22}Q_2^T. \tag{3.9}$$

The first term in (3.9) is spanned by the row vectors in $X_f$ and the second term is orthogonal to it. Thus, the orthogonal projection of $Y$ onto the null space of $X_f$

can be written as [30] [49]

$$YX_f^{\perp} = L_{22}Q_2^T. \tag{3.10}$$

Since $Q_2$ is orthogonal, (3.10) indicates

$$rank(YX_f^{\perp}) = rank(L_{22}).$$

As a result, the rank minimization problem for $YX_f^{\perp}$ can be rewritten as the rank minimization problem for $L_{22}$. Using the signal $\hat{x}(t)$ in (3.2), $X_f$ in (3.6) can be reconstructed using $\hat{x}$ as

$$\hat{X}_f = U_2\Theta \tag{3.11}$$

where $\Theta$ is the block diagonal matrix, including $\lambda$, defined by

$$\Theta = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ \lambda_2 & \vdots & \cdots & 0 \\ \vdots & 0 & \cdots & 0 \\ \lambda_M & \vdots & \ddots & 0 \\ 0 & \lambda_1 & \ddots & 0 \\ \vdots & \lambda_2 & \ddots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \vdots & \lambda_M & \ddots & 0 \\ \vdots & 0 & \cdots & \lambda_1 \\ \vdots & \vdots & \cdots & \lambda_2 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \ddots & \lambda_M \end{bmatrix}, \tag{3.12}$$

and $U_2$ is the data matrix including the input $u(t)$, defined by

$$U_2 = \begin{bmatrix} \rho(u(1)) & \cdots & \rho(u(n_2)) \\ \rho(u(2)) & \cdots & \rho(u(n_2+1)) \\ \vdots & \ddots & \vdots \\ \rho(u(n_1+1)) & \cdots & \rho(u(n_1+n_2)) \end{bmatrix}. \tag{3.13}$$

From (3.8), we have

$$rank\left(\begin{bmatrix} X_f \\ Y \end{bmatrix}\right) = rank\left(\begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix}\right).$$

Let $L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix}$. The following lemma explains the rank inequality condition for the block matrix $L$.

**Lemma 2** *The rank of the block matrix $L$ satisfies the following inequality:*

$$rank\left(\begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix}\right) \geq rank(L_{11}) + rank(L_{22})$$

**Proof** Suppose that $p = rank(L_{11})$ and $q = rank(L_{22})$. Then $L_{11}$ has $p$ linearly independent columns, $a_1, \cdots, a_p$, and $L_{22}$ has $q$ linearly independent columns, $d_1, \cdots, d_q$. Let $c_i$ denote the columns of $L_{21}$. Consider the columns of the block matrix $L$

$$\begin{pmatrix} a_1 \\ c_1 \end{pmatrix}, \cdots \begin{pmatrix} a_p \\ c_p \end{pmatrix}, \cdots \begin{pmatrix} 0 \\ d_1 \end{pmatrix}, \cdots \begin{pmatrix} 0 \\ d_q \end{pmatrix}.$$

Then one can show that these $p + q$ column vectors are linearly independent by contradiction. Suppose these column vectors are linearly dependent. Then there exist numbers, $\alpha_1, \cdots, \alpha_p$ and $\beta_1, \cdots, \beta_q$, not all zero, such that

$$\sum_{i=1}^{p} \alpha_i \begin{pmatrix} a_i \\ c_i \end{pmatrix} + \sum_{i=1}^{q} \beta_i \begin{pmatrix} 0 \\ d_i \end{pmatrix} = 0.$$

This gives the two equations:

$$\sum_{i=1}^{p} \alpha_i a_i = 0 \tag{3.14}$$

and

$$\sum_{i=1}^{p} \alpha_i c_i + \sum_{i=1}^{q} \beta_i d_i = 0. \tag{3.15}$$

Since $a_i$ are linearly independent, $\alpha_i = 0$ from (3.14). Then $\beta_i = 0$ from (3.15) because $d_i$ are linearly independent. These are contradicted to the assumption. Therefore, the block matrix $L$ has at least $p + q$ linearly independent columns. □

With Lemma 2, we have

$$rank(L_{11}) + rank(L_{22}) \leq rank\left(\begin{bmatrix} X_f \\ Y \end{bmatrix}\right). \tag{3.16}$$

Since $rank(L_{11}) = rank(X_f)$ from (3.8), (3.16) can be written as

$$rank(X_f) + rank(L_{22}) \leq rank\left(\begin{bmatrix} X_f \\ Y \end{bmatrix}\right).$$

From (3.11), we have

$$rank(X_f) = rank(U_2\Theta) = constant.$$

In this section, the rank minimization problem for $L_{22}$ is relaxed to the upper bound minimization problem for $rank(L_{22})$, which is equivalent to the minimization problem for $rank\left(\begin{bmatrix} X_f \\ Y \end{bmatrix}\right)$. As a result, with the parametrization based on Condition 1, system parameters for a lower order model, consistent with the input and output measurement data, will be estimated by minimizing $rank\left(\begin{bmatrix} X_f \\ Y \end{bmatrix}\right)$ under the constraints developed based on Condition 2.

### 3.3.2 Rank minimization for intermediate signal reconstruction

In this section, a rank minimization problem for the reconstruction of the intermediate signal $x(t)$ in Figure 3.1 is summarized and the optimization problem is constructed. In Section 3.2, the rank minimization problem for $H$ is reformulated as the rank minimization problem of the data matrix $\begin{bmatrix} \hat{X}_f \\ Y \end{bmatrix}$. With the parametrization and constraints explained in Section 3.2, an optimization problem can be written as

**Optimization Problem 1**

$$\begin{aligned}
&\textit{Consider}\\
&\quad \textit{variable } \lambda \textit{ in (3.1)}\\
&\quad \textit{to create } \hat{x} \textit{ in (3.2) and } \Theta \textit{ in (3.12),}\\
&\textit{Minimize}\\
&\quad \textit{rank}\left(\begin{bmatrix} \hat{X}_f \\ Y \end{bmatrix}\right),\\
&\quad \textit{where } \hat{X}_f = U_2\Theta, \textit{ and } Y \textit{ is given in (3.5),}\\
&\quad \textit{with } U_2 \textit{ defined in (3.13),}\\
&\textit{subject to}\\
&\quad (\hat{x}(i) - \hat{x}(j))(\hat{x}(i) - \hat{x}(j) - u(i) + u(j)) \leq 0\\
&\quad \forall i > j.
\end{aligned}$$

Optimization Problem 1 results in the optimal solution for the system parameter $\theta$ that is used to construct the intermediate signal $x$ using the relationship in (3.2), automatically satisfying Condition 1-I. In Optimization Problem 1, the reconstructed signal $\hat{x}$ is generated in such a way that a static nonlinear function satisfies the monotonically non-decreasing and scaling conditions in Condition 2, and provides a lower order model for the linear dynamic system satisfying Condition 1-II by minimizing $rank\left(\begin{bmatrix} \hat{X}_f \\ Y \end{bmatrix}\right)$. Unfortunately, the rank constraint in Optimization Problem 1 is not convex. Minimizing the nuclear norm instead of the rank of the matrix is a convex relaxation of the rank minimization problem [26] [27]. The motivation for this nuclear norm relaxation is that the nuclear norm is the convex envelope of the rank function on the set of matrices with norms less than 1 as shown in Lemma 1. Using the nuclear norm relaxation for rank minimization, Optimization Problem 1 will be reformulated as a convex problem. First, let us express the constraints in Optimization Problem 1 as Linear Matrix Inequalities (LMIs). Let $\delta x = [\delta x(1) \;\cdots\; \delta x(k_{max})]^T$, where $\delta x(k) = \hat{x}(i) - \hat{x}(j)$ and $\delta u = [\delta u(1) \;\cdots\; \delta u(k_{max})]^T$, where $\delta u(k) = u(i) - u(j)$ for all $i > j$ and $k_{max} = \sum_{k=1}^{N-1} k$. Let $\Delta X = diag(\delta x)$ and $\Delta U = diag(\delta u)$, where $diag(\delta x)$ is the diagonal matrix whose diagonal elements are the elements of $\delta x$. Then, the constraints in Optimization Problem 1 can be written as

$$\Delta X(\Delta X - \Delta U) \leq 0$$

Nuclear norm minimization coupled with linear constraints lead to a Semidefinite Programming (SDP) problem that converts a non-convex problem to a convex optimization problem by defining a feasible convex set. This is easier to solve and the solution is close to the solution of the original non-convex problem [26] [27]. Using SDP relaxation, Optimization Problem 1 can be rewritten as the following convex optimization problem:

**Optimization Problem 2**

*Consider*

*variable $\lambda$ in (3.1)*

*to create $\hat{x}$ in (3.2) and $\Theta$ in (3.12),*

*Minimize*

$$\left|\left| \begin{bmatrix} \hat{X}_f \\ Y \end{bmatrix} \right|\right|_*$$

*where $\hat{X}_f = U_2\Theta$, and $Y$ is given in (3.5),*

*with $U_2$ defined in (3.13),*

*subject to*

$$\Delta X(\Delta X - \Delta U) \leq 0$$

*where*

$$||H||_* = \sum_{i=1}^{r} \sigma_i(H)$$

*is the nuclear norm of $H$.*

The following lemma is used to compute the nuclear norm while preserving the convexity of the optimization problem. The following result indicates that Optimization Problem 2 can be solved via a convex optimization problem.

**Lemma 3** *[26] For $X \in \Re^{m\times n}$ and $t \in \Re$, we have $||X||_* \leq t$ if and only if there* *exist matrices* $A \in \Re^{m \times m}$ *and* $Z \in \Re^{n \times n}$ *such that*

$$\begin{bmatrix} A & X \\ X^T & Z \end{bmatrix} \geq 0$$

$$\textit{and}$$

$$tr(A) + tr(Z) \leq 2t.$$

**Proof** : [26] Let $A$ and $Z$ satisfy the relations in Lemma 3, and let $X = U\Sigma V^T$ be the SVD of $X$. Here, $\Sigma$ is of size $r$, where $r$ is the rank of $X$. We have

$$tr \begin{bmatrix} UU^T & -UV^T \\ -VU^T & VV^Y \end{bmatrix} \begin{bmatrix} A & X \\ X^T & Z \end{bmatrix} \geq 0$$

since the trace of the product of two positive semidefinite (PSD) matrices is always non-negative. This yields

$$trUU^TA - trUV^TX^T - trVU^TX + trVV^TZ \geq 0. \tag{3.17}$$

Since columns of $U$ are orthonormal, we can always add more columns to complete them to a full basis, i.e., there exists $U^T$ such that $[U\tilde{U}][U\tilde{U}]^T = I$, or $UU^T + \tilde{U}\tilde{U}^* = I$, so $||UU^T|| \leq 1$. So we get

$$|trUU^TA| \leq \sum_i \lambda_i(UU^T)\lambda_i(Y) \leq TrA.$$

Similarly, for $V$ we have $trVV^TZ \leq trZ$. Also, $trVU^TX = trV\Sigma V^* = tr\Sigma$. Using these facts, and (3.17) above, we get

$$trA + trZ - tr\Sigma \geq 0,$$
$$tr\Sigma \leq \tfrac{1}{2}(trA + trZ),$$
$$tr\Sigma = ||X||_* \leq t.$$

Suppose $||X||_* \leq t$. We will show $A$ and $Z$ can be chosen to satisfy the relations in Lemma 3. Let $A = U\Sigma U^T + \gamma I$ and $Z = V\Sigma V^T + \gamma I$, then

$$trA + trZ = 2tr\Sigma + r(p+q) = 2||X||_* + \gamma(p+q),$$

so if we choose $\gamma = \dfrac{2}{p+q}$, we will have $trA + trZ = 2t$. Also note that

$$\begin{aligned} \begin{bmatrix} A & X \\ X^T & Z \end{bmatrix} &= \begin{bmatrix} U\Sigma U^T & U\Sigma V^T \\ V\Sigma U^T & V\Sigma V^Y \end{bmatrix} + \gamma \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} U \\ V \end{bmatrix} \Sigma [U^T \; V^T] + \gamma I, \end{aligned}$$

which is PSD. □

Lemma 3 indicates that the condition $||X||_* \leq t$ can be represented as an LMI [26].

## 3.4 Numerical example

In this section, numerical examples of Hammerstein system identification using the proposed identification method are presented. An excitation signal $u(t)$ is zero mean white noise with a standard deviation of 3. The output disturbance $v(t)$ is filtered white noise, where the filtering properties are unknown. For the system identification, two sets of data (ten different measurements for each set) are generated from the Hammerstein system. In the first data set, SNR varies from $10dB$ to $20dB$ and in the second data set, SNR is greater than $20dB$. For both data sets, $m = [min(u(t)) \;\; -3 \;\; -1 \; 0 \; 1 \; 3 \; max(u(t))]$, and $n_a = 2$ and $n_b = 2$ with 1 step time delay are used to model the static nonlinear function and the linear dynamic system respectively. $W_1 = W_2 = I$ are chosen in this example. In order to solve the SDP problem (Optimization Problem 2), SeDuMi (Self-Dual-Minimization) [59] and YALMIP (Yet Another LMI Parser) [50] are used. The specifications of the Hammerstein system are as follows:

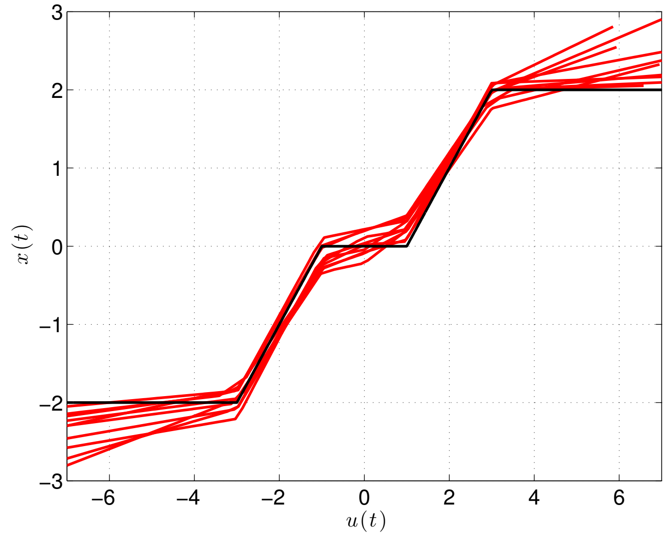
$$\begin{cases} \text{Linear dynamical system:} \\ G(q) = \dfrac{0.1994q^{-1} - 0.1804q^{-2}}{1 - 1.886q^{-1} + 0.9048q^{-2}} \\ \text{Static nonlinearity:} \\ \qquad f(u(t)) = \begin{cases} 2 & \text{if} \quad x(t) > 3 \\ u(t)+1 & \text{if} \quad 1 < x(t) \leq 3 \\ 0 & \text{if} \quad |x(t)| \leq 1 \\ u(t)-1 & \text{if} \quad -3 \leq x(t) < -1 \\ -2 & \text{if} \quad x(t) < -3 \end{cases} \\ \text{Noise dynamics:} \\ H(q) = \dfrac{1 + 0.5q^{-1}}{1 - 0.85q^{-1}} \end{cases}.$$

The estimation results are shown in Figure 3.2, Figure 3.3, Figure 3.4, and Figure 3.5. Figure 3.2 and Figure 3.3 show the simulation results for the first data set (SNR of $10dB - 20dB$). Figure 3.4 and Figure 3.5 show the simulation results for the second data set (SNR $> 20dB$). As shown in Figure 3.4 and Figure 3.5, the proposed algorithm provides excellent identification results for data with SNR greater than $20dB$. The pole location and the characteristics of the static nonlinearity are very well captured. As shown in Figure 3.2 and Figure 3.3, when SNR varies from $10dB$ to $20dB$, the proposed algorithm provides satisfactory identification results. If SNR is less than $10dB$, the proposed algorithm does not always provide satisfactory estimation results.

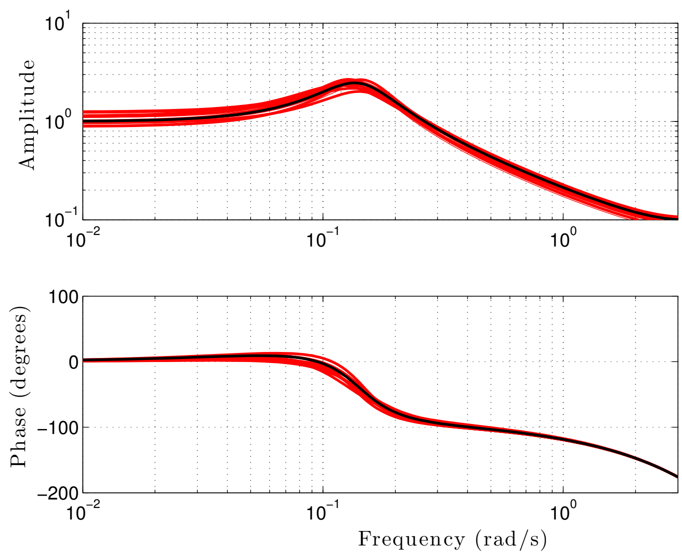
## 3.5 Application to Hard Disk Drive (HDD) thermal actuator identification

### 3.5.1 Thermal actuator identification problem formulation

In this section, a nonlinear dynamic model of a thermal actuator in a hard disk drive is identified using a Hammerstein system identification technique. Hammerstein structure is chosen to represent a thermal actuator because of its efficiency in modeling systems with actuator nonlinearity or other nonlinear effects that can




(a)




(b)

**Figure 3.2**: The plot of the identified static nonlinearity function (top figure). The Bode plot of the identified linear dynamic system (bottom figure). The black solid line indicates the real Hammerstein system. The (colored) dashed lines indicate estimated systems by using ten different data. The SNR of each data in the set is between $10dB$ and $20dB$.

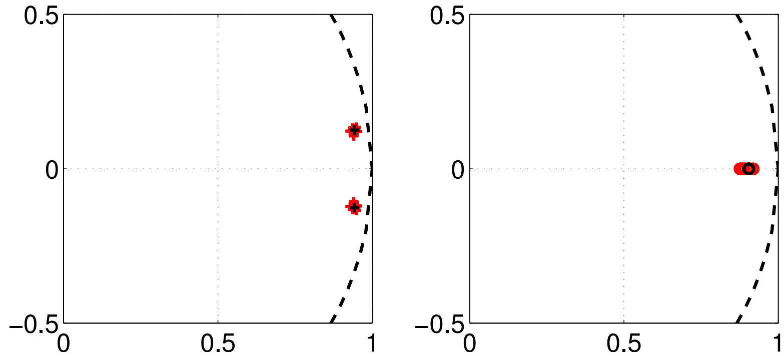
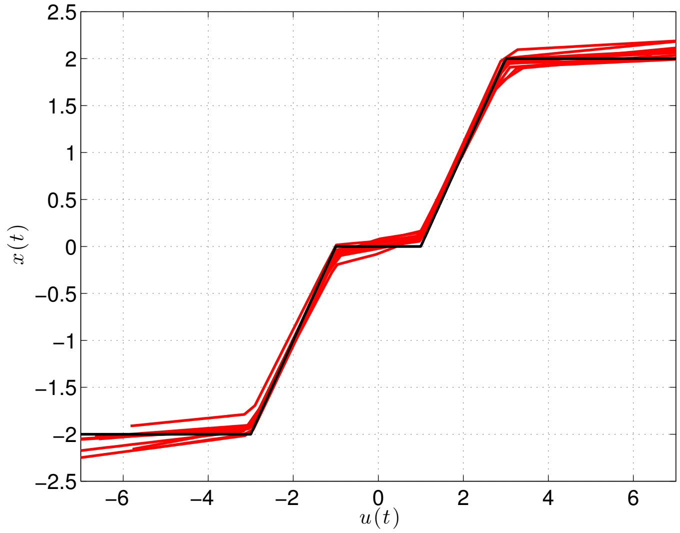



**Figure 3.3**: Pole (top figure) and zero (bottom figure) locations of the identified linear dynamical system. The black cross and circle indicate the real linear dynamical system. The colored crosses and circles indicate estimated linear dynamical systems. SNR varies from $10dB$ to $20dB$.

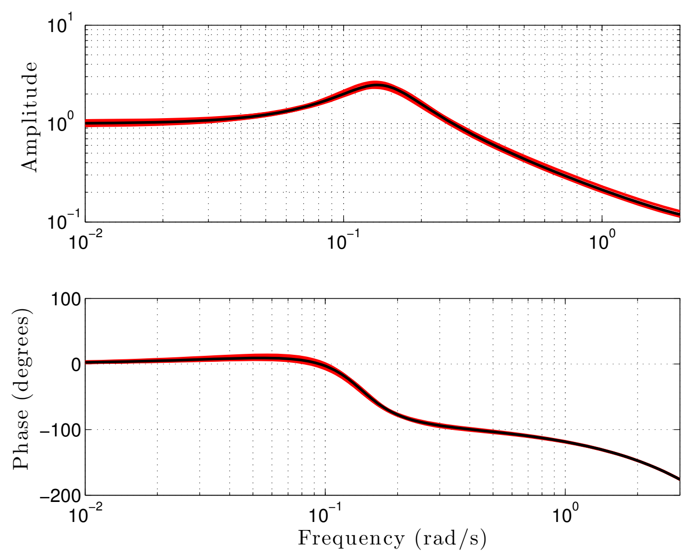
be brought to the system input [29]. For the purpose of identification, the input voltage $u(t)$ to the thermal actuator and the flying height variation output $z(t)$ were measured through spin stand experiments (refer to [11]). The parameter estimation problem for the Hammerstein model is formulated in a slightly different in order to account for the known characteristics of the given static nonlinearity in the thermal actuator in this section.

Figure 3.6 shows a side view of the slider and the disk for this case. As shown in Figure 3.6, the read/write element and the resistance heater are positioned at the trailing edge of the slider. Activating the resistance heater, one can reduce the head-disk clearance by $\Delta d$. Hence, the write current induced pole tip protrusion can be compensated by activating the resistance heater during reading.

In order to capture the linear dynamics in a thermal actuator, the quadratic relationship between the voltage input and the power output in a thermal actuator was assumed in [11]. Although quadratic dependence is motivated by the quadratic relationship between power and voltage for a resistance component, nonlinearity due to height dependent thermal conductivity may occur. Thus, in this experimental study, the nonlinear relationship between the voltage input and the power output is assumed unknown and we will identify the nonlinearity as well as the linear dynamics of a thermal actuator. The schematic diagram of a thermal actuator




(a)




(b)

**Figure 3.4**: The plot of the identified static nonlinearity function (top figure). The Bode plot of the identified linear dynamic system (bottom figure). The black solid line indicates the real Hammerstein system. The (colored) dashed lines indicate estimated systems by using ten different data. The SNR of each data in the set is greater than $20dB$.

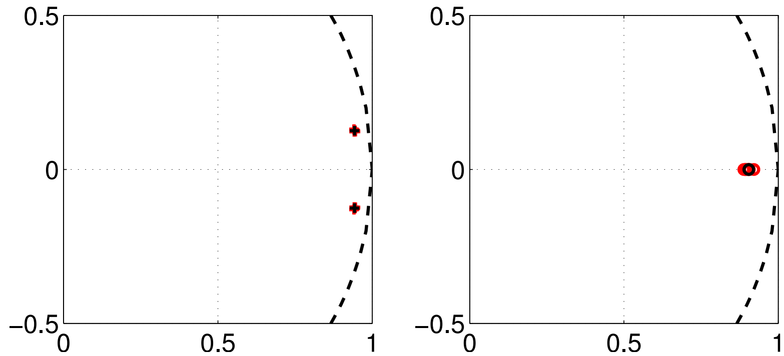



**Figure 3.5**: Pole (top figure) and zero (bottom figure) locations of the identified linear dynamical system. The black cross and circle indicate the real linear dynamical system. The colored crosses and circles indicate estimated linear dynamical systems. SNR is greater than $20dB$.

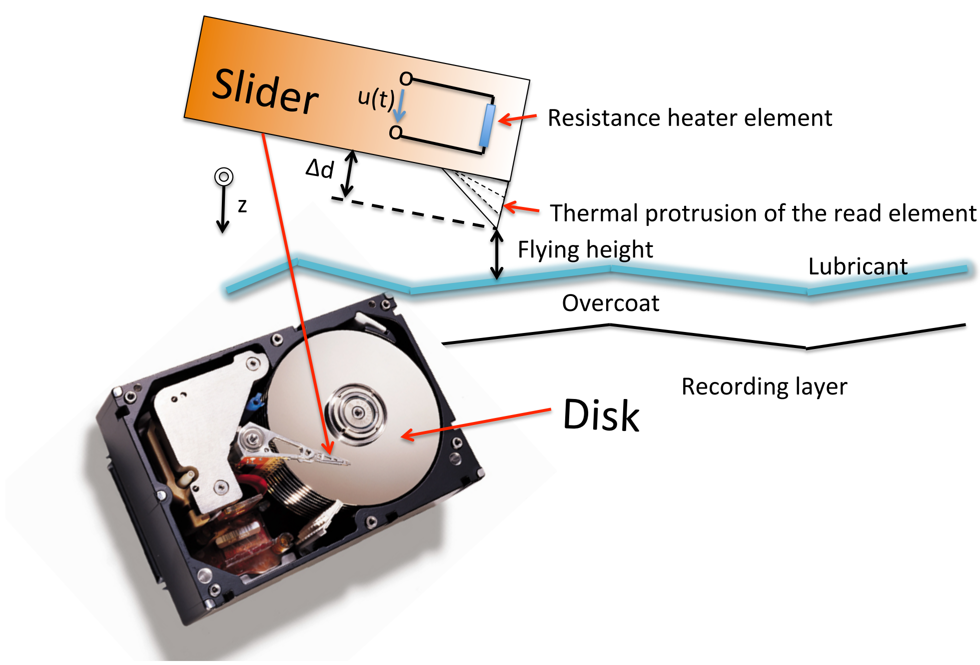



**Figure 3.6**: Hard disk drive and the slider with a resistance heater element for thermal flying height control.

represented by the Hammerstein structure is shown in Figure 3.7. Regarding Fig-

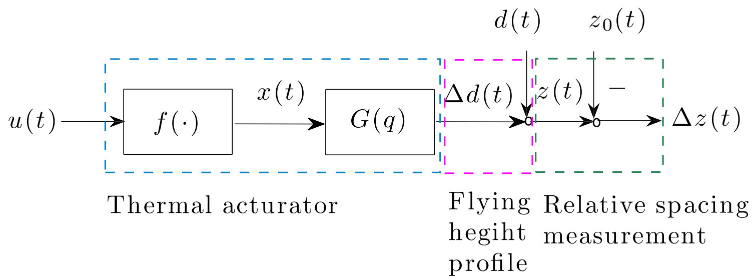



**Figure 3.7**: Thermal actuator represented by a Hammerstein system that consists of a static nonlinear block followed by a linear dynamic block.

ure 3.7, $u(t)$ is the input voltage, $x(t)$ is the unmeasurable intermediate signal that indicates the input power to the system, $z(t)$ is the absolute spacing, $\Delta z(t)$ is the spacing variation relative to an initial flying height $z_0(t)$, $\Delta d(t)$ is the head/disk clearance, and $d(t)$ is disturbance. The absolute spacing $z(t)$ is not measurable. However, the spacing variation relative to an initially unknown flying height $z_0$ can be measured where

$$\Delta z = z - z_0. \tag{3.18}$$

The contribution of the actuator to the flying height change can be estimated by performing two experiments: one without an external input signal as a reference measurement and another experiment using an input signal that excites the system.

The objective of this experimental study is to identify nonlinear dynamics of a thermal actuator system in HDDs using the Hammerstein model structure. Let $\hat{x} = [\hat{x}_1 \ \cdots \ \hat{x}_N]^T$. Then

$$\begin{aligned} &\hat{x} = \rho\lambda \\ &where \\ &\rho = [\rho(u(1)) \ \cdots \ \rho(u(N))]^T. \end{aligned} \tag{3.19}$$

where $\lambda$ is the amplitude parameter in (3.1) and $\rho(u(t))$ is defined in (3.3). Since a non-negative staircase voltage input is used for the identification of the thermal actuator, a non-negative monotonically increasing condition will be applied to the system parameters of the input static nonlinearity as follows:

$$0 \leq \lambda_1 < \cdots < \lambda_M. \tag{3.20}$$

The relationship between the intermediate signal $x(t)$ and the output $z(t)$ can be described by the convolution as

$$z(t) = \sum_{i=0}^{\infty} g(i)x(t-i) + d(t).$$

Due to Condition 1 (finite McMillan degree), the Hankel matrix defined as

$$H = \begin{bmatrix} g(1) & \cdots & g(N/2) \\ g(2) & \cdots & g(N/2+1) \\ \vdots & \ddots & \vdots \\ g(N/2) & \cdots & \cdots g(N-1) \end{bmatrix}, \tag{3.21}$$

has a $rank(H) \leq n$. The order of the linear dynamical system is determined by the $rank(H)$ as $H$ is simply the product of the extended observability and controllability matrices [30]. A lower order model, consistent with the input and output signals can be estimated by minimizing the rank of $H$. Let

$$X = \begin{bmatrix} \hat{x}(1) & \hat{x}(0) & \cdots & \hat{x}(2-N) \\ \hat{x}(2) & \hat{x}(1) & \cdots & \hat{x}(1-N) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}(N) & \hat{x}(N-1) & \cdots & \hat{x}(1) \end{bmatrix} \tag{3.22}$$

and

$$g = [g(0)\ g(1)\ \cdots\ g(N-1)]^T. \tag{3.23}$$

Then the estimate of the output $z$ can be written as

$$\begin{aligned} \hat{z} &= [\hat{z}(1)\ \hat{z}(2)\ \cdots\ \hat{z}(N)]^T \\ &= Xg. \end{aligned} \tag{3.24}$$

Let $\theta = [g^T \lambda^T]^T$. With $\theta$, let us define a positive semidefinite symmetric matrix

$\Theta = \theta\theta^T$ as

$$\Theta = \begin{bmatrix} g(0)^2 & \cdots & g(0)\lambda(M) \\ \vdots & \ddots & \vdots \\ g(N-1)g(0) & \cdots & g(N-1)\lambda(M) \\ \lambda(1)g(0) & \cdots & \cdots\lambda(1)\lambda(M) \\ \vdots & \ddots & \vdots \\ \lambda(M)g(0) & \cdots & \lambda(M)^2 \end{bmatrix}. \tag{3.25}$$

For the purpose of normalization, the following condition will be applied to the system parameters of linear dynamics:

$$\sum_{k=1}^{N} \Theta(k) = g(0)^2 + \cdots + g(N)^2 = \alpha$$

where $\alpha$ is a user-chosen normalization coefficient. Due to the structure of $\Theta$, the monotonically increasing condition of $\lambda$ in (3.20) is relaxed to

$$\begin{aligned} &0 \leq \Theta(N+1, N+1) < \cdots < \Theta(N+M, N+M) \\ &\textit{which is equivalent to} \\ &0 \leq \lambda_1^2 < \cdots < \lambda_M^2. \end{aligned}$$

Without loss of generalization, this monotonicity assumption on the unknown static nonlinearity with a normalization coefficient $\alpha$ of the linear dynamic system guarantees a solution for an FIR linear system. From (3.19) and (3.24), the model output $\hat{z}$ is defined by

$$\hat{z} = \sum_{k=1}^{min(t,N)} T_{t-k+1,k} \tag{3.26}$$

where

$$\begin{aligned} T &= \rho\lambda g^T \\ &= \rho\Theta(N+1 : N+M, 1 : N). \end{aligned}$$

Based on its structure, it is clear that $\Theta$ is a *rank* 1 matrix if there is no noise in the data. For cases where there is noise, system parameters will be found by minimizing $rank(\Theta)$. Because $\Theta$ is a square positive semidefinite matrix, minimizing its trace is the closest approximation of the rank minimization that can be efficiently solved [27]. With the given system parametrization, the optimization problem for estimation of system parameters can be written as

**Optimization Problem 3**

$$\begin{aligned}
&\textit{Consider}\\
&\quad \textit{variable } \Theta \textit{ in (3.25) to define}\\
&\quad \hat{z} = \sum_{k=1}^{min(t,N)} T_{t-k+1,k}\\
&\quad \textit{where } T = \rho\Theta(N+1:N+M, 1:N)\\
&\quad \textit{with } \rho \textit{ in (3.19).}\\
&\textit{Minimize}\\
&\quad w_1||z(t) - \hat{z}(t)||_2 + w_2 trace(\Theta)\\
&\textit{subject to}\\
&\quad \sum_{k=1}^{N} \Theta(k) = \alpha \textit{ and}\\
&\quad 0 \leq \Theta(N+1, N+1) < \cdots < \Theta(N+M, N+M).
\end{aligned}$$

In Optimization Problem 3 above, $\alpha$ is a normalization coefficient, and $w_1$ and $w_2$ are weighting factors. Also, the rank minimization of $H$ is included in the trace minimization of $\Theta$.

Due to the over-parametrization used to define $\Theta$ in (3.25), we need to separate the parameters of the linear dynamical system $g$ and the parameter of the static nonlinearity $\lambda$. Singular Value Decomposition (SVD) is used in this section to separate these system parameters. The SVD of $\Theta$ is given as

$$\Theta = U\Sigma V^T \tag{3.27}$$

where $U$ and $V$ are orthogonal matrices, $U = V$ due to the structure of $\Theta$, and $\Sigma$ is a rectangular diagonal matrix. The positive diagonal entries of $\Sigma$ are called singular values. From (3.27), the parameter vector $\theta = [g^T \lambda^T]^T$, where $\Theta = \theta\theta^T$ can be calculated by

$$\begin{aligned}
\theta &= \sqrt{\sigma_1} U(:, 1)\\
g &= \theta(1 : N)\\
\lambda &= \theta(N+1 : N+M)
\end{aligned} \tag{3.28}$$

providing an optimal *rank* 1 approximation of $\Theta$.

### 3.5.2 Result of thermal actuator identification

The flying height was estimated based on servo sector measurements (128 servo sectors at 7200 rpm). Four different step inputs are applied to the thermal actuator system and 20 flying height variation outputs for each step input are measured and averaged. Then the inputs and outputs are concatenated to generate a staircase input and output as shown in Figure 3.8 and Figure 3.11 respectively. $M = 9$ with $m = [0\ .05\ .1\ .15\ .2\ .25\ .3\ .35\ max(u(t))]$ is used to model the input static nonlinearity. Once $\lambda$ is estimated (subsequently $\hat{x}$), a second order linear transfer function model is used to model the linear dynamics. The estimation results are shown in Figure 3.8, Figure 3.9, Figure 3.10, and Figure 3.11. The estimated intermediated power signal $\hat{x}$ is shown in Figure 3.8, the comparison between the measured output $y$ and simulated output $\hat{y}$ is shown in Figure 3.11, and the identified static nonlinearity and the identified linear dynamic system are shown in Figure 3.9 and Figure 3.10 respectively.

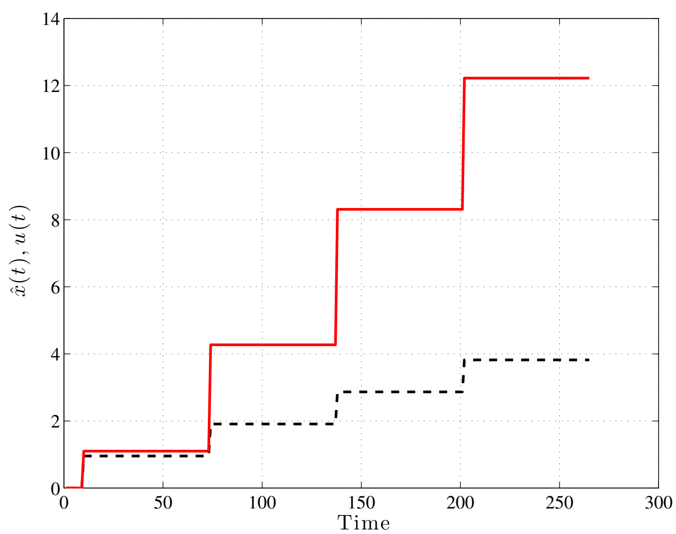



**Figure 3.8**: Concatenated input signal $u$ (black dotted line, 10× magnified) and estimated intermediate signal $\hat{x}$ (colored solid line).

When the voltage input is low, the quadratic relationship between the input voltage and the intermediate power signal is well preserved. As the input increases (the read/write head gets close to the disk), the characteristics of the static nonlinearity changes due to the head's extreme proximity to the disk. The experimental

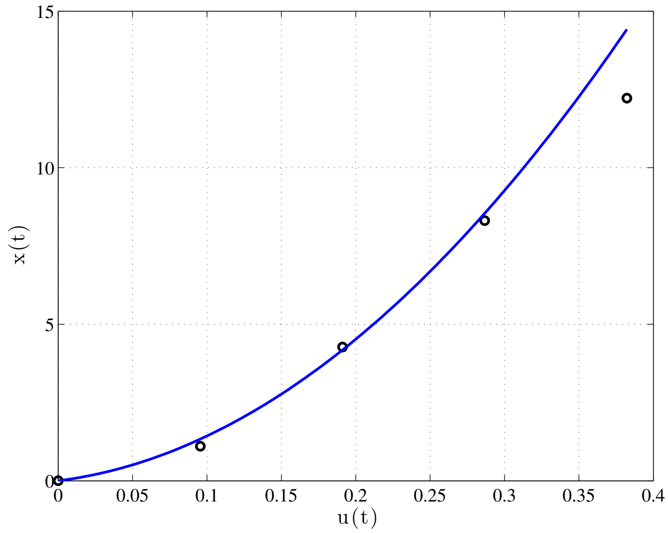



**Figure 3.9**: Identified input static nonlinear block (black circles) and its quadratic approximation (colored solid line).

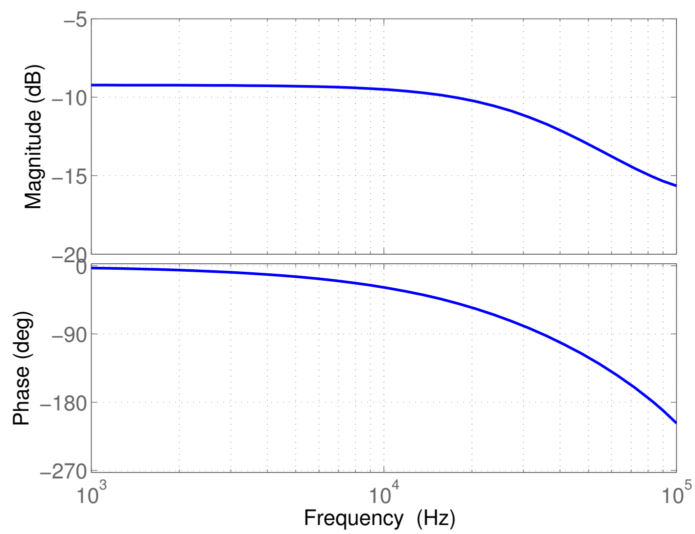



**Figure 3.10**: Identified linear dynamic block.

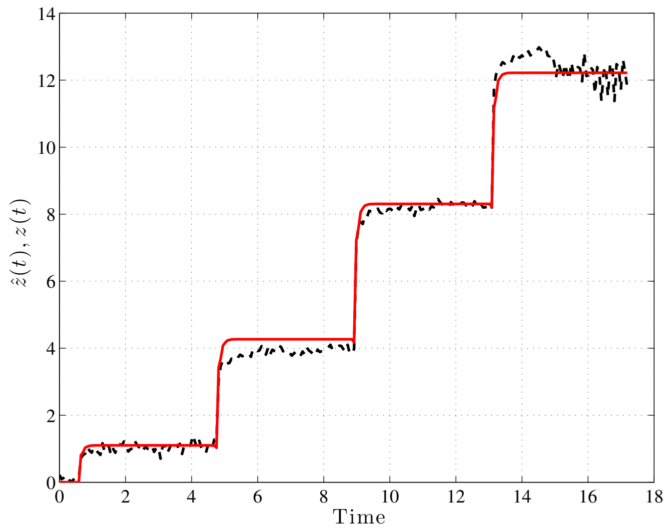



**Figure 3.11**: Concatenated measured output $y$ (black dotted line) and simulated output $\hat{y}$ (colored solid line).

study shows that the Hammerstein model captures the nonlinear dynamics of the thermal actuator sufficiently well and the proposed identification method provides an efficient way to identify Hammerstein system parameters.

## 3.6 Conclusion

In this chapter, the Hammerstein system parameter identification problem is formulated as a nuclear norm minimization problem. First, the system parameter identification problem is formulated as a rank minimization problem to reconstruct the intermediate signal between the static nonlinearity and the linear dynamics in a Hammerstein system. This non-convex optimization problem is then reformulated as a convex optimization problem using a nuclear norm relaxation. Once the system parameters for the static nonlinearity are estimated, an intermediate signal can be created to facilitate the identification of the linear dynamic system. The main assumption used in this study is that static nonlinearity is monotonically non-decreasing in order to guarantee a unique combination of a static nonlinear block and Finite Impulse Response (FIR) linear block. The proposed identification method is applied to simulation data and a slightly modified version of this method is applied to identify a nonlinear dynamic model of a thermal actuator in a hard disk drive. The numerical simulation and experimental result shows the effectiveness of the proposed identification method. The materials in Chapter 3 have been accepted for publication in Y. Han and R. de Callafon, Hammerstein system identification using nuclear norm minimization, Automatica, to appear 2012. The dissertation author was the primary investigator and author of this paper.

# 4

# Open-loop Identification of Wiener Systems

## 4.1 Introduction

A Wiener system has a block oriented structure where a linear dynamical system and a output static nonlinearity are separated, as shown in Figure 4.1. Wiener structure is a good approach to modeling systems with sensor nonlinearity or other nonlinear effects that can be brought to the system output [25] [29] [35] [67]. It has been shown that Wiener models can be used to effectively capture various nonlinear dynamics, such as chemical processes and biological systems [10] [29] [33]. The identification of Wiener systems involves estimating the parameters describing the linear dynamical and the output static nonlinear blocks from the measured input and output data. A comprehensive overview of block-oriented nonlinear system identification, including Wiener systems, can be found in [29]. The most common assumptions used in Wiener system identification are the Gaussian assumption of the input signal and the invertibility of the static nonlinearity. These assumptions are popular because, if the input signal is Gaussian noise, the identification of the linear dynamical block can be separated from the identification of the static nonlinear function based on separability assumption [7] [8] [22] [31] and parameterization of the output static nonlinearity is possible for the inverse of the given static nonlinearity. However, the Gaussian input assumption is too restrictive for practical application and the invertibility of the static nonlinearity assumption excludes hard nonlinearities, such as saturation, common in control systems.

Recently, a system identification method was introduced based on the sector bound property of static nonlinearity using Quadratic Programming (QP) in [68]. This monotonicity assumption on the unknown static nonlinearity guarantees a solution for an Finite Impulse Response (FIR) linear system and leads to possible nonparametric identification of static nonlinear function [55].

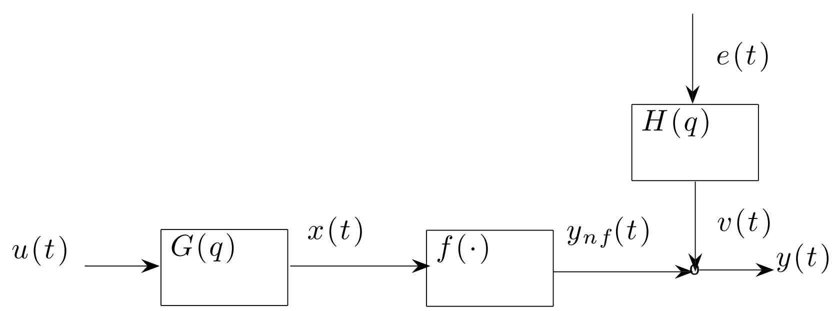



**Figure 4.1**: Wiener system with output noise.

In this study, the monotonicity assumption on the unknown static nonlinearity is utilized for nonparametric identification of static nonlinear function. The main contribution of this study is that the proposed method is robust to output noise and neither the Gaussian assumption of the input signal nor the invertibility of the static nonlinearity is necessary.

## 4.2 Problem description

The system to be modeled is a Wiener system as shown in Figure 4.1. The purpose of this study is to propose a method to identify the unknown linear dynamical systems $G(q)$ and a static nonlinear function $f(\cdot)$ from a finite number of observations of the data $u(t)$ and $y(t)$. This is done in a novel way by the reconstruction of the intermediate signal $x(t)$ and the noise free output signal $y_{nf}(t)$ with conditions on the finite dimensional dynamical representation of the linear systems $G(q)$ and the memoryless static nonlinearity $f(\cdot)$. The output disturbance $v(t)$ is filtered zero-mean white noise independent of the input signal, where the filtering properties are unknown. As mentioned before, finding the simplest model within a feasible model set restricted by convex constraints can often be formulated as a rank minimization problem [27]. Based on this idea, in this study, the rank minimization problem is used to formulate a convex optimization problem via Semidefinite Programming (SDP) relaxation. The system parameters will be estimated by finding a feasible model consistent with the input and output data, and satisfying the following basic properties of the Wiener system:

**Condition 3**

*I. The static nonlinear function has no memory:*
*The current output* $y_{nf}(t)$ *only depends on*
*the current input* $x(t)$.
*II. The linear dynamical system has a finite, but*
*unknown, McMillan degree* $n$:
$x(t) = \phi^T(t)\theta$, *where*
$\phi^T(t) = [u(t) \ \cdots \ u(t-n_b) \ x(t-1) \ \cdots \ x(t-n_a)]$,
$\theta$ *is the linear system parameter,*
*and* $n \leq max(n_b - 1, n_a)$.

The intermediate signal $x(t)$ and the noise free output signal $y_{nf}(t)$ in Figure 4.1 are not measurable. The unknown signals will be parametrized and the estimation of the unknown coefficients will be formulated as a SDP problem. Let $\hat{x}(t)$ be the reconstructed signal of $x(t)$ and $\hat{y}_{nf}(t)$ be the reconstructed signal of $y_{nf}(t)$. The SDP problem will be formulated in such a way that $\hat{x}(t)$ and $\hat{y}_{nf}(t)$ are related via a memoryless static nonlinearity, $u(t)$ and $\hat{x}(t)$ are related via a linear dynamical system with the smallest McMillan degree, and $||y - \hat{y}_{nf}||_2$ is minimized under Condition 3. Once $\hat{x}(t)$ and $\hat{y}_{nf}(t)$ have been reconstructed, the identification of $G(q)$ from $u(t)$ to $\hat{x}(t)$ can be solved with a standard Prediction Error (PE) identification method in [49] and the identification of $f(\cdot)$ from $\hat{x}(t)$ to $\hat{y}_{nf}(t)$ can be solved via the Least Squares (LS) method.

## 4.3 System parametrization

### 4.3.1 Input-output map of the linear dynamical system

Let $g(k),\ k = 0, 1, \cdots$ be a causal sequence of unit impulse responses of $G(q)$. The relationship between the input $u(t)$ and the intermediate signal $x(t)$ can be described by the convolution as

$$x(t) = \sum_{k=0}^{\infty} g(k)u(t-k).$$

Due to Condition 3 (finite McMillan degree), the Hankel matrix defined as

$$H = \begin{bmatrix} g(1) & \cdots & g(N/2) \\ g(2) & \cdots & g(N/2+1) \\ \vdots & \ddots & \vdots \\ g(N/2) & \cdots & \cdots g(N-1) \end{bmatrix}, \tag{4.1}$$

has a $rank(H) \leq n$. The order of the linear dynamical system is determined by the $rank(H)$ as $H$ is simply the product of the extended observability and controllability matrices [30]. A lower order model, consistent with the input and output signals can be estimated by minimizing the rank of $H$. Let

$$\hat{x} = [\hat{x}(1)\ \hat{x}(2)\ \cdots\ \hat{x}(N)]^T$$

and

$$U = \begin{bmatrix} u(1) & u(0) & \cdots & u(2-N) \\ u(2) & u(1) & \cdots & u(1-N) \\ \vdots & \vdots & \ddots & \vdots \\ u(N) & u(N-1) & \cdots & u(1) \end{bmatrix}. \tag{4.2}$$

With

$$g = [g(0)\ g(1)\ \cdots\ g(N-1)]^T, \tag{4.3}$$

the finite sequence of the input

$$u = [u(1)\ u(2)\ \cdots\ u(N)]^T$$

and the estimate of the intermediate signal $x$ can be written as

$$\hat{x} = Ug.$$

### 4.3.2 Characteristics of static nonlinearity

Let $\hat{y}_{nf}$ be the noise free output defined as

$$\hat{y}_{nf} = [\hat{y}_{nf}(1)\ \cdots\ \hat{y}_{nf}(N)]^T. \tag{4.4}$$

In this chapter, a monotonically non-decreasing static nonlinearity with the maximum slope of 1 is considered as in Chapter 3:

**Condition 4**

*I. The static nonlinear function is monotonically non-decreasing with the maximum slope of* 1*:*
$(\hat{y}_{nf}(i) - \hat{y}_{nf}(j))(\hat{y}_{nf}(i) - \hat{y}_{nf}(j) - \hat{x}(i) + \hat{x}(j)) \le 0$
$\forall i > j$.

In Condition 4,

$$\hat{y}_{nf}(i) - \hat{y}_{nf}(j) \ge 0 \Rightarrow \hat{y}_{nf}(i) - \hat{y}_{nf}(j) \le \hat{x}(i) - \hat{x}(j)$$

or

$$\hat{y}_{nf}(i) - \hat{y}_{nf}(j) \le 0 \Rightarrow \hat{y}_{nf}(i) - \hat{y}_{nf}(j) \ge \hat{x}(i) - \hat{x}(j).$$

In both cases,

$$\hat{x}(i) - \hat{x}(j) = 0 \Rightarrow \hat{y}_{nf}(i) - \hat{y}_{nf}(j) = 0$$

or

$$\hat{x}(i) - \hat{x}(j) \ne 0 \Rightarrow \frac{\hat{y}_{nf}(i) - \hat{y}_{nf}(j)}{\hat{x}(i) - \hat{x}(j)} \le 1.$$

Condition 4 implies that once $\hat{x}(t)$ is chosen, $\hat{y}_{nf}(t)$ is determined as

$$\hat{y}_{nf} = \alpha(t)\hat{x}(t),\ 0 \le \alpha(t) \le 1.$$

This implies that the cross-covariance function between $\hat{x}(t)$ and $\hat{y}_{nf}(t)$ only depends on the static nonlinearity, characterized by $\alpha(t)$, and the auto-covariance of $\hat{x}(t)$, not $\tau$ as

$$\begin{aligned} R_{yx}(\tau) &= \frac{1}{N}\sum_{t=1}^{N}\hat{y}_{nf}(t)\hat{x}(t-\tau) \\ &= \frac{1}{N}\sum_{t=1}^{N}\alpha(t)\hat{x}(t)\hat{x}(t-\tau). \end{aligned}$$

Thus, Condition 4 guarantees that the intermediate signal $\hat{x}(t)$ and the output $\hat{y}_{nf}(t)$ are related by a static nonlinear function.

## 4.4 Parameter estimation

In this section, a rank minimization problem with the memoryless constraint on the static nonlinearity for the reconstruction of the intermediate signal $x(t)$ and the noise free output signal $y_{nf}(t)$ in Figure 4.1 is summarized and the optimization problem is constructed. With the parametrization and constraints explained in the previous section, an optimization problem can be written as follows:

**Optimization Problem 4**

*Consider*

*variables* $g$ *in (4.3) and* $\hat{y}_{nf}$ *in (4.4)*

*Define*

$\hat{x} = Ug$, *with* $U$ *in (4.2)*

*Minimize*

$w_1 \cdot ||y - \hat{y}_{nf}||_2 + w_2 \cdot rank\ H$, *with* $H$ *in (4.1)*

*subject to*

$(\hat{y}_{nf}(i) - \hat{y}_{nf}(j))(\hat{y}_{nf}(i) - \hat{y}_{nf}(j) - \hat{x}(i) + \hat{x}(j)) \leq 0$

$\forall i > j$

*where*

$w_1$ *and* $w_2$ *are weighting factors*

Optimization Problem 4 results in the optimal solution for the system parameter $g$ that is used to construct the intermediate signal $x$ and the noise free output $y_{nf}$. In Optimization Problem 4, the reconstructed signals $\hat{x}$ and $\hat{y}_{nf}$ are generated in such a way that a static nonlinear function satisfies the monotonically non-decreasing condition, the linear dynamical system has the minimum order, and the prediction error is minimized under the chosen weighting and constraints.

Unfortunately, the rank condition and the constraint in Optimization Problem 4 are not convex. In this section, a new variable $\Theta$ is defined in order to convert the non-convex optimization problem to an approximated convex optimization problem, resulting in a Semidefinite Programming (SDP) problem. This SDP problem is easier to solve and the solution is close to the solution of the original non-convex problem [27].

Let $\theta = [g^T \hat{y}_{nf}^T]^T$. With $\theta$, let us define a positive semidefinite symmetric matrix $\Theta = \theta\theta^T$ as

$$\Theta = \begin{bmatrix} g(0)g(0) & \cdots & \cdots & g(0)\hat{y}_{nf}(N) \\ \vdots & \vdots & \vdots & \vdots \\ g(N)g(0) & \vdots & \vdots & g(N)\hat{y}_{nf}(N) \\ \hat{y}_{nf}(1)g(0) & \cdots & \cdots & \hat{y}_{nf}(1)\hat{y}_{nf}(N) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{y}_{nf}(N)g(0) & \cdots & \cdots & \hat{y}_{nf}(N)\hat{y}_{nf}(N) \end{bmatrix}. \tag{4.5}$$

Based on its structure, it is clear that $\Theta$ is a *rank* 1 matrix if there is no noise in the data. For cases where there is noise, system parameters will be found by minimizing $rank(\Theta)$. Because $\Theta$ is a square positive semidefinite matrix, minimizing its trace is the closest approximation of the rank minimization that can be efficiently solved. Without loss of generalization, the maximum slope 1 of the static nonlinearity combined with the minimization of $trace(\Theta)$ serves as a normalization condition on the static nonlinearity, so that the static gain of the Wiener system is modeled by the static gain of the linear system $G(q)$. Due to the over-parametrization of $\Theta$, it is impossible to access $\hat{y}_{nf}$ directly through $\Theta$. However, $\Theta$ contains information of $\hat{y}_{nf}\hat{y}_{nf}^T$. Thus, minimizing $||y - \hat{y}_{nf}||_2$ is relaxed to minimizing $||yy^T - \hat{y}_{nf}\hat{y}_{nf}^T||_F$, where $||\cdot||_F$ is a Frobenius norm. With $\Theta$, let us express the quadratic constraints in Optimization Problem 3 as Linear Matrix Inequalities (LMIs). Let

$$\delta Y = \begin{bmatrix} \hat{y}_{nf}(2) - \hat{y}_{nf}(1) & \cdots & 0 \\ 0 & \cdots & 0 \\ 0 & \vdots & 0 \\ 0 & \cdots & \hat{y}_{nf}(N) - \hat{y}_{nf}(1) \end{bmatrix}$$

be a diagonal matrix whose diagonal entries include $\hat{y}_{nf}(i) - \hat{y}_{nf}(j), \ \forall i > j$ and

$$\delta X = \begin{bmatrix} \hat{x}(2) - \hat{x}(1) & 0 & \cdots & 0 \\ 0 & \hat{x}(3) - \hat{x}(2) & \cdots & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & \cdots & \hat{x}(N) - \hat{x}(1) \end{bmatrix}$$

be a diagonal matrix whose diagonal entries include $\hat{x}(i) - \hat{x}(j), \ \forall i > j$. Then $diag(\delta Y) = \Delta Y y$, where

$$\Delta Y = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 \cdots & 0 \\ 0 & 0 & 1 & 0 \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & 0 & 0 \cdots & \cdots 0 & 1 \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & \vdots \\ 0 & \vdots & \cdots & \cdots & \vdots \\ 1 & 0 & \cdots & \cdots & 0 \end{bmatrix}$$

and $diag(\delta X) = \Delta X h$, where

$$\Delta X = \begin{bmatrix} u_2 & u_1 & 0 & \cdots & 0 \\ u_3 & u_2 & u_1 & 0 \cdots & 0 \\ u_3 & u_2 & u_1 & 0 \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ u_N & \cdots & \cdots & \cdots & u_1 \end{bmatrix} - \begin{bmatrix} u_1 & 0 & \cdots & \cdots & 0 \\ u_2 & u_1 & 0 & \cdots & 0 \\ u_1 & 0 & \cdots & \cdots & 0 \\ u_3 & u_2 & u_1 & 0 \cdots & 0 \\ u_2 & u_1 & 0 \cdots & \cdots & 0 \\ u_1 & 0 & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ u_1 & 0 & \cdots & \cdots & 0 \end{bmatrix}$$

where $u_i$ is used instead of $u(i)$ for notational brevity. Then, the constraints in Optimization Problem 4 can be written as

$$\delta Y^T \delta Y - \delta Y^T \delta X \leq 0.$$

where

$$\delta Y^T \delta Y = diag(diag(\Delta Y \tilde{\tilde{\Theta}} \Delta Y^T))$$
$$\text{where } \tilde{\tilde{\Theta}} = \Theta(N+1:2N, N+1:2N)$$

and

$$\delta Y^T \delta X = diag(diag(\Delta X \tilde{\Theta} \Delta Y^T))$$
$$\text{where } \tilde{\Theta} = \Theta(N+1:2N, 1:N)$$

where the notation $(k,:)$ and $(:,k)$ are used to denote the $k^{th}$ row and the $k^{th}$ column in a matrix respectively. Here $diag(x)$ indicates a square matrix with the elements of a vector $x$ on the diagonal, and $diag(X)$ indicates the main diagonal of a matrix $X$.

Using SDP relaxation, Optimization Problem 4 can be rewritten as the following convex optimization problem:

**Optimization Problem 5**

$$
\begin{aligned}
&\textit{Consider} \\
&\quad \textit{variable symmetric} \;\; \Theta \\
&\textit{Minimize} \\
&\quad w_1 \cdot ||\tilde{\tilde{\Theta}} - yy^T||_F + w_2 \cdot trace(\Theta) \\
&\textit{subject to} \\
&\delta Y^T \delta Y - \delta Y^T \delta X \leq 0 \\
&\Theta \geq 0 \\
&\textit{where} \\
&\quad \delta Y^T \delta Y = diag(diag(\Delta Y \tilde{\tilde{\Theta}} \Delta Y^T)) \\
&\quad \delta Y^T \delta X = diag(diag(\Delta X \tilde{\Theta} \Delta Y^T)) \\
&\textit{where} \\
&\quad \tilde{\tilde{\Theta}} = \Theta(N+1:2N, N+1:2N) \\
&\quad \tilde{\Theta} = \Theta(N+1:2N, 1:N) \\
&\quad || \cdot ||_F \;\; \textit{is a Frobenius norm} \\
&\textit{where} \\
&\quad w_1 \;\; \textit{and} \;\; w_2 \;\; \textit{are weighting factors.}
\end{aligned}
$$

Due to the over-parametrization used to define $\Theta$ in (4.5), we need to separate the parameters of the linear dynamical system $g$ and the noise free output $\hat{y}_{nf}$. Singular Value Decomposition (SVD) is used in this study to separate the system parameters. The SVD of $\Theta$ is given as

$$\Theta = U \Sigma V^T \tag{4.6}$$

where $U_{2N \times 2N}$ and $V_{2N \times 2N}$ are orthogonal matrices, $U_{2N \times 2N} = V_{2N \times 2N}$ due to the structure of $\Theta$, and $\Sigma_{2N \times 2N}$ is a rectangular diagonal matrix. The positive diagonal entries of $\Sigma$ are called singular values. From (4.6), the parameter vector $\theta = [g^T \hat{y}_{nf}^T]^T$, where $\Theta = \theta \theta^T$ can be calculated by

$$
\begin{aligned}
\theta &= \sqrt{\sigma_1} U(:, 1) \\
g &= \theta(1:N) \\
\hat{y}_{nf} &= \theta(N+1:2N)
\end{aligned}
$$

providing an optimal $rank$ 1 approximation of $\Theta$.

## 4.5 Numerical example

In this section, a numerical example of Wiener system identification using the proposed identification method is presented. A Pseudo Random Binary Sequence (PRBS) excitation signal, defined as

$$u(t) = 4 \cdot sign(randn(N, 1))$$

where

$$u(t) = \begin{cases} +1 & w.p. \quad \frac{1}{2} \\ -1 & w.p. \quad \frac{1}{2} \end{cases},$$

is used as the input. The output disturbance $v(t) = H(q)e(t)$ is filtered zero-mean white noise independent of the input signal, where the filtering properties, $H(q)$, are not estimated or not need to be known. For the system identification, twenty sets of estimation data with 100 samples are generated from the Wiener system with the following specifications:

$$\begin{cases} \text{Linear dynamical system:} \\ G(q) = \dfrac{0.0997q^{-1} - 0.0902q^{-2}}{1 - 1.886q^{-1} + 0.9048q^{-2}} \\ \text{Static nonlinearity:} \\ \quad f(x(t)) = \begin{cases} .5 & \text{if} \quad x(t) > .5 \\ x(t) & \text{if} \quad |x(t)| \leq .5 \\ -.5 & \text{if} \quad x(t) < -.5 \end{cases} \\ \text{Noise dynamics:} \\ H(q) = \dfrac{1 + 0.5q^{-1}}{1 - 0.85q^{-1}} \end{cases}.$$

The input and output signals are shown in Figure 4.2. In order to solve the Semidefinite Programming (SDP) problem (Optimization Problem 5), SEDUMI [59] and YALMIP [50] are used. The estimation results are shown in Figure 4.3, Figure 4.4 and Figure 4.5. As shown in Figure 4.3 and Figure 4.4, both pole and zero locations are well estimated. As shown in Figure 4.5, $\pm.5$ saturation is well identified.

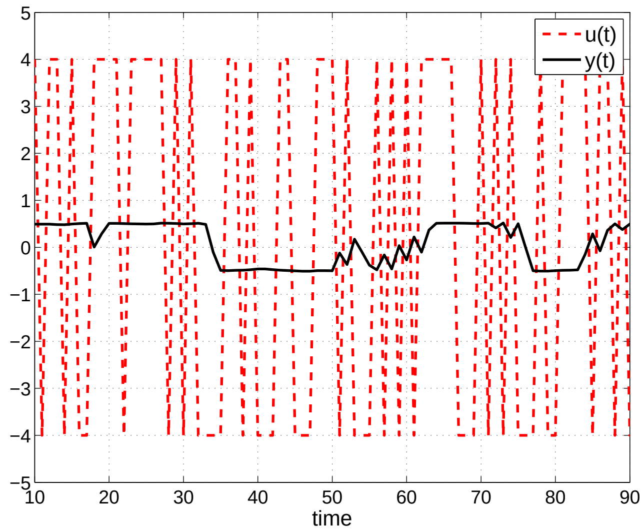



**Figure 4.2**: The input and output signals.

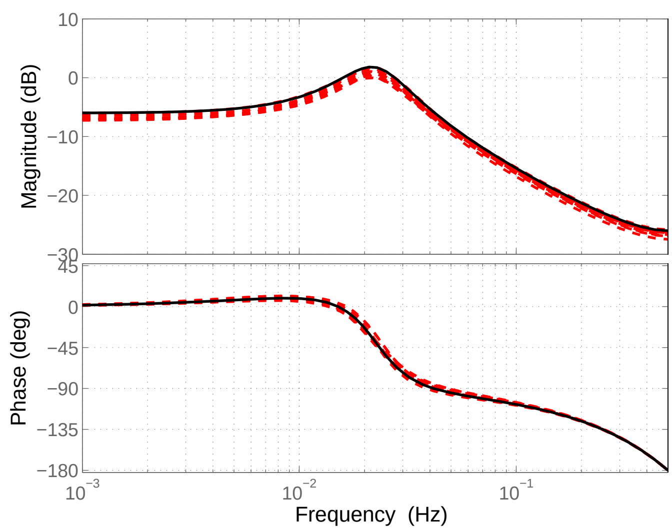



**Figure 4.3**: The Bode plot of the identified linear dynamical system. The black solid line indicates the real linear dynamical system. The (colored) dashed lines indicate estimated linear dynamical systems by using twenty different sets of data. The SNR of each data set is greater than 50dB.

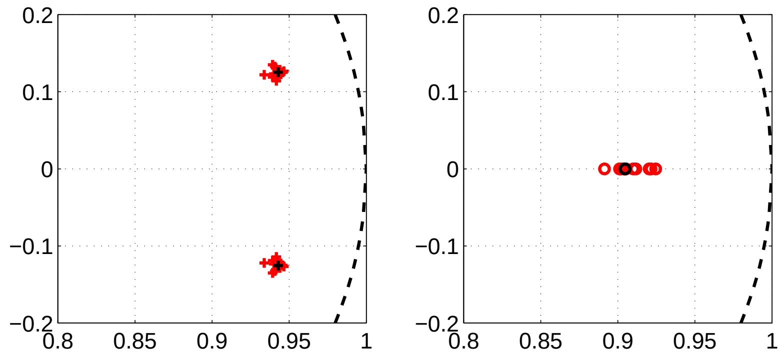



**Figure 4.4**: Pole (left figure) and zero (right figure) locations of the identified linear dynamical system. The black cross and circle indicate the real linear dynamical system. The colored crosses and circles indicate estimated linear dynamical systems. The SNR of each data set is greater than 50dB.

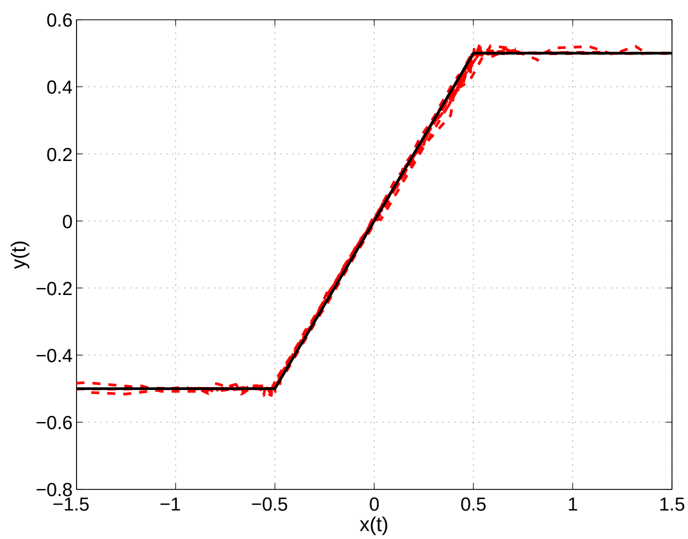



**Figure 4.5**: The plot of the identified static nonlinear function. The black solid line indicates the real static nonlinear function. The (colored) dashed lines indicate estimated static nonlinear functions by using twenty different sets of data. The SNR of each data set is greater than 50dB.

## 4.6 Conclusion

In this chapter, the Wiener system identification problem is formulated as a Semidefinite Programming (SDP) problem to reconstruct the intermediate signal and noise free output. The system parameter identification problem is formulated as a rank minimization problem by imposing the monotonically non-decreasing condition on the static nonlinear function. This non-convex optimization problem is then reformulated as a convex optimization problem via SDP relaxation by using over-parametrization. The proposed method is robust to output noise and neither the Gaussian assumption of the input signal nor the invertibility of the static nonlinearity is necessary. Singular Value Decomposition (SVD) is used to separate the linear system parameters and the noise free output signal. Once the intermediate signal and noise free output signal are reconstructed, the identification of the linear dynamical system and the static nonlinear function become trivial. The proposed identification method is applied to simulation data from a Wiener system. The numerical simulation result shows the effectiveness of the proposed identification method. The materials in Chapter 4 have been accepted for publication in Y. Han and R. de Callafon, 2012, Identification of a Wiener System via Semidefinite Programming, 16th IFAC Symposium on System Identification 2012, Brussels, Belgium. The dissertation author was the primary investigator and author of this paper.

# 5

# Open-loop Identification of Wiener-Hammerstein Systems

## 5.1 Introduction

Wiener-Hammerstein systems are dynamical systems characterized by a series connection of three parts: a linear dynamical system, a static nonlinearity and another linear dynamical system, as shown in Figure 5.1. This structure

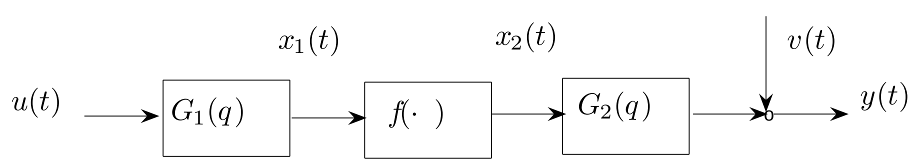



**Figure 5.1**: Wiener-Hammerstein system consisting of the cascade of a linear dynamic block $G_1(q)$, a static non-linear block $f(\cdot)$ and another linear dynamic block $G_2(q)$.

can be used to represent sensor systems, electromechanical systems in robotics, mechatronics, biological and chemical systems [29] [39]. Early works on Wiener-Hammerstein system identification can be found in [7] [8]. In this early research, the correlation analysis-based identification method under Gaussian excitation has been proposed. The authors in [15] introduced a time-domain identification method based on the Maximum Likelihood principle. The authors in [12] presented a simple technique for recursive identification of the Wiener-Hammerstein model with extension to the multi-input single-output (MISO) case. More recent work can be found in [2] [17] [22] [52] [54] [57] [65]. The authors in [54] proposed an identification method using the polynomial nonlinear state space (PNLSS) approach. The authors in [2] presented a method iteratively identifying the linear system and the Hammerstein system by minimizing the square norm of output prediction error and by using the orthogonal decomposition subspace method (ORT).

Recently, a system identification method was introduced based on the sector bound property of static nonlinearity using Quadratic Programming (QP) and semidefinite programming (SDP) relaxation in [57] [68]. In [57], the identification problem was formulated as a non-convex QP. A convex SDP relaxation is then formulated and solved to obtain a sub-optimal solution to the original non-convex QP. However, the formulation of the problem is based on the existence of the inverse of the second dynamic system, which cannot be generally guaranteed.

In this chapter, the SDP relaxation approach by [57] is extended by using rank minimization to propose a Wiener-Hammerstein system identification method which does not require invertibility of any sub-systems. As mentioned before, choosing the simplest model in the set of feasible models that is described by convex constraints can often be expressed as a rank minimization problem [27]. Based on this idea, in this chapter a Wiener-Hammerstein system identification problem is formulated as a rank minimization problem and the non-convex rank minimization problem is then formulated as a convex problem via SDP relaxation.

The objective of this study is to formulate a procedure that allows the characterization and identification of the three parts in a a Wiener-Hammerstein system individually based on the finite number of input $u(t)$ and the output $y(t)$ observations. In this study, this is accomplished by reconstructing unmeasurable intermediate signals $x_1(t)$ and $x_2(t)$ that satisfy the conditions on the finite dimensional dynamical representation of the linear systems $G_1(q)$ and $G_2(q)$, and the memoryless static nonlinearity $f(\cdot)$. Once the intermediate signals $x_1(t)$ and $x_2(t)$ are reconstructed, the identification and characterization of the three parts becomes trivial.

## 5.2 Problem description

The system to be modeled is a Wiener-Hammerstein system as shown in Figure 5.1. In this study, each block is parametrized separately. System parameters for each block will be estimated simultaneously by finding feasible models consistent with the input and output data, and by satisfying the following basic properties of the Wiener-Hammerstein system:

**Condition 5**

*I. The static nonlinear function has no memory:*
*The current output $x_2(t)$ only depends on*
*the current input $x_1(t)$.*

*II. The first linear dynamic system has a finite, but*
*unknown, McMillan degree $n_1$:*
$x_1(t) = \phi_1^T(t)\theta_1$, *where*
$\phi_1^T(t) = [u(t) \ \cdots \ u(t-n_b) \ x_1(t-1) \ \cdots \ x_1(t-n_a)]$,
$\theta_1$ *is the first linear system parameter,*
*and* $n_1 \leq max(n_b - 1, n_a)$.

*III. The second linear dynamic system has a finite, but*
*unknown, McMillan degree $n_2$:*
$y(t) = \phi_2^T(t)\theta_2$, *where*
$\phi_2^T(t) = [x_2(t) \ \cdots \ x_2(t-n_d) \ y(t-1) \ \cdots \ y(t-n_c)]$,
$\theta_2$ *is the second linear system parameter,*
*and* $n_2 \leq max(n_d - 1, n_c)$.

The intermediate signals $x_1(t)$ and $x_2(t)$ in Figure 5.1 are not measurable, and the properties in Condition 5 are used to formulate a procedure to reconstruct $x_1(t)$ and $x_2(t)$. The unknown signals $x_1(t)$ and $x_2(t)$ will be parametrized, and the estimation of the unknown coefficients will be formulated as a SDP problem. Let $\hat{x}_1(t)$ be the reconstructed signal of $x_1(t)$, $\hat{x}_2(t)$ be the reconstructed signal of

$x_2(t)$, and $\hat{y}(t)$ be the model output. The SDP problem will be formulated in such a way that $\hat{x}_1(t)$ and $\hat{x}_2(t)$ are related via a memoryless static nonlinearity, $u(t)$ and $\hat{x}_1(t)$ are related via a linear dynamical system with the smallest McMillan degree, and $||y - \hat{y}||_2$ is minimized under Condition 5. Once $\hat{x}_1(t)$ and $\hat{x}_2(t)$ have been reconstructed, the identification of $G_1(q)$ from $u(t)$ to $\hat{x}_1(t)$, and the identification of $G_2(q)$ from $\hat{x}_2(t)$ to $y(t)$ can be solved with the standard Prediction Error (PE) identification method in [49] and the identification of $f(\cdot)$ from $\hat{x}_1(t)$ to $\hat{x}_2(t)$ can be solved via the Least Squares (LS) method.

## 5.3 System parametrization

### 5.3.1 Input-output map of the first dynamic system

In order to formulate the parameter estimation problem, a finite impulse response (FIR) model is used to model the first dynamic system $G_1$. Let $g_k,\ k = 0, 1, \cdots$ be a causal sequence of unit impulse responses for $G_1(q)$. The relationship between the input $u(t)$ and the intermediate signal $x_1(t)$ can be described by the convolution as

$$x_1(t) = \sum_{k=0}^{\infty} g_k u(t-k).$$

Due to Condition 5 (finite McMillan degree), the Hankel matrix defined as

$$H = \begin{bmatrix} g(1) & \cdots & g(N/2) \\ g(2) & \cdots & g(N/2+1) \\ \vdots & \ddots & \vdots \\ g(N/2) & \cdots & \cdots g(N-1) \end{bmatrix} \tag{5.1}$$

has a $rank(H) \leq n_1$. The order of the linear dynamical system is determined by the $rank(H)$ as $H$ is simply the product of the extended observability and controllability matrices [30]. Let

$$\hat{x}_1 = [\hat{x}_1(1)\ \hat{x}_1(2)\ \cdots\ \hat{x}_1(N)]^T$$

and

$$U = \begin{bmatrix} u(1) & u(0) & \cdots & u(2-N) \\ u(2) & u(1) & \cdots & u(1-N) \\ \vdots & \vdots & \vdots & \\ u(N) & u(N-1) & \cdots & u(1) \end{bmatrix}. \tag{5.2}$$

With the system parameter

$$g = [g_0 \; g_1 \; \cdots \; g_{N-1}]^T \tag{5.3}$$

to be estimated, $\hat{x}_1$ can be written as

$$\hat{x}_1 = Ug. \tag{5.4}$$

The finite order sequence of $g_k,\ k = 0, 1, \cdots, N-1$, for a lower order model for $G_1$ can be estimated by minimizing the rank of $H$ in (5.1) [27]. Here, the rank minimization of $H$ is used only to minimize the order of $G_1$. The FIR approximation of $G_1$ is used to formulate a convex optimization problem to estimate system parameters in Section 5.4. Once $\hat{x}_1(t)$ has been reconstructed, the identification of $G_1(q)$ from $u(t)$ to $\hat{x}_1(t)$ can be solved with the standard Prediction Error (PE) identification method in [49]. Based on (5.4), the error is defined by

$$\begin{aligned} e(t) &= x_1(t) - \hat{x}_1(t) \\ &= x_1(t) - \sum_{k=0}^{L-1} g_k u(t-k). \end{aligned}$$

Thus,

$$\begin{aligned} ||e(t)||_2^2 &= \sum_{t=1}^{N} \left[ x_1(t) - \sum_{k=0}^{L-1} g_k u(t-k) \right]^2 \\ &= \sum_{t=1}^{N} x_1^2(t) - 2\sum_{k=0}^{L-1} g_k R_{x_1 u}(k) \\ &\quad + \sum_{k=0}^{L-1}\sum_{l=0}^{L-1} g_k g_l R_{uu}(k-l) \end{aligned}$$

where

$$R_{x_1u}(k) = \sum_{t=1}^{N} x_1(t)u(t-k)$$
$$R_{uu}(k) = \sum_{t=1-k-l}^{N-k-l} u(t)u(t+k).$$

If $L$ tends toward infinity, the $g_k$ obtained by minimizing the $rank(H)$ will satisfy $|g_k| << 1$ for $k \geq L$, resulting in $R_{x_1u}(k) \rightarrow \sum_{l=0}^{L-1} g_l R_{uu}(l-k)$. Then,

$$\lim_{N\rightarrow\infty, L\rightarrow\infty} ||e(t)||_2^2 = 0 \tag{5.5}$$

As a result, the estimate $\hat{x}_1(t)$ in (5.4) will converge to $x_1(t)$ provided that $N \rightarrow \infty$ and $L \rightarrow \infty$.

### 5.3.2 Modeling of static nonlinearity

In this section, a $n_f^{th}$ order polynomial function is used to model static nonlinearity. With the polynomial basis functions, $\hat{x}_2(t)$ is defined by

$$\hat{x}_2(t) = \lambda_0 + \lambda_1 \hat{x}_1(t) + \lambda_2 \hat{x}_1^2(t) + \cdots + \lambda_{n_f} \hat{x}_1^{n_f}(t). \tag{5.6}$$

Weierstrass's Theorem in Section 2.2 guarantees that the polynomial approximation $\hat{x}_2(t)$ in (5.6) will converge to $x_2(t)$ as $n_f$ tends toward to infinity for an arbitrary interval. There could be many possible combinations of $(\hat{x}_1(t), \hat{x}_2(t))$ that satisfy Condition 5 and (5.6). In order to limit the number of possible selections of $(\hat{x}_1(t), \hat{x}_2(t))$, it is assumed that the static nonlinearity is monotonically non-decreasing with the maximum slope of 1 as in Chapters 1 and 2:

**Condition 6**

*I. The static nonlinear function is monotonically non-decreasing with the maximum slope of* 1*:*
$(\hat{x}_2(i) - \hat{x}_2(j))(\hat{x}_2(i) - \hat{x}_2(j) - \hat{x}_1(i) + \hat{x}_1(j)) \leq 0$
$\forall i > j.$

The maximum slope of the static nonlinearity is a user-chosen value, thus can be adjusted by a user. In this section, we assumed that $\lambda_0 = 0$ and $\lambda_1 = 1$. The assumptions are not necessary for the proposed method, but chosen for notational brevity and normalization. Without loss of generalization, this monotonicity assumption on the unknown static nonlinearity combined with the given assumptions ($\lambda_0 = 0$, $\lambda_1 = 1$, and the maximum slope $\leq 1$) guarantees a solution for an FIR linear system for $G_1$ and serves as a normalization condition on the static nonlinearity. Based on (5.4) and (5.6), $\hat{x}_2 = [\hat{x}_2(1) \ \cdots \ \hat{x}_2(N)]^T$ can be calculated as

$$\hat{x}_2 = Ug + X_1\lambda \tag{5.7}$$

where

$$\lambda = [\lambda_2 \ \cdots \ \lambda_{n_f}]^T \tag{5.8}$$

and

$$X_1 = \begin{bmatrix} \hat{x}_1^2(1) & \cdots & \hat{x}_1^{n_f}(1) \\ \vdots & \vdots & \vdots \\ \hat{x}_1^2(N) & \cdots & \hat{x}_1^{n_f}(N) \end{bmatrix}. \tag{5.9}$$

An iterative approach will be used to update the higher order nonlinear terms of $\hat{x}_1$ in (5.9) that are included in the description of $\hat{x}_2$ in (5.7).

### 5.3.3 Modeling of the second dynamic system

Since the output of the Wiener-Hammerstein system is measured, a rational transfer function is used to model the second dynamic system $G_2$. The model output $\hat{y}(t)$ is defined as

$$\begin{aligned} \hat{y}(t) &= G_2(q)\hat{x}_2(t) \\ &= \frac{D(q)}{C(q)}\hat{x}_2(t) \end{aligned}$$

where

$$\begin{aligned} C(q) &= 1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c} \\ D(q) &= d_0 + d_1 q^{-1} + \cdots + d_{n_d} q^{-n_d}. \end{aligned} \tag{5.10}$$

Using the system parameters in (5.10), the linear difference equation between the output $\hat{y}(t)$ and the intermediate signal $\hat{x}_2(t)$ is defined as

$$\hat{y}(t) = -\sum_{k=1}^{n_c} c_k y(t-k) + \sum_{k=0}^{n_d} d_k \hat{x}_2(t-k).$$

Let

$$\begin{aligned} \Gamma(t) &= \sum_{k=0}^{n_d} d_k \hat{x}_2(t-k) \\ &= \sum_{k=0}^{n_d} d_k U(t-k,:)g + \sum_{k=0}^{n_d} d_k X_1(t-k,:)\lambda \end{aligned} \tag{5.11}$$

and

$$\begin{aligned} T &= Ugd^T + X_1 \lambda d^T \\ &= \begin{bmatrix} T_{1,1} & \cdots & T_{1,n_d+1} \\ \vdots & \vdots & \vdots \\ T_{N,1} & \cdots & T_{N,n_d+1} \end{bmatrix} \end{aligned}$$

using $U$ in (5.2), $g$ in (5.3), $\lambda$ in (5.8) and $X_1$ in (5.9), where

$$d = [d_0 \;\cdots\; d_{n_d}]^T, \tag{5.12}$$

and the notations $(k,:)$ and $(:,k)$ are used to denote the $k^{th}$ row and the $k^{th}$ column in a matrix respectively. Then $\Gamma(t)$ in (5.11) can be rewritten as

$$\Gamma(t) = \sum_{k=1}^{min(t,n_d)} T_{t-k+1,k}.$$

Using the given parameterization, the model output vector $\hat{y} = [\hat{y}(1) \;\cdots\; \hat{y}(N)]^T$ can be written as

$$\hat{y} = Yc + \Gamma \tag{5.13}$$

where

$$c = [c_1 \;\cdots\; c_{n_c}]^T, \tag{5.14}$$

$$Y = \begin{bmatrix} -y(1-1) & \cdots & -y(1-n_c) \\ \vdots & \vdots & \vdots \\ -y(N-1) & \cdots & -y(N-n_c) \end{bmatrix}, \tag{5.15}$$

and

$$\Gamma = [\Gamma(1) \;\cdots\; \Gamma(N)]^T. \tag{5.16}$$

With the system parametrization in Section 5.3 based on Condition 5, system parameters for a lower order model for $G_1$, consistent with the input and output measurement data, can be estimated by minimizing $||y - \hat{y}||_2$, with $\hat{y}$ in (5.13) and $rank(H)$, with $H$ in (5.1) simultaneously under the constraints developed based on Condition 6.

# 5.4 Parameter estimation

## 5.4.1 Optimization problem

With the system parametrization and constraints explained in Section 5.3, the optimization problem to obtain system parameters can be written as

**Optimization Problem 6**

*Consider variables*

$g_{N\times 1}$ *in (5.3)*

$\lambda_{n_f - 1\times 1}$ *in (5.8)*

$c_{n_c\times 1}$ *in (5.14)*

$d_{n_d+1\times 1}$ *in (5.12)*

*and define*

$\hat{x}_1 = Ug$ *in (5.4)*

$\hat{x}_2 = Ug + X_1\lambda$ *in (5.7)*

$\hat{y} = Yc + \Gamma$ *in (5.13)*

*minimize*

$w_1 rank(H) + w_2||y - \hat{y}||_2,$ *with $H$ in (5.1)*

*subject to*

$(\hat{x}_2(i) - \hat{x}_2(j))(\hat{x}_2(i) - \hat{x}_2(j) - \hat{x}_1(i) + \hat{x}_1(j)) \leq 0$

$\forall i > j.$

In Optimization Problem 6 above, $w_1$ and $w_2$ are weighting factors. Optimization Problem 6 is a non-convex quadratic programming (QP) problem. Semidefinite programming (SDP) relaxation is a standard approach to solve non-convex QP problems. A SDP relaxation procedure converts a non-convex optimization problem to a convex optimization problem by defining a feasible convex set, which is easier to solve and whose solution is close to the solution of the original non-convex optimization problem.

### 5.4.2 Semidefinite programming relaxation

In order to convert the non-convex Optimization Problem 6 to a convex optimization problem, the over-parametrization technique is used in this chapter. Let us define a system parameter matrix $\Theta$ that includes system parameters $g, \lambda$ and $d$. Let us define the parameter $\theta$ as

$$\theta = [g^T \quad \lambda^T \quad d^T]^T \tag{5.17}$$

and then define the over-parametrized parameter matrix $\Theta$ as

$$\Theta = \theta \cdot \theta^T = \begin{bmatrix} g_0 g_0 & \cdots & \lambda_{n_f} g_0 & \cdots & d_{n_d} g_0 \\ g_0 g_1 & \cdots & \lambda_{n_f} g_1 & \cdots & d_{n_d} g_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_0 \lambda_{nf} & \cdots & \lambda_{n_f} \lambda_{n_f} & \cdots & d_{n_d} \lambda_{n_f} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_0 d_{n_d} & \cdots & \lambda_{n_f} d_{n_d} & \cdots & d_{n_d} d_{n_d} \end{bmatrix}. \tag{5.18}$$

An arbitrary gain may be distributed among the static nonlinearity and the two linear dynamic systems. In order to avoid an ambiguous gain, the scaling of the first dynamic system can be fixed by setting $\sum_{k=0} g_k^2 = 1$. The scaling of the static nonlinear function is fixed as explained in Condition 6. With the system parameter matrix $\Theta$, the constraint in Optimization Problem 6 can be rewritten as a linear matrix inequity (LMI) condition as

$$\Delta X_2^T \Delta X_2 - \Delta X_2^T \Delta X_1 \leq 0 \tag{5.19}$$

where

$$\Delta X_2^T \Delta X_2 = diag(diag(\delta X_1 \tilde{\Theta} \delta X_1^T)), \tag{5.20}$$

and

$$\Delta X_2^T \Delta X_1 = diag(diag(\delta X_1 \tilde{\tilde{\Theta}} \delta U^T)). \tag{5.21}$$

Here,

$$\delta X_1 = \begin{bmatrix} X_1(2,:) - X_1(1,:) \\ X_1(3,:) - X_1(2,:) \\ \vdots \\ X_1(N,:) - X_1(N-1,:) \\ \vdots \\ X_1(N,:) - X_1(1,:) \end{bmatrix}, \tag{5.22}$$

with $X_1$ in (5.9),

$$\delta U = \begin{bmatrix} U(2,:) - U(1,:) \\ U(3,:) - U(2,:) \\ \vdots \\ U(N,:) - U(N-1,:) \\ \vdots \\ U(N,:) - U(1,:) \end{bmatrix}, \tag{5.23}$$

with $U$ in (5.2), $\tilde{\Theta} = \Theta(N+1 : N+M, N+1 : N+M)$, and $\tilde{\tilde{\Theta}} = \Theta(N+1 : N+M, 1 : N)$. Here $diag(x)$ indicates a square matrix with the elements of a vector $x$ on the diagonal, and $diag(X)$ indicates the main diagonal of a matrix $X$. Using the system parameter matrix $\Theta$, the simulated output $\hat{y}$ in (5.13) is defined as

$$\hat{y} = Yc + \Gamma$$

where

$$\begin{aligned} T &= U\Theta(1 : N, N+n_f : N+n_f+n_d) \\ &+ X_1\Theta(N+1 : N+n_f-1, N+n_f : N+nf+n_d) \end{aligned}$$

Here, $\Theta$ in (5.18), satisfying the LMI in (5.19), is a rank 1 matrix for noiseless cases. However, in order to account for the noise effect, the condition can be relaxed to a rank inequality condition as $rank(\Theta) \leq \gamma$, where $\gamma$ is a positive constant. An optimization problem with rank inequality conditions is hard to solve. One simple and effective way, applicable when the matrix is symmetric positive semidefinite, is to use its trace in place of its rank. The motivation for the use of its trace is that if the matrix $\Theta$ is a symmetric and positive semidefinite, its singular values are the same as its eigenvalues. Therefore, the nuclear norm reduces to trace, and the nuclear norm is the convex envelope of the rank function on the set of matrices with norms less than 1. Thus, the trace inequality condition is the closest convex approximation to the original rank inequity condition that can be efficiently solved [26]. As a result, $trace(\Theta) \leq p$, where $p$ is a positive constant, is used instead of the rank inequality condition. The positive constant $p$ can be tuned by investigating estimation results. Due to this SDP relaxation used to formulate Optimization Problem 6 and the user-chosen value $p$, the constraint I in Condition 6 is relaxed to

$$(\hat{x}_2(i) - \hat{x}_2(j))(\hat{x}_2(i) - \hat{x}_2(j) - (1 + r)(\hat{x}_1(i) - \hat{x}_1(j))) \leq q$$
$$\forall i > j$$

where $r$ and $q$ are small magnitude positive constants determined by the user-chosen value $p$ and a noise level. Also, the rank minimization on $H$ in Optimization Problem 6 is eliminated since this condition is included into the $trace(\Theta) \leq p$. Finally, the non-convex Optimization Problem 6 is reformulated as a SDP convex optimization problem as

**Optimization Problem 7**

*Consider variables*

$\Theta_{N+M+n_d \times N+M+n_d}$ *in (5.18)*

$c_{n_c \times 1}$ *in (5.14)*

*and define*

$\hat{y} = Yc + \Gamma$ *in (5.13)*

*minimize*

$||y - \hat{y}||_2$

*subject to*

$trace(\Theta) \leq p$

$\Delta X_2^T \Delta X_2 - \Delta X_2^T \Delta X_1 \leq 0$

$\Theta \geq 0$

$$\sum_{k=1}^{N} \Theta(k,k) = 1$$

*with* $\Delta X_2^T \Delta X_2$ *in (5.20) and* $\Delta X_2^T \Delta X_1$ *in (5.21).*

In Optimization Problem 7, it is assumed that the user-specified structure variables $n_c, n_d$, and $n_f$ are known. Once the optimal $\Theta$ in (5.18) is obtained, the optimal $\theta$ (5.17) can be obtained by conducting a Singular Value Decomposition (SVD). The singular vector corresponding to the largest singular value is the optimal solution for $\theta$.

### 5.4.3 Iterative approach

Obviously, the proposed identification method requires prior information of $\hat{x}_1$ to obtain $X_1$ in (5.9). Let us define a new system parameter $\phi$ that includes the parameter $\theta$ in (5.17) and the parameter $c$ in (5.14) as $\phi = [\theta^T \; c^T]^T$. With the initialization $X_1^1 = zeros(N, n_f - 1)$ (this means $\hat{x}_1^1(t) = u(t)$) and the previous parameter estimation $\hat{\phi}^{k-1}$, we propose the following iterative method:

Step 1: Construct the necessary matrices for the
optimization problem formulation.
($\delta U$ in (5.23), $\delta X_1$ in (5.22), $Y$ in (5.15),
$\Gamma$ in (5.16), $\Delta X_2^T \Delta X_2$ in (5.20), $\Delta X_2^T \Delta X_1$ in (5.21)).

Step 2: Solve Optimization Problem 7 to obtain
$\Theta$ in (5.18) and $d$ in (5.14).

Step 3: Conduct a SVD on $\Theta$ in (5.18) to obtain $\theta$ in (5.17)
and define $\hat{x}_1$ in (5.4).

Step 4: Update $X_1^k$ in (5.9) using $\hat{x}_1$ estimated in Step 3.

Step 5: Stopping criterion of the algorithm.
If $||\hat{\phi}^k - \hat{\phi}^{k-1}||/||\hat{\phi}^{k-1}|| < \varepsilon$ , stop.
Otherwise, go to Step 1.

Step1 creates the matrices necessary for constructing Optimization Problem 7. Step 2 actually solves Optimization Problem 7 to obtain $\Theta$ in (5.18) and $c$ in (5.14). Step 3 conducts a SVD to obtain $\theta$ in (5.17). Step 4 updates the prior information to construct Optimization Problem 7. Step 4 formulates a stopping criterion for the algorithm by looking at the relative parameter error.

As long as the classes of models used for the estimation contain the true models for static nonlinearity and for linear dynamic systems, and the assumptions on static nonlinearity are indeed true, $\hat{x}_1(t)$ and $\hat{x}_2(t)$ will converge to $x_1(t)$ and $x_2(t)$ provided $N, L$ and $n_f$ are large enough at each iteration step based on (5.5) and Weierstrass's Theorem.

## 5.5 Benchmark problem

The system to be modeled is an electronic nonlinear system with a Wiener-Hammerstein structure that was built by [63]. The first linear dynamic system $G_1$ is designed as a third order Chebyshev filter (pass-band ripple of 0.5 dB and cut off frequency of 4.4 kHz). The second linear dynamic system $G_2$ is designed as a third order inverse Chebyshev filter (stop-band attenuation of 40 dB starting at

5 kHz). This system has a transmission zero in the frequency band of interest. This can complicate the identification significantly, because the inversion of such a characteristic is difficult.

The proposed iterative identification method is applied to the benchmark problem. In this benchmark, the estimation data are the first part of the measured input $u(t)$ and output $y(t)$ $(t = 1, 2, \cdots, 100000)$, and the test data are given by the remaining part of the measured input $u(t)$ and output $y(t)$ $(t = 100001, \cdots, 188000)$. The goal of the benchmark is to identify a nonlinear model using the estimation data. Next, this model is used to simulate the output $y_{sim}(t)$ of the system on the test set. $n_c = 3, n_d = 3$, and $n_f = 5$ are used in this study. In order to solve the SDP problem (Optimization Problem 7), SEDUMI [59] and YALMIP [50] are used. The estimation results are shown in Figure 5.2, Figure 5.3, Figure 5.4, Figure 5.5, and Table 5.1. Table 5.1 shows the mean value ($\mu$), the standard deviation ($s$), and the root mean square (RMS) value ($e_{RMS}$) of the simulation error (time domain) for the estimation data and the test data obtained by using the proposed method, and the comparison with the results from [54] and [2]. Each value is calculated based on the following equation:

Test data

1. The mean value of the simulation error:

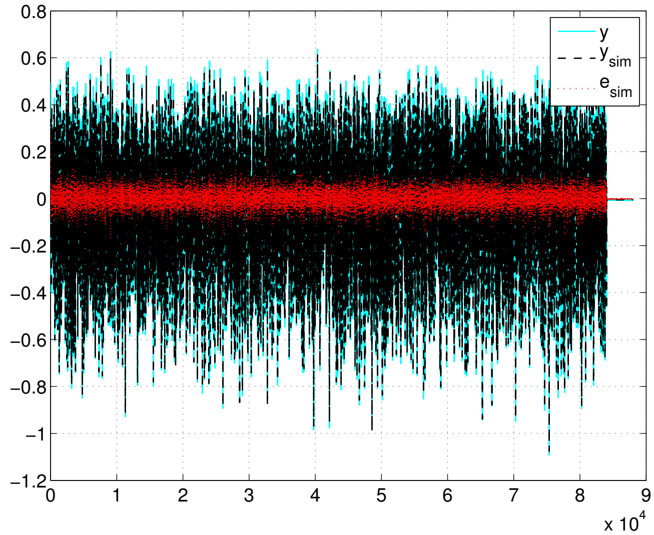
$$\mu = \frac{1}{87000} \sum_{t=101001}^{188000} e_{sim}(t)$$

2. The standard deviation of the simulation error:

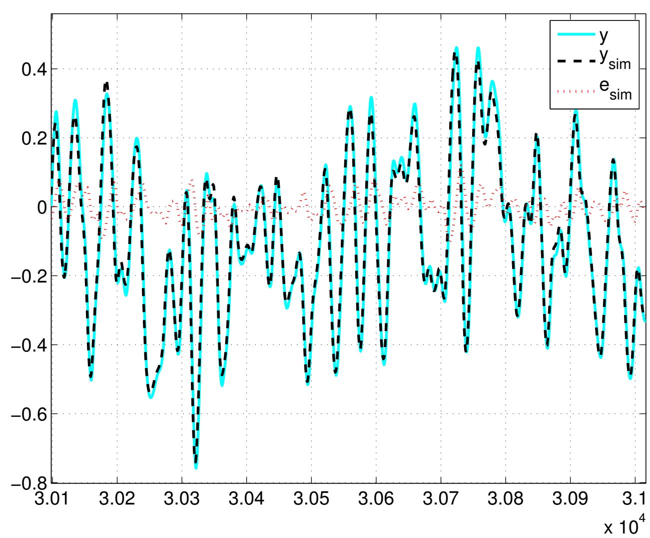
$$s = \sqrt{\frac{1}{87000} \sum_{t=101001}^{188000} (e_{sim}(t) - \mu)^2}$$

3. The root mean square (RMS) value of the error:

$$e_{RMS} = \sqrt{\frac{1}{87000} \sum_{t=101001}^{188000} e_{sim}^2(t)}$$

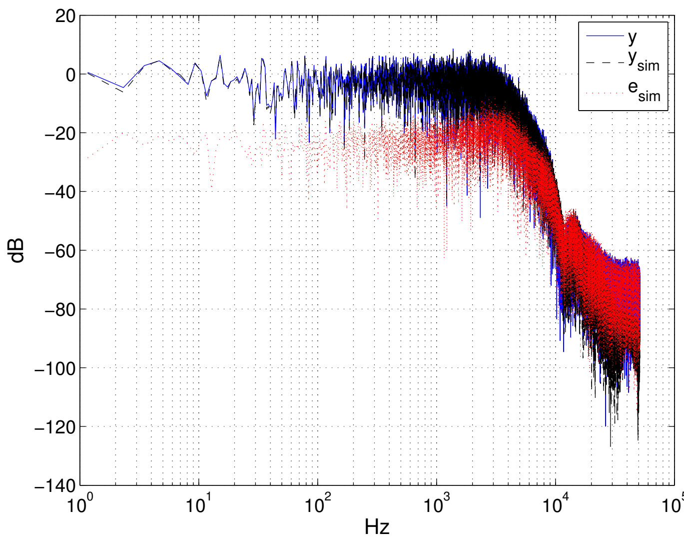



(a)

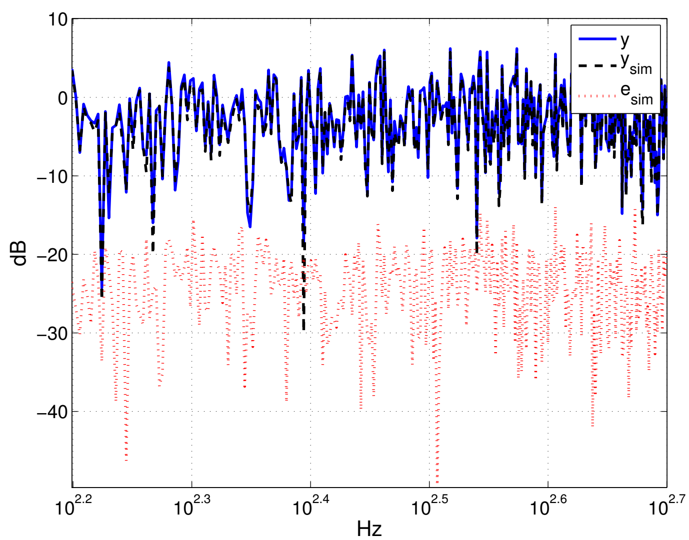



(b)

**Figure 5.2**: Modeled output $y_{sim}$, test data $y$, and the simulation error $e_{sim}$ in the time domain (top figure). The magnified figure of the top figure (bottom figure).




(a)




(b)

**Figure 5.3**: Modeled output $y_{sim}$, test data $y$, and the simulation error $e_{sim}$ in the frequency domain (top figure). The magnified figure of the top figure (bottom figure).

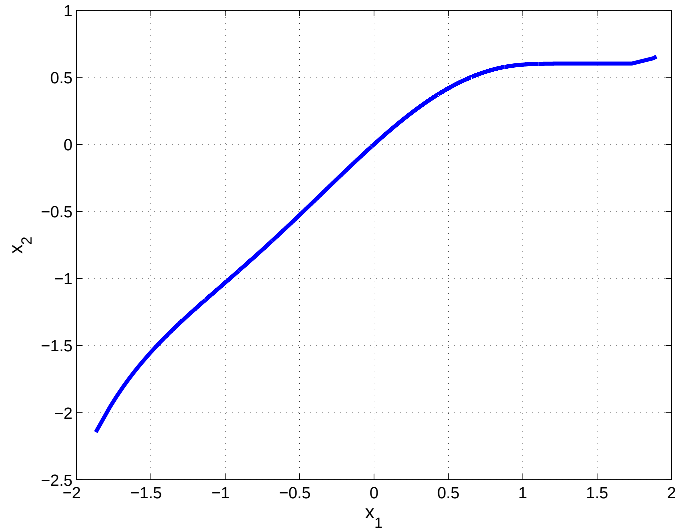



**Figure 5.4**: Identified static nonlinear function, $f(\cdot)$.

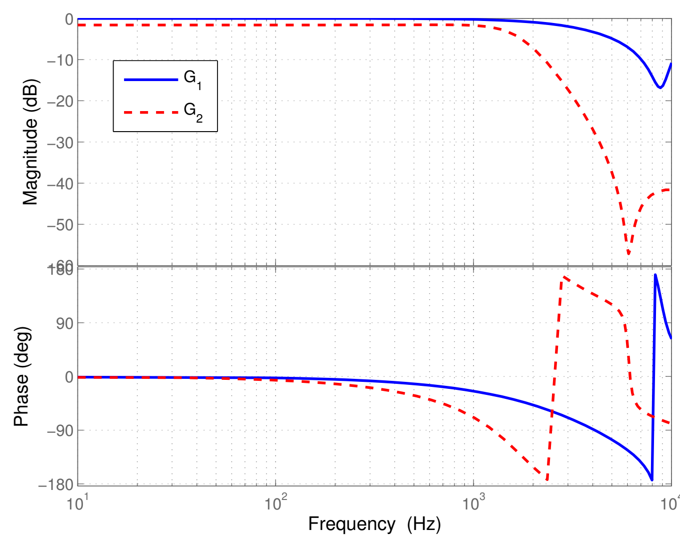



**Figure 5.5**: Identified dynamical systems, $G_1$ and $G_2$.

Estimation data

1. The mean value of the simulation error:
$$\mu = \frac{1}{99000}\sum_{t=1001}^{100000} e_{sim}(t)$$
2. The standard deviation of the simulation error:
$$s = \sqrt{\frac{1}{99000}\sum_{t=1001}^{100000} (e_{sim}(t) - \mu)^2}$$
3. The root mean square (RMS) value of the error:
$$e_{RMS} = \sqrt{\frac{1}{99000}\sum_{t=1001}^{100000} e_{sim}^2(t)}$$

**Table 5.1**: Characteristics of the simulation error.

| Method | Parameters | Estimation data | Test data |
|---|---|---|---|
| The proposed method | $\mu$ | 0.0011 V | 0.0015V |
| | $s$ | 0.0345 V | 0.0345 V |
| | $e_{RMS}$ | 0.0345 V | 0.0345 V |
| PNLSS Paduart (2009) | $\mu$ | 0.031 mV | 0.048 mV |
| | $s$ | 0.359 mV | 0.415 mV |
| | $e_{RMS}$ | 0.360 mV | 0.418 mV |
| [2] | $\mu$ | -0.0051 V | -0.0038 V |
| | $s$ | 0.0332 V | 0.0333 V |
| | $e_{RMS}$ | 0.0336 V | 0.0335 V |

## 5.6 Conclusion

In this chapter, an iterative convex optimization algorithm is proposed to identify Wiener-Hammerstein systems. A non-convex rank minimization problem is formulated first, and then the non-convex rank minimization problem is reformulated as a convex optimization problem using a SDP relaxation technique. In the proposed identification method, the first linear dynamic system, the static nonlinear function, and the second linear dynamic system are parameterized as an FIR model, a polynomial function, and a rational transfer function respectively. For the modeling of static nonlinearity, the monotonically non-deceasing condition was applied to limit the number of possible selections for intermediate signals. As two unmeasurable intermediate signals are included in the system description, the over-parameterization technique is used and the parameter estimation problem is solved iteratively. At each step of iteration, the over-parametrized parameters are estimated and then separated by using the singular value decomposition (SVD). The proposed method is applied to the benchmark problem and the estimation result shows the effectiveness of the proposed algorithm. The materials in Chapter 5 have been accepted for publication in Y. Han and R. de Callafon, Identification of Wiener-Hammerstein Benchmark Model using Convex Optimization, Control Engineering Practice Special Issue, to appear 2012. The dissertation author was the primary investigator and author of this paper.

# 6

# Closed-loop Identification of Hammerstein Systems

## 6.1 Introduction

Closed-loop system identification techniques are useful for control relevant identification. From the control design point of view, the use of data gathered from closed-loop experiments provides advantages for designing control systems to satisfy typical control performance requirements, such as stability. A feedback connection of a Hammerstein system and a controller provides a good approach to modeling systems with static actuator nonlinearity or input saturation during closed-loop experiments. There has been much research on the problem of identifying Hammerstein systems in a open-loop setting [4] [14] [29] [32] [34] [36] [37] [53] , while much less attention has been paid to the problem of identifying Hammerstein systems in a closed-loop setting. One of the early works dealing with closed-loop Hammerstein system identification can be found in [6]. In this work, Beyer et al. proposed a closed-loop identification method for Hammerstein systems using the LS method, the GLS method and the maximum likelihood method. In addition, Linard et al. [46] extended closed-loop identification methods (a two-stage method and using right coprime factorizations) for linear dynamic systems to nonlinear dynamic systems and De Bruyne et al. [19] presented gradient expressions for a closed-loop parametric identification scheme. However, these methods are based on linearization of a nonlinear map between time domain signals. Recently, van Wingerden and Verhaegen [62] presented an algorithm to identify MIMO Hammerstein systems under open and closed-loop conditions. They formulated an optimized predictor based subspace identification algorithm in the dual space. Laurain et al. [45] presented an IV method dedicated to closed-loop Hammerstein systems. Comprehensive studies of block-oriented nonlinear system identification can be found in [29].

In the identification of Hammerstein systems, equation error type models are commonly used because the parameter estimation can be reduced to an ordinary least squares (LS) problem that can be solved with a convex optimization. In this chapter, we focus on the Output Error (OE) identification of Hammerstein systems in a closed-loop setting. Closed-loop identification is often used for control-relevant identification where the goal is to estimate models suitable for robust control design. It is then often only necessary to model the plant dynamics, not noise properties. So it would be natural to use an output error model structure [28].

It is well known that the direct use of input/output data, generated from a closed-loop setting, results in biased estimation if no noise model is estimated due to the correlation between input and noise. The main contribution of the study in this chapter is that we propose a method that allows us to solve a nonlinear OE minimization problem as an iterative linear optimization problem that is robust to the correlation between input and noise. The basic idea is to express the iterative nonlinear parameter estimation as an iterative nuclear norm minimization problem based on gradient expression similar to the method in [9]. Convergence of the iterative steps guarantees a local minimum of the OE minimization problem.

## 6.2 Problem description

Figure 6.1 shows the Hammerstein system in a closed-loop setting considered in this study. For identification purposes, the reference input $r(t)$ and the

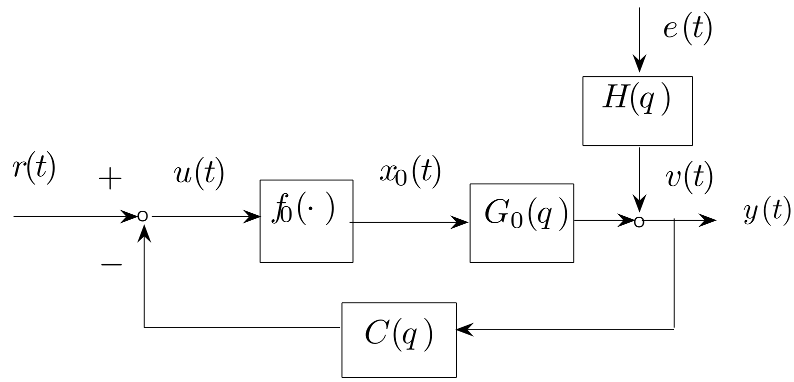



**Figure 6.1**: Closed-loop Hammerstein system.

controller $C(q)$ are known, the input $u(t)$ and output $y(t)$ are measured, whereas the intermediate signal $x_0(t)$, the static nonlinearity $f_0(\cdot)$ and the linear dynamic system $G_0(q)$ are unknown. The disturbance $v(t)$ is a filtered white noise, where the filtering properties are unknown. The purpose of this study is to propose an OE identification method for the consistent estimation of the static nonlinearity $f_0(\cdot)$ and the linear dynamic system $G_0(q)$ in a closed-loop setting on the basis of the measured signals, the input $u(t)$ and output $y(t)$, and the knowledge of the controller $C(q)$ . In this study, the following conditions are assumed for identifiability:

**Condition 7**

*I. The closed-loop system is stable.*

*II. $r(t)$ and $v(t)$ are not correlated.*

*III. All signals are stationary.*

*VI. The input to the linear dynamic system is persistently exciting.*

*V. The controller is known.*

## 6.3 System parametrization

From the Hammerstein system in a closed-loop setting shown in Figure 6.1, the output $y(t)$ generated from a closed-loop Hammerstein system is defined as

$$\begin{aligned} y(t) &= G_0(q)x_0(t) + v(t) \\ &= q^{-td}\frac{B_0(q)}{A_0(q)}x_0(t) + v(t) \end{aligned}$$

where $v(t)$ is colored noise, $x_0(t) = f_0(u(t))$, and $td$ indicates the number of steps of time delay of the system. It is assumed that there is at least one step time delay in $G_0(q)$. The noise free output $\hat{y}(t)$ is defined by

$$\hat{y}(t) = \frac{B_0(q)}{A_0(q)}x(t - td)$$

where

$$x(t) = f\left(\hat{u}(t)\right), \ \hat{u}(t) = r(t) - \hat{y}_c(t), \ \hat{y}_c(t) = C(q)\hat{y}(t). \tag{6.1}$$

In order to define $\hat{y}(t)$, one only needs $x(t - td), \cdots, x(t - n_b - td)$ and $\hat{y}(t - 1), \cdots, \hat{y}(t - n_a)$. Subsequently, in order to define $\hat{y}_c(t) = C(q)\hat{y}(t)$, where

$$C(q) = \frac{d_0 + \cdots + d_{n_d}q^{-n_d}}{1 + c_1q^{-1} + \cdots + c_{n_c}q^{-n_c}}, \tag{6.2}$$

one only needs $\hat{y}(t), \cdots, \hat{y}(t - n_d)$ and $\hat{y}_c(t - 1), \cdots, \hat{y}_c(t - n_c)$. As a result, noise free input and output signals are generated by the known reference signal $r(t)$ only.

In this section, the input static nonlinearity $f_0(\cdot)$ is modeled as a piecewise linear function using piecewise triangle functions as shown in Section 2.2. In feedback control systems, non-smooth static nonlinearity, such as saturation, is common. A piecewise linear approximation is an excellent way to estimate such nonlinearity for feedback control systems since it can achieve good approximation with only a small number of parameters.

Let $\hat{x}(t, \lambda) = f(\hat{u}(t), \lambda)$ be the approximation of $x(t)$, where $\hat{u}(t)$ is the noise free input and $\lambda$ is the amplitude parameter

$$\lambda = \begin{bmatrix} \lambda_1 & \cdots & \lambda_M \end{bmatrix}^T \tag{6.3}$$

on $m = [m_1(= min(u(t))) \ \cdots \ m_M(= max(u(t)))]^T$. In each segment of the $m$-axis, the resulting linear function is defined by two overlapping triangle functions in the segment. Thus, $\hat{x}(t, \lambda)$ can be written as

$$\hat{x}(t, \lambda) = \rho(\hat{u}(t))\lambda$$

where $\rho(\hat{u}(t))$ is defined as

$$\rho(\hat{u}(t)) = \left[\cdots \ 0 \ \frac{m_{k+1} - \hat{u}(t)}{m_{k+1} - m_k} \ \frac{\hat{u}(t) - m_k}{m_{k+1} - m_k} \ 0 \ \cdots\right]$$
$$\text{for } \ m_k \leq \hat{u}(t) < m_{k+1}$$

where $m_k$ and $m_{k+1}$ are the center locations of the triangle basis functions. Let $G(q, \phi)$ be the estimation of $G_0(q)$ with the system parameter

$$\phi = [a_1 \ \cdots \ a_{n_a} \ b_0 \ \cdots \ b_{n_b}] \tag{6.4}$$

such that

$$G(q, \phi) = q^{-td}\frac{B(q, \phi)}{A(q, \phi)}$$

where

$$A(q, \phi) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a},$$
$$B(q, \phi) = b_0 + b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}.$$

With the parameters $\phi$ in (6.4) and $\lambda$ in (6.3), the noise free OE model output $\hat{y}(t)$ now can be written as

$$\hat{y}(t, \phi, \lambda) = \frac{B(q, \phi)\hat{x}(t - td, \lambda)}{A(q, \phi)} = \frac{B(q, \phi)\rho(\hat{u}(t - td))\lambda}{A(q, \phi)}. \tag{6.5}$$

Realizing that $\rho(\hat{u}(t - td))\lambda$ in (6.5) is a linear combination of the time shifted noise free input signal weighted by $\lambda_k, k = 1, \cdots, M$, it can be verified that $B(q, \phi)\rho(\hat{u}(t - td))\lambda$ in (6.5) can be written in a linear combination of time shifted inputs weighted by the parameter

$$\tilde{\theta} = [b_0\lambda_1 \ \cdots \ b_0\lambda_M \ \cdots \ b_{n_b}\lambda_1 \ \cdots \ b_{n_b}\lambda_M]^T .$$

In this parametrization, an arbitrary gain may be distributed between the static nonlinearity and the linear dynamic system [16] [67]. In order to avoid an ambiguous gain, the scaling of either the linear dynamic system or the static nonlinearity can be fixed. In this study, we choose to normalize the scaling of the linear dynamic system by fixing $b_0 = 1$. As a result, we will define the parameter

$$\begin{aligned} &\theta = [a_1 \ \cdots \ a_{n_a} \ \lambda_1 \ \cdots \ \lambda_M \ \cdots \ b_{n_b}\lambda_1 \ \cdots \ b_{n_b}\lambda_M]^T \in R_{s\times 1}, \\ &s = n_a + M \cdot (n_b + 1) \end{aligned} \tag{6.6}$$

as the parameter to be identified, leading to the shorthand notation

$$\hat{y}(t,\theta) = \frac{T(q, \hat{u}(t-td), \theta)}{A(q,\theta)} \tag{6.7}$$

where $\theta_1, \cdots, \theta_{n_a}$ are used to capture $a_1, \cdots, a_{n_a}$ and $\theta_{n_a+1}, \cdots, \theta_{n_a+M\cdot n_b}$ are used to capture $b_0, \cdots, b_{n_b-1}$, and $\lambda_1, \cdots, \lambda_M$. With the chosen system parameter $\theta$ in (6.6), the output error is defined as

$$\varepsilon(t,\theta) = y(t) - \hat{y}(t,\theta). \tag{6.8}$$

## 6.4 Parameter estimation

### 6.4.1 OE minimization

Due to the nonlinear parameter dependency of $\varepsilon(t,\theta)$ in (6.8), an output error (OE) model requires a nonlinear optimization (iterative search) to find at least a local minimum. Let

$$E(\theta) = [\varepsilon(1,\theta) \ \cdots \ \varepsilon(N,\theta)]^T$$

where $\varepsilon(t,\theta)$ is given in (6.8). Then, the parameter estimation is given by

$$\begin{aligned} &\hat{\theta}_{OE}^N = \arg\min_\theta V^N(\theta) \\ &\text{where} \\ &V^N(\theta) = \frac{1}{2N} E^T(\theta) E(\theta). \end{aligned} \tag{6.9}$$

The minimum of $V^N(\theta)$ in (6.9) can be obtained by solving

$$\frac{dV^N(\theta)}{d\theta} = \frac{1}{N} E^T(\theta) \frac{dE(\theta)}{d\theta} = \vec{0} \tag{6.10}$$

where $\vec{0}$ represents a zero vector.

**Theorem 1** *Let* $y^L(t) = L(q,\theta)y(t)$ *and* $\rho^L(\hat{u}(t)) = L(q,\theta)\rho(\hat{u}(t))$, *where* $\hat{u}(t)$ *denotes the noise free input. Define the filtered regressor as*

$$\Phi^L = \left[ \begin{array}{cc} \Phi_a^L & \Phi_{b\lambda}^L \end{array} \right] \tag{6.11}$$

*where*

$$\Phi_a^L = \left[ \begin{array}{ccc} -y^L(0) & \cdots & -y^L(1-a_{n_a}) \\ -y^L(1) & \cdots & -y^L(2-a_{n_a}) \\ \vdots & \ddots & \vdots \\ -y^L(N-1) & \cdots & -y^L(N-a_{n_a}) \end{array} \right]$$

*and*

$$\Phi_{b\lambda}^L = \left[ \begin{array}{ccc} \rho^L(\hat{u}(1-td)) & \cdots & \rho^L(\hat{u}(1-td-n_b)) \\ \vdots & \vdots & \vdots \\ \rho^L(\hat{u}(N-td)) & \cdots & \rho^L(\hat{u}(N-td-n_b)) \end{array} \right]. \tag{6.12}$$

*Then, solving*

$$\frac{dV^N(\theta)}{d\theta} = \vec{0}$$

*is equivalent to solving*

$$\psi^T(\theta) \cdot [Y^L - \Phi^L\theta] = \vec{0}$$

*where*

$$\psi^T(\theta) = -\frac{dE(\theta)}{d\theta}^T \tag{6.13}$$

*and*

$$Y^L = [y^L(1) \ \cdots \ y^L(N)]^T. \tag{6.14}$$

**Proof** From (6.10) and (6.13),

$$\frac{dV^N(\theta)}{d\theta} = -\frac{1}{N}E^T(\theta)\psi(\theta).$$

Since $E(\theta) = Y - \hat{Y}(\theta)$, where $Y = [y(1) \ \cdots \ y(N)]^T$ and $\hat{Y} = [\hat{y}(1,\theta) \ \cdots \ \hat{y}(N,\theta)]^T$,

$$E^T(\theta)\psi(\theta) = \psi^T(\theta) \cdot [Y - \hat{Y}(\theta)].$$

Let

$$Y_A^L = \left[ \begin{array}{ccc} Y_A^L(1) & \cdots & Y_A^L(N) \end{array} \right]^T$$

where $Y_A^L(t) = A(q,\theta)y^L(t)$,

$y^L(t) = L(q,\theta)y(t)$ and $L(q,\theta) = \dfrac{1}{A(q,\theta)}$,

and

$$\hat{Y}_T^L(\theta) = \left[ \begin{array}{ccc} \hat{Y}_T^L(1,\theta) & \cdots & \hat{Y}_T^L(N,\theta) \end{array} \right]^T$$

where $\hat{Y}_T^L(t,\theta) = T^L(q, \hat{u}(t-td), \theta)$

and $T^L(q, \hat{u}(t-td), \theta) = L(q,\theta)T(q, \hat{u}(t-td), \theta)$.

Then

$$[Y - \hat{Y}(\theta)] = [Y_A^L - \hat{Y}_T^L(\theta)].$$

With $Y^L$ in (6.14) and $\Phi^L$ in (6.11), which allow an additional filtering of the output $y(t)$ and the regressor $\Phi$ with a filter $L(q,\theta)$, solving

$$\frac{dV^N(\theta)}{d\theta} = \vec{0}$$

is equivalent to solving

$$\psi^T(\theta) \cdot [Y^L - \Phi^L\theta] = \vec{0}. \tag{6.15}$$

□

Theorem 1 implies that one can compute a local minimum by explicitly solving $\psi^T(\theta) \cdot [Y^L - \Phi^L\theta] = \vec{0}$ for $\theta$ which is also known as Instrumental Variables (IV) estimate [9] [56].

### 6.4.2 Calculation of the instrument

In this section, the instrument $\psi^T(\theta)$ in (6.13) is calculated as

$$\psi^T(\theta) = -\frac{dE(\theta)}{d\theta}^T = \frac{d\hat{Y}(\theta)}{d\theta}^T$$

where

$$\frac{d\hat{Y}(\theta)}{d\theta} = \left[ \begin{array}{ccc} \frac{d\hat{y}(1,\theta)}{d\theta_1} & \cdots & \frac{d\hat{y}(N,\theta)}{d\theta_1} \\ \vdots & \cdots & \vdots \\ \frac{d\hat{y}(1,\theta)}{d\theta_s} & \cdots & \frac{d\hat{y}(N,\theta)}{d\theta_s} \end{array} \right]^T.$$

The following lemma summarizes the calculation of the instrument $\psi^T(\theta)$.

**Lemma 4** *The instrument $\psi^T(\theta)$ in (6.13) is defined by*

$$\psi^T(\theta) = [\psi_a \ \psi_{b\lambda}]$$

*where*

$$\psi_a = \begin{bmatrix} -\hat{y}^L(0) & \cdots & -\hat{y}^L(1-n_a) \\ -\hat{y}^L(1) & \cdots & -\hat{y}^L(2-n_a) \\ \vdots & \ddots & \vdots \\ -\hat{y}^L(N-1) & \cdots & -\hat{y}^L(N-n_a) \end{bmatrix} \tag{6.16}$$

$$+ \begin{bmatrix} d\hat{y}_1(0) & \cdots & d\hat{y}_{n_a}(1-td) \\ d\hat{y}_1(1) & \cdots & d\hat{y}_{n_a}(2-td) \\ \vdots & \ddots & \vdots \\ d\hat{y}_1(N-td) & \cdots & d\hat{y}_{n_a}(N-td) \end{bmatrix} \tag{6.17}$$

*and*

$$\psi_{b\lambda} = \Phi_{b\lambda}^L + \begin{bmatrix} d\Phi_1(1) & \cdots & d\Phi_{M\times(n_b+1)}(1) \\ \vdots & \vdots & \vdots \\ d\Phi_1(N) & \cdots & d\Phi_{M\times(n_b+1)}(N) \end{bmatrix}. \tag{6.18}$$

$\hat{y}^L(t)$ *in (6.16) is defined as* $\hat{y}^L(t) = L(q,\theta)\hat{y}(t,\theta)$. $d\hat{y}_i(t)$ *in (6.17) is defined as*

$$d\hat{y}_i(t) = \frac{1}{A(q,\theta)} \frac{dT(q,\hat{u}(t),\theta)}{d\theta_i} \tag{6.19}$$

*where* $\frac{dT(q,\hat{u}(t),\theta)}{da_i}$ *in (6.19) is defined as*

$$\frac{dT(q,\hat{u}(t),\theta)}{da_i} = \sum_{k=1}^{(n_b+1)\cdot M} \frac{d\Phi_{b\lambda}(t,k)}{d\hat{u}(t-p)} \frac{d\hat{u}(t-p)}{da_i} \tilde{\theta}_k. \tag{6.20}$$

$\Phi_{b\lambda}^L$ *in (6.18) is given in (6.12) and* $d\Phi_i(t)$ *in (6.18) is defined as*

$$d\Phi_i(t) = \frac{1}{A(q,\theta)} \sum_{k=1}^{(n_b+1)\cdot M} \frac{d\Phi_{b\lambda}(t,k)}{d\hat{u}(t-td-p)} \frac{d\hat{u}(t-td-p)}{d\tilde{\theta}_i} \tilde{\theta}_k. \tag{6.21}$$

$\frac{d\Phi_{b\lambda}(t,:)}{d\hat{u}(t-p)}$ *in (6.20) and (6.21) is defined as*

$$\frac{d\Phi_{b\lambda}(t,:)}{d\hat{u}(t-p)} = \left[ \frac{d\rho(\hat{u}(t))}{d\hat{u}(t)} \ \cdots \ \frac{d\rho(\hat{u}(t-n_b))}{d\hat{u}(t-n_b))} \right]$$

*where*

$$\frac{d\rho(\hat{u}(t))}{d\hat{u}(t)} = \left[ \cdots \ 0 \ \frac{-1}{m_{k+1}-m_k} \ \frac{1}{m_{k+1}-m_k} \ 0 \ \cdots \right]$$

*when* $m_k \leq \hat{u}(t) < m_{k+1}$, $p = ceil\left(\frac{k}{M}\right)$ *that rounds the elements of* $\frac{k}{M}$ *to the nearest integers greater than or equal to* $\frac{k}{M}$, *and* $\tilde{\theta} = \theta(n_a + 1 : s)$.

**Proof** (i) Calculation of $\left.\frac{d\hat{y}(t,\theta)}{d\theta_i}\right|_{i=1:n_a}$. From (6.7), $\hat{y}(t,\theta) = \frac{T(q,\hat{u}(t-td),\theta)}{A(q,\theta)}$. For brevity of notation, hereafter $\theta$ will be omitted if it does not lead to confusion in notation. If we take the derivative of $\hat{y}(t)$ with respect to $\theta_{i=1:n_a}$, we obtain

$$\left.\frac{d\hat{y}(t)}{d\theta_i}\right|_{i=1:n_a} = -q^{-i}\frac{T(q,\hat{u}(t-td))}{A(q)^2} + \frac{1}{A(q)}\frac{dT(q,\hat{u}(t-td))}{d\theta_i}. \tag{6.22}$$

For brevity of notation, let $a_i = \theta_{i=1:n_a}$. Then, with (6.19), (6.22) is written as

$$\frac{d\hat{y}(t)}{da_i} = -\hat{y}^L(t-i) + d\hat{y}_i(t-td).$$

With the noise free input $\hat{u}(t)$ in (6.1) and the controller $C(q)$ in (6.2), the derivative of $\hat{u}(t)$ with respect to $a_i$ is defined as

$$\frac{d\hat{u}(t)}{da_i} = -\sum_{j=1}^{n_c} c_j \frac{d\hat{u}(t-j)}{da_i} - \sum_{k=0}^{n_d} d_k \frac{d\hat{y}(t-k)}{da_i}. \tag{6.23}$$

(ii) Calculation of $\frac{d\hat{y}(t)}{d\tilde{\theta}_i}$. From (6.7) and (6.12),

$$\hat{y}(t,\theta) = \frac{T(q,\hat{u}(t-td),\theta)}{A(q)} = \Phi_{b\lambda}^L(t,:)\tilde{\theta}$$

where $\tilde{\theta} = \theta(n_a + 1 : s)$. If we take the derivative of $\hat{y}(t)$ with respect to $\tilde{\theta}_i$, we obtain

$$\begin{aligned} \frac{d\hat{y}(t)}{d\tilde{\theta}_i} &= \Phi_{b\lambda}^L(t,i) + \\ &\frac{1}{A(q)} \sum_{k=1}^{(n_b+1)\cdot M} \frac{d\Phi_{b\lambda}(t,k)}{d\hat{u}(t-td-p)} \frac{d\hat{u}(t-td-p)}{d\tilde{\theta}_i} \tilde{\theta}_k \end{aligned} \tag{6.24}$$

where $\Phi_{b\lambda}^L(t,i)$ is an element of $\Phi_{b\lambda}^L$ defined in (6.12). Then, with (6.21), (6.24) is written as

$$\frac{d\hat{y}(t)}{d\tilde{\theta}_i} = \Phi_{b\lambda}^L(t,i) + d\Phi_i(t).$$

Similar to (6.23), the derivative of $\hat{u}(t)$ with respect to $\tilde{\theta}_i$ in (6.24) is defined as

$$\frac{d\hat{u}(t)}{d\tilde{\theta}_i} = -\sum_{j=1}^{n_c} c_j \frac{d\hat{u}(t-j)}{d\tilde{\theta}_i} - \sum_{k=0}^{n_d} d_k \frac{d\hat{y}(t-k)}{d\tilde{\theta}_i}.$$

□

### 6.4.3 Convex optimization and parameter separation

Due to the independent parametrization of the static nonlinear block and the linear dynamic block, the Hammerstein system is over-parametrized by the modified parameter vector given in (6.6). As the parameter estimate $\hat{\theta}_{OE}^{N}$ has the same structure, we need to separate the parameters of the linear dynamic system in

$$\eta = \begin{bmatrix} a_1 \cdots a_{n_a} & b_1 \cdots b_{n_b} \end{bmatrix}^T$$

and the parameters for the piecewise linear approximation of $f(\cdot)$ in

$$\lambda = \begin{bmatrix} \lambda_1 & \cdots & \lambda_M \end{bmatrix}^T.$$

First the parameter vector $\theta$ is reorganized into $\Gamma_{b\lambda}$ given by

$$\begin{aligned} \Gamma_{b\lambda} &= \begin{bmatrix} \theta_{n_a+1} & \cdots & \theta_{n_a+M} \\ \vdots & \vdots & \vdots \\ \theta_{n_a+(n_b-1)\times M} & \cdots & \theta_{n_a+n_b\times M} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \cdots & \lambda_M \\ \vdots & \vdots & \vdots \\ b_{n_b}\lambda_1 & \cdots & b_{n_b}\lambda_M \end{bmatrix}. \end{aligned}$$

The rank of $\Gamma_{b\lambda}$ is equal to 1 for noiseless cases. In order to account for the effect of noise, the parameter estimation problem will be formulated by minimizing $rank(\Gamma_{b\lambda})$ in the set of $\psi^T(\theta) \cdot [Y^L - \Phi^L\theta] = \vec{0}$ in (6.15). The optimization problem can be written as:

**Optimization Problem 8**

$$
\begin{aligned}
&\textit{Consider}\\
&\quad \textit{variable } \theta\\
&\quad \Gamma_{b\lambda} = \begin{bmatrix} \theta_{n_a+1} & \cdots & \theta_{n_a+M} \\ \vdots & \vdots & \vdots \\ \theta_{n_a+(n_b-1)\times M} & \cdots & \theta_{n_a+n_b\times M} \end{bmatrix}\\
&\textit{Minimize}\\
&\quad rank(\Gamma_{b\lambda})\\
&\textit{subject to}\\
&\quad \psi^T(\theta)\cdot[Y^L-\Phi^L\theta]=\vec{0} \textit{ in (6.15)}
\end{aligned}
$$

Optimization Problem 8 results in the optimal solution for the system parameter $\theta$ that satisfies the constraint $\psi^T(\theta)\cdot[Y^L-\Phi^L\theta]=\vec{0}$ in (6.15) and provides the minimum rank $\Gamma_{b\lambda}$. This means the smaller singular values of $\Gamma_{b\lambda}$ will be minimized compared to the largest singular value and simplify the *rank* 1 approximation of $\Gamma_{b\lambda}$. Unfortunately, the rank minimization in Optimization Problem 8 is not convex. Minimizing the nuclear norm instead of the rank of the matrix is a convex relaxation of the rank minimization problem. The motivation for this nuclear norm relaxation is that the nuclear norm is the convex envelope of the rank function on the set of matrices with norms less than or equal to 1. Thus, by solving the nuclear norm minimization problem, we obtain a lower bound on the optimal value of the original rank minimization problem [26] [27]. Using the nuclear norm relaxation for rank minimization, Optimization Problem 8 will be reformulated as a convex optimization problem.

**Optimization Problem 9**

$$
\begin{aligned}
&\textit{Consider}\\
&\quad \textit{variable } \theta\\
&\quad \Gamma_{b\lambda} = \begin{bmatrix} \theta_{n_a+1} & \cdots & \theta_{n_a+M} \\ \vdots & \vdots & \vdots \\ \theta_{n_a+(n_b-1)\times M} & \cdots & \theta_{n_a+n_b\times M} \end{bmatrix}\\
&\textit{Minimize}\\
&\quad ||\Gamma_{b\lambda}||_*\\
&\textit{subject to}\\
&\quad \psi^T(\theta)\cdot[Y^L - \Phi^L\theta] = \vec{0} \textit{ in (6.15)}\\
&\textit{where}\\
&\quad ||\Gamma_{b\lambda}||_* = \sum_{i=1}^{r} \sigma_i(\Gamma_{b\lambda})\\
&\textit{is the nuclear norm of } \Gamma_{b\lambda}.
\end{aligned}
$$

Once $\hat{\theta}_{OE}^N$ is estimated by solving Optimization Problem 9, $\hat{a} = [\hat{a}_1 \ \cdots \ \hat{a}_{n_a}]^T$ is easily obtained. The parameter vectors $\hat{b} = \left[\hat{b}_1 \ \cdots \ \hat{b}_{n_b}\right]^T$ and $\hat{\lambda}$ can be separated using the singular value decomposition (SVD) [4]. The singular value decomposition of $\hat{\Gamma}_{b\lambda}$ is given as

$$
\begin{aligned}
\hat{\Gamma}_{b\lambda} &= \begin{bmatrix} \hat{\lambda}_1 & \cdots & \hat{\lambda}_M \\ \vdots & \vdots & \vdots \\ \hat{b}_{n_b}\hat{\lambda}_1 & \cdots & \hat{b}_{n_b}\hat{\lambda}_M \end{bmatrix}\\
&= U\Sigma V^T
\end{aligned}
$$

where $U_{(n_b+1)\times(n_b+1)}$ and $V_{M\times M}$ are orthogonal matrices, and $\Sigma_{(n_b+1)\times M}$ is a rectangular diagonal matrix. The positive diagonal entries of $\Sigma$ are called singular values. With the constraint, $b_0 = 1$, the parameter vectors $\hat{\eta}$ and $\hat{\lambda}$ can be calculated by

$$
\begin{aligned}
\hat{\eta} &= \begin{bmatrix} \hat{a}^T & \hat{b}^T \end{bmatrix}\\
&= \begin{bmatrix} \hat{\theta}_{IV}^N(1:n_a)^T & U(:,1)/U(1,1) \end{bmatrix}\\
\hat{\lambda}^T &= \sigma_1 V^T(1,:)\cdot U(1,1)
\end{aligned}
\tag{6.25}
$$

where $\sigma_1$ is the largest singular value and $U(1,1)$ denotes the first nonzero element of $U(:,1)$, where the notation (1,:) and (:,1) are used to denote the first row and the first column in a matrix respectively. In this way, the optimal system parameter vectors $\hat{b}$ and $\hat{\lambda}$ are obtained by minimizing the matrix Frobenius norm given by

$$[\hat{\lambda}, \hat{b}] = \arg \min_{\lambda \in R^M,\ b \in R^{n_b+1}} ||\Gamma_{b\lambda} - \hat{b}\hat{\lambda}^T||_F^2.$$

### 6.4.4 Iterative approach

In the constraint $\psi^T(\theta) \cdot [Y^L - \Phi^L \theta] = \vec{0}$ in Optimization Problem 9, $\psi^T(\theta)$ depends on the solution $\theta$. Thus, it cannot be used to compute $\hat{\theta}_{OE}^N$ directly. However, the (parameter dependent) instrument $\psi_k(\theta)$ can be calculated based on the previous parameter estimate $\hat{\theta}_{k-1}^N$. Let us summarize the iterative procedure to compute an OE parameter estimate $\hat{\theta}_{OE}^N$. With an initial parameter estimate $\hat{\theta}_{OE}^N = \hat{\theta}_k^N$ to model the static nonlinearity $\hat{f}$ and the linear dynamic system $\hat{G}$, one could employ an iterative solution that consists of the following computational steps:

**Step 1** : Separate $\hat{\theta}_k^N$ into $\hat{\eta}$ and $\hat{\lambda}$ in (6.25) and generate noise free input $\hat{y}(t)$ using (6.5) and noise free output $\hat{u}(t)$ using (6.1).

**Step 2** : Define the filter $L(q, \hat{\theta}_{k-1}^N) = A(q, \hat{\theta}_{k-1}^N)^{-1}$. If the filter $L(q, \hat{\theta}_{k-1}^N)$ is unstable, project the poles outside the unit circle inside the unit circle.

**Step 3** : Define $\Phi_{k-1}^L$ in (6.11), $\psi_{k-1}^T$ in (6.13) , and filtered output vector $Y_{k-1}^L$ in (6.14).

**Step 4** : Compute Optimization Problem 9 to obtain $\hat{\theta}_k^N$.

**Step 5** : Stopping criterion of the algorithm. If $\|\hat{\theta}_k^N - \hat{\theta}_{k-1}^N\| / \|\hat{\theta}_{k-1}^N\| < \varepsilon,$ stop. Otherwise, go to **Step 1**.

In the above steps, the stable filter $L(q)$, the filtered output vector $Y^L$, the filtered regressor $\Phi^L$ and the instrument $\psi^T$ are updated using $\hat{\theta}_{k-1}^N$ during the iterations over $k$. **Step 1** creates the noise free signals generated from closed-loop simulations. In **Step 2** the filter $L(q)$ is updated to provide the correct filtering for signals used in **Step 3**. In **Step 3**, the regressor and the instrument are calculated based on the gradient expression. **Step 4** is the actual computation of Optimization Problem 9 and **Step 5** formulates a stopping criterion for the algorithm by looking at the relative parameter error.

## 6.5 Numerical study

In this section, two numerical examples ($Case$ 1 and $Case$ 2) of nonlinear system identification using the proposed iterative nuclear norm minimization method are presented. The configuration in $Case$ 1 is the same as the example that appeared in [45] except that the controller gain is reduced for the closed-loop system stability. $Case$ 2 is similar to $Case$ 1, but the input static nonlinearity is replaced by a saturation nonlinearity in order to compare the efficiency of the proposed method for different static nonlinearities. An excitation signal $r(t)$ follows a uniform distribution with values between $-2$ and 2. The output disturbance $v(t)$ is filtered white noise. Twenty sets of estimation data with 2000 samples are generated for the system identification. The $M = 19$ grid points are equally spaced between $min(u(t))$ and $max(u(t))$ to model static nonlinearity for $Case$ 1. $M = 5$, with $m = [min(u(t)) \;\; -1 \;\; 0 \;\; 1 \;\; max\,(u(t))]^T$ is used for $Case$ 2. An $2^{nd}$ order model with 1 step time delay is used to model the linear dynamic system. The configuration of the simulation is shown in Figure 6.1.

$$
\begin{cases}
Case\ 1 \\
\quad f_0(u(t)) = sin(u(t)) - 0.5sin(2u(t)) + 0.4sin(3u(t)) \\
Case\ 2 \\
\quad f_0(u(t)) = \begin{cases} 1 & \text{if} \quad u(t) > 1 \\ u(t) & \text{if} \quad |u(t)| \le 1 \\ -1 & \text{if} \quad u(t) < -1 \end{cases} \\
G_0(q) = \dfrac{0.0997q^{-1} - 0.0902q^{-2}}{1 - 1.8858q^{-1} + 0.9048q^{-2}} \\
C(q) = 0.1\dfrac{10.75 - 9.25q^{-1}}{1 - q^{-1}} \\
H(q) = \dfrac{1 + 0.5q^{-1}}{1 - 0.85q^{-1}}
\end{cases}
$$

The results of applying the proposed iterative nuclear norm identification method to the closed-loop time domain data are shown in Figure 6.2 and Figure 6.3. The results show that the proposed method is very efficient in not only identifying non-smooth static nonlinearity due to the use of triangle basis functions, but also in identifying smooth static nonlinearity.

## 6.6 Application to LTO-3 tape drive servo actuator identification

In this section, the proposed iterative nuclear norm minimization method is applied to the experimental closed-loop time domain data from the servo actuator in a Quantum LTO-3 tape drive in order to identify the actuator dynamics and static nonlinearity existing in the closed-loop experiment. In this experiment, the tape drive was running at $4m/s$ causing periodic disturbances due to Lateral Tape Motion (LTM). An excitation signal $r$ was added to the output signal (the only change is from $u(t) = r(t) - y_c(t)$ in Figure 6.1 to $u(t) = C(q)(r(t) - y(t))$ in Figure 6.5) and the excitation level was chosen such that the control signal $u(t)$ to the plant was being saturated during the experiment. A total of $1,406,251$ actuator output measurements, in the form of a Position Error Signal (PES) at $16bit$ resolution, was measured for $70.3126sec$ sampled at $20kHz$. The controller $C(q)$ implemented during experiments is known. Only $N = 10,000$ (for $0.05sec$) data was used for the system identification. $M = 5$ (the total number of grid points) with $m = [min(u) \ -5 \ 0 \ 5 \ max(u)]$ is used to model static nonlinearity (we can start with $M > 5$ and remove unnecessary grid points as we go) and an $8^{th}$ order model with 1 step time delay is used to model the linear dynamic system. The configuration of the experiment is shown in Figure 6.5. The results of applying the proposed iterative nuclear norm minimization method to the closed-loop time domain data from the servo actuator in a Quantum LTO-3 tape drive is shown in Figure 6.6 and Figure 6.7.

Knowing that the LTO-3 drive has a saturation of $\pm 5V$ on the control input, it can be observed from Figure 6.6-(a) that the input saturation is properly estimated. In addition, several resonance modes have been estimated in the linear dynamic response of the actuator as indicated in Figure 6.6-(b). The resulting closed-loop Hammerstein system represents the LTO-3 actuator under control with input saturation and the simulation of this closed-loop Hammerstein system shows excellent agreement with the experiment data, as shown in Figure 6.7.

## 6.7 Conclusions

For Hammerstein system identification in a closed-loop setting, an iterative nuclear norm minimization method minimizing the output error based on gradient expression is proposed. This method allows us to solve a nonlinear OE minimization problem as an iterative linear optimization problem. In the proposed method, an instrument is calculated with filtered noise-free signals and their gradients where the filter is derived by a priori knowledge of the pole locations of the linear dynamic system, and where the noise free signals are computed from simulated closed-loop input and output signals generated by the known reference signal. For accurate computation of the closed-loop signals and the filter, an iterative procedure that updates the knowledge of the static nonlinearity and the linear dynamic system is used. Convergence of the iterative steps guarantees a local minimum of the OE minimization problem. The simulation and experimental studies show the effectiveness of the proposed algorithms in closed-loop identification of Hammerstein systems. The materials in Chapter 6 have been published in Y. Han and R. de Callafon, Output Error Identification of Closed-loop Hammerstein Systems, IEEE Conference on Decision and Control 2011, Orlando, US, and Y. Han and R. de Callafon, Closed-loop Identification of Hammerstein Systems Using Iterative Instrumental Variables, IFAC World Congress 2011, Milano, Italy. The dissertation author was the primary investigator and author of these papers.

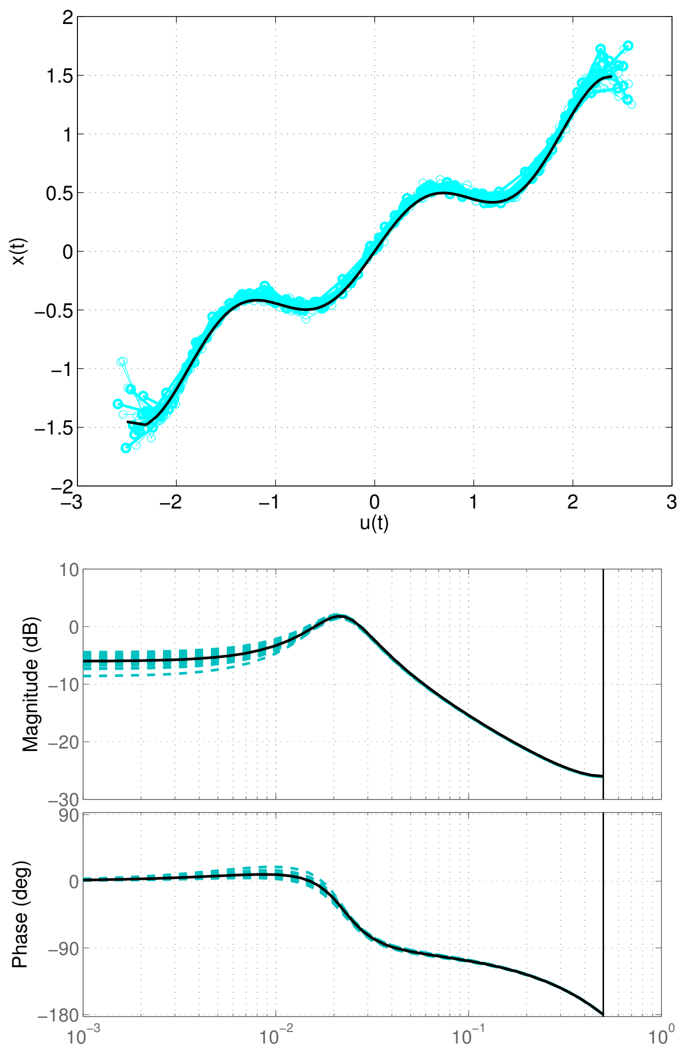



**Figure 6.2**: *Case* 1: The plot of the identified static nonlinearity function (top figure). The Bode plot of the identified linear dynamic system (bottom figure). The black solid line indicates the real Hammerstein system. The (colored) dashed lines indicate estimated systems by using twenty different sets of data. The SNR of each data set is greater than $20dB$.

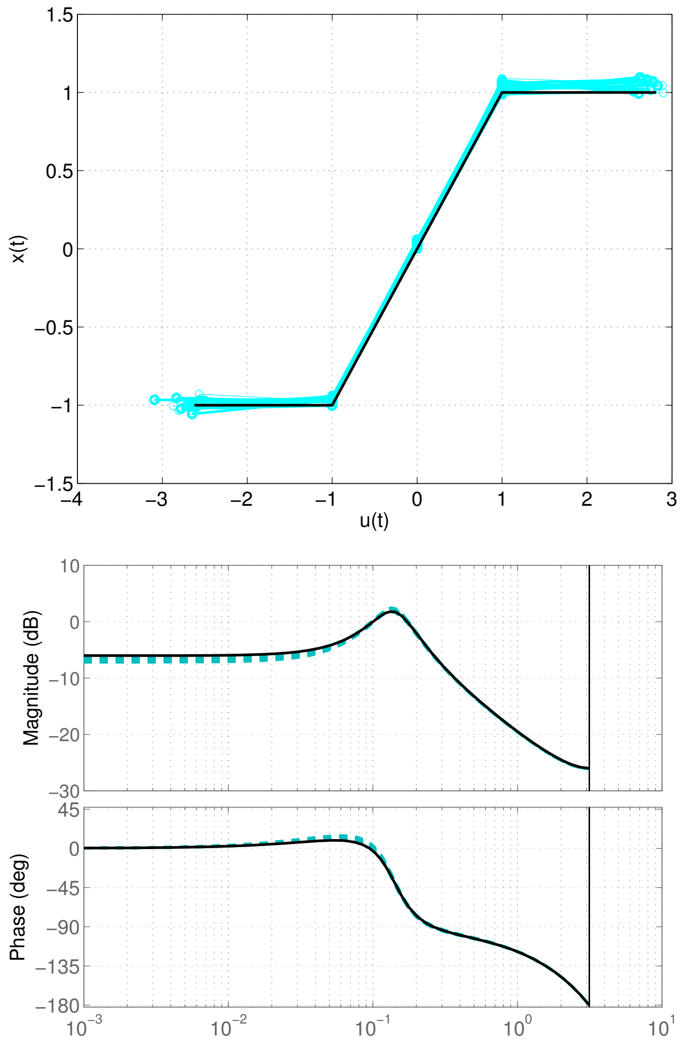



**Figure 6.3**: *Case* 2: The plot of the identified static nonlinearity function (top figure). The Bode plot of the identified linear dynamic system (bottom figure). The black solid line indicates the real Hammerstein system. The (colored) dashed lines indicate estimated systems by using twenty different sets of data. The SNR of each data set is greater than $20dB$.

**Figure 6.4**: Quantum LTO-3 tape drive.

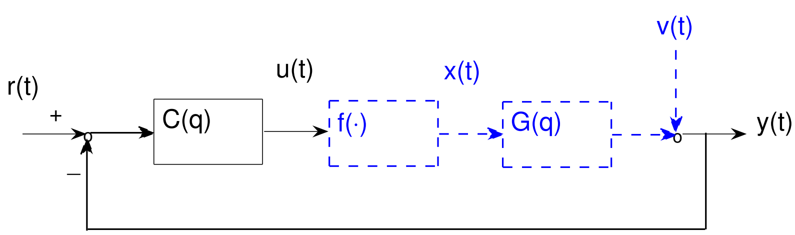
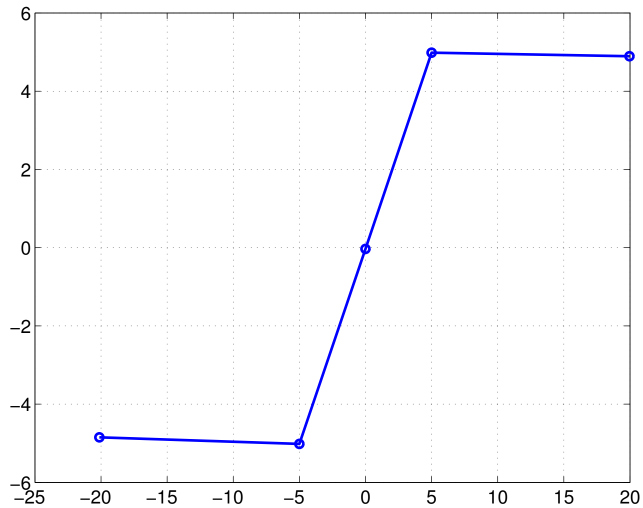
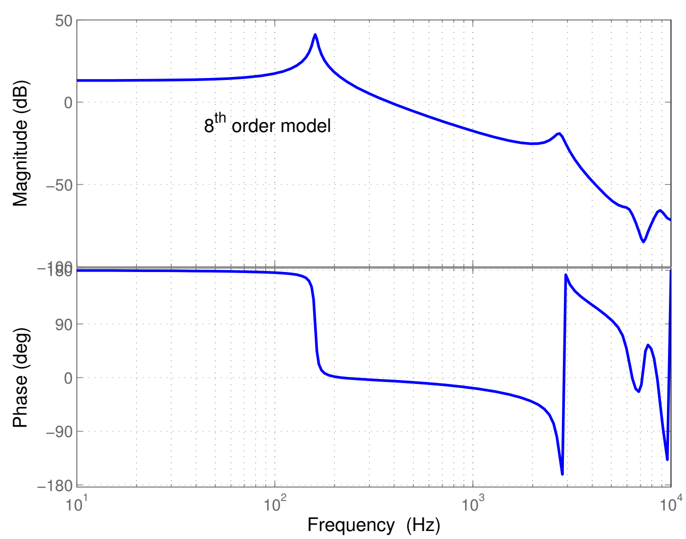



**Figure 6.5**: Closed-loop experimental setup of a Quantum LTO-3 tape drive. The excitation signal $r(t)$ and the linear controller $C(q)$ are known. The input to the static nonlinearity $u(t)$ and the output $y(t)$ are measured. The static nonlinearity $f(\cdot)$ and linear dynamic system $G(q)$ are unknown and need to be estimated under a unknown colored disturbance $v(t)$ (the dotted line indicates unknown parts and the solid line indicates the known parts).

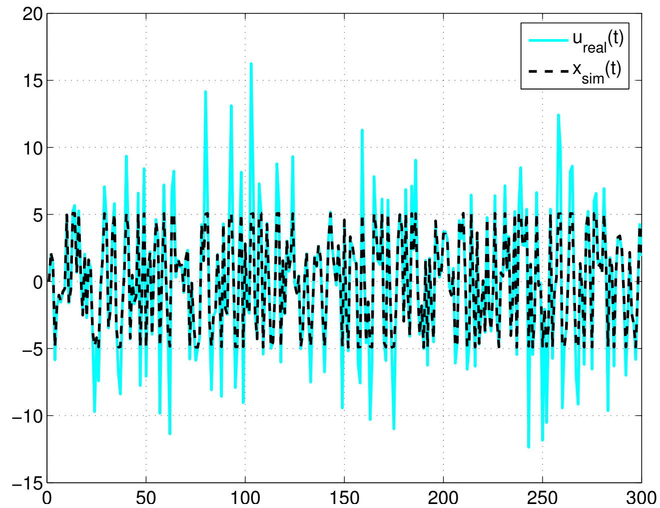


(a)

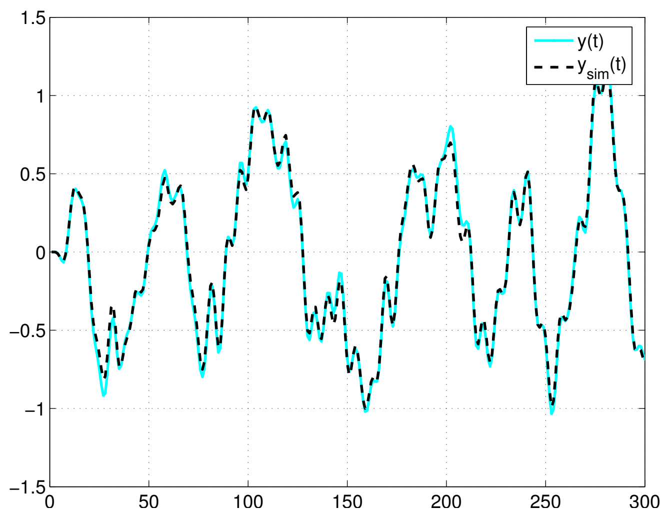



(b)

**Figure 6.6**: (a) The plot of the identified static nonlinearity function $\hat{f}(\cdot)$. (b) The Bode plot of the identified linear dynamic system $\hat{G}(q)$.




(a)




(b)

**Figure 6.7**: (a) The plot of the measured input signal $u(t)$ and the simulated intermediate signal $x_{sim}(t)$. The $\pm 5V$ input saturation is very nicely estimated. (b) The plot of the measured output signal $y(t)$ and the simulated output signal $y_{sim}(t)$.

# 7

# Conclusions and future research

## 7.1 Conclusion

In this dissertation, a new approach for block-oriented system identification has been studied. Identification of block-oriented nonlinear systems has been an active research area for the last several decades. In this study, the main focus has been placed on tackling the inaccessibility of measurement of intermediate signals in block-oriented nonlinear systems via rank minimization. The system parameter estimation problem is formulated as a rank minimization problem or as the combination of prediction error and rank minimization problems by constraining a finite dimensional time dependency between signals and by using a monotonicity of static nonlinearity. This allows us to reconstruct non-measurable intermediate signals. Since the rank minimization problem is non-convex, the use of the nuclear norm instead of the rank is proposed in order to define a semidefinite programming problem. It was proven that by solving the nuclear norm minimization problem, we can minimize a lower bound on the optimal value of the original rank minimization problem. This convex problem is easier to solve and the solution is close to the solution of the original non-convex problem. The idea of constraining rank for unmeasurable intermediate signal reconstruction can be applied, with some modifications, to any block-oriented system.

## 7.2 Future work

In order to impose a static relationship between signals, the monotonically non-decreasing condition on a static nonlinear function is extensively used in this study. However, the computational cost of imposing the monotonically non-decreasing condition could be high and the usability of the method for other static nonlinearity could be limited. Finding a way to define new constraints that are less expensive to compute and guarantee a static relationship between signals would be preferable.

# Bibliography


[1] J. Abony, R. Babuska, M. Ayala Botto, F. Szeifert, and L. Nagy. Identification and control of nonlinear systems using fuzzy Hammerstein models. *Industrial and Engineering Chemistry Research*, 39:4302–4314, 2000.

[2] H. Ase, T. Katayama, and H. Tanaka. A state-space approach to identification of Wiener-Hammerstein benchmark model. In *15th IFAC Symposium on System Identification*, pages 1092–1097, Saint-Malo, France, 2009.

[3] E. W. Bai and M. Fu. A blind approach to hammerstein model identification. In *40th IEEE Conference on Decision and Control*, pages 4794–4799, Orlando, Florida, 2001.

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

[5] A. Balestrino, A. Landi, M. Ould-Zmirli, and L. Sani. Automatic nonlinear auto-tuning method for hammerstein modeling of electrical drives. *IEEE Transactions on Industrial Electronics*, 48:645–656, 2001.

[6] J. Beyer, G Gens, and J Wernstedt. Identification of nonlinear systems in closed loop. In *Proc. 5th IFAC Symposium on Identification and System Parameter Estimation*, volume 2, pages 661– 667, Darmstadt, Germany, 1979.

[7] S. A. Billings and S. Y. Fakhouri. Identification of a class of nonlinear systems using correlation analysis, 1978.

[8] S. A. Billings and S. Y. Fakhouri. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18:15–26, 1982.

[9] R. S. Blom and P. M. J. Van den Hof. Multivariable frequency domain identification using IV-based linear regression. In *49th IEEE Conference on Decision and Control*, pages 1148–1153, Atlanta, GA, USA, 2010.

[10] J. Bobet, E. R. Gossen, and R.B. Stein. A comparison of models of force production during stimulated isometric ankle dorsiexion in humans. *IEEE*

*Transactions on Neural Systems and Rehabilitation Engineering*, 13:444–451, 2005.

[11] U. Boettcher, H. Li, R.A. de Callafon, and F.E Talk. Dynamic flying height adjustment in hard disk drives through feedforward control. *IEEE Transactions on Magnetics*, 47:1823–1829, 2011.

[12] M. Boutayed and M. Darouach. Recursive identification method for MISO Wiener-Hammerstein model. *IEEE Transactions on Automatic Control*, 40:287–291, 1995.

[13] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge, Cambridge, 2004.

[14] F. Chang and R. Luus. A noniterative method for identification using Hammerstein models. *IEEE Transactions on Automatic Control*, 16:464–468, 1971.

[15] C. H. Chen and S. D. Fassois. Maximum likelihood identification of stochastic Wiener-Hammerstein-type non-linear systems. *Mechanical Systems and Signal Processing*, 6:135–153, 1992.

[16] C.T. Chou and M. Verhaegen. An indirect approach to closed-loop identification of Wiener models. In *Proc. American Control Conference*, pages 3451–3455, Philadelphia, PA, 1999.

[17] P. Crama and J. Schoukens. Initial estimates of Wiener and Hammerstein systems using multisine excitation. *IEEE Trans. Instrum. Meas.*, 50:1791–1795, 2001.

[18] P. Crama and J. Schoukens. Hammerstein-Wiener system estimator initialization. In *Proc. ISMA 2002*, Leuven, Belgium, 2002.

[19] F. De Bruyne, B.O.D. Anderson, N. Linard, and M. Gevers. On closed identification with a tailor-made parametrization. In *Proc. American Control Conference*, pages 3177–3181, Philadelphia, PA, 1998.

[20] C. De Luca. Myoelectrical manifestations of localized muscular fatigue in humans. *Crit Rev Biomed Eng.*, 11:251–279, 1984.

[21] E. J. Dempsey and D. T. Westwick. Identification of hammerstein models with cubic spline nonlinearities. *IEEE Transactions on Biomedical Engineering*, 51:237–245, 2004.

[22] M. Enqvist and L. Ljung. Linear approximations of nonlinear FIR systems for separable input process. *Automatica*, 41:459–473, 2005.

[23] E. Eskinat, S. H. Johnson, and W. L. Luyben. Use of hammerstein models in identification of nonlinear systems. *AIChE Journal*, 37:255–268, 1991.

[24] T. Falck, Suykens, J.A.K., J. Schoukens, and B. De Moor. Nuclear norm regularization for overparametrized Hammerstein Systems. In *49th IEEE Conference on Decision and Control*, pages 7202–7207, Atlanta, GA, 2010.

[25] Y. Fang and T. Chow. Orthogonal wavelet neural networks applying to identification of Wiener model. *IEEE Transactions on Circuits and Systems*, 47:591–593, 2000.

[26] M. Fazel, H. Hindi, and S. Boyd. Rank minimization heuristic with application to minimum order system approximation. In *Proc. American Control Conference*, pages 4734–4739, Arlington, VA, 2001.

[27] M. Fazel, H. Hindi, and S. Boyd. Rank minimization and applications in system theory. In *Proc. American Control Conference*, pages 3273–3278, Boston, Massachusetts,, 2004.

[28] U. Forssell and L. Ljung. Identification of unstable systems using output error and box-jenkins model structures. *IEEE Transactions on Automatic Control*, 45:137–141, 2000.

[29] F. Giri and E.W. Bai. *Block Oriented Nonlinear System Identification*. Springer-Verlag, Berlin, 2010.

[30] I. Goethals, K. Pelckmans, J.A.K. Suykens, and B. De Moor. Subspace identification of Hammerstein systems using least squares support vector machines. *IEEE Transactions on Automatic Control, Special Issue on System Identification*, 50:1509–1519, 2005.

[31] W. Greblicki. Nonparametric identification of Wiener systems. *IEEE Transactions on information theory*, 38:1487–1493, 1992.

[32] W. Greblicki and M. Pawlak. Identification of discrete Hammerstein systems using kernel regression estimates. *IEEE Trans. Automatic Control*, 31:74– 77, 1986.

[33] P.L. Gribble and D.J. Ostry. Origins of the power law relation between movement velocity and curvature: Modeling the effects of muscle mechanics and limb dynamics. *Journal of Neurophysiology*, 76:2853–3860, 1996.

[34] R. Haber. Parametric identification of nonlinear dynamic systems based on nonlinear crosscorrelation functions. In *IEE Proceedings*, Pt. D, No.6, 1988.

[35] A. Hagenblad, L. Ljung, and A. Wills. Maximum likelihood identification of Wiener models. *Automatica*, 44:2697–2705, 2008.

[36] N. D. Haist, F. Chang, and R. Luus. Nonlinear identification in the presence of correlated noise using a Hammerstein model. *IEEE Transactions on Automatic Control*, 18:553–555, 1973.

[37] T.C. Hsia. *System Identification, Least-Squares Methods*. Lexington Books, Lexington, MA, 1977.

[38] K. J. Hunt, M. Munih, N. de N. Donaldson, and F. M. D. Barr. Investigation of the hammerstein hypothesis in the modeling of electrically stimulated muscle. *IEEE Transactions on Biomedical Engineering*, 45:998–1009, 1998.

[39] I. W. Hunter and M. J. Korenberg. The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, 55:135–144, 1986.

[40] Ding. J., A.S. Wexler, and S.A. Binder-Macleod. A mathematical model that predicts the force-frequency relationship of human skeletal muscle. *Muscle Nerve*, 26:477–485, 2002.

[41] L. Jia, M-S. Chiu, Ge S.S., and Z. Wang. Adaptive neuro-fuzzy identification method of Hammerstein model. In *IEEE Conference on Cybernetics and Intelligent Systems*, Singapore, 2004.

[42] L. Jia, M.S. Chiu, and S.S. Ge. Iterative identification of neuro-fuzzy-based Hammerstein model with global convergence. *Industrial and engineering chemistry research*, 44:1823–1831, 2005.

[43] F. Jurado. A method for the identification of solid oxide fuel cells using a Hammerstein model. *Journal of Power Sources*, 154:145–152, 2006.

[44] J. Kim and K. Konstantinou. Digital predistortion of wideband signals based on power amplifier model with memory. *IEE Electronics Letters*, 37:1417–1418, 2001.

[45] V. Laurain, M. Gilson, and H. Garnier. Refined instrumental variable methods for identifying Hammerstein models operating in closed loop. In *48th IEEE Conference on Decision and Control*, pages 3614–3619, Shanghai, China, 2009.

[46] N. Linard, B.O.D. Anderson, and F. De Bruyne. Closed loop identification of nonlinear systems. In *Proc. 36th IEEE Conference on Decision and Control*, pages 2998–3003, San Diego, CA, 1997.

[47] Z. Liu and L. Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 31:1235–1256, 2009.

[48] L. Ljung. System identification toolbox user guide.

[49] L. Ljung. *System identification: theory for the user*. Prentice-Hall, Inc, Upper Saddle River, New Jersey, 1999.

[50] J. Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *IEEE International Symposium on CACSD Conference*, pages 284–289, Taipei, Taiwan, 2004.

[51] D. N. Miller and R. A. de Callafon. Subspace identification from classical realization methods. In *15th IFAC Symposium on System Identification*, pages 102–107, Saint-Malo, France, 2009.

[52] B.Q. Mu and H.F. CHEN. Recursive identication for Wiener-Hammerstein system. In *Proc. 30th Chinese Control Conference*, Yantai, China, 2011.

[53] K.S. Narendra and P.G. Gallman. An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Transactions on Automatic Control*, 11:546–550, 1966.

[54] J. Paduart, L. Lauwers, R. Pintelon, and J. Schoukens. Identification of a Wiener-Hammerstein system using the polynomial nonlinear state space approach. In *15th IFAC Symposium on System Identification*, pages 1080–1085, Saint-Malo, France, 2009.

[55] J. M. Reyland. *Towards Wiener system identification with minimum a priori information*. Ph.D. Thesis, University of Iowa., 2011.

[56] T. Söderstöm and P. Stoica. Instrumental variable methods for system identification. *Circuits, Systems and Signal Processing*, 21:1–9, 2002.

[57] K. C. Sou, A. Megretski, and L. Daniel. Convex relaxation approach to the identification of the Wiener-Hammerstein model. In *47th IEEE Conference on Decision and Control*, pages 1375–1382, Cancun, Mexico, 2008.

[58] R. Srinivasan, R. Rengaswamy, S. Narasimhan, and R. Miller. Control loop performance assessment. 2. Hammerstein model approach for stiction diagnosis. *Industrial and engineering chemistry research*, 44:6719–6728, 2005.

[59] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11-12:625–653, 1999.

[60] S. W. Sung. System identification method for hammerstein processes. *Ind. Eng. Chem. Res.*, 41:4295–4302, 2002.

[61] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer Academic Publishers, 1996.

[62] J.W. van Wingerden and M. Verhaegen. Closed loop identification of MIMO Hammerstein models using ls-svm. In *15th IFAC Symposium on System Identification*, Saint-Malo, France,, 2009.

[63] G. Vandersteen. Identification of linear and nonlinear systems in an errors-in-variables least squares and total least squares framework., 1997. Phd- thesis, Vrije Universiteit Brussel.

[64] J. Vörös. An iterative method for Hammerstein- Wiener systems parameter identification. *Journal of Electrical Engineering*, 55:328–331, 2004.

[65] J. Vörös. An iterative method for Wiener-Hammerstein systems parameter identification. *Journal of Electrical Engineering*, 58:114–117, 2007.

[66] J. Wang, A. Sano, T. Chen, and B. Huang. Identification of Hammerstein systems without explicit parameterisation of non-linearity. *International Journal of Control*, 82:937–952, 2009.

[67] T. Wigren. Recursive prediction error identification using the nonlinear Wiener model. *Automatica*, 29:1011–1025, 1993.

[68] Q. Zhang, A. Iouditski, and L Ljung. Identification of Wiener system with monotonous nonlinearity. In *14th IFAC Symposium on System Identification*, pages 166–171, Newcastle, Australia, 2006.

[69] Y. Zhu. Identification of Hammerstein models for control using ASYM. *International Journal of Control*, 73:1692–1702, 2000.