Identification of Parallel Block-Oriented Models starting from the Best Linear Approximation

Thesis submitted in fulfillment of the requirements of the degree of Doctor in Engineering by

Maarten Schoukens

March 2015

Advisors: Prof. Dr. ir. Yves Rolain
           Prof. Dr. ir. Gerd Vandersteen
Identification of Parallel Block-Oriented Models starting from the Best Linear Approximation

Thesis submitted in fulfillment of the requirements of the degree of Doctor in Engineering by

Maarten Schoukens

Advisors: Prof. Dr. ir. Yves Rolain
          Prof. Dr. ir. Gerd Vandersteen

Members of the jury: Prof. Dr. ir. Lennart Ljung
                     Linköping University, Sweden
                     Prof. Dr. ir. Fouad Giri
                     University of Caen Basse-Normandie, France
                     Prof. Dr. ir. Joos Vandewalle
                     Katholieke Universiteit Leuven, Belgium
                     Prof. Dr. ir. Rik Pintelon
                     Vrije Universiteit Brussel, Belgium
                     Prof. Dr. ir. Annick Hubin (President)
                     Vrije Universiteit Brussel, Belgium
                     Prof. Dr. ir. Johan Deconinck (Vice-president)
                     Vrije Universiteit Brussel, Belgium
                     Dr. ir. Francesco Ferranti (Secretary)
                     Vrije Universiteit Brussel, Belgium
Abstract

Many systems over different fields of engineering exhibit a nonlinear dynamic behavior. Block-oriented models are often used to model nonlinear dynamical systems (e.g. chemical processes, power amplifiers, physiological systems, control valves and thermal microsystems). The dynamics and the nonlinearity of a system are modeled using two types of blocks: linear time invariant (LTI) blocks and static nonlinear (SNL) blocks. One can obtain many different block-oriented model structures to model a wide range of system behaviors by connecting the LTI and SNL blocks in various ways. For instance Hammerstein (series connection of SNL with an LTI), Wiener (LTI-SNL), Wiener-Hammerstein (LTI-SNL-LTI) and Hammerstein-Wiener (SNL-LTI-SNL) are often studied in the literature. The aim of this thesis is to develop and validate new methods to identify parallel block-oriented models. Parallel block-oriented models can be used to model multichannel topology systems, and to extend the flexibility of single branch models significantly.

First, the more simple cases of parallel Hammerstein and parallel Wiener systems are considered. The best linear approximation (BLA) of such systems is analyzed and used to obtain sound modeling techniques for these systems. Next, the proposed methods are extended towards parallel Wiener-Hammerstein models. Furthermore, the decoupling of multivariate polynomials is studied to simplify the representation of the static nonlinear
block that is present in a parallel Wiener or a parallel Wiener-Hammerstein model. Finally the proposed identification techniques are validated on real-life systems such as high frequency power amplifiers and biomedical systems.
During the last five years I had the opportunity to pursue a PhD at the ELEC department, VUB. Five years is a long period. There are many people that I worked together with during these five years, many colleagues that made working at ELEC a pleasant experience, many people that helped me, guided me, and supported me. In short, there are many people that deserve my gratitude. I thank you all.

I would like to thank my advisors Yves and Gerd for the open discussions we had, the support that you have given me, and the confidence that you have put in me over the years.

The comments and remarks of the members of the jury and the discussion at the private defense made this thesis a better work. I would like to thank you for this, it is much appreciated.

I thank the Research Foundation Flanders (FWO-Vlaanderen) for the financial support that made this research possible.

I spent three times a period abroad during my PhD (Uppsala, Linköping and Göteborg, all in Sweden). I learned a lot, and I had the opportunity to meet and to work together with many nice people during these stays. I thank prof. Er-Wei Bai, prof. Kristiaan Pelckmans, prof. Martin Enqvist, prof. Christian Fager and dr. Mattias Thorsell for their warm welcome. I would also like to thank my many colleagues in Sweden, you made my stays in Sweden a great experience!
Acknowledgments

I thank the ELEC department for creating a working environment where open discussions, collaboration and friendship amongst the colleagues are encouraged.

I thank my colleagues at the ELEC department. Thank you for the many discussions, collaborations and for the academical, technical and administrative support that you have given me. Also, thank you for the nice events we organized: the movie nights, dinners, LAN nights, futsal training, squash etc... I would like to thank my former and current office mates in particular for the many laughs, pleasant discussions and the friendship you gave me.

I would like to thank my parents and my sisters. You were there to help me in many different ways: advice, encouragement and little distractions. Thank you.

Finally, I would like to thank Anna. The list why is too long to be included in these acknowledgments.

Maarten

Brussels
April 2015
## Contents

Abstract \hspace{1cm} i

Acknowledgments \hspace{1cm} iii

Contents \hspace{1cm} v

List of Symbols \hspace{1cm} xiii

List of Abbreviations \hspace{1cm} xv

1 Introduction \hspace{1cm} 1
   1.1 System Identification \hspace{1cm} 1
      1.1.1 Data Collection \hspace{1cm} 2
      1.1.2 Model Structure Selection \hspace{1cm} 3
      1.1.3 Model Estimation \hspace{1cm} 4
      1.1.4 Model Selection and Model Validation \hspace{1cm} 4
   1.2 Nonlinear System Identification \hspace{1cm} 6
      1.2.1 Volterra Kernels \hspace{1cm} 6
      1.2.2 NARX and NARMAX Models \hspace{1cm} 7
      1.2.3 Nonlinear State Space Models \hspace{1cm} 8
      1.2.4 Block-Oriented Models \hspace{1cm} 9
   1.3 Continuous or Discrete Time \hspace{1cm} 10
## Contents

1.4 Research Objectives ........................................... 11  
1.5 Overview and Outline ......................................... 12

2 Best Linear Approximation .................................. 15

2.1 Introduction .................................................. 15  
2.2 Definition .................................................... 15  
  2.2.1 Defining the Signal Class, System Class and Noise Framework ............... 16  
  2.2.2 Best Linear Approximation ................................ 19  
2.3 BLA Estimation ............................................... 21  
  2.3.1 Nonparametric Estimation ............................... 22  
  2.3.2 Parametric Estimation ................................. 24  
  2.3.3 Consistency of the Parametric Estimate .............. 26  
2.4 The BLA for Gaussian Signals ......................... 26  
  2.4.1 Bussgang’s Theorem .................................. 27  
  2.4.2 BLA for other Classes of Excitation Signals ....... 28  
  2.4.3 BLA in the Presence of Even Nonlinearities ........ 28  
2.5 Experiment Design ......................................... 29  
2.6 Example System: a Hardening Spring ................... 30  
2.7 Conclusion and Discussion .................................. 32

3 Hammerstein and Wiener ..................................... 33

3.1 Introduction .................................................. 33  
3.2 The Hammerstein and Wiener System Class ............ 35  
3.3 Identifiability of the Hammerstein and Wiener System ... 36  
3.4 Best Linear Approximation for Gaussian Inputs ....... 37  
3.5 Hammerstein and Wiener Identification Procedure .... 39  
  3.5.1 Estimating the LTI block ............................... 39  
  3.5.2 Estimating the Static Nonlinearity: Hammerstein .... 40  
  3.5.3 Estimating the Static Nonlinearity: Wiener .......... 41  
  3.5.4 Persistence of Excitation .............................. 41  
  3.5.5 Consistency .............................................. 42  
  3.5.6 Nonlinear Optimization ............................... 44
### Contents

7.9 Simulation Example ........................................ 173
  7.9.1 System and Signals .................................. 173
  7.9.2 Model Estimation: SVD Approach ................... 175
  7.9.3 Model Estimation: MAVE Approach ................. 176
7.10 Conclusion and Discussion .............................. 181

8 Parallel Wiener-Hammerstein ............................. 185
  8.1 Introduction ........................................... 185
  8.2 System .................................................. 186
  8.3 Identifiability of a Parallel Wiener-Hammerstein System 188
  8.4 Best Linear Approximation ............................ 191
  8.5 Identification through BLA Decomposition .......... 193
    8.5.1 Estimating a BLA for Different Operating Conditions 195
    8.5.2 Parametrizing the BLAs ........................... 195
    8.5.3 Decomposing the BLAs ............................ 196
    8.5.4 Partitioning the Poles and Zeros ................. 199
    8.5.5 Improving the Estimated Nonlinearity ............ 202
    8.5.6 Persistence of Excitation ........................ 203
    8.5.7 Consistency of the Initial Estimates ............ 204
  8.6 Faster Branch-per-Branch Identification ............ 206
  8.7 Computational Aspects .................................. 209
  8.8 Nonlinear Optimization ............................... 211
  8.9 Simulation Example .................................... 212
    8.9.1 System and Signals ................................ 212
    8.9.2 Model Estimation .................................. 213
  8.10 Conclusion and Discussion ........................... 216

9 Applications .................................................. 221
  9.1 Introduction ........................................... 221
  9.2 Gluoregulatory System ................................. 221
    9.2.1 Problem Formulation .............................. 222
    9.2.2 Considered Model Structures ...................... 223
9.2.3 The Insulin-Glucose Modeling Problem: Multisine Input ........................................ 223
9.2.4 The Insulin-Glucose Modeling Problem: A more Realistic Case ................................. 229
9.2.5 Discussion ........................................ 231
9.3 Wiener-Hammerstein Benchmark ................................................................. 233
  9.3.1 Experimental Setup .................................................................................. 233
  9.3.2 Estimating the Parameters ................................................................. 234
  9.3.3 Validation .............................................................................................. 234
  9.3.4 Comparison of the Proposed Approach ................................................. 237
  9.3.5 Conclusion .............................................................................................. 240
9.4 Parallel Wiener-Hammerstein ................................................................. 240
  9.4.1 Measurement Setup ................................................................................ 240
  9.4.2 Input Design .......................................................................................... 241
  9.4.3 Model Estimation ..................................................................................... 241
  9.4.4 Model Validation ..................................................................................... 244
  9.4.5 Study of the Initialization Procedure ..................................................... 249
  9.4.6 Conclusion .............................................................................................. 250
9.5 Modeling a High Frequency Doherty Amplifier ............................................. 251
  9.5.1 The Doherty System ................................................................................ 251
  9.5.2 Measurement Setup ................................................................................ 252
  9.5.3 Model Estimation for the Doherty Amplifier ............................................ 254
  9.5.4 Model Validation ..................................................................................... 255
  9.5.5 Conclusion .............................................................................................. 257
9.6 Conclusion and Discussion .............................................................................. 257

10 Conclusion and Future Research Directions .................................................. 259
  10.1 Main Contributions and General Conclusion ............................................. 259
  10.2 Future Research Directions ....................................................................... 263
    10.2.1 Improving the Block-Oriented Modeling Tools: Structure Detection and Estimation ..... 263
    10.2.2 Towards a General Framework for Gray-Box Identification ...................... 266
10.2.3 Nonlinear Parameter Varying Identification . . . . . . 267
10.2.4 Advancing the Nonlinear Modeling and Predistortion
of High Frequency Power Amplifiers . . . . . . . . . . 268

List of Publications 271
Bibliography 279
List of Symbols

\( a \)  
scalar

\( \textbf{a} \)  
vector

\( \textbf{A} \)  
matrix or tensor

\( E\{.\} \)  
expected value operator

\( f_s \)  
sampling frequency

\( F \)  
\# of excited frequencies

\( f(.) \)  
static nonlinear function

\( g(.) \)  
static nonlinear function

\( G_{bla} \)  
resulting LTI model of the BLA

\( G(q) \)  
LTI system operator

\( G(e^{j\omega T_s}) \)  
LTI system \( G \) evaluated at the frequency \( \omega \) on the unit circle in the complex plane

\( H(q) \)  
LTI system operator

\( H(e^{j\omega T_s}) \)  
LTI system \( H \) evaluated at the frequency \( \omega \) on the unit circle in the complex plane

\( K \)  
regressor matrix

\( M \)  
\# of realizations

\( N \)  
\# of samples

\( \mathbb{N}_0 \)  
set of natural numbers, zero included

\( P \)  
\# of periods

\( q^{-1} \)  
backward shift operator
List of Symbols

\( R \)  
\# of setpoints

\( \mathbb{R} \)  
set of real numbers

\( \mathbb{R}^{n \times m} \)  
\( n \) by \( m \) matrix of real numbers

\( S \)  
system class

\( S(q) \)  
LTI system operator

\( S(e^{j\omega T_s}) \)  
LTI system \( S \) evaluated at the frequency \( \omega \) on the unit circle in the complex plane

\( S_{UU}(e^{j\omega T_s}) \)  
power spectrum (or spectral density) of \( u(t) \)

\( S_{YY}(e^{j\omega T_s}) \)  
power spectrum (or spectral density) of \( y(t) \)

\( S_{YU}(e^{j\omega T_s}) \)  
cross power spectrum (or spectral density) of \( u(t) \) and \( y(t) \)

\( T_s \)  
sampling period \((1/f_s)\)

\( T \)  
transformation matrix

\( \cdot^T \)  
transpose operator

\( u(t) \)  
discrete time input signal

\( U(k) \)  
discrete Fourier transform of \( u(t) \)

\( U \)  
input excitation class

\( V_N(\theta) \)  
cost function depending on \( \theta \)

\( V_{L\alpha,k_1,k_2,...,k_{\alpha-1}} \)  
frequency domain Volterra kernel of degree \( \alpha \)

\( y_0(t) \)  
discrete time noiseless output signal

\( y(t) \)  
discrete time noisy output signal

\( Y(k) \)  
discrete Fourier transform of \( y(t) \)

\( \theta \)  
parameter vector

\( \sigma^2_{G_{\text{tot}}.t} \)  
variance of the total distortion

\( \sigma^2_{G_{\text{noise}}.n} \)  
variance of the noise distortion

\( \omega \)  
angular frequency

\( \omega_k \)  
angular frequency corresponding to frequency line \( k \) of the DFT spectrum

\( \| \cdot \| \)  
\( l^2 \) (Euclidean) norm

\( \hat{\cdot} \)  
denotes an estimate
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLA</td>
<td>Best Linear Approximation</td>
</tr>
<tr>
<td>BIBO</td>
<td>Bounded Input Bounded Output</td>
</tr>
<tr>
<td>LTI</td>
<td>Linear Time Invariant</td>
</tr>
<tr>
<td>MAVE</td>
<td>Minimum Average Variance Estimation</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple-Input-Multiple-Output</td>
</tr>
<tr>
<td>MISO</td>
<td>Multiple-Input-Single-Output</td>
</tr>
<tr>
<td>NARX</td>
<td>Nonlinear AutoRegressive model with eXogenous inputs</td>
</tr>
<tr>
<td>NARMAX</td>
<td>Nonlinear AutoRegressive Moving Average model with eXogenous inputs</td>
</tr>
<tr>
<td>QBLA</td>
<td>Quadratic BLA</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>rms</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>SISO</td>
<td>Single-Input-Single-Output</td>
</tr>
<tr>
<td>SNL</td>
<td>Static NonLinear</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
</tbody>
</table>
Introduction

This thesis targets the identification of block structured models consisting of the series and parallel connection of linear dynamic and nonlinear static building blocks.

This clearly calls for a balance between accuracy, cost of model extraction and model complexity. In the last years the state of the art is slowly moving away from linear time invariant (LTI) models to nonlinear ones when the application calls for efficiency increases, whether energetic or performance based.

This thesis is situated in the field of nonlinear system identification. To start with, an introduction of the general system identification framework is given in Section 1.1. A short overview of some of the most common nonlinear model structures is given next (Section 1.2). The research objectives of this thesis are discussed in Section 1.4. Finally an outline of the text is given in Section 1.5.

1.1 System Identification

System identification is a discipline of (engineering) science that infers a model for a dynamic system from measured, noisy input-output data collected from a system. The influence of disturbing noise sources on the behavior of the model is eliminated as much as possible during the iden-
1. Introduction

tification process through the introduction of prior knowledge about these perturbations. Theoretically speaking, the goal is to obtain the system model with a reduced amount of measurements and a low model complexity. In practice, this goal boils down to finding a model that describes the system behavior 'sufficiently well', without increasing the model complexity 'too much'. Söderström and Stoica [1989]; Ljung [1999] and Pintelon and Schoukens [2012] provide a good overview of the general frameworks and methods that are used for system identification, mainly in the LTI framework.

Based on these classical references, one finds out that a standard system identification cycle consists of four main steps: the data collection, the model structure selection, the model estimation and the model validation. Each step is discussed briefly below.

1.1.1 Data Collection

During the data collection step, the user designs the experiment and performs it. A set of input excitation signals needs to be selected such that the data are informative enough to estimate the parameters of the model without noticeable modification of the system behavior. This design sets the bandwidth of the signal, the probability density distribution of the signal, the measurement sampling frequency, and the experiment length. Further choices also include whether band-limited or zero-order hold measurements are performed and whether the experiments should be performed in open or closed loop. Finally the measurement of the input \( u(t) \) and the output \( y(t) \) needs to be performed.

A more tricky choice aims to design the excitation power. In a theoretic LTI setting, 'more is better' as higher power results in a larger signal-to-noise ratio. Practically, things are much more complicated as both power and bandwidth will excite the linear and nonlinear aspects of the system. More power may create much more distortion than it will increase the signal-to-noise ratio. Again, a practical balance is to be sought here.
1.1.2 Model Structure Selection

The selection of the model is a user choice based on the prior knowledge about the model and on the intended usage purpose of the model (control, simulation, system predistortion, obtaining a deeper understanding of the system, ...). This decision is called the model structure selection step. The model structure selection includes the choice between white-box, gray-box or black-box models, parametric or nonparametric models, etc...

As a general rule, it is best to include as much prior knowledge about the system as is possible. This is not always easy, though. Including too much prior knowledge might complicate the model structure beyond its intended purpose, may complicate the estimation of the model parameters significantly or may even introduce bias error in the final result. For example, a white-box model for the insulin-glucose regulatory system includes a lot of prior knowledge obtained by physicians. It describes the behavior of the system very well but it is too complex for most practical control purposes (Marconato et al. [2014]).

One could even wonder if it is always necessary to try to describe a nonlinear system as well as possible. Often the complexity of the model is also an important criterion. Model complexity needs to be traded-off versus model accuracy. Model errors are definitely present when such a trade-off is made. Nevertheless, the importance of this step can hardly be over-emphasized as it directly conditions practical applicability.

Especially for nonlinear systems, the model structure selection step can influence the final result significantly. The class of nonlinear systems is very diverse and extensive. Finding a model structure that can describe nearly all nonlinear systems is very difficult, albeit feasible, and would result in an unnecessarily complicated model structure. Therefore, the user needs to select the model structure with care and validate this choice later during the model validation step (see section 1.1.4). Some of the more important classes of black-box nonlinear model structures are introduced in Section 1.2.

Sometimes 'less is more', and some initial linearized modeling efforts
1. **Introduction**

can guide the user in the model structure selection step for the nonlinear system under test (Lauwers et al. [2006]; Schoukens et al. [2015a]).

1.1.3 **Model Estimation**

The number of model parameters and their values need to be estimated such that the model matches the data as ‘closely’ as possible. This match is expressed using a norm called the cost function, and the model parameters are obtained by minimizing this norm. One of the most common cost functions is the quadratic cost function, also called least squares:

\[
\frac{1}{N} \sum_{t=1}^{N} (y(t) - \hat{y}(t, \theta))^2,
\]

where \(\hat{y}(t, \theta)\) is the modeled output depending on the parameter vector \(\theta\) and \(N\) is the total number of samples in the signal.

This cost function can be minimized in closed form using linear least squares if the model is linear in the parameters. If the model is nonlinear in the parameters, the solution can be obtained by splitting the problem into a series of separate subproblems that are linear in the parameters, or by using a nonlinear optimization algorithm (such as the Levenberg-Marquardt algorithm (Fletcher [1991]; Pintelon and Schoukens [2012])). Such nonlinear optimization algorithms need good initial values to start the optimization from.

Since the class of nonlinear systems is so broad, it is almost certain that the exact system will not have an exact model representation. Model errors will be present (the selected model structure is not ‘rich’ enough to describe the true behavior of the nonlinear system). Therefore it is important that the developed identification algorithms are robust with respect to the presence of such model errors.

1.1.4 **Model Selection and Model Validation**

The fourth and last step of a system identification cycle is the model selection and model validation step. An estimated model can fit the data that
1.1. System Identification

are used for the estimation well, but it can fail prediction completely on any other data set taken in similar experiment conditions. This means that a good model needs to remain valid for a class of experiment conditions. To check the quality of an estimated model, one needs to check if the model is valid for its purpose. One also would like to possibly compare it to other models of different complexity obtained on the same information. A model selection and validation step is therefore of great importance if practical applicability is targeted.

In this work, we will use a validation data sequence to determine the model quality. The validation data set is obtained from the system to be modeled. It is not used during the estimation step. Based on the error obtained between the measured and the simulated data using this data sequence as an excitation, we can assess if the model achieves the quality requirements. Model extrapolation outside the range of the estimation data set should in general be avoided during this validations step, even though it can sometimes be used to check the robustness of the model to extensions of the class of experimental conditions.

Many other model selection and validation tools exist in the literature (Ljung [1999]; Pintelon and Schoukens [2012]). Some methods do not need the use of an extra validation dataset. These methods include, but are not limited to, Akaike’s information criterion and the whiteness test of the model residuals. A parametric model can also be compared to its non-parametric counterpart (e.g. a rational transfer function with the frequency response function (FRF) it is describing).

If the model does not achieve the quality requirements set by the user, the identification procedure starts again, possibly by estimating new, more complex, models from the same model class or even another model class on the same data. If the available data are a limiting factor, more measurements on the system under test need to be performed.
1. Introduction

1.2 Nonlinear System Identification

This section gives a short overview of a subset of nonlinear model structures used for black-box nonlinear system identification that are relevant in the context of this text. The Volterra kernel representation, NARX and NARMAX, nonlinear state space and block-oriented model structures are considered to this end. One can think of many other nonlinear modeling approaches e.g.: restoring force models (Masri and Caughey [1979]) and neuro-fuzzy models (Lin and Lee [1991]; Wang and Mendel [1992]; Jang [1993]).

1.2.1 Volterra Kernels

One of the most well known representations for nonlinear systems is the Volterra kernel representation (Schetzen [1980]). The idea behind it is that the linear impulse response is extended to a nonlinear one in a multilinear way. A multilinear contribution of degree $\alpha$ is called a $\alpha$-th order kernel in this context. The model output consists of the sum of the outputs of the kernels of different degree. The output of a (causal) Volterra kernel $v_{\alpha}(k_1, k_2, \ldots, k_{\alpha})$ of degree $\alpha$ in discrete time is given by:

$$y_{\alpha}(t) = \sum_{k_1, k_2, \ldots, k_{\alpha}=0}^{n} v_{k_1, k_2, \ldots, k_{\alpha}} u(t - k_1)u(t - k_2) \ldots u(t - k_{\alpha}), \quad (1.2)$$

where $k_1, k_2, \ldots, k_{\alpha} \in \mathbb{N}_0$, and $n$ is the memory depth of the Volterra kernel. The kernel $v_{k_1, k_2, \ldots, k_{\alpha}}$ is symmetrized such that the order of the delays $k_1, k_2, \ldots, k_{\alpha}$ is of no importance (Schetzen [1980]).

This can also be represented in the frequency domain for a periodic excitation as (Pintelon and Schoukens [2012]):

$$Y_{\alpha}(k) = \sum_{k_1, k_2, \ldots, k_{\alpha}=1-N/2+1}^{N/2-1} V_{L_k,k_1,k_2,\ldots,k_{\alpha}-1}^\alpha U(L_k) U(k_1) \ldots U(k_{\alpha}-1), \quad (1.3)$$

where $L_k = k - k_1 - k_2 - \ldots - k_{\alpha-1}$. $N$ is the total number of samples in one period. $V_{L_k,k_1,k_2,\ldots,k_{\alpha}-1}^\alpha$ is a symmetrized frequency domain representation of the Volterra kernel of degree $\alpha$. $Y_{\alpha}(k)$ is obtained as the Discrete Fourier
Transform (DFT) of $y_\alpha(t)$:

$$Y_\alpha(k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} u(t) e^{-j2\pi tk/N}, \quad (1.4)$$
$$y_\alpha(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} U(k) e^{j2\pi tk/N}. \quad (1.5)$$

The major problem encountered estimating a Volterra kernel model is the large number of parameters that needs to be estimated. The number of parameters grows exponentially with the degree of the kernel. Volterra kernels come in handy to analyze and compare different model structures as in Westwick and Kearney [2003], where different block-oriented systems are compared as far as their modeling capacity is concerned based on their equivalent Volterra kernel representation.

### 1.2.2 NARX and NARMAX Models

NARX (Nonlinear AutoRegressive model with eXogenous inputs) and NARMAX (Nonlinear AutoRegressive Moving Average model with eXogenous inputs) can be seen as nonlinear extensions of the linear ARX and ARMAX model structures. The input-output relation of a NARMAX model is given by (Leontaritis and Billings [1985]; Billings [2013]):

$$y(t) = f(y(t-1), y(t-2), \ldots, y(t-n_y), u(t), u(t-1), \ldots, u(t-n_u), e(t-1), e(t-2), \ldots, e(t-n_e)) + e(t), \quad (1.6)$$

where $f(.)$ is a static nonlinear function, $e(t)$ is the perturbing noise sequence, and $n_y$, $n_u$ and $n_e$ are the maximum lags for the system output, input and noise. Note that delayed versions of the noise enter in the nonlinear function $f(.)$. This is not the case for the NARX model structure. The input-output relation of a NARX model is given by (Billings [2013]):

$$y(t) = f(y(t-1), y(t-2), \ldots, y(t-n_y), u(t), u(t-1), \ldots, u(t-n_u)) + e(t). \quad (1.7)$$
1. Introduction

The NARMAX and NARX model structure can describe a wide range of nonlinear systems (see Chen and Billings [1989]; Billings [2013]). The main advantage of the NARX and the NARMAX model structure over the Volterra kernel representation is parsimony. A NARX or NARMAX model structure will often need much less model parameters to obtain a good description of the input-output behavior of the system if a good model structure is selected. NARX and NARMAX structures are also good candidate model structures to model systems that exhibit nonlinear feedback behavior.

Several model structure detection techniques exist in the literature to select the structure of the nonlinear function \( f(.) \). However, the resulting models remain unstructured: the nonlinear behavior and the linear behavior are strongly entangled in the model description. A thorough discussion and overview of different NARMAX estimation, structure detection and model analysis methods can be found in Billings [2013] and the references therein.

1.2.3 Nonlinear State Space Models

Nonlinear state space models provide a very general structure that can be used to describe a nonlinear system. The discrete time single-input-single-output (SISO) input-output equation of an \( n_a \)-th order nonlinear state space model is given by (Paduart et al. [2010]):

\[
\begin{align*}
\mathbf{x}(t+1) &= f(\mathbf{x}(t), u(t)), \\
y(t) &= g(\mathbf{x}(t), u(t)),
\end{align*}
\]

where \( \mathbf{x} \in \mathbb{R}^{n_a \times 1} \) represent the states and the input \( u(t) \) and the output \( y(t) \) are scalars. Such a state space model structure can easily be extended for multiple-input-multiple-output (MIMO) descriptions.

The black-box identification of nonlinear state space systems received quite some attention (see Suykens et al. [1995]; Paduart [2008]; Paduart et al. [2010]; Schön et al. [2011]; Marconato et al. [2012a]). Nonlinear state space models can describe a wide range of nonlinear systems (Paduart [2008]). They are a good candidate to model systems that exhibit nonlin-
ear feedback behavior. The main disadvantage of a black-box state space model is the lack of parsimony and structure in the model. This is especially true when polynomial representations are used for the nonlinearities $f(.)$ and $g(.)$ (Paduart [2008]). There are several models that obtain structured state space descriptions by using gray-box modeling approaches or by transforming a black-box state space estimate into a structured one (Van Mulders [2012]; Van Mulders et al. [2013]; Noël and Kerschen [2013]). This results in a structured and parsimonious model, possibly with a physical meaning of the parameters.

1.2.4 Block-Oriented Models

Block-oriented model structures consist of two types of building blocks: linear time invariant (LTI) blocks and static nonlinear (SNL) blocks. These blocks can be interconnected to form the system model that needs to be estimated next. Series connection, parallel connection and feedback connection are possible ways to interconnect these building blocks (Giri and Bai [2010]).

The most simple series connections result in the Hammerstein (SNL-LTI) and Wiener (LTI-SNL) block structures (Hammerstein [1930]; Wiener [1966]; Narendra and Gallman [1966]; Billings [1980]; Hunter and Korenberg [1986]; Westwick and Kearney [2003]; Giri and Bai [2010]; Mzyk [2014]). These can be extended further to the Wiener-Hammerstein (LTI-SNL-LTI) and Hammerstein-Wiener (SNL-LTI-SNL) block structures (Billings [1980]; Westwick and Kearney [2003]; Giri and Bai [2010]; Mzyk [2014]). These block structured models are, as their name suggests, highly structured and parsimonious, but are not as general as the Volterra kernel models, NARMAX models or nonlinear state space models. This lack of generality can be solved by increasing the model complexity, while the structured and parsimonious nature of the models are maximally preserved. One way to increase the model complexity is to connect several branches in parallel. This results for instance in parallel Wiener, parallel Hammerstein or parallel Wiener-Hammerstein models (Palm [1979]; Billings [1980]; Korenberg [2010]).
1. Introduction

[1991]; Westwick and Kearney [2003]; Schoukens et al. [2015c]). Another way to increase the model complexity is by introducing feedback in the model structure as in Paduart [2008]; Schoukens et al. [2008a]; Vanbeylen [2013].

The main advantages of block-oriented models lie in their structured and parsimonious nature. The complexity of the model depends on the model structure. The model selection of a good model structure is not an easy problem without the availability of prior knowledge. It requires a trade-off between the quality of the model and the complexity of the model structure. Some model structure selection techniques are discussed in Haber and Unbehauen [1990]; Lauwers et al. [2006]; Schoukens et al. [2015a]. Another advantage of block-oriented models is that the identification of such a model can be separated into different steps using a suitable experiment setup (Billings and Fakhouri [1978]; Crama et al. [2004]; Enqvist [2005]; Rochdi et al. [2010]; Ikhouane and Giri [2014]).

1.3 Continuous or Discrete Time

Based on Schoukens et al. [2015b]

The identification algorithms that are presented in this work will focus on (but are in most cases not limited to) the identification of a discrete time model using discrete time data. Most systems under test are continuous time systems. The input and output of such a continuous time system are sampled, and often a discrete time model is estimated for prediction, or discrete time controller design purposes. The problem of using discrete time representations for sampled continuous time LTI systems is already well studied (Goodwin et al. [2010]; Yuz and Goodwin [2014]). Recently, some results have been obtained for the case of sampled nonlinear time invariant systems. The parametric representation of sampled state space systems is studied in Yuz and Goodwin [2005, 2014], and the case of sampled Volterra systems is studied in Koeppel and Schwingshackl [2004].
Can we represent a sampled nonlinear system of a certain structure, using a infinite order discrete-time model with the same structure, given a class of input excitation signals? It turns out that this is not always possible due to the sampled nature of the signals (Schoukens et al. [2015b]).

A sampled continuous time Wiener, Hammerstein, Wiener-Hammerstein or Volterra system can always be represented exactly by a discrete time Wiener, Hammerstein, Wiener-Hammerstein or Volterra model respectively when band-limited input signals are used and the sampling frequency is sufficiently high.

When the system is excited with zero-order hold input signals, a sampled continuous time system cannot always be represented exactly using a discrete time model with the same structure. Sampled Wiener, Hammerstein, and Volterra systems do not pose a problem and can be modeled exactly using discrete time Wiener, Hammerstein and Volterra models respectively. Sampled Wiener-Hammerstein systems cannot always be modeled exactly using a discrete time Wiener-Hammerstein model. However, the approximation error can be made arbitrary small if the input and output signals are sampled sufficiently fast.

Therefore, we can conclude that it is key to sample the input and the output sufficiently fast. A sample rate that is ten to twenty times higher than the highest excited frequency will avoid the presence of aliasing terms in most cases, and this will also typically result in an acceptable approximation error.

### 1.4 Research Objectives

The objective of this work is to develop identification methods for nonlinear systems. These identification methods will be designed to maximally meet the following properties:

- The model structure can describe a broad range of nonlinear systems.
- The model structure is parsimonious.
1. Introduction

- The model is easy to analyze and to interpret.

- The identification methods obtain 'good' results, even when the system does not belong to the model class.

The choice for nonlinear block-oriented models is motivated by these objectives. Nonlinear block-oriented models are, due to their structured nature, parsimonious and fairly easy to analyze and to interpret. However, most block-oriented models have a lack of general applicability. The most well known block-oriented structures, the Hammerstein and the Wiener structures, are really limited when it comes to practical applicability. This work goes beyond the Hammerstein and Wiener block-oriented structures using parallel cascade models. The main goal is to develop and validate estimation algorithms for parallel Wiener-Hammerstein models that remain robust to the presence of model errors.

The development of such a parallel Wiener-Hammerstein identification algorithm led to other research questions. How are single branch Wiener-Hammerstein systems identified? Which Wiener-Hammerstein identification method is best suited for the identification of parallel Wiener-Hammerstein systems? How can we identify parallel Hammerstein and parallel Wiener structures? How can we decouple multivariate polynomials to increase the parsimony of the model structure? And finally, how can the previously developed approaches be combined into a parallel Wiener-Hammerstein identification algorithm? These questions are answered in the following chapters.

1.5 Overview and Outline

In Chapter 2 we review one of the basic tools that will be used throughout the text, the Best Linear Approximation (BLA). This tool will be used in the following chapters to estimate the linear dynamics that are present in the nonlinear system around an operating point.
1.5. Overview and Outline

Chapter 3 introduces BLA-based Hammerstein and Wiener identification methods. It also introduces the basic modeling concepts that are used everywhere else.

The problem of identifying a Wiener-Hammerstein system is discussed in Chapter 4. Four different approaches to estimate the parameters of a Wiener-Hammerstein model are listed. The first approach (Section 4.5) uses genetic algorithms to obtain the estimate of the two LTI blocks that are present in a Wiener-Hammerstein model. It is based on the results that are presented in Schoukens et al. [2014d]. The second method (Section 4.6) uses a basis function expansion of the BLA to obtain the model of the two LTI blocks as in Tiels et al. [2014]. The third approach, presented in Section 4.7, separates the BLA in a front and back LTI contribution in a nonparametric way to obtain an estimate of the front and back dynamics (Schoukens et al. [2014b]). The last approach, in Section 4.8, was first introduced in Schoukens et al. [2014a] and Tiels et al. [2015]. It uses a special class of specifically tailored input signals to extract the front and back dynamics of the Wiener-Hammerstein system directly.

Chapter 5 makes a little sidestep away from the research objectives that are discussed in the previous section. It considers the problem of identifying a Hammerstein-Wiener system. The work presented in this chapter is an extension to the work that is presented in Schoukens et al. [2012b].

The extension of single branch block-oriented models to parallel block-oriented models occurs in Chapters 6 and 7. Chapter 6 discusses the identification of parallel Hammerstein systems. It is based on the results of Schoukens et al. [2010] and Schoukens et al. [2011]. Chapter 7 discusses the identification of parallel Wiener systems, based on the work in Schoukens and Rolain [2012b] and Schoukens et al. [2013a]. Also the decoupling of multivariate polynomials is discussed in Chapter 7 based on the results of Schoukens and Rolain [2012a].

The results obtained in the previous chapters are combined in Chapter 8 for the identification of parallel Wiener-Hammerstein systems. This chapter is based on the work published in Schoukens et al. [2013b] and Schoukens et al. [2015c].
1. Introduction

The proposed identification methods are applied to several measurement examples in Chapter 9 to show the good results and the practical applicability of the proposed identification methods, even when the system does not belong to the model class. This chapter is partially based on the results that are presented in Marconato et al. [2014], Schoukens et al. [2014b] and Schoukens et al. [2015c].

Finally, the conclusions of this work and some future research directions are discussed in Chapter 10.
2.1 Introduction

Most real-world systems do not behave completely linearly. Nevertheless, a linear model often explains a significant part of the behavior of a (weakly) nonlinear system. This approximative linear model also provides the user with a better insight into the behavior of the system under test. It motivates the use of a framework that approximates the behavior of a nonlinear system by a linear time invariant model under well-chosen system-specific boundary conditions. Throughout this text, we will use the Best Linear Approximation (BLA) framework (Ljung [2001]; Enqvist [2005]; Enqvist and Ljung [2005]; Pintelon and Schoukens [2012]) to estimate a linear approximation for a nonlinear system given a fixed set of such boundary conditions.

Section 2.2 introduces the theoretical framework of the BLA. Next, Section 2.3 discusses the use of the so-called robust method to estimate the BLA of a nonlinear system. Finally, the results of this chapter are illustrated by an example and are discussed in Sections 2.6 and 2.7.

2.2 Definition

The theoretical framework of the BLA results in a model whose structure is shown in Figure 2.1. It includes four components: one boundary condition,
2. Best Linear Approximation

Figure 2.1: The BLA of a nonlinear system for a given class of excitation signals $U$ consists of the resulting LTI model $G_{bla}(q)$, the unmodeled nonlinear contributions $y_s(t)$, and the additive noise source $n_y(t)$ that is assumed to be present at the output of the nonlinear system. The input excitation $u(t)$ belongs to the signal class $U$.

As mentioned above, the BLA of a system depends on the signal class $U$ that is used. This work considers $U$ to be the Riemann equivalence class of asymptotically normally distributed excitation signals (see also Schoukens et al. [2009a]; Pintelon and Schoukens [2012]). $U$ mostly contains Gaussian noise sequences, but considering the Riemann class allows one to extend $U$ to also contain periodic signal sets, which will prove very useful. When this class of signals is considered to provide an excitation signal, the BLA of many block-oriented systems becomes a simple function of the linear dynamics that are present in that system. All signals are assumed to be stationary for the remainder of this work.

Definition 2.1. Riemann equivalence class of asymptotically normally distributed excitation signals. Consider a signal $u(t)$ with a power spectrum $S_{UU}(e^{j\omega T_s})$. $S_{UU}(e^{j\omega T_s})$ is piecewise continuous, with a finite number of discontinuities. A random signal $u(t)$ belongs to the Riemann equivalence class if it obeys by any of the following statements:

1. $u(t)$ is a Gaussian noise signal with power spectrum $S_{UU}(e^{j\omega T_s})$. 

16
2.2. Definition

2. \( u(t) \) is a random multisine or random phase multisine (Pintelon and Schoukens [2012]) such that:

\[
\frac{1}{N} \sum_{k=k_1}^{k_2} E \{ |U(k)|^2 \} = \frac{1}{2\pi} \int_{\omega_{k_1}}^{\omega_{k_2}} S_{UU}(e^{j\nu T_s}) \, d\nu + O(N^{-1}),
\]

with \( \omega_k = k \frac{2\pi f_s}{N} \), \( k \in \mathbb{N} \), \( 0 < \omega_{k_1} < \omega_{k_2} < \pi f_s \), \( f_s \) is the sampling frequency.

A random phase multisine \( u(t) \) is a periodic signal with period length \( \frac{N}{f_s} \) defined in Pintelon and Schoukens [2012] as:

\[
u(t) = N^{-1/2} \sum_{k=-N/2+1}^{N/2-1} U_k e^{j(2\pi k \frac{t}{N} + \varphi_k)}, \quad k \neq 0
\]

\[
u(t) = N^{-1/2} \sum_{k=1}^{N/2-1} 2U_k \cos(2\pi k \frac{t}{N} + \varphi_k),
\]

(2.1)

where \( U_{-k} = U_k \) and \( \varphi_{-k} = -\varphi_k \) to obtain a real signal \( \nu(t) \). The phases \( \varphi_k \) are random variables that are independent over the frequency and are a realization of a random process on \([0, 2\pi]\), such that \( E\{e^{j\varphi_k}\} = 0 \). For instance, the random phases can be uniformly distributed between \([0, 2\pi]\).

The (real) amplitude \( U_k \) is set in a deterministic way by the user. \( U_k \) is uniformly bounded by \( M_U \) \( (U_k \leq M_U < \infty) \). Random phase multisines have the advantage of being periodic signals. This avoids the adverse effects of spectral leakage for a proper choice of the period length \( \frac{N}{f_s} \). They also offer full control over the applied amplitude spectrum to the user. The Riemann equivalence ensures that a random phase multisine is asymptotically \( (N \to \infty) \) Gaussian distributed (Pintelon and Schoukens [2012]).

**Assumption 2.1. System class:** \( \mathcal{S} \) is the class of nonlinear systems such that, when excited by a random phase multisine:

\[
\exists C_1, \text{s.t.} \sum_{\alpha=1}^{\infty} M_{G^\alpha} M_{U^\alpha}^\alpha \leq C_1 < \infty,
\]

(2.2)

with \( M_{G^\alpha} = \max |V_{\alpha,k_1,k_2,\ldots,k_{\alpha-1}}^\alpha| \) as defined in Section 1.2.1.
2. **Best Linear Approximation**

Assumption 2.1 postulates the existence of a uniformly bounded Volterra series whose output converges in mean square sense to the output of the non-linear system belonging to the system class $S$ (see Pintelon and Schoukens [2012] for more detail).

Although this assumption is mainly focused on the convergence of the series, some parallels can be made with the BIBO (Bounded Input Bounded Output) stability criterion. A single Volterra kernel is BIBO stable if (Schetzen [1980]):

$$\sum_{k_1, k_2, \ldots, k_\alpha = -\infty}^{+\infty} |v_{k_1, k_2, \ldots, k_\alpha}| < \infty.$$  \hspace{1cm} (2.3)

Eq. (2.3) is part of the set of Dirichlet conditions (Dirichlet [1829]) for the existence of the frequency domain Volterra kernels. The output of the Volterra series consists of the sum of the outputs of the Volterra kernels of different degree. This sum needs to converge to have a BIBO stable Volterra series. Some minimal knowledge of the input signal is required to set a restriction on the separate Volterra kernels (Schetzen [1980]). This leads to an expression similar to the one in Assumption 2.1. In short, if a Volterra series is composed of BIBO stable Volterra kernels, the condition in Assumption 2.1 guarantees the BIBO stability of that Volterra series for the considered class of input signals.

To give the reader physical insight in the definition of $S$, we provide some examples on the type of nonlinear behavior that is included in the system class $S$. $S$ includes for example systems with hard saturation and dead-zone nonlinearities, but excludes for example chaotic systems and hysteric systems.

**Assumption 2.2. Noise framework:** An additive, colored zero-mean noise source $v(t)$ with a finite variance $\sigma^2$ is present at the output of the system only:

$$y(t) = y_0(t) + v(t).$$  \hspace{1cm} (2.4)

This noise $v(t)$ is assumed to be independent of the known input $u(t)$. $y(t)$ is the actual output signal and a subscript 0 denotes the exact (unknown) value.
2.2. Definition

This noise framework will be used in the remainder of the text, unless explicitly stated otherwise.

The choice for the noise framework that is described in Assumption 2.2 is a simplified representation of reality. This simplification can lead to biased estimates when there are other noise sources present, located at other positions inside the system (Hagenblad et al. [2008]). A more realistic noise framework can be obtained by introducing multiple noise sources, or by placing the noise source at a different location in the considered system structure. This more realistic noise framework comes often at the cost of more complex identification algorithm. Multiple noise sources are, for instance, considered in Hagenblad et al. [2008]; Wills et al. [2013]; Lindsten et al. [2013]; Wahlberg et al. [2014].

2.2.2 Best Linear Approximation

The BLA model of a nonlinear system is an LTI approximation of the behavior of that system. It is best in mean square sense for a fixed class of input signals $U$ only. The BLA is defined in Enqvist [2005]; Enqvist and Ljung [2005]; Pintelon and Schoukens [2012] as:

$$G_{bla}(q) \doteq \arg \min_{G(q)} E \{ \left| \tilde{y}(t) - G(q) \tilde{u}(t) \right|^2 \} ,$$

(2.5)

where $E \{ . \}$ denotes the expected value operator, while the zero-mean signals $\tilde{u}(t)$ and $\tilde{y}(t)$ are defined as:

$$\tilde{u}(t) \doteq u(t) - E \{ u(t) \} ,$$

(2.6)

$$\tilde{y}(t) \doteq y(t) - E \{ y(t) \} .$$

(2.7)

This definition is equal to the definition of the linear time invariant second order equivalent model defined in Ljung [2001]; Enqvist [2005]; Enqvist and Ljung [2005] when the stability and causality restrictions imposed there are omitted.

The output residuals $n_t(t)$ of the model are given by:

$$n_t(t) = y(t) - G_{bla}(q) u(t)$$

(2.8)
2. Best Linear Approximation

These residuals represent the total distortion that is present at the output of the system. The total distortion can be split in two contributions based on their nature as is depicted in Figure 2.1. The nonlinear distortion $y_s(t)$ represents the unmodeled nonlinear contributions, while the noise distortion $n_y(t)$ is the additive noise that is assumed to be present at the system output. The nonlinear distortion and the noise distortion can be calculated separately as:

$$y_s(t) = y_0(t) - G_{bla}(q)u(t),$$

$$n_y(t) = y(t) - y_0(t),$$

where $y_0(t)$ is the unknown noiseless output. The nonlinear distortion $y_s(t)$ is uncorrelated with the input $\tilde{u}(t)$ for a linear correlation analysis. The nonlinear distortion $y_s(t)$ is not independent of the input $\tilde{u}(t)$ however (Pintelon and Schoukens [2012]). The noise source $n_y(t)$ on the contrary is typically assumed to be independent of the input $u(t)$.

It is shown in Enqvist [2005]; Enqvist and Ljung [2005]; Pintelon and Schoukens [2012] that the minimizer of eq. (2.5) is easily obtained as follows:

$$G_{bla}(e^{j\omega T_s}) = \frac{S_{\tilde{Y}\tilde{U}}(e^{j\omega T_s})}{S_{\tilde{U}\tilde{U}}(e^{j\omega T_s})},$$

where $S_{\tilde{Y}\tilde{U}}(e^{j\omega T_s})$ represents the cross-power spectrum or cross-power spectral density of $\tilde{u}$ and $\tilde{y}$. $S_{\tilde{U}\tilde{U}}(e^{j\omega T_s})$ is the auto-power spectrum (or power spectral density) of $\tilde{u}$. This can be further simplified in the case periodic excitation signals are considered. As is shown in Pintelon and Schoukens [2012] the minimizer then becomes:

$$G_{bla}(e^{j\omega k T_s}) = E\left\{ \frac{\tilde{Y}(k)}{\tilde{U}(k)} \right\}.$$  

Here $\tilde{Y}(k)$ and $\tilde{U}(k)$ are the leakage-free DFT spectra of the periodic signals $\tilde{y}(t)$ and $\tilde{u}(t)$ respectively. The expectation is taken over all realizations of $\tilde{U}(k)$ belonging to the signal class $\mathbb{U}$ and the resulting $\tilde{Y}(k)$.

Some important observations are based on the definition of the BLA:

1. The obtained model $G_{bla}(e^{j\omega T_s})$ does not only depend on the nonlinear system, but also on the class of input signals $\mathbb{U}$ that is used. The class
2.3. BLA Estimation

\( \mathbb{U} \) fixes both the probability density function (e.g. Gaussian inputs, uniform input distribution, etc.) and the power spectrum (or power spectral density) of the signals that are used to estimate it.

2. The obtained model \( G_{bla}(e^{j\omega T_s}) \) is only valid at excited frequencies. The model does not predict the spectral components that may appear due to spectral regrowth, or the harmonic content that is generated by the nonlinear system. These contributions are covered by the nonlinear distortion source \( y_s(t) \).

3. The nonlinear distortion \( y_s(t) \) depends both on the nonlinear system and on the input signal \( u(t) \). In fact, the nonlinear distortion can be better understood intuitively if one treats it as a random distortion source that realizes for each input signal in the class \( \mathbb{U} \) at hand rather than with time.

The last observation deserves an additional explanation. If a periodic signal is applied to a nonlinear system, the system reacts identically to each period when in steady state. A steady state periodic output is hence obtained. This periodic output contains the nonlinear distortion. As a consequence, the nonlinear distortion is the same for every period.

When the input signal is changed to another element of the class \( \mathbb{U} \), the contribution of the nonlinear distortion changes as well. This is why the nonlinear distortion can be better understood intuitively if one considers it to be a random distortion source that realizes w.r.t. the specific input signal at hand rather than with time.

2.3 BLA Estimation

The BLA can be estimated using different methods. Here, the so-called robust method is recapitulated in the presence of output noise only (Schoukens et al. [2004, 2005]; Pintelon and Schoukens [2012]; Schoukens et al. [2012a]). More details and other approaches, e.g. the so-called fast method or the
2. Best Linear Approximation

Figure 2.2: Measurement procedure for the robust method. $P$ steady state periods of $M$ signal realizations $u_i(t) \in U$ of a chosen class $U$ of input signals are measured.

The robust method in other noise frameworks, can be found in Pintelon and Schoukens [2012] and Schoukens et al. [2012a].

The robust method results in a nonparametric estimate of the three constituents of the BLA, the FRF part $G_{bla}(e^{j\omega kT_s})$, the variance of the total distortion $n_t(t)$ labeled $\sigma^2_t(k)$ and the variance of the noise distortion $n_y(t)$ labeled $\sigma^2_n(k)$. A nonparametric estimate is obtained at the excited frequency lines only. If needed, a parametric estimate of the BLA can be obtained in a second step based on these nonparametric estimates.

2.3.1 Nonparametric Estimation

The method that is presented here follows the lines of Schoukens et al. [2012a]. It is valid in an output noise framework only. The input $u(t)$ is assumed to be exactly known. The method requires multiple input-output measurements of the system excited successively by more than one signal. These signals $u_i(t)$ are drawn from a class of signals $U$ and each signal $u_i(t)$ is labeled as a ‘realization’ of the signal class $U$.

The $M$ different independent input signal realizations $u_i(t)$ excite the system in turn. $P$ steady state periods of each signal $u_i(t)$ are measured, as is shown in Figure 2.2. The $M$ different realizations are necessary to estimate the total distortion level $\sigma^2_t(k)$, and to obtain an averaged version...
2.3. BLA Estimation

$G_{bla}(e^{j\omega_k T_s})$. The $P$ different periods of one input $u_i(t)$ are used to obtain an estimate of the output noise level $\sigma_n^2(k)$ that is present in the measurements.

Both the nonlinear distortion $y_n(t)$ and the noise distortion $n_y(t)$ are stochastic contributions (Pintelon and Schoukens [2012]). The nonlinear distortion realizes over the signal realizations, while the noise distortion changes over the periods and over the realizations all together. Hence, averaging techniques can be applied to reduce and estimate the nonlinear and noise distortions separately. Starting from eq. (2.12) (in the case of additive noise on the output only) we obtain:

\[
G^{[m,p]}(e^{j\omega_k T_s}) = \frac{\tilde{Y}^{[m,p]}(k)}{\tilde{U}^{[m]}(k)} = G_{bla}(e^{j\omega_k T_s}) + \frac{Y_a^{[m]}(k)}{\tilde{U}^{[m]}(k)} + \frac{N_y^{[m,p]}(k)}{\tilde{U}^{[m]}(k)}.
\] (2.13)

\[
\hat{Y}(k)\text{ and }\tilde{U}(k)\text{ are the leakage-free DFT spectra of the periodic signals }\tilde{y}(t)\text{ and }\tilde{u}(t)\text{ respectively (see eq. (2.6) and eq. (2.7))}.\text{ Both the second and the third term are stochastic contributions. The second term contains the non-linear disturbances. The third term contains the output noise disturbance. By averaging over the different periods, one can reduce the influence of the output noise on the estimate of }G_{bla}.\text{ Taking the variance over the different periods results in the sample variance of the noise }\hat{\sigma}_G^{2}G^{[m]}, \text{ which is an estimate of the variance of the noise:}
\]

\[
G^{[m]}(e^{j\omega_k T_s}) = \frac{1}{P} \sum_{p=1}^{P} G^{[m,p]}(e^{j\omega_k T_s}),
\] (2.15)

\[
\hat{\sigma}_G^{2}G^{[m]}(k) = \sum_{p=1}^{P} \frac{|G^{[m,p]}(e^{j\omega_k T_s}) - G^{[m]}(e^{j\omega_k T_s})|^2}{P(P - 1)}.
\] (2.16)

To reduce the influence of the nonlinear distortions on the estimate of $G_{bla}$, and to reduce the influence of the noise even further, the estimate is averaged over the different realizations next. This results in the estimate $\hat{G}_{bla}$:

\[
\hat{G}_{bla}(e^{j\omega_k T_s}) = \frac{1}{M} \sum_{m=1}^{M} G^{[m]}(e^{j\omega_k T_s}).
\] (2.17)
2. Best Linear Approximation

The sample variance of the total distortion, and the sample variance of the noise are given by:

\[
\hat{\sigma}_{G_{bla,t}}^2(k) = \sum_{m=1}^{M} \left| G[m](e^{j\omega k T_x}) - \hat{G}_{bla}(e^{j\omega k T_x}) \right|^2 \frac{1}{M(M-1)},
\]

(2.18)

\[
\hat{\sigma}_{G_{bla,n}}^2(k) = \frac{1}{M^2} \sum_{m=1}^{M} \hat{\sigma}_{G[m]}^2(k).
\]

(2.19)

The sample variance of the total distortion \(\hat{\sigma}_{G_{bla,t}}^2\) includes the sample variance of the nonlinear distortions and the sample variance of the noise.

The variance on the measurement of one realization of the output averaged over the \(P\) measured periods is given by:

\[
\hat{\sigma}_{Y[m]_t}^2(k) = M \hat{\sigma}_{G_{bla,t}}^2(k) \left| U[m](k) \right|^2,
\]

(2.20)

\[
\hat{\sigma}_{Y[m]_n}^2(k) = M \hat{\sigma}_{G_{bla,n}}^2(k) \left| U[m](k) \right|^2.
\]

(2.21)

The difference between both variances is a measure for the nonlinear distortion:

\[
\hat{\sigma}_{Y[m]_s}^2(k) = \hat{\sigma}_{Y[m]_t}^2(k) - \hat{\sigma}_{Y[m]_n}^2(k).
\]

(2.22)

The system behaves quite linearly if there is a big gap between the magnitude of the output spectrum \(Y[m](k)\) and \(\hat{\sigma}_{Y[m]_s}(k)\), and very nonlinear if this gap is very small.

2.3.2 Parametric Estimation

A downside of a nonparametric approach is that the number of input signal realizations that are necessary to obtain results with a sufficiently low variability over the frequencies can be quite high, depending on the available signal-to-distortion ratio. A parametric model of the BLA is readily obtained when a rational form model \(\hat{G}_{bla}\) for \(G_{bla}\) is estimated. This reduces the number of parameters to be estimated in the model. Hence, the variability of the fewer estimated parameters is reduced. The simulated FRF is more smooth as the number of parameters in the rational form is typically much lower than the number of frequencies in the FRF. This means
2.3. BLA Estimation

that less input signal realizations are required to achieve the same model variability for $\hat{G}_{bla}$.

A parametric model is also more convenient for simulation and control purposes. It offers extra opportunities to better understand the system behavior using the pole-zero representation.

The model considered here is a rational function in the backwards shift operator $q^{-1}$:

$$\hat{G}_{bla}(q, \theta) = \frac{b_0 + b_1 q^{-1} + \ldots + b_{nb} q^{-nb}}{a_0 + a_1 q^{-1} + \ldots + a_{na} q^{-na}}. \quad (2.23)$$

The parameter vector $\theta \in \mathbb{R}^{(nb + na + 2) \times 1}$ contains the parameters $b_0, \ldots, b_{nb}, a_0, \ldots, a_{na}$. Since one parameter can be chosen freely because of the scaling invariance of the transfer function, only $nb + na + 1$ independent parameters need to be estimated. The order of the parametric model in eq. (2.23) can for example be determined using a signal theoretic measure such as the minimum description length (MDL) criterion (Pintelon and Schoukens [2012]).

The parametrization for continuous time models is completely analog, but now $s = j\omega$ is used as a free variable instead of $q^{-1}$.

A maximum likelihood framework is chosen to identify $\hat{G}_{bla}(q)$. As the estimator uses the previously estimated sample mean and the prior knowledge of the sample variance of the estimated nonparametric BLA, a sample maximum likelihood estimator results (Schoukens et al. [1997]; Söderström et al. [2010]; Pintelon and Schoukens [2012]):

$$V_{SML}(\theta) = \frac{F}{\sigma^2_{\hat{G}_{bla},t}(k)} \sum_{k=1}^{F} |\hat{e}(k, \theta)|^2, \quad (2.24)$$

$$\hat{e}(k, \theta) = \hat{G}_{bla}(e^{j\omega_k T_s}) - \hat{G}_{bla}(e^{j\omega_k T_s}, \theta), \quad (2.25)$$

where $\hat{G}_{bla}(e^{j\omega_k T_s}, \theta)$ is the spectral representation of the parametric model $\hat{G}_{bla}(q, \theta)$. $F$ is the number of excited frequencies that are present in the excitation signal, while $j\omega_1, \ldots, j\omega_F$ represents these excited frequencies.

One could also choose to directly estimate a parametric model of the BLA based on the measured input-output data. However, in that case
2. Best Linear Approximation

the estimates (2.18) and (2.19) cannot be obtained. Instead a parametric noise model needs to be estimated. This parametric noise model is then an estimate of \( \sigma^2_{\text{bla},t}(k) \), and not only the noise distortion as it is the case in the classical LTI identification framework. Hence, the noise model needs to be selected carefully if one chooses to estimate a parametric estimate of the BLA directly, as discussed in Schoukens et al. [2007].

2.3.3 Consistency of the Parametric Estimate

**Assumption 2.3. Model set:** \( G_{\text{bla}}(e^{j\omega T_s}) \) (eq. (2.11)) is contained in the reachable model set described by eq. (2.23).

**Theorem 2.1.** The parametrized BLA \( \hat{G}_{\text{bla}}(q, \theta) \) in eq. (2.23) is a consistent (convergence with probability 1) estimate of eq. (2.11) when the number of measured samples \( N \) tends to infinity, where the number of excited frequencies \( F \) is proportional to \( N \) \( (F = O(N)) \), and the number of input signal realizations \( M \geq 4 \) under the assumption that the BLA is part of the model set (Assumption 2.3) and the assumption of zero mean normally distributed errors.

**Proof.** See Section 10.7 and Theorems 10.3 and 9.21 in Pintelon and Schoukens [2012].

Since a nonparametric noise model is used during the identification, a minimum of 4 independent realizations \( M \) is required to obtain convergence of the parametric BLA estimate to its expected value (see Theorem 10.3 in Pintelon and Schoukens [2012]). This can be relaxed if a parametric rather than a nonparametric noise model is estimated.

2.4 The BLA for Gaussian Signals

The BLA can be used to extract the dynamics that are present in a nonlinear system. In many interesting cases the estimated BLA is a simple function of the linear dynamics that are present in the block-oriented system under
study when a specific input signal class, e.g. the class of Gaussian input signals, is used.

This section discusses some properties of the BLA when a Gaussian signal is applied to a static nonlinearity. First, Bussgang’s theorem is explained in Section 2.4.1. Section 2.4.2 investigates the effect of altering the class of excitation signals on the estimated BLA. The special case of systems with even nonlinearities only is considered in detail in Section 2.4.3.

2.4.1 Bussgang’s Theorem

Bussgang’s theorem is stated in Bussgang [1952] as follows:

**Theorem 2.2.** For two Gaussian signals, the crosscorrelation function taken after one of them has undergone nonlinear amplitude distortion is identical, except for a factor of proportionality, to the crosscorrelation function taken before the distortion.

**Proof.** See Bussgang [1952].

In practice this theorem allows us to replace a static nonlinearity that is excited by a Gaussian input by a multiplication with a constant in a correlation analysis of a nonlinear system.

\[ y(t) = f(u(t)), \quad (2.26) \]
\[ R_{yu}(\tau) = \alpha R_{uu}(\tau), \quad (2.27) \]

where

\[ R_{yu}(\tau) = E(y(t)u(t-\tau)), \quad (2.28) \]
\[ R_{uu}(\tau) = E(u(t)u(t-\tau)). \quad (2.29) \]

In the frequency domain this becomes:

\[ S_{YU}(e^{j\omega T_s}) = \alpha S_{UU}(e^{j\omega T_s}). \quad (2.30) \]
2.4.2 BLA for other Classes of Excitation Signals

Bussgang’s theorem has been extended to other classes of signals, besides the Gaussian class, using the concept of separable processes that is introduced in Nuttall [1958]. Gaussian processes, sine wave and phase modulated signals are shown to be separable processes in Nuttall [1958]. Furthermore, McGraw and Wagner [1968] have shown that the signals belonging to the class of elliptically symmetric distributions are also separable. A more in depth discussion of separable processes for nonlinear system identification can be found in Enqvist and Ljung [2005].

A biased estimate of the BLA (compared to the Gaussian BLA) is obtained when the input signals of the static nonlinearity are neither Gaussian input signals, nor input signals that are not listed as separable processes. This is studied in more detail for Wiener-Hammerstein systems in Wong et al. [2012]. It turns out that the bias error decreases in the cases studied in Wong et al. [2012] for an increasing impulse response length of the input LTI block $H(q)$ when maximum length binary sequences are applied at the input.

One can also design special input signals to extract more information from the nonlinear system using the BLA. In Schoukens et al. [2014a] a so-called phase coupled multisine signal is used to extract and to separate the front and the back dynamics of a Wiener-Hammerstein system. This is discussed in more detail in Section 4.8.

2.4.3 BLA in the Presence of Even Nonlinearities

Bussgang’s theorem shows that the BLA of a static nonlinearity that is excited by a Gaussian input signal is equal to a constant gain. This constant gain depends on the considered static nonlinearity, the power spectrum $S_{UU}(e^{j\omega T_s})$ of the input signal $u(t)$, and hence as well on the variance of the input signal $u(t)$.

A gain $\alpha = 0$ is obtained in the special case of a static nonlinearity $f(u(t))$ that is even with respect to the mean of its Gaussian input $u(t)$. This results in a BLA that is equal to zero.
2.5. Experiment Design

If the static nonlinearity is preceded or succeeded by an LTI block (as for Wiener or Hammerstein systems), these dynamics will not show in the estimated BLA. The problem is then that the BLA becomes zero, hence the linear dynamics that are preceding or succeeding the static nonlinearity are not identifiable. This problem can be avoided by designing an input signal with a different offset value (mean value), or, if no offset can be applied to the input of the system (AC coupling), by using a non-Gaussian signal with an asymmetric distribution. Remember that the estimated BLA can be biased as asymmetrically distributed non-Gaussian input signals are used.

2.5 Experiment Design

The estimation of the BLA with the robust method as described in Section 2.3.1 requires $M$ realizations and $P$ periods of a periodic signal. Random phase multisines are typically used in this work. During the design of a random phase multisine with a flat amplitude, the amplitude and the excited frequencies need to be set by the user. This means that the user needs to carefully select the amplitude and excited frequencies of the input signal, and choose how many periods $P$ and realizations $M$ are needed for the BLA estimation. This section formulates some guidelines for the user.

It is important that the input signal excites the frequency band of interest of the system. The length of one period, and hence also the length of the measurements, depend on the desired frequency resolution. This frequency resolution should be taken small enough such that all the resonances of the system are captured well. It is good practice to select the amplitude of the input signals used during the identification close to the amplitude of the input signals that will be used to excite the system in practice.

The number of steady-state periods $P$ and the number of realizations $M$ need to be chosen such that the variances of the BLA are ‘small enough’, this depends of course on the application of the model and on the desires of the user. A minimum of $P = 2$ periods and $M = 4$ realizations is advised (or even $M = 7$ depending on the desired properties of the estimator (Pintelon
and Schoukens [2012])). Both increasing $P$ and $M$ will decrease the noise variance on the BLA (eq. (2.19)), while only increasing $M$ will decrease the variance due to the nonlinearities on the BLA. Therefore, it is best to increase the number of realizations $M$ rather than increasing the number of periods $P$, if the goal is to obtain a high quality BLA estimate. However, it can be useful to increase the number of periods $P$, for instance if one wants to make a clear distinction between the variance due to the noise and the variance due to the nonlinear behavior of the system.

On top of this, several methods that are discussed in the following chapters will require that the users estimates the BLA at $R$ setpoints of the system. Also the number $R$ and the distribution of these setpoints need to be selected by the user. Section 6.5.1 discusses the choice of the number of setpoints and their distribution.

2.6 Example System: a Hardening Spring

As an example, the robust method is applied on the so-called Silverbox system (Marconato et al. [2012b]). The Silverbox is an electrical circuit that mimics a mass-damper-spring system with a nonlinear hardening spring.

The system is excited by a random phase multisine at three different amplitudes (0.045, 0.167 and 0.472 $V_{rms}$), sampled at 8 kHz, with 4096 points per period. All frequencies between DC and 1 kHz are excited, excluding DC. 20 realizations are measured at each amplitude. Three steady state periods of each realization are measured. The BLA is obtained using the robust method discussed before. The result is shown in Figure 2.3. Only the frequencies from 0 to 300 Hz are shown in Figure 2.3 to visualize the shift of the resonance frequency better.

Note that the resonance that is present in the BLA of the nonlinear Silverbox system shifts to lower frequencies for decreasing power levels. The resonance peak also becomes higher and that indicates that the damping decreases too. We can also observe in Figure 2.3 that the system response becomes also more linear for a decreasing excitation amplitude. The total variance (squares in Figure 2.3) is much higher than the noise variance
2.6. Example System: a Hardening Spring

Figure 2.3: The BLA of the Silverbox system. The results for the three different amplitudes are shown in blue (highest amplitude), green and red (lowest amplitude). The FRF is shown by the full line, the noise distortion sample variance estimate is shown by the dots, and the total distortion sample variance is shown by the squares. The variances are scaled with respect to one realization of the input signal.

(stars in Figure 2.3) for all cases. The total variance is hence dominated by the nonlinear distortion. This total distortion decreases for a decreasing amplitude.

A linear model clearly cannot model the nonlinear distortions of a nonlinear system. The expected value of the model residual variance obtained for an estimated LTI model are therefore expected to coincide with the variance of the total distortion.

A nonlinear model on the other hand models the nonlinear distortions. The expected value of the model residual variance of a nonlinear model is ideally coinciding with the noise distortion variance. We clearly see that for the Silverbox example a nonlinear model can perform significantly better than any LTI model. The ultimate goal of the ideal and exact nonlinear model is therefore to reduce the variance of the model residuals to the level of the measurement noise.
2.7 Conclusion and Discussion

The BLA framework approximates a nonlinear system by a different LTI model for each operation condition. Stated differently, the estimated BLA FRF approximates the system for a given excitation signal class only.

Note that the variance of the noise distortion and the total distortion (which contains both the noise and nonlinear distortion) are estimated together with the BLA. These estimates can be used to analyze the nonlinear behavior of the system quantitatively or qualitatively, and allows one to validate a nonlinear model when this is obtained. In the following chapters the BLA will be used to generate initial parameter estimates for nonlinear block-oriented models and to compare the different proposed approaches to each other and to the ideal model leaving only perturbation noise in the estimation residual.
3.1 Introduction

The Hammerstein and the Wiener structure are two of the more simple block-oriented model structures. A Hammerstein structure (Hammerstein [1930]) consists of a static nonlinear block followed by an LTI block (Figure 3.1). The Hammerstein structure is used to model nonlinear systems for which the static nonlinearity is present at the input of the system, such as actuators in general but also chemical processes and physiological systems (Giri and Bai [2010]). A Wiener structure (Wiener [1966]) consists of an LTI block followed by a static nonlinear block (Figure 3.2). A Wiener structure is often used to model systems where the nonlinear behavior is present at the output of the system. Some examples of such systems are systems with sensor nonlinearities, overflow valves and some physiological systems (Giri and Bai [2010]).

Many different Hammerstein identification algorithms exist in the literature and are listed in Giri and Bai [2010]. The first Hammerstein identification algorithm was introduced in Narendra and Gallman [1966], and further improved in Bai and Li [2004] and Bai and Li [2010]. Several other Hammerstein identification algorithms were developed in the last decades. A non-exhaustive list is given below where the methods are classified depending on their properties: nonparametric (or semi-parametric) SISO identifi-
cation algorithms (Mzyk [2014]), parametric SISO approaches (Chang and Luus [1971]; Crama et al. [2004]; Schoukens et al. [2007]), and blind SISO identification algorithms (Bai and Fu [2002]; Vanbeylen et al. [2008]). The MIMO Hammerstein case is considered in Verhaegen and Westwick [1996]; Goethals et al. [2005].

Also several Wiener system identification algorithms are listed in Giri and Bai [2010]. A wide variety of Wiener identification algorithms have been developed in the last decades. Some are grouped here based on their properties. A non-exhaustive list of methods contains: nonparametric (or semi-parametric) identification algorithms (Lindsten et al. [2013]; Mzyk [2014]), parametric approaches (Billings and Fakhouri [1977]; Hunter and Korenberg [1986]; Wigren [1993]; Crama and Schoukens [2001]; Westwick and Kearney [2003]), blind identification algorithms (Vanbeylen et al. [2009]), and for systems that contain backlash nonlinearities (Giri et al. [2014]). The MIMO Wiener system case is considered in Westwick and Verhaegen [1996]; Janczak [2007]. Most of the approaches consider that the noise source is present at the output of the system only, however Hagenblad et al. [2008]; Lindsten et al. [2013] allow for process noise to be present in the middle of the Wiener system, in between the linear dynamics and the nonlinearity.

This chapter does not introduce any new research, it illustrates the general concepts on a more simple block structure. These concepts are used later in Chapters 4, 6, 7, and 8.

This chapter is structured as follows. First, the class of Hammerstein systems and the class of Wiener systems are introduced in Section 3.2. Next, the identifiability of a Hammerstein system and a Wiener system is discussed in Section 3.3. The specific behavior of the BLA of these simple block-oriented systems is studied in Section 3.4. A Hammerstein and a Wiener identification algorithm is explained in Section 3.5. A simulation example is used to illustrate the good performance of the proposed estimators in Section 3.6. Finally, some conclusions are drawn in Section 3.8.
3.2 The Hammerstein and Wiener System Class

A Hammerstein system is a series (or tandem) connection of a static nonlinear block and an LTI block as is shown in Figure 3.1. The noiseless output \( y_0(t) \) of a Hammerstein system is given by:

\[
y_0(t) = S(q)r(t), \tag{3.1}
\]
\[
r(t) = f(u(t)). \tag{3.2}
\]

Herein, \( f(\cdot) \) is a static nonlinear function, \( S(q) \) is an LTI system, and the signals \( y_0(t) \), \( r(t) \) and \( u(t) \) are scalar signals.

A Wiener system is a series connection of an LTI block and a static nonlinear block as shown in Figure 3.2. The output of a Wiener system is given by:

\[
y_0(t) = g(x(t)), \tag{3.3}
\]
\[
x(t) = H(q)u(t), \tag{3.4}
\]

where \( g(\cdot) \) is a static nonlinear function, \( H(q) \) is an LTI system, and the signals \( y_0(t) \), \( x(t) \) and \( u(t) \) are scalar signals.
The static nonlinearities $f(.)$ and $g(.)$ are assumed to be a linear combination of basis functions (e.g. polynomial). This gives the following expression for the Hammerstein system class:

$$f(u(t)) = \sum_{i=1}^{n_f} \gamma_i f_i(u(t)),$$

or in the case of a Wiener system:

$$g(x(t)) = \sum_{i=1}^{n_g} \gamma_i g_i(x(t)),$$

where the coefficients $\gamma_i$ are grouped in the coefficient vector $\gamma \in \mathbb{R}^{n_f \times 1}$ in case of the Hammerstein system and $\gamma \in \mathbb{R}^{n_g \times 1}$ in case of the Wiener system. $n_f$ is the total number of nonlinear basis functions $f_i$ and $n_g$ is the total number of nonlinear basis functions $g_i$.

The LTI subsystems $S(q)$ and $H(q)$ are assumed to be a rational function of the backward shift operator $q^{-1}$:

$$S(q) = \frac{b_0 + b_1 q^{-1} + \cdots + b_{nb} q^{-nb}}{a_0 + a_1 q^{-1} + \cdots + a_{na} q^{-na}},$$

$$H(q) = \frac{d_0 + d_1 q^{-1} + \cdots + d_{nd} q^{-nd}}{c_0 + c_1 q^{-1} + \cdots + c_{nc} q^{-nc}},$$

where the coefficients $b_i$ and $a_i$ are grouped in the coefficient vectors $b \in \mathbb{R}^{(nb+1) \times 1}$ and $a \in \mathbb{R}^{(na+1) \times 1}$. The $l^2$-norm of the coefficient vector $a$ is assumed to be equal to 1, and the first nonzero element of $a$ is assumed to be positive. The coefficients $d_i$ and $c_i$ are grouped in the coefficient vectors $d \in \mathbb{R}^{(nd+1) \times 1}$ and $c \in \mathbb{R}^{(nc+1) \times 1}$. The $l^2$-norm of the coefficient vector $c$ is assumed to be equal to 1, and the first nonzero element of $c$ is assumed to be positive to compensate for the degeneration that is present in the rational transfer function parametrization.

### 3.3 Identifiability of the Hammerstein and Wiener System

The LTI block and the static nonlinear block of a Hammerstein system can only be identified up to an unknown gain, unless one has access to the
intermediate signal \( r(t) \). If a gain \( \beta \) is introduced for both \( S(q) \) and \( f(u) \) one obtains:

\[
y(t) = \tilde{S}(q) \tilde{r}(t) = \beta S(q) \tilde{r}(t), \tag{3.9}
\]
\[
\tilde{r}(t) = \tilde{f}(u(t)) = \frac{1}{\beta} f(u(t)). \tag{3.10}
\]

The input-output behavior of a Hammerstein system containing \( \tilde{S}(q) \) and \( \tilde{f}(u(t)) \) is equal to a Hammerstein system containing \( S(q) \) and \( f(u(t)) \) as building blocks.

A normalized estimate can be obtained by dividing the coefficients \( \gamma_i \) and multiplying the numerator coefficients of \( S(q) \) with the \( l^2 \)-norm of the coefficient vector \( \gamma \).

A completely analog reasoning can be applied to the Wiener system class.

Although theoretically any gain exchange \( \beta \) results in the same input-output behavior of the system, bad choices of this gain \( \beta \) can lead to numerical issues. It is best to avoid a high dynamic range over the different parameters that are present in the model. Also a high dynamic range over the different signals \( (u(t), r(t) \text{ and } y(t)) \) should be avoided if possible.

## 3.4 Best Linear Approximation for Gaussian Inputs

This section shows that the BLA of a Hammerstein and a Wiener system is a scaled version of the LTI dynamics that are present in \( S(q) \) or \( H(q) \) respectively when input signals belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals are used.

**Assumption 3.1. Input signal:** The input signal belongs to the Riemann equivalence class of asymptotically normally distributed excitation signals as defined in Definition 2.1.

**Theorem 3.1.** The BLA \( G_{bla}(e^{j\omega T_0}) \) of a Hammerstein system excited by an input signal belonging to the Riemann equivalence class of asymptotically normally
3. Hammerstein and Wiener

Normally distributed excitation signals (Assumption 3.1) is given by:

\[ G_{bla}(e^{j\omega T_s}) = \alpha S(e^{j\omega T_s}), \quad \alpha \in \mathbb{R}. \]  \hfill (3.11)

**Proof.** The BLA is given by eq. (2.11):

\[
G_{bla}(e^{j\omega T_s}) = \frac{S\hat{Y}\hat{U}(e^{j\omega T_s})}{S\hat{U}\hat{U}(e^{j\omega T_s})},
\]

\[ = S(e^{j\omega T_s}) \frac{S\hat{R}\hat{U}(e^{j\omega T_s})}{S\hat{U}\hat{U}(e^{j\omega T_s})}, \]

\[ = \alpha S(e^{j\omega T_s}), \]  \hfill (3.14)

where \( \frac{S\hat{Y}\hat{U}(e^{j\omega T_s})}{S\hat{U}\hat{U}(e^{j\omega T_s})} = \alpha \) due to Bussgang’s theorem, in combination with Assumption 3.1.

\[ \square \]

**Theorem 3.2.** The BLA \( G_{bla}(e^{j\omega T_s}) \) of a Wiener system excited by an input signal belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals (Assumption 3.1) is given by:

\[ G_{bla}(e^{j\omega T_s}) = \alpha H(e^{j\omega T_s}), \quad \alpha \in \mathbb{R}. \]  \hfill (3.15)

**Proof.** \( x(t) \) belongs to the Riemann equivalence class of asymptotically normally distributed excitation signals since it is a filtered version of \( u(t) \).

The BLA is given by eq. (2.11):

\[
S\hat{X}\hat{X}(e^{j\omega T_s}) = H(e^{j\omega T_s}) \tilde{H}(e^{j\omega T_s}) S\hat{U}\hat{U}(e^{j\omega T_s}),
\]

\[ S\hat{Y}\hat{X}(e^{j\omega T_s}) = \tilde{H}(e^{j\omega T_s}) S\hat{Y}\hat{U}(e^{j\omega T_s}), \]  \hfill (3.17)

\[ G_{bla}(e^{j\omega T_s}) = \frac{S\hat{Y}\hat{U}(e^{j\omega T_s})}{S\hat{U}\hat{U}(e^{j\omega T_s})}, \]

\[ = H(e^{j\omega T_s}) \frac{S\hat{Y}\hat{X}(e^{j\omega T_s})}{S\hat{X}\hat{X}(e^{j\omega T_s})}, \]

\[ = \alpha H(e^{j\omega T_s}), \]  \hfill (3.20)

where \( \frac{S\hat{Y}\hat{X}(e^{j\omega T_s})}{S\hat{X}\hat{X}(e^{j\omega T_s})} = \alpha \) due to Bussgang’s theorem, in combination with Assumption 3.1.

\[ \square \]
Note that the shape of \( G_{bla}(e^{j\omega T_s}) \) is independent of the power spectrum of the input signal. All \( G_{bla}(e^{j\omega T_s}) \) share the same shape up to a gain that is constant over all frequencies. The coefficient \( \alpha \) depends on many factors: the LTI subsystem \( S(q) \) or \( H(q) \), the static nonlinear subsystems \( f(u(t)) \) or \( g(x(t)) \), the power spectrum \( S_{UU}(e^{j\omega T_s}) \) of the input signal \( u(t) \), and hence as well on the variance of the input signal \( u(t) \).

**Assumption 3.2. Non-zero gain:** The gain \( \alpha \) present in eq. (3.11) and (3.15) are non-zero.

It can happen that the BLA is equal to zero, or in other words \( \alpha = 0 \). This is the case when the nonlinear function \( f(.) \) (or \( g(.) \)) is even around the expected value of \( u(t) \) (or \( x(t) \)). Different DC setpoints can be tried to avoid a zero gain \( \alpha \).

### 3.5 Hammerstein and Wiener Identification Procedure

The Hammerstein and Wiener identification algorithm that is presented in this section is based on the results of Schoukens et al. [2007]. First the BLA of the system is estimated. Next, the static nonlinearity is modeled using a linear-in-the-parameters model for the static nonlinearity (e.g. polynomial). This approach results in a consistent estimate of a Hammerstein system or Wiener system under weak assumptions. Finally, it can be worthwhile to use a nonlinear optimization when a finite number of data samples is used to refine the estimates further.

#### 3.5.1 Estimating the LTI block

In a first step the BLA of the Hammerstein (or Wiener) system is estimated. The BLA results in an estimate of the LTI block of the Hammerstein (or Wiener) system \( \hat{S} = \hat{G}_{bla} \) or the Wiener system \( \hat{H} = \hat{G}_{bla} \) up to an unknown gain \( \hat{\alpha} \) as was discussed earlier in Theorems 3.1 and 3.2. The estimate
can be both nonparametric ($\hat{G}_{bla}(e^{j\omega T})$) as in Section 2.3.1) or parametric ($\hat{G}_{bla}(q, \theta)$) as in Section 2.3.2).

**3.5.2 Estimating the Static Nonlinearity:**

**Hammerstein**

Next, a model for the static nonlinearity is needed. The static nonlinearity $f(u(t))$ is modeled as a weighted sum of nonlinear basis functions (see eq. (3.5)). The estimation of the static nonlinearity expanded in the basis is an estimation problem that is linear in the parameters $\gamma_i$:

$$y(t) = S(q) f(u(t)),$$

$$= \sum_{i=1}^{n_f} \gamma_i S(q) f_i(u(t)).$$

The regressor matrix $K \in \mathbb{R}^{N \times n_f}$ is constructed based on the previously estimated LTI block $\hat{S}(q)$:

$$K = \begin{bmatrix}
\hat{S}(q)f_1(u(1)) & \hat{S}(q)f_2(u(1)) & \cdots & \hat{S}(q)f_{n_f}(u(1)) \\
\hat{S}(q)f_1(u(2)) & \hat{S}(q)f_2(u(2)) & \cdots & \hat{S}(q)f_{n_f}(u(2)) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{S}(q)f_1(u(N)) & \hat{S}(q)f_2(u(N)) & \cdots & \hat{S}(q)f_{n_f}(u(N)) 
\end{bmatrix}$$

The least squares solution for the coefficient $\hat{\gamma}_i$ is given by:

$$\hat{\gamma} = (K^T K)^{-1} K^T y,$$

$$y = [y(1) \ y(2) \ \ldots \ y(N)]^T,$$

$$\hat{\gamma} = [\hat{\gamma}_1 \ \hat{\gamma}_2 \ \ldots \ \hat{\gamma}_{n_f}]^T.$$ 

In practice, the solution is obtained using a QR decomposition of $K$. To improve the numerical conditioning of the matrix, the columns of $K$ are divided by their $l^2$-norm. The conditioning of $K$ can also be improved by using orthogonal nonlinear basis functions (e.g. Hermite polynomials). A bad conditioning of the matrix can, for instance, be caused by the collinearity of the matrix, or due to the high dynamic range of the elements that
are present in the matrix. The scaling of the columns can solve the matrix conditioning issues due to a high dynamic range of the matrix entries over the different columns. Collinearity issues can be solved by using an appropriate set of basis functions to construct the matrix $K$.

### 3.5.3 Estimating the Static Nonlinearity: Wiener

The static nonlinearity $g(x(t))$ is modeled using a weighted sum of nonlinear basis functions (see eq. (3.6)). The estimation of the static nonlinearity expanded in the basis is linear in the parameters $\gamma_i$:

$$g(x) = \sum_{i=1}^{n_g} \gamma_i g_i(\hat{H}(q)u(t)), \quad (3.27)$$

The regressor matrix $K$ is constructed based on the previously estimated LTI block $\hat{H}(q)$ as follows:

$$K = \begin{bmatrix} g_1(\hat{H}(q)u(1)) & g_2(\hat{H}(q)u(1)) & \ldots & g_{n_g}(\hat{H}(q)u(1)) \\ g_1(\hat{H}(q)u(2)) & g_2(\hat{H}(q)u(2)) & \ldots & g_{n_g}(\hat{H}(q)u(2)) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(\hat{H}(q)u(N)) & g_2(\hat{H}(q)u(N)) & \ldots & g_{n_g}(\hat{H}(q)u(N)) \end{bmatrix}, \quad (3.28)$$

The least squares solution for the coefficient $\gamma_i$ is given by:

$$\gamma = (K^T K)^{-1} K^T y, \quad (3.29)$$

$$y = [y(1) \ y(2) \ \ldots \ y(N)]^T, \quad (3.30)$$

$$\gamma = [\gamma_1 \ \gamma_2 \ \ldots \ \gamma_{n_g}]^T. \quad (3.31)$$

In practice, just like for the Hammerstein case, the solution is obtained using a QR decomposition of $K$. To improve the numerical conditioning of the matrix, the columns of $K$ are divided by their $l^2$-norm.

### 3.5.4 Persistence of Excitation

Assumption 3.3. Persistence of excitation: The system is persistently excited by the signal $u(t)$. 

---

41
The assumption that the excitation is persistent is a very common assumption in system identification. This section discusses what persistence of excitation means for the proposed identification procedure.

The first step in the identification algorithm is to identify the parametric BLA of the nonlinear Hammerstein (or Wiener) system. It is important that the BLA identifies the dynamics that are present in the system correctly. Therefore, the number of excited frequencies in the input signal $u(t)$ needs be at least equal or preferably higher than $\frac{n_k+n_a+1}{2}$ (or $\frac{n_d+n_c+1}{2}$ for the Wiener case).

The static nonlinearity needs also to be estimated. For this identification step to work, the matrix $K$ in eq. (3.23) and eq. (3.28) needs to be of full rank. Put in other words, the nonlinear basis functions $f_i(u(t))$ (or $g_i(\hat{x}(t))$) need to be linearly independent over the domain of the input signal $u(t)$ (or $\hat{x}(t)$). Consequently, the amplitude range that is covered by $u(t)$ (or $\hat{x}(t)$) needs to be sufficiently large.

### 3.5.5 Consistency

To model the system exactly, the system needs to be contained in the reachable model set. Otherwise model errors will be present and the estimated parameters will not converge to the true parameters of the system. This requirement is formalized in the following assumption.

**Assumption 3.4. Model set:** The exact model of the system is contained in the reachable model set.

**Theorem 3.3.** The proposed estimator is a consistent (with probability 1 for $N \to \infty$) estimator for the class of Hammerstein systems under the following assumptions:

- the common assumption of zero-mean additive noise at the output only (Assumption 2.2),
- the excitation signal $u(t)$ is assumed to belong to the Riemann equivalence class of asymptotically normally distributed signals (Assumption 3.1),
3.5. Hammerstein and Wiener Identification Procedure

- the BLA is non-zero (Assumption 3.2),
- the input signal $u(t)$ is persistently exciting the system (Assumption 3.3),
- the system is contained in the reachable model set (Assumption 3.4).

This is needed for the estimated parameters to converge to the true parameters of the system up to an unknown gain $\beta$ (see Section 3.3).

Proof. The first step in the algorithm as proposed above estimates the BLA of the Hammerstein system. As was shown in Theorems 3.1 and 2.1 this results in a consistent estimate of the LTI block that is present in the Hammerstein system under the Assumptions 2.2, 3.1, 3.2 and 3.3.

In the last step of the estimation algorithm, the static nonlinearity is estimated. This estimation problem is linear in the parameters, and it is solved with a linear least squares approach. Hence, the estimate of the static nonlinearity is consistent under Assumptions 2.2 and 3.3.

The estimated parameters are therefore a consistent estimate and converge to the true parameters of the system up to an unknown gain $\beta$ under Assumption 3.4 (see Section 3.3).

\[ \square \]

Theorem 3.4. The proposed estimator is a consistent (with probability 1 for $N \rightarrow \infty$) estimator for the class of Wiener systems under the following assumptions:

- the common assumption of zero-mean additive noise at the output only (Assumption 2.2),
- the excitation signal $u(t)$ is assumed to belong to the Riemann equivalence class of asymptotically normally distributed signals (Assumption 3.1),
- the BLA is non-zero (Assumption 3.2),
- the input signal $u(t)$ is persistently exciting the system (Assumption 3.3),
the system is contained in the reachable model set (Assumption 3.4). This is needed for the estimated parameters to converge to the true parameters of the system up to an unknown gain $\beta$ (see Section 3.3).

Proof. The proof is completely analog to the proof given in Theorem 3.3.

3.5.6 Nonlinear Optimization

Joining all the previous estimation steps allows one to obtain the model parameter vector as a the estimation of a partition of the parameter vector $\theta$ in non-overlapping subsets. Note that the parameter vector $\theta$ now contains all the parameters of the model: $\gamma$, $b$, and $a$, $\theta \in \mathbb{R}^{(n_b+n_a+n_f+2) \times 1}$ in the case of the Hammerstein system class, $\gamma$, $d$, and $c$, $\theta \in \mathbb{R}^{(n_d+n_c+n_g+2) \times 1}$ in the case of the Wiener system class. Although this results in a consistent estimate when the number of data points $N$ tends to infinity (and hence the number of excited frequencies $F$ tends to infinity too, see Theorem 2.1), this typically yields a sub-optimal estimate for a finite number of data samples: the variability is too high.

To decrease the variability of the estimated parameters, one can fine-tune all the parameters simultaneously in a final nonlinear-in-the-parameters estimation step. The optimized parameters are obtained by calculating:

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{N} (y(t) - \hat{y}(t, \theta))^2,$$

(3.32)

where $\hat{y}(t, \theta)$ is the modeled output, depending on the parameters $\theta$.

This cost function unfortunately is non-convex with respect to the parameters $\theta$. A Levenberg-Marquardt algorithm (Levenberg [1944]; Marquardt [1963]; Fletcher [1991]; Pintelon and Schoukens [2012]) is used to minimize the cost function in a numerically stable way. This algorithm converges to the local minimum of the cost function that is 'closest' to the initial parameter values. Hence, good initial values of the parameters are very important to ensure the good quality of the final estimates. Such initial estimates can be obtained using the algorithm that is explained in Section 3.5.
The Jacobian of the cost function is rank deficient due to the identifiability issues (see Section 3.3). The related degenerations (due to the gain exchanges inside the LTI block and between the LTI block and the static nonlinearity) in the Jacobian needs to be taken into account during the optimization.

Eq. (3.32) obtains the optimized parameters using an unweighted time-domain cost function. A weighted cost function in the time or frequency domain can also be used to refine the parameters $\theta$. The estimated sample variance of the noise (see eq. (2.21)) can be used as a weighting function in a frequency domain cost function.

Note that the cost function in eq. (3.32) results in a maximum likelihood estimate if the additive noise is white Gaussian noise. A weighted version (using either a nonparametric noise model or a monic parametric noise model) of the cost function (eq. (3.32)) needs to be used to obtain a (sample) maximum likelihood estimate in the case of additive colored Gaussian noise. A maximum likelihood estimator is asymptotically efficient (Cramér [1649]; Ljung [1999]; Pintelon and Schoukens [2012]). This means that the maximum likelihood estimator achieves the lowest asymptotic mean squared error possible with a consistent estimator.

3.6 Simulation Example: Hammerstein

3.6.1 System and Signals

The proposed Hammerstein identification algorithm is illustrated on a simulation example. The simulated system is a Hammerstein system. The LTI block $S(q)$ is a third order Chebychev filter with a 3dB in-band ripple and a cut-off frequency at $0.3f_s$. The static nonlinearity $f(u)$ is a polynomial of degree 5:

$$S(q) = \frac{0.0029 + 0.0087q^{-1} + 0.0087q^{-2} + 0.0029q^{-3}}{1 - 2.5655q^{-1} + 2.3439q^{-2} - 0.7553q^{-3}}$$

$$f(u) = u + 0.10u^2 - 0.2u^3 + 0.05u^4 + 0.04u^5$$

The static nonlinearity is visualized in Figure 3.6.
White additive Gaussian noise with a standard deviation of 0.01 is added to the output of the system.

A random phase multisine is used as an input signal. 7 independent random phase realizations are applied to the system. Each realization is measured during 2 steady state periods, one period consists of 256 samples. All frequencies in the range \(0, \frac{f_s}{2}\) are excited. The random phases are uniformly distributed between \([0, 2\pi]\). The rms-value of the input signals is equal to 1. This results in an output signal-to-noise ratio of 40 dB.

### 3.6.2 Model Estimation

The BLA of the system is estimated in a first identification step using the robust method (see Section 2.3.1) followed by a parametrization of the nonparametric BLA (see Section 2.3.2). The estimated BLA is shown in Figure 3.3 together with the total and noise variances. The parametrized estimate of the BLA is used in the second step to obtain an estimate of the polynomial static nonlinearity. Finally, the parameters are refined in a nonlinear optimization step, optimizing all the parameters simultaneously.

The estimated model is validated using a different realization of the random phase multisine. This one has not been used during the estimation. The time domain representation of the modeled and measured validation output and residual error \((\hat{y}(t) - y(t))\) are shown in Figure 3.4. The frequency domain validation output and error are shown in Figure 3.5. It is clear from Figure 3.5 that the output of the Hammerstein system is quite nonlinear, there is only a 5 dB gap between the output spectrum and the total distortion variance. The estimated static nonlinearity coincides almost perfectly with the true nonlinearity, as can be seen in Figure 3.6.

The rms-value of the model error \(\text{rms}(\hat{y}(t) - y(t))\) is equal to 0.103 for the model after nonlinear optimization. The obtained model before the nonlinear optimization step resulted in an rms-value of the model error equal to 0.158. The rms-value of the error of the optimized model is roughly equal to the rms-value of the additive noise that is present at the output of the system.
3.7. Simulation Example: Wiener

3.7.1 System and Signals

The proposed Wiener identification algorithm is illustrated on a simulation example. The simulated system is a Wiener system. The LTI block \( H(q) \) is a third order low-pass Butterworth filter with a a cut-off frequency at 0.4\( f_s \), the static nonlinearity \( f(u) \) is a polynomial of degree 5:

\[
S(q) = \frac{0.0181 + 0.0543q^{-1} + 0.0543q^{-2} + 0.0181q^{-3}}{1 - 1.7600q^{-1} + 1.1829q^{-2} - 0.2781q^{-3}} \quad (3.35)
\]
\[
f(u) = u + 0.4u^2 - 0.5u^3 - 0.2u^4 + 0.2u^5 \quad (3.36)
\]

The static nonlinearity is visualized in Figure 3.10.

Random phase multisines are used as input signals. 7 realizations are applied to the system. Each realization is measured during 2 steady state periods, each period is 256 samples long. All frequencies in the range \([0, f_s/2]\) are excited. The random phases are uniformly distributed in the interval \([0, 2\pi]\). The rms-value of the input signal is equal to 1.
Figure 3.4: (a) Time domain measurement of the output and the output of the estimated Hammerstein model during validation. The noisy system output is depicted by the full blue line. The model output is represented by the dashed green line. (b) Time domain residuals $y(t) - \hat{y}(t)$ before (red) and after (green) optimization.
3.7. Simulation Example: Wiener

![Validation Output: Frequency Domain](image)

**Figure 3.5:** Frequency domain validation of the optimized Hammerstein model. The noisy system output is depicted by the full blue line. The model error before optimization is represented by the red pluses. The model error after optimization is represented by the green circles. The variance of the total distortion that is present at the output is shown by the squares, the noise distortion is represented by the stars.

White additive Gaussian noise with a standard deviation of 0.001 is added to the output of the system. This results in an output signal-to-noise ratio of 50 dB.

### 3.7.2 Model Estimation

The BLA of the system is estimated in a first identification step. The estimated BLA, together with the total and noise variance, is shown in Figure 3.7. The parametrized estimate of the BLA is used in the next step to obtain an estimate of the static nonlinearity.

The approach that is presented here is similar to the Wiener identification method that is presented in Billings and Fakhouri [1977].

The estimated model is validated using a newly realized random phase multisine. This signal is not used during the estimation of the model. The time domain measured signal, the model output and the residual error \( \hat{y}(t) - y(t) \) are shown in Figure 3.8. The frequency domain model vali-
3. **Hammerstein and Wiener**

![Figure 3.6:](image)

(a) The true static nonlinearity of the Hammerstein system (blue circles), and the estimated static nonlinearity of the optimized Hammerstein model (green dots). (b) The error between the true and the estimated static nonlinearity of the Hammerstein system.
3.8. Conclusion and Discussion

Figure 3.7: The estimate of the BLA of the Wiener system. The nonparametric estimate is represented by a dashed blue line. The full red line represents the parametrized BLA estimate. The estimate of the variance of the total distortion level of the averaged BLA is depicted with the squares. The stars represent the estimated variance of the noise distortion level.

The rms-value of the model error $\text{rms}(\hat{y}(t) - y(t))$ is equal to $1.03 \times 10^{-3}$ for the model after nonlinear optimization. The obtained model before the nonlinear optimization step results in an rms-value of the model error equal to $5.92 \times 10^{-3}$. The rms-value of the error of the optimized model is roughly equal to the rms-value of the additive noise that is present at the output of the system.

3.8 Conclusion and Discussion

This chapter introduces a simple 2-step identification procedure for Hammerstein and Wiener systems. First, the BLA of the system is estimated.
Figure 3.8: (a) Time domain measurement of the output and the output of the estimated Wiener model during validation. The noisy system output is depicted by the full blue line. The model output is represented by the dashed green line. (b) Time domain residuals $y(t) - \hat{y}(t)$ before (red) and after (green) optimization.
Next, the static nonlinearity is estimated using a linear-in-the-parameters static nonlinearity model. The proposed estimator is strongly consistent for the number of data points $N \to \infty$. To decrease the variance of the estimated parameters when a finite amount of data is used, one can fine-tune all the parameters simultaneously in a final nonlinear-in-the-parameters estimation step. The proposed methods are illustrated on a simulation example.
Figure 3.10: (a) The true static nonlinearity of the Wiener system (blue circles), and the estimated static nonlinearity of the optimized Wiener model (green dots). (b) The error between the true and the estimated static nonlinearity of the Wiener system.
Wiener-Hammerstein

4.1 Introduction

A Wiener-Hammerstein model has a structure that is a bit more involved than the Wiener or the Hammerstein model structure. The static nonlinearity is now sandwiched in between two LTI blocks (Figure 4.1). The presence of the two LTI blocks results in a problem that is harder to identify.

The problem of identifying a Wiener-Hammerstein system received quite some attention in the literature too. Many approaches use the BLA, or a similar correlation analysis, as a starting point for the algorithm (Billings and Fakhouri [1978]; Hunter and Korenberg [1986]; Crama and Schoukens [2005]; Lauwers [2011]; Tan et al. [2012]; Sjöberg and Schoukens [2012]; Sjöberg et al. [2012]; Westwick and Schoukens [2012]; Schoukens et al. [2014b]; Vanbeylen [2014]; Tiels et al. [2014, 2015]; Schoukens et al. [2014a]).

The maximum likelihood solution for Wiener-Hammerstein systems with an additive, colored zero-mean noise source that is present at the output of the system only is presented in Chen and Fassois [1992]. Vandersteen et al. [1997] estimate the two LTI blocks in a nonparametric way using a ‘large’ multisine signal combined with ‘small’ single sine experiments. Kibangou and Favier [2006] start from the diagonal of the Volterra kernels of the system to perform the identification. The MISO Wiener-Hammerstein case is discussed in Boutayeb and Darouach [1995].
This chapter is built up as follows. First, the Wiener-Hammerstein system class is introduced in Section 4.2. The identifiability of a Wiener-Hammerstein system is discussed in Section 4.3. The BLA of a Wiener-Hammerstein system is studied in Section 4.4. Next, four different Wiener-Hammerstein identification algorithms are explained in Sections 4.5, 4.6, 4.7 and 4.8. The first approach uses genetic algorithms to obtain the estimate of the two LTI blocks that are present in a Wiener-Hammerstein model. The second method (Section 4.6) uses a basis function expansion of the BLA to obtain the model of the two LTI blocks as in Tiels et al. [2014]. The third approach (Section 4.7) separates the BLA in a front and back LTI contribution in a nonparametric way to obtain an estimate of the front and back dynamics (Schoukens et al. [2014b]). The last approach (Section 4.8) uses a special class of specifically tailored input signals to extract the front and back dynamics of the Wiener-Hammerstein system directly (Schoukens et al. [2014a]; Tiels et al. [2015]). Finally, some conclusions are drawn in Section 4.10.

4.2 System

A Wiener-Hammerstein system is a series connection of an LTI block $H(q)$, a static nonlinear block $f(x(t))$, and a second LTI block $S(q)$ as shown in Figure 4.1. The output of a Wiener-Hammerstein system is given by:

\begin{align*}
y_0(t) &= S(q)r(t), \quad (4.1) \\
r(t) &= f(x(t)), \quad (4.2) \\
x(t) &= H(q)u(t). \quad (4.3)
\end{align*}
4.3 Identifiability Issues for the Wiener-Hammerstein Model

Herein, \( f(\cdot) \) is a static nonlinear function, \( H(q) \) and \( S(q) \) are LTI systems, and the signals \( y_0(t), r(t), x(t) \) and \( u(t) \) are scalar signals.

The static nonlinearity \( f(\cdot) \) is assumed to be a linear combination of basis functions (e.g. polynomial):

\[
f(x(t)) = \sum_{i=1}^{n_f} \gamma_i f_i(x(t)),
\]

where the coefficients \( \gamma_i \) are grouped in the coefficient vector \( \gamma \in \mathbb{R}^{n_f \times 1} \), and \( n_f \) is the total number of nonlinear basis functions.

The LTI subsystems \( H(q) \) and \( S(q) \) are assumed to be a rational function of the backward shift operator \( q^{-1} \):

\[
S(q) = \frac{b_0 + b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}}{a_0 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}},
\]

\[
H(q) = \frac{d_0 + d_1 q^{-1} + \ldots + d_{n_d} q^{-n_d}}{c_0 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}},
\]

where the coefficients \( b_i, a_i, d_i \) and \( c_i \) are grouped in the coefficient vectors \( b \in \mathbb{R}^{(n_b+1) \times 1}, a \in \mathbb{R}^{(n_a+1) \times 1}, d \in \mathbb{R}^{(n_d+1) \times 1} \) and \( c \in \mathbb{R}^{(n_c+1) \times 1} \). The \( l^2 \)-norm of the denominator vectors \( a \) and \( c \) are assumed to be equal to 1, and the first nonzero element of \( a \) and \( c \) are assumed to be positive to compensate for the degeneration that is present in the rational transfer function parametrization.

4.3 Identifiability Issues for the Wiener-Hammerstein Model

Different identifiability issues pop up during the identification of a Wiener-Hammerstein model (Vandersteen et al. [1997]). One can think of different Wiener-Hammerstein models that result in an identical input-output behavior. First of all, a gain can be swapped between the different blocks.

An arbitrary gain factor \( \beta_H \) can be moved from the front LTI block \( H(q) \) to the static nonlinearity \( f(x) \), and another arbitrary gain \( \beta_S \) can be exchanged between the back LTI block \( S(q) \) and the static nonlinearity
4. WIENER-HAMMERSTEIN

$f(x)$. This results in a double degeneracy in the parameters.

\[
y(t) = \tilde{S}(q)\tilde{r}(t) = \beta_S S(q)\tilde{r}(t), \tag{4.7}
\]

\[
\tilde{r}(t) = \tilde{f}(\tilde{x}(t)) = \frac{1}{\beta_S} f\left(\frac{1}{\beta_H} \tilde{x}(t)\right), \tag{4.8}
\]

\[
\tilde{x}(t) = \tilde{H}(q)u(t) = \beta_H H(q)u(t). \tag{4.9}
\]

Secondly, a phase shift degeneration is also present in the linear subsystems. Indeed, an arbitrary fractional sample delay $\tau$ can be introduced in the two LTI blocks of the system without any change in the response. To this end $H(q)$ is to be delayed by $\tau$ while $S(q)$ is delayed by $-\tau$.

\[
\tilde{H}(q) = \beta_H H(q)q^{-\tau},
\]

\[
\tilde{S}(q) = \beta_S S(q)q^\tau. \tag{4.10}
\]

This results in the same input-output behavior. The reason for the presence of the degeneration is that in the time domain such a phase shift is simply a delay of the signal at the output of the first LTI block $\tilde{H}(q)$. This does not affect the behavior of the static nonlinearity at all, as it has no memory and is time invariant. The delay that was caused by the first LTI block can therefore be exactly compensated by an inverse delay in the second LTI block $\tilde{S}(q)$.

This shows that the LTI blocks obtained during the identification of a Wiener-Hammerstein model can be scaled and delayed versions of the true LTI blocks that are assumed to be present in the system. The estimated static nonlinear block will be a scaled version of the true one.

Although theoretically the gain exchanges $\beta_S$ and $\beta_H$ result in the same input-output behavior of the system, bad choices of these gains $\beta_S$ and $\beta_H$ can lead to numerical issues. It is best to avoid a high dynamic range over the different parameters that are present in the model. Also a high dynamic range over the different signals ($u(t), x(t), r(t)$ and $y(t)$) should be avoided if possible.

58
4.4 Best Linear Approximation for Gaussian Inputs

This section shows that the BLA of a Wiener-Hammerstein system is a scaled version of the combined LTI dynamics that are present in \( H(q) \) and \( S(q) \) when input signals belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals are used.

**Assumption 4.1. Input signal:** The input signal belongs to the Riemann equivalence class of asymptotically normally distributed excitation signals as defined in Definition 2.1.

**Theorem 4.1.** The BLA of a Wiener-Hammerstein system excited by an input signal belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals (Assumption 4.1) is given by:

\[
G_{\text{bla}}(e^{j\omega T_s}) = \alpha H(e^{j\omega T_s})S(e^{j\omega T_s}), \quad \alpha \in \mathbb{R}.
\]

**Proof.** \( x(t) \) belongs to the Riemann equivalence class of asymptotically normally distributed excitation signals since it is a filtered version of \( u(t) \).

The BLA is given by eq. (2.11):

\[
S_{\tilde{X}\tilde{X}}(e^{j\omega T_s}) = H(e^{j\omega T_s})H(e^{j\omega T_s})S_{\tilde{U}\tilde{U}}(e^{j\omega T_s}),
\]

\[
S_{\tilde{Y}\tilde{U}}(e^{j\omega T_s}) = S(e^{j\omega T_s})S_{\tilde{R}\tilde{U}}(e^{j\omega T_s}),
\]

\[
S_{\tilde{R}\tilde{X}}(e^{j\omega T_s}) = H(e^{j\omega T_s})S_{\tilde{R}\tilde{U}}(e^{j\omega T_s}),
\]

\[
G_{\text{bla}}(e^{j\omega T_s}) = \frac{S_{\tilde{Y}\tilde{U}}(e^{j\omega T_s})}{S_{\tilde{U}\tilde{U}}(e^{j\omega T_s})},
\]

\[
= H(e^{j\omega T_s})S(e^{j\omega T_s}) \frac{S_{\tilde{R}\tilde{X}}(e^{j\omega T_s})}{S_{\tilde{X}\tilde{X}}(e^{j\omega T_s})},
\]

\[
= \alpha H(e^{j\omega T_s})S(e^{j\omega T_s}),
\]

where \( \frac{S_{\tilde{R}\tilde{X}}(e^{j\omega T_s})}{S_{\tilde{X}\tilde{X}}(e^{j\omega T_s})} = \alpha \) due to Bussgang’s theorem, in combination with Assumption 4.1.

**Note:** The shape of \( G_{\text{bla}}(e^{j\omega T_s}) \) is independent of the power spectrum of the input signal. All \( G_{\text{bla}}(e^{j\omega T_s}) \) share the same shape up to a gain.
that is constant over all frequencies. The coefficient $\alpha$ depends on many factors: the LTI subsystems $H(q)$ and $S(q)$, the static nonlinear subsystems $f(x(t))$, the power spectrum $S_{UU}(e^{j\omega T_s})$ of the input signal $u(t)$, and hence as well on the variance of the input signal $u(t)$.

4.5 Identification using Discrete Optimization

Based on Schoukens et al. [2014d]

This section presents the results that were introduced in Schoukens et al. [2014d]. First, the results of Sjöberg and Schoukens [2012] are recapitulated. Sjöberg and Schoukens [2012] present a Wiener-Hammerstein identification method that includes a brute-force scan to identify the linear dynamics that are present at the front and the back of the system. This brute-force scan is replaced by a discrete optimization approach in Schoukens et al. [2014d] to speed up the identification algorithm.

Vanbeylen [2014] proposes a fractional approach to replace the brute-force scan of the algorithm presented in Sjöberg and Schoukens [2012]. Such a fractional approach replaces the discrete optimization by a continuous optimization. This allows the approach to handle high model orders, and to handle the case where two real poles are modeled as a complex pole pair by the parametric BLA estimate. The main disadvantage of the proposed approach is that it needs a reparametrization step of the LTI blocks after the optimization of the fractional powers. Due to this second parametrization step, little or no speed-up is observed in practice.

4.5.1 Brute-force Identification by Pole-Zero Allocation Scan

The algorithm proposed in Sjöberg and Schoukens [2012] starts with the estimation of the poles and zeros of the BLA of the nonlinear system. When input signals belonging to the Riemann equivalence class of asymptotically
4.5. Identification using Discrete Optimization

normally distributed excitation signals are used, the BLA of a Wiener-Hammerstein system is given by (see Theorem 4.1):

\[ G_{BLA} = \alpha H(q) S(q), \quad (4.18) \]

\[ = \alpha \prod_{i=1}^{n_b+n_d} (z_i - q^{-1}) \prod_{i=1}^{n_a+n_c} (p_i - q^{-1}). \quad (4.19) \]

The poles \( p_i \) and the zeros \( z_i \) of the parametrized BLA are the combined poles and zeros of the LTI blocks \( H(q) \) and \( S(q) \) of the Wiener-Hammerstein system.

These poles and zeros of the BLA need to be assigned to either \( H(q) \) or \( S(q) \):

\[ G_{BLA} = \alpha \hat{H}(q) \hat{S}(q), \quad (4.20) \]

\[ \hat{H}(q) = \frac{\prod_{i=1}^{n_b+n_d} (z_i - q^{-1})^{\beta_{z_i}}}{\prod_{i=1}^{n_a+n_c} (p_i - q^{-1})^{\beta_{p_i}}}, \quad (4.21) \]

\[ \hat{S}(q) = \frac{\prod_{i=1}^{n_b+n_d} (z_i - q^{-1})^{1-\beta_{z_i}}}{\prod_{i=1}^{n_a+n_c} (p_i - q^{-1})^{1-\beta_{p_i}}}. \quad (4.22) \]

\( \beta_{p_i} \) and \( \beta_{z_i} \) are binary parameters that assign the pole or zero to either \( \hat{H}(q) \) \((\beta_{p_i} = 1)\), or to \( \hat{S}(q) \) \((\beta_{p_i} = 0)\). The parameters \( \beta_{p_i} \) and \( \beta_{z_i} \) are grouped in the binary parameter vector \( \beta \in \{0, 1\}^{(n_a+n_b+n_c+n_d) \times 1} \).

There is a finite number of different realizations of the parameter vectors \( \beta \). All possible realizations of the parameter vector are evaluated. To evaluate a realization of \( \beta \), a static nonlinearity is estimated for every realization \( k \). This problem is linear in the parameters if the static nonlinearity \( \hat{f}_k(x(t)) \) is described by a linear combination of basis functions. It is solved using linear least squares regression, resulting in the estimate \( \hat{f}_k \). Finally, all the estimated models are ranked based on their mean squares error \( e_k \):

\[ e_k = \frac{1}{N} \sum_{t=1}^{N} (y(t) - \hat{y}_k(t))^2, \quad (4.23) \]

\[ \hat{y}_k(t) = \hat{S}_k(q) [\hat{f}_k (\hat{H}_k(q) [u(t)])]. \quad (4.24) \]

where \( N \) is the number of data points in the measured input-output record. The model with the lowest error is selected.
The total number of realizations of $\beta$ depends on the number of poles and the number of zeros, $n_p$ and $n_z$ respectively. It equals the total number of least squares regressions that needs to be performed to scan all possible pole-zero allocations. The total number is minimum $2^{n_p+n_z}$ and maximum $2^{n_p+n_z}$. The minimum number of realizations is obtained when all poles and zeros are part of a complex conjugate pair, the maximum number of realizations is obtained when all poles and zeros are real. This number increases very rapidly with the model order, which makes the brute-force scan approach in Sjöberg and Schoukens [2012] computationally expensive.

It is important to note that each complex conjugate pole or zero pair is assigned either to $S(q)$ or $H(q)$ in the combinations. The single elements of the pair are never assigned separately. This is required as the LTI blocks consist of transfer functions with real coefficients only. This also introduces a possible disadvantage of the approach: it can happen that two real poles (or zeros) are combined into a complex conjugate pole (or zero) pair during the parametrization of the BLA. This pair cannot be correctly assigned if these poles (or zeros) originate from the two different LTI subsystems.

### 4.5.2 Pole-Zero Allocation using Discrete Optimization

The discrete optimization-based approach proposed in this chapter speeds up the process of finding the best pole and zero allocation in a Wiener-Hammerstein model. Instead of performing a brute-force scan off all possible realizations of the parameter vector $\beta$, the optimization of the binary parameter vector is performed using a genetic algorithm. Genetic algorithms belong to the larger class of evolutionary algorithms, which generate solutions to optimization problems using techniques inspired by natural evolution, such as mutation, selection and crossover (Back [1996]; Ahn [2006]). Genetic algorithms are used in several fields such as bioinformatics, computational science, engineering, economics, chemistry, mathematics and physics.

The static nonlinearity is estimated each time the cost function is eval-
4.5. Identification using Discrete Optimization

uated for a certain value of the parameter vector $\beta$, as explained in the previous section.

Instead of evaluating all possible pole-zero combinations, only a subset will be evaluated using the genetic algorithm. This results in a significant speedup of the identification algorithm, as will be shown below.

### 4.5.3 Simulation Example

A Monte Carlo simulation is performed to show the good results in accuracy and efficiency that are obtained using the discrete optimization approach. In each simulation a Wiener-Hammerstein system is created with a polynomial static nonlinearity of degree 3:

$$r(t) = 3x(t) + \gamma_2 x(t)^2 + \gamma_3 x(t)^3. \quad (4.25)$$

The coefficient of the linear term is equal to 3, the 2nd and 3rd degree coefficients $\gamma_2$ and $\gamma_3$ are uniformly distributed in the range $[-0.25, 0.25]$. The simulation is performed for different LTI orders. The front LTI block and the back LTI block are Chebychev filters of the same order ($n_a = n_b = n_c = n_d = [5, 6, 7, 8]$). The front LTI block is a Chebychev type 1 filter with an in band ripple of 3 dB and a cut-off frequency that is uniformly distributed in the range $[0.025f_s, 0.125f_s]$. The back LTI block is a Chebychev type 2 filter with a stop band ripple at 50 dB and a cut-off frequency that is uniformly distributed in the range $[0.025f_s, 0.125f_s]$. The cut-off frequencies of the LTI blocks are independent of each other.

The brute-force algorithm and the discrete optimization approach both use the same cost function. The brute-force algorithm evaluates this cost function for every possible parameter vector $\beta$. The discrete optimization approach evaluates this cost function for a much smaller set of parameter vectors. The genetic algorithm that is present in the Matlab Optimization toolbox (‘ga’ function name) is used. The optimization settings used for the simulations are shown in Table 4.1. An important setting of the genetic algorithm is the population size. If the population size is too small, the algorithm might get stuck in a local optimum, a too large population size slows down the algorithm since it needs to perform many function evaluations.
4. Wiener-Hammerstein

**Table 4.1:** Settings of the Matlab Optimization toolbox function ‘ga’.

<table>
<thead>
<tr>
<th>Setting Name</th>
<th>setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>PopulationType</td>
<td>bitstring</td>
</tr>
<tr>
<td>CreationFcn</td>
<td>gacreationuniform</td>
</tr>
<tr>
<td>CrossoverFcn</td>
<td>crossoverscattered</td>
</tr>
<tr>
<td>MutationFcn</td>
<td>mutationuniform</td>
</tr>
<tr>
<td>Generations</td>
<td>50</td>
</tr>
<tr>
<td>PopulationSize</td>
<td>see Table 4.2</td>
</tr>
<tr>
<td>StallGenLimit</td>
<td>5</td>
</tr>
<tr>
<td>TolFun</td>
<td>1e-20</td>
</tr>
</tbody>
</table>

**Table 4.2:** Monte Carlo simulation results: average time and success rates.

<table>
<thead>
<tr>
<th>$n_a = n_b = n_c = n_d$</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Size</td>
<td>200</td>
<td>400</td>
<td>600</td>
<td>800</td>
</tr>
<tr>
<td>brute-force Scan</td>
<td>8.73s</td>
<td>12.9s</td>
<td>149s</td>
<td>247s</td>
</tr>
<tr>
<td>Discrete optimization (GA)</td>
<td>4.12s</td>
<td>7.30s</td>
<td>14.7s</td>
<td>18.9s</td>
</tr>
<tr>
<td>Success Rate</td>
<td>98%</td>
<td>97%</td>
<td>95%</td>
<td>95%</td>
</tr>
</tbody>
</table>

The true zeros and poles of the LTI blocks are used, instead of the zeros and poles of the estimated BLA, to simplify the Monte Carlo simulation. Periodic Gaussian noise with a standard deviation equal to one is used as an input signal to evaluate the cost function. The system response is measured during one steady state period, one period consists of 4096 samples. 100 simulations are performed for each model order. The low degree of the nonlinearity in the system, and the limited number of data points that is used during the cost function evaluation ensure that the Monte Carlo simulation can be performed in a reasonable amount of time.

The discrete optimization approach clearly speeds up the identification process, as is shown in Table 4.2. A speed-up factor equal to 10 is obtained for the higher model orders, while a speed-up factor equal to 2 (a bit less than 2 for the model order 6 case) is obtained for the lower model orders. The success rate shows the percentage of the cases where the discrete optimization and the brute-force scan ended up in the same (global) minimum value of the cost function. A success rate of 95% to 100% is achieved in all cases. An even higher speed-up can be achieved, but this comes at the cost
of a lower success rate.

4.5.4 Conclusion

We have presented a fast identification algorithm for Wiener-Hammerstein systems. The problem is reformulated such that a binary parameter vector can be used to represent the combinations of poles and zeros in the front and the back LTI blocks. A genetic algorithm is used to perform the discrete optimization. The simulation results show that the proposed method is able to identify a Wiener-Hammerstein system ten times faster than the brute-force scanning approach for the high order models. The proposed method is still almost twice as fast as the brute-force method for lower order models. The proposed method is more efficient than the brute-force scan approach, while it achieves a very similar accuracy level.

4.6 Identification via Basis Function Expansion

Based on Tiels et al. [2014]

In this method the user selects a set of basis functions for the static non-linearity (polynomial, piece-wise linear, ...) and a second set of basis functions for the LTI blocks (tap delays, Laguerre filters, Takenaka-Malmquist basis functions, ... (Heuberger et al. [2005])). The basis functions of the LTI blocks can be derived from the BLA (Tiels and Schoukens [2011]) of the system, or can be selected based on prior knowledge.

First, the problem of estimating a Wiener-Hammerstein model is reformulated as a simplified problem that is bilinear in the parameters. Such a problem can easily be solved using an iterative scheme (Bai and Liu [2005]). Next, the parameters of the front LTI block are obtained using either a linearization or a dimension reduction approach. The biggest advantage of this identification approach lies in its simplicity. Besides that, little restrictions are imposed on the input signals. On the other hand, the user has the
possibility to introduce some prior knowledge during the model estimation step. The results of this section are presented in Tiels et al. [2014].

4.6.1 Problem Formulation

The noiseless output of a Wiener-Hammerstein system is given in eq. (4.1). When a basis function expansion for the LTI blocks is used the system equations (4.1)-(4.3) can be written as follows:

\[
y_0(t) = \sum_{i=1}^{n_S} \beta_i S_i(q) r(t), \quad (4.26)
\]
\[
r(t) = \sum_{j=1}^{n_f} \gamma_j f_j(x(t)), \quad (4.27)
\]
\[
x(t) = \sum_{k=1}^{n_H} \delta_k H_k(q) u(t), \quad (4.28)
\]

\(H_k(q)\) and \(S_i(q)\) represent the basis functions used for the front and the back LTI block. The respective number of basis functions for the front and the back LTI block is set to \(n_H\) and \(n_S\). At this point, three sets of parameters are to be estimated in the model: the weights \(\beta_i, \delta_k\) used to obtain the LTI blocks, and the coefficients \(\gamma_j\) used in the description of the static nonlinear block.

The SISO static nonlinearity of eq. (4.27) can be replaced by a MISO static nonlinearity as follows to simplify the estimation problem:

\[
y_0(t) = \sum_{i=1}^{n_S} \beta_i S_i(q) r(t), \quad (4.29)
\]
\[
r(t) = \sum_{j=1}^{n_q} \lambda_j g_j(x_1(t), \ldots, x_{n_H}(t)), \quad (4.30)
\]
\[
x_k(t) = H_k(q) u(t), \quad (4.31)
\]

where the nonlinear basis functions \(g_j(x_1(t), x_2(t), \ldots, x_{n_H}(t))\) are now MISO nonlinear basis functions (e.g. multivariate polynomials). The MISO nonlinear function includes the coefficients \(\delta_k\) used to model the front LTI block \(H(q)\) before. The estimation problem is now simplified to a problem that is bilinear in the parameters \((\beta_i, \lambda_j)\). Indeed the model output is given
4.6. Identification via Basis Function Expansion

by:

\[ y_0(t) = \sum_{i=1}^{n_S} \sum_{j=1}^{n_q} \beta_i \lambda_j S_i(q) g_j(H_1(q)u(t), \ldots, H_{n_H}(q)u(t)) . \]  

(4.32)

Only the parameters \( \beta_i \) and \( \lambda_j \) are unknown once a set of basis functions is chosen. Eq. (4.32) corresponds to the structure that is shown in Figure 4.2.

![Figure 4.2: The expanded Wiener-Hammerstein structure.](image)

**Figure 4.2:** The expanded Wiener-Hammerstein structure. \( H_k(q) \) and \( S_i(q) \) are the known LTI basis functions, \( g(x_1, \ldots, x_{n_H}) \) is the unknown MISO static nonlinearity, \( \beta_i \) are the unknown weights of the LTI basis functions at the output.

The proposed method is related to the advanced method in Sjöberg et al. [2012]. The main difference lies in the use of a basis function expansion for \( S(q) \), based on the poles of the BLA, rather than a basis function expansion for \( S(q)^{-1} \), based on the zeros of the BLA. Furthermore, only one MIMO polynomial is needed to describe the static nonlinearity instead of two MISO polynomials. When formulated like this, the proposed model structure is able to describe parallel Wiener-Hammerstein systems as well, but for now, the focus will be on single branch Wiener-Hammerstein systems.

### 4.6.2 Iterative Estimation Scheme

An iterative estimation scheme (Bai and Liu [2005]) is proposed in the literature to obtain good estimates of the \( \beta_i \) and \( \lambda_j \) parameters in eq (4.32). An overparametrization approach that is similar to the one that is proposed in Bai and Liu [2005] cannot be used here since it is observed from numerical simulations that the regressor matrix related to the overparametrized problem is not of full rank (Tiels et al. [2014]).
Once the $\beta_i$ are fixed, the estimation of the $\lambda_j$ parameters becomes a problem that is linear in the parameters. Hence, it can readily be solved using a linear least squares estimator. The estimates of the problem that is linear in the parameters. Hence, it can readily be solved

 estimates:

$$\hat{\lambda}[p] = (K_{\lambda[p]}^T K_{\lambda[p]})^{-1} K_{\lambda[p]}^T y, \quad (4.33)$$

$$K_{\lambda[p]} = \begin{bmatrix} \sum_{i=1}^{n_S} \beta_i^{[p-1]} S_i(q) g_1(x(1)) & \ldots & \sum_{i=1}^{n_S} \beta_i^{[p-1]} S_i(q) g_n(y(x(1))) \\ \sum_{i=1}^{n_S} \beta_i^{[p-1]} S_i(q) g_1(x(2)) & \ldots & \sum_{i=1}^{n_S} \beta_i^{[p-1]} S_i(q) g_n(y(x(2))) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n_S} \beta_i^{[p-1]} S_i(q) g_1(x(N)) & \ldots & \sum_{i=1}^{n_S} \beta_i^{[p-1]} S_i(q) g_n(y(x(N))) \end{bmatrix},$$

$$x(t) = [x_1(t) \ldots x_{NH}(t)],$$

$$y = [y(1) y(2) \ldots y(N)]^T,$$

$$\hat{\lambda}[p] = [\hat{\lambda}_1[p] \ldots \hat{\lambda}_{ng}[p]]^T.$$

The estimates $\beta_i^{[p]}$ are obtained based on the estimates $\hat{\lambda}_j^{[p]}$ using a different linear least squares estimator:

$$\hat{\beta}^{[p]} = (K_{\beta[p]}^T K_{\beta[p]})^{-1} K_{\beta[p]}^T y, \quad (4.34)$$

$$K_{\beta[p]} = \begin{bmatrix} \sum_{j=1}^{n_g} \hat{\lambda}_j^{[p]} S_1(q) g_j(x(1)) & \ldots & \sum_{j=1}^{n_g} \hat{\lambda}_j^{[p]} S_{ng}(q) g_j(x(1)) \\ \sum_{j=1}^{n_g} \hat{\lambda}_j^{[p]} S_1(q) g_j(x(2)) & \ldots & \sum_{j=1}^{n_g} \hat{\lambda}_j^{[p]} S_{ng}(q) g_j(x(2)) \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{n_g} \hat{\lambda}_j^{[p]} S_1(q) g_j(x(N)) & \ldots & \sum_{j=1}^{n_g} \hat{\lambda}_j^{[p]} S_{ng}(q) g_j(x(N)) \end{bmatrix},$$

$$x(t) = [x_1(t) \ldots x_{NH}(t)],$$

$$y = [y(1) y(2) \ldots y(N)]^T,$$

$$\hat{\beta}^{[p]} = [\hat{\beta}_1^{[p]} \ldots \hat{\beta}_{ng}^{[p]}]^T.$$
4.6. Identification via Basis Function Expansion

The estimates $\hat{\beta}_i^{[p]}$ are normalized such that the $l^2$-norm of $\hat{\beta}_i^{[p]}$ is equal to 1 to avoid identifiability issues. The estimates of the $\hat{\beta}_i$ (or $\hat{\lambda}_j$) parameters are initialized to some user selected initial values $\hat{\beta}_i^{[0]}$ (or $\hat{\lambda}_j^{[0]}$).

For every iteration $p$ a MISO nonlinearity needs to be estimated. This can be quite costly when a large number of basis functions $g_j$ is used. Note that the number of basis functions $g_j$ typically depends on the number of basis functions $H_k(q)$. The convergence of this method has not yet been studied. The proposed method obtains good results on the simulation example shown in Section 4.6.5.

In practice, the solution of the linear least squares problems are obtained using a QR decomposition of $K_{\beta^{[p]}}$ and $K_{\lambda^{[p]}}$. To improve the numerical conditioning of the matrix, the columns of $K_{\beta^{[p]}}$ and $K_{\lambda^{[p]}}$ are normalized. Each column is therefore divided by its $l^2$-norm.

4.6.3 Returning to a Single Branch Model

The iterative algorithm results in the estimates $\hat{\beta}_i$ and $\hat{\lambda}_j$. The $\hat{\beta}_i$ are the estimate of the LTI block that is present at the output of the system. The $\hat{\lambda}_j$ on the other hand are a combined estimate of the LTI block that is present at the input of the Wiener-Hammerstein system ($\delta_k$ parameters), and the SISO static nonlinearity of the Wiener-Hammerstein system ($\gamma_j$ parameters). This results in a MISO static nonlinearity that is preceded by a set of LTI basis functions. Here we discuss two different approaches that can be used to obtain an estimate of the front LTI block, and the SISO static nonlinearity based on the $\hat{\lambda}_j$ estimates separately. The first approach linearizes the estimated static nonlinearity and uses the gains resulting from this linearization as $\delta_k$ estimates. The second approach uses a dimension reduction approach called minimum average variance estimation (MAVE) (Xia et al. [2002]; Lyzell et al. [2012]; Schoukens et al. [2013a]) to obtain the $\delta_k$ estimates in a different way.
4. Wiener-Hammerstein

4.6.3.1 Projection through Linearization

An estimate \( \hat{r}(t) \) of the intermediate signal \( r(t) \) is obtained by the previously explained iterative estimation scheme. Here we estimate the coefficients \( \delta_k \) (see eq. (4.28)) by modeling the signal \( \hat{r}(t) \) as a weighted sum of the signals \( x_k(t) \):

\[
\hat{r}(t) \approx \sum_{k=1}^{n_H} \delta_k x_k(t).
\]  (4.35)

This corresponds with a linearization of the estimated static nonlinearity.

The coefficients \( \delta_k \) are obtained through a linear least squares estimation. The regressor matrix \( K \) contains the internal signals \( x_k(t) \) that are known in advance as the basis functions are fixed:

\[
K = \begin{bmatrix}
  x_1(1) & x_2(1) & \ldots & x_{n_H}(1) \\
  x_1(2) & x_2(2) & \ldots & x_{n_H}(2) \\
  \vdots & \vdots & \ddots & \vdots \\
  x_1(N) & x_2(N) & \ldots & x_{n_H}(N)
\end{bmatrix}
\]  (4.36)

The least squares solution for the coefficient \( \delta_k \) is given by:

\[
\hat{\delta} = (K^T K)^{-1} K^T \hat{r},
\]  (4.37)

\[
\hat{r} = [\hat{r}(1) \  \hat{r}(2) \  \ldots \  \hat{r}(N)]^T,
\]  (4.38)

\[
\hat{\delta} = [\hat{\delta}_1 \  \hat{\delta}_2 \  \ldots \  \hat{\delta}_{n_H}]^T,
\]  (4.39)

where

\[
\hat{r}(t) = \sum_{j=1}^{n_q} \hat{\lambda}_j g_j(x_1(t), \ldots, x_{n_H}(t)).
\]  (4.40)

In practice, the solution is obtained using a QR decomposition. To improve the numerical conditioning of the matrix, the columns of \( K \) are normalized. Each column is therefore divided by its \( l^2 \)-norm.

The nonlinearity is modeled as a sum of nonlinear basis functions (eq. (4.27)). The estimation of the coefficients of the expansion is a problem that is linear in the parameters once the LTI blocks are known. This is similar to the estimation of the static nonlinearity in Section 3.5. The estimation is performed using a linear least squares approach.
4.6. Identification via Basis Function Expansion

One problem here is that the linearization approach in this section does not work if the least squares estimate $\hat{\delta}$ returns a zero vector. This is the case if the BLA of the static nonlinearity given the signals $x(t)$ and $r(t)$ is equal to zero at all excited frequencies.

4.6.3.2 Projection with MAVE

The dimension reduction method that is used in this section is the minimum average variance estimation (MAVE) method (Xia et al. [2002]; Lyzell et al. [2012]). MAVE uses an iterative forward regression approach. MAVE is computationally more expensive than the linearization approach of the previous section, but rather than linearizing the static nonlinearity, a non-parametric model of the static nonlinearity is estimated.

To apply MAVE off the shelf, the problem should be reformulated as follows:

$$y(t) = f(B^T \varphi(t)) + e(t),$$

where $e(t)$ is an unknown disturbing noise source, $\varphi(t) \in \mathbb{R}^{n_d \times 1}$ contains the known regression variables, $f(\cdot)$ is a static nonlinear function, $B \in \mathbb{R}^{n_d \times n_{rd}}$, and $B^T B = I_{n_{rd}}$ is a $n_{rd} \times n_{rd}$ identity matrix. The reduced dimension $n_{rd}$ is smaller than the number of regressors $n_d$.

MAVE is a forward regression approach that estimates the matrix $B$ present in (4.41) without requiring any prior knowledge about the static nonlinearity $f(\cdot)$. It was first proposed in (Xia et al. [2002]) as a semi-parametric approach. The coefficients of the matrix $B$ are estimated, while a nonparametric model approximates the nonlinearity by its local linear expansion. Ideally, MAVE minimizes the following cost function:

$$\min_{B: B^T B = I_{n_{rd}}} E \left\{ \frac{1}{2} y(t) - E \left[ y(t) \Big| B^T \varphi(t) \right] \right\}^2,$$

To avoid the calculation of $E \left[ y(t) \Big| B^T \varphi(t) \right]$ the inner expected value is approximated by a local linear approximation around a set of points $\varphi_t$ in the support of the regression variables $\varphi(t)$:

$$E \left[ y(t) \Big| B^T \varphi(t) \right] \approx \alpha_t + \beta^T_B (\varphi(t) - \varphi_t),$$

71
where $\alpha_{\tau}$ is a scalar and $\beta_{\tau} \in \mathbb{R}^{n_{rd} \times 1}$. Here, we choose $\varphi_{\tau}$ to be the measured/simulated input of the static nonlinearity, where $\tau$ ranges from 1 to the number of samples $N$. This results in the following optimization problem:

$$
\begin{aligned}
\minimize_{\alpha_{\tau}, \beta_{\tau}, B: B^T B = I_{n_{rd}}} & \frac{1}{2} \sum_{\tau=1}^{N} \sum_{t=1}^{N} w_{t\tau} \| y(t) - \alpha_{\tau} - \beta_{\tau}^T \tilde{\varphi}_{\tau}(t) \|_2^2, \\
\end{aligned}
$$

(4.44)

$$
\tilde{\varphi}_{\tau}(t) = \varphi(t) - \varphi_{\tau},
$$

(4.45)

where $w_{t\tau}$ are nonnegative weights defining the support of the different local linear models. Here, these weights are defined by the Epanechnikov kernel (Epanechnikov [1969]):

$$
w_{t\tau} = K_h(\tilde{\varphi}_{\tau}(t)),
$$

(4.46)

$$
K(\varphi) = \frac{3}{2} \max \left(0, 1 - \| \varphi \|_2^2\right),
$$

(4.47)

$$
K_h(\varphi) = K(\varphi/h)/h,
$$

(4.48)

where $h$ sets the bandwidth of the kernel. $h$ is selected using the $k$-nearest-neighbors principle, and is updated in the next steps when an initial estimate $\hat{B}$ is obtained (Lyzell et al. [2012]).

One of the issues is that this problem is nonconvex. Lyzell et al. [2012] introduces a convex relaxation using a nuclear norm regularization to solve this minimization problem.

The problem of estimating the coefficients $\delta_k$ can be formulated as eq. (4.44). The matrix $B$ is a $n_H \times 1$ matrix in this case. The vector $\varphi(t)$ is given by:

$$
\varphi(t) = [x_1(t) \ x_2(t) \ \ldots \ x_{n_H}(t)]^T.
$$

(4.49)

The estimated matrix $B$ contains the $\delta_k$ coefficients.

Similar to the linearization approach, the nonlinearity is modeled by a sum of nonlinear basis functions (eq. (4.27)). The estimation of the static
nonlinearity as a sum of nonlinear basis functions becomes a problem that is linear in the parameters once the LTI blocks are known. This is similar to the estimation of the coefficients of the static nonlinearity in Section 3.5. The estimation is performed using a linear least squares approach.

4.6.4 Rational Form Representation

The estimated model represents the front and back LTI blocks $H(q)$ and $S(q)$ as a sum of basis functions (see eq. (4.27)). In some cases it can be beneficial to represent the LTI blocks using a rational transfer function. A rational transfer function allows for instance to easily optimize the poles of $\hat{H}(q)$ and $\hat{S}(q)$ in a nonlinear optimization step. Approximate pole-zero cancellations in the estimates of $H(q)$ and $S(q)$ can be removed when the basis function representation is transformed into a rational transfer function model to lower the model complexity (Tiels et al. [2014]). This transformation also comes at the cost of losing the orthogonality properties that can be achieved using a basis function expansion representation (Heuberger et al. [2005]).

4.6.5 Simulation Example

4.6.5.1 System and Signals

The basis function expansion Wiener-Hammerstein identification algorithm is illustrated on a simulation example. The simulated system is a Wiener-Hammerstein system. The front LTI block $H(q)$ is a third order Chebychev filter with a 3dB in-band ripple and a cut-off frequency at $0.10f_s$. The back LTI block $S(q)$ is a second order Chebychev filter with a 1dB in-band ripple and a cut-off frequency at $0.075f_s$. The static nonlinearity $f(x)$ is a
The static nonlinearity is shown in Figure 4.3.

White additive Gaussian noise with a standard deviation of 0.01 is added to the output of the system.

A random phase multisine is used as input signal. One period of the steady state response of the system to one realization is simulated. One period consists of 2048 samples. All frequencies in the range \(0, f_s/2\] are excited. The random phases are uniformly distributed between \([0, 2\pi]\]. The rms-value of the input signals is equal to 1. This results in a signal-to-noise ratio at the output of 30 dB.

### 4.6.5.2 Model Estimation

The parametric BLA of the system is estimated in a first identification step (see Section 2.3.2). The estimated BLA is shown in Figure 4.4. The
parametrized estimate of the BLA is used in the second step to obtain an estimate of the basis function expansion. Next, the parameters of the basis function expansion Wiener-Hammerstein model are estimated using an iterative algorithm. The iterative algorithm is initialized by setting the coefficients $\lambda[0]_j = 0.1$. Afterwards, the estimated basis function expansion model is projected back to a Wiener-Hammerstein model. Finally, the parameters of the Wiener-Hammerstein model are refined in an nonlinear optimization step, optimizing all the parameters simultaneously.

Note that since only one realization of a random phase multisine is used in this simulation, no nonparametric noise model can be obtained during the BLA estimation step that is used in this work, this would be possible if a more advanced method such as the local polynomial approach (Pintelon et al. [2011a,b]) is used. Although the BLA is parametrized without a correct weighting, a high quality estimate of the system dynamics is still obtained in Figure 4.4.

The iterative estimation procedure results in a basis function expansion Wiener-Hammerstein model. The rms-value of the residuals $(y(t) - \hat{y}(t))$ of this model during validation is equal to 0.0106 (see also Figure 4.6 (a)).
This corresponds to the additive noise that is present at the output of the system.

The rms-value of the residuals increase to 0.0293 after the projection of the basis function expansion model to a single branch Wiener-Hammerstein model using the projection through linearization (see also Figures 4.5 and 4.6), projection with MAVE results in a slightly lower rms-value of 0.0280. This is still well below the residual error rms-value of 0.1042 that is obtained by the BLA.

After optimization the rms-value of the residuals drops back to the noise standard deviation 0.0100 (see also Figures 4.5 and 4.6 (b)).

4.7 Identification by a Nonparametric Separation of the BLA

Based on Schoukens et al. [2014b]

The main idea of this method is to obtain an estimate of the dynamic blocks by a nonparametric decomposition of the of the BLA of the Wiener-Hammerstein system. Next, the nonparametric estimates are parametrized. Once the front and back dynamics are known, the estimation of the static nonlinearity boils down to a simple linear least squares problem. Finally, a nonlinear optimization of all the parameters taken together is performed to reduce the variability of the model estimate.

An initial nonparametric LTI estimate has several advantages. First, an initial estimate of the model can be generated with significantly less user interaction than its parametric counterpart. No model order selection is required in this step. Next, this approach solves some challenges that are linked to the assignment of the identified poles and zeros to either the front or the back LTI block. This assignment is present in Lauwers [2011]; Sjöberg et al. [2012]; Westwick and Schoukens [2012]; Tan et al. [2012]; Schoukens et al. [2014d], and mainly remains an open problem. Here, the otherwise tedious parametrization of the BLA and the so-called quadratic BLA (Schoukens et al. [2008b]; Westwick and Schoukens [2012]) are no
4.7. Identification by a Nonparametric Separation of the BLA

Figure 4.5: (a) Time domain simulation of the system output and the output of the estimated Wiener-Hammerstein model during validation. The noisy system output is depicted by the full blue line (hardly visible). The model output is represented by the dashed green line. (b) Time domain residuals $y(t) - \hat{y}(t)$ of the Wiener-Hammerstein model before (red) and after (green) optimization.
Figure 4.6: (a) Frequency domain validation of the Wiener-Hammerstein model. The noisy system output is depicted by the full blue line. The model error after the iterative estimation procedure is shown by the red pluses. (a & b) The model error after the projection but before optimization is represented by the green circles. (b) The model error after optimization is represented by the magenta triangles.
longer required. Therefore, the pole assignment problem does not occur. As opposed to Lauwers [2011]; Sjöberg et al. [2012]; Westwick and Schoukens [2012]; Tan et al. [2012]; Schoukens et al. [2014d] the current method allows overlapping pole/zero combinations in the two LTI blocks of the system to be present. Furthermore, the model order selection that is needed during the parametrization step is applied separately to each of the two LTI blocks, rather than being applied separately to the BLA and to the quadratic BLA. This problem is of lower dimension and is physically more easy to handle and understand.

A disadvantage of this approach is that a sufficiently high number of data samples is needed to obtain a good estimate of the BLA and the quadratic BLA. Furthermore, the class of input signals is restricted to the Riemann equivalence class of white Gaussian excitations for this method. Fortunately, the proposed algorithm proves to be robust with respect to the use of colored excitations.

The next section introduces the quadratic BLA of a Wiener-Hammerstein system (Section 4.7.1). The splitting of the BLA and quadratic BLA over the LTI blocks of the Wiener-Hammerstein model is discussed next (Section 4.7.2). The parametrization of the nonparametric estimates of the LTI blocks and the estimation of the static nonlinearity are discussed in Section 4.7.3 and 4.7.4. Finally, the consistency of the proposed method is proven in Section 4.7.6 and the method is illustrated on a simulation example in Section 4.7.7.

4.7.1 Quadratic BLA of a Wiener-Hammerstein system

An extension to the BLA, the so-called quadratic BLA, or QBLA, is proposed in Schoukens et al. [2008b]; Westwick and Schoukens [2012]. The QBLA of a Wiener-Hammerstein system is obtained by estimating the BLA between the squared input $u^2(t)$, and the residuals $y_s(t)$ of the modeled output $y_{BLA}(t)$ of the BLA and the true output $y(t)$ as is shown in Figure 4.7 for a Wiener-Hammerstein system.
Figure 4.7: Schematic representation of the QBLA estimation for a Wiener-Hammerstein system. The QBLA is obtained by estimating the BLA between the squared input $p(t)$ and the residuals of the BLA model $y_s(t)$.

Definition 4.1. **Quadratic best linear approximation (QBLA).** The QBLA of a nonlinear system is the linear system $G_{QBLA}(q)$ with input $u^2(t)$ and output $y_s(t)$ that minimizes the mean squared error between measurement and model (Schoukens et al. [2008b]; Westwick and Schoukens [2012]):

$$G_{QBLA}(q) = \arg\min_{G(q)} E \left\{ (y_s(t) - G(q)\tilde{p}(t))^2 \right\},$$

where the expectation is taken with respect to the random realization of $\tilde{p}(t)$. The signals $y_s(t)$ and $\tilde{p}(t)$ are defined as:

$$y_s(t) = \tilde{y}(t) - G_{BLA}(q)\tilde{u}(t),$$
$$\tilde{u}(t) = u(t) - E\{u(t)\},$$
$$\tilde{y}(t) = y(t) - E\{y(t)\},$$
$$\tilde{p}(t) = \tilde{u}^2(t) - E\{\tilde{u}^2(t)\}.$$

The BLA is used to extract the overall linear dynamics present in the system under test. The combined information obtained by the BLA and the QBLA is used to distribute the overall dynamics over the front LTI block and the back LTI block present in a Wiener-Hammerstein model. The next section describes the procedure followed to do this.

The QBLA of a Wiener-Hammerstein system obtained with an input signal belonging to the Riemann equivalence class of Gaussian signals is,
just like the BLA, a simple function of the two dynamic blocks that are present in the Wiener-Hammerstein system under test.

**Theorem 4.2.** The QBLA of a Wiener-Hammerstein system for white inputs belonging to the Riemann equivalence class of Gaussian excitations is given by:

\[
G_{QBLA}(e^{j\omega T_s}) = \alpha_{QBLA}[H \ast H](e^{j\omega T_s})S(e^{j\omega T_s}),
\]

(4.55)

where \(\alpha_{QBLA}\) depends on the static nonlinearity and the power spectrum of the input signal. The operator \([H \ast H](e^{j\omega T_s})\) denotes the circular convolution of \(H\) with itself in the frequency domain evaluated at frequency \(j\omega\).

**Proof.** See Schoukens et al. [2008b].

The convolution of \(H(e^{j\omega T_s})\) with itself appears in the expression of the QBLA (eq. (4.55)) because the output of the front LTI subsystem is transformed by the static nonlinearity that is present in the Wiener-Hammerstein system. The QBLA is equal to zero if no quadratic term is present in \(f(x)\). If this is the case, a higher order extension of the BLA can be considered, for example a so called cubic BLA, and a similar reasoning as presented in this paragraph can be applied (see Schoukens et al. [2008b]).

### 4.7.2 Splitting the BLA

To split the BLA estimate into the front and back LTI blocks a nonparametric approach is used based on a nonlinear optimization scheme. The advantage of this approach is that the user does not have to take order decisions needed to parametrize the BLA (or the QBLA) as is the case otherwise (Sjöberg et al. [2012]; Tan et al. [2012]; Westwick and Schoukens [2012]; Schoukens et al. [2014d]). The parametrization of the QBLA is a tedious job since the number of poles and zeros tends to be high. This is due to the self-convolution of the front dynamics. On top of that, the variance of the estimate of the QBLA tends to be high too. This is due to the fact that the QBLA is estimated using the residuals that are obtained from the BLA (Schoukens et al. [2008b]; Westwick and Schoukens [2012]).
Furthermore, overlapping pole/zero combinations in \( H(q) \) and \( S(q) \) can be tackled using the approach proposed here, while this is an unsolved problem in some previous approaches (Westwick and Schoukens [2012]; Tan et al. [2012]; Sjöberg et al. [2012]).

The estimates for the LTI subsystems should be able to reconstruct the nonparametric BLA and QBLA estimate. With the knowledge obtained from Sections 4.4 and 4.7.1, the following least squares cost function can be defined:

\[
L_F = \frac{1}{M} \sum_{k=1}^{F} \left[ \frac{\hat{G}_{BLA}(e^{j\omega_k T_s}) - \hat{H}(e^{j\omega_k T_s}) \hat{S}(e^{j\omega_k T_s})}{\hat{\sigma}_{\hat{G}_{BLA},t}(k)} \right]^2 + \left[ \frac{\hat{G}_{QBLA}(e^{j\omega_k T_s}) - \left[ \hat{H} \ast \hat{H} \right](e^{j\omega_k T_s}) \hat{S}(e^{j\omega_k T_s})}{\hat{\sigma}_{\hat{G}_{QBLA},t}(k)} \right]^2.
\]  

(4.56)

Here, \( M \) is the total number of realizations used for measuring the BLA \( \hat{G}_{BLA}(e^{j\omega_k T_s}) \) and the QBLA \( \hat{G}_{QBLA}(e^{j\omega_k T_s}) \), and \( F \) is the number of excited frequency lines. The corresponding sample variances are labeled \( \hat{\sigma}_{\hat{G}_{BLA},t}(k) \) and \( \hat{\sigma}_{\hat{G}_{QBLA},t}(k) \). They serve as a weighting function to ensure that both terms are normalized over the frequencies with respect to the noise, and with respect to each other.

\( \hat{H}(e^{j\omega_k T_s}) \) and \( \hat{S}(e^{j\omega_k T_s}) \) are obtained as the minimizer of the cost function (4.56). A Levenberg-Marquardt nonlinear optimization algorithm (Fletcher [1991]; Pintelon and Schoukens [2012]) is needed to solve the problem that is nonlinear in the parameters.

The \( \hat{H}(e^{j\omega_k T_s}) \) and \( \hat{S}(e^{j\omega_k T_s}) \) estimates are initialized with the help of the BLA and QBLA estimates, as is proposed in Schoukens et al. [2008b]. The convolution of \( H \) with itself in the frequency domain smoothens the shape of the FRF. The QBLA will therefore provide a decent initial value for \( \hat{S} \). Since the BLA of a Wiener-Hammerstein system is given by eq. (4.11), the initial value of \( \hat{H} \) is chosen as in eq. (4.58):

\[
\hat{S}_{\text{init}}(e^{j\omega_k T_s}) = \hat{G}_{QBLA}(e^{j\omega_k T_s}), \quad (4.57)
\]

\[
\hat{H}_{\text{init}}(e^{j\omega_k T_s}) = \frac{\hat{G}_{BLA}(e^{j\omega_k T_s})}{\hat{G}_{QBLA}(e^{j\omega_k T_s})}. \quad (4.58)
\]
4.7. Identification by a Nonparametric Separation of the BLA

A lower bound on the covariance matrix $\hat{\Sigma}_{\hat{H}, \hat{S}}$ of these estimates can be found by taking the inverse of the Fisher information matrix:

$$\Sigma_{\hat{H}, \hat{S}} = \left[ J_{\hat{H}, \hat{S}}^H J_{\hat{H}, \hat{S}} \right]^{-1},$$

(4.59)

where $J^H$ denotes the Hermitian operator, and with $J_{\hat{H}, \hat{S}}$ the derivatives of the weighted residuals in cost function (4.56) with respect to $\hat{H}$ and $\hat{S}$.

4.7.3 Parametrization of the LTI Blocks

For many applications a parametric model for the linear subsystems $H$ and $S$ is desired. Hence, the parametrized estimates $\hat{H}(q, \hat{\theta}_H)$ and $\hat{S}(q, \hat{\theta}_S)$ are sought for starting from the nonparametric estimates $\hat{H}(e^{j\omega T_s})$ and $\hat{S}(e^{j\omega T_s})$.

A problem for the straightforward parametrization of the nonparametric FRF in this case is the presence of a delay exchange between the front LTI block and the back LTI block. Such a delay exchange does not affect the input-output behavior of the model as is shown in Section 4.3, but it can make the parametrization more tedious. This problem can be solved by parametrizing the two LTI subsystems simultaneously, taking into account the delay ambiguity problem. The estimates are then obtained by:

$$\begin{bmatrix} \hat{\theta}_H \\ \hat{\theta}_S \\ \hat{\tau} \end{bmatrix} = \arg\min_{\theta_H, \theta_S, \tau} \sum_{k=1}^{F} \left[ \frac{H(e^{j\omega k T_s}) e^{j2\pi \frac{k}{N} \tau} - \hat{H}(q, \theta_H)}{\hat{\sigma}_H(k)} \right]^2 + \left[ \frac{\hat{S}(e^{j\omega k T_s}) e^{-j2\pi \frac{k}{N} \tau} - \hat{S}(q, \theta_S)}{\hat{\sigma}_S(k)} \right]^2,$$

(4.60)

where $\hat{\sigma}_H(k)$ and $\hat{\sigma}_S(k)$ are calculated from the covariance matrix $\hat{\Sigma}_{\hat{H}, \hat{S}}$ obtained in the previous section (eq. (4.59)). This estimate can be initialized, for instance, by the ELiS (Estimator for Linear Systems) toolbox (Kollar [2004-2009]; Pintelon and Schoukens [1990, 2012]) and further refined using a nonlinear optimization of the combined parameters of $H$, $S$, and the time delay $\tau$. 
4. Wiener-Hammerstein

Note that the obtained nonparametric FRF models of the front and back LTI block do not need to be parametrized at all. The user can use the nonparametric FRF as a model for the LTI block. The parametrization step proves to be helpful to obtain a better initial estimate of the LTI models and a faster nonlinear optimization when all the model parameters are taken together.

4.7.4 Estimating the static nonlinearity

The nonlinearity is modeled by a linear combination of nonlinear basis functions:

\[ f(x(t)) = \sum_{i=1}^{n_f} \gamma_i f_i(x(t)). \]  \hfill (4.61)

The estimation of the coefficients \( \gamma_i \) is a problem that is linear in the parameters once the LTI blocks are known. Similarly to the estimation of the static nonlinearity in Section 3.5, the estimation is performed using a linear least squares estimator.

4.7.5 Persistence of Excitation

**Assumption 4.2. Persistence of excitation:** The system is persistently excited by the signal \( u(t) \).

The assumption that the excitation is persistent is a very common assumption in system identification. This section discusses what persistence of excitation means for the proposed identification procedure.

The first step in the identification algorithm is to identify the nonparametric BLA and QBLA of the nonlinear Wiener-Hammerstein system, this step requires a white Gaussian input signal. In a later step, the dynamics of each LTI-block are parametrized together with a delay estimation. Therefore, the number of frequencies in the nonparametric FRFs needs be at least equal or preferably higher than \( \frac{n_b+n_a+2}{2} \) for the front LTI block and \( \frac{n_d+n_c+2}{2} \) for the back LTI block.
The static nonlinearity also needs to be estimated. For this identification step to work the nonlinear basis functions $f_i(\hat{x}(t))$ need to be linearly independent over the domain of the input signal $\hat{x}(t)$. Consequently, the amplitude range that is covered by $\hat{x}(t)$ needs to be sufficiently large.

### 4.7.6 Consistency

To model the system exactly, the system needs to be contained in the reachable model set. Otherwise model errors will be present and the estimated parameters will not converge to the true parameters of the system. This requirement is formalized in the following assumption.

**Assumption 4.3. Model set:** The exact model of the system is contained in the reachable model set.

**Theorem 4.3.** The nonparametric method proposed in Sections 4.7.2 and 4.7.4 is a consistent estimator when the number of samples $N$ and the number of realizations $M$ are both tending to infinity ($N, M \to \infty$) under the following assumptions:

- the common assumption of zero-mean additive noise at the output only (Assumption 2.2),
- the excitation signal $u(t)$ is assumed to belong to the Riemann equivalence class of asymptotically normally distributed signals (Assumption 4.1),
- the persistence of excitation is assumed (Assumption 4.2),
- the system is contained in the reachable model set (Assumption 4.3) for the estimated parameters to converge to the true parameters of the system up to the degenerations that are present in the model representation (see Section 4.3).

**Proof.** It is shown in Schoukens et al. [2008b]; Pintelon and Schoukens [2012] that the BLA and QBLA are consistent estimates of equations (4.11) and (4.55) respectively if white Gaussian input signals are used (Assumption
4. **Wiener-Hammerstein**

4.1), when the number of samples \( N \) and the number of realizations \( M \) are both tending to infinity. Hence, the expected value of the different terms of the cost function defined in eq. (4.56) is minimal in a time delayed and scaled version of the true values of \( H(e^{j\omega T_s}) \) and \( S(e^{j\omega T_s}) \), as explained in Section 4.3. These estimates are referred to as \( \tilde{H}(e^{j\omega T_s}) \) and \( \tilde{S}(e^{j\omega T_s}) \) as in eq. (4.10) of this proof.

The expected value of the first term of the cost function defined in eq. (4.56) is given by:

\[
E \left\{ \frac{1}{M} \sum_{k=1}^{F} \left| \hat{G}_{BLA}(e^{j\omega T_s}) - \tilde{H}(e^{j\omega T_s})\tilde{S}(e^{j\omega T_s}) \right|^2 \right\}, \tag{4.62}
\]

where \( F = O(N) \) is the total number of frequencies, and the expectation is taken with respect to the noise present in the BLA estimate. Define

\[
\hat{G}_{BLA}(e^{j\omega T_s}) = G_{BLA}(e^{j\omega T_s}) + n_{BLA}(e^{j\omega T_s}),
\]

\[
\Delta_{HS}(e^{j\omega T_s}) = G_{BLA}(e^{j\omega T_s}) - \tilde{H}(e^{j\omega T_s})\tilde{S}(e^{j\omega T_s}),
\]

where \( G_{BLA}(e^{j\omega T_s}) \) for a Wiener-Hammerstein system is defined in eq. (4.11). Then eq. (4.62) for \( N \) growing to infinity can be written as:

\[
\lim_{N \to \infty} \sum_{k=1}^{F} \left( \frac{\left| \Delta_{HS}(e^{j\omega T_s}) \right|^2}{M\hat{\sigma}_{G_{BLA},t}^2(k)} + E \left\{ \frac{n_{BLA}(e^{j\omega T_s})^2}{M\hat{\sigma}_{G_{BLA},t}^2(k)} \right\} \right) + 2\text{real} \left( E \left\{ \frac{\text{real} \left[ \Delta_{HS}(e^{j\omega T_s}) n_{BLA}(e^{j\omega T_s}) \right]}{M\hat{\sigma}_{G_{BLA},t}^2(k)} \right\} \right) = \lim_{N \to \infty} \sum_{k=1}^{F} \left( \frac{\left| \Delta_{HS}(e^{j\omega T_s}) \right|^2}{M\hat{\sigma}_{G_{BLA},t}^2(k)} + \frac{\sigma_{G_{BLA},t}^2(k)}{M\hat{\sigma}_{G_{BLA},t}^2(k)} + 2\text{real} \left( E \left\{ \frac{\text{real} \left[ \Delta_{HS}(e^{j\omega T_s}) n_{BLA}(e^{j\omega T_s}) \right]}{M\hat{\sigma}_{G_{BLA},t}^2(k)} \right\} \right) \right) \tag{4.63}
\]

The limit for \( N \) growing to infinity is taken such that \( n_{BLA} \) and \( \hat{\sigma}_{G_{BLA},t}^2 \) are independently distributed (Pintelon and Schoukens [2012]). When the number of realizations \( M \) and the number of data points \( N \) both tend to infinity \( \hat{G}_{BLA}(e^{j\omega T_s}) \) tends to \( G_{BLA}(e^{j\omega T_s}) \). The estimate of the variance
4.7. Identification by a Nonparametric Separation of the BLA

\[ M \hat{\sigma}^2_{BLA,t} \] also converges to the true variance \( M \sigma^2_{BLA,t} \):

\[
\lim_{M \to \infty} M \left( \hat{\sigma}^2_{BLA,t} - \sigma^2_{BLA,t} \right) = 0,
\]

and the noise \( n_{BLA} \) on \( \hat{G}_{BLA} \) becomes independent of \( \hat{\sigma}_{BLA} \):

\[
\lim_{M,N \to \infty} \left( \sum_{k=1}^{F} \left| \Delta_{HS} \left( e^{j\omega k T_s} \right) \right|^2 M \hat{\sigma}^2_{BLA,t}(k) + \frac{\sigma^2_{BLA,t}(k)}{M \hat{\sigma}^2_{BLA,t}(k)} \right)
\]

\[
+ 2 \text{real} \left( E \left\{ \text{real} \left[ \Delta_{HS} \left( e^{j\omega k T_s} \right) n_{BLA} \left( e^{j\omega k T_s} \right) \right] \right\} \right) \]

\[
\lim_{M,N \to \infty} \left( \sum_{k=1}^{F} \frac{\left| \Delta_{HS} \left( e^{j\omega k T_s} \right) \right|^2}{M \sigma^2_{BLA,t}(k)} \right).
\]

The third term in the right hand side of eq. (4.65) is zero since the noise \( n_{BLA} \) is zero-mean and independent of \( \Delta_{HS} \). The second term is independent of the estimates \( \hat{H} \) and hence \( \hat{S} \). Eq. (4.66) is zero in \( \bar{H}(e^{j\omega k T_s}) \) and \( \bar{S}(e^{j\omega k T_s}) \) (as defined in eq. (4.10)). A similar reasoning can be performed for the second term of the cost function eq. (4.56). This shows the consistency of the complete parameter set.

The estimate of the static nonlinear block is consistent under Assumptions 2.2, 4.2, and 4.3 when \( M \) and \( N \) both tend to infinity since the nonparametric estimates of the two LTI blocks are consistent, and since the estimation of the static nonlinearity is performed using a linear least squares estimation with zero-mean additive noise on the output only.

This shows that the proposed nonparametric identification method for Wiener-Hammerstein systems is consistent. The parametrization of the LTI blocks and the optimization of all the parameters together is not required to be consistent, but it will in most practical cases increase the efficiency of the estimated model. When a parametrization step of the BLA and QBLA is introduced before splitting the BLA, only the number of data points \( N \) needs to grow to infinity to prove consistency.

**Theorem 4.4.** The method proposed in Sections 4.7.2 and 4.7.4 is a consistent estimator when the number of samples \( N \) tends to infinity \((N \to \infty)\)
when parametrized estimates $\hat{G}_{\text{BLA}}(q, \hat{\theta}_{\text{BLA}})$ and $\hat{G}_{\text{QBLA}}(q, \hat{\theta}_{\text{QBLA}})$ of the BLA and QBLA respectively are used under the following assumptions:

- the common assumption of zero-mean additive noise at the output only (Assumption 2.2),
- the excitation signal $u(t)$ is assumed to belong to the Riemann equivalence class of asymptotically normally distributed signals (Assumption 4.1),
- the persistence of excitation is assumed (Assumption 4.2),
- the system is contained in the reachable model set (Assumption 4.3) for the estimated parameters to converge to the true parameters of the system up to the degenerations that are present in the model representation (see Section 4.3).

Proof. The parametric estimates $\hat{G}_{\text{BLA}}(q, \hat{\theta}_{\text{BLA}})$ and $\hat{G}_{\text{QBLA}}(q, \hat{\theta}_{\text{QBLA}})$ are consistent estimates of $G_{\text{BLA}}(e^{j\omega T_s})$ and $G_{\text{QBLA}}(e^{j\omega T_s})$ respectively, when the number of data points $N$, and hence the number of frequencies, grow to infinity (Pintelon and Schoukens [2012]). The consistency of the method in Sections 4.7.2 and 4.7.4 can be proven in a similar fashion as presented in the proof of Theorem 4.3.

Remark: the consistency results can be extended to use higher order versions of the BLA than the QBLA, for instance a cubic BLA.

4.7.7 Simulation Example

4.7.7.1 System and Signals

The Wiener-Hammerstein identification by a nonparametric separation of the BLA is illustrated on a simulation example. The simulated system is a Wiener-Hammerstein system. The front LTI block $H(q)$ is a third order Chebychev filter with a 5dB in-band ripple and a cut-off frequency at 0.10$f_s$. The back LTI block $S(q)$ is a third order Chebychev type 2 filter with a
30dB attenuation at $0.16f_s$. The static nonlinearity $f(x)$ is a polynomial of degree 3:

$$H(q) = \frac{0.0358 + 0.1075q^{-1} + 0.1075q^{-2} + 0.0358q^{-3}}{1 - 1.4676q^{-1} + 1.3414q^{-2} - 0.5871q^{-3}}, \quad (4.67)$$

$$S(q) = \frac{0.0430 + 0.0064q^{-1} + 0.0064q^{-2} + 0.0430q^{-3}}{1 - 1.9586q^{-1} + 1.4014q^{-2} - 0.3439q^{-3}}, \quad (4.68)$$

$$f(x) = x + 0.2x^2 - 0.05x^3. \quad (4.69)$$

The static nonlinearity is shown in Figure 4.8.

White additive Gaussian noise with a standard deviation of 0.001 is added to the output of the system.

A random phase multisine is used as an input signal. Two periods of the steady state response to 100 realizations are simulated. One period consists of 512 samples. All frequencies in the range $]0, f_s/2[$ are excited. The random phases are uniformly distributed between $[0, 2\pi]$. The rms-value of the input signals is equal to 1. This results in a signal-to-noise ratio at the output of 50 dB.
4.7.7.2 Model Estimation

The nonparametric BLA and QBLA of the system are estimated in a first identification step (see Sections 2.3.2 and 4.7.1). The estimated BLA and QBLA are shown in Figure 4.9. These nonparametric estimates are used to obtain a nonparametric estimate of the LTI blocks present in the Wiener-Hammerstein system (Figure 4.10). These LTI block estimates are parametrized next, and the static nonlinearity is estimated. Finally, the parameters of the Wiener-Hammerstein model are refined in an nonlinear optimization step, optimizing all the parameters simultaneously. Figure 4.10 shows that the nonparametric LTI block estimate provides already a good estimate of the LTI blocks when the are compared to the final optimized parametric estimates.

The estimated model is validated using a different realization of the random phase multisine. This realization has not been used during the estimation. The time domain representation of the measured validation output and the residual error ($\hat{y}(t) - y(t)$) are shown in Figure 4.11 and Figure 4.12. The frequency domain representation of the validation output and the magnitude of the complex error between the measurement and the modeled output are shown in Figure 4.13.

The initial model estimate results in residual errors which are well below the output spectrum (about 20 dB). The residual errors are almost coinciding with the total distortion level. After optimization the residual error coincides with the estimated noise distortion level as is shown in Figure 4.13. This means that the nonlinearity is almost completely modeled.

4.8 Identification using Phase Coupled Multisines

Based on Schoukens et al. [2014a] and Tiels et al. [2015]

This method uses a specific class of input signals, the class of phase coupled multisines, to obtain an estimate of the front and the back LTI blocks.
4.8. Identification using Phase Coupled Multisines

**Figure 4.9:** The estimate of the BLA (red) and QBLA (blue) of the simulated Wiener-Hammerstein system. The total variance estimate obtained with the BLA and the QBLA is shown by the squares (blue and red). The noise variance estimate obtained with the BLA and QBLA is depicted by the dots (blue and red).

**Figure 4.10:** The nonparametric (dashed line) and optimized parametric (full line) estimate of the LTI blocks present in the Wiener-Hammerstein system. The front LTI block estimate is shown in red, the back LTI block estimate is shown in blue.
4. Wiener-Hammerstein

**Figure 4.11:** Time domain simulation of the system output and the output of the estimated Wiener-Hammerstein model after optimization. The noisy system output is depicted by the full blue line. The model output is represented by the dashed green line. Note that there is hardly a noticeable difference.

**Figure 4.12:** Time domain residuals $y(t) - \hat{y}(t)$ before (red) and after (green) optimization of the parametrized Wiener-Hammerstein model.
4.8. Identification using Phase Coupled Multisines

Figure 4.13: Spectrum of the measured and modeled output of the estimated Wiener-Hammerstein system. The noisy system output is depicted by the full blue line. The model error before optimization is represented by the red pluses. The model error after optimization is represented by the green circles. The variance of the total distortion that is present at the output is shown by the black squares, the noise distortion is represented by the black stars.

of a Wiener-Hammerstein system. The class of phase coupled multisine signals is introduced in the next section. Next, it is shown how to extract an estimate of the LTI blocks of the Wiener-Hammerstein system under test starting from the BLA that is obtained with phase coupled multisines. Finally, a simulation example shows the good performance of the proposed method.

4.8.1 BLA with Phase Coupled Multisines

The phases of a standard multisine, as described in Section 2.2.1, are independent over the frequency. A phase coupled random phase multisine introduces a coupling between the phases of spectral lines at different frequencies.

Definition 4.2. Phase Coupled Random Phase Multisine. A signal $u(t)$ is a phase coupled multisine if it is a multisine that only excites the
harmonics $k$ for which

$$\pm k \in \left\{ \frac{d}{2} + di, \frac{d}{2} + di + s \right\}, \quad i = 0, 1, \ldots, i_{\text{max}}. \quad (4.70)$$

The excited harmonics $\left\{ \frac{d}{2} + di, \frac{d}{2} + di + s \right\}$ are called a frequency couple. Each frequency couple gets an independently and identically distributed random phase $\phi_{\frac{d}{2} + di} = \phi_{\frac{d}{2} + di + s}$ with the property $E\left\{ e^{j\phi_{\frac{d}{2} + di}} \right\} = 0$ assigned. The even integer $d \geq 4$ determines the frequency resolution, and $s = k_s d + 1$, $k_s \in \mathbb{N}_1$ determines the shift.

For simplicity of notation we assume here that the static nonlinearity of the system is a monomial of third degree $f(x) = x^3$. The extraction of the BLA of a Wiener-Hammerstein system using coupled random phase multisine input signals is introduced in Tiels et al. [2015] for systems with more general polynomial nonlinearities. The method is based on the following observation. The systematic output contributions that appear in the BLA at a frequency $k$ need to have a phase that is equal to the input phase $\phi_k$ plus a constant term that depends on the LTI blocks of the system. The BLA hence contains these systematic contributions. All the other contributions in the noiseless output are the so-called stochastic nonlinear distortions (Pintelon and Schoukens [2012]). The BLA contributions in the output $Y_{bla}$ can be described by four different cases: consider $m, l \in \frac{d}{2} + di, i = 0, 1, \ldots, i_{\text{max}}$:

1. At frequency $j\omega_m$:

$$Y_{bla}(m) = H(e^{j\omega_m T_s}) S(e^{j\omega_m T_s}) U(m) \sum_l |H(e^{j\omega_l T_s})|^2 |U(l)|^2$$

$$+ H(e^{j\omega_m + s T_s}) S(e^{j\omega_m T_s}) U(m)$$

$$\sum_l H(e^{j\omega_l T_s}) \bar{H}(e^{j\omega_l + s T_s}) U(l) \bar{U}(l + s)$$

$$+ O(N^{-1}). \quad (4.71)$$
4.8. Identification using Phase Coupled Multisines

2. At frequency $j\omega_{m+s}$:

\[
Y_{bla}(m + s) = H(e^{j\omega_{m+s}T_s})S(e^{j\omega_{m+s}T_s})U(m + s) \sum_l |H(e^{j\omega_l T_s})|^2 |U(l)|^2 \\
+ H(e^{j\omega_{m}T_s})S(e^{j\omega_{m}T_s})U(m) \\
\sum_l H(e^{j\omega_{l+s}T_s})\bar{H}(e^{j\omega_l T_s})U(l + s)\bar{U}(l) \\
+ O(N^{-1}).
\]  

(4.72)

3. At frequency $j\omega_{m-s}$:

\[
Y_{bla}(m - s) = H(e^{j\omega_{m}T_s})S(e^{j\omega_{m-s}T_s})U(m) \\
\sum_l H(e^{j\omega_l T_s})\bar{H}(e^{j\omega_{l+s}T_s})U(l)\bar{U}(l + s) \\
+ O(N^{-1}).
\]  

(4.73)

4. At frequency $j\omega_{m+2s}$:

\[
Y_{bla}(m + 2s) = H(e^{j\omega_{m+2s}T_s})S(e^{j\omega_{m+2s}T_s})U(m + s) \\
\sum_l H(e^{j\omega_{l+s}T_s})\bar{H}(e^{j\omega_l T_s})U(l + s)\bar{U}(l) \\
+ O(N^{-1}).
\]  

(4.74)

The sum over $l$ runs over all harmonics $l \in \frac{d}{2} + di$, $i = 0, 1, \ldots, i_{\text{max}}$.

The contributions $Y_{bla}(m - s)$ and $Y_{bla}(m + 2s)$ are of particular interest. Consider the contribution $Y_{bla}(m - s)$. The reasoning for the contribution $Y_{bla}(m + 2s)$ is similar. The summation over $l$ can be replaced by a complex constant $\alpha$ due to the particular construction of the phase relation described above:

\[
Y_{bla}(m - s) = \alpha H(e^{j\omega_{m}T_s})S(e^{j\omega_{m-s}T_s})U(m) + O(N^{-1}).
\]

(4.75)

Note that the system is not excited at the frequencies $j\omega_{m-s}$. Therefore the BLA at the frequencies $j\omega_{m-s}$ is defined as:

\[
G_{bla}(e^{j\omega_{m-s}T_s}) = E\left\{ \frac{Y(m - s)}{U(m)} \right\},
\]

(4.76)

\[
= \alpha H(e^{j\omega_{m}T_s})S(e^{j\omega_{m-s}T_s}) + O(N^{-1}).
\]

(4.77)
The BLA at the frequencies $j\omega_{m+2s}$ is given by:

$$G_{bla}(e^{j\omega_{m+2s}T_s}) = E\left\{ \frac{Y(m + 2s)}{U(m + s)} \right\},$$

$$= \alpha H(e^{j\omega_{m+Ts}})S(e^{j\omega_{m+2s}T_s}) + O(N^{-1}).$$

The expected value is taken with respect to the random phase realization of the input $U$.

Two new FRFs can be derived from $G_{bla}(e^{j\omega_{m-2s}T_s})$ and $G_{bla}(e^{j\omega_{m+2s}T_s})$:

$$G_R(e^{j\omega_{m-2s}T_s}) = G_{bla}(e^{j\omega_{m-2s}T_s}),$$

$$G_R(e^{-j\omega_{m-2s}T_s}) = \bar{G}_{bla}(e^{j\omega_{m+2s}T_s}),$$

and

$$G_L(e^{-j\omega_{m+2s}T_s}) = \bar{G}_{bla}(e^{j\omega_{m-2s}T_s}),$$

$$G_L(e^{j\omega_{m+2s}T_s}) = G_{bla}(e^{j\omega_{m+2s}T_s}).$$

This results in the following expressions of $G_R$ and $G_L$:

$$G_R(e^{j\omega_kT_s}) = \alpha H(e^{j\omega_k+Ts})S(e^{j\omega_kT_s}) + O(N^{-1}),$$

$$G_L(e^{j\omega_kT_s}) = \alpha H(e^{j\omega_k-Ts})S(e^{j\omega_kT_s}) + O(N^{-1}).$$

Notice that the $S$ dynamics are shifted with $-s$ with respect to the $H$ dynamics in $G_L$, while the $S$ dynamics are shifted with $+s$ with respect to the $H$ dynamics in $G_R$. This will be used during the estimation to separate the front and the back dynamics that are present in the Wiener-Hammerstein system.

### 4.8.2 Estimation

The FRFs $G_R$ and $G_L$ are, due to the shift of the $S$ dynamics, not complex conjugate anymore at the positive and negative frequencies. Hence, the numerator and denominator coefficients of the parametrized rational transfer function model are complex coefficients. The parameters are estimated
using a weighted least squares estimator estimator (Peeters et al. [2001]; Pintelon and Schoukens [2012]; Schoukens et al. [2014a]):

\[ V_R(\theta) = \frac{1}{F} \sum_{k=k_{\text{min}}}^{k_{\text{max}}} \frac{|G_R(e^{j\omega_k T_s}) - G_R(q, \theta)|^2}{\hat{\sigma}_{G,R,k}^2} \], \hspace{1cm} (4.86)

where \( F \) is the number of excited frequency lines, \( k_{\text{min}} \) is the lowest excited frequency and \( k_{\text{max}} \) is the highest excited frequency. \( \hat{\sigma}_{G,R,k}^2 \) is the variance of the total distortion of \( G \) at frequency \( j\omega_k \). A similar cost function can be defined for \( G_L \).

The separation of the dynamics of the BLA into the front dynamic block \( H(q) \) and back dynamic block \( S(q) \) is obtained by comparing the estimated poles and zeros of \( G_R(q, \theta) \) and \( G_L(q, \theta) \). Some poles (or zeros) of \( G_R \) and \( G_L \) will coincide, these are the poles (or zeros) of the non-shifted system dynamics present in \( S(q) \), while other poles will be shifted with respect to each other, these are the poles (or zeros) present in the shifted system dynamics \( H(q) \). The poles and zeros of \( G_R \) and \( G_L \) can be assigned to either \( H(q) \) or \( S(q) \) with this insight.

The nonlinearity is modeled by a sum of nonlinear basis functions:

\[ f(x(t)) = \sum_{i=1}^{n_f} \gamma_i f_i(x(t)). \] \hspace{1cm} (4.87)

This estimation is a problem that is linear in the parameters once the LTI blocks are known, similarly to the estimation of the static nonlinearity in Section 3.5. The estimation is performed using a linear least squares approach.

### 4.8.3 Simulation Example

This simulation example illustrates the shift of the poles and the zeros when the BLA of a Wiener-Hammerstein system is obtained using phase coupled multisine input signals.

#### 4.8.3.1 System and Signals

The simulated system is a Wiener-Hammerstein system. The front LTI block \( H(q) \) is a third order Chebychev type 2 filter with a 30dB attenuation
at $0.10 f_s$. The back LTI block $S(q)$ is a second order Chebychev filter with a $1\text{dB}$ in-band ripple and a cut-off frequency at $0.075 f_s$. The static nonlinearity $f(x)$ is a polynomial of degree 3:

$$H(q) = \frac{0.0257 - 0.0130q^{-1} - 0.0130q^{-2} + 0.0257q^{-3}}{1 - 2.3728q^{-1} + 1.9291q^{-2} - 0.5309q^{-3}},$$  \hspace{1cm} (4.88)$$

$$S(q) = \frac{0.0040 + 0.0080q^{-1} + 0.0080q^{-2}}{1 - 1.8795q^{-1} + 0.9298q^{-2}},$$  \hspace{1cm} (4.89)$$

$$f(x) = x + 0.2x^3.$$  \hspace{1cm} (4.90)$$

The static nonlinearity is shown in Figure 4.14.

White additive Gaussian noise with a standard deviation of 0.001 is added to the output of the system.

A phase coupled multisine is used as an input signal. One period of the steady state response to 100 realizations is simulated. One period consists of 4096 samples. The variables in eq. (4.70) are set to: $d = 4$, $s = 65$ and $i_{max} = 255$. This results in an approximate excited frequency range between 0 and $f_s/4$. The random phases are uniformly distributed between $[0, 2\pi]$. The rms-value of the input signals is equal to 1. This results in a signal-to-noise ratio at the output of approximately 50 dB. The sampling frequency is set to 4096 Hz.
4.8.3.2 Model Estimation

The BLA of the system is estimated in a first identification step using the robust method (see Section 2.3.1). The FRFs $G_L$ and $G_R$ are extracted and parametrized next. The parametrized $G_L$ and $G_R$ estimates are shown in Figure 4.15.

The poles and zeros of the shifted FRFs are shown in Figures 4.16 and 4.17. A clear shift can be observed for the poles and zeros of the front LTI block $H(q)$, while the poles and zeros of the back LTI block $S(q)$ are coinciding. A clear pole and zero allocation can be made based on these figures. The system is only excited up to $f_s/4$, the zero estimates lying beyond that frequency are of low quality. No conclusions can be made based on their location.

4.9 Nonlinear Optimization

Joining all the previous estimation steps for the identification of a Wiener-Hammerstein system allows one to obtain the model parameter vector as
Figure 4.16: Poles of the parametric estimate of the shifted FRFs. The true poles of $H(q)$ are shown in red, the true poles of $S(q)$ are shown in magenta, the poles of the left shifted FRF are shown in green, and the poles of the right shifted FRF are shown in blue.

Figure 4.17: Zeros of the parametric estimate of the shifted FRFs. The true zeros of $H(q)$ are shown in red, the true zeros of $S(q)$ are shown in magenta, the zeros of the left shifted FRF are shown in green, and the zeros of the right shifted FRF are shown in blue. Some zeros are outside the figure borders, they are indicated by the arrows.
the estimation of a partition of the parameter vector $\theta$ in non-overlapping subsets as is explained in Section 3.5.6. Note that the parameter vector $\theta \in \mathbb{R}^{(n_b+n_a+n_f+n_d+n_c+4)\times 1}$ now contains all the parameters of the model: $\gamma$, $b$, $a$, $d$, and $c$. The Jacobian of the cost function is rank deficient due to the identifiability issues (see Section 4.3). The related degenerations (2 inside the LTI blocks and 2 between the LTI blocks and the static nonlinearity) in the Jacobian need to be taken into account during the optimization.

Good initial estimates for the nonlinear optimization can be obtained using the algorithms that are explained in Sections 4.5, 4.6, 4.7 and 4.8.

4.10 Conclusion and Discussion

This chapter introduces four novel identification methods for Wiener-Hammerstein systems. The first approach uses genetic algorithms to obtain the estimate of the two LTI blocks that are present in a Wiener-Hammerstein model. The second method uses a basis function expansion of the BLA to obtain the model of the two LTI blocks. The third approach separates the BLA in a front and back LTI contribution in a nonparametric way to obtain an estimate of the front and back dynamics. The last approach uses a special class of specifically tailored input signals (phase coupled multisine) to extract the front and back dynamics of the Wiener-Hammerstein system directly.

Each method has its strengths and drawbacks. The nonparametric approach and the phase coupled multisine method require more measurements to obtain good results. The parametric approaches can suffer from pole-zero cancellations in the $H(q)S(q)$ combination.

The good performance of the proposed methods has been illustrated on simulation examples.
Chapter 5

Hammerstein-Wiener

Based on Schoukens et al. [2012b]

5.1 Introduction

Hammerstein and Wiener models are discussed in the previous chapters. A more general block structure is the Hammerstein-Wiener model (static nonlinear block - LTI block - static nonlinear block), as shown in Figure 5.1. Hammerstein-Wiener models are used in a wide range of applications such as chemical processes (Giri and Bai [2010]), ionospheric dynamics (Palanthandalam-Madapusi et al. [2005]), submarine detection (Abrahamsson et al. [2007]), and RF (radio frequency) power amplifier modeling (Taringou et al. [2010]). Different identification approaches are proposed to identify a Hammerstein-Wiener model or models that are very similar to Hammerstein-Wiener block structures. Amongst those are: iterative approaches (Zhu [2002]; Voros [2004]), overparametrization methods (Bai [1998]), frequency domain methods (Crama and Schoukens [2004]; Brouri et al. [2014a]), subspace methods (Goethals et al. [2005]), stochastic algorithms (Wang and Ding [2008]), blind approaches (Bai [2002]), and maximum likelihood based methods (Wills et al. [2013]). The identification of Hammerstein-Wiener systems with a backlash input nonlinearity is discussed in Brouri et al. [2014b].
This chapter introduces a new formulation of the Hammerstein-Wiener model estimation as is presented in Schoukens et al. [2012b]. The results that are obtained in this chapter are less general than the more recent results that are reported in Wills et al. [2013]. However, the methods that are described in this chapter can be used to initialize the nonlinear optimization that is used in Wills et al. [2013]. This might be of particular interest when the noise is dominantly present in the middle of the Hammerstein-Wiener system.

Firstly, the identification setup is described. Secondly, a new formulation of the Hammerstein-Wiener identification problem is introduced and a comparison with a popular approach described in (Bai [1998]) is made. Next, two identification methods are proposed and implemented: an over-parametrization method and an iterative approach. Finally, the performance of the different methods is shown on a simulation example.

5.2 The Hammerstein-Wiener System Class

A Hammerstein-Wiener system is a series connection of a static nonlinearity followed by an LTI block, and another static nonlinear block as is shown in Figure 5.1. The noiseless output of a Hammerstein-Wiener system is given by:

\[ y_0(t) = g(r_0(t)), \]
\[ r_0(t) = H(q)x(t), \]
\[ x(t) = f(u(t)), \]

where \( f(.) \) and \( g(.) \) are static nonlinear functions, \( H(q) \) is a stable LTI system, and the signals \( y_0(t), r_0(t), x(t) \) and \( u(t) \) are scalar signals. The outputs of \( f(.) \) and \( g(.) \) are assumed to be bounded for bounded inputs.

The input nonlinearity \( f(u) \) of the system is assumed to be a linear
5.2. The Hammerstein-Wiener System Class

Figure 5.1: A Hammerstein-Wiener system. The static nonlinearities are represented by $f(u)$ and $g(r)$, the linear time invariant subsystem is represented by $H(q)$.

A combination of basis functions (e.g. polynomial):

$$x(t) = f(u(t)) = \sum_{i=1}^{n_f} \gamma_i f_i(u(t)), \quad (5.4)$$

where the coefficients $\gamma_i$ are grouped in the coefficient vector $\gamma \in \mathbb{R}^{n_f \times 1}$, with $n_f$ the total number of nonlinear basis functions used.

The output nonlinearity $g(r_0)$ of the system is assumed to be invertible. $g^{-1}(y_0)$ is also assumed to be representable by a linear combination of basis functions (e.g. polynomial):

$$y_0(t) = g(r_0(t)), \quad r_0(t) = g^{-1}(y_0(t)) = \sum_{j=1}^{n_g} \lambda_j g_j(y(t)), \quad (5.5)$$

where the coefficients $\lambda_j$ are grouped in the coefficient vector $\lambda \in \mathbb{R}^{n_g \times 1}$, with $n_g$ the total number of nonlinear basis functions used. This assumption limits the class of systems that can be modeled. However, it is not an uncommon assumption for the identification of Wiener and Hammerstein-Wiener systems (Kalafatis et al. [1997]; Crama and Schoukens [2004]; Goethals et al. [2005]).

The linear time invariant block $H(q)$ is assumed to be a rational function of the discrete time backward shift operator $q^{-1}$:

$$H(q) = \frac{B(q)}{A(q)} = \frac{b_0 + b_1 q^{-1} + b_2 q^{-2} + \ldots + b_{n_b} q^{-n_b}}{a_0 + a_1 q^{-1} + a_2 q^{-2} + \ldots + a_{n_a} q^{-n_a}}, \quad (5.6)$$

where the coefficients $b_i$ and $a_i$ are grouped in the coefficient vectors $b \in \mathbb{R}^{(n_b+1) \times 1}$ and $a \in \mathbb{R}^{(n_a+1) \times 1}$. 
The reader should realize that the sub-blocks in this structure cannot be uniquely identified from input-output data only. Infinitely many indistinguishable realizations can be obtained for the model by exchanging gain factors between the LTI subsystem and the static nonlinearities. To obtain a unique solution a normalization can be introduced, as is explained in Section 5.4.

### 5.3 Disturbing Noise Framework

![Diagram](image)

**Figure 5.2:** A Hammerstein-Wiener system, with white additive noise. The static nonlinearities are represented by the \( f \) and \( g \) blocks, the linear time invariant subsystem \( H(q) \) is represented by the \( B(q) \) and \( A(q) \) blocks. \( v(t) \) is the white additive noise source.

The noise is assumed to be added inside the LTI-block in the Hammerstein-Wiener model. It captures both the modeling error and the measurement noise. The LTI-block is split into two parts, as is shown on Figure 5.2. This results in the following output equations:

\[
\begin{align*}
y(t) &= g(r(t)), \\
r(t) &= \frac{1}{A(q)}(w(t) + v(t)), \\
w(t) &= B(q)x(t), \\
x(t) &= f(u(t)).
\end{align*}
\]

The noise source \( v(t) \) is assumed to be white zero-mean Gaussian noise.

### 5.4 System Normalization

To have a unique system description of a Hammerstein-Wiener system, some normalizations have to be introduced. Consider the system model as pre-
5.5. The Identification Problem

The identification problem presented in Figure 5.2. This description consists of 4 different blocks. A gain exchange is possible between these blocks without a change in the output signal. This introduces 2 degrees of freedom in the system description. To obtain a unique system description, 2 constraints are to be introduced on the parameter vectors $a$, $b$, $\gamma$ and $\lambda$. To this end, it is assumed that:

$$\|\gamma\|_2 = 1, \quad \gamma = [\gamma_1, \gamma_2, \ldots, \gamma_{nf}]^T,$$

$$\|\lambda\|_2 = 1, \quad \lambda = [\lambda_1, \lambda_2, \ldots, \lambda_{ng}]^T,$$

$$\left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2 = 1, \quad a = [a_0, a_1, \ldots, a_{na}]^T,$$

$$b = [b_0, b_1, \ldots, b_{nb}]^T. \quad (5.11)$$

Without loss of generality, the first nonzero element of $a$, $\lambda$ and $\gamma$ is forced to be positive, to fix the signs of the parameters.

This results in a unique system description in which all signals $u(t)$, $x(t)$, $w(t)$, $v(t)$, $s(t)$, $r(t)$, and $y(t)$ are uniquely defined.

5.5 The Identification Problem

The problem of identifying the parameters of a Hammerstein-Wiener system can be rewritten as is shown in Figure 5.3 using the assumptions made in Section 5.2. The intermediate signal $s(t)$ of the system may be written in two ways as a function of the input $u(t)$ and the output signal $y(t)$:

$$s(t) = \sum_{j=0}^{nb} b_j \sum_{k=1}^{nf} \gamma_k f_k (u(t-j)) + v(t), \quad (5.12)$$

and

$$s(t) = \sum_{i=0}^{na} a_i \sum_{l=1}^{ng} \lambda_l g_l (y(t-i)), \quad (5.13)$$

Subtracting eq. (5.12) and eq. (5.13) gives:

$$0 = \sum_{i=0}^{na} a_i \sum_{l=1}^{ng} \lambda_l g_l (y(t-i)) - \sum_{j=0}^{nb} b_j \sum_{k=1}^{nf} \gamma_k f_k (u(t-j)) - v(t),$$

$$v(t) = \sum_{i=0}^{na} a_i \lambda_l g_l (y(t-i)) - \sum_{j=0}^{nb} b_j \gamma_k f_k (u(t-j)). \quad (5.14)$$
The identification of the Hammerstein-Wiener system is performed by minimizing the following cost function:

\[
V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \epsilon(t)^2, \quad (5.15)
\]

\[
\epsilon(t) = \sum_{i=0}^{n_a} \sum_{l=1}^{n_g} \hat{a}_i \hat{\lambda}_l g_l (y(t-i)) + \sum_{j=0}^{n_b} \sum_{k=1}^{n_f} \hat{b}_j \hat{\gamma}_k f_k (u(t-j)). \quad (5.16)
\]

\(\theta\) contains the parameter vectors \(\hat{a}, \hat{b}, \hat{\gamma},\) and \(\hat{\lambda}\). The cost function expression is bilinear in the parameters. This bilinear problem can be converted into a linear problem by using an overparametrization approach (Section 5.7) or an iterative optimization approach (Section 5.8) (Bai and Liu [2005]).

### 5.6 Analogy with a Previous Approach

Hammerstein-Wiener like structures were already studied in literature. One of the most popular approaches for their identification is presented in Bai [1998]. This section studies the analogy between the problem formulation proposed in Section 5.5, and the original approach proposed in Bai [1998].

The system studied in Bai [1998] is described by the following input-output equation:

\[
y(t) = \sum_{i=1}^{n_a} a_i \sum_{l=1}^{n_g} \lambda_l g_l (y(t-i)) + \sum_{j=0}^{n_b} b_j \sum_{k=1}^{n_f} \gamma_k f_k (u(t-j)) + v(t). \quad (5.17)
\]
This represents a system which is related to, but not exactly equal to, a Hammerstein-Wiener block structured system. The block structure that is associated to eq. (5.17) is shown in Figure 5.4. As can be seen, this structure is almost a Hammerstein-Wiener block structure, but there are some differences. The signal $w(t)$ in Figure 5.4 contributes directly to the output signal, without being distorted by the output nonlinearity $g$. This is not the case for a 'true' Hammerstein-Wiener system as is shown in Figure 5.1.

\begin{align}
\sum_{l=1}^{n_g} a_l g_l(y(t)) &= - \sum_{i=1}^{n_a} a_i \sum_{l=1}^{n_g} \lambda_l g_l(y(t-i)) + \sum_{j=0}^{n_b} b_j \sum_{k=1}^{n_f} \gamma_k f_k(u(t-j)) + v(t).
\end{align}

Eq. (5.18) is very similar to eq. (5.17). The only difference is the presence (eq. (5.18)) or absence (eq. (5.17)) of the dependence of the left hand side of the equation on the output nonlinearity $g$. Remember that eq. (5.18) is obtained for a Hammerstein-Wiener system, where the output nonlinearity is assumed to be invertible.
5.7 Overparametrization Approach

The problem of identifying the Hammerstein-Wiener system can now be solved using an overparametrization approach, following the lines of Bai [1998]. First an overparametrized estimation problem that is linear in the new set of parameters (containing products of the original parameters) is solved in Section 5.7.1. Next the original parameters $a, b, \gamma, \lambda$ defined above are obtained starting from these new parameters as obtained in section 5.7.2.

5.7.1 Overparametrization of the Identification Problem

Eq. (5.18) is bilinear in the parameters $a, b, \gamma, \lambda$, but can be rewritten in an overparametrized form that is linear in a new set of parameters that consists of products of the original ones:

$$v(t) = \sum_{i=0}^{n_a} \sum_{l=1}^{n_g} a_i \lambda_l g_l(y(t-i)) - \sum_{j=0}^{n_b} \sum_{k=1}^{n_f} b_j \gamma_k f_k(u(t-j)),$$

$$= \Phi(t) \theta,$$  \hspace{1cm} (5.19)

where $\Phi(t) \in \mathbb{R}^{1 \times (n_f(n_b+1)+n_g(n_a+1))}$ and $\theta \in \mathbb{R}^{(n_f(n_b+1)+n_g(n_a+1)) \times 1}$ are given by:

$$\Phi(t) = [-f_1[u(t)], -f_2[u(t)], \ldots, -f_{n_f}[u(t)], -f_1[u(t-1)], \ldots, -f_{n_f}[u(t-n_b)],$$

$$g_1[y(t)], g_2[y(t)], \ldots, g_{n_g}[y(t)], g_1[y(t-1)], \ldots, g_{n_g}[y(t-n_a)]],$$

$$\theta = [b_0 \gamma_1, b_0 \gamma_2, \ldots, b_0 \gamma_{n_f}, b_1 \gamma_1, \ldots, b_{n_b} \gamma_{n_f},$$

$$a_0 \lambda_1, a_0 \lambda_2, \ldots, a_0 \lambda_{n_g}, a_1 \lambda_1, \ldots, a_{n_a} \lambda_{n_g}]^T.$$  \hspace{1cm} (5.20)

This equation is linear in the parameters $\theta$, and can be solved in total least squares sense (Van Huffel and Lemmerling [2002]) by taking the right singular vector corresponding to the smallest singular value of $\Phi \in \mathbb{R}^{1 \times (n_f(n_b+1)+n_g(n_a+1))}$. 

110
5.7. Overparametrization Approach

\( \mathbb{R}^{N \times (n_f(n_b+1) + n_g(n_a+1))} \). This corresponds to finding the nullspace of \( \Phi \).

\[
\Phi = [\Phi^T(1), \ldots, \Phi^T(t), \ldots, \Phi^T(N)]^T,
\]

\[
\Phi = U\Sigma V^T = \sum_l \sigma_l u_l v_l^T,
\]

\[
L = (n_b + 1) \times n_f + (n_a + 1) \times n_g,
\]

where \( u_i \in \mathbb{R}^{N \times 1} \) is the \( i \)-th left singular vector, and \( v_i \in \mathbb{R}^{(n_f(n_b+1) + n_g(n_a+1)) \times 1} \) is the \( i \)-th right singular vector. Given the singular value decomposition, where \( \sigma_L \) is the smallest singular value, the estimate of the parameter vector \( \hat{\theta} \) is given by:

\[
\hat{\theta} = v_L.
\]

When the nullspace has dimension 1, this normalized parameter vector with a \( L^2 \)-norm equal to one is unique. Without normalization, a scaling factor could be applied to all the parameters, resulting in a same system description. Normalization yields a unique solution for the overparametrized problem that is linear in the parameters (see Section 5.4).

5.7.2 Estimating the Model Parameters

Once an estimate \( \hat{\theta} \) is found, the parameters are rearranged in the matrices \( \hat{\theta}_{b\gamma} \in \mathbb{R}^{n_f \times (n_b+1)} \) and \( \hat{\theta}_{a\lambda} \in \mathbb{R}^{n_g \times (n_a+1)} \):

\[
\hat{\theta}_{b\gamma} = \hat{\gamma} \hat{b}^T = \begin{bmatrix}
\hat{b}_0 \gamma_1 & \cdots & \hat{b}_{n_b} \gamma_1 \\
\vdots & \ddots & \vdots \\
\hat{b}_0 \gamma_{n_f} & \cdots & \hat{b}_{n_b} \gamma_{n_f}
\end{bmatrix},
\]

\[
\hat{\theta}_{a\lambda} = \hat{\lambda} \hat{a}^T = \begin{bmatrix}
\hat{a}_0 \lambda_1 & \cdots & \hat{a}_{n_a} \lambda_1 \\
\vdots & \ddots & \vdots \\
\hat{a}_0 \lambda_{n_g} & \cdots & \hat{a}_{n_a} \lambda_{n_g}
\end{bmatrix}.
\]

The solutions of \( \hat{a}, \hat{b}, \hat{\gamma}, \hat{\lambda} \) are then found by calculating a rank-1 approximation of these matrices. This can again be done using the singular
value decomposition. The left and right singular vectors corresponding to the largest singular value of $\hat{\theta}_{b\gamma}$ and $\hat{\theta}_{a\lambda}$ yield the requested estimates, as shown in Bai [1998]:

\[
\hat{\theta}_{b\gamma} = \min_{l=1}^{\text{min}(n_b,n_f)} \sigma_l \mu_l \nu_l^T,
\]

(5.26)

\[
\hat{\theta}_{a\lambda} = \min_{l=1}^{\text{min}(n_a,n_g)} \delta_l \xi_l \zeta_l^T,
\]

(5.27)

where $\mu_l \in \mathbb{R}^{n_f \times 1}$ is the $l$-th left singular vector, $\nu_l \in \mathbb{R}^{(n_b+1)\times 1}$ is the $l$-th right singular vector of $\hat{\theta}_{b\gamma}$. Vector $\xi_l \in \mathbb{R}^{n_a \times 1}$ is the $l$-th left singular vector, and $\zeta_l \in \mathbb{R}^{(n_a+1)\times 1}$ is the $l$-th right singular vector of $\hat{\theta}_{a\lambda}$. The singular values of $\hat{\theta}_{b\gamma}$ and $\hat{\theta}_{a\lambda}$ are given by $\sigma_l$ and $\delta_l$ respectively. The parameters $\hat{a}$, $\hat{b}$, $\hat{\gamma}$, $\hat{\lambda}$ are given by:

\[
\hat{a} = \delta_1 \zeta_1,
\]

(5.28)

\[
\hat{b} = \sigma_1 \nu_1,
\]

(5.29)

\[
\hat{\gamma} = \mu_1,
\]

(5.30)

\[
\hat{\lambda} = \xi_1.
\]

(5.31)

5.7.3 Normalizing the Model Parameters

The estimated parameters $\hat{a}$, $\hat{b}$, $\hat{\gamma}$, $\hat{\lambda}$, are normalized according to the following normalization scheme:

1. Let $s_a, s_\gamma, s_\lambda$ denote the sign of the first nonzero element of $\hat{a}$, $\hat{\gamma}$, and $\hat{\lambda}$.

2. Let $\alpha$ denote the joint parameter vector of $\hat{a}$ and $\hat{b}$:

\[
\alpha = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}.
\]

(5.32)
5.8 Iterative Approach

The normalized parameter vectors $\hat{a}_n, \hat{b}_n, \hat{\gamma}_n, \hat{\lambda}_n$ are then given by:

\[
\hat{\gamma}_n = s_\gamma \frac{\hat{\gamma}}{\| \hat{\gamma} \|_2},
\]

\[
\hat{\lambda}_n = s_\lambda \frac{\hat{\lambda}}{\| \hat{\lambda} \|_2},
\]

\[
\hat{a}_n = s_q \frac{\hat{\alpha}}{\| \hat{\alpha} \|_2} \sqrt{\frac{\| \hat{\lambda} \|_2}{\| \hat{\gamma} \|_2}},
\]

\[
\hat{b}_n = s_q s_\lambda s_\gamma \frac{\hat{\beta}}{\| \hat{\alpha} \|_2} \sqrt{\frac{\| \hat{\gamma} \|_2}{\| \hat{\lambda} \|_2}},
\]

where $\| \cdot \|_2$ denotes the $l^2$-norm of a vector.

5.8 Iterative Approach

The iterative method is a well known approach for Hammerstein systems (Narendra and Gallman [1966]; Crama et al. [2004]; Bai and Li [2010]). It was first introduced by Narendra and Gallman [1966]. An iterative approach can also be used for Wiener (Wang and Ding [2011]) or Hammerstein-Wiener (Voros [2004]) systems. In this section, an iterative approach is used to estimate the model parameters of a Hammerstein-Wiener system, starting from eq. (5.14).

5.8.1 Problem Formulation

Eq. (5.14) can be rewritten as:

\[
v(t) = \sum_{i=1}^{n_q} \sum_{l=1}^{n_q} a_i \lambda_{il} g_i(y(t-i)) - \sum_{j=0}^{n_f} \sum_{k=1}^{n_f} b_j \gamma_k f_k(u(t-j)),
\]

\[
= \mathbf{a}^T \mathbf{G}(t) \mathbf{\lambda} - \mathbf{b}^T \mathbf{F}(t) \mathbf{\gamma},
\]
where $G(t) \in \mathbb{R}^{(n_a+1) \times n_g}$ and $F(t) \in \mathbb{R}^{(n_b+1) \times n_f}$ are given by:

$$F(t) = \begin{bmatrix}
    f_1(u(t)) & f_2(u(t)) & \cdots & f_{n_f}(u(t)) \\
    f_1(u(t-1)) & f_2(u(t-1)) & \cdots & f_{n_f}(u(t-1)) \\
    \vdots & \vdots & \ddots & \vdots \\
    f_1(u(t-n_b)) & f_2(u(t-n_b)) & \cdots & f_{n_f}(u(t-n_b))
\end{bmatrix}.$$  \hspace{1cm} (5.38)

$$G(t) = \begin{bmatrix}
    g_1(y(t)) & g_2(y(t)) & \cdots & g_{n_g}(y(t)) \\
    g_1(y(t-1)) & g_2(y(t-1)) & \cdots & g_{n_g}(y(t-1)) \\
    \vdots & \vdots & \ddots & \vdots \\
    g_1(y(t-n_a)) & g_2(y(t-n_a)) & \cdots & g_{n_g}(y(t-n_a))
\end{bmatrix}.$$  \hspace{1cm} (5.39)

This results in the following cost function $V_N(\hat{a}, \hat{b}, \hat{\gamma}, \hat{\lambda})$:

$$V_N(\hat{a}, \hat{b}, \hat{\gamma}, \hat{\lambda}) = \frac{1}{N} \sum_{t=1}^{N} \left( \hat{a}^T G(t) \hat{\lambda} - \hat{b}^T F(t) \hat{\gamma} \right)^2.$$  \hspace{1cm} (5.40)

### 5.8.2 Iterative Algorithm

The cost function $V_N(\hat{a}, \hat{b}, \hat{\gamma}, \hat{\lambda})$ in eq. (5.40) can be minimized using an iterative approach. The parameter values of iteration $i$ $(\hat{a}_i, \hat{b}_i, \hat{\gamma}_i, \hat{\lambda}_i)$ are obtained as the result of a four-step iterative approach:

0. The initialization of the iterative approach is done using random parameter values.

1. Estimate updated values of $\hat{a}_i$ and $\hat{b}_i$, using the estimates of $\hat{\gamma}_{i-1}$ and $\hat{\lambda}_{i-1}$ of the previous iteration.

$$\left( \hat{a}_i, \hat{b}_i \right) = \arg \min_{\hat{a}, \hat{b}} \{ V_N(\hat{a}, \hat{b}, \hat{\gamma}_{i-1}, \hat{\lambda}_{i-1}) \}. \hspace{1cm} (5.41)$$

2. Normalize updated parameter vectors of $\hat{a}_i$ and $\hat{b}_i$, according to the normalization rules introduced in 5.4.

3. Update the values of $\hat{\gamma}_i$ and $\hat{\lambda}_i$, using the estimates of $\hat{a}_i$ and $\hat{b}_i$ obtained in step 2 in the current iteration.

$$\left( \hat{\gamma}_i, \hat{\lambda}_i \right) = \arg \min_{\hat{\gamma}, \hat{\lambda}} \{ V_N(\hat{a}_i, \hat{b}_i, \hat{\gamma}, \hat{\lambda}) \}. \hspace{1cm} (5.42)$$
4. Normalize the updated parameter vectors of $\hat{\gamma}_i$ and $\hat{\lambda}_i$, following the normalization rules introduced in 5.4.

5. Stop if convergence is reached according to Fletcher [1991]:

$$\| V_N(\hat{a}_i, \hat{b}_i, \hat{\gamma}_i, \hat{\lambda}_i) - V_N(\hat{a}_{i-1}, \hat{b}_{i-1}, \hat{\gamma}_{i-1}, \hat{\lambda}_{i-1}) \|_2 < \delta, \quad (5.43)$$

where $\delta$ is a lower bound set by the user. Jump back to step 1, with $i = i + 1$ if convergence is not reached.

Eq. (5.41) boils down to a least squares estimation that is linear in the parameters. When $\hat{\gamma}_{i-1}$ and $\hat{\lambda}_{i-1}$ are known, eq. (5.40) can be rewritten as:

$$V_N(\hat{a}, \hat{b}, \hat{\gamma}_{i-1}, \hat{\lambda}_{i-1}) = \frac{1}{N} \sum_{t=1}^{N} \left( \hat{a}^T G(t) \hat{\lambda}_{i-1} - \hat{b}^T F(t) \hat{\gamma}_{i-1} \right)^2,$$

$$= \frac{1}{N} \sum_{t=1}^{N} \left( \hat{a}^T G_{\lambda_{i-1}}(t) - \hat{b}^T F_{\gamma_{i-1}}(t) \right)^2,$$

$$= \frac{1}{N} \sum_{t=1}^{N} (K(t) \hat{\theta})^2,$$  \quad (5.44)

where:

$$F_{\gamma_{i-1}}(t) = F(t) \hat{\gamma}_{i-1}, \quad F_{\gamma_{i-1}}(t) \in \mathbb{R}^{(n_b+1) \times 1},$$

$$G_{\lambda_{i-1}}(t) = G(t) \hat{\lambda}_{i-1}, \quad G_{\lambda_{i-1}}(t) \in \mathbb{R}^{(n_a+1) \times 1},$$

$$K(t) = \begin{bmatrix} -F_{\gamma_{i-1}}(t) & G_{\lambda_{i-1}}^T(t) \end{bmatrix},$$

$$\hat{\theta} = \begin{bmatrix} \hat{b}_i^T & \hat{a}_i^T \end{bmatrix}^T.$$

(5.45)

This problem is linear in the parameters and can be solved using a total least squares approach as above, by taking the right singular vector corresponding to the smallest singular value of $K \in \mathbb{R}^{N \times (n_b+n_a+2)}$:

$$K = \begin{bmatrix} K^T(1), \ldots, K^T(t), \ldots, K^T(N) \end{bmatrix}^T.$$  \quad (5.46)

The problem of minimizing eq. (5.42) is very similar, and can be solved using the same approach.
5. HAMMERSTEIN-WIENER

5.9 Bias of the Estimate

This section investigates the presence of a bias on the estimates that are obtained by minimizing the cost function in eq. (5.15). Consider the expected value of the eq. (5.15):

\[ \bar{V} (\theta) = E \{ V_N (\theta) \} . \]  

(5.47)

The difference between \( \bar{V} (\theta) \) and \( \bar{V} (\theta_0) \) results in:

\[ \bar{V} (\theta) - \bar{V} (\theta_0) = 2E \{ (\epsilon (t) - \epsilon_0(t)) \epsilon_0(t) \} + E \{ (\epsilon (t) - \epsilon_0(t))^2 \} , \]

(5.48)

where \( \epsilon_0(t) \) is the value of \( \epsilon(t) \) evaluated for the true parameters \( \theta_0 \). The signal \( \epsilon(t) \) can be split in a part that is independent of the noise \( v(t) \), and a part that is dependent on \( v(t) \):

\[ \epsilon(t) = \sum_{i=0}^{n_i} \sum_{l=1}^{n_g} \hat{a}_i \hat{\lambda}_l g_l (y(t - i)) - \sum_{j=0}^{n_f} \sum_{k=1}^{n_f} \hat{b}_j \hat{\gamma}_k f_k (u(t - j)) , \]

(5.49)

\[ = \sum_{l=1}^{n_g} \hat{a}_0 \hat{\lambda}_l g_l (y(t)) + \sum_{i=1}^{n_g} \sum_{l=1}^{n_g} \hat{a}_i \hat{\lambda}_l g_l (y(t - i)) - \sum_{j=0}^{n_f} \sum_{k=1}^{n_f} \hat{b}_j \hat{\gamma}_k f_k (u(t - j)) , \]

(5.50)

where only \( \sum_{l=1}^{n_g} \hat{a}_0 \hat{\lambda}_l g_l (y(t)) \) is depending on \( v(t) \). The signals \( \tilde{y}(t) \) and \( \tilde{u}(t) \) are independent of \( v \) at time instant \( t \):

\[ \tilde{y}(t) = \sum_{i=1}^{n_g} \sum_{l=1}^{n_g} \hat{a}_i \hat{\lambda}_l g_l (y(t - i)) , \]

(5.51)

\[ \tilde{u}(t) = \sum_{j=0}^{n_f} \sum_{k=1}^{n_f} \hat{b}_j \hat{\gamma}_k f_k (u(t - j)) . \]

(5.52)

The difference \( \epsilon(t) - \epsilon_0(t) \) is hence given by:

\[ \epsilon(t) - \epsilon_0(t) = \sum_{l=1}^{n_g} \Delta \hat{a}_0 \Delta \hat{\lambda}_l g_l (y(t)) + \Delta \tilde{y}(t) + \Delta \tilde{u}(t) , \]

(5.53)

\[ \Delta \hat{a}_0 = \hat{a}_0 - a_0 , \]

(5.54)

\[ \Delta \hat{\lambda}_l = \hat{\lambda}_l - \lambda_l , \]

(5.55)

\[ \Delta \tilde{y}(t) = \tilde{y}(t) - \tilde{y}_0(t) , \]

(5.56)

\[ \Delta \tilde{u}(t) = \tilde{u}(t) - \tilde{u}_0(t) , \]

(5.57)
where \( \tilde{y}_0(t) \) and \( \tilde{u}_0(t) \) are the values of \( \tilde{y}(t) \) and \( \tilde{u}(t) \) evaluated for the true parameters \( \theta_0 \).

The signal \( \epsilon_0(t) \) is equal to the noise \( v(t) \) (see eq. (5.14) and eq. (5.15)). This results in the following expression for the first term of eq. (5.48):

\[
E \{ (\epsilon(t) - \epsilon_0(t)) \epsilon_0(t) \} = E \left\{ \left( \sum_{l=1}^{n_g} \Delta \hat{a}_0 \lambda_l g_l (y(t)) + \Delta \tilde{y}(t) + \Delta \tilde{u}(t) \right) v(t) \right\},.
\]

The expected value of the terms \( \Delta \tilde{y}(t) v(t) \) and \( \Delta \tilde{u}(t) v(t) \) is zero since \( \Delta \tilde{y}(t) \) and \( \Delta \tilde{u}(t) \) depend only on input-output data up to time \( t-1 \). \( \Delta \tilde{y}(t) \) and \( \Delta \tilde{u}(t) \) are therefore independent of \( v(t) \). This does not hold for the term \( \sum_{l=1}^{n_g} \Delta \hat{a}_0 \lambda_l g_l (y(t)) v(t) \). Hence, this term can introduce a bias on the estimates that are obtained by minimizing eq. (5.15).

The second term of eq. (5.48) \( E \{ (\epsilon(t) - \epsilon_0(t))^2 \} \) is strictly positive if \( \theta \) and \( \theta_0 \) originate from different models. Hence, this term does not introduce a bias on the estimates.

The bias on the estimates is illustrated on a simulation example in the next section.

### 5.10 Unbiased Estimates: Monic Denominator

The previous section shows that the minimization of the cost function (eq. (5.15)) leads to a biased estimate. The bias is due to the term shown in eq. (5.59). It is clear from eq. (5.59) that this term becomes zero if \( \Delta \hat{a}_0 \) is zero. This can be obtained by an appropriate normalization by enforcing a monic denominator model.

A monic denominator model fixes the direct term equal to one \( (a_0 = 1) \). To achieve this \( a_0 \) is no longer estimated, it is simply fixed to one, and the
normalization of the model parameters is changed:

\[
\hat{\gamma}_n = s_{\gamma} \frac{\hat{\gamma}}{\|\hat{\gamma}\|_2}, \quad (5.60)
\]

\[
\hat{\lambda}_n = s_{\lambda} \frac{\hat{\lambda}}{\|\hat{\lambda}\|_2}, \quad (5.61)
\]

\[
\hat{b}_n = s_{\lambda} s_{\gamma} \hat{b} \frac{\|\hat{\gamma}\|_2}{\|\hat{\lambda}\|_2}. \quad (5.62)
\]

The estimation of the parameters cannot be performed with the algorithms as they are described in Sections 5.7 and 5.8. A nonlinear optimization similar to the one described in Section 3.5.6 is used instead. The parameters can be initialized using the two algorithms that are described in Sections 5.7 and 5.8. Note that the iterative algorithm (Section 5.8) could be adapted such that a monic denominator model is used.

5.11 Simulation Example

A simulation experiment is performed and the results of the different identification methods are discussed. The biased nature of the estimator, in case no monic denominator model is used, is illustrated. Finally, the unbiased nature of the estimates using a monic denominator model is illustrated.

5.11.1 Illustration of the Estimators

The method is illustrated on the simulation of a Hammerstein-Wiener system with additive zero mean white Gaussian noise at the signal \( w(t) \) as in Figure 5.2.

The static nonlinear block at the input is a polynomial of third degree given below:

\[
x = 0.9759u + 0.0976u^2 + 0.1952u^3. \quad (5.63)
\]

The inverse function of the static nonlinear block at the output is a polynomial of third degree:

\[
r = 0.9998y + 0.0100y^2 + 0.0200y^3. \quad (5.64)
\]
5.11. Simulation Example

The front and the back static nonlinearity are shown in Figure 5.5 and Figure 5.6 respectively.

The linear dynamics of the system are produced by a low-pass Chebychev filter of order 10, with a passband ripple of 10dB, a gain of 34dB and a cut-off frequency at 0.1\textit{f}_s.

The disturbance noise \( v(t) \) is white Gaussian noise with a standard
deviation of $5 \times 10^{-8}$. This results in a noise level at the output that lies 20 to 30dB lower than the output spectrum, as is shown in Figure 5.7. White Gaussian noise with a standard deviation of 1 is used as an input excitation signal. The 'measured' signals contain 1024 samples and the sampling frequency, $f_s$, is normalized to one. The initial conditions of the system are set to zero.

A validation simulation is performed using a different input realization. The results are shown in Figure 5.7. The norm of the simulated output signal is 231, the norm of the noise at the output is 9.72, the norm of the error of the output of the iterative model is 10.4, the norm of the error of the output of the model obtained with the overparametrization approach is 21.8. The estimates shown for the iterative method are obtained after 10 iterations. This shows that both methods provide very good estimates of the system, even in the presence of noise.

The estimate of the overparametrization approach has a larger error than the iterative approach. This can be explained by the higher number
of parameters that are estimated during the overparametrized step. The difference between both approaches is less clear when lower orders for the linear filters \( B(q) \) and \( A(q) \) are used. One could consider to use the estimate of the overparametrization approach as a starting point for the iterative estimation algorithm.

### 5.11.2 Bias of the Estimates

A simple simulation example is studied to illustrate the bias on the estimates. The simulated system has no input nonlinearity, the inverse of the output nonlinearity is given by a third order monomial:

\[
x = u, \quad (5.65)
\]

\[
r = y^3. \quad (5.66)
\]

The LTI block is a first order rational transfer function:

\[
H(q) = \frac{0.6667}{0.6667 - 0.3333q^{-1}}. \quad (5.67)
\]

The estimated model has the same model structure as the simulated system. A Monte Carlo simulation is performed for different noise levels. 1000 models are estimated in each Monte Carlo simulation. The system is excited by white Gaussian noise with a standard deviation of 1. Each random noise sequence contains \( 2^{14} \) samples. The initial conditions of the system are set to zero. The coefficients of the front and the back nonlinearity are normalized to 1 in the normalization step. Only the coefficients of the LTI block vary over the estimated models. The bias is obtained as:

\[
\text{bias} \hat{\theta} = \left| \frac{1}{1000} \sum_{i=1}^{1000} \hat{\theta}_i - \theta \right|. \quad (5.68)
\]

A clear bias can be observed on the parameter estimates in Table 5.1 and Figure 5.8. For larger standard deviations of the noise \( v(t), \sigma_v \), the bias is much larger than the standard deviation on the estimate. The standard deviation shown in Table 5.1 and Figure 5.8 is the standard deviation of one estimate (one Monte Carlo realization), not the standard deviation of the mean of the estimates.
Table 5.1: Monte Carlo simulation results. The bias and standard deviation on the estimated parameters for an increasing standard deviation of the noise $v(t)$ ($\sigma_v$). The results are shown for $b_0$, $a_0$ and $a_1$.

<table>
<thead>
<tr>
<th>$\sigma_v$</th>
<th>$10^{-3}$</th>
<th>$5 \cdot 10^{-3}$</th>
<th>$10^{-2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias $b_0$</td>
<td>266.6 $10^{-9}$</td>
<td>8.734 $10^{-6}$</td>
<td>34.65 $10^{-6}$</td>
</tr>
<tr>
<td>bias $a_0$</td>
<td>130.6 $10^{-9}$</td>
<td>3.421 $10^{-6}$</td>
<td>8.184 $10^{-6}$</td>
</tr>
<tr>
<td>bias $a_1$</td>
<td>201.3 $10^{-9}$</td>
<td>7.022 $10^{-6}$</td>
<td>30.55 $10^{-6}$</td>
</tr>
<tr>
<td>$\sigma_{b_0}$</td>
<td>135.2 $10^{-9}$</td>
<td>639.9 $10^{-9}$</td>
<td>1.271 $10^{-6}$</td>
</tr>
<tr>
<td>$\sigma_{a_0}$</td>
<td>204.3 $10^{-9}$</td>
<td>1.035 $10^{-6}$</td>
<td>2.181 $10^{-6}$</td>
</tr>
<tr>
<td>$\sigma_{a_1}$</td>
<td>157.1 $10^{-9}$</td>
<td>754.6 $10^{-9}$</td>
<td>1.565 $10^{-6}$</td>
</tr>
<tr>
<td>$\sigma_v$</td>
<td>$5 \cdot 10^{-2}$</td>
<td>$10^{-1}$</td>
<td>$5 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>bias $b_0$</td>
<td>878.4 $10^{-6}$</td>
<td>3.516 $10^{-3}$</td>
<td>84.87 $10^{-3}$</td>
</tr>
<tr>
<td>bias $a_0$</td>
<td>161.6 $10^{-6}$</td>
<td>700.3 $10^{-6}$</td>
<td>8.258 $10^{-3}$</td>
</tr>
<tr>
<td>bias $a_1$</td>
<td>796.4 $10^{-6}$</td>
<td>3.148 $10^{-3}$</td>
<td>78.87 $10^{-3}$</td>
</tr>
<tr>
<td>$\sigma_{b_0}$</td>
<td>6.339 $10^{-6}$</td>
<td>12.28 $10^{-6}$</td>
<td>67.06 $10^{-6}$</td>
</tr>
<tr>
<td>$\sigma_{a_0}$</td>
<td>10.65 $10^{-6}$</td>
<td>20.96 $10^{-6}$</td>
<td>94.32 $10^{-6}$</td>
</tr>
<tr>
<td>$\sigma_{a_1}$</td>
<td>7.924 $10^{-6}$</td>
<td>15.65 $10^{-6}$</td>
<td>74.32 $10^{-6}$</td>
</tr>
</tbody>
</table>

Figure 5.8: Monte Carlo simulation results. The bias (circles - full line) and standard deviation (crosses - dashed line) on the estimated parameters for an increasing standard deviation $\sigma_v$ of the noise $v(t)$. The results for $b_0$, $a_0$ and $a_1$ are shown in red, blue and green respectively.
5.11.3 Monic Denominator Model

A simulation example similar to the one in the previous section is studied to illustrate the unbiased nature of the estimates. The simulated system has no input nonlinearity, the inverse of the output nonlinearity is given by a third order monomial:

\[ x = u, \quad (5.69) \]
\[ r = y^3. \quad (5.70) \]

The LTI block is a first order rational transfer function with a monic denominator:

\[ H(q) = \frac{1}{1 - 0.5q^{-1}}. \quad (5.71) \]

The estimated model has the same model structure as the simulated system and uses a monic denominator model (see Section 5.10). A Monte Carlo simulation is performed with a standard deviation of the noise equal to \( 5 \times 10^{-1} \). 1000 models are estimated in the Monte Carlo simulation. The system is excited by white Gaussian noise with a standard deviation of 1. Each random noise sequence contains \( 2^{14} \) samples. The initial conditions of the system are set to zero. The coefficients of the front and the back nonlinearity are normalized to 1 in the normalization step. Only the coefficients of the LTI block vary over the estimated models. The bias is obtained as in eq. (5.68).

It can be observed in Table 5.2 that the bias is always smaller than one standard deviation on the estimate. This illustrates the unbiased nature of the estimates. The standard deviation shown in Table 5.2 is the standard deviation of one estimate (one Monte Carlo realization), not the standard deviation of the mean of the estimates.

5.12 Conclusion and Discussion

The problem of estimating a Hammerstein-Wiener model is re-formulated. Using this formulation, both an overparametrization and an iterative approach are proposed to estimate the model parameters, starting from input
Table 5.2: Monte Carlo simulation results. The bias and standard deviation on the estimated parameters for an increasing standard deviation of the noise $v(t)$ ($\sigma_v$). The results are shown for $b_0$ and $a_1$.

<table>
<thead>
<tr>
<th>$\sigma_v$</th>
<th>$5 \times 10^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias $\hat{b}_0$</td>
<td>$0.037 \times 10^{-4}$</td>
</tr>
<tr>
<td>bias $\hat{a}_1$</td>
<td>$0.799 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\hat{\sigma}_{b_0}$</td>
<td>$1.305 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\hat{\sigma}_{a_1}$</td>
<td>$1.016 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

and output data only. The proposed method assumes the presence of a specific noise model, and also assumes that the output nonlinearity is invertible. The linear time invariant block is modeled using a difference equation, while the nonlinearities are modeled as a linear combination of nonlinear static basis functions. The bias on the estimates is analyzed. It is shown that an unbiased estimate can be obtained if a monic denominator model is used. The good performance of both methods is illustrated in a simulation example together with the behavior of the bias on the estimates and the unbiased nature of the estimates using a monic denominator model.
Parallel Hammerstein

Based on Schoukens et al. [2010] and Schoukens et al. [2011]

6.1 Introduction

Hammerstein systems consist of a static nonlinearity connected in tandem to a linear time invariant (LTI) dynamic system. These systems are used to model nonlinear systems for which the nonlinearity is present mainly at the input (Doyle et al. [2002]; Haber and Keviczky [1999]), such as power amplifiers, chemical processes and physiological systems (Giri and Bai [2010]). The problem of a Hammerstein model is that it suffers from a lack of general applicability because it is not flexible enough to cope with systems that are ’close’ to a Hammerstein behavior.

The extension that we propose to the Hammerstein system is based on the following observation: the input signal of some systems can follow different paths that are finally combined in some way to form the system output. For example in RF power amplifiers, the main part of the input signal will be amplified to the output in a nonlinear way through the main signal path. A Hammerstein model is well suited for this type of behavior. However, some distorted part of the signal can also find its way to the power supply terminal connection and can eventually appear at the output after reflection and modulation. This forms a second path that can again be
described as belonging to the Hammerstein class.

The intuitive feeling therefore is that such systems can be better modeled by replacing the single Hammerstein model by a more general structure consisting of a parallel connection of more than one Hammerstein system. These models are sometimes called Uryson models in the literature, see for example Doyle et al. [2002]; Gallman [1975].

A popular approach that is used to identify a parallel Hammerstein system is to feed the input signal to a parallel connection of some non-linear basis functions that are each followed by a finite impulse response filter (FIR) to realize a branch. The corresponding LTI multiple-input-single-output identification problem is then solved using one’s favorite FIR identification method in Gadringer et al. [2007]. In earlier work (Gallman [1975]), Hermite polynomials were used as orthogonal basis functions to model the static nonlinearity for Gaussian input signals.

These approaches result in a linear least squares identification problem, which is easy to solve. However, they suffer from two drawbacks: i) the user gets no physical insight in the number of parallel paths in the device under test (DUT), as the number of parallel branches in the model is set by the degree of the nonlinearity, ii) for systems with long memory effects, a large number of parameters is needed due to the FIR-nature of the dynamic model. The model is therefore not parsimonious.

Both a semi-parametric and a parametric identification method for parallel Hammerstein systems are introduced in Schoukens et al. [2010] and Schoukens et al. [2011]. The main contributions of these methods are:

1. The number of branches is independent of the degree of nonlinearity, and they possibly possess a physical interpretation.

2. The number of branches can be selected based on the results obtained by the identification method, leading to parsimony in the branch count.

3. The LTI blocks can be modeled using rational models, leading to parsimony in the LTI parametrization.
4. This chapter focuses on the discrete time case, but the method that is presented here can be extended for the continuous time case.

This chapter follows the lines of Schoukens et al. [2010] and Schoukens et al. [2011]. The first part describes the system class and discusses its identifiability (Sections 6.2 and 6.3). The specific behavior of the BLA of a Hammerstein system is studied in Section 6.4. Next, the identification procedure is discussed in some detail in Section 6.5. Finally, an illustration on a simulation example shows the good performance of the estimator in Section 6.6.

### 6.2 The Parallel Hammerstein System Class

![Figure 6.1: A parallel Hammerstein system with 3 parallel branches. $S[k](q)$ is the LTI system of branch $k$, $f[k](u)$ is the static nonlinearity of branch $k.](image)

A parallel Hammerstein system connects several Hammerstein subsystems in parallel as is shown in Figure 6.1. The noiseless output of a parallel Hammerstein system is given by:

$$y_0(t) = \sum_{k=1}^{n_{br}} y_k(t), \quad (6.1)$$

$$y_k(t) = S[k](q)r_k(t), \quad (6.2)$$

$$r_k(t) = f[k](u(t)). \quad (6.3)$$
Herein, $f^{[k]}(.)$ is the static nonlinear function and $S^{[k]}(q)$ is the LTI system that are associated to branch $k$. The number of parallel branches is given by $n_{br}$. The signals $y_0(t)$, $y_k(t)$, $r_k(t)$ and $u(t)$ are scalar signals.

The static nonlinearities $f^{[k]}(.)$ are assumed to be obtained as a linear combination of basis functions (e.g. polynomial):

$$f^{[k]}(u(t)) = \sum_{i=1}^{n_f} \gamma_i^{[k]} f_i(u(t)), \quad (6.4)$$

where the coefficients $\gamma_i^{[k]}$ are grouped in the coefficient vector $\gamma^{[k]} \in \mathbb{R}^{n_f \times 1}$, and $n_f$ is the total number of nonlinear basis functions.

The LTI subsystem $S^{[k]}(q)$ is assumed to be a rational function of the backward shift operator $q^{-1}$:

$$S^{[k]}(q) = \frac{b_0^{[k]} + b_1^{[k]} q^{-1} + \cdots + b_{n_{b,k}}^{[k]} q^{-n_{b,k}}}{a_0^{[k]} + a_1^{[k]} q^{-1} + \cdots + a_{n_{a,k}}^{[k]} q^{-n_{a,k}}}, \quad (6.5)$$

where the coefficients $b_i^{[k]}$ and $a_i^{[k]}$ are grouped in the coefficient vectors $b^{[k]} \in \mathbb{R}^{(n_{b,k}+1) \times 1}$ and $a^{[k]} \in \mathbb{R}^{(n_{a,k}+1) \times 1}$. The $l^2$-norm of the coefficient vector $a^{[k]}$ is assumed to be fixed to a value that is equal to 1, and the first nonzero element of $a^{[k]}$ is assumed to be positive to compensate for the degeneration that is present in the rational transfer function parametrization.

### 6.3 Identifiability of a Parallel Hammerstein System

The LTI blocks and the static nonlinear blocks contained in a parallel Hammerstein system can only be identified up to an unknown gain for each branch, unless one has access to the intermediate signals $r_k(t)$. If a gain $\beta_k$ is introduced for both $S^{[k]}(q)$ and $f^{[k]}(u)$ one obtains:

$$y_k(t) = \tilde{S}^{[k]}(q) \tilde{r}_k(t) = \beta_k S^{[k]}(q) \tilde{r}_k(t), \quad (6.6)$$

$$\tilde{r}_k(t) = \tilde{f}^{[k]}(u(t)) = \frac{1}{\beta_k} f^{[k]}(u(t)). \quad (6.7)$$
6.3. Identifiability of a Parallel Hammerstein System

The input-output behavior of a parallel Hammerstein system that is composed using $\tilde{S}[k](q)$ and $\tilde{f}[k](u(t))$ is mathematically exactly the same as a parallel Hammerstein system containing $S[k](q)$ and $f[k](u(t))$ as building blocks.

![Diagram of a parallel Hammerstein system with 3 parallel branches subject to a linear transformation $T$.](image)

**Figure 6.2:** A parallel Hammerstein system with 3 parallel branches subject to a linear transformation $T$.

An additional identifiability issue appears here due to the parallel nature of the parallel Hammerstein systems. Starting from input-output data only, infinitely many equivalent models can be obtained by adding a regular linear transformation to one of the models. This introduces a full rank linear transformation matrix between the outputs of the static nonlinearities $f[k](u(t))$ and the inputs of the back LTI blocks $S[k](q)$. This full rank linear transformation is described by a matrix $T$. Since $T$ is regular its inverse transformation $T^{-1}$ can also be included in between the LTI blocks and the static nonlinear blocks as is illustrated in Figure 6.2. This results in a new set of LTI and static nonlinear blocks that are a linear combination of the original ones:

$$\tilde{S}[k](q) = \sum_{i=1}^{n_{br}} t_{k,i} S[i](q), \quad (6.8)$$

$$\tilde{f}[k](u(t)) = \sum_{i=1}^{n_{br}} t_{k,i} f[i](u(t)), \quad (6.9)$$

where

$$T = \begin{bmatrix} t_{1,1} & \cdots & t_{1,n_{br}} \\ \vdots & \ddots & \vdots \\ t_{n_{br},1} & \cdots & t_{n_{br},n_{br}} \end{bmatrix}. \quad (6.10)$$
6. Parallel Hammerstein

and

\[ \tilde{T} = T^{-1} = \begin{bmatrix} \tilde{t}_{1,1} & \cdots & \tilde{t}_{1,n_{br}} \\ \vdots & \ddots & \vdots \\ \tilde{t}_{1,1} & \cdots & \tilde{t}_{n_{br},n_{br}} \end{bmatrix}. \] (6.11)

The input-output behavior of a parallel Hammerstein system composed by the transformed \( \tilde{S}^{[k]}(q) \) and \( \tilde{f}^{[k]}(u(t)) \) is exactly the same as that of a parallel Hammerstein system containing \( S^{[k]}(q) \) and \( f^{[k]}(u(t)) \) as building blocks.

Although theoretically the gain exchanges \( \beta_k \) and the transformation \( T \) result in the same input-output behavior of the system, bad choices of these gains \( \beta_k \) and the transformation \( T \) can lead to numerical issues. It is best to avoid a high dynamic range over the different parameters that are present in the model. Also a high dynamic range over the different signals \( (u(t), r_k(t), y_k(t) \) and \( y(t) \) ) should be avoided if possible.

6.4 Best Linear Approximation

This section shows that the BLA of a parallel Hammerstein system is a linear combination of the LTI dynamics that are present in the different LTI blocks \( S^{[k]}(q) \) when input signals belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals are used as an excitation signal.

**Assumption 6.1. Input signal:** The input signal belongs to the Riemann equivalence class of asymptotically normally distributed excitation signals as defined in Definition 2.1.

**Theorem 6.1.** The BLA \( G_{bla}(e^{j\omega T_s}) \) of a parallel Hammerstein system excited by an input signal belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals (Assumption 6.1) is given by:

\[ G_{bla}(e^{j\omega T_s}) = \sum_{k=1}^{n_{br}} \alpha_k S^{[k]}(e^{j\omega T_s}), \quad \alpha_k \in \mathbb{R}. \] (6.12)
6.4. Best Linear Approximation

Proof. The BLA of a Hammerstein system is given by (3.11) \( G_{bla}(e^{j\omega T_s}) = \alpha S(e^{j\omega T_s}) \) under Assumption 6.1. This holds for each separate branch in the parallel Hammerstein system. The output \( y \) is given by a summation of the outputs \( y_k \) of each parallel Hammerstein branch. Hence, the BLA of a parallel Hammerstein system is given by the sum of the BLA of each parallel branch.

Note that the shape of \( G_{bla}(e^{j\omega T_s}) \) depends on the power spectrum of the input signal. The shape of \( G_{bla}(e^{j\omega T_s}) \) can change if a class of input signals with a different power spectrum is used. The coefficients \( \alpha_k \) depend on many factors: the LTI subsystem \( S[k](q) \), the static nonlinear subsystems \( f[k](u(t)) \), the power spectrum \( S_{UU}(e^{j\omega T_s}) \) of the input signal \( u(t) \), and hence as well on the variance of the input signal \( u(t) \).

Assumption 6.2. Non-zero gain: The BLA \( \alpha_k S[k](q) \) of every branch \( k \) has a non-zero gain \( \alpha_k \).

It can happen that the BLA of one of the branches of the parallel Hammerstein system is equal to zero, or in other words \( \alpha_k = 0 \). This is the case when the nonlinear function \( f[k](u(t)) \) is even around the expected value of \( u(t) \). In this case, a BLA of a reduced order is obtained that does not contain the dynamics of branch \( k \). This assumption excludes that the static nonlinearity \( f[k](u(t)) \) of branch \( k \) is symmetric with respect to the DC setpoint of the input signal \( u \). Different DC setpoints can be tried to avoid a zero gain \( \alpha_k \).

An important observation with respect to eq. (6.12) is that the input dependent gain \( \alpha_k \) only appears in the numerator:

\[
G_{bla}(q) = \frac{\sum_{k=1}^{n_{br}} \alpha_k B_s^{[k]}(q) \prod_{j=1, j \neq k}^{n_{br}} A_s^{[j]}(q)}{\prod_{k=1}^{n_{br}} A_s^{[k]}(q)}, \tag{6.13}
\]

where

\[
S[k](q) = \frac{B_s^{[k]}(q)}{A_s^{[k]}(q)}. \tag{6.14}
\]
This means that under Assumption 6.2, the poles of the identified BLA are also the poles of the LTI blocks that are present in the system. The zeros of the BLA of a parallel Hammerstein system may change when the amplitude, power spectrum, or the offset (DC value) of the input signal changes. Indeed, the denominator stays the same when the gains $\alpha_k$ change, but the numerator coefficients depend on the gains $\alpha_k$.

6.5 Identification

The proposed approach starts with an estimation of the BLA of the considered system for different operating conditions. The different operating conditions are obtained using input signals with different power spectra. This includes the use of different magnitudes, different offsets, or different coloring of the power spectra. This results in the consistent estimate of the overall dynamics of the nonlinear parallel Hammerstein system.

Next, the measured BLAs are parametrized using a different LTI model for each operating condition. A common denominator model is used for all operating conditions simultaneously. This is indeed possible, as Theorem 6.1 assures that the poles of the different measured BLAs are the same.

Starting from the parametrized BLAs, a decomposition of the overall dynamics at the different operating conditions is calculated. It uses the singular value decomposition (SVD) of a matrix constructed using the numerator coefficients of the parametrized BLAs obtained at the different operating conditions. This step results in an estimate of the number of branches that is present in the parallel Hammerstein system. The number is obtained based on the estimated rank of the decomposed matrix. The dynamics $S^{[k]}(q)$ that are present in each branch are estimated next.

6.5.1 Estimating a BLA for Different Operating Conditions

The nonparametric FRF estimate of the BLA at operating condition $i_r$ is labeled $\hat{G}^{[r]}_{bla}(e^{j\omega T_s})$. It is obtained by the robust BLA estimation method
6.5. Identification

(see Section 2.3.1). Both the FRF and the sample variance \( \hat{\sigma}^2_G[\text{ir}_k] \) of the BLA are obtained at each excited frequency. The latter is used to determine the weighting factor for the parametric model estimation of the BLA. This process is explained in more detail in Pintelon and Schoukens [2012]; Schoukens et al. [2012a, 2013b].

The distribution of the operating conditions, or setpoints, need to be chosen such that the system behavior over the different setpoints is sufficiently large. This requires some insight into the system behavior. An exploratory scan over many different setpoints can be performed if such insight is not present. The number of setpoints needs to be chosen such that there are more BLAs measured than the number of parallel branches in the system. The number of parallel branches in the system is not always known beforehand. It is possible to measure the BLA at more setpoints if it turns out that the set of measurements is not rich enough.

The number of realizations and periods can vary over the different setpoints. One can measure more realizations for the setpoints where the BLA behaves more nonlinear.

6.5.2 Parametrizing the BLAs

The measured nonparametric BLAs \( \hat{G}_{\text{bla}}^{[\text{ir}]}(e^{j\omega_T}) \) at the \( R \) different operating conditions are parametrized simultaneously using a common denominator model. To perform the estimation, a weighted total least squares initialization is used (Pintelon et al. [1998]). It is followed by a sample maximum likelihood estimation (Pintelon et al. [2011b]; Pintelon and Schoukens [2012]). The frequency dependent estimation weights for the FRF are inversely proportional to the estimated sample variances \( \hat{\sigma}^2_G[\text{ir}_k] \) of the BLAs for the \( R \) different operating conditions of the system. This results in a parametrized version of the different BLAs \( \hat{G}_{\text{bla}}^{[\text{ir}]}(q, \hat{\theta}_{\text{bla}}) \):

\[
\hat{G}_{\text{bla}}^{[\text{ir}]}(q, \hat{\theta}_{\text{bla}}) = \frac{\hat{d}_{0}^{[\text{ir}]} + d_{1}^{[\text{ir}]} q^{-1} + \ldots + d_{n_d}^{[\text{ir}]} q^{-n_d}}{\hat{c}_{0} + c_{1} q^{-1} + \ldots + c_{n_c} q^{-n_c}},
\]

where the denominator coefficients are shared by the BLAs for the different operating conditions \( i_r \), while the numerator coefficients vary with the input
6. Parallel Hammerstein

operating condition \( i_r \). \( \hat{\theta}_{bla} \) contains all the denominator coefficients \( \hat{c}_i \) and all the numerator coefficients \( \hat{d}_i^{[i_r]} \) of the BLAs for the different operating conditions \( i_r \). The model order of the parametrized BLAs can be selected using standard model structure selection techniques (see for instance Ljung [1999]).

The use of a common denominator parametrization is of key importance for this approach, it is used in the next section for the decomposition of the BLAs. It is based on the important insight that the poles of the BLA of a parallel Hammerstein system don’t move for changing Gaussian input excitations.

6.5.3 Decomposing the BLAs

The overall frequency dynamics need to be distributed over the different LTI systems that are present in the branches at the back of the parallel Hammerstein model. This section presents a decomposition of the numerator coefficients of the overall dynamics of the BLA into a set of basis vectors that describe the space spanned by the numerator vectors. These basis vectors are an estimate of the dynamics of each parallel branch.

A difference with the previous approaches in Schoukens et al. [2011]; Schoukens and Rolain [2012b] is that the numerators of the estimated BLAs are decomposed, rather than the nonparametric BLA transfer functions. This is possible and adequate since a common denominator model is used for the parametrized BLAs. This new method avoids a frequency sampling step of the parametric BLAs, and a re-parametrization of the decomposed BLA dynamics. The process is explained in the sequel of this section.

First, a matrix \( \hat{D} \) is constructed containing the stacked estimated numerator coefficients of the BLAs at the different operating conditions:

\[
\hat{D} = \begin{bmatrix}
\hat{d}_0^{[1]} & \hat{d}_1^{[1]} & \cdots & \hat{d}_{n_d}^{[1]} \\
\hat{d}_0^{[2]} & \hat{d}_1^{[2]} & \cdots & \hat{d}_{n_d}^{[2]} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{d}_0^{[R]} & \hat{d}_1^{[R]} & \cdots & \hat{d}_{n_d}^{[R]}
\end{bmatrix}
\]  

(6.16)
6.5. Identification

The underlying distortion-free version of this matrix, $D$, is of low rank. The maximum rank of the matrix $D$, for $R$, $n_d > n_{br}$ is $n_{br}$. Using eq. (6.13), one can write the $D$ matrix as:

$$D = AB,$$

\[A = \begin{bmatrix}
\alpha_1^{[1]} & \alpha_2^{[1]} & \cdots & \alpha_{n_{br}}^{[1]} \\
\alpha_1^{[2]} & \alpha_2^{[2]} & \cdots & \alpha_{n_{br}}^{[2]} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_1^{[R]} & \alpha_2^{[R]} & \cdots & \alpha_{n_{br}}^{[R]}
\end{bmatrix},\]

\[B = \begin{bmatrix}
b_0^{[1]} & b_1^{[1]} & \cdots & b_{n_d}^{[1]} \\
b_0^{[2]} & b_1^{[2]} & \cdots & b_{n_d}^{[2]} \\
\vdots & \vdots & \ddots & \vdots \\
b_0^{[n_{br}]} & b_1^{[n_{br}]} & \cdots & b_{n_d}^{[n_{br}]}
\end{bmatrix},\]

where $\alpha_k^{[i]}$ is the gain of the $k$-th branch of the $i$-th BLA, and $b_i^{[k]}$ is the $i$-th degree coefficient of the polynomial $B_k^{[k]}(q) \prod_{j=1, j \neq k}^{n_{br}} A_j^{[j]}(q)$ (see eq. (6.13)).

The matrix $B$ depends only on the dynamics that are present in the different branches of the system. The matrix $A$ depends both on the system and on the input signal.

**Assumption 6.3. Rank of $D$:** The rank of the BLA numerator matrix $D$ is equal to the number of parallel branches in the system.

The proposed identification method is based on a decomposition of the BLA over the different branches of the parallel Hammerstein model. For the method to work, this decomposition should be able to separate the dynamics of each branch. This assumes that the numerator of the dynamics of one branch ($S_k^{[k]}(q)$) is linearly independent from the numerators of the dynamics of the other branches of the parallel Hammerstein system.

The SVD of $\hat{D}$ yields an orthonormal basis for the space spanned by the $\hat{D}$-matrix:

$$\hat{D} = U_{bla} \Sigma_{bla} V_{bla}^T,$$

where the columns of $V_{bla}$ are the right singular vectors which act as an orthonormal basis for the row space of the $\hat{D}$-matrix, $\Sigma_{bla}$ is a diagonal
matrix containing the singular values, and the columns of $U_{bla}$ constitute an orthonormal basis for the left hand side space.

The column vectors in $V_{bla}$ provide an estimate of the numerator coefficients for each branch $k$:

\[
\hat{S}^{[k]}(q) = \frac{\hat{\delta}_0^{[k]} + \hat{\delta}_1^{[k]} q^{-1} + \ldots + \hat{\delta}_n_d^{[k]} q^{-n_d}}{\hat{c}_0 + \hat{c}_1 q^{-1} + \ldots + \hat{c}_n_c q^{-n_c}},
\]

(6.21)

where $\hat{\delta}_j^{[k]}$ is the element of the $j$-th row and $k$-th column of the matrix $V_{bla}$.

The rank of the matrix $D$ corresponds to the number of parallel branches $n_{br}$ that is necessary to describe the system by Assumption 6.3. This rank can be obtained by applying a rank estimation algorithm on the singular value matrix $\Sigma_{bla}$ (Rolain et al. [1997]), that is obtained from the noisy matrix $\hat{D}$. To do so, the column covariance matrix $C_D$ of $\hat{D}$ is needed. This column covariance matrix is obtained from the covariance of the parameters estimated in the BLA parametrization step. The whitened matrix $D_{white}$ is given by:

\[
D_{white} = \hat{D} C_D^{-1/2}.
\]

(6.22)

The estimated rank of the noisy matrix $\hat{D}$ corresponds to the number of singular values of $D_{white}$ that are higher than 1 (Rolain et al. [1997]). The reader is referred to Rolain et al. [1997] for more details about the rank estimation method and its hypotheses.

### 6.5.4 Estimating the Static Nonlinearity

After modeling the linear dynamic part, a model for the static nonlinearity is needed. The static nonlinearities $f^{[k]}(u(t))$ are modeled by a weighted sum of nonlinear basis functions (see eq. (6.4)). The estimation of the static nonlinearity when it is expanded in the basis $f_i(u(t))$ is an estimation
problem that is linear in the parameters $\gamma_i$:

$$y(t) = \sum_{k=1}^{n_{br}} S^{[k]}(q)f^{[k]}(u(t)), \quad (6.23)$$

$$= \sum_{k=1}^{n_{br}} \sum_{i=1}^{n_f} \gamma_i S^{[k]}(q)f_i(u(t)). \quad (6.24)$$

The regressor matrix $K \in \mathbb{R}^{N \times n_f n_{br}}$ is constructed based on the previously estimated LTI blocks $\hat{S}^{[k]}(q)$:

$$K = \begin{bmatrix} K^{[1]} & K^{[2]} & \ldots & K^{[n_{br}]} \end{bmatrix}, \quad (6.25)$$

where $K^{[k]} \in \mathbb{R}^{N \times n_f}$ is given by:

$$K^{[k]} = \begin{bmatrix} \hat{S}^{[k]}(q)f_1(u(1)) & \hat{S}^{[k]}(q)f_2(u(1)) & \ldots & \hat{S}^{[k]}(q)f_{n_f}(u(1)) \\ \hat{S}^{[k]}(q)f_1(u(2)) & \hat{S}^{[k]}(q)f_2(u(2)) & \ldots & \hat{S}^{[k]}(q)f_{n_f}(u(2)) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{S}^{[k]}(q)f_1(u(N)) & \hat{S}^{[k]}(q)f_2(u(N)) & \ldots & \hat{S}^{[k]}(q)f_{n_f}(u(N)) \end{bmatrix}. \quad (6.26)$$

The least squares solution for the coefficient $\hat{\gamma}_i$ is given by:

$$\hat{\gamma} = (K^T K)^{-1} K^T y, \quad (6.27)$$

$$y = [y(1) \ y(2) \ \ldots \ \ y(N)]^T, \quad (6.28)$$

$$\hat{\gamma} = \begin{bmatrix} \hat{\gamma}_1^{[1]} & \hat{\gamma}_2^{[1]} & \ldots & \hat{\gamma}_1^{[n_{br}]} & \hat{\gamma}_2^{[n_{br}]} & \ldots & \hat{\gamma}_1^{[n_f]} & \ldots & \hat{\gamma}_n^{[n_f]} \end{bmatrix}^T. \quad (6.29)$$

In practice, the solution is obtained using a QR decomposition of $K$. To improve the numerical conditioning of the matrix, the columns of $K$ are normalized. Each column is therefore divided by its $l^2$-norm.

### 6.5.5 Persistence of Excitation

**Assumption 6.4. Persistence of excitation:** The input signal $u(t)$ is assumed to be persistently exciting for the system.

The assumption that the excitation is persistent is a very common assumption in system identification. This section discusses what persistence of excitation means for the proposed identification procedure.
The first step in the identification algorithm is to identify the parametric BLA of the nonlinear parallel Hammerstein system. It is important that the BLA identifies the dynamics that are present in the system correctly. Therefore, the number of excited frequencies in the input signal $u(t)$ needs be at least equal or preferably higher than $\frac{n_d n_c + 1}{2}$.

The MIMO static nonlinearity needs also to be estimated. For this identification step to work, the matrix $K$ in eq. (6.25) needs to be of full rank. Put in other words, the nonlinear basis functions $f_i(u(t))$ need to be linearly independent over the domain of the input signal $u(t)$. Consequently, the range of amplitudes that is present in $u(t)$ needs to be sufficiently large.

Furthermore, the rank constraint of Assumption 6.3 does not only have consequences for the system. It also determines the choice of the different setpoints of the input signals. These setpoints are chosen to ensure that the rank of the matrix $D$ is equal to $n_{br}$.

### 6.5.6 Consistency

To model the system exactly, the system needs to be contained in the reachable model set. Otherwise model errors will be present and the estimated parameters will not converge to the true parameters of the system. This requirement is formalized in the following assumption.

**Assumption 6.5. Model set:** The exact model of the system is contained in the reachable model set.

**Theorem 6.2.** The proposed estimator is a consistent (with probability 1 for $N \to \infty$) estimator of the class of parallel Hammerstein systems under the following assumptions:

- the common assumption of zero-mean additive noise at the system output only is fulfilled (Assumption 2.2),

- the excitation signal $u(t)$ is assumed to belong to the Riemann equivalence class of asymptotically normally distributed signals (Assumption 6.1),
• the input signal $u(t)$ is persistently exciting the system (Assumption 6.4),

• the system is contained in the reachable model set (Assumption 6.5). This is needed for the estimated parameters to converge to the true parameters of the system up to the degenerations that are present in the model representation (see Section 6.3),

• all the dynamics that are present in the system are also present in the BLA of that system (Assumption 6.2).

Proof. Due to Assumptions 6.2 and 6.4 and Theorems 6.1 and 2.1, the poles of the BLA are a consistent estimate of the poles of the dynamics that are present in each branch. Hence, the matrix $\hat{D}$ defined in eq. (6.16) is of low rank for a model in the reachable model class (Assumption 6.5). The rank of the matrix $\hat{D}$ is a consistent estimate of the number of parallel branches that is present in the system (see eq. (6.17) and Assumption 6.3). The columns of the matrix $V_{bla}$ that correspond to the significant singular values are a consistent estimate for the numerators of the BLA. Hence the zeros and poles that are present in each branch are estimated consistently, up to the degeneration of the model structure that is explained in Section 6.3.

In the last step of the estimation algorithm, the static nonlinearity of each branch is estimated using eq. (6.27). This estimation problem is linear in the parameters, and it is solved with a linear least squares approach. The estimate of the static nonlinearity is consistent under Assumptions 2.2 and 6.4.

The estimated parameters are consistent and converge to the true parameters under Assumption 6.5 up to the degenerations of the model structure as explained in Section 6.3.

Remark 6.1. It has been observed in practice that the rank determination still works well for small values of $R$ (smaller than $n_d$, larger than $n_{br}$) and a finite number of samples $N$ and realizations $M$. 

139
6.5.7 Nonlinear Optimization

Joining all the previous estimation steps allows one to obtain the model parameter vector as a the estimation of a partition of the parameter vector $\theta$ in non-overlapping subsets as is explained in Section 3.5.6. Note that the parameter vector $\theta \in \mathbb{R}^{(n_{br}+n_{d}+n_{f}+1)+n_{c}+1} \times 1$ now contains all the parameters of the model: $\gamma$, $\delta$, and $c$. The Jacobian of the cost function is rank deficient due to the identifiability issues (see Section 6.3). The related degenerations $(n_{br} + n_{br}^2)$ in the Jacobian need to be taken into account during the optimization.

6.6 Simulation Example

6.6.1 System and Signals

The proposed parallel Hammerstein identification algorithm is illustrated on a simulation example. The simulated system is a 2-branch parallel Hammerstein system. The first LTI block $S[1](q)$ is a third order Chebychev type 1 filter with a 5dB in-band ripple and a cut-off frequency at $0.08 f_s$, the static nonlinearity $f[1](u)$ is a polynomial of degree 5:

$$S[1](q) = \frac{0.0029 + 0.0087q^{-1} + 0.0087q^{-2} + 0.0029q^{-3}}{1 - 2.5655q^{-1} + 2.3439q^{-2} - 0.7553q^{-3}}.$$  \hspace{1cm} (6.30)

$$f[1](u) = u + 0.05u^2 - 0.1u^3 + 0.01u^4 - 0.002u^5.$$  \hspace{1cm} (6.31)

The second LTI block $S[2](q)$ is a third order Chebychev type 2 filter with a 50dB attenuation at 0.25$f_s$, the static nonlinearity $f[2](u)$ is a polynomial of degree 5:

$$S[2](q) = \frac{0.0029 + 0.0087q^{-1} + 0.0087q^{-2} + 0.0029q^{-3}}{1 - 2.5655q^{-1} + 2.3439q^{-2} - 0.7553q^{-3}}.$$  \hspace{1cm} (6.32)

$$f[2](u) = u + 0.01u^2 + 0.05u^3 + 0.005u^4 - 0.001u^5.$$  \hspace{1cm} (6.33)

The static nonlinearities are visualized in Figure 6.3.

White additive Gaussian noise with a standard deviation of 0.0001 is added to the output of the system.
6.6. Simulation Example

Figure 6.3: The true static nonlinearities of the first (green) and second (blue) branch of the parallel Hammerstein system.

A random phase multisine is used as an input signal. 10 realizations at 5 different amplitudes are applied to the system. Each realization is measured during 2 steady state periods, one period consists of 4096 samples. All frequencies in the range $]0, f_s/2[$ are excited. The random phases are uniformly distributed between $[0, 2\pi[$. The rms-value of the input signals is logarithmically spaced from 0.1 to 2. This results in a signal-to-noise ratio at the output ranging from 55 dB to 85 dB depending on the input amplitude.

The system is sampled at a sampling frequency $f_s$ of 4096 Hz.

6.6.2 Model Estimation

The BLAs of the system at the different amplitudes are estimated in a first identification step using the robust method (see Section 2.3.1) followed by a common denominator parametrization of the nonparametric BLA (see Section 6.5.2). The estimated BLAs are shown in Figure 6.4. A clear change in the shape of the BLAs for changing input rms-values can be observed. This information is used in the second step to obtain an estimate of the dynamics that are present in each branch. Next, the static nonlinearity of
each branch is estimated. Finally, the parameters are refined in an nonlinear optimization step, optimizing all the parameters simultaneously.

The estimated model is validated using a different realization of the random phase multisine with the lowest and the highest rms-value of the input. These realizations have not been used during the estimation. The time domain representation of the measured validation output and the residual error ($\hat{y}(t) - y(t)$) are shown in Figure 6.5 and Figure 6.6. The frequency domain representation of the validation output and the magnitude of the complex error between the measurement and the modeled output are shown in Figure 6.7. It is clear from Figure 6.7 that the output of the parallel Hammerstein system is very nonlinear for the highest input rms-value. The output spectrum and the total distortion variance for the highest rms-value of the input are coinciding. The system behaves quite linear for the lowest rms-value of the input: there is a 40 dB gap between the output spectrum and the total distortion level.

In both cases the initial model estimate results in residual errors which are well below the output spectrum (about 20 and 40 dB for the lowest and the highest excitation level respectively). The residual errors are almost
coinciding with the total distortion level for the lowest excitation level, and are well below the total distortion level for the highest excitation level. This shows that the initial model estimate models the nonlinear behavior of the device over a broad range of the input excitation level. After optimization the residual error coincides with the estimated noise distortion level for both the low and high input excitation levels.

6.7 Conclusion and Discussion

This chapter introduces a simple identification procedure for parallel Hammerstein systems. First the BLA of the system is estimated at different operating points of the system (e.g. power level of the input signal). Next, the measured BLAs are decomposed to obtain an estimate of the linear dynamics that are present in each branch of the model. Finally, the static nonlinearity is estimated using a linear-in-the-parameters static nonlinearity model. The proposed estimator is consistent with probability 1 for the number of data points $N \to \infty$. To decrease the variance of the estimated parameters when a finite amount of data is used, one can fine-tune all the parameters simultaneously in a final nonlinear-in-the-parameters estimation step. The proposed method is illustrated on a simulation example.

The advantage of the presented approach is that the number of parallel branches in the model does not depend on the number of basis functions that is used to model the nonlinear behavior of the system (such as in Gadringer et al. [2007]) but rather on the number of parallel signal paths that is identified to be present in the system.
Figure 6.5: Time domain validation of the parallel Hammerstein model for a low input rms-value (a) and a high input rms-value (b). The noisy system output is depicted by the full blue line, the output of the optimized model is depicted by the dashed green line.
Figure 6.6: Time domain validation of the parallel Hammerstein model before and after optimization for a low input rms-value (a) and a high input rms-value (b). The model error before optimization is represented by the red line. The model error after optimization is represented by the green line.
Figure 6.7: Frequency domain validation of the parallel Hammerstein model before and after optimization for a low input rms-value (a) and a high input rms-value (b). The noisy system output is depicted by the full blue line. The model error before optimization is represented by the red dots. The model error after optimization is represented by the green dots. The variance of the total distortion that is present at the output is shown by top black dots, the noise distortion is represented by the bottom black dots.
Chapter 7

Parallel Wiener

7.1 Introduction

To improve the flexibility of the Wiener model structure, parallel cascades can be introduced. This allows to model power dependent dynamics in the system. Even more, it is shown that parallel Wiener models can approximate all fading memory systems arbitrarily well (Boyd and Chua [1985]) if the number of branches is sufficiently high. This chapter introduces two new methods to estimate the parameters of a parallel Wiener model starting from measured input-output data only.

A parallel cascade Wiener system applies the input signal $u(t)$ to different LTI blocks in parallel. These different filtered versions of $u(t)$ are transformed to the output of the parallel Wiener system by a multiple-input-single-output static nonlinearity. A three-branch parallel cascade Wiener system is shown in Figure 7.1. The identification of the LTI blocks in a parallel Wiener system is not trivial due to the static nonlinear combination of the outputs $x_k(t)$ of all the LTI subsystems that is present in the output $y(t)$ of the nonlinear system.

Previously proposed parallel Wiener estimation methods suffer from some disadvantages. Some methods rely on an estimate of the Volterra kernel of the system under test (Kibangou and Favier [2009]). This requires a very large amount of data for the identification. Other methods are limited
to the use of finite impulse response models for the linear subsystems (Lyzell and Enqvist [2012]; Westwick and Kearney [1997]; Korenberg [1991]), or allow for infinite impulse response filters (Aljamaan et al. [2014]) but rely on the user to select these filters based on prior knowledge. These approaches typically results in an unwanted high number of parallel branches.

A first method is presented in Section 7.5 based on the results that are presented in Schoukens and Rolain [2012b] and Schoukens et al. [2015c]. The proposed method consists of a 3 step approach to obtain initial estimates. First, the nonlinearity needs to be uncovered. Thereto, different linearized models for different power levels of the input of the nonlinear system are estimated. This requires experiments to be conducted at different levels of the input signal. To estimate the FRF of the linear models, the parametric BLA will be used. Next, initial estimates of the dynamics of the LTI blocks of the parallel Wiener model are obtained by performing a decomposition of the joint frequency dynamics of the estimated BLAs. Third, the MISO static nonlinearity needs to be estimated. Therefore, a set of basis functions is chosen to model the nonlinearity, for instance multivariate polynomials, of which the coefficients need to be estimated.

The method proposed in Section 7.6 combines the advantages of the best linear approximation (BLA) of a nonlinear system (Schoukens et al. [1998]; Enqvist and Ljung [2005]) and dimension reduction techniques (Lyzell et al. [2012]; Xia et al. [2002]) to estimate a parsimonious parallel Wiener model (based on Schoukens et al. [2013a]). The advantages of this approach are the small amount of data that is needed to estimate a good quality model, and the ability to estimate an infinite impulse response model for the LTI blocks. This results in a low number of parallel branches in the model, as dimension reduction techniques are used to select a set of dominant branches.

The main contributions of these methods are:

1. The number of branches is independent of the degree of nonlinearity, and they possibly possess a physical interpretation.
2. The number of branches can be selected based on the results obtained
by the identification method, leading to parsimony in the branch count.

3. The LTI blocks can be modeled using rational models, leading to parsimony in the LTI parametrization.

4. This chapter focuses on the discrete time case, but the method that is presented here can be extended for the continuous time case.

The two methods presented in this chapter assume that the input is elliptically distributed (McGraw and Wagner [1968]; Enqvist [2005]), which is a disadvantage compared to some of the distribution free methods (Lyzell et al. [2012]; Tiels and Schoukens [2011]). On the other hand, the method presented in Tiels and Schoukens [2011] suffers from the presence of an high number of parallel branches in the estimated model. The method proposed in Lyzell et al. [2012] uses a similar dimension reduction approach as is presented in this chapter, but starts from different delayed versions of the input. This results in a higher number of parallel branches before the dimension reduction, compared to the proposed approach.

The problem of decoupling multivariate polynomials is tackled in Section 7.7. Multivariate polynomials are often used to model nonlinear behavior, for instance in parallel Wiener models. These multivariate polynomials are mostly hard to interpret, due to the presence of crossterms. These polynomials also have a high amount of coefficients, and calculation of an inverse of a multivariate polynomial with crossterms is cumbersome. Section 7.7 proposes a method to eliminate the crossterms of a multivariate polynomial using a linear input transformation. It is shown how every homogeneous polynomial described using tensors can be transformed to a canonical form using multilinear algebraic decomposition methods. Such tensor decomposition methods have already been used in nonlinear system modeling to reduce the complexity of Volterra models (Favier and Bouilloc [2009]; Nowak and Van Veen [1996]). Since every polynomial can be written as a sum of homogeneous polynomials, this method results in a decoupled description of any multivariate polynomial, allowing for a model description that is easier to interpret, easier to use in a design and easier to invert.
7. Parallel Wiener

7.2 System

Figure 7.1: A parallel Wiener system with 3 parallel branches. $H^{[k]}(q)$ is the LTI system associated to branch $k$, $f(x_1, x_2, x_3)$ is a MISO static nonlinearity.

A parallel Wiener system is a generalized parallel connection of a several Wiener subsystems as shown in Figure 7.1. Instead of one SISO static nonlinearity for each branch, one MISO nonlinearity is considered in a parallel Wiener system. The noiseless output of a parallel Wiener system is given by:

$$y_0(t) = f(x_1(t), \ldots, x_{n_{br}}(t)),$$
$$x_k(t) = H^{[k]}(q) u(t),$$

where $f(.)$ is a MISO static nonlinear function and $H^{[k]}(q)$ is the LTI system associated to branch $k$. The number of parallel branches is given by $n_{br}$. The signals $y_0(t)$, $x_k(t)$ and $u(t)$ are scalar signals.

The static nonlinearity $f(.)$ is assumed to be obtained as a linear combination of basis functions (e.g. polynomial):

$$f(x_1(t), \ldots, x_{n_{br}}(t)) = \sum_{i=1}^{n_f} \gamma_i f_i(x_1(t), \ldots, x_{n_{br}}(t)),$$

where the coefficients $\gamma_i$ are grouped in the coefficient vector $\gamma \in \mathbb{R}^{n_f \times 1}$, and $n_f$ is the total number of nonlinear basis functions.
7.3 Identifiability of a Parallel Wiener System

The LTI subsystem $H[k](q)$ is assumed to be a rational function of the backward shift operator $q^{-1}$:

$$H[k](q) = \frac{b_0[k] + b_1[k]q^{-1} + \cdots + b_{nb,k}[k]q^{-nb,k}}{a_0[k] + a_1[k]q^{-1} + \cdots + a_{na,k}[k]q^{-na,k}},$$

(7.4)

where the coefficients $b_i[k]$ and $a_i[k]$ are grouped in the coefficient vectors $b[k] \in \mathbb{R}^{(nb,k+1) \times 1}$ and $a[k] \in \mathbb{R}^{(na,k+1) \times 1}$. The $l^2$-norm of the coefficient vector $a[k]$ is assumed to be fixed to a value that is equal to 1, and the first nonzero element of $a[k]$ is assumed to be positive to compensate for the degeneration that is present in the rational transfer function parametrization.

7.3 Identifiability of a Parallel Wiener System

The LTI blocks and the static nonlinear blocks contained in a parallel Wiener system can only be identified up to an unknown gain, unless one has access to the intermediate signals $x_k(t)$. If a gain $\beta_k$ is introduced for both $H[k](q)$ and $f(.)$ one obtains:

$$y_0(t) = \tilde{f}(x_1(t), \ldots, \tilde{x}_k(t), \ldots, x_{nb'},(t)),$$

$$= f(x_1(t), \ldots, \frac{1}{\beta_k} x_k(t), \ldots, x_{nb'},(t)),$$

$$\tilde{x}_k(t) = \tilde{H}^{[k]}(q)u(t) = \beta_k H[k](q)u(t).$$

(7.5)

(7.6)

(7.7)

The input-output behavior of a parallel Wiener system that is composed using $\tilde{H}^{[k]}(q)$ and $\tilde{f}(.)$ is mathematically exactly the same as a parallel Wiener system containing $H[k](q)$ and $f(.)$ as building blocks.

An additional identifiability issue appears here due to the parallel nature of the parallel Wiener systems. Starting from input-output data only, infinitely many equivalent models can be obtained by adding a regular linear transformation to one of the models. This introduces a full rank linear transformation matrix between the inputs of the static nonlinearity $f(.)$ and the outputs of the LTI blocks $H[k](q)$. This full rank linear transformation is described by a matrix $T \in \mathbb{R}^{nb' \times nb'}$. Since $T$ is regular its inverse
Parallel Wiener

Figure 7.2: A parallel Wiener system with 3 parallel branches subject to a linear transformation $T$.

Transformation $T^{-1}$ can also be included in between the LTI blocks and the static nonlinear block as is illustrated in Figure 7.2. This results in a new set of LTI blocks and static nonlinear blocks that are a linear combination of the original ones:

$$\tilde{H}[k](q) = \sum_{i=1}^{n_{br}} t_{k,i} H[i](q), \quad (7.8)$$

$$\tilde{f}(\tilde{x}_1(t), \ldots, \tilde{x}_{n_{br}}(t)) = f(x_1(t), \ldots, x_{n_{br}}(t)), \quad (7.9)$$

$$\tilde{x}_k(t) = \sum_{i=1}^{n_{br}} t_{k,i} x_i(t), \quad (7.10)$$

where

$$T = \begin{bmatrix} t_{1,1} & \cdots & t_{1,n_{br}} \\ \vdots & \ddots & \vdots \\ t_{n_{br},1} & \cdots & t_{n_{br},n_{br}} \end{bmatrix}, \quad (7.11)$$

and

$$\tilde{T} = T^{-1} = \begin{bmatrix} \tilde{t}_{1,1} & \cdots & \tilde{t}_{1,n_{br}} \\ \vdots & \ddots & \vdots \\ \tilde{t}_{n_{br},1} & \cdots & \tilde{t}_{n_{br},n_{br}} \end{bmatrix}. \quad (7.12)$$

The input-output behavior of a parallel Wiener system composed by the transformed $\tilde{H}[k](q)$ and $\tilde{f}(.)$ is exactly the same as that of a parallel Wiener system containing $H[k](q)$ and $f(.)$ as building blocks.

Although theoretically the gain exchanges $\beta_k$ and the transformation $T$ result in the same input-output behavior of the system, bad choices of
these gains $\beta_k$ and the transformation $T$ can lead to numerical issues. It is best to avoid a high dynamic range over the different parameters that are present in the model. Also a high dynamic range over the different signals $(u(t), x_k(t), y(t))$ should be avoided if possible.

### 7.4 Best Linear Approximation

This section shows that the BLA of a parallel Wiener system is a linear combination of the LTI dynamics that are present in the different LTI blocks $H^{[k]}(q)$ when input signals belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals are used as an excitation signal.

**Assumption 7.1. Input signal:** The input signal belongs to the Riemann equivalence class of asymptotically normally distributed excitation signals as defined in Definition 2.1.

Furthermore, it is assumed that the static nonlinearity of the parallel Wiener system can be represented by a multivariate polynomial. This is needed to obtain an analytic expression of the BLA of a parallel Wiener system excited by an input signal belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals.

**Assumption 7.2. MISO nonlinearity:** The MISO static nonlinearity $f(.)$ of the parallel Wiener system can be approximated arbitrarily well in least squares sense by a multivariate polynomial of finite degree.

**Theorem 7.1.** The BLA $G_{bla}(e^{j\omega T_s})$ of a parallel Wiener system excited by an input signal belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals (Assumption 7.1) and with a MISO static nonlinearity $f(.)$ of the parallel Wiener system that can be approximated arbitrarily well in least squares sense by a MISO polynomial (Assumption 7.2) is given by:

$$G_{bla}(e^{j\omega T_s}) = \sum_{k=1}^{n_{br}} \alpha_k H^{[k]}(e^{j\omega T_s}), \quad \alpha_k \in \mathbb{R}. \quad (7.13)$$
Proof. The static nonlinearity of the parallel Wiener system can be represented by a multivariate polynomial (Assumption 7.2). Any multivariate polynomial can be decoupled as a sum of univariate polynomials using a linear input transformation $T_f$ (Tiels and Schoukens [2013]). Note that the number of inputs of the decoupled static nonlinearity can be higher than the original coupled static nonlinearity. This results in the following output equations:

\[ y(t) = \sum_{k=1}^{n_{\text{dec}}} y_k(t), \quad (7.14) \]
\[ y_k(t) = \hat{f}^{[k]}(\hat{x}_k(t)), \quad (7.15) \]
\[ \hat{x}_k(t) = \hat{H}^{[k]}(q) u(t), \quad (7.16) \]

where:

\[ \hat{x}(t) = T_f x(t), \quad (7.17) \]

where $T_f \in \mathbb{R}^{n_{\text{dec}} \times n_{\text{br}}}$, and $\hat{x}(t)$ and $x(t)$ are given by:

\[ \hat{x}(t) = [\hat{x}_1(t) \ldots \hat{x}_k(t) \ldots \hat{x}_{n_{\text{dec}}}(t)]^T, \quad (7.18) \]
\[ x(t) = [x_1(t) \ldots x_k(t) \ldots x_{n_{\text{br}}}(t)]^T, \quad (7.19) \]

where $n_{\text{dec}}$ is the number of branches in the decoupled parallel Wiener representation. In other words:

\[ \hat{x}_k(t) = \sum_{m=1}^{n_{\text{br}}} t_{f,km} x_m(t) \quad (7.20) \]
\[ = \sum_{m=1}^{n_{\text{br}}} t_{f,km} H^{[m]}(q) u(t), \quad (7.21) \]

where $t_{f,km}$ is the element on the $k$-th row and $m$-th column of matrix $T_f$.

The BLA of a Wiener system is given by (3.15):

\[ G_{\text{bla}}(e^{j\omega T_s}) = \alpha H(e^{j\omega T_s}). \]

The output $y$ is the sum of the outputs $y_k$ of each decoupled parallel Wiener branch. Hence, the BLA of a parallel Wiener system is given by the sum
of the BLA of each parallel branch:

\[
G_{bla}(e^{j\omega T_s}) = \sum_{k=1}^{n_{dec}} \hat{\alpha}_k \hat{H}[k](e^{j\omega T_s}),
\]

(7.22)

\[
= \sum_{k=1}^{n_{dec}} \hat{\alpha}_k \sum_{m=1}^{n_{br}} t_{f,km} H_m(e^{j\omega T_s}),
\]

(7.23)

\[
= \sum_{m=1}^{n_{br}} \alpha_m H[m](e^{j\omega T_s}).
\]

(7.24)

Note that the shape of \( G_{bla}(e^{j\omega T_s}) \) depends on the power spectrum of the input signal. The shape of \( G_{bla}(e^{j\omega T_s}) \) can change if a class of input signals with a different power spectrum is used. The coefficients \( \alpha_k \) depend on many factors: the LTI subsystem \( H[k](q) \), the static nonlinear subsystems \( f[k](u(t)) \), the power spectrum \( S_{UU}(e^{j\omega T_s}) \) of the input signal \( u(t) \), and hence as well on the variance of the input signal \( u(t) \).

**Assumption 7.3. Non-zero gain:** The gains \( \alpha_k \) present in eq. (7.13) are non-zero.

It can happen that the BLA of one of the branches of the parallel Wiener system is equal to zero, or in other words \( \alpha_k = 0 \). This is the case when the nonlinear function \( f(.) \) is even around the expected value of \( x_k(t) \), given the other input signals \( x_1(t), \ldots, x_{n_{br}}(t) \). In this case, a BLA of a reduced order is obtained that does not contain the dynamics of branch \( k \). Different DC setpoints can be tried to avoid a zero gain \( \alpha_k \).

An important observation with respect to eq. (7.13) is that the input dependent gain \( \alpha_k \) only appears in the numerator:

\[
G_{bla}(q) = \frac{\sum_{k=1}^{n_{br}} \alpha_k B_h^{[k]}(q) \prod_{j=1,j\neq k}^{n_{br}} A_h^{[j]}(q)}{\prod_{k=1}^{n_{br}} A_h^{[k]}(q)},
\]

(7.25)

where

\[
H[k](q) = \frac{B_h^{[k]}(q)}{A_h^{[k]}(q)},
\]

(7.26)
This means that under Assumption 7.3, the poles of the identified BLA are also the poles of the LTI blocks that are present in the system. The zeros of the BLA of a parallel Wiener system may change when the amplitude, power spectrum, or the offset (DC value) of the input signal changes. Indeed, the denominator stays the same when the gains $\alpha_k$ change, while the numerator coefficients depend on the gains $\alpha_k$.

### 7.5 Identification through BLA Decomposition

*Based on Schoukens and Rolain [2012b] and Schoukens and Rolain [2012c]*

The approach proposed in Schoukens and Rolain [2012b,c]; Schoukens et al. [2015c] starts with an estimation of the BLA of the considered system for different operating conditions. It is very similar to the approach presented in Chapter 6. The different operating conditions are obtained using input signals with different power spectra. This includes the use of different magnitudes, different offsets, or different coloring of the power spectra.

Next, the measured BLAs are parametrized using a different LTI model for each operating condition. A common denominator model is used for all operating conditions simultaneously. This is indeed possible, as Theorem 7.1 assures that the poles of the different measured BLAs are the same. A consistent estimate of the overall dynamics that are present in the nonlinear parallel Wiener system results.

Starting from the parametrized BLAs, a decomposition of the overall dynamics at the different operating conditions is calculated. It uses the singular value decomposition (SVD) of a matrix constructed using the numerator coefficients of the parametrized BLAs obtained at the different operating conditions. This enables the estimation of the number of branches that is present in the parallel Wiener system. The number is obtained based on the estimated rank of the decomposed matrix. The dynamics $H^{[k]}(q)$ that are present in each branch are estimated next.
Sections 7.5.1, 7.5.2 and 7.5.3 are identical to Sections 6.5.1, 6.5.2 and 6.5.3 presented in the parallel Hammerstein chapter.

### 7.5.1 Estimating a BLA for Different Operating Conditions

The nonparametric FRF estimate of the BLA at operating condition \( i_r \) is labeled \( \hat{G}_{bla}^{[ir]}(e^{j\omega_k T_s}) \). It is obtained by the robust BLA estimation method (see Section 2.3.1). Both the FRF and the sample variance \( \hat{\sigma}^2_{\hat{G}_{bla}^{[ir]}}(k) \) of the BLA are obtained at each excited frequency. The latter is used to determine the weighting factor for the parametric model estimation of the BLA. This process is explained in more detail in Pintelon and Schoukens [2012]; Schoukens et al. [2012a, 2013b].

### 7.5.2 Parametrizing the BLAs

The measured nonparametric BLAs \( \hat{G}_{bla}^{[ir]}(e^{j\omega_k T_s}) \) at the \( R \) different operating conditions are parametrized simultaneously using a common denominator model. To perform the estimation, a weighted total least squares initialization is used (Pintelon et al. [1998]). It is followed by a sample maximum likelihood estimation (Pintelon et al. [2011b]; Pintelon and Schoukens [2012]). The frequency dependent estimation weights for the FRF are inversely proportional to the estimated sample variances \( \hat{\sigma}^2_{\hat{G}_{bla}^{[ir]}}(k) \) of the BLAs for the \( R \) different operating conditions of the system. This results in a parametrized version of the different BLAs \( \hat{G}_{bla}^{[ir]}(q, \hat{\theta}_{bla}) \):

\[
\hat{G}_{bla}^{[ir]}(q, \hat{\theta}_{bla}) = \frac{\hat{d}_0^{[ir]} + \hat{d}_1^{[ir]} q^{-1} + \ldots + \hat{d}_{nd}^{[ir]} q^{-nd}}{\hat{c}_0 + \hat{c}_1 q^{-1} + \ldots + \hat{c}_{nc} q^{-nc}},
\]

(7.27)

where the denominator coefficients are shared by the BLAs for the different operating conditions \( i_r \), while the numerator coefficients vary with the input operating condition \( i_r \). \( \hat{\theta}_{bla} \) contains all the denominator coefficients \( \hat{c}_i \) and all the numerator coefficients \( \hat{d}_i^{[ir]} \) of the BLAs for the different operating conditions \( i_r \). The model order of the parametrized BLAs can be selected
using standard model structure selection techniques (see for instance Ljung [1999]).

7.5.3 Decomposing the BLAs

The overall frequency dynamics need to be distributed over the different LTI systems that are present in the branches at the front of the parallel Wiener model. This section presents a decomposition of the numerator coefficients of the overall dynamics of the BLA into a set of basis vectors that describe the space spanned by the numerator vectors. These basis vectors are an estimate of the dynamics of each parallel branch.

A difference with the previous approaches in Schoukens et al. [2011]; Schoukens and Rolain [2012b] is that the numerators of the estimated BLAs are decomposed, rather than the nonparametric BLA transfer functions. This is possible and adequate since a common denominator model is used for the parametrized BLAs. This new method avoids a frequency sampling step of the parametric BLAs, and a re-parametrization of the decomposed BLA dynamics. The process is explained in the sequel of this section.

First, a matrix $\hat{D}$ is constructed containing the stacked estimated numerator coefficients of the BLAs at the different operating conditions:

$$\hat{D} = \begin{bmatrix}
\hat{d}_0^{[1]} & \hat{d}_1^{[1]} & \ldots & \hat{d}_{n_d}^{[1]} \\
\hat{d}_0^{[2]} & \hat{d}_1^{[2]} & \ldots & \hat{d}_{n_d}^{[2]} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{d}_0^{[R]} & \hat{d}_1^{[R]} & \ldots & \hat{d}_{n_d}^{[R]}
\end{bmatrix}.$$  \hfill (7.28)

The underlying distortion-free version of this matrix, $D$, is assumed to be of low rank. The maximum rank of the matrix $D$, for $R, n_d > n_{br}$ is $n_{br}$.
Using eq. (7.25), one can write the $D$ matrix as:

$$D = AB, \quad (7.29)$$

$$A = \begin{bmatrix} \alpha_1^{[1]} & \alpha_2^{[1]} & \ldots & \alpha_{n_{br}}^{[1]} \\ \alpha_1^{[2]} & \alpha_2^{[2]} & \ldots & \alpha_{n_{br}}^{[2]} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{[R]} & \alpha_2^{[R]} & \ldots & \alpha_{n_{br}}^{[R]} \end{bmatrix}, \quad (7.30)$$

$$B = \begin{bmatrix} b_0^{[1]} & b_1^{[1]} & \ldots & b_{n_d}^{[1]} \\ b_0^{[2]} & b_1^{[2]} & \ldots & b_{n_d}^{[2]} \\ \vdots & \vdots & \ddots & \vdots \\ b_0^{[n_{br}]} & b_1^{[n_{br}]} & \ldots & b_{n_d}^{[n_{br}]} \end{bmatrix}, \quad (7.31)$$

where $\alpha_k^{[i]}$ is the gain of the $k$-th branch of the $i$-th BLA, and $b_i^{[k]}$ is the $i$-th degree coefficient of the polynomial $B_s^{[k]}(q) \prod_{j=1, j \neq k}^{n_{br}} A_s^{[j]}(q)$ (see eq. (7.25)).

The matrix $B$ depends only on the dynamics that are present in the different branches of the system. The matrix $A$ on the other hand depends both on the system and on the input signal.

**Assumption 7.4. Rank of $D$:** The rank of the BLA numerator matrix $D$ is equal to the number of parallel branches in the system.

The proposed identification method is based on a decomposition of the BLA over the different branches of the parallel Wiener model. For the method to work, this decomposition should be able to separate the dynamics of each branch. This assumes that the numerator of the dynamics of one branch ($H^{[k]}(q)$) is linearly independent from the numerators of the dynamics of the other branches of the parallel Wiener system.

The SVD of $\hat{D}$ yields an orthonormal basis for the space spanned by the $D$-matrix:

$$\hat{D} = U_{bla} \Sigma_{bla} V_{bla}^T, \quad (7.32)$$

where the columns of $V_{bla}$ are the right singular vectors which act as an orthonormal basis for the row space of the $D$-matrix, $\Sigma_{bla}$ is a diagonal matrix containing the singular values, and the columns of $U_{bla}$ constitute an orthonormal basis for the left hand side space.
7. Parallel Wiener

The column vectors in $V_{bla}$ provide an estimate of the numerator coefficients for each branch $k$:

$$\hat{H}^{[k]}(q) = \frac{\hat{\delta}_0^{[k]} + \hat{\delta}_1^{[k]} q^{-1} + \ldots + \hat{\delta}_n^{[k]} q^{-n_d}}{\hat{c}_0 + \hat{c}_1 q^{-1} + \ldots + \hat{c}_{n_c} q^{-n_c}},$$  

(7.33)

where $\hat{\delta}_j^{[k]}$ is the element of the $j$-th row and $k$-th column of the matrix $V_{bla}$.

The rank of the matrix $D$ corresponds to the number of parallel branches $n_{br}$ that is necessary to describe the system by Assumption 7.4. This rank can be obtained by applying a rank estimation algorithm on the singular value matrix $\Sigma_{bla}$ (Rolain et al. [1997]), that is obtained from the noisy matrix $\hat{D}$. To do so, the column covariance matrix $C_D$ of $\hat{D}$ is needed. This column covariance matrix is obtained from the covariance of the parameters estimated in the BLA parametrization step. The whitened matrix $D_{white}$ is given by:

$$D_{white} = \hat{D} C_D^{-1/2}.$$  

(7.34)

The estimated rank of the noisy matrix $\hat{D}$ corresponds to the number of singular values of $D_{white}$ that are higher than 1 (Rolain et al. [1997]). The reader is referred to Rolain et al. [1997] for more details about the rank estimation method and its hypotheses.

7.5.4 Estimating the Static Nonlinearity

After modeling the linear dynamic part, a model for the static nonlinearity is needed. The static nonlinearity $f(.)$ is modeled by a weighted sum of multivariate nonlinear basis functions (see eq. (7.3)). The estimation of the static nonlinearity when it is expanded in the basis $f_i(.)$ is an estimation problem that is linear in the parameters $\gamma_i$:

$$y(t) = f(x_1(t), \ldots, x_{n_{br}}(t)), $$  

(7.35)

$$= \sum_{i=1}^{n_f} \gamma_i f_i(x_1(t), \ldots, x_{n_{br}}(t)).$$  

(7.36)
7.5. Identification through BLA Decomposition

The regressor matrix \( K \in \mathbb{R}^{N \times n_f} \) is constructed based on the previously estimated LTI blocks \( \hat{H}^{[k]}(q) \):

\[
\hat{x}_k(t) = \hat{H}^{[k]}(q)u(t). \tag{7.37}
\]

This results in the regressor matrix \( K \):

\[
K = \begin{bmatrix}
    f_1(\hat{x}_1(1), \ldots, \hat{x}_{n_{br}}(1)) & \ldots & f_{n_f}(\hat{x}_1(1), \ldots, \hat{x}_{n_{br}}(1)) \\
    f_1(\hat{x}_1(2), \ldots, \hat{x}_{n_{br}}(2)) & \ldots & f_{n_f}(\hat{x}_1(2), \ldots, \hat{x}_{n_{br}}(2)) \\
    \vdots & \ddots & \vdots \\
    f_1(\hat{x}_1(N), \ldots, \hat{x}_{n_{br}}(N)) & \ldots & f_{n_f}(\hat{x}_1(N), \ldots, \hat{x}_{n_{br}}(N))
\end{bmatrix}. \tag{7.38}
\]

The least squares solution for the coefficient \( \hat{\gamma}_i \) is given by:

\[
\hat{\gamma} = (K^TK)^{-1}K^Ty, \tag{7.39}
\]

\[
y = [y(1) \ y(2) \ \ldots \ y(N)]^T, \tag{7.40}
\]

\[
\hat{\gamma} = [\hat{\gamma}_1 \ \hat{\gamma}_2 \ \ldots \ \hat{\gamma}_{n_f}]^T. \tag{7.41}
\]

In practice, the solution is obtained using a QR decomposition of \( K \). To improve the numerical conditioning of the matrix, the columns of \( K \) are normalized. Each column is therefore divided by its \( l^2 \)-norm.

### 7.5.5 Persistence of Excitation

**Assumption 7.5. Persistence of excitation:** The input signal \( u(t) \) is assumed to be persistently exciting for the system.

The assumption that the excitation is persistent is a very common assumption in system identification. This section discusses what persistence of excitation means for the proposed identification procedure.

The first step in the identification algorithm is to identify the parametric BLA of the nonlinear parallel Wiener system. It is important that the BLA identifies the dynamics that are present in the system correctly. Therefore, the number of excited frequencies in the input signal \( u(t) \) needs be at least equal or preferably higher than \( \frac{nd+n_{c}+1}{2} \).

The MISO static nonlinearity needs also to be estimated. For this identification step to work, the matrix \( K \) in eq. (7.38) needs to be of full rank.
Put in other words, the nonlinear basis functions $f_i(\hat{x}_1(t), ..., \hat{x}_{n_{br}}(t))$ need to be linearly independent over the domain of the intermediate signals $\hat{x}_1(t), ..., \hat{x}_{n_{br}}(t)$. Consequently, the range of amplitudes that is present in $\hat{x}_1(t), ..., \hat{x}_{n_{br}}(t)$ needs to be sufficiently large.

Furthermore, the rank constraint of Assumption 7.4 does not only have consequences for the system. It also determines the choice of the different setpoints of the input signals. These setpoints are chosen to ensure that the rank of the matrix $D$ is equal to $n_{br}$.

### 7.5.6 Consistency

To model the system exactly, the system needs to be contained in the reachable model set. Otherwise model errors will be present and the estimated parameters will not converge to the true parameters of the system. This requirement is formalized in the following assumption.

**Assumption 7.6. Model set:** The exact model of the system is contained in the reachable model set.

The consistency proof is completely analog to the proof given in Section 7.5.6 for the parallel Hammerstein case.

**Theorem 7.2.** The proposed estimator is a consistent (with probability 1 for $N \to \infty$) estimator of the class of parallel Wiener systems under the following assumptions:

- the common assumption of zero-mean additive noise at the system output only is fulfilled (Assumption 2.2),

- the excitation signal $u(t)$ is assumed to belong to the Riemann equivalence class of asymptotically normally distributed signals (Assumption 7.1),

- the input signal $u(t)$ is persistently exciting the system (Assumption 7.5),
7.5. Identification through BLA Decomposition

- the system is contained in the reachable model set (Assumption 7.6). This is needed for the estimated parameters to converge to the true parameters of the system up to the degenerations that are present in the model representation (see Section 7.3),

- all the dynamics that are present in the system are also present in the BLA of that system (Assumption 7.3).

Proof. Due to Assumptions 7.3 and 7.5 and Theorems 7.1 and 2.1, the poles of the BLA are a consistent estimate of the poles of the dynamics that are present in each branch. Hence, the matrix $\hat{D}$ defined in eq. (6.16) is of low rank for a model in the reachable model class (Assumption 7.6). The rank of the matrix $\hat{D}$ is a consistent estimate of the number of parallel branches that is present in the system (see eq. (7.29) and Assumption 7.4). The columns of the matrix $V_{bla}$ that correspond to the significant singular values are a consistent estimate for the numerators of the BLA. Hence the zeros and poles that are present in each branch are estimated consistently, up to the degeneration of the model structure that is explained in Section 7.3.

In the last step of the estimation algorithm, the static nonlinearity of each branch is estimated using eq. (7.39). This estimation problem is linear in the parameters, and it is solved with a linear least squares approach. The estimate of the static nonlinearity is consistent under Assumptions 2.2 and 7.5.

The estimated parameters are consistent and converge to the true parameters under Assumption 7.6 up to the degenerations of the model structure as explained in Section 7.3.

Remark 7.1. It has been observed that the rank determination still works well in practice for small values of $R$ (smaller than $n_d$, larger than $n_{br}$) and a finite number of samples $N$ and realizations $M$. 

163
7. Parallel Wiener

7.6 Identification through Dimension Reduction

*Based on Schoukens et al. [2013a]*

A second approach to identify parallel Wiener systems uses a combination of the BLA and dimension reduction techniques, more specifically MAVE (see Section 4.6.3.2). The BLA provides a set of basis functions that can be used to model the dynamics of the parallel Wiener system, while the MAVE algorithm is used to recombine this set of basis functions such that a parsimonious description of the system is obtained. The user can choose to use other basis functions than the ones obtained with the BLA. This provides the possibility to introduce some prior knowledge into the estimation algorithm.

The BLA estimates the transfer function of the dynamics that are present in the nonlinear system. This transfer function can be represented by $\hat{G}_{bla}(q, \hat{\theta})$ using a rational function of $q^{-1}$. For simplicity, it is assumed that no coinciding poles are present in the $\hat{G}_{bla}(q, \hat{\theta})$, and that $\hat{G}_{bla}(q, \hat{\theta})$ is proper. This rational form can be decomposed in a sum of rational forms of order 1 using the partial fraction expansion, as is expressed below:

$$\hat{G}_{bla}(q, \hat{\theta}) = \frac{\hat{b}_0 + \hat{b}_1 q^{-1} + \ldots + \hat{b}_{nb} q^{-nb}}{\hat{a}_0 + \hat{a}_1 q^{-1} + \ldots + \hat{a}_{na} q^{-na}},$$

where $\hat{\theta}$ is a constant in the case of proper systems and $\hat{r}_i$ is the residual associated to each pole $\hat{p}_i$. The terms of this partial fraction expansion can be used as a set of basis functions to reconstruct the original LTI blocks of the parallel Wiener system.

The terms of complex conjugate poles can be combined in order to have
7.6. Identification through Dimension Reduction

fewer terms:

$$\hat{G}_{bla}(q, \hat{\theta}) = \frac{\hat{r}_1}{1 - \hat{p}_1q^{-1}} + \ldots + \frac{\hat{r}_{n_r}}{1 - \hat{p}_{n_r}q^{-1}}$$

$$+ \frac{\hat{r}_{n_r+1,0} + \hat{r}_{n_r+1,1}q^{-1}}{1 + \hat{p}_{n_r+1,1}q^{-1} + \hat{p}_{n_r+1,2}q^{-2} + \ldots}$$

$$+ \frac{\hat{r}_{n_d-1,0} + \hat{r}_{n_d-1,1}q^{-1}}{1 + \hat{p}_{n_d-1,1}q^{-1} + \hat{p}_{n_d-1,2}q^{-2} + \hat{k}};$$

$$= \hat{P}_1(q) + \hat{P}_2(q) + \ldots + \hat{P}_{n_d-1}(q) + \hat{k},$$

(7.43)

where the terms 1 to \(n_r\) correspond to the basis functions associated to the real poles, and the terms \(n_r + 1\) to \(n_d - 1\) correspond to the basis functions associated to the complex conjugate poles, where \(n_d\) is previously defined in (4.41).

The regression variables \(\varphi(t)\) for the dimension reduction step can be constructed starting from the partial fraction expansion of \(\hat{G}_{bla}(q, \hat{\theta})\). Denote \(\hat{P}_i(q)\) as the \(i\)-th basis function of the partial fraction expansion. The associated regression variables are:

$$\varphi(t) = \begin{bmatrix} \hat{P}_1(q)u(t) & \ldots & \hat{P}_{n_d-1}(q)u(t) & u(t) \end{bmatrix}^T.$$  \hspace{1cm} (7.44)

This results in the matrix \(\Phi \in \mathbb{R}^{N \times n_d}\):

$$\Phi = \begin{bmatrix} \varphi(1) & \varphi(2) & \ldots & \varphi(N) \end{bmatrix}^T.$$  \hspace{1cm} (7.45)

Next, the MAVE dimension reduction method (Xia et al. [2002]; Lyzell et al. [2012]) is applied to the constructed regression matrix \(\Phi\) and the measured output \(y(t)\). This results in the matrix \(\hat{B}\), an estimate of the true \(B\) matrix as defined in (4.41). Hence, the in dimension reduced set of basis functions is given by:

$$\hat{H}(q) = \hat{B}^T \hat{P}(q),$$  \hspace{1cm} (7.46)

$$\begin{bmatrix} \hat{H}_1(q) & \ldots & \hat{H}_{n_{rd}}(q) \end{bmatrix} = \hat{B}^T \begin{bmatrix} \hat{P}_1(q) & \ldots & \hat{P}_{n_d-1}(q) & 1 \end{bmatrix},$$  \hspace{1cm} (7.47)

where \(n_{rd} < n_d\), as defined in (4.41). This low rank set of basis functions \(\hat{H}(q)\) acts as the estimate of the LTI blocks in the parallel Wiener model.

The estimate of the static nonlinearity can be obtained as is explained in Section 7.5.4.
A parallel Wiener system model contains a MISO static nonlinear block at its output (Figure 7.1). Multivariate polynomials are often used to describe this static nonlinearity (Schetzen [1980]; Boyd and Chua [1985]). These polynomials generally contain crossterms. These hamper an easy interpretation of the model, and make it hard to invert the nonlinearity. The use of the model in a design calls for cross-term elimination to obtain a decoupled description of the static nonlinearity. Multilinear algebraic methods, more specifically tensor decomposition methods (Harshman [1970]; Andersson and Bro [2000]), can transform a multivariate homogeneous polynomial into a canonical description. Tensor decomposition methods are already used in nonlinear system modeling to reduce the complexity of Volterra models (Favier and Bouilloc [2009]; Nowak and Van Veen [1996]).

The results that are presented in this section are based on Schoukens and Rolain [2012a]. These results were a first step in tackling the decoupling problem. More elegant solutions are recently presented in Tiels and Schoukens [2013]; Schoukens et al. [2014c]; Dreesen et al. [2015].

To make the method more understandable, we start with a polynomial of degree 2, generalize it to degree 3, and finally extend the method to an arbitrary degree. Only 3 inputs are considered in the following equations for the ease of notation. The results can easily be generalized to an arbitrary number of inputs.

### 7.7.1 Homogeneous Polynomial of Degree 2

If a homogeneous polynomial of degree 2 with 3 inputs \(x_1, x_2, x_3\) is considered, the polynomial \(y = f(x_1, x_2, x_3)\) is given by:

\[
y = a_{11}x_1^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{22}x_2^2 + a_{23}x_2x_3 + a_{33}x_3^2.
\]

(7.48)
7.7. Decoupling the MISO Static Nonlinearity

Or in matrix notation:

\[ y = x^T Ax, \quad (7.49) \]

with:

\[
\begin{bmatrix}
    x_1 \\
    x_2 \\
    x_3
\end{bmatrix},
\begin{bmatrix}
    a_{11} & \frac{a_{12}}{2} & \frac{a_{13}}{2} \\
    \frac{a_{12}}{2} & a_{22} & \frac{a_{23}}{2} \\
    \frac{a_{13}}{2} & \frac{a_{23}}{2} & a_{33}
\end{bmatrix}
\]

(7.50)

The problem of eliminating the crossterms of the multivariate polynomial is reduced to the problem of diagonalizing a square symmetric matrix \( A \). This reduces to an eigenvalue decomposition of \( A \) (Golub and Van Loan [1996]). Two matrices \( V \) and \( D \) result, where \( V \) is an orthogonal matrix containing the eigenvectors, and \( D \) is a diagonal matrix containing the real eigenvalues. Equation (7.49) can be rewritten as:

\[ y = x^T VDV^T x. \quad (7.51) \]

Defining \( \bar{x} = V^T x \) equation (7.51) is reduced to:

\[ y = \bar{x}^T D \bar{x} = d_{11} \bar{x}_1^2 + d_{22} \bar{x}_2^2 + d_{33} \bar{x}_3^2, \quad (7.52) \]

where the crossterms are eliminated.

**7.7.2 Homogeneous Polynomial of Degree 3**

If a homogeneous polynomial of degree 3 is considered with 3 inputs \( x_1, x_2, x_3 \), the output \( y \) of the polynomial is given by:

\[
\begin{align*}
y &= a_{111} x_1^3 + a_{112} x_1^2 x_2 + a_{113} x_1^2 x_3 + \cdots \\
&= a_{112} x_1^2 x_2 + a_{123} x_1 x_2 x_3 + a_{133} x_1 x_3^2 + \cdots \\
&= a_{222} x_2^3 + a_{223} x_2^2 x_3 + a_{233} x_2 x_3^2 + a_{333} x_3^3.
\end{align*}
\]

(7.53)

To represent this in a similar way as before, one can use a three dimensional array or a tensor of order 3:

\[ y = A \times_1 x \times_2 x \times_3 x, \quad (7.54) \]
with:

\[
x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad A_{1,j,k} = \begin{bmatrix}
\frac{a_{111}}{3} & a_{112} & \frac{a_{113}}{3} \\
\frac{a_{112}}{3} & \frac{a_{122}}{3} & \frac{a_{123}}{6} \\
\frac{a_{113}}{3} & \frac{a_{123}}{6} & \frac{a_{133}}{3}
\end{bmatrix},
\]

(7.55)

\[
A_{2,j,k} = \begin{bmatrix}
\frac{a_{112}}{3} & \frac{a_{122}}{3} & \frac{a_{123}}{6} \\
\frac{a_{122}}{3} & a_{222} & \frac{a_{223}}{3} \\
\frac{a_{123}}{6} & \frac{a_{223}}{3} & \frac{a_{333}}{3}
\end{bmatrix},
\]

(7.56)

\[
A_{3,j,k} = \begin{bmatrix}
\frac{a_{113}}{3} & \frac{a_{123}}{6} & \frac{a_{133}}{3} \\
\frac{a_{123}}{6} & \frac{a_{223}}{3} & \frac{a_{323}}{3} \\
\frac{a_{133}}{3} & \frac{a_{323}}{3} & a_{333}
\end{bmatrix},
\]

(7.57)

where the multiplication of a tensor \( T \) with matrices \( U, V \) and \( W \) is defined as:

\[
C = T \times_1 U \times_2 V \times_3 W,
\]

(7.58)

\[
C_{\alpha\beta\gamma} = \sum_{i,j,k=1}^{l,m,n} t_{ijk} u_{\alpha i} v_{\beta j} w_{\gamma k},
\]

(7.59)

where:

\[
C \in \mathbb{R}^{p \times q \times r}, \quad T \in \mathbb{R}^{l \times m \times n},
\]

\[
U \in \mathbb{R}^{p \times l}, \quad V \in \mathbb{R}^{q \times m}, \quad W \in \mathbb{R}^{r \times n}.
\]

The coefficient tensor \( A \) is a symmetric tensor, since it is invariant under all permutations on the indices. This means that the problem of eliminating the crossterms in a homogeneous polynomial of degree 3 is now reduced to a problem of finding a rotation matrix \( V \) such that:

\[
y = A \times_1 x \times_2 x \times_3 x = D \times_1 \bar{x} \times_2 \bar{x} \times_3 \bar{x},
\]

(7.60)

where each element \( D_{ijk} \) that does not lie on the hyperdiagonal is zero: \( D_{ijk} \in \mathbb{R} \) if \( i = j = k \) and \( D_{ijk} = 0 \) elsewhere.
This problem can be solved for example by using the parafac method (Harshman [1970]), as implemented in the n-way toolbox presented in Andersson and Bro [2000], or any other similar methods (Brachat et al. [2010]; Sorber et al. [2014]). This results in a three-dimensional tensor $D$, with only nonzero elements on its hyperdiagonal, and a transformation matrix $V$ such that:

$$\bar{x} = V^T x.$$  \hfill (7.61)

It is important to note that the transformation matrix $V$ does not have to be square. It can be that to represent a general tensor $A \in \mathbb{R}^{3 \times 3 \times 3}$, a diagonal tensor of higher dimension is needed e.g. a $D \in \mathbb{R}^{d \times d \times d}$ diagonal tensor (Brachat et al. [2010]; Harshman [1970]). This results in a transformation matrix $V \in \mathbb{R}^{3 \times d}$, where $d$ is the rank of the tensor $A$ (Harshman [1970]). In other words, the decoupled representation of a coupled polynomial can have more inputs (more parallel branches) than the original coupled representation. The rank $d$ of tensor $A$ is selected by performing a scan over a range of possible values of $d$.

### 7.7.3 Homogeneous Polynomial of Degree $n$

A homogeneous polynomial of degree $n$ can always be expressed as a multilinear product of an $n$-dimensional symmetric tensor $A$ and an input vector $x$, similar to the description of homogeneous polynomials of degree 3 as described in Section 7.7.2. The problem of eliminating the crossterms is again reduced to a problem of finding a rotation matrix $V$ such that all non-hyperdiagonal elements of the new coefficient tensor $D$ are zero. This problem can be solved again using tensor decomposition methods (Brachat et al. [2010]; Harshman [1970]; Andersson and Bro [2000]). This results in:

$$y = A \times_1 x \times_2 x \times_3 x \ldots \times_n x,$$

$$\quad = D \times_1 \bar{x} \times_2 \bar{x} \times_3 \bar{x} \ldots \times_n \bar{x},$$  \hfill (7.62)

$$\bar{x} = V^T x,$$  \hfill (7.63)
where $A$ and $D$ are $n$-dimensional tensors, and:

$$V \in \mathbb{R}^{n \times d}, \ x \in \mathbb{R}^{n \times 1}, \ \bar{x} \in \mathbb{R}^{d \times 1}. \quad (7.64)$$

The rank $d$ of tensor $A$ is selected by performing a scan over a range of possible values of $d$. This rank can be higher than the size $n$ of the tensor. This means that decoupling a coupled polynomial can increase the number of inputs of that polynomial.

### 7.7.4 Polynomial of Degree $n$

A multivariate polynomial $f$ of degree $n$ can always be written as a sum of $n$ homogeneous polynomials $g_1, g_2, \ldots, g_n$ of degree 1 to $n$:

$$f(x_1, \ldots, x_{n_{br}}) = g_1(x_1, \ldots, x_{n_{br}}) + \ldots + g_n(x_1, \ldots, x_{n_{br}}). \quad (7.65)$$

With the technique described in Section 7.7.3 all these separate homogeneous polynomials $g_i$ can be transformed to a canonical form.

$$y_i = g_i(x_1, \ldots, x_{n_{br}}),$$

$$= A_i \times_1 x \times_2 x \times_3 x \ldots \times_i x,$$

$$= D_i \times_1 \bar{x}_i \times_2 \bar{x}_i \times_3 \bar{x}_i \ldots \times_i \bar{x}_i, \quad (7.66)$$

$$\bar{x}_i = V_i^T x. \quad (7.67)$$

Where $x$ is the original input vector, $\bar{x}_i$ the input vector after transformation, $V_i$ the rotation matrices, $D_i$ the hyperdiagonal coefficient tensor, and $A_i$ the original coefficient tensor for the homogeneous polynomial of degree $i$.

A downside of this approach is that new input signals need to be introduced for each homogeneous polynomial $g_i$, which increases the total number of inputs of the original polynomial $f$. This problem is solved in more recent decoupling approaches. A first approach uses first order approaches (Dreesen et al. [2015]). This approach is not restricted to polynomial representations of the coupled static nonlinearity. A second approach Tiels and
Decoupling the MISO Static Nonlinearity

Schoukens [2013]; Schoukens et al. [2014c] uses tensor decomposition similar to the one that is explained here. The main difference between this tensor decoupling approach and the more recent one of Tiels and Schoukens [2013]; Schoukens et al. [2014c] is that all homogeneous polynomials of different degree are combined into one big tensor. Diagonalizing this tensor results in the necessary transformation matrix $V$. However, only one transformation for all degrees of the polynomial is obtained, instead of one transformation matrix for every degree of the polynomial. This results in a more compact representation of the decoupled static nonlinearity.

Note that the number of parallel branches that are obtained after the decoupling can be higher than the number of parallel branches of the coupled representation of the model. This depends on the rank of the tensor that needs to be diagonalized.

### 7.7.5 Parallel Wiener Transformation

A parallel Wiener model contains a multiple input single output nonlinearity at the output as shown in Figure 7.3. These nonlinearities are often described using a multivariate polynomial. The crossterms present in this multivariate polynomial can now be eliminated with the method as described in Section 7.7.4. A 2-branch, third order parallel Wiener model shown in Figure 7.3 is taken as an example.

The first step is to split the nonlinearity in the different homogeneous polynomials $g_i$. Next, the linear transformations $V_i$ can be calculated, with which the crossterms of each homogeneous polynomial are eliminated. These linear transformations can be placed in the block schematics as an additional block in between the linear time invariant (LTI) blocks $H^{[j]}(q)$ and the static nonlinear blocks $g_i$. The result of step one and two is shown in Figure 7.4. New static nonlinear blocks $x^i$ and LTI blocks $H^{[j]}_i(q)$ can be introduced when the linear transformation $V_i$ is applied on the original LTI blocks $H^{[j]}(q)$ and static nonlinear blocks $g_i$.

The elements in the different $D_i$ matrices are the new coefficients of the decomposed polynomial. Based on the magnitude of these coefficients,
the user can see which paths contribute most to the output. Also some model complexity reduction decisions can be made based on the relative magnitude of these coefficients.

The final model without crossterms in the static nonlinearity, is shown in Figure 7.5.

A similar method can be used to decouple a multivariate polynomial in a parallel Wiener-Hammerstein model structure or in polynomial nonlinear state space models.

The number of branches in the model is increased significantly, but the nonlinearity is described using only decoupled monomials. This allows a model description that is easier to interpret by a designer and easier to invert.

![Figure 7.3: Parallel Wiener model with 2 branches. The static nonlinearity is modeled by a multivariate polynomial $f$, the linear time invariant blocks are represented by $H^1$ and $H^2$.](image)

7.8 Nonlinear Optimization

Joining all the previous estimation steps allows one to obtain the model parameter vector as the estimation of a partition of the parameter vector $\theta$ in non-overlapping subsets, as is explained in Section 3.5.6. Note that the parameter vector $\theta \in \mathbb{R}^{(n_{br}(n_{d}+1)+n_f+n_c+1) \times 1}$ now contains all the parameters of the model: $\gamma$, $\delta$, and $c$. The Jacobian of the cost function is rank deficient due to the identifiability issues (see Section 7.3). The related degenerations ($n_{br} + n_{br}^2$ for the coupled model and $2n_{br}$ for the decoupled model) in the Jacobian need to be taken into account during the optimization.
7.9 Simulation Example

7.9.1 System and Signals

The proposed parallel Wiener identification algorithms are illustrated on a simulation example. The simulated system is a 3-branch parallel Wiener system. The LTI blocks $H^[k](q)$ are fourth order Chebychev type 1 band-pass filters with a 2dB in-band ripple and cut-off frequencies at $0.025f_s$ and $0.1375f_s$, $0.1375f_s$ and $0.25f_s$, $0.25f_s$ and $0.3625f_s$ respectively:

\[
H^[1](q) = \frac{0.0632 - 0.1263q^{-2} + 0.0632q^{-4}}{1 - 3.0325q^{-1} + 3.7168q^{-2} - 2.2493q^{-3} + 0.5789q^{-4}}, \tag{7.68}
\]

\[
H^[2](q) = \frac{0.0632 - 0.1263q^{-2} + 0.0632q^{-4}}{1 - 1.2030q^{-1} + 1.6473q^{-2} - 0.8923q^{-3} + 0.5789q^{-4}}, \tag{7.69}
\]

\[
H^[3](q) = \frac{0.0632 - 0.1263q^{-2} + 0.0632q^{-4}}{1 + 1.2030q^{-1} + 1.6473q^{-2} + 0.8923q^{-3} + 0.5789q^{-4}}. \tag{7.70}
\]
Figure 7.5: Parallel Wiener model with 2 branches and a third degree polynomial function describing the static nonlinearity after elimination of the crossterms in the multivariate polynomial.

The static nonlinearity \( f(x_1, x_2, x_3) \) is a polynomial of degree 3:

\[
\begin{align*}
  f(x_1, x_2, x_3) &= 0.9x_1 + x_2 + x_3 + 0.3x_1^2 \\
  &= 0.15x_2^2 + 0.03x_1^3 + 0.09x_2^3 + 0.03x_3^3.
\end{align*}
\]

White additive Gaussian noise with a standard deviation of 0.0001 is added to the output of the system.

A random phase multisine is used as an input signal. 16 realizations at 5 different amplitudes are applied to the system. Each realization is measured during one steady state period, consisting of 512 samples. All the frequencies in the range \([0, f_s/2]\) are excited. The random phases are uniformly distributed between \([0, 2\pi]\). The rms-value of the input signals is logarithmically spaced from 0.1 to 1. This results in a signal-to-noise ratio at the output ranging from 55 dB to 75 dB depending on the input amplitude. The identification through dimension reduction method (Section 7.6) uses
only the multisine input signals with the highest input rms-value.

The system is sampled at a sampling frequency $f_s$ of 512 Hz.

### 7.9.2 Model Estimation: SVD Approach

The BLAs of the system at the different amplitudes are estimated in a first identification step using the robust method (see Section 2.3.1) followed by a common denominator parametrization of the nonparametric BLA (see Section 7.5.2). The estimated BLAs are shown in Figure 7.6. A small change in the shape of the BLAs for changing input rms-values can be observed. This information is used in the second step to obtain an estimate of the dynamics that are present in each branch. Next, the static nonlinearity of each branch is estimated. Finally, the parameters are refined in a nonlinear optimization step, optimizing all the parameters simultaneously.

The estimated model is validated using a different realization of the random phase multisine of the highest rms-value of the input. This realization has not been used during the estimation. The time domain representation of the measured validation output and the residual error ($\hat{y}(t) - y(t)$) are shown in Figure 7.7 and Figure 7.8. The frequency domain representation...
Table 7.1: Validation of the models obtained with both the SVD and the MAVE approach before and after optimization. The rms value of the output residuals ($\text{rms}(y - \hat{y})$) is shown.

<table>
<thead>
<tr>
<th>input rms value</th>
<th>0.1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD method (init)</td>
<td>0.00472</td>
<td>0.08317</td>
</tr>
<tr>
<td>SVD method (optimized)</td>
<td>0.00011</td>
<td>0.00010</td>
</tr>
<tr>
<td>MAVE method (init)</td>
<td>/</td>
<td>0.00643</td>
</tr>
<tr>
<td>MAVE method (optimized)</td>
<td>/</td>
<td>0.00011</td>
</tr>
</tbody>
</table>

of the validation output and the magnitude of the complex error between the measurement and the modeled output are shown in Figure 7.9. It is clear from Figure 7.9 that the output of the parallel Wiener system is very nonlinear for the highest input rms-value: there is only a 10 dB gap between the output spectrum and the total distortion variance for the highest rms-value of the input. The system behaves quite linearly for the lowest rms-value of the input: there is a 30 to 40 dB gap between the output spectrum and the total distortion level.

In both cases, the initial model estimate results in residual errors which are well below the output spectrum (about 40 and 15 dB for the lowest and the highest excitation level respectively). Note that the level of the residual errors is just below the total distortion level. This shows that the initial model estimate models the nonlinear behavior of the device over a broad range of the input excitation level, but only captures a small part of the nonlinear behavior of the system in this simulation example. After optimization the residual error coincides with the noise distortion level that is added during the simulations for both the low and high input excitation levels (see Table 7.1). This shows that the final model captures the nonlinear behavior of the system almost completely.

7.9.3 Model Estimation: MAVE Approach

The BLA of the system for the highest amplitudes only is estimated in a first identification step using the robust method (see Section 2.3.1) followed by a parametrization of the nonparametric BLA (see Section 2.3.2).
Figure 7.7: Time domain validation of the parallel Wiener model obtained with the SVD approach for a low input rms-value (a) and a high input rms-value (b). The noisy system output is depicted by the full blue line, the output of the optimized model is depicted by the dashed green line. Note that there is hardly a noticeable difference.
Figure 7.8: Time domain validation of the parallel Wiener model obtained with the SVD approach before and after optimization for a low input rms-value (a) and a high input rms-value (b). The model error before optimization is represented by the red line. The model error after optimization is represented by the green line.
Figure 7.9: Frequency domain validation of the parallel Wiener model obtained with the SVD approach before and after optimization for a low input rms-value (a) and a high input rms-value (b). The noisy system output is depicted by the full blue line. The model error before optimization is represented by the red dots. The model error after optimization is represented by the green dots. The variance of the total distortion that is present at the output is shown by the black dots.
The estimated BLA is shown in Figure 7.10. The parametrized BLA is decomposed using the partial fraction expansion followed by a dimension reduction step using MAVE. Next, the static nonlinearity of each branch is estimated. Finally, the parameters are refined in a nonlinear optimization step, optimizing all the parameters simultaneously. Note that a different dataset (belonging to the same excitation class) is used here compared to the simulation example that is shown in the previous section.

The estimated model is validated using a different realization of the random phase multisine of the highest rms-value of the input. These realizations have not been used during the estimation. The time domain representation of the measured validation output and the residual error \((\hat{y}(t) - y(t))\) are shown in Figure 7.11 and Figure 7.12. The frequency domain representation of the validation output and the magnitude of the complex error between the measurement and the modeled output are shown in Figure 7.13.

The initial model estimate results in residual errors which are well below the output spectrum (about 50 dB). The residual errors are also well below
7.10 Conclusion and Discussion

The total distortion level (about 30 dB). This shows that the initial model estimate captures a large part of the nonlinear behavior of the system in this simulation example. The rms-value of the error before and after optimization is 0.00643 and 0.00011 respectively (see Table 7.1). The model error before optimization with this approach is well below the model error before optimization obtained with the SVD approach (see Table 7.1). The model error after optimization coincides with the rms-value of the additive noise that is present at the output of the system.

7.10 Conclusion and Discussion

Two parametric approaches to identify parallel Wiener systems are presented. The first method is a 3-step approach, the second uses the MAVE dimension reduction method.

The first method uses a 3-step approach to estimate initial estimates, together with a nonlinear optimization of these initial estimates to improve model efficiency. Firstly, different linearized models of the nonlinear system
7. Parallel Wiener

Figure 7.12: Time domain validation of the parallel Wiener model obtained with the MAVE approach before and after optimization for a high rms-value of the input. The model error before optimization is represented by the red line. The model error after optimization is represented by the green line.

Figure 7.13: Frequency domain validation of the parallel Wiener model obtained with the MAVE approach before and after optimization for an input with a high rms-value. The noisy system output is depicted by the full blue line. The model error before optimization is represented by the red dots. The model error after optimization is represented by the green dots. The variance of the total distortion that is present at the output is shown by the black dots.
are estimated at different input power levels. This is done using the BLA using either Gaussian noise or random phase multisine excitations. Secondly, a decomposition of the frequency dynamics of the BLAs results in an initial estimate of the linear time invariant blocks that are present in the parallel Wiener model structure. Finally, the coefficients of the nonlinear basis functions describing the static nonlinear block are estimated using a least squares estimator.

The second approach uses the combination of the BLA and the MAVE dimension reduction method. The BLA estimates the poles and zeros present in the overall system dynamics, and the MAVE dimension reduction method selects which poles and zeros belong to which branch. The proposed approach needs good-quality estimates of the poles present in the parallel Wiener system to obtain accurate estimates of the LTI blocks present in the system. This approach obtains good results while it uses a much lesser amount of measured data than the SVD approach. This comes at the cost of a more complex, but still user friendly, identification algorithm.

A major advantage of these methods compared to other approaches is that no information of the Volterra kernels is required, and the number of branches is independent of the degree of nonlinearity. The number of branches rather depends on the number of independent signal paths that are present in the system. Furthermore, this number of branches can be selected based on results of the identification method. Finally, the LTI blocks can be modeled using rational transfer functions, and both discrete- and continuous time models are possible.

This chapter also describes a method to eliminate the crossterms that are present in a multivariate polynomial using a linear transformation of the inputs, as can be used for instance in parallel Wiener or Wiener-Hammerstein models. This allows for a model description that is easier to interpret and to invert. A downside of the method is the introduction of new input signals for each degree in the polynomial. This is solved in more recent methods that use similar approaches.
Parallel Wiener-Hammerstein

8.1 Introduction

Parallel Hammerstein and parallel Wiener model structures are restricted to systems with dominant output dynamics (Hammerstein) or input dynamics (Wiener). A further increase in flexibility of the model structure is obtained here by considering the parallel Wiener-Hammerstein structure. This chapter presents a method to identify parallel Wiener-Hammerstein systems, whose structure is shown in Figure 8.1. Previously published methods (Baumgartner and Rugh [1975]; Wysocki and Rugh [1976]; Billings and Fakhouri [1979]) studied a subclass of the parallel Wiener-Hammerstein structure. Identification methods based on repeated sine measurements (Baumgartner and Rugh [1975]; Wysocki and Rugh [1976]), or white Gaussian inputs (Billings and Fakhouri [1979]) are available for this model structure. In Palm [1978, 1979] it is shown that a wide class of Volterra systems can be approximated arbitrary well using a parallel Wiener-Hammerstein model structure. However, no method is presented there to identify such models.

The identification method presented in Billings and Fakhouri [1979] uses Gaussian excitation signals as the methods presented in this chapter do. However, this method is a generalization of a Wiener-Hammerstein identification algorithm based on a parametrized version of higher order correlation
functions between input and output (Billings and Fakhouri [1978]). This approach has been compared in Schoukens et al. [2014b] with two other approaches (Westwick and Schoukens [2012]; Schoukens et al. [2014b]), and it was outperformed by these alternatives. The main problem of the method seems to be its sensitivity to perturbation noise.

The parallel Wiener-Hammerstein identification approach proposed here combines the parallel Hammerstein and parallel Wiener identification methods presented in Schoukens et al. [2011]; Schoukens and Rolain [2012b] and Chapters 6 and 7 with a specific initialization approach for Wiener-Hammerstein systems presented in Sjöberg and Schoukens [2012]. This chapter is based on, and hereby extends, the results of Schoukens et al. [2013b, 2015c].

This chapter introduces two identification methods for parallel Wiener-Hammerstein systems. The consistency of the proposed methods is discussed, as well as their computational complexity. The proposed identification methods are illustrated on a simulation example.

### 8.2 System

![Figure 8.1: A parallel Wiener-Hammerstein system with 3 parallel branches.](image)

$S^{[k]}(q)$ and $H^{[k]}(q)$ are the LTI systems of branch $k$, $f^{[k]}(x_k)$ is the static nonlinearity of branch $k$.

A parallel Wiener-Hammerstein system is a parallel connection of several Wiener-Hammerstein subsystems as shown in Figure 8.1. The noiseless
8.2. System

Output of a parallel Wiener-Hammerstein system is given by:

\[ y_0(t) = \sum_{k=1}^{n_{br}} y_k(t), \quad (8.1) \]

\[ y_k(t) = S[k](q) r_k(t), \quad (8.2) \]

\[ r_k(t) = f[k](x_k(t)), \quad (8.3) \]

\[ x_k(t) = H[k](q) u(t), \quad (8.4) \]

where \( f[k](x_k(t)) \) is the SISO static nonlinear function of branch \( k \). \( H[k](q) \) and \( S[k](q) \) are the LTI systems that are present in branch \( k \). The number of parallel branches is given by \( n_{br} \). The signals \( y_0(t) \), \( y_k(t) \), \( r_k(t) \), \( x_k(t) \) and \( u(t) \) are scalar signals.

The static nonlinearities \( f[k](x_k(t)) \) are assumed to be obtained as a linear combination of some set of basis functions (e.g. polynomial):

\[ f[k](x_k(t)) = \sum_{i=1}^{n_f} \gamma_i[k] f_i(x_k(t)), \quad (8.5) \]

where the coefficients \( \gamma_i[k] \) are grouped in the coefficient vector \( \gamma[k] \in \mathbb{R}^{n_f \times 1} \), and \( n_f \) is the total number of nonlinear basis functions used to represent the nonlinearity.

The LTI subsystems \( H[k](q) \) and \( S[k](q) \) are assumed to be rational functions of the backward shift operator \( q^{-1} \):}

\[ H[k](q) = \frac{B_h[k]}{A_h[k](q)}, \quad (8.6) \]

\[ = \frac{b_h[k] + b_h[1] q^{-1} + \ldots + b_h[n_{bh,k}] q^{-n_{bh,k}}}{a_h[0] + a_h[1] q^{-1} + \ldots + a_h[n_{ah,k}] q^{-n_{ah,k}}}, \]

\[ S[k](q) = \frac{B_s[k]}{A_s[k](q)}, \quad (8.7) \]

\[ = \frac{b_s[k] + b_s[1] q^{-1} + \ldots + b_s[n_{bs,k}] q^{-n_{bs,k}}}{a_s[0] + a_s[1] q^{-1} + \ldots + a_s[n_{as,k}] q^{-n_{as,k}}}, \]

where the coefficients \( b_{h,i}^k \) and \( a_{h,i}^k \) are grouped in the coefficient vectors \( b_h^k \in \mathbb{R}^{(n_{bh,k}+1) \times 1} \) and \( a_h^k \in \mathbb{R}^{(n_{ah,k}+1) \times 1} \). The coefficients \( b_{s,i}^k \) and \( a_{s,i}^k \) are
grouped in the coefficient vectors \( b_s^{[k]} \in \mathbb{R}^{(n_{bs,k}+1) \times 1} \) and \( a_s^{[k]} \in \mathbb{R}^{(n_{as,k}+1) \times 1} \). The \( l^2 \)-norm of the coefficient vectors \( a_h^{[k]} \) and \( a_s^{[k]} \) is assumed to be fixed to a value that is equal to 1, and the first nonzero element of \( a_h^{[k]} \) and \( a_s^{[k]} \) are assumed to be positive to compensate for the degeneration that is present in the rational transfer function parametrization.

The parallel Wiener-Hammerstein system class that is used here is a more general system class than the \( S_M \) system class that is used in Baumgartner and Rugh [1975]; Wysocki and Rugh [1976]; Billings and Fakhouri [1979] in the following sense. The \( S_M \) model has \( M \) parallel branches, and the \( m \)-th branch contains a monomial nonlinearity that is fixed and equal to \( (.)^m \). This restricts the model to have a polynomial nonlinearity only, and to contain only one branch for each degree of this polynomial nonlinearity. Thus a parallel Wiener-Hammerstein system containing two parallel branches, each with different LTI subsystems, and with different polynomial nonlinearities can, in general, not be modeled by a \( S_M \) model. The methods that are presented in this chapter also make some extra assumptions on the parallel Wiener-Hammerstein system in the following sections. However, even when these assumptions are met, the considered system class still allows for a much more complicated nonlinear system behavior than the \( S_M \) class does.

### 8.3 Identifiability of a Parallel Wiener-Hammerstein System

The LTI blocks and the static nonlinear blocks contained in a parallel Wiener-Hammerstein system can only be identified up to an unknown gain, unless one has access to the intermediate signals \( x_k(t) \) and \( r_k(t) \). If the gains \( \beta_{Hk} \) and \( \beta_{Sk} \) are introduced for \( H^{[k]}(q) \), \( S^{[k]}(q) \) and \( f^{[k]}(u) \) one ob-
8.3. Identifiability of a Parallel Wiener-Hammerstein System

\[ y_k(t) = \tilde{S}^{[k]}(q) \tilde{r}_k(t) = \beta_{Sk} S^{[k]}(q) \tilde{r}_k(t), \]  
(8.8)

\[ \tilde{r}_k(t) = \tilde{f}^{[k]}(\tilde{x}_k(t)) = \frac{1}{\beta_{Sk}} f\left(\frac{1}{\beta_{Hk}} \tilde{x}_k(t)\right), \]  
(8.9)

\[ \tilde{x}_k(t) = \tilde{H}^{[k]}(q) u(t) = \beta_{Hk} H^{[k]}(q) u(t). \]  
(8.10)

The input-output behavior of a parallel Wiener-Hammerstein system that is composed using \( \tilde{H}^{[k]}(q) \), \( \tilde{S}^{[k]}(q) \) and \( \tilde{f}^{[k]}(u(t)) \) is exactly the same as a parallel Wiener-Hammerstein system containing \( H^{[k]}(q) \), \( S^{[k]}(q) \) and \( f^{[k]}(u(t)) \) as building blocks.

A phase shift degeneration is also present in the linear subsystems (as was the case for a Wiener-Hammerstein system in Section 4.3). Indeed, an arbitrary delay \( \tau_k \) can be introduced in the two LTI blocks of branch \( k \) of the system without any change in the response. To this end \( H^{[k]}(e^{j\omega T_s}) \) is to be delayed by \( \tau_k \) while \( S^{[k]}(e^{j\omega T_s}) \) is delayed by \(-\tau_k\).

\[ \tilde{H}^{[k]}(q) = \beta_{Hk} H^{[k]}(q) q^{-\tau_k}, \]
\[ \tilde{S}^{[k]}(q) = \beta_{Sk} S^{[k]}(q) q^{\tau_k}. \]  
(8.11)

This results in the same input-output behavior. The reason for the presence of the degeneration is that in the time domain such a phase shift is simply a delay of the signal at the output of the first LTI block \( \tilde{H}^{[k]}(q) \) of branch \( k \). This does not affect the behavior of the static nonlinearity at all, as it has no memory and is time invariant. The delay that was caused by the first LTI block can therefore be exactly compensated by an inverse delay in the second LTI block \( \tilde{S}^{[k]}(q) \).

An additional identifiability issue appears here due to the parallel nature of the parallel Wiener-Hammerstein systems. Starting from input-output data only, infinitely many equivalent models can be obtained by introducing a regular linear transformation to the models. In the block diagram of the model, this introduces a full rank linear transformation matrix between the outputs of the front LTI blocks \( H^{[k]}(q) \) and the inputs of the static nonlinearities \( f^{[k]}(x_k(t)) \) and between the outputs of the static nonlinearities \( f^{[k]}(x_k(t)) \) and the inputs of the back LTI blocks \( S^{[k]}(q) \). These full rank
8. Parallel Wiener-Hammerstein

Figure 8.2: A parallel Wiener-Hammerstein system with 3 parallel branches subject to two linear transformations $T_H$ and $T_S$.

Linear transformations are described by the matrices $T_H$ and $T_S$. Since they are regular, their inverse transformations $T_H^{-1}$ and $T_S^{-1}$ exist and can also be included in between the LTI blocks and the static nonlinear blocks as illustrated in Figure 8.2. The linear transformations can also be included in the LTI blocks and the static nonlinear blocks. This results in a new set of LTI blocks that are a linear combination of the original ones. The SISO static nonlinearities of the different branches $k$ are transformed into one MIMO static nonlinearity which contains crossterms combining the different branches. The resulting coupled Wiener-Hammerstein system is shown in Figure 8.3. The derivation of the transformed LTI blocks and the transformed static nonlinearity is similar to the derivation shown in Sections 6.3 and 7.3.

It can be argued that the linear transformation matrices represent no true identifiability issue of the parallel Wiener-Hammerstein system class since they transform the originally decoupled SISO static nonlinearities into a coupled MIMO static nonlinearity. Here we do consider them as an identifiability issue since they appear in the identification algorithms as will be shown in Sections 8.5 and 8.6. This identifiability issue can be taken care of in a later step by decoupling the MIMO static nonlinearity into several SISO static nonlinearities (see Tiels and Schoukens [2013]; Schoukens et al. [2014c]; Dreesen et al. [2015] and Section 7.7).

Although theoretically the gain exchanges $\beta_{Hk}$ and $\beta_{Sk}$ and the transformations $T_H$ and $T_S$ result in the same input-output behavior of the system, bad choices of these gains and these transformations can lead to numerical issues. It is best to avoid a high dynamic range over the different
parameters that are present in the model. Also a high dynamic range over the different signals \( u(t), x_{k}(t), r_{k}(t), y_{k}(t) \) and \( y(t) \) should be avoided if possible.

Figure 8.3: A parallel coupled Wiener-Hammerstein system with 3 parallel branches.

8.4 Best Linear Approximation

This section shows that the BLA of a parallel Wiener-Hammerstein system can be written as a linear combination of the LTI dynamics that are present in the different branches \( H^{[k]}(q)S^{[k]}(q) \) when input signals belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals are used as an excitation signal.

Assumption 8.1. **Input signal:** The input signal belongs to the Riemann equivalence class of asymptotically normally distributed excitation signals as defined in Definition 2.1.

Theorem 8.1. The BLA \( G_{bla}(e^{j\omega T_s}) \) of a parallel Wiener-Hammerstein system excited by an input signal belonging to the Riemann equivalence class of asymptotically normally distributed excitation signals (Assumption 8.1) is given by:

\[
G_{bla}(e^{j\omega T_s}) = \sum_{k=1}^{n_h} \alpha_k H^{[k]}(e^{j\omega T_s})S^{[k]}(e^{j\omega T_s}), \quad \alpha_k \in \mathbb{R}.
\]  

(8.12)

Proof. The BLA of a Wiener-Hammerstein system is given by (4.11) \( G_{bla}(e^{j\omega T_s}) = \alpha H(e^{j\omega T_s})S(e^{j\omega T_s}) \) under Assumption 8.1. This holds for each separate
branch in the parallel Wiener-Hammerstein system. The output $y$ is given by a summation of the outputs $y_k$ of each parallel Wiener-Hammerstein branch. Hence, the BLA of a parallel Wiener-Hammerstein system is given by the weighted sum of the BLA of each parallel branch.

Note that the shape of $G_{bla}(e^{j\omega T_s})$ depends on the power spectrum of the input signal. The shape of $G_{bla}(e^{j\omega T_s})$ can change if a class of input signals with a different power spectrum is used. The coefficients $\alpha_k$ depend on many factors: the LTI subsystems $H[k](q)$ and $S[k](q)$, the static nonlinear subsystems $f[k](x_k(t))$, the power spectrum $S_{UV}(e^{j\omega T_s})$ of the input signal $u(t)$ and, hence, as well on the variance of the input signal $u(t)$.

**Assumption 8.2. Non-zero gain:** The BLA $\alpha_k H[k](q) S[k](q)$ of every branch $k$ has a non-zero gain $\alpha_k$.

It can happen that the BLA of one of the branches of the parallel Wiener-Hammerstein system is equal to zero, or in other words $\alpha_k = 0$. This is the case when the nonlinear function $f[k](x_k(t))$ is even around the expected value of $x_k(t)$. In this case, a BLA of a reduced order is obtained that does not contain the dynamics of branch $k$. This assumption excludes that the static nonlinearity $f[k](x_k(t))$ of branch $k$ is symmetric with respect to the DC setpoint of the signal $x_k(t)$. Different DC setpoints can be tried to avoid a zero gain $\alpha_k$ if this should happen and the system allows this kind of variation.

**Assumption 8.3. Pole-zero cancellations:** The combined dynamics $H[k](q) S[k](q)$ of the $k$-th branch do not contain any pole-zero cancellation for any branch $k$.

A pole that appears in the front LTI block of a branch, can be canceled by a zero that is present in the back LTI block of the same branch. They will not be detected during the parametrization of the BLA. This assumption is quite common for Wiener-Hammerstein identification algorithms, see for example the two identification algorithms presented in Sjöberg et al. [2012]. However, there exist different Wiener-Hammerstein and $S_M$ identification algorithms that do not need this assumption (Baumgartner and Rugh [2012]).
8.5 Identification through BLA Decomposition

[1975]; Wysocki and Rugh [1976]; Billings and Fakhouri [1979]; Schoukens et al. [2014b]). This is possible using a more advanced correlation analysis (Billings and Fakhouri [1979]; Schoukens et al. [2014b]), or a more restrictive class of input signals (Baumgartner and Rugh [1975]; Wysocki and Rugh [1976]).

An important observation with respect to eq. (8.12) is that the input dependent gain $\alpha_k$ only appears in the numerator:

$$G_{bla}(q) = \frac{\sum_{k=1}^{n_{br}} \alpha_k B^{[k]}_{hs}(q) \prod_{j=1,j\neq k}^{n_{br}} A^{[j]}_{hs}(q)}{\prod_{k=1}^{n_{br}} A^{[k]}_{hs}(q)},$$  \hspace{1cm} (8.13)

where

$$B^{[k]}_{hs}(q) = B^{[k]}_h(q) B^{[k]}_s(q),$$  \hspace{1cm} (8.14)

$$A^{[k]}_{hs}(q) = A^{[k]}_h(q) A^{[k]}_s(q).$$  \hspace{1cm} (8.15)

This means that under Assumption 8.2, the poles of the identified BLA are also the poles of the LTI blocks that are present in the system. The zeros of the BLA of a parallel Wiener-Hammerstein system may change when the amplitude, power spectrum, or the offset (DC value) of the input signal changes. Indeed, the denominator stays the same when the gains $\alpha_k$ change, but the numerator coefficients depend on the gains $\alpha_k$.

8.5 Identification through BLA

**Decomposition**

*Based on Schoukens et al. [2014c] and Schoukens et al. [2015c]*

This section is based on the results that are reported in Schoukens et al. [2015c]. That paper combines and extends the approaches presented in Schoukens et al. [2011]; Schoukens and Rolain [2012b] to decompose the dynamics over the different branches of a parallel Hammerstein and parallel Wiener systems with an initialization approach that splits the dynamics into the front and back LTI blocks of a Wiener-Hammerstein system as presented
in Sjöberg and Schoukens [2012]. Other approaches to split the dynamics of a Wiener-Hammerstein system exist in the literature Billings and Fakhouri [1978]; Westwick and Schoukens [2012]; Schoukens et al. [2014b], but are more complex to implement and seem to be more sensitive to the presence of noisy data (e.g. corrupted by perturbation noise).

The proposed approach starts with an estimation of the BLA of the considered system for different operating conditions (Section 8.5.1). The different operating conditions are obtained using input signals with different power spectra. This includes the use of different magnitudes, different offsets, or different coloring of the power spectra. A consistent estimate of the overall dynamics that are present in the nonlinear parallel Wiener-Hammerstein system results.

The measured BLAs are parametrized in a second step (Section 8.5.2) using a different LTI model for each operating condition. A common denominator model is used for all operating conditions simultaneously. This is indeed possible, as Theorem 8.1 assures that the poles of the different measured BLAs are the same.

Starting from the parametrized BLAs, a decomposition of the overall dynamics at the different operating conditions is calculated in Section 8.5.3. It uses the singular value decomposition (SVD) of a matrix constructed using the numerator coefficients of the parametrized BLAs obtained at the different operating conditions. This step also results in an estimate of the number of branches that is present in the parallel Wiener-Hammerstein system. This number is obtained based on the estimated rank of the decomposed matrix. The dynamics $H^{[k]}(q)S^{[k]}(q)$ that are present in each branch are estimated next, up to the identifiability issues that are discussed in Section 8.3.

Finally, a modified version of the algorithm that is used in Sjöberg and Schoukens [2012] is proposed in Section 8.5.4 to partition the dynamics $H^{[k]}(q)S^{[k]}(q)$ over the different blocks of the parallel Wiener-Hammerstein model, and to estimate the static nonlinearity that is present in the model.

Sections 8.5.1, 8.5.2 and 8.5.3 are identical to Sections 6.5.1, 6.5.2 and 6.5.3 presented in the parallel Hammerstein chapter, but are repeated here.
8.5. Identification through BLA Decomposition

for the convenience of the reader.

8.5.1 Estimating a BLA for Different Operating Conditions

The nonparametric FRF estimate of the BLA at operating condition \( i_r \) is labeled \( \hat{G}_{bla}^{[i_r]}(e^{j\omega_k T_s}) \). It is obtained by the robust BLA estimation method (see Section 2.3.1). Both the FRF and the sample variance \( \hat{\sigma}^2_{\hat{G}_{bla}^{[i_r]}}(k) \) of the BLA are obtained at each excited frequency. The latter is used to determine the weighting factor for the parametric model estimation of the BLA. This process is explained in more detail in Pintelon and Schoukens [2012]; Schoukens et al. [2012a, 2013b].

8.5.2 Parametrizing the BLAs

The measured nonparametric BLAs \( \hat{G}_{bla}^{[i_r]}(e^{j\omega_k T_s}) \) at the \( R \) different operating conditions are parametrized simultaneously using a common denominator model. To perform the estimation, a weighted total least squares initialization is used (Pintelon et al. [1998]). It is followed by a sample maximum likelihood estimation (Pintelon et al. [2011b]; Pintelon and Schoukens [2012]). The frequency dependent estimation weights for the FRF are inversely proportional to the estimated sample variances \( \hat{\sigma}^2_{\hat{G}_{bla}^{[i_r]}}(k) \) of the BLAs for the \( R \) different operating conditions of the system. This results in a parametrized version of the different BLAs \( \hat{G}_{bla}^{[i_r]}(q, \hat{\theta}_{bla}) \):

\[
\hat{G}_{bla}^{[i_r]}(q, \hat{\theta}_{bla}) = \frac{\hat{d}_0^{[i_r]} + \hat{d}_1^{[i_r]}q^{-1} + \ldots + \hat{d}_{nd}^{[i_r]}q^{-nd}}{\hat{c}_0 + \hat{c}_1q^{-1} + \ldots + \hat{c}_{nc}q^{-nc}},
\]

where the denominator coefficients are shared by the BLAs for the different operating conditions \( i_r \), while the numerator coefficients vary with the input operating condition \( i_r \). \( \hat{\theta}_{bla} \) contains all the denominator coefficients \( \hat{c}_i \) and all the numerator coefficients \( \hat{d}_i^{[i_r]} \) of the BLAs for the different operating conditions \( i_r \). The model order of the parametrized BLAs can be selected using standard model structure selection techniques (see for instance Ljung [1999]).
8.5.3 Decomposing the BLAs

The overall frequency dynamics need to be distributed over the different LTI systems that are present in the branches at the front and the back of the parallel Wiener-Hammerstein model. This section presents a decomposition of the numerator coefficients of the overall dynamics of the BLA into a set of basis vectors that describe the space spanned by the numerator vectors. These basis vectors are an estimate of the dynamics of each parallel branch.

A difference with the previous approaches in Schoukens et al. [2011] or Schoukens and Rolain [2012b] is that the numerators of the estimated BLAs are decomposed, rather than the nonparametric BLA transfer functions. This is possible and adequate since a common denominator model is used for the parametrized BLAs. This new method avoids the need for a frequency sampling step of the parametric BLAs, and a re-parametrization of the decomposed BLA dynamics. The process used here is explained in the sequel of this section.

First, a matrix \( \hat{D} \) is constructed containing the stacked estimated numerator coefficients of the BLAs at the different operating conditions:

\[
\hat{D} = \begin{bmatrix}
\hat{d}_0^{[1]} & \hat{d}_1^{[1]} & \ldots & \hat{d}_{n_d}^{[1]} \\
\hat{d}_0^{[2]} & \hat{d}_1^{[2]} & \ldots & \hat{d}_{n_d}^{[2]} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{d}_0^{[R]} & \hat{d}_1^{[R]} & \ldots & \hat{d}_{n_d}^{[R]}
\end{bmatrix}.
\]  

(8.17)

The underlying distortion-free version of this matrix, \( D \), is of low rank. The maximum rank of the matrix \( D \), for \( R, n_d > n_{br} \) is \( n_{br} \). Using eq. (8.13),
one can write the \( D \) matrix as:

\[
D = AB, \quad \text{(8.18)}
\]

\[
A = \begin{bmatrix}
\alpha_1^{[1]} & \alpha_2^{[1]} & \cdots & \alpha_{n_{br}}^{[1]} \\
\alpha_1^{[2]} & \alpha_2^{[2]} & \cdots & \alpha_{n_{br}}^{[2]} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_1^{[R]} & \alpha_2^{[R]} & \cdots & \alpha_{n_{br}}^{[R]}
\end{bmatrix}, \quad \text{(8.19)}
\]

\[
B = \begin{bmatrix}
b_0^{[1]} & b_1^{[1]} & \cdots & b_{n_d}^{[1]} \\
b_0^{[2]} & b_1^{[2]} & \cdots & b_{n_d}^{[2]} \\
\vdots & \vdots & \ddots & \vdots \\
b_0^{[n_{br}]} & b_1^{[n_{br}]} & \cdots & b_{n_d}^{[n_{br}]}
\end{bmatrix}, \quad \text{(8.20)}
\]

where \( \alpha_k^{[i]} \) is the gain of the \( k \)-th branch of the \( i \)-th BLA, and \( b_i^{[k]} \) is the \( i \)-th degree coefficient of the polynomial \( B_i^{[k]}(q) \prod_{j=1,j\neq k}^{n_{br}} A_j^{[j]}(q) \) (see eq. (8.13)).

The matrix \( B \) depends only on the dynamics that are present in the different branches of the system. The matrix \( A \) depends both on the system and on the input signal used.

**Assumption 8.4. Rank of \( D \):** The rank of the BLA numerator matrix \( D \) is equal to the number of parallel branches in the system.

The proposed identification method is based on a decomposition of the BLA over the different branches of the parallel Wiener-Hammerstein model. For the method to work, this decomposition should be able to separate the dynamics of each branch. This assumes that the numerator of the dynamics of one branch \( (H^{[k]}(q)S^{[k]}(q)) \) is linearly independent from the numerators of the dynamics of the other branches of the parallel Wiener-Hammerstein system.

The SVD of \( \hat{D} \) yields an orthonormal basis for the space spanned by the \( \hat{D} \)-matrix:

\[
\hat{D} = U_{bla} \Sigma_{bla} V_{bla}^T, \quad \text{(8.21)}
\]

where the columns of \( V_{bla} \) are the right singular vectors which act as an orthonormal basis for the row space of the \( \hat{D} \)-matrix, \( \Sigma_{bla} \) is a diagonal
matrix containing the singular values, and the columns of $U_{bla}$ constitute an orthonormal basis for the left hand side space.

The column vectors in $V_{bla}$ provide an estimate of the numerator coefficients of the combined front and back dynamics for each branch $k$:

$$
\hat{G}^{[k]}(q) = \frac{\hat{\delta}_0^{[k]} + \hat{\delta}_1^{[k]} q^{-1} + \ldots + \hat{\delta}_{nd}^{[k]} q^{-nd}}{\hat{c}_0 + \hat{c}_1 q^{-1} + \ldots + \hat{c}_{nc} q^{-nc}},
$$

(8.22)

where $\hat{\delta}_j^{[k]}$ is the element of the $j$-th row and $k$-th column of the matrix $V_{bla}$.

The decomposition of the $\hat{D}$ matrix results in the branch dynamics estimates $\hat{G}^{[k]}(q)$. These dynamics do not correspond to the true system dynamics, but rather are a linear combination of the true system dynamics. The true system dynamics cannot be retrieved based on the input-output data only (see Section 8.3). As a result a MIMO static nonlinearity will be estimated, while the system is composed of one SISO static nonlinearity for each branch.

The rank of the matrix $D$ corresponds to the number of parallel branches $n_{br}$ that is necessary to describe the system by Assumption 8.4. This rank can be obtained by applying a rank estimation algorithm on the singular value matrix $\Sigma_{bla}$ (Rolain et al. [1997]), that is obtained from the noisy matrix $\hat{D}$. To do so, the column covariance matrix $C_D$ of $\hat{D}$ is needed. This column covariance matrix is obtained from the covariance of the parameters estimated in the BLA parametrization step. The whitened matrix $D_{white}$ is given by:

$$
D_{white} = \hat{D} C_D^{-1/2}.
$$

(8.23)

The estimated rank of the noisy matrix $\hat{D}$ corresponds to the number of singular values of $D_{white}$ that are higher than 1 (Rolain et al. [1997]). The reader is referred to Rolain et al. [1997] for more details about the rank estimation method and its hypotheses.
8.5.4 Partitioning the Poles and Zeros

This section presents an algorithm to partition the dynamics of each branch $\hat{G}_{i_{br}}(q)$ over the front $\hat{H}^{[k]}(q)$ and the back $\hat{S}^{[k]}(q)$ dynamics. The basic idea is pretty simple: try every partition of poles and zeros in the different LTI blocks, estimate the static nonlinear block with a fixed set of nonlinear basis functions, and finally select the model that minimizes the simulation error for the considered set of inputs.

8.5.4.1 Generating all Pole and Zero Partitions

Assumption 8.5. Common poles: The front dynamic block of branch $k$ ($k = 1, \ldots, n_{br}$) and the back dynamic block of branch $j$ ($j = 1, \ldots, n_{br}$) have no common poles, whenever $k \neq j$.

This assumption allows one to assign each estimated BLA pole to either the front or the back dynamics. A pole that is present in two different branches only appears once in the BLA. This does not pose a problem, if that pole is originating from either the front or the back LTI blocks due to the common denominator approach. However, this creates a problem when that pole is present once in the front LTI block of one branch and once in the back LTI block of another branch since it can only be assigned to either the front or the back LTI blocks of the parallel Wiener-Hammerstein model.

A first step in the algorithm is to generate all possible pole and zero partitions for the different LTI blocks. The poles and zeros to be distributed are the ones obtained from the branch dynamic estimated before. Let $\hat{G}^{[k]}(q)$ be the dynamics of branch $k$ of the parallel Wiener-Hammerstein model. Under Assumption 8.5, every pole and zero of $\hat{G}^{[k]}(q)$ has to be assigned to either the front or the back LTI block of the $k$-th branch. Some of the computational aspects of this approach are discussed in Section 8.7. Complex pole and/or zero pairs are allocated pairwise to impose real coefficients in the transfer function model. The common denominator approach is preserved during the partitioning procedure. The construction of the front and
the back dynamic systems of the branch $k$ is then obtained as follows:

\[
\hat{G}^{[k]}(q) = \gamma^{(z^k_j, p_i)} \hat{H}^{(z^k_j, p_i)}(q) \hat{S}^{(z^k_j, p_i)}(q),
\]

\[
= \gamma^{(z^k_j, p_i)} \frac{\hat{B}^{(z^k_j)}(q) \hat{B}^{(z^k_j)}(q)}{\hat{A}^{(p_i)}(q) \hat{A}^{(p_i)}(q)}
\]

for all possible pole partitions $\{p_i\}$, and for all possible zero partitions $\{z^k_j\}$ of branch $k$. In eq. (8.25) the subscript $h$ denotes the front dynamic block, and subscript $s$ denotes the back dynamic block. $\gamma^{(z^k_j, p_i)}$ denotes a gain factor that depends on the particular pole and zero partition.

### 8.5.4.2 Estimating the Static Nonlinearity

The static nonlinearity is estimated for every possible pole-zero partition $\{p_i, z^k_j\}$ of every branch $k$.

This estimation is linear in the parameters when the nonlinearity is expressed as a linear combination of nonlinear basis functions (such as multivariate polynomial basis functions, piecewise linear basis functions, or radial basis function networks with a fixed width and a fixed center):

\[
\hat{r}^k(t) = \sum_{i=1}^{n_f} \hat{\gamma}^{[k]}_i f_i(\hat{x}^1(t), \ldots, \hat{x}^{n_{br}}(t)),
\]

where $\hat{\gamma}^{[k]}_i$ is the coefficient belonging to the $i$-th basis function $f_i$ for the $k$-th output $\hat{r}^k(t)$ of the MIMO static nonlinearity, $\hat{x}^i(t)$ is the $j$-th input of the MIMO static nonlinearity, and $n_f$ is the number of nonlinear basis functions that is selected by the user.

First, the intermediate signals $\hat{x}^{(p_i, z^k_j)}$ for pole partition $\{p_i\}$ and every possible zero partition $\{z^k_j\}$ of every branch $k$ are obtained:

\[
\hat{x}^{(p_i, z^k_j)}(t) = \frac{\hat{B}^{(z^k_j)}(q)}{\hat{A}^{(p_i)}(q)} u(t),
\]

\[
\hat{x}^{(p_i, z^k_j)}(t) = \begin{bmatrix} \hat{x}^{(p_i, z^1_j)}(t) & \ldots & \hat{x}^{(p_i, z^{n_{br}}_j)}(t) \end{bmatrix}^T,
\]

\[
\hat{x}^{(p_i, z^k_j)}(t) = \begin{bmatrix} z^1_j & z^2_j & \ldots & z^{n_{br}}_j \end{bmatrix}.
\]
Next, the MIMO nonlinearity is estimated from the intermediate signals \( \hat{x}\{p_i, z_j\} \) generated through the output filters of all the branches \( k \) to the measured output. A regressor matrix \( K^{\{p_i, z_j\}} \in \mathbb{R}^{N \times n_{br} n_f} \) is constructed using a fixed, user selected set of nonlinear basis functions \( f_1 \) to \( f_{n_f} \). For one partition of poles and zeros \( \{p_i, z_j\} \) one obtains:

\[
K^{\{k, p_i, z_j\}} = \begin{bmatrix}
\frac{\hat{B}_{s}^{(z_j)}(q)}{A_{s}^{(p_i)}(q)} f_1(\hat{x}\{p_i, z_j\}(1)) & \cdots & \frac{\hat{B}_{s}^{(z_j)}(q)}{A_{s}^{(p_i)}(q)} f_{n_f}(\hat{x}\{p_i, z_j\}(1)) \\
\vdots & \ddots & \vdots \\
\frac{\hat{B}_{s}^{(z_j)}(q)}{A_{s}^{(p_i)}(q)} f_1(\hat{x}\{p_i, z_j\}(N)) & \cdots & \frac{\hat{B}_{s}^{(z_j)}(q)}{A_{s}^{(p_i)}(q)} f_{n_f}(\hat{x}\{p_i, z_j\}(N)) 
\end{bmatrix},
\]

where \( N \) is the total number of data points used.

The coefficients of the nonlinear basis functions for the partition \( \{p_i, z_j\} \) are obtained using a linear least squares estimation:

\[
\hat{\gamma}\{p_i, z_j\} = (K^{\{p_i, z_j\}}^T K^{\{p_i, z_j\}})^{-1} K^{\{p_i, z_j\}}^T y,
\]

where \( y = [y(1) \ y(2) \ \ldots \ y(N)]^T \).

In practice, the solution is obtained using a QR decomposition. To improve the numerical conditioning of the matrix, the columns of \( K^{\{p_i, z_j\}} \) are normalized. Each column is therefore divided by its \( l^2 \)-norm.

### 8.5.4.3 Pole-Zero Pattern Selection

The simulation error \( \hat{e}\{p_i, z_j\} \) that is present between the modeled output and the measured output is computed. The partition that results in the lowest root mean square error of the simulation error is selected. From this point on, the front and the back LTI blocks, \( \hat{H}^{[k]}(q) \) and \( \hat{S}^{[k]}(q) \), and the coefficients of the static nonlinearity \( \hat{\gamma}_i^{[k]} \) are all estimated.
The modeled output \( \hat{y}(t) \) is obtained as follows:

\[
\hat{x}_k(t) = \hat{H}^{[k]}(q)u(t), \quad (8.32)
\]

\[
\hat{r}_k(t) = \sum_{i=1}^{n_f} \hat{\gamma}_i^{[k]} f_i(\hat{x}_1(t), \ldots, \hat{x}_{nb_r}(t)), \quad (8.33)
\]

\[
\hat{y}(t) = \sum_{k=1}^{nb_r} \hat{S}^{[k]}(q)\hat{r}_k(t), \quad (8.34)
\]

where the signals are as in Figure 8.3. The model parameters of the static nonlinearity are stored in the parameter vector \( \gamma \in \mathbb{R}^{n_f \times 1} \). The common denominator coefficients of the different \( \hat{S}^{[k]}(q) \) LTI blocks are stored in \( a_h \in \mathbb{R}^{(n_{ah}+1) \times 1} \) and the common denominator coefficients of the \( \hat{S}^{[k]}(q) \) LTI blocks are stored in \( a_s \in \mathbb{R}^{(n_{as}+1) \times 1} \). The numerator coefficients of the different \( \hat{S}^{[k]}(q) \) LTI blocks are stored in \( b_h \in \mathbb{R}^{nb_r(n_{bh}+1) \times 1} \) and the numerator coefficients of the \( \hat{S}^{[k]}(q) \) LTI blocks are stored in \( b_s \in \mathbb{R}^{nb_r(n_{bs}+1) \times 1} \).

### 8.5.5 Improving the Estimated Nonlinearity

The number of parameters used by a MIMO static nonlinear model that is linear in the parameters tends to grow very fast. It grows combinatorially in the case of a multivariate polynomial for an increasing number of inputs and outputs, and for an increasing model complexity (e.g. the degree of the multivariate polynomial). Static nonlinear models that are nonlinear in the parameters, such as neural networks, can be less sensitive to this problem if properly tuned. For a standard feed-forward neural network with one hidden layer and a linear output layer, the number of parameters grows linearly with the number of input and outputs, and linearly with the complexity (number of neurons) of the neural network. Furthermore, a feed-forward neural network with one hidden layer is shown to be a universal approximator of continuous functions Cybenko [1989]; Hornik et al. [1989]; Hornik [1991].

An initial estimate of the nonlinear behavior and the LTI blocks that are present in the parallel Wiener-Hammerstein model can be obtained using one set of nonlinear basis functions resulting in a model that is linear in the parameters, e.g. using multivariate polynomials. In a second step, the
static nonlinearity can be re-estimated using another MIMO static nonlinear model, e.g. using a neural network, to increase the model flexibility without increasing the number of parameters too much. The decision whether or not to perform this refinement step is left to the user. This step is easily performed as the intermediate signals $\hat{x}_i$ and $\hat{r}_i$, defined in Figure 8.3, can be obtained using the model estimated in Section 8.5.4.3. The initial guess of the parameters of this second parametrization can then be further refined in a final complete optimization step, as described in Section 8.8.

8.5.6 Persistence of Excitation

Assumption 8.6. Persistence of excitation: The input signal $u(t)$ is assumed to be persistently exciting for the system.

The assumption that the excitation is persistent is a very common assumption in system identification. This section discusses what persistence of excitation means for the proposed identification procedure.

The first step in the identification algorithm is to identify the parametric BLA of the nonlinear parallel Wiener-Hammerstein system. It is important that the BLA identifies the dynamics that are present in the system correctly. Therefore, the number of excited frequencies in the input signal $u(t)$ needs be at least equal or preferably higher than $n_d + n_c + 1$.

The MIMO static nonlinearity needs also to be estimated. For this identification step to work, the matrix $K^{\{p_i,z_j\}}$ in eq. (8.29) needs to be of full rank. Put in other words, the nonlinear basis functions $f_i(\hat{x}_1(t), \ldots, \hat{x}_{n_{br}}(t))$ need to be linearly independent over the domain of the intermediate signals $\hat{x}_1(t), \ldots, \hat{x}_{n_{br}}(t)$. Consequently, the range of amplitudes that is present in $\hat{x}_1(t), \ldots, \hat{x}_{n_{br}}(t)$ needs to be sufficiently large.

Furthermore, the rank constraint of Assumption 8.4 does not only have consequences for the system. It also determines the choice of the different setpoints of the input signals. These setpoints are chosen to ensure that the rank of the matrix $D$ is equal to $n_{br}$.

203
8. Parallel Wiener-Hammerstein

8.5.7 Consistency of the Initial Estimates

This section shows the consistency of the proposed estimator when a linear-in-the-parameters nonlinearity model is used to describe the MIMO static nonlinearity.

To model the system exactly, the system needs to be contained in the reachable model set. Otherwise model errors will be present and the estimated parameters will not converge to the true parameters of the system. This requirement is formalized in the following assumption.

**Assumption 8.7. Model set:** The exact model of the system is contained in the reachable model set.

**Theorem 8.2.** The proposed estimator is a strongly consistent (convergence with probability 1 for \( N \to \infty \)) estimator of the class of parallel Wiener-Hammerstein systems under the following assumptions:

- the common assumption of zero-mean additive noise at the system output only is fulfilled (Assumption 2.2),
- the excitation signal \( u(t) \) is assumed to belong to the Riemann equivalence class of asymptotically normally distributed signals (Assumption 8.1),
- the input signal \( u(t) \) is persistently exciting the system (Assumption 8.6),
- the system is contained in the reachable model set (Assumption 8.7). This is needed for the BLA decomposition step, and for the estimated parameters to converge to the true parameters of the system up to the degenerations that are present in the model representation (see Section 8.3),
- all the dynamics that are present in the system are also present in the BLA of that system (Assumptions 8.2 and 8.3),
- the poles that are present in \( H[k](q) \) are not present in \( S[j](q) \), \( \forall k \neq j \) (Assumption 8.5).
8.5. Identification through BLA Decomposition

Proof. Due to Assumptions 8.7, 8.2, 8.3, 8.6 and Theorems 8.1 and 2.1, the poles of the BLA are a consistent estimate of the poles of the system representing the dynamics that are present in each branch. The matrix $\hat{D}$, defined in eq. (8.17), is of low rank for a model in the reachable model class (Assumption 8.7). The rank of the matrix $\hat{D}$ is a consistent estimate of the number of parallel branches that is present in the system (see eq. (8.18) and Assumption 8.4). The columns of the matrix $V_{bla}$ that correspond to the significant singular values are a consistent estimate for the numerators of the BLA. Hence the zeros and poles that are present in each branch are estimated consistently, up to the degeneration of the model structure that is explained in Section 8.3.

The last step of the estimation algorithm estimates the MIMO static nonlinearity (eq. (8.30)) for every possible pole-zero allocation. This problem is linear in the parameters, and it is solved with a linear least squares approach. Under Assumption 8.5, the poles and zeros that are allocated in this step are consistent estimates of the true poles and zeros that are present in the system, up to the degenerations of the model structure, as discussed in the previous paragraphs.

The estimate of the static nonlinearity is consistent for the pole-zero allocation that corresponds to the pole-zero allocation of the true system under Assumption 8.7. A bias error will be present for the other pole-zero allocations, since the selected pole-zero allocation does not correspond to the exact pole-zero allocation of the system. Thus, this step results in a consistent estimate of the LTI blocks and the static nonlinearity when considering the pole-zero allocation that results in the smallest estimation error.

The estimated parameters are consistent and converge to the true parameters under Assumption 8.7 up to the degenerations of the model structure as explained in Section 8.3.

Remark 8.1. It has been observed that the rank determination still works well in practice for small values of operating conditions $R$ (smaller than $n_d$, larger than $n_{br}$) and a finite and practically relevant number of samples $N$.
8. Parallel Wiener-Hammerstein

and realizations $M$.

8.6 Faster Branch-per-Branch Identification

A downside of the identification algorithm that is presented in Section 8.5 is the high computational load (see Section 8.7). This section presents an identification algorithm that speeds up the estimation of the parallel Wiener-Hammerstein model significantly. The main difference lies in the pole partitioning step. Where the algorithm in Section 8.5 performs the partitioning for all the branches simultaneously, this identification algorithm will consider one branch at a time.

The BLA estimation step (Sections 8.5.1 and 8.5.2) and the BLA decomposition step (Section 8.5.3) are identical to what is done in the previous algorithm. Here we focus on the pole-zero partitioning step.

![Diagram](image)

Figure 8.4: A parallel coupled Wiener-Hammerstein system with 3 parallel branches where the first branch is partitioned only.

The first branch is based on $\hat{G}^{[1]}(q)$, resulting from the SVD of the matrix $D$ (see eq. (8.17)). This describes the linear dynamics corresponding to the largest singular value of the matrix $D$. Hence, one can expect that it contains the dominant (non)linear behavior of the system. The pole-zero partitioning is performed only for the $\hat{G}^{[1]}(q)$ dynamics. The other dynamics are still taken into account, but only in a linear manner, as is shown in Figure 8.4. The static nonlinearity is estimated for all possible
8.6. Faster Branch-per-Branch Identification

partitions \( \{z_j^1\} \) and \( \{p_i\} \). The regressor matrix \( K\{p_i, z_j^1\} \) is constructed as:

\[
K\{p_i, z_j^1\} = \begin{bmatrix}
K^{(1, p_i, z_j^1)} & K^{(1, G)}
\end{bmatrix},
\]  

(8.35)

\[
K^{(1, p_i, z_j^1)} = \begin{bmatrix}
\hat{B}^{(z_j^1)}(q) f_1(\hat{x}_{1}^{(p_i, z_j^1)}(1)) & \cdot & \hat{B}^{(z_j^1)}(q) f_n(\hat{x}_{1}^{(p_i, z_j^1)}(1)) \\
\vdots & \ddots & \vdots \\
\hat{B}^{(z_j^1)}(q) f_1(\hat{x}_{1}^{(p_i, z_j^1)}(N)) & \cdot & \hat{B}^{(z_j^1)}(q) f_n(\hat{x}_{1}^{(p_i, z_j^1)}(N))
\end{bmatrix},
\]

\[
K^{(1, G)} = \begin{bmatrix}
\hat{G}^{[2]}(q) u(1) & \cdot & \hat{G}^{[n_{br}]}(q) u(1) \\
\vdots & \ddots & \vdots \\
\hat{G}^{[2]}(q) u(N) & \cdot & \hat{G}^{[n_{br}]}(q) u(N)
\end{bmatrix},
\]

following the notation of Section 8.5.4.2. The coefficients of the nonlinear basis functions and the gains \( \beta^{[k]} \) for the partition \( \{p_i, z_j^1\} \) are obtained using a linear least squares estimation:

\[
\begin{bmatrix}
\hat{\gamma}^{(p_i, z_j^1)} \beta^{[2]} \ldots \beta^{[n_{br}]}
\end{bmatrix} = \left( K^{(p_i, z_j^1)T} K^{(p_i, z_j^1)} \right)^{-1} K^{(p_i, z_j^1)T} y,
\]

(8.36)

\[
y = [y(1) \ y(2) \ldots \ y(N)]^T.
\]

In practice, the solution is obtained using a QR decomposition. To improve the numerical conditioning of the matrix, the columns of \( K\{p_i, z_j^1\} \) are normalized. Each column is therefore divided by its \( l^2 \)-norm. This results in an optimal zero partition \( \{z_{opt}^1\} \) and \( \{p_{opt}\} \) and an estimate of the static nonlinearity. Only the optimal zero allocation \( \{z_{opt}^1\} \) is retained, the other parameters (\( \{p_{opt}\}, \) the gains \( \beta^{[k]} \), and the static nonlinearity estimate) are discarded. Note that the partitioning is optimal for the partitioning of one branch only, it is not necessarily the optimal partitioning for the parallel Wiener-Hammerstein model as a whole.

Next, the second branch is considered. We perform again the pole-zero allocation scan over all possibilities \( \{z_j^2\} \) and \( \{p_i\} \), given the previously obtained zero allocation \( \{z_{opt}^1\} \) of the first branch, as is shown in Figure 8.5. Now a MIMO static nonlinearity (2-input 2-output) is estimated in every
Figure 8.5: A parallel coupled Wiener-Hammerstein system with 3 parallel branches where the second branch is partitioned only, given the estimate of the first branch.

step:

\[
K\{p_i, z_{opt}^1, z_j^2\} = \left[ K\{1, p_i, z_{opt}^1, z_j^2\} K\{2, p_i, z_{opt}^1, z_j^2\} \right] K\{2,G\}, \tag{8.38}
\]

\[
K\{1, p_i, z_{opt}^1, z_j^2\} = \begin{bmatrix}
\frac{B_{1}^{(z_{opt}^1)}}{A_{p_i}^{(p_i)}} f_1(x_1^{(p_i, z_{opt}^1)}(1), x_2^{(p_i, z_j^2)}(1)) & \cdots & \frac{B_{1}^{(z_{opt}^1)}}{A_{p_i}^{(p_i)}} f_n f_1(x_1^{(p_i, z_{opt}^1)}(1), x_2^{(p_i, z_j^2)}(1)) \\
\vdots & \ddots & \vdots \\
\frac{B_{1}^{(z_{opt}^1)}}{A_{p_i}^{(p_i)}} f_1(x_1^{(p_i, z_j^2)}(N), x_2^{(p_i, z_j^2)}(N)) & \cdots & \frac{B_{1}^{(z_{opt}^1)}}{A_{p_i}^{(p_i)}} f_n f_1(x_1^{(p_i, z_{opt}^1)}(N), x_2^{(p_i, z_j^2)}(N))
\end{bmatrix}, \tag{8.39}
\]

\[
K\{2, p_i, z_{opt}^1, z_j^2\} = \begin{bmatrix}
\frac{B_{2}^{(z_j^2)}}{A_{p_i}^{(p_i)}} f_1(x_1^{(p_i, z_{opt}^1)}(1), x_2^{(p_i, z_j^2)}(1)) & \cdots & \frac{B_{2}^{(z_j^2)}}{A_{p_i}^{(p_i)}} f_n f_1(x_1^{(p_i, z_{opt}^1)}(1), x_2^{(p_i, z_j^2)}(1)) \\
\vdots & \ddots & \vdots \\
\frac{B_{2}^{(z_j^2)}}{A_{p_i}^{(p_i)}} f_1(x_1^{(p_i, z_j^2)}(N), x_2^{(p_i, z_j^2)}(N)) & \cdots & \frac{B_{2}^{(z_j^2)}}{A_{p_i}^{(p_i)}} f_n f_1(x_1^{(p_i, z_{opt}^1)}(N), x_2^{(p_i, z_j^2)}(N))
\end{bmatrix}, \tag{8.40}
\]

\[
K\{2,G\} = \begin{bmatrix}
\hat{G}^{[3]}(q) u(1) & \cdots & \hat{G}^{[m_{br}]}(q) u(1) \\
\vdots & \ddots & \vdots \\
\hat{G}^{[3]}(q) u(N) & \cdots & \hat{G}^{[m_{br}]}(q) u(N)
\end{bmatrix}.
\]

The coefficients of the nonlinear basis functions for the partition \(\{p_i, z_{opt}^1, z_j^2\}\)
and the gains $\beta[k]$ are obtained using a linear least squares estimation:

$$
\hat{\gamma}(p_i,z_{1opt},z_{2j}) \hat{\beta}[3] \ldots \hat{\beta}[n_{br}] = \left( K^{(p_i,z_{1opt},z_{2j})} K^{(p_i,z_{1opt},z_{2j})} \right)^{-1} K^{(p_i,z_{1opt},z_{2j})} y.
$$

(8.41)

$$
y = [y(1) \ y(2) \ldots \ y(N)]^T.
$$

(8.42)

A similar reasoning can be applied for the next branches. When the last branch is considered, the static nonlinear estimate and the pole allocation are preserved. This results in the final estimate of the model parameters.

Discarding the pole partition at every step is not strictly necessary. However, by discarding and re-estimating every time the common denominator structure is preserved. One could also decide to fix the pole partition of all branches to the one that is obtained for the first branch. This would also preserve the common denominator structure. The advantage of such an approach is that even less possible pole-zero allocations need to be scanned. The disadvantage lies in a possibly worse pole allocation.

The gain in computation time obtained by this modified parallel Wiener-Hammerstein identification algorithm is discussed in Section 8.7.

This modified identification approach is no longer consistent. However, due to the SVD approach to obtain the dynamics $\hat{G}[k](q)$ of the parallel branches, one can still expect a good performance for the proposed algorithm.

## 8.7 Computational Aspects

The major part of the workload of the proposed estimation algorithm in Section 8.5 lies in the partitioning of the poles and zeros. Remember that all possible pole-zero partitions are tried in this step (Section 8.5.4). For each partition, a linear least squares estimation needs to be performed. This can be quite demanding with respect to the computation time. To be more specific, consider a BLA with $n$ poles and $n$ zeros. The number of combinations $n_{comb}$ that needs to be scanned is bounded by:

$$
2^n 2^{n_{br} \frac{n}{2}} \leq n_{comb} \leq 2^n 2^{n_{br}n},
$$

(8.43)
where \( n_{br} \) is the number of parallel branches of the model.

The upper limit is reached when only real poles and zeros are present in the decomposition of the BLA, while the lower limit is reached when all poles and zeros of the BLA decomposition appear in complex conjugate pairs. Typically, most poles and zeros do appear in complex conjugate pairs. In practical cases, the actual number of combinations to be scanned will therefore be closer to the lower limit.

For example, consider a BLA of order \( n_d = n_c = 10 \) in both numerator and denominator, and a 2-branch model. This results in a maximum number of combinations equal to \( 2^{10} \cdot 2^{10} \), which is about one billion combinations. Fortunately, the minimum number is only 32768. Scanning all possible combinations in the upper limit is clearly not feasible. Scanning all possible combinations for the lower limit of this example is possible, although it remains expensive.

The number of combinations that needs to be scanned can be reduced further by making some extra assumptions or by including prior knowledge about the system. A common assumption is that the linear subsystems should be proper. This reduces the number of combinations to be scanned significantly:

\[
\sum_{k=0}^{n/2} \left( \frac{n!}{k!\left(\frac{n}{2} - k\right)!} \right)^{n_{br}} \leq n_{comb} \leq \sum_{k=0}^{n} \left( \frac{n!}{k!(n-k)!} \right)^{n_{br}}
\]  

(8.44)

Considering the same example as above, this results in maximum of 184756, and minimum of 252 combinations. Scanning all possible combinations in the upper limit is feasible in about a day (considering that trying one possibility takes about 0.5 seconds). Scanning all possible combinations of the lower limit is fortunately done in a couple of minutes.

The order of the separate LTI-blocks can be fixed in advance. This also reduces the number of combinations that need to be tested. Also, the speed of the algorithm can be improved further by using parallel computing techniques that are nowadays present in, for instance, Matlab and Mathematica.

The modified branch-per-branch parallel Wiener-Hammerstein identification method (see Section 8.6) has a much lower computational load. Since
the different branches are treated separately, the number of combinations to be scanned is now no longer exponential but linear in the number of parallel branches:

\[ n_{br}2^n \leq n_{comb} \leq n_{br}2^{2n}. \] (8.45)

Consider again the previous example: a BLA of order \( n_d = n_c = 10 \) in both numerator and denominator, and a 2-branch model. This results, with the modified algorithm, in a maximum number of combinations equal to \( 2 \cdot 2^{20} \), which is about two million combinations (a factor 500 less!). The minimum number is only 2048, a factor 16 less compared to the standard algorithm. Scanning all possible combinations in the upper limit starts to become feasible, but still takes a vast amount of time. Scanning all possible combinations for the lower limit of this example is done in a couple of minutes. The number of combinations to scan in the modified algorithm can be further reduced by making some extra assumptions or by including prior knowledge about the system. A common assumption is again that the linear subsystems should be proper as is already illustrated above.

### 8.8 Nonlinear Optimization

Joining all the previous estimation steps allows one to obtain the model parameter vector as the estimation of a partition of the parameter vector \( \theta \) in non-overlapping subsets as is explained in Section 3.5.6. Note that the parameter vector \( \theta \in \mathbb{R}^{(n_{br}(n_d+2)+n_f+n_c+2)\times 1} \) now contains all the parameters of the model: \( \gamma, a_h, a_s, b_h \) and \( b_s \). The Jacobian of the cost function is rank deficient due to the identifiability issues (see Section 6.3). The related degenerations \( (2n_{br} + 2n_{br}^2) \) in the Jacobian need to be taken into account during the optimization.
8.9 Simulation Example

8.9.1 System and Signals

The proposed parallel Wiener-Hammerstein identification algorithms are illustrated on a simulation example. The simulated system is a 2-branch parallel Wiener-Hammerstein system. The dynamics of the first branch are given by $H^{[1]}(q)$ and $S^{[1]}(q)$. $H^{[1]}(q)$ is a second order Chebychev type 1 filter with a 3dB in-band ripple and a cut-off frequency at 0.15 $f_s$. $S^{[1]}(q)$ is also a second order Chebychev type 1 filter with a 3dB in-band ripple and a cut-off frequency at 0.08 $f_s$. The static nonlinearity $f^{[1]}(x_1)$ is a polynomial of degree 3:

\[
H^{[1]}(q) = \frac{0.0860 + 0.1721q^{-1} + 0.0860q^{-2}}{1 - 1.0794q^{-1} + 0.5655q^{-2}} ,
\]

\[
S^{[1]}(q) = \frac{0.0194 + 0.0387q^{-1} + 0.0194q^{-2}}{1 - 1.6592q^{-1} + 0.7970q^{-2}} ,
\]

\[
f^{[1]}(x_1) = x_1 + 0.05x_1^2 + 0.1x_1^3.
\]

The dynamics of the second branch are given by $H^{[2]}(q)$ and $S^{[2]}(q)$. $H^{[2]}(q)$ is a second order Chebychev type 1 filter with a 5dB in-band ripple and a cut-off frequency at 0.125 $f_s$. $S^{[2]}(q)$ is a second order Chebychev type 2 filter with a 50dB attenuation at 0.25 $f_s$. The static nonlinearity $f^{[2]}(x_2)$ is a polynomial of degree 3:

\[
H^{[2]}(q) = \frac{0.0451 + 0.0902q^{-1} + 0.0451q^{-2}}{1 - 1.3860q^{-1} + 0.7069q^{-2}} ,
\]

\[
S^{[2]}(q) = \frac{0.0085 + 0.0057q^{-1} + 0.0085q^{-2}}{1 - 1.7766q^{-1} + 1.7992q^{-2}} ,
\]

\[
f^{[2]}(x_2) = u + 0.01x_2^2 + 0.05x_2^3.
\]

The static nonlinearities $f^{[1]}$ and $f^{[2]}$ are shown in Figure 8.6.

White additive Gaussian noise with a standard deviation of 0.001 is added to the output of the system.

A random phase multisine is used as an input signal. 16 realizations at 5 different amplitudes are applied to the system. Each realization is measured during one steady state period consisting of 4096 samples. All
8.9. Simulation Example

Figure 8.6: The static nonlinearities $f^{[1]}$ (blue) and $f^{[2]}$ (red) that are present in the simulated parallel Wiener-Hammerstein system.

frequencies in the range $]0, f_s/2[$ are excited. The random phases are uniformly distributed between $[0, 2\pi]$. The rms-value of the input signals is logarithmically spaced from 0.1 to 2. This results in a signal-to-noise ratio at the output ranging from 30 dB to 58 dB depending on the input amplitude.

The system is sampled at a sampling frequency $f_s$ of 4096 Hz.

8.9.2 Model Estimation

The BLAs of the system at the different amplitudes are estimated in a first identification step using the robust method (see Section 2.3.1) followed by a common denominator parametrization of the nonparametric BLA (see Section 8.5.2). The estimated BLAs are shown in Figure 8.7. Only a small change in the shape of the BLAs for variations of the input rms-values can be observed. This information proves to be sufficient to obtain an estimate of the dynamics that are present in each branch. 2 or 3 parallel branches are detected in Figure 8.8 by the singular value decomposition of the white numerator matrix $D_{\text{white}}$ (Section 8.5.3). The third singular value is just above the detection threshold (equal to 1, or 0 dB). This can
8. Parallel Wiener-Hammerstein

Figure 8.7: The parametrized estimate of the BLAs of the parallel Wiener-Hammerstein system for the different input rms-values. A minor change in the shape of the BLAs for changing input rms-values can be observed.

be due to the low number of BLAs that are used and due to the noise on the measurements. A 2-branch model is proceeded with. Next, the static nonlinearity of each branch is estimated. Finally, the parameters are refined in a nonlinear optimization step, optimizing all the parameters simultaneously.

The model orders of the LTI blocks are fixed to the correct model orders that are present in the system (numerators and denominators of order 4, both for the front and the back LTI blocks). The proposed identification methods can scan all model orders themselves, but they need very high quality data to select the correct model order. Especially the branch-per-branch identification performs worse than the standard identification method with respect to the model order selection. The model order selection is much easier after the nonlinear optimization.

The estimated models are validated using a different realization of the random phase multisine of the lowest and the highest rms-value of the input. These realizations have not been used during the estimation. The time domain representation of the measured validation output and the residual error ($\hat{y}(t) - y(t)$) are shown in Figure 8.9 and Figure 8.10. The frequency
domain representation of the validation output and the magnitude of the complex error between the measurement and the modeled output are shown in Figure 8.11. It is clear from Figure 8.11 that the output of the parallel Wiener-Hammerstein system is quite nonlinear for the highest input rms-value. There is a gap of 15 to 20 dB between the output spectrum and the total distortion variance for the highest rms-value of the input. The system behaves quite linear for the lowest rms-value of the input, there is a 40 dB gap between the output spectrum and the total distortion level.

The standard identification method (Section 8.5) and the branch-per-branch identification method (Section 8.6) obtain the same initial estimate. They obtain exactly the same pole-zero partition over the different branches. Hence, they also obtain exactly the same validation results as is shown in Table 8.1.

The initial model estimate results in residual errors which are well below the output spectrum (about 40 and 60 dB for the lowest and the highest excitation level respectively). The residual errors are almost coinciding with the noise distortion (Table 8.1), and are well below the total distortion level for the highest excitation level (Figure 8.11). This shows that the initial model estimate models the nonlinear behavior of the device over a broad range.
Table 8.1: Validation of the models obtained with the standard and the branch-per-branch method before and after optimization. The rms value of the output residuals (\(\text{rms}(y - \hat{y})\)) are shown.

<table>
<thead>
<tr>
<th>input rms value</th>
<th>0.1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>standard method (init)</td>
<td>0.00099</td>
<td>0.00208</td>
</tr>
<tr>
<td>standard method (optimized)</td>
<td>0.00099</td>
<td>0.00105</td>
</tr>
<tr>
<td>branch-per-branch (init)</td>
<td>0.00099</td>
<td>0.00208</td>
</tr>
<tr>
<td>branch-per-branch (optimized)</td>
<td>0.00099</td>
<td>0.00105</td>
</tr>
</tbody>
</table>

range of the input excitation level. After optimization the residual error is further reduced. It coincides with the noise distortion level that is added during the simulations for both the low and high input excitation levels (see Table 8.1).

8.10 Conclusion and Discussion

Two identification methods for parallel Wiener-Hammerstein systems are presented which start from input-output data only. In the first step, the best linear approximation is estimated for different input excitation levels. In the second step, the dynamics are decomposed over a number of parallel orthogonal branches. Next, the dynamics of each branch are partitioned into a linear time invariant subsystem at the input and a linear time invariant subsystem at the output of each branch of the model. The static nonlinear block is also estimated during this step using a model that is linear in the parameters.

The two proposed approaches differ in this partitioning step, where the first algorithm partitions all the branches simultaneously, the second algorithm only partitions one branch at the time. This results in a much faster partitioning for the second algorithm. This comes at the cost of a possibly worse performance of the estimation.

Finally, a nonlinear least squares optimization of the parameters of all blocks together can be performed to refine the estimates. The consistency, and the computational complexity of the proposed initialization approach are discussed.
Figure 8.9: Time domain validation of the parallel Wiener-Hammerstein model obtained with the standard identification procedure for a low rms-value of the input (a) and a high rms-value of the input (b). The noisy system output is depicted by the full blue line, the output of the optimized model is depicted by the dashed green line. Note that there is hardly a noticeable difference.
Figure 8.10: Time domain validation of the parallel Wiener-Hammerstein model obtained with the standard identification procedure before and after optimization for a low rms-value of the input (a) and a high rms-value of the input (b). The model error before optimization is represented by the red line. The model error after optimization is represented by the green line.
Figure 8.11: Frequency domain validation of the parallel Wiener-Hammerstein model obtained with the standard identification procedure before and after optimization for a low rms-value of the input (a) and a high rms-value of the input (b). The noisy system output is depicted by the full blue line. The model error before optimization is represented by the red dots. The model error after optimization is represented by the green dots. The variance of the total distortion that is present at the output is shown by top black dots.
Applications

9.1 Introduction

Some of the methods that are discussed in the previous chapters are applied to two real life applications and two dedicated hardware realizations of (parallel) Wiener-Hammerstein systems. Two real life applications are selected in different scientific disciplines. The glucoregulatory system in Section 9.2 shows the usefulness of the methods in the (bio)medical sector. The Doherty amplifier in Section 9.5 illustrates the usefulness of the methods for the modeling of microwave amplifiers. Specifically developed hardware is used to compare the methods developed here to state of the art methods from the literature. The Wiener-Hammerstein benchmark is tackled in Section 9.3. The parallel Wiener-Hammerstein demonstrator system is modeled in Section 9.4. These examples illustrate the robustness of the methods and the practical usefulness of the developed approaches.

9.2 Glucoregulatory System

Based on Marconato et al. [2014]

In this section, several advanced data-driven nonlinear identification techniques are compared on the same specific problem: a simplified glucoregulatory system modeling example. This problem represents a challenge
in the development of an artificial pancreas for T1DM treatment. This application demands good nonlinear models to design accurate closed-loop controllers to regulate the glucose level in the blood. Both block-oriented and state space models are used to describe both the dynamics and the nonlinear behavior of the insulin-glucose system, and the advantages and drawbacks of each method are pointed out. The obtained nonlinear models are accurate in simulating the patient’s behavior, and some of them are also sufficiently simple to be considered in the implementation of a model-based controller to develop the artificial pancreas.

Within the class of block-oriented models, single branch and parallel Wiener models are selected. Within the state space models, polynomial nonlinear state space and neural network based state space models are considered. The focus here will be put on the results obtained with the single branch Wiener model.

9.2.1 Problem Formulation

The application considered in this work is the identification of the glucoregulatory system. The idea is to take a first step towards the estimation of a model that can be used to implement a controller for the development of the artificial pancreas for T1DM patients.

The glucoregulatory system normally has two inputs: the insulin and the meal intake. The goal is to model the behavior of one branch of the two-input system, namely the insulin-glucose subsystem. Note that, for a glucoregulatory system, the insulin is indeed the control variable. Hence, it is natural to start the modeling by focusing on the most important insulin-glucose branch.

The Meal model (Dalla Man et al. [2007]) is employed here as a simulation model to generate the data. This previously developed complete model is used to generate experimental data that are repeatable and to avoid the use of human tests. The meal intake (the second input signal in the Meal model) is kept equal to zero to produce the input-output data for the insulin-glucose subsystem, i.e. the branch of the glucoregulatory
system on which we focus in this section. Note that the physical model is only used to reproduce the patient’s input-output relationship for data generation purposes, while in the remainder the simpler behavioral models will be estimated to model the system input-output behavior, based on the generated data.

9.2.2 Considered Model Structures

The first model structure that is considered is the Wiener model. It is identified using the techniques that are described in Chapter 3. Next to a linear-in-the-parameters polynomial model for the static nonlinearity, a nonlinear-in-the-parameters neural network is also used as an alternative representation of the static nonlinearity. This model will be denoted as Wiener-NN in the remainder of this section. Both models obtain good results on the example that is studied in the next subsections, even though the system clearly does not belong to the model class.

The second model structure under consideration is a Wiener-Schetzen model. This is a type of parallel Wiener model. It is identified using the approach described in Tiels and Schoukens [2011].

Finally, two types of nonlinear state space models are considered: a polynomial nonlinear state space (PNLSS) and a neural network nonlinear state space (NN-NLSS) model. They are identified using the approaches that are described in Paduart et al. [2010] and in Marconato et al. [2012a] respectively.

9.2.3 The Insulin-Glucose Modeling Problem: Multisine Input

In this first example, the input signal $u(t)$ shown in Figure 9.1 is used to excite the system. It is a random phase multisine, with a sampling time of 20 minutes and a period length of 2000 points. The corresponding output signal $y(t)$ is also shown in Figure 9.1. The random phase multisine excitation signal is chosen since it helps to reveal and study the dynamics
of the system, while a more realistic type of signal will be considered later on in Section 9.2.4.

The multisine input signal is applied at 12 different operating points of the system, ranging from 100 to 550 pmol/min. Hence 12 different offset levels of the input (i.e. basal insulin) are considered. At every operating point two periods of the signal are available. Two datasets are generated taking into account two different input rms levels: the first one is used as estimation set to build the models with, and the second one is used as a validation set to test the performance of the models on previously unused data.

Note that the large time scale of the signals considered in this example is used only for the simulation and study of the different nonlinear models. It is part of the experiment design to choose it such that the dynamics of the system will indeed be present in the data.

9.2.3.1 Results for the BLA

The results obtained by the different methods are quantified in terms of the following relative error criterion, that is calculated at each of the 12 operating points for each estimated model:

\[ e_{rel}(\%) = \frac{\|y(t) - \hat{y}(t)\|}{\|y(t) - \bar{y}\|} \cdot 100, \]  

(9.1)

where \(\hat{y}(t)\) is the modeled output, \(\bar{y}\) is the mean value of the output taken over the time at each operating point, and \(\|\cdot\|\) is the Euclidean norm.

The number of model parameters is also an important outcome for the specific application, since it gives an indication of the model complexity. This should be kept as low as is possible for control purposes.

First the results of a linear model obtained with the BLA are discussed. The magnitude of the frequency response of a third order BLA (order 3 in both numerator and denominator) for the 12 operating points is shown in Figure 9.2. Some changes in the amplitude and shape of the BLA can be observed at the different operating points. This is due to the nonlinear behavior of the system.
Figure 9.1: (a) Excitation signal: random phase multisine (one period) at one operating point of the system. (b) Output signal, obtained by simulating the Meal model with the excitation signal above as input.
9. Applications

Figure 9.2: Magnitude response of the parametrized transfer function for 12 operating points.

Table 9.1: Results of the BLA computed at operating point 7 on the validation set, for the different operating points (o/p).

<table>
<thead>
<tr>
<th>o/p</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_{rel}$ (%)</td>
<td>78.3</td>
<td>70.4</td>
<td>60.0</td>
<td>47.1</td>
<td>37.3</td>
<td>19.4</td>
<td>7.0</td>
<td>27.5</td>
<td>52.8</td>
<td>80.0</td>
<td>108.2</td>
<td>137.0</td>
</tr>
</tbody>
</table>

To illustrate the performance of the linear model, the BLA that is estimated for operating point 7 is used as a linear model for the system. The validation set for each operation point is used as an input signal to obtain $e_{rel}$ for all the operating points. The corresponding relative error of the BLA of operating point 7 is reported in Table 9.1.

The BLA performs very well only around operating point 7, as could be expected. It fails completely to predict the system behavior in the regions further away from it as the excitation signal class used for identification and validation are now very different. Note the presence of relative error values that are larger than 100% at operating points 11 and 12. This is due to the fact that the linear model error is larger than the output signal. Similar conclusions can be drawn when the BLA model is taken at another operating point. Therefore, it can be safely stated that a single LTI model
9.2. Glucoregulatory System

Table 9.2: Characteristics of the nonlinear models. Two types of nonlinear functions are used: polynomials (poly) and tanh(·) sigmoid functions (tanh). \( d \) represents the degree of the polynomials, while \( n \) is the number of neurons in the hidden layer of the NN. For the NN-based NLSS, \( n_f \) and \( n_g \) are the number of neurons used in the state and output equation respectively. Remark: the linear part in the Wiener-Schetzen model consists of 10 parallel branches.

<table>
<thead>
<tr>
<th>Model</th>
<th>BLA order</th>
<th>Nonlinearity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiener</td>
<td>4</td>
<td>poly, ( d = 4 )</td>
</tr>
<tr>
<td>Wiener-NN</td>
<td>3</td>
<td>tanh, ( n = 4 )</td>
</tr>
<tr>
<td>Wiener-Schetzen</td>
<td>4</td>
<td>poly, ( d = 2 )</td>
</tr>
<tr>
<td>PNLSS</td>
<td>3</td>
<td>poly, ( d = 3 )</td>
</tr>
<tr>
<td>NN-NLSS</td>
<td>3</td>
<td>tanh, ( n_f = 3 ), ( n_g = 4 )</td>
</tr>
</tbody>
</table>

is not suited to describe the overall input-output behavior of the system.

9.2.3.2 Results for the BLA

The obtained BLA is now used as a first step to build nonlinear models that describe the glucoregulatory system in a more accurate way, at all operating points simultaneously. The results of the Wiener models and the other nonlinear models are discussed here. Table 9.2 summarizes the main features of the nonlinear models obtained by applying the different methods on the insulin-glucose modeling problem.

The obtained results are shown in Table 9.3. The performance of the different nonlinear models is compared in terms of relative error on the validation set and in terms of the number of parameters. The BLA results are also reported as a reference.

A first observation is that the proposed nonlinear identification methods allow one to significantly improve the results obtained with the BLA. In particular, the average relative error over the 12 operating points is reduced approximately by a factor of 3 for all nonlinear models. Moreover, with the obtained nonlinear models it is possible to consequently get low error results at all operating points. Remember that this was not the case for the linear model.
9. Applications

Table 9.3: Results of the nonlinear models: multisine excitation. The averaged relative error taken over the 12 operating points on the validation set is reported, together with the minimum and maximum relative error values (in brackets) and the number of model parameters. As a comparison, also the results of a third order BLA are shown.

<table>
<thead>
<tr>
<th>Model</th>
<th>Average $e_{rel}$ (min – max)</th>
<th>%</th>
<th># parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiener</td>
<td>21.1 (8.2 – 33.8)</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Wiener-NN</td>
<td>14.6 (3.8 – 37.0)</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Wiener-Schetzen</td>
<td>25.2 (2.2 – 46.1)</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>PNLSS</td>
<td>19.4 (8.9 – 44.9)</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>NN-NLSS</td>
<td>17.2 (7.2 – 30.8)</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>BLA</td>
<td>60.4 (7.0 – 137.0)</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

A side note is required for the number of parameters used in the Wiener-Schetzen model: the large number of parameters is due to the fact that the output of the obtained model consists of a weighted sum of the outputs of three different nonlinear submodels. This was necessary in order to build a single nonlinear model that has a good performance at all operating points.

Finally, one can argue that a Wiener model can also be described by a general NLSS representation. Following this theoretical reasoning, it may appear surprising to note that the results obtained for the NN-NLSS are worse than for the Wiener-NN model. A possible explanation is that the higher error values for the NN-NLSS model are due to the presence of local minima trapping during the nonlinear optimization step.

This shows that it is important for the user to select a good model structure. A good model structure is not necessarily a very general one, but rather one that only has the complexity needed to describe the system under study. The more local minima are introduced by a (more complex) model structure, the higher the quality of the initial estimates needs to be to yield a satisfactory model.
9.2.4 The Insulin-Glucose Modeling Problem: A more Realistic Case

To reproduce a more realistic scenario, the input signal shown in Figure 9.3 is used as an excitation. It consists of a random phase multisine superimposed on a band-limited pulse signal. The multisine component of the signal represents the small variations of the insulin input, while the pulses represent the insulin bolus injected at specific moments (e.g. meal-time, corrections for high glucose levels, and so on). This method of creating a band-limited pulse signal and combining it with a multisine is a procedure that brings the excitation closer to a realistic signal but remains extremely usable for the purpose of system identification (Widanage et al. [2011]).

This signal is again applied at 12 different operating points. Two periods of the signal are available, each containing 40000 samples with a sampling period equal to 1 minute. The results obtained with the new dataset are summarized in Table 9.4.

The models are re-estimated with the new dataset. Due to the very large amount of data in the new dataset, PNLSS and NN-based NLSS models could not be obtained, because the required computational time becomes prohibitive. The reason for this lies in the very high computational complexity that is required for the nonlinear optimization step present in both methods. The calculation of the Jacobian is carried out in a recursive way at each iteration of the Levenberg-Marquardt algorithm, and this results in a drastic increase of the total computational time (Paduart [2008]; Marconato et al. [2014]).

It can be observed that also for this example the nonlinear models yield very good results: the average error values are four times smaller than the error of the BLA.

Figure 9.4 shows the true output and the modeled output for the Wiener-NN for two regions: operating point 1, for which the model performs worst (highest error value: 27.4%), and operating point 7, for which the model performs best (lowest error value: 4.2%). It can be observed that the shape of the signal is modeled good in both cases, although a small error appears
Figure 9.3: Excitation signal: band-limited pulses combined with a random phase multisine signal, at one operating point of the system.

Table 9.4: Results of the nonlinear models: band-limited pulses and multisine excitation. The averaged relative error taken over the 12 operating points on the validation set is reported, together with the minimum and maximum relative error values (in brackets) and the number of model parameters. As a comparison, also the results of a fourth order BLA are shown.

<table>
<thead>
<tr>
<th>Model</th>
<th>Average $e_{rel}$ (min – max)</th>
<th>%</th>
<th># parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiener</td>
<td>14.3 (6.5 – 23.6)</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Wiener-NN</td>
<td>12.8 (4.2 – 27.4)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Wiener-Schetzen</td>
<td>11.9 (5.8 – 22.3)</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>BLA</td>
<td>54.1 (13.7 – 105.7)</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>
9.2. Glucoregulatory System

Figure 9.4: Results of the Wiener-NN model: band-limited pulses and multi-sine excitation. True output (solid blue line) and modeled output (dashed green line) for operating points 1 and 7 are shown.

at the peaks of the output of operating point one.

9.2.5 Discussion

When comparing the different nonlinear identification methods that have been applied to model the insulin-glucose system, the neural network nonlinearity seems to be better suited to describe the nonlinearity of the system than the polynomial functions. This can be observed when comparing the
Wiener and Wiener-NN models on the one hand, and the PNLSS and the NN-based NLSS models on the other hand. Models containing sigmoid nonlinearities are characterized by lower average error and lower minimum error values. This could be the case due to the saturating behavior of the sigmoid function. Note, however, that for the neural network nonlinearity a nonlinear optimization is required, while a simple least squares problem is sufficient to solve for the polynomials.

Wiener models are simple models that are quite parsimonious in terms of number of parameters, while the NLSS and (especially) the Wiener-Schetzen approaches result in general in more complex models, with a significantly higher number of parameters.

Wiener-Schetzen models can yield very high accuracy (at specific operating points), although in the multisine excitation case their overall performance is the worst. However, they obtain the best results in terms of average error in the more realistic scenario, with the pulses and multisine excitation signal.

The NLSS methods achieve low errors, but they show severe limitations when the size of the dataset used for the estimation becomes very large. Moreover, they seem to be more sensitive to the presence of local minima.

In summary, when looking at both error values and number of parameters, the Wiener-NN approach seems to offer the best trade-off between model accuracy and simplicity in both examples. Wiener models seem to represent a good option for the implementation of a model-based controller as a fundamental block in the development of the artificial pancreas for T1DM patients. Remember that the model used in this specific application should in fact be very accurate to guarantee that no harm is caused to the patient, and simple enough to make the controller implementation easy.

Note that the good results of the Wiener models are obtained although some of the assumptions proposed in Chapter 3 are violated. The system does not belong to the model class, and the band-limited pulse excitation signal does not belong to the Riemann equivalence class of asymptotically normally distributed excitation signals. This illustrates the robustness of the proposed Wiener identification method.
9.3 Wiener-Hammerstein Benchmark

Based on Schoukens et al. [2014b]

The Wiener-Hammerstein identification method presented in Section 4.7 is validated on a set of real measurement data obtained from a physical system. More precisely, a benchmark problem that was studied during the IFAC SYSID conference in 2009 is considered (Schoukens et al. [2009b]).

9.3.1 Experimental Setup

The benchmark data are generated from a nonlinear electronic circuit designed to have a Wiener-Hammerstein structure, as is shown in Figure 9.5. The first LTI block is a third order Chebychev low-pass filter with 0.5 dB ripple and a cut-off frequency at 4.4 kHz. The second LTI block is a third order inverse Chebychev low-pass filter with a -40 dB stop band starting at 5 kHz. The static nonlinearity is a resistor-diode network. The system is excited by low-pass filtered Gaussian noise, with cut-off frequency set at 10 kHz. The input and output signals are measured with a sampling frequency equal to 51.2 kHz. The measured dataset is split in two parts: a first set of $10^5$ samples to estimate the parameters of the model, and a second set of $88\times10^3$ samples to validate the estimated model. Note that this experimental setup does not satisfy all the assumptions made in Section 4.7. More specifically the input of the Wiener-Hammerstein system is not white. However, the proposed algorithm still works well since the cut-off frequency of the low-pass filters in the system is approximately half of the bandwidth of the input signal. This means that the input can be
considered to be white within the bandwidth of the Wiener-Hammerstein system.

9.3.2 Estimating the Parameters

The first step in the identification procedure is to identify the BLA and QBLA of the system, as is explained in Section 4.7. For this purpose the $10^5$ samples of the identification dataset are partitioned in smaller sets of $N = 10^3$ samples per set, resulting in $M = 100$ realizations. From these sets $\hat{G}_{BLA}(e^{j\omega_k T_s})$ and $\hat{G}_{QBLA}(e^{j\omega_k T_s})$ are estimated.

The dynamics of the two LTI blocks of the system are estimated by splitting the BLA nonparametrically. This results in an initial nonparametric estimate of the dynamics $\hat{H}(e^{j\omega_k T_s})$ and $\hat{S}(e^{j\omega_k T_s})$ shown in Figure 9.6. As discussed in Section 4.3, a linear phase shift can be present in the nonparametric estimate. This complicates the parametrization (see Section 4.7.3). Both the front and the back LTI blocks are parametrized using a rational function in the $z$-domain of third order in both the numerator and the denominator. The amplitude and the phase spectrum of both the nonparametric and the parametrized LTI blocks are shown in Figure 9.6. Note the presence of a linear phase shift between the nonparametric and parametric estimate in Figure 9.6.

The nonlinearity is parametrized by a piecewise linear function with 30 knots, where the position of the knots on the $x$-axis is fixed. The $x$-position of the knots is chosen such that the number of measured samples between two consecutive knots is constant.

9.3.3 Validation

In Schoukens et al. [2009b] the mean value $\mu$, the standard deviation $s$ and the root mean square $e_{rms}$ of the simulation error, show the performance of the method used. All quantities are calculated both for the estimation and the validation dataset. The results are shown in Table 9.5. Figure 9.7 shows the modeled output and the model error in the time and the frequency domain.
9.3. Wiener-Hammerstein Benchmark

**Figure 9.6:** Estimated magnitude spectrum (top) and phase spectrum (bottom) of the LTI blocks. Both the nonparametric (dashed line) and the parametric estimate (solid line) of the front LTI block (blue) and of the LTI block at the back (red) are shown.
Table 9.5: Mean value of the error $\mu$, the standard deviation of the error $s$, and the $e_{rms}$ value of the error for both the estimation (est.) and the validation (val.) dataset before and after the final full optimization step, and for a linear BLA model of order 6. All the values are shown in mV.

<table>
<thead>
<tr>
<th></th>
<th>optimized</th>
<th>initial</th>
<th>BLA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est.</td>
<td>val.</td>
<td>est.</td>
</tr>
<tr>
<td>$\mu$</td>
<td>$2 \times 10^{-8}$</td>
<td>0.023</td>
<td>0.133</td>
</tr>
<tr>
<td>$s$</td>
<td>0.270</td>
<td>0.278</td>
<td>31.53</td>
</tr>
<tr>
<td>$e_{rms}$</td>
<td>0.270</td>
<td>0.279</td>
<td>31.53</td>
</tr>
</tbody>
</table>

It is clear from Figure 9.7 and Table 9.5 that the method obtains a very good result on the benchmark measurements. The initial estimate provides a good starting value with an rms error of 31.6 mV. When compared to the rms error of 55.9 mV obtained for the linear BLA model the error is almost halved. This indicates that the nonlinear behavior is well captured by the initial Wiener-Hammerstein model. After optimization, the error is lowered to 0.278 mV. This shows that the method obtains good initial values, which can be used next as starting values of a nonlinear optimization scheme. This result shows as well that the nonlinear least squares optimization converges to a good quality (local) minimum.

The results obtained by the proposed approach are similar to other Wiener-Hammerstein modeling approaches such as (Sjöberg et al. [2012]; Westwick and Schoukens [2012]; Van Mulders [2012]; Lauwers [2011]), as can be seen in Table 9.6. The number of parameters of the approach presented in this work is higher than the number of parameters used by Westwick and Schoukens [2012]. This is due to a different parametrization of the static nonlinear block.

A slightly better performance is obtained with the method proposed in Section 4.7 when compared to the method that is presented in Wills and Ninness [2012] where a generalized Hammerstein-Wiener model structure is used and the approach presented in Van Mulders et al. [2010] where a polynomial nonlinear state space model is transformed into a Wiener-Hammerstein structure.
9.3. Wiener-Hammerstein Benchmark

![Time Domain Output](image_url)

*Figure 9.7:* Measured output (blue) in the time and frequency domain, and the model errors before (red) and after full optimization (green).

9.3.4 Comparison of the Proposed Approach

This section compares the method presented in Section 4.7 with the methods presented in Billings and Fakhouri [1978] and Westwick and Schoukens [2012]. All three methods use the BLA and the QBLA to generate initial values for a Wiener-Hammerstein model.
Table 9.6: The rms value of the validation error $e_{rms}$ and the number of free parameters $n_\theta$ obtained with different modeling methods: the BLA, transformed PNLSS (Van Mulders et al. [2010]; Van Mulders [2012]), incremental optimization (Tan et al. [2012]), generalized Hammerstein-Wiener (Wills and Ninness [2012]), pole-zero scan (Sjöberg et al. [2012]), pole-zero split (Sjöberg et al. [2012]; Lauwers [2011]), parametric BLA and QBLA (Westwick and Schoukens [2012]; Van Mulders [2012]) and the nonparametric BLA and QBLA approach (Section 4.7 of Schoukens et al. [2014b]).

<table>
<thead>
<tr>
<th>method</th>
<th>$e_{rms}$ (mV)</th>
<th>$n_\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLA</td>
<td>55.865</td>
<td>13</td>
</tr>
<tr>
<td>Transformed PNLSS</td>
<td>0.81</td>
<td>30</td>
</tr>
<tr>
<td>Incremental optimization</td>
<td>0.679</td>
<td>25</td>
</tr>
<tr>
<td>Generalized H-W</td>
<td>0.481</td>
<td>47</td>
</tr>
<tr>
<td>Pole-zero scan</td>
<td>0.33</td>
<td>30</td>
</tr>
<tr>
<td>Pole-zero split</td>
<td>0.30</td>
<td>64</td>
</tr>
<tr>
<td>Parametric BLA, QBLA</td>
<td>0.286</td>
<td>26</td>
</tr>
<tr>
<td>Nonparametric BLA, QBLA</td>
<td>0.279</td>
<td>44</td>
</tr>
</tbody>
</table>

9.3.4.1 Setup

The different methods are applied to simulated data obtained by adding noise to the data from the SYSID 2009 identification benchmark (Schoukens et al. [2009b]). The benchmark setup is the same as presented in Section 9.3.1. A 300 trial Monte Carlo simulation was performed as in Westwick and Schoukens [2012]. In each trial white Gaussian noise was added to the output of the benchmark data. The standard deviation of the noise is scaled such that a signal-to-noise ratio of 30 dB is obtained at the output.

The FRF of the BLA and the QBLA is obtained in the same way as is explained in Section 9.3.2. For the method presented in Billings and Fakhouri [1978] the BLA is parametrized using a sixth order discrete time model, while the QBLA is parametrized using a ninth order model. The direct term of the numerator is allowed to be nonzero. Only the first 120 frequency lines (DC excluded) are considered during the parametrization. It turns out that spectral lines at higher frequencies lie below the noise floor.
9.3.4.2 Results and Discussion

In 85% of the trials, the initial values of the method presented in Section 4.7 resulted in a model with a validation rms error that is equal to the standard deviation of the additive noise at the output. The method presented in Billings and Fakhouri [1978] resulted in a good model in 52% of the cases. In Westwick and Schoukens [2012], a good model was found for 65% of the trials within the first two splits proposed by the selection procedure outlined there. 93% of the trials result in a good model when the results of the method proposed in Section 4.7 and the method presented in Billings and Fakhouri [1978] are combined. In most trials where the proposed approach failed, the obtained nonparametric FRF was either too noisy to obtain a good initial parametric model, or the initial parametric model was unstable.

The initial values generated with the method presented in Billings and Fakhouri [1978] are generated based on the estimated coefficients of the parametrized BLA and QBLA. This method seems to be quite sensitive to the value of estimated coefficients. This can explain the high noise sensitivity of this method. The method presented in Westwick and Schoukens [2012] requires a good estimate of the poles and zeros that are present in the BLA and QBLA. This estimate is not easy to obtain due to the presence of the extra output noise added in this simulation example and the nonlinear behavior of the Wiener-Hammerstein benchmark system. Hence, in quite a lot of the cases the proposed approach does not come up with a good model.

The method presented in Section 4.7 does not depend on the parametrization of the BLA and the QBLA. The parametrization happens in a second step, where the two LTI-blocks of the model are parametrized instead of the BLA and the QBLA. This seems to increase the robustness of the method with respect to disturbing noise as is indicated by the results on this simulation example.
9.3.5 Conclusion

The result of this section shows that the proposed Wiener-Hammerstein identification algorithm of Section 4.7 performs well on measured data. The system almost belongs to the model class that is considered here: the system and the model have the same basic structure, but the system is still a hardware implementation of this structure and therefore is prone to (small) nonidealities. This means that it is difficult to find a perfect model for, for instance, the diode-resistor nonlinearities that are present in the system.

The different identification steps of the nonparametric separation algorithm (Section 4.7) are illustrated. The performance is compared with other, similar, Wiener-Hammerstein identification methods in a Monte Carlo simulation setup that is based on the measured benchmark data.

9.4 Parallel Wiener-Hammerstein

Based on Schoukens et al. [2015c]

A real-world measurement based identification is performed to illustrate the good performance of the method proposed in Section 8.5. First, the measurement setup is introduced. Next, the different steps of the model estimation procedure are shown. Finally, the validation results are discussed.

9.4.1 Measurement Setup

The device under test (DUT) is an electronic circuit designed to have a 2-branch parallel Wiener-Hammerstein structure. The front and back LTI blocks of each branch are third order continuous time filters. The static nonlinearity of each branch is realized with a diode-resistor network.

The measurement setup is similar to the setup used in Schoukens and Rolain [2012c]. The signals are generated by an arbitrary waveform generator (AWG), the Agilent/HP E1445A, sampling at 625 kHz. An internal
9.4. Parallel Wiener-Hammerstein

low-pass filter with a cut-off frequency of 250 kHz is used as a reconstruction filter for the input signal. The in- and output signals of the DUT are measured by the alias protected acquisition channels (Agilent/HP E1430A) sampling at 78 kHz. The AWG and acquisition cards are synchronized with the AWG clock, and hence the acquisition is phase coherent to the AWG. Leakage errors are hereby easily avoided. Finally, buffers are added between the acquisition cards and the in- and output of the DUT to avoid that the measurement equipment would distort the measurements.

9.4.2 Input Design

The generated input signal \( u(t) \) is a random phase multisine containing \( N = 131072 \) samples with a flat amplitude spectrum. The excited band ranges from \( \frac{f_s}{N} \) to \( f_{\text{max}} = 20 \) kHz:

\[
    u(t) = A \sum_{n=1}^{n_{\text{max}}} \cos(2\pi n \frac{f_s}{N} t + \phi_n),
\]

where \( n_{\text{max}} \) is the integer number closest to \( N \frac{f_s}{f_{\text{max}}} \). The phases \( \phi_n \) are independent uniformly distributed random variables ranging from \([0, 2\pi]\). 20 independent random phase realizations of the multisines are used at each input level to determine the BLA using the robust method (see Section 2.3.1). The input signal is scaled to obtain 5 different rms values that are linearly distributed between 100 mV and 1 V.

The signals are measured at a sampling frequency of 78 kHz, which is 8 times slower than the sampling frequency used at the generator side. This results in measured input and output signals that contain \( N = 16384 \) measured samples per period.

9.4.3 Model Estimation

This section shows how the different steps of the BLA decomposition based parallel Wiener-Hammerstein estimation algorithm (Section 8.5) are applied to the measurement example. First, the BLA of the system is measured and parametrized. Next, the estimated dynamics are distributed over the
9. Applications

Figure 9.8: The parametrized BLA for the lowest (blue) to the highest (magenta) rms level of the excitation. The FRF of the BLAs for the different rms values is shown by the full lines. The total variance on the BLAs is shown by the triangles. The noise variance on the BLAs is shown by the circles.

different LTI blocks that are present in the model. Finally, the nonlinearity is estimated and a nonlinear optimization of all the parameters of the model is performed.

9.4.3.1 BLA Estimation and Parametrization

The BLA is estimated and parametrized as discussed in Sections 8.5. The BLAs are parametrized with a discrete time rational transfer function model with a common denominator. The numerators and denominator are both of order 12. The FRFs of the parametrized BLAs are shown in Figure 9.8. Figure 9.8 also shows the noise variance and the total variance on the estimated BLAs. The total variance is the variance that accounts for the nonlinear behavior of the system and the noise that is present in the measurements (Pintelon and Schoukens [2012]; Schoukens et al. [2012a]). The small variation that can be observed in the shape of the FRF of the BLAs will prove to be sufficiently informative to decompose the dynamics over the parallel branches.

242
9.4.3.2 Splitting the Dynamics

The estimated dynamics are decomposed over the different parallel branches. Two parallel branches are retrieved by the SVD of the numerator matrix to model the system under test. The decomposed dynamics are then partitioned over the front and the back LTI blocks of the parallel Wiener-Hammerstein model. To do so, all the possible pole-zero combinations are scanned. It is assumed that all the LTI-blocks in the model are proper to reduce the number of possible combinations. As a result, a total of 140817 combinations are scanned. The mean squared simulation error is used as an error criterion. The error is evaluated using one realization taken at both the lowest and the highest input excitation level of the estimation data.

The lowest error after the pole-zero allocation scan is obtained with a model that has 4 poles and 4 zeros in the front LTI blocks, and 8 poles and 8 zeros in the back LTI blocks. This candidate model did not converge to a good local minimum after the final optimization step that is described in Section 8.8. The second lowest error after the initial pole-zero scan (before the optimization step) is obtained with a model that has 6 poles and 6 zeros in the front LTI blocks, and 6 poles and 6 zeros in the back LTI blocks. This corresponds to the hardware realization of the system under test. This model is selected to be refined in the next steps.

9.4.3.3 Estimating the Static Nonlinearity

A multivariate polynomial nonlinearity of order 7 is estimated during the partitioning of the dynamics to the front and the back LTI blocks. To increase the modeling power of the static nonlinear block, this polynomial nonlinearity is replaced by a 2-input 2-output neural network after the separation of the dynamics. The neural network has one hidden layer that contains 10 tanh(.) activation functions, and a linear output layer. A tanh(.) nonlinear function captures the saturation behavior in the system very well. Afterwards, a final simultaneous optimization of all the parameters is performed to further refine the estimated model.
9. Applications

9.4.4 Model Validation

The estimated model is validated using two different signal types: random phase multisines of different magnitudes, and a growing envelope filtered Gaussian noise signal.

9.4.4.1 Multisine Validation

The model is validated with a random phase multisine realization that is not used during the identification. The experiments are taken at 5 different rms values that are linearly distributed between 0.1 V and 1 V. The quality of the model is shown in Table 9.7 using three figures of merit: the rms value of the simulation error \( \text{rms}(e) \), the absolute mean value of the simulation error \( \mu_e \), and the standard deviation of the simulation error \( \sigma_e \), as defined below:

\[
\begin{align*}
\text{rms}(e) &= \sqrt{\frac{1}{N} \sum_{t=1}^{N} e^2(t)}, \quad (9.3) \\
\mu_e &= \left| \frac{1}{N} \sum_{t=1}^{N} e(t) \right|, \quad (9.4) \\
\sigma_e &= \sqrt{\frac{1}{N-1} \sum_{t=1}^{N} (e(t) - \mu_e)^2}, \quad (9.5)
\end{align*}
\]

where \( e(t) \) is the difference between the measured output \( y(t) \) and the simulated output \( \hat{y}(t) \).

The obtained model outperforms the BLA for every rms value of the input, as can be seen from Table 9.7. Note that a different BLA is used for every rms value of the input, while only one parallel Wiener-Hammerstein model is used for all the different rms values of the input. The rms error is a combination of the standard deviation of the simulation error, and the mean value of the simulation error.

The BLA is a linear approximation of the system, and cannot model the nonlinearities that are present in the system. The BLA can therefore not model the rms dependent constant contribution to the output that is generated by the nonlinearities. This explains the much larger mean error.
Table 9.7: Validation error on a multisine signal

<table>
<thead>
<tr>
<th></th>
<th>Validation error (mV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rms(u)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Par WH</td>
<td>0.30</td>
</tr>
<tr>
<td>WH</td>
<td>2.91</td>
</tr>
<tr>
<td>NARX</td>
<td>3.20</td>
</tr>
<tr>
<td>NOE</td>
<td>2.63</td>
</tr>
<tr>
<td>BLA</td>
<td>1.34</td>
</tr>
</tbody>
</table>

\[ \mu_e \]

\[ \text{of the model output obtained with the BLA. Also the varying nonlinear contributions in the output cannot be explained by a linear model, and will contribute to the standard deviation of the simulation error. This explains the higher standard deviation of the simulation error.} \]

The parallel Wiener-Hammerstein model approximates the static nonlinearities that are present in the system quite well. Figure 9.9 shows that the model error of the parallel Wiener-Hammerstein model is 30 to 40 dB lower than the total variance on the output, and it is only 10 dB higher than the output noise variance level. This shows that the proposed identification method captures the nonlinear behavior of the system very well.

The output of the parallel Wiener-Hammerstein model is compared to the results obtained by a Wiener-Hammerstein model in Table 9.7. This Wiener-Hammerstein model is estimated using the method proposed in Sjöberg and Schoukens [2012], which is similar to the parallel Wiener-Hammerstein modeling method of Section 8.5. It uses a neural network with one
hidden layer that contains 10 tanh(.) activation functions and a linear output layer as a static nonlinearity. The Wiener-Hammerstein model is able to obtain a model error that is lower than the one obtained by the BLAs at the different excitation levels, but the errors are still 10 to 20 times larger than the errors obtained by the parallel Wiener-Hammerstein model.

The parallel Wiener-Hammerstein model is also compared to a neural network NARX model in Table 9.7. The NARX input-output relationship is given by Billings [2013]:

\[
y(t) = f(u(t), \ldots, u(t-n_b), y(t-1), \ldots, y(t-n_a)) + e(t),
\]

where \(n_b, n_a = 12\), \(f(.)\) is a static nonlinear function, and \(e(t)\) is white additive noise. Here, \(f(.)\) is described by a neural network with one hidden layer that contains 25 tanh(.) activation functions and a linear output layer. The estimation of the NARX model is performed using the Matlab Neural Network Toolbox using the so-called series-parallel architecture. The NARX model performs quite well, its performance is similar to that of to the Wiener-Hammerstein model. The error obtained with the parallel Wiener-Hammerstein model is still 10 to 20 times smaller than the errors of the NARX model.

The result that is obtained with the NARX model is further improved using a nonlinear output error model (NOE in Table 9.7). Here, the delayed instances of the measured (noisy) outputs are no longer used in the regressor matrix, they are replaced by delayed instances of the noiseless output:

\[
y_0(t) = f(u(t), \ldots, u(t-n_b), y_0(t-1), \ldots, y_0(t-n_a)) \\
y(t) = y_0(t) + e(t),
\]

where \(y_0\) denotes the noiseless output. This corresponds to the parallel architecture in the Matlab Neural Network Toolbox. The estimation of the parameters is performed using the Matlab Neural Network Toolbox. This results in an error which is over 30% smaller than the error of the NARX model. However, the parallel Wiener-Hammerstein model still outperforms the NOE model (see Table 9.7).
9.4. Parallel Wiener-Hammerstein

Figure 9.9: Measured and simulated output spectrum of a validation dataset at the middle excitation level (rms value of the input equal to 0.55 V). The measured output is shown in blue. The model error of the parallel Wiener-Hammerstein model is shown with the green plus symbols. The model error of the BLA is shown with the red plus symbols. The noise level at the system output is shown with the bottom black circles. The total distortion level at the output is shown with the top black triangles.

9.4.4.2 Growing Envelope Validation

A second validation signal is used to assess the model quality over a broad amplitude range of the input signal. The selected input is a filtered Gaussian noise signal with an envelope that grows linearly over time:

\[ u(t) = \frac{2k}{N}(H(q)r(t)), \]

(9.8)

where \( r(t) \) represents zero-mean white Gaussian noise with a standard deviation equal to one, and \( H(q) \) is a sixth order low-pass Chebychev filter with a cut-off frequency located at 20 kHz and a passband ripple of 0.5 dB. Note that this is a generalization of the input signals that are used during the estimation. During the last part of the growing envelope input signal, the excitation amplitude is higher than the magnitude of the signals used in the estimation of the model. The instantaneous rms value of the last portion of the growing envelope input signal is 1.4 V, where the maximum rms value
during the estimation step was limited to 1 V. This shows that the obtained model is even capable of extrapolating, although it is not advisable to rely on this property.

The parallel Wiener-Hammerstein model outperforms the BLA again. The results obtained for the different models are shown in Table 9.8 and in Figure 9.10. The BLA is obtained for an input rms value of 0.775 V. It is also clear from the obtained results that the model still performs well in the last quarter of the growing envelope input (after 0.15 seconds). This is the region where the model extrapolates. This proves the robustness of the obtained parallel Wiener-Hammerstein model with a neural network nonlinearity for this specific example. The rms errors of the BLA are about 10 to 20 times larger (20 to 26 dB) than the errors of the parallel Wiener-Hammerstein model. The Wiener-Hammerstein model, the NARX model and the NOE model are again able to obtain model errors that are lower than the model error of the BLA, but the errors are still about 5 to 10 times larger (20 dB) than the errors of the parallel Wiener-Hammerstein model.
9.4. Parallel Wiener-Hammerstein

Table 9.8: Validation error on a growing envelope signal

<table>
<thead>
<tr>
<th>Validation error (mV)</th>
<th>total</th>
<th>quarter 1</th>
<th>quarter 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rms(u)</td>
<td>822.30</td>
<td>179.07</td>
<td>522.89</td>
</tr>
<tr>
<td>rms(e)</td>
<td>σ_e</td>
<td>μ_e</td>
<td>σ_e</td>
</tr>
<tr>
<td>Par WH</td>
<td>2.66</td>
<td>2.64</td>
<td>0.36</td>
</tr>
<tr>
<td>WH</td>
<td>20.20</td>
<td>20.20</td>
<td>0.03</td>
</tr>
<tr>
<td>NARX</td>
<td>18.77</td>
<td>18.53</td>
<td>3.01</td>
</tr>
<tr>
<td>NOE</td>
<td>22.93</td>
<td>22.32</td>
<td>0.82</td>
</tr>
<tr>
<td>BLA</td>
<td>55.74</td>
<td>46.70</td>
<td>30.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Validation error (mV)</th>
<th>quarter 3</th>
<th>quarter 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>rms(u)</td>
<td>889.42</td>
<td>1268.1</td>
</tr>
<tr>
<td>rms(e)</td>
<td>σ_e</td>
<td>μ_e</td>
</tr>
<tr>
<td>Par WH</td>
<td>1.86</td>
<td>1.84</td>
</tr>
<tr>
<td>WH</td>
<td>17.32</td>
<td>17.29</td>
</tr>
<tr>
<td>NARX</td>
<td>17.76</td>
<td>17.54</td>
</tr>
<tr>
<td>NOE</td>
<td>17.50</td>
<td>15.45</td>
</tr>
<tr>
<td>BLA</td>
<td>55.78</td>
<td>33.12</td>
</tr>
</tbody>
</table>

9.4.5 Study of the Initialization Procedure

A good initial estimate is a key factor to start the further optimization of the parameters if a high quality model is to be obtained. In this section we run the proposed algorithm until it arrives at the model selection step that is described in Section 8.5.4.3. The models that correspond to the 100 best pole-zero allocations are optimized, and the models corresponding to 100 random pole-zero allocations are also optimized separately. All the pole-zero allocations that are considered have 6 poles and 6 zeros in the front LTI blocks and 6 poles and 6 zeros in the back LTI blocks to match with the system under test. The Levenberg-Marquardt optimization algorithm is stopped after 500 iterations, or sooner when convergence is reached.

It is clear from the results shown in Figure 9.11 that the probability to obtain a good final model is higher when the best initial estimates are
Figure 9.11: Boxplot of the rms error obtained after optimization using different initialization schemes. The boxplot on the left is obtained using a random pole-zero allocation over the LTI blocks of the model. The boxplot on the right uses the 100 pole-zero allocation resulting in the best candidate models.

....

9.4.6 Conclusion

Similarly to the Wiener-Hammerstein benchmark in Section 9.4, the result of this section shows that the proposed parallel Wiener-Hammerstein identification algorithm of Section 8.5 performs well on measured data. The system almost perfectly belongs to the reachable model set: the system and the model have the same basic structure, but the system is a hardware implementation of this structure and therefore contains (small) nonidealities. This means that it is difficult to find a perfect model for the system in general and the diode-resistor nonlinearities that are present in the system in particular.
The parallel Wiener-Hammerstein model outperforms the other model structures that are estimated on the same dataset. This can be explained by the good correspondence between the parallel Wiener-Hammerstein model structure and the system. This shows that it is important to try different model structures to obtain a good model for a nonlinear system. Very general model structures, such as the NARX model class, are not always the optimal choice.

The importance of a good initialization of the model parameters before nonlinear optimization is illustrated in Section 9.4.5.

9.5 Modeling a High Frequency Doherty Amplifier

The nonlinear system considered in this section is a high frequency Doherty amplifier. This type of amplifier is currently attracting a lot of interest because of its favorable compromise between power efficiency and linearity. Nevertheless, the presence of nonlinear distortion remains a severe problem in the practical applicability of the device.

A good nonlinear model and a nonlinear inverse for the model are required to compensate for the distortion induced. This modeling problem leads to some specific modeling challenges.

The system is introduced in Section 9.5.1. The measurement setup used to gather the experimental data is explained in Section 9.5.2. The specific choices and details of the model estimation are discussed next in Section 9.5.3. Finally, the estimated model is validated in Section 9.5.4 and some conclusions are drawn in Section 9.5.5.

9.5.1 The Doherty System

The nonlinear system that is studied in this example is a high frequency Doherty power amplifier. This type of amplifier was first presented in Doherty [1936] as a linear power amplifier. Nowadays, the call for high efficiency is pulling RF (radio frequency) power amplifiers, including Doherty power
amplifiers, towards a nonlinear operating mode to obtain higher power efficiencies (Ghannouchi and Hammi [2009]). This comes at the cost of an increased nonlinear distortion.

The structure of a Doherty power amplifier is shown in Figure 9.12. A Doherty power amplifier consists of two amplifiers, implemented as two transistors. A first transistor implements the 'large' amplifier that amplifies the bulk of the signal (carrier amplifier). The second transistor, the 'small' amplifier, amplifies the peaks present in the input signal (peaking amplifier). At the first glance this seems to result in a parallel system structure. However, the true system behavior is more complex than this. The behavior of the carrier amplifier output changes when the peaking amplifier starts working. This introduces a feedback behavior in the system and complicates the overall behavior of a Doherty amplifier significantly.

The Doherty power amplifier that is studied in this work is described in detail in Özen and Fager [2014].

### 9.5.2 Measurement Setup

The measurement setup for the Doherty power amplifier measurements is quite complex, a simplified description of the setup is shown in 9.13 and discussed here. First the excitation signal is generated at low frequencies with a sampling frequency of 900 MHz. In the following stage this signal is upconverted to higher frequencies (around 3.45 GHz) and band-limited by a reconstruction filter. Next, the signal is amplified such that the de-
sired input power for the Doherty amplifier is obtained. The input and the output signals are attenuated before the measurement is performed. The measurement of the amplifier uses a sampling frequency of 10GHz and covers a bandwidth of 2 GHz.

The measured bandpass signals are transformed back to their low-pass equivalents. This results in complex time-domain signals. The standard procedure for the modeling of high frequency power amplifiers is to take the low-pass input and output signals to model an amplifier (Ghannouchi and Hammi [2009]). This allows to obtain models that are much cheaper to evaluate and much more simple to implement.

The downconverted excitation signals at the input of the Doherty amplifier are complex time-domain periodic multitone signals. The excited frequencies range from -150MHz to 150MHz. Each period contains 17778 samples, with a sampling frequency of 2GHz. 100 successive periods per realization are measured. There are 20 realizations for each of the 5 different input rms levels. The highest power level is four times larger than the lowest one. The power levels of the excitation are only known up a constant gain due to the different pre-amplification and attenuation stages and the lack of a power calibration. This unknown scale factor is fixed for all the measurements. The power grows linearly over the five different power lev-
9. Applications

The first 19 realizations are used for the estimation, the last realization is kept apart and will be used during the validation step.

9.5.3 Model Estimation for the Doherty Amplifier

A 2-branch parallel Wiener-Hammerstein model is estimated using the branch-per-branch parallel Wiener-Hammerstein identification algorithm presented in Section 8.6. This identification algorithm is modified to allow the complex valued time domain input-output data.

The BLAs are parametrized using an FIR model with 10 taps and complex coefficients. FIR models are used because discrete time rational transfer function models suffered from unstable pole estimates. Unstable poles jeopardize the possibility of a time domain based simulation.

The static nonlinearity is modeled using a polynomial of degree 7 with complex coefficients defined in the low-pass domain (Landin [2012]). Such a polynomial does not generate the higher harmonics that are generated by the standard polynomial basis functions. These harmonics fall outside the measured bandwidth and this type of behavior is not captured in the low-pass data. Furthermore, only odd functions are considered. Again, the frequency content generated by even functions is not captured in the low-pass data. Some examples of the polynomial basis function and their low-pass equivalent polynomial basis functions are given below:

\[
x^3 \rightarrow x|x|^2, \quad (9.9)
\]
\[
x_1x_2^2 \rightarrow x_1|x_2|^2 \text{ and } \bar{x}_1x_2x_2, \quad (9.10)
\]

where \( \bar{x} \) denotes the complex conjugate of \( x \).

Furthermore, all the delay that is present between the measured input and output data is removed during a postprocessing step. To take this into account properly, a delay is included in the model estimation.

To obtain a fair comparison with some other models, the parallel Wiener-Hammerstein model is optimized for one realization of the highest input rms level only. The other models are all estimated using this data record.
9.5.4 Model Validation

The model is validated using the 20th realization that has not been used during the estimation. Although the model is optimized for one input rms level only, it is validated on all of them. Note that the other input rms levels are all lower than the one that is used for the model optimization.

The relative error $e_{rel}$ is used to quantify the quality of the model:

$$e_{rel} = \frac{\text{rms}(y(t) - \hat{y}(t))}{\text{rms}(y(t))},$$

(9.11)

where $y(t)$ is the measured output and $\hat{y}(t)$ is the modeled output.

The parallel Wiener-Hammerstein model obtains good results as can be seen in Figure 9.14 and Table 9.9. The error of the parallel Wiener-Hammerstein model is well below the output level. The model describes about 90% of the nonlinear behavior of the power amplifier since the error is about 20dB below the total distortion and spectral regrowth levels (this is the spectral output content present at the non-excited frequencies in Figure 9.14).

The estimated parallel Wiener-Hammerstein model performs very well when it is compared to some of the standard and state of the art behavioral modeling techniques for high frequency power amplifiers. Table 9.9 shows the performance of different modeling techniques for the different rms levels of the input. The performance of the parallel Wiener-Hammerstein model is compared to the BLA (note that for each excitation level a different BLA is used), a low-pass equivalent polynomial model (degree 7), a memory polynomial model (MP) (of degree 7, memory depth 11), a generalized memory polynomial (GMP) (of degree 7, memory depth 11, mixing depth 2), and a Volterra model (degree 7, memory depth 3). A detailed description of these models can be found in Tehrani et al. [2010].

The GMP model is the most advanced model present in the comparison and obtains the best results in this comparison together with the parallel Wiener-Hammerstein model. The advantage of the GMP model is that it can be obtained through a simple linear least squares estimation, while the identification procedure for the parallel Wiener-Hammerstein model is more
Figure 9.14: Frequency domain validation of the parallel Wiener-Hammerstein model for the highest rms-value of the input. The noisy system output is depicted by the full blue line. The model error is represented by the green dots. The variance of the total distortion that is present at the output is shown by top black triangles (only on the excited frequency lines), the noise distortion is represented by the bottom black circles (only on the excited frequency lines).

<table>
<thead>
<tr>
<th>Input level</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLA</td>
<td>0.047</td>
<td>0.057</td>
<td>0.063</td>
<td>0.067</td>
<td>0.071</td>
</tr>
<tr>
<td>Polynomial</td>
<td>0.070</td>
<td>0.063</td>
<td>0.061</td>
<td>0.060</td>
<td>0.061</td>
</tr>
<tr>
<td>MP</td>
<td>0.051</td>
<td>0.035</td>
<td>0.023</td>
<td>0.020</td>
<td>0.021</td>
</tr>
<tr>
<td>GMP</td>
<td>0.054</td>
<td>0.036</td>
<td>0.021</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>Volterra</td>
<td>0.060</td>
<td>0.042</td>
<td>0.029</td>
<td>0.026</td>
<td>0.030</td>
</tr>
<tr>
<td>Par. WH</td>
<td>0.056</td>
<td>0.036</td>
<td>0.021</td>
<td>0.014</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Table 9.9: Validation error

Relative validation error: $e_{rel}$

involved. The disadvantage of the GMP model is its lack of structure (while this is one of the strong points of the parallel Wiener-Hammerstein model). This makes the model more costly to evaluate.
9.5.5 Conclusion

The parallel Wiener-Hammerstein modeling techniques proposed in Section 8.6 are applied on a challenging dataset obtained from a high frequency Doherty power amplifier. A Doherty power amplifier exhibits both a parallel and a feedback system behavior. The parallel Wiener-Hammerstein model performed very well despite the clear discrepancy between the model class and the system. It obtained results similar to other state of the art high frequency power amplifier modeling techniques.

9.6 Conclusion and Discussion

This chapter shows that the applied methods obtain good results on different applications. Both real life applications (Glucoregulatory system, Doherty amplifier), and hardware implementations of specific nonlinear model structures (Wiener-Hammerstein benchmark, parallel Wiener-Hammerstein) are considered. The obtained models are always of high quality. This illustrates the robustness of the methods used as the system does not always belong to the reachable model class. It also illustrates the practical usefulness of the developed approaches and considered model classes.
Chapter 10

Conclusion and Future Research Directions

This chapter highlights the main contributions of this thesis in Section 10.1. Secondly, some possible future research directions are discussed in Section 10.2.

10.1 Main Contributions and General Conclusion

This thesis targets the identification of block structured models consisting of the series and parallel connection of linear dynamic and nonlinear static building blocks. A favorable balance between accuracy, cost of model extraction and model complexity is pursued.

In 2010 the state of the art in block-oriented modeling was sampled in Giri and Bai [2010]. The chapters of this book deal mainly with the identification of Wiener and Hammerstein system models, only a minority of the contributions discusses Hammerstein-Wiener or Wiener-Hammerstein structures. This thesis pushes the state of the art beyond the content of Giri and Bai [2010] in the following directions:

- four novel Wiener-Hammerstein identification methods are explored
10. Conclusion and Future Research Directions

in Chapter 4,

• a novel Hammerstein-Wiener identification algorithm is studied in Chapter 5,

• identification algorithms for parallel Hammerstein (Chapter 6) and parallel Wiener systems (Chapter 7) are developed,

• the decoupling of MIMO polynomial nonlinearities for block-oriented models is discussed in Chapter 7,

• two novel identification methods for parallel Wiener-Hammerstein systems are presented in Chapter 8.

The developed methods are always compared and situated with respect to the state of the art and with previously developed identification algorithms for similar model structures. The strong points, the weak points and the assumptions of the proposed identification methods are highlighted and discussed. The good performance and the practical usefulness of the methods are illustrated using simulation based and measurement based examples. The measurement based examples show that it is a sound idea to use linearized models of a nonlinear system as a starting point for the estimation of nonlinear models. This approach proves to be robust to the presence of nonidealities in the system. It continues to work well even when the system does not belong to the reachable model class considered in the estimation.

The results in the previous chapters give an answer to the research questions that were posed in the first chapter:

Q1: How are single branch Wiener-Hammerstein systems identified? The four Wiener-Hammerstein identification methods use very different identification techniques to obtain an estimate of a Wiener-Hammerstein system. The first one uses a scan of all possible pole-zero allocations. The second one focuses on an iterative procedure based on a set of basis functions extracted from the BLA. The third method solves an optimization problem based on the nonparametric BLA and QBLA. Finally, the fourth one uses
specially designed multisine inputs to extract the front and back LTI blocks of the Wiener-Hammerstein system. They all offer an improvement to the current state of the art techniques even though the proposed approaches are sometimes based on similar techniques and ideas in the literature.

**Q2: Which Wiener-Hammerstein identification method is best suited for the identification of parallel Wiener-Hammerstein systems?** As we learned in Chapter 8, the zeros of the BLA of a parallel Wiener-Hammerstein system are to be distributed over the branches and also over the front and back dynamics of each branch. The algorithm, presented in Sjöberg and Schoukens [2012] and improved in Section 4.5, is selected due to its simplicity and due to its capacity to assign the zeros and poles of the BLA to the LTI subsystems present in the parallel Wiener-Hammerstein model. It is possible to use other methods as well, as is discussed in Section 10.2.

**Q3: How can we identify parallel Hammerstein and parallel Wiener structures?** The state of the art for the identification of parallel Hammerstein and parallel Wiener systems is discussed in Chapters 6 and 7. These chapters also introduce a new procedure for the identification of parallel branch models based on the decomposition of the BLA. Chapter 7 also introduces a parallel Wiener identification method based on dimension reduction. Both approaches prove to obtain good results.

**Q4: How can we decouple multivariate polynomials to increase the parsimony of the model structure?** The decoupling of MIMO polynomial nonlinearities is discussed in Chapter 7. It turns out that the coefficients of the polynomial can be grouped in different tensors. A decoupled description of the polynomial is obtained by diagonalizing these tensors using for example the parafac decomposition as has been done here.

**Q5: How can the previously developed approaches be combined into a parallel Wiener-Hammerstein identification algorithm?** Chapter 8 introduces two novel identification algorithms for parallel Wiener-Hammerstein systems based on the results obtained in Chapters 4, 6 and 7. The proposed approach starts with an estimation and parametrization of the BLA of the considered system for different operating conditions. Starting from the parametrized BLAs, a decomposition of the overall dynamics at the
different operating conditions is calculated using the singular value decomposition. Finally, the dynamics of each branch are partitioned over the different blocks of the parallel Wiener-Hammerstein model, and the static nonlinearity that is present in the model is estimated.

In some cases the work that is presented in the previous chapters triggered active research collaborations within the department ELEC. This gave rise to research output that is not presented in this thesis:

- the work on the decoupling of MIMO polynomial nonlinearities for block-oriented models discussed in Chapter 7 was the start of a sequence of papers that discuss this topic in depth (Schoukens and Rolain [2012a]; Tiels and Schoukens [2013]; Schoukens et al. [2014c]; Hollander et al. [2015]; Dreesen et al. [2015]),

- the observations made in Schoukens et al. [2015c] (also presented in Chapter 8) about the common denominator nature of the BLA of a parallel Wiener-Hammerstein system are a starting point for the work that is presented in Schoukens et al. [2015a] about model structure detection for block-oriented models.

To conclude, this thesis developed new identification methods for a variety of block-oriented structures. The good performance and the practical usefulness of the methods are illustrated on simulation based and measurement based examples. The parallel Wiener-Hammerstein block-oriented structure is the most advanced structure that is considered in this thesis. The identification methods that are developed for this structure provide a significant step forward in the field of block-oriented modeling. However, the developed methods are far from perfect. Some of the methods, such as the parallel Wiener-Hammerstein identification methods or the nonparametric Wiener-Hammerstein identification methods, are data-hungry. Also, none of the model structures that are considered in this thesis contain feedback. Finally, it is not straightforward for the user to decide how to include more prior knowledge in the identification procedure then the model structure. These drawbacks, together with other observations, form the basis for some possible future research directions.
10.2 Future Research Directions

This section discusses some possible future research directions related on or based on the work that is presented in this thesis. Some of the research ideas are more general, while others are rather specific. The four general research directions I selected more specifically because they look the most interesting and promising to me are:

1. Improving the block-oriented modeling tools: structure detection, estimation and feedback.
2. Towards a general framework for gray-box identification.
4. Advancing the nonlinear modeling of high frequency power amplifiers.

These four research directions are elaborated on in the following sections.

10.2.1 Improving the Block-Oriented Modeling Tools: Structure Detection and Estimation

A first research direction is a direct continuation of the research that is presented in this thesis. The goal is to improve the developed methods such that less user interaction and less data are needed to obtain a good model estimate. This research track can be split in four parts: structure detection, reducing the amount of data needed for the estimation, including feedback into the parallel block-oriented modeling methods and reducing the user interaction during the estimation.

10.2.1.1 Structure Detection

A good model structure detection algorithm is needed to guide the user towards a good model structure. Such an algorithm should be able to make objective decisions about the model structure, preferably without the need for special experiments and certainly without requiring expert knowledge.
from the user. The first steps to obtain such a tool are taken in Lauwers et al. [2006]; Schoukens et al. [2015a], where only well-known linear modeling tools are used to make decisions about the structure of the nonlinear system. Again, it is striking that the linear tools are used as a step up to the solution of the nonlinear problem setting. This is pretty much the same as for the use of the BLA as a key component to block-oriented model estimation.

10.2.1.2 Reducing the Amount of Data Needed for Identification

Several of the identification algorithms that are presented in the previous chapters require a large amount of data and experiments to obtain good results. The goal here is to develop some methods that reduce the amount of data needed for the model estimation step, for instance, in Section 7.6 using dimension reduction techniques. These techniques can also be used to improve other identification problems. For instance, the dimension reduction based Wiener-Hammerstein identification method in Section 4.6 can be extended to parallel Wiener-Hammerstein systems. This would remove the need for measuring the BLA at many different operating points in the algorithms presented in Sections 8.5 and 8.6.

Note, however, that it is also important that an identification method can handle large datasets. For many applications, building and testing the measurement setup is the most expensive part of a measurement campaign. It is not necessarily true that the actual data collection is the most costly step. This means that it is not a problem, and often even the standard, to obtain large datasets for identification in many applications (e.g. measurements on electronic circuits, or measurements on vibrating systems). Many methods have the rather annoying property to scale rather poorly for an increased size of the dataset. While it would be useful to use all data available to extract the model.
10.2.1.3 Including Feedback in Parallel Block-Oriented Models

There are already some identification methods for feedback block-oriented structures. However they are restricted to single branch models (Schoukens et al. [2008a]; Paduart [2008]), or, in case of the most advanced method, to a single branch model in parallel with a linear subsystem (Vandersteen and Schoukens [1999]; Vanbeylen [2013]).

The extension of parallel block-oriented modeling techniques towards parallel block-oriented model structures that also allow for feedback in the parallel branches is quite challenging. It is not straightforward to obtain good initial estimates for such a structure, neither is it easy to ensure a stable estimate or to keep the complexity of such a structure under control.

10.2.1.4 Reducing the User Interaction: Nonparametric Methods and Regularization

The identification of a nonlinear model often requires a highly educated user in the field of identification. If dissemination to a public of specialists of the system rather than of identification is targeted, nonparametric methods can be used to reduce the number of user decisions and lower the usage threshold. The problem with many nonparametric methods is that they need much more data than parametric methods to obtain a model of similar quality.

Regularization is used nowadays to obtain high quality nonparametric impulse response and frequency response estimates of linear systems with little data, little user interaction and very reasonable assumptions on the system Chen et al. [2012]. A first step for the extension of these methods towards nonlinear systems, more specifically Volterra kernels, is made in Birpoutsoukis and Schoukens [2015]. A future research track could include the use of these methods for the nonparametric or semiparametric identification of block-oriented models.
10. Conclusion and Future Research Directions

10.2.2 Towards a General Framework for Gray-Box Identification

This thesis introduced different identification techniques. Every identification method was specifically developed with a certain block-oriented structure in mind. This research direction wants to go beyond this.

An experienced practitioner has often a lot of prior knowledge about the system that needs to be modeled. However, most nonlinear identification techniques cannot take into account this prior knowledge unless a specific identification method is developed for that specific application. This research track wants to create a general framework to include prior knowledge about the application in the nonlinear system identification method. This leads to a gray-box identification problem: the structure, or a large part of the structure, is fixed, but the parameters need to be estimated. I believe that the practitioners will have much less reluctance to accept a nonlinear model for their application if they see that it corresponds better to their prior knowledge.

Typical prior knowledge that can easily be obtained from a practitioner is the position and the type of (static) nonlinearities that are present in the nonlinear system. A circuit designer can point at some of the transistors or at some nonlinear capacitors that are present in the design as a source of the nonlinear distortion. A designer of airplanes can point at the bolted connections or some vibration dampers with hard stops as a source of non-linearity. How can we include this information in a general framework? Can we develop a method that generates initial estimates of the model parameters that can be used for further nonlinear optimization? Or can we start from a general black-box approach and end up with a model that fits the prior knowledge that was given to us by the practitioners?

A first possible approach tries to include the prior knowledge at the start of the model estimation step. The main difficulty here is to develop a method that can generate initial estimates of the model parameters for a very wide range of models in a very flexible model set. This requires a lot of insight about how different interconnections of nonlinearities and linear
dynamics behave on the one hand, but also on how this behavior can be exploited for a system identification purpose on the other hand.

A second approach to tackle this problem starts from a very general black-box model. The prior knowledge can be introduced gradually into this model. This can be for instance be obtained using linear and nonlinear transformations of a polynomial nonlinear state space model. The decomposition methods for multivariate polynomials (see Section 7.7) can be a useful tool to achieve this goal.

10.2.3 Nonlinear Parameter Varying Identification

The work of this thesis is focused on obtaining good nonlinear models for nonlinear systems. Where the model class needs to include a broad range of nonlinear systems, the model structure should be parsimonious and the estimated model should also be easy to analyze and to interpret. The final model that is discussed in this thesis, the parallel Wiener-Hammerstein model structure, makes a good trade-off between these requirements.

By allowing the model to be a parameter varying model a wider class of systems can be modeled, and a different trade-off can be made. A combination of block-oriented models and parameter variability largely preserves the parsimony of the model structure. The parameter varying structure can be used to model external influences on the system (e.g. temperature influences), but it can also be used to model systems that have a moving operating point. In some cases it could for instance be sufficient to model a parallel Wiener-Hammerstein system using a parameter varying Wiener-Hammerstein model. Nonlinear parameter varying models are already studied in literature, for instance in Previdi and Lovera [2004]. This research direction could extend the state of the art for nonlinear parameter varying identification, and investigate possible applications for this type of models.
10.2.4 Advancing the Nonlinear Modeling and Predistortion of High Frequency Power Amplifiers

Models that are used for predistortion of high frequency power amplifiers are typically single-input-single-output models, considering only the forward signal path. To reduce the computational load to simulate the output of the model, the envelope of the input-output signals is modeled. Many different RF power amplifier model structures are used in the literature. The most common ones are static nonlinear models, (generalized) memory polynomial models, Volterra series based models, block-oriented models and several combinations of the aforementioned models (Ghannouchi and Hammi [2009]; Tehrani et al. [2010]).

We propose to use block-oriented models, because of their user friendly structures. The nonlinear and dynamic behavior of the system are both modeled using different and well separated blocks that can be combined in various ways to adapt to the complexity of the system. One resulting advantage is the parsimony of the model when compared with Volterra series based models. The block-oriented models that are used to model and to predistort power amplifiers are mainly of the Wiener, Hammerstein, and parallel Hammerstein type. This is mainly due to the very simple identification algorithms that are available for them.

The specificities of the proposed MIMO nonlinear block-oriented model are:

1. The model allows for parallel signal paths (parallel branches).

2. The model allows for nonlinear feedback.

3. The model can be estimated from data that are obtained from a system operating in feedback.

Parallel signal paths are often observed physically in nonlinear systems. For RF power amplifiers specifically, parallel signal paths result from the presence of parasitic signal paths: the main path transfers the signal
through the main transistor to the output, while a reflection of a part of
the signal on the power supply is also transferred to the output in a second
path. The presence of multiple paths can even be imposed by the design;
as is for instance the case for Doherty power amplifiers (Özen and Fager
[2014]).

Nonlinear feedback is inherent to transistor models due to the presence
of parasitic capacitances in the transistor itself and its periphery. The model
can be estimated from data that are obtained from a system in feedback.
RF devices operate typically in feedback due to the loading of the device
at its output or input, or due to the presence of reflections and backward
signal paths in the device itself, or even often by design.

Estimating a MIMO model of an RF system has several benefits. The
main benefits are:

1. A deeper understanding of the system behavior.

2. The models can be cascaded while multitone excitation signals are
   used.

3. The possibility for multi-input linearization.

Typical telecommunication signals are very similar to multitone input sig-
nals. More realistic models are obtained by performing the model identi-
fication using these multitone input signals. This leads to a better under-
standing and approximation of the system behavior in real operation.

Due to the use of multitone input signals and the estimation of a MIMO
nonlinear model, the model can be scaled from a single component to the
complete system simply by cascading the nonlinear models. The possibility
to cascade the nonlinear models can be used to obtain better designs for the
full communication channel and to simulate the full communication chan-
nel from amplifier to transceiver. Previous methods (such as x-parameters
(Vye [2010])) are more restrictive on the type of excitation signals. Hence,
they result in a model that is further away from real-world operation, or
is strongly depending on the initial values of the parameter optimization
(such as for neural network based models).
MIMO models enable multi-input linearization if the power supply port is also considered as an input port. A more power efficient and linear power amplifier behavior will be obtained by combining envelope tracking techniques (through the power supply port) with digital predistortion (through the signal port). This can only be realized using the knowledge obtained by a MIMO power amplifier model.
List of Publications

Articles in Journals


List of Publications


Papers under review:


Chapters in Books


Articles in Proceedings of Scientific Conferences


Other

International Workshops


**Local Conferences**


List of Publications


Local Workshops


Seminars

List of Publications


Bibliography


280


Bibliography


Bibliography


Bibliography


R. Pintelon, P. Guillaume, G. Vandersteen, and Y. Rolain. Analyses, Development and Applications of TLS Algorithms in Frequency-Domain Sys-


M. Schoukens, C. Lyzell, and M. Enqvist. Combining the best linear approximation and dimension reduction to identify the linear blocks of parallel

292


J. Sjöberg, L. Lauwers, and J. Schoukens. Identification of Wiener-Hammerstein models: Two algorithms based on the best split of a linear


K. Tiels and J. Schoukens. From coupled to decoupled polynomial representations in parallel Wiener-Hammerstein models. In *Decision and Control*


