Linköping Studies in Science and Technology. Licentiate Thesis.
No. 1605

Estimation of Inverse Models Applied to Power Amplifier Predistortion

Ylva Jung

Division of Automatic Control
Department of Electrical Engineering
Linköping University, SE-581 83 Linköping, Sweden
http://www.control.isy.liu.se
ylvju@isy.liu.se

Linköping 2013

This is a Swedish Licentiate's Thesis.

Swedish postgraduate education leads to a Doctor's degree and/or a Licentiate's degree.
A Doctor's Degree comprises 240 ECTS credits (4 years of full-time studies).
A Licentiate's degree comprises 120 ECTS credits,
of which at least 60 ECTS credits constitute a Licentiate's thesis.

Linköping Studies in Science and Technology. Licentiate Thesis.
No. 1605

Estimation of Inverse Models
Applied to Power Amplifier Predistortion

Ylva Jung

ylvju@isy.liu.se
www.control.isy.liu.se
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping
Sweden

ISBN 978-91-7519-571-1 ISSN 0280-7971 LIU-TEK-LIC-2013:39

Copyright © 2013 Ylva Jung

Printed by LiU-Tryck, Linköping, Sweden 2013

Till Daniel

Abstract

Mathematical models are commonly used in technical applications to describe the behavior of a system. These models can be estimated from data, which is known as *system identification*. Usually the models are used to calculate the output for a given input, but in this thesis, the estimation of *inverse models* is investigated. That is, we want to find a model that can be used to calculate the input for a given output. In this setup, the goal is to minimize the difference between the input and the output from the cascaded systems (system and inverse). A good model would be one that reconstructs the original input when used in series with the original system.

Different methods for estimating a system inverse exist. The inverse model can be based on a forward model, or it can be estimated directly by reversing the use of input and output in the identification procedure. The models obtained using the different approaches capture different aspects of the system, and the choice of method can have a large impact. Here, it is shown in a small linear example that a direct estimation of the inverse can be advantageous, when the inverse is supposed to be used in cascade with the system to reconstruct the input.

Inverse systems turn up in many different applications, such as sensor calibration and power amplifier (PA) predistortion. PAs used in communication devices can be nonlinear, and this causes interference in adjacent transmitting channels, which will be noise to anyone that transmits in these channels. Therefore, linearization of the amplifier is needed, and a prefilter is used, called a *predistorter*. In this thesis, the predistortion problem has been investigated for a type of PA, called *outphasing power amplifier*, where the input signal is decomposed into two branches that are amplified separately by highly efficient nonlinear amplifiers, and then recombined. If the decomposition and summation of the two parts are not perfect, nonlinear terms will be introduced in the output, and predistortion is needed.

Here, a predistorter has been constructed based on a model of the PA. In a first method, the structure of the outphasing amplifier has been used to model the distortion, and from this model, a predistorter can be estimated. However, this involves solving two nonconvex optimization problems, and the risk of obtaining a suboptimal solution. Exploring the structure of the PA, the problem can be reformulated such that the PA modeling basically can be done by solving two least-squares (LS) problems, which are convex. In a second step, an analytical description of an ideal predistorter can be used to obtain a predistorter estimate. Another approach is to compute the predistorter without a PA model by estimating the inverse directly. The methods have been evaluated in simulations and in measurements, and it is shown that the predistortion improves the linearity of the overall power amplifier system.

Populärvetenskaplig sammanfattning

Matematiska beskrivningar, här kallade modeller, används i många tekniska tillämpningar. Ett exempel är utveckling av bilar, där man med simuleringar kan utvärdera olika designval på ett kostnadseffektivt sätt. Ett annat är flygtillämpningar där riktiga tester på flygplanet skulle kunna leda till fara för piloten. Dessa modeller kan skattas från uppmätt data från systemet, och detta förfarande kallas systemidentifiering. Ett system är den avgränsade del av världen som vi är intresserade av, i exemplen ovan bilen och flygplanet. I systemidentifiering är målet att finna en modell som så bra som möjligt beskriver utsignalen, baserat på tidigare in- och utsignaler som har kunnat mätas. I denna avhandling undersöks hur *inversa* modeller kan skattas. Här ska inversen användas i kombination med det ursprungliga systemet, med målet att utsignalen från de seriekopplade systemen (det ursprungliga och dess invers) ska vara densamma som insignalen.

Skattning av inversa system kan göras på flera sätt. Inversen kan baseras på en modell av systemet som sedan inverteras, eller skattas direkt genom att insignalen och utsignalen byter plats i systemidentifieringsproblemet. Hur inversen skattas påverkar modellen genom att olika egenskaper hos systemet fångas, och detta kan därför ha en stor inverkan på slutresultatet. I ett litet förenklat exempel visas att det kan löna sig att skatta inversen direkt när den ska användas i serie med systemet för att återskapa insignalen.

Linjärisering av effektförstärkare är ett exempel där inversa system används. Effektförstärkare används i många tillämpningar, bland annat mobiltelefoni, och dess uppgift är att förstärka en signal vilket är ett steg i överföringen av information. I exemplet med mobiltelefoner kan det exempelvis vara en persons röst som är signalen, vilken ska överföras från telefonen via luften och vidare till mottagaren. Effektförstärkare kan vara olinjära, vilket medför att de sprider effekt till närliggande frekvensband. För den som ska sända i dessa frekvensband uppfattas detta som brus, och det finns gränser för hur mycket spridning som får ske. För att uppfylla dessa krav på spridning krävs alltså ofta linjärisering. Genom att modellera förstärkarens olinjäriteter och invertera dem kan man få ett system som inte sprider effekt i frekvensbandet. I detta sammanhang säger man att en förkompensering, kallad fördistortion, används.

I denna avhandling tillämpas fördistortion på en typ av effektförstärkare, som kallas outphasing-förstärkare. Detta är en olinjär effektförstärkarstruktur som delar upp signalen i två delar, där delarna förstärks av effektsnåla förstärkare för att sedan adderas ihop. Om denna uppdelning och summation inte är perfekta uppstår olinjäriteter, och fördistortion krävs. Fördelen med olinjära effektförstärkare är att dessa kan göras mer effektsnåla, vilket direkt speglas i exempelvis batteritiden för en mobiltelefon.

Här presenteras flera olika metoder för att ta fram en fördistortion. Metoderna har utvärderats på fysiska förstärkare i mätningar, vilka visar att en förbättring kan uppnås vid användning av fördistortion.

Acknowledgments

I want to start by expressing my deepest gratitude to my supervisor Dr. Martin Enqvist. Your never-ending knowledge, patience and encouragement is rather remarkable. The (almost) infinite amount of comments and questioning on every detail is very much appreciated. Thank you so much for the help and the time.

I also want to thank Prof. Lennart Ljung for letting me join this great group. The way you and your successor Prof. Svante Gunnarsson lead the group in large and also handle smaller matters is impressing. All administrative help from Ninna Stensgård and her predecessor Åsa Karmelind is also appreciated.

Without Dr. Jonas Fritzin and Prof. Atila Alvandpour, I would not have gotten into this field of research, I appreciate the nice cooperation and especially Jonas for answering all my questions about the hardware.

Many thanks to Lic. Patrik Axelsson, Lic. Daniel Eriksson, M.Sc. Manon Kok, Lic. Roger Larsson, Dr. Christian Lyzell and M.Sc. Maryam Sadeghi Reineh for proofreading parts of this thesis, your comments have been very valuable to clarify and improve the thesis. I am also grateful to the LaTeX gurus Dr. Gustaf Hendeby and Dr. Henrik Tidefelt for making the great template so I don't have to do much more than writing. And whenever I do need help, Lic. (soon to be Dr.) Daniel Petersson always comes to the rescue. Thank you so much!

I am very happy to have happened to end up in the group of Automatic Control. The fika-breaks are always nice, whether the discussions concern work matters or something completely different. I hope there will be more great beer drunk and barbecues had, and that the girl-power group will continue and make us even more fit and powery. Special thanks to my former room mate Patrik Axelsson for taking care of me when I was new and lost and helping me with whatever, and to my present room mate Maryam Sadeghi Reineh for dragging me along to fika ;-)

Thanks to life (though not appreciated at the time) for making me see what is important and that this is not necessarily what I had in mind.

This work has been supported by the Excellence Center at Linköping-Lund in Information Technology (ELLIIT), the Center for Industrial Information Technology at Linköping University (CENIIT) and the Swedish Research Council (VR) Linneaus Center CADICS, which is gratefully acknowledged.

I am also very grateful that I have family and friends outside of work. Friends in Linköping, those who left, those from back home and from over the world, I'm so glad that you are still in my life. I just wish there were more time for us and shorter distances!

And most of all, to Daniel, the one who puts up with me the most and knows how to get me up in the morning. I'm so glad you're mine! Thank you for your patience, love and encouragement :)

Linköping, August 2013

Ylva Jung

Contents

Notation **xv**

1 Introduction **1**
1.1 Research Motivation 1
1.2 Outline 3
1.3 Contributions 4

I System Inversion

2 Introduction to Model Estimation **9**
2.1 System Identification 10
2.2 Transfer Function Models 11
2.3 Prediction Error Method 12
2.4 Linear Regression 13
2.5 Least-squares Method 14
2.6 The System Identification Procedure 15

3 Introduction to System Inversion **17**
3.1 Inversion by Feedback 18
3.1.1 Feedback and Feedforward Control 19
3.1.2 Iterative Learning Control 20
3.1.3 Exact Linearization 21
3.2 Analytic Inversion 22
3.2.1 Problems Occurring with System Inversion 22
3.2.2 Postinverse and Preinverse 23
3.2.3 Volterra Series 24
3.3 Inversion by System Simulation 26
3.3.1 Separation of a Nonlinear System 26
3.3.2 Hirschorn's Method 27

4 Estimation of Inverse Models **33**
4.1 System Inverse Estimation 34

4.2 Inverse Identification of LTI Systems 35
4.3 An Illustrative Linear Dynamic Example 37
4.4 Inverse Identification of Nonlinear Systems 39

II Power Amplifier Predistortion

5 Power Amplifiers 47
5.1 Power Amplifier Fundamentals 47
5.1.1 Basic Transmitter Functionality 48
5.2 Power Amplifier Characterization 50
5.2.1 Gain 51
5.2.2 Efficiency 51
5.2.3 Linearity 52
5.3 Classification of Power Amplifiers 55
5.3.1 Transistors 55
5.3.2 Linear Amplifiers 56
5.3.3 Switched Amplifiers 57
5.3.4 Other Classes 58
5.4 Outphasing Concept 58
5.5 Linearization of Power Amplifiers 61
5.5.1 Volterra series 62
5.5.2 Block-oriented Models 62
5.5.3 Outphasing Power Amplifiers 63

6 Modeling Outphasing Power Amplifiers 65
6.1 An Alternative Outphasing Decomposition 65
6.2 Nonconvex PA Model Estimator 67
6.3 Least-squares PA Model Estimator 69
6.4 PA Model Validation 71
6.5 Convex vs Nonconvex Formulations 82
6.6 Noise Influence 82
6.7 Memory Effects and Dynamics 84

7 Predistortion 85
7.1 A DPD Description 85
7.2 The Ideal DPD 87
7.3 Nonconvex DPD Estimator 88
7.4 Analytical DPD Estimator 89
7.5 Inverse Least-Squares DPD Estimator 90
7.6 Simulated Evaluation of Analytical and LS Predistorter 94
7.7 Recursive Least-Squares and Least Mean Squares 99

8 Predistortion Measurement Results 101
8.1 Signals Used for Evaluation 101
8.2 Measurement Setup 103
8.3 Evaluation of Nonconvex Method 104

8.3.1 Measured Performance of EDGE Signal 105
8.3.2 Measured Performance of WCDMA Signal 105
8.3.3 Summary 107
8.4 Evaluation of Least Squares PA and Analytical Inversion Method . 109
8.4.1 Measured Performance of WCDMA Signal 110
8.4.2 Measured Performance of LTE Signal 111
8.4.3 Evaluation of Polynomial Degree 115
8.4.4 Summary 115

9 Concluding Remarks 117
9.1 Conclusions 117
9.2 Open Questions 118

A Power Amplifier Implementation 119
A.1 +10.3 dBm Class-D Outphasing RF Amplifier in 90 nm CMOS . . 119
A.2 +30 dBm Class-D Outphasing RF Amplifier in 65 nm CMOS 121

Bibliography 125

Notation

Outphasing Amplifiers

Notation	Meaning
$\Delta_\psi(s_1, s_2)$	$\arg(s_1) - \arg(s_2)$, angle difference between outphasing signals, defined on page 65
Δ_ψ	same as $\Delta_\psi(s_1, s_2)$
$\Delta_\psi(s_{1,P}, s_{2,P})$	angle difference between predistorted outphasing input signals
$\Delta_\psi(y_{1,P}, y_{2,P})$	angle difference between predistorted outphasing output signals
ξ_k	angle difference between $\tilde{s}_k$ and s_k, defined in (6.6)-(6.7), page 67, and Figure 6.1, page 66
f_k	phase distortion in the amplifier branch k, defined in (6.9)
g_1, g_2	gain factors of each branch in PA, should ideally be $g_1 = g_2 = g_0$
h_k	phase predistorter functions in the amplifier branch k, defined in (7.1)
s_k	outphasing input signals, decomposed in standard way (5.11)
$s_{k,P}$	predistorted outphasing input signal in branch k, decomposed with identical gain factors using (5.11)
$\tilde{s}_k$	outphasing input signal in branch k, decomposed with nonidentical gain factors using (6.3)
y_k	outphasing output signal in branch k, decomposed with nonidentical gain factors using (6.3)
$y_{k,P}$	predistorted outphasing output signal in branch k, decomposed with nonidentical gain factors using (6.3)
$\hat{x}$	an estimate of the value of x

Power Amplifier Glossary

Notation	Definition
ACLR, ACPR	adjacent channel leakage (power) ratio, a linearity measure that describes the amount of power spread to neighboring channels, page 52.
AM-AM, AM-PM	amplitude modulation to amplitude modulation or phase modulation, respectively, a plot mapping the output amplitude (or phase distortion) to the input amplitude to determine the distortion induced by the circuit, for example a power amplifier, page 52.
combiner	the circuit that handles the addition of signals in, for example, Figure 5.13, page 59.
dBc	decibel to carrier, the power ratio of a signal to a carrier signal, expressed in decibels.
dBm	power level expressed in dB referenced to one milliwatt, so that zero dBm equals one mW and one dBm is one decibel greater (about 1.259 mW).
DE, PAE	drain efficiency and power added efficiency are efficiency measures for power amplifiers, page 51.
DLA, ILA	direct and indirect learning architectures are two approaches to estimate a power amplifier predistorter, see Method B and Method C on page 34.
DPD	digital predistortion, a linearization technique for power amplifiers that modifies the input to counteract power amplifier distortion from nonlinearities and dynamics, page 61.
DR	dynamic range, defining the ratio of the maximum and minimum output amplitudes an amplifier can achieve, page 60.
IQ	a signal separation into an imaginary part (quadrature, Q) vs real part (in-phase, I), page 48.
LO	local oscillator, a circuit that produces a continuous sine wave. Usually drives a mixer in a transmitter/receiver, page 48.
mixer	translates the signal up or down to another frequency, page 48 and Figure 5.2.
outphasing, LINC	an outphasing amplifier, also called linear amplification with nonlinear components, is a nonlinear amplifier structure.
PA	power amplifier, used to increase the power of a signal, so that the output is a magnified replica of the input.
RF	radio frequency, ranging between 3 kHz and 300 GHz.
SCS	signal component separator, (here) decomposes the signal into outphasing signals according to (5.11).

Abbreviations A-O

Abbreviation	Meaning
AC	Alternating current
ACLR	Adjacent channel leakage ratio
ACPR	Adjacent channel power ratio
AM	Amplitude modulation
AM-AM	Amplitude modulation to amplitude modulation
AM-PM	Amplitude modulation to phase modulation
BJT	Bipolar junction transistor
CMOS	Complementary metal-oxide-semiconductor
DAC	Digital-to-analog converter
DB	Digital baseband
DC	Direct current
DE	Drain efficiency
DLA	Direct learning architecture
DPD	Digital predistortion or predistorter
DR	Dynamic range
EDGE	Enhanced data rates for GSM evolution
EVM	Error vector magnitude
FPGA	Field programmable gate array
FET	Field-effect transistor
FIR	Finite impulse response
FM	Frequency modulation
GSM	Global system for mobile communications
GPRS	General packet radio service
IIR	Infinite impulse response
ILA	Indirect learning architecture
ILC	Iterative learning control
IQ	in-phase component (I, real part) vs quadrature component (Q, imaginary part)
LINC	Linear amplification with nonlinear components
LMS	Least mean squares
LO	Local oscillator
LS	Least squares
LTE	Long term evolution
LTI	Linear time invariant
LUT	Look-up table
MIMO	Multiple-input multiple-output
MOSFET	Metal-oxide-semiconductor field-effect transistor
NMOS	N-channel metal-oxide-semiconductor

Abbreviations P-Z

Abbreviation	Meaning
PA	Power amplifier
PAE	Power added efficiency
PAPR	Peak-to-average power ratio
PD	Predistortion or predistorter
PEM	Prediction-error (identification) method
PM	Phase modulation
PMOS	P-channel metal-oxide-semiconductor
PVT	Process, voltage and temperature
PWM	Pulse-width modulated
RBW	Resolution bandwidth
RF	Radio frequency
RLS	Recursive least squares
RMS	Root mean square
RX	Receiver
SCS	Signal component separator
SISO	Single-input single-output
SLS	Separable least-squares
TX	Transmitter
WCDMA	Wideband code-division multiple access

1

Introduction

Inverse systems and models thereof show up in numerous applications. This entails a need for estimation of models of inverse systems. The concept of building models based on measured data is called system identification, and many theoretical results exist concerning the properties of the estimated models. However, less is known when the goal is to estimate the inverse. Should it be based on a forward model, or should the inverse be estimated directly? Inverse models produced in different ways will capture different properties of the system and more insights are needed.

In this chapter, a short research motivation will be given, followed by an outline of the thesis. Then follows an overview of the contributions of the thesis, and some clarifications of the author's role in the work.

1.1 Research Motivation

Power amplifiers (PAs) are used in many applications, such as communication devices (mobile phones) and loudspeakers. In a hand-held device such as a mobile phone, the power efficiency is an important property as it will reflect directly on the battery time. In order to match the increasing demand for lower power consumption, nonlinear power amplifiers have been developed. These nonlinear PAs can be made more power efficient than linear ones, but introduce other problems. A nonlinear device will not only transmit power in the frequency band where the input signal is, but also risks spreading power to neighboring transmitting channels. For anyone transmitting in these frequency bands, this will be perceived as noise. Therefore, there are standards describing the amount of power that is allowed to be spread to adjacent frequencies. So, for the power amplifier to be useful, linearization is needed, limiting the interference in the neighboring channels. Since the distorted output of the power amplifier is an amplified version

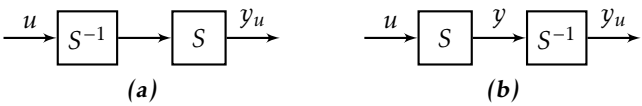


Figure 1.1: *An inverse S^{-1} of the system S is used to undo the effects of the system S such that $y_u = u$. In (a), a preinverse is used, where the inverse S^{-1} is applied before the system S, and in (b), a postinverse is applied, where the order of the system and the inverse is reversed.*

of the input, it is preferable to work with the input. Thus, the goal is to find a prefilter that inverts the nonlinearities, called a *predistorter*.

Small loudspeakers, in mobile phones for example, can also show a nonlinear behavior due to limitations in the movement of the cone. This will distort the sound and make listening to music less agreeable. The idea behind a compensation of the nonlinearities is similar to that of the power amplifier predistortion, since only the input is available for modification. Once the signal has been converted to sound, it cannot be altered.

In the power amplifier and loudspeaker applications, the goal is to find a prefilter that inverts the nonlinearities introduced by the power amplifier or loudspeaker, but the same type of inversion problems can be found also in other areas. In sensor applications it is rather a postdistortion that is needed. If the sensor itself has dynamics or a nonlinear behavior, the sensor output is not the true signal but will also contain some sensor contamination. This has to be handled at the sensor output since this is where the user can get access to the signal.

The need for calibration is also relevant in other applications, such as *analog-to-digital converters* (ADCs). In an ADC, an analog (continuous) input signal is converted to a digital output, which is limited to a number of discrete values. A small error in the analog input risks causing a larger error in the output, since the discrete signal is limited to certain values.

Inversion of systems also appear in other areas, not directly connected to pre- or postinversion. One application where models of both the system S and its inverse S^{-1} are used is robotics. The forward kinematics, describing how to compute the robot tool pose as a function of the joint variables, is used for control as well as the inverse kinematics, how to compute joint configuration from a given tool pose.

In all of the above applications, the question is how to find an inverse S^{-1} to the system S. The application will determine if it is a *pre*inverse or a *post*inverse that is desired. In Figure 1.1, the two different approaches are illustrated.

If an inverse cannot be found analytically, it can be estimated. This opens up for questions regarding this inverse estimation. Different methods can be applied. Either it can be based on an inverted model of the system itself, or the method can try to estimate the inverse directly. That the choice of estimation method matters is motivated by Example 1.1.

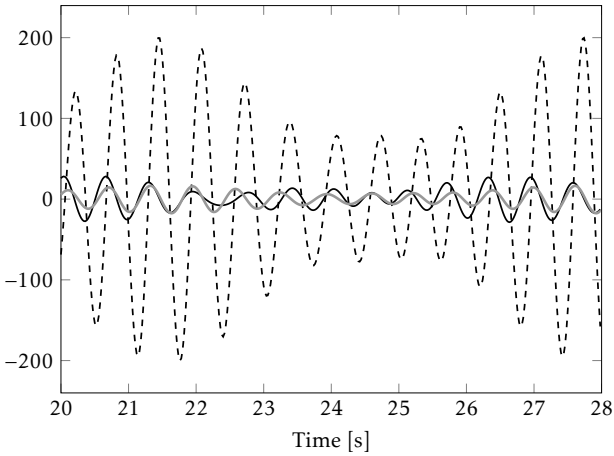


Figure 1.2: *The input u (black solid line), and the reconstructed input y_u using an inverted estimated forward model (black dashed line) and the inverse model estimated directly (gray solid line). The estimation of the inverse (gray) cannot perfectly reconstruct the input (black solid), but is clearly better than the inverted forward model (dashed).*

Example 1.1: Introductory example

Consider a *linear time-invariant* (LTI) system. The goal is to reconstruct the input by modifying the measured output. When the structure of the inverse is set, in this case to a *finite impulse response* (FIR) system, what is the best way to estimate it? Should the inverse be estimated directly or should an inverted model of the system itself be used? These two approaches have been applied to noise-free data, and the results are presented in Figure 1.2. We see here that the two models, both descriptions of the system inverse, capture very different aspects of the system, and that the method chosen can have a large impact. This example is described in more details in Section 4.3.

1.2 Outline

The thesis is divided into two parts. The first introduces system inversion and the estimation of inverse models. The second part concerns using estimated inverse models for power amplifier predistortion.

Part I – System inversion gives a background to the problem of estimating inverse models. A short introduction to model estimation is provided in Chapter 2 and a background to system inversion in Chapter 3. In Chapter 4, some ideas concerning the estimation of inverse models are presented, and three basic approaches are explained. In particular, some conclusions concerning linear, time-invariant systems are presented.

In *Part II – Power amplifier predistortion*, the estimation of inverse models is applied to outphasing power amplifiers. Here, the goal is to find an inverse such that the output of the power amplifier is an amplified replica of the input, counteracting the distortion caused by the amplifier. An introduction to power amplifier functionality and characterization is given in Chapter 5 as well as an overview of earlier predistortion methods. This chapter also contains a description of the outphasing power amplifier, which is a nonlinear amplifier structure that needs predistortion, and for which the predistorter methods in this thesis were produced. Modeling approaches for the power amplifier are presented in Chapter 6 and methods for finding a predistorter in Chapter 7. The predistortion methods are evaluated on real power amplifiers in Chapter 8.

The thesis is concluded by Chapter 9 where some conclusions and a discussion on ideas for future research are presented. Some additional information about the power amplifiers used is given in the appendix.

1.3 Contributions

The contributions in this thesis are in two areas, power amplifier predistortion and the more general field of estimating models of inverse systems.

The power amplifier predistortion was first presented in

> Jonas Fritzin, Ylva Jung, Per N. Landin, Peter Händel, Martin Enqvist, and Atila Alvandpour. Phase predistortion of a Class-D outphasing RF amplifier in 90nm CMOS. *IEEE Transactions on Circuits and Systems-II: Express Briefs*, 58(10):642–646, October 2011a.

where a novel model structure for the outphasing power amplifiers was used. A predistorter that changes only the phases of the outphasing signals was shown to successfully reduce the distortion introduced by the power amplifier. The proposed model and predistorter structures were produced in close collaboration between the paper's first three authors. The theoretical motivation of the predistorter model has been developed by the author of this thesis.

The nonconvex predistortion method presented in the above publication was then developed into a method that makes use of the structure of the outphasing power amplifier. It basically consists of solving least-squares problems, which are convex, and performing an analytical inversion, and it is suitable for online implementation. This is presented in

> Ylva Jung, Jonas Fritzin, Martin Enqvist, and Atila Alvandpour. Least-squares phase predistortion of a +30dbm Class-D outphasing RF PA in 65nm CMOS. *IEEE Transactions on Circuits and Systems-I: Regular papers*, 60(7):1915–1928, July 2013.

The derivation of this least-squares predistortion method has mainly been done by the author of this thesis, whereas the paper's second author has been responsible for the power amplifier and hardware issues. In addition to the reformulation of the nonconvex problem, the paper provides a theoretical description of

an ideal outphasing predistorter, that is, one that does not change neither the amplitude nor the phase of the output. This involves a mathematical description of the branch decomposition and the impact of unbalanced amplification in the two branches.

Inverse models can either be estimated directly, or based on a model of the (forward) system. Some insights into different approaches to estimate models of inverse systems are discussed in Chapter 4. These results concerning the estimation of inverse models have been accepted for presentation at the 52nd IEEE Conference on Decision and Control (CDC):

> Ylva Jung and Martin Enqvist. Estimating models of inverse systems. In *52nd IEEE Conference on Decision and Control (CDC)*, Florence, Italy, To appear, December 2013.

The paper also contains the postinverse application of Hirschorn's method presented in Section 3.3.2.

The contents of Appendix A are included here for the sake of completeness and are not part of the contributions of this thesis. The power amplifiers and the characterization thereof were done at the Division of Electronic Devices, Department of Electrical Engineering at Linköping University, Linköping, Sweden, by Jonas Fritzin, Christer Svensson and Atila Alvandpour.

Part I

System Inversion

2

Introduction to Model Estimation

In many cases it is costly, tedious or dangerous to perform real experiments, but we still want to extract information somehow. The limited part of the world that we are interested in is called a *system*. This system can be pretty much anything. It can for example be interesting for a car manufacturer to know how the car will react to a change in the accelerator. Or in a paper mill, how the moist content of the wood will affect the quality of the paper. For a diabetic it is essential to know how the blood sugar level depends on food intake and exercise. A pilot needs to know how an airplane reacts to the control of different rudders, and in economics it is necessary to know how a change in the interest rate will influence the customers' willingness to borrow or save money. What we see as a system depends on the application. In the car analogy, the system can be only the engine, or the whole car. In the blood sugar level we can either be interested only in how food intake effects the glucose levels, or how exercise contributes.

In many of these applications one does not want to perform experiments directly, but instead start the evaluation using simulations. This leads to a need for *models* of the systems. One way is to use *physical modeling* where the models are based on what we know of the system by using the knowledge of, for example, the forces, moments, flows, etc. In the engine example, it is possible to calculate the output and the connection between the accelerator and the engine torque. This method is sometimes called *white box modeling*. Another modeling approach is to gather data from the system and construct a model based on this information. This approach is called *system identification* and will be presented in this chapter.

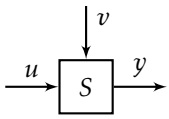


Figure 2.1: *A system S with input u, output y, and disturbance v. For the blood glucose example, the system S is the patient, or rather a part of the body's metabolism system, the input u could represent food intake, the output y is the measured blood glucose level and the disturbance v is for example an infection that effects the body's insulin sensitivity.*

2.1 System Identification

System identification deals with the problem of identifying properties of a system. More specifically, it treats the problem of using measured data to extract a mathematical model of a system we are interested in. The introduction and notation presented here is based on Ljung [1999], but other standard references include Pintelon and Schoukens [2012] and Söderström and Stoica [1989]. Since we are dealing with sampled data, t will be used to denote the time index. Also, for notational convenience, the sample time T_s will be assumed to be one, so that $y(tT_s) \triangleq y(t)$ and $y((t+1)T_s) \triangleq y(t+1)$ is the measurement after $y(t)$, but this can of course easily be adapted to other choices of T_s.

The observable signals that we are interested in are called *outputs*, denoted $y(t)$, and in the examples above this can be the car speed/engine velocity, or the glucose level in the blood. The system can also be affected by different sources that we are in control of – the accelerator or the food intake – called *inputs*, $u(t)$. Other external sources of stimuli that we cannot control or manipulate are called *disturbances*, $v(t)$, – such as a steep uphill affecting the car or a fever or infection which effect the insulin sensitivity. Some disturbances are measurable and for others the effects can be noted, but the signal itself cannot be measured. The different concepts are presented in Figure 2.1.

A system has a number of properties connected to it. A system is *linear* if its output response to a linear combination of inputs is the same linear combination of the output responses of the individual inputs. That is

$$f(\alpha x + \beta y) = f(\alpha x) + f(\beta y) = \alpha f(x) + \beta f(y),$$

with x and y independent variables and α and β real-valued scalars. The first equality makes use of the additivity (also called the superposition property), and the second the homogeneity property. A system that is not linear is called nonlinear. Since this includes "everything else", it is hard to do a classification and come to general conclusions. Most results in system identification are therefore developed for linear systems, or some limited subset of nonlinear systems. The system is *time invariant* if its response to a certain input signal does not depend on absolute time. A system is said to be *dynamical* if it has some memory or history, i.e., the output does not only depend on the current input but also previous inputs and outputs. If it depends only on the current input, it is *static*.

In system identification, the goal is to use the known input data u and the measured output data y to construct a model of the system S. Here, only *single-input single-output* (SISO) systems are considered, but the ideas can most of the time be adapted to *multiple-input multiple-output* (MIMO) systems. It is usually neither possible nor desirable to find a model that describes the whole system and all its properties, but rather one wants to construct a model which captures and can describe some interesting subset thereof, needed for the application. It is up to the user to define such criteria as to what needs to be captured by the model.

2.2 Transfer Function Models

One way to present a *linear time invariant* (LTI) system is via the *transfer function model*

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t) \tag{2.1}$$

where q is the shift operator, such that $qu(t) = u(t+1)$ and $q^{-1}u(t) = u(t-1)$, and $e(t)$ is a white noise sequence. $G(q,\theta)$ and $H(q,\theta)$ are rational functions of q and the coefficients in θ, where θ consists of the unknown parameters that describe the system. Depending on the choice of polynomials in $G(q,\theta)$ and $H(q,\theta)$, different structures can be obtained. The most general structure is

$$A(q)y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t) \tag{2.2}$$

where the polynomials are described by

$$X(q) = 1 + x_1 q^{-1} + \cdots + x_{n_x} q^{-n_x} \quad \text{for} \quad X = A,\ C,\ D,\ F,$$

and n_x is the order of the polynomial and a possible delay n_k in $B(q)$,

$$B(q) = b_{n_k} q^{-n_k} + \cdots + x_{n_k+n_b-1} q^{-(n_k+n_b-1)},$$

such that there can be a delay between input and output. This structure is often too general, and one or several of the polynomials will be set to unity. Depending on the polynomials used, different commonly used structures will be obtained. When the noise is assumed to enter directly at the output, such as white measurement noise, or when we are not interested in modeling the noise, the structure is called an *output error* (OE) model, which can be written

$$y(t) = \frac{B(q)}{F(q)}u(t) + e(t),$$

i.e., the polynomials $A(q)$, $C(q)$ and $D(q)$ have all been set to unity. Many such structures exist (see Ljung [1999] for more examples) and are called *black-box models*, since the model structure reflects no physical insight but acts like a black box on the input, and delivers an output. One strength of these structures is that

they are flexible and, depending on the choice of $G(q,\theta)$ and $H(q,\theta)$, they can cover many different cases.

A model which does not belong to the black-box model structure, and is not completely obtained from physical knowledge of the system is called a *gray-box model*. This can for example be a physical structure with unknown parameters, such as an unknown resistance in an elsewise known circuit. It can also be a some properties of the data that can be explored in the choice of model structure. The latter is done in the power amplifier modeling in Chapter 6.

2.3 Prediction Error Method

In order to say something about the system, we need a model that can predict what will happen next. At the present time instant t, we have collected data from previous time instants $t-1$, $t-2,\dots$, and this can be used to predict the output. The *one-step-ahead predictor* of (2.2) is

$$\hat{y}(t) = \frac{D(q)B(q)}{C(q)F(q)}u(t) + \left[1 - \frac{D(q)A(q)}{C(q)}\right]y(t), \tag{2.3}$$

and depends only on previous output data. The unknown parameters in the polynomials $A(q)$, $B(q)$, $C(q)$, $D(q)$ and $F(q)$ are gathered in the *parameter vector* θ,

$$\theta = [a_1 \dots a_{n_a}\ b_{n_k} \dots b_{n_k+n_b-1}\ c_1 \dots c_{n_c}\ d_1 \dots d_{n_d}\ f_1 \dots f_{n_f}]^T.$$

The predictor $\hat{y}(t)$ is often written $\hat{y}(t|\theta)$ to point out the dependence on the parameters in θ.

By defining the *prediction error*

$$\varepsilon(t) = y(t) - \hat{y}(t|\theta), \tag{2.4}$$

a straightforward modeling approach is to try to find the parameter vector $\hat{\theta}$, that minimizes this difference,

$$\hat{\theta} = \arg\min_{\theta} V(\theta), \tag{2.5a}$$

$$V(\theta) = \frac{1}{N}\sum_{t=1}^{N} l(\varepsilon(t)) \tag{2.5b}$$

where $l(\cdot)$ is a scalar valued, usually positive, function. Finding the parameters by this minimization is called a *prediction-error (identification) method* (PEM). This idea is illustrated in Figure 2.2.

Except for special choices of the model structures $G(q,\theta)$ and $H(q,\theta)$ and the function $l(\varepsilon)$ in (2.5b), there is no analytical way of finding the minimum of the minimization problem (2.5a). Numerical solutions have to be relied upon, which means that a local optimum might be found instead of the global one if the cost function is nonconvex, with more than one minimum. For results on the convergence of the parameters and other properties of the estimate, such as consistency and variance, see Ljung [1999].

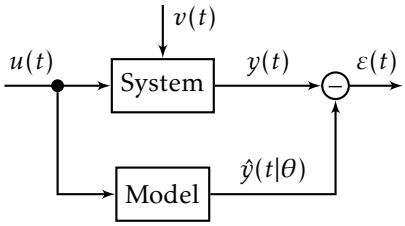


Figure 2.2: *An illustration of the idea behind system identification.*

2.4 Linear Regression

Another common way to describe the relationship between input and output of an LTI system is through a *linear difference equation* where the present output, $y(t)$, depends on previous inputs, $u(t-n_k), \ldots, u(t-n_k-n_b+1)$, and outputs, $y(t-1), \ldots, y(t-n_a)$, as well as the noise and disturbance contributions. This can for example be done for (2.2) when $C(q)$, $D(q)$ and $F(q)$ are set to unity, so that $G(q,\theta)$ and $H(q,\theta)$ in (2.1) correspond to

$$G(q,\theta) = \frac{B(q)}{A(q)}, \quad H(q,\theta) = \frac{1}{A(q)}$$

with

$$\begin{aligned} A(q) &= 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a} \\ B(q) &= b_{n_k} q^{-n_k} + \cdots + b_{n_k+n_b-1} q^{-(n_k+n_b-1)}. \end{aligned}$$

The linear difference equation is then

$$y(t)+a_1 y(t-1)+\cdots+a_{n_a} y(t-n_a) = b_{n_k} u(t-n_k)+\cdots+b_{n_k+n_b-1} u(t-n_k-n_b+1)+e(t),$$

and we can write

$$A(q)y(t) = B(q)u(t) + e(t). \tag{2.6}$$

This particular structure is called *auto-regressive with external input* (ARX). Another special case is when the output only depends on past inputs, such that $n_a = 0$ in (2.6). This is called a *finite impulse response* (FIR) structure.

The predictor for an ARX model is

$$\begin{aligned} \hat{y}(t|\theta) = &-a_1 y(t-1) - \cdots - a_{n_a} y(t-n_a) + \\ &b_{n_k} u(t-n_k) + \cdots + b_{n_k+n_b-1} u(t-n_k-n_b+1). \end{aligned} \tag{2.7}$$

By gathering all the known elements into one vector, the *regression vector*,

$$\phi(t) = [-y(t-1), \ldots, -y(t-n_a)\ u(t-n_k), \ldots, u(t-n_k-n_b+1)]^T$$

and the unknown elements into the parameter vector,

$$\theta = [a_1 \ldots a_{n_a}\ b_{n_k} \ldots b_{n_k+n_b-1}]^T,$$

the predictor (2.7) can be written as a *linear regression*

$$\hat{y}(t|\theta) = \phi^T(t)\theta, \tag{2.8}$$

that is, the unknown parameters in θ enter the predictor linearly.

2.5 Least-squares Method

With the function $l(\cdot)$ in (2.5b) chosen as a quadratic function,

$$l(\varepsilon) = \frac{1}{2}\varepsilon^2,$$

and the predictor described by a linear regression, as in (2.8), we get

$$V(\theta) = \frac{1}{2N}\sum_{t=1}^{N}\left[y(t) - \phi^T(t)\theta\right]^2, \tag{2.9}$$

called the *least-squares* (LS) criterion. A good thing about this criterion is that it is quadratic in θ, which means that the problem is convex and the minimum can be calculated analytically. The minimum is obtained for

$$\hat{\theta}^{LS} = \left[\frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi^T(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\phi(t)y(t), \tag{2.10}$$

called the *least-squares estimator*. See Draper and Smith [1998] for a more thorough description of the LS method and its properties.

Apart from the guaranteed convergence to the global optimum, a benefit with LS solutions is that there exist many efficient numerical methods to solve them. The *recursive least-squares* (RLS) method can be used to solve the numerical optimization recursively [Björck, 1996]. Another option is the *least mean square* (LMS) method, which can make use of the linear regression structure of the optimization problem, developed in (2.8).

Separable Least-squares

For some model structures, the parameter vector can be divided into two parts, $\theta = [\rho^T\ \eta^T]^T$, so that one part enters the predictor linearly and the other nonlinearly, i.e.,

$$\hat{y}(t|\theta) = \hat{y}(t|\rho,\eta) = \phi^T(t,\eta)\rho.$$

Hence, for a fixed η, the predictor is a linear function of the parameters in ρ. The identification criterion is then

$$V(\theta) = V(\rho,\eta) = \frac{1}{2N}\sum_{t=1}^{N}\left[y(t) - \phi^T(t,\eta)\rho\right]^2$$

and this is an LS criterion for any given η. Often, the minimization is done first for the linear ρ and then the nonlinear η is solved for. The nonlinear minimization problem now has a reduced dimension, where the reduction depends on the dimensions of the linear and nonlinear parameters. This method is called *separable least-squares* (SLS) as the LS part has been separated out, leaving a nonlinear problem of a lower dimension, see Ljung [1999, p. 335-336].

2.6 The System Identification Procedure

The process of constructing a model from data consists of a number of steps, which often have to be performed a number of times before a suitable model can be obtained.

1. A data set is needed, usually containing input and output data. The data should be "good enough", so that it excites the desired properties of the system. This is called persistency of excitation.

2. Different model structures should be examined, to evaluate which structure best captures the properties of the data. These structures should fulfill certain demands, such that two sets of parameters do not lead to the same model. This property is called *identifiability*.

3. A measurement of "goodness", such as the criterion (2.5), has to be selected to decide which models best describe the data.

4. The model estimation step is where the parameters in θ are determined. In the LS method, this would consist of inserting the data into (2.10), and in the PEM case, the minimization of (2.5) for a certain choice of predictor structure $\hat{y}(t|\theta)$ in (2.3).

5. Model validation. In this step, different models should be evaluated to determine if the models obtained are good enough. The evaluation should be done on a new set of data, *validation data*, to ensure that the model is useful not only for the data for which it was estimated. Two important components of the model validation are the comparisons between measured data and model output as well as the residual analysis, where the statistics of the unmodeled properties of the data are evaluated.

Some of these steps contain a large user influence, whereas others might be set or rather straightforward. The choice of model structure and model order, such as n_a and n_b in (2.7), is often hard and needs to be repeated a number of times before a suitable model can be found.

3

Introduction to System Inversion

Inverse systems are used in many applications, more or less visibly. One application example of this is power amplifiers in communication devices, which are often nonlinear, causing interference in adjacent transmitting channels [Fritzin et al., 2011a]. This interference will be noise to anyone that transmits in these channels, and there are measures describing the amount of power that is allowed to be spread to adjacent frequencies. So to be useful, linearization of the amplifier is needed, limiting the interference in the neighboring channels. However, one does not want to work with the amplified signal, but rather with the input signal to the system, that is, before the signal is amplified. A prefilter that inverts the nonlinearities, called a *predistorter*, is thus preferable.

In sensor applications it is rather a postdistortion that is needed. If the sensor itself has dynamics or a nonlinear behavior, the sensor output is not the true signal but will also contain some sensor contamination. This would have to be handled at the sensor output, since this is where the user can get access to the signal.

In the area of robotics, there is a need for control such that the robot achieves the demands on precision. Smaller and lighter robots reduces the need for large motors, and also the cost and wear of the robot. However, this also introduces new problems such as larger oscillations and increases the demands on the control performance. In robotic control applications, a common strategy is to use feedback to control the joint positions. The last part of the robot, however, connecting the tool to the robot, is often controlled using open-loop control. Models of both the forward and the inverse kinematics are used for control.

In the above applications, finding the inverse of the system is a crucial point; how should the input to or the output from the system be modified to obtain the desired dynamics from input to output? Each application entails its own restrictions and special conditions to attend to, and in this chapter, some aspects of

system inversion are discussed. For a nonlinear system, the inversion is nontrivial, and different approaches can be used. A selection of methods is presented here.

In this thesis, it is assumed that an inverse exists, that is, there is a one-to-one relation between input and output. This property is called bijectivity. Furthermore, we assume that the system and the inverse can both be written analytically, see Example 3.1 for a case when this is not the valid. Both the system and the inverse are assumed to be stable and causal (see for example Rugh [1996]). In this chapter, the main focus is on inversion, and a model of the system is supposed to be known, either by physical modeling or by system identification. Different approaches to estimate inverse models will be presented in Chapter 4.

Example 3.1: Nonexisting Analytical Inverse

Consider the system

$$y(t) = e^{x(t)} + \sin(x(t))$$

for $|x(t)| < 0.5\pi$. The function $e^x + \sin(x)$ is monotonic on $[-\frac{\pi}{2}\ \frac{\pi}{2}]$, and thus also invertible. However, no analytic expression of the inverse exists, and a numerical inverse will have to be used.

Here, the methods are described in either continuous or discrete time. Different frameworks are usually most easily described in one domain or the other, hence the mixed use in this chapter. Also, the systems are often continuous whereas the controllers are implemented in discrete time. The explicit dependency on time will sometimes be left out for notational convenience.

3.1 Inversion by Feedback

The behavior of a system can be modified in a multitude of ways, often with the goal of making the output follow a desired trajectory, called *reference signal*, r. In the automatic control society the main choices are feedback and feedforward control. For the linear case many different control strategies exist, the perhaps most common of which is the PID, consisting of a proportional (P), an integral (I) and a derivative (D) part. The P, I and D parameters of the controller can be trimmed to obtain a desired behavior of the controlled system [Åström and Hägglund, 2005].

In this section, a few feedback strategies will be introduced. An iterative control approach that can be used for linear and nonlinear systems is the *iterative learning control* (ILC). ILC works on systems with a repetitive input signal, such as a robot that performs the same task over and over again. It makes use of the output from the last repetition and tries to improve this so that the output better follows the reference signal. Another feedback solution for nonlinear systems is the exact input-output linearization, that makes use of a known model of the system to obtain overall linear dynamics, determined by the user.

Though the classical view of feedback control is not that of system inversion, this is indeed one interpretation; the feedback system produces the input that

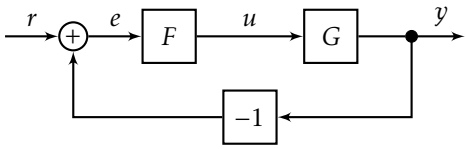


Figure 3.1: *A feedback controller F applied to the system G.*

leads to a desired output. This is also the goal of an inverse system, to produce an input by use of an output. We will start this chapter by covering a few control strategies.

3.1.1 Feedback and Feedforward Control

Feedback control refers to a measured output of a system that is used to determine the input to said system. A standard solution is to look at the difference between the reference $r(t)$ and the output $y(t)$, called *control error* $e(t) = r(t) - y(t)$. This signal can be used for control of the system. For example, if the control error is negative, a conclusion can be that the input u is too small, and should be increased, and vice versa. Many control strategies based on this idea have been constructed and are commonly used in industry. The idea is presented in Figure 3.1 where a feedback controller F is applied to the system G.

On the other hand, if we know something about how the system will transform the input, we might want to use this to counteract later effects. This is the concept of *feedforward control*, where the reference signal is altered and sent to the system, or fed forward. Often, feedforward and feedback control are used together to get the advantages of both approaches. Figure 3.2 shows a block diagram where the feedback loop in Figure 3.1 has been expanded to include a feedforward loop with the feedforward controller F_f. A common requirement is that the output should have a softer behavior than the reference, and this can be achieved via the filter G_m, which denotes the desired dynamics. The ideal choice of the feedforward controller is $F_f = G_m/G$. If feedforward control is used alone, with no feedback loop, it is often called *open-loop control*.

Feedback control can handle phenomena like disturbances and model uncertainties, since it is based on the true output and not only the input and a model of the system. It can also handle unstable systems, which is not possible for a pure feedforward (open-loop) control, but a bad feedback loop may cause instability.

Feedforward control has the advantage of not needing any measurements but the drawback is that ideal feedforward control (using $F_f = G_m/G$) requires perfect knowledge of the system, and that both G and G_m/G are stable. Also, there is no possibility to compensate for disturbances. However, if the disturbances are perfectly known or measurable, feedforward control from the disturbances can be applied and the disturbance compensated for. These are of course limiting assumptions. A benefit with feedforward control is that two cascaded stable systems will always be stable, and a bad controller can therefore not destabilize the system.

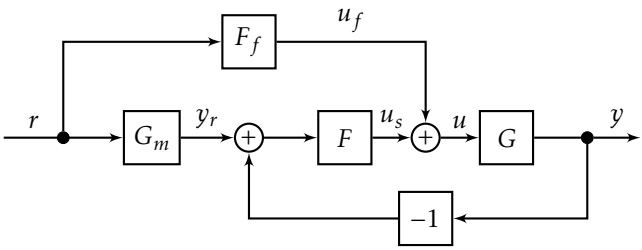


Figure 3.2: *Feedforward controller F_f and feedback controller F applied to a system G. G_m is used here to describe the desired dynamics between reference and output.*

3.1.2 Iterative Learning Control

As discussed in the introduction, *iterative learning control* (ILC) can be seen as an iterative inversion method [Markusson, 2001]; the goal is to find the input that leads to the desired output. In this section, the basic concepts of ILC will be described, but for a more thorough analysis see for example Wallén [2011], Moore [1993] and the references therein. The ILC concept comes from the industrial robot application, where the same task or motion is performed repeatedly. The idea is to use the knowledge of how the controller performed in the last repetition and improve the performance in each iteration.

The system S in this setting is described by the input u, the output y, and the reference r over a finite time interval. The task is assumed to be repeated, so that the reference r and the starting point are the same for each iteration. The time index is t, where $t \in [0, N-1]$ for each repetition, and each repetition is of length N. A basic first order ILC algorithm is described by

$$u_{k+1}(t) = Q(q)\left(u_k(t) + L(q)e_k(t)\right) \tag{3.1}$$

where

$$e_k(t) = r(t) - y_k(t)$$

and k is the iteration index, and indicates how many times the task has been repeated. Here, q is the shift operator such that $q^{-1}u(t) = u(t-1)$ and $Q(q)$ and $L(q)$ denote linear or nonlinear operators, chosen by the user. It is important that this choice leads to convergence and an input where the output achieves the desired performance. Also, the learning should be fast enough. There are structured ways to determine $Q(q)$ and $L(q)$, which can be based on a model of the system. The concepts of stability and convergence of ILC systems are treated in, for example, Wallén [2011]. It can be shown that ILC is robust to model errors, such that for a linear system, a relative model error of 100% can be tolerated [Markusson, 2001]. Even a rather simple model can therefore perform well.

Iterative methods are used in many applications, also outside the control community. The common factor is that the information found in the output y is used to improve the input, but the algorithm is not necessarily similar to (3.1). One

application where iterative solutions are often used is *analog-to-digital converter* (ADC) correction, such as in Soudan and Vogel [2012].

3.1.3 Exact Linearization

In exact linearization (also called input-output linearization) [Sastry, 1999], the output from a nonlinear system S,

$$\begin{aligned} \dot{x} &= f(x) + g(x)u \\ y &= h(x), \end{aligned} \tag{3.2}$$

which is affine in u, is differentiated enough times to obtain a relation between the differentiated output $y^{(n)}$ and the input, u. Differentiating y with respect to time, we obtain

$$\begin{aligned} \dot{y} &= \frac{\partial h}{\partial x} f(x) + \frac{\partial h}{\partial x} g(x)u \\ &= L_f h(x) + L_g h(x)u, \end{aligned}$$

where $L_f h(x)$ and $L_g h(x)$ are the Lie derivatives of h with respect to f and g, respectively. If $L_g h(x) \neq 0$, a relation between the differentiated output $\dot{y}$ and the input u has been obtained and an input,

$$u = \frac{1}{L_g h(x)}(-L_f h(x) + r)$$

can be calculated that leads to a linear relation between output and reference, $\dot{y} = r$. If $L_g h(x) = 0$, a second differentiation can be done,

$$\begin{aligned} \ddot{y} &= \frac{\partial L_f h}{\partial x} f(x) + \frac{\partial L_f h}{\partial x} g(x)u \\ &= L_f^2 h(x) + L_g L_f h(x)u, \end{aligned}$$

from which a control law can be calculated if $L_g L_f h(x) \neq 0$. In this manner, one can continue until there is a direct relation between $y^{(\gamma)}$ and r through the control law

$$\begin{aligned} u &= \frac{1}{L_g L_f^{\gamma-1} h(x)}(-L_f^{\gamma} h(x) + r) \\ &= \alpha(x) + \beta(x)r. \end{aligned} \tag{3.3}$$

Here, γ is the smallest integer for which $L_g L_f^i \equiv 0$ for $i = 0, 1, \ldots, \gamma - 2$ and $L_g L_f^{\gamma-1} h(x) \neq 0$ and it is called the relative degree of the system.

The system (3.2) with control input (3.3) now describes a system with linear dynamics. Thus, linear theory can be used to obtain the desired dynamics, G_m, chosen by the user, and the linear feedback loop can be combined with the nonlinear one. The overall system from r to y (the nonlinear system with the nonlinear

and linear feedback) will thus be linear, and the dynamics will be described by the transfer function G_m.

Exact linearization requires knowledge of all the states, and is therefore often used in combination with a nonlinear observer. This can lead to a complicated feedback loop. Here, it is assumed that any zero dynamics present are stable. The above system and the derivation of the feedback loop is described in continuous time. A discrete-time description can also be done, as presented in Califano et al. [1998].

3.2 Analytic Inversion

In the above feedback loops, only the system itself, or a model thereof, is used to produce an inverse. No explicit inversion is done. Another approach is to perform an analytic inversion of the system, which can be applied at the input to, or the output from, the system, see Figure 1.1. The output from this cascaded system should have the desired dynamics. If the goal is to make the output exactly the same as the reference, a "true" inverse has to be found. But even for other cases, the inversion can be seen as a case where the unwanted nonlinear and linear dynamics have been inverted. For example, in the exact linearization case, the nonlinear and dynamical behavior of the system are inverted, and in the end a system with some user-defined linear dynamics is obtained. This approach has already been used in the feedforward controller $F_f = G_m/G$, where the system G is inverted.

Finding a system inverse can be done in multiple ways. One method for finding an inverse to dynamic systems uses Volterra series, which is a nonlinear extension of the impulse response concept from the linear case. This leads to an analytical inverse. Other systems that might be analytically invertible are block-oriented systems, which consist of a static nonlinearity and a linear dynamic system. A brief overview of Volterra series will be presented here together with a short discussion on the use of preinverse and postinverse and problems that occur with inversion.

3.2.1 Problems Occurring with System Inversion

For a stable and minimum-phase LTI system G, it is rather straightforward to find an inverse G^{-1}. However, if these conditions are not fulfilled, we quickly run into problems, even for linear systems. Any nonminimum-phase zeros of the original system will become unstable poles of the inverse system. However, if the system is nonminimum phase, the inverse can be used if noncausal filtering is allowed. If a delay can be allowed, time-reversed input and output sequences can be used together with a matching, stable inverse [Markusson, 2001].

Another trouble with inverse systems concerns whether the system is proper or not. A proper transfer function is one where the order of the denominator is greater than or equal to that of the numerator. A strictly proper transfer function is one where the order of the denominator is greater than that of the numerator.

The amplification of a proper system always approaches a value as the frequency goes to infinity. If the transfer function is strictly proper, the amplification will approach zero at high frequencies. For a transfer function that is not proper, however, the amplification will approach infinity when the frequency approaches infinity. That is, high frequency contents will be amplified. This means that the inverse of a strictly proper system will be improper.

Here, the goal is not to cover all problems with the inversion of systems, but to give some insights to the problems that can occur.

3.2.2 Postinverse and Preinverse

As is commonly known, the ordering of a linear system does not matter, i.e., the output from $A*B$ equals the output from $B*A$ when A and B are linear dynamical systems. This property is called commutativity. However, this does not apply to nonlinear systems, as shown in Example 3.2.

Example 3.2: Noncommutativity of Nonlinear Systems

Consider the two functions

$$f_1(x) = 2x \quad \text{and} \quad f_2(x) = x^2.$$

If the order of the systems is f_1, f_2, the output is $y_{12} = 4u^2$ and with the reversed order, the output is $y_{21} = 2u^2 \neq y_{12}$.

Thus, for nonlinear systems, the output depends on the order of the systems. For some nonlinear systems, this is not true and the systems can change order without changing the output. One example where two nonlinear systems commute, is where one of the systems is the inverse of the other as in Example 3.3 for a Hammerstein-Wiener system. When an exact inverse exists, the preinverse and the postinverse are the same. However, it is often not possible to determine the exact inverse, and an approximate inverse has to be used. This approximate function does not necessarily commute with the system.

Another example of nonlinear systems that commute are the Volterra series and the p-th order Volterra inverse that will be described in the next section. But, in general, the commutative property does not apply to nonlinear systems. See Mämmelä [2006] for an extended discussion on commutativity in linear and nonlinear systems.

Example 3.3: Analytical Inversion

Consider the Hammerstein system with the static nonlinearity

$$f_H(x) = x^3$$

which is invertible for all x, followed by the minimum-phase linear dynamic system

$$G_H(s) = \frac{s+1}{s+2},$$

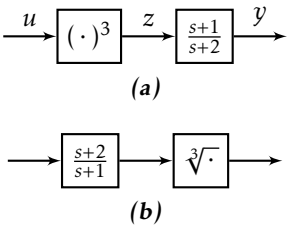


Figure 3.3: *(a) A Hammerstein system with invertible static nonlinearity followed by a linear, stable minimum-phase dynamical system. For such a system, an analytical inverse exists, as shown in (b).*

as shown in Figure 3.3a. For this system, an analytical inverse exists, namely the Wiener system

$$G_W(s) = \frac{s+2}{s+1}, \quad f_W(x) = \sqrt[3]{x},$$

see Figure 3.3b. This inverse is also an example of where a nonlinear system and its inverse are commutative, that is, the two systems can be placed in whichever order. This Wiener system can thus be used as a *pre*inverse or a *post*inverse.

Different approximate modeling approaches, which will be further considered in Chapter 4, lead to either a preinverse or a postinverse that are not necessarily equal. Which one that is requested is connected to the application, such that for power amplifier linearization a preinverse is desired, and for sensor calibration a postinverse. However, for power amplifier predistortion, the commutativity property is often considered approximately valid, and the pre- and postinverses are used interchangeably without further consideration [Abd-Elrady et al., 2008, Paaso and Mämmelä, 2008].

3.2.3 Volterra Series

In the linear systems theory, a common way to describe the output, $y(t)$, of the system affected by the input $u(t)$, is by the impulse response $g(\cdot)$,

$$y(t) = \int_{-\infty}^{\infty} g(\tau)u(t-\tau)\mathrm{d}\tau, \tag{3.4}$$

usually with the added constraints that the system is causal and the input zero for $t < 0$, so that the integral is limited to $[0, t]$. It can also be described by the corresponding Laplace relation

$$Y(s) = G(s)U(s) \tag{3.5}$$

where $Y(s)$ and $U(s)$ are the Laplace transformed versions of $y(t)$ and $u(t)$, respectively, and $G(s)$ is the transfer function. This is not possible for nonlinear systems.

However, if the nonlinear system is time invariant with certain restrictions, an input-output relation can be determined. These conditions include convergence of the infinite sums and integrals that occur [Sastry, 1999], but will not be further considered here. The input-output relation can be described by

$$y(t) = \int_{-\infty}^{\infty} h_1(\tau_1)u(t-\tau_1)\mathrm{d}\tau_1 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)u(t-\tau_1)u(t-\tau_2)\mathrm{d}\tau_1\mathrm{d}\tau_2 + \dots$$
$$+ \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} h_n(\tau_1,\dots,\tau_n)u(t-\tau_1)\dots u(t-\tau_n)\mathrm{d}\tau_1\dots\mathrm{d}\tau_n + \dots \tag{3.6}$$

where

$$h_n(\tau_1,\dots,\tau_n) = 0 \quad \text{for any } \tau_j < 0, \quad j = 1,2,\dots,n.$$

The relation (3.6) is called a Volterra series (sometimes Volterra-Wiener series) and the functions $h_n(\tau_1,\dots,\tau_n)$ are called the *Volterra kernels* of the system. The expression (3.6) can also be written as

$$y(t) = \mathbf{H}_1[u(t)] + \mathbf{H}_2[u(t)] + \dots + \mathbf{H}_n[u(t)] + \dots \tag{3.7}$$

where

$$\mathbf{H}_n[u(t)] = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} h_n(\tau_1,\dots,\tau_n)u(t-\tau_1)u(t-\tau_2)\dots u(t-\tau_n)\mathrm{d}\tau_1\dots\mathrm{d}\tau_n \tag{3.8}$$

is called an n-th order *Volterra operator*.

When considering an LTI *single input-single output* (SISO) system, the Volterra series reduces to the standard form, and the kernel $h_1(\cdot)$ in (3.6) corresponds to $g(\cdot)$ in (3.4). See for example Schetzen [1980] for a more thorough description of Volterra series. The counterpart of the transfer function is based on the multivariable Fourier transform,

$$H_p(j\omega_1,\dots,j\omega_p) = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} h_p(\tau_1,\dots,\tau_p)e^{-j(\omega_1\tau_1+\dots+\omega_p\tau_p)}\mathrm{d}\tau_1\dots\mathrm{d}\tau_p \tag{3.9}$$

called the p-th order kernel transform. The inverse relation is

$$h_p(\tau_1,\dots,\tau_p) = \frac{1}{(2\pi)^p}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} H_p(j\omega_1,\dots,j\omega_p)e^{j(\omega_1\tau_1+\dots+\omega_p\tau_p)}\mathrm{d}\omega_1\dots\mathrm{d}\omega_p. \tag{3.10}$$

In analogy to the linear case, these functions are sometimes referred to as higher order transfer functions. The discrete counterpart of the Volterra operators (3.8) is [Tummla et al., 1997]

$$\mathbf{H}_n[u(t)] = \sum_{i_1=-\infty}^{\infty}\dots\sum_{i_n=-\infty}^{\infty} h^{(n)}_{i_1,i_2,\dots,i_n}u(n-i_1)\dots u(n-i_n). \tag{3.11}$$

This version is often used in data-based modeling, where the models are based on sampled data.

p-th Order Volterra Inverse

A p-th order inverse, $\mathbf{H}_{(p)}^{-1}$, is defined as a system that, when connected in series with the nonlinear system $\mathbf{H}$ results in a system, $\mathbf{Q}$, in which the first-order Volterra kernel is a unit pulse and the other Volterra kernels are zero, $q_k = 0, k = 2, \ldots, p$. The Volterra kernels for $k > p$ might however be nonzero but are generally considered to be negligible [Zhu et al., 2008]. The inverse, $\mathbf{H}_{(p)}^{-1}$, can be determined by using the Volterra series (assumed known) of the system, and the desired output. This is done in a sequential way by first finding the first order Volterra operator, $\mathbf{H}_{(1)}^{-1}$, and then solving for the higher order Volterra operators $\mathbf{H}_{(n)}^{-1}, n = 2, \ldots, p$ which then only depend on the system $\mathbf{H}$ and lower order operators of the inverse, see Schetzen [1980, Chapter 7] for a thorough discussion.

The ordering of the system $\mathbf{H}$ and the inverse $\mathbf{H}_{(p)}^{-1}$ will affect the output, but it can be shown [Schetzen, 1980] that the first p Volterra operators of the connected systems are the same. The order of the system $\mathbf{H}$ and the inverse $\mathbf{H}_{(p)}^{-1}$ can thus be interchanged and the postinverse $\mathbf{H}_{(p)}^{-1}$ can also be used as a preinverse, if only nonlinearities up to order p are of interest.

3.3 Inversion by System Simulation

Some approaches to avoid the explicit inversion of a system are based on a simulation of the true system, without including any feedback from the actual system. The exact linearization described in Section 3.1.3 can be modified such that it uses a simulated output in the feedback loop. Another approach is to decompose the original system to avoid the explicit inversion of the nonlinear system. The idea with these inversion methods is that the inverse can be used as a preinverse or a postinverse, thus avoiding a feedback loop.

3.3.1 Separation of a Nonlinear System

A way to avoid the explicit inversion of a nonlinear system is presented in Markusson [2001, p. 51]. There, the nonlinear system S is separated into a linear part, L, and a nonlinear part, N, where operator notation is used. The inverse of $S = L + (S - L) = L + N = L(I + L^{-1}N)$, can then be written $S^{-1} = (I + L^{-1}N)^{-1}L^{-1}$. We have thus obtained a postinverse S^{-1} such that

$$S^{-1}S = (I + L^{-1}N)^{-1}L^{-1}(L + N) = (I + L^{-1}N)^{-1}(I + L^{-1}N) = I.$$

This can also be used as a preinverse, since

$$SS^{-1} = (L + N)(I + L^{-1}N)^{-1}L^{-1} = L(I + L^{-1}N)(I + L^{-1}N)^{-1}L^{-1} = I.$$

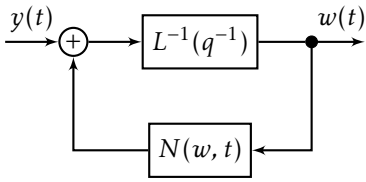


Figure 3.4: *An inversion method that only uses the inverse of the linear part L of a nonlinear system $S = L + N$.*

The inverse $(I + L^{-1}N)^{-1}L^{-1}$ can be obtained in a feedback loop with the nonlinear part N in the feedback and the linear inverse L^{-1} in the forward path (compare to the sensitivity function for LTI systems), see Figure 3.4. It follows that the nonlinear part N does not have to be explicitly inverted, and that only the linear part L is to be inverted. The output from the inverted system is denoted $w(t)$ to separate it from the true input $u(t)$, since different initial conditions of the true system and the model will produce an output that is not exactly equal to the input. Unknown initial states are discussed in Markusson [2001, p. 45], in a *maximum likelihood* (ML) setting.

3.3.2 Hirschorn's Method

Another approach to invert nonlinear systems is Hirschorn's method, where exact linearization is used in order to construct a linear system [Hirschorn, 1979]. Given that the model is good enough, it should be possible to use the model not only in the feedback, as it is used in the construction of u in (3.3), but also as a simulation model.

Preinversion

If instead of the measured output from the system, the output from the simulated model is fed back to the controller, see Figure 3.5, the overall system (from reference r to output y_s) will by construction be linear with the dynamics G_m. Also, the input calculated for this (simulated) system leads to the desired dynamics, and the same input signal can be used also for the true system. The system from r to u will be denoted $S^\dagger$. A pure open loop controller is thus obtained, as in Hirschorn [1979], see Figure 3.6, and this is called *Hirschorn's method*. The simulated feedback can also be interpreted as an observer with no measurement inputs.

Postinversion

Let the nonlinear system be denoted S and the precompensation be denoted $S^\dagger$, since it is not really an inverse of S, but rather creates a system that, in series with S will be linear. The dynamics of the overall linear system is G_m.

The method described above can be seen as an inversion of the nonlinearities of the system – the output from the overall system will be linear with dynamics

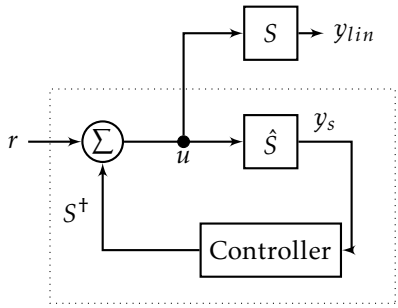


Figure 3.5: *A block diagram of Hirschorn's method, where the system S is replaced by a model $\hat{S}$ in the exact linearization feedback loop. The input signal calculated in this way is then also applied to the real system S. The simulation system and feedback loop that leads to an overall linear behavior between r and y_s is denoted $S^\dagger$. The input to $S^\dagger$ is the reference r and the output is the control signal u.*

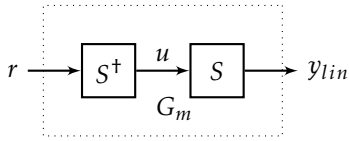


Figure 3.6: *The predistortion block $S^\dagger$ obtained using Hirschorn's method in series with the real system leads to an overall linear behavior between r and y_{lin}.*

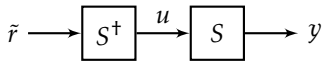


Figure 3.7: *The (possibly fictitious) reference signal $\tilde{r}$ can be seen as input to the block $S^\dagger$, creating the input u to the nonlinear system S.*

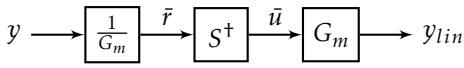


Figure 3.8: *Hirschorn's method applied as postdistortion, when the output can be assumed to be created according to Figure 3.7. The block $S^\dagger$ cannot simply be applied at the output y, but has to be manipulated to obtain a linear behavior between u and y_{lin}.*

G_m chosen by the user. This is based on the assumption that the model is accurate enough, of course. This is a setup where preinverse and postinverse are not interchangeable; Hirschorn's method tells us only how to determine the input to the nonlinear system such that the reference-to-output has the linear dynamics G_m, not how to manipulate the output to make it a linear response to the input. If it is this postinverse that is wanted, a different setup is needed.

It is known that $S^\dagger$ in cascade with S leads to a linear system G_m, so that

$$y = G_m r \tag{3.12}$$

with r the reference, cf. Figure 3.6. The goal is to obtain a linear response to u by using a postinverse on the output y. Assume that u was actually created by a prefilter, $S^\dagger$, with u as output and the fictitious signal $\tilde{r}$ as input, as in Figure 3.7. An estimate of this signal can then be obtained by

$$\bar{r} = \frac{1}{G_m} y, \tag{3.13}$$

where, if no transients or noise are present, $\bar{r} = \tilde{r}$. An estimate of the input u, called $\bar{u}$, can be obtained by filtering $\bar{r}$ by $S^\dagger$. Now, to obtain the desired dynamics, $\bar{u}$ must be filtered by the linear function G_m, see Figure 3.8. The cascade of these three blocks ($1/G_m$, $S^\dagger$ and G_m), thus make up a postdistorter that leads to a linear response between u (not available for manipulation) and y_{lin} in Figure 3.9. This method is illustrated in Example 3.4.

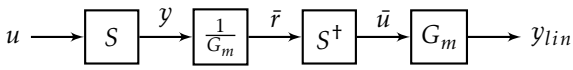


Figure 3.9: *Hirschorn's method used as postdistortion. The postinverse consists of the three blocks $1/G_m$, $S^\dagger$ and G_m. Used in this way, the overall behavior between u and y_{lin} will be linear.*

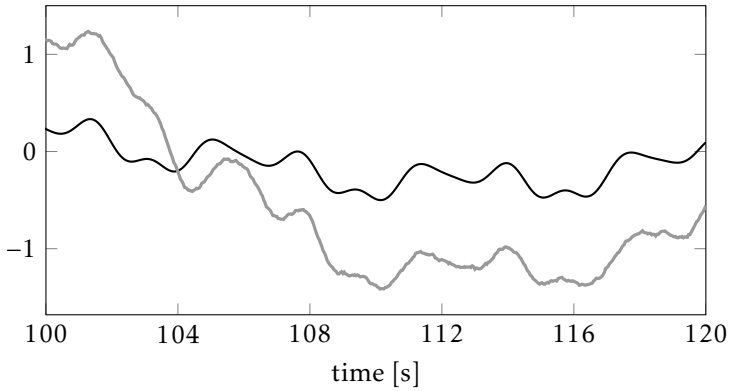


Figure 3.10: *The output from the nonlinear system* (3.14) *in gray and the desired dynamics from* G_m (3.15) *in black.*

Example 3.4: Hirschorn's Postinverse

Consider the nonlinear system

$$\begin{aligned} \dot{x}_1 &= -x_1^3 + x_2 + w_1 \\ \dot{x}_2 &= -x_2 + u + w_2 \\ y &= x_1 \end{aligned} \tag{3.14}$$

with process noise $w_i \in N(0, 0.05)$ and a multisine input. The nonlinear feedback

$$u = -3x_1^5 + 3x_1^2 x_2 + x_2 + \tilde{u}$$

leads to a linear system $\ddot{y} = \tilde{u}$. Now, linear theory can be applied and pole placement has been used to get an overall system response from reference r to output y corresponding to the one from

$$G_m(s) = \frac{1}{s^2 + 5s + 6}. \tag{3.15}$$

The output from the nonlinear system (3.14) is plotted in Figure 3.10 together with the output from the desired dynamics G_m.

A preinverse $S^\dagger$ has been constructed as in Figure 3.5. $S^\dagger$ has been used as a preinverse, as well as a postinverse for evaluation purposes. The results are shown in Figure 3.11. Here, it is clear that the desired preinverse and postinverse are not the same, and that $S^\dagger$ cannot straight away be used as a postinverse. If instead, the output y is filtered by the cascaded systems $1/G_m$, $S^\dagger$ and G_m, as in Figure 3.8, the result improves considerably, as shown in Figure 3.11. The remaining errors are primarily caused by the noise. For noise-free data, the preinverse performs perfectly whereas the postinverse has some minor errors.

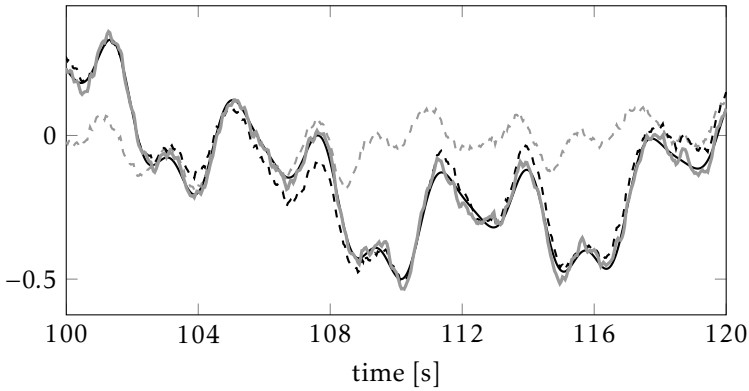


Figure 3.11: *A Hirschorn postinverse applied to the system* (3.14)*. The output from* G_m (3.15) *is plotted in solid black and the output when* $S^\dagger$ *was used as a preinverse in dashed black. The output from the system in series with the inverse* $S^\dagger$ *is plotted in dashed gray when* $S^\dagger$ *is used as a postinverse with no extra filtering. When the postdistortion is constructed according to Figure 3.8, the result improves considerably. Here, the postinverse consists of three blocks,* $1/G_m$*,* $S^\dagger$ *and* G_m *and the postdistorted output is plotted in solid gray. Note the scale difference from Figure 3.10.*

4

Estimation of Inverse Models

An inverse model is here estimated with the purpose of using it in cascade with the system itself, as an inverter, and a good inverse model in this setting would be one that, when used in series with the original system, reconstructs the original input, see Figure 4.1.

In estimation, one should usually estimate the system in the setting it should be used, concerning for example the choice of input and the experimental conditions [Ljung, 1999, Gevers and Ljung, 1986, Pintelon and Schoukens, 2012]. It is important to choose the input signal to capture the significant characteristics of the system. Usually, it ought to resemble the conditions in which it is intended to be used, but what does that correspond to in this case? Will it be the input spectrum that decides the weighting, or is it rather the output spectrum that should be weighted to reflect the relative importance of the model fit, since the output is in this case seen as the input to the inverse system to be estimated? Another important topic in system identification is the choice of loss function, V in (2.5b).

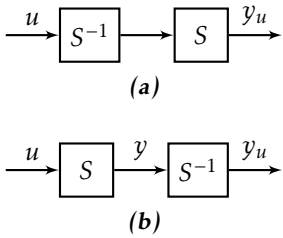


Figure 4.1: *The intended use of the estimated inverses. Figure (a) shows predistortion, where the inverse S^{-1} is applied before the system S, and (b) shows postdistortion, where the order is reversed.*

Table 4.1: *Inputs and outputs to the identification procedure, using the different methods.*

	Input	Output	Requires	Model
METHOD A	u	y		forward
METHOD B	u	u	$\hat{S}$	inverse
METHOD C	y	u		inverse

It should reflect the goal of the identification, and, depending on how it is chosen, different properties of the estimated model will be emphasized. In this setup, the goal is to make use of these degrees of freedom and the flexibility of the model to obtain an accurate input estimate. In this chapter, some aspects of system inverse estimation are discussed. The contents are also presented in Jung and Enqvist [2013].

4.1 System Inverse Estimation

In system identification, the goal is to achieve as good a model as possible to explain the behavior of y by a prediction or simulation $\hat{y}(t|\theta)$, which depends on the estimated model parameters θ and the input u. This is done using measured data, usually input data $u(t)$ and output data $y(t)$, (2.3), see Chapter 2. Here, a model describing the system itself will be referred to as a *forward* model and a model describing the inverse will be called an *inverse* model.

The inverse model is estimated with the purpose of using it in series with the system itself, as an inverter, see Figure 4.1. In this setup, the goal is to minimize the difference between the input u and the output from the cascaded systems, y_u. A good model in this setting would be one that, when used in series with the original system, regains the original input, so that $y_u = u$.

There are three main approaches to the estimation of an inverse of a system S, described in more detail below.

METHOD A In a first step, the forward model $\hat{S}$ is estimated in the standard way, with input data u and output data y. Step two is to invert the resulting model to obtain an approximate inverse $\hat{S}^{-1}$.

METHOD B In a pre-step, the forward model $\hat{S}$ is estimated in the standard way, with input data u and output data y. This model is used in series with an inverse model, $\hat{S}^{-1}$, and the inverse model parameters are estimated in this setting, by trying to minimize the difference between the input u and the simulated, distorted output y_u.

METHOD C The identification is done in one step, by identifying the inverse directly, using input data y and output data u.

The inputs and outputs to the different approaches are summarized in Table 4.1.

The identification in the first approach, METHOD A, is the standard one, as described in, for example, Ljung [1999] and Pintelon and Schoukens [2012], and

the inversion is discussed in Åström and Hägglund [2005] in the feedforward control application. The use of feedforward control based on an inverse model of the system in the presence of plant uncertainty is discussed in Devasia [2002]. A good thing with METHOD A is that the identification uses standard methods, but on the other hand, an inversion is required, and the weighting of the model fit is not necessarily optimal for the use intended here.

The second approach, METHOD B, is often used in power amplifier predistortion [Fritzin et al., 2011a, Abd-Elrady et al., 2008, Paaso and Mämmelä, 2008]. In this application, it is also called *direct learning architecture* (DLA). The quality of the inverse and the forward models are closely coupled, and two choices are available. Since it is often preferable to obtain a rather simple inverse model (for example in the predistorter case), this restriction can also be applied to the forward model, so that the same model structure is used for the forward and the inverse models. Another approach is to use a more complex forward model, making sure that as much as possible of the system behavior is captured, and then let the inverse model be less complex. The choice in the end comes down to the implementation – if the forward model has to be implemented, also this model needs to have a limited complexity. A good thing with this approach is that the estimation of the inverse is done with no noise present, but it also requires two, possibly nonconvex, minimizations with the risk of obtaining local minima. The quality of the inverse also clearly depends on the quality of the forward model.

The third approach, METHOD C, is also called *indirect learning architecture* (ILA) in power amplifier predistortion applications. It has been evaluated in PA predistortion applications in Abd-Elrady et al. [2008] and Paaso and Mämmelä [2008]. For this approach to be applicable in predistortion, it is assumed that the predistorter and the postdistorter are interchangeable (commutativity), see also Section 3.2.2. An advantage with this method is that the inverse is estimated in the setting in which it is going to be used, and that the weighting is possibly better than for METHOD A. A drawback is that the measured output is used as input, which risks causing a biased estimate [Amin et al., 2012]. It can be an easier approach, since the estimation is done in one step. Furthermore, there is no need to construct a model for the forward system that will later be discarded.

In power amplification predistortion applications, METHOD C (ILA) is more commonly used than METHOD B (DLA), as investigated in Paaso and Mämmelä [2008]. In Paaso and Mämmelä [2008], comparisons performed indicate that the DLA performs better in the simulation setup used, whereas in Abd-Elrady et al. [2008] the ILA seems to perform slightly better.

4.2 Inverse Identification of LTI Systems

To simplify the discussion, we will start by looking at LTI dynamical systems. The model estimation is done in open loop and assuming the output was created according to

$$y(t) = G_0(q)u(t) + H_0(q)e_0(t) \tag{4.1}$$

where G_0 is the true system, H_0 is the true noise dynamics and e_0 is a white noise sequence.

In system identification, the goal is often to find the minimizing argument of a function of the prediction error $\varepsilon(t, \theta)$

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t, \theta)^2 = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \left[y(t) - \hat{y}(t|\theta) \right]^2, \tag{4.2}$$

where $y(t)$ is the measured output and $\hat{y}(t|\theta)$ is the predicted output given the model parameters θ. Here, we use a fixed noise model $H_* \equiv 1$ such that the prediction is described by $\hat{y}(t|\theta) = G(q, \theta)u(t)$. Looking at the identification from a frequency domain point of view, the minimization criterion in (4.2) can asymptotically be written as [Ljung, 1999, (8.71) p. 266]

$$\hat{\theta} = \arg\min_{\theta} \int_{-\pi}^{\pi} \left| G_0(e^{i\omega}) - G(e^{i\omega}, \theta) \right|^2 \Phi_u(\omega) \mathrm{d}\omega \tag{4.3}$$

where $G(e^{i\omega}, \theta)$ is the model and $\Phi_u(\omega)$ is the spectrum of the input signal. The estimation will thus be done in a way to emphasize the model fit in frequency bands where the transfer function and the input spectrum are large enough to have a significant impact on the total criterion. The minimization is done with respect to the product of model fit ($|G_0 - G|^2$) and input spectrum. If the input is white noise (flat spectrum), it is thus more important to obtain a good model fit at frequencies with a large transfer function magnitude.

If instead the goal is to estimate the inverse model to be used as described in Section 4.1, the minimization criterion in the time domain can be written

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \left[u(t) - \frac{1}{G(q, \theta)} y(t) \right]^2 \tag{4.4}$$

and the frequency domain equivalent to (4.4), when y is noise-free, is

$$\hat{\theta} = \arg\min_{\theta} V_{\mathrm{inv}}(\theta). \tag{4.5}$$

The loss function is

$$
\begin{aligned}
V_{\text{inv}}(\theta) &= \int_{-\pi}^{\pi} \left| \frac{1}{G_0(e^{i\omega})} - \frac{1}{G(e^{i\omega},\theta)} \right|^2 \Phi_y(\omega)\mathrm{d}\omega \\
&= \int_{-\pi}^{\pi} \left| \frac{1}{G_0(e^{i\omega})} - \frac{1}{G(e^{i\omega},\theta)} \right|^2 |G_0(e^{i\omega})|^2 \Phi_u(\omega)\mathrm{d}\omega \\
&= \int_{-\pi}^{\pi} \left| 1 - \frac{G_0(e^{i\omega})}{G(e^{i\omega},\theta)} \right|^2 \Phi_u(\omega)\mathrm{d}\omega \qquad (4.6) \\
&= \int_{-\pi}^{\pi} \left| G(e^{i\omega},\theta) - G_0(e^{i\omega}) \right|^2 \frac{\Phi_u(\omega)}{|G(e^{i\omega},\theta)|^2}\mathrm{d}\omega \qquad (4.7)
\end{aligned}
$$

using $\Phi_y = |G_0(e^{i\omega})|^2\Phi_u$ if no noise is present. The loss function in (4.7) is similar to the weighting for the input error case where $H = G$ so that $y(t) = Gu + Ge = G(u+e)$, that is, the error enters the system at the same place as the input [Åström and Eykhoff, 1971].

Comparing the minimization criterion for the forward estimation in (4.3) to the one for the inverse estimation in (4.6), the weighting is clearly different. In the forward case, a relative model error at a frequency where the system amplification is small, will affect the criterion much less than a model error at a frequency where the system amplification is large. In the inverse estimation case, a relative model error will have the same effect on the criterion for two frequencies with the same input spectral density, and does not depend on the system amplification at that frequency. The weighting, and thus the model fit, between the different frequencies will be shifted to better reflect the importance of a good fit also at frequencies with a small transfer function magnification.

The time domain criterion (4.4) thus leads to the frequency domain description (4.6), and the weighting is automatically done to match the use of the inverse model estimate. Here, only the case when the system and its inverse are both stable and causal will be investigated. See Section 3.2.1 for a brief discussion on the problems involved in system inversion.

4.3 An Illustrative Linear Dynamic Example

Let us look at a small example. The goal is to obtain a system inverse to be used in series with the original system in order to retrieve the input, see Figure 4.1a. The input u and the noise-free output y are measured. The system has two resonance frequencies, at $\omega = 1$ rad/s and $\omega = 10$ rad/s. The magnitudes of the two resonance peaks are very different, with the first one a hundred times larger than the second one. The true system, G_0 is described by

$$
G_0(s) = \frac{10}{s^4 + 1.1s^3 + 101.1s^2 + 11s + 100} \qquad (4.8)
$$

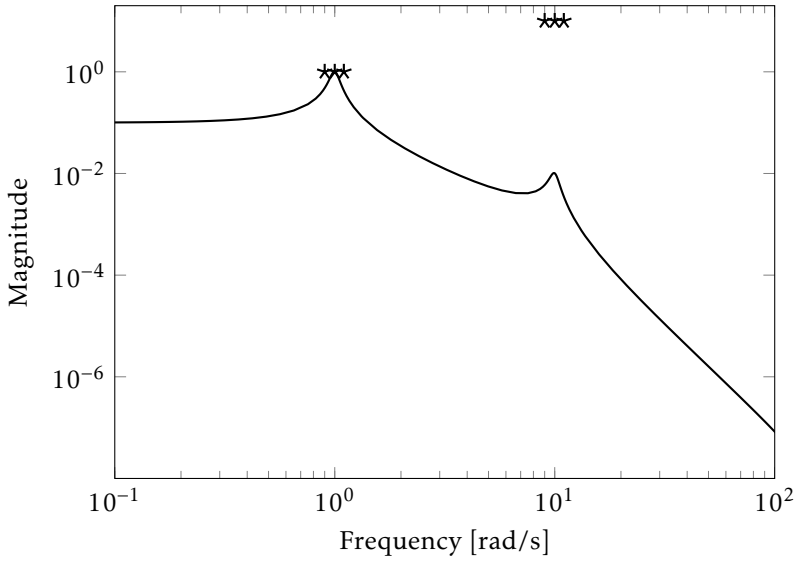


Figure 4.2: *The Bode magnitude plot of G_0 in (4.8) in the solid line. The stars mark the amplitude of the multisine input (u in (4.9)) components at each frequency.*

and the Bode magnitude diagram is shown in Figure 4.2. The input consists of three sinusoids around each of the two resonance peaks such that the input power is concentrated in two bands, centered around the resonance frequencies, i.e.,

$$u = \sum_{k=1}^{6} a_k \sin(\omega_k t + \phi_k) \tag{4.9}$$

with $a_k = 1$ for $k = 1, 2, 3$, $a_k = 10$ for $k = 4, 5, 6$, $\omega_k = \left\{0.9,\ 1,\ 1.1,\ 9,\ 10,\ 11\right\}$ and $\phi_k \sim U[-\pi\ \ \pi]$. The input amplitude and the frequency points are illustrated by the stars in Figure 4.2. The sampling time is $T_s = 0.02\ s$ and $N = 10\,000$ simulated measurements have been collected.

With the goal of using an FIR model as a prefilter to recover the input u, two models have been estimated, using METHOD A and METHOD C in Section 4.1. An FIR model depends only on previous input signals, as described on page 13. As the system is linear, the ordering of the two systems does not matter, and the preinverse and postinverse are interchangeable.

First, a forward model has been estimated as an output error (OE) model using System Identification Toolbox in MATLAB [Ljung, 2003], with [nb nf nk] = [1 3 0]. This model has then been inverted resulting in an FIR model with 4 terms, according to METHOD A. The approximative inverse using METHOD C is an FIR model with 4 terms, i.e. [nb nf nk] = [4 0 0], and will have a very different weighting. Hence, the two inverses will catch different behaviors of the system. The system G_0, (4.8), is a fourth order system whereas the model is third order. Thus,

the model cannot perfectly model the system but should be able to capture one resonance peak and the overall behavior of the system.

As can be seen in the Bode magnitude plot in Figure 4.3, the METHOD A model has a much better fit around $\omega = 1$ rad/s and almost perfectly models the resonance peak, but completely misses the second resonance peak at $\omega = 10$ rad/s. The inverse estimate, the METHOD C model, on the other hand, does not manage to catch either of the resonance peaks in a satisfactory way but catches both of the resonance frequencies. That is, the amplification at $\omega = 1$ and 10 rad/s is well captured, but not the resonance peaks around. Estimating the forward model in the standard way will clearly focus on the frequencies where the product of model fit ($|G_0 - G|^2$, connected to the transfer function amplification) and input spectrum is large. When this system approximation is then inverted, according to METHOD A, the errors around $\omega = 10$ rad/s will become prominent.

The results in the time and frequency domains are presented in Figures 4.4 and 4.5. In the time domain plot in Figure 4.4, it is clear that the METHOD C model better reconstructs the input than the METHOD A model. In Figure 4.5, the periodograms of the reconstructed inputs are shown, zoomed in around the input frequencies. At the lower frequency around $\omega = 1$ rad/s, the METHOD A model captures the input almost perfectly, but around $\omega = 10$ rad/s, the reverse is true and the METHOD C model performs better.

As shown in this small example, there are clearly occasions when it is advantageous to estimate an approximate inverse directly as opposed to estimating the forward model and then inverting it.

4.4 Inverse Identification of Nonlinear Systems

It is hard to say anything about the estimation of inverse systems for a general nonlinear system. Depending on the type of nonlinearity and how it enters, the effects will be different. So the choice of how to estimate the inverse is closely connected to the system itself.

For a linear system, a binary input signal is enough to extract all information in an identification experiment. One example of where this can be used is in identification of Hammerstein systems, which are block-oriented systems where a static nonlinearity is followed by a linear dynamical system. The linear part of the Hammerstein system can be modeled perfectly by using a binary input signal. In a second experiment, where the input is no longer binary, the model of the linear system can be used to simplify the estimation of the nonlinearity. The estimates of the nonlinearity and the linear dynamics can be inverted and a Wiener system is obtained, according to METHOD A, assuming a stable, minimum-phase dynamical system and an invertible nonlinearity. This is similar to Example 3.3 on page 23 where an exact inverse could be found (but there the system was assumed known).

One way of finding an inverse to a more general nonlinear system is by using Hirschorn's method, described in Section 3.3.2. A question is how to estimate this system inverse. In the case where the structure of the nonlinear system is

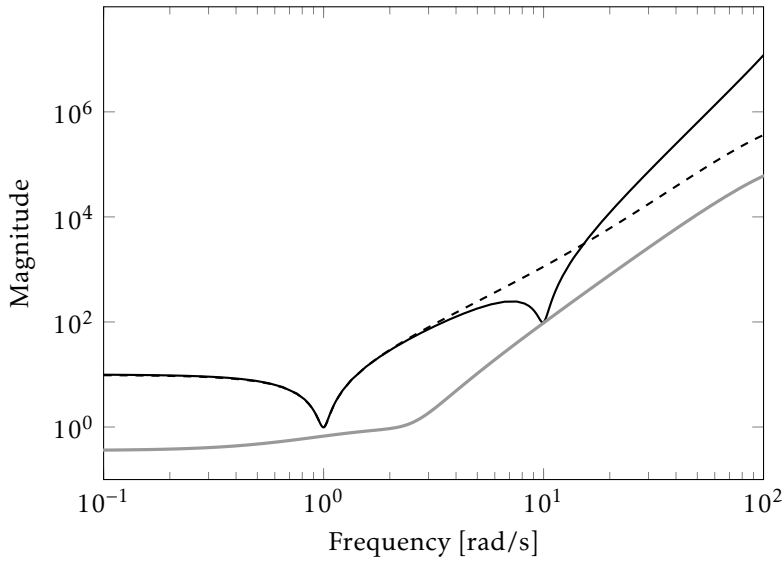


Figure 4.3: *The Bode magnitude response of G_0^{-1} (black solid line), the inverted forward model, METHOD A, (black dashed line) and the inverse model estimate using METHOD C (gray solid line). The inverted forward model perfectly catches the resonance peak at $\omega = 1$ rad/s, whereas the direct estimation of the inverse does not model either of the resonance peaks in a satisfactory way. The METHOD C model instead has an accurate modeling of both peak frequencies, that is, it manages to accurately model the amplification at $\omega = 1$ and 10 rad/s, but not the resonance peaks.*

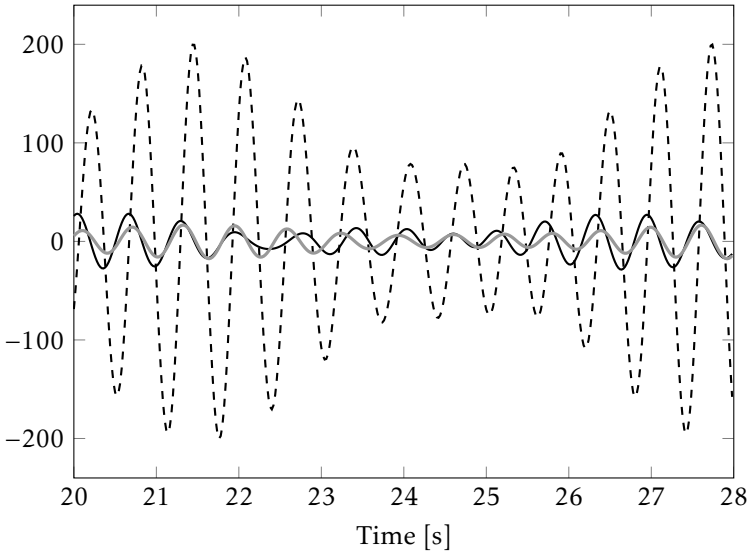


Figure 4.4: *The input u (black solid line), and the reconstructed input y_u using the inverted forward model (black dashed line) and the inverse model estimate using* METHOD C *(gray solid line). The estimation of the inverse cannot perfectly reconstruct the input, but is clearly better than the inverted forward model.*

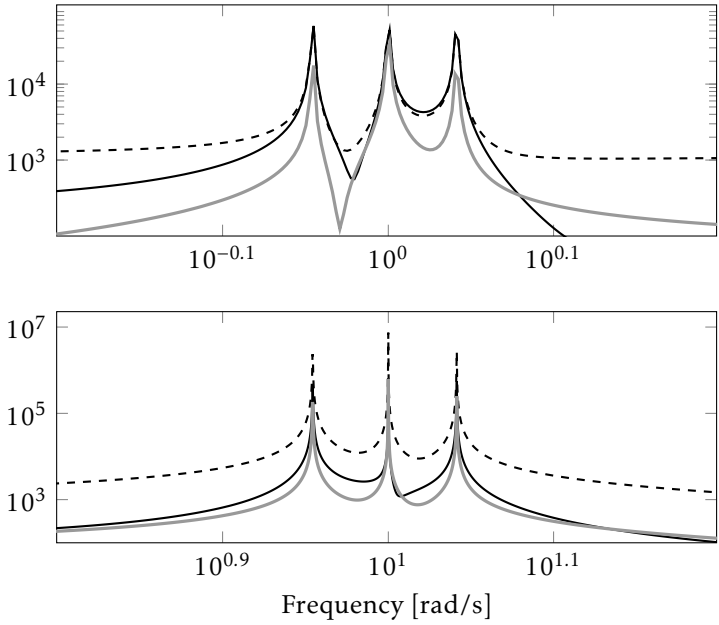


Figure 4.5: *Periodogram of the input u (black solid line), and the reconstructed input y_u using the inverted forward model (black dashed line) and the inverse model estimate using* METHOD C *(gray solid line) around $\omega = 1$ rad/s (top) and $\omega = 10$ rad/s (bottom). It is clear also in the frequency domain that the forward model better captures the behavior around $\omega = 1$ rad/s than the inverse estimation, but the reverse is true around $\omega = 10$ rad/s.*

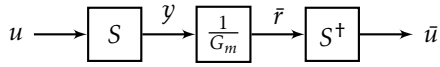


Figure 4.6: *Estimation of Hirschorn inverse model. By filtering the output y through the inverse dynamics of the desired dynamics G_m, an estimate of the reference signal $\bar{r}$ can be obtained.*

known, but where there are unknown parameters that need to be estimated, the identification can be done in several ways, just as described in Section 4.1.

METHOD A would correspond to measuring the input u and the output y, and identifying the unknown parameter values in the standard (forward) way. This estimated model could then be used to provide the inverse, since if a model of the forward system is available, a model of the inverse system is as well.

Since the exact linearization framework provides us with an inverse once the forward model is known, METHOD B does not really have an equivalence in this case – once the forward model is known, the exact inverse to match it is also known. In the general case, this forward model could be used to estimate an approximate inverse.

METHOD C would correspond to estimating the inverse $S^\dagger$ directly. The order of the inverse and the output are reversed in METHOD C, so that y is used as input and u as output. In Hirschorn's method, the inverse takes the reference r as input and the output is the control signal u. So, in order to find the inverse of $S^\dagger$, we would need u as output and the reference r as input. But, as the data was collected in open loop with no pre- or postdistorter, the signal r is not available, only u and y. Now, as in Section 3.3.2, assume that the system was actually preceded by a system $S^\dagger$, fed by a fictitious reference signal $\bar{r}$, and that the overall behavior from $\bar{r}$ to y is in fact linear with dynamics described by G_m. If this is true, then the signal $\bar{r}$ would be obtained by filtering y with $1/G_m$, and the system $S^\dagger$ can be identified using $\bar{r}$ as input and u as output, see Figure 4.6. So, this equals finding the inverse by using (a filtered version of) the output y as the input and u as output as in METHOD C. A benefit with Hirschorn's method is that it provides a parameterized inverse, so that the structure of this inverse system is already known.

Part II

Power Amplifier Predistortion

5

Power Amplifiers

An electronic amplifier, or *power amplifier* (PA) is used to increase the power of a signal, so that the output is a magnified replica of the input. There are many different constructions of amplifiers, and they can be characterized by different measures such as gain, efficiency and linearity. Amplifiers are commonly used in many applications, such as audio applications and telecommunications, both in base stations and hand-held devices.

This chapter provides a basis to understand the amplifier related problems described in later chapters. It is by no means a complete description of PAs, but should be enough to understand this thesis. It also introduces the concepts of predistortion and linearization as well as the outphasing PA.

5.1 Power Amplifier Fundamentals

Today, wireless communication is used everywhere to transfer information. An important part of the technology is the possibility to transmit and receive the information, and the devices used are called *transmitter* (TX) and *receiver* (RX). The transmitter converts the information to an electrical signal suitable for the transmission in the given medium (in this case air, but in standard communication this can be a wire, fiber-optics, etc.). At the other end of the transmitting medium, a device is needed to receive the message and convert it into the original form – the receiver. This process of sending and propagating an information signal over a medium is called a *transmission*.

It is often desired that the equipment should be able to both send and receive information (a phone for example, where one can speak and listen), that is, a device that contains both a transmitter and a receiver. Such a circuit is called a *transceiver*. The physical circuit is connected to a *chip*.

By combining the receiver and transmitter into a transceiver, the circuits can

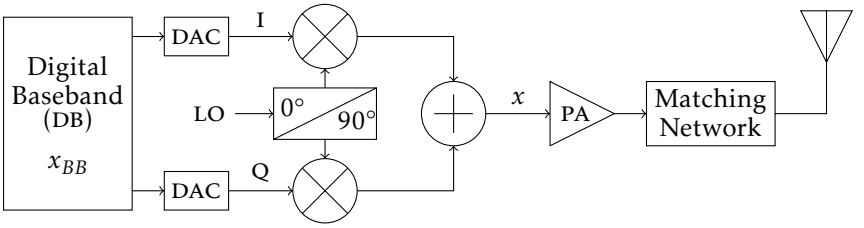


Figure 5.1: *Block diagram of a direct-conversion transmitter. The baseband signal (x_{BB}) is upconverted to radio frequencies by the modulator and passes through a* PA *before being sent to the antenna.*

be used for multiple purposes, reducing the number of components (and thus the cost) as well as the size of the chip, leading to more functionality per area. Such shareable components are antennas, oscillators, amplifiers, tuned networks and filters, frequency synthesizers and power supplies [Frenzel, 2003].

5.1.1 Basic Transmitter Functionality

A standard transmitter includes a *digital baseband* (DB), *digital-to-analog converters* (DACs), mixers (X) (further explained in Example 5.1), two *local oscillators* (LOs) that are 90° out of phase, a combiner, a power amplifier and a matching network before the antenna. The signal of interest, x_{BB}, is split into an in-phase channel, I, and a quadrature channel, Q,

$$x_{BB}(t) = I(t) + jQ(t) \tag{5.1}$$

by the DB, corresponding to the real (I) and imaginary (Q) parts of the signal, to generate two independent signals. Complex signals are commonly used in different modulation techniques in communications applications, see for example Frenzel [2003]. The I and Q signals are upconverted to the *radio frequency* (RF, ranging between 3 kHz and 300 GHz) carrier frequency, ω_c, and recombined, see Figure 5.1. The upconversion is done by a quadrature modulator, usually implemented by two mixers and two LO signals with a phase difference of 90°. The power of the recombined output signal,

$$x(t) = r(t)\cos(\omega_c t + \alpha(t)) \tag{5.2}$$

where

$$r(t) = \sqrt{I^2(t) + Q^2(t)} \tag{5.3}$$

and

$$\alpha(t) = \arctan(Q(t)/I(t)) \tag{5.4}$$

is often too low for transmission, and it has to pass through a power amplifier before being sent to the antenna.

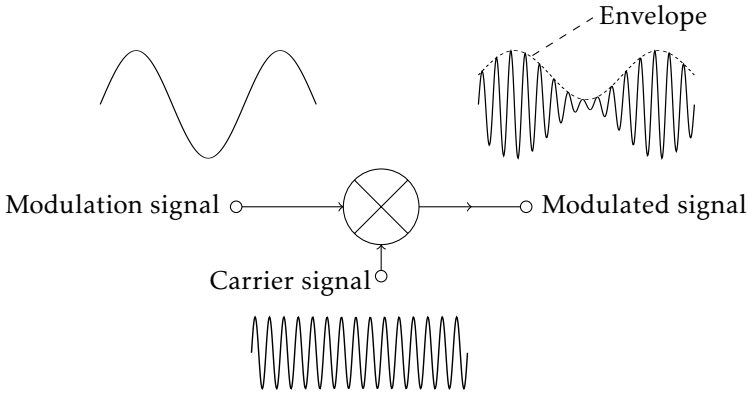


Figure 5.2: *Amplitude modulation. The information in the modulation signal is upconverted in the mixer to the carrier frequency (frequency of the carrier signal) and the shape (envelope) of the modulated signal contains the original information in the modulation signal.*

Example 5.1: Amplitude modulation

Modulation is the process of varying the properties of a high-frequency signal, the *carrier signal* (usually a sine wave) with a *modulation signal* that contains the information to be transmitted. The modulation can be performed using a mixer, a component that multiplies the two (possibly shifted) inputs. When *amplitude modulation* (AM) is used, the information can be found in the amplitude of the modulation signal. The imaginary line that connects the peaks of the modulated signal is the information signal, and is called the *envelope*. Other common analog modulation techniques include *phase modulation* (PM) and *frequency modulation* (FM). Here, the envelope of the signal is kept constant but the phase shift or the frequency, respectively, of the carrier frequency is varied. These modulation techniques can also be combined into more complex modulation techniques.

For the example in Figure 5.2, the modulation (information) signal is a sine wave. The carrier is a sine wave of much higher frequency, and the modulated output is a high frequency signal where the shape of the envelope contains the information in the modulation signal.

The amplitude modulation in Example 5.1 is an analog modulation scheme that can be used for continuous signals. If the baseband signal is digital, a digital modulation is needed, which will be introduced in Example 5.2.

Example 5.2: Digital modulation

One digital modulation scheme is *phase-shift keying* (PSK) that changes, modulates, the phase of the carrier signal. A digital modulation uses a finite number of distinct signals to represent digital data. In PSK, the phase is unique for each signal section, or *symbol*, that is transmitted. The demodulator, at the receiver end, should interpret the signal and map it back to the original symbol. This

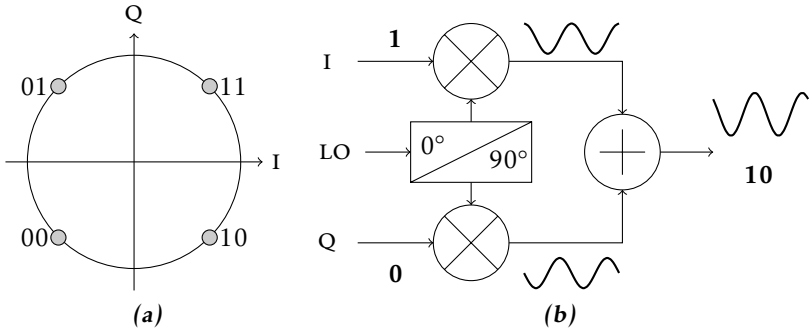


Figure 5.3: *(a) Constellation diagram for quadrature phase-shift keying, a digital modulation scheme. The four symbols represent the bits 00, 01, 11 and 10. (b) shows an example where the symbol 10 is to be transmitted. The* I *part is 1 and the* Q *part is 0. The bits are modulated by a carrier signal, a sinusoidal with a 90° phase shift between the* I *and* Q *parts, and the signals are added. Typically, the zero is coded as* -1*. The phase of the output is unique and can be mapped back to the* I *and* Q *parts, as seen in Figure 5.4.*

requires the receiver to be able to compare the phase of the received signal to a reference signal. Such a system is termed coherent.

One type of digital PSK modulation is *quadrature phase-shift keying* (QPSK) which uses four phases, and can encode two data bits per symbol. In a constellation diagram, the QPSK scheme has four points spread out around a circle, as seen in Figure 5.3a.

We will here look at an example where the symbol to be transmitted is 10. The IQ decomposition is done such that the odd-numbered bit (1) is the I component and the even-numbered bit (0) is the Q component, as seen in Figure 5.3b. The bits are modulated by the carrier signal, a sinusoidal with a 90° phase shift between the I and Q branches, and the signals are added. The resulting signal is unique, as seen in the bottom row of Figure 5.4, and can be mapped back to the I and Q components.

5.2 Power Amplifier Characterization

The choice of PA is a trade-off between different properties such as output power, efficiency and linearity, and will depend on the application. If power efficiency is an important property, such as in handheld devices where it will reflect directly on the battery time, a lower linearity might be accepted, whereas an audio amplifier, always connected to the power net, might focus more on the linearity and gain than on the efficiency. Any number of PAs can be cascaded in order to combine the benefits of each step.

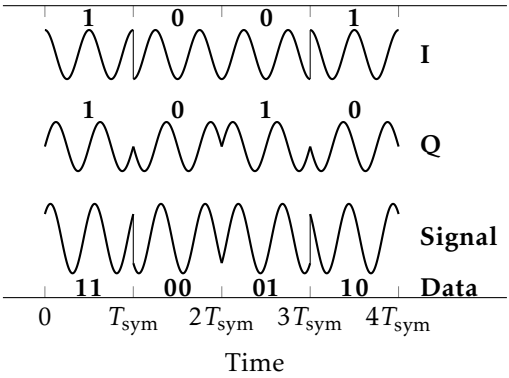


Figure 5.4: *The modulated signals in the IQ modulation, where the two carrier waves are sinusoidal with a 90° phase shift. The odd-numbered bits encode the in-phase (I) component and the even-numbered bits encode the quadrature (Q) component. The total signal is shown at the bottom, together with the mapping. The digital data transmitted by this signal is 1 1 0 0 0 1 1 0.* T_{sym} *is the symbol duration.*

5.2.1 Gain

An amplifier is of course supposed to amplify the input signal, and this property is described by the gain. The gain of an amplifier expresses the relationship between the input and the output [Frenzel, 2003], and is usually described by the voltage gain, A_V,

$$A_V = \frac{V_{out}}{V_{in}}, \tag{5.5}$$

where V_{in} and V_{out} are the input and output voltages, respectively. It can also be expressed by the power gain, A_P,

$$A_P = \frac{P_{out}}{P_{in}},$$

where P_{in} and P_{out} are the input and output powers, respectively, see Figure 5.5. The gain is usually expressed in decibels (dB), so that the power gain is

$$A_P = 10 \log_{10} \left(\frac{P_{out}}{P_{in}} \right). \tag{5.6}$$

5.2.2 Efficiency

Another important property of a PA is the efficiency, which describes the amount of power needed to perform the amplification. A part of the input power will be dissipated in the circuit and can be counted as losses. The efficiency of a PA will

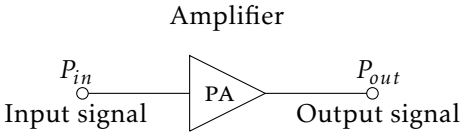


Figure 5.5: *Amplifier with input and output. The power gain is* $A_P = \frac{P_{out}}{P_{in}}$.

directly affect the battery time for a cell phone for example, and a high efficiency is desired.

The output efficiency, η, of a PA is defined as the ratio between the output power at the fundamental frequency, P_{out}, and the DC supply power of the last amplifier stage, P_{DC}, [Cripps, 2006]

$$\eta = \frac{P_{out}}{P_{DC}}, \tag{5.7}$$

and is often denoted *drain efficiency* (DE). Another efficiency measure is the *power added efficiency* (PAE),

$$\text{PAE} = \frac{P_{out} - P_{in}}{P_{DC}}, \tag{5.8}$$

where P_{DC} now represents the total power consumption of all amplifier stages constituting the whole PA [Razavi, 1998].

5.2.3 Linearity

By assigning transmissions different frequency bands, many transmissions can be done at the same time. For this setup to work, each of these transmissions must send only in the allotted slot, or *channel*. A radio transmission is allocated a frequency band with a certain bandwidth, ω_b, around a center frequency, f_c, where power may be transmitted. Any power falling outside the boundaries will cause disturbances in the neighboring channels. Broadening of the spectrum can be caused by, for example, nonlinearities in the PA. So to be practically useful in radio communications, PAs need to be linear. This means that the signal should be amplified in such a way that the output is an exact replica of the input but with a larger amplitude, and not be transferred to other frequencies. This is not possible in practice, and the level of linearity, or rather nonlinearity, is quantified by measures such as spectral mask, *adjacent channel power ratio* (ACPR) and *error vector magnitude* (EVM).

Spectral mask A spectral mask is a nonlinearity measure describing the amount of power that is allowed to be spread to adjacent frequencies. It is usually specified in decibel to carrier (dBc, the power ratio of a signal to a carrier signal, expressed in decibels) or in power levels given in dBm (power expressed in dB with one milliwatt as reference) in a specified bandwidth at defined frequency

Table 5.1: *Spectral mask limitations for an EDGE signal*

Offset [kHz]	100	200	250	400	600	1000
Limit [dBc]	0	-30	-33	-54	-60	-60

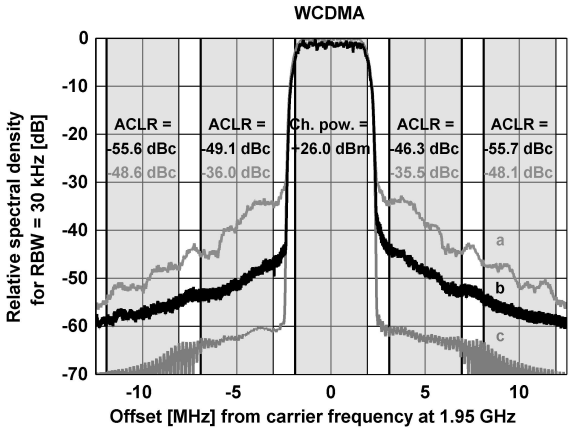


Figure 5.6: *Spectrum at 1.95 GHz for (a) measured output without DPD, (b) measured output with predistortion (linearization) and (c) the input signal for a WCDMA signal. The measured ACLR are printed in gray for the original output signal (without predistortion) and in black for the predistorted output. The gray shadows represent the passband in which the integration takes place.*

offsets [Fritzin, 2011]. See Table 5.1 for an example of the spectral mask limits for an EDGE signal.

Adjacent Channel Power Ratio The ACPR is a measure that, like the spectral mask, describes the amount of power spread to neighboring channels. It is defined as the power in a passband away from the main signal divided by the power in a passband within the main signal [Anritsu, 2013]. The power at frequencies that are not in the main signal is the power transmitted in neighboring channels, i.e., the distortion caused by nonlinearities. Another measure is the *alternate channel power ratio*, which is defined as the ratio between the power in a passband two channels away from the main signal, over the power within the main signal.

The bandwidths and limits are connected to the standard used (for example WCDMA and LTE). For a WCDMA signal, the ACPR can be calculated by integrating the spectrum over a bandwidth of ω_b = 3.84 MHz at ±5 MHz distance from the

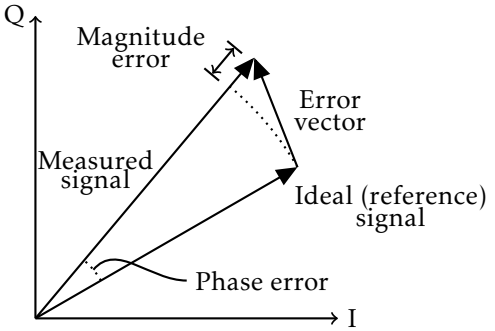


Figure 5.7: *Error vector magnitude (EVM) and related quantities.*

center frequency, as

$$\mathrm{ACPR} = \frac{\int_{f_c+l\cdot 5-1.92}^{f_c+l\cdot 5+1.92} \mathrm{WCDMA}_{\mathrm{spectrum}}\,\mathrm{d}f}{\int_{f_c-1.92}^{f_c+1.92} \mathrm{WCDMA}_{\mathrm{spectrum}}\,\mathrm{d}f}. \tag{5.9}$$

Here, f_c is the center frequency in the main signal and $l = \pm 1$ for the adjacent and $l = \pm 2$ for the alternate channel power ratio. ACPR is also named *adjacent power leakage ratio* (ACLR). An example of the ACLR can be seen in Figure 5.6.

Error Vector Magnitude The *error vector magnitude* (EVM) is a description of the quality of a signal with both magnitude and phase, such as the IQ signals as described in Section 5.1. The error vector is defined as the difference between the ideal signal and the measured signal [Agilent, 2013], see Figure 5.7.

Gain Compression, AM-AM and AM-PM At some point, a change in input amplitude does not result in a corresponding change in output amplitude, as illustrated in Figure 5.8. This phenomenon is called *gain compression*. This leads to nonlinearities in the output, since different amplitudes of the input will be amplified in different ways.

Other nonlinearity measures describing the amplitude and phase distortion are the *amplitude modulation to amplitude modulation* (AM-AM) and the *amplitude modulation to phase modulation* (AM-PM). The AM-AM maps the input amplitude to the output amplitude (similar to the gain compression graph in Figure 5.8) and deviations from the straight line will result in output distortion. The AM-PM maps the input amplitude to the output phase, where an increasing input amplitude results in an additional output phase shift [Cripps, 2006].

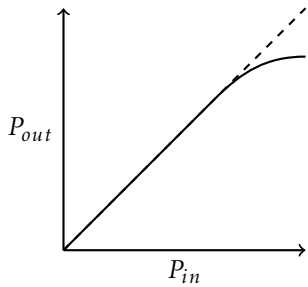


Figure 5.8: *Gain compression due to saturation in an amplifier transistor. The dashed line represents the ideal operation of the amplifier, while the solid line is the true output of the* PA *and a consequence of gain compression.*

5.3 Classification of Power Amplifiers

There are many different types of amplifiers, but they can be divided into two basic types; linear and switched amplifiers, see for example Frenzel [2003] and Jaeger and Blalock [2008] for a more thorough description of the different PA classes and the circuitry to implement them. Classical PAs usually assume both the input and the output to be sinusoidal, which limits the efficiency. If this assumption is disregarded, higher efficiency can be achieved [Razavi, 1998]. Here, the different classes are described.

5.3.1 Transistors

An important part of power amplifier implementation are the transistors, and we will start with a short overview of transistor functionality. A transistor is a device that uses a small signal to control a much larger signal. The two basic types of transistors are *bipolar junction transistors* (BJT s) and *field-effect transistors* (FETs). The structure of the commonly used FETs using semiconducting material has led to the name *metal-oxide-semiconductor field-effect transistor* (MOSFET). Depending on how the silicon is doped, the FETs can be either of p-type (PMOS) or n-type (NMOS), and thus have different conduction capabilities with respect to the applied voltages at the transistor terminals. Doping is the process of intentionally introducing impurities into an extremely pure semiconductor for the purpose of modulating its electrical properties. *Complementary metal-oxide-semiconductor* (CMOS) is a technology that typically uses complementary and symmetrical pairs of p-type and n-type MOSFETs for logic functions.

The FETs have three terminals, labeled gate (G), source (S) and drain (D), and a voltage at the gate controls the current between source and drain, see Figure 5.9. See for example Jaeger and Blalock [2008] for more insights into the workings and construction of transistors. For an NMOS transistor, a high voltage at the gate leads to a large current between source and drain, and for a small gate voltage, there is no current. For a PMOS transistor the relations are reversed, and a small gate voltage leads to a large current between source and drain, and a large gate

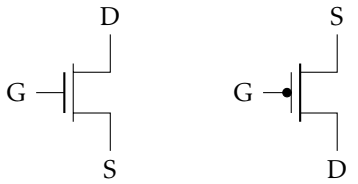


Figure 5.9: *The symbols of* NMOS *(left) and* PMOS *(right) and the associated ports. The ports are labeled gate (G), source (S) and drain (D).*

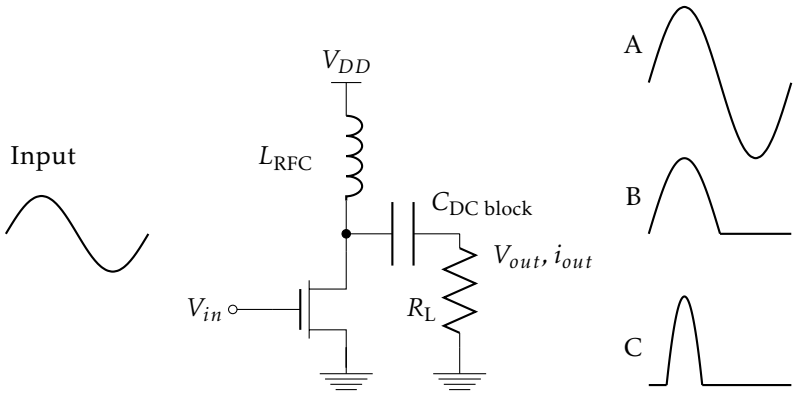


Figure 5.10: *Generic Class A/B/C power amplifier. The biasing of the transistor determines the conduction angle of the* PA*, as illustrated in the amplification of a sinewave input (left). The conduction angles are (from top to bottom)* 360° *for the Class A,* 180° *for the Class B and* 90° *for the Class C here.*

voltage opens the circuit and no current flows. Common uses for transistors are as amplifiers and switches, depending on the circuitry surrounding them.

5.3.2 Linear Amplifiers

Linear amplifiers provide an amplified replica of the input. The drawback is that linear amplifiers often require a high power level and provide a rather low efficiency, as they operate far from their maximum output power where the linearity is limited.

Class A Amplifiers

A Class A amplifier operates linearly over the whole input and output range. It is said to conduct for 360° of an input sine wave, that is, it will amplify for the whole of the input cycle, see Figure 5.10. Since the device is always conducting, a lot of power will be dissipated and the maximum achievable output efficiency is low, only 50%.

Class B Amplifiers

In a Class B amplifier, the device is biased so that it only conducts for half of the input cycle, i.e., it has a conduction angle of 180°, see Figure 5.10. In this region the amplifier is linear, and at the rest of the input it is turned off, and the efficiency reaches $\eta = \pi/4 \approx 78.5\%$, with η defined in (5.7).

Class B amplifiers are often connected in a push-pull circuit, so that two amplifiers are connected, each of them conducting for half of the cycle, and together they conduct for the whole 360°. The efficiency is still the same, and in theory this will be a completely linear amplifier. In practice, however, if the biasing of the two amplifiers is not perfect, this will cause cross-over distortion at the time of switching between the two amplifiers [Jaeger and Blalock, 2008].

Class AB Amplifiers

The Class AB amplifier uses the same idea as the Class B configuration with two amplifiers, but the amplifiers are slightly overlapping such that the cross-over distortion is minimized. Each amplifier thus has a larger conduction angle than the 180° of a Class B amplifier, but less than the full 360° of a Class A amplifier. This reduction of cross-over distortion is at the expense of efficiency.

Class C Amplifiers

Class C amplifiers have a conduction angle smaller than 180°, typically between 90° and 150°, see Figure 5.10. This causes a very distorted output consisting of short pulses, and the amplifier usually has some form of resonant circuit connected to recover the original sine wave.

5.3.3 Switched Amplifiers

The low efficiency of linear amplifiers is caused by the high power dissipation due to constant conduction. Switched amplifiers consist of transistors that are either *on* (conducting) or *off* (nonconducting). In the off state (cutoff state), no current flows so there is (almost) no dissipation. When the transistor is conducting, the resistance across it is very low, and so is the power dissipation.

The output of a switched amplifier is a square wave, which is passed through a filter to obtain a sinusoidal signal.

Class D Amplifiers

A Class D amplifier consists of two transistors that alternately are on and off. The output is a *pulse-width modulated* (PWM) signal, which can be filtered to obtain the fundamental sine wave, see Figure 5.11. With ideal switches and ideal series resonant network (C_1 and L_1) stopping all frequencies but the fundamental tone, the theoretical maximum efficiency is 100%.

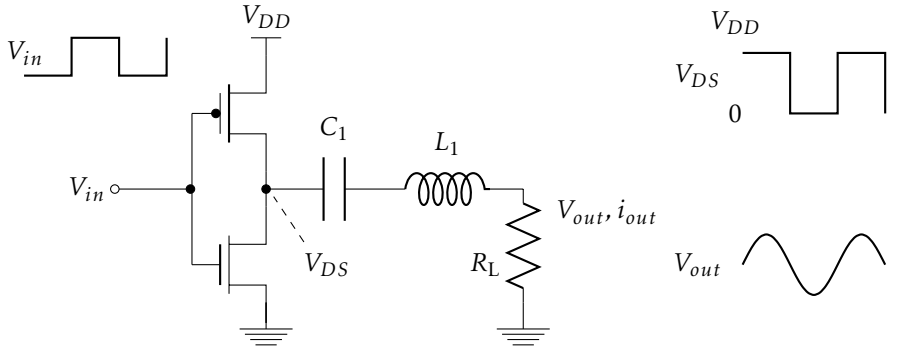


Figure 5.11: *Class D power amplifier.*

Class E Amplifiers

In a Class E amplifier, only one transistor is used (compared to the two for Class D). By choosing a suitable load matching network, the drain current and voltage can be shaped to not overlap each other, making the theoretical efficiency 100%.

5.3.4 Other Classes

There exist many other classes including Class F (a variation of the Class E amplifier) and Class S (a variation of switching amplifier using pulse-width modulation), see for example Frenzel [2003].

5.4 Outphasing Concept

An outphasing amplifier is based on the idea that a nonconstant envelope signal, with amplitude and phase information, can be decomposed into two constant envelope signals with phase information only. The two signals can then be amplified separately by two nonlinear and highly efficient amplifiers and recombined, as presented in Cox [1974] and Chireix [1935]. The output signal will be amplitude and phase modulated, just like the input signal. Another name for the outphasing concept is *linear amplification with nonlinear components* (LINC).

The outphasing concept is illustrated in Figure 5.12. Here, a nonconstant envelope-modulated signal

$$s(t) = r(t)e^{j\alpha(t)} = r_{\max}\cos(\varphi(t))e^{j\alpha(t)}, \quad 0 \le r(t) \le r_{\max} \tag{5.10}$$

where $r_{\max}$ is a real-valued constant, and α and φ are angles, is used to create two constant-envelope signals, $s_1(t)$ and $s_2(t)$. This is done in the *signal component*

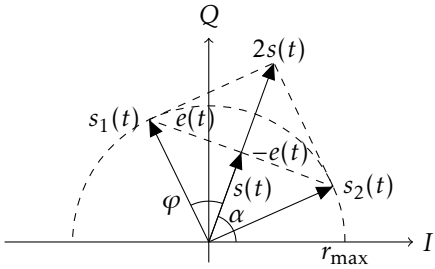


Figure 5.12: *Outphasing concept and signal decomposition.*

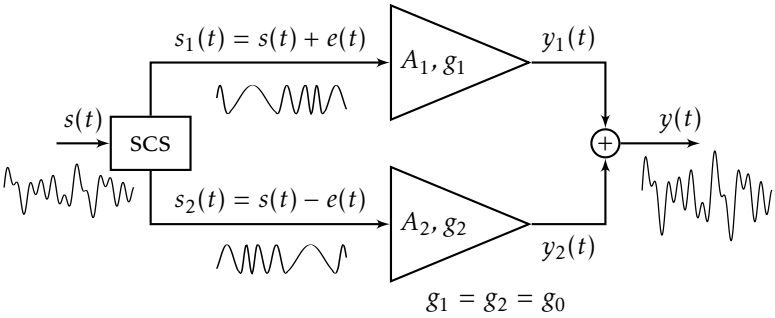


Figure 5.13: *Illustration of ideal power combining (the plus sign) of the two constant-envelope signals. The signals are amplified separately by two nonlinear amplifiers, A_1 and A_2, and recombined to an amplified replica of the input $s(t)$.*

separator (SCS) in Figure 5.13 as

$$\begin{aligned} s_1(t) &= s(t) + e(t) = r_{\max} e^{j\alpha(t)} e^{j\varphi(t)} \\ s_2(t) &= s(t) - e(t) = r_{\max} e^{j\alpha(t)} e^{-j\varphi(t)} \\ e(t) &= j s(t) \sqrt{\frac{r_{\max}^2}{r^2(t)} - 1}. \end{aligned} \tag{5.11}$$

The outphasing signals $s_1(t)$ and $s_2(t)$ contain the original signal, $s(t)$, and a quadrature signal, $e(t)$, and are suitable for amplification by switched amplifiers like Class D/E. By separately amplifying the two constant-envelope signals and combining the outputs of the two individual amplifiers as in Figure 5.13, the output signal is an amplified replica of the input signal.

In theory, the two quadrature signals will cancel each other perfectly in the combiner, but in practice, implementation imperfections and asymmetries will cause distortion. Letting g_1 and g_2 denote two positive real-valued gain factors, in each branch $s_1(t)$ and $s_2(t)$, and δ denote a phase mismatch in the path for $s_1(t)$,

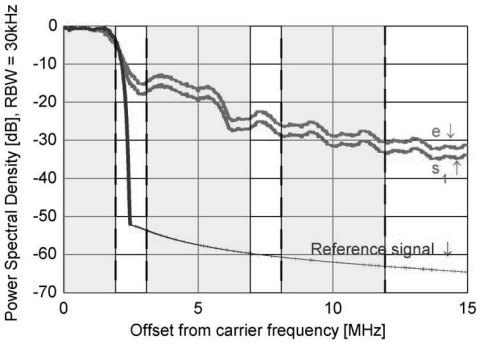


Figure 5.14: *The bandwidth of the quadrature signal $e(t)$, and thus the outphasing signals $s_1(t) = s(t) + e(t)$ and $s_2(t) = s(t) - e(t)$, is much larger than that of the original signal $s(t)$. Any remainders of the quadrature signal caused by* PA *imperfections will thus lead to degraded* ACLR *and reduced margins to the spectral mask. From Fritzin [2011].*

it is clear from

$$\begin{aligned} y(t) &= g_1 e^{j\delta} s_1(t) + g_2 s_2(t) \\ &= [g_1 e^{j\delta} + g_2] s(t) + [g_1 e^{j\delta} - g_2] e(t), \end{aligned} \tag{5.12}$$

that besides the amplified signal, a part of the quadrature signal remains. As the bandwidth of the quadrature signal, $e(t)$, is larger than the original signal, $s(t)$, see Figure 5.14, this would lead to a degraded ACLR and reduced margins to the spectral mask [Birafane and Kouki, 2005, Birafane et al., 2010, Romanò et al., 2006].

The phase and gain mismatches between $s_1(t)$ and $s_2(t)$ must be minimized in order not to allow a residual quadrature component to distort the spectrum or limit the *dynamic range* (DR),

$$c_{DR} = 20 \log_{10}\left(\frac{\max(|y(t)|)}{\min(|y(t)|)}\right) = 20 \log_{10}\left(\frac{|g_1 + g_2|}{|g_1 - g_2|}\right), \tag{5.13}$$

of the PA [Birafane and Kouki, 2005]. The DR defines the ratio of the maximum and minimum output amplitudes the PA can achieve. However, all phases and amplitudes within the DR can be reached by changing the phases of the outphasing signals $s_1(t)$ and $s_2(t)$.

Since an outphasing amplifier only uses two states (on or off), it will not experience problems like the conventional PAs such as gain compression (see Section 5.2.3), where the peak amplitudes are clipped. Instead, the smallest amplitudes will not be properly amplified in outphasing PAs, since any mismatch of the amplifier gains will make it impossible for $s_1(t)$ and $s_2(t)$ to cancel each other, compare Figures 5.12 and 5.15. Thus, the DR in an outphasing PA limits the spectral performance when amplifying modulated signals.

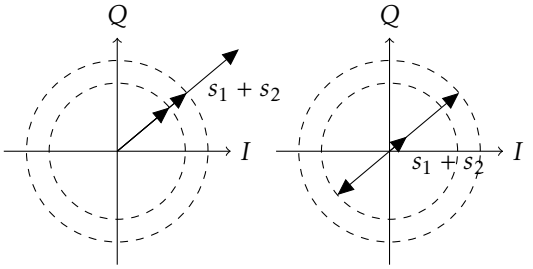


Figure 5.15: *The outphasing concept when the gain factors g_1 and g_2 are not identical. In the left figure, the outphasing signals are parallel and the resulting output is the maximal one. In the right figure, the nonidentical gain factors cannot cancel each other, and some remains are left. The dynamic range (the ratio between the maximal and minimal amplitudes, see* (5.13)*) of the power amplifier will determine the limit of small amplitude clipping.*

As the output of a Class D stage can be considered as an ideal voltage source whose output voltage is independent of the load [Yao and Long, 2006], i.e., the output is connected to either V_{DD} or *GND*, the constant gain approximations g_1 and g_2 are appropriate and make Class D amplifiers suitable for nonisolating combiners like transformers [Xu et al., 2010]. The implementation of the combiner (the plus sign in Figure 5.13) can be done in a multitude of ways, see for example Fritzin [2011] and the references therein.

5.5 Linearization of Power Amplifiers

The increased use of nonlinear amplifiers in an attempt to improve efficiency also requires new linearization methods. As described in Chapter 3, there are different approaches to do linearization. Since it is desirable to work with the original signal, and not with the amplified output of the PA, a prefilter is desired, also called a *predistorter* [Kenington, 2000]. Originally, these predistorters consisted of small analog circuits, but now they are often implemented in a *look-up table* (LUT) or a *digital signal processor* (DSP). Such an implementation is called a *digital predistorter* (DPD). The idea behind predistortion is presented in Figure 5.16. The predistortion can be divided into two parts, the construction of the predistorter functions and the implementation of the obtained DPD.

The implementation of predistortion methods entails further considerations, and as concluded in Guan and Zhu [2010], "different methodologies or implementation structures will lead to very different results in terms of complexity and cost from the viewpoint of hardware implementation". An implementation using a look-up table will grow quickly with the resolution of the DPD, and thus needs a large chip area, but avoids the necessity of calculations needed in a polynomial implementation (leading to a larger power consumption). The implementation issues have not been considered in this thesis.

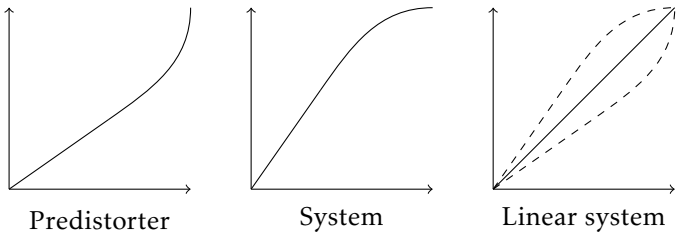


Figure 5.16: *The main idea behind predistortion is to compensate for future nonlinearities and dynamics so that the overall system is linear.*

5.5.1 Volterra series

The theory of p-th order Volterra inverses, introduced in Section 3.2.3, allows for the simpler postinverse (see Section 4.1) to be calculated and then used as the desired preinverse. This is used in the predistortion, or linearization, of for example RF power amplifiers. See also Section 3.2.2 for a discussion on preinverse versus postinverse.

Since Volterra series consist of an infinite sum of integrals, the use of general Volterra theory is rather limited. To reduce the complexity a pruned, or truncated, version of the Volterra series is often used, where the memory length and/or the order of nonlinearity is limited. This heavily reduces the complexity of the sum, but the computational growth is still exponential/polynomial in memory length/order of nonlinearity, limiting the practical use of Volterra series.

Using pruned Volterra series as a means for modeling and predistortion of high-power amplifiers is presented in Tummla et al. [1997] and is shown to work for simulated data with memory length of 1 and nonlinearity order of 7. In Zhu et al. [2008], pruning techniques have been applied to drastically reduce the number of terms in the (discrete time) Volterra series and the method was applied to experimental data. Here, a memory length of 2 and an order of nonlinearity of 11 was used. Volterra based predistorters have also been implemented in *field programmable gate array* (FPGA), shown in Guan and Zhu [2010]. An FPGA is a circuit that can be configured by the user and are used to implement complex digital computations.

5.5.2 Block-oriented Models

Since general nonlinear systems are very difficult to model, a common assumption is that the dynamics are linear, and that the nonlinearity is static, which gives a block-oriented model. This will be the case when there is, for example, a nonlinear actuator (due to saturation) in a control application.

A Hammerstein system consists of a static nonlinear system followed by a linear dynamic system and in a Wiener system, the static nonlinearity is at the output of the linear dynamics, see also Example 3.3. One way to broaden the use of the Hammerstein system is to use a more general *parallel Hammerstein* system, where multiple Hammerstein systems are branched. This structure is often

used in modeling of power amplifiers, where a basic assumption is that the main part of the signal is amplified in a nonlinear way through the PA, and distortions are added to the output. The number of branches in the parallel Hammerstein structure determines the complexity of the model.

In Gilabert et al. [2006], a Wiener model of the PA has been used in combination with a Hammerstein structure predistorter, with memoryless nonlinearities followed by linear blocks using *finite impulse response* (FIR) and *infinite impulse response* (IIR) filters. The implementation of a Hammerstein predistorter in FPGA technique is discussed in Xu et al. [2009] using a WCDMA input signal.

5.5.3 Outphasing Power Amplifiers

In outphasing PAs, there is no linearity between the individual outphasing signals, and any gain or phase mismatch between the two signal paths will cause spectral distortion, see for example Birafane and Kouki [2005] and Romanò et al. [2006]. Typical requirements are approximately $0.1 - 0.5$ dB in gain matching and $0.2^\circ - 0.4^\circ$ in phase matching, which is very hard to achieve [Zhang et al., 2001].

The gain mismatch could be eliminated by adjusting the voltage supplies in the output stage [Moloudi et al., 2008], but this would require an extra, adjustable voltage source on the chip, which is undesirable. For the outphasing amplifier, all amplitudes (within the dynamic range) and phases can be achieved by tuning the outphasing signals $s_1(t)$ and $s_2(t)$, see Figures 5.12 and 5.13. This can be used in the predistortion, so that the two signals are adjusted in a way to compensate for gain errors and possibly other unwanted effects in the PA.

Earlier predistortion methods for outphasing PAs compensate for the gain and phase mismatches in the signal branches. In Myoung et al. [2008], a mismatch detection algorithm has been evaluated using four test signals. These two-tone signals are used to calculate the amplitude and phase mismatches of the amplifier using a closed-form expression, later used for predistortion. Chen et al. [2011] presents a *signal component separator* (SCS) implementation with a built-in compensation for branch mismatches in phase and amplitude. The SCS performs the decomposition of the original signal $s(t)$ into the outphasing signals $s_1(t)$ and $s_2(t)$, (5.11). By taking gain and phase mismatches into account, the SCS has a built-in predistorter.

Helaoui et al. [2008] discuss the impact of the combiner on the outphasing PA performance. The choice of combiner is a trade-off between linearity and power consumption. Nonlinearities can be introduced by a nonisolated combiner such that the output distortion depends on the input power. These nonlinearities were successfully reduced by the use of a predistorter.

The solutions in Myoung et al. [2008] and Chen et al. [2011] consider the gain mismatch between the two branches and compute the ideal phase compensation when the outputs are approximated as two signals with constant amplitudes. This is possible when there is no interaction between the amplifier stages. In this thesis, the outputs are still considered as two constant amplitude signals generating amplitude and phase distortion. Furthermore, an amplitude dependent phase distortion, occurring due to the interaction and signal combining of

the amplifiers' outputs, is also considered.

Parts of the results in Chapters 6-8 can also be found in Fritzin et al. [2011a] and Jung et al. [2013]. The nonconvex algorithm, presented in Fritzin et al. [2011a], has in Landin et al. [2012] been developed to include a method for finding good initial values to the nonlinear optimization. However, the basic problem of nonconvexity has not been solved there and local minima still risk posing problems in the optimization. In Jung et al. [2013], the nonconvex formulation has been reformulated into a convex method. In this method, the PA model is estimated in a least-squares setting and an analytical calculation of the predistorter is used. Furthermore, a theoretical characterization of an outphasing PA is presented and form a basis for an ideal DPD. This characterization has also been used to obtain an estimate thereof.

6

Modeling Outphasing Power Amplifiers

In this chapter, one way of modeling of the outphasing power amplifier using knowledge of the physical structure of outphasing amplifiers is presented. It consists of a new decomposition of the outphasing signals making use of the knowledge of the uneven amplification in the two branches, as well as a way to incorporate the possible nonlinearities in the branches.

Despite the fact that the PA is analog and the baseband model is in discrete time, the notation t is used to indicate the dependency on time. Based on the context, t may thus be a continuous or discrete quantity and denote the time or the time indexation. For notational convenience, the explicit dependency on time will be omitted in parts of this chapter and the following one.

6.1 An Alternative Outphasing Decomposition

As mentioned in Chapter 5, the PA output signal $y(t)$ is a distorted version of the input signal. The nonlinearities are due to *(i)* the nonidentical gain factors g_1 and g_2, and *(ii)* nonlinear distortion in the amplifier branches. First, a novel decomposition will be described, accounting for the nonidentical gain factors g_1 and g_2, followed by a description of how these can be used in the modeling of the outphasing power amplifier. Since it is desired that the predistorter should invert all effects of the PA except for the gain, the signals can be assumed to be normalized such that

$$\max_t |s(t)| = \max_t |y(t)| = 1. \tag{6.1}$$

As described in Figure 5.12, the amplitude information of the original input signal $s(t)$ can be found in the angle between $s_1(t)$ and $s_2(t)$. Let

$$\Delta_\psi(s_1, s_2) = \arg(s_1) - \arg(s_2) \tag{6.2}$$

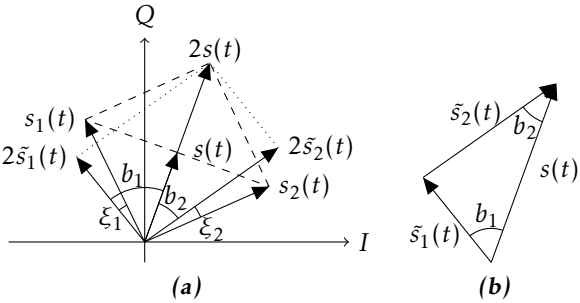


Figure 6.1: *(a) Decomposition of the input signal $s(t)$ into $s_1(t)$ and $s_2(t)$ when $g_1 = g_2 = g_0 = 0.5$ and into $\tilde{s}_1(t)$ and $\tilde{s}_2(t)$ when decomposed as in (6.3) with nonidentical gain factors g_1 and g_2. (b) Trigonometric view of the decomposition of $s(t)$ using nonidentical gain factors. Note that $|\tilde{s}_k| = g_k, k = 1, 2$.*

denote the phase difference of the outphasing signals $s_1(t)$ and $s_2(t)$. Since the amplitude of the nondecomposed signal in the outphasing system is determined by $\Delta_\psi(s_1, s_2)$, this difference can be used instead of the actual amplitude in many cases. For notational convenience, Δ_ψ will be used instead of $\Delta_\psi(s_1, s_2)$, unless specified otherwise. Here, all phases are assumed unwrapped.

To describe the distortions caused by the imperfect gain factors, consider again the decomposition of $s(t)$ into $s_1(t)$ and $s_2(t)$ in (5.11). This is only valid when $g_1 = g_2$ but we can use an alternative decomposition of $s(t)$ into $\tilde{s}_1(t)$ and $\tilde{s}_2(t)$ such that

$$\tilde{s}_1(t) + \tilde{s}_2(t) = s(t), \tag{6.3a}$$

$$|\tilde{s}_k| = g_k, \quad k = 1, 2, \quad \text{and} \tag{6.3b}$$

$$\arg(\tilde{s}_1) \geq \arg(\tilde{s}_2). \tag{6.3c}$$

Assuming knowledge of g_1 and $g_2 = 1 - g_1$ and given $s(t)$, the signals $\tilde{s}_1(t)$ and $\tilde{s}_2(t)$ can be computed from (6.3). Let

$$b_1 = \arg(\tilde{s}_1) - \arg(s)$$

and

$$b_2 = \arg(s) - \arg(\tilde{s}_2)$$

denote the angles between the decomposed signals and $s(t)$ as shown in Figure 6.1a.

Figure 6.1b shows that the decomposition can be viewed as a trigonometric problem and application of the law of cosines gives

$$g_2^2 = g_1^2 + |s|^2 - 2g_1|s|\cos(b_1) \tag{6.4}$$

and

$$g_1^2 = g_2^2 + |s|^2 - 2g_2|s|\cos(b_2). \tag{6.5}$$

The angles b_1 and b_2 that define $\tilde{s}_1(t)$ and $\tilde{s}_2(t)$ can be computed from these expressions and can be viewed as functions of Δ_ψ since $|s| = r_{\max}\cos(\Delta_\psi/2)$. This means that the angles

$$\xi_1(\Delta_\psi) \triangleq \arg(\tilde{s}_1) - \arg(s_1) = b_1 - \frac{1}{2}\Delta_\psi \tag{6.6}$$

and

$$\xi_2(\Delta_\psi) \triangleq \arg(\tilde{s}_2) - \arg(s_2) = \frac{1}{2}\Delta_\psi - b_2 \tag{6.7}$$

can also be viewed as functions of Δ_ψ.

When the goal is to model the phase distortions in the two branches, this alternative way of defining the decomposition reflects the physical behavior better than the standard outphasing decomposition in (5.11). The output $y(t)$ can be decomposed in the same way to $y_1(t)$ and $y_2(t)$, taking the gain factors g_1 and g_2 into account.

6.2 Nonconvex PA Model Estimator

A first step on the way to model the outphasing PA is to observe that although the two branches are identical in theory, once implemented in hardware this will not be the case. Since the signals $s_1(t)$ and $s_2(t)$ are amplified by two different amplifiers, there might be a small amplification difference resulting in a gain offset between these signals, as well as a time delay stemming from the fact that $s_1(t)$ and $s_2(t)$ take different paths to the power combiner. With this insight, a first model structure with a gain mismatch between g_1 and g_2 and a phase shift δ in one branch is proposed. This leads to a model structure described by

$$y(t) = g_1 e^{j\delta} s_1(t) + g_2 s_2(t), \tag{6.8}$$

where g_1, g_2 and δ are real-valued constants.

When adding more complex behavior to the model structure, the structure of the physical PA must still be kept in mind. The separation of the two branches is still valid, but each branch can be affected by other factors than the gain difference and possible phase shift. As the amplitudes of the outphasing signals are fixed, a phase dependent distortion in each branch is proposed.

To model an amplitude dependent phase shift while keeping in mind the constant amplitude of the signals $s_1(t)$ and $s_2(t)$, a model structure with an exponential function can be used. An amplitude-dependent phase distortion in $y_k(t), k = 1, 2$ (the two amplifier branches) can be written as

$$y_k(t) = g_k e^{j f_k(\Delta_\psi)} s_k(t), \quad k = 1, 2, \tag{6.9a}$$

$$y(t) = y_1(t) + y_2(t), \tag{6.9b}$$

as in Figure 6.2. Here, f_1 and f_2 are two real-valued functions describing the phase distortion

$$\arg(y_k) - \arg(s_k) = f_k(\Delta_\psi), \quad k = 1, 2, \tag{6.10}$$

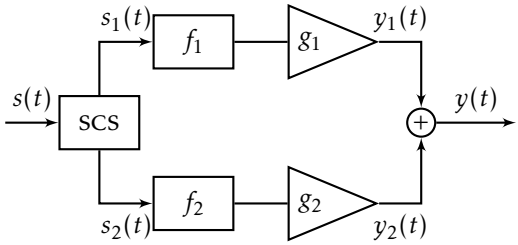


Figure 6.2: *A schematic picture of the amplifier branches setup. Note that the functions f_k, $k = 1, 2$, are not functions of the input to the block only but are used to show the general functionality of the* PA *with the separation of the two branches.*

in each signal path. Furthermore, g_1 and g_2 are the gain factors in each amplifier branch. Hence, an ideal PA would have $f_1 = f_2 = 0$ and $g_1 = g_2 = g_0$ and any deviations from these values will cause nonlinearities in the output signal and spectral distortion as previously concluded.

The functions f_1 and f_2 describing the phase distortion in the separate branches can be described by arbitrary basis functions. Here, polynomials

$$\hat{f}_k = p(\eta_k, \Delta_\psi) = \sum_{i=0}^{n} \eta_{k,i} \Delta_\psi^i, \quad k = 1, 2, \tag{6.11}$$

where

$$\eta_k = \begin{pmatrix} \eta_{k,0} & \eta_{k,1} \cdots & \eta_{k,n} \end{pmatrix}^T,$$

have been used as parameterized versions of the functions f_k, motivated by the Stone-Weierstrass theorem, see Rudin [1976, Theorem 7.26].

The model parameters in the given model structure are estimated by minimizing a quadratic cost function [Ljung, 1999] as in

$$\hat{\theta} = \arg\min_{\theta} V(\theta), \tag{6.12}$$

$$V(\theta) = \sum_{t=1}^{N} \left| y(t) - \hat{y}(t, \theta) \right|^2 \tag{6.13}$$

with

$$\hat{y}(t, \theta) = g_1 e^{j\, p(\eta_1, \Delta_\psi(s_1, s_2))} s_1(t) + g_2 e^{j\, p(\eta_2, \Delta_\psi(s_1, s_2))} s_2(t) \tag{6.14}$$

where $\theta = [g_1 \quad g_2 \quad \eta_1^T \quad \eta_2^T]^T \in \mathbb{R}^{2n+4}$, $y(t)$ is the measured output data and $\hat{y}(t, \theta)$ is the modeled output. The model (6.14) can be compared to the structure (6.9), where $y(t) = g_1 e^{j f_1(\Delta_\psi)} s_1(t) + g_2 e^{j f_2(\Delta_\psi)} s_2(t)$. This structure leads to a nonlinear and nonconvex optimization problem, so the minimization algorithm might find a local optimum instead of a global. In order to obtain a good minimum in a nonconvex optimization problem, it is essential to have good initial

values, and one way to obtain these is presented in Landin et al. [2012]. For further discussions on convexity and nonconvexity, see Section 6.5.

Here, a model of the PA was estimated by minimizing a quadratic cost function measuring the difference between the measured and modeled output signal. This estimation problem involves solving a nonconvex optimization problem. However, using the knowledge of the structure of the outphasing amplifier, there is an alternative way which essentially only involves solving standard least-squares problems, presented in the next section.

6.3 Least-squares PA Model Estimator

The output distortions originate both from imperfect gain factors and nonlinearities in the amplifiers. Once the gain factor impact has been accounted for, the amplifier nonlinearities can be modeled. This means that the modeling optimization problem can also be rewritten as a *separable least squares* (SLS) problem, also presented in Jung et al. [2013]. A separable least squares problem is when one set of parameter enters the model linearly and one set nonlinearly. Given the nonlinear parameters, the linear part can be solved for efficiently, leaving a nonlinear problem of a lower dimension [Ljung, 1999]. See also Section 2.5 for a short introduction to SLS problems.

Often, the minimization is done first for the linear part and then the nonlinear parameters are solved for and this nonlinear minimization problem now has a reduced dimension. Here, the idea is to use knowledge of the gain factors to make a nonlinear transformation of the data using the decomposition (6.3). Once this decomposition is done, the minimization can be rewritten as a *least-squares* (LS) problem in the phase distortion in the two branches. This is not the usual SLS method since it involves a nonlinear transformation of the data, but the basic idea of separating out the nonlinear parameters to obtain a LS problem still applies. We will here explore two ways of estimating the gain factors g_1 and g_2. One is based on the dynamic range of the PA and the other is based on a parameter gridding of possible values of g_1 and g_2.

Assuming the gain factors to be known, we know what the phases of the outputs from the two outphasing branches must be in order for the two signals to sum up to the measured output $y(t)$. It is now possible to decompose the output $y(t)$ into $y_1(t)$ and $y_2(t)$, using the decomposition in Section 6.1. What is left to determine is the phase distortion in the branches, described by the functions f_k. Since the gain factor influence is handled by the alternative decomposition of $y(t)$, the phase distortion is now described by the difference between the phase of the input $s_k(t)$ and the output $y_k(t), k = 1, 2$ and this can be formulated as a least-squares problem.

Consider first the two gain factors g_1 and $g_2 = 1 - g_1$, where the relation between them comes from the normalization (6.1). Let

$$\begin{aligned} g_1 &= g_0 \pm \Delta_g, \\ g_2 &= g_0 \mp \Delta_g, \end{aligned} \tag{6.15}$$

where $\Delta_g \geq 0$ represents the gain imbalance between the amplifier stages and $g_0 = 0.5$. Inserting (6.15) into (5.13) gives

$$c_{DR} = 20 \log_{10}\left(\frac{g_0}{\Delta_g}\right). \tag{6.16}$$

Hence, the imbalance term Δ_g can be computed as

$$\Delta_g = g_0 \cdot 10^{-c_{DR}/20}, \tag{6.17}$$

making it possible to find approximations of g_1 and g_2 from the dynamic range of the output signal. The value of c_{DR} can be estimated from measurements as the ratio between the maximum and minimum output amplitudes. The estimate is noise sensitive, but this can be handled by averaging multiple realizations. These approximations are valid for input signals with large peak to minimum power ratios, like WCDMA and LTE, where the PA generates an output signal including its peak and minimum output amplitudes, i.e., its full dynamic range. If this is not fulfilled or the noise influence is too large, an alternative approach is to evaluate a range of values of g_1 and $g_2 = 1 - g_1$ and then solve the PA modeling problem for each pair of gain factors, as in the usual SLS approach.

Once the gain factors have been determined, $s(t)$ can be decomposed into $\tilde{s}_1(t)$ and $\tilde{s}_2(t)$, and $y(t)$ into $y_1(t)$ and $y_2(t)$ using (6.3) to (6.5). Furthermore, the standard outphasing decomposition of $s(t)$ into $s_1(t)$ and $s_2(t)$ as in (5.11) will be used in the sequel.

Since the gain factor mismatch has been accounted for, it is now possible to determine the impact of the nonlinearities on the two branches. The phase distortion in each signal path caused by the amplifiers can thus be modeled from measurements of $s(t)$ and $y(t)$. Here, polynomials

$$p(\eta_k, \Delta_\psi) = \sum_{i=0}^{n} \eta_{k,i} \Delta_\psi^i, \quad k = 1, 2,$$

have been used as parameterized versions of the functions f_k, as in (6.11). Estimates $\hat{\eta}_{k,i}$ of the model parameters $\eta_{k,i}$ have been computed by minimizing a quadratic cost function, i.e.,

$$\hat{\eta}_k = \arg\min_{\eta_k} V_k(\eta_k), \quad k = 1, 2, \tag{6.18}$$

where

$$V_k(\eta_k) = \sum_{t=1}^{N} \Big(\arg\left(y_k(t)\right) - \arg\left(\tilde{s}_k(t)\right) - p\left(\eta_k, \Delta_\psi(s_1(t), s_2(t))\right)\Big)^2, \tag{6.19}$$

and

$$\eta_k = \begin{pmatrix} \eta_{k,0} & \eta_{k,1} \cdots & \eta_{k,n} \end{pmatrix}^T.$$

The cost function (6.19) can be motivated by the fact that the true functions f_k satisfy (6.10) when the amplifier is described by (6.9). Minimization of V_1 and V_2

are standard least-squares problems, which guarantees that the global minimum will be found [Ljung, 1999].

Once the LS problem is solved for each setup of g_1 and g_2, the problem of finding the best setup is now reduced to a one dimensional (possibly nonconvex) optimization problem over $g_1(g_2 = 1 - g_1)$, which is much easier to solve than the original, multidimensional problem. A problem this small can be solved at a small computational cost.

The parameter estimates $\hat{\eta}_k$ define function estimates

$$\hat{f}_k(z) = p(\hat{\eta}_k, z), \quad k = 1, 2, \tag{6.20}$$

that, together with the gain factor estimates $\hat{g}_1$ and $\hat{g}_2$ describe the power amplifier behavior. The different steps are also described in Part A – Estimation of PA model in Algorithm 1, page 91.

The alternative decomposition described in Section 6.1 depends on the gain factors g_1 and g_2 via a nonlinear relation, but with these given, the problem is reduced to a LS-problem in the phase as in (6.19). If the gain factor estimation is done using the DR as in (6.15) and (6.17), this will result in two LS-problems to solve, and gridding of g_1 will result in $\left(\frac{g_{\max}-g_{\min}}{p_M} + 1\right)$ LS problems. The values $g_{\min}$ and $g_{\max}$ bounds the values of g_1 and g_2 that one wants to evaluate and p_M is the precision, so that $g_1 \in [g_{\min}, g_{\min} + p_M, \ldots, g_{\max}]$ and $g_2 = 1 - g_1$. Compare to Algorithm 1, page 91, for notation. This is not the standard SLS method, since a nonlinear transformation of the data is done before solving the LS problem, but the separation of the linear and nonlinear parameters applies. This separation reduces the optimization to a number of LS problems and a nonlinear optimization in only one dimension, g_1 ($g_2 = 1 - g_1$ due to the normalization (6.1)). This is clearly a reduction from the nonlinear optimization in $2n + 4$ dimensions of the original problem.

6.4 PA Model Validation

As an evaluation of the different approaches presented above, the models have been compared. The Figures 6.3-6.6 present the amplitude and phase of the measured output and the model output. The amplitude error $|y - \hat{y}|$ and the phase error $\arg(y) - \arg(\hat{y})$ are also included. The first simple model in (6.8), using only the gain factors g_1 and g_2 and a phase shift δ, is presented in Figure 6.3. The more complex model structure (6.14) is presented in Figures 6.4, 6.5 and 6.6, using the different modeling methods. The model obtained by the nonconvex approach as in (6.12)-(6.14) is presented in Figure 6.4. The LS method using (6.18)-(6.19) and the dynamic range to obtain the gain factors is presented in Figure 6.5. In Figure 6.6, the LS method using gridding of g_1 over a range of values and then determining the best fit is presented.

The more complex models perform very well, and rather similarly. This is easier to see in Figure 6.7, where the errors for the different modeling methods are plotted together. Though the models all perform well, there are still errors. These errors are largest where the input amplitude is small, such as around time

Table 6.1: *PA Model Validation*

Method	g_1	g_2	$\|y - \hat{y}\|_2^2$, (6.13)
Delay only, model structure (6.8)	0.4911	0.5089	62.99
Nonconvex	0.4986	0.5014	0.9985
LS, grid	0.50	0.50	1.119
LS, DR	0.4994	0.5006	0.9781

152 μs and 156.5 μs. The result of the DR LS model is also presented in an IQ plot in Figure 6.8 where the signals are plotted in the complex plane. Also in this plot, the model shows a very good behavior, with a slightly worse performance for small amplitudes.

The gain factor estimates are presented in Table 6.1 together with the cost function (6.13) for the different methods. As seen in the rightmost column where $|y - \hat{y}|_2^2$, (6.13), is presented, the added model complexity with nonlinearities makes a large improvement in the model fit. The LS method using DR and the nonconvex method achieve rather similar results with the gridding LS method slightly behind. The results of the nonconvex method depends on the number of iterations used in the optimization.

Except for the first simple model, the other methods perform very similarly with a very good fit to validation data. This clearly shows that the nonlinear extension to the model has a significant impact on the model properties. This also means that the choice of method comes down to other considerations than the fit. The lack of guarantees of convergence to a global minimum of nonconvex optimization methods is a reason to avoid the method described in Section 6.2. If the LS method is chosen, this also entails the choice of gridding or using the dynamic range. Gridding is more robust against noise, since the DR estimation is done using only two measurements (the one with minimal and the one with maximal amplitude), so noise at either of these data points will have a large impact. A drawback with gridding is the risk of missing the "true" value, if the precision p_M (difference in g_1 and g_2) is chosen too large. A decreasing p_M, on the other hand, will increase the number of LS problems that need to be solved. Benefits and drawbacks for the dynamic range method are the opposite.

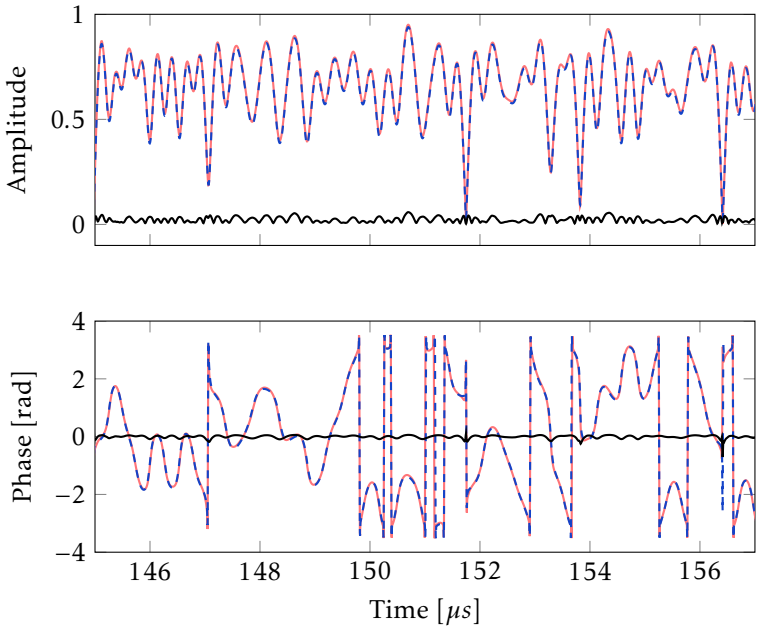


Figure 6.3: *Model validation of the model produced using the first structure* (6.8)*, with gain factors g_1 and g_2 and a phase shift δ only. The upper plot shows the amplitude of the measured signal (solid pink), the model output (dashed blue) and the error (black). The lower plot shows the phase.*

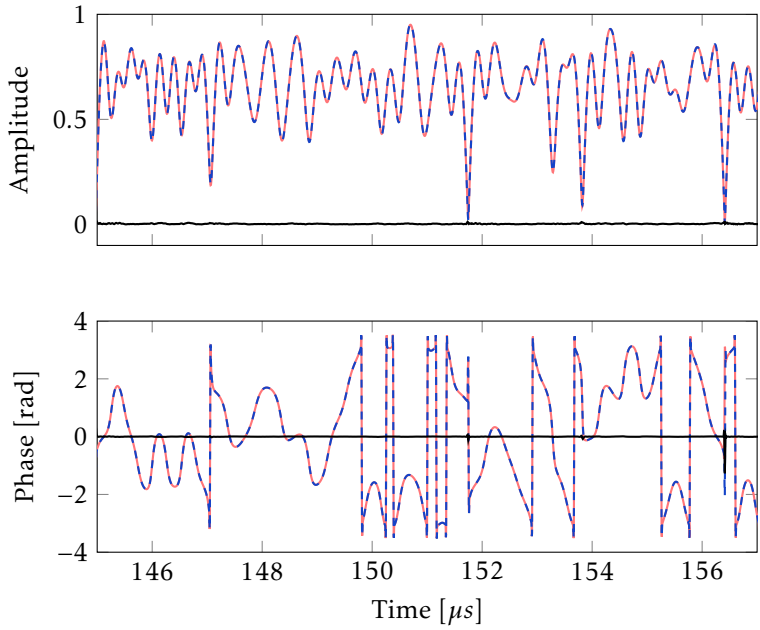


Figure 6.4: *Model validation of the model produced using the original, non-convex, optimization in* (6.12)-(6.14)*. The upper plot shows the amplitude of the measured signal (solid pink), the model output (dashed blue) and the error (black). The lower plot shows the phase.*

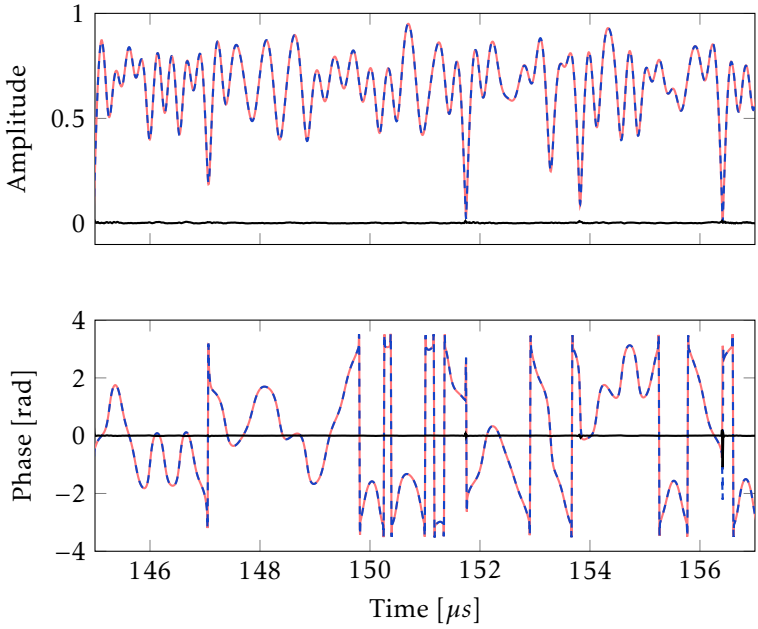


Figure 6.5: *Model validation of the model produced using the convex method in* (6.18)-(6.19) *and the dynamic range has been used to determine the gain factors as in* (6.17) *and* (6.15)*. The upper plot shows the amplitude of the measured signal (solid pink), the model output (dashed blue) and the error (black line). The lower plot shows the phase.*

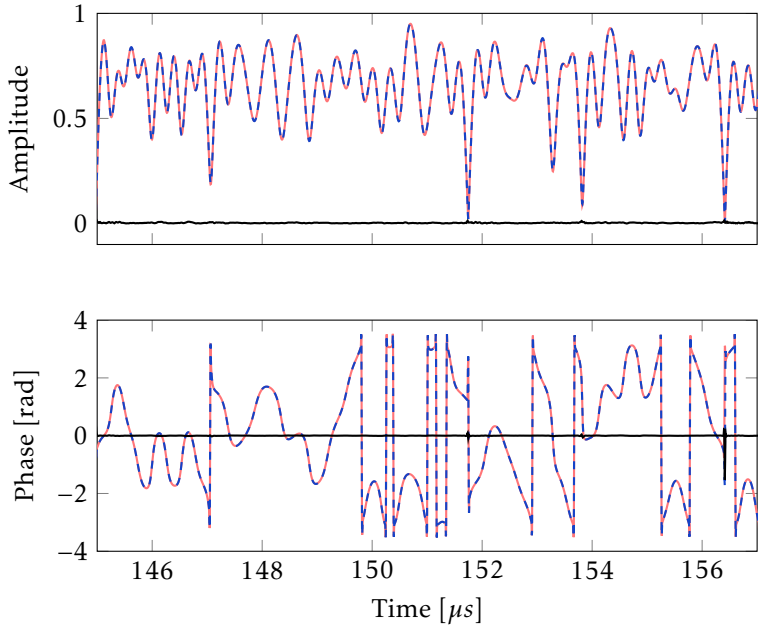


Figure 6.6: *Model validation of the model produced using the convex method in* (6.18)-(6.19) *and* g_1 *has been gridded in* $[g_{min}, g_{max}] = [0.4, 0.6]$ *with precision* $p_M = 0.005$. *The upper plot shows the amplitude of the measured signal (solid pink), the model output (dashed blue) and the error (black line). The lower plot shows the phase.*

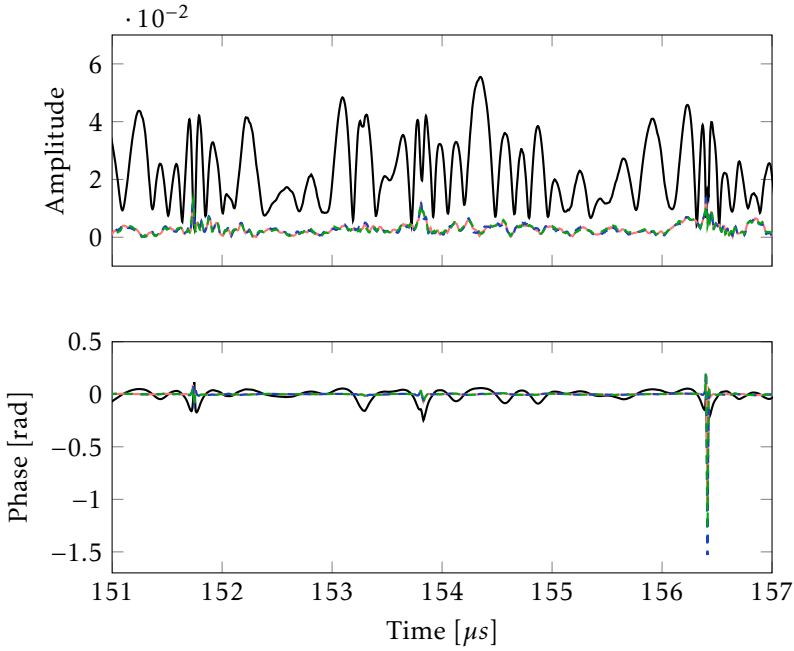


Figure 6.7: *A summary of the model errors of the different models. The upper plot shows the amplitude error* $|y - \hat{y}|$ *and the lower plot shows the phase errors* $\arg(y) - \arg(\hat{y})$*. The simple model* (6.8) *is plotted in black, the* LS *methods using* DR *in solid pink and gridding in dashed blue. The model obtained by the nonconvex method is plotted in a green dashed line. The three models describing a nonlinear behavior perform very well and in a very similar way, as seen in the figure where the lines are almost on top of each other.*

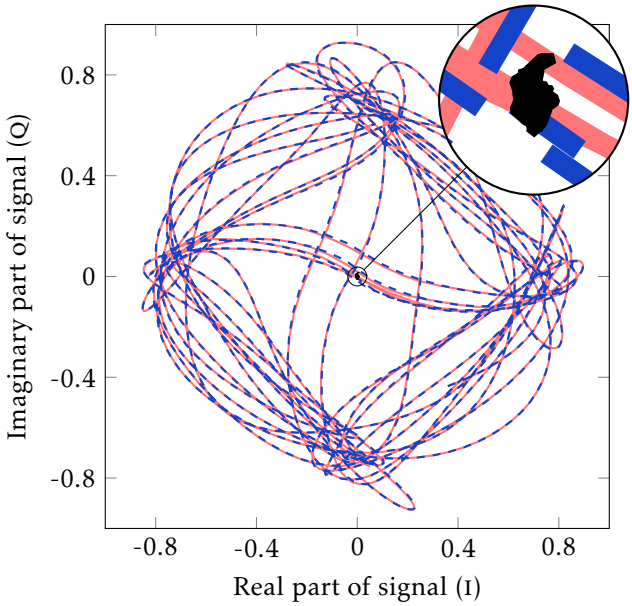


Figure 6.8: *IQ plot (imaginary part, Q, vs real part, I) of the measured signal (solid pink) and the model output (dashed blue) and the error $y - \hat{y}$ (black). The model was estimated by the LS method using DR to estimate g_1 and g_2. The zoom-in in the upper right corner is a ten times amplification of the error signal.*

The estimated phase distortion functions, $\hat{f}_1$ and $\hat{f}_2$, from the models can be plotted as functions of Δ_ψ and the results for a WCDMA signal for the different methods are rather similar. The function $\hat{\hat{f}}$ describes the phase change between the two outphasing signals at the output, and thus the amplitude change of the output. The phase distortion functions $\hat{\hat{f}}$ are presented in Figure 6.9 as deviations from the ideal phase distortion, which should be as close to zero as possible. The ideal phase distortion includes the compensation for nonequal gain factors. By this, it is clear that at amplitudes close to zero (Δ_ψ close to π), a zero distortion will not be possible for nonequal gain factors. In Figure 6.10, the different functions $f_k, k = 1, 2$, are shown for the different methods. The methods achieve rather similar results, but at the expense of the number of computations in the nonconvex approach, where 25 000 function evaluations have been performed to achieve the optimum.

Even though the methods result in similar validation results, the largest differences are found close to the edges of the interval. In the WCDMA signal, 99.1% of the measured data points have $0.8 \leq \Delta_\psi \leq 3.0$, so the focus of the fit is where the most data points are. Compared to Figure 5.12 and (6.2), it is clear that the data points with a very large Δ_ψ (close to π) have a very small amplitude, and errors in the phase distortion modeling might not affect as much as the data points with a small Δ_ψ (large amplitude). It can thus be concluded that it could be more important to obtain a good model for small values of Δ_ψ than for large values (something that could be achieved by weighting functions). It can also be noted that, if the amplitude of the input had been used instead of the angle $\Delta_\psi = \arg(s_1) - \arg(s_2)$, more weight would have been put at the largest amplitudes. This is not done now since a large input amplitude equals a small Δ_ψ and vice versa.

In polynomial fitting, the agreement with the function f is often bad at the outer parts of the interval to be approximated. If one can choose the points at which the polynomial is to be fitted, Chebyshev points should be chosen, with more points at the outskirts of the interval [Dahlquist and Björck, 2008, p. 377-379]. Here, we are fitting a polynomial using the method of least squares, but the same reasoning holds. To obtain a smaller error at the peak power, more data points could have been collected there. Instead, the least-squares fitting focuses on fitting the overall performance, and hence more effort is made to obtain a small error in the parts where there is a larger point density. For the signals used in this thesis, this area of larger point density is in the center of the interval, where an improvement will be clearly seen in for example Figure 8.9. We will return to this subject in Chapter 8 when evaluating the predistortion results.

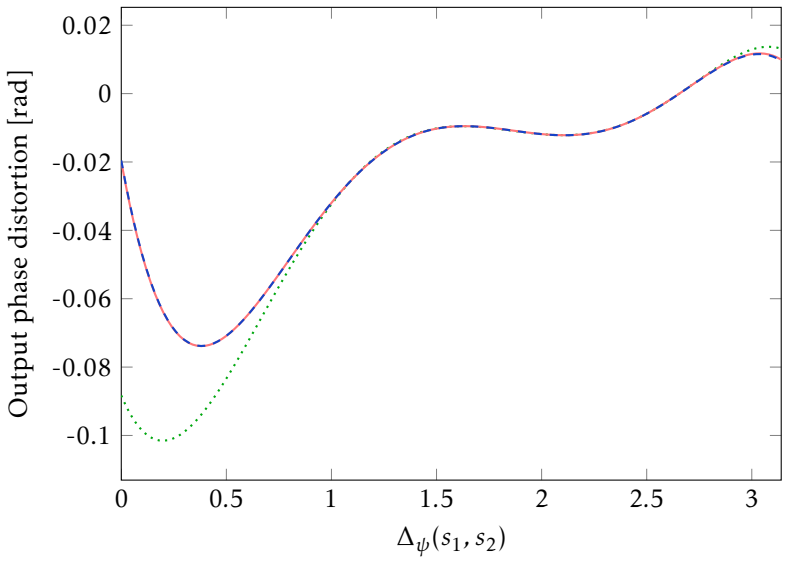


Figure 6.9: *Simulated output phase distortion of the models from the non-convex method (dotted green) and the* LS *methods using* DR *(dashed blue) and gridding (pink) (the two model outputs are almost completely on top of each other). The lines describe the modeled phase difference as a function of the input signal amplitudes, that is, taking the different gain factors into account. The three methods evaluated estimate the phase shift almost equally for the middle range where most of the data points are (99.1% have* $0.8 \leq \Delta_\psi \leq 3.0$*), but the differences are visible at the edges.*

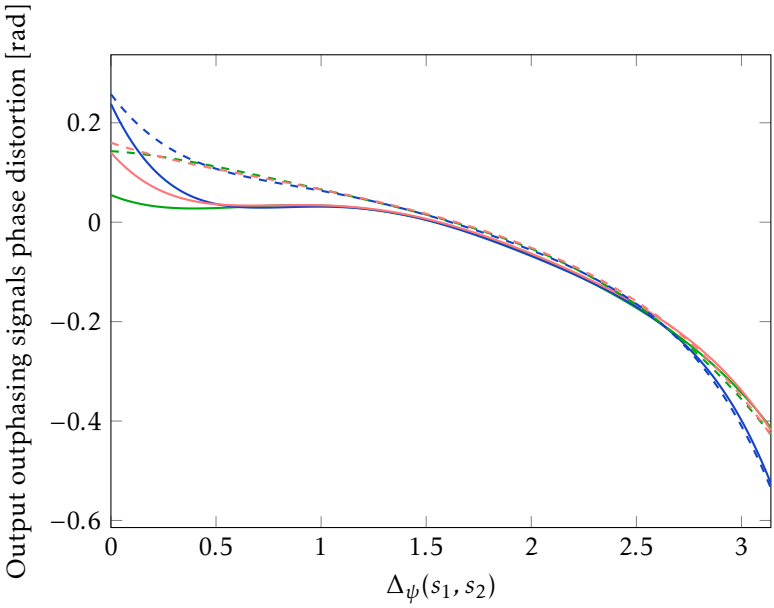


Figure 6.10: *Simulated outphasing output phase distortion of the models from the nonconvex method (green) and the* LS *methods using* DR *(blue) and gridding (pink). The lines describe the modeled phase in each branch as a function of the input signal amplitudes. Branch one is plotted in solid lines and branch two in dashed lines.*

6.5 Convex vs Nonconvex Formulations

The minimization of the cost function (6.12)-(6.14) is a nonconvex optimization problem in $2n + 4$ dimensions with possible presence of local minima. Nonconvex optimization problems can either be solved by a local optimization method or a global one. A local optimization method minimizes the cost function over points close to the current point, and guarantees convergence to a local minimum only. Global methods find the global minimum, at the expense of efficiency [Boyd and Vandenberghe, 2004]. Hence, even under ideal conditions (noise-free data, true PA described exactly by one model with the proposed structure), there is no guarantee that the nonconvex approach will produce an optimal model of the PA in finite time. The least-squares approach in (6.18)-(6.19) does exactly this and results in a closed-form expression for the parameter estimate. This is a major advantage since it removes the need for error-prone sub-optimality tests and possible time-consuming restarts of the search algorithm. Additionally, the computation time for the iterative, nonconvex, and potentially sub-optimal solution is significantly longer compared to the least-squares method.

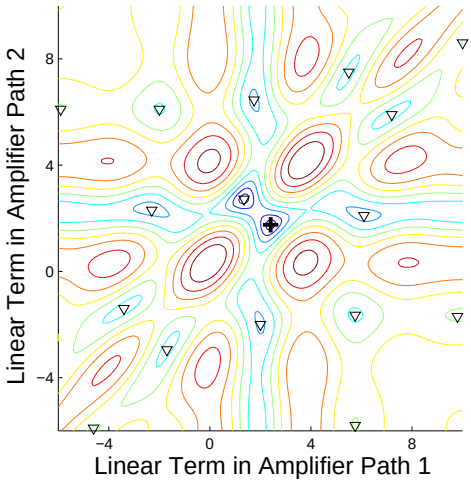
A two dimensional projection of the cost functions to be minimized, (6.13) in the nonlinear formulation and (6.19) in the LS reformulation, can be seen in Figure 6.11. All parameters but two have been fixed at the optimum, and the linear term in each amplifier branch ($\eta_{k,1}$ in (6.11)) has been varied. Clearly, there is a risk of finding a local minimum in the nonconvex formulation illustrated in (a) whereas there is only one (global) optimum in the least-squares formulation in (b).

The local minima in themselves might not be a problem if they are good enough to produce a well performing DPD, but there are no guarantees that this is the case. Typically, a number of different initial points need to be tested in order to get a reasonable performance.

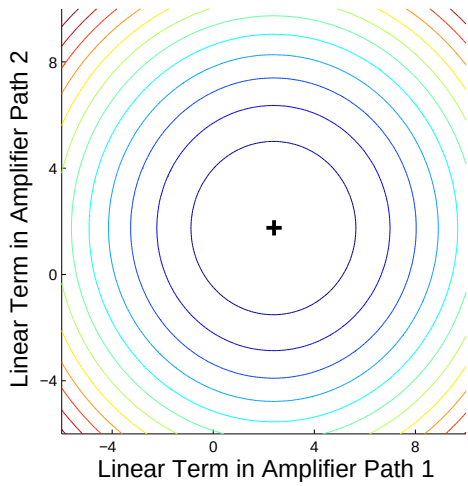
6.6 Noise Influence

Noise is always present in measurements, and the noise will effect the models. The algorithms presented in this chapter are sensitive to noise especially in two steps; the normalization $g_1 + g_2 = 1$ in (6.1) and the calculation of c_{DR} in (5.13). Both these calculations are based on very few measurements, one for the normalization (the largest amplitude) and two for the DR calculation (the smallest and the largest amplitudes), so noise at these instances might have a large influence on the estimation, and thus the performance of the predistorter.

The measurements used for the modeling and model validation in this chapter were recorded using the same measurement setup and power amplifier that will be used in Section 8.4. To avoid the influence of measurement noise, the same input was applied a number of times, K, and the output was measured, whereupon the average over the different realizations was calculated. In measurements used for the PA model estimation described here, $K = 10$. No automatic synchronization between input and measured output is done, so a manual syn-



(a)



(b)

Figure 6.11: *Two dimensional projections of the cost functions of (a) the original nonconvex optimization problem* (6.12)-(6.14) *and (b) the least-squares reformulation,* (6.18)-(6.19) *using the dynamic range for the estimation of* g_1 *and* g_2*. All but two parameters in each amplifier branch have been fixed at the optimal value, and the linear terms* ($\eta_{k,1}$ *in* (6.11)) *are varied. In (a), the visible local minima are marked with* $\triangledown$ *and the minimum obtained clearly depends on the initial point of the local optimization. In the least-squares formulation illustrated in (b), there is only one minimum (the global one) and convergence is guaranteed. The + marks the global minimum.*

chronization has to be performed. This also means that the sample times of the output differ between different measurement sets and that the synchronization between input and output is not the same for different data sets. When looking at the different data sets, the most dominant noise effect seems to stem from this time mismatch, which is evenly distributed around the mean value. The noise levels in general are very low.

6.7 Memory Effects and Dynamics

A more complex model structure has also been investigated by adding memory, that is to say that the output depends not only on the current input but also on the previous inputs, as in the model structure

$$p_{mem}(\alpha, \bar{\beta}_{n_m}(s)) = \sum_{m=0}^{n_m} \sum_{j=0}^{n} \alpha_{mj} \beta(s(t-m))^j, \tag{6.21}$$

with a memory depth n_m, where

$$\bar{\beta}_{n_m}(s) = \Big(\beta\big(s(t-m)\big)\Big)_{m=0}^{n_m}. \tag{6.22}$$

This approach did not lead to a better fit in the model validation, nor did it give any significant improvement in predistortion.

If dynamics are present in the PA, it is not unreasonable to assume that they would appear in the combiner, since the amplifier components in each branch can be assumed to contribute with little dynamics. This would mean that we have a parallel Hammerstein system with two parallel nonlinear, static branches (the amplifiers) followed by a dynamic system (the combiner). To investigate how such dynamics would effect the method described above, a dynamical system has been simulated at the output of a static model. The model was estimated using the LS method with DR. The dynamical system was a first order system with different values of the time constant in the range $[0.2T_s\ 5T_s]$, where T_s is the sample time. The same identification method was then applied to this data. In this case, the decomposition of the output using an estimate of g_1 and g_2 (obtained by dynamic range or gridding), is no longer a good approximation of the system, and the method will not perform in a satisfactory way. Thus, further investigation of how to include dynamics is needed.

7

Predistortion

Power amplifiers in communication devices are often nonlinear and/or dynamic, which causes interference in adjacent transmitting channels. To reduce this interference, linearization is needed. This is preferably done at the input, so that a prefilter inverts the nonlinearities/dynamics. This prefilter is called a *predistorter* (PD). Originally, these predistorters consisted of small analog circuits, but now they are often implemented in a *look-up table* (LUT) or a *digital signal processor* (DSP). Such an implementation is called a *digital predistorter* (DPD).

For the outphasing amplifiers evaluated in this thesis, the gain mismatch could be eliminated by adjusting the voltage supplies in the output stage, but this would require an extra adjustable voltage source on the chip, which is undesirable. Instead, the goal is to find a predistorter that uses only the phases of the two outphasing signals. By adjusting the outphasing signals, it is possible to achieve all amplitudes (within the dynamic range) and phases, and this idea will be explored in the construction of a predistorter.

In this chapter, a description of an ideal DPD will be presented and different methods to obtain it will be described. As a first step, the evaluation of the predistorters will be based on a model of the PA (described in Chapter 6), on simulated data only. In Chapter 8, the predistorters will be evaluated on real measurement data.

7.1 A DPD Description

With the description of the power amplifier in (6.9)-(6.10), it is clear that an ideal PA would have $f_1 = f_2 = 0$ and $g_1 = g_2 = g_0 = 0.5$ and any deviations from these values will cause nonlinearities in the output signal and spectral distortion. In order to compensate for these effects, a DPD can be used to modify the input outphasing signals to the two amplifier branches, i.e., $s_1(t)$ and $s_2(t)$.

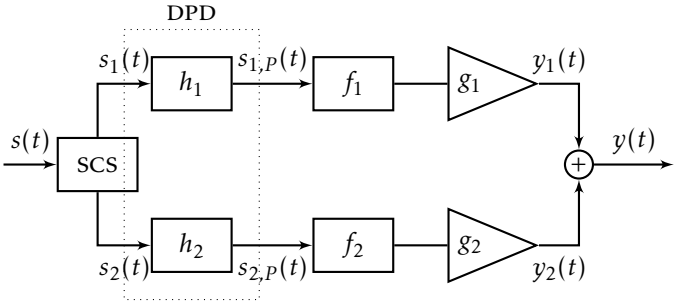


Figure 7.1: *A schematic picture of the amplifiers with predistorters. Note that the functions f_k and h_k, $k = 1, 2$, are not functions of the input to the block only, but are used to show the general functionality of the* PA *and the* DPD *with the separation of the two branches.*

Since the outputs of the Class D stages (the amplifiers in each branch) have constant envelopes, the DPD may only change the phase characteristics of the two input outphasing signals. With this in mind, a DPD that produces the predistorted signals

$$s_{k,P}(t) = e^{j\,h_k(\Delta_\psi)} s_k(t), \quad k = 1, 2, \tag{7.1}$$

to the two amplifier branches is proposed. Here, h_1 and h_2 are two real-valued functions that depend on the phase difference between the two signal paths. By modifying the signals in each branch using the DPD in (7.1), shown in Fig. 7.1, the predistorted PA output $y_P(t)$ can be written

$$y_P = \underbrace{g_1 e^{j\,f_1(\Delta_\psi(s_{1,P}, s_{2,P}))} s_{1,P}}_{\triangleq y_{1,P}} + \underbrace{g_2 e^{j\,f_2(\Delta_\psi(s_{1,P}, s_{2,P}))} s_{2,P}}_{\triangleq y_{2,P}}. \tag{7.2}$$

The output is thus a sum of the two predistorted branches. In each branch $k = 1, 2$, the phase of the input is changed to counteract the effects of the nonequal gain factors and the PA nonlinearities. Each branch is predistorted separately and sent to the outphasing PA.

We will start by describing the effects of the predistorter on the output. The phase difference between the two paths after the predistorters is described by

$$\begin{aligned}\Delta_\psi(s_{1,P}, s_{2,P}) &= \arg(s_{1,P}) - \arg(s_{2,P}) \\ &= [\arg(s_1) + h_1(\Delta_\psi)] - [\arg(s_2) + h_2(\Delta_\psi)] \\ &= \Delta_\psi + h_1(\Delta_\psi) - h_2(\Delta_\psi) \triangleq \tilde{h}(\Delta_\psi),\end{aligned} \tag{7.3}$$

and the phase difference between the two paths at the (predistorted) outputs by

$$\begin{aligned}
\Delta_\psi(y_{1,P}, y_{2,P}) &= \arg(y_{1,P}) - \arg(y_{2,P}) \\
&= \Big[\arg(s_{1,P}) + f_1(\Delta_\psi(s_{1,P}, s_{2,P}))\Big] - \Big[\arg(s_{2,P}) + f_2(\Delta_\psi(s_{1,P}, s_{2,P}))\Big] \\
&= \Big[\arg(s_1) + h_1(\Delta_\psi) + f_1(\tilde{h}(\Delta_\psi))\Big] - \Big[\arg(s_2) + h_2(\Delta_\psi) + f_2(\tilde{h}(\Delta_\psi))\Big] \\
&= \Delta_\psi + h_1(\Delta_\psi) - h_2(\Delta_\psi) + f_1(\tilde{h}(\Delta_\psi)) - f_2(\tilde{h}(\Delta_\psi)) \\
&= \tilde{h}(\Delta_\psi) + f_1(\tilde{h}(\Delta_\psi)) - f_2(\tilde{h}(\Delta_\psi)) \\
&\triangleq \tilde{f}(\tilde{h}(\Delta_\psi)).
\end{aligned} \tag{7.4}$$

These phase differences correspond to the amplitude of the signal, since it is known that $|s| = \cos(\Delta_\psi/2)$, cf. Figure 5.12. The absolute phase change in each branch is given by

$$\arg(y_{k,P}) = \arg(s_k) + h_k(\Delta_\psi) + f_k(\Delta_\psi(s_{1,P}, s_{2,P})) \tag{7.5}$$

for $k = 1, 2$. We now have a model structure describing how the phases of each outphasing signal, and thus the amplitude and phase of the output, depend on the characteristics g_1, g_2, f_1 and f_2 of the PA and the predistorter functions h_1 and h_2.

7.2 The Ideal DPD

As mentioned above, the PA output signal $y(t)$ is a distorted version of the input signal. An ideal DPD should compensate for this distortion and result in a normalized output signal $y_P(t) = y_{1,P}(t) + y_{2,P}(t)$ that is equal to the input signal $s(t) = 0.5s_1(t) + 0.5s_2(t)$. In the ideal case when $g_1 = g_2 = g_0 = 0.5$, this is obtained when $y_1(t) = 0.5s_1(t)$ and $y_2(t) = 0.5s_2(t)$. However, this is not possible to achieve when $g_k \neq 0.5, k = 1, 2$. In this case, the ideal values for $y_{1,P}(t)$ and $y_{2,P}(t)$ are instead $\tilde{s}_1(t)$ and $\tilde{s}_2(t)$, as described in (6.3). These signals define an alternative decomposition of $s(t)$ such that the gain mismatch is accounted for.

Assume now that an ideal DPD (7.1) is used together with the PA (6.9). In this case, the equalities

$$y_{1,P}(t) = \tilde{s}_1(t) \tag{7.6}$$

and

$$y_{2,P}(t) = \tilde{s}_2(t) \tag{7.7}$$

hold, which results in

$$y_P(t) = y_{1,P}(t) + y_{2,P}(t) = \tilde{s}_1(t) + \tilde{s}_2(t) = s(t).$$

That is, when the ideal DPD is applied to the PA, the original input will be retrieved. This assumes that the model perfectly describes the PA. Some more conclusions can be drawn about the ideal DPD by looking at the amplitudes and the phases of the input and the output. In order not to distort the amplitude at

the output, the phase difference between $y_{1,P}(t)$ and $y_{2,P}(t)$ must be equal to the one between $\tilde{s}_1(t)$ and $\tilde{s}_2(t)$, i.e.,

$$\begin{aligned}\Delta_\psi(y_{1,P}, y_{2,P}) &= \Delta_\psi(s_{1,P}, s_{2,P}) = \arg(\tilde{s}_1) - \arg(\tilde{s}_2) = \\ &= \Big[\arg(s_1) + \xi_1(\Delta_\psi)\Big] - \Big[\arg(s_2) + \xi_2(\Delta_\psi)\Big] \\ &= \Delta_\psi + \xi_1(\Delta_\psi) - \xi_2(\Delta_\psi) \triangleq \tilde{\xi}(\Delta_\psi). \end{aligned} \tag{7.8}$$

Hence, inserting (7.8) into (7.4) gives

$$\tilde{f}(\tilde{h}(\Delta_\psi)) = \tilde{\xi}(\Delta_\psi) \quad \Leftrightarrow \quad \tilde{h}(\Delta_\psi) = \tilde{f}^{-1}(\tilde{\xi}(\Delta_\psi)), \tag{7.9}$$

assuming that $\tilde{f}$ is invertible. Furthermore, for (7.6) and (7.7) to hold, that is, $y_{1,P} = \tilde{s}_1$ and $y_{2,P} = \tilde{s}_2$, we require that the phases of the two signals are equal,

$$\arg(y_{k,P}) = \arg(\tilde{s}_k), \quad k = 1, 2. \tag{7.10}$$

Now, we have a description of how the predistorter will affect the output as well as of how the gain factors g_1 and g_2 changes the desired outphasing output signals. The phase condition (7.10) combined with (7.3), (7.5) as well as (6.6) or (6.7), respectively (for each branch), gives

$$\arg(s_k) + h_k(\Delta_\psi) + f_k(\tilde{h}(\Delta_\psi)) = \arg(s_k) + \xi_k(\Delta_\psi), \; k = 1, 2.$$

That is, the predistorter functions h_k is the only unknown in each branch and can be solved for. This results in

$$\begin{aligned} h_k(\Delta_\psi) &= -f_k(\tilde{h}(\Delta_\psi)) + \xi_k(\Delta_\psi) \\ &= -f_k(\tilde{f}^{-1}(\tilde{\xi}(\Delta_\psi))) + \xi_k(\Delta_\psi) \end{aligned} \tag{7.11}$$

for $k = 1, 2$. Here, (7.9) has been used in the last equality.

Hence, using the predistorters (7.11) in (7.1), the output $y(t)$ will be an amplified replica of the input signal $s(t)$, despite the gain mismatch and nonlinear behavior of the amplifiers.

7.3 Nonconvex DPD Estimator

A first approach to identify the predistorter is to notice that the goal is to minimize the difference between the normalized input and the normalized predistorted output. This can be written down in a straightforward way as solving the minimization criterion

$$\hat{\theta}_{\text{DPD}} = \underset{\theta_{\text{DPD}}}{\operatorname{argmin}} \sum_{t=1}^{N} \left| s(t) - \hat{y}_{\text{P}}(t, \theta_{\text{DPD}}) \right|^2, \tag{7.12}$$

$$\hat{y}_{\text{P}}(t, \theta_{\text{DPD}}) = \hat{g}_1 e^{j\,p(\hat{\eta}_1, \Delta_\psi(s_{1,P}, s_{2,P}))} s_{1,P}(t) + \hat{g}_2 e^{j\,p(\hat{\eta}_2, \Delta_\psi(s_{1,P}, s_{2,P}))} s_{2,P}(t), \tag{7.13}$$

where

$$s_{k,P}(t) = e^{j\,p(\eta_{k,\mathrm{DPD}},\Delta_\psi(s_1,s_2))} s_k(t), \quad k = 1, 2, \tag{7.14}$$

and $\theta_{\mathrm{DPD}} = [\eta_{1,\mathrm{DPD}}^T \quad \eta_{2,\mathrm{DPD}}^T]^T \in \mathbb{R}^{2n+2}$. The signal $\hat{y}_P(t)$ is the output from a PA model, using a predistorted input, as in Figure 7.1, where the amplifiers are replaced by the obtained models thereof. The DPD is thus identified based on a model of the forward system, according to METHOD B in Section 4.1. The forward model was approximated by polynomials, $\hat{f}_k(\Delta_\psi) = p(\eta_k, \Delta_\psi)$, according to (6.11), and this is used in (7.12)-(7.13) to explicitly point out the dependence on the model parameters. When identifying the DPD model, the model structure was assumed to be the same as for the PA model, see (6.11), motivated by the Stone-Weierstrass theorem (Theorem 7.26 in Rudin [1976]), so that

$$\hat{h}_k(\Delta_\psi) = p(\eta_{k,\mathrm{DPD}}, \Delta_\psi) = \sum_{i=0}^{n_h} \eta_{k,i,\mathrm{DPD}} \Delta_\psi^i, \quad k = 1, 2, \tag{7.15}$$

where

$$\eta_{k,\mathrm{DPD}} = \begin{pmatrix} \eta_{k,0,\mathrm{DPD}} & \eta_{k,1,\mathrm{DPD}} \cdots & \eta_{k,n,\mathrm{DPD}} \end{pmatrix}^T.$$

The resulting estimated parameter vector $\hat{\theta}_{\mathrm{DPD}}$ contains the DPD model parameters.

This formulation leads to a nonconvex optimization problem and is thus at a risk of obtaining a suboptimal solution if the optimization algorithm finds a local minimum. To restart the algorithm at different initial points is a possible way to reduce the risk of getting stuck in a local minimum instead of the global minimum, but this solution would not be useful in an online implementation, see also Section 6.5 for a discussion on convex and nonconvex optimization.

7.4 Analytical DPD Estimator

The ideal DPD outlined in Section 7.2 requires knowledge of the PA model, and once the PA characteristics g_1, g_2, f_1 and f_2 are known (or estimated), the predistorter functions can be determined. The first step to construct a DPD is thus to obtain a model of the PA, as described in Chapter 6. This method follows METHOD A in Section 4.1, where a model of the system itself is used to analytically produce an inverse.

The parameter estimates $\hat{\eta}_k$ define function estimates (6.20)

$$\hat{f}_k(z) = p(\hat{\eta}_k, z), \quad k = 1, 2,$$

from which an estimate

$$\hat{\tilde{f}}(z) = z + \hat{f}_1(z) - \hat{f}_2(z) \tag{7.16}$$

of the function $\tilde{f}$ from (7.4) can be computed. Provided that this function can be inverted numerically, estimates $\hat{h}_k$ of the ideal phase correction functions can be computed as in (7.11), i.e.,

$$\hat{h}_k(\Delta_\psi) = -\hat{f}_k(\hat{\tilde{f}}^{-1}(\tilde{\xi}(\Delta_\psi))) + \xi_k(\Delta_\psi) \tag{7.17}$$

for $k = 1, 2$, where Δ_ψ is given by (6.2) and $\tilde{\xi}$, ξ_1 and ξ_2 by (7.8), (6.6) and (6.7), respectively.

Hence, the complete DPD estimator consists of the selection of gain factors g_1 and g_2, see Sections 6.1 and 6.3. Also, the two least-squares estimators given by (6.18), a numerical function inversion in order to obtain $\hat{\tilde{f}}^{-1}$ and the expressions for the phase correction functions in (7.17) make part of the complete DPD estimator. The DPD estimation can either be done at each point in time, or (as has been done here) by evaluating the function for the range of possible Δ_ψ and saving this nonparametric, piecewise constant function.

The DPD estimator will result in two functions $\hat{h}_1$ and $\hat{h}_2$ which take Δ_ψ as argument, and by using these as in (7.1), the predistorted input signals $s_{1,P}(t)$ and $s_{2,P}(t)$ can be calculated for arbitrary data. Measurement results for a validation data set, not used during the modeling, will be presented in Chapter 8.

The algorithm thus consists of two main parts, A – Estimation of PA model and B – Calculation of DPD functions. Part A consists of three subparts where the first, A.I, produces candidates for the gain factors g_1 and g_2 by either using the DR by gridding possible values. A.II produces LS estimates of the nonlinear functions $\hat{f}_1$ and $\hat{f}_2$ for each pair of g_1 and g_2 and in A.III, the best performing model is chosen among all the candidates. In Part B, the DPD functions $\hat{h}_1$ and $\hat{h}_2$ are calculated. The different steps are described in more detail in Algorithm 1.

7.5 Inverse Least-Squares DPD Estimator

In the deduction of the predistorter described above, the ideal DPD was deduced using analytical relationships between the input and the desired output, following the basic METHOD A described in Section 4.1, page 34. By instead choosing METHOD C, we want to estimate the inverse directly. This means that the system input $s(t)$ (or rather $s_1(t)$ and $s_2(t)$) will be considered as output to the identification, and $y(t)$ (or $y_1(t)$ and $y_2(t)$) as input.

Since g_1 and g_2 can be found rather easily (through the dynamic range or gridding), these can still be assumed to be known, so the decomposition of $y(t)$ into $y_1(t)$ and $y_2(t)$ can be performed using (6.3). In each branch $k = 1, 2$ we thus have

$$\arg(s_k) = \arg(y_k) - h_k\left(\Delta_\psi(y_1, y_2)\right). \tag{7.18}$$

The left hand side is the input, which is known. The first term on the right hand side represents what we have measured, using the decomposition (6.3). The second term represents how the outphasing outputs should be modified to match the input, a *postdistorter*. The only unknown is thus the predistorter functions h_1 and h_2 in the two branches. By approximating these as polynomials,

$$\hat{h}_k \approx p(\zeta_k, \Delta_\psi(y_1, y_2)) = \sum_{i=0}^{n_h} \zeta_{k,i} \Delta_\psi^i(y_1, y_2), \quad k = 1, 2, \tag{7.19}$$

where

$$\zeta_k = \begin{pmatrix} \zeta_{k,0} & \zeta_{k,1} \cdots & \zeta_{k,n} \end{pmatrix}^T.$$

Algorithm 1 LS modeling and analytical DPD method

Require: model order n, method for choice of g_1 and g_2, precision of PA model (p_M) and inverse (p_I), estimation data.

{**A – Estimation of PA model**}

1: Normalize the output $y(t) = \frac{y(t)}{max(|y(t)|)}$

2: Calculate Δ_ψ $\forall t$ according to (6.2).

{**A.I – Estimation of gain factor candidates g_1 and g_2**}

3: **if** Use Dynamic Range to determine g_1 and g_2 **then**

4: Calculate c_{DR} using (5.13), and Δ_g using (6.17).

5: Calculate possible choices of g_1, g_2 according to (6.15).

6: **else** {g_1 and g_2 over a range of values}

7: Grid $g_1 \in [g_{\min}, g_{\max}]$ with precision p_M and let $g_2 = 1 - g_1$.

8: **end if**

{**A.II – Estimation of nonlinearity function candidates $\hat{f}_1$ and $\hat{f}_2$**}

9: **for** all pairs of g_1, g_2 **do**

10: Create $\tilde{s}_k = g_k e^{j \arg(\tilde{s}_k)}$ and $y_k = g_k e^{j \arg(y_k)}$, $k = 1, 2$ using (6.4) to (6.7).

11: Find η_k using (6.18) and calculate $\hat{f}_k$, $k = 1, 2$ using (6.20).

12: Simulate the output $\hat{y}_{g_1,g_2}(t) = g_1 e^{j \hat{f}_1(\Delta_\psi)} s_1(t) + g_2 e^{j \hat{f}_2(\Delta_\psi)} s_2(t)$.

13: Calculate error $V_g(g_1, g_2) = \sum_t |y(t) - \hat{y}_{g_1,g_2}(t)|^2$.

14: **end for**

{**A.III – Choose best forward model, $\hat{g}_1, \hat{g}_2, \hat{f}_1$ and $\hat{f}_2$**}

15: Select $\hat{g}_1 = \arg\min_{g_1} V_g(g_1, 1 - g_1)$, $\hat{g}_2 = 1 - \hat{g}_1$ and the corresponding $\hat{f}_1$ and $\hat{f}_2$.

{**B – Calculation of DPD functions $\hat{h}_1$ and $\hat{h}_2$**} {Create a Look Up Table (LUT) for different values of Δ_ψ by creating an intermediate signal s}

16: Grid $\Delta_\psi \in [0, \pi]$ with precision p_I.

17: **for** each value of Δ_ψ **do**

18: Create $s = \cos(\Delta_\psi/2)$ according to (5.10) assuming $\alpha = 0$ and $r_{\max} = 1$ ($\varphi = \Delta_\psi/2$).

19: Create s_1 and s_2 according to (5.11) and $\tilde{s}_1$ and $\tilde{s}_2$ using (6.3) to (6.5).

20: Find $\tilde{\xi}$ using (7.8), (6.6) and (6.7).

21: Calculate $\hat{\tilde{f}}(\tilde{\xi})$ using (7.16).

22: **end for**

23: Invert $\hat{\tilde{f}}(\tilde{\xi})$ numerically to get $\hat{\tilde{f}}^{-1}$. This can e.g. be done by calculating $\tilde{f}(\tilde{\xi})$ for a number of values of $\tilde{\xi} \in [0, \pi]$, grid $\hat{\tilde{f}}(\tilde{\xi})$ and match with the $\tilde{\xi}$ that gives the closest value.

24: **for** each value of Δ_ψ in line 16 **do**

25: Find estimate $\hat{h}_k(\Delta_\psi)$ according to (7.17).

26: **end for**

as was done for the PA model, the parameters corresponding to the h_k-functions can be found.

The estimates $\hat{\zeta}_{k,i}$ of the model parameters have been computed by minimizing a quadratic cost function, i.e.,

$$\hat{\zeta}_k = \arg\min_{\zeta_k} V_k^h(\zeta_k), \quad k = 1, 2, \tag{7.20}$$

where

$$V_k^h(\zeta_k) = \sum_{t=1}^{N} \Big(\arg\left(y_k(t)\right) - \arg\left(s_k(t)\right) - p\left(\zeta_k, \Delta_\psi(y_1(t), y_2(t))\right)\Big)^2. \tag{7.21}$$

The parameter estimates $\hat{\zeta}_k$ define inverse function estimates

$$\hat{h}_k(z) = p(\hat{\zeta}_k, z), \quad k = 1, 2, \tag{7.22}$$

that can be used as a DPD. As discussed in Chapter 3, this method assumes commutativity of the two systems (system and inverse), so that the inverse which was estimated at the output of the power amplifier, a *postdistorter*, can also be used at the input as a *predistorter*. The method is summarized in Algorithm 2.

Algorithm 2 Inverse LS DPD method

Require: model order n_h, method for choice of g_1 and g_2, precision of gain factors (p_M), estimation data.

Normalize the output $y(t) = \frac{y(t)}{max(|y(t)|)}$

{**I – Estimation of gain factor candidates g_1 and g_2**}
if Use Dynamic Range to determine g_1 and g_2 **then**
  Calculate c_{DR} using (5.13), and Δ_g using (6.17).
  Calculate possible choices of g_1, g_2 according to (6.15).
else {g_1 and g_2 over a range of values}
  Grid $g_1 \in [g_{\min}, g_{\max}]$ with precision p_M and let $g_2 = 1 - g_1$.
end if

{**II – Estimation of nonlinearity function candidates $\hat{h}_1$ and $\hat{h}_2$**}
for all pairs of g_1, g_2 **do**
  Create $y_k = g_k e^{j \arg(y_k)}$ using (6.4) to (6.7) and s_k using (5.11).
  Calculate $\Delta_\psi(y_1, y_2)\ \forall t$ according to (6.2).
  Find $\hat{\zeta}_k$ using (7.20) and calculate $\hat{h}_k, k = 1, 2$ using (7.22).
  Simulate the input $\hat{s}_{g_1,g_2}(t) = e^{j\,\hat{h}_1(\Delta_\psi(y_1,y_2))} y_1(t) + e^{j\,\hat{h}_2(\Delta_\psi(y_1,y_2))} y_2(t)$
  Calculate error $V_g(g_1, g_2) = \sum_t |s(t) - \hat{s}_{g_1,g_2}(t)|^2$
end for

{**III – Choose best inverse model, $\hat{h}_1$ and $\hat{h}_2$**}
Select $\hat{g}_1 = \arg\min_{g_1} V_g(g_1, 1 - g_1)$, $\hat{g}_2 = 1 - \hat{g}_1$ and the corresponding $\hat{h}_1$ and $\hat{h}_2$.

Table 7.1: DPD Model Validation

Method	$\lvert s - \hat{y}_p \rvert_2^2$
Analytical	0.0532
LS	1.008
Gain factors only	65.5

7.6 Simulated Evaluation of Analytical and LS Predistorter

The goal here is to evaluate the performance of the predistorter methods in simulations, and determine how well the different methods achieve an inversion. One way is to look at the AM-AM modulation to assess how much the amplitude of the predistorted output is distorted. For an outphasing PA, this is connected to the phase difference $\Delta_\psi(y_{1,P}, y_{2,P})$ of the outphasing outputs $y_{1,P}$ and $y_{2,P}$.

The predistorter methods in Sections 7.4 and 7.5 are evaluated using a model of the amplifier as "the truth". The model is presented in Chapter 6, where the gain factors were estimated using the DR and the nonlinearities using the LS approach, see Section 6.3 and the model validation in Section 6.4 and Figure 6.5. The same validation data have been used in order to evaluate the different predistorter methods. Evaluation on a real PA will be presented in Chapter 8.

Test 1 – Inversion Evaluation

We will start by looking at the AM-AM modulation to determine how much the amplitude of the predistorted output is changed. The deviation from the ideal phase difference at the output (i.e., the output amplitude) with and without predistortion is presented in Figure 7.2. Both the analytical method and the LS method clearly reduce the phase shift introduced by the PA. Figure 7.3 shows the estimated deviation from the ideal phase for each signal branch with and without predistortion, with rather similar performance for the two DPD methods.

The values of the cost function (7.12) are presented in Table 7.1 for the two methods. The result using only the estimation of the gain factors and the alternative decomposition (using the knowledge of the nonequal gain factors) is also presented. It is clear that incorporating the nonlinearities improves the performance. For cases when the gain factors differ more from the ideal $g_1 = g_2 = 0.5$ than in this case ($g_1 = 0.4986$ and $g_2 = 0.5014$), the alternative decomposition (6.3) will have a larger improvement on the modeling than in this case, when the difference is small.

For the LS method, the fit is almost perfect in the middle range, which is to be expected since a polynomial is used (see discussion in Section 6.4, page 79). Also the number of measurements is unevenly spread out over Δ_ψ with most data in the middle, only 0.9% of the estimation data have $\Delta_\psi < 0.8$ or $\Delta_\psi > 3$. For the

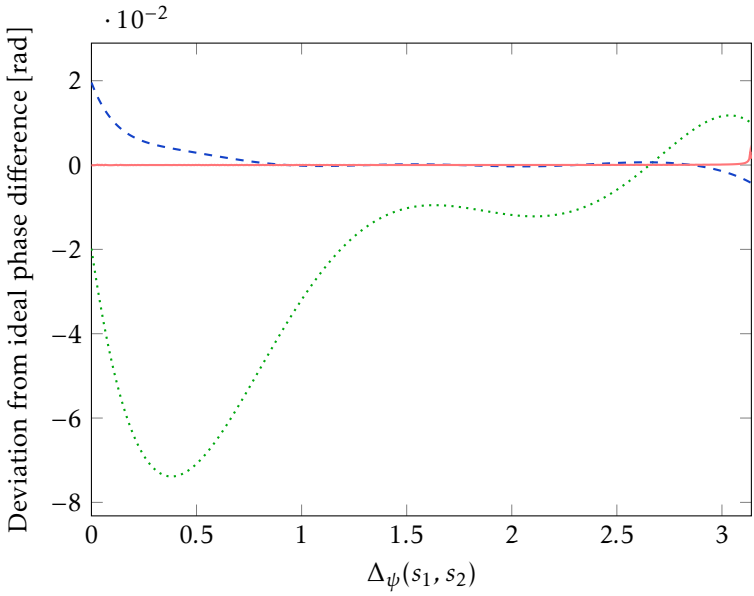


Figure 7.2: *Simulated predistorter evaluation for a model with polynomial degree $n = 5$ using the WCDMA input signal (see Chapter 8). The signals are generated using the DPD functions and the PA model. For an ideal PA, there is no amplitude distortion, that is, the phase difference of the outphasing signals is the same at the output and the input. The deviation from this ideal phase difference for the (modeled, not predistorted) output signal $\hat{y}$ is shown in dotted green and the predistorted output signals $\hat{y}_P$ in pink and blue. The pink line shows the result using the analytical inversion as described in Sections 7.2 and 7.4 and the dashed blue line shows the result of the LS approach in Section 7.5, with predistorter degree $n_h = 5$ in (7.19). The two methods both perform very well in a large interval.*

analytical solution, one can see an inversion error close to $\Delta_\psi = \pi$. This is a consequence of the nonequal gain factors, $\Delta_\psi = \pi$ should represent a complete opposition of the two outphasing signals such that the output amplitude is zero. If $g_1 \neq g_2$ however, this is not possible and no phase combination of the two outphasing signals will lead to a zero-amplitude output. A power amplifier with a large dynamic range (DR, difference between the gain factors g_1 and g_2) will have a very small distortion close to $\Delta_\psi = \pi$, whereas a PA with a small DR will show this distortion in a larger region. The errors of the two methods when compared to validation data are shown in Figure 7.4. Also in this plot, it can be seen that both methods reduce the power amplifier distortion, and that the analytical inversion performs slightly better.

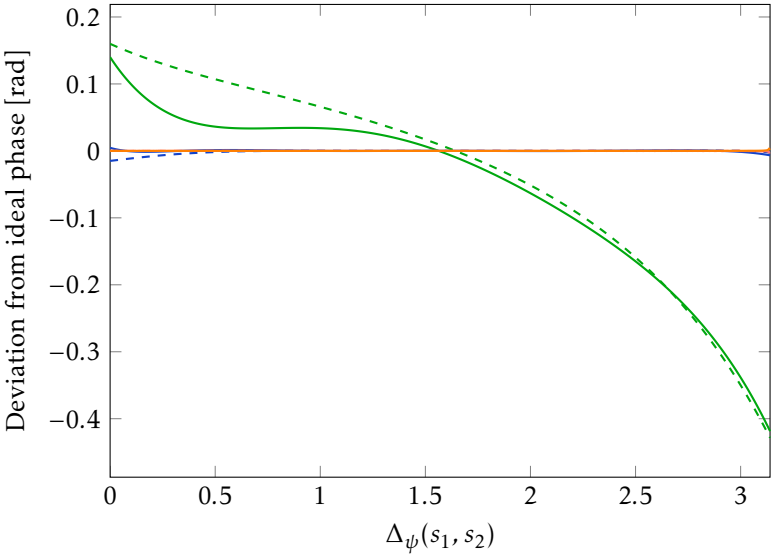


Figure 7.3: *Simulated predistorter evaluation for a model with polynomial degree $n = 5$ using the WCDMA input signal. The signals are generated using the DPD functions and the PA model. The deviation from the ideal phase for the (modeled, not predistorted) output outphasing signals $\hat{y}_1$ and $\hat{y}_2$ are shown in green and the predistorted output signals $\hat{y}_{1,P}$ and $\hat{y}_{2,P}$ in pink and blue. The pink lines show the results using the analytical inversion as described in Sections 7.2 and 7.4 and the blue lines show the result of the LS approach in Section 7.5. Branch one is plotted in solid lines and branch two in dashed lines.*

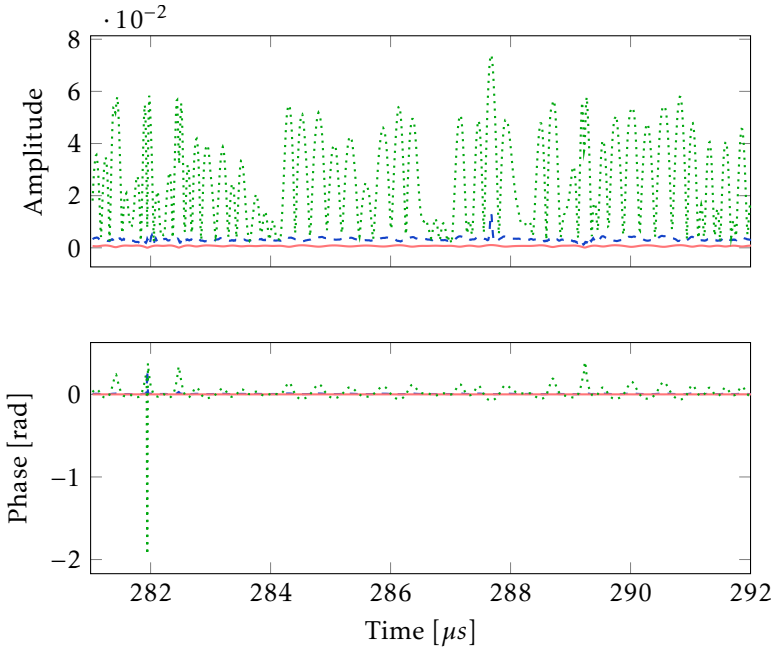


Figure 7.4: *The upper plot shows the amplitude error, $|s - \hat{y}_p|$, and the lower plot shows the phase error, $\arg(s) - \arg(\hat{y}_p)$, for the two* DPD *methods. The analytical method is in pink and the* LS *in blue. As a comparison, the errors for the original, unpredistorted signal $y(t)$ are also plotted in green.*

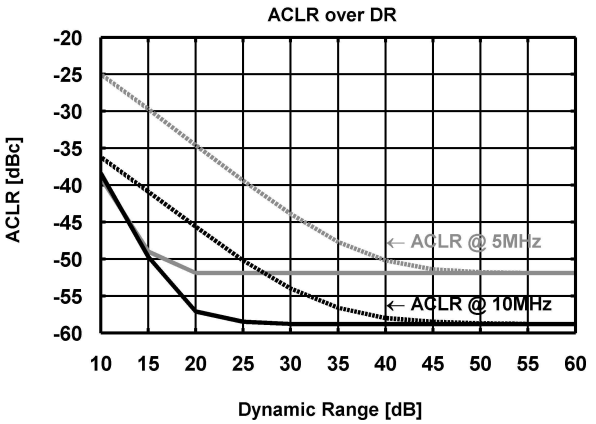


Figure 7.5: *Simulated ACLR at 5 MHz and 10 MHz offset with DPD (solid line) and without (dashed line) for the WCDMA signal.*

Test 2 – Impact of ACLR on Predistorter Performance

As previously explained, the result of a limited dynamic range is that all amplitude and phase errors occurring outside the DR cannot be corrected. The signal clipping in an outphasing PA occurs at small amplitudes, while in a conventional linear PA, the peak amplitudes are clipped. Thus, the DR in an outphasing PA limits the spectral performance when amplifying modulated signals. To investigate the performance limits of the predistorter, simulations have been done using two amplifiers with a given DR (no phase distortion), with and without DPD. In Figure 7.5, the ACLR over DR at 5 MHz and 10 MHz for the WCDMA signal are plotted with and without DPD. Here, the phase error between the outphasing signals is assumed to be zero. For a PA with a DR of 25 dB the differences in ACLR between the nonpredistorted and predistorted outputs are 8-13 dB. When the DR is 25 dB the optimal theoretical ACLR is achieved after DPD. For a PA with 45 dB of DR, the difference between when a DPD is used or not is negligible.

Summary

In this simulated evaluation, both DPD methods achieve an improvement, compared to the original power amplifier output. The analytical inversion leads to slightly better results at the cost of a higher computational complexity. The look-up table for the analytical DPD has $2*3142$ elements (with precision $p_I = 0.001$ in Algorithm 1), and the polynomials contain $2*6$ coefficients ($n_h = 5$ in (7.19)). Using a higher polynomial degree could lead to improved results for the LS method, and a smaller LUT might lead to a small degradation of the analytical method. As implementation issues are out of scope for this thesis, the methods are not optimized for implementation and therefore these considerations have not been further pursued.

7.7 Recursive Least-Squares and Least Mean Squares

Here, a few aspects of a possible future implementation of the DPD methods are presented. In addition to the guaranteed convergence, least-squares formulations also have the advantage that there are many efficient numerical methods for solving this type of problems. They can be solved recursively by, for example, the *recursive least-squares* (RLS) method [Björck, 1996] making them suitable for an online implementation. An even less complex parameter estimation algorithm is the *least mean square* (LMS) method, which can make use of the linear regression structure of the optimization problem, developed here in (6.11) and (6.19). LMS has been used for RF PA linearization in Montoro et al. [2007] and implemented in *field programmable gate array* (FPGA) technology, as shown in Gilabert et al. [2009].

With a recursive implementation of the algorithm, it is even more important that the algorithm can be proved to converge to good values, as no monitoring of the performance should be necessary in order for the method to be useful in practice. This also means that a nonconvex solution as in (6.12)-(6.13) is not suitable for online implementation since it cannot guarantee convergence to good enough minima. In an offline application, the possibility to restart the optimization could be added but, together with the lack of a bound on the number of iterations, this does not seem like a good solution for an online version. Using well explored methods like RLS or LMS would result in a low-complexity implementation, and though it is hard to judge the exact complexity of the iterative implementation that would be needed for the online version of nonconvex solution, it is clear that it would be very hard to find a simpler one than for the low-complexity LMS version of the convex method.

Since circuitry will behave differently depending on the settings under which it operates, it is important to be robust to such conditions. This is covered in the concept of *process, voltage and temperature variations* (PVT variations). One way to handle the PVT variations and changes in the setting, such as aging, would be to use a method with a forgetting factor, reducing the influence of older measurements [Ljung, 1999]. The RLS and LMS solutions assume the changes in the operating conditions to be slow.

8

Predistortion Measurement Results

The models presented in Chapter 6 and the predistorters in Chapter 7 are based on measured data from a power amplifier. In Chapter 7, the methods' ability to invert the nonlinearities was investigated, using a forward model as a "true" system. In this chapter, the methods will be evaluated on real measurements. The predistorters are applied to a new data set, *validation data*, that is not the same as the signal used for estimation. To start off, a short introduction to the signal types used and the measurement setup will be presented.

8.1 Signals Used for Evaluation

The predistortion methods have been evaluated for the different signal types EDGE, WCDMA and LTE. Mobile communication technologies are often divided into generations, and the new devices of today are the fourth generation, 4G. The first generation, 1G, was the first analog mobile radio systems of the 1980s. 2G was the first digital mobile systems and 3G the first mobile systems handling broadband data.

Enhanced data rates for GSM evolution (EDGE) is a mobile phone technology with higher bit rates than *general packet radio service* (GPRS) [Ahlin et al., 2006], and has been called 2.75G since it did not quite reach the 3G standards. The carrier frequency used is 2 GHz, and the bandwidth is 200 kHz. *Wideband code division multiple access* (WCDMA) is a third generation (3G) mobile phone technology, and is one of the 3G mobile communications standards [Frenzel, 2003]. The carrier frequency used is 2 GHz, and the bandwidth is 5 MHz. The bandwidth of the *long term evolution* (LTE) signal is variable, and can be adjusted between 1 and 20 MHz. It is sometimes called 4G or 3.9G since it does not completely satisfy the 4G requirements Dahlman et al. [2011].

The WCDMA and LTE have large peak-to-minimum power ratio, i.e., the PA

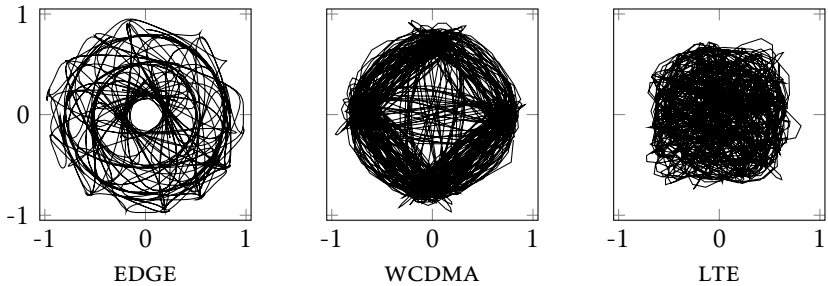


Figure 8.1: *IQ plots (imaginary part, Q, vs real part, I) of signal realizations of the EDGE, WCDMA and LTE standards in the complex plane. The sampling frequency in the modeling data sets is four times higher than that shown here.*

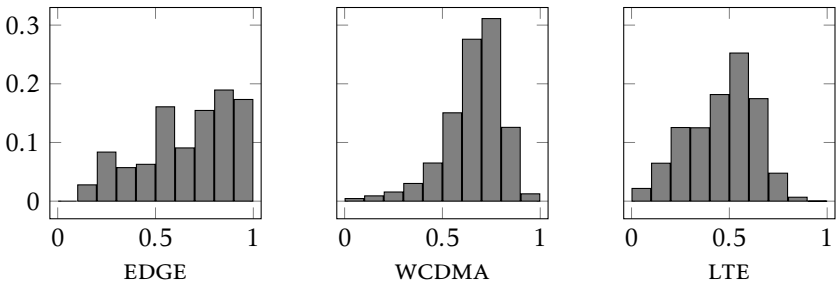


Figure 8.2: *Histograms of the distribution of the input amplitude of signal realizations of the EDGE, WCDMA and LTE standards. This difference in input distribution effects the peak-to-average power ratio, and it also implicitly determines the weighting of the fit of the polynomials, see also the discussion on polynomial fitting on page 79.*

output signals include the minimum and maximum amplitudes (the full dynamic range). For these signals, the DR of the PA will effect the output signal, by clipping the smallest amplitudes. For EDGE, the signal amplitude is never close enough to zero to be effected by the PA DR, and no clipping will occur. Realizations of each signal type (EDGE, WCDMA and LTE) are shown in Figure 8.1 as IQ-plots. Histograms of the distribution of the input amplitude are shown in Figure 8.2. The distribution also implicitly determines the weighting of the fit of the polynomials, see also the discussion on polynomial fitting on page 79. One characteristic of a signal is the *peak-to-average power ratio* (PAPR). A signal with a high PAPR sets high standards on the linearity of the PA, since a large range of input signal amplitudes has to be amplified.

The signals used are created as random signals with predefined characteristics.

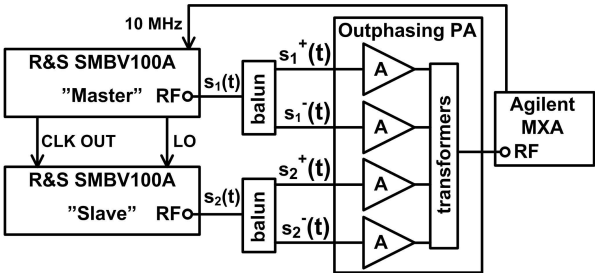


Figure 8.3: Measurement setup for IQ-data with two Master-Slave-configured SMBV signal generators [Rohde & Schwarz].

8.2 Measurement Setup

The measurements that will be discussed in Section 8.3 have been performed using an SMU200A signal generator with two phase-coherent RF outputs and an arbitrary waveform generator where the input signals ($s_1(t)$ and $s_2(t)$) and the predistorted input signals ($\tilde{s_1}(t)$ and $\tilde{s_2}(t)$) were stored. For the measurements that will be discussed in Section 8.4, two R&S SMBV100A signal generators with phase-coherent RF outputs and arbitrary waveform generators with maximum IQ sample rate of 150 MHz have been used. Figure 8.3 shows the measurement setup.

The outphasing power amplifiers used in the measurements have been developed by Jonas Fritzin et al. and are briefly described in Appendix A and in more detail in Fritzin [2011].

Sampling

The sampling rate in Section 8.4 was 92.16 MHz in the measurements, six times the original sampling frequency of the signal. The impact of baseband filtering and limited bandwidth is investigated in Gerhard and Knöchel [2005a,b], where it was concluded that to obtain an optimal signal/distortion ratio over the entire bandwidth, a compromise between the sampling frequency and the filter characteristics has to be made. Here, we have evaluated the required bandwidth/sampling rate based on measurements with two signal generators and one combiner, no PA was used. Increasing the sampling frequency from the original 15.36 MHz to 30.72 MHz and 61.44 MHz, the ACLR is improved, see Table 8.1.

Thus, for the specific tests performed here, the ACLR at 5 and 10 MHz can be improved by 6-9 dB and 4-8 dB, respectively, when increasing the sampling rate up to four times the original sampling rate of 15.36 MHz. Further increasing the sampling frequency, up to 92.16 MHz, shows no significant change.

Table 8.1: *Measured Spectral Performance at 1.95 GHz for WCDMA and LTE Uplink Signals for Different Sampling Frequencies.*

	Measured Parameter	15.36 MHz	30.72 MHz	61.44 MHz
WCDMA	ACLR @ 5 MHz [dBc]	-44	-50	-52
	ACLR @ 10 MHz [dBc]	-48	-52	-56
LTE	ACLR @ 5 MHz [dBc]	-34	-43	-46

8.3 Evaluation of Nonconvex Method

In this section, the nonconvex approach presented in Sections 6.2 and 7.3 has been evaluated. The PA model has been obtained by minimizing the nonconvex cost function in (6.13) and the corresponding DPD by minimizing (7.12). The method involves solving two nonconvex optimization problems, and corresponds to METHOD B in Section 4.1. This method has been evaluated on the PA described in Appendix A.1 and Fritzin [2011].

The predistortion methods were evaluated on a physical chip. The measurement setup was optimized and the branch amplifiers were tuned to achieve the best performance possible. The phase offset between $s_1(t)$ and $s_2(t)$ in the baseband was adjusted to minimize phase mismatch (ideally 180 ° between the two RF inputs for nonmodulated $s_1(t)$ and $-s_2(t)$ in Figure A.2, i.e. maximum output power for a continuous signal). Since this is not a reasonable assumption in a real-life application, an additional phase error of 3 ° was added in one of the branches.

Measurements of input $s(t)$ and output $y(t)$ of length N_{id} were collected K times, and an average was taken to avoid the influence of measurement noise. This data was used to model the power amplifier. Based on this PA model, a predistorter model was produced. Polynomials with order n have been used as parameterized versions of the PA nonlinearities and of order n_h for the predistorter functions. The predistorted input signals, $s_{1,P}$ and $s_{2,P}$, were then computed (in MATLAB) for a validation input signal of length N_{val}. The predistorted outphasing input signals were sent to the PA, resulting in a predistorted output. The additional phase error was still applied during the predistorter validation.

For the computation of the model parameters, a large number of algorithms are available for solving a nonlinear optimization problem. Here, the MATLAB routine `fminsearch`, based on the Nelder-Mead simplex method, was used. The estimation and validation data sets contain N_{id} and N_{val} samples, respectively. The input and output sampling frequencies are denoted f_{s} and $f_{\text{s,out}}$, respectively. To minimize the influence of measurement noise, the signals were measured K times, and a mean was calculated. The data collection parameters are shown in Table 8.2. Since the WCDMA is a more wide-band signal than the EDGE signal, the number of samples N_{id} and N_{val} were chosen larger.

Table 8.2: *Data Collection, Nonconvex Method*

	N_{id}	N_{val}	f_{s}	$f_{\text{s,out}}$	K
EDGE	40 001	80 001	8.67 MHz	34.68 MHz	150
WCDMA	153 600	153 600	61.44 MHz	61.44 MHz	200

Table 8.3: *Measured Spectral Performance of the EDGE Signal*
(a) With no phase error and no DPD.
(b) For a 3° phase error and no DPD.
(c) When DPD is applied to (b).

Freq.	**Freq. offset**	**Spec.**	**Meas. (a)**	**Meas. (b)**	**Meas. (c)**
2 GHz	400 kHz	-54 dB	-54.4 dB	-53.5 dB	-65.9 dB
	600 kHz	-60 dB	-60.3 dB	-59.9 dB	-68.2 dB

8.3.1 Measured Performance of EDGE Signal

EDGE is a rather narrow-band signal with a *peak-to-average power ratio* (PAPR) of 3.0 dB. The spectrum of the estimation input data set is shown in Figure 8.4(d). The output of a perfectly matched PA in Figure 8.4(a) fulfills the requirements, but without any margins to the spectral mask. The spectral mask is a nonlinearity measurement that describes the amount of power that is allowed to be spread to the neighboring channels. The requirements for an EDGE signal are summarized in Table 5.1 and illustrated in Figure 8.4. As the phase error cannot be assumed to be 0° in a transceiver, a phase error of 3° was added and led to a violated spectral mask as in Figure 8.4(b).

When predistortion was applied to a validation data set, not used for estimation, the linearity improves, as seen in Figure 8.4(c). The PA model was of order $n = 5$ and the predistorter of order $n_h = 5$. The measured power at 400 and 600 kHz offsets were -65.9 and -68.2 dB, with margins of 11.9 and 8.2 dB, respectively. The average power at 2 GHz was +7 dBm with 22 % PAE and *root mean square* (RMS) EVM of 2 %. The measured performance of the amplifier for an EDGE signal is summarized in Table 8.3.

8.3.2 Measured Performance of WCDMA Signal

The PAPR of the WCDMA signal was 3.2 dB and the spectrum of the estimation data set is shown in Figure 8.5(d). Figure 8.5(a) shows the measured WCDMA spectrum at 2 GHz, with minimized phase mismatch and no predistortion. When the same phase error of 3° as for the EDGE signal was added to simulate reasonable phase settings, a distorted spectrum as in Figure 8.5(b) was measured. The ACLR is an integrated measure that describes the power spread to adjacent channels.

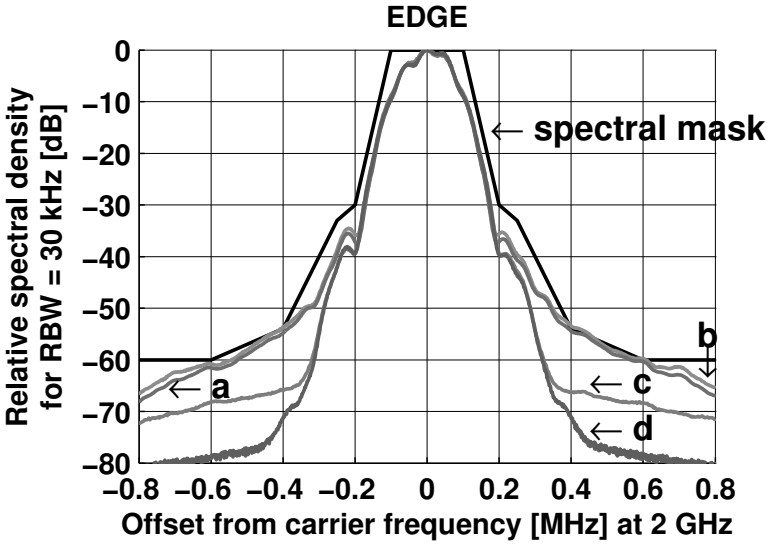


Figure 8.4: *Measured* EDGE *spectrum at 2 GHz.*
(a) Output spectrum without phase error between $s_1(t)$ and $s_2(t)$.
(b) Output spectrum with 3° phase error between $s_1(t)$ and $s_2(t)$.
(c) Output spectrum when DPD *is applied to (b).*
(d) Output spectrum of the estimation signal. The spectrum of the validation signal was similar.

Table 8.4: *Measured Spectral Performance of the WCDMA Signal*
(a) With no phase error and no DPD.
(b) For a 3° phase error and no DPD.
(c) When DPD is applied to (b).

Freq.	ACLR	Spec.	Meas. (a)	Meas. (b)	Meas. (c)
1 GHz	5 MHz	-33 dBc	-40.6 dBc	-39.4 dBc	-53.6 dBc
	10 MHz	-43 dBc	-59.8 dBc	-56.2 dBc	-60.3 dBc
2 GHz	5 MHz	-33 dBc	-43.4 dBc	-38.0 dBc	-50.2 dBc
	10 MHz	-43 dBc	-53.9 dBc	-50.9 dBc	-52.2 dBc

At 1 GHz and 2 GHz, the power amplifier fulfills the requirements, also with the additional phase error, as seen in Table 8.4.

The phase predistortion method, with $n = 5$ and $n_h = 4$, for a validation signal, improves the measured ACLR. A spectrum is shown in Figure 8.5(c). The channel power at 2 GHz was +6.3 dBm with PAE of 22 % and RMS composite EVM of 1.4 % (0.6 % after DPD). The measured performance of the amplifier for a WCDMA signal is summarized in Table 8.4.

8.3.3 Summary

The nonconvex predistortion method clearly improves the PA performance for both EDGE and WCDMA signals, even when an extra phase error is added. The measured spectral performance at 400 kHz offset and the ACLR at 5 MHz is comparable to state-of-the-art EDGE [Mehta et al., 2010] and WCDMA [Huang et al., 2010] transmitters.

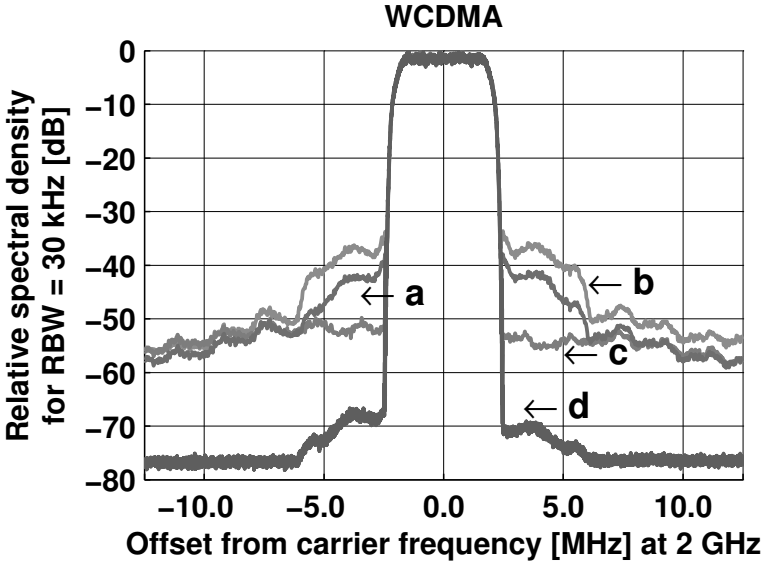


Figure 8.5: *Measured* WCDMA *spectrum at 2 GHz.*
(a) Output spectrum without phase error between $s_1(t)$ and $s_2(t)$.
(b) Output spectrum with 3° phase error between $s_1(t)$ and $s_2(t)$.
(c) Output spectrum when DPD *is applied to (b).*
(d) Output spectrum of the estimation signal. The spectrum of the validation signal was similar.

Table 8.5: *Data Collection, Least-Squares and Analytical Method*

	N_{id}	N_{val}	f_{s}	$f_{\text{s,out}}$	K
WCDMA	100 000	100 000	92.1 MHz	92.1 MHz	10
LTE	100 000	100 000	92.1 MHz	92.1 MHz	10

8.4 Evaluation of Least Squares PA and Analytical Inversion Method

In this section, the least-squares modeling of the PA, using the DR to estimate g_1 and g_2, has been applied. An analytical inversion has been used to construct the predistorter functions, as in METHOD A in Section 4.1. The PA modeling is described in Section 6.3, the DPD in Section 7.4 and the method is summarized in Algorithm 1, page 91. This method has been evaluated on the PA described in Appendix A.2 and Fritzin et al. [2011c].

The measurement setup was optimized and the branch amplifiers were tuned to achieve the best performance possible. For the measurements without predistortion, the phase offset between $s_1(t)$ and $s_2(t)$ in the baseband was adjusted to minimize phase mismatch (ideally $0\,^\circ$ between nonmodulated $s_1(t)$ and $s_2(t)$, that is, maximum output power for a continuous signal). Moreover, the IQ-delay between the signal generators was adjusted for optimal performance [Rohde & Schwarz].

Measurements of input $s(t)$ and output $y(t)$ were collected K times, and an average was taken to avoid the influence of measurement noise. This averaged data set was used to model the PA, and based on the PA model, a predistorter model was produced. Polynomials with order n have been used as parameterized versions of the PA nonlinearities and based on this model, an approximation of the ideal predistorter has been constructed. The predistorted input signals, $s_{1,P}$ and $s_{2,P}$, were then computed (in MATLAB) for a validation input signal. The predistorted outphasing input signals were sent to the PA, resulting in a predistorted output.

The estimation and validation data sets contain N_{id} and N_{val} samples, respectively. The input and output sampling frequencies are denoted f_{s} and $f_{\text{s,out}}$, respectively. The data collection parameters are shown in Table 8.5. In all following experiments, the DPD estimates $\hat{h}_k, k = 1, 2$, have been calculated for 3142 uniformly distributed points ($p_I = 0.001$ in Algorithm 1). This LUT has been used in the construction of the predistorted outphasing input signals. For each input phase difference Δ_ψ, the outphasing input signals $s_1(t)$ and $s_2(t)$ were adjusted according to the nearest neighbor principle.

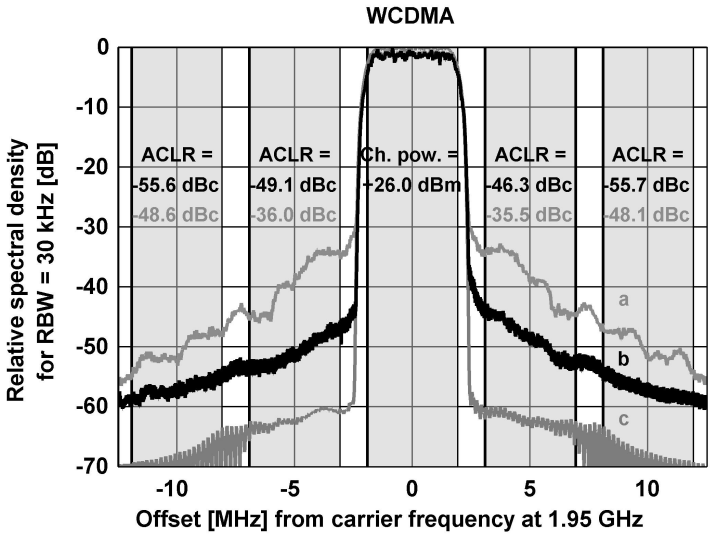


Figure 8.6: *Measured WCDMA spectrum at 1.95 GHz.*
(a) Measured WCDMA spectrum without DPD. The measured ACLR is printed in gray.
(b) When DPD is applied to (a). The measured ACLR is printed in black.
(c) Spectrum of estimation signal. Spectrum of validation signal was similar.

8.4.1 Measured Performance of WCDMA Signal

The PAPR of the WCDMA uplink signal was 3.5 dB. The spectrum of the estimation data is shown in Figure 8.6(c). For the WCDMA signal at 1.95 GHz without predistortion, the measured ACLR at 5 MHz and 10 MHz offsets were -35.5 dBc and -48.1 dBc, respectively. The spectrum is shown in Figure 8.6(a). The estimation output data $y(t)$ were used in the predistortion method to extract the model parameters, with $n = 5$. The ACLR is a measure describing the amount of leakage into adjacent channels that can be tolerated, and the standards for WCDMA are -33 dBc and -43 dBc at 5 MHz and 10 MHz offsets, respectively.

The predistorted input signals, $s_{1,P}(t)$ and $s_{2,P}(t)$, were computed for the validation input signal, resulting in an output spectrum as shown in Figure 8.6(b). The power spectral densities of the predistorted input is similar to that of the nonpredistorted input signal, and therefore not included (similarly for the LTE signal). With predistortion, the measured ACLR at 5 MHz and 10 MHz offsets were -46.3 dBc and -55.6 dBc, respectively. Thus, the measured ACLR at 5 MHz and at 10 MHz offsets were improved by 10.8 dB and 7.5 dB, respectively. The average power at 1.95 GHz was +26.0 dBm with 16.5 % PAE. It is clear that the predistortion reduces the spectral leakage.

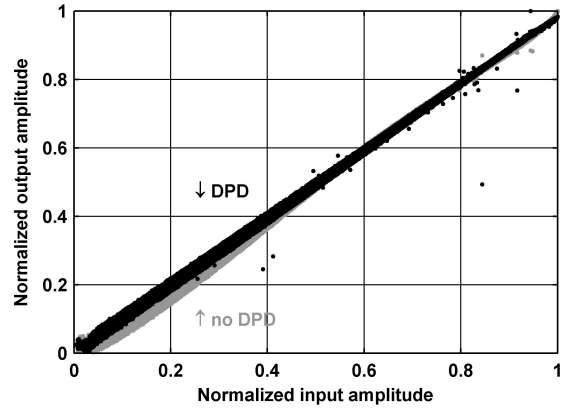
Figure 8.7 shows the measured AM-AM (output amplitude vs. input amplitude) and AM-PM (phase change vs. input amplitude) characteristics with and without DPD for the WCDMA signal. The upper figure shows the amplitude mod-

ulation, and should ideally be a straight line from lower left corner (0,0) to the upper right (1,1), such that the output amplitude equals the input amplitude for the whole range of the signal. If this is not the case, there will be amplitude distortions. Here, the improvement can be seen in normalized amplitudes smaller than 0.4. The lower plot shows the phase distortion, and the ideal is zero. It can be seen that the DPD reduces the phase distortion for normalized amplitudes in the range $0.05 \lesssim |s| \lesssim 0.95$. For amplitudes close to one, the distortion is slightly worse with a predistorter than without. This is due to the polynomial fit of the PA model, which has a best fit in the middle region where the density of data points is largest.

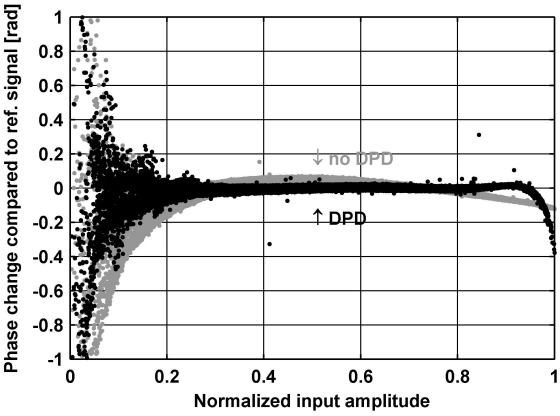
8.4.2 Measured Performance of LTE Signal

The PAPR of the LTE uplink signal was 6.2 dB and the spectrum of the estimation data sets is shown in Figure 8.8(c). For the LTE signal at 1.95 GHz without predistortion, the measured ACLR at 5 MHz offset was -34.1 dBc. The spectrum is shown in Figure 8.8(a). The estimation output data $y(t)$ were used in the predistortion method to extract the model parameters with $n = 5$. The predistorted input signals, $s_{1,P}(t)$ and $s_{2,P}(t)$, were computed for the validation input signal, resulting in an output spectrum as shown in Figure 8.8(b). With the predistorted spectrum in Figure 8.8(b), a small asymmetry can be observed, which was expected due to the asymmetrical frequency spectrum of the reference signal. With predistortion, the measured ACLR at 5 MHz offset was -43.5 dBc. Thus, the measured ACLR at 5 MHz offset was improved by 9.4 dB. The average power at 1.95 GHz was +23.3 dBm with 8.0 % PAE.

Figure 8.9 shows the measured AM-AM and AM-PM characteristics with and without DPD for the LTE signal. The amplitude mapping in the upper figure should ideally be a straight line from the lower left corner to the upper right one, and the bottom figure should be zero always. The figure shows that the amplitude and phase errors are significantly reduced for small amplitudes, with a normalized amplitude $|s| \lesssim 0.4$.



(a)



(b)

Figure 8.7: *(a) Measured* AM-AM *characteristics (output amplitude vs. input amplitude) with* DPD *(black) and without* DPD *(gray) for* WCDMA *signal. (b) Measured* AM-PM *characteristics (phase change vs. input amplitude) with* DPD *(black) and without* DPD *(gray) for* WCDMA *signal.*

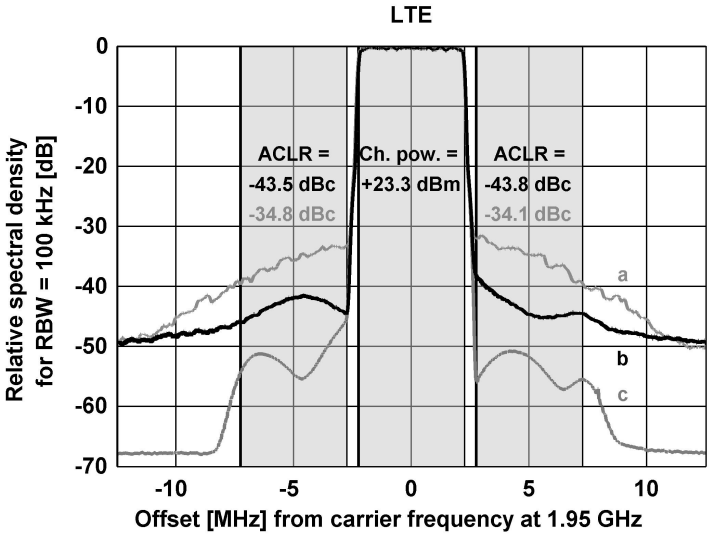
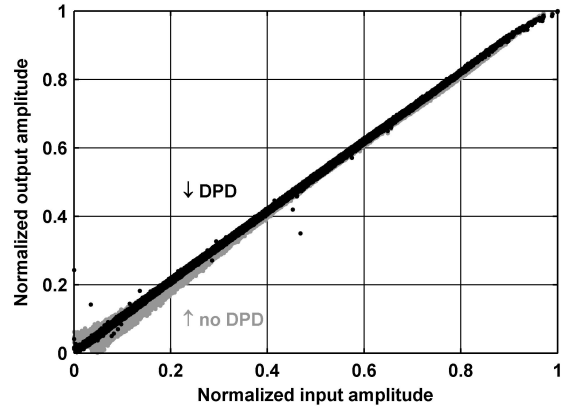
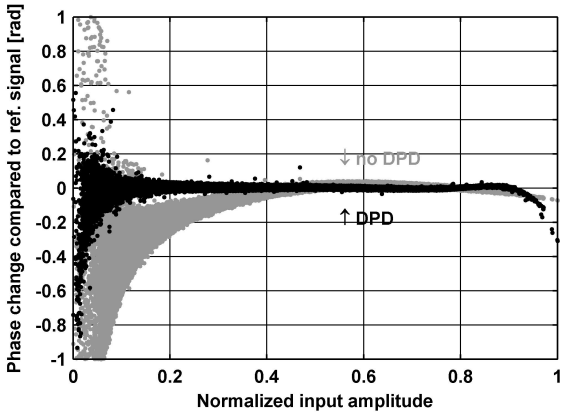


Figure 8.8: *Measured LTE spectrum at 1.95 GHz.*
(a) Measured LTE spectrum without DPD. The measured ACLR is printed in gray.
(b) When DPD is applied to (a). The measured ACLR is printed in black.
(c) Spectrum of estimation signal. Spectrum of validation signal was similar.



(a)



(b)

Figure 8.9: *(a) Measured* AM-AM *characteristics (output amplitude vs. input amplitude) with* DPD *(black) and without* DPD *(gray) for* LTE *signal. (b) Measured* AM-PM *characteristics (phase change vs. input amplitude) with* DPD *(black) and without* DPD *(gray) for* LTE *signal.*

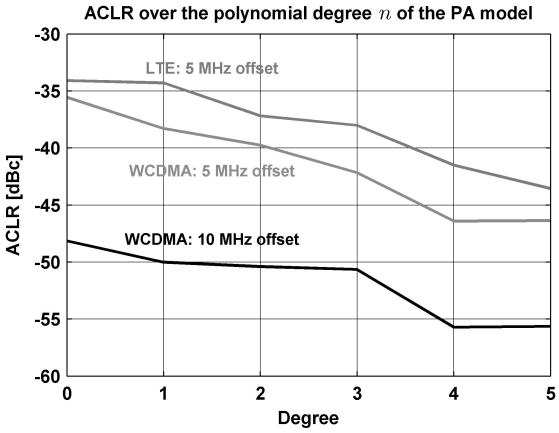


Figure 8.10: *Measured ACLR depending on the polynomial degree n of the PA model. Degree $n = 0$ represents the performance without predistortion. The nonlinear modeling and distortion clearly improves the performance by reducing the ACLR.*

8.4.3 Evaluation of Polynomial Degree

A small evaluation of the impact of polynomial degree in the PA model has been performed, and the result is presented in Figure 8.10. It is clear that the added nonlinear terms improves the ACLR and reduces the spectral leakage. Polynomials with orders above $n = 5$ did not further improve the results significantly. A discussion on the impact of the choice of data points used in the LS problem can be found in Section 6.4 on page 79.

8.4.4 Summary

The measured performance of the PA for modulated signals is summarized in Table 8.6. The table shows measured ACLR with DPD, without DPD, and the required (Req) ACLR for the WCDMA [3GP] and the LTE [3GPP] standards. In measurements at 1.95 GHz, the DPD proved to be successful and improved the WCDMA ACLR at 5 MHz and 10 MHz offsets by 10.8 dB and 7.5 dB, respectively. The LTE ACLR at 5 MHz offset was improved by 9.4 dB. Thus, the predistortion method improves the measured ACLR to have at least 12.6 dB of margin to the requirements [3GP, 3GPP]. The measured ACLR at 5 MHz is comparable to state-of-the-art WCDMA transceivers [Huang et al., 2010].

To compare the DPD performance to the achievable ACLR, a small simulation study has been performed. Assuming a PA with 35 dB of dynamic range (neglecting phase distortions), i.e. assuming $g_1 = 0.509$ and $g_2 = 0.491$, and a polynomial degree of $n = 5$, the computed achievable ACLR at 5 MHz and 10 MHz is ~3 dB better compared to the measurements with the WCDMA signal. Similarly, the computed achievable ACLR at 5 MHz is ~2 dB better compared to the measurements with the LTE signal.

Table 8.6: *Measured Spectral Performance at 1.95 GHz for WCDMA and LTE Uplink Signals with Predistortion (using $n = 5$) and without.*

	Measured Parameter	**Req**	**Without DPD**	**With DPD**
WCDMA	ACLR @ 5 MHz [dBc]	-33	-35.5	-46.3
	ACLR @ 10 MHz [dBc]	-43	-48.1	-55.6
LTE	ACLR @ 5 MHz [dBc]	-30	-34.1	-43.5

As discussed in Section 6.4 on page 79, the polynomial fit is best in the middle, and in intervals where there is most data points. For the signals in this thesis, that is in the center of the interval, see Figure 8.2 for the distribution of the different signal types used. As seen in Figures 8.7 and 8.9, this is where the predistorter improves the performance. The predistorter is based on inversion of the PA models estimated using least squares. Since the inversion is almost perfect, see Figure 7.2 for the analytical inversion, the misfit at the smallest and largest input amplitudes can be assumed to be correlated with the polynomial fit of the PA model. The nonlinearity functions can be compared for different signal types, and though the overall appearance is very similar, a small shift can be seen, such that the fit has been adapted to the signal type. That is, for an LTE signal, the functions $\hat{f}_k$ differ a bit from the ones estimated for a WCDMA signal. This can be seen for lower amplitudes in particular, where the LTE signal has a higher signal density than the WCDMA.

9

Concluding Remarks

In this chapter, conclusions and some discussions on possible research ideas for the future are provided.

9.1 Conclusions

In this thesis, some different aspects concerning the estimation of inverse models have been discussed.

In system identification, the model should typically be estimated in the setting in which it will be used. This idea has been further investigated for the inverse model estimation, where different approaches are used in applications. Here, an inverse model has been estimated with the purpose of using it in cascade with the system itself, as an inverter. A good inverse model in this setting would be one that, when used in series with the original system, reconstructs the original input. This problem has been treated in applications, such as power amplifier predistortion, but theoretical insights as to how it should be done have been lacking. Different methods can lead to good results, but it can be shown that the characteristics captured by the methods differ. It is important to know how the choice of method effects the inverse model. For a noise-free linear time-invariant system, it is shown here that the weighting of the identification will be adjusted to better reflect this intended use when the inverse is estimated directly instead of based on a forward model. This has also been illustrated by a small example.

Inverse systems are used in many applications, and here, outphasing power amplifier predistortion has been investigated. The goal is to obtain a predistorter that counteracts the nonlinearities introduced by the amplifier. In outphasing PAs, the signal is decomposed into two branches, so that highly efficient, nonlinear amplifiers can be used. This structure makes it hard to use conventional

predistortion methods, but enables a theoretical description of the outphasing PA and the matching ideal predistorter. Here, a first method based on two non-convex optimization problems has been further developed using the structure of the outphasing amplifier. The improved method basically consists of two least-squares problems and an analytic inversion, and can be adapted to online implementation. It has been shown that the methods reduce the nonlinearities and the leakage into adjacent channels.

9.2 Open Questions

For amplifier predistortion, an interesting extension to the methods presented here is to include dynamics. In this thesis, two different approaches have been mentioned that did not improve the modeling results. Since nonlinear systems is "everything that is not linear", there are many ways to include nonlinear dynamics, and one of them not working (improving the modeling) does not mean that some other way will not. Even though the measurements did not indicate a large dynamic influence, extending the method to include possible dynamics would extend the field of application.

Since the measurements were performed in a rather ideal setup and then averaged over multiple realizations, the noise influence was minor, but the influence under less ideal conditions could be evaluated. Now, the noise has a large impact in the normalization and the estimation of gain factors in the two branches, since these depend on only one and two measurements, respectively. This could be made more robust by looking at multiple measurements.

The construction of a least-squares method is one step towards a possible online implementation, but further adaption could also be done. This includes the choice of whether the DPD should be implemented as a polynomial or in a LUT solution. Furthermore, if the method can be allowed to use some calibration time to adapt the parameters to the device at hand, this calls for another solution than if the method has to find the parameters during operation. However, this would more concern hardware implementation questions.

Concerning the research on estimation of inverse systems, only a small first investigation has been performed, and many open questions remain. These include a more thorough analysis of the properties of the estimators, as well as the noise influence in the different approaches. An extension to the nonlinear case would also be an interesting, but challenging topic.

A

Power Amplifier Implementation

The outphasing power amplifiers used for the measurements presented in Chapter 8 and the power amplifier modeling in Chapter 6 have been constructed by Jonas Fritzin, Christer Svensson and Atila Alvandpour at the Division of Electronic Devices, Linköping University, Linköping, Sweden. The results and pictures in this chapter are all measured and reproduced with the authors' permission and are published here for sake of completeness.

As described in Section 5.2, a power amplifier can be characterized by different measures, such as the efficiency and the gain. For the PA beginner, a quick review of these concepts and the others in Section 5.2 could be useful. See also the Glossary in the preamble (page xvi).

The power amplifiers are of outphasing-type. The amplifier in each branch is a Class D amplifier, based on inverters, that switches between V_{DD} and *GND*.

A.1 +10.3 dBm Class-D Outphasing RF Amplifier in 90 nm CMOS

The chip used for validation of the nonconvex method in Section 8.3 can be seen in the chip photo in Figure A.1 and the sketch in Figure A.2. The PA is a Class D outphasing amplifier with an inverter-based output stage and an on-chip transformer as power combiner. More specifics can be found in Fritzin [2011] and Fritzin et al. [2011a].

Figure A.3a shows the measured maximum output power (P_{out}), the *drain efficiency* (DE) and the *power-added efficiency* (PAE) over frequency for the power amplifier. V_{DD} and V_{bias} were 1.3 V and 0.65 V, respectively. The 3 dB bandwidth was 2 GHz (1-3 GHz). The output power at 2 GHz was +10.3 dBm with DE and PAE of 39 % and 33 %, respectively, with a gain of 23 dB from the buffers to the

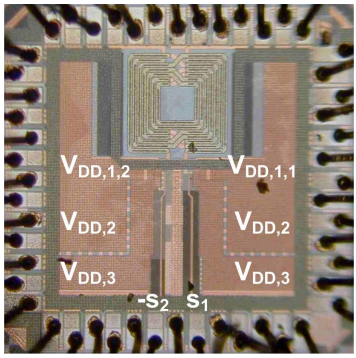


Figure A.1: *Photo of the chip with size 1x1mm^2.*

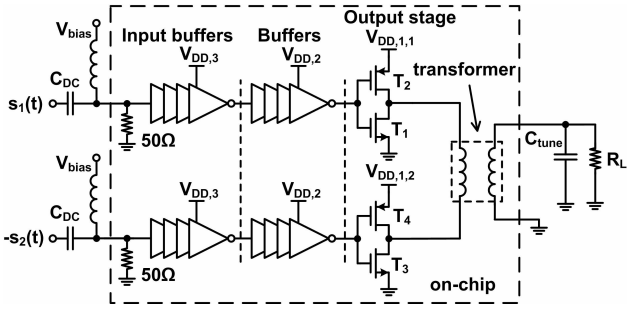
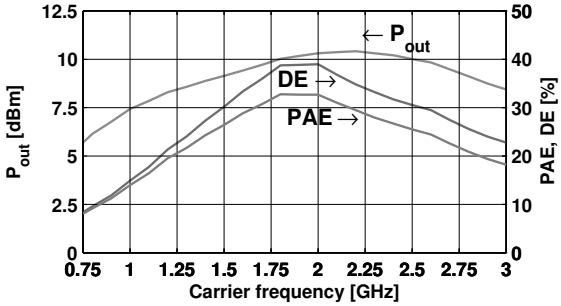
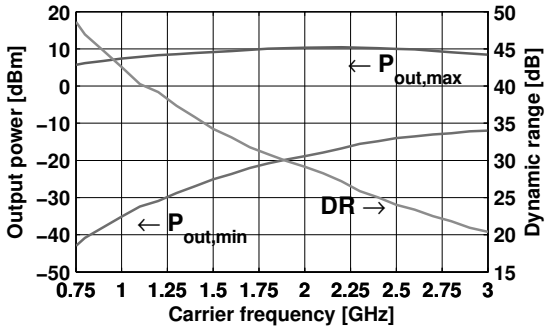


Figure A.2: *Implemented outphasing amplifier with inverters in the output stage.*



(a)



(b)

Figure A.3: *(a) Measured output power (P_{out}), DE and PAE over frequency. (b) Measured maximum output power, $P_{out,max}$, minimum output power, $P_{out,min}$, and dynamic range, DR, over frequency.*

output. The minimum and maximum output power and DR of the PA are plotted in Figure A.3b, where $P_{\text{out,max}} = P_{\text{out}}$ in Figure A.3a.

A.2 +30 dBm Class-D Outphasing RF Amplifier in 65 nm CMOS

The PA used for validation in Section 8.4 is described in more detail in Fritzin et al. [2011c] , but some basic characteristics can be found here. The chip photo can be seen in Figure A.4. Figure A.5 shows the outphasing PA, based on a Class D amplifier stage utilizing a cascode configuration illustrated in Figure A.6a. This configuration improves the life-time of the transistors by achieving a low on-resistance in the on-state and distributing the voltage stress in the off state which assures that the *root mean square* (RMS) electric fields across the gate oxide is kept low. The output stage is driven by an AC-coupled low-voltage driver operating at 1.3 V, V_{DD1}, to allow a 5.5 V, V_{DD2}, supply without excessive device voltage stress as discussed in Fritzin et al. [2011b] and Fritzin et al. [2011c]. The chip was attached

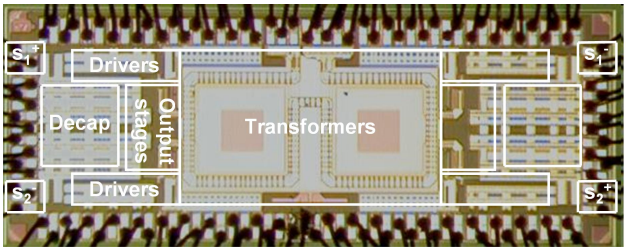


Figure A.4: *Photo of the chip with size 2.5x1.0mm^2. The photo has the same orientation as the simplified PA schematic in Figure A.5.*

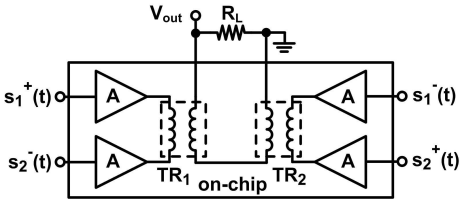


Figure A.5: *The implemented Class-D outphasing RF PA using two transformers to combine the outputs of four amplifier stages.*

to an FR4 PCB and connected with bond-wires.

The measured output power, drain efficiency and power-added efficiency over frequency and outphasing angle, φ in (5.11) (where $\varphi = 2\Delta_\psi$), for $V_{\mathrm{DD1}} = 1.3\,\mathrm{V}$ and $V_{\mathrm{DD2}} = 5.5\,\mathrm{V}$ is shown in Figures A.7. The output power at 1.95 GHz was +29.7 dBm with a PAE of 26.6 % (including all drivers). The PA had a peak to minimum power ratio of ~35 dB and the gain was 26 dB from the drivers to the output. The DC power consumption of the smallest drivers was considered as input power.

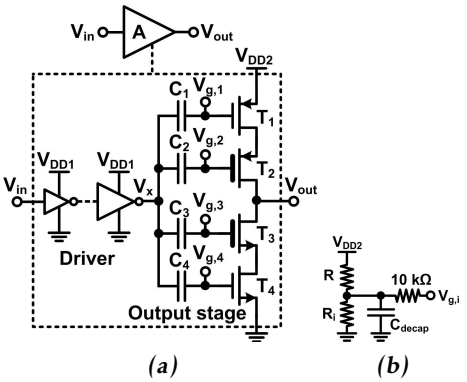
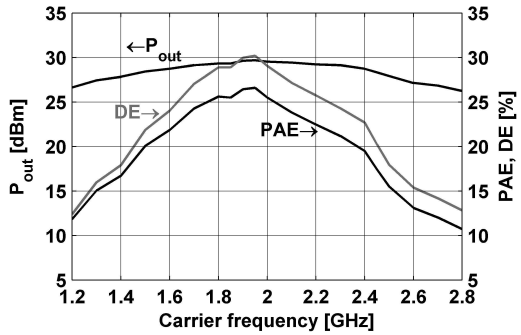
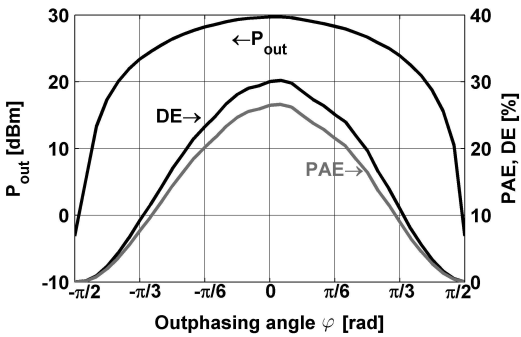


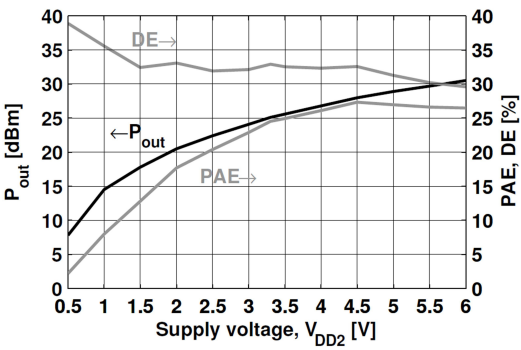
Figure A.6: *(a) The Class-D stage used in the outphasing PA Fritzin et al. [2011c]. C_1-C_4 are MIM capacitors. (b) Off-chip biasing resistors, R and R_i.*



(a)



(b)



(c)

Figure A.7: *Measured P_{out}, DE and PAE for V_{DD1} = 1.3 V and V_{DD2} = 5.5 V [Fritzin et al., 2011c]:*
(a) over carrier frequency.
(b) over outphasing angle, φ, at 1.95 GHz.
(c) Measured P_{out}, DE and PAE over V_{DD2} for V_{DD1} = 1.3 V at 1.95 GHz.

Bibliography

3GP. TS 25.101 v10.2.0 (2011-06). 3rd Generation Partnership Project; Technical specification group radio access network; user equipment (UE) radio transmission and reception (FDD), Release 10. Cited on page 115.

3GPP. TS 36.101 v10.3.0 (2011-06). 3rd Generation Partnership Project; Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); user equipment (UE) radio transmission and reception, Release 10. Cited on page 115.

Emad Abd-Elrady, Li Gan, and Gernot Kubin. Direct and indirect learning methods for adaptive predistortion of IIR Hammerstein systems. *Elektrotechnik & Informationstechnik*, 125(4):126–131, April 2008. Cited on pages 24 and 35.

Agilent. Agilent PN 89400-14, using error vector magnitude measurements to analyze and troubleshoot vector-modulated signals - *Product Note* 2000. http://cp.literature.agilent.com/litweb/pdf/5965-2898e.pdf. Accessed January, 2013. Cited on page 54.

Lars Ahlin, Jens Zander, and Ben Slimane. *Principles of Wireless Communications*. Studentlitteratur, 2006. ISBN 91-44-03080-0. Cited on page 101.

Shoaib Amin, Efrain Zenteno, Per N. Landin, Daniel Rönnow, Magnus Isaksson, and Peter Händel. Noise impact on the identification of digital predistorter parameters in the indirect learning architecture. In *Swedish Communication Technologies Workshop (SWE-CTW)*, pages 36–39, Lund, Sweden, October 2012. Cited on page 35.

Anritsu. Adjacent channel power ratio (ACPR) - *Application Note, Rev. A*. February 2001. http://www.us.anritsu.com/downloads/files/11410-00264.pdf, accessed January, 2013. Cited on page 53.

Karl J. Åström and Pieter Eykhoff. System identification - a survey. *Automatica*, 7:123–162, 1971. Cited on page 37.

Karl J. Åström and Tore Hägglund. *Advanced PID Control*. ISA - Instrumentation, Systems, and Automation Society, Second edition, 2005. ISBN 1-55617-942-1. Cited on pages 18 and 35.

Ahmed Birafane and Ammar B. Kouki. Phase-only predistortion for LINC amplifiers with Chireix-outphasing combiners. *IEEE Transactions on Microwave Theory and Techniques*, 53(6):2240–2250, June 2005. Cited on pages 60 and 63.

Ahmed Birafane, Mohamed El-Asmar, Ammar B. Kouki, Mohamed Helaoui, and Fadhel M. Ghannouchi. Analyzing LINC systems. *IEEE Microwave Magazine*, 11(5):59–71, August 2010. Cited on page 60.

Åke Björck. *Numerical Methods for Least Squares Problems*. Siam, 1996. Cited on pages 14 and 99.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. Cited on page 82.

Claudia Califano, Salvatore Monaco, and Dorothé Normand-Cyrot. On the discrete-time normal form. *IEEE Transactions on Automatic Control*, 43(11): 1654–1658, November 1998. Cited on page 22.

Tsan-Wen Chen, Ping-Yuan Tsai, Jui-Yuan Yu, and Chen-Yi Lee. A sub-mW all-digital signal component separator with branch mismatch compensation for OFDM LINC transmitters. *IEEE Journal of Solid-State Circuits*, 46(11):2514–2523, November 2011. Cited on page 63.

Hektbi Chireix. High power outphasing modulation. *IRE*, 23:1370–1392, November 1935. Cited on page 58.

Donald C. Cox. Linear amplification with nonlinear components. *IEEE Transactions on Communication*, COM-23:1942–1945, December 1974. Cited on page 58.

Steve C. Cripps. *RF Power Amplifiers for Wireless Communications*. Artech House, Second edition, 2006. ISBN 1-59693-018-7. Cited on pages 52 and 54.

Erik Dahlman, Stefan Parkvall, and Johan Sköld. *4G LTE/LTE-Advanced for Mobile Broadband*. Elsevier, 2011. ISBN 978-0-12-385489-6. Cited on page 101.

Germund Dahlquist and Åke Björck. *Numerical Methods in Scientific Computing, Vol I*. Siam, 2008. ISBN 978-0-898716-44. Cited on page 79.

Santosh Devasia. Technical notes and correspondance - should model-based inverse inputs be used as feedforward under plant uncertainty. *IEEE Transactions on Automatic Control*, 47(11):1865–1871, November 2002. Cited on page 35.

Norman R. Draper and Harry Smith. *Applied Regression Analysis*. John Wiley & Sons, Third edition, 1998. ISBN 0-471-17082-8. Cited on page 14.

Louis E. Frenzel. *Principles of Electronic Communication Systems*. McGraw-Hill, Second edition, 2003. ISBN 0-07-828131-8. Cited on pages 48, 51, 55, 58, and 101.

Jonas Fritzin. *CMOS RF Power Amplifiers for Wireless Communications*. Linköping Studies in Science and Technology. Dissertations. No 1399, Linköping University, Linköping, Sweden, SE-581 83 Linköping, Sweden, November 2011. Cited on pages 53, 60, 61, 103, 104, and 119.

Jonas Fritzin, Ylva Jung, Per N. Landin, Peter Händel, Martin Enqvist, and Atila Alvandpour. Phase predistortion of a Class-D outphasing RF amplifier in 90nm CMOS. *IEEE Transactions on Circuits and Systems-II: Express Briefs*, 58(10): 642–646, October 2011a. Cited on pages 17, 35, 64, and 119.

Jonas Fritzin, Christer Svensson, and Atila Alvandpour. A +32dBm 1.85GHz Class-D outphasing RF PA in 130nm CMOS for WCDMA/LTE. In *IEEE European Solid-State Circuits Conference (ESSCIRC)*, pages 127–130, Helsinki, Finland, September 2011b. Cited on page 121.

Jonas Fritzin, Christer Svensson, and Atila Alvandpour. A wideband fully integrated +30dBm Class-D outphasing RF PA in 65nm CMOS. In *IEEE International Symposium on Integrated Circuits (ISIC)*, pages 25–28, Singapore, Singapore, December 2011c. Cited on pages 109, 121, 123, and 124.

Walter Gerhard and Reinhard Knöchel. Prediction of bandwidth requirements for a digitally based WCDMA phase modulated outphasing transmitter. In *The European Conference on Wireless Technology*, pages 97–100, Paris, France, October 2005a. Cited on page 103.

Walter Gerhard and Reinhard Knöchel. LINC digital component separator for single and multicarrier W-CDMA signals. *IEEE Transactions on Microwave Theory and Techniques*, 53(1):274–282, January 2005b. Cited on page 103.

Michel Gevers and Lennart Ljung. Optimal experimental designs, with respect to the intended model application. *Automatica*, 22(5):543–554, September 1986. Cited on page 33.

Pere L. Gilabert, Daniel D. Silveira, Gabriel Montoro, and Gottfried Magerl. RF-power amplifier modeling and predistortion based on a modular approach. In *European Microwave Integrated Circuits Conference*, pages 265–268, Manchester, UK, September 2006. Cited on page 63.

Pere L. Gilabert, Eduard Bertran, Gabriel Montoro, and Jordi Berenguer. FPGA implementation of an LMS-based real-time adaptive predistorter for power amplifiers. In *Circuits and Systems and TAISA Conference, 2009. NEWCAS-TAISA '09. Joint IEEE North-East Workshop on*, Toulouse, France, June-July 2009. Cited on page 99.

Lei Guan and Anding Zhu. Low-cost FPGA implementation of Volterra series-based digital predistorter for RF power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, 58(4):866–872, April 2010. Cited on pages 61 and 62.

Mohamed Helaoui, Slim Boumaiza, and Fadhel M. Ghannouchi. On the outphasing power amplifier nonlinearity analysis and correction using digital predistortion technique. In *IEEE Radio and Wireless Symposium (RWS)*, pages 751–754, Orlando, FL, USA, January 2008. Cited on page 63.

Ronald M. Hirschorn. Invertibility of multivariable nonlinear control systems. *IEEE Transactions on Automatic Control*, 24(6):855–865, December 1979. Cited on page 27.

Qiuting Huang, Jürgen Rogin, Xinhua Chen, David Tschopp, Thomas Burger, Thomas Christen, Dimitris Papadopolous, Ilian Kouchev, Chiara Martelli, and Thomas Dellsperger. A tri-band SAW-less WCDMA/HSPA RF CMOS transceiver, with on-chip DC-DC converter connectable to battery. In *International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pages 60–61, San Fransisco, CA, USA, February 2010. Cited on pages 107 and 115.

Richard C. Jaeger and Travis N. Blalock. *Microelectronic Circuit Design*. McGraw-Hill, Third edition, 2008. ISBN 978-0-07-110203-2. Cited on pages 55 and 57.

Ylva Jung and Martin Enqvist. Estimating models of inverse systems. In *52nd IEEE Conference on Decision and Control (CDC)*, Florence, Italy, To appear, December 2013. Cited on page 34.

Ylva Jung, Jonas Fritzin, Martin Enqvist, and Atila Alvandpour. Least-squares phase predistortion of a +30dbm Class-D outphasing RF PA in 65nm CMOS. *IEEE Transactions on Circuits and Systems-I: Regular papers*, 60(7):1915–1928, July 2013. Cited on pages 64 and 69.

Peter B Kenington. *High-Linearity RF Amplifier Design*. Artech House, 2000. ISBN 1-58053-143-1. Cited on page 61.

Per N. Landin, Jonas Fritzin, Wendy Van Moer, Magnus Isaksson, and Atila Alvandpour. Modeling and digital predistortion of Class-D outphasing RF power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, 60(6): 1907–1915, June 2012. Cited on pages 64 and 69.

Lennart Ljung. *System Identification, Theory for the User*. Prentice Hall PTR, Second edition, 1999. ISBN 0-13-656695-2. Cited on pages 10, 11, 12, 15, 33, 34, 36, 68, 69, 71, and 99.

Lennart Ljung. *System Identification, Toolbox, User's Guide*. MathWorks, Sixth edition, 2003. Cited on page 38.

Aarne Mämmelä. Commutation in linear and nonlinear systems. *Frequenz*, 60 (5-6):92–94, June 2006. Cited on page 23.

Ola Markusson. *Model and System Inversion with Applications in Nonlinear System Identification and Control*. TRITA-S3-REG-0201, Royal Institute of Technology, Stockholm, Sweden, SE-100 44 Stockholm, Sweden, 2001. Cited on pages 20, 22, 26, and 27.

Jaimin Mehta, Vasile Zoicas, Oren Eliezer, R. Bogdan Staszewski, Sameh Rezeq, Mitch Entezari, and Poras Bolsara. An efficient linearization scheme for a digital polar EDGE transmitter. *IEEE Transactions on Circuits and Systems-II: Express Briefs*, 57(3):193–197, March 2010. Cited on page 107.

Shervin Moloudi, Koji Takanami, Michael Youssef, Mohyee Mikhemar, and Asad Abidi. An outphasing power amplifier for software-defined radio transmitter. In *International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pages 568–569, San Fransisco, CA, USA, February 2008. Cited on page 63.

Gabriel Montoro, Pere L. Gilabert, Eduard Bertran, Albert Cesari, and José A. Garcia. An LMS-based adaptive predistorter for cancelling nonlinear memory effects in RF power amplifiers. In *Microwave Conference, 2007. APMC 2007. Asia-Pacific*, Bangkok, Thailand, December 2007. Cited on page 99.

Kevin L. Moore. *Iterative Learning Control for Deterministic Systems*. Springer Verlag, 1993. ISBN 978-1-4471-1914-2. Cited on page 20.

Seong-Sik Myoung, Il-Kyoo Kee, Jong-Gwan Yook, Kyutae Lim, and Joy Laskar. Mismatch detection and compensation method for the LINC system using a closed-form expression. *IEEE Transactions on Microwave Theory and Techniques*, 56(12):3050–3057, December 2008. Cited on page 63.

Henna Paaso and Aarne Mämmelä. Comparison of direct learning and indirect learning predistortion architechtures. In *IEEE International Syposium on Wireless Communication Systems (ISWCS)*, pages 309–313, Reykjavik, Iceland, October 2008. Cited on pages 24 and 35.

Rik Pintelon and Johan Schoukens. *System Identification - A Frequency Domain Approach*. IEEE Press and John Wiley & Sons, Second edition, 2012. ISBN 978-0-470-64037-1. Cited on pages 10, 33, and 34.

Behzad Razavi. *RF Microelectronics*. Prentice Hall, 1998. ISBN 0-13-887571-5. Cited on pages 52 and 55.

Rohde & Schwarz. Application note, 1GP67: Phase adjustment of two MIMO signal sources with option B90. Cited on pages 103 and 109.

Luca Romanò, Luigi Panseri, Carlo Samori, and Andrea L. Lacaita. Matching requirements in LINC transmitters for OFDM signals. *IEEE Transactions on Circuits and Systems-I: Regular Papers*, 53(7):1572–1578, July 2006. Cited on pages 60 and 63.

Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill Book Co., Third edition, 1976. ISBN 0-07-085613-3. Cited on pages 68 and 89.

Wilson J. Rugh. *Linear System Theory*. Prentice-Hall, Second edition, 1996. ISBN 0-13-441205-2. Cited on page 18.

Shankar Sastry. *Nonlinear Systems – Analysis, Stability and Control*. Springer Verlag, New York, 1999. ISBN 0-387-98513-1. Cited on pages 21 and 25.

Martin Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems*. John Wiley & Sons, New York, 1980. ISBN 0-471-04455-5. Cited on pages 25 and 26.

Torsten Söderström and Petre Stoica. *System Identification*. Prentice Hall, 1989. ISBN 0-13-881236-5. Cited on page 10.

Michael Soudan and Christian Vogel. Correction structures for linear weakly time-varying systems. *IEEE Transactions on Circuits and Systems-I: Regular papers*, 59(9):2075–2084, September 2012. Cited on page 21.

Murali Tummla, Michael T. Donovan, Bruce E. Watkins, and Robert North. Volterra series based modeling and compensation of nonlinearities in high power amplifiers. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2417–2420 vol. 3, Munich, Germany, April 1997. Cited on pages 25 and 62.

Johanna Wallén. *Estimation-Based Iterative Learning Control*. Linköping Studies in Science and Technology. Dissertations. No 1358, Linköping University, Linköping, Sweden, SE-581 83 Linköping, Sweden, February 2011. Cited on page 20.

Gaoming Xu, Taijun Liu, Yan Ye, and Tiefeng Xu. FPGA implementation of augmented hammerstein predistorters for RF power amplifier linearization. In *Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*, pages 481–484, Beijing, China, October 2009. Cited on page 63.

Hongtao Xu, Yorgos Palaskas, Ashoke Ravi, and Krishnamurthy Soumyanath. A highly linear 25dBm outphasing power amplifier in 32nm CMOS for WLAN application. In *IEEE European Solid-State Circuits Conference (ESSCIRC)*, pages 306–309, Seville, Spain, September 2010. Cited on page 61.

Jingshi Yao and Stephen I. Long. Power amplifier selection for LINC application. *IEEE Transactions on Circuits and Systems-II: Express Briefs*, 53(8):763–766, August 2006. Cited on page 61.

Xuejun Zhang, Lawrence E. Larson, Peter M. Asbeck, and Peter Nanawa. Gain/phase imbalance-minimization techniques for LINC transmitters. *IEEE Transactions on Microwave Theory and Techniques*, 49(12):2507–2516, June 2001. Cited on page 63.

Anding Zhu, Paul J. Draxler, Jonmei J. Yan, Thomas J. Brazil, Donald F. Kimball, and Peter M. Asbeck. Open-loop digital predistorter for RF power amplifiers using dynamic deviation reduction-based Volterra series. *IEEE Transactions on Microwave Theory and Techniques*, 56(7):1524–1534, July 2008. Cited on pages 26 and 62.

Licentiate Theses
Division of Automatic Control
Linköping University

P. Andersson: Adaptive Forgetting through Multiple Models and Adaptive Control of Car Dynamics. Thesis No. 15, 1983.

B. Wahlberg: On Model Simplification in System Identification. Thesis No. 47, 1985.

A. Isaksson: Identification of Time Varying Systems and Applications of System Identification to Signal Processing. Thesis No. 75, 1986.

G. Malmberg: A Study of Adaptive Control Missiles. Thesis No. 76, 1986.

S. Gunnarsson: On the Mean Square Error of Transfer Function Estimates with Applications to Control. Thesis No. 90, 1986.

M. Viberg: On the Adaptive Array Problem. Thesis No. 117, 1987.

K. Ståhl: On the Frequency Domain Analysis of Nonlinear Systems. Thesis No. 137, 1988.

A. Skeppstedt: Construction of Composite Models from Large Data-Sets. Thesis No. 149, 1988.

P. A. J. Nagy: MaMiS: A Programming Environment for Numeric/Symbolic Data Processing. Thesis No. 153, 1988.

K. Forsman: Applications of Constructive Algebra to Control Problems. Thesis No. 231, 1990.

I. Klein: Planning for a Class of Sequential Control Problems. Thesis No. 234, 1990.

F. Gustafsson: Optimal Segmentation of Linear Regression Parameters. Thesis No. 246, 1990.

H. Hjalmarsson: On Estimation of Model Quality in System Identification. Thesis No. 251, 1990.

S. Andersson: Sensor Array Processing; Application to Mobile Communication Systems and Dimension Reduction. Thesis No. 255, 1990.

K. Wang Chen: Observability and Invertibility of Nonlinear Systems: A Differential Algebraic Approach. Thesis No. 282, 1991.

J. Sjöberg: Regularization Issues in Neural Network Models of Dynamical Systems. Thesis No. 366, 1993.

P. Pucar: Segmentation of Laser Range Radar Images Using Hidden Markov Field Models. Thesis No. 403, 1993.

H. Fortell: Volterra and Algebraic Approaches to the Zero Dynamics. Thesis No. 438, 1994.

T. McKelvey: On State-Space Models in System Identification. Thesis No. 447, 1994.

T. Andersson: Concepts and Algorithms for Non-Linear System Identifiability. Thesis No. 448, 1994.

P. Lindskog: Algorithms and Tools for System Identification Using Prior Knowledge. Thesis No. 456, 1994.

J. Plantin: Algebraic Methods for Verification and Control of Discrete Event Dynamic Systems. Thesis No. 501, 1995.

J. Gunnarsson: On Modeling of Discrete Event Dynamic Systems, Using Symbolic Algebraic Methods. Thesis No. 502, 1995.

A. Ericsson: Fast Power Control to Counteract Rayleigh Fading in Cellular Radio Systems. Thesis No. 527, 1995.

M. Jirstrand: Algebraic Methods for Modeling and Design in Control. Thesis No. 540, 1996.

K. Edström: Simulation of Mode Switching Systems Using Switched Bond Graphs. Thesis No. 586, 1996.

J. Palmqvist: On Integrity Monitoring of Integrated Navigation Systems. Thesis No. 600, 1997.
A. Stenman: Just-in-Time Models with Applications to Dynamical Systems. Thesis No. 601, 1997.
M. Andersson: Experimental Design and Updating of Finite Element Models. Thesis No. 611, 1997.
U. Forssell: Properties and Usage of Closed-Loop Identification Methods. Thesis No. 641, 1997.
M. Larsson: On Modeling and Diagnosis of Discrete Event Dynamic systems. Thesis No. 648, 1997.
N. Bergman: Bayesian Inference in Terrain Navigation. Thesis No. 649, 1997.
V. Einarsson: On Verification of Switched Systems Using Abstractions. Thesis No. 705, 1998.
J. Blom, F. Gunnarsson: Power Control in Cellular Radio Systems. Thesis No. 706, 1998.
P. Spångéus: Hybrid Control using LP and LMI methods – Some Applications. Thesis No. 724, 1998.
M. Norrlöf: On Analysis and Implementation of Iterative Learning Control. Thesis No. 727, 1998.
A. Hagenblad: Aspects of the Identification of Wiener Models. Thesis No. 793, 1999.
F. Tjärnström: Quality Estimation of Approximate Models. Thesis No. 810, 2000.
C. Carlsson: Vehicle Size and Orientation Estimation Using Geometric Fitting. Thesis No. 840, 2000.
J. Löfberg: Linear Model Predictive Control: Stability and Robustness. Thesis No. 866, 2001.
O. Härkegård: Flight Control Design Using Backstepping. Thesis No. 875, 2001.
J. Elbornsson: Equalization of Distortion in A/D Converters. Thesis No. 883, 2001.
J. Roll: Robust Verification and Identification of Piecewise Affine Systems. Thesis No. 899, 2001.
I. Lind: Regressor Selection in System Identification using ANOVA. Thesis No. 921, 2001.
R. Karlsson: Simulation Based Methods for Target Tracking. Thesis No. 930, 2002.
P.-J. Nordlund: Sequential Monte Carlo Filters and Integrated Navigation. Thesis No. 945, 2002.
M. Östring: Identification, Diagnosis, and Control of a Flexible Robot Arm. Thesis No. 948, 2002.
C. Olsson: Active Engine Vibration Isolation using Feedback Control. Thesis No. 968, 2002.
J. Jansson: Tracking and Decision Making for Automotive Collision Avoidance. Thesis No. 965, 2002.
N. Persson: Event Based Sampling with Application to Spectral Estimation. Thesis No. 981, 2002.
D. Lindgren: Subspace Selection Techniques for Classification Problems. Thesis No. 995, 2002.
E. Geijer Lundin: Uplink Load in CDMA Cellular Systems. Thesis No. 1045, 2003.
M. Enqvist: Some Results on Linear Models of Nonlinear Systems. Thesis No. 1046, 2003.
T. Schön: On Computational Methods for Nonlinear Estimation. Thesis No. 1047, 2003.
F. Gunnarsson: On Modeling and Control of Network Queue Dynamics. Thesis No. 1048, 2003.
S. Björklund: A Survey and Comparison of Time-Delay Estimation Methods in Linear Systems. Thesis No. 1061, 2003.

M. Gerdin: Parameter Estimation in Linear Descriptor Systems. Thesis No. 1085, 2004.

A. Eidehall: An Automotive Lane Guidance System. Thesis No. 1122, 2004.

E. Wernholt: On Multivariable and Nonlinear Identification of Industrial Robots. Thesis No. 1131, 2004.

J. Gillberg: Methods for Frequency Domain Estimation of Continuous-Time Models. Thesis No. 1133, 2004.

G. Hendeby: Fundamental Estimation and Detection Limits in Linear Non-Gaussian Systems. Thesis No. 1199, 2005.

D. Axehill: Applications of Integer Quadratic Programming in Control and Communication. Thesis No. 1218, 2005.

J. Sjöberg: Some Results On Optimal Control for Nonlinear Descriptor Systems. Thesis No. 1227, 2006.

D. Törnqvist: Statistical Fault Detection with Applications to IMU Disturbances. Thesis No. 1258, 2006.

H. Tidefelt: Structural algorithms and perturbations in differential-algebraic equations. Thesis No. 1318, 2007.

S. Moberg: On Modeling and Control of Flexible Manipulators. Thesis No. 1336, 2007.

J. Wallén: On Kinematic Modelling and Iterative Learning Control of Industrial Robots. Thesis No. 1343, 2008.

J. Harju Johansson: A Structure Utilizing Inexact Primal-Dual Interior-Point Method for Analysis of Linear Differential Inclusions. Thesis No. 1367, 2008.

J. D. Hol: Pose Estimation and Calibration Algorithms for Vision and Inertial Sensors. Thesis No. 1370, 2008.

H. Ohlsson: Regression on Manifolds with Implications for System Identification. Thesis No. 1382, 2008.

D. Ankelhed: On low order controller synthesis using rational constraints. Thesis No. 1398, 2009.

P. Skoglar: Planning Methods for Aerial Exploration and Ground Target Tracking. Thesis No. 1420, 2009.

C. Lundquist: Automotive Sensor Fusion for Situation Awareness. Thesis No. 1422, 2009.

C. Lyzell: Initialization Methods for System Identification. Thesis No. 1426, 2009.

R. Falkeborn: Structure exploitation in semidefinite programming for control. Thesis No. 1430, 2010.

D. Petersson: Nonlinear Optimization Approaches to $\mathcal{H}_2$-Norm Based LPV Modelling and Control. Thesis No. 1453, 2010.

Z. Sjanic: Navigation and SAR Auto-focusing in a Sensor Fusion Framework. Thesis No. 1464, 2011.

K. Granström: Loop detection and extended target tracking using laser data. Thesis No. 1465, 2011.

J. Callmer: Topics in Localization and Mapping. Thesis No. 1489, 2011.

F. Lindsten: Rao-Blackwellised particle methods for inference and identification. Thesis No. 1480, 2011.

M. Skoglund: Visual Inertial Navigation and Calibration. Thesis No. 1500, 2011.

S. Khoshfetrat Pakazad: Topics in Robustness Analysis. Thesis No. 1512, 2011.

P. Axelsson: On Sensor Fusion Applied to Industrial Manipulators. Thesis No. 1511, 2011.

A. Carvalho Bittencourt: On Modeling and Diagnosis of Friction and Wear in Industrial Robots. Thesis No. 1516, 2012.

P. Rosander: Averaging level control in the presence of frequent inlet flow upsets . Thesis No. 1527, 2012.

N. Wahlström: Localization using Magnetometers and Light Sensors. Thesis No. 1581, 2013.
R. Larsson: System Identification of Flight Mechanical Characteristics. Thesis No. 1599, 2013.