# Contents

**Preface** vii  
**Beknopte samenvatting** ix  
**Glossary** xi  
Symbols and Notations ........................................ xi  
Abbreviations .................................................. xii  

## 1 Introduction 1  
1.1 General remarks .......................................... 1  
1.1.1 How nonlinear can it get? ............................... 2  
1.1.2 Approximating nonlinear systems ..................... 6  
1.1.3 Different systems, different models ................... 8  
1.2 Outline of the thesis ...................................... 12  
1.3 Publications .............................................. 14  

## Appendix 16  
1.A Convergence of derivatives of the output signals .......... 16  
1.A.1 In mean square sense ................................... 16  
1.A.2 Uniform convergence ................................... 17  

## 2 Introduction to Block oriented Systems 19  
2.1 Block oriented systems .................................... 20  
2.1.1 Linear time invariant systems ......................... 20  
2.1.2 Static nonlinearities .................................. 20  
2.1.3 Commonly used structures ............................. 21  
2.2 The excitation signal: the multisine ....................... 23  
2.2.1 Nonlinear systems and periodical signals ............. 24  
2.3 The Related Linear Dynamic System ....................... 25  
2.3.1 The RLDS of a Wiener-Hammerstein system .......... 26  
2.4 Noise .................................................... 30  

## Appendix 31  
2.A The Fourier Transform .................................... 31  
2.A.1 Why is the frequency domain so important? .......... 31  

## 3 Hammerstein systems and Wiener systems 33  
3.1 Introduction .............................................. 34  
3.2 Simple identification of Hammerstein systems .......... 34  
3.2.1 Informal presentation of the method ................. 34  
3.2.2 About the estimate of the linear part ............... 35  
3.2.3 About the estimate of the nonlinear part .......... 36
### Contents

3.3 Iterative identification of Hammerstein systems .......................... 39
   3.3.1 The algorithm ........................................... 40
   3.3.2 Demonstration on a measurement example .................... 43
3.4 Estimating Hammerstein systems using multiple experiments .......... 46
   3.4.1 Definitions ................................................ 46
   3.4.2 Demonstration on measurements .......................... 48
3.5 Parameterization issues .............................................. 48
   3.5.1 Choosing a parameterization for the static nonlinearity .... 50
   3.5.2 Choosing a parameterization for the linear part ........... 53
3.6 Simple identification of Wiener systems ................................ 53
   3.6.1 About the estimate of the linear part ...................... 54
   3.6.2 About the estimate of the nonlinear part .................. 54
3.7 Iterative identification of Wiener systems ............................ 55
   3.7.1 The algorithm ............................................. 56
   3.7.2 Demonstration on a measurement example .................... 58
3.8 Estimating Wiener systems using multiple experiments .............. 60
   3.8.1 Modification in the method ............................... 60
   3.8.2 Demonstration on measurements .......................... 63
3.9 On the importance of the frequency domain representation ........... 65
3.10 Comparison with other methods .................................... 67
3.11 Conclusion ...................................................... 68
Appendix .............................................................. 69
   3.A Introduction to the convergence analysis ...................... 69
   3.B Convergence analysis for the Hammerstein estimation algorithm .... 69
      3.B.1 Introduction ............................................. 69
      3.B.2 Convergence in the nonparametric case .................. 72
      3.B.3 Convergence in the parametric case .................... 73
      3.B.4 Convergence when multiple experiments are used ......... 78
   3.C Convergence analysis for the Wiener estimation algorithm ....... 81
      3.C.1 Introduction ............................................. 81
      3.C.2 Convergence for the parametric case .................... 83
      3.C.3 Convergence when multiple experiments are used ......... 86
   3.D Product of columns of $B(U)$ or $B(Y)$ .......................... 88
      3.D.1 Product of two identical columns ....................... 88
      3.D.2 Product of two different columns ....................... 88
   3.E Consistency of the Hammerstein estimation algorithm .......... .... 89
      3.E.1 Notations and Assumptions ................................ 89
      3.E.2 Convergence of (3.5) to the noiseless solution ........... 92
      3.E.3 Convergence of (3.9) to the noiseless solution .......... 93
      3.E.4 Convergence of (3.7) to the noiseless solution .......... 93
   3.F Consistency of the Wiener estimation algorithm ................. .... 94
      3.F.1 Notations and Assumptions ................................ 94
      3.F.2 Convergence of Step 4.1 to the noiseless solution .......... 96
      3.F.3 Convergence of Step 4.3 to the noiseless solution .......... 96
      3.F.4 Discussion ................................................ 96

4 Hammerstein-Wiener systems ............................................ 99
   4.1 Introduction to the identification method ........................ 100
4.2 What is so particular about Hammerstein-Wiener systems? 101
4.3 The Estimation Procedure 104
  4.3.1 Parameterization of the Static Nonlinearities 105
  4.3.2 The Algorithm 106
4.4 The Algorithm applied to Measurements 109
  4.4.1 The Measurement Setup 110
  4.4.2 Results 110
4.5 Comparison with other methods 113
4.6 Degenerated Hammerstein-Wiener systems 115
4.7 Conclusion 117

Appendix 118
  4.A Analysis of the iterative procedure 118
    4.A.1 Notations and assumptions 118
    4.A.2 The Analysis 119

5 Wiener-Hammerstein systems 125
  5.1 Introduction 126
    5.1.1 Parameterization 127
    5.1.2 Degenerations in the Wiener-Hammerstein structure 128
  5.2 The identification algorithm 129
    5.2.1 Estimating $f$ 129
    5.2.2 Estimating $S$ 130
    5.2.3 Estimating $R$ 131
    5.2.4 Definition of the estimation algorithm 131
  5.3 Implementation details 133
    5.3.1 Initializing $\hat{R}^{[l]}$ 133
    5.3.2 Requirements for the input signal 134
    5.3.3 Choosing the sets of frequencies 135
    5.3.4 Damping the transition from $\hat{S}^{[k-1]}$ to $\hat{S}^{[k]}$ 136
  5.4 Measurement example 137
    5.4.1 Setup and parameterization 137
    5.4.2 Results of the initialization algorithm 138
  5.5 Comparison with existing methods 141
  5.6 Optimizing the global cost function 142
    5.6.1 Output Error Framework 142
    5.6.2 Errors in variables Framework 145
  5.7 Conclusion 145
Appendix 146
  5.A The Jacobian 146

6 Fast Approximate Identification of Nonlinear Systems 149
  6.1 Introduction 150
  6.2 Model Structure 150
  6.3 Identification Procedure 153
    6.3.1 Identification of the best linear approximation 153
    6.3.2 Identifying the nonlinear model parameters 156
    6.3.3 Model selection 157
  6.4 Impact of the Disturbing Noise 157
  6.5 Experimental Verification 157
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.5.1</td>
<td>Nonlinear feedback system</td>
<td>158</td>
</tr>
<tr>
<td>6.5.2</td>
<td>System with bidirectional dynamics</td>
<td>159</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Hammerstein system</td>
<td>161</td>
</tr>
<tr>
<td>6.6</td>
<td>Conclusions</td>
<td>161</td>
</tr>
<tr>
<td>Appendix</td>
<td></td>
<td>164</td>
</tr>
<tr>
<td>6.A</td>
<td>Motivation of the feedback branch</td>
<td>164</td>
</tr>
<tr>
<td>6.B</td>
<td>Identification of the parallel structure</td>
<td>165</td>
</tr>
<tr>
<td>6.C</td>
<td>Computing the Volterra Kernels of a Nonlinear Feedback Structure</td>
<td>166</td>
</tr>
<tr>
<td>7</td>
<td>Measuring Volterra Kernels of Microwave Devices</td>
<td>169</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>170</td>
</tr>
<tr>
<td>7.2</td>
<td>Problem statement</td>
<td>171</td>
</tr>
<tr>
<td>7.3</td>
<td>Illustration of the problem</td>
<td>172</td>
</tr>
<tr>
<td>7.4</td>
<td>Basic Idea</td>
<td>173</td>
</tr>
<tr>
<td>7.5</td>
<td>Experimental verification</td>
<td>174</td>
</tr>
<tr>
<td>7.5.1</td>
<td>Measurement</td>
<td>174</td>
</tr>
<tr>
<td>7.5.2</td>
<td>Extracted Models</td>
<td>176</td>
</tr>
<tr>
<td>7.5.3</td>
<td>Comparison of the extracted models for the single tone and dual tone measurements</td>
<td>177</td>
</tr>
<tr>
<td>7.5.4</td>
<td>Comparison of the estimated kernel values for the MAR 6</td>
<td>179</td>
</tr>
<tr>
<td>7.6</td>
<td>Intermediate conclusion</td>
<td>182</td>
</tr>
<tr>
<td>7.7</td>
<td>Consequences</td>
<td>182</td>
</tr>
<tr>
<td>7.8</td>
<td>Conclusions</td>
<td>183</td>
</tr>
<tr>
<td>Appendix</td>
<td></td>
<td>185</td>
</tr>
<tr>
<td>7.A</td>
<td>Current and Voltage at the input of the DUT of Figure 7.2</td>
<td>185</td>
</tr>
<tr>
<td>8</td>
<td>Wave Phase Calibration</td>
<td>187</td>
</tr>
<tr>
<td>8.1</td>
<td>Introduction</td>
<td>188</td>
</tr>
<tr>
<td>8.2</td>
<td>The Network Analyzer for Nonlinear Systems</td>
<td>188</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Downconversion</td>
<td>189</td>
</tr>
<tr>
<td>8.2.2</td>
<td>Frequency Response Calibration</td>
<td>191</td>
</tr>
<tr>
<td>8.2.3</td>
<td>Absolute Amplitude Calibration</td>
<td>192</td>
</tr>
<tr>
<td>8.2.4</td>
<td>Wave Phase Calibration</td>
<td>193</td>
</tr>
<tr>
<td>8.3</td>
<td>The Measurements and Identification Results</td>
<td>195</td>
</tr>
<tr>
<td>8.4</td>
<td>The Measurement Problem</td>
<td>200</td>
</tr>
<tr>
<td>8.5</td>
<td>A Proposal for a Wave Phase Calibration</td>
<td>203</td>
</tr>
<tr>
<td>8.6</td>
<td>Simulations</td>
<td>204</td>
</tr>
<tr>
<td>8.6.1</td>
<td>Setup</td>
<td>204</td>
</tr>
<tr>
<td>8.6.2</td>
<td>Parameterization issues</td>
<td>207</td>
</tr>
<tr>
<td>8.6.3</td>
<td>The Calibration Procedure and Results</td>
<td>207</td>
</tr>
<tr>
<td>8.7</td>
<td>Conclusion</td>
<td>210</td>
</tr>
<tr>
<td>9</td>
<td>Conclusions and Ideas for Future Research</td>
<td>211</td>
</tr>
<tr>
<td>9.1</td>
<td>Identification</td>
<td>211</td>
</tr>
<tr>
<td>9.2</td>
<td>Measurements</td>
<td>212</td>
</tr>
<tr>
<td>9.3</td>
<td>Future Work</td>
<td>212</td>
</tr>
<tr>
<td>9.3.1</td>
<td>Extension to MIMO</td>
<td>212</td>
</tr>
<tr>
<td>9.3.2</td>
<td>Extension to closed loop nonlinear systems</td>
<td>213</td>
</tr>
<tr>
<td>9.3.3</td>
<td>Gain insight usable for design</td>
<td>213</td>
</tr>
</tbody>
</table>
Preface

Some personal notes

I was attracted to engineering because of the fascination that predicting reality from abstract equations held for me. To me, it is still some kind of miracle that useful statements and predictions about our world can be formulated using such an abstract thing as calculus built on some axioms. There was another component, too: I wanted to build these insights into something new. In this thesis, I’ve had the opportunity to learn a lot more about how those “magical” models can be built. Initially, I also hoped to apply some of that knowledge to practical designs to fulfill my taste for applied sciences. However, identification proved to be a much richer topic than anticipated and I ended up doing “only” identification.

Acknowledgments

Luckily, writing a PhD is not a solitary activity. While researching and writing the material presented here, the support of many people has been as welcome as it was necessary.

There is more to life than scientific research. As the remainder of this book deals exclusively with science, the balance will be restored by mentioning first those who supported me in my private life: my wife Tamara and my family.

This work would not have been the same if it had not been for the helpful remarks of patient proofreaders (in alphabetical order): Rik Pintelon, Yves Rolain and Johan Schoukens. The guidance and advice they offered is very much appreciated.

I have also had the opportunity to get to know a lot of nice colleagues in the department: going to the store and buying cookies for the department has always been a nice change from work!

Identification without measurements is not really complete and I have used Yves Rolain’s and Wendy Van Moer’s measurement system a lot. Thank you for implementing this software and helping me getting it to run. Also, in Chapter 7 I was allowed to use some of their measurement data, for which I would like to thank them explicitly.

Last but not least, the logistics: a friendly and competent technical staff, nice and efficient secretaries and the Funds for Scientific Research Flanders (FWO) deserve their place in my list of thanks.

And because a second time won’t hurt: thanks to all, especially my wife for bearing with me during the redaction of this manuscript.
Beknopte samenvatting

Iedereen heeft vast nog herinneringen aan de regel van drie: als één kilo bloem €0.50 kost, kost 2kg bloem €1.00. Dit noemt men lineair: de prijs en de hoeveelheid bloem worden op dezelfde manier geschaald. Nochtans kost een zak van 5kg bloem geen €2.50, maar bijvoorbeeld slechts €2.25. Dit is een niet lineair verschijnsel. In het dagelijkse leven zijn er veel voorbeelden van niet lineaire systemen die we als natuurlijk of rechtvaardig ervaren. Een ander niet lineair systeem dat iedereen kent, zijn de belastingen: het is niet zo dat wanneer het bruto loon verdubbelt, het netto loon ook met een factor twee stijgt. Dit is een eenvoudig niet lineair systeem: het bevat geen geheugen vermits de belastingen voor het jaar 2003 onafhankelijk zijn van de belastingen in 2002.

Niet lineaire systemen vindt men overal, onder meer in de elektronica branche komen zulke systemen veelvuldig voor. Redeneren over niet lineaire systemen is moeilijk omdat men niet langer de regel van drie mag gebruiken. Daarom worden gemakshalve veel niet lineaire systemen dikwijls door een lineair benaderd. Dit heeft als voordeel dat het gemakkelijker wordt om het gedrag ervan te analyseren. Maar een dergelijke vereenvoudigd design is dikwijls niet verantwoord om economische of ecologische redenen. Een voorbeeld daarvan is de zender in een mobiele telefoon. Om langer met eenzelfde batterij te kunnen bellen, moet de schakeling minder verbruiken. De meest energiezuinige schakeling is een sterk niet lineair systeem dat niet goed benaderd kan worden door een lineair model. Men moet ze niet lineair modelleren om tot een optimaal ontwerp te komen.

In deze thesis werd onderzocht hoe modellen voor niet lineaire systemen opgebouwd kunnen worden. Dergelijke modellen worden veel gebruikt, bijvoorbeeld in het ontwerpen van vliegtuigen, wagens of chips. Hun gebruik laat toe om dure experimenten te vervangen door een goedkope computersimulatie. Om de modellen te bouwen observeert men het gedrag van een werkelijk systeem, en vervolgens wordt een niet lineair model voorgesteld. Dat model wordt dan fijngeregeld zodat het de waarnemingen zo goed mogelijk verklaart.

Er zijn echter heel veel mogelijkheden te onderzoeken om het beste model te bepalen. Om het probleem tot een vertebrabe brok te herleiden, worden daarom vereenvoudigingen ingebracht. Men kijkt slechts naar een beperkte klasse van modellen die gemakkelijk te bepalen en te gebruiken zijn. Deze zijn opgebouwd uit lineaire delen met geheugen en niet lineaire delen zonder geheugen.

De thesis ontwikkelde niet alleen theoretische ideeën, een belangrijk gedeelte ervan is gewijd aan experimentele toepassingen: er werden metingen verricht in de GSM frequentie band. In deze omstandigheden treden er echter bijkomende moeilijkheden op. Onze meetinstrumenten zijn als een slechte vuile bril met gekleurde glazen. Ze vervormen en verkleuren de realiteit. Om goede metingen te bekomen moeten we eerst deze fouten compenseren en pas dan kunnen we ermee aan de slag om een model te bouwen.
Glossary

Symbols and Notations

All notations and symbols are defined in Chapter 2 and locally in the different chapters. The following tables show general rules that were followed for notations in this book. Signals in the time domain use lower case. Upper case letters are used for the corresponding Fourier spectra:

\[ u(t) \] input signal at the time \( t \)

\[ U(\Omega) \] Fourier spectrum of the input

\[ y(t) \] output signal at the time \( t \)

\[ Y(\Omega) \] Fourier spectrum of the output

\( v, w \) unmeasurable time domain signals between two blocks

\( V, W \) Fourier spectra of the unmeasurable signals between two blocks

Parameter vectors are denoted with greek letters:

\( \theta \) parameters of the input LTI system of a Wiener-Hammerstein system or parameters of the linear system of a Wiener, Hammerstein or Hammerstein-Wiener system

\( \varphi \) parameters of the static nonlinearity inside a Wiener, a Hammerstein or a Wiener-Hammerstein system or parameters of the input static nonlinearity of a Hammerstein-Wiener system

\( \zeta \) parameters of the output LTI system of a Wiener-Hammerstein system

\( \psi \) parameters of the output nonlinearity of a Hammerstein-Wiener system

The following adornments are associated with parameters:

\( \hat{\varphi} \) estimate of \( \varphi \)

\( \varphi^{(k)} \) estimate of \( \varphi \) at iteration \( k \)
Transfer functions of LTI systems are also written as upper case letters:

\( R \) transfer function of the input linear system of a Wiener or a Wiener-Hammerstein system

\( S \) transfer function of the linear system inside a Hammerstein or Hammerstein-Wiener system or of the output linear system of a Wiener-Hammerstein system

### Abbreviations

**AWG** Arbitrary Waveform Generator

**ADC** Analog to Digital Converter

**CW** Continuous Wave

**DAC** Digital to Analog Converter

**DFT** Discrete Fourier Transform

**DUT** Device Under Test

**EIV** Errors in Variables

**FFT** Fast Fourier Transform

**FRF** Frequency Response Function

**IF** Intermediate Frequency

**LTI** Linear Time Invariant

**MIMO** Multiple Input, Multiple Output

**NARMAX** Nonlinear Auto-Regressive Moving Average with eXogenous inputs

**RLDS** Related Linear Dynamic System. The RLDS is the best linear approximation of a nonlinear system.

**SISO** Single Input, Single Output

**SNR** Signal to Noise Ratio

**ZOH** Zero Order Hold
1 Introduction

1.1 General remarks

Building complex systems demands that the composing subsystems are well known. Modeling and identification offer a solution for this request: given a set of measurements, mathematical equations are constructed that (hopefully) allow to predict or simulate the performance of a device before building it. This is an essential cost saver, both in production materials that are not wasted and in time that is not spent to build a prototype that a simulation would have shown to fail. This time that is not lost is translated into a reduced time to market. Also, these models can give insight in the working of the device and help to understand which modifications are necessary to obtain a desired behavior.

If sufficient prior knowledge and physical insight is present, a so-called white-box model can be built. This means that by looking to the physics of the device, a series of equations describing the device is proposed. These equations contain unknown parameters that are estimated by fitting the model to the measurements. However, as the size of the system grows, a second force comes into play: the complexity that can be handled is limited. It would be impractical to describe a microprocessor at the transistor level, to give an example. At the other extreme are black-box models: no prior knowledge is used and a model structure is proposed. Then the element of the set of models that best explains the measurements is retained as the model. The set of models is usually a family of systems described by unknown parameters, so that choosing the aforementioned element is done by minimizing some cost function with respect to the parameters. Black-box models describe the behavior of a system.

Linear models have been used for a long time and their identification is a mature
science. However linear modeling is not always sufficient: each model is only an approximation of reality and many systems are in fact nonlinear. For many inputs, linear approximations are valid (and very useful, too), but as the input range is increased, the nonlinear effects become so important that they can’t be ignored anymore.

One domain where the transition from linear models to nonlinear models is necessary is the domain of high frequency electronics. In order to extend the battery life of mobile devices, the electronic components are used in their nonlinear region because their efficiency is increased that way. This makes nonlinear models at high frequencies necessary.

1.1.1 How nonlinear can it get?

Nonlinear behavior comes in all shapes and colors. This section will review some of them, without worrying at the moment how they could be modeled or identified. It will be shown how the properties of the models are balanced against the underlying constraints and assumptions made on the Device Under Test (DUT). Mostly, nonlinear models are an approximation to reality. It turns out that the selected convergence criterion strongly influences the class of systems that can be approximated.

1.1.1.1 Open loop nonlinear systems

If the output of a system is a nonlinear function of its input, but not of its output, the nonlinear system is called an open loop nonlinear system. A discrete time open loop nonlinear system could be written as

\[ y(t) = F(u(t), u(t - 1), \ldots) \]

Arguably the most observed nonlinearity in electronics is saturation. Saturation occurs when the input power increases without corresponding increase in output power: this is sketched in Figure 1.1 where the output power of a saturating nonlinear system is shown in function of the input power. Such saturation is observed if \( F \) is a saturating function like \( \arctan x \) or \( \frac{x}{1+|x|} \). Obviously, the nonlinear function \( F \) is not limited to continuous saturating functions but could be anything: non invertible, discontinuous…
1.1 General remarks

1.1.1.2 Closed loop nonlinear systems

In contrast to the open loop nonlinear systems, closed loop nonlinear systems include the output in the nonlinear input output relation:

\[ y(t) = F(u(t), u(t-1), \ldots, y(t-1), y(t-2), \ldots) \]

These systems can exhibit a highly nonlinear behavior like subharmonics, bifurcations or chaos. In the literature over chaotic systems (Lakshmanan and Murali, 1996), the nonlinear system is studied while varying some parameter \( \mu \). In this section, the following nonlinear system is used:

\[ y(k+1) = -\frac{1}{5}y^3(k-1) + \frac{11}{20}y(k-1) - \frac{1}{5}y^3(k-3) + \frac{11}{20}y(k-3) + u(k-3) - \frac{2\mu}{5} \]  

This system is said to follow a route to chaos as \( \mu \) changes. The system (1.1) follows the so-called period-doubling route: as \( \mu \) is increased, the period of the output for a periodic input will increase (double). It turns out that for small \( \mu = 1 \), the period of the output is the same as the period of the input (both signals are shown in Figure 1.2). In the left plot of Figure 1.2, only a small part of the time signal is shown for clarity. The actual simulation used data records of 2048 points, from which the 1024 first points are thrown away to eliminate the transients. The spectrum in Figure 1.2 is computed using the points 1025 to 2048 of the simulation.

Figure 1.1: A saturating amplifier: the dots are the true values, the grey line is \( P_{\text{out}} = \alpha P_{\text{in}} \) (the input output relation if the system had been linear).
Subharmonics appear when $\mu$ is further increased: in Figure 1.3 the period of the output is longer than the period of the input (for $\mu = 2$). This period doubling is proven by the presence of subharmonics at $\frac{f_e}{2}$ in the right plot of Figure 1.3. This is the first bifurcation from the stable solution in Figure 1.2. Note that harmonics of $f_e = \frac{f_S}{4}$ fall at $2f_e = \frac{f_S}{2}$, $3f_e = \frac{3f_S}{4}$, $4f_e = f_S$ and so on. Because of aliasing, these harmonics show up at $\frac{f_S}{2} = 2f_e$, $\frac{f_S}{4} = f_e$ and 0. Hence, the energy at e.g. $\frac{f_S}{8} = \frac{f_e}{2}$ is truly a subharmonic and not some higher harmonic that has been aliased. Figure 1.4 shows that if $\mu$ is increased even more ($\mu = 2.5$), the period doubles again.

Chaos sets in when $\mu$ is increased even more. In Lakshmanan and Murali (1996), chaos is defined as a complicated and aperiodic motion which is highly sensitively dependent upon initial conditions in deterministic nonlinear systems. Figure 1.5 contains the simulation results of (1.1) for $\mu = 4.5$. 
1.1 General remarks

Figure 1.4: Left plot: Output (black) of (1.1) ($\mu = 2.5$) for a periodic input (grey). Right plot: spectrum of the output, $f_e$ is the frequency of the input.

Figure 1.5: Left plot: Output (black) of (1.1) ($\mu = 4.5$) for a periodic input (grey). Right plot: spectrum of the output, $f_e$ is the frequency of the input.
1 Introduction

1.1.2 Approximating nonlinear systems

The previous section has shown that the field of nonlinear systems is huge. Therefore, it is very difficult (if not impossible) to propose a model structure for all nonlinear systems at once: either it is not flexible enough (e.g. the model can’t produce subharmonics, even though the system does) or overly flexible and thus prone to overfitting (e.g. using a model that could produce subharmonics to approximate a system that “simply” saturates). Thus, depending on the system, different approaches exist to propose models for nonlinear dynamic systems. Some of these approaches will be presented shortly here, focusing on the class of systems that can be modeled. Since no model structure can cover all possible nonlinear systems, approximations will be introduced. However, these approximations (and their approximation properties) depend on the input signal and the approximation criterion. The set of excitation signals $E$ is specified first, followed by the discussion of the approximation criteria in Section 1.1.2.2.

1.1.2.1 The class of excitation signals $E$

In this work, normally distributed random excitations with a user defined power spectrum $S_u(f)$ will be used. These excitation signals have two interesting properties: they are omnipresent, and all their higher order moments exist and can be easily computed. Within this class, the focus will be mainly on periodic signals (random multisines, see Definition 1.4), but all results in this chapter are valid for the broader class of excitations. The overall structure of $E$ is given in Figure 1.6.

All excitations will be specified as discrete time signals. Although nothing is said about the behavior of the excitation between the sample points, the reader should be aware that it affects the model that will be obtained from the sampled data.
1.1 General remarks

**Definition 1.1** The spectrum generating function $S_u(f)$ is a uniformly bounded real positive function with a countable number of discontinuities.

**Definition 1.2** A Gaussian noise excitation is a random sequence $u(t)$, $t = 0, 1, \ldots$ drawn from a zero mean normally distributed process with a user defined power spectral density $S_u(f)$ (Gibson, 1997).

**Definition 1.3** A signal $u(t)$ is a periodic noise excitation if

$$u(t) = N^{-\frac{1}{2}} \sum_{k=-N}^{N} \hat{U}\left(\frac{k}{N}\right) e^{j(2\pi k \frac{t}{N} + \phi_k)}$$

with $\phi_{-k} = -\phi_k$, $\hat{U}(f) \geq 0$ and $\hat{U}(f = 0) = 0$. $\hat{U}\left(\frac{k}{N}\right)$ and $\phi_k$ are the realizations of independent (jointly, and over $k$) random processes satisfying the following condition: $\hat{U}(f) \geq 0$ has bounded moments of any order ($< \infty$), $E[e^{j\phi_k}] = 0$ and $E[\hat{U}^2(f)] = S_u(f)$.

**Definition 1.4** A signal $u(t)$ is a random phase multisine if

$$u(t) = N^{-\frac{1}{2}} \sum_{k=-N}^{N} \hat{U}\left(\frac{k}{N}\right) e^{j(2\pi k \frac{t}{N} + \phi_k)}$$

with $\phi_{-k} = -\phi_k$, $\hat{U}(f) = \sqrt{S_u(f)} \geq 0$ and $\hat{U}(f = 0) = 0$. The phases $\phi_k$ are the realizations of an independent (over $k$) uniformly distributed random process on $[0, 2\pi)$. $u(t)$ is normalized to have a constant power for every $N$ ($\infty$ included).

The set of signals considered in this introduction is the union of the previously defined signals:

**Definition 1.5** A signal $u$ with power spectrum $S_u(f)$ belongs to $E$ if $u$ is either a Gaussian noise excitation (Definition 1.2), periodic noise (Definition 1.3) or a random phase multisine (Definition 1.4).

Remarks:

- The periodic signals (Definitions 1.3 and 1.4) are asymptotically ($N \to \infty$) normally distributed in the time domain.
• The phase condition in Definitions 1.3 and 1.4 can be relaxed. The phases can be restricted to a discrete set as long as $E[e^{j\phi}] = 0$. This allows for example to include orthogonal frequency domain modulation (OFDM) where such random multisines are intensively used.

### 1.1.2.2 Approximation criteria

There are different ways for a model to approximate a real system. This is expressed by the convergence criterion which defines the notion of “close to the true value”. In this chapter, two different convergences will be used: convergence in mean square sense and uniform convergence.

**Definition 1.6** The model $\hat{F}(u, y, \theta)$ converges to a system $F(u, y)$ in mean square sense if $\forall \epsilon > 0, \exists M_\epsilon$, independent of $\theta$ such that for $M$ the number of parameters in $\theta$:

$$\text{for } M > M_\epsilon, \exists \theta_M \Rightarrow E\left[\left| F(u, y) - \hat{F}(u, y, \theta_M) \right|^2 \right] < \epsilon$$

where the mean is computed over the set $E$ of excitations.

**Definition 1.7** The model $\hat{F}(u, y, \theta)$ converges uniformly to a system $F(u, y)$ if $\forall \epsilon > 0, \exists M_\epsilon$ such that for $M$ the number of parameters in $\theta$:

$$\text{for } M > M_\epsilon, \exists \theta_M \Rightarrow \forall u \in E, \forall t : \left| F(u(t), y(t)) - \hat{F}(u(t), y(t), \theta_M) \right| < \epsilon$$

Note that uniform convergence implies mean square convergence.

### 1.1.3 Different systems, different models

Now that the signals and approximation criteria are defined, models and their approximation properties for nonlinear systems can be introduced. Table 1.1 is a summary of different models and their approximation properties for nonlinear systems. The following paragraphs provide the details that did not fit into this condensed view.

#### 1.1.3.1 Models for open loop nonlinear systems

The open loop systems that will be considered here are those that can be approximated arbitrarily well in least squares sense by a Volterra series (1.3) for the class of excitations
### Table 1.1: Convergence properties of different nonlinear model classes.

<table>
<thead>
<tr>
<th>Model</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiener models</td>
<td>- Mean square convergence to Wiener systems with saturations and discontinuities. Derivatives of signals converge for band limited input.</td>
</tr>
<tr>
<td></td>
<td>- Uniform convergence for Volterra systems (Wiener systems with analytic nonlinearities) for a limited system dependent input range (similar to Taylor expansion, derivatives of the system converge, too)</td>
</tr>
<tr>
<td></td>
<td>- Uniform convergence for fading memory systems (Wiener systems with continuous discontinuities) over a user chosen input range. Derivatives of signals converge for band limited input.</td>
</tr>
<tr>
<td>Recurrent Neural Network</td>
<td>Uniform convergence for continuous open loop nonlinear systems. No result was found in the literature about the convergence of the derivatives.</td>
</tr>
<tr>
<td>Bilinear models</td>
<td>Uniform convergence for continuous closed loop nonlinear systems for bounded $u(t)$ and $x(t)$ for a finite time. No result was found in the literature about the convergence of the derivatives.</td>
</tr>
<tr>
<td>NARMAX models</td>
<td>Uniform convergence for continuous closed loop nonlinear systems for bounded $u(t)$ and $x(t)$. No result was found in the literature about the convergence of the derivatives.</td>
</tr>
</tbody>
</table>
\[ y(t) = \int_{-\infty}^{t} h_1(t - \tau)u(\tau)d\tau + \int_{-\infty}^{t} \int_{-\infty}^{t} h_2(t - \tau_1, t - \tau_2)u(\tau_1)u(\tau_2)d\tau_1d\tau_2 + \ldots \quad (1.3) \]

In (1.3), the nonlinear behavior is created by using powers of \( u(t) \) and the dynamic nature is caught in the convolution integrals. If the Volterra kernels \( h_n(\tau_1, \ldots, \tau_n) \) tend to zero as \( \tau_i \to \infty \),

\[ \forall i = 1, 2, \ldots, n : \lim_{\tau_i \to \infty} h_n(\tau_1, \ldots, \tau_n) = 0 \quad (1.4) \]

the system is called a Wiener system. The requirement on the kernels (1.4) means that the system has a finite memory: the influence of the initial conditions vanishes asymptotically. For these systems, the steady state response to a periodic input is a periodic signal with the same period as the input. The nonlinearity may be discontinuous. Since only the input enters the computation of \( y(t) \) in (1.3), phenomena like bifurcations can’t exist for Wiener systems. Each Wiener system can be represented by the block structure shown in Figure 1.7 if the Multiple Input, Multiple Output (MIMO) static nonlinearity has enough inputs and outputs.

Wiener models (Schetzen, 1980) converge in mean square sense to the class of Wiener systems given by (1.3) and (1.4). Of course, the mean square convergence imposes certain restrictions on the usability of the model: the output of the system at the discontinuities cannot be predicted. However, the model is still very usable: the output signal and its derivatives for band-limited inputs converge to their true values in mean square sense inside a limited frequency band (see Appendix 1.A). Nonetheless, the derivatives of the model characteristics do not necessarily converge to their true values.

\[ ^{1} \text{It is not really required that the input signal be a Gaussian input: it suffices that all higher order moments exist and be finite, which is the case for Gaussian signals.} \]
If the nonlinearity is continuous, the system is called a fading memory system (Boyd and Chua, 1985) and a uniformly converging approximation of the system can be obtained for bounded inputs. The bounds inside which the approximation is valid are chosen by the user. However, derivatives of the model still must not be used as derivatives of the system’s characteristics.

If the nonlinearity is further constrained to weak nonlinearities, i.e. nonlinearities where the static nonlinear block in Figure 1.7 is an analytic function, then uniform convergence of the model and its derivatives inside a limited frequency band to the system is obtained. However, the bounds where this approximation is valid depends on the system and can’t be set by the user.

The nonlinear systems that will be studied in this book can be described by Wiener systems, though except for Chapter 7 the Volterra kernels will not be identified directly.

1.1.3.2 Models for closed loop nonlinear systems

In Nonlinear Auto-Regressive Moving Average with eXogenous inputs (NARMAX) modeling, the output is a nonlinear combination of past inputs and outputs:

\[ y(k) = f(y(k-1), y(k-2), \ldots, y(k-n_a), u(k-1), u(k-2), \ldots, u(k-n_b)) \]  

(1.5)

If the response function \( f \) of the system in (1.5) is finitely realizable and a linear model of the system exists when operated in a region around the chosen equilibrium point, the nonlinear discrete-time system can be approximated uniformly by a NARMAX model for bounded inputs and outputs (Leontaritis and Billings, 1985). When using NARMAX, a structure for \( f \) must be chosen. Often, a polynomial representation is chosen (with terms like i.e. \( y(k-1)u(k-1) \) or \( y(k-1)^2u(k-2) \)). Of course, the number of terms grows very fast as the degree of the polynomial increases or as \( n_a \) and \( n_b \) increase (Piroddi and Spinelli, 2003). Because past outputs \( y(k-l) \) are included in the model, NARMAX models can also model bifurcation.

Neural networks are another general nonlinear function approximator. Many variants exist, but they are all built on the repetition of a basic structure: the neuron that combines several (past) inputs and possibly past outputs with a nonlinear activation function. A neural network consists of many of these neurons, organized in layers. So, additionally to the choice of the activation function of the neuron (hinging hyperplanes,
radial basis functions, . . . ), the “organization” of the neural network may be chosen (how
many layers, how many past inputs, how many past outputs). Due to the high number
of parameters to identify, regularization is needed to avoid overfitting as otherwise,
the parameters are used to follow the noise instead of modeling the system. Another
approach is to terminate the iterative search for the parameter values early, i.e. before
the (local) minimum of the cost function is reached. Sjöberg (1995) has shown that this
is similar to regularization. The convergence properties of neural networks are discussed
in Barron (1993); Pinkus (1999); Sontag (1993).

Bilinear models are mostly used in the framework of state space models:

\[
\begin{align*}
\dot{x}(t) &= Ax(t) + Bu(t) + F(u(t) \otimes x(t)) \\
y(t) &= Cx(t) + Du(t)
\end{align*}
\]

where \(x(t)\) is the state vector, \(u(t)\) the input vector (in the most general case) and
\(y(t)\) the output vector (also in the most general case). The matrices \(A, B, C, D\) and
\(F\) have the appropriate dimensions. The symbol \(\otimes\) denotes the Kronecker product.
Bilinear models are somewhat similar to linear models: when either the input or the
state vector is kept constant, they behave like a linear model. A bilinear model can
approximate any nonlinear system in a finite time interval for bounded inputs if the
system is continuous and causal (Mohler and Kolodziej, 1980). However note that these
approximation property only holds for continuous time (Fliess and Normand-Cyrot,
1982).

1.2 Outline of the thesis

A proper definition of the subset of the nonlinear systems that will be studied is given in
Chapter 2: basically, the studied systems will be composed of static nonlinearities and
linear dynamic systems. This allows for a very natural extension of the linear system
identification techniques. Indeed, all identification algorithms that will be presented
here start by estimating a linear model of the nonlinear system.

With these two building blocks, a huge number of combinations can be assembled.
The combinations that have already received a lot of attention in the literature will
be studied in Chapters 3, 4, 5 and 6. The aims of Chapters 3, 4 and 5 are somewhat
different from those of Chapter 6. In Chapters 3, 4 and 5, the focus is on precisely identifying the different building blocks of the nonlinear system. The rationale behind this approach is the desire to be able to pinpoint inside a real system the origin of the overall characteristics (e.g. bandwidth or dominant nonlinear behavior). This means that these chapters are close to the white-box modeling approach: conclusions about the overall system can be only drawn if it corresponds to the chosen model structure. The approach in Chapter 6 is closer to black-box modeling as there, the goal is to enhance the prediction abilities of a linear model with no additional measurement effort and only a small number of additional computations.

The decision to approach nonlinear systems from the angle of these basic building blocks was motivated by the assumption that the blocks map to basic components of a system. The original goal of this research was to identify models that could be used to give hints to designers. An example will better explain what was intended: given the amplifier of Figure 1.8, which part of the circuit limits the bandwidth?

Assuming that the transition frequency $f_T$ of the transistor is much higher than the time constants associated with the resistors and the capacitors, a Wiener-Hammerstein model (Chapter 5) could approximate this circuit: a static nonlinear element (the transistor) sandwiched between two linear circuits (the biasing at the input and at the output). Obtaining a model of these three parts allows to decide where the bandwidth is limited.

Proposing an identifiable model structure and an identification algorithm is only a first step, because all these methods require data to work on. Measuring nonlinear systems
proves to be a daunting task, where some problems had to be overcome. These problems along with proposals for solutions can be found in Chapters 7 and 8. The first problem, shown in Chapter 7 is that the excitation signal can not always be forced upon the DUT and that problems may arise for nonlinear system identification. The second problem that is specific to certain application domains, is one of calibration: at high frequencies, the measurement systems suffer from a high phase distortion which hinders nonlinear modeling, as Chapter 8 shows.

1.3 Publications

The chapters in this thesis are the result of work that has been presented either at conferences or as journal papers.

Parts of Chapter 3 have been published in the IEEE Transactions on Instrumentation and Measurement (Crama and Schoukens, 2001). The iterative enhancement of the estimates has been presented at a conference (Crama et al., 2003) and the convergence proof for the Hammerstein estimation algorithm has been accepted for publication as a Technical Communique in Automatica:


Chapter 4 has been presented at a conference (Crama and Schoukens, 2002b) and a journal paper including the convergence analysis has been accepted for publication as a brief paper by Automatica:


Chapter 5 has been presented at a conference (Crama and Schoukens, 2002a). It has also been accepted for publication in the IEEE Transactions on Measurement and
Instrumentation.


Chapter 6 has been published in Automatica (Schoukens et al., 2003a).


The major part of Chapter 7 has been published in the IEEE Transactions on Microwave Theory and Techniques (Crama et al., 2002).


The measurements leading to the redaction of Chapter 8 have been presented at a conference (Crama and Rolain, 2002) and the proposed wave phase calibration has been presented on another occasion (Crama and Schoukens, 2003).


1 Introduction

1.A Convergence of derivatives of the output signals

1.A.1 In mean square sense

In this section, it will be shown that when a model converges in mean square sense to the true system, then the derivatives of the output converge also in mean square sense inside a limited frequency band to the derivatives of the true output.

Suppose a periodic input signal $u(t)$ with period $T$ is applied to a Wiener system $S_0$, yielding a periodic (after the transients have died out) output $y_0$:

$$y_0(t) = S_0[u(t)]$$

Now, when a model $\hat{S}$ is identified for $S_0$ and converges in mean square sense to $S_0$, one obtains, e.g. for the continuous-time case:

$$\frac{1}{T} \int_0^T (\hat{y}(t) - y_0(t))^2 \, dt < \epsilon \quad (1.6)$$

As $y_0(t)$ and $\hat{y}(t)$ are periodic signals, both can be expressed as Fourier series:

$$y_0(t) = \sum_{k=-\infty}^{\infty} Y_0,k e^{j\omega_k t}$$

$$\hat{y}(t) = \sum_{k=-\infty}^{\infty} \hat{Y}_k e^{j\omega_k t}$$

Due to Parceval’s theorem, (1.6) can be also written as

$$\frac{1}{T} \sum_{k=-\infty}^{\infty} \left| Y_0,k - \hat{Y}_k \right|^2 < \epsilon \quad (1.7)$$

It follows immediately that $\forall k_{\text{max}}$:

$$\sum_{k=-k_{\text{max}}}^{k_{\text{max}}} \left| Y_0,k - \hat{Y}_k \right|^2 < T\epsilon \quad (1.8)$$

The derivative of these signals is easily computed:

$$\frac{dy_0}{dt} = \sum_{k=-\infty}^{\infty} Y_0,k j\omega_k e^{j\omega_k t}$$
1.A Convergence of derivatives of the output signals

\[
\frac{dy}{dt} = \sum_{k=-\infty}^{\infty} \hat{Y}_k e^{j\omega_k t}
\]

The error for the derivative of the output, when limited to a certain bandwidth \(k_{\text{max}}\) is thus given by:

\[
\frac{1}{T} \int_{0}^{T} \left( \sum_{k=k_{\text{max}}}^{k_{\text{max}}} j\omega_k (Y_{0,k} - \hat{Y}_k) e^{j\omega_k t} \right)^2 dt = \frac{1}{T} \sum_{k=-k_{\text{max}}}^{k_{\text{max}}} \left| j\omega_k (Y_{0,k} - \hat{Y}_k) \right|^2 \\
\leq \frac{1}{T} \left| \omega k_{\text{max}} \right|^2 \sum_{k=-k_{\text{max}}}^{k_{\text{max}}} \left| Y_{0,k} - \hat{Y}_k \right|^2 \\
< \frac{1}{T} \left| \omega k_{\text{max}} \right|^2 T \epsilon \\
= \left| \omega k_{\text{max}} \right|^2 \epsilon
\] (1.9)

It follows that if the mean square output simulation error is bounded (1.6), there also exists a bound on the mean square error of the derivatives of the output signal when a limited frequency band is considered (1.9).

1.A.2 Uniform convergence

A similar argument can be developed for the case where the model converges uniformly to the true system. One has

\[
|y_0(t) - \hat{y}(t)| = \left| \sum_{k=-\infty}^{\infty} Y_{0,k} e^{j\omega_k t} - \sum_{k=-\infty}^{\infty} \hat{Y}_k e^{j\omega_k t} \right| < \delta
\] (1.10)

The error \(|Y_{0,k} - \hat{Y}_k|\) of each Fourier coefficient is bounded, as the total error power is \(\delta^2\):

\[
|Y_{0,k} - \hat{Y}_k| < \frac{\sqrt{2} \delta}{2}; \forall k
\] (1.11)

Computing the derivatives as in the previous subsection, the error of the derivatives of the output signal inside a limited frequency band (defined by \(k_{\text{max}}\)) can be computed as follows:

\[
\left| \frac{dy_0}{dt} - \frac{d\hat{y}}{dt} \right| = \left| \sum_{k=-k_{\text{max}}}^{k_{\text{max}}} j\omega_k Y_{0,k} e^{j\omega_k t} - \sum_{k=-k_{\text{max}}}^{k_{\text{max}}} j\omega_k \hat{Y}_k e^{j\omega_k t} \right|
\]
\[ = \left| \sum_{k=-k_{\text{max}}}^{k_{\text{max}}} j\omega_k \left( Y_{0,k} - \hat{Y}_k \right) e^{j\omega_k t} \right| \]
\[ \leq \sum_{k=-k_{\text{max}}}^{k_{\text{max}}} |j\omega_k| \left| Y_{0,k} - \hat{Y}_k \right| |e^{j\omega_k t}| \]
\[ \leq \omega_{k_{\text{max}}} \sum_{k=-k_{\text{max}}}^{k_{\text{max}}} \left| Y_{0,k} - \hat{Y}_k \right| \]
\[ < \omega_{k_{\text{max}}} 2k_{\text{max}} \sqrt{2\delta} = \omega_{k_{\text{max}}} k_{\text{max}} \sqrt{2\delta} \quad (1.12) \]

(1.12) shows that if the output converges uniformly (1.10), the derivatives of the output converge uniformly in a limited frequency band, too.
2 Introduction to Block oriented Systems

The study of nonlinear systems is like the study of non-elephant biology: the field that remains to study is much bigger than the part that has been excluded. The aim of this introduction is first to limit the scope to a reasonable size. Additionally, some basic concepts will be introduced, along with the notations.
2 Introduction to Block oriented Systems

2.1 Block oriented systems

Except where noted otherwise, all studied nonlinear models will have a block-oriented structure: different building blocks will be assembled in specific configurations and studied. These structures are widely used and the most common have a specific name. In this thesis, only Single Input, Single Output (SISO) systems will be studied.

2.1.1 Linear time invariant systems

This section introduces briefly the Linear Time Invariant (LTI) systems. These systems have been used for a long time and their identification has been studied very thoroughly (Ljung, 1999; Pintelon and Schoukens, 2001; Söderström and Stoica, 1989). Most, if not all, of the models treated in this thesis are combinations of LTI systems with static nonlinear blocks. The LTI systems, represented by Figure 2.1 in all schematics, introduce the dynamic behavior that makes the systems “interesting”.

Definition 2.1 A system is called LTI system if its input-output relation satisfies the following conditions:

- The system must be linear:

\[
\begin{align*}
y_1(t) &= H[u_1(t)] \\
y_2(t) &= H[u_2(t)]
\end{align*}
\]

\[\Rightarrow \forall \alpha_1, \alpha_2 : \alpha_1 y_1(t) + \alpha_2 y_2(t) = H[\alpha_1 u_1(t) + \alpha_2 u_2(t)]\]

- The system must be time invariant:

\[y(t) = H[u(t)] \Leftrightarrow y(t - \tau) = H[u(t - \tau)]\]

2.1.2 Static nonlinearities

The output of a static nonlinearity depends only on the present input value, but not on past (or future) inputs. These are the simplest nonlinear system imaginable. They
are represented by a mathematical function which maps the input value at instant $\tau$ to the output value at the same instant. This concept is embodied in Figure 2.2 which symbolizes this nonlinear mapping function.

### 2.1.3 Commonly used structures

The basic blocks can be assembled in series, yielding Hammerstein or Wiener systems when two blocks are used and Hammerstein-Wiener or Wiener-Hammerstein systems when three blocks are used. When identifying such model structures, the difficulty lies in the fact that the signals between two blocks cannot be measured.

#### 2.1.3.1 The Hammerstein system

Hammerstein systems are studied in Chapter 3. As can be seen in Figure 2.3, the input $u(t)$ is first distorted by a static nonlinearity $f$. The unmeasured output $w(t)$ of the static nonlinearity is eventually filtered by the linear system whose Frequency Response Function (FRF) is $S$.

This model has already been used in model predictive control (Wang and Henriksen, 1994; Zhu, 2000). It is also used to model the distortion of nonlinear amplified digital communication signals in satellite and microwave links followed by a linear channel (Greblicki, 1996; Prakriya and Hatzinakos, 1997).

#### 2.1.3.2 The Wiener system

Wiener systems are studied together with Hammerstein systems in Chapter 3. They are grouped together because of their similarity illustrated by comparing Figure 2.3 with
2 Introduction to Block oriented Systems

Figure 2.4: A Wiener system

Figure 2.5: A Hammerstein-Wiener system

Figure 2.4. In the Wiener system, the input signal $u(t)$ is filtered by the LTI system $R$ first. The resulting signal $v(t)$ is distorted afterwards by the static nonlinearity $f$. These are the same operations as for a Hammerstein system, but in a reversed order.

Wiener systems have been used for control (Pajunen, 1992; Zhu, 1999b), in the chemical process industry (Pardo et al., 1998; Norquay et al., 1999; Zhu, 1999a), for power amplifiers (Chrisikos et al., 1998; Christopher Clark et al., 1998) and for the study of seismic data (Taleb et al., 2001).

2.1.3.3 The Hammerstein-Wiener system

A Hammerstein-Wiener system (Figure 2.5) is a combination of a Hammerstein system followed by a Wiener system: the linear part of both systems is concentrated into one linear block surrounded by to static nonlinear blocks. It presents an important difference with the Wiener or the Hammerstein system. This difference, along with an identification method for such systems is discussed in Chapter 4.

The Hammerstein-Wiener system has been studied in Bai (1998, 2002a); Bauer and Ninness (2002); Zhu (2002) and is used for model-based predictive control (Bloemen et al., 2001). The first nonlinearity $f$ could for example be the nonlinearity of an actuator in a plant, followed by the LTI behavior of this plant and its nonlinear characteristic $g$.

2.1.3.4 The Wiener-Hammerstein system

The system in this section that is the most difficult to estimate is the Wiener-Hammerstein system. This is because of the presence of the two LTI systems $R$ and $S$. If there
2.2 The excitation signal: the multisine

The random phase multisine that will be used throughout this book has been defined in Definition 1.4. Besides the theoretical results explained in Section 2.3 linked to the random phase multisine (though they are asymptotically valid for Gaussian signals also), this input signal has other advantages:

- Because it is periodical, the measurements can be averaged over successive periods to reduce the measurement or process noise. This increases the Signal to Noise Ratio (SNR).

- The Fourier transform of the signals can be computed with the Discrete Fourier Transform (DFT) (see Definition 2.2).

- The input signal can be adapted to the identification task: the excitation frequency band can be chosen by setting the amplitudes $U_k$ of certain frequencies to 0. It is also possible to infer some properties of the nonlinear DUT with special multisines (Vanhoenacker et al., 2001). Furthermore, even the probability density function wasn’t the static nonlinearity $f$ it would even be impossible to separate those two LTI systems. The identification of these systems is studied in Chapter 5. There are already a number of papers dedicated to this subject: Bershad *et al.* (2000, 2001); Boutayeb and Darouach (1995); Chen and Fassois (1992); Tan and Godfrey (2002a, 2003); Vandersteen *et al.* (1997).

An amplifier can be thought of as a Wiener-Hammerstein system: the packaging and biasing circuits are essentially the LTI systems around a central nonlinearity that concentrates all the nonlinear behavior of the different elements of the amplifier. An example of this can be found in Silva *et al.* (2001) where a traveling wave tube amplifier is modeled with a Wiener-Hammerstein-like model.

2.2 The excitation signal: the multisine

The random phase multisine that will be used throughout this book has been defined in Definition 1.4. Besides the theoretical results explained in Section 2.3 linked to the random phase multisine (though they are asymptotically valid for Gaussian signals also), this input signal has other advantages:

- Because it is periodical, the measurements can be averaged over successive periods to reduce the measurement or process noise. This increases the Signal to Noise Ratio (SNR).

- The Fourier transform of the signals can be computed with the Discrete Fourier Transform (DFT) (see Definition 2.2).

- The input signal can be adapted to the identification task: the excitation frequency band can be chosen by setting the amplitudes $U_k$ of certain frequencies to 0. It is also possible to infer some properties of the nonlinear DUT with special multisines (Vanhoenacker *et al.*, 2001). Furthermore, even the probability density function
of the input signal used in the identification process can be made to match the probability density function of the signals applied to the DUT under normal operating conditions (Schoukens and Dobrowiecki, 1998). This is important because the model of a nonlinear system is optimized for the class of signals that were used to identify it (unless the chosen model structure can exactly represent the true system), so if a useful model is to be obtained, the input signals used for the identification of the system must be close to the real input signals.

### 2.2.1 Nonlinear systems and periodical signals

Using a sine wave \( u(t) = A \sin(2\pi f_0 t) \) as excitation and \( y(t) = f(u(t)) = u(t) + u(t)^2 + u(t)^3 \) as nonlinearity, some general facts about nonlinear systems will be derived. It will be seen that the even contribution \( u^2 \) is very different from the odd contribution \( u^3 \).

\[
\begin{align*}
y(t) &= u(t) + u(t)^2 + u(t)^3 = A \sin(2\pi f_0 t) + A^2 \sin^2(2\pi f_0 t) + A^3 \sin^3(2\pi f_0 t) \\
&= A \sin(2\pi f_0 t) + A^2 \cos(2\pi 2f_0 t) + 3A^3 \sin(2\pi f_0 t) - A^3 \sin(2\pi 3f_0 t) \\
&= \underbrace{\frac{1}{2}}_{1} + A \sin(2\pi f_0 t) + \underbrace{\frac{3A^3}{4}}_{3} \sin(2\pi f_0 t) - A^2 \cos(2\pi 2f_0 t) - \underbrace{\frac{A^3}{4}}_{3} \sin(2\pi 3f_0 t)
\end{align*}
\]

In (2.1) and (2.2), the digits under the braces show the degree of the nonlinearity creating the nonlinear contribution. In (2.2), the terms have been sorted so that the Fourier coefficients of \( y(t) \) can be seen easily. The linear contribution behaves as expected: the input and the output frequencies are equal. Both nonlinear parts cause new lines to
appear. The difference between the quadratic $u^2$ and the cubic $u^3$ contributions can also be seen in Figure 2.7: $u^2$ causes energy to appear elsewhere than the excited frequency in the input, while $u^3$ creates contributions elsewhere and at the excited frequency, too.

When the signal is a multiharmonic signal, e.g. $u(t) = A_1 \sin(2\pi f_0 t) + A_2 \sin(2\pi 3f_0 t) + A_3 \sin(2\pi 5f_0 t)$, the frequencies in the output spectrum are found by looking for all combinations $f_i + f_j$ over the positive and negative frequencies ($-10f_0, -8f_0, -6f_0, \ldots, 8f_0, 10f_0$ in this example). A similar rule holds for the cubic nonlinearities: all combinations $f_i + f_j + f_k$ have to be considered again, resulting in excited frequencies $-15f_0, -13f_0, -11f_0, \ldots, 11f_0, 13f_0, 15f_0$ at the output for this example. For such an excitation, called an odd multisine because only odd multiples of $f_0$ are present in the input, the even $u^2$ contributions can be separated easily from the odd contributions $u^3$, by looking at either the even or the odd multiples of $f_0$ in the output spectrum. It is possible to define multisines where different sets of multiples of $f_0$ are excited (Vanhoenacker et al., 2001). A last example that will be given here because it is used in Chapter 5 is the special-odd multisine where $f_0, 3f_0, 9f_0, 11f_0, \ldots, 8n + 1, 8n + 3, \ldots$ are excited.

Taking all the combinations $f_i + f_j$ for the nonlinear term $u^2$ results in even multiples of $f_0$ in the output spectrum, as for the odd multisine. However, in the output spectrum, there are now two sets of odd multiples of $f_0$: those that are only excited by $u^3$ (e.g. $5f_0$ or $7f_0$) and those that are also excited by $u$ (e.g. $f_0$ or $3f_0$). This property of the special-odd multisines is used in Chapter 5.

### 2.3 The Related Linear Dynamic System

Block-oriented systems without feedback excited by periodical signals have a periodical output with the same period. What can be said about the ratio of the Fourier coefficients at the same frequency at the output and the input? In other words, what are the properties of an approximation of the nonlinear system by an LTI system? This section summarizes some results answering this question, which has been studied in the literature (Schoukens et al., 1998a; Pintelon and Schoukens, 2001).

The nonparametric FRF of a nonlinear system is given by

$$G_{NP}(\Omega) = \frac{Y(\Omega)}{U(\Omega)} = G_R(\Omega) + G_S(\Omega) \quad (2.3)$$
Working with random phase multisines allows to study how $G_{NP}$ varies for different realizations of the random phase, while keeping the amplitudes of the Fourier coefficients of the input constant. The contributions to $G_{NP}$ that are independent of the actual phase realization of the input are gathered in $G_R$. All other contributions of the nonlinearities are called “stochastic contributions” and are grouped in $G_S$ in (2.3) (Schoukens et al., 1998a). Note that $G_R(\Omega)$ can be expressed as a function of the frequency, while $G_S(\Omega)$ which collects the random contributions, depending on the phase realization, is a random variable.

It has been shown by Schoukens et al. (1998a) that $G_S(\Omega)$ has zero mean:

$$E_\phi[G_S(\Omega)] = 0$$  \hspace{1cm} (2.4)

where $E_\phi[G_S(\Omega)]$ is the average of $G_S(\Omega)$ taken over the realizations of the phases of the input multisine.

This means that averaging $G_{NP}$ over many realizations of the phase of the input signal yields $G_R$. Instead of an averaging procedure which is expensive in measurement time, a parametric model can be fit to $G_{NP}$. It has been shown (Schoukens et al., 1998a) that because of the properties of $G_S$, this estimate will converge strongly to $G_R$ as $N \to \infty$.

The Related Linear Dynamic System (RLDS) $G_R$ can be divided again into two parts:

$$G_R(\Omega) = G_0(\Omega) + G_B(\Omega)$$  \hspace{1cm} (2.5)

In (2.5), $G_0(\Omega)$ is the Volterra kernel of order 1 and $G_B(\Omega)$ are the nonlinear bias contributions. The nonlinear bias contributions $G_B(\Omega)$ are that part of the nonlinear contributions that do not change when a multisine with the same power spectrum and a different random phase realization is applied to the nonlinear system. However, $G_B$ varies with the input power.

### 2.3.1 The RLDS of a Wiener-Hammerstein system

This section presents a main result related to the RLDS. It is included in this introduction because its conclusion (2.6) is central in the identification of Hammerstein, Wiener and Wiener-Hammerstein systems.

For a Wiener-Hammerstein system, the RLDS is (Pintelon and Schoukens, 2001;
2.3 The Related Linear Dynamic System

Schoukens et al., 1998a)

\[ G_R(\Omega) = C(|RU|, f)R(\Omega)S(\Omega) + O\left(\frac{1}{N}\right) \quad (2.6) \]

(2.6) will be derived here as preparation for a similar computation in Section 4.2. The computation consists of 3 steps:

1. Compute the Volterra kernels.
2. Decide which Volterra kernels contribute to the RLDS.
3. Explain the \( O(N^{-1}) \) term in (2.6).

Note that for simplicity, the computations are done in the continuous time domain \((\Omega = \omega)\). If the discrete time domain \((\Omega = e^{j\omega})\) is preferred, then sums of the form \(\omega_{k_1} + \omega_{k_2}\) have to be interpreted as meaning \(e^{j\omega_{k_1} + j\omega_{k_2}}\). A second point of interest is that some \(U(\omega)\) that are used in the following equations may be 0. One instance of this is \(U(0) = 0\): the operating point of the modeled systems is never taken into account in this book.

For a Wiener-Hammerstein system as in Figure 2.6 with a polynomial static nonlinearity

\[ f(v) = \sum_{a=1}^{q} \varphi_a v^a \]

and whose linear systems have a transfer function \(R(\omega)\) and \(S(\omega)\), the Volterra kernel of order \(\alpha\) is given by

\[ H_\alpha(\omega_{k_1}, \omega_{k_2}, \ldots, \omega_{k_a}) = S\left(\sum_{a=1}^{\alpha} \omega_{k_a}\right) \varphi_\alpha R(\omega_{k_1}) R(\omega_{k_2}) \ldots R(\omega_{k_a}) \quad (2.7) \]

The theory of Volterra kernels (Schtzen, 1980) defines how the Volterra kernel of (2.7) influences the output:

\[ Y(\omega) = Y_1(\omega) + Y_2(\omega) + \ldots Y_q(\omega) \]

\[ = H_1(\omega) + \sum_{\omega_{k_1} = -N}^{N} H_2(\omega - \omega_{k_1}, \omega_{k_1}) U(\omega - \omega_{k_1}) U(\omega_{k_1}) + \]

\[ + \sum_{\omega_{k_1} = -N}^{N} \sum_{\omega_{k_2} = -N}^{N} H_3(\omega - \omega_{k_1} - \omega_{k_2}, \omega_{k_1}, \omega_{k_2}) U(\omega - \omega_{k_1} - \omega_{k_2}) U(\omega_{k_1}) U(\omega_{k_2}) + \]
Now, by averaging $G_{NP}$ over the phase realizations of the input signals (in Pintelon and Schoukens (2001), more classes of input signals are studied but this is outside the scope of this introduction), the contribution $G_{R,\alpha}$ of each Volterra kernel $H_\alpha$ can be computed:

$$G_{R,\alpha}(\omega) = \mathbb{E}_\phi \left[ \frac{Y_2(\omega)}{U(\omega)} \right] \tag{2.9}$$

The value of $G_{R,\alpha}(\omega)$ depends on the parity of $\alpha$. First $\alpha = 2n + 1$ odd will be studied, followed by $\alpha = 2n$ even. For $\alpha = 2n + 1$ odd, the phase of the terms inside the multiple sum (2.8) is given by (2.10) (the phase $\phi_U(\omega)$ of $U(\omega)$ is subtracted because of the division in (2.9)):

$$\phi_{H_{2n+1}}(\omega - \sum_a \omega_{ka}) + \phi_U(\omega - \sum_a \omega_{ka}) + \phi_U(\omega_{k1}) + \ldots + \phi_U(\omega_{k2n}) - \phi_U(\omega) \tag{2.10}$$

In (2.10), the phase of $\phi_{H_{2n+1}}(\omega - \sum_a \omega_{ka})$ is deterministic, and there is an even number of random phases $\phi_U$. Hence, some terms will have a deterministic phase: those where each $\omega_{ka}$ is paired with an $\omega_{kb} = -\omega_{ka}$, as $U(\omega)$ is the spectrum of a real valued time signal $u(t)$ and thus $\phi_U(\omega_{ka}) = -\phi_U(-\omega_{ka})$. A consequence of this particular selection of $\omega_{ka}$ values is that $\omega - \sum_a \omega_{ka} = \omega$, resulting in:

$$\phi_{H_{2n+1}}(\omega - \sum_a \omega_{ka}) + \phi_U(\omega - \sum_a \omega_{ka}) + \phi_U(\omega_{k1}) + \ldots + \phi_U(\omega_{k2n}) - \phi_U(\omega) = \phi_{H_{2n+1}}(\omega - \sum_a \omega_{ka}) + \phi_U(\omega - \sum_a \omega_{ka}) - \phi_U(\omega) = \phi_{H_{2n+1}}(\omega - \sum_a \omega_{ka}) + \phi_U(\omega) - \phi_U(\omega) = \phi_{H_{2n+1}}(\omega - \sum_a \omega_{ka})$$

which is deterministic, as was stated in the beginning of this paragraph.

It can be shown that the even $\alpha = 2n$ Volterra kernels do not contribute to the RLDS, i.e. $G_{R,2n}(\omega) = \mathbb{E}_\phi \left[ \frac{Y_{2n}(\omega)}{U(\omega)} \right] = 0$. This can be seen by looking at the phase of the terms inside the multiple sum of (2.8) and subtracting the phase $\phi_U(\omega)$ of $U(\omega)$ (because of the
The previous equation is only an approximation, in reality, \(C(\lvert RU\rvert, f)\) depends on the frequency because of the different factors (3 and 6) in (2.13). However, since \(\lvert U(\omega)\rvert\) is \(O(N^{-\frac{1}{2}})\) (see Definition 1.4), the error is the \(O(N^{-1})\) term in (2.6). As such, it will disappear if enough spectral components are present in the excitation.
2.4 Noise

The identification methods for the systems will be tested on measurements wherever possible. This means that the data will be corrupted with noise. But, save for some exceptions, the presence of noise will be ignored (see Figures 2.3, 2.4, 2.5 and 2.6 where no noise is considered). However, the problem posed by noise is acknowledged implicitly throughout this book: all presented methods avoid operations which are known to yield bad results in the presence of noise.

A well-known example of a bad practice with respect to the noise properties is the estimate of a linear system with transfer function $\frac{B(\Omega)}{A(\Omega)}$ by minimizing the cost function (Levy, 1959)

$$\sum_{l=1}^{N} \left| B(\Omega_l) U(\Omega_l) - A(\Omega_l) Y(\Omega_l) \right|^2$$

Such cost functions will be avoided here. So, even though no explicit attention is devoted to noise issues most of the time, the proposed algorithms are sound, as e.g. Appendices 3.E and 3.F show.
2.A The Fourier Transform

The periodicity of the excitation signals make the frequency domain attractive, because the efficient Fast Fourier Transform (FFT) algorithm can be used to go from the time domain to the frequency domain.

**Definition 2.2** The DFT of a signal $u(t)$ that has been sampled at the moments $0, 1, 2, 3 \ldots N - 1$ is

$$U(l) = \mathcal{F}(u(t)) = \frac{1}{N} \sum_{t=0}^{N-1} u(t) e^{-j2\pi tl/N}$$

Its inverse is

$$u(t) = \mathcal{F}^{-1}(U(l)) = \sum_{l=0}^{N-1} U(l) e^{j2\pi tl/N}$$

Throughout this book, $\Omega$ will be used as the frequency variable, because the algorithms can be applied to continuous time modeling ($\Omega = j\omega$) or to discrete time systems ($\Omega = e^{j\omega}$).

2.A.1 Why is the frequency domain so important?

2.A.1.1 Filtering

Many of the operations on signals are done in the frequency domain, as there is one operation which is much easier if the signals are already in the frequency domain: the avoidance (or selection) of certain nonlinear contributions. In the frequency domain, this is done easily by putting the unneeded Fourier coefficients to 0. An example of the importance of this will be given in Section 3.9. Another example is the identification of Wiener-Hammerstein systems in Chapter 5 that depends on the possibility to extract only the nonlinear contributions inside a signal.

2.A.1.2 Averaging $G_{NP}$

The estimation algorithms of Wiener systems, Hammerstein systems, Hammerstein-Wiener systems or Wiener-Hammerstein systems all use the $G_R$, starting from its nonparametric representation $G_{NP}$. One way to obtain $G_R$ from $G_{NP}$ is

$$G_R(\Omega) = E_\phi[G_{NP}(\Omega)]$$
The other way is to fit a parametric transfer function model through $G_{NP}$. These two approaches may be applied together, starting with the averaging step to reduce the number of data (see Algorithm 3 or Algorithm 6).

In the time domain, the average of all impulse responses would have to be computed. However, the impulse response is not readily available, so one model would have to be estimated for each input signal followed by the averaging step. This would involve more computations.
3 Hammerstein systems and Wiener systems

The simplest nonlinear systems that can be built with two blocks are the Hammerstein system (Figure 2.3) and the Wiener system (Figure 2.4). In a Hammerstein system, the input first undergoes a nonlinear distortion in a static nonlinear block before being filtered in an LTI block. In a Wiener system the order of the blocks is inverted: the input is filtered by an LTI block and then distorted by the static nonlinear block.

These two structures are grouped in one chapter because their identification procedures are very similar. The methods in this chapter present the basic ideas for most of the other identification methods of this thesis. They aim to give a fast first approximation of their characteristics. More complex estimators that build on these initializations will be studied in another chapter.
3.1 Introduction

The two systems that are studied in this chapter are very similar. First comes the Hammerstein system: starting from the basic estimation method in Section 3.2, two variations on this method will be given in Sections 3.3 and 3.4. The parameterization of the estimates is common to all methods in this chapter, and is presented in Section 3.5. The same structure is repeated for the Wiener systems in Sections 3.6, 3.7 and 3.8. A comparison with other work is given in Section 3.10. The main goal was to obtain simple methods that could be used as starting values for other more complex algorithms. Hence, the convergence properties and the statistic properties of the estimators are not included in the main text, but rather in the appendix. For the same reason, all methods will be explained as if there was neither process nor measurement noise, unless a particular point needs to be made, as in Section 3.2.3.

3.2 Simple identification of Hammerstein systems

This section explains a first approach to the Hammerstein estimation problem. The presented method is simple, but worth a section on itself nonetheless as the basic ideas and guiding principles of this research can be explained here on a clear example. The method is demonstrated on measured data used in Crama and Schoukens (2001). In order to keep the focus on the method, a description of the measurements has been delayed until Section 3.3.2 and the discussion of the parameterization choices is deferred until Section 3.5.

3.2.1 Informal presentation of the method

In a Hammerstein system, there is only one linear system. Comparing Figure 2.3 with Figure 2.6, it is obvious that a Hammerstein system is a special case of the Wiener-Hammerstein system where the linear input system $R$ is trivial: $R(\Omega) = 1$. For the Hammerstein system, (2.6) defining the RLDS simplifies to

$$G_R(\Omega) = C(U, f)S(\Omega) + O(N^{-1}) \quad (3.1)$$

This means that if the RLDS of the Hammerstein system can be obtained, its linear
3.2 Simple identification of Hammerstein systems

The $\varphi_k$ are weights applied to the basis functions.

Figure 3.1: Splitting the static nonlinearity in basis functions. The $\varphi_k$ are weights applied to the basis functions.

Figure 3.2: Two different Hammerstein systems with identical observable behavior.

part $S$ is known within a constant. This step is explained in more detail in Section 3.2.2.

To estimate the static nonlinearity, the Hammerstein system is represented as a sum of $q$ basic Hammerstein systems (see Figure 3.1), using the estimate of the linear part. This parameterization yields a minimization problem which is linear in the parameters. The reasons and advantages of this choice are explained further in Section 3.2.3.

3.2.2 About the estimate of the linear part

This is the first step of the estimation algorithm: the RLDS of a Hammerstein system is equal to its linear part, multiplied by an unknown scaling constant (3.1). However, this scaling constant is not an obstacle because estimating the absolute gain of the linear part is an unidentifiable problem, anyway. This is illustrated in Figure 3.2 that shows two Hammerstein systems which look equal to an observer who cannot measure the internal signal $w$. Hence, this issue will be ignored and the gain of the linear system will be normalized to 1 at an arbitrary frequency chosen by the user. Because of this arbitrary gain distribution, each plot in this chapter will be silently rescaled to allow for a meaningful comparison between the true and the estimated values.
The goal of this step is to recover the RLDS of the Hammerstein system which is equal to $S$. The only quantity related to the RLDS that is easily available is the nonparametric FRF of the system (2.3). Recall from Section 2.3 that this FRF contains not only the RLDS $G_R$ but also the stochastic contributions $G_S$. This is illustrated in Figure 3.3. A superficial look at Figure 3.3 suggests that the performed measurements were very noisy. However, this is not the case, as the plot of the standard deviation of the nonparametric FRF proves. The standard deviation in Figure 3.3 has been compute by measuring 16 periods with the same excitation signal.

When a parametric model is fitted consistently to the data points of $G_{NP}$, the stochastic contributions are asymptotically ($N \rightarrow \infty$, with $N$ the number of excited frequencies) rejected because the limited number of parameters cannot follow their random variations. Figure 3.4 shows how a parametric estimate approximates the RLDS. Note that the errors $G_R - \hat{G}_R$ are well below the stochastic contributions.

### 3.2.3 About the estimate of the nonlinear part

Once the linear part $S$ is known, the static nonlinearity $f$ can be estimated. This has to be done carefully: it may be tempting to invert $S(\Omega)$ to estimate the unknown intermediate signal $w$ first (see Figure 2.3). This would allow to graph the function $f$ nonparameterically. However, this is avoided for two reasons:

![Figure 3.3: Comparison of the nonparametric FRF $G_{NP}$ (black dots) with the RLDS $G_R$ (full line) and the stochastic contributions $G_S$ (pluses) of a Hammerstein system and the standard deviation of the disturbing noise obtained from repeated experiments with the same excitation signal (grey dots).](image)
3.2 Simple identification of Hammerstein systems

Figure 3.4: The estimation error (dashed line: $G_R - \hat{G}_R$) of the RLDS (full line) compared with the estimate of $G_S$ (pluses) and the standard deviation of $G_{NP}$ (grey dots)

- To get the time domain values of $w$, its Fourier coefficients have to be known at all frequencies. This is not the case because the linear part $S$ is only known inside the frequency band where the excitation was located. The nonlinearity $f$ produces harmonics outside the excitation band, which can only be estimated by extrapolating the model of $S$ at frequencies where it was not measured.

- The linear part $S$ may attenuate the signal at certain frequencies. When the inverse of $S$ is applied to $Y$, the attenuated frequencies (where the SNR is often the worst if for example additive white noise is assumed) are amplified most. This means that the estimated $w$ will be very sensitive to noise distortions.

To avoid these problems, another approach was chosen: the Hammerstein system is represented as a sum of $q$ basic Hammerstein systems (see Figure 3.1). The input is passed through each of these basic Hammerstein systems and the weights $\varphi_a$ are chosen to approximate in least squares the measured output. This approximation is done in the frequency domain so that only the frequency band where the model of the linear part is valid is considered. Because of this choice, the minimization problem to compute $\varphi_a$ reduces to a linear least squares problem.

For each of the basis functions $f_a$, the contribution $\hat{Y}_a$ to the output is computed as

$$W_a(\Omega) = \mathcal{F}(f_a(u)); a = 1, 2, \ldots, q \quad (3.2)$$
Figure 3.5: Comparison of the estimate of the static nonlinearity (black line, circles: breaking points of the piecewise linear characteristic) with the true value of $f$ (black line). The scale to the right of the plot is used to zoom in on the difference (grey line) between the model and the true value.

$$\hat{Y}_a(\Omega) = S(\Omega)W_a(\Omega) \quad (3.3)$$

The weighted sum of these contributions should approximate $Y$ in least squares sense for the set of measured frequencies inside the excitation band. Hence the following cost function is minimized with respect to $\varphi$:

$$V = \sum_{l=1}^{N} \left| Y(\Omega_l) - \sum_{a=1}^{q} \varphi_a \hat{Y}_a(\Omega_l) \right|^2 \quad (3.4)$$

This cost function is easily minimized, since (3.4) is quadratic in the parameters $\varphi$. Also, the problems mentioned at the beginning of this section have been avoided, since no inverse FRF is needed.

Different choices can be made for the basis functions. Due to the saturating nature of $f$ in the present example, a polynomial approximation is less suited, so a piecewise linear function was used in Figure 3.5. The estimated parameters $\varphi$ are the function values associated with each breaking point. More information on this topic can be found in Section 3.5.

The algorithm for a simple identification of Hammerstein systems can be summarized as follows:

**Algorithm 1.** Simple identification of a Hammerstein system.
3.3 Iterative identification of Hammerstein systems

The simple identification of Hammerstein systems is not perfect: the output error in Figure 3.6 is bigger than the standard deviation of the output measurements. A simple extension of the previous algorithm would be to use the estimate of the static nonlinearity to improve the estimate of the linear part. This idea leads to an iterative algorithm.

After the formal definition of the algorithm, the process will be demonstrated on measured data in Section 3.3.2. The convergence of the optimization method is analyzed in Appendix 3.B.

Step 1.1 Estimate a parametric model of the RLDS, $\hat{S} = G_R$.

Step 1.2 Simulate $q$ different basic Hammerstein systems (3.3).

Step 1.3 Estimate $\varphi$ to approach the measured output in a least squares sense with a weighted sum of the basic Hammerstein systems (3.4).

The result of the identification process is shown in Figure 3.6, which shows that the model is able to predict the output for another data set as well as for the identification data set.

Figure 3.6: Output simulation error for Algorithm 1 for validation data (black) and the identification data (grey). The dots are the output measurements; the pluses are the output simulation error; the lines are the standard deviation of the output measurements.
3.3.1 The algorithm

As the goal of the iterative scheme is to improve on the results of the simple identification scheme presented in Section 3.2, it starts with the method described by Algorithm 1. If the true input of the linear part was known, the estimate $\hat{S}$ would not be disturbed by the stochastic nonlinear contributions. An approximation of $w$ can be obtained by applying the estimated static nonlinearity $\hat{f}$ to the input. If this results in a better estimate of $S$, then $f$ can be estimated again, using the improved estimate of $S$.

3.3.1.1 Estimating the linear part

The initial estimate of the linear part at the start of the iterative process is obtained by fitting a parametric model to $G_{NP}$. This is the same as Step 1.1 of Algorithm 1:

$$\hat{\theta}^{[0]} = \arg \min_\theta N \sum_{l=1} |Y(\Omega_l) - S(\Omega_l; \theta)|^2$$  \hspace{1cm} (3.5)

This estimation step can be done with any method in the time or in the frequency domain. In this case, the software package ELiS (Kollár et al., 1991, 2002) for Matlab was used. The advantage of working in the frequency domain is that it is easy to select only those frequencies that were excited. The input and output signals are periodic, hence the Fourier coefficients represent these signals and filtering is as simple as selecting the correct coefficients.

It is important to fit a parametric model to the linear system: the stochastic contributions act as noise on the estimator and their influence is reduced in the parametric step. If this suppression of the stochastic contributions is not done, the iterative algorithm converges in one step to a wrong solution. This will be proven formally in Appendix 3.B.

Once the iterative process has been started, the previous estimate $\hat{f}^{[k-1]}$ can be used to estimate the internal signal:

$$\hat{W}^{[k-1]} = \mathcal{F}\left(\hat{w}^{[k-1]}(t)\right) = \mathcal{F}\left(\hat{f}^{[k-1]}(u(t))\right)$$  \hspace{1cm} (3.6)

Using (3.6) to estimate $\hat{S}_{NI}^{[k]}$, reduces the stochastic nonlinear contributions in the next estimate of the linear part, as Figure 3.7 shows that the estimate of the stochastic
3.3 Iterative identification of Hammerstein systems

Figure 3.7: The estimate $\hat{S}^{[1]}_{\text{NP}}$ (black dots); the RLDS (full line) and the $\hat{S}^{[0]}_{\text{NP}}$ (grey dots) together with the estimate of the stochastic contributions $G_R - \hat{S}^{[1]}_{\text{NP}}$ (black pluses) and the estimate of the stochastic contributions $G_R - \hat{S}^{[0]}_{\text{NP}}$ (grey pluses).

contributions $G_R - \hat{S}^{[1]}_{\text{NP}}$ is lower than $G_R - \hat{S}^{[0]}_{\text{NP}}$:

$$\hat{S}^{[k]}_{\text{NP}}(\Omega) = \frac{Y(\Omega)}{W^{[k-1]}(\Omega)}; k \geq 1$$

$$\hat{\theta}^{[k]} = \arg\min_\theta \sum_{l=1}^N \left\| \frac{Y(\Omega_l)}{W^{[k-1]}(\Omega_l)} - S_\theta(\Omega_l; \theta) \right\|^2$$

(3.7)

3.3.1.2 Estimating the static nonlinear part

The same idea as in Section 3.2.3 can be used here:

$$\hat{Y}^{[k]}_a(\Omega) = \hat{S}^{[k]}_{\text{NP}}(\Omega)W_a(\Omega); a = 1, 2, \ldots, q$$

(3.8)

$$V^{[k]} = \sum_{l=1}^N \left| Y(\Omega_l) - \sum_{a=1}^q \hat{\varphi}^{[k]}_a \hat{Y}^{[k]}_a(\Omega_l) \right|^2$$

(3.9)

Note that in (3.8), the signals defined in (3.2) can be reused at each iteration step without modification as they depend only on the input data and the choice of basis functions.

Algorithm 2. Iterative identification of a Hammerstein system

Step 2.1 Compute the RLDS and then compute an estimate of the static nonlinearity (3.5) and (3.4).

Step 2.2 $k \leftarrow 1$. 

41
Step 2.3 Using $\hat{f}^{[k-1]}$, estimate $\hat{w}^{[k-1]}(t)$ (3.6).

Step 2.4 Build a parametric model $\hat{S}^{[k]}$ of the linear system that approaches $y(t)$ best with $\hat{w}^{[k-1]}(t)$ as input. Normalize the gain to 1 at a user chosen frequency.

Step 2.5 Use $\hat{S}^{[k]}$ to obtain a parametric model of the static nonlinearity $\hat{f}^{[k]}$.

Step 2.6 If the convergence criterion is not met, $k \leftarrow k + 1$ and jump to Step 2.3.

In Step 2.2 and Step 2.6, the notation $k \leftarrow x$ means “assign the value of $x$ to $k$”. The convergence criterion in Step 2.6 may be freely chosen. In the simulation experiments that were done, the output simulation error decreases significantly in the first 10 steps (more or less), after which it reaches a minimum. The observed variations of the output simulation error are smooth. Two different stop criteria were tried: stopping after a fixed number of iterations or stopping once the relative cost function (3.9) variation falls below a threshold. Both gave similar results. The energy of the output simulation error can be used as a cost function. Instead of using the output simulation error of the training data, the output simulation error of validation data could be used to avoid overfitting. However, this is not absolutely necessary since the number of parameters is very small compared to the number of measurements (less than 50 parameters, 800 frequency bins are used in the cost functions).

Applying Algorithm 2 to measurement data yields a good result, as the output simulation error for validation data clearly drops: Figure 3.8 shows that the best model
3.3 Iterative identification of Hammerstein systems

makes a better prediction than the model obtained from Algorithm 1. However, the fit
is not perfect, as the output error for the validation data around 2kHz is still significantly
larger than the standard deviation of the output noise. This is probably the result of two
different causes: the parameters of $f$ and $S$ are optimized separately and there was also
some disturbing noise at the input of the measurement system. The nonlinear estimator
that would optimize the parameters of both $f$ and $S$ simultaneously is a special case
of the nonlinear estimator that will be presented for Wiener-Hammerstein systems in
Chapter 5. This nonlinear search can be initialized with the results of this section.

3.3.2 Demonstration on a measurement example

The test system used in Vandersteen (1997) has been chosen as a test object, because
the signal $w$ could be measured. While obviously these values were not used for the
identification process, they are useful to compare the obtained results with the true
values. The measurement data have already been used to demonstrate various steps of
Algorithm 1 and Algorithm 2. This section discusses how the measurements were made,
and gives the parameterization choices for the linear and the nonlinear parts.

3.3.2.1 The measurements

The system was excited by an arbitrary function generator (HP E1445A) and measured
with two samplers (HP E1437A).

As required by the described method, a random phase multisine was used to excite
the DUT. The excited band stretched from the lowest available frequency grid line
($\approx 76.3$Hz) to approximately 10kHz. The sampling frequency was 156.25kHz, with
a period of 2048 points. This results in a frequency resolution of $\approx 76.3$Hz. The excitation signal was a full multisine (each of the 130 frequency lines in the frequency
band of interest was excited) with a flat amplitude spectrum. 16 periods were measured
of which only the 13 last periods were used to guarantee that the measurements were
done under steady state conditions.

Notice that the sampling frequency was chosen to be much bigger than the highest
excitation frequency. This makes the computations easier afterwards: the static non-
linearity introduces harmonics at higher frequencies. These harmonics could disturb
the frequencies of interest (DC $\rightarrow 10$kHz) by aliasing if no precautions are taken. By
leaving some headroom in the spectrum, no aliasing appears during the calculations. Alternatively, upsampling techniques can be used (Crochiere and Rabiner, 1983).

3.3.2.2 Parameterization

The model structure for this example is defined in this section. Section 3.5 contains a discussion about other possibilities.

The linear part was identified using a discrete time model ($\Omega = z^{-1}$) with 3 zeros and 3 poles:

$$S_\theta(z^{-1}) = \frac{\sum_{n=0}^{3} \theta_n z^{-n}}{1 + \sum_{n=1}^{3} \theta_{n+3} z^{-n}}$$

The nonlinear part was modeled with a piecewise linear function. This function had 25 breaking points, distributed over the input range such that each interval where the basis functions (triangles) are different from 0 covers the same number of samples.

3.3.2.3 Results

The results at different steps of the algorithm have already been shown. The most important of these plots is perhaps Figure 3.8 which shows that the model is able to predict adequately a validation data set that was not used to identify the model. Two
3.3 Iterative identification of Hammerstein systems

additional figures, Figure 3.9 and Figure 3.10 compare the true values with the best model that was obtained. Note the zero in the transfer function around 6kHz. If the linear system had been inverted, this zero would have blown up the noise far above the signal.

Finally, Figure 3.11 shows that the normalized output simulation error drops smoothly. The normalized output simulation error is the square root of the energy of the output error divided by the energy of the measured output. The algorithm stopped after 13 iterations because the relative variation was smaller than 0.0001dB (0.001%).
3.4 Estimating Hammerstein systems using multiple experiments

The iterative Algorithm 2 can be enhanced for the case where many different excitation signals are measured. The stochastic contributions $G_S$ can be diminished not only by fitting a parametric model to the nonparametric FRF, but also by averaging $\hat{S}_{NP}$ over different phase realizations of the input signal. This lowering of the stochastic contributions obviously improves the convergence of Algorithm 2 as the first guess of $S$ will be closer to its true value. But there is even more information that can be used: the variance of the stochastic contributions can be used as a weight function when the linear model is estimated (3.7). This is a variation on Algorithm 2, so this section will be kept short, focusing on the changes that need to be made.

3.4.1 Definitions

Many different random phase multisines with the same power spectrum can be produced by choosing another realization for the phases in (1.2). The use of different excitation signals not only allows to reduce the stochastic contributions (Schoukens et al., 1998a; Pintelon and Schoukens, 2001) but also permits to quantify their effect.

To accommodate the modifications, some notations need to be changed, too. Suppose that $E$ multisines with different phase realizations were used in the experiment, then the following equations define the modified procedure which is explained in Algorithm 3. The basic idea remains the same as in Algorithm 2: start the algorithm by estimating the RLDS, assign it to the linear part of the system then estimate the static nonlinearity. Iterate by alternatively freezing $f$ while estimating $S$ and freezing $S$ while estimating $f$. This is formalized in Algorithm 3.

\[
\hat{S}_{NP}^{[0]}(\Omega) = \frac{1}{E} \sum_{m=1}^{E} \frac{Y_m(\Omega)}{U_m(\Omega)} \quad (3.10)
\]

\[
\sigma_{NP}^{[0]}(\Omega)^2 = \frac{1}{E - 1} \sum_{m=1}^{E} \left| \frac{Y_m(\Omega)}{U_m(\Omega)} - \hat{S}_{NP}^{[0]}(\Omega) \right|^2 \quad (3.11)
\]

\[
\hat{\theta}^{[0]} = \arg \min_{\theta} \sum_{i=1}^{N} \frac{\left| \frac{\hat{S}_{NP}^{[0]}(\Omega_i)}{\sigma_{NP}^{[0]}(\Omega_i)} \right|^2}{\sigma_{NP}^{[0]}(\Omega_i)^2} \quad (3.12)
\]
3.4 Estimating Hammerstein systems using multiple experiments

\[ \hat{W}^0_m(\Omega_l) = U_m(\Omega); m = 1, \ldots, E \]
\[ W_{m,a}(\Omega_l) = \mathcal{F}(w_{m,a}(t)) = \mathcal{F}(f_a(u_m(t))); m = 1, \ldots, E \]

\[ \hat{\phi}^k = \arg \min_{\phi} \sum_{m=1}^E \sum_{l=1}^N \left| \sum_{a=1}^q \phi_a W_{m,a}(\Omega_l) \hat{S}^k(\Omega_l) - Y_m(\Omega_l) \right|^2 \] (3.13)

\[ \hat{W}^k_m(\Omega_l) = \sum_{a=1}^q \hat{\phi}^k_a W_{m,a}(\Omega_l); m = 1, \ldots, E \] (3.14)

\[ \hat{S}^k_{NP}(\Omega_l) = \frac{1}{E} \sum_{m=1}^E \frac{Y_m(\Omega)}{W_m^{k-1}(\Omega)}; k \geq 1 \] (3.15)

\[ \sigma^k_{NP}(\Omega)^2 = \frac{1}{E - 1} \sum_{m=1}^E \left| \frac{Y_m(\Omega)}{W_m^{k-1}(\Omega)} - \hat{S}^k_{NP}(\Omega) \right|^2 \] (3.16)

\[ \hat{\theta}^k = \arg \min_{\theta} \sum_{l=1}^N \left| \frac{\hat{S}^k_{NP}(\Omega_l)}{\sigma^k_{NP}(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2 \] (3.17)

If the estimate \( \hat{\phi}^k \) converges to the true values (it will be shown in Appendix 3.B under which assumptions this happens), \( \sigma^k_{NP} \) will tend to 0. To avoid the problem, the weighting in (3.17) is scaled: this only modifies the value of the cost function, but not the value of its minimizer.

Algorithm 3. Iterative procedure to produce initial estimates of Hammerstein systems

Step 3.1 Compute \( G_R \) and then compute an estimate of the static nonlinearity:
\[ \hat{S}^0 = G_R; \hat{f}^0(u) = f(\hat{\phi}^0). \]
\( G_R \) is obtained from a parametric model fitted to (3.10), weighted with (3.11): (3.12). \( \hat{f}^0 \) is computed with (3.13).

Step 3.2 \( k \leftarrow 1 \).

Step 3.3 For each input signal, estimate the values of the signal \( w(t) \) using the most recent estimate of \( f(u) \): (3.14).

Step 3.4 Build a parametric model \( \hat{S}^k \) of the linear system that approaches \( y_1(t), \ldots, y_E(t) \) best with \( \hat{w}_1^k(t), \ldots, \hat{w}_E^k(t) \) as input. Average out the stochastic nonlinearities introduced by the errors in the estimate of the static nonlinearity as in (3.15). This parametric model is obtained by minimizing a cost function that includes the uncertainty created by the stochastic nonlinearity in its weighting (3.16) as in (3.17).

Step 3.5 Use \( \hat{S}^k \) to obtain a parametric model of the static nonlinearity \( \hat{f}^k \): (3.13).

Step 3.6 If the convergence criterion is not met and if \( \|\sigma^k_{NP}\| > \epsilon_{user}, k \leftarrow k + 1 \)
and jump to Step 3.3.

$\epsilon_{\text{user}}$ is a user-settable threshold to avoid a division by 0 in (3.17).

### 3.4.2 Demonstration on measurements

The measurements used in Section 3.3.2 were repeated with 4 different input signals (all with the same power spectrum, but with another realization of the random phase). This section compares the results obtained with Algorithm 3 applied to these measurements with the results from Section 3.3.2. The same parameterization from Section 3.3.2.2 has been used in this example.

The result of Step 3.1 is compared to the result of Step 2.1 in Figure 3.12: the first estimate of the linear part is slightly better. Figure 3.13 compares the best estimate produced by both algorithms and shows that the algorithms converge to similar results. However, Figure 3.14 shows that Algorithm 3 converges faster than Algorithm 2. However, this comes at a price: the measurement time has been multiplied by 4.

### 3.5 Parameterization issues

Now that the first method has been defined, it makes sense to give some details about the parameterizations used in these algorithms. This is a section in its own right because
3.5 Parameterization issues

Figure 3.13: Model error $S - \hat{S}$ for Algorithm 3 (black dashed line) compared with $S - \hat{S}$ for Algorithm 2 (grey dashed line). The true value of $S$ (full line) is included as a reference.

Figure 3.14: The normalized output simulation error of Algorithm 3 (black circles) and Algorithm 2 (grey circles).
the material is applicable to the identification methods in the other sections.

3.5.1 Choosing a parameterization for the static nonlinearity

The freedom of choice of basis function for the static nonlinearity entails a responsibility: the basis functions must be chosen and the user has to make sure that all the necessary data is collected for the parameterization to be identifiable.

The static nonlinearity was modeled with a piecewise linear function with 25 breaking points. These breaking points are chosen automatically for equal support (Hagenblad, 1999). This is done to identify a model for the static nonlinearity that takes into account the amplitude distribution of the input signal: more parameters will be used for the input range where many points are sampled. This results in a better estimate as can be seen in Figure 3.15: the estimate with the uniformly distributed breaking points oscillates for the input region where only few data points are present. Note that the breaking points are fixed a priori. Optimizing the choice of these breaking points would make the cost function more involved (Roll, 2003).

The advantage of a piecewise linear approximation is that strong nonlinear characteristics can be modeled easily (if enough breaking points are chosen). The disadvantage is that the final model may have a wild behavior: Figure 3.5 minimizes the cost function, but would yield poor results for validation measurements.
Another set of basis functions are the polynomial functions (or orthogonal polynomials) $u^a$. These have a smooth appearance, but may display large oscillations for strong saturations if the order is chosen too high (see Figure 3.16). This can be avoided if Hermite polynomials are chosen as basis functions (Råde and Westergren, 1990). Since a random phase multisine is approximately normally distributed, Hermite polynomials will form a nearly orthogonal basis. The basis functions now are (for an input having zero mean):

\[
\begin{align*}
H_0(x) &= 1 \\
H_1(x) &= x \\
H_{n+1}(x) &= 2xH_n(x) - 2nH_{n-1}(x); \quad n \geq 1 \\
f_a &= H_a\left(\frac{x}{\sqrt{2 \sum x^2}}\right); \quad a = 1, 2, \ldots, q
\end{align*}
\]

With Hermite’s polynomials as basis functions, a polynomial of order 17 can be used to represent $f$ with good results: Figure 3.17 and Figure 3.18 show that a good fit can be obtained with much lower oscillations of the polynomial approximation.

In this chapter, piecewise linear approximations of the static nonlinearity were mostly used to allow saturation: their advantage is that the choice of breaking points can be automated easily and that the “locality” of the basis functions with respect to the input values yields a better numerical conditioning of the problem than straightforward polynomials. It would have been possible also to use orthogonal polynomials, but these have to be chosen in function of the amplitude distribution of the signals used for the

Figure 3.16: Polynomial approximation of $f$ (grey dots) of order 9 (full line) and of order 17 (dashed line).
identification. Using a piecewise linear approximation is thus preferred because of its simplicity and numerical robustness.

### 3.5.1.1 Maintaining identifiability

The set of frequencies $\Omega_f$ used for the identification of $f$ (see Step 2.5 and (3.9)) can be different from the set of frequencies $\Omega_S$ used for the identification of the linear part (see Step 2.4 and (3.7)). Obviously, the set $\Omega_S$ used to identify the linear block $S$ in (3.7) may only contain $\Omega : U(\Omega) \neq 0$, as otherwise a division by zero would occur. If, for instance, an odd multisine (see Section 2.2) is used as input, there are $\Omega$ values that are not used in (3.7), but which might be used in (3.9). It is important to include all $\Omega$ values that are necessary to make the minimization of the cost function in (3.9) uniquely solvable.

Imagine an experiment where the following Hammerstein system is excited with a random phase multisine with power at 1Hz, 3Hz and 5Hz:

$$\begin{cases}
f(u) &= u + u^2 + u^3 \\
S(j\omega) &= \frac{1}{1 + j\omega}
\end{cases} \quad (3.18)$$

The periodic output will contain components at 0Hz, 1Hz, 2Hz, 3Hz, ..., 14Hz and 15Hz. The linear system will be identifiable at the frequencies 1Hz, 3Hz and 5Hz in the
initial step. Further suppose that the three basis functions used to identify the static nonlinearity are \( u, u^2 \) and \( u^3 \). The basic function \( u^2 \) produces harmonics only on the even frequencies (0Hz, 2Hz, 4Hz, 6Hz, 8Hz and 10Hz). If the same frequencies (1Hz, 3Hz, 5Hz) are used in both cost functions minimized to identify the linear system and the static nonlinear system, the value of \( \varphi_2 \) —associated with \( u^2 \)— has no influence on the value of the cost function, and hence it is undefined. The solution to this problem is to use the complete set of frequencies inside the excitation frequency band: 1Hz, 2Hz, 3Hz, 4Hz and 5Hz. The other frequencies are rejected because they would imply an extrapolation of the linear model in the first step. From the second step on, \( \hat{S} \) could be estimated for the higher frequencies (6Hz, 7Hz, ... 15Hz) also, as an estimate of \( \hat{W}^{[k-1]} \) can be computed in Step 2.3 at these higher frequencies, too. This idea has not been pursued because the excitation level at these frequencies cannot be controlled practically by the experiment and a lower SNR will be noted at those lines where only stochastic contributions will appear.

Another example is the approximation of the same Hammerstein system from (3.18) with a piecewise linear function. Since the DC component of the output is not taken into account, an arbitrary constant may be added to all parameters without changing the value of the cost function (3.9).

### 3.5.2 Choosing a parameterization for the linear part

The parameterization of the linear part can be done in the same way as it would be done for linear modeling. Hence, the order of the model for \( S = G_R \) can be estimated with traditional linear methods. This has already been studied, see for example Pintelon and Schoukens (2001).

### 3.6 Simple identification of Wiener systems

Because of the similarity between a Wiener system (Figure 2.4) and a Hammerstein system, a similar procedure to Algorithm 1 is proposed: estimate the RLDS which gives the linear part \( R \), then use this estimate of the linear part to extract the static nonlinearity.
3.6.1 About the estimate of the linear part

The estimate of the linear part may be computed in exactly the same way as for Hammerstein systems in Section 3.2.2. The RLDS of a Wiener system is the same as its linear system \( R \) multiplied by an unknown scaling constant, because a Wiener system is a Wiener-Hammerstein system with trivial output linear part \( S = 1 \), so that (2.6) becomes

\[
G_R(\Omega) = C(U(\Omega)R(\Omega), f)R(\Omega) + O(N^{-1})
\]

Because of the gain degeneration (an equivalent figure to Figure 3.2 could be drawn), the unknown constant in (3.19) is not important, and the estimate of the linear part can be normalized to a gain of 1 at a user chosen frequency. Therefore, the estimate of \( R(\Omega) \) is given by

\[
\hat{R}_{\text{NP}}(\Omega) = \frac{Y(\Omega)}{U(\Omega)}
\]

\[
\hat{\theta} = \arg \min_\theta \sum_{l=1}^{N} \left| \hat{R}_{\text{NP}}(\Omega_l) - R_\theta(\Omega_l; \theta) \right|^2
\]

3.6.2 About the estimate of the nonlinear part

For a Wiener system, estimating the static nonlinearity is simpler than for Hammerstein systems: the excited frequencies of the unknown input of the static nonlinearity \( V \) are the same as the excitation components at the input \( U \) of the nonlinear system. This stems from the fact that \( R \) is a linear system, so no energy is transferred to unexcited lines. Consequently, the time values \( v \) can be reconstructed without extrapolating the estimate of the linear system:

\[
v(t) = \mathcal{F}^{-1}(V(\Omega)) = \mathcal{F}^{-1}(R(\Omega)U(\Omega))
\]

Once \( v \) has been computed, a graph of the pairs \((v, y)\) can be drawn. This allows the user to choose an appropriate method to identify the static nonlinearity. As this chapter is about simple approaches to initialize the identification process, the function will be
approximated using basis functions:

\[ f_\varphi = \sum_{a=1}^{q} \varphi_a f_a \quad (3.21) \]

The model of the static nonlinearity is linear in its parameters. These parameters \( \varphi \) will be obtained by minimizing a cost function that fits the function to the pairs \((v, y)\) in a least squares sense:

\[ \hat{Y}_a(\Omega_l) = \mathcal{F}(f_a(\hat{v}(t))); a = 1, 2, \ldots, q \quad (3.22) \]

\[ \hat{\varphi} = \arg \min_{\varphi} \sum_l \left| Y(\Omega_l) - \sum_{a=1}^{q} \varphi_a \hat{Y}_a(\Omega_l) \right|^2 \quad (3.23) \]

The cost function (3.23) sums over all measured samples. Minimizing this cost function is trivial, as it is quadratic in its parameters.

**Algorithm 4.** Simple identification of a Wiener system.

**Step 4.1** Estimate a parametric model of the RLDS, \( \hat{R} = G_R \) (3.20).

**Step 4.2** Compute \( \hat{v} \) (3.21), then \( \hat{Y}_a \) (3.22).

**Step 4.3** Estimate \( \hat{\varphi} \) (3.23).

As this algorithm is very similar in ideas to Algorithm 1 and will be used as a starting step for an iterative estimation algorithm in Section 3.7, a demonstration on measurements is not included here.

**3.7 Iterative identification of Wiener systems**

Again, the solution obtained with Algorithm 4 is not perfect and an iterative improvement is proposed: supposing that the static nonlinearity can be inverted, the output signal of the linear part can be estimated, allowing to compute a new estimate of the linear part which should be less disturbed by the stochastic contributions. This improved estimate of the linear part will result in a better estimate of the static nonlinearity. The convergence properties of this scheme are analyzed in Appendix 3.C.
3.7.1 The algorithm

The iterative algorithm is started with Algorithm 4, followed by alternatively estimating \( v(t) \) using the most recent estimate of the static nonlinearity or the most recent estimate of the linear part. Estimating \( v(t) \) with the estimate of \( f \) involves an inversion of the nonlinear characteristic. Hence an assumption needs to be made:

**Assumption 1.1** The static nonlinearity \( f(v) \) is invertible for \( v \) in the range \( R_v \) that was excited during the experiment.

3.7.1.1 Estimating the linear part

The start of the algorithm comes from Algorithm 4. The RLDS of the nonlinear system is the first estimate of \( R \):

\[
\hat{R}^{[0]}_{NP}(\Omega) = \frac{Y(\Omega)}{U(\Omega)}
\]

\[
\hat{\theta}^{[0]} = \arg \min_{\theta} \sum_{l=1}^{N} \left| \hat{R}^{[0]}_{NP}(\Omega_l) - R_\theta(\Omega_l; \theta) \right|^2 \tag{3.24}
\]

\[
\hat{R}^{[0]}(\Omega) = R_\theta(\Omega; \hat{\theta}^{[0]})
\]

Once an estimate of \( f \) has been computed, an attempt is made to reduce the stochastic nonlinear contributions by estimating \( \hat{V}^{[k;a]}(\Omega) \). \( \hat{V}^{[k;a]} \) is defined by (3.28) in Section 3.7.1.2.

\[
\hat{R}^{[k]}_{NP}(\Omega) = \frac{\hat{V}^{[k;a]}(\Omega)}{U(\Omega)}
\]

\[
\hat{\theta}^{[k]} = \arg \min_{\theta} \sum_{l=1}^{N} \left| \hat{R}^{[k]}_{NP}(\Omega_l) - R_\theta(\Omega_l; \theta) \right|^2 \tag{3.25}
\]

\[
\hat{R}^{[k]}(\Omega) = R_\theta(\Omega; \hat{\theta}^{[k]})
\]

3.7.1.2 Estimating the nonlinear part

Using the estimate \( \hat{R}^{[k]}(\Omega) \) of the linear part, the unknown signal \( \hat{V}^{[k;b]} \) can be estimated:

\[
\hat{V}^{[k;b]}(\Omega) = \hat{R}^{[k]}(\Omega)U(\Omega) \tag{3.26}
\]

Up to here, nothing new has been added with respect to Section 3.6.2. However, it is
3.7 Iterative identification of Wiener systems

more convenient to estimate the inverse of $f$ because at each iteration step, the unknown signal will be estimated using the output $y(t)$ and $f^{[k]}$ to yield $\hat{V}^{[k:a]}$ for Section 3.7.1.1. Hence, the inverse of $f$ is parameterized as a linear combination of basis functions and the cost function (3.23) is redefined as

$$V_{Y,a}(\Omega) = \mathcal{F}(f_a(y(t)))$$
$$\hat{\varphi}^{[k]} = \arg\min_{\varphi} \sum_{l=1}^{N} \left| \sum_{a=1}^{q} \varphi_a V_{Y,a}(\Omega_l) - \hat{V}^{[k:a]}(\Omega_l) \right|^2$$ (3.27)

Note that (3.27) is still quadratic in the parameters $\varphi$, so that minimizing the cost function is trivial. Also, the cost function takes into account only the frequency band where the model of the linear part is known (not extrapolated).

Now that the inverse of $f$ has been estimated, it is easy to estimate the input $\hat{V}^{[k:a]}$ of $f$ to be used for the linear estimation step in the next iteration (3.25):

$$\hat{V}^{[k:a]}(\Omega) = \mathcal{F} \left( \sum_{a=1}^{q} \varphi_a^{[k-1]} f_a(y(t)) \right) ; k \geq 1 \quad (3.28)$$

The algorithm can now be defined, subject to Assumption 1.1:

**Algorithm 5.** Iterative estimation of a Wiener system:

**Step 5.1** Compute the estimate of the RLDS as first estimate of $R$ (3.24).

**Step 5.2** $k \leftarrow 0$.

**Step 5.3** Estimate the unknown signal $\hat{V}^{[k:b]}$ (3.26).

**Step 5.4** Estimate the inverse of the static nonlinearity (3.27).

**Step 5.5** $k \leftarrow k + 1$.

**Step 5.6** If the stop criterion is met, stop the iteration and jump to Step 5.9.

**Step 5.7** Estimate $\hat{V}^{[k:a]}$ (3.28).

**Step 5.8** Estimate $\hat{R}^{[k]}$ (3.25). Normalize the gain of $\hat{R}^{[k]}$ to 1 at a user chosen frequency, then continue with Step 5.3.

**Step 5.9** Invert the estimate of the inverse of $f$.

The stop criterion mentioned in Step 5.6 can be freely chosen. The simplest stop criterion is to limit the number of iterations. It is impractical to compute the output simulation error and stop once this value decreases below a threshold or varies too slowly, because the inverse of $f$ is estimated instead of $f$. In other words, computing
the output simulation error entails inverting the estimate of \( \hat{f}^{[k]} \) at each iteration: even if \( f \) is invertible (Assumption 1.1), some \( \hat{f}^{[k]} \) might not be invertible. Also, since the parameters are not computed directly as minimizers of the output simulation error, there is no obvious relationship between reaching the minimum of (3.27) and minimizing the output simulation error.

### 3.7.2 Demonstration on a measurement example

The test system used in Vandersteen (1997) can also be used as a test object. The measurements are similar to those from Section 3.3.2. These will be used to illustrate the various steps of Algorithm 5 for the parameterization detailed in the next subsection.

#### 3.7.2.1 Parameterization

The parameters that were identified for this example are defined in this section. Section 3.5 contains a discussion about other possibilities that is also applicable here.

The linear part was identified using a discrete time model (\( \Omega = z^{-1} \)) with 3 zeros and 3 poles:

\[
R_\theta(z^{-1}) = \frac{\sum_{n=0}^{3} \theta_n z^{-n}}{1 + \sum_{n=1}^{3} \theta_{n+3} z^{-n}}
\]

The inverse of the nonlinear part was modeled with a piecewise linear function. This function had 25 breaking points, distributed over the output range \([\min(y(t)), \max(y(t))]\) such that each interval where the basis functions (triangles) are different from 0 covers the same number of samples. The estimated parameters were the values of \( v \) which \( f \) mapped to the breaking points value. This choice of basis functions also makes the inversion at the end (Step 5.9) trivial, as the only thing that needs to be done is to exchange the role of the breaking points and the parameter values.

#### 3.7.2.2 Results

The nonparametric estimate \( G_{NP} \) of the FRF can be compared to the RLDS: Figure 3.19 shows that the wild behavior of \( G_{NP} \) cannot be explained by noise alone. The Step 5.1 is motivated by Figure 3.20 that shows how a parametric estimate of the RLDS reduces
Figure 3.19: Comparison of the nonparametric FRF $G_{NP}$ (dots) of a Wiener system with the RLDS $G_R$ (full line) and of the stochastic contributions $G_S = G_{NP} - G_R$ (pluses) with the standard deviation of $G_{NP}$ (grey dots).

the impact of the stochastic contributions $G_S$ on $\hat{G}_R$.

An illustration of Step 5.4 can be found in Figure 3.21 and Figure 3.22: the first figure shows that the static nonlinearity is estimated very well from the start, while the other shows that the stochastic contributions in $\hat{R}_{NP}^{[1]}$ are lower than $G_S$ in $G_{NP}$, due to the estimate of $\hat{V}^{[0;a]}(\Omega)$.

The result of Algorithm 5 is shown in Figure 3.23, Figure 3.24 and Figure 3.25. The two first of these figures compare the identified models with the real values of the Wiener system. Note that the model of the static nonlinearity was extrapolated to cover the complete range of the measured real values, as the model extends beyond the extreme breaking points in Figure 3.23. The last figure, Figure 3.25 shows that the cost function that is minimized in (3.27) and used in the stop criterion of Step 5.6 decreases at every iteration. The identification process was stopped once two consecutive cost function values did not differ by more than $0.0001\text{dB}$. As the inverse of $f$ is identified, the cost function isn’t the output simulation error, but the difference between the two estimates $\hat{V}^{[k;a]}$ (3.28) and $\hat{V}^{[k;b]}$ (3.26) that are formed at each iteration step.
3.8 Estimating Wiener systems using multiple experiments

Similar to the identification procedure of the Hammerstein systems, the knowledge of the variance of the stochastic contributions can be incorporated into the estimation algorithm. This information can be exploited to improve the efficiency of the estimation process. In Step 5.1 and Step 5.8, where a parametric model for the linear part is estimated, a frequency dependent weighting can be introduced. This section will analyze what changes in the identification method when this weighting is introduced.

3.8.1 Modification in the method

Many different random phase multisines with the same power spectrum can be generated by choosing another realization for the phases in (1.2). The use of different excitation signals not only allows to reduce the stochastic contributions (Schoukens et al., 1998a; Pintelon and Schoukens, 2001) but also permits to quantify their effect. To accommodate the modifications, some notations need to be changed, too. Suppose that $E$ different multisines were used in the experiment, then the following equations define the modified procedure which is explained in Algorithm 6.
3.8 Estimating Wiener systems using multiple experiments

Figure 3.21: The static nonlinearity $f$ (black line) compared with its model $\hat{f}^{[0]}$ (line, the circles are the breaking points of the piecewise linear model). The scale to the right of the plot is used to zoom in on the difference (grey line) between the true value and the model.

Figure 3.22: $\hat{R}^{[1]}_{NP}$ (black dots) is closer to the RLDS $G_R$ (line) than $\hat{R}^{[0]}_{NP}$ (grey dots): the stochastic contributions $G_R - \hat{R}^{[1]}_{NP}$ (black pluses) are lower than $G_R - \hat{R}^{[0]}_{NP}$ (grey pluses).

Figure 3.23: The best model (black line, circles: breaking points) obtained by Algorithm 5 for $f$ (black line). The scale to the right of the plot is used to zoom in on the difference (grey line) between the true value and the model.

Figure 3.24: The best model (line) obtained by Algorithm 5 for $R$ (grey dots). The model error $R - \hat{R}$ is shown as a grey line.
The iterative Wiener estimation algorithm can be restated as follows for the case were $E$ different excitation signals are used:
Algorithm 6. Iterative procedure to produce initial estimates of Wiener systems

Step 6.1 Compute $G_R$ and then compute an estimate of the static nonlinearity (see (3.31) and (3.32)): $\hat{R}^{[0]} = G_R; \hat{f}^{[0]}(y) = f_\varphi (y; \hat{\varphi}^{[0]}$).

Step 6.2 $k \leftarrow 1$.

Step 6.3 Estimate the unmeasurable internal signals $\hat{v}_1^{[k]}, \ldots, \hat{v}_E^{[k]}$: (3.33).

Step 6.4 Build a parametric model $\hat{R}^{[k]}$ of the linear system that approaches $\hat{V}_1^{[k]} \ldots, \hat{V}_E^{[k]}$ best with $u_1(t), \ldots, u_E(t)$ as input (3.36).

Step 6.5 Use $\hat{R}^{[k]}$ of Step 6.4 to estimate $\hat{v}_{1 \ldots E}^{[k;b]}(t)$ by filtering of $u_{1 \ldots E}(t)$: $\hat{v}_{1 \ldots E}^{[k;b]}(t) = \hat{R}^{[k]}[u_{1 \ldots E}(t)]$.

Step 6.6 Consider the couples $\left(\hat{v}_{1 \ldots E}^{[k;b]}(t), y_{1 \ldots E}(t)\right)$ and fit a parametric model through the obtained static nonlinearity (3.32).

Step 6.7 If the convergence criterion is not met and if $\left\|\sigma_{NP}^{[k]}\right\| > \epsilon_{user}$, $k \leftarrow k + 1$ and jump to Step 6.3.

$\epsilon_{user}$ is a user-settable threshold to avoid a division by 0 in (3.36).

3.8.2 Demonstration on measurements

The measurements described in Section 3.7.2 have been repeated six times. For each experiment, another random phase realization for the input signal $u(t)$ was chosen, but the rest of the measurement setup was not changed. The following figures compare the results from Algorithm 5 with Algorithm 6.

The estimate at the start of Algorithm 6 of $\hat{R}^{[0]}$ is better than the estimate of $\hat{R}^{[0]}$ of Algorithm 5: Figure 3.26 compares both estimation errors and shows the weight used in this estimation step for Algorithm 6. Notice that the weight is in essence flat: the reason why Algorithm 6 converges faster than Algorithm 5 stems from the averaging of the stochastic nonlinear contributions.

The final results of both algorithms are compared in Figure 3.27: they are equal. Even though Figure 3.28 shows that the result is obtained in less iterations, probably the increase in measurement time means overall nothing is gained.
Figure 3.26: Comparing the model error $R - \hat{R}^{(0)}$ for Algorithm 6 (black dashed line) and Algorithm 5 (grey dashed line) with $R$ (full line). The weighting used in Step 6.4 is shown as black dots.

Figure 3.27: The model error $R - \hat{R}$ for Algorithm 6 (black dashed line) compared with the model error of Algorithm 5 (grey dashed line) and the true value of $R$ (full line).
3.9 On the importance of the frequency domain representation

This text tends to emphasize frequency domain methods. The main reason for this is the importance of being able to eliminate the stochastic nonlinear contributions. In theory, Step 5.8 can be implemented with any identification method that is able to estimate a consistent parametric model of a linear system. However, as Figure 3.29 shows, it pays off to filter out the stochastic nonlinear contributions inside the excitation band. This is trivial if the data is represented in the frequency domain, which favors frequency domain identification methods.

A Wiener system without noise was simulated to produce the results of Figure 3.29. An odd excitation was used and the static nonlinearity contained an important even component (but was still invertible). Figure 3.29 shows the evolution of the cost function for two different implementations of Algorithm 5. The reason why the implementation using the frequency domain identification method converges faster (if the number of iterations is counted) is that no even stochastic nonlinear contributions are disturbing the estimate of the linear part, while they are disturbing this estimate for the time domain identification methods.

If the time is measured, however, the implementation of the time domain identification (the ident toolbox for Matlab) is faster than the frequency domain toolbox (fdident,
Figure 3.29: Evolution of the cost function for Algorithm 5 using time domain identification (black circles) and frequency domain identification methods (grey circles).

Table 3.1: Execution times for Algorithm 5 using time domain identification methods or frequency domain identification methods

<table>
<thead>
<tr>
<th></th>
<th>Time Domain</th>
<th>Frequency Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of iterations</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Total execution time (s)</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>Execution time of one iteration (s)</td>
<td>1.4</td>
<td>3.6</td>
</tr>
<tr>
<td>Time spent in Step 5.8 (s)</td>
<td>0.96</td>
<td>2.9</td>
</tr>
</tbody>
</table>

The best implementation would probably be to take advantage of the frequency domain for the ease with which filtering can be obtained and transform back the data to the time domain for the estimation step to profit from the speed of the implementation. For the results of Table 3.1, there was already some sort of filtering for the time domain implementation: if the spectrum above the highest exciting component is not removed, the implementation using the time domain identification converges to higher values of the cost function. Thus, filtering is already done in a limited way. Extending it to keep only the excited frequencies would probably not slow it down too much.

However, there exist methods to speed up ELiS, which is not optimized for speed. Schoukens et al. (1998b) shows that significant gains can be made for higher model orders.
3.10 Comparison with other methods

The identification of Hammerstein and Wiener systems has already been studied in many papers. In these methods, three different types are encountered: nonparametric (Greblicki, 1997, 2000; Lang, 1997), semi-parametric (Celka et al., 2001; Hagenblad, 1999; Pawlak et al., 2000) and parametric (Bai, 2002b; Bai and Fu, 2002; Crama and Schoukens, 2001; Giri et al., 2001; Verhaegen and Westwick, 1996; Vörös, 1997; Westwick and Verhaegen, 1996). The group of the semi-parametric methods differs from the group of parametric methods by their representation of the linear part of the Wiener system. Semi-parametric methods describe the linear part by its finite impulse response, which implies that many parameters are estimated, and hence that long data records are needed. None of the already existing methods for generating initial estimates, except Vörös (1997) for Hammerstein systems, iterate to improve their initial estimates.

Section 3.B.2 shows that if the linear part is estimated with a nonparametric scheme (the proof for Wiener systems is similar to that for Hammerstein systems), the iteration procedure converges (in one step) to a solution that is different from the true model.

In Verhaegen and Westwick (1996); Westwick and Verhaegen (1996) subspace methods are used. The advantage of the subspace methods is that they address the MIMO case. In Bruls et al. (1999), an iterative improvement of this method has been presented, based on the principle of separable least squares (Golub and Pereyra, 1973).

The main point of this chapter is the iteration process to generate enhanced estimates. This iterative process converges to the true system (for SISO systems) if the estimation step of the linear dynamic part is consistent (among other assumptions, the details are in Section 3.B and Section 3.C). Because the subspace method is a consistent estimator, the frequency domain identification method can also be replaced by the subspace method (or any other consistent estimator) in Algorithm 5 and Algorithm 2, still leading to the same estimates as the frequency domain approach. However, the advantage of the frequency domain approach is that in the case of disturbing noise, a nonparametric noise model can easily be used in the cost functions, even if only one excitation signal is measured, using a technique similar to Schoukens et al. (2003b). This enhances the efficiency.

When more than one excitation signal may be measured, then the estimate of the linear part can be weighted with the deviations introduced by the nonlinear behavior. This results in an efficient estimator. This weighting could also be introduced in subspace
methods (McKelvey et al., 1996; Pintelon, 2002).

The method presented in Bai (2002b) is simpler for Hammerstein systems when the static nonlinearity is expressed with few parameters (1 or 2) but has special requirements on the input amplitude distribution when more parameters are needed to describe the static nonlinearity. The method presented in Bai and Fu (2002) handles blind identification but is restricted to systems driven by a zero order hold excitation. This requirement on the excitation signal is essential.

The main addition of this chapter with respect to Vörös (1997) is the convergence analysis of the iteration scheme. The presented method has been defined completely and, based on theoretical results (3.1), it is shown how to start up the iterative procedure.

3.11 Conclusion

This chapter has shown how two basic nonlinear systems, namely Hammerstein and Wiener systems may be approximately identified with simple methods. The focus was on simplicity, rather than precision: the effects of noise are not taken into account in the different algorithms. For those algorithms that are iterative, an analysis of the convergence properties has been provided.
3.A Introduction to the convergence analysis

The convergence analysis for both algorithms (Hammerstein system estimation or Wiener system estimation) have many things in common. The analysis will be worked out completely for a Hammerstein system, while for the Wiener system only the necessary changes will be explained.

3.B Convergence analysis for the Hammerstein estimation algorithm

3.B.1 Introduction

3.B.1.1 The system

The studied system is a Hammerstein system (Figure 2.3), excited with a random phase multisine of constant power. The notations that are used in this analysis are summarized in Table 3.2. The convergence of a similar algorithm has already been studied in Stoica (1981): a Hammerstein system with even static nonlinearity ($u^2$) and finite impulse response length cannot be identified with an iterative algorithm with a relaxation approach where a single cost function is used to identify the parameters of the linear and the nonlinear part. This is given as an analytical counterexample to show that such an iterative algorithm doesn’t converge in all cases. However, by putting the presented method in the framework described in Schoukens et al. (1998a), the reason for the existence of such a counterexample becomes clear. This counterexample is explicitly prohibited by the assumption (common to both the Wiener identification and the Hammerstein identification algorithm) that their best linear approximation $G_R$ is different from 0. The method proposed in this paper minimizes two different cost functions: one for the linear and one for the nonlinear part in each iteration.

This analysis consists of two steps: first, it will be proven that the proposed iterative algorithm converges to a wrong solution when a nonparametric pointwise estimate of the FRF of the linear system $S$ is used. The second step shows that a consistent parametric estimate of $S$ yields an improved estimate of the parameters of the static nonlinearity, and shows the convergence of the iterative algorithm to the true system in the absence
Table 3.2: Notations used in the analysis of the convergence of the Hammerstein estimation algorithm

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>Number of excitation frequencies</td>
</tr>
<tr>
<td>$p$</td>
<td>The number of parameters $\hat{\theta}^{[k]}$ describing the linear system at iteration $k$</td>
</tr>
<tr>
<td>$q$</td>
<td>The number of parameters $\hat{\varphi}^{[k]}$ describing the static nonlinearity at iteration $k$</td>
</tr>
<tr>
<td>$U$</td>
<td>Components of the Fourier transform of the input (at the excitation frequencies)</td>
</tr>
<tr>
<td>$B(U)$</td>
<td>$2N \times q$ matrix containing the Fourier coefficients of the basis functions used in the estimation process applied to the input (one basis function in each column). The rows $1 \rightarrow N$ contain the positive frequencies, followed by the negative frequencies in the rows $N + 1 \rightarrow 2N$.</td>
</tr>
<tr>
<td>$W_0 = B(U) \varphi_0$</td>
<td>$2N \times 1$ vector containing the Fourier coefficients of the intermediate signal (at the excitation frequencies) $^a$</td>
</tr>
<tr>
<td>$\text{diag}(S_0)$</td>
<td>Diagonal $2N \times 2N$ matrix containing the transfer function of the linear system at the excitation frequencies</td>
</tr>
<tr>
<td>$Y = \text{diag}(S_0) W_0$</td>
<td>$2N \times 1$ vector of the output Fourier coefficients (at the excitation frequencies)</td>
</tr>
<tr>
<td>$\hat{W}^{[k]} = B(U) \hat{\varphi}^{[k]}$</td>
<td>Estimate (at iteration $k$) of the static nonlinearity, computed in the frequency domain</td>
</tr>
</tbody>
</table>

$^a$Since the same matrix of basis functions is used to define the system and to identify it, Assumption 2.1 is verified.
of disturbing noise and model errors.

**Assumption 2.1** The class of models used for the estimation contains the true model for the static nonlinearity as well as for the linear system.

**Assumption 2.2** The excitation \( u(t) \) is a random phase multisine with power \( P \) (independently of \( N \)), distributed over a sufficiently large number of frequencies to ensure a persistent excitation (even for \( N \to \infty \)).

An example of such an excitation is (3.37), though in the most general case, all Fourier coefficients of \( u(t) \) do not need to have the same amplitude.

\[
u(t) = \sum_{l=1}^{N} \sqrt{2P / N} \cos(l\omega_0 \cdot t + \phi_l)
\]  

(3.37)

**Assumption 2.3** The model of the static nonlinearity is linear in its parameters and each of the basis functions \( B_k \) applied to the random phase multisine input has as output a multisine with random phase and Fourier coefficients that are an \( O\left(\frac{1}{\sqrt{N}}\right) \). An example of such basis functions are \( u^n \).

**Assumption 2.4** The number of parameters \( p \) used to model the linear system and the number of parameters \( q \) used to model the static nonlinearity \( f(u) \) are kept constant when the number \( N \) of excitation lines in the spectrum increases. This means also that \( \frac{p}{N} \) is \( O(N^{-1}) \) and \( \frac{q}{N} \) is \( O(N^{-1}) \).

**Assumption 2.5** The algorithm starts by computing a consistent parametric estimate of the best linear approximation of the Hammerstein system. This estimate is proportional to the true linear system \( S_0 \) with an error term that is an \( O\left(\frac{1}{\sqrt{N}}\right) \). It is assumed that this error is “small enough” as specified in Section 3.B.3.2.

**Assumption 2.6** The best linear approximation \( G_R \) of the Hammerstein system is not identically 0, i.e. the static nonlinearity is not even in the input domain: \( f(u) \neq f(-u) \) for some \( u \).

**Assumption 2.7** The static nonlinearity of the true system is invertible in the input domain\(^1\).

**Assumption 2.8** The following matrix product is well-conditioned and invertible \( \forall k \),

\(^1\)This assumption is *not* necessary to apply Algorithms 2 and 3, but only to analyze their convergence.
\( \forall N (\infty \text{ included}): \)
\[
B(U)^H \text{diag} \left( \hat{S}^{[k]} \right)^H \text{diag} \left( \hat{S}^{[k]} \right) B(U)
\]

### 3.B.2 Convergence in the nonparametric case

Suppose that after iteration \( k \) the estimated parameters of the static nonlinearity are given by \( \hat{\varphi}^{[k]} \), then the nonparametric estimate of \( S_0 \) is given by (with \( B(U)_l; \) the \( l \)-th row of \( B(U) \))
\[
\hat{S}^{[k]}_{NP} (l) = \frac{Y (l)}{W^{[k]} (l)} \frac{Y (l)}{B(U)_l; \hat{\varphi}^{[k]}}
\]
\[\Leftrightarrow \text{diag} \left( \hat{S}^{[k]}_{NP} \right) = \text{diag}(Y) \text{diag}(B(U) \hat{\varphi}^{[k]})^{-1} \quad (3.38)\]

\( E^{[k+1]} \) is the output error of the model
\[
E^{[k+1]} = Y - \text{diag} \left( \hat{S}^{[k]}_{NP} \right) B(U) \hat{\varphi}^{[k+1]}
\]
\[
= Y - \text{diag}(Y) \text{diag}(B(U) \hat{\varphi}^{[k]})^{-1} B(U) \hat{\varphi}^{[k+1]}
\]
\[
= Y - \text{diag}(B(U) \hat{\varphi}^{[k]})^{-1} \text{diag}(B(U) \hat{\varphi}^{[k+1]}) Y
\]
\[
= \left[ \text{diag}(B(U) \hat{\varphi}^{[k]}) - \text{diag}(B(U) \hat{\varphi}^{[k+1]}) \right] \text{diag}(B(U) \hat{\varphi}^{[k]})^{-1} Y \quad (3.40)
\]
\[
= \text{diag}(B(U) \left( \hat{\varphi}^{[k]} - \hat{\varphi}^{[k+1]} \right)) \text{diag}(B(U) \hat{\varphi}^{[k]})^{-1} Y \quad (3.41)
\]

Throughout the computation, the fact that all matrices are diagonal is used, and to obtain (3.40), \( Y \) has been multiplied and divided pointwise by the same vector \( B(U) \hat{\varphi}^{[k]} \). This can be written as
\[
Y = \text{diag}(B(U) \hat{\varphi}^{[k]}) \text{diag}(B(U) \hat{\varphi}^{[k]})^{-1} Y.
\]
In order to estimate the parameters of the static nonlinearity for the next iteration \( \hat{\varphi}^{[k+1]} \), \( \| E^{[k+1]} \|^2 \) is minimized with respect to \( \hat{\varphi}^{[k+1]} \). In (3.41), the output error \( E^{[k+1]} \) is computed in function of \( \hat{\varphi}^{[k+1]} \) and \( \hat{\varphi}^{[k]} \). \( \hat{\varphi}^{[k+1]} = \hat{\varphi}^{[k]} \) minimizes \( \| E^{[k+1]} \|^2 \) because then \( E^{[k+1]} = 0 \) (see (3.41)). Of course, plugging \( \hat{\varphi}^{[k+1]} \) into (3.38) to estimate \( \hat{S}^{[k+1]}_{NP} \) yields the same value as \( \hat{S}^{[k]}_{NP} \).

Hence, the algorithm has converged, but to a value different from \( \varphi_0 \) if the initial estimate of the parameters of the static nonlinearity \( \hat{\varphi}^{[0]} \neq \varphi_0 \).
3.B.3 Convergence in the parametric case

This section analyzes the convergence of the estimation procedure when a parametric model is fitted to the nonparametric pointwise FRF $\hat{S}_{NP}^k$. A complete iteration will be studied to relate the estimation error of the linear system at iteration $k + 1$ to that at iteration $k$. The convergence follows then from (3.60).

3.B.3.1 Influence of the estimation error of the linear system on the estimate of the static nonlinearity

Suppose that a parametric model $\hat{S}^k$ has been fitted to $\hat{S}_{NP}^k$ starting from $\hat{W}^k$ and $Y$ and that this estimate is then used in (3.39) instead of $\hat{S}_{NP}^k$. This defines a new output error $F^{[k+1]}$ whose norm can be minimized with respect to $\hat{\varphi}^{[k+1]}$:

$$F^{[k+1]} = \text{diag}(S_0) \mathbf{B}(U) \varphi_0 - \text{diag}(\hat{S}^k) \mathbf{B}(U) \hat{\varphi}^{[k+1]}$$  

(3.42)

Each estimated quantity in (3.42) can be expressed as the sum of the correct parameters and an error term:

$$\text{diag}(\hat{S}^k) = \text{diag}(S_0) + \text{diag}(\hat{\delta}^k_S)$$  

(3.43)

$$\hat{\varphi}^{[k+1]} = \varphi_0 + \hat{\delta}^{[k+1]}_\varphi$$  

(3.44)

Substitute (3.43) and (3.44) into (3.42) to obtain:

$$F^{[k+1]} = \text{diag}(S_0) \mathbf{B}(U) \varphi_0 - \text{diag}(S_0) \mathbf{B}(U) \varphi_0 - \text{diag}(\hat{\delta}^k_S) \mathbf{B}(U) \varphi_0$$

$$- \text{diag}(\hat{S}^k) \mathbf{B}(U) \delta^{[k+1]}_\varphi$$  

(3.45)

The least squares solution $\hat{\delta}^{[k+1]}_\varphi$ minimizing $\|F^{[k+1]}\|^2$ (3.45) is given by (3.46).

$$\hat{\delta}^{[k+1]}_\varphi = - \left\{ \mathbf{B}(U)^H \text{diag}(\hat{S}^k)^H \text{diag}(\hat{S}^k) \mathbf{B}(U) \right\}^{-1}$$

$$\mathbf{B}(U)^H \text{diag}(\hat{S}^k)^H \text{diag}(\hat{\delta}^k_S) \mathbf{B}(U) \varphi_0$$  

(3.46)

The different parts of the product in (3.46) will be analyzed to estimate their order of magnitude for different values of the number $N$ of excited frequencies.
From Appendix 3.D and Assumption 2.8, it follows that

\[ B(U)^H \text{diag}(\hat{S}^{[k]})^H \text{diag}(\hat{S}^{[k]}) B(U) = \begin{bmatrix} O(1) & \cdots & O(1) \\ \vdots & \ddots & \vdots \\ O(1) & \cdots & O(1) \end{bmatrix}_{q \times q} = O(1)_{q \times q} \]  

(3.47)

is a well-conditioned invertible matrix with \( O(1) \) elements of dimensions \( q \times q \).

The second part of (3.46) can be rewritten using the diagonal structure of \( \text{diag}(\hat{S}^{[k]}) \):

\[
B(U)^H \text{diag}(\hat{S}^{[k]})^H \text{diag}(\hat{\delta}^{[k]}_S) B(U) \varphi_0 \]
\[
= \underbrace{B(U)^H \text{diag}(\hat{S}^{[k]})^H \text{diag}(\hat{S}^{[k]}) B(U) \varphi_0}_{O\left(\frac{1}{\sqrt{N}}\right)_{q \times 2N}} \underbrace{\text{diag} O\left(\frac{1}{\sqrt{N}}\right)_{2N \times 1}}_{O\left(\frac{1}{N}\right)_{q \times 2N}} 
\]

(3.48)

In (3.48), the matrix multiplication in front of \( \hat{\delta}^{[k]}_S \) results in a \( q \times 2N \) matrix with elements of \( O\left(\frac{1}{N}\right) \) (because of Assumption 2.3 and since the diagonality of two of the three terms of the matrix product guarantees that no terms of \( O\left(\frac{1}{N}\right) \) will be added together). Taking (3.47) and (3.48) together yields

\[
\hat{\delta}^{[k+1]}_\varphi = O(1)_{q \times q} O\left(\frac{1}{N}\right)_{q \times 2N} \hat{\delta}^{[k]}_S = O\left(\frac{1}{N}\right)_{q \times 2N} \hat{\delta}^{[k]}_S
\]

(3.49)

In (3.49), Assumption 2.4 is used to estimate the order of magnitude of the matrix product: \( q \ll N \) means that only a fixed number of terms of \( O\left(\frac{1}{N}\right) \) will be added together, yielding an \( O\left(\frac{1}{N}\right) \).

### 3.B.3.2 Influence of the estimation error of the static nonlinearity on the estimate of the linear system

The error of the estimate of the static nonlinearity \( \hat{\delta}^{[k+1]}_\varphi \) propagates to the estimate of the linear system at iteration \( k + 1 \), because the estimated intermediate signal \( \hat{W}^{[k+1]} \neq W_0 \).

\( \hat{W}^{[k+1]} \) is the simulated output of the identified static nonlinearity:

\[
\hat{W}^{[k+1]} = B(U) \hat{\varphi}^{[k+1]} = B(U) (\varphi_0 + \hat{\delta}^{[k+1]}_\varphi)
\]

(3.50)
3.B Convergence analysis for the Hammerstein estimation algorithm

\[ w_0 \xrightarrow{f^{-1}_0} u \xrightarrow{\hat{f}^{[k+1]}} \hat{\omega}^{[k+1]} \]

Figure 3.30: Static nonlinear relationship \( h^{[k+1]} \) between \( w_0 \) and \( \hat{\omega}^{[k+1]} \).

\[
\hat{\omega}^{[k+1]} = W_0 + O \left( \frac{1}{\sqrt{N}} \right)_{2N \times q} O \left( \frac{1}{N} \right)_{q \times 2N} \hat{\delta}^{[k]}_S
\]

\[
= W_0 + O \left( \frac{1}{N^{\frac{3}{2}}} \right)_{2N \times 2N} \hat{\delta}^{[k]}_S \tag{3.51}
\]

The estimated intermediate signal \( \hat{W}^{[k+1]} \) can also be written as a static nonlinearity \( h^{[k+1]}(w_0(t)) \) applied to \( W_0 \), as in Figure 3.30 because there is a static invertible nonlinear relationship between \( U \) and \( W_0 \) (see Assumption 2.7) and a static nonlinear relationship between \( U \) and \( \hat{W}^{[k+1]} \). Hence, \( \hat{W}^{[k+1]} \) can be split into two contributions as in (3.53): a contribution \( \hat{W}^{[k+1]}_{\Delta R} \) proportional to \( W_0 \) coming from the Related Linear Dynamic System of the static nonlinearity \( h^{[k+1]}(w_0(t)) \) defined in (3.52) and a stochastic contribution \( \hat{W}^{[k+1]}_{\Delta S} \) that isn’t correlated with \( \hat{W}^{[k+1]}_{\Delta R} \) when the Fourier coefficients of both signals at the same frequency are considered. In (3.53), \( \hat{\alpha}^{[k+1]} \) is independent of the frequency because the RLDS of a static nonlinearity for a random multisine excitation is a static system.

\[
\hat{W}^{[k+1]}(l) = \mathcal{F} \left( h^{[k+1]}(w_0(t)) \right)
\]

\[
= \hat{W}^{[k+1]}_{\Delta R}(l) + \hat{W}^{[k+1]}_{\Delta S}(l)
\]

\[
= \hat{\alpha}^{[k+1]}W_0(l) + \hat{W}^{[k+1]}_{\Delta S}(l) \tag{3.53}
\]

with \( l \) including positive and negative frequencies.

When estimating a Hammerstein system, there always remains a gain indetermination. In the proposed estimation method, the algorithm scales the gain of the linear system to a user-defined value at a chosen frequency after the parametric estimation of the linear system. Of course, the estimate of the static nonlinearity is adapted to preserve the overall gain of the system. This is equivalent to scaling the internal variable \( \hat{W}^{[k+1]} \) before estimating the linear system’s transfer function. Hence, after scaling (3.53), an important property of the error term in (3.51) can be established: this error term
\[
O\left(\frac{1}{N^\frac{3}{2}}\right)_{2N \times 2N} \hat{\delta}_S^{[k]} = O\left(\frac{1}{\alpha^{k+1}}\right)_{2N \times 2N} \hat{W}_{\Delta S}^{[k+1]} \] is not correlated frequency per frequency with \(W_0\). This leads to
\[
\hat{W}_{\Delta S}^{[k+1]} = O\left(\frac{1}{N^\frac{3}{2}}\right)_{2N \times 2N} \hat{\delta}_S^{[k]} \tag{3.54}
\]

\[
K = \sum_{l=-N}^{N} \left| Y(l) - \hat{S}_{[k+1]}(l) \left(W_0(l) + \hat{W}_{\Delta S}^{[k+1]}(l)\right)\right|^2 \tag{3.55}
\]

When the parametric model \(\hat{S}_{[k+1]}\) is estimated, the cost function of (3.55) is minimized with respect to the parameters \(\hat{\theta}^{[k+1]}\). At the minimum of the cost function, the derivative of \(K\) with respect to \(\hat{\theta}^{[k+1]}\) is 0:
\[
\frac{\partial S}{\partial \hat{\theta}^{[k+1]}} H \text{diag} \left(W_0 + \hat{W}_{\Delta S}^{[k+1]}\right) H \left(\text{diag} \left(W_0 + \hat{W}_{\Delta S}^{[k+1]}\right) \hat{S}_{[k+1]} - Y\right) = 0 \tag{3.56}
\]

Because the cost function (3.55) uses both the positive and negative frequencies, there is no need for a \(\text{Re}(\cdot)\) operator when the derivative (3.56) is computed.

\[
S_0 = S(\Omega; \theta_0) = S(\Omega; \hat{\theta}^{[k+1]} - \hat{\delta}^{[k+1]}) = \hat{S}_{[k+1]} - \frac{\partial S}{\partial \hat{\theta}^{[k+1]}} \hat{\delta}^{[k+1]} + O \left(\hat{\delta}^{[k+1]}\right)^2
\]

\[
= \hat{S}_{[k+1]} - \hat{\delta}_S^{[k+1]} \tag{3.57}
\]

It is assumed that \(\hat{S}_{[k+1]}\) is close enough to \(S_0\) (see Assumption 2.5) so that the linear approximation in (3.57) is valid. This is justified because the consistent estimate of \(\hat{S}^{[0]}\) yields a result \(\hat{S}^{[0]} = S_0 + O \left(\frac{1}{\sqrt{N}}\right)\). Hence, if the frequency content of the excitation signal is rich enough \((N \gg 1)\), the iteration process will start close enough to the true solution and the second order effects can be neglected.

Substitution of (3.57) into (3.56) together with \(Y = \text{diag}(W_0) S_0\) yields
\[
0 = \frac{\partial S}{\partial \hat{\theta}^{[k+1]}} H \text{diag} \left(W_0 + \hat{W}_{\Delta S}^{[k+1]}\right) H \left(\text{diag} \left(W_0 + \hat{W}_{\Delta S}^{[k+1]}\right) \left(S_0 + \hat{\delta}_S^{[k+1]}\right) - Y\right)
\]

\[
\Leftrightarrow 0 = \frac{\partial S}{\partial \hat{\theta}^{[k+1]}} H \text{diag} \left(W_0 + \hat{W}_{\Delta S}^{[k+1]}\right) H \left[\text{diag}(W_0) S_0 + \text{diag} \left(\hat{W}_{\Delta S}^{[k+1]}\right) S_0 + \text{diag} \left(W_0 + \hat{W}_{\Delta S}^{[k+1]}\right) \frac{\partial S}{\partial \hat{\theta}^{[k+1]}} \hat{\delta}_S^{[k+1]} - Y\right]
\]
\[
\hat{\delta}_{\theta}^{[k+1]} = -\left[ \frac{\partial S}{\partial \hat{\theta}^{[k+1]}} H \text{diag}(W_0 + \hat{W}_s^{[k+1]}) H \text{diag}(W_0 + \hat{W}_s^{[k+1]}) \frac{\partial S}{\partial \hat{\theta}^{[k+1]}} \right]^{-1}
\]

\[
\quad \times \frac{\partial S}{\partial \hat{\theta}^{[k+1]}} H \text{diag}(W_0 + \hat{W}_s^{[k+1]}) H \text{diag}(\hat{W}_s^{[k+1]}) S_0 \quad (3.58)
\]

Now, (3.54) can be plugged into (3.58) to see the influence of increasing \(N\) values on the parameter estimation error \(\hat{\delta}_{\theta}^{[k+1]}\):

\[
\hat{\delta}_{\theta}^{[k+1]} = \left[ O(1)_{p \times 2N} \text{diag} \left( O \left( \frac{1}{N} \right) \right) O(1)_{2N \times p} \right]^{-1}
\]

\[
\quad \times O(1)_{p \times 2N} \text{diag}(S_0) \text{diag}(W_0 + \hat{W}_s^{[k+1]}) H \hat{W}_s^{[k+1]}
\]

\[
= O(1)_{p \times p} O(1)_{p \times 2N} \text{diag}(W_0 + \hat{W}_s^{[k+1]}) H \hat{W}_s^{[k+1]}
\]

\[
= O(1)_{p \times 2N} \left[ \text{diag}(W_0)^H \hat{W}_s^{[k+1]} + \text{diag}(\hat{W}_s^{[k+1]})^H \hat{W}_s^{[k+1]} \right] \quad (3.59)
\]

In (3.59), there is a product of uncorrelated Fourier coefficients \(2\) and \(3\) (3.53), hence the matrix multiplication with the \(O(1)_{p \times 2N}\) grows as an \(O \left( \sqrt{N} \right)\) instead of \(O(N)\). Hence a factor \(O \left( \frac{1}{\sqrt{N}} \right)\) should be added \(4\):

\[
\hat{\delta}_{\theta}^{[k+1]} = O \left( \frac{1}{\sqrt{N}} \right) O(1)_{p \times 2N} \text{diag} \left( O \left( \frac{1}{\sqrt{N}} \right) \right) O \left( \frac{1}{N^2} \right)_{2N \times 2N} \hat{\delta}_{\theta}^{[k]}
\]

\[
+ O(1)_{p \times 2N} \text{diag} \left( O \left( \frac{1}{N^2} \right)_{2N \times 2N} \hat{\delta}_{\theta}^{[k]} \right) H \left( O \left( \frac{1}{N^2} \right)_{2N \times 2N} \hat{\delta}_{\theta}^{[k]} \right)
\]

\[
= O \left( \frac{1}{\sqrt{N}} \right) O(1)_{p \times 2N} O \left( \frac{1}{N^2} \right)_{2N \times 2N} \frac{\partial S}{\partial \hat{\theta}^{[k]}} \hat{\delta}_{\theta}^{[k]}
\]

\[
+ O(1)_{p \times 2N} \text{diag} \left( O \left( \frac{1}{N^2} \right)_{2N \times 2N} \frac{\partial S}{\partial \hat{\theta}^{[k]}} \hat{\delta}_{\theta}^{[k]} \right) H \left( O \left( \frac{1}{N^2} \right)_{2N \times 2N} \frac{\partial S}{\partial \hat{\theta}^{[k]}} \hat{\delta}_{\theta}^{[k]} \right)
\]

\[
\left\| \hat{\delta}_{\theta}^{[k+1]} \right\| = \left\| O \left( \frac{1}{\sqrt{N}} \right) O(1)_{p \times 2N} O \left( \frac{1}{N} \right)_{2N \times p} \hat{\delta}_{\theta}^{[k]} + O(1)_{p \times 2N} O(1)_{2N \times p} O \left( \hat{\delta}_{\theta}^{[k]} \right)^2 \right\|
\]

\[
\leq \left\| O \left( \frac{1}{\sqrt{N}} \right) \hat{\delta}_{\theta}^{[k]} \right\| + \left\| O \left( \hat{\delta}_{\theta}^{[k]} \right)^2 \right\| \quad (3.60)
\]

When the algorithm converges, it follows from (3.60) how this happens: by construc-
tion, $\hat{\delta}_b^{[0]} = O\left(\frac{1}{\sqrt{N}}\right)$ and starting from iteration 0,

$$\left\|\hat{\delta}_b^{[0]}\right\| = O\left(\frac{1}{\sqrt{N}}\right)$$

(3.61)

$$\Rightarrow \left\|\hat{\delta}_b^{[1]}\right\| \leq O\left(\frac{1}{\sqrt{N}}\right)\left\|\hat{\delta}_b^{[0]}\right\| + \left\|O\left(\hat{\delta}_b^{[0]}\right)^2\right\|$$

$$= O\left(\frac{1}{N}\right) + O\left(\frac{1}{N}\right) = O\left(\frac{1}{N}\right)$$

(3.62)

$$\Rightarrow \left\|\hat{\delta}_b^{[2]}\right\| \leq O\left(\frac{1}{\sqrt{N}}\right)\left\|\hat{\delta}_b^{[1]}\right\| + \left\|O\left(\hat{\delta}_b^{[1]}\right)^2\right\| = O\left(\frac{1}{N^{3/2}}\right)$$

(3.63)

As can be seen by comparing (3.61), (3.62) and (3.63), at each iteration step, the error is multiplied with $O\left(\frac{1}{\sqrt{N}}\right)$. This is not a formal convergence proof since at each iteration $k$, the $O\left(\frac{1}{\sqrt{N}}\right)$ may be different, so that possibly $\lim_{k \to \infty} O\left(\frac{1}{\sqrt{N}}\right) = 1$, meaning that the error would become smaller at first (for $N$ large enough), but not disappear completely.

### 3.B.4 Convergence when multiple experiments are used

This section analyzes the convergence of Algorithm 3. There is no difference between Algorithm 2 and Algorithm 3 for the estimation procedure for the static nonlinear part, hence, Section 3.B.3.1 will not be repeated here.

#### 3.B.4.1 Influence of the estimation error of the static nonlinearity on the estimate of the linear system

The equations of Section 3.B.3.2 will be adapted to the introduction of the weighting with the nonlinear stochastic contributions. Note that when the algorithm converges, $\sigma_{NP}^{[k]}$ converges to 0 (because no stochastic nonlinearities remain once the static nonlinear part has been estimated). To avoid a division by 0, $\left\|\sigma_{NP}^{[k]}\right\| = 1$ is imposed. This scaling changes only the value of the cost function (3.17), but not the value of its minimizer. A consequence of this scaling is that the elements of $\sigma_{NP}^{[k]} = O\left(\frac{1}{\sqrt{N}}\right)$.

The convergence analysis uses the properties of the stochastic nonlinear contributions: Schoukens et al. (1998a); Pintelon and Schoukens (2001) show that they act like disturbances added at the output. These disturbances have zero mean and are uncorrelated frequency per frequency with the input signal. Hence, the estimate of the variations of the stochastic nonlinear contributions can be used as a weighting function in the same
3.B Convergence analysis for the Hammerstein estimation algorithm

Figure 3.31: The relation between $\hat{W}^{[k+1]}$ and $Y$ can be represented as a Hammerstein system.

Way as the variance of the output noise would be used (3.17).

Assuming a static nonlinear relationship between $\hat{W}^{[k+1]}$ and $W_0$, Figure 3.31 shows that $Y$ is the output of a Hammerstein system whose input is $\hat{W}^{[k+1]}$. This means that the weighting of (3.17) may be used because the stochastic nonlinear contributions of this system whose variance is $\sigma_{NP}^{[k+1]}$ behave as additive noise at the output.

Section 3.B.3.2 needs to be changed only from (3.55):

$$K = \sum_{l=-N}^{N} \frac{|\hat{S}_{NP}^{[k+1]}(l) - S(\Omega, \theta)|^2}{\sigma_{NP}^{[k+1]}(l)^2}$$

$$\frac{\partial K}{\partial \theta} = 0 \Leftrightarrow \frac{\partial S}{\partial \theta} H \text{diag} \left( \sigma_{NP}^{[k+1]} \right)^{-2} \left( \hat{S}_{NP}^{[k+1]} - \hat{S}_{NP}^{[k+1]} \right) = 0$$

$$\frac{\partial S}{\partial \theta} H \text{diag} \left( \sigma_{NP}^{[k+1]} \right)^{-2} \left( \frac{\partial S}{\partial \theta} \hat{\delta}_{NP}^{[k+1]} - \hat{\delta}_{NP}^{[k+1]} \right) = 0 \quad (3.64)$$

The influence of $\hat{\delta}_{NP}^{[k]}$ on (3.64) needs to be studied (using (3.54)). The averaged nonparametric FRF $\hat{S}_{NP}^{[k+1]}$ is obtained by computing the mean of the nonparametric FRFs computed from the measurements of the different realizations of the excitation. It should be added that because there is no measurement nor process noise, dividing the spectra before the averaging process does not introduce any bias.

$$\hat{S}_{NP}^{[k+1]} = \frac{1}{E} \sum_{m=1}^{E} \text{diag} \left( \hat{W}^{[k+1]}_m \right)^{-1} Y^m$$

Dividing and multiplying the left hand side by $\hat{W}^{[k+1]}_m$, together with averaging it over
Because \( \text{diag}(\hat{W}_m^{[k+1]})^{-1} \delta_{S,\text{NP}}^{[k+1]} = \text{diag}(\hat{W}_m^{[k+1]})^{-1} \hat{W}_m^{[k+1]} \delta_{S,\text{NP}}^{[k+1]} \),

\[
\begin{align*}
\delta_{S,\text{NP}}^{[k+1]} &= -\frac{1}{E} \sum_{m=1}^{E} \text{diag}(\hat{W}_m^{[k+1]})^{-1} \text{diag}(S_0) \text{diag}(\hat{W}_m^{[k+1]}) \\
&= \text{diag}\left(O\left(\frac{1}{\sqrt{N}}\right)\right)^{-1} \text{diag}(S_0) \text{O}\left(\frac{1}{N^2}\right)_{2N \times 2N} \frac{\partial S}{\partial \hat{\theta}^{[k]}} \delta_{\hat{\theta}}^{[k]} \tag{3.65} \\
&= \text{diag}\left(O\left(\frac{1}{\sqrt{N}}\right)\right)^{-1} \text{diag}(S_0) \text{O}\left(\frac{1}{N^2}\right)_{2N \times 2N} \frac{\partial S}{\partial \hat{\theta}^{[k]}} \delta_{\hat{\theta}}^{[k]} \tag{3.66}
\end{align*}
\]

In (3.65), the underlined product is a matrix product where the rows of the left factor have a random phase character. Hence the sum over the matrix elements grows only as \( O\left(\sqrt{N}\right) \), not \( O\left(N\right) \).

\[
\begin{align*}
\delta_{S,\text{NP}}^{[k+1]} &= O\left(\sqrt{N}\right) \text{O}\left(\frac{1}{N}\right)_{2N \times p} \delta_{\hat{\theta}}^{[k]} = O\left(\frac{1}{\sqrt{N}}\right)_{2N \times p} \delta_{\hat{\theta}}^{[k]} \tag{3.67}
\end{align*}
\]

Plugging (3.67) into (3.64) yields an equation relating \( \delta_{\hat{\theta}}^{[k+1]} \) to \( \delta_{\hat{\theta}}^{[k]} \):

\[
\begin{align*}
\left(\frac{\partial S}{\partial \hat{\theta}^{[k+1]}}\right)^H \text{diag}\left(\sigma_{\text{NP}}^{[k+1]}\right)^{-2} \frac{\partial S}{\partial \hat{\theta}^{[k+1]}} \delta_{\hat{\theta}}^{[k+1]} &= \\
\left(\frac{\partial S}{\partial \hat{\theta}^{[k+1]}}\right)^H \text{diag}\left(\sigma_{\text{NP}}^{[k+1]}\right)^{-2} \text{O}\left(\frac{1}{\sqrt{N}}\right)_{2N \times p} \delta_{\hat{\theta}}^{[k]} \\
\delta_{\hat{\theta}}^{[k+1]} &= \left(O(1)_{p \times 2N} O(N) \text{diag}(O(1)) O(1)_{2N \times p}\right)^{-1}
\end{align*}
\]
3.C Convergence analysis for the Wiener estimation algorithm

At each iteration, the error on the parameter estimate of the linear part decreases if \( N \) is chosen large enough, as (3.68) shows.

\[
\begin{align*}
O(1)_{p\times2N} O(N) \text{diag}(O(1)) O\left(\frac{1}{\sqrt{N}}\right) \delta^{[k]}_\theta \\
= O\left(N^2\right)^{-1}_{p\times p} O(N)_{p\times2N} O\left(\frac{1}{\sqrt{N}}\right) \delta^{[k]}_\theta \\
= O\left(N^2\right)^{-1}_{p\times p} O\left(N^{\frac{3}{2}}\right)_{p\times p} \delta^{[k]}_\theta \\
= O\left(\frac{1}{\sqrt{N}}\right)_{p\times p} \hat{\delta}^{[k]}_\theta
\end{align*}
\]

(3.68)

3.C Convergence analysis for the Wiener estimation algorithm

3.C.1 Introduction

The assumptions are restated in Section 3.C.1.2 and the notations are defined in Table 3.3. The proof of immediate convergence to a solution different from the true system when no parametric estimate of the linear system \( R \) is computed does not show anything new compared with Section 3.B.2 and will not be repeated. Only the parametric case will be developed in Section 3.C.2 following the lines of Section 3.B.3.

3.C.1.1 The system

The studied system is a Wiener system (Figure 2.4), excited with a random phase multisine of constant power. The notations that are used in this analysis are summarized in Table 3.3.

3.C.1.2 Assumptions

The assumptions of Section 3.B.1 are adapted to the Wiener system estimation algorithm: instead of the static nonlinearity, its inverse is estimated, hence Assumption 3.1 is reformulated and Assumption 3.6 is added. Because Assumption 3.6 implies that \( f(v) \) is not even, an assumption like Assumption 2.6 need not be included.

**Assumption 3.1** The class of models used for the estimation contains the true model for the inverse of the static nonlinearity as well as for the linear system.
Table 3.3: Notations used in the analysis of the convergence of the Wiener estimation algorithm

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>Number of excitation frequencies</td>
</tr>
<tr>
<td>$p$</td>
<td>The number of parameters $\hat{\theta}[k]$ describing the linear system at iteration $k$</td>
</tr>
<tr>
<td>$q$</td>
<td>The number of parameters $\hat{\varphi}[k]$ describing the static nonlinearity at iteration $k$</td>
</tr>
<tr>
<td>$U$</td>
<td>Components of the Fourier transform of the input (at the excitation frequencies)</td>
</tr>
<tr>
<td>$\text{diag}(R_0)$</td>
<td>Diagonal $2N \times 2N$ matrix containing the transfer function of the linear system at the excitation frequencies</td>
</tr>
<tr>
<td>$V_0 = \text{diag}(R_0)U$</td>
<td>Fourier coefficients of the intermediate signal (at the excitation frequencies)</td>
</tr>
<tr>
<td>$B(Y)$</td>
<td>$2N \times q$ matrix containing the Fourier coefficients of the basis functions used in the estimation process applied to the output (one basis function in each column)</td>
</tr>
</tbody>
</table>

\[
B(Y) = [B_1(Y) \ B_2(Y) \ldots \ B_q(Y)]
\]

\[
B(Y)_{l,i} = O \left(\frac{1}{\sqrt{N}}\right) C_{l,i} e^{j\phi_{l,i}}
\]

<table>
<thead>
<tr>
<th>Expression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_0 = B(Y) \varphi_0$</td>
<td>Fourier coefficients of the intermediate signal (at the excitation frequencies)(^a)</td>
</tr>
<tr>
<td>$\hat{V}^{[k]} = B(Y) \hat{\varphi}^{[k]}$</td>
<td>Estimate (at iteration $k$) of the static nonlinearity, computed in the frequency domain</td>
</tr>
</tbody>
</table>

\(^a\)Since the same matrix of basis functions is used to define the system and to identify it, Assumption 3.1 is verified.
Assumption 3.2 The excitation $u(t)$ is a random phase multisine with power $P$ (independently of $N$), distributed over a sufficient big number of frequencies to ensure a persistent excitation (even for $N \to \infty$).

Assumption 3.3 The model of the inverse of the static nonlinearity is linear in its parameters and each of the basis functions $B_k$ applied to the random phase multisine input has as output a multisine with random phase and whose Fourier coefficients are an $O\left(\frac{1}{\sqrt{N}}\right)$.

Assumption 3.4 The number of parameters $p$ used to model the linear system and the number of parameters $q$ used to model the inverse of the static nonlinearity are kept constant when the number $N$ of excitation lines in the spectrum increases. This means also that $\frac{p}{N}$ and $\frac{q}{N}$ are $O\left(\frac{1}{N}\right)$.

Assumption 3.5 The algorithm starts by computing a consistent parametric estimate of the best linear approximation of the Wiener system. This estimate is proportional to the true linear system $R_0$ with an error term that is an $O\left(\frac{1}{\sqrt{N}}\right)$. It is assumed that this error is small enough so that the error $\tilde{\delta}_R^{[k]}$ on the estimate of $R$ is linear in $\tilde{\delta}_\theta^{[k]}$ (see (3.76) in Section 3.C.2.2).

Assumption 3.6 The static nonlinearity $f(v)$ is invertible.

Assumption 3.7 The following matrix product is well-conditioned and invertible $\forall N$ ($\infty$ included):

$$B(Y)^H B(Y)$$

3.C.2 Convergence for the parametric case

3.C.2.1 Influence of the estimation error of the linear system on the estimate of the static nonlinearity

$$\hat{R}_{NP}^{[k]} = \hat{V}^{[k]} U$$

(3.69)
Suppose that a parametric model $\hat{R}^{[k]}$ has been fitted to $\hat{R}^{[k]}_{\text{NP}}$ (3.69) starting from $U$ and $\hat{V}^{[k]}$. In order to estimate the inverse of the static nonlinearity (see Figure 3.32), the norm of the error $F^{[k+1]}$ will be minimized with respect to $\hat{\varphi}^{[k+1]}$:

\[
F^{[k+1]} = \text{diag}(\hat{R}^{[k]}) U - B(Y) \hat{\varphi}^{[k+1]} \\
= \text{diag}(R_0) U + \text{diag}(\hat{\delta}^{[k]}_R) U - B(Y) \varphi_0 - B(Y) \hat{\delta}^{[k+1]}_\varphi \\
= \text{diag}(\hat{\delta}^{[k]}_R) U - B(Y) \hat{\delta}^{[k+1]}_\varphi
\]

(3.70)

$\hat{\delta}^{[k+1]}_\varphi$ from (3.70) minimizes the norm of $F^{[k+1]}$. As in Section 3.B.3.1, the order of magnitude of $\hat{\delta}^{[k+1]}_\varphi$ can be estimated:

\[
\hat{\delta}^{[k+1]}_\varphi = \left( B(Y)^H B(Y) \right)^{-1} B(Y)^H \text{diag}(U) \hat{\delta}^{[k]}_R
\]

(3.70)

\[
\hat{\delta}^{[k+1]}_\varphi \approx \text{diag} \left( O \left( \frac{1}{\sqrt{N}} \right) \right) \hat{\delta}^{[k]}_R
\]

\[
\approx O \left( \frac{1}{N} \right) \hat{\delta}^{[k]}_R
\]

3.C.2.2 Influence of the estimation error of the static nonlinearity on the estimate of the linear system

$\hat{V}^{[k+1]}$ is the result of the simulation of the inverted static nonlinearity applied to $Y$:

\[
\hat{V}^{[k+1]} = B(Y) \hat{\varphi}^{[k+1]} = B(Y) \left( \varphi_0 + \hat{\delta}^{[k+1]}_\varphi \right) \\
= V_0 + O \left( \frac{1}{\sqrt{N}} \right)_{2N \times q} O \left( \frac{1}{N} \right)_{q \times 2N} \hat{\delta}^{[k]}_R \\
= V_0 + O \left( \frac{1}{N^2} \right)_{2N \times 2N} \hat{\delta}^{[k]}_R
\]

(3.71)

On the other hand, Figure 3.33 shows that there is a static nonlinear relationship between $V_0$ and $\hat{V}^{[k+1]}$:

\[
\hat{V}^{[k+1]} = \alpha^{[k+1]} V_0 + \hat{V}^{[k+1]}_{\Delta S}
\]

(3.72)
When (3.71) and (3.72) are combined after scaling the result of (3.72), the following result can be obtained for $\hat{V}_{k+1}$ which is uncorrelated frequency per frequency with $V_0$:

$$
\hat{V}_{k+1} = O \left( \frac{1}{N^2} \right)_{2N \times 2N} \hat{\delta}_R^{[k]}
$$

(3.73)

Now, when $\hat{V}_{k+1}$ is used to estimate the linear system $R_0$, the following cost function is minimized:

$$
K = \sum_{l=-N}^{N} \left| V_0(l) + \hat{V}_{k+1}^{[\Delta S]}(l) - \hat{R}_{k+1}(l)U(l) \right|^2
$$

(3.74)

In the minimum of the cost function defined in (3.74), the derivative of $K$ with respect to the parameters $\theta$ is 0:

$$
0 = \left( \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \right)^H \text{diag}(U)^H \left( V_0 + \hat{V}_{k+1}^{[\Delta S]} - \text{diag}(\hat{R}_{k+1})U \right)
$$

$$
\Leftrightarrow 0 = \left( \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \right)^H \text{diag}(U)^H \left( \hat{V}_{k+1}^{[\Delta S]} - \text{diag}(U) \hat{\delta}_R^{[k+1]} \right)
$$

(3.75)

The parameters $\hat{\theta}^{[k+1]}$ can be estimated from (3.75). Using Assumption 3.5, $\hat{\delta}_R^{[k+1]}$ can be rewritten:

$$
\hat{\delta}_R^{[k+1]} = \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \hat{\delta}_R^{[k+1]}
$$

(3.76)

Hence, (3.75) becomes

$$
0 = \left( \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \right)^H \text{diag}(U)^H \left( \hat{V}_{k+1}^{[\Delta S]} - \text{diag}(U) \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \hat{\delta}_R^{[k+1]} \right)
$$

$$
\Leftrightarrow \hat{\delta}_R^{[k+1]} = \left[ \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \text{diag}(U)^H \text{diag}(U) \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \right]^{-1} \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \text{diag}(U)^H \hat{V}_{k+1}^{[\Delta S]}
$$

(3.77)
Figure 3.34: The relation between $U$ and $\hat{V}^{[k+1]}$ can be represented as a Wiener system.

\[
\delta_\theta^{[k+1]} = \left[ O(1)_{p \times p} \text{diag} \left( O \left( \frac{1}{\sqrt{N}} \right)_{2N \times 1} \right) \text{diag} \left( O \left( \frac{1}{\sqrt{N}} \right)_{2N \times 1} \right) O(1)_{p \times p} \right]^{-1} \\
O(1)_{p \times 2N} \text{diag} \left( O \left( \frac{1}{\sqrt{N}} \right)_{2N \times 1} \right) O \left( \frac{1}{N^{\frac{3}{2}}} \right)_{2N \times 2N} O(1)_{2N \times p} \delta_\theta^{[k]} \\
= O \left( \frac{1}{\sqrt{N}} \right)_{p \times 2N} O \left( \frac{1}{N} \right)_{2N \times p} \delta_\theta^{[k]} \\
= O \left( \frac{1}{\sqrt{N}} \right)_{p \times p} \delta_\theta^{[k]} 
\]

The orders of magnitude of the terms in (3.77) can be introduced in (3.78) to see if the estimation error decreases. In (3.78), the underlined terms are uncorrelated (3.73) so that the sum in the matrix product grows as an $O \left( \sqrt{N} \right)$, hence the elements of the matrix multiplication (which implies a summation over $2N$ elements) are $O \left( \frac{1}{N} \right)$ (3.79).

When the algorithm converges, (3.80) shows that each iteration step divides the estimation error by $\sqrt{N}$.

### 3.C.3 Convergence when multiple experiments are used

This section analyzes the convergence of Algorithm 6. Again, as only the estimate of the linear part is slightly different, the contents of Section 3.C.2.1 is not repeated here.

#### 3.C.3.1 Influence of the estimation error of the static nonlinearity on the estimate of the linear system

The equations of Section 3.C.2.2 need to be adapted to reflect the change in the cost function from (3.25) to (3.36). For the same reasons as in Section 3.B.4.1, $\left\| \sigma^{[k]} \right\| = 1$.

Figure 3.34 shows that the relationship between $U$ and $\hat{V}^{[k+1]}$ can be described by a Wiener system. Hence, the stochastic nonlinear contributions which disturb the parametric estimate of $\hat{R}^{[k+1]}$ can be considered as noise that is added at the output of
the system (Schoukens et al., 1998a; Pintelon and Schoukens, 2001). This means that weighting the estimate as in (3.36) make sense, and hence the cost function which is minimized to obtain \( \hat{\theta}^{[k+1]} \) is:

\[
K = \sum_{l=-N}^{N} \frac{\left| \hat{R}^{[k+1]}_\text{NP} - R(\Omega, \theta) \right|^2}{\sigma^{[k+1]}_\text{NP}(l)^2}
\]

In \( \hat{\theta}^{[k+1]} \), the derivative of this cost function is 0:

\[
\frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \text{diag} \left( \sigma^{[k+1]}_\text{NP} \right)^{-2} \left( \hat{R}^{[k+1]} - \hat{R}^{[k+1]}_\text{NP} \right) = 0 \Leftrightarrow \\
\frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \text{diag} \left( \sigma^{[k+1]}_\text{NP} \right)^{-2} \left( \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \hat{\delta}^{[k+1]} - \hat{\delta}^{[k+1]}_\text{NP} \right) = 0 \quad (3.81)
\]

The value of \( \hat{\delta}^{[k+1]}_\text{NP} \) in (3.81) has to be studied to be able to relate \( \hat{\delta}^{[k+1]} \) with \( \hat{\delta}^{[k]} \):

\[
\hat{R}^{[k+1]}_\text{NP} = R_0 + \hat{\delta}^{[k+1]}_\text{NP} = \frac{1}{E} \sum_{m=1}^{E} \text{diag}(U^m)^{-1} \left( V^m_0 + \hat{V}^{[k+1]}_\Delta \right)
\]

\[
\Leftrightarrow \hat{\delta}^{[k+1]}_\text{NP} = \frac{1}{E} \sum_{m=1}^{E} \text{diag}(U^m)^{-1} \hat{V}^{[k+1]}_\Delta
\]

\[
= \frac{1}{E} \sum_{m=1}^{E} \text{diag} \left( O \left( \frac{1}{\sqrt{N}} \right) \right)^{-1} O \left( \frac{1}{N^2} \right)_{2N \times 2N} \frac{\partial R}{\partial \hat{\theta}^{[k]}} \hat{\delta}^{[k]}_\theta
\]

\[
= O \left( \sqrt{N} \right) O \left( \frac{1}{N} \right)_{2N \times p} \hat{\delta}^{[k]}_\theta = O \left( \frac{1}{\sqrt{N}} \right)_{2N \times p} \hat{\delta}^{[k]}_\theta \quad (3.82)
\]

In (3.82), the underlined product involves a sum over \( 2N \) uncorrelated terms, hence their value grows only as \( O \left( \sqrt{N} \right) \), not \( O \left( N \right) \). Plugging (3.83) into (3.81) yields:

\[
\hat{\delta}^{[k+1]}_\theta = \left( \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \right)^H \text{diag} \left( \sigma^{[k+1]}_\text{NP} \right)^{-2} \left( \frac{\partial R}{\partial \hat{\theta}^{[k+1]}} \hat{\delta}^{[k+1]} - \hat{\delta}^{[k+1]}_\text{NP} \right)
\]

\[
= \left( O(1)^H_{2N \times p} \text{diag} \left( O \left( \frac{1}{\sqrt{N}} \right) \right)^{-2} O(1)_{2N \times p} \right)^{-1}
\]

\[
O(1)^H_{2N \times p} \text{diag} \left( O \left( \frac{1}{\sqrt{N}} \right) \right)^{-2} O \left( \frac{1}{\sqrt{N}} \right)_{2N \times p} \hat{\delta}^{[k]}_\theta
\]

\[
= O \left( N^2 \right)_{p \times p} O(N)_{p \times 2N} O \left( \frac{1}{\sqrt{N}} \right)_{2N \times p} \hat{\delta}^{[k]}_\theta
\]
\[ O(N^2)_{p \times p}^{-1} O(N^2)_{p \times p} \delta[k] = O\left( \frac{1}{\sqrt{N}} \right)_{p \times p} \delta[k] \] (3.84)

The convergence of the iterative algorithm for the identification of Wiener systems with a set of different input signals follows from (3.84): at each iteration, the error is divided by \( \sqrt{N} \).

### 3.D Product of columns of \( B(U) \) or \( B(Y) \)

Because of Assumption 2.3, an element of \( B(U) \) can be written as (3.85). Assumption 3.3 allows analogous expressions for \( B(Y) \).

\[ B(U)_{l,m} = O\left( \frac{1}{\sqrt{N}} \right) C_{l,m} e^{i\phi_{l,m}} \] (3.85)

#### 3.D.1 Product of two identical columns

\[
B(U)^H_{:,i} B(U)_{:,i} = \sum_{l=1}^{2N} O\left( \frac{1}{\sqrt{N}} \right) C_{l,i} e^{-j\phi_{l,i}} O\left( \frac{1}{\sqrt{N}} \right) C_{l,i} e^{j\phi_{l,i}} \\
= O\left( \frac{1}{N} \right) \sum_{l=1}^{2N} C_{l,i}^2 \\
= O(1) \] (3.86)

#### 3.D.2 Product of two different columns

\[
B(U)^H_{:,i} B(U)_{:,m} = \sum_{l=1}^{2N} O\left( \frac{1}{\sqrt{N}} \right) C_{l,i} e^{-j\phi_{l,i}} O\left( \frac{1}{\sqrt{N}} \right) C_{l,m} e^{j\phi_{l,m}} \\
= O\left( \frac{1}{N} \right) \sum_{l=1}^{2N} C_{l,i} C_{l,m} e^{j\phi_{l,m} - j\phi_{l,i}} \\
\leq O(1)
\]
3.E Consistency of the Hammerstein estimation algorithm

In this section, the consistency of Algorithm 2 will be studied in the Output Error framework as shown in Figure 3.35. It will be proven that under certain circumstances, the method is consistent. Stated informally, this means that as more data are collected, the estimates come closer to the true value.

The consistency of Algorithm 2 will be proven by showing that the steps of Algorithm 2 converge to the noiseless values when the number of components in the excitation \( N \to \infty \). Because of the convergence of Algorithm 2 that was demonstrated in Appendix 3.B in the noise free case, it will follow that under the union of all assumptions, Algorithm 2 converges to a consistent estimate for \( N \to \infty \). The three steps of the consistency proof are built on Pintelon and Schoukens (2001).

3.E.1 Notations and Assumptions

Since the consistency of Algorithm 2 is to be studied, the notations introduced there will be reused. Recall that Algorithm 2 first estimates the linear part \( S \), then the nonlinear part \( f \). The linear part \( S \) is estimated in the very first step by minimizing the cost function \( K^{[0]} \) with respect to the parameters \( \theta \):

\[
K^{[0]}_N(\theta, Y) = \frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y(\Omega_l)}{U(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2
\]

The iterative scheme then alternates between estimating the static nonlinear part \( f \) and the linear part \( S \). The static nonlinearity \( f \) is estimated by minimizing \( V^{[k]} \) with respect to \( \varphi \):

\[
\hat{Y}^{[k]}_a(\Omega) = \hat{S}^{[k]}_a(\Omega) \mathcal{F}(f_a(u(t))) ; a = 1, 2, \ldots, q
\]
\[ V_N^{[k]}(\varphi, Y) = \frac{1}{N} \sum_{l=1}^{N} \left| Y(\Omega_l) - \sum_{a=1}^{q} \varphi_a \hat{Y}_a^{[k]}(\Omega_l) \right|^2 \]

The linear part is estimated by minimizing \( K^{[k]} \) with respect to \( \theta \):

\[ K_N^{[k]}(\theta, Y) = \frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y(\Omega_l)}{W_l^{[k-1]}(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2 \]

The cost functions \( K_N^{[0]}(\theta, Y) \), \( K_N^{[k]}(\theta, Y) \), \( V_N^{[k]}(\varphi, Y) \) and their higher order derivatives with respect to \( \theta \) and \( \varphi \) may not exist for some values of the model parameters \( \theta \) and \( \varphi \). To stay within the bounds of the theorems proven in Pintelon and Schoukens (2001), regular sets \( \theta \) and \( \varphi \) of \( \theta \) and \( \varphi \) are constructed where \( K_N^{[0]}(\theta, Y) \), \( K_N^{[k]}(\theta, Y) \), \( V_N^{[k]}(\varphi, Y) \) and their higher order derivatives with respect to \( \theta \) and \( \varphi \) exist and are finite. By construction, these sets are closed and bounded compact. The minimizers are then defined as

\[
\hat{\theta}^{[0]}(Y) = \arg \min_{\theta \in \theta} K_N^{[0]}(\theta, Y)
\]
\[
\hat{\theta}^{[k]}(Y) = \arg \min_{\theta \in \theta} K_N^{[k]}(\theta, Y)
\]
\[
\hat{\varphi}^{[k]}(Y) = \arg \min_{\varphi \in \varphi} V_N^{[k]}(\varphi, Y)
\]

**Assumption 4.1** The excitation signal is a normalized random multisine (see Definition 1.4).

**Assumption 4.2** The model set described by the parameterization of \( f \) and \( S \) contains the true system’s static nonlinearity \( f \) and linear part \( S \) respectively.

**Assumption 4.3** The noise \( M \) is assumed to have the following properties \((k, l \in [0, N - 1])\)

1. \( \mathbb{E}[M(\Omega_l)] = 0. \)
2. \( \mathbb{E}[M(\Omega_l)M(\Omega_k)^*] = \sigma_M^2(\Omega_l)\delta_{kl}. \)
3. \( \mathbb{E}[M(\Omega_l)|M(\Omega_k)|^2] = 0. \)
4. \( \mathbb{E}[|M(\Omega_l)|^2|M(\Omega_k)|^2] = \begin{cases} 
\sigma_M^2(\Omega_k)\sigma_M^2(\Omega_l), & k \neq l \\
2\sigma_M^4(\Omega_k), & k = l \end{cases}. \)
3.E Consistency of the Hammerstein estimation algorithm

Assumption 4.4 There exists an $N_0$ such that $\forall N > N_0$ including $N = \infty$, the expected value of the cost function $K_{N}^{[k]}(\theta) = E\left[K_{N}^{[k]}(\theta, Y)\right]$ has a unique global minimum $\tilde{\theta}^{[k]}(Y_0)$, which is an interior point of $\theta$.

Assumption 4.5 There exists an $N_0$ such that $\forall N > N_0$ including $N = \infty$, the expected value of the cost function $V_{N}^{[k]}(\phi) = E\left[V_{N}^{[k]}(\phi, Y)\right]$ has a unique global minimum $\tilde{\phi}^{[k]}(Y_0)$, which is an interior point of $\phi$.

Assumption 4.6 The noise $M$ is independent of the input $U$ and the undisturbed output $Y_0$.

Assumption 4.7 The following cost function is continuous with respect to $\theta$ and $\phi$ in $\theta$ and $\phi$ respectively:

$$K_{N}^{[0]} = \frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y(\Omega_l)}{U(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2$$

Assumption 4.8 The following cost function is continuous with respect to $\theta$ and $\phi$ in $\theta$ and $\phi$ respectively:

$$V_{N}^{[k]} = \frac{1}{N} \sum_{l=1}^{N} \left| Y(\Omega_l) - \sum_{a=1}^{q} \hat{\phi}_a^{[k]} \hat{Y}_a^{[k]}(\Omega_l) \right|^2$$

Assumption 4.9 The following cost function is continuous with respect to $\theta$ and $\phi$ in $\theta$ and $\phi$ respectively:

$$K_{N}^{[k]} = \frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y(\Omega_l)}{W^{[k-1]}(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2$$

These assumptions are taken from Schoukens et al. (1998a) (Pintelon and Schoukens (2001) is also a reference on this subject). Assumption 4.1 is fulfilled for input signals described in Definition 1.4. The properties of complex normally distributed noise meet Assumption 4.3. Assumptions 4.4 and 4.5 are identical, but each is applied to a particular step of the iterative process that minimizes two cost functions: both require a unique minimum to exist for the noiseless data. Note that the identification of a Hammerstein system cannot be unique unless the gain degeneration is taken into account (see Section 3.2.2 and Figure 3.2). This has not been repeated in the assumptions because
Algorithm 2 states it explicitly in Step 2.4.

### 3.E.2 Convergence of (3.5) to the noiseless solution

Algorithm 2 is started with \( \hat{\theta}_0 \), the minimizer of (3.5):

\[
\frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y(\Omega_l)}{U(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2 = \frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y_0(\Omega_l) + M(\Omega_l)}{U(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2
\]

The assumptions 4.1–4.9 match the assumptions 7.2 (Assumption 4.1), 7.4 and 7.5 (Assumption 4.3), 7.6 (Assumption 4.7), 7.7 (Assumption 4.4) given in Pintelon and Schoukens (2001, Ch. 7, pp. 191–192). Because \( K_N^{[0]} \) is quadratic in the measurements \( Y \), Theorem 7.21 of Pintelon and Schoukens (2001, Ch. 7, p. 196) can be applied to (3.87) and \( \hat{\theta}_0 \) converges strongly to the minimizer of \( E[K_N^{[0]}] \):

\[
\text{a.s. } \lim_{N \to \infty} \hat{\theta}_0 = \tilde{\theta}_0
\]

The mean of the cost function (3.87) has the same minimizer as the noiseless cost function (3.5):

\[
E[K_N^{[0]}] = E \left[ \frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y_0(\Omega_l) + M(\Omega_l)}{U(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2 \right]
\]

\[
= E \left[ \frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y_0(\Omega_l)}{U(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2 \right] + E \left[ \frac{1}{N} \sum_{l=1}^{N} |M(\Omega_l)|^2 \right] +
\]

\[
E \left[ 2 \text{Re} \left( \frac{1}{N} \sum_{l=1}^{N} \left( \frac{Y_0(\Omega_l)}{U(\Omega_l)} - S_\theta(\Omega_l; \theta) \right) M(\Omega_l)^* \right) \right]
\]

\[
= \frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y_0(\Omega_l)}{U(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2 + \frac{1}{N} \sum_{l=1}^{N} E[|M(\Omega_l)|^2] +
\]

\[
\frac{2}{N} \text{Re} \left( \sum_{l=1}^{N} E \left[ \frac{Y_0(\Omega_l)}{U(\Omega_l)} - S_\theta(\Omega_l; \theta) \right] M(\Omega_l)^* \right)
\]

\[
= \frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y_0(\Omega_l)}{U(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2 + \frac{1}{N} \sum_{l=1}^{N} E[|M(\Omega_l)|^2]
\]

because \( E[|M(\Omega_l)|^2] \) does not depend on the parameter values and because \( E[M(\Omega_l)] = 0 \). Hence \( \tilde{\theta}_0 \) is also the solution that would be obtained for noiseless measurements.
3.E Consistency of the Hammerstein estimation algorithm

3.E.3 Convergence of (3.9) to the noiseless solution

The following step in Algorithm 2 is to estimate the parameters \( \varphi \) of the static nonlinearity \( f \).

\[
\hat{Y}_a^k(\Omega) = \hat{S}^k(\Omega)W_a(\Omega); a = 1, \ldots, q
\]

\[
V^k = \sum_{l=1}^{N} \left| Y(\Omega_l) - \sum_{a=1}^{q} \hat{\varphi}_a[k] \hat{Y}_a^k(\Omega_l) \right|^2
\]

\[
= \sum_{l=1}^{N} \left| Y_0(\Omega_l) + M(\Omega_l) - \sum_{a=1}^{q} \hat{\varphi}_a[k] \hat{Y}_a^k(\Omega_l) \right|^2
\]

This is achieved by computing the contribution of each basis function to the output in (3.8). Because \( \hat{Y}_a^k(\Omega) \) depends on a previous estimate (i.e., \( \hat{\theta}^k \)) that converges strongly to the noiseless solution, Theorem 7.21 of Pintelon and Schoukens (2001) is not applicable. However, adding Assumption 4.8 allows to apply Corollary 7.22 of Pintelon and Schoukens (2001) which again states that as the number of excitation frequencies \( N \to \infty \), \( \varphi[k] \) converges to the minimizer of \( E[V^k_N] \):

\[
\text{a.s.} \lim_{N \to \infty} \varphi[k] = \tilde{\varphi}[k]
\]

The noise appears in a similar way in (3.87) and (3.88) so that \( \tilde{\varphi}[k] \) is equal to the minimizer of the noiseless cost function for similar reasons.

3.E.4 Convergence of (3.7) to the noiseless solution

The next step in Algorithm 2 is to estimate the parameters \( \theta \) of the linear part \( S \).

\[
\hat{S}_{NP}^k(\Omega) = \frac{Y(\Omega)}{W^{[k-1]}(\Omega)}; k \geq 1
\]

\[
\hat{\theta}[k] = \arg \min_{\theta} \sum_{l=1}^{N} \left| \frac{Y(\Omega_l)}{W^{[k-1]}(\Omega_l)} - S_\theta(\Omega_l; \theta) \right|^2
\]

This is achieved by computing the output \( \hat{W}^{[k-1]} \) of the static nonlinearity \( f \) using the most recent estimate \( \hat{\varphi}[k-1] \). Because \( \hat{W}^{[k-1]}(\Omega) \) depends on a previous estimate (i.e., \( \hat{\varphi}[k-1] \)) that converges strongly to the noiseless solution, Theorem 7.21 of Pintelon and Schoukens (2001) is not applicable. However, adding Assumption 4.9 allows to apply
Corollary 7.22 of Pintelon and Schoukens (2001) which again states that as the number of excitation frequencies $N \to \infty$, $\hat{\theta}^{[k]}$ converges to the minimizer of $E\left[K_{N}^{[k]}\right]$:

$$\text{a.s.} \lim_{N \to \infty} \hat{\theta}^{[k]} = \tilde{\theta}^{[k]}$$

Except for replacing $U$ with $\tilde{W}^{[k-1]}$, $E\left[K_{N}^{[k]}\right]$ and $E\left[K_{N}^{[0]}\right]$ are the same, so that one can easily see that $\tilde{\theta}^{[k]}$ is equal to the minimizer of the corresponding noiseless cost function.

### 3.F Consistency of the Wiener estimation algorithm

In this section, the consistency of Algorithm 4 will be studied in the Output Error framework as shown in Figure 3.36 analogously to Appendix 3.E. It will be proven that under certain circumstances, the method is consistent.

The proof consists of two steps, corresponding to the two identification steps of the method (Step 4.1 and Step 4.3). The consistency of both steps will be proven and the consistency of Algorithm 4 follows from these results.

#### 3.F.1 Notations and Assumptions

Since the consistency of Algorithm 4 is to be studied, the notations introduced there will be reused. Recall that Algorithm 4 first estimates the linear part $R$, then the nonlinear part $f$. The linear part $R$ is estimated in Step 4.1 by minimizing the cost function $K_{N}(\theta)$ with respect to the parameters $\theta$ (3.20):

$$K_{N}(\theta, Y) = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Y(\Omega_i)}{U(\Omega_i)} - R_{\theta}(\Omega_i; \theta) \right|^2$$
Then the static nonlinearity $f$ is estimated by minimizing $V_N(\varphi)$ with respect to $\varphi$ (3.23):

$$V_N(\varphi, Y) = \frac{1}{N} \sum_{n=1}^{N} \left| Y(\Omega_l) - \sum_{a=1}^{q} \varphi_a \hat{Y}_a(\Omega_l) \right|^2$$

For the same reasons as in Appendix 3.E.1, closed and bounded compact sets $\theta$ and $\varphi$ are constructed where the cost functions $K_N(\theta, Y)$, $V_N(\varphi, Y)$ and their higher order derivatives exist with respect to $\theta$ and $\varphi$.

**Assumption 5.1** The excitation signal is a normalized random multisine (see Definition 1.4).

**Assumption 5.2** The model set described by the parameterization of $f$ and $R$ contains the true system’s static nonlinearity $f$ and linear part $R$ respectively.

**Assumption 5.3** The noise $M$ is assumed to have the following properties ($k, l \in [0, N - 1]$)

1. $E[M(\Omega_l)] = 0$.

2. $E[M(\Omega_l)M(\Omega_k)^*] = \sigma^2_M(\Omega_l) \delta_{kl}$.

3. $E[M(\Omega_l)M(\Omega_k)^2] = 0$.

4. $E[|M(\Omega_l)|^2|M(\Omega_k)|^2] = \begin{cases} \sigma^2_M(\Omega_k)\sigma^2_M(\Omega_l), & k \neq l \\ 2\sigma^4_M(\Omega_k), & k = l \end{cases}$.

**Assumption 5.4** There exists an $N_0$ such that $\forall N > N_0$ including $N = \infty$, the expected value of the cost function $K_N(\theta) = E[K_N(\theta, Y)]$ has a unique global minimum $\hat{\theta}(Y_0)$, which is an interior point of $\theta$.

**Assumption 5.5** There exists an $N_0$ such that $\forall N > N_0$ including $N = \infty$, the expected value of the cost function $V_N(\varphi) = E[V_N(\varphi, Y)]$ has a unique global minimum $\hat{\varphi}(Y_0)$, which is an interior point of $\varphi$.

**Assumption 5.6** The noise $M$ is independent of the input $U$ and the undisturbed output $Y_0$.

**Assumption 5.7** The following cost function is continuous with respect to $\theta$ and $\varphi$ in $\theta$ and $\varphi$ respectively:

$$K_N(\theta) = \frac{1}{N} \sum_{l=1}^{N} \left| \frac{Y(\Omega_l)}{U(\Omega_l)} - R_{\theta}(\Omega_l; \theta) \right|^2$$
Assumption 5.8 The following cost function is continuous with respect to \( \theta \) and \( \varphi \) in \( \theta \) and \( \varphi \) respectively:

\[
V_N(\varphi) = \frac{1}{N} \sum_{l=1}^{N} \left| Y(\Omega_l) - \sum_{a=1}^{q} \hat{\varphi}_a \hat{Y}_a(\Omega_l) \right|^2
\]

These assumptions are the same as in Appendix 3.E.1.

3.F.2 Convergence of Step 4.1 to the noiseless solution

Since the assumptions 5.1–5.8 are similar to 4.1–4.9 and (3.20) is similar to (3.5), the same result as in Appendix 3.E.2 is obtained:

\[
\text{a.s. lim }_{N \to \infty} \hat{\theta} = \tilde{\theta}
\]

and \( \tilde{\theta} \) is also equal to the minimizer of the noiseless cost function.

3.F.3 Convergence of Step 4.3 to the noiseless solution

As in (3.9), (3.23) depends on previously estimated parameters, so that Assumption 5.8 is necessary to be able to use Corollary 7.22 of Pintelon and Schoukens (2001). The conclusion is the same

\[
\text{a.s. lim }_{N \to \infty} \hat{\varphi} = \tilde{\varphi}
\]

and \( \tilde{\varphi} \) is also equal to the noiseless minimizer.

3.F.4 Discussion

Note that while Algorithm 4 is consistent, the same does not hold for the iterative variant, Algorithm 5, because \( \text{E}[V_N(\varphi)] \) is not minimized by \( \varphi_e(N) \) from Assumption 5.5.

In the case of the iterative algorithm, the inverse of the static nonlinearity is estimated. Consequently, \( M \) is present in \( V_{Y,a}(\Omega_l) \) that will be split up into its true (undisturbed value) and the remaining part which contains \( M \):

\[
V_{Y,a}(\Omega_l) = V_{0,Y,a}(\Omega_l) + \delta_a(\Omega_l)
\]
with in general $\mathbb{E}[\delta_a(\Omega_l)] \neq 0$.

With the assumptions and notations from both Appendix 3.F.1 and Section 3.7:

$$
\mathbb{E}[V_N(\varphi)] = \mathbb{E}\left[ \frac{1}{N} \sum_{l=1}^{N} \sum_{a=1}^{q} \varphi_a V_{Y,a}(\Omega_l) - \hat{V}^{[k;b]}(\Omega_l) \right]^2
$$

$$
\mathbb{E}[V_N(\varphi)] = \mathbb{E}\left[ \frac{1}{N} \sum_{l=1}^{N} \sum_{a=1}^{q} \varphi_a V_{Y,a}(\Omega_l) - \hat{V}^{[k;b]}(\Omega_l) \right]^2 + \frac{1}{N} \mathbb{E}\left[ \sum_{l=1}^{N} \sum_{a=1}^{q} \varphi_a \delta_a(\Omega_l) \right]^2 + 
$$

$$
\frac{2}{N} \mathbb{E}\left[ \sum_{l=1}^{N} \text{Re} \left( \sum_{a=1}^{q} \varphi_a \delta_a^*(\Omega_l) \left( \sum_{a=1}^{q} \varphi_a V_{0,Y,a}(\Omega_l) - \hat{V}^{[k;b]}(\Omega_l) \right) \right) \right]
$$

$$
\frac{2}{N} \sum_{l=1}^{N} \text{Re} \left( \mathbb{E} \left[ \sum_{a=1}^{q} \varphi_a \delta_a^*(\Omega_l) \right] \left( \sum_{a=1}^{q} \varphi_a V_{0,Y,a}(\Omega_l) - \hat{V}^{[k;b]}(\Omega_l) \right) \right)
$$

Because of the underlined terms in (3.89), $\varphi_e(N) \neq \arg \min_{\varphi} \mathbb{E}[V_N(\varphi)]$ so that the estimator is not consistent.

However, the fact that there is bias does not mean that for given measurements, the bias is always larger than the uncertainty on the parameters caused by the stochastic nonlinear contributions. In the measurement example presented in this chapter, the initialization $k = 0$ of Algorithm 5 could be made consistent (as it would consist of the steps in Algorithm 4), but the following steps (i.e. for $k \geq 1$) will not be consistent. But, as Figure 3.22 shows, for a finite measurement record, and in this particular case, the biased algorithm yields better results. Iterating in Algorithm 5 allows to suppress the stochastic nonlinear contributions, resulting in a better estimate of the linear part.
4 Hammerstein-Wiener systems

Hammerstein-Wiener systems contain two nonlinear blocks and an LTI block in cascade, connected as shown in Figure 2.5. They are particular because their RLDS is different from the underlying linear system under certain circumstances. The reasons for this will be investigated in this chapter. An iterative estimation procedure will be proposed for the Hammerstein-Wiener systems. This iterative procedure is analyzed, allowing to predict how the estimation error at iteration $k$ propagates to iteration $k + 1$. 
4.1 Introduction to the identification method

Before any theoretical discussion about the bias of the RLDS of the Hammerstein-Wiener system, the basic steps of the identification method will be given here.

Recall that in Figure 2.5, there is only one linear block: \( S \). Taking inspiration from the identification procedure for Hammerstein systems, it is feasible to estimate the intermediate signal \( W \), using the input data and the estimates of \( f \) and \( S \). Looking at the estimation of Wiener systems, \( W \) could also be computed with the output data and the estimate of the inverse of \( g \).

Roughly speaking, the identification of a Hammerstein-Wiener system could be attempted as follows (refer also to Figure 4.1):

1. Estimate the RLDS to estimate \( S \).
2. Write \( W \) as a weighted sum of basic nonlinear contributions filtered by \( \hat{S} \). These weights are the parameters that appear linearly in the estimate of \( W \) (because \( \hat{S} \) is a linear operator).
3. Write \( W \) as a weighted sum of basis functions applied to \( Y \).
4. Try and make both estimates of \( W \) equal (in a least squares sense). This yields an optimization problem that is easily solved, as the cost function is quadratic in the parameters (the weights from 2 and 3).

Hammerstein-Wiener systems are not special cases of Wiener-Hammerstein systems, so that the RLDS of Hammerstein-Wiener systems has to be computed first. Ideally, it
4.2 What is so particular about Hammerstein-Wiener systems?

Given the results (2.6), (3.1) and (3.19), it is tempting to consider the RLDS as a good estimate of \( S \). It appears that the two nonlinear blocks \( f \) and \( g \) can bias the RLDS in such a way that it becomes different from the underlying LTI system \( S \). This section computes the third Volterra kernel (Schetzen, 1980) of a Hammerstein-Wiener system defined by Figure 2.5 and

\[
\begin{align*}
  v & = \sum_{i=1}^{n} \varphi_i u^i \\
  y & = \sum_{i=1}^{n} \psi_i w^i
\end{align*}
\]

(4.2)

Using this Volterra kernel, the bias contribution to the RLDS is computed. The result of this computation, (4.6) will prove that indeed, (4.1) is wrong in the most general case. All computations in this section will be done for discrete time systems, hence \( \Omega = e^{j\omega} \).

With \( f(u), g(w) \) defined by (4.2), the third Volterra kernel of the Hammerstein-Wiener system \( H_3(e^{j\omega_1}, e^{j\omega_2}, e^{j\omega_3}) \) is given by

\[
H_3(e^{j\omega_1}, e^{j\omega_2}, e^{j\omega_3}) = \psi_1 S \left( e^{(j\omega_1+j\omega_2+j\omega_3)} \right) \varphi_3 + \\
- \frac{2}{3} \psi_2 S \left( e^{(j\omega_1+j\omega_3)} \right) S \left( e^{j\omega_2} \right) \varphi_2 \varphi_1 + \frac{2}{3} \psi_2 S \left( e^{(j\omega_1+j\omega_3)} \right) S \left( e^{j\omega_2} \right) \varphi_2 \varphi_1 + \\
- \frac{2}{3} \psi_2 S \left( e^{(j\omega_2+j\omega_3)} \right) S \left( e^{j\omega_1} \right) \varphi_2 \varphi_1 + \psi_3 S \left( e^{j\omega_1} \right) S \left( e^{j\omega_2} \right) S \left( e^{j\omega_3} \right) \varphi_1^3
\]

(4.3)

Putting \( l = k_1 + k_2 + k_3 \), the contribution of \( H_3 \) to the output at frequency \( l \) is obtained:

\[
Y_3(l) = \sum_{k_1} \sum_{k_2} U(l-k_1-k_2)U(k_1)U(k_2) \left[ \psi_1 S \left( e^{j\omega_1} \right) \varphi_3 + \\
- \frac{2}{3} \psi_2 S \left( e^{(j\omega_1+j\omega_2)} \right) S \left( e^{j\omega_3} \right) \varphi_2 \varphi_1 + \frac{2}{3} \psi_2 S \left( e^{(j\omega_1+j\omega_3)} \right) S \left( e^{j\omega_2} \right) \varphi_2 \varphi_1 + \\
- \frac{2}{3} \psi_2 S \left( e^{(j\omega_2+j\omega_3)} \right) S \left( e^{j\omega_1} \right) \varphi_2 \varphi_1 + \psi_3 S \left( e^{j\omega_1} \right) S \left( e^{j\omega_2} \right) S \left( e^{j\omega_3} \right) \varphi_1^3
\]

(4.4)
\[
\frac{2}{3} \psi_2 S (e^{j(\omega_1 - \omega_{k_1})}) S (e^{j\omega_{k_1}}) \varphi_2 \varphi_1 + \\
\frac{2}{3} \psi_2 S (e^{j(\omega_1 - \omega_{k_2})}) S (e^{j\omega_{k_2}}) \varphi_2 \varphi_1 + \\
\frac{2}{3} \psi_2 S (e^{j(\omega_1 + \omega_{k_1} + \omega_{k_2})}) S (e^{j(\omega_1 - \omega_{k_1} - \omega_{k_2})}) \varphi_2 \varphi_1 + \\
\psi_3 S (e^{j(\omega_1 - \omega_{k_1} - \omega_{k_2})}) S (e^{j\omega_{k_1}}) S (e^{j\omega_{k_2}}) \varphi_3^3
\]

(4.4)

Applying (2.3) to the third order terms in (4.4) yields the contributions of the 3rd Volterra kernel to the nonparametric FRF:

\[
S_{3,\text{NP}}(l) = \frac{Y_3(l)}{U(l)} = \sum_{k_1} \sum_{k_2} \frac{U(l - k_1 - k_2)U(k_1)U(k_2)}{U(l)} \left[ \psi_1 S (e^{j\omega}) \varphi_3^3 + \\
\frac{2}{3} \psi_2 S (e^{j(\omega_1 - \omega_{k_1})}) S (e^{j\omega_{k_1}}) \varphi_2 \varphi_1 + \\
\frac{2}{3} \psi_2 S (e^{j(\omega_1 - \omega_{k_2})}) S (e^{j\omega_{k_2}}) \varphi_2 \varphi_1 + \\
\frac{2}{3} \psi_2 S (e^{j(\omega_1 + \omega_{k_1} + \omega_{k_2})}) S (e^{j(\omega_1 - \omega_{k_1} - \omega_{k_2})}) \varphi_2 \varphi_1 + \\
\psi_3 S (e^{j(\omega_1 - \omega_{k_1} - \omega_{k_2})}) S (e^{j\omega_{k_1}}) S (e^{j\omega_{k_2}}) \varphi_3^3 \right]
\]

(4.5)

The bias contributions in \( S(l) \) are those terms in (4.5) which are independent of the input signal phase realization. These are the terms where the phases of \( U(k_1) \) and \( U(k_2) \) cancel each other, which is the case for \( k_1 = -k_2 \).

Because \( u(t) \) is a real signal, \( \angle U(k_1) = -\angle U(-k_1) \). This means that

\[
\frac{U(l - k_1 - (-k_1))U(k_1)U(-k_1)}{U(l)} = |U(k_1)|^2
\]

which is a real value whose phase (by definition) is independent of the phase realization of the input signal. As in (2.12), \( |U(l)|^2 \) appears 3 times and \( |U(k_1 \neq l)|^2 \) appears 6 times in (4.5). This leads to the \( O(N^{-1}) \) in (4.6).

Keeping only the terms in (4.5) that are independent of the phase realization of \( U \), the bias contributions are computed to be

\[
S_{3,B}(e^{j\omega}) = 6 \psi_1 S (e^{j\omega}) \varphi_3 \sum_{k_1} |U(k_1)|^2 + 
\]

102
4.2 What is so particular about Hammerstein-Wiener systems?

Figure 4.2: Hammerstein-Wiener system used to simulate the results of Figure 4.3

\[ v = u + 0.5u^2 \]
\[ y = w - 0.8w^2 \]

These computations are the application to the specific case of the 3rd order nonlinear contributions of a Hammerstein-Wiener system of the more general expressions found in Schoukens et al. (1998a). The first term in (4.6) is proportional to \( S \), while the last two terms depend quadratically on \( S \). These contributions are different from zero if both \( \varphi_2 \neq 0 \) and \( \psi_2 \neq 0 \).

In practice, the introduced error seems to be small enough as it is the sum over the frequency of complex quantities with varying phases. Adding these results in an averaging effect, which keeps the error small. Anyway, this error is proportional to the power of the input signal, allowing to limit its influence if necessary. A simulation was done to show the size of this error: 4000 different random phase multisines with the same power spectrum were applied to the system shown in Figure 4.2. All nonparametric FRFs (2.3) of these simulations were collected, but only their mean and their standard deviation are shown in Figure 4.3. The Figure 4.3 shows that the difference between the RLDS (the mean of all nonparametric FRFs) and the true linear system inside the Hammerstein-Wiener system is of the same order of magnitude as the stochastic contributions. Once the static nonlinearities are estimated (Section 4.3.2.2), this knowledge is used to
lower the unwanted bias contributions (this step could even be bootstrapped with an estimate of the linear part coming from measurements with a lower input power). If one of the static nonlinearities is odd (e.g. $\varphi_{2n} = 0$), the unwanted terms in (4.6) disappear, and the RLDS is proportional to the underlying linear system.

### 4.3 The Estimation Procedure

The proposed method is an iterative process, where each iteration consists of two steps: identifying a parametric model for a linear system using the linear system identification algorithms (Ljung, 1999; Pintelon and Schoukens, 2001; Söderström and Stoica, 1989; Viberg, 1995) and solving a linear least squares problem to identify the static nonlinearities. The linear system is improved in the next iteration using the estimate of the static nonlinearities of the previous iteration, while the estimate of the parameters of the static nonlinearities depends on the parameters of the linear system identified in the same iteration.

The basic idea behind this proposal is the observation that the RLDS is close to the true linear system, and for a given linear part the estimation of the static nonlinearities is an easy task. Once the static nonlinearities are estimated, the internal signals $v$ and $w$ can be estimated. The rationale behind this step it that by estimating the internal signals, the impact of the nonlinearities on the linear estimation step is nearly eliminated.
This will reduce the unwanted bias and stochastic contributions, resulting in a better estimate of the linear part. Since it can be expected that a better estimate of the linear part yields an improved estimate of the static nonlinearity, the algorithm is made iterative.

4.3.1 Parameterization of the Static Nonlinearities

The nonlinearities are modeled linearly in the parameters so that their estimation reduces to a linear least squares estimation:

\[ v(t) = f(u(t)) = \sum_{l=1}^{q_1} \varphi_l f_l(u(t)) \] (4.7)

The weights of the basis functions are gathered in a vector to ease notations later on:

\[ \varphi = [\varphi_1, \varphi_2, \ldots, \varphi_{q_1}]^T \]

The basis functions in (4.7) can be freely chosen by the user as long as \( f(u) \) is linear in the parameters. Possible choices of \( f_l(u) \) include \( u^l \) or triangular functions with a priori fixed centers. With these triangular functions, it is possible to obtain a piecewise linear approximation of the nonlinear characteristic. The discussion about the choice of basis functions is similar to Section 3.5.1, hence it will not be repeated here.

The static nonlinearity \( g(w) \) is not identified directly, instead the inverse \( w = g^{-1}(y) \) is modeled to avoid a nonlinear optimization problem. The inverse of the output nonlinearity is parameterized using basis functions (which may be different from the ones used for the input nonlinearity \( v = f(u) \)):

\[ w(t) = g^{-1}(y(t)) = \sum_{l=1}^{q_2} \psi_l g_l(y(t)) \] (4.8)

As for \( \varphi \), the parameters are gathered in a vector:

\[ \psi = [\psi_1, \psi_2, \ldots, \psi_{q_2}]^T \]

To make this parameterization sensible, an assumption needs to be made:

**Assumption 6.1** \( g \) is invertible in \([\min y, \max y]\).
4.3.1.1 Parameterization of the linear part

The linear part is parameterized as a rational form with \( p \) parameters:

\[
S_\theta(e^{j\omega}) = \frac{\sum_{l=0}^{n_B} b_l e^{j\omega l}}{1 + \sum_{l=1}^{n_A} a_l e^{j\omega l}} = \frac{\sum_{l=0}^{n_B} \theta_l e^{j\omega l}}{1 + \sum_{l=n_B+1}^{p-1} \theta_l e^{j(l-n_B)\omega}}
\]

4.3.2 The Algorithm

When estimating a Hammerstein-Wiener system (or any block structured model with internal signals that cannot be measured), there can be several degenerations. These are due to the fact that the gain of the system can be arbitrarily distributed over the different building blocks: a Hammerstein-Wiener system with \( \tilde{f}(u) = \lambda f(u) \), \( \tilde{S}(\omega) = \mu S(\omega) \) and \( \tilde{g}(w) = g\left(\frac{w}{\lambda \mu}\right) \) is the same as a Hammerstein-Wiener system with \( f(u) \) as first nonlinearity, \( S(\omega) \) as linear dynamic system and \( g(w) \) as output nonlinearity (Figure 4.4). This shows that input-output measurements allow to reconstruct the different blocks’ characteristic only up to an unknown scaling constant.

4.3.2.1 Estimating the Linear Part

At iteration \( k \), the estimates of the static nonlinearities of iteration \( k - 1 \) are used to estimate the signals \( v(t) \) and \( w(t) \). These two signals are then used to identify a linear model:

\[
\hat{W}^{[k-1]}(l) = \hat{S}^{[k]}(e^{j\omega}) \hat{V}^{[k-1]}(l)
\]

This identification can be done with any estimation method: in this case, frequency domain identification methods were used to take advantage of the periodic character of the excitation (arbitrary selection of the frequency band that will be used in the
identification step, as explained further), but the prediction error framework could have been used as well (Ljung, 1999).

This parametric estimation step is very important: because of errors in the estimate of the static nonlinearities, the estimates of \( v(t) \) and \( w(t) \) are disturbed by a small bias due to the errors in the estimates of the nonlinear parts and by stochastic nonlinear contributions that behave like noise. If a nonparametric model of the linear system is built by dividing the Fourier coefficients of the output by the Fourier coefficients of the input, the algorithm converges immediately to a biased estimate of the Hammerstein-Wiener system. The smoothing effect of the parametric estimation step of the linear system is indispensable for the convergence of the method: the effect of the stochastic contributions is removed (since the estimate of the linear system is computed with a consistent estimator (Pintelon and Schoukens, 2001)). The remaining bias contributions are lowered in the following substep, when the static nonlinearities are estimated.

\[
\hat{V}^{[k-1]}(l) = \mathcal{F} (\hat{v}^{[k-1]}(t)) = \mathcal{F} (\hat{f}^{[k-1]}(u(t)))
\]

\[
\hat{W}^{[k-1]}(l) = \mathcal{F} (\hat{w}^{[k-1]}(t)) = \mathcal{F} (\hat{g}^{[k-1]-1}(y(t)))
\]

\[
\hat{S}^{[k]}_{\text{NP}}(l) = \frac{\hat{W}^{[k-1]}(l)}{\hat{V}^{[k-1]}(l)}
\]

\[
\hat{\theta}^{[k]} = \arg \min_{\theta} \sum_{l} \left| S_{\theta}(e^{j\omega l}; \theta) - \hat{S}^{[k]}_{\text{NP}}(l) \right|^2
\]

\[
\hat{S}^{[k]}(e^{j\omega}) = S_{\theta}(e^{j\omega}; \hat{\theta}^{[k]})
\]

For the special case where \( k = 0 \), there are no previous estimates available: some starting values are needed. The following equations are an application of (2.3). As there is only one linear system in the Hammerstein-Wiener system, and since the gain is not important due to the gain degeneration, an estimate of the linear system inside the Hammerstein-Wiener system can be obtained with (2.3).

\[
\hat{S}^{[0]}_{\text{NP}}(l) = \frac{X(l)}{U(l)}
\]

\[
\hat{\theta}^{[0]} = \arg \min_{\theta} \sum_{l} \left| S_{\theta}(e^{j\omega l}; \theta) - \hat{S}^{[0]}_{\text{NP}}(l) \right|^2
\]

\[
\hat{S}^{[0]}(e^{j\omega}) = S_{\theta}(e^{j\omega}; \hat{\theta}^{[0]})
\]
The iterative algorithm is started with $\hat{S}^{[0]}(\omega)$ defined in (4.15). As explained in Section 4.2, this starting value is biased in certain cases. However, this fact has never prevented the algorithm to converge to the correct values in noiseless simulations on systems without model error and to good results on noisy measurements.

### 4.3.2.2 Estimation of the Static Nonlinearities

Once the first step of iteration $k$ is completed, the estimate of the static nonlinearities can be updated. This is done with a method similar to the one used in Crama and Schoukens (2001) for Hammerstein system identification.

The parameterization of the static nonlinearities (4.7), (4.8) has been chosen to be able to express an estimation problem that is linear in the parameters. The underlying idea is to compute the intermediate signal $w(t)$ in two different ways, once starting from $u$ (yielding $\hat{W}_u$) and once from $y$ (yielding $\hat{W}_y$). Firstly $\hat{W}_u$ is computed as a linear combination of the input filtered by the actual Hammerstein estimate $\hat{S}^{[k]}[f_i(u(t))]$:

$$\hat{W}_u(l) = \sum_{m=1}^{q_1} \hat{S}^{[k]}(e^{j\omega l}) \varphi_m \mathcal{F}\{f_m(u(t))\}$$  \hspace{1cm} (4.16)

Secondly, $\hat{W}_y$ is written as

$$\hat{W}_y(l) = \sum_{m=1}^{q_2} \psi_m \mathcal{F}\{g_m(y(t))\}$$  \hspace{1cm} (4.17)

Both (4.16) and (4.17) have been written in the frequency domain, once more to take advantage of the periodic character of the input signal: $\hat{W}_u$ can reliably represent $w(t)$ only in the frequency band where the model of the linear system was estimated. Outside this frequency band, the model $\hat{S}^{[k]}$ is extrapolated and hence unreliable. This filtering is (in the frequency domain for periodic excitations) reduced to the proper selection of the Fourier coefficients at the frequencies of interest.

Both expressions for $W(\omega)$ ((4.16) and (4.17)) are matched to each other by minimizing

$$e(\omega) = \hat{W}_y(\omega) - \hat{W}_u(\omega)$$  \hspace{1cm} (4.18)
in least square sense with respect to $\varphi$ and $\psi$:

$$V(\varphi, \psi) = \sum_{l=1}^{N} |e(\omega_l)|^2$$  \hspace{1cm} (4.19)

which results in a linear least squares problem.

This minimization has to be done with an additional constraint to avoid the trivial solution $\varphi_l = 0, \psi_l = 0$.

**Algorithm 7.** Iterative procedure to produce initial estimates of Hammerstein-Wiener systems

**Step 7.1** Estimate $\hat{S}_{\text{NP}}^{[0]}$ (4.14).

**Step 7.2** Estimate $\hat{S}^{[0]}$ (4.15).

**Step 7.3** Estimate the static nonlinearities by minimizing $V(\varphi, \psi)$ (4.19) subject to $\|\varphi\|^2 + \|\psi\|^2 = 1$.

**Step 7.4** $k \leftarrow 1$.

**Step 7.5** Estimate the intermediate signals $\hat{V}^{[k]}$ (4.9) and $\hat{W}^{[k]}$ (4.10), using $\hat{\varphi}^{[k-1]}$ and $\hat{\psi}^{[k-1]}$.

**Step 7.6** Estimate $\hat{S}_{\text{NP}}^{[k]}$ (4.11).

**Step 7.7** Fit a parametric model to $\hat{S}_{\text{NP}}^{[k]}$ to compute $\hat{\theta}^{[k]}$ and $\hat{S}^{[k]}$ (4.13) (normalized to a gain of 1 at a user-chosen frequency).

**Step 7.8** Estimate the parameters $\hat{\varphi}^{[k]}$ and $\hat{\psi}^{[k]}$ of the static nonlinearities by minimizing $V(\varphi, \psi)$ (4.19) subject to $\|\varphi\|^2 + \|\psi\|^2 = 1$.

**Step 7.9** If the convergence criterion is not met, $k \leftarrow k + 1$ and jump to Step 7.5.

The identification process iterates until a convergence criterion is met. There is no unique choice for this: a threshold on the variation of a cost function (e.g. the output error) or a maximum number of iterations that should not be exceeded are two possible examples.

**4.4 The Algorithm applied to Measurements**

This section presents the measurement setup used to test the proposed method on real-world data.
4.4.1 The Measurement Setup

A laboratory setup has been assembled, consisting of two custom-made static nonlinearities and a Brüel & Kjær Octave Filter set with a center frequency of 125Hz. The highest frequency component of the input spectrum in the experiments has been kept below 200Hz, to make sure that the nonlinearities would indeed be static. The input signal was given the special amplitude distribution shown in Figure 4.5, using the method described in Schoukens and Dobrowiecki (1998). This was done to get a better model of the input static nonlinearity $f$. When no special precautions are taken, the amplitude distribution function looks more like a Gaussian probability density function, resulting in a decreased model quality for extreme values of the input $u$.

The measurements were controlled by a computer using an arbitrary waveform generator (HP E1445A) and sampled with two Analog to Digital Converter (ADC) cards (HP E1437A). The sampling frequency was 9.765625kHz, well above the Nyquist frequency for the used excitation. One period of 65536 samples was measured.

4.4.2 Results

For the measured input and output signal, a first estimate of the measured Hammerstein-Wiener system was computed. Both static nonlinearities were modeled using a piecewise linear function. The 25 breakpoints were chosen automatically for equal support (Hagenblad, 1999). The interval between the minimum and the maximum value
4.4 The Algorithm applied to Measurements

Figure 4.6: Identification result (black line) for the input nonlinearity $f(u)$ of the measured Hammerstein-Wiener system compared to the true value (grey dots). The scale to the right of the plot is used to zoom in on the difference (grey line) between the true value and the model.

of the input of $f(u)$ was divided in bins in such a way that in each bin there would be the same number of input samples. This ensures that each parameter of the static nonlinearities is supported by the same number of measurements, which avoids numerical problems.

Figure 4.6, Figure 4.7 and Figure 4.8 show the results of the algorithm on the measurement data. On each of these plots, the full line represents the identified model and the dots represent the true values: on our laboratory setup, we were able to measure the intermediate signals $v(t)$ and $w(t)$ to test the identification scheme, they were not used during the identification. The gain degeneration mentioned in Section 4.3 was resolved by scaling the model in each plot in order to make the comparison between the true system and the model the easiest possible. These figures all show that the result is very good: the models predict the true systems very well. The biggest imperfections in Figure 4.6 come from the nonideal breakpoint selection: the true system is a piecewise linear function, but the breakpoints of the model do not fall on the breakpoints of the true system. This causes the model to cut the corners of the characteristic, as can be seen most clearly in Figure 4.6 for input values around $-0.1$. These results are obtained after only 4 iterations. After 4 iterations, the difference of the cost function (4.19) is less than 0.001dB (0.01%) and the iteration is stopped.
Figure 4.7: Comparison of the amplitude of the identification result (black line) for the linear system inside the measured Hammerstein-Wiener system to the true system (grey dots) together with the difference between both (black pluses).

Figure 4.8: Identification result (black line) for the output nonlinearity $g(w)$ of the measured Hammerstein-Wiener system compared to the true value (grey dots). The scale to the right of the plot is used to zoom in on the difference (grey line) between the true value and the model.
4.5 Comparison with other methods

Hammerstein-Wiener systems (or similar) have already been identified using overparameterization (Bai, 1998), in the context of blind system identification (Bai, 2002a) and as an optimization problem minimized with relaxation (Zhu, 2002). This chapter presents some differences with these previous papers: it exploits the possibilities of the frequency domain representation for perfect filtering. It is also easy to add a nonparametric frequency dependent weighting to the linear system estimation step (4.12) or to the nonlinear part estimation step (4.19).

The method was demonstrated on measurements for Hammerstein-Wiener systems. These measurements demonstrate that the iterative process converges to acceptable estimates even when there is noise (inherent to the measurement).

The system studied in Bai (1998) is not quite the same system as the Hammerstein-Wiener system definition used here. When Figure 2.5 and Figure 4.9 are compared, the position of the output nonlinearity $g$ is different: in Figure 4.9, $g$ is inside the feedback part of the linear system, whereas in Figure 2.5 $g$ follows the linear part. In Bai (1998), the parameterization is chosen to make the estimation error bilinear in the parameters. The cost function is then minimized by introducing a new parameter for each of the products of 2 original parameters. In a second step, the original parameter set is computed using an SVD. The disadvantage of this method is that because of the overparameterization, a huge amount of parameters needs to be estimated in the

![Figure 4.9: Nonlinear system studied in Bai (1998): $f$ and $g$ are static nonlinearities, $b_i$ and $a_i$ are the coefficients of the denominator and the numerator of a linear dynamic system and $z^{-1}$ is the delay operator.](image-url)
first step and some work is needed to adapt it to the Hammerstein-Wiener structure as defined in Figure 2.5. However, it can be adapted easily for recursive estimation.

Since Bai (2002a) deals with blind identification, an assumption about the excitation signals is made there: the input needs to be a zero-order hold sequence and the measurements of the output need to be oversampled.

The differences with Zhu (2002) are at a deeper level. The main difference stems from the fact that in Zhu (2002) one cost function containing all parameters is minimized using a relaxation technique (i.e. some parameters are kept fixed while the function is minimized with respect to others). Here, two different cost functions are minimized consecutively: one for the linear part (with standard linear system identification techniques) and another for the static nonlinearities. Another difference is the weighting of the cost function: in Zhu (2002), the cost function is multiplied with the denominator of the linear system’s transfer function (ARX modeling). Because this would overweight the frequency range where the denominator gets big (hence where the output measurements are small and thus the signal to noise ratio is worst in case of for example white noise distortions), a high order model is built first, followed by a model reduction step. Additionally, the method presented in Zhu (2002) is consistent for white noise $m$ entering the system at a very special place, shown in Figure 4.10. Algorithm 7 is not consistent.
4.6 Degenerated Hammerstein-Wiener systems

Wiener systems or Hammerstein systems are degenerated cases of a Hammerstein-Wiener system: one of the static nonlinearities defaults to the identity. In this section, Algorithm 7 is applied to the measurements from Chapter 3.

What is investigated here is the behavior of Algorithm 7 when the proposed model structure is much richer than needed. It appears that the algorithm is well-behaved: the static nonlinearities that are actually the identity functions are recognized as such. In both cases, the static nonlinearities will be modeled as piecewise linear functions with 25 breakpoints.

First, the Hammerstein case is studied in Figure 4.11, Figure 4.13 and Figure 4.12. As can be seen, the Hammerstein system structure appears clearly because $g(w) = w$ in Figure 4.12.

The same can be seen for the Wiener system that was also measured and identified in Chapter 3: Figure 4.14 shows that the input static nonlinearity is correctly estimated as the identity function, while Figure 4.15 and Figure 4.16 show that the Wiener system is correctly identified.
Figure 4.13: Estimated (black line) and true (grey dots) linear part of a Hammerstein system estimated with Algorithm 7. The difference between both is indicated with pluses.

Figure 4.14: Estimated (black line) and true (grey dots) static nonlinearity at the input of a Wiener system estimated with Algorithm 7. The scale to the right of the plot is used to zoom in on the difference (grey line) between the true value and the model.

Figure 4.15: Estimated (black line) and true (grey dots) trivial static nonlinearity at the output of a Wiener system estimated with Algorithm 7. The scale to the right of the plot is used to zoom in on the difference (grey line) between the true value and the model.
4.7 Conclusion

This chapter extends the work of Chapter 3 to another nonlinear structure: the Hammerstein-Wiener systems. The biggest difference with respect to Wiener systems or Hammerstein systems for Hammerstein-Wiener systems is the fact that the RLDS isn’t always equal to the linear system inside the nonlinear structure. However, the identification results show that the iterative approach of Algorithm 7 works. This algorithm is analyzed in Appendix 4.A. It has also been shown that Algorithm 7 can identify simpler structures as well, without using the additional freedom to fit the noise.

Figure 4.16: Estimated (black line) and true (grey dots) linear part of a Wiener system estimated with Algorithm 7. The difference between both is indicated with pluses.
4. A Analysis of the iterative procedure

In the noise-less case and when there are no model errors, the iterative method presented in Algorithm 7 converges to the true value. The analysis is very similar to the convergence proof in Appendix 3.B. Section 4.A.1 presents the necessary formal background, while Section 4.A.2 is the kernel of the analysis.

4.A.1 Notations and assumptions

The notations are summarized in Table 4.1. Some assumptions are necessary to analyze the convergence of the iterative method:

Assumption 7.1 There are no modeling errors:

\[ \exists \varphi_0, \theta_0, \psi_0 : \text{diag}(S_\theta(e^{j\omega}; \theta_0)) A(U) \varphi_0 = B(Y) \psi_0 + Y \]

with \( A(U) \) and \( B(Y) \) defined in Table 4.1.

Assumption 7.2 The excitation \( u(t) \) is a random phase multisine with power \( P \) (independently of \( N \)), distributed over a sufficiently large number of frequencies \( N \) in the frequency band of interest to ensure a persistent excitation (even for \( N \to \infty \)).

Assumption 7.3 The models of the static nonlinearity are linear in their parameters and each of the basis functions, when excited with a random phase multisine (see Definition 1.4) has a random phase multisine as output. An example of such basis functions are the Fourier coefficients of \( u^n \).

Assumption 7.4 The number of excitation lines \( N \) is much higher than the number of parameters: \( N \gg (q_1 + p + q_2) \). The number of parameters \( (q_1, p \text{ and } q_2) \) is kept constant when the number \( N \) of excitation lines in the spectrum increases.

Assumption 7.5 \( \forall k \), the model estimated for \( \hat{f}^{[k]} \) is invertible\(^1\) over its input range.

Assumption 7.6 The true output static nonlinearity \( g_0 \) is invertible.

Assumption 7.7 The algorithm starts by computing a consistent parametric estimate of the best linear approximation of the Hammerstein-Wiener system. This estimate should be proportional to the underlying linear system \( S_0 \), but may be biased (4.6). It is assumed that this error is small enough (see Section 4.2 and Section 4.A.2.2).

\(^1\)This assumption is not necessary to apply Algorithm 7, but only to analyze its convergence.
Assumption 7.8 The inverse of the output static nonlinearity $g$ can be described by
$W = Y + B(Y) \psi_0$.

Assumption 7.9 The following matrix product is well-conditioned and invertible $\forall k$, $\forall N$ ($\infty$ included):

$\hat{C}^{[k],H} \hat{C}^{[k]}$

An example of an excitation satisfying Assumption 7.2 is (4.20), though in the most general case, all Fourier coefficients of $u(t)$ need not have the same amplitude.

$$u(t) = \sum_{l=1}^{N} \sqrt{\frac{2P}{N}} \cos(l \omega_0 t + \phi_l)$$

(4.20)

Assumption 7.6 makes sure that at least one of the static nonlinearities is not even in the domain where it has been excited. Otherwise, the RLDS (defined in Section 4.2) is 0 or contains only the unwanted bias contribution discussed in Section 4.2, (4.6).

Assumption 7.8 is not as restrictive as it may seem. It simply introduces a constraint similar to $\|\hat{\phi}^{[k]}\|^2 + \|\hat{\psi}^{[k]}\|^2 = 1$ (Step 7.8). By forcing one of the parameter values to 1, the trivial solution $\hat{\phi}^{[k]} = 0$, $\hat{\psi}^{[k]} = 0$ is avoided. The coefficient of any other basis function could have been chosen.

**4.A.2 The Analysis**

A complete iteration step will be studied to express the relationship between the estimation error $\hat{\delta}^{[k]}$ and $\hat{\delta}^{[k+1]}$. It will be seen that the drop in estimation error is obtained in Step 7.7 where a parametric transfer function model is fit to $\hat{S}^{[k]}_{NP}$. The analysis may be summarized as follows:

1. (4.24) in Section 4.A.2.1 shows that Step 7.8 at iteration $k$ keeps the estimation error of $\hat{\phi}^{[k]}$ and $\hat{\psi}^{[k]}$ at the same order of magnitude as the estimation error of $\hat{\theta}^{[k]}$.

2. (4.31) in Section 4.A.2.2 shows that when a parametric model is fit to $\hat{S}^{[k+1]}_{NP}$ in Step 7.7 at iteration $k + 1$, the estimation error of $\hat{\theta}^{[k+1]}$ is reduced.

The conclusion of these 2 steps is that $\hat{\delta}^{[k]} \rightarrow 0$, $\hat{\delta}^{[k]} \rightarrow 0$ and $\hat{\delta}^{[k]} \rightarrow 0$ as the error does not increase in step 1 and is reduced in step 2.
Table 4.1: Notations used in the convergence analysis

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>number of frequencies that have been measured.</td>
</tr>
<tr>
<td>$q_1$</td>
<td>number of parameters used to describe the static nonlinearity at the input.</td>
</tr>
<tr>
<td>$p$</td>
<td>number of parameters used to describe the linear system.</td>
</tr>
<tr>
<td>$q_2$</td>
<td>number of parameters used to describe the static nonlinearity at the output.</td>
</tr>
<tr>
<td>$\hat{\phi}^{[k]}$</td>
<td>estimate of the parameters of the static nonlinearity at the input ($f$) at iteration $k$. This is a $q_1 \times 1$ vector.</td>
</tr>
<tr>
<td>$\hat{\theta}^{[k]}$</td>
<td>estimate of the parameters of the linear system ($S$) at iteration $k$. This is a $p \times 1$ vector.</td>
</tr>
<tr>
<td>$\hat{\psi}^{[k]}$</td>
<td>estimate of the parameters of the static nonlinearity at the output ($g$) at iteration $k$. This is a $q_2 \times 1$ vector.</td>
</tr>
<tr>
<td>$S_0(\omega)$</td>
<td>true transfer function of the linear system inside the Hammerstein-Wiener system.</td>
</tr>
<tr>
<td>$\hat{S}^{[k]}(\omega)$</td>
<td>parametric estimate of the linear part at iteration step $k$: $\hat{S}^{[k]}(\omega) = S_\theta(\omega; \hat{\theta}^{[k]})$.</td>
</tr>
<tr>
<td>$\text{diag}(\hat{S}^{[k]})$</td>
<td>$N \times N$ diagonal matrix built from the $N \times 1$ vector $\hat{S}^{[k]}(\omega)$.</td>
</tr>
<tr>
<td>$A(U)$</td>
<td>$N \times q_1$ matrix containing the contributions of the basis functions of the static nonlinearity at the input, such that $A(U) \varphi_0 = V$.</td>
</tr>
<tr>
<td>$B(Y)$</td>
<td>$N \times q_2$ matrix containing the contributions of the basis functions of the inverse of the static nonlinearity at the output, such that $Y + B(Y) \psi_0 = W$.</td>
</tr>
<tr>
<td>$\hat{C}^{[k]}$</td>
<td>$N \times (q_1 + q_2)$ matrix containing the columns of $A(U)$ and $B(Y)$: $\hat{C}^{[k]} = \begin{bmatrix} \text{diag}(\hat{S}^{[k]}) \cdot A(U); -B(Y) \end{bmatrix}$.</td>
</tr>
<tr>
<td>$\hat{\eta}^{[k]}$</td>
<td>$(q_1 + q_2) \times 1$ vector containing the parameters $\hat{\phi}^{[k]}$ and $\hat{\psi}^{[k]}$: $\hat{\eta}^{[k]} = \text{stack}(\hat{\phi}^{[k]}, \hat{\psi}^{[k]})$.</td>
</tr>
<tr>
<td>$\hat{\delta}_{\phi}^{[k]}$</td>
<td>estimation error of $\hat{\phi}^{[k]} = \varphi_0 + \hat{\delta}_{\phi}^{[k]}$.</td>
</tr>
<tr>
<td>$\hat{\delta}_{\theta}^{[k]}$</td>
<td>estimation error of $\hat{\theta}^{[k]} = \theta_0 + \hat{\delta}_{\theta}^{[k]}$.</td>
</tr>
<tr>
<td>$\hat{\delta}_{\psi}^{[k]}$</td>
<td>estimation error of $\hat{\psi}^{[k]} = \psi_0 + \hat{\delta}_{\psi}^{[k]}$.</td>
</tr>
</tbody>
</table>
4.A Analysis of the iterative procedure

4.A.2.1 Propagation of an estimation error of the linear system to the estimate of the static nonlinearities

Consider Step 7.5 of iteration \(k\): the estimation error \(\hat{\delta}^{[k]}_{\theta}\) will introduce an error into \(\hat{W}^{[k]}_u\). Supposing that the error \(\hat{\delta}^{[k]}_{\theta}\) is small (Assumption 7.7), the error on the estimate of the linear system’s transfer function can be expressed as:

\[
\hat{\delta}^{[k]}_{\theta} = \frac{\partial S^{[\theta]}_g}{\partial \hat{\delta}^{[k]}_{\theta}}
\]  
(4.21)

The parameters of the static nonlinearity are estimated as defined in Step 7.8, taking into account Assumption 7.8:

\[
\hat{\varphi}^{[k]}, \hat{\psi}^{[k]} = \arg \min_{\varphi, \psi} \| \text{diag}(\hat{S}^{[k]}_g) \cdot A(U) \cdot \varphi - Y - B(Y) \cdot \psi \|^2
\]

Using the notations from Table 4.1, this equation can be rewritten as

\[
\hat{\eta}^{[k]} = \arg \min_{\eta} \| \hat{C}^{[k]}_{\eta} \hat{\eta}^{[k]} - Y \|^2
\]  
(4.22)

This least squares problem is solved in the usual way:

\[
\hat{\eta}^{[k]} = \text{Re} \left( \hat{C}^{[k],H} \hat{C}^{[k]} \right)^{-1} \text{Re} \left( \hat{C}^{[k],H} Y \right)
\]

where \(\text{Re}(\alpha + j\beta) = \alpha\) computes the real part of a complex number.

By writing \(\hat{\eta}^{[k]} = \eta_0 + \hat{\delta}^{[k]}_{\eta}\) in (4.22), the estimation error of the parameters of the static nonlinearity can be obtained. The following expression is linked with the estimation error of the linear part with (4.21):

\[
\eta_0 + \hat{\delta}^{[k]}_{\eta} = \text{Re} \left( \hat{C}^{[k],H} \hat{C}^{[k]} \right)^{-1} \text{Re} \left( \hat{C}^{[k],H} Y \right)
\]

\[
\Leftrightarrow \text{Re} \left( \hat{C}^{[k],H} \hat{C}^{[k]} \right) \hat{\delta}^{[k]}_{\eta} = \text{Re} \left( \hat{C}^{[k],H} \left( Y - \hat{C}^{[k]} \eta_0 \right) \right)
\]

\[
\Leftrightarrow \hat{\delta}^{[k]}_{\eta} = \text{Re} \left( \hat{C}^{[k],H} \hat{C}^{[k]} \right)^{-1} \text{Re} \left( \hat{C}^{[k],H} \left( Y - \hat{C}^{[k]} \eta_0 \right) - \left[ \text{diag}(\hat{\delta}^{[k]}_{S}) \cdot A(U) ; 0_{N \times q_2} \right] \eta_0 \right)
\]

\[
\Leftrightarrow \hat{\delta}^{[k]}_{\eta} = -\text{Re} \left( \hat{C}^{[k],H} \hat{C}^{[k]} \right)^{-1} \text{Re} \left( \hat{C}^{[k],H} \text{diag}(\hat{\delta}^{[k]}_{S}) \cdot A(U) ; 0_{N \times q_2} \right) \eta_0
\]

\[
\Leftrightarrow \hat{\delta}^{[k]}_{\eta} = -\text{Re} \left( \hat{C}^{[k],H} \hat{C}^{[k]} \right)^{-1} \text{Re} \left( \hat{C}^{[k],H} \text{diag}(\hat{\delta}^{[k]}_{S}) \cdot A(U) \cdot \varphi_0 \right)
\]

\[
\Leftrightarrow \hat{\delta}^{[k]}_{\eta} = -\text{Re} \left( \hat{C}^{[k],H} \hat{C}^{[k]} \right)^{-1} \text{Re} \left( \hat{C}^{[k],H} \text{diag}(\hat{\delta}^{[k]}_{S}) \cdot A(U) \cdot \varphi_0 \right)
\]
\[ \Leftrightarrow \hat{\delta}_{\eta}^{[k]} = -\text{Re} \left( \hat{C}^{[k]H} \hat{C}^{[k]} \right)^{-1} \text{Re} \left( \hat{C}^{[k]H} \text{diag}(A(U) \varphi_0) \frac{\partial S_0}{\partial \theta} \hat{\delta}_{\eta}^{[k]} \right) \]

In this equation, the order of magnitudes may be introduced:

\[ \hat{\delta}_{\eta}^{[k]} = \text{Re} \left( \left( \frac{1}{\sqrt{N}} \right)^N O \left( \frac{1}{\sqrt{N}} \right)^N \right)^{-1} \text{Re} \left( \left( \frac{1}{\sqrt{N}} \right)^N \text{diag} \left( O \left( \frac{1}{\sqrt{N}} \right) \right) O(1)_{N \times p} \right) \hat{\delta}_{\theta}^{[k]} \quad (4.23) \]

\[ = O(1)_{(q_1+q_2) \times (q_1+q_2)} \left( \frac{1}{\sqrt{N}} \right) \left( \frac{1}{\sqrt{N}} \right)_{N \times (q_1+q_2)} O(1)_{N \times p} \hat{\delta}_{\theta}^{[k]} \quad (4.24) \]

In (4.23) the order of magnitude for \( \hat{C}^{[k]} \) comes from Assumption 7.3, while \( \frac{\partial S_0}{\partial \theta}^{[k]} \) is independent of \( N \). A similar computation with more comments can be found in Appendix 3.B. (4.24) shows that \( \hat{\delta}_{\eta}^{[k]} \) and \( \hat{\delta}_{\theta}^{[k]} \) are of the same order of magnitude as \( \hat{\delta}_{\eta}^{[k]} \), regardless of \( N \).

### 4.4.2.2 Influence of the estimation error of the static nonlinearities on the estimate of the linear system

The linear part at iteration \( k + 1 \) is estimated by guessing the input \( \hat{V}^{[k]} \) and the output \( \hat{W}^{[k]} \) of the linear part, using the estimates \( \hat{\varphi}^{[k]} \) and \( \hat{\psi}^{[k]} \).

\[ \hat{S}_{\text{NP}}^{[k+1]} = \frac{\hat{W}^{[k+1]}}{\hat{V}^{[k+1]}} = \frac{Y + B(Y) \hat{\psi}^{[k]}}{A(U) \hat{\varphi}^{[k]}} = S_0 + \hat{\delta}_{S,\text{NP}}^{[k+1]} \quad (4.25) \]

\[ \text{diag} \left( S_0 + \hat{\delta}_{S,\text{NP}}^{[k+1]} \right) \left( A(U) \varphi_0 + A(U) \hat{\delta}_{\varphi}^{[k]} \right) = Y + B(Y) \psi_0 + B(Y) \hat{\delta}_{\psi}^{[k]} \]

\[ \Leftrightarrow \text{diag} \left( \hat{\delta}_{S,\text{NP}}^{[k+1]} \right) A(U) \left( \varphi_0 + \hat{\delta}_{\varphi}^{[k]} \right) + \text{diag} \left( S_0 \right) A(U) \hat{\delta}_{\varphi}^{[k]} = B(Y) \hat{\delta}_{\psi}^{[k]} \]

\[ \Leftrightarrow \hat{\delta}_{S,\text{NP}}^{[k+1]} = \text{diag} \left( A(U) \hat{\varphi}^{[k]} \right)^{-1} \left( B(Y) \delta_{\psi}^{[k]} - \text{diag} \left( S_0 \right) A(U) \hat{\delta}_{\varphi}^{[k]} \right) \]

\[ \Leftrightarrow \hat{\delta}_{S,\text{NP}}^{[k+1]} = -\text{diag} \left( A(U) \hat{\varphi}^{[k]} \right)^{-1} C_0 \hat{\delta}_{\eta}^{[k]} \quad (4.26) \]

The relation between \( \hat{V}^{[k+1]} \) and \( \hat{W}^{[k+1]} \) has the same structure as a Hammerstein-Wiener system, as Figure 4.17 shows. Figure 4.17 is a Hammerstein-Wiener system if and only if \( f_0 \left( \hat{f}^{[k]-1} (\cdot) \right) \) and \( g^{[k]-1} (g_0 (\cdot)) \) are static nonlinearities. \( \hat{g}^{[k]-1} (g_0 (\cdot)) \) is
4.A Analysis of the iterative procedure

Figure 4.17: The values of $\hat{V}^{[k+1]}$ and $\hat{W}^{[k+1]}$ are linked together by a Hammerstein-Wiener system.

A static nonlinearity by definition, as it is the composition of two static nonlinearities. Assumption 7.5 is necessary, because if $\hat{f}^{[k]}(\cdot)$ was not invertible, then the mapping from $\hat{V}^{[k+1]}$ to $U$ (i.e. $\hat{f}^{[k]}(\cdot)$) would not be static, hence the relation between $\hat{V}^{[k+1]}$ and $V$ would not be static either.

This nature of the relation between $\hat{V}^{[k+1]}$ and $\hat{W}^{[k+1]}$ allows to study the properties of $\hat{\delta}_{S,NP}^{[k+1]}$: $\hat{\delta}_{S,NP}^{[k+1]}$ contains stochastic contributions and unwanted bias contributions (see Section 4.2). These are expected to drop as the iteration process goes on, as $f_0(\hat{f}^{[k]}(\cdot))$ and $g^{[k]}(g_0(\cdot))$ approach to identity. This will be formally shown in the next three paragraphs.

The order of magnitude of $\hat{\delta}_{S,NP}^{[k+1]}$ can be estimated by inserting (4.24) into (4.26):

$$\hat{\delta}_{S,NP}^{[k+1]} = \text{diag} \left( O \left( \frac{1}{\sqrt{N}} \right) \right)^{-1} O \left( \frac{1}{\sqrt{N}} \right) N \times (q_1 + q_2) \hat{\delta}_{\theta}^{[k]} = O(1)_{N \times p} \hat{\delta}_{\theta}^{[k]}$$ (4.27)

The stochastic contribution is given directly in (4.27) as a vector of uncorrelated random complex numbers of $O(\hat{\delta}_{\theta}^{[k]})$.

The unwanted bias contribution comes from terms where different elements of the vectors $\hat{\delta}_{\varphi}^{[k]}$ and $\hat{\delta}_{\psi}^{[k]}$ are multiplied as in (4.6). Hence its order of magnitude needs to be estimated separately:

$$\mu_2 \lambda_2 \lambda_1 \sum_{\omega_1} S(\omega - \omega_1) S(\omega_1) |U(\omega_1)|^2 = O \left( \hat{\delta}_{\psi}^{[k]} \right) O \left( \hat{\delta}_{\varphi}^{[k]} \right)^2 O \left( \frac{P}{\sqrt{N}} \right)^2 N$$

$$= O \left( \hat{\delta}_{\theta}^{[k]} \right)^3 O(P)$$ (4.28)

where $\lambda_i = O \left( \hat{\delta}_{\varphi}^{[k]} \right)$ are the coefficients of the input static nonlinearity of Figure 4.17 and $\mu_i = O \left( \hat{\delta}_{\psi}^{[k]} \right)$ are the coefficients of the output static nonlinearity of Figure 4.17.

In (4.28), the input power $P$ of the experiment is included to emphasize its importance in the bias term: if the bias term is too large, it can be reduced by repeating the
experiment with a smaller input power. Higher order unwanted bias terms will have a higher power of \( O\left(\hat{\delta}^{[k]}_\theta\right) \) and may be neglected.

**Impact on the parametric estimate \( \hat{S}^{[k+1]}_\theta \)** In Step 7.7, a parametric model is fit to \( \hat{S}^{[k+1]}_\theta \), by minimizing the cost function

\[
K = \sum_\omega |S_\theta(\omega; \theta) - \hat{S}^{[k+1]}_\theta|^2 = \sum_\omega |S_0 + \hat{\delta}^{[k+1]}_\theta - S_0 - \hat{\delta}^{[k+1]}_{S,\text{NP}}|^2
\]

The derivative of this cost function is zero at \( \hat{\theta}^{[k+1]} \):

\[
2\text{Re}\left( \frac{\partial S_\theta}{\partial \hat{\theta}^{[k+1]}} H \left( \frac{\partial S_\theta}{\partial \theta^{[k+1]}} \hat{\delta}^{[k+1]}_\theta - \hat{\delta}^{[k+1]}_{S,\text{NP}} \right) \right) = 0
\]

\[
\Rightarrow 2\text{Re}\left( \frac{\partial S_\theta}{\partial \theta^{[k+1]}} H \left( \hat{\delta}^{[k+1]}_\theta \right) \right) = 0
\]

\[
\Rightarrow \text{Re}\left( \frac{\partial S_\theta}{\partial \theta^{[k+1]}} H \hat{\delta}^{[k+1]}_\theta \right) = \text{Re}\left( \frac{\partial S_\theta}{\partial \theta^{[k+1]}} H \hat{\delta}^{[k+1]}_{S,\text{NP}} \right)
\] (4.29)

The results of (4.27), (4.28) can now be inserted into (4.29) to estimate the order of magnitude of the new estimation error \( \hat{\delta}^{[k+1]}_\theta \):

\[
\hat{\delta}^{[k+1]}_\theta = \text{Re}\left( \frac{\partial S_\theta}{\partial \theta^{[k+1]}} H \frac{\partial S_\theta}{\partial \theta^{[k+1]}} \right)^{-1} \text{Re}\left( \frac{\partial S_\theta}{\partial \theta^{[k+1]}} H \hat{\delta}^{[k+1]}_{S,\text{NP}} \right)
\]

\[
= \left( O(1)_{p \times N} O(1)_{N \times p} \right)^{-1} O(1)_{p \times N} \left( O\left( \hat{\delta}^{[k]}_\theta \right) + O\left( \hat{\delta}^{[k]}_{\theta} \right)^3 O(P) \right) \] (4.30)

\[
= O(N)^{-1} O\left( \sqrt{N} \right) O\left( \hat{\delta}^{[k]}_\theta \right) + O(N) O\left( \hat{\delta}^{[k]}_{\theta} \right)^3 O(P)
\]

\[
= O\left( \frac{1}{\sqrt{N}} \right) O\left( \hat{\delta}^{[k]}_\theta \right) + O\left( \hat{\delta}^{[k]}_{\theta} \right)^3 O(P)
\] (4.31)

The underlined term in (4.30) is a vector of length \( N \) with random elements (the stochastic contribution). This fact causes the matrix product to grow only as \( \sqrt{N} \) instead of \( N \). The convergence follows from (4.31) and Assumption 7.7: at each iteration, the estimation error of the linear part decreases if \( N \) is large enough (for the first term of the sum) and if the original solution was close enough to the true solution (for the second term of the sum).
5 Wiener-Hammerstein systems

Wiener-Hammerstein systems consist of two linear dynamic systems placed around a static nonlinearity. These models are difficult to identify due to the presence of two dynamic systems. Usually, a nonlinear estimation procedure is necessary to estimate the parameters of the different parts. This nonlinear estimation procedure needs good starting values to converge quickly and/or reliably to a global minimum. First, a method will be proposed to compute a first estimate based on one measurement record only. In a second step, these initializations will be used in a nonlinear search (i.e., one that optimizes the output error cost function directly).
5 Wiener-Hammerstein systems

5.1 Introduction

Wiener-Hammerstein systems model the nonlinearity as a series of fundamental building blocks: linear dynamic systems and a static nonlinearity. It is plain to see that the static nonlinearity introduces the nonlinear behavior, while the linear dynamic systems allow to model the memory that might be present in the system. Figure 2.6 shows the Wiener-Hammerstein structure as well as the symbols for the subsystems and the signal names used in this chapter.

There has been a lot of interest for the identification of the Wiener-Hammerstein structure. Nonlinear estimation procedures for Wiener-Hammerstein systems have already been proposed (Boutayeb and Darouach, 1995; Chen and Fassois, 1992). Both papers stress the importance of the choice of good initialization values. However, no indication is given how such initial values could be obtained. Other methods have been developed that identify the Wiener-Hammerstein structure without the need for initial values (Bershad et al., 2001; Korenberg and Hunter, 1986; Leith et al., 2003; Tan and Godfrey, 2002a, 2003; Vandersteen et al., 1997), however these methods suffer from some limitations like the large number of measurements to perform (Vandersteen et al., 1997) or the tractable model complexity (Bershad et al., 2001; Leith et al., 2003; Tan and Godfrey, 2002a, 2003). This chapter presents a method that tries to produce good starting values without too many restrictions on the excitation signal or the model complexity, using only one measurement record.

The estimation algorithm uses the same ideas as the previously defined algorithms. The problem is broken down into subproblems for which well known solutions exist: identifying linear systems and solving linear least squares problems. This goal is reached by keeping all parts of the nonlinear system constant except one. The RLDS gives some information about the linear systems $R$ and $S$ to start with. Then, by parameterizing the static nonlinearity with basis functions, estimate its parameters by solving a linear least squares problem. Afterwards, iterate so that improvements in the estimate of i.e. $R$ aid in the identification of the others. After the definition of the parameterization and miscellaneous other notations in Section 5.1.1, the formal definition of the identification algorithm is given after informally presenting the different steps in Section 5.2. However, identifying Wiener-Hammerstein systems is much more difficult than the previous systems, so the presentation of Section 5.2 is necessarily a very high-level view of the
identification method: the implementation details can be found in Section 5.3. The proposed identification method is demonstrated on a measurement example in Section 5.4. The convergence of this method has not been studied formally, but cannot be proven with assumptions similar to Assumption 2.1–2.7.

### 5.1.1 Parameterization

In the estimation process, the linear transfer functions \( R \) and \( S \) will be parameterized as follows (with respectively \( p_1 \) and \( p_2 \) parameters):

\[
R(l; \theta) = \frac{\sum_{a=0}^{n_\theta} \theta_a \Omega_l^a}{1 + \sum_{a=1}^{m_\theta} \theta_{a+n_\theta} \Omega_l^a}; \quad n_\theta + m_\theta + 1 = p_1
\]

with \( \Omega_l = e^{j2\pi \frac{l}{f_0}} \) in the discrete-time case or \( \Omega_l = j2\pi \frac{l}{f_0} \) in the continuous-time case. Likewise, \( S \) is parameterized as

\[
S(l; \zeta) = \frac{\sum_{a=0}^{n_\zeta} \zeta_a \Omega_l^a}{1 + \sum_{a=1}^{m_\zeta} \zeta_{a+n_\zeta} \Omega_l^a}; \quad n_\zeta + m_\zeta + 1 = p_2
\]

A parametric model of the RLDS will be needed as well. It is obtained as follows, with \( \hat{G}_R \) the result of fitting a rational transfer function model to \( \hat{G}_{NP} \):

\[
G(l; \xi) = \frac{\sum_{a=0}^{n_\xi} \xi_a \Omega_l^a}{1 + \sum_{a=1}^{m_\xi} \xi_{a+n_\xi} \Omega_l^a}; \quad n_\xi + m_\xi + 1 = p_1 + p_2
\]

\[
\hat{\xi} = \arg \min_{\xi} \sum_{l \in L_U} |G(l; \xi) U(l) - Y(l)|^2
\]

\[
\hat{G}_R(l) = G \left( l; \hat{\xi} \right)
\]

Using the ideas from Schoukens et al. (2003b), the cost function in (5.3) can even be weighted to include the uncertainty introduced by the noise. This will not be pursued here, as such a weighting is not straightforward for all steps that will be necessary to
identify Wiener-Hammerstein systems.

The static nonlinearity \( f \) will be approximated by estimating the weights \( \varphi_a \) of \( q \) basis functions \( f_a \):

\[
f_\varphi(v) = \sum_{a=1}^{q} \varphi_a f_a(v)
\]

The notation \( \hat{x} \) will be used to express “the estimate of \( x \”).

The proposed excitation signal, a random phase multisine, is defined in (1.2). The user can choose the amplitude spectrum of the excitation (\( U_l \) in (1.2)) and the fundamental frequency \( f_0 \). The phases \( \phi_l \) are taken from a random distribution such that \( E[e^{j\phi_l}] = 0 \). Actually, for this method, there should be some unexcited spectral lines in the excitation spectrum: \( \exists l \in [\min L_U, \max L_U], U_l = 0 \) so that the nonlinear behavior can be easily observed at the system output. This requirement is explained further and motivated in Section 5.3.2.

The algorithm that will be proposed depends heavily on the ability to separate the part of the spectrum where only stochastic nonlinear contributions are present from the part where also bias or linear contributions appear. Because the excitation signal is periodic, these filtering operations are simply implemented as the selection of the proper set of Fourier coefficients. For this purpose, different sets of coefficients that will be used will be defined. The most straightforward is \( L_U \):

\[
l \in L_U \iff U_l \neq 0
\]

There are three other sets, \( L_R \), \( L_f \), and \( L_S \), that define which Fourier coefficients are taken into account for the identification of respectively \( \hat{R} \), \( \hat{f} \) and \( \hat{S} \). The correct definition of these sets is essential for the proper working of this algorithm. The definition of their values and the rationale for these choices can be found in Section 5.3.

### 5.1.2 Degenerations in the Wiener-Hammerstein structure

When identifying block-oriented structures like the Wiener-Hammerstein structures, it is necessary to keep in mind the possible degenerations of the structure. These are due to the fact that the internal signals (\( v(t) \) and \( w(t) \) in this case) cannot be measured. This means that the blocks can be estimated only up to an unknown scaling constant and an unknown delay (Boyd and Chua, 1983). These two Wiener-
Hammerstein systems have the same output for the same input: \((R(l), f(v), S(l))\) and 
\((\alpha_1 R(l) e^{j\delta l}, \alpha_2 f\left(\frac{v(l)}{\alpha_1}\right), \frac{1}{\alpha_2} S(l) e^{-j\delta l})\) (with \(\alpha_1 \alpha_2 \neq 0\) and \(\delta\) an arbitrary delay).

### 5.2 The identification algorithm

The presented estimation algorithm is an iterative algorithm. In each iteration, a set of steps are executed. Each of these steps will be presented independently, followed by a formal definition of the algorithm that will relate all the steps to each other. To separate the concepts from some necessary implementation details, the following presentations refer to Section 5.3 as necessary. The parts of the algorithm are presented in the order that seems most natural for the explanations, but this order is different from the order in the iteration.

Each of the substeps acts as if while estimating one of the parts of the Wiener-Hammerstein system, the others are known: e.g. when estimating \(f\), \(R\) and \(S\) are kept constant. This is why the algorithm presented here is considered to yield only first estimates: the parameters of \(R\), \(f\), \(S\) are not optimized simultaneously (Boutayeb and Darouach, 1995; Chen and Fassois, 1992), and the presented algorithm is not a relaxation algorithm because each estimation step minimizes another cost function.

#### 5.2.1 Estimating \(f\)

Suppose that \(R\) and \(S\) are already known. Then \(f\) can be estimated easily by applying the ideas from Crama and Schoukens (2001) for the estimation of the static nonlinearity for Hammerstein systems.

Based on the parameterization of \(f\) (5.5), these signals that follow from the Wiener-Hammerstein structure (see Figure 2.6) can be defined:

\[
\hat{w}_a(t) = f_a\left(\mathcal{F}^{-1}(R(l)U(l))\right); 1 \leq a \leq q
\]
\[
\hat{W}_a(l) = \mathcal{F}(\hat{w}_a(t)); 1 \leq a \leq q
\]
\[
\hat{Y}_a(l) = S(l)\hat{W}_a(l); 1 \leq a \leq q
\]

The parameters \(\varphi\) of \(f\) are estimated by approximating \(Y(l)\) with linear combinations
of $\hat{Y}_a(l); 1 \leq a \leq q$:

$$\hat{\phi} = \arg \min_{\varphi} \sum_{l \in \mathcal{L}_f} \left| \sum_{a=1}^{q} \varphi_a \hat{Y}_a(l) - Y(l) \right|^2$$

(5.7)

In (5.7), the output is approximated in least squares sense by a weighted sum of basis Wiener-Hammerstein systems built with $R$, a basis function of the static nonlinearity and $S$. The advantage of fitting $\hat{Y}_a(l)$ to $Y(l)$ instead of fitting $\hat{W}_a(l)$ to $\hat{W}_Y(l) = Y(l)S^{-1}(l)$ is that the linear system is neither extrapolated nor inverted (Crama and Schoukens, 2001). The sum over the frequencies is restricted by selecting only that part of the spectrum where $\hat{Y}_a$ makes sense as $S$ is known only for a limited set $\mathcal{L}_f$ of frequencies: the excitation band $[\min \mathcal{L}_U, \max \mathcal{L}_U]$. The periodic character of the signals is used to filter these easily.

### 5.2.2 Estimating $S$

Supposing that $R$ and $f$ are known, $\hat{W}$ can be computed:

$$\hat{W}(l) = \mathcal{F} \left( f \left( \mathcal{F}^{-1} \left( R(l)U(l) \right) \right) \right)$$

As in (5.7), no inversion of a linear system is necessary. Now that the input $\hat{W}(l)$ and the output $Y(l)$ of $S$ are known, standard linear system identification techniques (Pintelon and Schoukens, 2001; Ljung, 1999) can be used to estimate $\hat{S}(l)$. When parameterized as in (5.2), this can be written as

$$\hat{\zeta} = \arg \min_{\zeta} \sum_{l \in \mathcal{L}_S} \left| S \left( l; \zeta \right) \hat{W}(l) - Y(l) \right|^2$$

(5.8)

$$\hat{S}(l) = S \left( l; \hat{\zeta} \right)$$

(5.9)

As in (5.3), the cost function that is minimized in (5.8) could be weighted. The weights can be computed even if only one measurement record is available (Schoukens et al., 2003b).

The fitting of $\hat{S}$ in (5.9) is done over a selected set of frequencies $\mathcal{L}_S$. The selection of this set is important to get proper results. A discussion of this can be found in Section 5.3.3. In the actual implementation of the algorithm, it appeared that this step needs to be damped. This topic is discussed in Section 5.3.4.
5.2.3 Estimating $R$

If $S$ is known, $R$ may be estimated with (2.6):

$$\hat{R}_{NP}(l) = \frac{G_R(l)}{S(l)}$$ \hspace{1cm} (5.10)

The proportionality factor $C(U, R)$ is not taken into account because of the gain degeneration (Boyd and Chua, 1983): to resolve it, both estimates of the linear part are normalized to a unity gain at a user-chosen frequency (see Step 8.5 and Step 8.7). Hence, (5.10) captures the shape of $R$, and the gain will be normalized afterwards, eliminating the influence of $C(U, R)$.

A parametric model $R(l; \theta)$ (5.1) is fit to $\hat{R}_{NP}$, again on a limited set of frequencies $\mathcal{L}_R$ whose choice is explained in Section 5.3.3:

$$\hat{\theta} = \arg\min_{\theta} \sum_{l \in \mathcal{L}_R} |R(l; \theta) - \hat{R}_{NP}(l)|^2$$ \hspace{1cm} (5.11)

$$\hat{R}(l) = R(l; \hat{\theta})$$ \hspace{1cm} (5.12)

The cost function that is minimized in (5.11) is not weighted and it is not obvious which weighting is to be used, as contrarily to (5.7) and (5.8), (5.11) is not based on an output simulation error.

5.2.4 Definition of the estimation algorithm

The three previous sections need to be tied together: each supposes that two of the three systems are known and estimates the remaining part. However, none of the parts are really known, so an iterative scheme is proposed where the improvement of the estimate of one part will improve the estimate of the other parts. The only problem that remains is starting up this process: following the signal flow in the Wiener-Hammerstein system, the estimation algorithm is started up by making a first guess of $R$.

Algorithm 8.

Step 8.1 Estimate $\hat{G}_R$ as in (5.4).

Step 8.2 Initialize $\hat{R}^{[1]}$:

$$\hat{R}^{[1]}_{NP}(l) = \sqrt{\hat{G}_R(l)}; l \in \mathcal{L}_R$$
\[ \hat{\theta}^{[1]} = \arg \min_\theta \sum_{l \in \mathcal{L}_R} \left| R(l; \theta) - \hat{R}^{[1]}_{NP}(l) \right|^2 \]
\[ \hat{R}^{[1]}(l) = R(l; \hat{\theta}^{[1]}) \]

The symbol \( \sqrt{\hat{G}_R(l)} \) is used loosely in the previous equations. Refer to Section 5.3.1 for more details.

**Step 8.3** Choose \( \hat{S}^{[0]}_{NP} \) to conform to (2.6):
\[ \hat{S}^{[0]}_{NP}(l) = \frac{\hat{G}_R(l)}{\hat{R}^{[1]}(l)} \]

**Step 8.4** Estimate \( \hat{f}^{[1]} \) with \( \hat{R}^{[1]} \) and \( \hat{S}^{[0]}_{NP} \) as in (5.7):
\[ W_a^{[1]}(l) = \mathcal{F}\left( f_a\left( \mathcal{F}^{-1}\left( \hat{R}^{[1]}(l)U(l) \right) \right) \right); 1 \leq a \leq q \]
\[ Y_a^{[1]}(l) = \hat{S}^{[0]}_{NP}(l)W_a^{[1]}(l); 1 \leq a \leq q \]
\[ \hat{\varphi}^{[1]} = \arg \min_\varphi \sum_{l \in \mathcal{L}_f} \left| \sum_{a=1}^q \varphi_a Y_a^{[1]}(l) - Y(l) \right|^2 \]

**Step 8.5** Estimate \( \hat{S}^{[1]} \) with (5.9) and normalize the gain:
\[ \hat{W}^{[1]}(l) = \mathcal{F}\left( \sum_{a=1}^q \hat{\varphi}^{[1]}_a f_a\left( \mathcal{F}^{-1}\left( \hat{R}^{[1]}(l)U(l) \right) \right) \right) \]
\[ \hat{\zeta}^{[1]} = \arg \min_\zeta \sum_{l \in \mathcal{L}_S} \left| S(l; \zeta) \hat{W}^{[1]}(l) - Y(l) \right|^2 \]
\[ \hat{S}^{[1]}(l) = S(l; \hat{\zeta}^{[1]}) \]

**Step 8.6** \( k \leftarrow 2 \).

**Step 8.7** Estimate \( \hat{R}^{[k]}_{NP} \) with (5.10), then \( \hat{R}^{[k]} \) with (5.12) and normalize the gain:
\[ \hat{R}^{[k]}_{NP}(l) = \frac{\hat{G}_R(l)}{\hat{S}^{[k-1]}(l)} \]
\[ \hat{\theta}^{[k]} = \arg \min_\theta \sum_{l \in \mathcal{L}_R} \left| R(l; \theta) - \hat{R}^{[k]}_{NP}(l) \right|^2 \]
\[ \hat{R}^{[k]}(l) = R(l; \hat{\theta}^{[k]}) \]

**Step 8.8** Estimate \( \hat{f}^{[k]} \) with \( \hat{R}^{[k]} \) and \( \hat{S}^{[k-1]} \) as in (5.7) (completely analogous with
5.3 Implementation details

The algorithm given in the previous section is only a very high level view of the estimation process. If this algorithm is to converge to a sensible solution, many different parameters of the algorithm have to be chosen correctly.

5.3.1 Initializing $\hat{R}^{[1]}$

To start up the proposed method, two of the three parts of the Wiener-Hammerstein system must be given a value, as each substep concerns itself with the identification of one part only of the Wiener-Hammerstein system. Since (2.6) ties together $R$ and $S$, it is natural to start with them as this simplifies the task of proposing a starting point of the method that satisfies (2.6). Two straightforward values for $\hat{R}^{[1]}_{NP}$ can be proposed: all-pass $\hat{R}^{[1]}_{NP}(l) = 1$ with $\hat{S}^{[0]}_{NP}(l) = \hat{G}_R(l)$ or the other way around, $\hat{R}^{[0]}_{NP}(l) = \hat{G}_R(l)$ and $\hat{S}^{[0]}_{NP}(l) = 1$.

However, both these choices are too radical: when assessing model quality by looking at the output model error, the frequencies where the output spectrum has a high amplitude carry the most weight. Hence getting a correct model in the pass band means that a big part of the problem has been solved already. As a first approximation, the bandwidth of the RLDS is the bandwidth of the linear part $R$ or $S$ that is the lowest. Hence, if the true bandwidths of $R$ and $S$ are very different, there is a 50% probability of getting it wrong. The most natural solution would be to suppose that both linear systems are equal:

\[
\begin{align*}
\hat{G}_R(l) &= \hat{R}^{[1]}_{NP}(l) \hat{S}^{[0]}_{NP}(l) \\
\hat{R}^{[1]}_{NP}(l) &= \hat{S}^{[0]}_{NP}(l) \\
\end{align*}
\]

\[
\iff \hat{R}^{[1]}_{NP}(l) = \hat{S}^{[0]}_{NP}(l) = \sqrt{\hat{G}_R(l)} \quad (5.13)
\]

Of course, computing the square root of a complex transfer function has to be done with caution. Otherwise, the result could be a transfer function with jumps in the
phase, due to phase wrapping, that could not be reproduced with a rational transfer function model. In the implementation, the phase and the amplitude of $\hat{G}_R$ are treated separately. The square root of the amplitude is computed trivially. However, the phase of $\hat{G}_R$ is unwrapped before dividing it by 2.

These three possible starting choices will be compared on a measurement example. The proposed method seems to be quite robust as the final result is independent of the starting value. The main difference between the three is the number of iterations necessary to reach the optimum.

### 5.3.2 Requirements for the input signal

Estimating Wiener-Hammerstein systems is a difficult task mainly because there are two linear systems. It is well known that two linear systems in cascade cannot be separated using only input and output measurements because linear systems are commutative. Hence, $R$ can be identified separately from $S$ only because of the presence of $f$. However, for moderate input amplitudes, the nonlinear contributions are usually small compared to the linear contributions. Therefore, wherever there are linear contributions in the output spectrum, the nonlinear contributions are hard to see. Furthermore, because of the random character of the stochastic nonlinear contributions, they can easily be mistaken for noise. In order to be able to separate $R$ from $S$, a part of the spectrum must be kept free from linear contributions, so that at these unexcited frequencies the nonlinear contributions dominate. This goal is achieved by using the degree of freedom introduced by random phase multisines: not every line on the frequency grid needs to be excited. A possible choice is discussed in Section 5.3.3.
5.3 Implementation details

5.3.3 Choosing the sets of frequencies

The respective cost functions that are minimized at each identification step (5.9), (5.7) and (5.12) are calculated at restricted sets of frequencies $\mathcal{L}_R$, $\mathcal{L}_f$, $\mathcal{L}_S$. These sets need not be equal.

It is advised to choose $\mathcal{L}_U$ (5.6) as \{1, 3, 9, 11, $\ldots$, $8n+1, 8n+3,$ $\ldots$\} to have a special-odd multisine (Vanhoenacker et al., 2001). For such a choice of $\mathcal{L}_U$, $\mathcal{L}_f$ should be defined as:

\[ l \in \mathcal{L}_f \iff \min \mathcal{L}_U \leq l \leq \max \mathcal{L}_U \]

This means that the complete range of the excited frequency band can be used to identify the nonlinear part $f$ of the Wiener-Hammerstein system. This achieves two goals: (5.7) uses as much information as is available, while the limitation $\min \mathcal{L}_U \leq l \leq \max \mathcal{L}_U$ ensures that $\hat{S}$ is not extrapolated.

The choice of $\mathcal{L}_R$ is a result of (5.10): where $G_R$ becomes small, it might be because $S$ is small at these frequencies. Hence, in (5.10), two small quantities are divided by each other, resulting in a poor quality estimate. To avoid this, $\mathcal{L}_R$ should only include those frequencies where $G_R$ lies above the noise floor. An example of this is given in Section 5.4.

The choice of $\mathcal{L}_S$ depends on several factors. As a general rule, $\mathcal{L}_S$ and $\mathcal{L}_U$ should be disjoint but cover the same frequency range:

\[ l \in \mathcal{L}_S \Rightarrow \begin{cases} \min \mathcal{L}_U \leq l \leq \max \mathcal{L}_U \\ l \notin \mathcal{L}_U \end{cases} \quad (5.14) \]

If $\mathcal{L}_U$ and $\mathcal{L}_S$ are not disjoint the linear parts $R$ and $S$ will not be separated correctly. For $l \in \mathcal{L}_U$, linear and nonlinear contributions are added in $Y(l)$. Usually, the linear contributions have a much bigger amplitude than the nonlinear contributions. Therefore, (5.8) becomes (5.15) as only the linear contributions matter.

\[
\hat{\zeta}^{[k]} = \arg \min_\zeta \sum_{l \in \mathcal{L}_S} \left| S \left( l; \zeta \right) \hat{W}^{[k]}(l) - Y(l) \right|^2 \\
= \arg \min_\zeta \sum_{l \in \mathcal{L}_S \setminus \mathcal{L}_U} \left| S \left( l; \zeta \right) \hat{W}^{[k]}(l) - Y(l) \right|^2 + \sum_{l \in \mathcal{L}_S \cap \mathcal{L}_U} \left| S \left( l; \zeta \right) \hat{W}^{[k]}(l) - Y(l) \right|^2
\]
\[
\text{arg min}_\zeta \sum_{l \in L_S \setminus L_U} \left| S(l; \zeta) \hat{W}^k(l) - Y(l) \right|^2 + \sum_{l \in L_S \cap L_U} \left| S(l; \zeta) \hat{R}^k(l)U(l) - Y(l) \right|^2 + \\
2 \text{Re} \left( \sum_{l \in L_S \cap L_U} \left( S(l; \zeta) \hat{R}^k(l)U(l) - Y(l) \right) \left( S(l; \zeta) \left( \hat{W}^k(l) - \hat{R}^k(l)U(l) \right) \right)^* \right) \\
\approx \text{arg min}_\zeta \sum_{l \in L_S \cap L_U} \left| S(l; \zeta) \hat{R}^k(l)U(l) - Y(l) \right|^2 \\
\Rightarrow \hat{S}^k(l) \approx \frac{G_R(l)}{\hat{R}^k(l)}
\]

This does not introduce any new information, as Step 8.7 is completely similar to (5.16). Hence, any combination of \( \hat{R}^k \) and \( \hat{S}^k \) such that \( \hat{R}^k(l) \hat{S}^k(l) = \hat{G}_R(l) \) appears to be a minimum. This underdetermined problem causes Algorithm 8 to jump around in the parameter space, without stabilizing on a plausible solution.

Supposing that a special-odd excitation is chosen, this means that \( L_S \) could contain all even multiples of the basic frequency \( f_0 \) and the \( 8n + 5 \) and \( 8n + 7 \) (these are odd) multiples of the same frequency. If no a priori knowledge is available about the static nonlinearity, it is necessary to leave room for both even and odd detection lines in the frequency band of interest. These detection lines serve a double purpose: to prove that the measured system is indeed nonlinear, and to allow the identification of the output linear part \( S \). However, the even and odd detection lines may not be mixed inside \( L_S \). So, if there are significant (with respect to the noise) even nonlinear contributions, \( L_S = \{2, 4, \ldots, 2n\} \) satisfying (5.14) is the best choice, as it produces the biggest set of frequencies at which to identify \( S(l) \). However, if there are no significant even nonlinear contributions, then \( L_S = \{5, 7, \ldots, 8n + 5, 8n + 7\} \) also satisfies (5.14).

### 5.3.4 Damping the transition from \( \hat{S}^{[k-1]} \) to \( \hat{S}^k \)

In Step 8.9, a parametric transfer function model is fit to \( \hat{S}^{[k]}_{NP}(l) \):

\[
\hat{S}^{[k]}_{NP}(l) = \frac{Y(l)}{W^k(l)}; l \in L_S \\
\hat{\zeta}^{[k]} = \text{arg min}_\zeta \sum_{l \in L_S} \left| S(l; \zeta) - \left( \alpha \hat{S}^{[k-1]}_{NP}(l) + (1 - \alpha) \hat{S}^{[k]}_{NP}(l) \right) \right|^2
\]
The value of $\alpha$ is chosen so as to minimize the output simulation error. The implementation does not do a perfect optimization with respect to $\alpha$ but scans a small set of values between 0.05 (next to no damping) and 0.95 (strong damping). Because this is very time intensive, this damping is not applied at all iterations. It is expected that in the beginning of the iteration process, the algorithm makes wild jumps that should not be damped ($\alpha = 0$) because the initial values provided to Algorithm 8 (5.13) cannot be expected to be close to the true values. Whenever the output simulation error decreased at iteration $k - 2$ and $k - 1$, the different values of $\alpha$ are tried out to try and keep Algorithm 8 close to the found solution.

5.4 Measurement example

5.4.1 Setup and parameterization

The measurements were done as in Crama and Schoukens (2001). The only change was the Device Under Test (DUT): a Wiener-Hammerstein system was measured instead of a mere Wiener system or Hammerstein system. Besides the input and output of the DUT, two other signals were collected for validation purposes: $v$ and $w$ were measured also, but not used during the identification process.

For this particular DUT, there was a priori knowledge about the static nonlinearity: there were significant even contributions seen while writing Crama and Schoukens (2001). For this reason, an odd multisine was selected instead of a special-odd multisine to have a better frequency resolution in the same measurement time:

$$L_U = \{1, 3, \ldots, 2n + 1, \ldots, 523\}$$

The sampling frequency was 156250kHz and 8192 points were sampled (after anti alias filtering) per period, resulting in a frequency resolution $f_0 \approx 19\text{Hz}$. The excitation defined by $L_U$ goes from 0 to 10kHz, the rest of the spectrum is used as headroom for the nonlinear effects: due to the nonlinear character of the DUT, energy appears outside the excitation band. In order to avoid aliasing during the calculations, this headroom is necessary, though the signal could have been resampled after the measurement using classical signal processing techniques to create the headroom (Crochiere and Rabiner,
The RLDS $\hat{G}_R$ can be seen in Figure 5.2. As the amplitude of the RLDS drops sharply around 6kHz, $L_R$ was adapted to avoid this frequency band:

$$L_R = \{1, 2, \ldots, 288, 289, 319, 320, \ldots, 522, 523\}$$

Translated into frequencies, it means that $L_R$ covers the bands from 19Hz to 5.5kHz and from 6kHz to 10kHz. The even lines were used for $L_S$ since they were not part of the excitation spectrum $L_U$ (see Section 5.3.3):

$$L_S = \{2, 4, \ldots, 522\}$$

Both linear parts have been modeled using a rational transfer function model with three zeros and three poles ($n_\theta = n_\zeta = 3$, $m_\theta = m_\zeta = 3$) and the static nonlinearity was approximated by a 9th order polynomial ($q = 9$, $f_a(x) = x^9$).

### 5.4.2 Results of the initialization algorithm

Each of the identified systems has been compared with its true value which was obtained by measuring $v$ and $w$. This comparison is not easy because of the degeneration, hence Figure 5.3 and Figure 5.4 show the amplitudes after scaling to make a comparison easier.

The proposed method was tried with the three different starting values for $R_{NP}^{[0]}$. All
Figure 5.3: Best estimate of $\hat{R}$ (black line) compared to the true value $R_0$ (grey dots). The difference $|R_0 - \hat{R}|$ is shown as a grey line.

Figure 5.4: Best estimate of $\hat{S}$ (black line) compared to the true value $S_0$ (grey dots). The difference $|S_0 - \hat{S}|$ is shown as a grey line.
Figure 5.5: The true nonlinearity (grey dots) compared with the best model (black line)

Figure 5.6: Value of the output simulation error for three different starting points of the proposed method: $\sqrt{\hat{G}_R}$ (black line), $\hat{G}_R$ (broken black line), 1 (grey line)
trials converged to the same identified value. However, they did so at different speeds, as Figure 5.6 shows. In this particular example, S had the lowest bandwidth, leading to a slower convergence for the method started with $R_{\text{NP}}^{(0)} = \hat{G}_R$. Also, it appears that Algorithm 8 after reaching a minimum value for the simulation error, finally settles for a larger simulation error. Because this is an initialization algorithm, Algorithm 8 allows the cost function to increase, to allow for the presence of local minima. It does not make sense for an initialization to try to be too conservative and get stuck too close to $\hat{R}^{(0)}$, $\hat{f}^{(0)}$ and $\hat{S}^{(0)}$. Hence the implementation of the damping in Section 5.3.4 allows the cost function to grow rather than stopping Algorithm 8: the damping $\alpha$ is 0.95 at the most. If a value of 1 was allowed, Algorithm 8 would not exhibit this jump, but might also have stopped evolving much earlier, in a worse local minimum!

## 5.5 Comparison with existing methods

The proposed method generates starting values for nonlinear optimization schemes (Boutayeb and Darouach, 1995; Chen and Fassois, 1992), making a comparison with these works difficult as they do not exactly solve the same problem: Boutayeb and Darouach (1995); Chen and Fassois (1992) minimize one cost function (but may suffer from problems due to local minima) while this work focuses on breaking down the Wiener-Hammerstein estimation problem into simpler estimation problems (at the cost of minimizing three different cost functions instead of optimizing all parameters at once) to provide a starting value for such methods.

There are other methods that identify Wiener-Hammerstein systems (Bershad et al., 2001; Korenberg and Hunter, 1986; Tan and Godfrey, 2002a, 2003; Leith et al., 2003; Vandersteen et al., 1997). The advantage with respect to these methods is that the method works on systems that are not purely quadratic (Tan and Godfrey, 2002a) or cubic (Tan and Godfrey, 2003). Also, the use of random phase multisines in this method results in a higher number of frequencies at which $R$ is identified compared to Tan and Godfrey (2002a, 2003). The linear systems are not limited to a finite impulse response representation (Bershad et al., 2001; Leith et al., 2003) and one measurement record is sufficient for the identification (Vandersteen et al., 1997). The main difference with Korenberg and Hunter (1986) is that the output $w(t)$ of the static nonlinearity $f$ is not
Figure 5.7: A Wiener-Hammerstein system in the output error framework: the measured output $Y$ is disturbed by noise $m_y$. 

estimated by inverting the output linear part $S$. Also, some guidelines for the choice of the excitation signal and for starting values of the algorithm have been given.

5.6 Optimizing the global cost function

5.6.1 Output Error Framework

The model computed by Algorithm 8 is only an initialization for a nonlinear optimization (Boutayeb and Darouach, 1995; Chen and Fassois, 1992). The identification of a Wiener-Hammerstein system has been implemented in the output-error framework (see Figure 5.7), so the following cost function is minimized with respect to all parameters:

$$K = \frac{\max L_U}{\sum_{l=\min L_U} \left| \hat{S} \left( f \left[ \hat{R} U \right] \right) - Y(l) \right|^2}{\sigma^2_l(l)}$$ (5.17)

The derivatives of the cost function (5.17) are computed in Appendix 5.A. These derivatives have been used in a Levenberg-Marquardt algorithm (Fletcher, 1991) initialized with Algorithm 8.

This yields the following results: Figures 5.8, 5.9, 5.10, 5.11 and 5.12 compare the results of the Levenberg-Marquardt algorithm with the initial values (obtained from Algorithm 8). All these plots show that the initial values have been improved upon. In Figure 5.10, both traces are too close to each other to be seen clearly. In Figure 5.12, the phase error for both the initialization and the final results are huge around 6kHz, but this is not significant because the amplitude of the FRF is small at the same frequency.
5.6 Optimizing the global cost function

Figure 5.8: Amplitude of $R_0$ (grey dots) compared to $|R_0 - \hat{R}|$ for Algorithm 8 (grey line) and $|R_0 - \hat{R}|$ for the nonlinear estimator (black line).

Figure 5.9: Phase difference $|\angle R_0 - \angle \hat{R}|$ for Algorithm 8 (grey line) compared to $|\angle R_0 - \angle \hat{R}|$ for the nonlinear estimator (black line).

Figure 5.10: Static nonlinearity $f_0$ (grey dots) compared to $\hat{f}$ for Algorithm 8 (broken grey line) and $\hat{f}$ for the nonlinear estimator (black line).
Figure 5.11: Amplitude of $S_0$ (grey dots) compared to $|S_0 - \hat{S}|$ for Algorithm 8 (grey line) and $|S_0 - \hat{S}|$ for the nonlinear estimator (black line).

Figure 5.12: Phase difference $|\angle S_0 - \angle \hat{S}|$ for Algorithm 8 (grey line) compared to $|\angle S_0 - \angle \hat{S}|$ for the nonlinear estimator (black line).

Figure 5.13: A Wiener-Hammerstein system in the Errors in Variables framework.
5.6.2 Errors in variables Framework

The identification of Wiener-Hammerstein systems in the Errors in Variables (EIV) framework has been attempted. In the EIV framework, both the input and output is considered to be disturbed by uncorrelated noise (see Figure 5.13). Hence, the following cost function should be minimized with respect to the parameters $U_0(l)$, $\theta$, $\varphi$ and $\zeta$:

$$K = \max_{l=\min L_U} \left| U_0(l) - U(l) \right|^2 \sigma_U^2(l) + \left| \hat{S} \left( \hat{f} \left[ \hat{R}U_0 \right] \right) - Y(l) \right|^2 \sigma_Y^2(l)$$

(5.18)

However, this approach was abandoned because of its impracticality: (5.18) is highly nonlinear in $U_0(l)$. Hence these cannot be eliminated from (5.18) as they would be if the model of a linear system was considered (Pintelon and Schoukens, 2001). In conclusion, $K$ in (5.18) depends on a very high number of parameters, so that its minimization is impossible.

In Vandersteen et al. (1997), the identification of Wiener-Hammerstein systems in the EIV framework is made possible because a larger number of measurements are taken and the nonlinearity is parameterized such that the EIV cost function is quadratic in $U$. This makes it possible to eliminate $U$ analytically from the cost function (Cadzow, 1990; Golub and Le Veque, 1979; Pintelon and Schoukens, 2001).

5.7 Conclusion

A new method for the identification of Wiener-Hammerstein systems has been presented and tested on measurement data. This method has as main advantage that not too much data has to be collected. The result of the method has been used as an initialization for a nonlinear optimization of the parameters of the Wiener-Hammerstein system.
5 Wiener-Hammerstein systems

5.A The Jacobian

The nonlinear estimator presented in Section 5.6 uses the Levenberg-Marquardt algorithm (Fletcher, 1991) to minimize (5.17). For this, the Jacobian of the cost function must be computed. The Jacobian is a matrix containing \( p_1 + q + p_2 \) rows and as many columns as there are frequencies taken into account in (5.17).

The Jacobian will be computed in the following paragraphs, with an eye to make it easy to implement with Matlab. The expressions for the rows of the Jacobian are given supposing the parameterization used throughout this chapter:

\[
R(l; \theta) = \frac{\sum_{k=0}^{n_\theta} \theta_k \Omega^k_l}{1 + \sum_{k=1}^{m_\theta} \theta_{k+n_\theta} \Omega^k_l}
\]

\[
f(v; \varphi) = \sum_{k=1}^{q} \varphi_k f_k(v)
\]

\[
S(l; \zeta) = \frac{\sum_{k=0}^{n_\zeta} \zeta_k \Omega^k_l}{1 + \sum_{k=1}^{m_\zeta} \zeta_{k+n_\zeta} \Omega^k_l}
\]

The computation of these derivatives is straightforward, so only the final result will be given:

\[
\frac{\partial Y(l)}{\partial \theta_k} \bigg|_{0 \leq a \leq n_\theta} = S(l) \mathcal{F} \left( \frac{\partial f(v(t); \varphi)}{\partial v} \mathcal{F}^{-1} \left( \frac{\Omega^a_l}{1 + \sum_{k=1}^{m_\theta} \theta_{k+n_\theta} \Omega^k_l} \right) \right) |_l
\]

where \( \mathcal{F} (y) |_l \) means \( Y(l) \). Additionally, the products inside the \( \mathcal{F} (\cdot) \) and \( \mathcal{F}^{-1} (\cdot) \) operators are vectors:

\[
\mathcal{F}^{-1} (U \Omega^a) = \mathcal{F}^{-1} ([U(0) \Omega^a_0 U(1) \Omega^a_1 U(2) \Omega^a_2 \ldots]^T)
\]

The value of \( \mathcal{F}^{-1} (U \Omega^a) \) is a vector of the time domain values.

\[
\frac{\partial Y(l)}{\partial \theta_k} \bigg|_{a > n_\theta} = -S(l) \mathcal{F} \left( \frac{\partial f(v(t); \varphi)}{\partial v} \mathcal{F}^{-1} \left( \frac{\Omega^a_l}{(1 + \sum_{k=1}^{m_\theta} \theta_{k+n_\theta} \Omega^k_l)^2} \right) \right) |_l
\]

\[
\frac{\partial Y(l)}{\partial \varphi_k} = S(l) \mathcal{F} (f_a(v(t)))|_l
\]
\[ \frac{\partial Y(l)}{\partial \zeta_{0 \leq a \leq n_\zeta}} = W(l) \frac{\Omega_l^a}{1 + \sum_{k=1}^{m_\zeta} \zeta_{k+n_\zeta} \Omega_l^k} \]

\[ \frac{\partial Y(l)}{\partial \zeta_{a > n_\zeta}} = -W(l) S(l; \zeta) \frac{\Omega_l^a}{1 + \sum_{k=1}^{m_\zeta} \zeta_{k+n_\zeta} \Omega_l^k} \]
5 Wiener-Hammerstein systems
6 Fast Approximate Identification of Nonlinear Systems

This chapter builds on Chapters 3 and 5 to extend the classical identification methods for linear systems. The goal is to have a simple method that offers a better approximation of the DUT with little or no additional cost. A well chosen, general nonlinear model structure is proposed that is identified in a 2-step procedure. First, a best linear approximation is identified using the classical linear identification methods. In the second step, the nonlinear extensions are identified with a linear least squares method. The proposed model not only includes Wiener and Hammerstein systems, it is also suitable to model nonlinear feedback systems. The stability of the nonlinear model can be easily verified. The method is illustrated on experimental data.
6 Fast Approximate Identification of Nonlinear Systems

6.1 Introduction

The aim is to model nonlinear systems with a dominant linear behavior, the nonlinearities are considered to be (small) distortions. For that purpose block oriented nonlinear models are used as an alternative for the neural net approach, combining the power of the linear identification machinery with the needs to capture at least a significant part of the nonlinear distortions. Due to a well chosen model structure, the second step, that extends the linear model to include the nonlinear distortions, is almost for free because it consists of a linear least squares problem and no new experiments are needed. For that reason there is almost no cost to make a trial to identify the extension. If it turns out that we can reduce the remaining errors by a factor 2, 3 or more, a significant increase in model quality is obtained. If the trial fails, only a few seconds of additional computer time are lost. It should be clear that, eventually, it remains the responsibility of the user to accept or reject the final model.

The use of block-oriented models is not new (Bendat, 1998; Billings and Fakhouri, 1980, 1982; Billings and Tsang, 1990; Billings, 1980). Compared with these methods, the major contribution of the approach in this chapter is its simplicity. The identification of the nonlinear structure appears as a natural extension of the linear system identification approach. While it becomes possible to include a significant fraction of the nonlinear effects in a wide range of applications, the simplicity of the linear identification methodology is maintained.

In Section 6.2 we introduce and motivate the extended model and in Section 6.3 an identification procedure is proposed. After studying the properties of the method in Section 6.4, it is illustrated on a number of experiments, ranging from a nonlinear feedback structure, over a bidirectional chemical sensor, to a Hammerstein system.

6.2 Model Structure

Consider the SISO continuous or discrete time nonlinear dynamic system

\[ y_0(t) = g_{NL}(u(t)), \]
and the discrete observations (for a normalized sampling period):

\[ y(t) = y_0(t) + m(t), \quad t = 0, 1, \ldots, N \]

where \( m(t) \) is zero mean noise.

**Assumption 8.1** The input is assumed to be known exactly. \( m(t) \) is filtered white noise with finite moments up to order 4 and with absolutely summable cumulants up to order 4 (the “correlation” length of the noise is not too large).

These conditions are very similar to the classic time domain noise assumptions. We modified them a bit in order to include also the convergence results of the frequency domain framework.

The plant structure that will be used to describe the input-output relation for this system is given in Figure 6.1. As it can be seen, it consists of a number of parallel branches, each taking care of a typical nonlinear behavior. It can also be noticed that only one linear model block is allowed. The major reason for this choice is the need for fast and simple identification methods (see Section 6.3).

- The first branch (called linear branch) captures all linear dynamics. \( S \) is a rational form in \( s \) (continuous time systems), or in \( z^{-1} \) (discrete time systems), and \( c_{1,1} \) is a real number.

- The second branch (called Hammerstein branch) is a Hammerstein system. \( f_1 \) is
a static nonlinear system, and can be described using a simple polynomial model
\[ v_1 = \sum_{a=2}^{q_1} c_{1,a} u_1^a. \] The powers \( u_1^a \) can be replaced by a set of basis functions \( B_a(u_1) \) in order to get better numerical properties during the identification step. Note that the sum starts from 2 in order to avoid identifiability problems: \( a = 1 \) would result in the same contribution as the first branch.

- The third branch (called Wiener branch) is a Wiener system, where \( f_2 \) is again a static nonlinearity \( v_2 = \sum_{a=2}^{q_2} c_{2,a} u_2^a. \)

- The last branch has a special structure. It also consists of a static nonlinearity, but this time it is pre- and post filtered by the linear system \( S \). It is based on the \( p \)-th order inverse structure of a nonlinear system, and it turns out that it is very well suited to take care for nonlinear feedback dynamics as is explained in Appendix 6.A. The static nonlinearity is described by \( v_3 = \sum_{a=2}^{q_3} c_{3,a} u_3^a. \) Note that this time the linear term is included in the structure. This can be tolerated because no identifiability problem with the first branch will appear as long as \( S \) is a frequency dependent system. This branch is called the feedback branch.

Since we identify an approximate model, it is necessary to specify the class of input signals for which the approximation is valid. Moreover, the approximation criterion should be specified.

**Assumption 8.2** \( u(t) \) is assumed to be a random excitation with a user specified amplitude distribution (e.g. normal, binary, uniform) and power spectrum.

Note that more than one specification can be put into the set of excitations (e.g. uniformly and normally distributed signals). The wider the set, the more generally valid the model will be, but the lower the quality of the approximation.

**Approximation criterion:** The model parameters will be tuned in order to minimize the least squares distance between the measured and modeled output:

\[ V(\theta, y) = |y(t) - y(t, \theta)|^2 \]

For simplicity and without loss of generality, we do not add a frequency weighting to the approximation criterion. The reader should be aware that it is not obvious to make an optimal choice in the presence of nonlinear distortions. In this chapter a two step
procedure is proposed, and the optimal noise weighting differs in both steps as explained below.

6.3 Identification Procedure

Although the structure in Figure 6.1 looks quite complicated, it can be easily identified using a two step procedure, similar to Algorithms 1 and 4.

Algorithm 9.

Step 9.1 Identify the best linear approximation $\hat{G}_R(\Omega, \theta_{GR})$ of the nonlinear system using the classical linear identification procedures (Ljung, 1999; Pintelon and Schoukens, 2001; Söderström and Stoica, 1989). Note that from noise weighting point of view, a Box-Jenkins model would be the best for the time domain approach. The noise model will not only include the disturbing noise power spectrum, but also that of the unmodeled stochastic nonlinear distortions. A similar approach is possible in the frequency domain where the stochastic nonlinear distortions should be included in the noise model (Pintelon and Schoukens, 2001; Schoukens et al., 1998a). As mentioned before, we continue in this chapter with an unweighted least squares procedure. It can be noted that since these identification schemes can deal with unstable systems (with some precautions for the time domain algorithms), there is no need for $\hat{G}_R(\Omega, \theta_{GR})$ to be stable.

Step 9.2 Fix the linear block $S$ in Figure 6.1 everywhere to $\hat{G}_R(\Omega, \theta_{GR})$. Next identify the parameters in the nonlinear models. Note that this problem is linear-in-the-parameters.

Remark The two step procedure is not optimal, as it is known that it does not necessarily reach the global minimum. A nonlinear search, leaving all model parameters free, should be added. We do not consider this step here since the goal is to setup a fast method.

6.3.1 Identification of the best linear approximation

The aim of the first step is to identify the linear block $S(j\omega)$ (or $S(e^{j\omega})$ for discrete time systems). In Billings and Fakhouri (1980, 1982) and Schoukens et al. (1998a), it was
shown that the best linear approximation of a Hammerstein or Wiener system (branch 2 and 3), is asymptotically \( N \to \infty \) given within a constant by \( S(\Omega) \), if the excitation is normally distributed or a random multisine (Pintelon and Schoukens, 2001; Schoukens et al., 1998a). This result is generalized to a wider class of excitations by Billings and Fakhouri (1982). The fourth branch results in \( S(\Omega)^2 = S(\Omega)S(\Omega) \), the squared transfer function corresponding to the cascaded linear systems in the feedback branch. Hence the best linear approximation \( \hat{G}_R(\Omega, \theta_{GR}) \) of the overall structure in Figure 6.1 is asymptotically

\[
\hat{G}_R(\Omega, \theta_{GR}) = d_1 S(\Omega) + d_2 S(\Omega)^2
\]

with \( d_1 \) and \( d_2 \) two unknown real constants.

**Remarks**

- For the specified class of excitations, a nonlinear system can be replaced by a linear system plus a nonlinear noise source (Pintelon and Schoukens, 2001; Schoukens et al., 1998a). The identification of the parametric model \( \hat{G}_R(\Omega, \theta_{GR}) \) using the classical identification methods on the measured data \( u(t), y(t) \) is a smoothing step, that eliminates significantly the impact of the nonlinear noise on the best linear approximation.

- In practice we apply this approach also to excitation signals that do not belong to the specified class (normally distributed noise or random multisines). For such excitations, there is in general no proof available that the best linear approximation of the Wiener- and Hammerstein-branch is still proportional to \( S \). However, this is the best we can do within this framework.

- From (6.1) it follows that an error of order \( d_2 S^2 \) appears in addition to the desired result (which is \( d_1 S \)). However, in the second identification step (explained in Section 6.3.2), a first order correction (under Assumptions 8.3 and 8.4) of the error \( d_2 S(\Omega)^2 \) in (6.1) can be performed due to the presence of a linear term \( c_{3,1} S(\Omega)^2 \) in the third nonlinear branch.

**Assumption 8.3** The nonlinear output contributions are small compared to the linear output: define \( \max(|y_2|_2, |y_3|_2, |y_4|_2) = M_{NL} \) then \( M_{NL} \ll |y_1|_2 \).
This can be rephrased by noticing that the nonlinear distortions are a first order distortion:

\[
\max \left( \frac{|y_2|^2}{|y_1|^2}, \frac{|y_3|^2}{|y_1|^2}, \frac{|y_4|^2}{|y_1|^2} \right) = O \left( \frac{M_{NL}}{|y_1|^2} \right)
\]

with \( \frac{M_{NL}}{|y_1|^2} \ll 1 \).

**Assumption 8.4** if \( \left| \hat{S}(u) - S(u) \right| \_2 = O \left( \frac{M_{NL}}{|y_1|^2} \right) \), then

\[
\left| \hat{S} (f_1(u)) - S (f_1(u)) \right|_2 = O \left( \frac{M_{NL}}{|y_1|^2} \right)
\]

\[
\left| f_2 \left( \hat{S}(u) - f_2(S(u)) \right) \right|_2 = O \left( \frac{M_{NL}}{|y_1|^2} \right)
\]

\[
\left| \hat{S} \left( f_3 \left( \hat{S}(u) \right) \right) - S \left( f_3 \left( S(u) \right) \right) \right|_2 = O \left( \frac{M_{NL}}{|y_1|^2} \right)
\]

**Remarks**

1. This assumption is not too hard if the considered frequency band is well excited during the experiments, for example using white noise. However, if in some excited frequency bands, the nonlinear power would dominate the linear power due to a poor excitation (a frequency band is almost not excited), large errors can appear on a nonparametric estimate of \( G_{\text{NP}} \) due to the very poor signal to noise ratio. Parametric estimation methods that identify \( \hat{G}_R(\Omega, \theta_G) \) from the input \( u \) and output \( y \), like for example the output error method in Ljung (1999) or the frequency domain estimator in Pintelon and Schoukens (2001), have an intrinsic weighting in their cost function that is proportional to the power spectrum of \( u \), and consequently the critical frequency bands (that are almost not excited) are already suppressed by these methods.

2. Combining Assumption 8.4 with Assumption 8.3 results immediately in

\[
\left| \hat{S} (f_1(u)) - S (f_1(u)) \right|_2 = O \left( \frac{M_{NL}}{|y_1|^2} \right)^2
\]

\[
\left| f_2 \left( \hat{S}(u) - f_2(S(u)) \right) \right|_2 = O \left( \frac{M_{NL}}{|y_1|^2} \right)^2
\]
\[
\frac{|\hat{S}(f_3(\hat{S}(u))) - S(f_3(S(u)))|}{|y_1|} = O\left(\frac{M_{NL}}{|y_1|}\right)^2
\]

\subsection{6.3.2 Identifying the nonlinear model parameters}

Once the best linear approximation is identified, the parameters \(c_{i,a}\) in the nonlinear static models are estimated. Because the linear blocks are fixed to \(\hat{G}_R(\Omega, \theta_{GR})\), this is a simple least squares problem that is linear-in-the-parameters. The basic idea is to calculate for each parameter \(c_{i,a}\) the contribution to the output, order these as the columns in the matrix \(J\), and solve the linear least squares problem:

\[
[y] \approx J_t \theta_{NL} \quad \text{(time domain)}
\]

or

\[
[Y] \approx J_f \theta_{NL} \quad \text{(frequency domain)}
\]

with \(\theta_{NL} = [c_{1,1}, c_{1,2}, \ldots, c_{1,q_1}, \ldots, c_{3,1}, \ldots, c_{3,q_3}]^T\), and \([y], [Y]\) column vectors containing the time or frequency domain data. As an example, the detailed structure of \(J_t\) is given.

\[
J_t = [y_1, y_2, \ldots, y_{q_1}, y_2, 3, \ldots, y_{q_2}, 3, y_1, 4, \ldots, y_{q_3}, 4]
\]

with \(y_1\) a column vector containing the output of the linear branch, \(y_{a,2} = \hat{S}(u^a)\), \(y_{a,3} = \left(\hat{S}(u)^a\right)^a\), \(y_{a,4} = \hat{S}\left(\left(\hat{S}(u)^a\right)^a\right)\), where \(\cdot^a\) raises each element of the vector to the power \(a\).

\textbf{Remarks}

- Care should be taken to use robust numerical procedures to maintain good numerical conditioning in the second step.

- It is possible to add a noise weighting to the least squares problem to take for example the noise properties into account. However it should be clear that an ‘a priori’ estimated noise filter should be used that is fixed during the estimation step in order to keep ‘the-linear-in-the-parameters’ property that is the kernel of the
simplicity of the method. In the frequency domain, such fixed noise models can be easily obtained (Pintelon and Schoukens, 2001).

6.3.3 Model selection

Although the model selection problem at first seems very complicated, it is quite easily solved, due to the fact that a two step procedure is followed. In the first step, the order of the linear model $\hat{S}(\Omega)$ is selected using the classical order selection schemes. Next the active nonlinear branches should be selected. In all the examples below, we followed the same strategy. In a first run, all branches are switched on, using a nonlinearity of degree 5. Next the contribution of each branch to the output is verified (e.g. checking the output power of each branch, or by comparing the output power spectra). Next the dominating branches are selected, and the order of the static nonlinearity is tuned. It turned out that this procedure is very effective.

6.4 Impact of the Disturbing Noise

The impact of disturbing noise is explicitly studied in (Schoukens et al., 1998a) and in Chapter 3. There it is shown that under Assumption 8.1, the estimate $\hat{G}_R(\Omega, \theta_{G_R})$ of $S(\Omega)$ is consistent, even in the presence of nonlinear distortions, if the model $G_R(\Omega, \theta_{G_R})$ is flexible enough to capture the dynamics of $S(\Omega)$. The stochastic convergence of the 2nd step is straightforward (under the same noise conditions as the first step) since it is linear-in-the-parameters. The estimated parameters will converge to the parameters that would be found if there was no disturbing output noise present. If the nonlinear structure is rich enough to capture the nonlinear part of the system the true system will be found asymptotically, otherwise an approximation will be obtained.

6.5 Experimental Verification

In this section the results from three experiments are reported, illustrating the flexibility of the method with respect to different types of nonlinearity, and the robustness with respect to the amplitude distribution of the excitation signal.
6.5.1 Nonlinear feedback system

The first experiment is done on a nonlinear electrical circuit that is ideally described by the following nonlinear 2nd order differential equation:

\[ m \frac{d^2}{dt^2} y(t) + d \frac{d}{dt} y(t) + ay(t) = u(t) - by(t)^3 \]  \hspace{1cm} (6.2)

Of course the actual realized circuit is not in perfect agreement with (6.2), for example, we noticed in the measurements also the presence of a small quadratic term \( y(t)^2 \). A detailed study of this nonlinear system can be found in Pintelon and Schoukens (2001). In a first step, the system is identified using an odd random multisine (almost normally distributed periodic signal with only odd harmonics excited, see Pintelon and Schoukens (2001), next the identified model is validated using a normally distributed noise excitation with slowly increasing standard deviation.

In the first step, a continuous time linear model \( S \) of order 2 was selected and identified:

\[ S(s) = \frac{b_0}{a_0 + a_1 s + a_2 s^2} \]  \hspace{1cm} using 8192 data points. Next, we checked the output power for each branch as explained in Section 6.3.3. For simplicity we give the results in dB, normalizing...
6.5 Experimental Verification

![Graphs of signal and error over time](image)

Figure 6.3: Validation results: top: the measured output; bottom: grey: the linear simulated output error; black: the nonlinear simulated output error.

The linear output to 0dB. The following results were found: $P_{\text{Lin}} = 0\text{dB}$, $P_{\text{Ham}} = -55\text{dB}$, $P_{\text{Wien}} = -41\text{dB}$, $P_{\text{Fdbck}} = -24\text{dB}$. Only the feedback branch was retained in parallel with the linear model branch. A static nonlinearity of degree 5 is used.

In Figure 6.2, the identification result is shown. The measured output spectrum is compared to the modeled one, and this for the best linear approximation and the nonlinear model. The validation result (on 40700 data points) is shown in Figure 6.3. The measured output is again compared to the linear and nonlinear modeled output. From Figures 6.2 and 6.3, it is clearly seen that the errors drop with about a factor 10 when the linear model is replaced by the nonlinear one. Note that both models completely fail once the output is larger than 0.1 V, which was exactly the range that was used during the identification step. This shows that these approximate models, that are not built on physical principles, may not be extrapolated.

6.5.2 System with bidirectional dynamics

In the second experiment, an ‘electronic nose’ is modeled. The sensor is based on an adsorption/desorption process. Since both processes have different dynamics, the behavior of the system depends on the slope of the output. The system is excited with a binary signal (an inverse repeat periodic signal). Two data sets with different excitations were available. In each set 1260 data points are available. These data were obtained by
In order to show the flexibility of the modeling approach, we selected this time a
discrete time dynamic model for $S$ of order 5 (similar results were obtained with a
continuous time model). The scan of the nonlinear branches resulted in: $P_{\text{Lin}} = 0\text{dB},$
$P_{\text{Ham}} = \text{not identifiable with a binary input and put equal to zero,}$ $P_{\text{Wien}} = -27\text{dB},$
$P_{\text{Fdbck}} = -21\text{dB}$. From this, the Wiener branch and also the feedback branch were
selected. The degree of the static nonlinearity was tuned to 4 for both branches.

The first experiment was used to identify the model. The measured output is compared
to the identified one in Figure 6.4. From this figure it is seen that especially in the
lower frequency band a significant gain is made by switching from a linear model to a
nonlinear one. The validation results are shown in Figure 6.5. Also in this case it can be
clearly seen that the nonlinear model performs significantly better than the linear one.
Figure 6.5 shows that the peak error is 100mV, while the mean square error is of the
same order of magnitude as the 68.7mV obtained in Tan and Godfrey (2002b).
6.5.3 Hammerstein system

In a last experiment, we identify a Hammerstein device. A detailed description of the setup is given in Section 3.3.2 (Crama and Schoukens, 2001). The length of the data records was 2048 points. The first scan of the branches resulted in: \( P_{\text{Lin}} = 0\text{dB}, P_{\text{Ham}} = -14\text{dB}, P_{\text{Wien}} = -47\text{dB}, P_{\text{FdBck}} = -48\text{dB} \). Only the Hammerstein branch was retained in parallel with the linear model branch. The final degree of the static nonlinearity was chosen equal to 10.

The identification results are shown in Figure 6.6 and the validation results in Figure 6.7. Again a significant improvement is found by using the nonlinear model. It can also be observed that in this case, the model was also able to give a reasonable prediction of the output outside the excitation band (0-10kHz). It should be noted that the identified linear model (continuous time, order 3) is extrapolated in that case, since no direct information about the linear contribution in the band 10-30 kHz is available.

6.6 Conclusions

In this chapter, an approximate nonlinear modeling technique is proposed. The major goals of the method are: simple in use, fast to identify at a low experimental cost. Instead of going for the best possible nonlinear model, we explicitly preferred a reasonable
Figure 6.6: Comparing the identified output to the measured one. Grey pluses: $|Y - Y_{Lin}|$, black pluses: $|Y - Y_{NLin}|$, black circles: measured output.

Figure 6.7: Validation results: top: the measured output; bottom: grey: the linear simulated output error; black: the nonlinear simulated output error.
approximation that reduces the nonlinearity errors significantly, using a model that can be easily used during design, simulation, etc. To reach this goal, a nonlinear model structure that can be easily identified using a two step procedure is proposed. In the first step, the linear dynamics are estimated. In the next step the nonlinear branches in the model are identified. A simple model selection procedure allows to select the active nonlinear branches. The method allows to deal with Hammerstein, Wiener, and nonlinear feedback systems. Although it is not possible to give a proof for the general applicability of the method, our experience on a wide range of experiments coming from different fields showed the flexibility and the power of the method. Besides the results reported in this paper, we also applied the method successfully to model the nonlinear behavior of an operational amplifier (reduction of the errors with a factor 2-3), to approximate the friction of micro-movements (reduction of the error with a factor 2). Till now, we did not face a situation where the method failed completely. And even in this case, no real harm is made because only a few seconds of calculation time are wasted.
6.A Motivation of the feedback branch

Consider a nonlinear feedback system (Figure 6.8) with nonlinear systems $P$ and $Q$. The relation between the input and the output is then

$$ y = P[u - Q[y]] $$

or

$$ P^{-1}[y] = u - Q[y] $$

with

$$ P^{-1}[P[u]] = u $$

Rearranging this equation results in

$$ u = P^{-1}[y] + Q[y] = R[y] \quad (6.3) $$

or

$$ y = R^{-1}[u] \quad (6.4) $$

Using classical Volterra techniques (Schetzen, 1980, Sections 7.3 and 7.4), the third-order inverse approximation of $R^{-1}$ can be calculated:

$$ y \approx R_1^{-1}[u] + R_1^{-1}\left[ R_2 \left[ R_1^{-1}[u] \right] \right] + R_1^{-1}\left[ R_3 \left[ R_1^{-1}[u] \right] \right] + R_1^{-1}\left[ L_3 \left[ R_1^{-1}[u] \right] \right] \quad (6.5) $$

with $R_1^{-1}$, $R_2^{-1}$ and $R_3^{-1}$ the first, second and third order kernel of the inverse of the nonlinear system $R$, and $L_3$ a shorthand notation for a third order kernel applied to combinations of $R_1^{-1}[u]$. For the reader who is not familiar with Volterra theory, it can be noted that:
6.B Identification of the parallel structure

- $R_1^{-1}$ is a linear system. Linear systems have as property that for all scalars $k$:
  \[ R_1^{-1}[ku] = kR_1^{-1}[u]. \]
- $R_2$ is a second order Volterra system, and it has as property that $R_2[ku] = k^2R_2[u]$.
- $R_3$ and $L_3$ are third order Volterra systems, and they have as property that $R_3[ku] = k^3R_3[u]$.

The derivation of these equations can be found in Appendix 6.C.

The important observation that should be made here is that the nonlinear kernels in (6.5) have a linear pre- and post-filter $R_1^{-1} […]$. This is exactly the structure that is proposed in the feedback branch. The approximation that is made in the model structure of Figure 6.1 is that the dynamic nonlinear systems $R_2$, $R_3$, and $L_3$ are replaced by a static nonlinear system.

6.B Identification of the parallel structure

In this appendix we study the identification of the feedback structure under noiseless conditions (no disturbing noise present) and analyze the impact of the presence of the feedback branch on the identification procedure. The basic problem with this branch is that its best linear approximation is not proportional to $S$ but to $S^2$. At the end of the first step of the identification process (the linear approximation), the resulting best linear approximation $\hat{S}$ of the nonlinear system in Figure 6.1 is asymptotically given by

\[ \hat{S} = d_1S + d_2S^2 \]

and $\hat{S}$ will be used in the dynamic blocks during the second step where the coefficients of the static nonlinearities are estimated. This raises the question if a structure like that in Figure 6.1, where $S$ is replaced by $\hat{S}$ can still be made equal to the original system in Figure 6.1 with $S$ as dynamic block. To answer that question, we will make a first order distortion analysis under the Assumptions 8.3 and 8.4.

Consider the nonlinear structure in Figure 6.1 with $\hat{S}$ as linear block. The contribution $d_2S^2$ is due to the feedback nonlinear branch, and Assumption 8.3 implies that its linear output contribution $d_2S^2(u)$ will be an $O\left(\frac{M_{NL}}{|y_1|^2}\right)$. From Assumption 8.4, Remark 2, it follows that within higher order effects ($O\left(\left(\frac{M_{NL}}{|y_1|^2}\right)^2\right)$ or higher), we can replace in all
nonlinear branches the dynamic block $\hat{S}$ by $S$. This is not allowed in the first linear branch because there the replacement would give a first order error.

In the second identification step, the coefficients $c_{i,j}$ are estimated. For the nonlinear output contributions, a full equivalence (within the neglected higher order effects) between both structures (the first one based on $S$, the second one on $\hat{S}$) is obtained by putting

$$\hat{c}_{1,j}d_1 = c_{1,j} \quad \text{for the second branch}$$
$$\hat{c}_{2,j} (d_1)^j = c_{2,j} \quad \text{for the third branch}$$
$$\hat{c}_{3,j} (d_1)^{(j+1)} = c_{3,j}, j \neq 1 \quad \text{for the feedback branch} \quad (6.6)$$

The linear output contribution is matched by putting

$$d_1^2 \hat{c}_{3,1} = c_{3,1} - c_{1,1}d_2 \quad (6.7)$$

and

$$d_1 \hat{c}_{1,1} = 1 \quad (6.8)$$

With these choices, both structures have within the second order effects the same output. This is easily verified by comparing the (non)linear output contributions of both structures, once with the first set of coefficients and once with the second set.

Within the neglected second order effects, the solution of the least squares problem will be exactly these special choices of the coefficients as given in (6.6) till (6.8) since the best match also minimizes the cost function.

**6.C Computing the Volterra Kernels of a Nonlinear Feedback Structure**

This section contains the calculations leading to (6.5). The system is shown in Figure 6.8. All calculations are limited to the third order kernels, resulting in an approximation for small inputs. The composition of two nonlinear operators described by their Volterra kernels is discussed in Schetzen (1980, Section 7.2).
Starting from (6.3) and (6.4), the following approximation is obtained:

\[ P^{-1} \left[ R_1^{-1} [u] + R_2^{-1} [u] + R_3^{-1} [u] \right] + Q \left[ R_1^{-1} [u] + R_2^{-1} [u] + R_3^{-1} [u] \right] \approx u \]

Developing the function compositions, but limiting the expansion to terms of order 3 or less yields

\[
\begin{align*}
  u & \approx P_1^{-1} [R_1^{-1} [u]] + P_1^{-1} [R_2^{-1} [u]] + P_2^{-1} [R_1^{-1} [u], R_1^{-1} [u]] + \\
  & \quad P_1^{-1} [R_3^{-1} [u]] + 2P_2^{-1} [R_1^{-1} [u], R_2^{-1} [u]] + P_3^{-1} [R_1^{-1} [u], R_1^{-1} [u], R_1^{-1} [u]] + \\
  & \quad Q_1 [R_1^{-1} [u]] + Q_1 [R_2^{-1} [u]] + Q_2 [R_1^{-1} [u], R_1^{-1} [u]] + \\
  & \quad Q_1 [R_3^{-1} [u]] + 2Q_2 [R_1^{-1} [u], R_2^{-1} [u]] + Q_3 [R_1^{-1} [u], R_1^{-1} [u], R_1^{-1} [u]]
\end{align*}
\]

where \( P_1^{-1} [u] \) is the first order kernel of \( P^{-1} \). Satisfying this equation gives the three first Volterra Kernels of \( R^{-1} \):

\[
\begin{align*}
  R_1^{-1} [u] &= (P_1^{-1} + Q_1)^{-1} [u] \\
  R_2^{-1} [u] &= - (P_1^{-1} + Q_1)^{-1} \left[ P_2^{-1} [R_1^{-1} [u]] + Q_2 [R_1^{-1} [u], R_1^{-1} [u]] \right] \\
  &= -R_1^{-1} \left[ P_2^{-1} [R_1^{-1} [u]] + Q_2 [R_1^{-1} [u], R_1^{-1} [u]] \right] \\
  R_3^{-1} [u] &= -R_1^{-1} \left[ 2P_2^{-1} [R_1^{-1} [u], R_2^{-1} [u]] + P_3^{-1} [R_1^{-1} [u], R_1^{-1} [u], R_1^{-1} [u]] + \
  & \quad 2Q_2 [R_1^{-1} [u], R_2^{-1} [u]] + Q_3 [R_1^{-1} [u], R_1^{-1} [u], R_1^{-1} [u]] \right]
\end{align*}
\]
6 Fast Approximate Identification of Nonlinear Systems
7 Measuring Volterra Kernels of Microwave Devices

There are application domains where the measurement of Volterra kernels is not straightforward due to imperfections in the measurement setup: high frequency devices (operating above 1GHz) are an example of this. These are usually characterized using a series of single sines with variable power as an excitation signal. However, models obtained with such one-tone methods are not generalizable when there is an interaction between a linear source and the nonlinear DUT: using them to predict the output of the DUT for another source gives physically impossible results. This will be demonstrated on measurement data. Another method that uses two tone signals with very close frequencies is proposed as an alternative to one-tone tests. By varying the phase relation between the tones, the system’s nonlinear behavior is separated from this interaction. Note that such excitations can be generated using commercially available synthesizers.
7 Measuring Volterra Kernels of Microwave Devices

7.1 Introduction

This chapter is to be situated in a special application domain: the problem of measuring and modeling high frequency devices. At these high frequencies (around 1GHz and above), it is easier to interpret waves than voltages or currents. However, with the following definitions (Kurokawa, 1965), the relation between the more usual physical quantities and waves can be seen:

\[
\begin{aligned}
    a_1 &= \frac{v + Z_0 i}{2} \\
    b_1 &= \frac{v - Z_0 i}{2}
\end{aligned}
\]  

(7.1)

These waves have an interesting interpretation: they represent the flow of power along the wire (\(a_1\) is the forward propagating voltage wave and \(b_1\) is the corresponding backward propagating voltage wave). \(Z_0\) is called the characteristic impedance of the system. When the output impedance of the source and the input impedance of the DUT are complex conjugate (hence equal if the impedances are real), there is an optimal power transfer from the source to the DUT. If the input impedance of the DUT is different from the output impedance of the source (i.e. there is a mismatch between them), waves \(a_2\) and \(b_2\) impinging the DUT can be computed as follows (see Figure 7.1):

\[
\begin{aligned}
    a_1 + b_1 &= v = a_2 + b_2 \\
    a_1 - b_1 &= i = \frac{a_2 - b_2}{Z_1} \\
    \frac{a_1 + b_1}{Z_0} &= \frac{a_2 + b_2}{Z_1}
\end{aligned}
\]  \begin{align*}
    2a_2 &= a_1 + b_1 + \frac{Z_1}{Z_0} (a_1 - b_1) \\
    2b_2 &= a_1 + b_1 - \frac{Z_1}{Z_0} (a_1 - b_1)
\end{align*}

\[
\begin{aligned}
    a_2 &= \frac{Z_0 + Z_1}{2Z_0} a_1 + \frac{Z_0 - Z_1}{2Z_0} b_1 \\
    b_2 &= \frac{Z_0 - Z_1}{2Z_0} a_1 + \frac{Z_0 + Z_1}{2Z_0} b_1
\end{aligned}
\]  

(7.2)

\(A_1\) represents the Fourier coefficients of \(a_1\). Even though the following will be written in terms of waves, it applies as well to any measurable quantity.
7.2 Problem statement

In recent years, a growing interest appeared for modeling the nonlinear behavior of microwave circuits. Prototypes of nonlinear vectorial network analyzers became available (Van den Broeck and Verspecht, 1994; Van Moer and Rolain, 1999), and the first attempts were made to model the measured nonlinear behavior using Volterra series approaches, e.g. in Verbeyst and Vanden Bossche (1994); Rolain et al. (1998). In many cases, the aim is to model the relation between the fundamental incident wave $A_1(f)$ and the nonlinear distorted output waves $B_2(f)$, $B_2(2f)$ and $B_2(3f)$. Although this seems, at first glance, to be a simple task, it turns out that the interaction between the nonlinear circuit and the Continuous Wave (CW) generator (nonlinear source pull) complicates this picture significantly. Due to these effects, the system is not only excited by the pure carrier $A_1(f)$ alone, but also by its higher harmonic components $A_1(2f)$ and $A_1(3f)$.

In principle, it should be possible to measure and identify the full Volterra map that describes the relation between $A_1(f)$, $A_1(2f)$, $A_1(3f)$ and $B_2(f)$, $B_2(2f)$, $B_2(3f)$. Assuming that there are no higher order unwanted contributions (i.e. $A_1(kf) = 0; k > 3$), this would be achieved by scanning a five dimensional space: the five dimensions are the amplitude of each component (three in total) and the phase of the harmonics referred to the fundamental (two more dimensions). There are 5 dimensions instead of 6 (2 dimensions, the amplitude and the phase, for each of the 3 components) because adding a linear delay to the signal only shifts the time origin of the experiment. However, the experiment is not “rich” enough to get a generalizable estimate. The user sets only $A_1(f)$ on the CW source: $A_1(2f)$ and $A_1(3f)$ are unwanted contributions created by the source-DUT pair. As these components are determined by the setup, they do not vary independently of $A_1(f)$. This leads to an excitation that does not fill up the full five dimensional excitation space but remains instead on a one-dimensional curve in it.

When a model is extracted using this type of measurements, a good agreement between model and measurement is obtained. However, the resulting model is very unreliable as it does not allow any generalization. Even simple simulations like the prediction of the system output for a pure sine excitation at the input (putting $A_1(2f)$ and $A_1(3f)$ equal to zero) will fail. The extracted model uses the additional degrees of freedom contained in the input space to reach a better fit of the measured data: basically the
model does not only describe the nonlinear system behavior, but the combination of the generator and the system setup. Changing one of the components of this setup will hence change the model. It is clear that for simulation purposes this is unacceptable. For that reason the nonlinear source pull should be separated from the system model. The same phenomenon has been observed on simulations in Bai (2003). In this chapter, the quality of a model will be judged by the plausibility based on physical arguments of its response for an ideal sine wave at the input.

7.3 Illustration of the problem

Nonlinear source pull is an effect induced by the cascade of a source with non-zero output impedance and a nonlinear device. Even for a generator with an output impedance of $Z_0$ at all frequencies and whose output spectrum is a pure tone when connected to a linear impedance, there will still be non-zero waves at the harmonic frequencies that excite the DUT whose input impedance is nonlinear. Consider the system shown in Figure 7.2 and built up with a perfect generator driving an amplifier with a nonlinear input impedance. Even though the generator is supposed to be matched perfectly, harmonics can be shown to appear in the excitation spectrum, because of the variation of the input impedance of the DUT implied by the nonlinearity itself.

Using the circuit of Figure 7.2, the incident ($a_1$) and reflected ($b_1$) wave are easily calculated, if the square root is approximated by its Taylor series (the computations can be found in Appendix 7.A):

$$\begin{align*}
    a_1 &= A \sin(\omega t) \\
    b_1 &= -\frac{1}{2} \beta Z_0 A^2 \sin^2(\omega t) + \frac{1}{2} \beta^2 Z_0^2 A^3 \sin^3(\omega t)
\end{align*}$$

(7.3)
But, because of the nonlinear input impedance, the input impedance of the DUT changes (it becomes $\neq Z_0$) when the amplitude $A$ changes and now depends on the power of $a_1$, resulting in a power-dependent mismatch at the device input. Converting the waves to another impedance system $Z_{NL}$ with (7.2) will introduce the harmonics in the incident wave: (7.4) shows that the DUT will be excited by harmonics of the excitation wave, even though the generator is assumed to be perfect. This shows that when looking at the nonlinear behavior of a DUT, one has to take into account that even in a single tone setting, the DUT is always excited at the harmonic frequencies.

$$a_2 = \frac{Z_0 + Z_{NL}}{2Z_0} A \sin(\omega t) + \frac{Z_0 - Z_{NL}}{2Z_0} \left( -\frac{1}{2} \beta Z_0 A^2 \sin^2(\omega t) + \frac{1}{2} \beta^2 Z_0^2 A^3 \sin^3(\omega t) \right)$$

$$b_2 = \frac{Z_0 - Z_{NL}}{2Z_0} A \sin(\omega t) + \frac{Z_0 + Z_{NL}}{2Z_0} \left( -\frac{1}{2} \beta Z_0 A^2 \sin^2(\omega t) + \frac{1}{2} \beta^2 Z_0^2 A^3 \sin^3(\omega t) \right)$$

Even if a filter would be added in between the generator and the DUT, this would not solve the problem, as the impedance $Z_0$ would now be the output impedance of the filter. Note also that this mismatch is dependent on the input power and frequency; a linear matching network, albeit ideal, can therefore never be used to obtain a perfect match for all power, frequency pairs.

### 7.4 Basic Idea

From the identification point of view, the main trouble comes from the fixed phase relation between the different harmonics of the input signal. One could use an impedance tuner between the source and the DUT to vary the source pull and change the phase relation between the harmonics. However, this does not allow to control directly the phase of the harmonics unless complex harmonic tuners are used.

The solution that is proposed here is to excite the system with a two tone excitation with almost coinciding frequencies $f_1$ and $f_2$. This is very close to a CW excitation, but introduces one more user controllable degree of freedom in the choice of the input signal: the phase between the two components. Now, when mixing will occur between these components, the resulting phase will be influenced by the phase difference between the carriers at $f_1$ and $f_2$. 
Due to the nonlinear source pull, the incident waves will consist of $A_1(f_1)$, $A_1(f_2)$ and their (inter)modulation products. All these components contribute to the output. Restricting ourselves without loss of generality to 3rd degree nonlinearities, the contributions at the fundamental frequency $f_1$ have the following form: $A_1(f_1)$ (linear), $A_1(f_1)A_1(3f_1)A_1(-3f_1)$, $A_1(f_1)A_1(3f_2)A_1(-3f_2)$, $A_1(f_1)A_1(f_2)A_1(-f_2)$ and other third degree participations. Similar contributions appear at $f_2$: one linear $A_1(f_2)$ and $A_1(f_2)A_1(3f_2)A_1(-3f_2)$ together with other third degree participations. Completely different terms appear at the nearby frequency $3f_1 - 2f_2$, e.g.: $A_1(f_2)A_1(3f_1)A_1(-3f_2)$. The Volterra kernels (Schetzen, 1980) associated with these different third order products can be approximated by the same value because these contributions hit the nonlinear system at almost the same place on the multidimensional frequency grid ($f_1 \approx f_2$). However these contributions have a completely different phase relation that is easily changed by varying the phase of $A_1(f_1)$ and $A_1(f_2)$.

Theoretically speaking, any system that can be modeled using Volterra series can be identified using this method. However, for strong nonlinearities, the number of harmonics to take into account might make the measurements and computation cost prohibitive: the more harmonics are generated, the more different excitation signals have to be applied to the DUT and the more input products one has to compute.

### 7.5 Experimental verification

The prediction power of the extracted model will be assessed as follows: the model that is extracted using the proposed dual tone method is used to model the DUT’s response to a single sine wave at the same frequency. Next, to show that CW measurements indeed suffer from the described problem, the DUT’s response itself is measured at this frequency for a sinewave experiment and compared to the obtained model and to the output power predicted by the model extracted from these single tone measurements. Both models will be extracted similarly by minimizing (7.5).

#### 7.5.1 Measurement

The proposed method was tested on a MAR 6 amplifier from Mini-Circuits with the 1dB compression point at -6dBm input power at 900MHz. The measurements were made
7.5 Experimental verification

Figure 7.3: Experimental setup

Figure 7.4: Comparison of the phase of the second input harmonic for the MAR 6 setup for the measurements with a single tone excitation (black lines) and with dual tone excitations (one symbol for each of the 8 different phases)

on the same system (Figure 7.3) with two different excitation signals: a single tone at 900MHz and a dual tone excitation at 899.995MHz and 900.005MHz. This allows to compare how both methods perform. The input and output waves were measured for an input power going from -20dBm to 0dBm in 1dB steps. For each wave, the necessary frequency components were recorded with the vectorial nonlinear network analyzer described by Van den Broeck and Verspecht (1994).

A second experiment was performed on the MRFIC2006, an RF amplifier from Motorola. The measurements were done at 800MHz, from -16dBm input power till 5dB compression. The 1dB compression point of this amplifier is at -0.2dBm input power at 800MHz. The results for this second experiment were very analogous. So only Figure 7.8 assessing the prediction power of the extracted model will be included.

Figure 7.4 shows that the phase differences between the fundamental and the second harmonic (1800MHz: \(2f\) for the CW measurements, \(f_1 + f_2\) for the dual tone measurements) are spread over the whole range \([-\pi; \pi]\) for the dual tone experiments as intended, while the phases of the second component for the single tone experiments stay the same. The plotted phase is computed as the phase difference between the component
at 1800MHz and the square of the first component (900MHz for the single tone measurements, 899.995MHz for the dual tone measurements). This representation cancels out the influence of the phase of the fundamental and allows to see clearly the influence of the 8 phase steps used in the excitation signal. For low input powers, the second harmonics are very close to the noise floor, resulting in a random distribution of the measured phases. For the higher input powers, the second harmonics are clearly defined and it is obvious that the phase relation between the second harmonic and the fundamental stays fixed and that the variation comes from the different phase realizations of the dual tone excitation.

### 7.5.2 Extracted Models

The kernels were estimated with (7.5) as a cost function. This is a slightly modified weighted least squares estimator (Van Moer, 2001) where $B_2(i)$ is the measured output wave, $X(i)$ a $1 \times n$ vector containing the products of the measured input waves ($A_1$), $V$ is a $n \times 1$ vector with the unknown Volterra kernels (ordered in the same way as the products) and $w_i$ is the weighting function.

$$K = \sum_i \left| \frac{\log \frac{B_2(i)}{X(i)V}}{w_i^2} \right|^2$$  \hspace{1cm} (7.5)

In identification, it is common practice to select $w_i$ to match the level of uncertainty: $w_i$ is then equal to the standard deviation of the numerator of (7.5). Here, the weights are not computed analytically using the standard deviation of the measurements, but estimated from the measured values: each measurement has been repeated (64 times for the CW measurements and 3 times for the dual tone measurements) and the standard deviation of the equation error is computed using each repetition as a sample. This makes the weights a function of the estimated kernels. To simplify the equation, the kernel values from the previous estimation step are used. Because the proposed scheme is iterative, the estimation of the kernel values has to be initialized. The first estimates of the kernels are the values that minimize the classical unweighted least-squares cost function.

The data matrix $X$ in (7.5) contains data covering a wide range: each row is associated with the applied input power, while each column represents a different Volterra kernel.
7.5 Experimental verification

Figure 7.5: Comparison of the measured output power (dots) and the fitted output power for the single-tone experiments done on the MAR 6.

(hence, products of the measurements). Consequently, $X$ is poorly conditioned and using the log operator in (7.5) is better for the conditioning of the minimization of (7.5) (Van Moer, 2001).

### 7.5.3 Comparison of the extracted models for the single tone and dual tone measurements

With the data from the CW experiment, the Volterra kernels up to the 7th order contributing to the fundamental frequency at the output were estimated. These kernels will be called the single tone kernels, with “single tone” indicating that they were obtained from CW measurements. In Figure 7.5, the model was applied on the same experimental data that is used to identify the model and compared to the measured output. A perfect fit is obtained. The convergence of the kernel values is extremely fast: the optimum is reached usually in less than 5 iterations.

It is impossible to represent on a plot the different axes needed to compare the fit of the dual tone model with the dual tone measurements: there should be an axis for the input power, one for the phase realization, one for the frequency and one for the output power. So only a subset of the data is plotted in Figure 7.6 to illustrate the comments.

For high input powers, additional frequency lines close to the fundamental frequency are included: these are the lines excited by the nonlinear source pull that grow above
the noise when the input power increases. Their existence in the bottom two plots of Figure 7.6 proves that nonlinear source pull is indeed present.

The fit of the dual tone model isn’t as good as for the single tone case: the output predicted by the model for the measured input matches the measured outputs less closely. This was to be expected: the input signal is much richer, yet the same number of parameters is estimated. Note also that the measurement points where the fit appears to be much worse correspond to frequency lines that are not in the intended excitation spectrum. These lines result of the nonlinear source pull third order contributions (e.g.: $2f_1 - f_2$, $2f_2 - f_1$). These contributions are much smaller than the lines excited by the generator and have a bad signal-to-noise ratio. Since the estimator weights measurements according to their quality, the fit for these lines may be loose without a significant increase of the cost. For lower input powers (see the upper plots of Figure 7.6) where these side-lines are hidden in the noise, they aren’t even taken into account.

A simulation of the CW and dual tone model is performed in Figure 7.7 for a pure sine wave excitation, using the models extracted from the measurements (putting the additional input harmonics to zero in the simulation), using only the kernels $V(f_1)$,
7.5 Experimental verification

Figure 7.7: Measurements and simulated output of the identified models of the MAR 6 for a pure sine excitation: single tone measurements (grey dots), dual tone measurements (grey pluses), output of the model identified from single tone measurements (broken line), output of the model identified with a dual tone excitation (full line).

\[ V(f_1, f_1, -f_1), \quad V(f_1, f_1, f_1, -f_1, -f_1) \] and \[ V(f_1, f_1, f_1, -f_1, -f_1, -f_1) \]. Despite the perfect fit between CW model and measurement (all input harmonics included), a very poor prediction is obtained here, while the dual tone model predicts much better.

Figure 7.8 shows the same simulation as Figure 7.7 for the RFIC. The model extracted from dual tone measurements was extrapolated over 1.9dBm to match the CW power range and still looks plausible: the crosses indicate that the dual tone measurements cover a smaller power range than the thick line which shows the output of the dual tone model. On the other hand, the output power simulated with the model identified from the single tone model doesn’t look plausible at all, since expansion is predicted instead of the measured compression.

7.5.4 Comparison of the estimated kernel values for the MAR 6

Tables 7.1 and 7.2 show the estimated single tone and the dual tone kernel values for the MAR 6 DUT. To compare their relative importance, the power of each kernel for the highest measured input power of -2dBm, corresponding to 7.9dBm at the output, is included. This power is computed as the predicted output power when all other kernel values are set to 0. The bold face kernels in Table 7.1 are the ones used for the predictions in Figure 7.7.
Figure 7.8: Measurements and simulated output of the identified models of the RFIC for a pure sine excitation: single tone measurements (grey dots), dual tone measurements (grey pluses), output of the model identified from single tone measurements (broken line), output of the model identified with a dual tone excitation (full line).

Table 7.1: Numerical values of some of the estimated Volterra kernels for the MAR 6 amplifier

<table>
<thead>
<tr>
<th>Volterra Kernel</th>
<th>CW kernel value</th>
<th>Dual Tone kernel value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V(f_1)$</td>
<td>-0.2096387-5.565479i+</td>
<td>-2.955302+4.868292i+</td>
</tr>
<tr>
<td>$V(f_1, f_1, -f_1)$</td>
<td>-38.33139+2.631012i+</td>
<td>+13.32411-117.6923i-</td>
</tr>
<tr>
<td>$V(f_1, f_1, f_1, -f_1, -f_1)$</td>
<td>+8867.484+4979.624i+</td>
<td>-53.37062+1165.927i-</td>
</tr>
<tr>
<td>$V(f_1, f_1, f_1, f_1, -f_1, -f_1)$</td>
<td>+122711.0+5927.236i+</td>
<td>+649.5390-2447.282i-</td>
</tr>
<tr>
<td>$V(2f_1, -f_1)$</td>
<td>-499.2629-61.82980i+</td>
<td>+309.8736+69.61719i-</td>
</tr>
<tr>
<td>$V(3f_1, -f_1, -f_1)$</td>
<td>-12208.92+15037.13i+</td>
<td>+1188.959-765.3288i-</td>
</tr>
<tr>
<td>$V(f_1, f_1, f_1, 2f_1)$</td>
<td>+71593.87-18381.58i+</td>
<td>+3886.299-5189.947i-</td>
</tr>
<tr>
<td>$V(2f_1, f_1, -f_1, -f_1)$</td>
<td>+7312.453+22897.17i+</td>
<td>-8263.892-2.454021i-</td>
</tr>
<tr>
<td>$V(3f_1, f_1, -f_1, -f_1)$</td>
<td>+320332.0-409185.2i+</td>
<td>-3402.645+1220.194i-</td>
</tr>
<tr>
<td>$V(f_1, f_1, f_1, f_1, -2f_1)$</td>
<td>-1381997+1937081i+</td>
<td>-23616.06+20997.73i-</td>
</tr>
<tr>
<td>$V(2f_1, f_1, f_1, -f_1, -f_1)$</td>
<td>+507612.4-1422140i+</td>
<td>+46275.19-7148.897i-</td>
</tr>
</tbody>
</table>
Table 7.2: Numerical values of some of the estimated Volterra kernels for the MAR 6 amplifier

<table>
<thead>
<tr>
<th>Volterra Kernel</th>
<th>CW power (dBm)</th>
<th>Dual Tone power (dBm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V( f_1) )</td>
<td>12.5</td>
<td>13.8</td>
</tr>
<tr>
<td>( V( f_1, f_1, -f_1) )</td>
<td>1.9</td>
<td>19.0</td>
</tr>
<tr>
<td>( V( f_1, f_1, f_1, -f_1, -f_1) )</td>
<td>23.9</td>
<td>19.6</td>
</tr>
<tr>
<td>( V( f_1, f_1, f_1, -f_1, -f_1, -f_1) )</td>
<td>19.5</td>
<td>7.8</td>
</tr>
<tr>
<td>( V(2 f_1, -f_1) )</td>
<td>1.6</td>
<td>-2.1</td>
</tr>
<tr>
<td>( V(3 f_1, -f_1, -f_1) )</td>
<td>-7.4</td>
<td>-11.1</td>
</tr>
<tr>
<td>( V( f_1, f_1, f_1, -2 f_1) )</td>
<td>14.1</td>
<td>0.4</td>
</tr>
<tr>
<td>( V(2 f_1, f_1, -f_1, -f_1) )</td>
<td>13.8</td>
<td>12.1</td>
</tr>
<tr>
<td>( V(3 f_1, f_1, -f_1, -f_1, -f_1) )</td>
<td>0.7</td>
<td>-15.9</td>
</tr>
<tr>
<td>( V( f_1, f_1, f_1, -f_1, -2 f_1) )</td>
<td>24.8</td>
<td>2.4</td>
</tr>
<tr>
<td>( V(2 f_1, f_1, f_1, -f_1, -f_1, -f_1) )</td>
<td>26.9</td>
<td>11.8</td>
</tr>
</tbody>
</table>

The first line shows that we are in compression: the output power predicted by the linear kernel is higher than the measured output power. The higher order kernels will hence need to cancel a part of this linear contribution. Since these kernels tend to infinity for power increasing without bound, a saturation contribution as required here can only be obtained when kernel contributions are almost canceling. This explains the huge power values created by separate kernels, for instance the \( V( f_1, f_1, -f_1) \) kernel. However, for the higher order kernels, the contributions of the single tone kernels to the prediction are higher than the contributions of the dual tone kernels. Because of the phase variation, fitting the dual tone kernels to reach cancellation on a certain phase realization will lead to a huge error for another phase realization of the input signal at the same input power. Hence, the estimation algorithm keeps the power from these kernel values low, instead of trying to correct for model errors by increasing the values of the estimated kernels to compensate for the lack of higher order kernels. In the CW case, the estimator is able to lower the model errors by increasing the higher order kernel’s power as long as they nearly cancel out, because the excitation space is not completely filled by the input signal.

Note that for the MAR 6, a good estimation could have been obtained from single tone measurements, too. One could simply consider the source as being perfectly matched and identify the Volterra kernels linked with the fundamental only. However, this works only because the matching of the generator and the MAR 6 was so good that the source
pull could hardly be seen. For a larger nonlinear source pull, it becomes mandatory to
include the input harmonics, leading to the aforementioned problems. The advantage
of the dual tone experiments is to provide a robust method that can be used without
much prior knowledge (like knowing whether there is nonlinear source pull or not).

7.6 Intermediate conclusion

A simple method is proposed to measure and identify sensible models of the nonlinear
system behavior, even in the presence of nonlinear source pull, using a single generator
e.g. the Rohde & Schwarz SMIQ generator. As opposed to models identified from CW
experiments, these models allow generalization and can be used for simulation purposes.
This method has been applied successfully to experimental measured data.

7.7 Consequences

It has been seen that the input of the DUT is not always the excitation that was chosen
for the identification experiment. Furthermore, a certain measurement technique (CW
measurements) has been shown to be inadequate in such circumstances. Naturally, the
question arises how the identification algorithms proposed in this book behave in such
circumstances.

In this section, the identification methods of the previous chapters will be tested on
an example including a nonlinear distortion of the input signals. The goal is to see if
these methods are robust with respect to the problem described in this chapter. These
tests will be run on simulations: applying the methods described in Chapters 3, 4 and 5
to high frequency measurements as for dual tones is not possible due to measurement
problems that will be exposed in Chapter 8. The simulated setup is shown in Figure 7.9.
The quality of the models will be tested by using them to predict the output of the DUT
with another realization of the excitation signal than the one used for the identification.
Nonlinear distortions of the input signal are removed in the simulation and a noise-free
setup is considered. This is shown for a Hammerstein system in Figure 7.10. The results
of this test shown in Table 7.3 indicate that Algorithms 2, 5, 7 and 8 are not disturbed
by a nonlinear distortion of the input signal used for their identification: the standard
deviation $\sigma_\epsilon$ of the output simulation error $\epsilon$ for validation data is always lower than

182
7.8 Conclusions

The generated input signal can be distorted nonlinearly before reaching the DUT. Applying different realizations of multisines (dual tones being a very particular kind of them) to the system allows the identification algorithms to be used even in the presence of such nonlinear distortions of the excitation signal without a significant impact on the

<table>
<thead>
<tr>
<th>System</th>
<th>$\sigma_M$</th>
<th>SNR (dB)</th>
<th>$\sigma_\epsilon$</th>
<th>$\frac{\sigma_\epsilon}{\sigma_M}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiener</td>
<td>$20 \times 10^{-3}$</td>
<td>23</td>
<td>$820 \times 10^{-6}$</td>
<td>4</td>
</tr>
<tr>
<td>Hammerstein</td>
<td>$20 \times 10^{-3}$</td>
<td>23</td>
<td>$370 \times 10^{-6}$</td>
<td>2</td>
</tr>
<tr>
<td>Hammerstein-Wiener</td>
<td>$52 \times 10^{-3}$</td>
<td>23</td>
<td>$4.6 \times 10^{-3}$</td>
<td>9</td>
</tr>
<tr>
<td>Wiener-Hammerstein</td>
<td>$17 \times 10^{-3}$</td>
<td>23</td>
<td>$140 \times 10^{-6}$</td>
<td>1</td>
</tr>
</tbody>
</table>
simulation properties of the identified model.
7.A Current and Voltage at the input of the DUT of Figure 7.2

The value of $v(t)$ in Figure 7.2 can be computed straightforwardly.

$$\frac{2A \sin(\omega t) - v(t)}{Z_0} = \frac{v(t)}{Z_0} + \beta v^2(t)$$

This is a quadratic equation in $v(t)$:

$$\beta Z_0 v^2(t) + 2v(t) - 2A \sin(\omega t) = 0$$

The solutions are

$$v(t) = -1 \pm \frac{1}{\beta Z_0} \sqrt{1 + 2\beta Z_0 A \sin(\omega t)}$$

Choosing the positive root and approximating $\sqrt{1+x}$ with $1 + \frac{1}{2} x - \frac{1}{8} x^2 + \frac{1}{16} x^3$ yields

$$v(t) = -1 + 1 + A Z_0 \beta \sin(\omega t) - \frac{1}{2} (\beta Z_0 A \sin(\omega t))^2 + \frac{1}{2} (\beta Z_0 A \sin(\omega t))^3$$

$$= A \sin(\omega t) - \frac{1}{2} \beta Z_0 A^2 \sin^2(\omega t) + \frac{1}{2} \beta^2 Z_0^2 A^3 \sin^3(\omega t)$$

The current produced by the excitation source is

$$i(t) = \frac{2A \sin(\omega t) - v(t)}{Z_0} = \frac{A \sin(\omega t)}{Z_0} + \frac{1}{2} \beta Z_0 A^2 \sin^2(\omega t) - \frac{1}{2} \beta^2 Z_0^2 A^3 \sin^3(\omega t)$$

Substituting $v(t)$ and $i(t)$ in (7.1) yields the result seen in (7.3):

$$a_1 = \frac{v(t) + Z_0 i(t)}{2} = A \sin(\omega t)$$

$$b_1 = \frac{v(t) - Z_0 i(t)}{2} = -\frac{1}{2} \beta Z_0 A^2 \sin^2(\omega t) + \frac{1}{2} \beta^2 Z_0^2 A^3 \sin^3(\omega t)$$
8 Wave Phase Calibration

This chapter proposes a method to calibrate the phase of a network analyzer for nonlinear systems. Besides the relative calibration that compensates for the phase and amplitude distortion at one frequency, there exist also “absolute” calibration procedures that fully calibrate the shape of spectrally rich time signals at RF frequencies. There already exist methods to calibrate the amplitude and even the phase of network analyzers for nonlinear systems. However, these cannot be applied to the dense frequency grids used to construct excitation signals for the methods presented in the previous chapters. Generally, in order to calibrate a measurement device, a standard device with known properties has to be measured (e.g. a thru, i.e. connecting input and output together trivially), and a compensation is obtained comparing the measurements to the known values. In this chapter, a wave phase calibration is proposed where the a priori knowledge of the nonlinear standard element resides in its structure (rather than in the actual parameter values).
8 Wave Phase Calibration

8.1 Introduction

The methods presented in Chapters 3, 4, 5 and 6 use broadband excitations in the sense that the relative bandwidth of the signals is large $\frac{BW}{f_{\text{center}}}$ is close to 1. The input signal is a multisine with many (> 100) spectral components. The use of this signal has been demonstrated on low frequency measurements. However, when applied to RF measurements, a new challenging problem arose. To build a nonlinear model, the measured input and output time signals may not be distorted at all: it is vital that measured and exact signals agree in shape. To remove the systematic errors of the network analyzer a dedicated calibration is to be used.

First, the problem is introduced giving a description of the (uncalibrated) measurements and the subsequent identification results that prove that the systematic errors introduced by the network analyzer are not negligible. The network analyzer for nonlinear systems (Rolain et al., 1998; Van Moer and Rolain, 1999; Van Moer et al., 2000) used to perform these measurements is presented in Section 8.2. The measurement data and identification results in Section 8.3 illustrate the influence of the lack of calibration.

The second part of the chapter proposes a possible solution to the problem (Section 8.5) that is tested on simulations in Section 8.6.

8.2 The Network Analyzer for Nonlinear Systems

The network analyzer for nonlinear systems shown in Figure 8.2 can measure the incoming and the reflected wave spectra (Kurokawa, 1965) impinging the DUT in one take. The excitation signal used here is created by an Arbitrary Waveform Generator (AWG). Conceptually, it can be represented by a Zero Order Hold (ZOH) cascaded with a low-pass filter used to limit the bandwidth of the generator to $\frac{f_S}{2}$. The acquisition part of the device uses directional couplers to separate incoming and reflected waves. Only a small portion of the energy is fed to the acquisition. The measurement itself is built around the harmonic sampling principle: the frequency $f_D$ of the clock IF-Gen is chosen to take advantage of the aliasing effect. $f_D$ is chosen much lower than half the bandwidth of the signals, but such that no two discrete components of the periodic signal fold down to the same place through aliasing. This is illustrated in Figure 8.1. The whole spectrum is compressed in the Intermediate Frequency (IF) bandwidth of the samplers after the
8.2 The Network Analyzer for Nonlinear Systems

harmonic mixer. To be sure that the samplers are not saturated, variable attenuators (ATT in Figure 8.2) are placed in the signal path before the sampler itself. After amplification of the resulting IF signal, the signals are digitized using an ADC whose sampling frequency $f_S$ is synchronized on $f_D$.

The measurement data that will be used in Section 8.3 is calibrated using a frequency response calibration, followed by an (absolute) power calibration. These calibration steps will be explained in Sections 8.2.2 and 8.2.3. A third calibration step, the phase calibration in between frequency components (Section 8.2.4) could not be executed because the necessary calibration element does not exist (yet). Section 8.4 shows that this step is necessary, too. The calibration presented here is the strict minimum necessary for this chapter. The calibration of the network analyzer for nonlinear systems has been studied in detail in the literature for a sparse frequency grid (Van den Broeck and Verspecht, 1994; Van den Broeck, 1995; Van Moer, 2001).

8.2.1 Downconversion

Sampling signals directly at RF frequencies proves impossible due to the limited conversion speed of the available ADC’s. To overcome this physical limitation, the harmonic
sampling principle has been used: the full frequency band is compressed into an IF band where signals can be sampled easily. Subsampling makes use of aliasing: when a high frequency signal at frequency $f_H$ is sampled with a sampling frequency $f_D$, the resulting downconverted frequency is

$$f_{IF} = f_D \left( \frac{f_H}{f_D} - \left\lfloor \frac{f_H + \frac{f_D}{2}}{f_D} \right\rfloor \right)$$  \hspace{1cm} (8.1)$$

The result of (8.1) is $-\frac{f_D}{2} \leq f_{IF} < \frac{f_D}{2}$. If there are two frequencies with non zero amplitude that are shifted to the same $f_{IF}$ by the sampling process, the spectral content of these lines will overlap information is lost: the sum of the two Fourier coefficients cannot be separated into its components. If a signal with known frequency grid is measured and care is taken to avoid signal frequencies that have a coinciding $f_{IF}$, the original signal can be reconstructed from the downconverted signal. Note that care is still required as the components whose $f_{IF} < 0$ are complex conjugated with respect to the original spectral lines.
8.2 The Network Analyzer for Nonlinear Systems

Figure 8.3: The frequency response calibration.

8.2.2 Frequency Response Calibration

The frequency response calibration is illustrated in Figure 8.3. It will be assumed here that the couplers have an infinite directivity: the \( a \) coupler measures only the \( a \) wave. In practice, this is a too simplistic approach, but inclusion of the full calibration is straightforward and does only obscure the point that is made here. The calibration is performed using a thru as a calibration standard. The standard is inserted instead of a regular DUT. For an ideal thru \( A_1 = B_2 \) (see Figure 8.3). Only the channels measuring \( A_{1,IF} \) and \( B_{2,IF} \) are shown here, as only these will be used in what follows. Both channels will introduce a different distortion, labeled \( C_i \). Assume that

\[
\begin{align*}
A_{1,IF}(f) &= C_1(f)A_1(f) \\
B_{2,IF}(f) &= C_2(f)B_2(f)
\end{align*}
\]

If a linear model is estimated for the transfer function of the thru, the result should be 1, independent of the frequency.

\[
G_c(f) = \frac{B_c(f)}{A_c(f)} = 1
\]
However, for the measured data one obtains

\[ G_{\text{Thru}}^{m}(f_{\text{IF}}) = \frac{B_{\text{Thru}}^{2}(f_{\text{IF}})}{A_{\text{Thru}}^{1}(f_{\text{IF}})} \neq 1 \]

A correction is the proposed for the measurement of a DUT (assume that \( A_{\text{DUT}}^{1,\text{IF}} \) and \( B_{\text{DUT}}^{2,\text{IF}} \) are the measured waves for the DUT):

\[ G_{\text{DUT}}^{c}(f) = \frac{G_{\text{DUT}}^{m}(f)}{G_{\text{Thru}}^{m}(f)} \]

\[ (8.2) \]

(8.2) is the frequency response calibration. For linear modeling purposes, this relative calibration is sufficient for a network analyzer with couplers that have an infinite directivity, as the same distortion is introduced both at the input and output, and hence disappears in the division:

\[ G_{\text{Thru}}^{c}(f) = \frac{G_{\text{Thru}}^{m}(f)}{G_{\text{Thru}}^{m}(f)} = 1 \]

The frequency response calibration is only concerned with the ratio of the waves \( A_{1,\text{IF}} \) and \( B_{2,\text{IF}} \). In the following sections, the waves will be studied so that an expression of the calibration procedure in term of the waves is necessary. The frequency response calibration causes the systematic errors of the network analyzer to be the same for both channels. It will be assumed that \( A_{1,\text{IF}}(f) = A_{1,\text{IF}}^{c}(f) \) is not distorted and \( B_{2,\text{IF}}^{c}(f) \) is computed to conform with (8.2):

\[
\begin{cases}
A_{1,\text{IF}}^{c}(f) = A_{1,\text{IF}}(f) \\
B_{2,\text{IF}}^{c}(f) = \frac{B_{2,\text{IF}}(f)}{G_{\text{Thru}}^{m}(f)}
\end{cases}
\]

### 8.2.3 Absolute Amplitude Calibration

If a nonlinear model is to be built, a wave calibration is needed: the amplitude and the phase of the distortion of each channel should be characterized separately over the complete frequency band. Imagine that a static nonlinearity is measured, and assume for simplicity that \( C_{1}(f) = 4 \) and \( C_{2}(f) = 2 \):

\[
\begin{align*}
b_{2}(t) &= a_{1}(t) + (a_{1}(t))^{2} \\
a_{1,\text{IF}}(t) &= 4a_{1}(t) \\
b_{2,\text{IF}}(t) &= 2b_{2}(t)
\end{align*}
\]
8.2 The Network Analyzer for Nonlinear Systems

Figure 8.4: Absolute amplitude calibration.

\[
a_{1,\text{IF}}(t) = a_{1,\text{IF}}(t)
\]
\[
b_{2,\text{IF}}(t) = 2b_{2,\text{IF}}(t) = 4a_{1}(t) + 4(a_{1}(t))^2 = a_{1,\text{IF}}(t) + \frac{1}{4}(a_{1,\text{IF}}(t))^2
\]

Obviously, the linear model that would be obtained from \(a_{1,\text{IF}}(t)\) and \(b_{2,\text{IF}}(t)\) is still correct, but the coefficient of \((a_{1}(t))^2\) is scaled by \(\frac{1}{4}\), or an unknown amount in general.

The solution to this problem is an absolute wave amplitude calibration as in Figure 8.4: a traceable power meter is used to compare \(a_{1}(t)\) and \(a_{1,\text{IF}}(t)\) and to obtain a compensating gain factor. Repeating such a calibration at each frequency that was used in the measurement process allows to correct the amplitude distortion of the measurement channels. Note that since the absolute amplitude calibration is preceded by a relative calibration (the frequency response calibration in this chapter), the distortion of each channel is the same, so that only one channel needs to be calibrated in Figure 8.4

8.2.4 Wave Phase Calibration

The necessity of a wave phase calibration for nonlinear measurements can be shown by a similar argument. Suppose that a frequency response calibration and an absolute amplitude calibration have already been carried out. Then the only remaining distortion results from a phase shift \(\phi_1\) of the spectral line at \(\omega = 2\pi f\) and \(\phi_2\) at \(2\omega = 4\pi f\) (which
is the same for both channels because of the frequency response calibration). Assuming \( \phi_2 \neq 2\phi_1 \) (this excludes the case where the distortion is nothing but a pure time delay), one gets:

\[
\begin{align*}
a_1(t) &= \sin(\omega t) \\
b_2(t) &= \sin(\omega t) + \sin^2(\omega t) = \frac{1}{2} + \sin(\omega t) + \frac{1}{2} \sin(2\omega t - \frac{\pi}{2}) \\
a_{1,\text{IF}}(t) &= \sin(\omega t - \phi_1) \\
b_{2,\text{IF}}(t) &= \frac{1}{2} + \sin(\omega t - \phi_1) + \frac{1}{2} \sin(2\omega t - \frac{\pi}{2} - \phi_2) \\
a_{1,\text{IF}}(t) &= \sin(\omega t - \phi_1) \\
b_{2,\text{IF}}(t) &= \frac{1}{2} + a_{1,\text{IF}}(t) + \frac{1}{2} \sin(2\omega t - \frac{\pi}{2} - \phi_2) = \\
&= \frac{\cos(2\phi_1 - \phi_2) - 1}{2 \cos(2\phi_1 - \phi_2)} + a_{1,\text{IF}}(t) + \frac{1}{\cos(2\phi_1 - \phi_2)} \left( a_{1,\text{IF}}(t) \right)^2 + \\
&\quad \frac{1}{2} \cos(2\omega t - \frac{\pi}{2} - \phi_2) \tan(2\phi_1 - \phi_2)
\end{align*}
\]

It can be seen again that the coefficient of \( a_{1,\text{IF}}(t) \) in \( b_{2,\text{IF}}(t) \) is the same as the coefficient of \( a_1(t) \) in \( b_2(t) \). The linear model is not disturbed by a nonlinear phase distortion. However, as soon as \( \phi_2 \neq 2\phi_1 \) (i.e. not a pure time delay \( \tau: \phi = \omega\tau \)), the coefficient of \( (a_{1,\text{IF}}(t))^2 \) is different from the coefficient of \( (a_1(t))^2 \) and distortion appears.

To calibrate for that effect, an element is required that produces both harmonic components with a fixed reproducible and known phase shift. The only realizable solution up to now is to use a reference signal. Here, the phase is calibrated with a pulse generator. In Figure 8.5, a pulse from a known source is applied to the measurement system. This allows to calibrate the phase distortion at the frequencies where the source has been characterized: if the pulse has a repetition frequency of 600MHz, the phases is calibrated at multiples of 600MHz. Comparing measured and known values, the necessary correction can again be computed. Currently, such a source is available with pulse repetition frequencies ranging from 600MHz to 1.2GHz, and characterized on a frequency grid with a resolution of 2MHz. Hence, signals whose fundamental frequency is 600MHz, or 602MHz, or 604MHz and so on, can be fully calibrated. However, nothing is known about the phase relationship between the the frequency grid at 600MHz and the one at 602MHz. This means that measurements of a random phase multisine with a frequency resolution lower than 600MHz cannot be calibrated directly using this pulse
Figure 8.5: Wave phase calibration using a known pulse generator.

generator. To use a similar technique for a multisine with a frequency resolution of 2MHz, a generator with a fundamental frequency of 2MHz should be used. However, for such a device, the frequency range 500MHz to 1.5GHz uses harmonic indices 250 and higher. It is not feasible with the actual technology to obtain a pulse which is sufficiently sharp to generate a comb with sufficient energy in such a wide band and the broadband calibration for non-harmonically related multi-carrier signals of nonlinear measurements is merely still an open problem.

8.3 The Measurements and Identification Results

The DUT is an ZHL-42 amplifier of Mini-Circuits, Brooklyn, New York. The characteristics of this amplifier are shown in Table 8.1 (Mini-Circuits, 2000, p. 136). The excitation signal is generated by a Sony/Tektronix AWG710 with a sampling frequency of 4GHz.

The excitation signal applied to the DUT is chosen similarly as in Section 5.3.3. The available equipment limits the class of excitation signals that can be applied to the DUT. Ideally, the whole bandwidth of the DUT should be covered by the excitation
signal. However, a 2GHz bandwidth is the upper limit that can be reached with the Sony/Tektronix AWG710: the maximum clock frequency of this AWG is 4GHz. An additional problem is the effect of the ZOH character of the excitation: in the frequency domain, this means that the frequencies close to 2GHz (half the sampling frequency) are only weakly excited. This situation can be corrected by predistorting the input signal, as long as the frequency characteristic of the reconstruction filter is not too low. The compensation is possible because the shape of the reconstruction filter of the AWG can be measured beforehand. For the measurements, a special odd multisine between 500MHz and 1.5GHz (Vanhoenacker et al., 2001) was used: only two out of every three odd grid lines were excited. A constant (after correction for the amplitude of the reconstruction filter of the AWG) non-zero amplitude was given to frequencies between 502MHz and 1498MHz with a spacing of 2MHz. This corresponds (for a frequency resolution $f_0 = 2MHz$) to the grid lines 251, 253, 257, 258 and so on. Note that even though 502MHz is even, for the case $f_0 = 2MHz$, the frequency 502MHz corresponds to an odd grid line. The amplitude of the predistorted input signal is shown in Figure 8.6.

The measurements of the ZHL-42 with the network analyzer for nonlinear systems of Figure 8.2 are shown in Figures 8.7 and 8.8. Figure 8.8 shows that the device is nonlinear, because the distance between the excited and non excited frequencies decreased from 30dB at the input to 20dB at the output. The modeling results can be seen in Figures 8.9, 8.10 and 8.11. The input data fed to Algorithm 8 has been calibrated with a frequency response calibration and an absolute wave amplitude calibration. No wave phase calibration was performed because none was available. The identification results are discussed in Section 8.4.
8.3 The Measurements and Identification Results

Figure 8.6: Predistorted input loaded into the AWG.

Figure 8.7: Input measurements (left hand side): the spectrum is divided into excitation lines (dots), odd detection lines (crosses) and even (that also serve as detection lines) lines (pluses).
Figure 8.8: Output measurements (left hand side): the spectrum is divided into excitation lines (dots), odd detection lines (crosses) and even (that also serve as detection lines) lines (pluses).

Figure 8.9: Estimated input linear part $R$ (see Figure 2.6 for the definition of the Wiener-Hammerstein model) for the ZHL-42.
8.3 The Measurements and Identification Results

Figure 8.10: Estimated static nonlinear part $f$ (see Figure 2.6 for the definition of the Wiener-Hammerstein model) for the ZHL-42.

Figure 8.11: Estimated output linear part $S$ (see Figure 2.6 for the definition of the Wiener-Hammerstein model) for the ZHL-42.
Table 8.2: Comparison of the Wiener-Hammerstein model and the linear model for the ZHL-42 amplifier.

<table>
<thead>
<tr>
<th></th>
<th>Predicted output power</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excitation lines</td>
<td>Detection lines</td>
</tr>
<tr>
<td></td>
<td>(% of the output power)</td>
<td>(% of the output power)</td>
</tr>
<tr>
<td>Wiener-Hammerstein model</td>
<td>99.4</td>
<td>23.8</td>
</tr>
<tr>
<td>Linear model</td>
<td>97.0</td>
<td>2.2</td>
</tr>
</tbody>
</table>

8.4 The Measurement Problem

This section shows that the wave phase calibration is not only a problem of theoretical importance. The model from Section 8.3 is only marginally better than a straightforward linear model. Table 8.2 compares the Wiener-Hammerstein model with a linear model. The output simulation error at the frequencies of the excited lines is comparable for both models: both are doing well, they can explain almost all (97% at least) of the power at the excited frequencies at the output. The same comparison is done at the detection lines (those that were not excited). However, the prediction of the nonlinear model for the detection lines is disappointing: only 25% of the output power at these frequencies is explained by the nonlinear model. As the purpose of nonlinear models is to predict the nonlinear behavior of the DUT, this is a failure: the result is not worth the effort.

One of the reasons of this failure is the lack of a wave phase calibration. It is impossible to plot the phase distortion of the measurement channel to show how large this distortion is: if it could be done, then a wave phase calibration could be executed. Figure 8.13 gives an indication of the presence of the phase distortion. It shows the phase of the raw measured data (for the DUT) before frequency response calibration. Clearly, the distortion of both measurement channels is different. If their difference \( \phi_{b2} - \phi_{a1} \) is large, then at least one of them (\( \phi_{b2} \) or \( \phi_{a1} \)) is large, too. There are three contributions in the phase of \( B_2/A_1 \): the phase shift of the DUT, the phase difference between the measurement channels before the downconversion and the phase difference of the channels after the downconversion. The phase of the DUT cannot be separated from the phase difference induced by the channels upstream of the downconversion unit in Figure 8.13. However, the difference induced by the channels after the downconversion can be observed very easily: it causes the jumps in Figure 8.13. This results in a figure that is a “band”
8.4 The Measurement Problem

Instead of a line, and the width of this “band” indicates that the phase distortion after the downconversion is at least 30° (half of the width of the “band”).

The representation of the data points in Figure 8.13 was chosen specially to illustrate the following point: there is a kind of structure in the deviation from a smooth phase. It looks as if 4 data sets with a phase offset had been plotted on top of each other. This structure in the error stems from phase differences in the downconverted stage of \( a_1 \) and \( b_2 \). Consider Figure 8.2 again. Table 8.3 tracks some of the IF frequencies obtained after the downconversion for the input signal defined in Section 8.3. The downconversion frequency was chosen to be 16.0005MHz.

Even though frequencies are always positive, Table 8.3 contains negative frequencies in the second column. For the negative frequencies, measured spectral lines are complex conjugated in the downconversion process. Examination of the table shows that there are four groups of grid lines: those that are converted around 2MHz, those that are complex conjugated and converted around 2MHz, those that are converted around 6MHz and finally those that land complex conjugated around 6MHz. Each group of frequencies will see a different phase shift, that is influenced by the IF frequency (\( \approx 2\text{MHz} \) or 6MHz) and by the complex conjugation. If the \( a_1 \) downconversion channel has a different phase response than the \( b_2 \) downconversion channel, then each of these four groups will have a phase offset that is different from the three others.

Obviously, the phase difference between both channels introduces an error that will affect the identification of nonlinear models even if a frequency response calibration is performed. However, the identification of linear models will not be disturbed by this distortion (and nonlinear models will also have a low output simulation error at the frequencies where linear contributions dominate). Getting rid of the biggest error source is a priority before considering other possible errors.
Table 8.3: Shifting of some of the grid lines of the excitation signal as the signal is downconverted and sampled. The numbers in the headings of the column refer to where the signal is considered in Figure 8.12.

<table>
<thead>
<tr>
<th>1 [MHz]</th>
<th>2 [MHz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>502</td>
<td>5.9845</td>
</tr>
<tr>
<td>506</td>
<td>−6.016</td>
</tr>
<tr>
<td>514</td>
<td>1.984</td>
</tr>
<tr>
<td>518</td>
<td>5.984</td>
</tr>
<tr>
<td>526</td>
<td>−2.0165</td>
</tr>
<tr>
<td>530</td>
<td>1.9835</td>
</tr>
<tr>
<td>538</td>
<td>−6.017</td>
</tr>
<tr>
<td>542</td>
<td>−2.017</td>
</tr>
</tbody>
</table>

Figure 8.13: Phase of $\frac{B_2}{A_1}$. Dots, pluses, crosses and stars are the same dataset, but emphasize the structure of the phase error.
8.5 A Proposal for a Wave Phase Calibration

Since calibration of a measurement system is always based on the comparison of standard measurements and known element properties, a suitable standard element is to be sought first. For the proposed wave phase calibration, the knowledge consists of the structure of the nonlinear system rather than some known parameter and the calibration procedure is as follows: the nonlinear calibration element with known structure (but possibly with unknown parameters) is measured and a model of the phase distortions introduced by the measurement system is optimized until a good nonlinear model for the calibration element is obtained. Then the model of the measurement system is used to compensate for the phase distortion of the network analyzer.

The model used to describe the phase distortion of the measurement system is shown in Figure 8.14: the phase spectrum of the input signal $u(t)$ is first distorted in the high frequency path (by the directional coupler and the signal path up to the downconverter). This distortion is modeled as being caused by an LTI system $H_{RF}$. If one assumes a perfect sampler, the distortion of the sampler and the hardware present after the downconverter can be grouped in a second LTI $H_{IF}$. Note that two different LTI systems are needed to model the distortion, as $H_{RF}$ and $H_{IF}$ operate at different frequencies. The phase distortion of the measured phase spectra of $A_1$ and $B_2$ are equal because the data are assumed to be calibrated by the frequency response calibration. Prior application of the wave power calibration leaves only the phase to be calibrated.

The inserted standard is a nonlinear element that needs to agree reasonably with a model whose structure can be identified correctly and reliably. This is important because the phase distortion of the instrument will be obtained using the phase response of the estimated model. Hammerstein systems or Wiener systems are both possible candidates for this: their estimates are consistent and locally convergent. Because the quality of the phase calibration is lower where the linear part of the nonlinear standard attenuates the signal, ideally, the bandwidth of the standard has to be larger than the bandwidth.
of the measurement. The standard hence is required to behave as a static system with respect to the measurement bandwidth. A Wiener-Hammerstein system cannot be used as a standard here. The phase distortion $H_{RF}$ can not be separated from the input linear part $R$ of the Wiener-Hammerstein system. This introduces an indetermination, as both the measurement channel (i.e. $H_{RF}$) and the standard (i.e. $R$) cause the phase distortion at the input of the static nonlinearity $f$.

8.6 Simulations

With the proposed wave phase calibration, the problem has been shifted: instead of needing a pulse generator with a repetition frequency of 2MHz whose harmonics of order 250 up to 750 were still measurable, a nonlinear device with known structure is needed. As no such device is known for now\(^1\), the wave phase calibration was tested only on simulations. These tests will be used to explain the calibration method in detail.

8.6.1 Setup

The measurement instrument’s acquisition channels will be represented as in Figure 8.14, with $H_{RF}$ shown in Figure 8.15 and $H_{IF}$ shown in Figure 8.16. A random phase multisine with a similar spectrum as in Section 8.3 is used as excitation signal: it is an odd multisine with a frequency resolution of 2MHz, ranging from 500MHz to 7.5GHz where one odd line out of three is not excited. The downconversion frequency $f_D$ is 16.0005MHz and the sampling frequency $f_S$ is 20MHz. When the spectrum of the excitation signal and $f_D$ are taken into account, it appears that the excitation bandwidth only covers a small part of the IF bandwidth, shown by the thick lines in Figure 8.16.

Instead of a static nonlinearity, the Hammerstein structure has been chosen as standard nonlinear element. This will show the influence of the bandwidth of the linear part $S$ inside the Hammerstein structure. Two nonlinear standard elements with slightly different characteristics will be used. These are shown in Figures 8.17 and 8.18. Their exact definition is summarized in Table 8.4.

These two Hammerstein systems have been simulated consecutively in the setup of Figure 8.19 under identical conditions (same excitation spectrum, $f_D$, $f_S$, $H_{RF}$ and $H_{IF}$).

\(^1\)At the NIST, there exists a superconductor whose physical model and first measurements suggest that it is a static nonlinearity.
8.6 Simulations

Figure 8.15: True phase distortion of $H_{RF}$.

Figure 8.16: True phase distortion of $H_{IF}$.

Figure 8.17: Static nonlinear characteristic $f$ (left plot) and transfer function of the linear part $S$ (right plot) of the first standard nonlinear element.
Figure 8.18: Static nonlinear characteristic $f$ (left plot) and transfer function of the linear part $S$ (right plot) of the second standard nonlinear element

Table 8.4: Characteristics of the two Hammerstein systems used in the simulations.

<table>
<thead>
<tr>
<th>Static nonlinearity $f$</th>
<th>$S$ (4th order Chebychev filter)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 8.17 $u + \frac{u^2}{4} + \frac{u^3}{24} + \frac{u^4}{192} + \frac{u^5}{1920}$</td>
<td>2dB 4.096GHz</td>
</tr>
<tr>
<td>Figure 8.18 $u + 0.08u^2 - 0.3u^3 + 0.005u^4 + 0.07u^5$</td>
<td>3dB 5.3248GHz</td>
</tr>
</tbody>
</table>

Figure 8.19: Simulation setup.
The simulations are noise free.

The data $A_{1,\text{IF}}$ and $B_{2,\text{IF}}$ will be used to try and fit a model to $H_{\text{RF}}$ and $H_{\text{IF}}$ such that a satisfactory model can be obtained for the standard nonlinear element. The result of the calibration is a model for the phase distortions $H_{\text{RF}}$ and $H_{\text{IF}}$ which can be inverted to apply the necessary corrections to other measurements.

### 8.6.2 Parameterization issues

Both phase distortions (the “high” frequency phase distortion $H_{\text{RF}}$ before the sampler and the “low” frequency phase distortion $H_{\text{IF}}$ after the sampler) are approximated by polynomials. The parameterization for $H_{\text{IF}}$ is peculiar because only a small part of its frequency range interacts with the measurement, as Figure 8.16 shows: because of the choice of the downconversion frequency and the necessity to leave unused frequency grid lines to avoid that higher harmonics fold down on measured lines, the lines of interest cluster around certain frequencies (see also Table 8.3). The thick lines in Figure 8.16 indicate the frequency bands where the phase distortion needs to be calibrated. Hence, $H_{\text{IF}}$ is approximated by 5 polynomials (one for each region) of order 5, except the last that is approximated by a polynomial of order 9 (to try getting a phase error that would be the same everywhere). $H_{\text{RF}}$ is approximated by one polynomial of order 7.

### 8.6.3 The Calibration Procedure and Results

The trick behind the calibration is the fact that as long as the influence of $H_{\text{RF}}$ and $H_{\text{IF}}$ is different from a pure time delay, it will be impossible to find a good model for the nonlinear standard element. Thus, the problem is recast as an optimization problem where the cost (the output simulation error of the obtained model for the nonlinear standard element) is to be minimized with respect to the parameters of $H_{\text{RF}}$ and $H_{\text{IF}}$.

This optimization is done using the Optimization Toolbox for Matlab. For the presented simulation, the cost function is evaluated around 1500 times. For each cost function evaluation, the models for $H_{\text{RF}}$ and $H_{\text{IF}}$ are inverted (to guess $A_e$ and $B_e$) and $A_e$ and $B_e$ are used as the input data for Algorithm 2. The resulting output simulation error is the cost function that is optimized by the Optimization Toolbox.

This has been done for both Hammerstein systems (Figures 8.17 and 8.18) and Figures 8.20 and 8.21 show the results, displaying the remaining phase error after calibra-
Figure 8.20: Remaining error on the estimate of $H_{\text{RF}}$ (left plot obtained with the Hammerstein system from Figure 8.17, right plot obtained with the Hammerstein system from Figure 8.18).

Figure 8.21: Total phase error (left plot obtained with the Hammerstein system from Figure 8.17, right plot obtained with the Hammerstein system from Figure 8.18).
Table 8.5: Calibration result for the simulation Figure 8.19.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Calibration set</th>
<th>Predicted power (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>99.76</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>99.93</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>99.94</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>99.77</td>
</tr>
</tbody>
</table>

There is an indetermination in the calibration due to the time invariant nature of the standard element. Once $H_{RF}$ and $H_{IF}$ are compensated up to an unknown delay, the cost function can’t be lowered anymore since changing the delay has no effect. This unknown delay has been compensated for in Figure 8.20: there is no linear trend visible in the phase error $H_{0,RF} - H_{RF}$. However, the compensation for $H_{IF}$ could not be obtained, because of the gaps in the frequency axis where there is no model for the phase distortion. Hence, the total remaining phase error after calibration is shown in Figure 8.21. It can be seen in Figure 8.21 that the remaining error comes essentially from the estimate of $H_{RF}$, with one exception: further investigation showed that the trace with the huge error corresponds to the spectrum that sees $H_{IF}$ around 8MHz. Even with the higher order of the polynomial, no better approximation was obtained. Also, the influence of the bandwidth of $S$ can be seen in Figure 8.21: the error is concentrated in the attenuation band of $S$, which is broader for the first Hammerstein test system.

Two systems have been used instead of only one to be able to validate the obtained calibration. Using $A_{1,IF}$ and $B_{2,IF}$ simulated with the first Hammerstein system (Figure 8.17), a model of the phase distortion was built. This model of $H_{RF}$ and $H_{IF}$, obtained from data set 1, will be called calibration set 1 in Table 8.5. When this calibration is applied to the data simulated with the second Hammerstein system (Figure 8.18), a model of this second system can be estimated: this is the first row of Table 8.5. The data set 2, obtained from the simulation of the second Hammerstein system was used to produce another model of the same phase distortion, called calibration set 2. Table 8.5 shows that both calibration sets allow to invert the phase distortion of the other simulation such that a good model is obtained in both cases.

Formally, the wave phase calibration is defined as follows:

**Algorithm 10. Wave Phase Calibration**
Step 10.1 Measure a nonlinear standard element in the same conditions as the DUT (same excitation spectrum, $f_D$ and $f_S$).

Step 10.2 Parameterize the phase distortion (Section 8.6.2).

Step 10.3 Choose initial values for the distortion, i.e. $H_{RF} = 0$ and $H_{IF} = 0$.

Step 10.4 Invert the model of the phase distortion to guess the true input and output data of the nonlinear standard element.

Step 10.5 Estimate a model for the nonlinear standard element with the guessed input/output data.

Step 10.6 Unless the output simulation error of the nonlinear standard element’s model doesn’t drop anymore, propose new parameter values and continue with Step 10.4.

The implementation of Algorithm 10 uses the Optimization Toolbox for use with Matlab. This toolbox provides the update of the parameter values for $H_{RF}$ and $H_{IF}$ in Step 10.6. The criterion optimized by this toolbox is the output simulation error of the model obtained by Algorithm 2.

8.7 Conclusion

A method has been proposed to obtain a wave phase calibration for broad band non harmonically related signals at high frequencies. Its feasibility has been tested on simulations. In this area, there is still future work to do: a device with a known structure has to be found and the robustness with respect to noise has to be verified. The superconductor from NIST should be used first, but the low temperatures at which it operates make it unwieldy.
9 Conclusions and Ideas for Future Research

9.1 Identification

The identification of the class of nonlinear systems that consist of the cascade of 2 or 3 linear or static nonlinear blocks has been carried out. The proposed identification techniques for these systems is based on the foundations provided by linear system identification theory.

First, iterative identification procedures have been proposed for the simpler systems consisting of two blocks, namely the Wiener systems (a Linear Time Invariant (LTI) system followed by a static nonlinearity) and the Hammerstein systems (a static nonlinearity followed by an LTI system). These initializations converge to the true solution in the absence of noise and model errors. For Hammerstein systems, the iterative estimation algorithm can also be shown to be consistent. For Wiener systems, only the first non-iterative estimation step can be shown to be consistent.

For the 3 block models, two different types of cascades exist: the Hammerstein-Wiener system (an LTI system in between two static nonlinearities) and the Wiener-Hammerstein system (a static nonlinearity in between two LTI systems). The identification procedure for Hammerstein-Wiener systems has been shown to converge to the true system in the absence of noise and model errors. This estimator was shown to be inconsistent.

The difficult problem of estimating Wiener-Hammerstein systems has been addressed next. The main problem that had to be overcome is the separation of the two linear system responses. This has been achieved by using a special kind of excitation signal,
namely (special) odd multisines. The resulting model is then used successfully as the initialization of an output error estimator.

The previously identified structures are then connected in parallel to obtain a more general model structure. The goal here is twofold: first, proposing a valid model structure for the Device Under Test (DUT) and second, improving easily on the linear estimate. Indeed, by comparing the contribution of each of the three branches, it can be checked if the DUT can be best described by a Wiener, a Hammerstein or a Wiener-Hammerstein system. The obtained model structure can then be estimated more precisely with one of the previously defined methods. This structure is more general than the previous ones, as it is able to simulate certain nonlinear systems with feedback.

9.2 Measurements

The robustness of the methods with respect to the non ideality of the excitation signal has been verified. Additionally, the need for a measurement instrument that measures spectrally “rich” signals without distorting the time domain shape of the signals has been demonstrated. A calibration has been proposed to correct for shortcomings of measurement systems at high frequencies because the frequency response calibration that is used for linear modeling is insufficient for nonlinear applications. This method for calibrating the phase of broadband measurements has been tested on simulations.

9.3 Future Work

9.3.1 Extension to MIMO

All systems studied here are Single Input, Single Output (SISO). The extension to Multiple Input, Multiple Output (MIMO) poses interesting challenges, as e.g. the number of basis functions (and thus parameters to estimate) for the static nonlinearities grows exponentially. Techniques to prune the unnecessary basis functions should be investigated.
9.3 Future Work

Figure 9.1: How much energy is there in the stochastic nonlinear contributions for monotonic static nonlinearities inside these bounds?

9.3.2 Extension to closed loop nonlinear systems

This thesis started by restricting the model structure to a class of open loop nonlinear systems in order to be able to say something useful about the DUT (Chapters 1 and 2). The extension to closed loop nonlinear systems would allow to model more devices. A physical example of such a device is the pendulum.

9.3.3 Gain insight usable for design

Ideally, one would like to have to work only with linear devices. However, in reality, systems have a nonlinear behavior. Sometimes, it could be desired to make a design assuming that the DUT is linear and that the nonlinear contributions can be seen as noise. A question then arises: when the amplitude or the energy of the input to the DUT is increased, how much does the “noise” (actually the nonlinear contributions) increase?

This question can probably not be answered without additional information, like the probability density function of the input, the monotonicity of the nonlinear function and possibly other assumptions that would be needed.

Even better would be to be able to answer the same question when only an uncertain model of the static nonlinearity is available, e.g. one knows that the static nonlinearity is smooth, monotonous and inside the bounds shown in Figure 9.1. This would allow to decide how far the DUT may be pushed into its nonlinear region if only an approximation of the nonlinearity is available and if it is acceptable to have a certain percentage of nonlinear contributions disturbing a dominantly linear system.

9.3.4 Try the wave phase calibration on measurements

The calibration method proposed in Chapter 8 should be tested on measurements.


9.3.5 User friendliness

The identification algorithms proposed in this thesis have been implemented to share a reasonably similar interface. However, it is clear that their use requires some in-depth knowledge of the methods. Hence, they are not ready yet to be shipped to an average “push on the button” computer user. Especially for Wiener-Hammerstein systems, the identification procedure needs a custom built excitation signal that is to be automatically designed, based on the measurement setup of the user. The choice of some algorithm tuning parameters also requires further automation. Chapter 6 is a first step in the direction of an automatic extraction, as it can help with the model selection problem.
Bibliography


Silva, Christopher P., Andrew A. Moulthrop and Michael S. Muha (2001). Introduction to polyspectral modeling and compensation techniques for wideband communications systems. In: *58th ARFTG Conference Digest, San Diego, California, USA*. 

219


