Nonlinear MIMO communication systems : channel estimation and information recovery using Volterra models

Carlos Alexandre Rolim Fernandes

Université de Nice-Sophia Antipolis

Ecole Doctorale STIC

Sciences et Technologies de l'Information et de la Communication

THESE

pour obtenir le titre de

Docteur en Sciences

de l'Université de Nice Sophia Antipolis

Mention: Automatique, Traitement du Signal et des Images

présentée et soutenue par

Carlos Alexandre ROLIM FERNANDES

Nonlinear MIMO Communication Systems: Channel Estimation and Information Recovery using Volterra Models

Thèse dirigée en régime de cotutelle par Gérard FAVIER et João Cesar MOTA

Soutenue le 03 juillet 2009

Jury:

M. Gérard FAVIER	Directeur de Recherche au CNRS, I3S	Directeur de thèse
M. João Cesar M. MOTA	Professeur à l'Université Federale du Ceará	Directeur de thèse
M. Luc Deneire	Professeur à l'Université de Nice-Sophia Antipolis	Président
M. Eric Moreau	Professeur à l'Université de Toulon	Rapporteur
M. João Marcos T. Romano	Professeur à l'Université de Campinas	Rapporteur
M. Charles C. Cavalcante	Maître de conf. à l'Université Federale du Ceará	Examinateur invité
M. André L. F. de Almeida	Chercheur à l'Université Federale du Ceará	Examinateur invité

To Aline

Acknowledgements

My thesis has been developed under a double diploma (*cotutelle*) convention between the Université de Nice-Sophia Antipolis (UNSA), France, and the Universidade Federal do Ceará (UFC), Brazil. First, I would like to thank the CAPES agency of the Brazilian government for their financial support of my PhD by means of a scholarship.

This work would not have been possible without the support of many people who encouraged and helped me during my PhD. I wish to express my gratitude to my supervisor at the UNSA, Prof. Gérard Favier, for the continuous support during my Ph.D study and research. Without his patience, knowledge and advice, from the initial to the final moment, this thesis would not have been possible. I would equally like to thank Prof. João Cesar M. Mota, my advisor at UFC, who has assisted me in numerous ways. It is he who has made emerge my interest in scientific research since my first years in college. His guidance, energy and enthusiasm have given me the necessary motivation to go on even during the hard times, specially in the last year. He has also become a great friend during these years. I would also like to thank the rest of my thesis committee, Prof. Luc Deneire, Prof. João Marcos T. Romano, Prof. Eric Moreau, Prof. Charles C. Cavalcante and Dr. André L. F. de Almeida, for their constructive comments. I would equally like to thank Alain Kibangou for his useful remarks.

All my lab friends at the I3S Laboratory made it a convivial and nice place to work. So, I would like to thank all of my friends at I3S (not exhaustively): Pietro Bonizzi, Ronald Phlypo, PA Aguilar, Matthias Hesse, Laurent Galluccio, Xavier Luciani, Laure Amate, Marie-Andre Agostini, Lionel Nicolas and Vicenzo Angelino. Many other people that I have met in these years have marked my life and made it much better. I would like to express my gratitude to Flora, Olivier, Michele, Neco, Emmanuelle, Jaime, Katiúcia and Elton. Thanks for all the great and unforgettable moments.

I would also like to thank Hélida, for being the great woman that you are. Special thanks to my brother Estevão, for having taught me a lot and who has inspired me to follow this career. You have always been the greatest example for me. I also wish to thank deeply the rest of my family, specially my parents, Fernando and Cynthia who have educated me with lots of affection and tought me to be who I am. Thanks for always having believed in me.

At last, I wish to thank very carefully the most important people in my life: my girlfriend, wife, fiancé and best friend Aline. You are the most beautiful person that I have ever met. I can not explain how much important and precious you are to me. Thanks for supporting me unconditionally in all my decisions. I know you have given up many things to be at my side during all these years. This is why I dedicate this work to you.

Abstract

Due to the presence of nonlinear devices such as power amplifiers (PA) and optical instruments, the communication signals are sometimes corrupted by nonlinear distortions. In such cases, nonlinear models are used to provide an accurate signal representation, allowing the development of efficient signal processing techniques capable of eliminating or reducing these nonlinear distortions. In this context, the choice of the nonlinear system model plays a fundamental role. The Volterra model has since longtime been used to represent communication systems in presence of nonlinear distortions, with applications for modeling satellite communication links, orthogonal frequency division multiplexing (OFDM) systems and radio over fiber (ROF) channels.

The main objective of this thesis is to propose techniques for channel estimation and information recovery in multiple-input-multiple-output (MIMO) Volterra communication systems. This kind of MIMO model is able of modeling nonlinear communication channels with multiple transmit and receive antennas, as well as multi-user channels with a single transmit antenna for each user and multiple receive antennas. Channel estimation and equalization techniques are developed for three types of nonlinear MIMO communication systems: OFDM, ROF and Code division multiple access (CDMA)-ROF systems. According to the considered communication systems, different kinds of MIMO Volterra models are used. In the case of OFDM systems, we develop receivers that exploit the diversity provided by a proposed transmission scheme. In the case of time and space division multiple access (TDMA-SDMA) systems, a set of orthonormal polynomials is developed for increasing the convergence speed of a supervised adaptive MIMO Volterra estimation algorithm. Moreover, in order to develop signal processing techniques for MIMO Volterra communication channels in a blind scenario, we make use of tensor decompositions. By exploiting the fact that Volterra models are linear with respect to their coefficients, blind estimation and equalization of MIMO Volterra channels are carried out by means of the Parallel Factor (PARAFAC) tensor decomposition, considering TDMA-SDMA and CDMA communication systems.

Key-words: Nonlinear communication system, Volterra model, MIMO, channel estimation, information recovery, OFDM, Radio Over Fiber, CDMA, PARAFAC decomposition.

Résumé

Du à la présence de dispositifs non-linéaires comme des amplificateurs de puissance (PAs) et des instruments optiques, les signaux de communication sont parfois corrompus par des distorsions non-linéaires. Dans ce cas, des modèles non-linéaires sont utilisés pour fournir une description précise des signaux, permettant le développement de techniques de traitement du signal capables d'éliminer ou de réduire ces distorsions. Dans ce contexte, le choix du modèle non-linéaire a une importance majeure. Les modèles de Volterra sont depuis longtemps utilisés pour représenter les systèmes de communication en présence de distorsions non-linéaires, ayant des applications dans les systèmes de communication par satellite, les systèmes OFDM et les systèmes *radio over fiber* (ROF), entre autres.

Le principal objectif de cette thèse est de proposer des techniques d'estimation et de récupération d'information dans les systèmes de communication MIMO Volterra. Ce type de modèle MIMO peut être utilisé pour modéliser des canaux de communication avec de multiples antennes à la transmission et à la réception, ainsi que des canaux multi-utilisateurs avec de multiples antennes à réception et une antenne de transmission par utilisateur. Les techniques d'estimation et d'égalisation de canaux sont développées pour trois systèmes de communication non-linéaires différents: OFDM, ROF et ROF-CDMA, différents modèles MIMO Volterra étant utilisés selon l'application considérée. Dans le cas des systèmes du type OFDM, un nouveau schéma de transmission qui introduit de la redondance dans les signaux transmis, ainsi que des récepteurs exploitant cette redondance sont proposés. Dans le cas des systèmes TDMA-SDMA, un ensemble de polynômes orthonormaux est développé pour améliorer la vitesse de convergence de l'algorithme LMS pour l'estimation adaptative supervisée d'un système MIMO Volterra. D'autre part, le développement de récepteurs pour des systèmes de communication MIMO Volterra dans un schéma de transmission aveugle est réalisé à l'aide de décompositions tensorielles. Dans ce cas, en exploitant le fait que les modèles de Volterra sont linéaires vis-à-vis de leurs coefficients, des techniques d'estimation et d'égalisation de canaux MIMO Volterra basées sur la décomposition PARAFAC sont développées pour des systèmes de communication TDMA-SDMA et CDMA.

Mots-clés: Système de communication non-linéaire, modèle de Volterra, MIMO, estimation de canal, récupération d'information, OFDM, système Radio Sur Fibre, CDMA, décomposition PARAFAC.

Resumo

DEVIDO à presença de dispositivos não-lineares tais como amplificadores de potência (PAs) e equipamentos ópticos, sinais em sistemas de comunicações estão sujeitos a distorções não-lineares. Quando tais efeitos são importantes, modelos não-lineares são utilizados para fornecer uma descrição precisa dos sinais, permitindo o desenvolvimento de técnicas de processamento de sinais capazes de reduzir ou eliminar estas distorções. Dentro deste contexto, a escolha apropriada de um modelo de sistema não-linear é de grande importância. Os sistemas de Volterra são um dos modelos mais utilizados para representar sistemas de comunicações contaminados por distorções não-lineares, com aplicações na modelagem de enlaces satelitários, sistemas OFDM e *radio over fiber* (ROF), entre outros.

O principal objetivo desta tese é propor técnicas de estimação de canal e recuperação da informação em sistemas de comunicação MIMO Volterra. Tais modelos MIMO podem ser utilizados para representar sistemas de comunicação não-lineares com múltiplas antenas na transmissão e recepção, assim como canais multi-usuários com múltiplas antenas na recepção e uma antena de transmissão por usuário. As técnicas de estimação e equalização de canais contidas nesta tese são desenvolvidas para três diferentes tipos de sistemas de comunicação MIMO não-lineares: OFDM, ROF e ROF-CDMA, com diferentes tipos de modelos MIMO Volterra sendo utilizados de acordo como o sistema de comunicação considerado. Para o caso de sistemas OFDM, um esquema de transmissão que introduz diversidade nos sinais recebidos, assim como receptores que exploram esta diversidade, são propostos. Para o caso de sistemas TDMA-SDMA, um conjunto de polinômios ortogonais é desenvolvido para acelerar a convergência do algoritmo LMS durante a estimação supervisionada de sistemas MIMO Volterra. Além disso, para se desenvolver técnicas de processamento de sinais para canais de comunicação MIMO Volterra em um ambiente não-supervisionado, esta tese faz uso de decomposições tensoriais. Neste caso, técnicas de estimação e equalização de canais MIMO Volterra são desenvolvidas, tanto no caso de sistemas TDMA-SDMA, como no caso de sistemas CDMA, utilizando-se a decomposição PARAFAC e explorando o fato de os modelos de Volterra serem lineares com relação a seus coeficientes.

Palavras-chave: Sistema de comunicação não-linear, modelo de Volterra, MIMO, estimação de canal, recuperação de informação, OFDM, Rádio Sobre Fibra, CDMA, decomposição PARAFAC.

Contents

List of Figures

List of Tables

Mathematical Notation

x	Scalar variable - lower-case letters
$\mathbf{x}$	Vector - boldface lower-case letters
$\mathbf{X}$	Matrix - boldface capital letters
$\mathcal{X}$	Tensor (high-order array) - calligraphic letters
$[\mathbf{x}]_i$	i^{th} element of the vector $\mathbf{x}$
$[\mathbf{X}]_{i,j}$	$(i,j)^{th}$ element of the matrix $\mathbf{X}$
$[\mathcal{X}]_{i_1,i_2,\ldots,i_N}$	$(i_1,i_2,\ldots,i_N)^{th}$ element of the tensor $\mathcal{X}$
$[\mathbf{X}]_{\cdot,i}$	i^{th} column of the matrix $\mathbf{X}$
$[\mathbf{X}]_{i,\cdot}$	i^{th} row of the matrix $\mathbf{X}$
$\lvert x \rvert$	modulus of the scalar x
$\| \mathbf{x} \|_2$	l^2 norm of the vector $\mathbf{x}$
$\| \mathbf{X} \|_F$	Frobenius norm of the matrix $\mathbf{X}$
$\text{vec}\,[\mathbf{X}]$	Vectorization operator: stacks the columns of the matrix $\mathbf{X}$ into a single column vector

$\text{diag}\left[\mathbf{x}\right]$	Diagonal matrix built from the vector $\mathbf{x}$
$\text{diag}_i[\mathbf{X}]$	Diagonal matrix built from the i^{th} row of the matrix $\mathbf{X}$
$(\cdot)^T$	Transpose of a matrix or vector
$(\cdot)^*$	Conjugate of a matrix or vector
$(\cdot)^H$	Hermitian (transpose and conjugate) of a matrix or vector
$\mathbf{X}^\dagger$	Moore-Penrose pseudo-inverse of the matrix $\mathbf{X}$
$r_{\mathbf{X}}$	Rank of the matrix $\mathbf{X}$
$k_{\mathbf{X}}$	k-rank of the matrix $\mathbf{X}$
$\jmath$	Imaginary unit $(\sqrt{-1})$
$\mathbf{0}_{N,M}$	"All-zeros" matrix of dimensions $N \times M$
$\mathbf{1}_{N,M}$	"All-ones" matrix of dimensions $N \times M$
$\mathbf{I}_N$	Identity matrix of order N
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
$\mathbb{E}\{\cdot\}$	Mathematical expectation
$\otimes$	Kronecker product
$\otimes^k$	Power of order k of the Kronecker product
$\oslash$	Truncated Kronecker product
$\oslash^k$	Power of order k of the truncated Kronecker product
$\diamond$	Khatri-Rao (column-wise Kronecker) product
$\delta(i)$	Kronecker symbol, i.e. $\delta(0) = 1$ and $\delta(i) = 0$ for $i \neq 0$.

List of Acronyms

AWGN	Additive white Gaussian noise
ALS	Alternating Least Squares
ALS-DD-BI	ALS with Direct Decision and Block Initialization
BER	Bit-error-rate
CDMA	Code Division Multiple Access
DTMC	Discrete time Markov chain
EVD	Eigenvalue Decomposition
E/O	Electrical-optical
FFT	Fast Fourier Transform
FIR	Finite impulse response
ICI	Inter-carrier interference
ISI	Intersymbol interference
IFFT	Inverse Fast Fourier Transform
JDA	Joint Diagonalization Algorithm
LMS	Least Mean Square
LS	Least squares
MMSE	Minimum mean square error
MAI	Multiple access interference
MIMO	Multiple-input multiple-output

NMSE	Normalized mean squared error
OFDM	Orthogonal Frequency Division Multiplexing
PARAFAC	Parallel Factor
PAPR	Peak-to-average power ratio
PSK	Phase-shift keying
PA	Power amplifier
PDR	Power Diversity-based Receiver
PDF	Probability density function
QAM	Quadrature amplitude modulation
RAP	Radio access point
ROF	Radio Over Fiber
SNR	Signal-to-noise-ratio
SIMO	Single-input multiple-output
SISO	Single-input single-output
SDMA	Space Division Multiple Access
TDMA	Time Division Multiple Access
TPM	Transition probability matrix
ZF	Zero Forcing

CHAPTER 1

Introduction

In many practical situations, the input-output relationship of a system cannot be assumed to be linear. In these cases, nonlinear models are powerful tools for representing the system behavior. The Volterra series [160], developed by the italian mathematician Vito Volterra in 1887, is one of the most common representations of nonlinear systems [134]. It constitutes a class of polynomial models that can be viewed as a sort of extension of the linear convolution. The Volterra series has received a considerable attention from researchers of different areas. Among them, it should be highlighted the work of N. Wiener, on the modeling of nonlinear systems using Volterra series [164]. A main property of this kind of model is the fact that it is linear with respect to its parameters. Another great advantage is its ability of modeling the behavior of nonlinear real-life phenomena, specially its ability to capture "memory" effects. Due to this characteristic, applications of Volterra models can be encountered in many areas as, for instance, in biological and physiological systems [101, 92, 61], magnetic recording channels [75, 13] and engine transmission modeling [123]. See [61, 65] for applications of Volterra series in different areas. Moreover, Volterra models have many applications in the field of telecommunication, which is the domain of study of this thesis. In the sequel, we discuss the characterization of nonlinear communication channels using Volterra models.

Volterra modeling of nonlinear communication channels

Due to the presence of nonlinear devices such as power amplifiers (PA) and optical instruments, communication channels are sometimes corrupted by nonlinear distortions such as nonlinear intersymbol interference (ISI), nonlinear multiple access interference (MAI) and nonlinear inter-carrier interference (ICI). These nonlinear distortions can significantly deteriorate the signal reception, leading to poor system performance. In such cases, linear models fail to characterize the communication channel, providing inexact channel description. Nonlinear models can then be used to provide an accurate channel representation, allowing the development of efficient signal processing techniques capable of eliminating or reducing these nonlinear distortions. Considering the growing complexity of the current and upcoming communication systems, the development of receivers that extract the transmitted information has become an important and difficult task. In this context, the channel estimation is often an essential task in the design of such receivers.

The choice of the nonlinear model plays then a fundamental role. The Volterra model has since longtime being used to represent communication channels in presence of nonlinear distortions. In this context, one of the most important works is due to Benedetto *et al* [11], for the modeling and performance evaluation of nonlinear satellite communication links using Volterra series. Posterior works of S. Benedetto and E. Biglieri have significantly contributed to the modeling, estimation and equalization of nonlinear channels using Volterra models, mainly in the case of satellite communication channels [9, 14, 10]. In this case, the signals are transmitted from a ground station towards a satellite station and then retransmitted to a ground station, the satellite station usually employing a PA driven at or near saturation to obtain a power efficient transmission. At saturation, the PA exhibits a nonlinear characteristic, resulting in the introduction of nonlinear distortions. Other important works about nonlinear satellite communication channels can be found in [78, 157, 30, 69, 169].

Furthermore, all the systems employing PA are subject to nonlinear distortions. Orthogonal Frequency Division Multiplexing (OFDM) signals are especially vulnerable to PA nonlinear distortions due to their high peak-to-average power ratio (PAPR) [31, 126, 148, 166, 16, 6, 5]. The PAPR of a signal is defined as the ratio of its maximum squared amplitude to the average power [96]. This means that, if the PAPR of a signal is high, the maximal signal power is high compared to the average signal power, i.e. the signal has large envelope fluctuations. In this case, if the PA operates near the saturation region, nonlinear distortions will be introduced.

Volterra series have also important applications in the field of telecommunications

for modeling Radio Over Fiber (ROF) channels [55, 57, 114, 117]. The uplink transmission of such systems is done from a mobile station towards a Radio Access Point (RAP), where the transmitted signals are converted in optical frequencies and then retransmitted through optical fibers to a Central Base Station. Important nonlinear distortions can be introduced by the electrical-optical (E/O) conversion [54, 57].

Other applications of Volterra models in communication systems can be encountered in the literature, e.g. for modeling ultra-wideband (UWB) systems [165, 113], nonlinear acoustic echo paths [140, 139, 8], software-defined radio systems [172], analog to digital converters [76], Code division multiple access (CDMA) systems [23, 127, 26] and optical transmitters [138]. See [61] for more bibliography about applications of nonlinear models in communication systems.

An important phenomenon caused by nonlinear channels is called spectral broadening, which corresponds to spreading the spectrum of the transmitted signal. In fact, for a passband modulated signal, this phenomenon can be viewed as the sum of two different phenomena: the spectral broadening of the signal carrier and envelope. The broadening of the carrier implies that the received signal will have spectral components centered in frequencies different from the transmitted signal carrier. However, this phenomenon is canceled by placing bandpass filters (zonal filters) after the nonlinear devices. On the other hand, the spectral broadening of signal envelope implies that the frequency support of the received signal envelope is higher than the frequency support of the transmitted signal envelope. This may lead to a significant increase of the signal bandwidth. The spectral broadening of the signal envelope can be partially canceled by a bandpass filter. However, sometimes it can be interesting to maintain all the spectral components of the envelope signal in order to exploit this information at the receiver.

Nevertheless, most part of the works dealing with channel estimation and/or information recovery on Volterra communication channels consider the case of single-input single-output (SISO) systems. There are few works dealing with these problems in the context of nonlinear multiple-input-multiple-output (MIMO) communication channels, which is the main subject of this thesis. In what follows, the use of Volterra models in MIMO channels is discussed and the main objectives of this thesis are presented.

MIMO Volterra communication channels

MIMO transmission schemes with multiple transmit and receive antennas are well-known solutions to improve the spectral efficiency and/or the transmission rate

by increasing the spatial diversity [63, 151]. The multiple transmit and receive antenna schemes are very attractive due to the rising demand for transmission bandwidth and they are one of the key technologies to be considered in current and upcoming wireless communication systems [111].

In fact, in this thesis, the term MIMO will also be used to denote a multi-user system with a single transmit antenna for each user and multiple receive antennas. In this case, the antenna array is used for directional signal reception, exploiting the spatial diversity to increase the number of users transmitting at the same time and same frequency band [150, 141, 159, 145]. This kind of technology, also known as beamforming, is used to provide space division multiple access (SDMA). The antenna array optimizes the radio spectrum by taking advantage of the directional properties of antenna array. This kind of transmission scheme is normally used at the base station in uplink, the antenna array providing signals of the highest possible quality to each user.

As well as in SISO communication channels, MIMO channels are subject to nonlinear distortions due to the presence of nonlinear blocks such as radio frequency PA and E/O conversion device. MIMO communication systems subject to nonlinear distortions can be found in OFDM systems [67, 133], multiuser ROF systems [114, 117, 116], satellite systems [117], CDMA systems [127], wireless communication links [148] and in ultra-wide-band systems [113]. In such situations, the received signals are nonlinear mixtures of transmitted signals, possibly including their delayed versions.

In these cases, MIMO Volterra models are interesting tools to model the communication channel. Different versions of MIMO Volterra models can be defined, depending on the generality of the model. Some versions of MIMO Volterra models have been used by a number of authors in different areas [123, 1, 136, 124, 2]. However, to the best of our knowledge, few authors have proposed channel estimation or information recovery techniques for MIMO Volterra models in the context of communications systems [113, 127, 67, 133]. Moreover, the systems considered in [113, 127, 67, 133] do not correspond to the most general form of the MIMO Volterra model studied in this thesis. Channel estimation and equalization techniques for multiple-input-single-output (MISO) Volterra channels were proposed in [114, 117, 116]. On the other hand, nonlinear single-input-multiple-output (SIMO) channels were also treated by some authors [60, 43, 99].

Other kind of models can also be used for modeling nonlinear mixtures as, for instance, post nonlinear (PNL) mixtures [149, 170, 80], constituted of a linear instantaneous mixture followed by memoryless nonlinearities. In order to take realistic phenomena into account, convolutive post nonlinear (CPNL) mixtures

were introduced in [4]. The CPNL mixtures are constituted of a linear convolutive mixture followed by memoryless nonlinearities. When the nonlinearities of PNL and CPNL mixtures are polynomial, they can be viewed as a special case of MIMO Volterra system also known as MIMO Wiener system. Applications of PNL mixtures can be encountered in biomedical data recording [171, 170], sensor array processing [80, 110], etc. Works dealing with blind source separation of nonlinear mixtures using other kinds of mixtures models can be found in [158, 72]. However, it should be highlighted that many of these nonlinear source separation techniques can not be applied to nonlinear MIMO communication channels, because these systems have some characteristics and constraints that must be taken into account. Moreover, important results concerning the inversion of nonlinear MIMO polynomial mixtures were demonstrated in [20, 21].

The general objective of this thesis is to propose new techniques for channel estimation and information recovery in MIMO Volterra communication channels. This is motivated by the fact that the performance limits of approaches based on linear channels have made less advances in the last years than the approaches based on nonlinear channels. Although this seems to be an important and promising research field, there is a lack of works dealing with nonlinear MIMO communication channels. This is probably due to the fact that the development of such techniques is a quite difficult task, since nonlinear MIMO models are, in general, very complex.

In fact, the most part of the techniques proposed in this manuscript are for estimating MIMO Volterra channels. The reason is that, once the channel estimated, the recovery of the transmitted signals can be efficiently carried out by using techniques such as the Viterbi algorithm or the Wiener receiver [118, 74]. An accurate channel estimation is then very important for the performance of the receiver. Nevertheless, some techniques for information recovery in MIMO Volterra channels are also proposed in Chapters 3 and 6 of this thesis.

In the development of the proposed techniques, various approaches are considered. An overview of some of these approaches is then given in the sequel, providing a motivation of their use.

Some approaches considered in this work

All the communication channels considered in this work correspond to series-cascades of linear and nonlinear systems such as the Wiener, Hammerstein and Wiener-Hammerstein models. As shown in Chapter 2, in these cases, it is always possible to obtain a global Volterra representation of the channel. One of the main

advantages of these block-structured nonlinear systems is that they are characterized by less parameters than their global Volterra representations. On the other hand, an important drawback of such kind of models is that they are not linear with respect to their parameters, contrarily to the global Volterra representation.

The signal processing methods developed in this thesis are based on global MIMO Volterra representations of the communication channel, exploiting the fact that the model is linear with respect to its coefficients. This approach is to be used when the compensation of the nonlinear distortions is carried out at the receiver side, which provides other advantages with respect to pre-distortion schemes [39, 31, 148, 38, 120, 81, 3] that compensate the nonlinear distortions at the transmitter side: (i) global optimization of the problem, (ii) in an uplink transmission, the computational complexity associated with the signal processing for compensating the nonlinearities is at the base station and (iii) compensation of other possible channel nonlinearities.

Some of the channel estimation techniques proposed in this thesis make use of a training sequence known by both transmitter and receiver during the acquisition period. In this period, the receiver has access to the transmitted data for estimating the channel by means of a supervised estimation technique. Other approach considered in this manuscript is the blind (or non supervised) channel estimation, which is characterized by the absence of a training sequence. Several of these applications take advantage of this feature to improve spectral efficiency by using the time earlier spent in the training period to transmit information. In some cases the transmission of a training sequence is undesired or even impossible, such as in multi-point computer network and radio-digital transmission on microwave band. As we will see in Chapters 5 and 6, signal processing techniques for MIMO Volterra channels in a blind scenario can be developed by exploiting some kind of redundancy of the received signals. In these chapters, we exploit these redundancies by making use of tensor decompositions. Moreover, the transmission scheme is called semi-blind when the training sequence is composed of only a few pilot symbols at the beginning of the acquisition period, the end of the acquisition period being carried out by a blind technique.

Another interesting point concerning channel estimation and information recovery in nonlinear systems is the use of the phase-shift keying (PSK) modulation. PSK signals provide less nonlinear distortions than quadrature amplitude modulation (QAM) signals due to the fact that PSK symbols have less amplitude fluctuations than QAM symbols, which makes the use of PSK signals interesting for transmissions over nonlinear channels. As we will see in Chapter 5, when the input signals are PSK-modulated, a Volterra system can be rewritten with a smaller number of coefficients. The performance of PSK signals over nonlinear satellite channels was

investigated by several authors, e.g. [9, 78]. In fact, the properties of nonlinearly distorted PSK signals established in [100] have motivated the use of PSK signals in Chapters 5 and 6.

Thesis content

This thesis is divided into five chapters, the first one containing a survey of the main concepts and models exploited throughout the thesis, and the others containing the different contributions. The organization of the thesis is illustrated in Fig. 1.1, where the links between the chapters, applications, type of MIMO Volterra models and used approaches are indicated. The chapters are organized according to the type of Volterra model and the considered application. In fact, the approach and the type of MIMO Volterra model used in each chapter depend on the considered communication system.

Chapter 2, entitled **Volterra Models for Nonlinear MIMO Communication Systems**, provides an overview of MIMO Volterra models in the context of nonlinear communication systems, containing the basic material to be exploited throughout the thesis. The development of an equivalent baseband representation of a discrete-time SISO Volterra system is carried out. These developments constitute the basis for the discrete-time equivalent baseband MIMO Volterra systems presented in the sequel, which are the models used through this work to model MIMO communication channels. Some important block structured nonlinear models are also described, with their link to the MIMO Volterra system being developed. The main nonlinear MIMO communication systems considered in this thesis are discussed at the end of the chapter. This chapter presents Volterra system models in a way to put in evidence the approaches and methods considered for each communication system exploited in the following chapters, as shown in Fig. 1.1. Moreover, this chapter contains two theoretical contributions concerning the modeling of nonlinear communication channels by MIMO Volterra systems.

In **Chapter 3**, entitled **Estimation and Equalization of Nonlinear MIMO-OFDM Channels**, the MIMO Volterra models are applied to the estimation and equalization of nonlinear MIMO-OFDM channels. In this case, as the nonlinearity is introduced by the transmitter PAs, the communication channel is modeled as a special case of MIMO Volterra systems that does not contain products of different sources, i.e. the multiple sources are not nonlinearly mixed. This kind of model can be viewed as a parallel-cascade of multiple SIMO Volterra systems.

Two different models for the PA are considered: a memoryless polynomial model and a memory polynomial model, also known as diagonal Volterra model [38, 119,

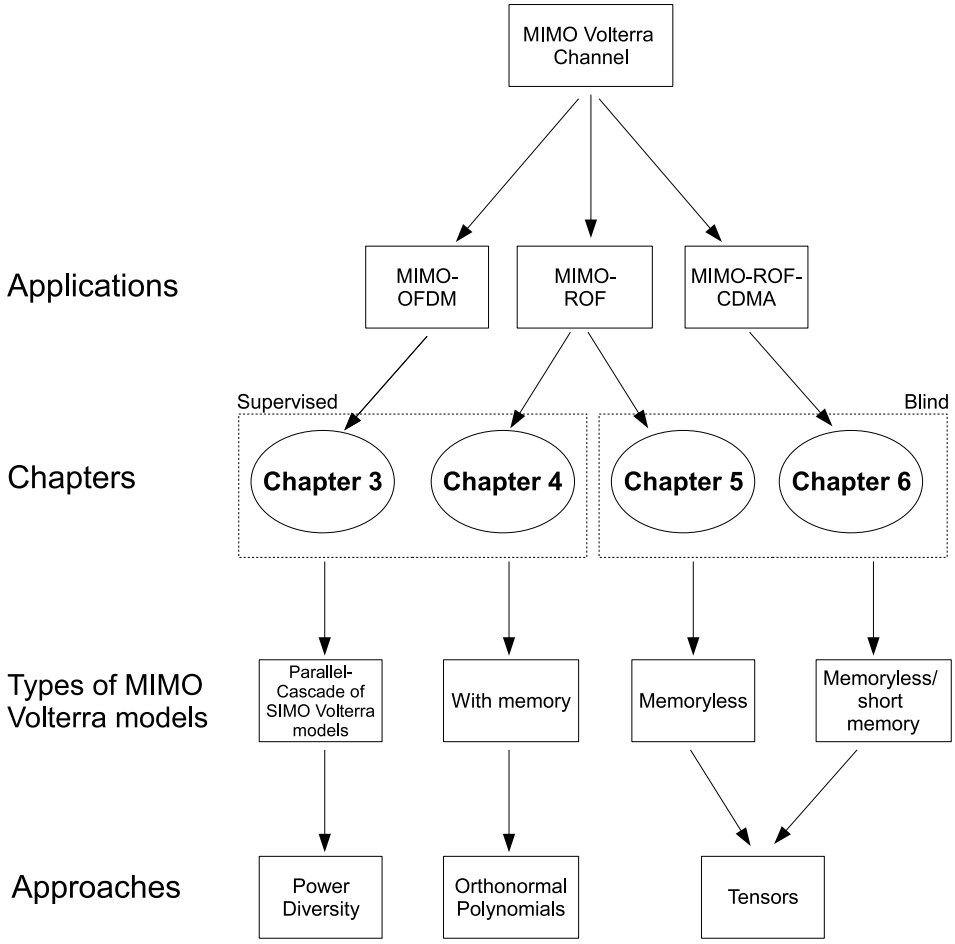

Figure 1.1: Links between the chapters, applications, types of MIMO Volterra models and used approaches.

39, 102, 66, 16, 40]. A technique for supervised channel estimation is first proposed for the memoryless polynomial PA case. Channel equalization techniques are then developed based on a transmission scheme that re-transmits all the symbols several times with a different transmission power each time. Due to the nonlinear nature of the channel, the redundancy added by this transmission scheme can be exploited to provide multi-access at the reception and, consequently, to recover the transmitted signals.

It is also demonstrated that when the PAs are modeled as memory polynomials, the received signals can be expressed in the same way as in the case of memoryless polynomial PAs. As a result, the channel estimation and equalization techniques proposed for polynomial PAs can be directly applied in the case of memory poly-

nomial PAs.

Chapter 4, entitled **Supervised Estimation of MIMO Volterra Channels Using Orthonormal Polynomials**, is dedicated to the development of estimation techniques for MIMO Volterra channels in a supervised scenario, considering time and space division multiple access (TDMA-SDMA) systems. This chapter considers the most general type of MIMO Volterra systems used in this thesis, where the received signals are nonlinear mixtures of the sources, with no constraint in the channel memory.

In adaptive schemes, the Least Mean Square (LMS) algorithm generally exhibits a slow convergence speed when identifying Volterra models, as the covariance matrix of the regression vector has, in general, a high eigenvalue spread when the input signals are QAM modulated [14]. In order to improve the conditioning of this covariance matrix and overcome this problem, the use of orthonormal polynomials is an efficient solution used by some authors in the SISO case [14, 108, 105]. In Chapter 4, we extend the procedure of construction and the use of these polynomials to the case of MIMO Volterra systems, allowing different probability density functions (PDFs) for the input signals and different memories with respect to the inputs. The supervised channel estimation technique proposed in this chapter is applied to the estimation of multisuer nonlinear ROF channels.

Chapter 5, entitled **Blind Estimation of Memoryless MIMO Volterra Channels Using Tensor Decomposition and Precoding**, also presents estimation methods for MIMO Volterra communication channels in the context of TDMA-SDMA systems. However, the techniques proposed in this chapter consider a blind scenario and treat the case of memoryless channels using tensor decompositions. By exploiting the fact that Volterra models are linear with respect to their coefficients, the blind estimation of MIMO Volterra channels is carried out by means of the Parallel Factor (PARAFAC) decomposition, also known as Canonical Decomposition (CANDECOMP). In fact, these methods are based on the PARAFAC decomposition of a tensor composed of channel output covariances. Such a decomposition is possible owing to a new precoding scheme developed for PSK signals modeled as Markov chains. Some conditions on the transition probability matrices (TPMs) of the Markov chains are established to introduce temporal correlation and satisfy statistical correlation constraints inducing the PARAFAC decomposition of the considered tensor.

A great advantage of using the PARAFAC decomposition is that it allows the blind channel identification and information recovery when the number of receive antennas is smaller than the number of terms of the Volterra series, contrarily to some previous works [127, 43, 113]. This is particularly interesting since the

Volterra filters may have a large number of parameters. Indeed, working with a number of receive antennas higher than or equal to the number of terms of the Volterra series imposes a strong constraint on the number of antennas to be used. Moreover, PARAFAC decomposition avoids the use of a pre-whitening step, an operation that increases the computational complexity and may degrade the channel estimation [168]. The channel estimation techniques proposed in this chapter are also applied to the estimation of multisuer nonlinear ROF channels.

Chapter 6, entitled **Estimation and Equalization of MIMO Volterra Channels in CDMA systems**, proposes various techniques for channel estimation and information recovery in MIMO Volterra systems in the context of a CDMA communication system based on the use of tensor decompositions. Indeed, signal processing techniques that use the PARAFAC decomposition are particularly interesting when used with CDMA systems.

This chapter deals with memoryless communication channels, as well as with short memory channels. In the memoryless case, a technique for joint blind channel estimation and information recovery is developed based on the PARAFAC decomposition of a third-order tensor composed of received signals, by exploring space, time and code diversities. Still in the case of a memoryless channel, a blind estimation technique is developed based on the PARAFAC decomposition of a fifth-order tensor composed of covariances of the received signals, considering that the transmitted signals are PSK modulated. Moreover, a blind identification method for a MIMO Volterra channel with short memory in a CDMA system is proposed. This method, based on the PARAFAC decomposition of a third-order tensor of channel output covariances, can be viewed as an extension of the ones developed in Chapter 5 for channels with short memory.

Thus, it should be highlighted that this thesis covers channel estimation and information recovery in different kinds of MIMO Volterra systems. Note that the techniques proposed in Chapters 3 and 4 are supervised, and the ones proposed in Chapters 5 and 6 are blind. In fact, as we will see later, Chapters 5 and 6 also propose some methods based on semi-blind transmission schemes. Note also that the proposed techniques use different kinds of approaches, according to the considered problem. At last, it can be viewed that the proposed techniques are applied to three different types of communication systems: wireless OFDM, ROF and CDMA-ROF systems.

Main contributions

The main contributions of this thesis can be summarized as follows:

Chapter 2

- Development of general expressions for equivalent baseband discrete-time MIMO Volterra channels (Section 2.2).
- Establishment of relationships between MIMO Wiener, Hammerstein and Wiener-Hammerstein models, and the MIMO Volterra model (Section 2.3).

Chapter 3

- Proposition of a technique for supervised estimation of SISO- and MIMO-OFDM channels with memoryless polynomial PAs (Sections 3.2.2 and 3.3.1).
- Proposition of Zero Forcing (ZF) and Minimum Mean Square Error (MMSE) receivers for SISO- and MIMO-OFDM channels with memoryless polynomial PAs (Sections 3.2.3 and 3.3.3.)
- Theorem 3.1 stating that a memory polynomial PA in a OFDM system can be expressed as a polynomial PA with coefficients that vary from one subcarrier to another (Section 3.4).
- Development of a link between the frequency domain received signals in a MIMO-OFDM system in terms of the global channel parameters in the case of memoryless and memory polynomial PAs (Section 3.4).

Chapter 4

- Development of a basis of orthonormal polynomials for equivalent baseband MIMO Volterra systems, allowing different probability density functions for the input signals and different memories with respect to the inputs (Section 4.3).
- Application of the proposed basis of orthonormal polynomials to the estimation of uplink nonlinear MIMO ROF channels (Sections 4.4 and 4.5).

Chapter 5

- Development of orthogonality constraints inducing the PARAFAC decomposition of a third-order tensor of received signal covariances in a memoryless MIMO Volterra channel (Sections 5.2 and 5.3).
- Proposition of a precoding scheme based on the use of discrete time Markov chains (DTMCs) so that the transmitted signals satisfy these orthogonality constraints (Section 5.4).

- Proposition of two techniques for blind estimation of memoryless MIMO Volterra channels, based on the use of the Alternating Least Squares (ALS) algorithm and a Joint Diagonalization Algorithm (JDA) (Section 5.5).

Chapter 6

- Proposition of two deterministic techniques for joint estimation and equalization of memoryless MIMO Volterra channels in a CDMA system (Section 6.1).
- Proposition of three stochastic techniques for blind estimation of memoryless MIMO Volterra channels in a CDMA system (Section 6.2).
- Development of a stochastic approach for blind estimation of MIMO Volterra channels with short memory in a CDMA system (Section 6.3).
- Application of the proposed techniques in uplink nonlinear MIMO-CDMA ROF channels (Section 6.4).

Publications

This thesis has originated the following publications:

Journal Papers

1. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "Estimation and Equalization of MIMO-OFDM channels with nonlinear Power Amplifiers", IEEE Transactions on Signal Processing, 2009. *To be Submitted*.
2. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "PARAFAC-Based Estimation and Equalization of Nonlinear Instantaneous MIMO Channels in CDMA Systems", Signal Processing, 2009. *To be Submitted*.
3. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "Blind Identification of Multiuser Nonlinear Channels Using Tensor Decomposition and Precoding", Signal Processing. *To Appear in 2009*.
4. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "Decision Directed Adaptive Blind Equalization Based on the Constant Modulus Algorithm", Signal, Image and Video Processing (SIViP), v. 1, n. 4, p. 333-346, Oct. 2007.

Conference papers

5. P. A. C. Aguilar, C. A. R. Fernandes, J. C. M. Mota and G. Favier, "Estimation and Equalization of OFDM channels with nonlinear memoryless Power Amplifiers", Brazilian Telecommunications Symposium (SBrT), Blumenau, Brazil, Sep. 29 - Oct. 2, 2009. *Submitted*

6. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "Tensor Based Receivers for Nonlinear Radio Over Fiber Uplinks in Multiuser CDMA Systems", IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Cannes, France, Sep. 15-18, 2008.

7. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "Tensor-Based Blind Identification of MIMO Volterra Channels in a Multiuser CDMA Environment", European Signal Processing Conference (EUSIPCO), Lausanne , Switzerland, Aug. 25-29, 2008.

8. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "Blind Estimation of Nonlinear Instantaneous Channels in Multiuser CDMA systems with PSK inputs", IEEE Signal Processing Advances in Wireless Communications (SPAWC) workshop, Recife, Brazil, Jul. 6-9, 2008.

9. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "A Modulation Code-Based Blind Receiver for Memoryless Multiuser Volterra Channels", ASILOMAR Conference on Signals, Systems, and Computers, Pacific Grove, CA, US, Nov. 4-7, 2007.

10. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "Input Orthogonalization Methods for Third-Order MIMO Volterra Channel Identification", Colloque GRETSI, Troyes, France, Sep. 11-14, 2007.

11. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "Blind Tensor-Based Identification of Memoryless Multiuser Volterra Channels Using SOS and Modulation Codes", European Signal Processing Conference (EUSIPCO), Poznan, Poland , Sep. 3-7, 2007.

12. C. A. R. Fernandes, G. Favier and J. C. M. Mota, "Blind Source Separation and Identification of Nonlinear Multiuser Channels using Second Order Statistics and Modulation Codes", IEEE Signal Processing Advances in Wireless Communications (SPAWC) workshop, Helsinki, Finland, Jun. 17-20, 2007.

13. C. A. R. Fernandes, A. Kibangou, G. Favier and J. C. M. Mota, "Identification of Nonlinear MIMO Radio Over Fiber Uplink Channels", International Telecommunications Symposium (ITS), Fortaleza, Brazil, Sep. 3-6, 2006.

The relationships between these publications and chapters of the thesis is the following: publications 1 and 5 are related to Chapter 3, publications 10 and 13 to Chapter 4, publications 3, 9, 10, 11 and 12 to Chapter 5, and publications 2, 6, 7 and 8 to Chapter 6. Publication 4 was developed during the period of the thesis, but its results are not included in this thesis. Its results have motivated some of the works of the present thesis.

CHAPTER 2

Volterra Models for Nonlinear MIMO Communication Channels

THIS thesis is fundamentally based on the development of techniques for channel estimation and information recovery in multiple-input multiple-output (MIMO) Volterra communication channels. This chapter presents a survey of the Volterra models used in this thesis for modeling nonlinear MIMO communication channels. As mentioned earlier, the multiple-inputs of these channels represent various sources transmitting at the same time and frequency band, which can correspond to multiple users or a single user with multiple transmit antennas. On the other hand, multiple-outputs represent the observations at the receiver, obtained through an antenna array. In fact, the multiple observations at the receiver can also be obtained by oversampling the received signals, however, this approach is not considered in this thesis.

We are particularly interested in three different applications of nonlinear MIMO communication channels: Orthogonal Frequency Division Multiplexing (OFDM), Radio-Over-Fiber (ROF) and ROF-Code division multiple access (ROF-CDMA) systems, which constitute important technologies used in current and upcoming telecommunication systems. An interesting issue about these systems is that their communication channels can be modeled as block-structured nonlinear systems, i.e. series-cascades of nonlinear and linear blocks, such as MIMO Wiener, Hammerstein and Wiener-Hammerstein models. As it will be shown later in the chapter, this kind of systems can be viewed as special cases of Volterra systems.

However, for simplifying the presentation of Volterra models, we first consider

single-input single-output (SISO) communication channels. By starting from an expression for the continuous-time passband received signals, we develop an expression for the continuous-time equivalent baseband received signals. In order to show the broadening of the spectral support provided by the Volterra system on the equivalent baseband received signals, the frequency domain representation of the Volterra channel is developed. The expression for the discrete-time equivalent baseband received signals is also introduced, constituting the basis for all the MIMO Volterra models used in this work. It is important to highlight that the signal processing techniques developed in this thesis exploit the fact that a Volterra model is linear with respect to its parameters. Due to this, vector representations for the discrete-time received signals that explicit this property are also introduced.

The rest of this chapter is organized as follows. Section 2.1 presents the single-input-single-output (SISO) communication channels modeled by Volterra models. In Section 2.2, the discrete-time equivalent baseband MIMO Volterra models are introduced. Section 2.3 describes some systems constituted of series-cascades of linear and nonlinear blocks. Section 2.4 describes some applications of MIMO Volterra models in communication systems and some conclusions about the present chapter are drawn in Section 2.5.

2.1 Volterra communication channels

In this section, we introduce the main Volterra systems used for modeling SISO communication channels, as well as some of their properties. The output $\check{x}(\xi)$ of a real-valued continuous-time SISO Volterra system of finite order $\check{K}$ can be represented by the following relationship:

$$\check{x}(\xi) = \sum_{k=1}^{\check{K}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \check{h}_k(\tau_1, \ldots, \tau_k) \prod_{i=1}^{k} \check{s}(\xi - \tau_i) d\tau_i \ , \tag{2.1}$$

where ξ is the continuous-time variable, $\check{h}_k(\tau_1, \ldots, \tau_k)$ is the real-valued continuous-time Volterra kernel of order k and $\check{s}(\xi)$ is the real-valued continuous-time input signal. Regarding the signals and systems considered through this thesis, the following assumptions will be considered in all the developments of this work:

(A1): The transmitted signals are stationary.

(A2): The systems are causal and time-invariant.

One of the main advantages of the Volterra model is that a wide range of practical physical systems can be approximated by (2.1). In particular, any finite memory nonlinear system satisfying A2 can be approximated using (2.1) [17]. Moreover, it must be remarked that, when $\check{K} = 1$, the Volterra model is equivalent to the convolution of input signal with the linear kernel $\check{h}_1(\tau_1)$.

A Volterra kernel $\check{h}_k(\tau_1, \ldots, \tau_k)$ is said to be symmetric if it is invariant to a permutation of the indices $\tau_1, \ldots, \tau_k$. Thus, as a permutation of the indices $\tau_1, \ldots, \tau_k$ does not change the product $\prod_{i=1}^{k} \check{s}(\xi - \tau_i)$, an asymmetric Volterra kernel can always be rewritten as a symmetric kernel [134]. For instance, let us consider a homogeneous Volterra system of order 2, i.e. a Volterra system containing only the quadratic kernel:

$$\check{x}(\xi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{h}_2(\tau_1, \tau_2) \check{s}(\xi - \tau_1) \check{s}(\xi - \tau_2) d\tau_1 d\tau_2 \ , \tag{2.2}$$

where the kernel $\bar{h}_2(\tau_1, \tau_2)$ is non-symmetric, i.e. $\bar{h}_2(\tau_1, \tau_2) \neq \bar{h}_2(\tau_2, \tau_1)$. Equation (2.2) can be rewritten as

$$\check{x}(\xi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \check{h}_2(\tau_1, \tau_2) \check{s}(\xi - \tau_1) \check{s}(\xi - \tau_2) d\tau_1 d\tau_2 \ , \tag{2.3}$$

with $\check{h}_2(\tau_1, \tau_2) = (\bar{h}_2(\tau_1, \tau_2) + \bar{h}_2(\tau_2, \tau_1))/2$ being symmetric. Thus, in this work, with no loss of generality, we will consider symmetric kernels.

In the case of a communication channel, $\check{x}(\xi)$ and $\check{s}(\xi)$ can be viewed respectively as the bandpass received and transmitted signals. In the sequel, we derive the continuous- and discrete-time equivalent baseband representations of the Volterra system (2.1).

2.1.1 The equivalent baseband Volterra channel

The digital signal processing techniques of this thesis are based on discrete-time equivalent baseband representations of the received signals, which are obtained by sampling the continuous-time equivalent baseband received signals. The development of a continuous-time equivalent baseband Volterra system is then needed. A real-valued continuous-time bandpass signal $\check{s}(\xi)$ and its continuous-time equiva-

lent baseband version $s(\xi)$ are related by the following expression:

$$\begin{aligned}\check{s}(\xi) &= \mathfrak{Re}\left\{s(\xi)e^{\jmath\,2\pi f_c\xi}\right\} \\ &= \frac{1}{2}\left[s(\xi)e^{\jmath\,2\pi f_c\xi} + s^*(\xi)e^{-\jmath\,2\pi f_c\xi}\right], \end{aligned} \tag{2.4}$$

where $\jmath = \sqrt{-1}$ is the imaginary unit and f_c the carrier frequency. The baseband signal $s(\xi)$, also known as complex-envelope, is assumed to be bandlimited, i.e. the Fourier transform and the power spectral density of $s(\xi)$ have finite support, vanishing outside $I = [-B, B]$. Moreover, it is assumed that B is much smaller than f_c ($B \ll f_c$).

Initially, let us consider a homogeneous Volterra system of order 2. Substituting (2.4) into (2.3), we get:

$$\begin{aligned}\check{x}(\xi) = \frac{1}{4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\check{h}_2(\tau_1,\tau_2)\left[s(\xi-\tau_1)e^{\jmath\,2\pi f_c(\xi-\tau_1)} + s^*(\xi-\tau_1)e^{-\jmath\,2\pi f_c(\xi-\tau_1)}\right] \\ \left[s(\xi-\tau_2)e^{\jmath\,2\pi f_c(\xi-\tau_2)} + s^*(\xi-\tau_2)e^{-\jmath\,2\pi f_c(\xi-\tau_2)}\right] d\tau_1 d\tau_2, \end{aligned} \tag{2.5}$$

which gives:

$$\begin{aligned}\check{x}(\xi) &= \frac{1}{4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\check{h}_2(\tau_1,\tau_2)\left[s(\xi-\tau_1)s(\xi-\tau_2)e^{\jmath\,2\pi(2f_c)\xi}e^{\jmath\,2\pi f_c(-\tau_1-\tau_2)}\right. \\ &+ s(\xi-\tau_1)s^*(\xi-\tau_2)e^{\jmath\,2\pi f_c(-\tau_1+\tau_2)} + s^*(\xi-\tau_1)s(\xi-\tau_2)e^{\jmath\,2\pi f_c(\tau_1-\tau_2)} \\ &+ \left. s^*(\xi-\tau_1)s^*(\xi-\tau_2)e^{-\jmath\,2\pi(2f_c)\xi}e^{\jmath\,2\pi f_c(\tau_1+\tau_2)}\right]. \end{aligned} \tag{2.6}$$

At the receiver, the signal $\check{x}(\xi)$ is filtered by a bandpass filter centered in f_c. Thus, if $B \ll f_c$, all the frequency components not centered at f_c are assumed to be suppressed by the bandpass filter[1]. As we can see in (2.6), the frequency components of $\check{x}(\xi)$ are centered at the frequencies 0 and $2f_c$. This means that the received signal is equal to zero after the bandpass filter. In fact, it can be demonstrated that all the spectral components generated by the even-order kernels are not centered at the carrier frequency f_c [77].

Similarly as (2.6), the output of a homogeneous Volterra system of order 3, i.e. a

[1]In fact, bandpass filters are also placed after nonlinear devices such as power amplifiers, suppressing all the frequency components lying outside the passband of the filter at this stage of the transmission.

Volterra system containing only the cubic kernel, can be rewritten as:

$$\begin{aligned}\check{x}(\xi) = \frac{1}{8}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\check{h}_3(\tau_1,\tau_2,\tau_3)\left[s(\xi-\tau_1)e^{\jmath 2\pi f_c(\xi-\tau_1)}+s^*(\xi-\tau_1)e^{-\jmath 2\pi f_c(\xi-\tau_1)}\right]\\ \left[s(\xi-\tau_2)e^{\jmath 2\pi f_c(\xi-\tau_2)}+s^*(\xi-\tau_2)e^{-\jmath 2\pi f_c(\xi-\tau_2)}\right]\\ \left[s(\xi-\tau_3)e^{\jmath 2\pi f_c(\xi-\tau_3)}+s^*(\xi-\tau_3)e^{-\jmath 2\pi f_c(\xi-\tau_3)}\right]d\tau_1 d\tau_2\tau_3, \quad (2.7)\end{aligned}$$

or,equivalently,

$$\begin{aligned}\check{x}(\xi) &= \frac{1}{8}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\check{h}_3(\tau_1,\tau_2,\tau_3)\\ &\quad\left[s(\xi-\tau_1)s(\xi-\tau_2)s(\xi-\tau_3)e^{\jmath 2\pi(3f_c)\xi}e^{\jmath 2\pi f_c(-\tau_1-\tau_2-\tau_3)}\right.\\ &+ s(\xi-\tau_1)s(\xi-\tau_2)s^*(\xi-\tau_3)e^{\jmath 2\pi f_c\xi}e^{\jmath 2\pi f_c(-\tau_1-\tau_2+\tau_3)}\\ &+ s(\xi-\tau_1)s^*(\xi-\tau_2)s(\xi-\tau_3)e^{\jmath 2\pi f_c\xi}e^{\jmath 2\pi f_c(-\tau_1+\tau_2-\tau_3)}\\ &+ s(\xi-\tau_1)s^*(\xi-\tau_2)s^*(\xi-\tau_3)e^{-\jmath 2\pi f_c\xi}e^{\jmath 2\pi f_c(-\tau_1+\tau_2+\tau_3)}\\ &+ s^*(\xi-\tau_1)s(\xi-\tau_2)s(\xi-\tau_3)e^{\jmath 2\pi f_c\xi}e^{\jmath 2\pi f_c(\tau_1-\tau_2-\tau_3)}\\ &+ s^*(\xi-\tau_1)s(\xi-\tau_2)s^*(\xi-\tau_3)e^{-\jmath 2\pi f_c\xi}e^{\jmath 2\pi f_c(\tau_1-\tau_2+\tau_3)}\\ &+ s^*(\xi-\tau_1)s^*(\xi-\tau_2)s(\xi-\tau_3)e^{-\jmath 2\pi f_c\xi}e^{\jmath 2\pi f_c(\tau_1+\tau_2-\tau_3)}\\ &+ \left.s^*(\xi-\tau_1)s^*(\xi-\tau_2)s^*(\xi-\tau_3)e^{-\jmath 2\pi(3f_c)\xi}e^{\jmath 2\pi f_c(\tau_1+\tau_2+\tau_3)}\right]d\tau_1 d\tau_2\tau_3. \quad (2.8)\end{aligned}$$

Taking the symmetry of the kernel $\check{h}_3(\tau_1,\tau_2,\tau_3)$ into account and noting that the first and the last terms inside the brackets in (2.8) are centered at $3f_c$ and, consequently, suppressed by bandpass filtering, we can write the following expression for the bandpass received signal after the bandpass filter:

$$\begin{aligned}\bar{x}(\xi) = \frac{3}{8}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\check{h}_3(\tau_1,\tau_2,\tau_3)\left[s(\xi-\tau_1)s(\xi-\tau_2)s^*(\xi-\tau_3)e^{\jmath 2\pi f_c(-\tau_1-\tau_2+\tau_3)}\right.\\ \left.e^{\jmath 2\pi f_c\xi}+s^*(\xi-\tau_1)s^*(\xi-\tau_2)s(\xi-\tau_3)e^{\jmath 2\pi f_c(\tau_1+\tau_2-\tau_3)}e^{-\jmath 2\pi f_c\xi}\right]d\tau_1 d\tau_2\tau_3, \quad (2.9)\end{aligned}$$

where we assumed that the bandpass filter is perfectly flat at the passband. Thus,

$$\begin{aligned}\bar{x}(\xi) = \mathfrak{Re}\left\{\left(\frac{3}{4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\check{h}_3(\tau_1,\tau_2,\tau_3)s(\xi-\tau_1)s(\xi-\tau_2)s^*(\xi-\tau_3)\right.\right.\\ \left.\left.e^{\jmath 2\pi f_c(-\tau_1-\tau_2+\tau_3)}d\tau_1 d\tau_2\tau_3\right)e^{\jmath 2\pi f_c\xi}\right\}. \quad (2.10)\end{aligned}$$

The equivalent baseband received signal is then given by:

$$x(\xi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h_3(\tau_1,\tau_2,\tau_3)s(\xi-\tau_1)s(\xi-\tau_2)s^*(\xi-\tau_3)d\tau_1 d\tau_3\tau_3, \quad (2.11)$$

where

$$h_3(\tau_1,\tau_2,\tau_3)=\frac{3}{4}\check{h}_3(\tau_1,\tau_2,\tau_3)e^{\jmath\,2\pi f_c(-\tau_1-\tau_2+\tau_3)}. \tag{2.12}$$

In a similar way, it can be shown that the equivalent baseband received signal for a Volterra system of order $2K+1$ can be written as [77]:

$$x(\xi)=\sum_{k=0}^{K}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}h_{2k+1}(\tau_1,\ldots,\tau_{2k+1})\prod_{i=1}^{k+1}s(\xi-\tau_i)\prod_{i=k+2}^{2k+1}s^*(\xi-\tau_i)\,d\tau_1 d\tau_2\ldots d\tau_{2k+1}, \tag{2.13}$$

where the equivalent baseband kernel is given by:

$$h_{2k+1}(\tau_1,\ldots,\tau_{2k+1})=\frac{C_{2k+1,k}}{2^{2k}}\;\check{h}_{2k+1}(\tau_1,\ldots,\tau_{2k+1})e^{\jmath 2\pi f_c(-\sum_{i=1}^{k+1}\tau_i+\sum_{i=k+2}^{2k+1}\tau_i)}. \tag{2.14}$$

with $C_{2k+1,k}$ defined in (A.6). Three characteristics of the equivalent baseband Volterra system (2.13) should be highlighted. The first one is that it includes only the odd-order kernels with one more non-conjugated term than conjugated terms. The second one is that the Volterra coefficients (2.14) are complex-valued, while the bandpass Volterra coefficients in (2.1) are real-valued. The last one is that, due to the asymmetry of indices on the phase of the exponential term in (2.14): $\varphi(\tau_1,...,\tau_{2k+1})=-\sum_{i=1}^{k+1}\tau_i+\sum_{i=k+2}^{2k+1}\tau_i$, the equivalent baseband Volterra kernels $h_{2k+1}(\tau_1,\ldots,\tau_{2k+1})$ are not symmetric.

2.1.2 Fourier transform of a Volterra channel output

The understanding of the behavior of a nonlinear system can be improved if the spectrum of the output signal is computed. In the sequel, the frequency domain representation of the equivalent baseband Volterra system (2.13) is developed. Using the inverse Fourier transform formula, we may rewrite (2.13) as:

$$x(\xi)=\sum_{k=0}^{K}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}h_{2k+1}(\tau_1,\ldots,\tau_{2k+1})\prod_{i=1}^{k+1}\left(\int_{-\infty}^{\infty}S(f_i)e^{\jmath\,2\pi f_i(\xi-\tau_i)}df_i\right)\prod_{i=k+2}^{2k+1}\left(\int_{-\infty}^{\infty}S^*(-f_i)e^{\jmath\,2\pi f_i(\xi-\tau_i)}df_i\right)d\tau_1\ldots d\tau_{2k+1}, \tag{2.15}$$

where $S(f)$ denotes the Fourier transform of $s(\xi)$. Denoting by $H_{2k+1}(f_1,\ldots,f_{2k+1})$ the multidimensional Fourier transform of $h_{2k+1}(\tau_1,\ldots,\tau_{2k+1})$ given by:

$$H_{2k+1}(f_1,\ldots,f_{2k+1}) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_{2k+1}(\tau_1,\ldots,\tau_{2k+1}) e^{-\jmath 2\pi\left(\sum_{i=1}^{2k+1} f_i\tau_i\right)} d\tau_1\ldots d\tau_{2k+1}, \quad (2.16)$$

equation (2.15) can be expressed as:

$$x(\xi) = \sum_{k=0}^{K}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} H_{2k+1}(f_1,\ldots,f_{2k+1})\prod_{i=1}^{k+1} S(f_i)\prod_{i=k+2}^{2k+1} S^*(-f_i) e^{\jmath 2\pi\left(\sum_{i=1}^{2k+1} f_i\right)\xi} df_1\ldots df_{2k+1}. \quad (2.17)$$

Defining $\upsilon_i = \upsilon_{i-1} + f_i$, for $i = 1, 2, ..., 2k+1$, with $\upsilon_0 = 0$, (2.17) can be written as:

$$x(\xi) = \int_{-\infty}^{\infty}\left(\sum_{k=0}^{K}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} H_{2k+1}(\upsilon_1,\upsilon_2-\upsilon_1,\ldots,\upsilon_{2k+1}-\upsilon_{2k}) \prod_{i=1}^{k+1} S(\upsilon_i-\upsilon_{i-1})\prod_{i=k+2}^{2k+1} S^*(-(\upsilon_i-\upsilon_{i-1}))d\upsilon_1\ldots d\upsilon_{2k}\right) e^{\jmath 2\pi\upsilon_{2k+1}\xi} d\upsilon_{2k+1}, \quad (2.18)$$

where we can recognize the inverse Fourier transform formula. Thus, we have:

$$X(f) = \sum_{k=0}^{K}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} H_{2k+1}(\upsilon_1,\upsilon_2-\upsilon_1,\ldots,f-\upsilon_{2k}) \prod_{i=1}^{k+1} S(\upsilon_i-\upsilon_{i-1})\prod_{i=k+2}^{2k} S^*(-(\upsilon_i-\upsilon_{i-1}))S^*(-(f-\upsilon_{2k}))d\upsilon_1\ldots d\upsilon_{2k}, \quad (2.19)$$

where $X(f)$ denotes the Fourier transform of $x(\xi)$ and υ_{2k+1} was replaced by f for the sake of simplifying the notation.

To get a better understanding of (2.19), let us consider a linear-cubic Volterra system ($K = 1$) and use the fact that $S(f)$ is zero outside $I = [-B, B]$:

$$X(f) = H_1(f)S(f) + \int_{-B}^{B}\int_{-2B}^{2B} H_3(\upsilon_1,\upsilon_2-\upsilon_1,f-\upsilon_2)S(\upsilon_1)S(\upsilon_2-\upsilon_1) S^*(-(f-\upsilon_2))d\upsilon_1 d\upsilon_2. \quad (2.20)$$

It can be viewed in (2.19) that if the Volterra system changes the phase of the transmitted signal frequency spectrum, a memory will then be introduced. Moreover, note that the frequency support of $X(f)$ is equal to $[-3B, 3B]$, which means that the output signal $x(\xi)$ may have spectral components outside the frequency support of the input signal $I = [-B, B]$.

It should be noted that a signal is never truly bandlimited in practice because a bandlimited signal would require infinite time support. Furthermore, due to causality, a system can not be truly bandlimited in practice neither. Thus, the bandwidth of a signal is usually considered as the range of frequencies where its Fourier transform has a power above a certain threshold. The signal bandwidth is then understood to be the width of the frequency range where the main part of its power is located. This means that, when the spectral broadening provided by the Volterra system is not important, the signal bandwidth may not change significantly.

Concerning the bandwidth of the bandpass filter considered in Section 2.1.1, two cases can be considered. The first one is when the bandpass filter bandwidth is wide enough to cover the spectral broadening provided by the Volterra system [78, 3]. In this case, it is generally assumed that the bandpass filter causes no significant signal distortion in the components centered at the frequency f_c, in such a way that this filter can be considered as transparent with respect to the equivalent baseband input signal. The second case is when the bandpass filter bandwidth is not large enough to cover the spectral broadening, i.e. the bandpass filter partially rejects the nonlinear interference introduced by the Volterra filter at the frequency f_c [9, 12, 68, 147]. In this case, the bandpass filter can not be considered as transparent with respect to the equivalent baseband input signals. In this thesis, we consider the first case.

2.1.3 Discrete-time equivalent baseband Volterra channel

In the sequel, a discrete-time representation of the equivalent baseband Volterra channel studied in Section 2.1.1 is developed. Discrete-time equivalent baseband Volterra models were developed in the context of communication systems for modeling bandwidth limited channels [11, 77, 9, 14, 137, 24].

We consider that the received signal in (2.13) is sampled with a rate of W, assumed to be higher than or equal to the Nyquist rate $2B$ of the input signal. Thus, assuming perfect synchronization, we get $x(n) = x(\xi)|_{\xi=n/W}$ $(n = 0, 1, ...)$. As pointed out earlier, nonlinear systems increase signal bandwidth, which means that it is not possible to reconstruct $x(\xi)$ from $x(n)$ using a sampling rate of

$W = 2B$. However, it was demonstrated that the Nyquist rate $2B$ of the input signal is sufficient to identify and compensate nonlinear systems [154, 104]. The input signal sampled at the Nyquist rate $W = 2B$ leads to [109]:

$$S(f) = \sum_{n=-\infty}^{\infty} s(n) e^{-\jmath 2\pi n f/2B}. \tag{2.21}$$

In the sequel, the Volterra kernel $h_{2k+1}(\tau_1, \ldots, \tau_{2k+1})$ is assumed to be bandlimited, i.e. $H(f_1, ..., f_{2k+1}) = 0$ for $|f_i| > B$, $\forall i = 1, 2, ..., 2k+1$. In fact, from (2.19), it can be viewed that the form of $H(f_1, ..., f_{2k+1})$ outside the hypercube $I \times \cdots \times I$ ($I = [-B, B]$) is of no consequence since $S(f)$ vanishes outside I [104]. Thus, sampling the Volterra kernel $h_{2k+1}(\tau_1, \ldots, \tau_{2k+1})$ at the first dimension with the Nyquist rate of input signal as:

$$h_{2k+1}(n, \tau_2, \ldots, \tau_{2k+1}) = \frac{1}{2B}\, h_{2k+1}(\tau_1, \tau_2, \ldots, \tau_{2k+1})|_{\tau_1 = \frac{n}{2B}}\,, \tag{2.22}$$

leads to:

$$H_{2k+1}(f, \tau_2, \ldots, \tau_{2k+1}) = \frac{1}{2B} \sum_{n=-\infty}^{\infty} h_{2k+1}(n, \tau_2, \ldots, \tau_{2k+1}) e^{-\jmath 2\pi n f/2B}. \tag{2.23}$$

Thus, sampling the Volterra kernel as:

$$h_{2k+1}(n_1, \ldots, n_{2k+1}) = \frac{1}{(2B)^{2k+1}}\, h_{2k+1}(\tau_1, \ldots, \tau_{2k+1})|_{\tau_1 = \frac{n_1}{2B}, \cdots, \tau_{2k+1} = \frac{n_{2k+1}}{2B}}\,, \tag{2.24}$$

leads to:

$$H_{2k+1}(f_1, ..., f_{2k+1}) = \frac{1}{(2B)^{2k+1}} \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_2=-\infty}^{\infty} h_{2k+1}(n_1, \ldots, n_{2k+1}) \\ e^{-\jmath 2\pi \left(\sum_{i=1}^{2k+1} n_i f_i\right)/2B}. \tag{2.25}$$

From (2.17), (2.21) and (2.25), the output signal sampled at the Nyquist rate of

input signal, i.e. $x(n) = x(\xi)|_{\xi=n/2B}$, can be written as:

$$x(n) = \frac{1}{(2B)^{2k+1}} \sum_{k=0}^{K} \int_{-B}^{B} \cdots \int_{-B}^{B} \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_{2k+1}=-\infty}^{\infty} h_{2k+1}(n_1, \ldots, n_{2k+1})$$
$$e^{-\jmath 2\pi \left(\sum_{i=1}^{2k+1} n_i f_i\right)/2B} \prod_{i=1}^{k+1} \sum_{n=-\infty}^{\infty} s(n) e^{-\jmath 2\pi n f_i/2B} \prod_{i=k+2}^{2k+1} \sum_{n=-\infty}^{\infty} s^*(n) e^{-\jmath 2\pi n f_i/2B}$$
$$e^{\jmath 2\pi \left(\sum_{i=1}^{2k+1} f_i\right) n/2B} df_1 \ldots df_{2k+1}, \quad (2.26)$$

which leads to

$$x(n) = \frac{1}{(2B)^{2k+1}} \sum_{k=0}^{K} \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_{2k+1}=-\infty}^{\infty} h_{2k+1}(n_1, \ldots, n_{2k+1}) \int_{-B}^{B} \cdots \int_{-B}^{B}$$
$$e^{-\jmath 2\pi \left(\sum_{i=1}^{2k+1} n_i f_i\right)/2B} \sum_{n_1'=-\infty}^{\infty} \cdots \sum_{n_{2k+1}'=-\infty}^{\infty} \prod_{i=1}^{k+1} s(n_i') \prod_{i=k+2}^{2k+1} s^*(n_i') e^{-\jmath 2\pi \left(\sum_{i=1}^{2k+1} n_i' f_i\right)/2B}$$
$$e^{\jmath 2\pi \left(\sum_{i=1}^{2k+1} f_i\right) n/2B} df_1 \ldots df_{2k+1}, \quad (2.27)$$

or, yet,

$$x(n) = \frac{1}{(2B)^{2k+1}} \sum_{k=0}^{K} \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_{2k+1}=-\infty}^{\infty} \sum_{n_1'=-\infty}^{\infty} \cdots \sum_{n_{2k+1}'=-\infty}^{\infty} h_{2k+1}(n_1, \ldots, n_{2k+1})$$
$$\prod_{i=1}^{k+1} s(n_i') \prod_{i=k+2}^{2k+1} s^*(n_i') \prod_{i=1}^{2k+1} \int_{-B}^{B} e^{\jmath 2\pi \left(n - n_i - n_i'\right) f_i/2B} df_i, \quad (2.28)$$

which implies:

$$x(n) = \sum_{k=0}^{K} \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_{2k+1}=-\infty}^{\infty} \sum_{n_1'=-\infty}^{\infty} \cdots \sum_{n_{2k+1}'=-\infty}^{\infty} h_{2k+1}(n_1, \ldots, n_{2k+1})$$
$$\prod_{i=1}^{k+1} s(n_i') \prod_{i=k+2}^{2k+1} s^*(n_i') \prod_{i=1}^{2k+1} \operatorname{sinc}\left(\pi \left[n - n_i - n_i'\right]\right), \quad (2.29)$$

where

$$\operatorname{sinc}(\theta) = \begin{cases} 1, & \text{for} \quad \theta = 0 \\ \frac{\sin\theta}{\theta}, & \text{otherwise.} \end{cases} \quad (2.30)$$

Equation (2.29) can then be rewritten as:

$$x(n) = \sum_{k=0}^{K} \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_{2k+1}=-\infty}^{\infty} h_{2k+1}(n_1, \ldots, n_{2k+1}) \prod_{i=1}^{k+1} s(n-n_i) \prod_{i=k+2}^{2k+1} s^*(n-n_i). \quad (2.31)$$

The output $x(n)$ of a causal complex-valued discrete-time baseband equivalent Volterra system of finite memory M can then be represented by the following relationship:

$$x(n) = \sum_{k=0}^{K} \sum_{m_1=0}^{M} \cdots \sum_{m_{2k+1}=0}^{M} h_{2k+1}(m_1, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} s(n-m_i) \prod_{i=k+2}^{2k+1} s^*(n-m_i). \quad (2.32)$$

Note that (2.32) corresponds to the output of a linear finite impulse response (FIR) filter of order $(M+1)$ when $K = 0$, the coefficients of which are given by the linear kernel $h_1(m_1)$, for $0 \leq m_1 \leq M$.

The discrete-time Volterra model (2.32) can be rewritten in a compact form as:

$$x(n) = \mathbf{h}^T \mathbf{w}(n), \quad (2.33)$$

where $\mathbf{w}(n) \in \mathbb{C}^{Q \times 1}$ is the nonlinear regression vector given by:

$$\mathbf{w}(n) = [\mathbf{s}^T(n) \quad \otimes_*^3 \mathbf{s}^T(n) \ \cdots \ \otimes_*^{2K+1} \mathbf{s}^T(n)]^T, \quad (2.34)$$

with the operator $\otimes_*^{2k+1}$ defined as:

$$\otimes_*^{2k+1} \mathbf{s}(n) \equiv \left[\otimes^{k+1} \mathbf{s}(n)\right] \otimes \left[\otimes^k \mathbf{s}^*(n)\right] \in \mathbb{C}^{(M+1)^{2k+1} \times 1}, \quad (2.35)$$

$\mathbf{s}(n) = [s(n) \ s(n-1) \ \ldots \ s(n-M)]^T \in \mathbb{C}^{(M+1)\times 1}$ being the linear regression vector, $\otimes^k \mathbf{s}(n)$ the power of order k of the *Kronecker product* of $\mathbf{s}(n)$ (see Appendix A) and Q the number of channel coefficients (nonlinear terms) of the Volterra filter, given by:

$$Q = \sum_{k=0}^{K} (M+1)^{2k+1}. \quad (2.36)$$

Note that $\mathbf{w}(n)$ contains all the products like $\prod_{i=1}^{k+1} s(n-m_i) \prod_{i=k+2}^{2k+1} s^*(n-m_i)$ of (2.32). Moreover, the vector $\mathbf{h} \in \mathbb{C}^{Q \times 1}$ containing the Volterra coefficients is

given by:

$$\mathbf{h} = \left[\left(\mathbf{h}^{(1)}\right)^T \left(\mathbf{h}^{(3)}\right)^T \dots \left(\mathbf{h}^{(2K+1)}\right)^T \right]^T, \tag{2.37}$$

with $\mathbf{h}^{(2k+1)} \in \mathbb{C}^{(M+1)^{2k+1}\times 1}$, for $0 \leq k \leq K$. Using (A.3), the q^{th} element of $\mathbf{h}^{(2k+1)}$ can be written as $h_q^{(2k+1)} = h_k(m_1, \dots, m_{2k+1})$, with

$$\begin{aligned} q &= m_{2k+1} + m_{2k}(M+1) + \cdots + m_2(M+1)^{2k-1} + m_1(M+1)^{2k} \\ &= 1 + \sum_{n=1}^{N} m_k(M+1)^{N-n}, \end{aligned} \tag{2.38}$$

for $0 \leq m_1, m_2, ..., m_{2k+1} \leq M+1$.

From (2.33), it can be viewed that the output $x(n)$ is linear with respect to the system parameters $h_k(m_1, \dots, m_k)$. This means that a Volterra filter can be viewed as a linear system where the output is a weighted sum of Q virtual-sources or quasi-sources, these sources being given by the nonlinear products $w_q(n)$ of the input, where $w_q(n)$ denotes the q^{th} element of the nonlinear regression vector $\mathbf{w}(n)$. In fact, this property corresponds to one of the main advantages of Volterra models, as it may simplify the estimation of the system coefficients and the recovery of the input signals.

2.1.4 Time-varying FIR filter interpretation

In order to get a better understanding of discrete-time Volterra filters, let us consider the following homogeneous real-valued Volterra system of order 2:

$$x(n) = \sum_{m_1=0}^{M} \left(\sum_{m_2=0}^{M} h_2(m_1, m_2) s(n-m_2) \right) s(n-m_1). \tag{2.39}$$

Note that (2.39) can be written as the output of a FIR filter:

$$x(n) = \sum_{m_1=0}^{M} f_2(m_1, n) s(n-m_1), \tag{2.40}$$

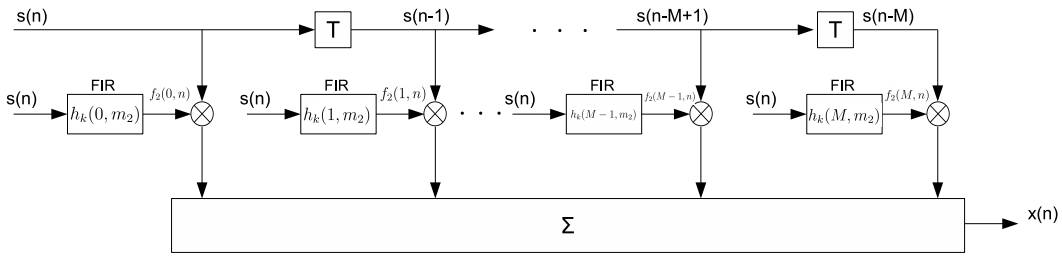

Figure 2.1: Homogeneous quadratic real-valued Volterra system viewed as a time-variant FIR filter.

the coefficients of which given by:

$$f_2(m_1, n) = \sum_{m_2=0}^{M} h_2(m_1, m_2) s(n - m_2), \tag{2.41}$$

for $0 \leq m_1 \leq M$. Moreover, the coefficient $f_2(m_1, n)$ can also be viewed as the output of a FIR filter with coefficients $h_2(m_1, m_2)$, for $0 \leq m_2 \leq M$, and input $s(n)$. This means that the output a homogeneous quadratic Volterra system can be written as the output of a FIR filter of order $(M+1)$, the m^{th} $(1 \leq m \leq M+1)$ coefficient of which being the output of another FIR filter of order $(M+1)$. This approach is illustrated by Fig. 2.1. Note that the coefficient that multiplies the delayed input $s(n-m)$ $(0 \leq m \leq M)$ is the output of the FIR with impulse response $[h_2(m,0)\ \ h_2(m,1) \cdots h_2(m,M)]$ and input $s(n)$. The considered Volterra filter can then be viewed as a linear FIR filter with time-variant coefficients that depend on the input signal $s(n)$.

A similar development can be done for a Volterra filter of any order of nonlinearity. For instance, the output of a homogeneous cubic Volterra system can be written as the output of a FIR filter, the coefficients of which being the output of a homogeneous quadratic Volterra system.

2.1.5 Triangular form

Let us go back to the general discrete-time baseband equivalent Volterra model (2.32). One of the main drawbacks of this models is its high number of coefficients. It can be viewed in (2.36) that the number of parameters Q grows exponentially with the nonlinearity order $2K+1$. However, it can be noted that some terms of the nonlinear regression vector (2.34) are redundant. They can be eliminated by

rewriting (2.32) as a triangular Volterra system:

$$x(n) = \sum_{k=0}^{K} \sum_{m_1=0}^{M} \sum_{m_2=m_1}^{M} \cdots \sum_{m_{k+1}=m_k}^{M} \sum_{m_{k+2}=0}^{M} \cdots \sum_{m_{2k+1}=m_{2k}}^{M} \tilde{h}_{2k+1}(m_1, m_2, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} s(n-m_i) \prod_{i=k+2}^{2k+1} s^*(n-m_i). \quad (2.42)$$

with

$$\tilde{h}_{2k+1}(m_1, m_2, \ldots, m_{2k+1}) = \underbrace{\sum_{\tilde{m}_1=0}^{M} \cdots \sum_{\tilde{m}_{k+1}=0}^{M}}_{\substack{\{\tilde{m}_1,\ldots,\tilde{m}_{k+1}\}\in \\ \pi(m_1,\ldots,m_{k+1})}} \underbrace{\sum_{\tilde{m}_{k+2}=0}^{M} \cdots \sum_{\tilde{m}_{2k+1}=0}^{M}}_{\substack{\{\tilde{m}_{k+2},\ldots,\tilde{m}_{2k+1}\}\in \\ \pi(m_{k+2}\ldots,m_{2k+1})}} h_{2k+1}(\tilde{m}_1, \tilde{m}_2, \ldots, \tilde{m}_{2k+1}), \quad (2.43)$$

where $\pi(m_1, \ldots, m_{k+1})$ denotes the set of all the permutations of $\{m_1, \ldots, m_{k+1}\}$. Note that, in (2.42), we have $m_1 \leq \cdots \leq m_{k+1}$ and $m_{k+2} \leq \cdots \leq m_{2k+1}$.

The vector form of a triangular Volterra model is given by:

$$x(n) = \tilde{\mathbf{h}}^T \mathbf{w}(n), \quad (2.44)$$

where the nonlinear regression vector $\mathbf{w}(n) \in \mathbb{C}^{Q\times 1}$ is given by:

$$\mathbf{w}(n) = [\mathbf{s}^T(n) \quad \oslash_*^3 \mathbf{s}^T(n) \ \cdots \ \oslash_*^{2K+1} \mathbf{s}^T(n)]^T \in \mathbb{C}^{Q\times 1}, \quad (2.45)$$

the operator $\oslash_*^{2k+1}$ being defined as:

$$\oslash_*^{2k+1} \mathbf{s}(n) = \left[\oslash^{k+1} \mathbf{s}(n)\right] \otimes \left[\oslash^{k} \mathbf{s}^*(n)\right] \quad (2.46)$$

and $\oslash^k \mathbf{s}(n)$ being the power of order k of the *truncated Kronecker product* of $\mathbf{s}(n)$ (see Appendix A). In this case, the vector $\tilde{\mathbf{h}} \in \mathbb{C}^{Q\times 1}$ contains the parameters of the triangular Volterra model and the number of parameters Q can be deduced from (A.6) in Appendix A:

$$Q = \sum_{k=0}^{K} C_{M+1,k} C_{M+1,k+1} \ , \quad \text{where} \ \ C_{i,p} = \frac{(i+p-1)!}{(i-1)!p!}. \quad (2.47)$$

As we will see later, if the input signals have a constant modulus, as in Phase Shift Keying (PSK) modulations, the triangular Volterra model (2.42) can be rewritten

with a smaller number of coefficients Q, since, in this case, some power terms of $s(n)$ can be viewed as power terms of smaller order.

2.2 MIMO Volterra channels

As mentioned earlier, MIMO communication channels are also subject to nonlinear distortions. In such cases, the MIMO Volterra models, described in this section, can be used for modeling the channel. In fact, the models described in the sequel correspond to the most general forms of MIMO Volterra systems considering discrete-time equivalent baseband representations. It should be highlighted that, assuming the multiple sources transmit at the same time and the same frequency band, the developments of Section 2.1 can be done in a similar way for MIMO Volterra channels. The applications of MIMO Volterra models in communication systems exploited in this thesis are discussed in Section 2.4.

For developing the expression of a discrete-time equivalent baseband MIMO Volterra model, we first consider a nonlinear Single-Input-Multiple-Output (SIMO) communication channel. That corresponds, for instance, to the case where a single user transmits to an array of R receive antennas. In this case, the link between each antenna element and the user can be modeled as a discrete-time baseband equivalent Volterra system like (2.32). Thus, the discrete-time signal received by the r^{th} antenna element can be expressed as:

$$x_r(n) = \sum_{k=0}^{K} \sum_{m_1=0}^{M} \cdots \sum_{m_{2k+1}=0}^{M} h_{2k+1}^{(r)}(m_1, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} s(n-m_i) \prod_{i=k+2}^{2k+1} s^*(n-m_i). \tag{2.48}$$

where $h_{2k+1}^{(r)}(m_1, \ldots, m_{2k+1})$ are the Volterra kernels coefficients associated with the r^{th} output.

In the case of MIMO channel, i.e. considering multiple transmit antennas or multiple-users, if the link between each source (Tx antenna or user) and each receive antenna is modeled as a Volterra system, the r^{th} output signal is then

expressed by:

$$x_r(n) = \sum_{t=1}^{T} \sum_{k=0}^{K} \sum_{m_1=0}^{M} \cdots \sum_{m_{2k+1}=0}^{M} h_{2k+1}^{(r)}(t, m_1, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} s_t(n - m_i) \prod_{i=k+2}^{2k+1} s_t^*(n - m_i). \quad (2.49)$$

where $s_t(n)$ is the t^{th} input signal ($1 \leq t \leq T$) and $h_k^{(r)}(t, m_1, \ldots, m_k)$ are the Volterra kernels coefficients associated with the r^{th} output and the t^{th} input signal.

Note that the discrete-time baseband equivalent MIMO Volterra system (2.49) have products of the input signals in the form $\prod_{i=1}^{k+1} s_t(n - m_i) \prod_{i=k+2}^{2k+1} s_t^*(n - m_i)$, which means that it does not contain products of different sources. This corresponds to the case where the signal of each source is corrupted by nonlinear intersymbol interference (ISI) and, then, the signal of all the sources are linearly mixed, i.e. the nonlinearities are applied to the input signals individually, before mixing the sources. This kind of model can be viewed as a parallel-cascade of T SIMO Volterra systems. Thus, although (2.49) corresponds to a MIMO Volterra model, it does not represent the case where the sources are nonlinearly mixed. In order to develop a generical expression for a MIMO Volterra model taking this case into account, let us consider the SIMO Volterra system (2.48) with the input signal $s(n)$ being the mixture of T different sources:

$$s(n) = \sum_{t=1}^{T} a_t s_t(n), \quad (2.50)$$

leading to:

$$x_r(n) = \sum_{k=0}^{K} \sum_{m_1=0}^{M} \cdots \sum_{m_{2k+1}=0}^{M} h_{2k+1}^{(r)}(m_1, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} \left(\sum_{t=1}^{T} a_t s_t(n - m_i) \right) \prod_{i=k+2}^{2k+1} \left(\sum_{t=1}^{T} a_t^* s_t^*(n - m_i) \right), \quad (2.51)$$

or, equivalently,

$$x_r(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{2k+1}=1}^{T} \sum_{m_1=0}^{M} \cdots \sum_{m_{2k+1}=0}^{M} \left[h_{2k+1}^{(r)}(m_1, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} a_{t_i} \prod_{i=k+2}^{2k+1} a_{t_i}^* \right] \prod_{i=1}^{k+1} s_{t_i}(n-m_i) \prod_{i=k+2}^{2k+1} s_{t_i}^*(n-m_i). \quad (2.52)$$

The MIMO Volterra model (2.52) contains products of different sources $\prod_{i=1}^{k+1} s_{t_i}(n-m_i) \prod_{i=k+2}^{2k+1} s_{t_i}^*(n-m_i)$, as it corresponds to the case where the nonlinearities are applied to the input signals after mixing the sources. Based on (2.52), the general discrete-time equivalent baseband MIMO Volterra model with T inputs and R outputs is defined as:

$$x_r(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{2k+1}=1}^{T} \sum_{m_1=0}^{M} \cdots \sum_{m_{2k+1}=0}^{M} h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, m_1, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} s_{t_i}(n-m_i) \prod_{i=k+2}^{2k+1} s_{t_i}^*(n-m_i). \quad (2.53)$$

where $h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, m_1, \ldots, m_k)$ are the Volterra kernels coefficients associated with the r^{th} output and the product $\prod_{i=1}^{k+1} s_{t_i}(n-m_i) \prod_{i=k+2}^{2k+1} s_{t_i}^*(n-m_i)$. When the diversity at the reception is provided by an antenna array, equation (2.53) can be viewed as a "spatial-temporal Volterra" model in contrast with the "temporal Volterra" model (2.32). Indeed, in this case, the Volterra kernels have multiple time $(t_1, \ldots, t_k)$ and space $(m_1, \ldots, m_k)$ indices, instead of only time indices, as in the SISO Volterra system (2.32), or multiple time indices and only one space index, as in the MIMO Volterra model (2.49).

As well as for the SISO case, the discrete-time MIMO Volterra model (2.53) can be rewritten in a triangular form, in such a way that the redundant terms are eliminated:

$$x_r(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{k+1}=t_k}^{T} \sum_{t_{k+2}=1}^{T} \cdots \sum_{t_{2k+1}=t_{2k}}^{T} \sum_{m_1=0}^{M} \cdots \sum_{m_{k+1}=m'_{k+1}}^{M} \sum_{m_{k+2}=0}^{M} \cdots \sum_{m_{2k+1}=m'_{2k+1}}^{M} \tilde{h}_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, m_1, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} s_{t_i}(n-m_i) \prod_{i=k+2}^{2k+1} s_{t_i}^*(n-m_i). \quad (2.54)$$

with

$$m_k^{'} = \begin{cases} m_{k-1} & \text{if} \quad t_k = t_{k-1}, \\ 0 & \text{if} \quad t_k \neq t_{k-1}, \end{cases} \tag{2.55}$$

It is also possible to define a MIMO Volterra model in order to take into account the fact that the system has different memories with respect to the inputs. Thus, a more general representation of discrete-time baseband equivalent MIMO Volterra systems can be written as:

$$x_r(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{k+1}=t_k}^{T} \sum_{t_{k+2}=1}^{T} \cdots \sum_{t_{2k+1}=t_{2k}}^{T} \sum_{m_1=0}^{M_{t_1}} \cdots \sum_{m_{k+1}=m'_{k+1}}^{M_{t_{k+1}}} \sum_{m_{k+2}=0}^{M_{t_{k+2}}} \cdots \sum_{m_{2k+1}=m'_{2k+1}}^{M_{t_{2k+1}}}$$
$$\tilde{h}_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, m_1, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} s_{t_i}(n - m_i) \prod_{i=k+2}^{2k+1} s_{t_i}^*(n - m_i), \tag{2.56}$$

where M_{t_i} is the memory of the system with respect to the t_i^{th} input, for $1 \leq t_i \leq T$. The system model (2.56) can be rewritten in a different way:

$$x_r(n) = \sum_{k=0}^{K} \sum_{\bar{m}_1=1}^{\bar{M}} \cdots \sum_{\bar{m}_{k+1}=\bar{m}_k}^{\bar{M}} \sum_{\bar{m}_{k+2}=1}^{\bar{M}} \cdots \sum_{\bar{m}_{2k+1}=\bar{m}_{2k}}^{\bar{M}} \tilde{h}_{2k+1}^{(r)}(\bar{m}_1, \ldots, \bar{m}_{2k+1})$$
$$\prod_{i=1}^{k+1} \bar{s}_{\bar{m}_i}(n) \prod_{i=k+2}^{2k+1} \bar{s}_{\bar{m}_i}^*(n), \tag{2.57}$$

where $\bar{M} = M_1 + \cdots + M_T + T$ and $\bar{s}_{\bar{m}_i}(n)$ corresponds to the $\bar{m}_i^{th}$ component of the linear input vector defined as:

$$\begin{aligned} \bar{\mathbf{s}}(n) &= [\bar{s}_1(n)\ \bar{s}_2(n)\ \ldots\ \bar{s}_{\bar{M}}(n)]^T \qquad \in \mathbb{C}^{\bar{M}\times 1} \\ &= [s_1(n) \ldots s_1(n - M_1) \cdots s_T(n) \ldots s_T(n - M_T)]^T . \end{aligned} \tag{2.58}$$

Note that the Volterra models (2.56) and (2.57) are equivalent and can be represented in the following compact way:

$$\mathbf{x}(n) = \tilde{\mathbf{H}}\mathbf{w}(n), \tag{2.59}$$

where $\mathbf{x}(n) = [x_1(n)\ x_2(n)\ \ldots\ x_R(n)]^T \in \mathbb{C}^{R\times 1}$ and the nonlinear input vector $\mathbf{w}(n) \in \mathbb{C}^{Q\times 1}$ is given by:

$$\mathbf{w}(n) = [\bar{\mathbf{s}}^T(n) \quad \oslash_*^3 \bar{\mathbf{s}}^T(n) \ \cdots \ \oslash_*^{2K+1} \bar{\mathbf{s}}^T(n)]^T \in \mathbb{C}^{Q\times 1}, \tag{2.60}$$

with $\bar{\mathbf{s}}(n)$ given by (2.58). Moreover, $\tilde{\mathbf{H}} = [\tilde{\mathbf{h}}_1\, \tilde{\mathbf{h}}_2\, \ldots\, \tilde{\mathbf{h}}_R]^T \in \mathbb{C}^{R\times Q}$, with $\tilde{\mathbf{h}}_r \in \mathbb{C}^{Q\times 1}$ $(1 \leq r \leq R)$ containing the coefficients of the triangular Volterra system associated with r^{th} output. In this case, the length of the Volterra filter is given by:

$$Q = \sum_{k=0}^{K} C_{\bar{M},k} C_{\bar{M},k+1}\,. \tag{2.61}$$

Discrete-time MIMO Volterra models have already being studied in a few works. However, some works deal with MIMO Volterra systems less generical than (2.53)-(2.57), e.g [123, 1, 124, 113, 127, 67, 133], and some of them deal with a real-valued version of (2.53), e.g [136, 2]. To the best of our knowledge, the complex-valued equivalent baseband MIMO Volterra models (2.53)-(2.57) are unexplored in the literature.

2.3 Block-structured nonlinear systems

The MIMO communication channels studied in this thesis can be viewed as special cases of Volterra filters constituted of series-cascades of nonlinear and linear systems. In this section, some of these block-structured nonlinear systems are studied. In particular, the Wiener, Hammerstein and Wiener-Hammerstein models are described. These block-structured nonlinear systems have important applications in many areas [65] and specially in telecommunication systems, as we will see in Section 2.4. The next developments concern discrete-time equivalent baseband MIMO models, the SISO case being a particular case for $T = R = 1$.

2.3.1 The Wiener model

A SISO Wiener system is the cascade of a linear finite impulse response (FIR) filter followed by a static nonlinearity. A MIMO Wiener system with T inputs and R outputs is composed of a linear mixer followed by memoryless nonlinearities, as shown in Fig. 2.2. In the sequel, we present the case where the linear part of the system is a convolutional mixer.

Let us denote by $z_r(n)$ $(1 \leq r \leq R)$ the outputs of the linear mixer and by $w_t^{(r)}(m)$ $(0 \leq m \leq M)$ the $(m+1)^{th}$ element of the impulse response associated with the t^{th} input and the r^{th} output. It is considered that the system has a fixed memory

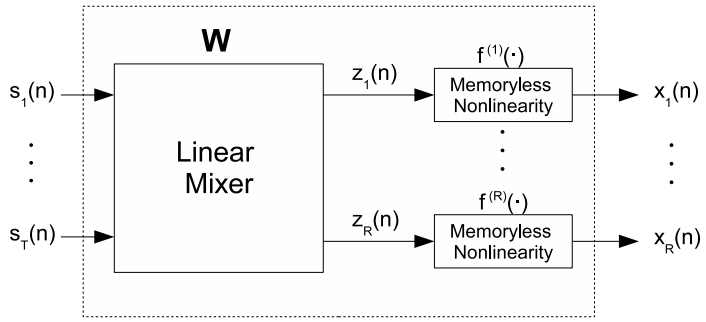

Figure 2.2: A MIMO Wiener system.

M with respect to all the inputs. So, we may write:

$$z_r(n) = \sum_{t=1}^{T} \sum_{m=0}^{M} w_t^{(r)}(m) s_t(n-m). \tag{2.62}$$

Denoting by $x_r(n)$ $(1 \leq r \leq R)$ the outputs of the MIMO Wiener system, we have:

$$x_r(n) = f^{(r)}\left(z_r(n)\right), \tag{2.63}$$

where $f^{(r)}(\cdot)$ $(1 \leq r \leq R)$ is a polynomial function. Considering an equivalent baseband representation, similarly as for the Volterra models (2.53)-(2.57), the functions $f^{(r)}(\cdot)$ can be represented by a polynomial of the form:

$$x_r(n) = \sum_{k=0}^{K} f_{2k+1}^{(r)} \left|z_r(n)\right|^{2k} z_r(n), \tag{2.64}$$

where $|\cdot|$ denotes the magnitude of a complex number and $\{f_1^{(r)}, f_3^{(r)}, \ldots, f_{2K+1}^{(r)}\}$ are the baseband equivalent coefficients of the polynomial function $f^{(r)}(\cdot)$. As in (2.53)-(2.57), the polynomial terms that do not have the form (2.64) correspond to spectral components lying outside the system bandwidth.

Substituting (2.62) into (2.64), we get:

$$x_r(n) = \sum_{k=0}^{K} f_{2k+1}^{(r)} \left| \sum_{t=1}^{T} \sum_{m=0}^{M} w_t^{(r)}(m) s_t(n-m) \right|^{2k} \sum_{t=1}^{T} \sum_{m=0}^{M} w_t^{(r)}(m) s_t(n-m), \tag{2.65}$$

or, equivalently,

$$x_r(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{2k+1}=1}^{T} \sum_{m_1=0}^{M} \cdots \sum_{m_{2k+1}=0}^{M} f_{2k+1}^{(r)} \prod_{i=1}^{k+1} w_{t_i}^{(r)}(m_i) s_{t_i}(n-m_i) \prod_{i=k+2}^{2k+1} \left[w_{t_i}^{(r)}(m_i) s_{t_i}(n-m_i) \right]^*. \quad (2.66)$$

So, by defining a global Volterra kernel as:

$$h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, m_1, \ldots, m_{2k+1}) \equiv f_{2k+1}^{(r)} \prod_{i=1}^{k+1} w_{t_i}^{(r)}(m_i) \prod_{i=k+2}^{2k+1} \left[w_{t_i}^{(r)}(m_i) \right]^*, \quad (2.67)$$

equation (2.66) can be rewritten as:

$$x_r(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{2k+1}=1}^{T} \sum_{m_1=0}^{M} \cdots \sum_{m_{2k+1}=0}^{M} h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, m_1, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} s_{t_i}(n-m_i) \prod_{i=k+2}^{2k+1} s_{t_i}^*(n-m_i). \quad (2.68)$$

From (2.68) and (2.67), it can be concluded that a MIMO Wiener system is equivalent to a MIMO Volterra model with separable kernels.

2.3.2 The Hammerstein model

A SISO Hammerstein system is composed of a memoryless nonlinear block followed by a linear FIR filter, while a MIMO Hammerstein model is composed of memoryless nonlinear blocks in parallel, followed by a linear mixer, as shown in Fig. 2.3. Note that the order of the blocks of a Hammerstein system is the inverse of that of a Wiener system.

Let $z_t(n)$ $(1 \leq t \leq T)$ be the outputs of the memoryless nonlinearities $f^{(t)}(\cdot)$ $(1 \leq t \leq T)$. Using the same assumption about the nonlinear functions considered for the Wiener model, we can write:

$$z_t(n) = \sum_{k=0}^{K} f_{2k+1}^{(t)} |s_t(n)|^{2k} s_t(n), \quad (2.69)$$

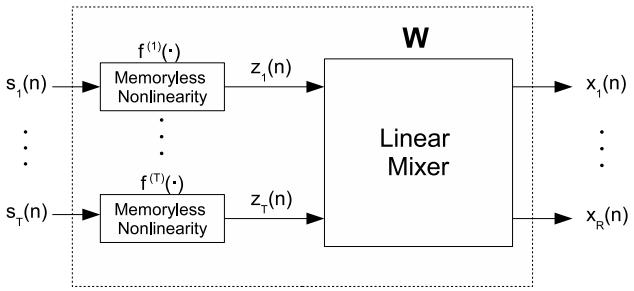

Figure 2.3: A MIMO Hammerstein system.

where $\{f_1^{(t)}, f_3^{(t)}, \ldots, f_{2K+1}^{(t)}\}$ are the baseband equivalent coefficients of the polynomial function $f^{(t)}(\cdot)$. Denoting by $w_t^{(r)}(m)$ $(0 \leq m \leq M)$ the $(m+1)^{th}$ element of the impulse response associated with the t^{th} input and the r^{th} output, we have:

$$x_r(n) = \sum_{t=1}^{T} \sum_{m=0}^{M} w_t^{(r)}(m) z_t(n-m). \tag{2.70}$$

Substituting (2.69) into (2.70), we get:

$$x_r(n) = \sum_{k=0}^{K} \sum_{t=1}^{T} \sum_{m=0}^{M} f_{2k+1}^{(t)} w_t^{(r)}(m) \left|s_t(n-m)\right|^{2k} s_t(n-m). \tag{2.71}$$

So, by defining the global Volterra kernel as:

$$h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, m_1, \ldots, m_{2k+1}) \equiv \begin{cases} f_{2k+1}^{(t_1)} w_{t_1}^{(r)}(m_1), & \text{if} \quad t_1 = \ldots = t_{2k+1} \\ & \text{and} \quad m_1 = \ldots = m_{2k+1}, \\ 0, & \text{else}, \end{cases} \tag{2.72}$$

it is possible to write the output of the MIMO Hammerstein system (2.71) as the output of the MIMO Volterra model (2.68). So, it can be concluded that a MIMO Hammerstein system can be viewed as a particular case of a diagonal MIMO Volterra system.

2.3.3 The Wiener-Hammerstein model

A SISO Wiener-Hammerstein system is composed of a static nonlinearity sandwiched between two linear FIR filters, and its MIMO version is composed of static nonlinear blocks in parallel, sandwiched between two linear mixers (see Fig. 2.4).

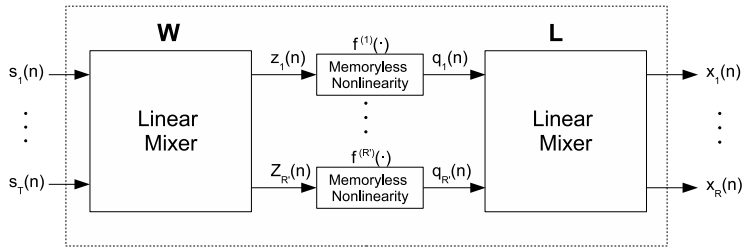

Figure 2.4: A MIMO Wiener-Hammerstein system.

Note that the Wiener and the Hammerstein models can be viewed as particular cases of a Wiener-Hammerstein system.

Let $R^{'}$ be the number of outputs of the first mixer, $z_{r'}(n)$ $(1 \leq r^{'} \leq R^{'})$ the outputs of the first mixer, $q_{r'}(n)$ $(1 \leq r^{'} \leq R^{'})$ the outputs of the nonlinear blocks, $w_t^{(r^{'})}(m)$ $(0 \leq m \leq M_w)$ the $(m+1)^{th}$ element of the impulse response of the first mixer associated with the t^{th} input and the $(r^{'})^{th}$ output, and M_w the memory of the first mixer. From (2.66), we can write:

$$q_{r'}(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{2k+1}=1}^{T} \sum_{m_1=0}^{M_w} \cdots \sum_{m_{2k+1}=0}^{M_w} f_{2k+1}^{(r^{'})} \prod_{i=1}^{k+1} w_{t_i}^{(r^{'})}(m_i) \prod_{i=k+2}^{2k+1} \left[w_{t_i}^{(r^{'})}(m_i) \right]^{*} \prod_{i=1}^{k+1} s_{t_i}(n-m_i) \prod_{i=k+2}^{2k+1} s_{t_i}^{*}(n-m_i). \quad (2.73)$$

On the other hand, the outputs of the MIMO Wiener-Hammerstein system can be expressed as:

$$x_r(n) = \sum_{r'=1}^{R^{'}} \sum_{m=0}^{M_l} l_{r'}^{(r)}(m) q_{r'}(n-m), \quad (2.74)$$

where $l_{r'}^{(r)}(m)$ $(0 \leq m \leq M_l)$ is the $(m+1)^{th}$ element of the impulse response of the second mixer associated with the $(r^{'})^{th}$ input and the r^{th} output, and M_l is

the memory of the second mixer. Substituting (2.73) into (2.74), we get:

$$x_r(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{2k+1}=1}^{T} \sum_{m_1=0}^{M_w} \cdots \sum_{m_{2k+1}=0}^{M_w} \sum_{r'=1}^{R'} \sum_{m=0}^{M_l} l_{r'}^{(r)}(m) f_{2k+1}^{(r')} \prod_{i=1}^{k+1} w_{t_i}^{(r')}(m_i) \prod_{i=k+2}^{2k+1} \left[w_{t_i}^{(r')}(m_i) \right]^* \prod_{i=1}^{k+1} s_{t_i}(n - m_i - m) \prod_{i=k+2}^{2k+1} s_{t_i}^*(n - m_i - m). \quad (2.75)$$

By defining:

$$h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, \bar{m}_1, \ldots, \bar{m}_{2k+1}) \equiv \underbrace{\sum_{m=0}^{M_l} \sum_{m_1=0}^{M_w} \cdots \sum_{m_{2k+1}=0}^{M_w}}_{\{m+m_1=\bar{m}_1\} \cap \cdots \cap \{m+m_{2k+1}=\bar{m}_{2k+1}\}} \sum_{r'=1}^{R'} l_{r'}^{(r)}(m) f_{2k+1}^{(r')} \left(\prod_{i=1}^{k+1} w_{t_i}^{(r')}(m_i) \right) \left(\prod_{i=k+2}^{2k+1} \left[w_{t_i}^{(r')}(m_i) \right]^* \right), \quad (2.76)$$

with $\bar{m}_{2k+1} = m_{2k+1} + m$, equation (2.75) can be rewritten as a MIMO Volterra filter:

$$x_r(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{2k+1}=1}^{T} \sum_{\bar{m}_1=0}^{M} \cdots \sum_{\bar{m}_{2k+1}=0}^{M} h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, \bar{m}_1, \ldots, \bar{m}_{2k+1}) \prod_{i=1}^{k+1} s_{t_i}(n - \bar{m}_i) \prod_{i=k+2}^{2k+1} s_{t_i}^*(n - \bar{m}_i), \quad (2.77)$$

where $M = M_w + M_l$.

Note that the MIMO Wiener, Hammerstein and Wiener-Hammerstein models have a global Volterra representation, allowing to conclude that they can be viewed as particular cases of the MIMO Volterra model. A link between Wiener, Hammerstein and Wiener-Hammerstein models, and the Volterra model was developed in [85] for SISO systems. The above developments linking these MIMO block-structured models and the MIMO Volterra model constitute then a generalization of the results of [85].

The main advantage of representing Wiener, Hammerstein and Wiener-Hammerstein systems in terms of the parameters of the subsystems, i.e. the linear mixers and memoryless nonlinearities, is that the total number of parameters of these subsystems is, in general, smaller than the number of coefficients of the glo-

bal Volterra system. On the other hand, the system output is not linear with respect to the parameters of these subsystems, as we can see in (2.66), (2.71) and (2.75), contrarily to the coefficients of the global Volterra system.

2.4 Applications in communication systems

This section is dedicated to present some applications of MIMO Volterra models in communication systems, based on the nonlinear MIMO models presented in the previous sections. These applications correspond to situations where the received signals are corrupted by nonlinear Intersymbol Interference (ISI), nonlinear Multiple Access Interference (MAI) and/or nonlinear Inter-carrier Interference (ICI). Applications of Volterra models in other kinds of MIMO communication systems can be found in [127, 148, 113]. In all the following examples, it is assumed perfect symbol synchronization and that the receive filter is matched to the transmit pulse shape filter.

2.4.1 MIMO channels with nonlinear power amplifiers

In general, all the wireless communication systems employing power amplifiers (PAs) are subject to nonlinear distortions. However, when the signal at the input of the PA is characterized by a high peak-to-average power ratio (PAPR), the introduced nonlinear distortions are particularly important. For these signals, the maximal signal amplitude is high compared to the root mean square (RMS) value. Thus, if the PA operates near the saturation region to obtain a good power efficiency, some components of the input signal fall at the saturation region due to the large fluctuations on the signal envelope. The PA exhibits a nonlinear characteristic at saturation, resulting in the introduction of nonlinear bandlimited distortions [169]. That may lead to significant signal distortion and system performance deterioration. For further details about the PA, see [131, 79] and references therein.

Some models can be encountered in the literature to represent the nonlinearity of the PA. The Saleh model represents the traveling wave tube (TWT) PA as a frequency independent memoryless nonlinearity characterized by the following amplitude to amplitude (AM/AM) and amplitude to phase (AM/PM) conversions [131]:

$$A(r) = \frac{\alpha_a r}{1 + \beta_a r^2} \tag{2.78}$$

and

$$\Phi(r) = \frac{\alpha_\phi r^2}{1 + \beta_\phi r^2}, \tag{2.79}$$

where r is the amplitude of the PA input signal, $A(r)$ and $\Phi(r)$ are respectively the amplitude and phase gain of the PA output signal, and α_a, β_a, α_ϕ and β_ϕ are positive scalar constants.

Radio frequency PAs can also be modeled using polynomial models such as Volterra systems. Considering memoryless models, the equivalent baseband input-output relationship of a PA is often represented by the memoryless memoryless polynomial model [38, 169, 119]:

$$x(n) = \sum_{k=0}^{K} f_{2k+1} \left|s(n)\right|^{2k} s(n). \tag{2.80}$$

If the polynomial coefficients f_{2k+1} are real-valued, the model (2.80) is strictly memoryless, which means that the PA introduces only amplitude distortion (AM/AM conversion). However, it was shown that, if the coefficients f_{2k+1} are complex-valued, the model (2.80) allows representing a more general class of models called quasi-memoryless PA [38, 119, 121]. This means that if the memory of the PA is short compared to the time variations of the input signal envelope, equation (2.80) may represent the output of a PA with amplitude and phase distortions (AM/AM and AM/PM conversions).

Nevertheless, when the bandwidth of the input signal is large, the memory of the PA can not be considered short with respect to the time variations of the input signal [38, 169]. More complex models must then be used to take the memory effects of the PA into account. In this case, among the nonlinear models usually considered in the literature for modeling the PA, the most general is given by the SISO Volterra model (2.42) [39, 38, 119]. Several special cases of the Volterra model can also be used for modeling the PA nonlinearities as the SISO Wiener, Hammerstein and Wiener-Hammerstein systems [38, 119, 25]. Moreover, PAs with memory effects are also often modeled as a diagonal Volterra model [38, 119, 39, 102, 66, 16, 40]:

$$x(n) = \sum_{k=0}^{K} \sum_{m=0}^{M} f_{2k+1}(m) \left|s(n-m)\right|^{2k} s(n-m). \tag{2.81}$$

This model is also referred to as memory polynomial model and can be viewed as a generalization of the SISO Hammerstein model.

Thus, all the wireless MIMO communication channels employing PAs are subject to nonlinear distortions. In theses cases, the MIMO Volterra model can be used for modeling the channel. In the sequel, two wireless communication systems modeled as MIMO Volterra models are presented.

OFDM systems

Orthogonal Frequency Division Multiplexing (OFDM) signals are characterized by a high PAPR [31, 126, 148, 166, 16, 6, 5], caused by the sum of several symbols with different phases and frequencies. As a consequence, the received signals in a OFDM system are particularly affected by the presence of a nonlinear PA. In this case, a nonlinear PA results in the introduction of nonlinear inter-carrier interference (ICI) between the subcarriers. Theoretical analysis and performance of OFDM signals in nonlinear channels have been widely studied in the literature [28, 32, 6, 5, 15, 16]. It should be highlighted that the nonlinear distortions in OFDM systems can be reduced by using peak power control techniques that reduce the PAPR of the transmitted signals [96]. However, this kind of technique will not be used in this thesis.

A single-user OFDM channel can then be modeled as a cascade of a nonlinear system, corresponding to the PA, followed by a linear FIR filter corresponding to the frequency selective fading wireless link. Besides, the global SISO-OFDM channel (PA + wireless link) can be modeled as a Volterra system [126].

MIMO transmission schemes can be used in OFDM systems to provide an efficient radio spectrum, allowing a good reuse of the same frequency range to increase the data rate and the system capacity. In this case, the global channel between each source (Tx antenna or user) and each receive antenna can be modeled as Volterra system and the global MIMO-OFDM channel can be written as a the MIMO Volterra model (2.49). Note that, due to the fact that the nonlinearity is at the transmitter, the nonlinear MIMO-OFDM channel does not contain products between terms of different sources. The signal of each source, corrupted by nonlinear ICI, is linearly mixed with the signal of the other sources.

Particularly, if the PA is characterized as a memoryless polynomial model, the global MIMO-OFDM channel can be modeled as a MIMO Hammerstein system. In this case, using the notation introduced in Section 2.3.1:

- $f_{2k+1}^{(t)}$ denotes coefficients of the polynomial function representing the PA of the t^{th} antenna element.

- $w_t^{(r)}(m), m = 0, 1, ...M$, denotes the channel impulse response of the wireless channel between the t^{th} user and the r^{th} receive antenna.

Channel estimation and equalization of nonlinear MIMO-OFDM channels are considered in Chapter 3 of this thesis. As it will be shown, the OFDM transmission/reception scheme allows a great simplification in the modeling of MIMO Volterra channels.

Satellite systems

In satellite communication systems, the signals are transmitted from a ground (earth) station towards a satellite station (uplink) and then retransmitted to a receive ground station (downlink). Due to power limitation, the satellite station usually employs a PA [30], often in the form of a traveling wave tube (TWT) or solid-state power amplifiers (SSPA), that is driven at or near saturation in order to obtain a power efficient transmission [9, 78, 157, 69], resulting in the introduction of nonlinear distortions.

The overall satellite channel, i.e. considering the cascade of the uplink, PA and downlink, was first modeled as an equivalent baseband SISO Volterra system by Benedetto *et al.* [11], its effectiveness for modeling this kind of channels being verified in [9]. In some cases, the Volterra model for the satellite channel incorporates the satellite pre- and post-filters [9, 12, 68, 147]. Satellite channels can also be modeled as a SISO Wiener-Hammerstein system, the wireless uplink and downlink being represented by FIR filters and the PA by a memoryless polynomial model [9, 79].

In order to improve the transmission spectral efficiency, the use of MIMO satellite systems has been considered by a number of works [135, 89, 90, 91, 167]. Concerning the structure of the MIMO satellite link, one of the following configurations can be considered: (i) one ground station with multiple transmit antennas transmits towards a satellite transponder with multiple antenna that retransmits towards another ground station with multiple receive antennas; (ii) the ground station with multiple transmit antennas transmits towards multiple satellites with a single antenna each one, that retransmits towards another ground station with multiple receive antennas. Moreover, we can also consider the case where the T sources correspond to various ground stations with a single transmit antenna, i.e. mobile unites transmitting towards a single receive station. In all these cases, the channel can be represented as a MIMO Wiener-Hammerstein model and, hence, as a MIMO Volterra model, with:

- the wireless uplink being modeled as a linear $T \times R^{'}$ mixer with channel impulse responses denoted by $w_t^{(r^{'})}(m), m = 0, 1, ...M_w, t = 1, ..., T, r^{'} = 1, ..., R^{'}$;

- $f_{2k+1}^{(r^{'})}$ denoting coefficients of the polynomial function representing the PA, $r^{'} = 1, ..., R^{'}$.

- the wireless downlink being modeled as a linear $R^{'} \times R$ mixer with impulse responses denoted by $l_{r^{'}}^{(r)}(m), m = 0, 1, ...M_w, r = 1, ..., R, r^{'} = 1, ..., R^{'}$.

Simulation results concerning nonlinear satellite channels are not provided in this thesis. However, the techniques for channel estimation and information recovery in MIMO Volterra models developed in the Chapters 4, 5 and 6 can be applied for this kind of communication systems.

2.4.2 Radio Over Fiber (ROF) channels

ROF links have found a new important application with their introduction in micro- and pico-cellular wireless networks [55, 57, 54, 116, 115]. Micro- and pico-cellular architectures provide to the system a better capacity, coverage and power consumption, specially in hot-spot areas. Thus, it can also improve the system reliability and Quality of Service. ROF links provide a cost-effective solution for important problems of such wireless systems as complexity and bandwidth limitation [88]. In ROF systems, the uplink transmission is done from a mobile station towards Radio Access Points (RAP), which are merely low-cost remote antenna stations consisting of an electro-optical converter and a transponder [132]. At the RAP, the transmitted signals are converted in optical frequencies by a laser diode and then retransmitted through optical fibers towards a central Base Station (BS), as summarized in Fig. 2.5. Most part of the signal processing, such as channel estimation, equalization, modulation and demodulation, is done at the BS [55, 57, 132].

Important nonlinear distortions are introduced by the laser diode at the electrical-optical (E/O) conversion device [55, 57, 54, 114, 116]. Gain compression characteristics combined with stimulated and spontaneous mechanisms of emission make the laser inherently nonlinear [55], the nonlinearity being categorized as static and dynamic [161]. However, only static nonlinearity is considered as a major concern. Dynamic nonlinearity plays an important role only in transient state, which are not of interest in communications. Moreover, other phenomena, as leakage current

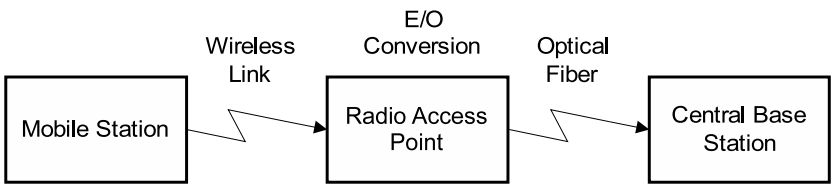

Figure 2.5: Radio Over Fiber Uplink system.

and axial hole burning may also be sources of nonlinearities [55]. The E/O nonlinearity in a ROF system is often modeled using the memoryless polynomial model [55, 57, 116, 114]. For more details about the ROF nonlinearities, see [161, 55] and references therein.

Concerning the optical link, chromatic dispersion is some of the main concerns with single-mode and multi-mode fibers. The transfer function of a fiber reflecting the chromatic dispersion, is given by [55]:

$$H(f) = e^{-\mathbf{j}\alpha l(f-f_o)^2} \tag{2.82}$$

where α is a dispersion coefficient, l is the fiber length and f_o is the optical carrier frequency. For a wavelength of 1310 nm, the chromatic dispersion of the fiber is not significant up to several hundreds of kilometers of fiber length and up to few GHz [55, 54, 162]. This means that the chromatic dispersion of the fiber is negligible and the nonlinear distortion arising from the E/O conversion process becomes then preponderant. Thus, the overall uplink channel can be viewed as a wireless link followed by an E/O conversion. In a single-user and a single receive antenna case, the wireless link can be modeled as a linear FIR filter and the overall ROF uplink channel as a SISO Wiener model [55, 54, 114] .

Aiming to supply the growing demand for system capacity, technologies such as smart antennas (or MIMO) and ROF transmission can be used together [88, 128]. In a ROF system with an antenna array at the RAP, the optical link between the RAP and the BS can be implemented using either multiple fibers or a single fiber with wavelength division multiplexing (WDM) [88, 128, 132]. The second case is particularly interesting, as the use of an antenna array can be done using the same optical components already installed for the ROF system [128]. In this case, the signal received at the RAP is multiplexed, a single optical carrier being assigned to each antenna element, and then transmitted over the optical fiber, followed by an optical carrier demultiplexing.

In a multiuser channel employing an antenna array at the RAP, the wireless link

can be modeled as a linear mixture and the overall ROF uplink as a MIMO Wiener model. In this case, using the notation introduced in Section 2.3.1:

- $w_t^{(r)}(m), m = 0, 1, ...M$, denotes the channel impulse response of the wireless channel between the t^{th} user and the r^{th} receive antenna;

- $f_{2k+1}^{(r)}$ denotes coefficients of the polynomial function representing the nonlinear E/O conversion device associated with the r^{th} antenna element.

Thus, using the developments of Section 2.3.1, the ROF uplink channel can be modeled as a MIMO Volterra filter like (2.68). Experimental measurements on a ROF link have shown that the channel frequency response can be considered as flat from 1.7 to 2.2 GHz [55, 54]. This means that, with adequate bandwidth, the baseband model of the ROF can be characterized by a memoryless MIMO Volterra filter ($M = 0$):

$$x_r(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{k+1}=t_k}^{T} \sum_{t_{k+2}=0}^{T} \cdots \sum_{t_{2k+1}=t_{2k}}^{T} h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}) \prod_{i=1}^{k+1} s_{t_i}(n) \prod_{i=k+1}^{2k+1} s_{t_i}^*(n). \tag{2.83}$$

Besides, ROF links have been considered by a number of works in the context of Code Division Multiple Access (CDMA) systems [56, 115, 116, 114], as well as for OFDM 802.11a systems [106, 88, 115, 132]. In fact, the representation of the overall ROF channel (5.40) also allows modeling the case where the T sources represent the number of transmitting antennas for one user.

The downlink channel of a ROF can be modeled similarly as the uplink. However, in this case, the wireless channel is placed after the E/O conversion [55]. The overall channel can then be viewed as a Hammerstein system. Besides, at the downlink, the received signals are more subject to nonlinear distortions due to PA saturation than at the uplink.

Channel estimation and information recovery in multiuser nonlinear uplink ROF channels are considered in Chapters 4 and 5, while the case of ROF-CDMA channels is treated in Chapter 6. In fact, the techniques proposed in these chapters are designed for the general MIMO Volterra models, the application of such techniques in ROF systems being considered in the simulations results of these chapters.

2.4.3 Signal processing techniques at the receiver

The digital signal processing techniques proposed in this thesis are based on global Volterra representations of the above studied nonlinear MIMO channels, exploiting the property that the output of a Volterra system is linear with respect to the channel coefficients. In fact, in all the chapters of this work, the knowledge of the global Volterra representation is sufficient to recovery the transmitted symbols, the estimation of the parameters of subsystems such as PA or wireless channel being unnecessary. Indeed, as we will see later, once the global Volterra channel is estimated, the recovery of the transmitted signals can be carried out by means of techniques such as the Viterbi algorithm, Zero Forcing (ZF) and Wiener receivers [118, 74].

This approach based on a global Volterra representation is to be used in the case where the compensation of the nonlinear distortions is carried out at the receiver side, which provides some advantages over pre-distortion schemes that try to compensate nonlinear distortions at the transmitter [39, 31, 148, 38, 120, 81, 3, 103, 129]. The first one is that it allows the global optimization of the problem, i.e. the joint compensation of the distortions due to the linear and nonlinear subsystems that constitute the Volterra channel. For instance, in a ROF system, the joint compensation of the distortions due to the wireless channel and E/O conversion is possible.

Another advantage of compensating the nonlinear distortions at the receiver side is that, in a multiuser uplink transmission, most part of the signal processing is done at the BS, no modification in the portable units being then necessary to accommodate the nonlinearity compensation [57, 115]. Indeed, if the compensation of the nonlinear distortions is done at the transmitter, the associated computational complexity may prohibit their use in small mobile units [67]. In fact, in the considered ROF system, the portable units do not need to be aware of the existence of the ROF uplink [115].

Moreover, the Volterra representation of the overall channel has the advantage of taking into account other possible channel nonlinearities, contrarily to pre-distortion schemes that generally compensates the nonlinear distortions of a single nonlinear block. For instance, the received signals in a ROF channel are also subject to nonlinearities due to PAs [55].

2.5 Conclusion

The main objective of this chapter is to provide a general overview about the modeling of nonlinear MIMO communication channels using Volterra models. The main properties of Volterra systems have been highlighted and, from the expression of a continuous-time passband received signal, we presented the expression of an equivalent baseband discrete-time SISO Volterra channel. Based on this channel representation, we developed some versions of equivalent baseband discrete-time MIMO Volterra models, which are used in this thesis for modeling nonlinear MIMO communication channels.

Furthermore, as these channels can be viewed as MIMO Wiener, Hammerstein or Wiener-Hammerstein models, it was shown that these block-structured nonlinear models can be viewed as special cases of MIMO Volterra models. Some applications of MIMO Volterra models in communication systems have also been described, in particular, MIMO-OFDM, ROF and satellite communication systems.

The developments carried out in this chapter are of great importance due to the lack of works dealing with nonlinear MIMO communication channels. Moreover, it should be highlighted that this chapter contains two main original contributions. The first one is the development of the general expressions (2.53)-(2.57) for equivalent baseband discrete-time MIMO Volterra channels. As earlier mentioned, some works have already used real-valued or less generical versions of MIMO Volterra systems. However, to the best of our knowledge, the equivalent baseband MIMO Volterra models developed in Section 2.2 are unexplored in the literature.

The second original contribution of this chapter is the development concerning MIMO Wiener, Hammerstein and Wiener-Hammerstein models, where we have demonstrated relationships between these block structured MIMO nonlinear systems and MIMO Volterra models. Due to this result, these nonlinear systems can be modeled as particular cases of MIMO Volterra systems.

In the next chapter, we propose techniques for channel estimation and equalization of nonlinear MIMO-OFDM channels described earlier. These techniques make use of the models introduced in the present chapter, considering two different models for the PA. Contrarily to the MIMO Volterra systems used in Chapters 4, 5 and 6, the model considered for the MIMO-OFDM channel does not correspond to a nonlinear mixture of the sources.

CHAPTER 3

Estimation and Equalization of Nonlinear MIMO-OFDM Systems

In this chapter, we develop techniques for estimation and equalization of nonlinear MIMO-OFDM channels. As it was viewed in Section 2.4, a transmitted signal in a OFDM system is characterized by a high peak-to-average power ratio (PAPR) [31, 126, 148, 166, 16, 6, 5], which causes the introduction of nonlinear interferences if the mean power of the transmitted signals is near the saturation region of the power amplifier (PA). It is well-known that, in a OFDM system, the frequency domain received signals do not suffer with interference from other information symbols if a cyclic-prefix is used. However, as it will be detailed later, nonlinear PAs introduce nonlinear inter-carrier interference (ICI) in the received signals, which may significantly deteriorate the recovery of the information symbols. Moreover, in the MIMO case, the received signals are also corrupted by Multiple Access Interference (MAI). A device that cancels or reduces these interferences, such as an equalizer, is then needed. It should be mentioned that the equalizers proposed in this chapter are developed to cancel ICI and MAI, instead of traditional equalizers that cancel inter-symbol interference (ISI).

It should be highlighted that the nonlinear distortions in OFDM systems can be reduced by using a peak power control technique that reduces the PAPR of the OFDM signals [96]. However, this kind of approach will not be considered in this chapter.

In this chapter, we consider two different models for the PA: the memoryless polynomial model (2.80) and the memory polynomial model (2.81), also known

as diagonal Volterra model. When its coefficients are real-valued, the memoryless polynomial model allows the characterization of only amplitude distortions. In this case, the PA is said to be strictly memoryless. However, if its coefficients are complex-valued and the memory of the PA is short compared to the time variations of the input signal envelope, the memoryless polynomial model allows the characterization of amplitude and phase distortions [38, 119, 121]. In this case, the PA is said to be quasi-memoryless. On the other hand, the memory polynomial model [148, 38, 119, 39, 102, 66, 16] allows the characterization of the PA when the memory is not short with respect to the time variations of the input signal envelope.

For simplifying the understanding of the present chapter, we first consider the case of a SISO-OFDM channel with a memoryless polynomial PA (Section 3.2). The presented channel estimation and equalization techniques are based on a global representation of the channel, i.e. a "Volterra-type" channel that characterizes the cascade of the nonlinear PA and the linear wireless link. An expression for the least squares (LS) estimate of the global channel parameters is first given, by using some known pilot symbols. Then, two channel equalization techniques are proposed assuming that these global channel coefficients are known. These techniques, called Zero-Forcing Power Diversity-based Receiver (ZF PDR) and Minimum Mean Square Error (MMSE) PDR, are based on a transmission scheme that re-transmits all the symbols several times with a different transmission power each time. As it will be demonstrated, due to the nonlinear nature of the global channel, the power diversity can be used to provide multi-channels at the reception. Techniques such as the ZF and MMSE receivers can then be used to separate the transmitted symbol from the nonlinear interferences. In fact, the PDRs can be viewed as source separation techniques. The main drawback of the PDRs is the fact that the transmission rate is divided by the repetition factor, i.e. the number of times that every symbol is transmitted. However, as it will be viewed later in the chapter, in many cases it is possible to use a repetition factor equal to 2.

In Section 3.3, we generalize the PDRs for the case of a MIMO-OFDM channel with PAs represented by memoryless polynomial models. In this case, the PAs of the sources (transmit antennas) are possibly characterized by different coefficients. Due to the fact that the nonlinearities are at the transmitters, the signal of each source is corrupted with nonlinear ICI and, then, the signal of all the sources are linearly mixed. We show that the spatial diversity provided by an antenna array at the reception can be exploited to separate the signals of the multiple sources, but it can not be exploited to separate the information signals from the nonlinear interferences. Thus, in this case, the PDRs jointly exploit the spatial diversity, provided by an antenna array, and transmission power diversity, provided by the

transmission scheme, to recover the information symbols.

Section 3.4 demonstrates two results concerning MIMO-OFDM channels assuming memory polynomial PAs. The first one is that a memory polynomial PA in a OFDM system can be expressed as a memoryless polynomial PA with coefficients that vary from one subcarrier to another. Based on this result, we show that the expressions of the frequency domain received signals in terms of the global channel parameters are the same in the case of memoryless and memory polynomial PAs. As a consequence, the channel estimation and equalization techniques proposed in Section 3.3 can be directly applied in the case of memory polynomial PAs. In other words, the case of memory polynomial PAs can be treated in the same way as the case of memoryless polynomial PAs if we use an approach based on the global channel representation.

Most of the techniques proposed to cancel or reduce PA nonlinear distortions in OFDM systems are based on pre-distortion schemes [31, 129, 3, 81, 120, 38, 39, 103]. However, as we saw in Section 2.4.3, this approach has some drawbacks. Techniques for channel estimation and equalization of nonlinear SISO-OFDM channels at the receiver side based on an equivalent baseband Volterra model were proposed in [126]. An iterative equalization method consisting in the estimation of the interference at the receiver and its further cancelation for nonlinear time-varying channels is presented in [41]. In fact, most of the techniques for nonlinear interference rejection at the receiver side of OFDM systems are based on iterative methods as, for instance, [42, 22]. Moreover, it should be highlighted that the theoretical characterization of nonlinear distortions in SISO-OFDM systems with nonlinear PAs has been widely studied in the literature [28, 32, 6, 5, 15, 16].

However, the methods for nonlinear SISO-OFDM systems can not be directly applied to the MIMO case, since all MIMO transmission links exhibit their own nonlinear transfer function, a MIMO detection being then needed [133]. The equalization of nonlinear MIMO-OFDM channels has been treated by a few authors. Several digital nonlinear compensation methods are proposed in [133]. An iterative PA nonlinearity cancelation (PANC) technique is proposed in [67], based on an iterative technique for maximum likelihood detection of nonlinearly distorted symbols [152]. As we will see in the simulation results, the drawback of this iterative method is that it does not work well with memory polynomial PAs, contrarily to the proposed PDRs.

This chapter is organized as follows. Section 3.1 describes a SISO-OFDM communication system assuming a linear PA. In Section 3.2, channel estimation and equalization techniques are proposed for a SISO-OFDM system with a memoryless polynomial PA. Section 3.3 extends these techniques for the case of a MIMO-

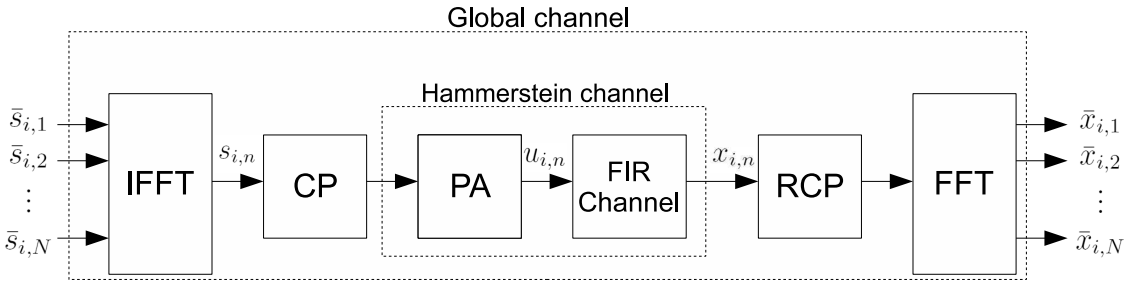

Figure 3.1: Discrete-time equivalent baseband SISO-OFDM system.

OFDM channel with memoryless polynomial PAs. In Section 3.4, we demonstrate the equivalence between the MIMO-OFDM channel in the case of memoryless and memory polynomial PAs when the global channel representation is used. In Section 3.5, we evaluate the performance of these techniques by means of simulations and the conclusions about the chapter are drawn in Section 3.6.

3.1 Linear SISO-OFDM channel

A simplified scheme of the considered discrete-time equivalent baseband SISO-OFDM system is shown in Fig. 3.1. Let N be the number of subcarriers and $\bar{\mathbf{s}}(i) = [\bar{s}_{i,1} \cdots \bar{s}_{i,N}]^T \in \mathbb{C}^{N\times 1}$ the i^{th} vector of frequency domain symbols, where i denotes the transmission block number and $\bar{s}_{i,n}$ the frequency domain symbol at the n^{th}subcarrier and i^{th} transmission block. The frequency domain symbol $\bar{s}_{i,n}$ is assumed to contain the source information and to belong to a QAM alphabet. In this chapter, all the variables with an overline correspond to frequency domain signals. The Fast Fourier Transform (FFT) matrix of dimension N is denoted by $\mathbf{V} \in \mathbb{C}^{N\times N}$, with

$$[\mathbf{V}]_{p,q} = e^{-\mathbf{j}2\pi(p-1)(q-1)/N}, \quad \text{for} \quad 1 \leq p, q \leq N, \tag{3.1}$$

or, equivalently:

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{(N-1)} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{(N-1)} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)(N-1)} \end{pmatrix}, \tag{3.2}$$

where $\omega = e^{-\mathrm{j}2\pi/N}$. The i^{th} vector of time domain symbols $\mathbf{s}(i) = [s_{i,1} \cdots s_{i,N}]^T \in \mathbb{C}^{N\times 1}$ is obtained by means of an Inverse Fast Fourier Transform (IFFT) of the vector of frequency domain symbols as:

$$\mathbf{s}(i) = \mathbf{V}^H \bar{\mathbf{s}}(i). \tag{3.3}$$

Considering that the time domain symbols $s_{i,n}$, for $1 \leq n \leq N$, at the output of the IFFT block are in a time-series, a cyclic prefix of length M_{cp} is inserted in the symbols $s_{i,n}$ in order to ensure that the subcarriers are orthogonal, avoiding intersymbol interference (ISI) and intercarrier interference (ICI). However, this is accomplished only if the time dispersion from the channel is smaller than the duration of the cyclic prefix. In fact, the cyclic prefix is a copy of the last symbols $s_{i,N}$ at the beginning of the transmission block, inserted in the following way:

$$\mathbf{s}^{(cp)}(i) = [s_{i,(N-M_{cp}+1)} \cdots s_{i,N} \;\; \mathbf{s}^T(i)]^T \in \mathbb{C}^{(N+M_{cp})\times 1}, \tag{3.4}$$

or, equivalently:

$$\mathbf{s}^{(cp)}(i) = \Omega\, \mathbf{s}(i), \tag{3.5}$$

where $\Omega \in \mathbb{C}^{(N+M_{cp})\times N}$ is a cyclic prefix insertion matrix given by:

$$\Omega = \begin{bmatrix} \mathbf{0}_{M_{cp},(N-M_{cp})} & \mathbf{I}_{M_{cp}} \\ & \mathbf{I}_N \end{bmatrix}, \tag{3.6}$$

with $\mathbf{0}_{M_{cp},(N-M_{cp})}$ being the null matrix of dimension $M_{cp} \times (N - M_{cp})$ and $\mathbf{I}_N$ the identity matrix of order N.

The time domain symbols are then transmitted through a frequency-selective fading wireless channel with impulse response denoted by w_m, for $0 \leq m \leq M$, where M is the wireless channel memory. The impulse response w_m is assumed to be time-invariant over I_B transmission blocks, that is, for $i = 1, ..., I_B$. At the receiver, the cyclic prefix is removed from the received signals $x_{i,n}$ $(1 \leq n \leq N+M_{cp})$ in the following way:

$$\mathbf{x}(i) = \Theta\, \mathbf{x}^{(cp)}(i), \tag{3.7}$$

where

$$\mathbf{x}^{(cp)}(i) = [x_{i,1} \cdots x_{i,(N+M_{cp})}]^T \in \mathbb{C}^{(N+M_{cp})\times 1}, \tag{3.8}$$

$$\mathbf{x}(i) = [x_{i,(M_{cp}+1)} \cdots x_{i,(N+M_{cp})}]^T \in \mathbb{C}^{N\times 1}, \tag{3.9}$$

and $\Theta \in \mathbb{C}^{N\times(N+M_{cp})}$ is a cyclic prefix removing matrix given by:

$$\Theta = [\mathbf{0}_{N,\,M_{cp}} \quad \mathbf{I}_N]. \tag{3.10}$$

Note that $\mathbf{x}(i)$ is the i^{th} vector of time domain received signals after removing the cyclic prefix.

The transfer function of a linear PA can be represented by a scalar gain that can be absorbed by the channel impulse response w_m. Assuming that length of the cyclic prefix is higher than or equal to the channel memory ($M_{cp} \geq M$), the wireless channel can be represented by a circular convolution:

$$x_{i,(n+M_{cp})} = \sum_{m=0}^{M} w_m s^{(cir)}_{i,(n-m)}, \tag{3.11}$$

for $1 \leq n \leq N$, where $s^{(cir)}_{i,n}$ is a circular version of the time domain signals $s_{i,n}$, i.e.

$$s^{(cir)}_{i,n} = \begin{cases} s_{i,n}, & \text{for} \quad 1 \leq n \leq N, \\ s_{i,N+n}, & \text{for} \quad 1-N \leq n \leq 0. \end{cases} \tag{3.12}$$

Equation (3.11) can be expressed in a vector form as:

$$\mathbf{x}(i) = \mathbf{W}\mathbf{s}(i), \tag{3.13}$$

where $\mathbf{W} \in \mathbb{C}^{N\times N}$ is the circulant channel matrix given by:

$$\mathbf{W} = \begin{bmatrix} w_0 & 0 & \cdots & 0 & w_M & w_{M-1} & \cdots & w_1 \\ w_1 & w_0 & \cdots & 0 & 0 & w_M & \cdots & w_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ w_{M-1} & w_{M-2} & \cdots & w_0 & 0 & \cdots & 0 & w_M \\ w_M & w_{M-1} & \cdots & w_1 & w_0 & 0 & \ldots & 0 \\ 0 & w_M & \cdots & w_2 & w_1 & w_0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & w_M & \cdots & w_1 & w_0 \end{bmatrix}. \tag{3.14}$$

The FFT of the received signals is then calculated as:

$$\bar{\mathbf{x}}(i) = \mathbf{V}\mathbf{x}(i), \tag{3.15}$$

where $\bar{\mathbf{x}}(i) \in \mathbb{C}^{N\times 1}$ is the i^{th} vector of frequency domain received signals. Substi-

tuting (3.3) and (3.13) into (3.15), we get:

$$\bar{\mathbf{x}}(i) = \mathbf{V}\mathbf{W}\mathbf{V}^H\bar{\mathbf{s}}(i). \tag{3.16}$$

It can be shown that a circulant matrix is diagonalized by a FTT matrix, i.e. $\Lambda = \mathbf{V}\mathbf{W}\mathbf{V}^H$, where $\Lambda \in \mathbb{C}^{N\times N}$ is a diagonal matrix containing the eigenvalues of $\mathbf{W}$ [64]. The n^{th} eigenvalue of $\mathbf{W}$ represents the channel frequency response at subcarrier n. Thus, we have:

$$\bar{\mathbf{x}}(i) = \Lambda\bar{\mathbf{s}}(i), \tag{3.17}$$

which shows the orthogonality of the OFDM channel when a cyclic prefix is used.

3.2 SISO-OFDM channel with memoryless polynomial PA

In this section, channel estimation and equalization techniques for a SISO-OFDM channel with a memoryless polynomial PA are proposed. As mentioned in Section 2.4.1, when the PA is represented by a memoryless polynomial model, the global channel (PA + wireless) can be modeled as a Hammerstein system. However, as we will be shown in the sequel, the OFDM transmission scheme provides a simple expression for the received signal in terms of the global channel coefficients. It should be mentioned that the proposed equalizer is to be placed after the FFT block in Fig. 3.1. The considered channel is characterized in Section 3.2.1 and the channel estimation method is presented in Section 3.2.2. Then, in Section 3.2.3 the proposed channel equalization techniques are developed.

3.2.1 Channel characterization

Let us assume that the PA is represented by a memoryless polynomial model of order $2K+1$, the equivalent baseband coefficients denoted by f_{2k+1}, for $0 \leq k \leq K$. Denoting by $u_{i,n}$ $(1 \leq n \leq N + M_{cp})$ the time domain symbols after the PA, we

may write for $1 \leq n \leq N$:

$$\begin{aligned} u_{i,(n+M_{cp})} &= \sum_{k=0}^{K} f_{2k+1}|s_{i,n}|^{2k}s_{i,n} \\ &= \sum_{k=0}^{K} f_{2k+1}\psi_{2k+1}(s_{i,n}), \end{aligned} \tag{3.18}$$

where the operator $\psi_{2k+1}(\cdot)$ is defined as:

$$\psi_{2k+1}(a) = |a|^{2k}a. \tag{3.19}$$

Note that the signal $u_{i,n}$ also contains a cyclic prefix: $\{u_{i,1}, \cdots, u_{i,M_{cp}}\}$ is equal to $\{u_{i,(N-M_{cp}+1)}, \cdots, u_{i,N}\}$. Let us define i^{th} vector of time domain symbols after the PA as:

$$\mathbf{u}(i) = [u_{i,(M_{cp}+)1} \cdots u_{i,(M_{cp}+N)}]^T \in \mathbb{C}^{N\times 1}. \tag{3.20}$$

Thus, we have:

$$\mathbf{u}(i) = \sum_{k=0}^{K} f_{2k+1}\Psi_{2k+1}(\mathbf{s}(i)), \tag{3.21}$$

where

$$\Psi_{2k+1}(\mathbf{a}) = [\psi_{2k+1}(a_1) \cdots \psi_{2k+1}(a_N)]^T \in \mathbb{C}^{N\times 1}, \tag{3.22}$$

for $\mathbf{a} = [a_1 \cdots a_N] \in \mathbb{C}^{N\times 1}$.

In this case, the i^{th} vector of time domain received signals after removing the cyclic prefix is given by:

$$\mathbf{x}(i) = \mathbf{W}\mathbf{u}(i). \tag{3.23}$$

Taking the FFT of (3.23), the i^{th} vector of frequency domain received signals can be expressed as:

$$\begin{aligned} \bar{\mathbf{x}}(i) &= \mathbf{V}\mathbf{W}\mathbf{u}(i) \\ &= \mathbf{V}\mathbf{W}\mathbf{V}^H\bar{\mathbf{u}}(i) = \Lambda\bar{\mathbf{u}}(i), \end{aligned} \tag{3.24}$$

where $\Lambda = \mathbf{V}\mathbf{W}\mathbf{V}^H \in \mathbb{C}^{N\times N}$ and

$$\bar{\mathbf{u}}(i) = \mathbf{V}\mathbf{u}(i) \tag{3.25}$$

is the frequency domain version of $\mathbf{u}(i)$. From (3.21), we may then write:

$$\begin{aligned}\bar{\mathbf{u}}(i) &= \sum_{k=0}^{K} f_{2k+1}\mathbf{V}\Psi_{2k+1}(\mathbf{s}(i)) \\ &= \sum_{k=0}^{K} f_{2k+1}\bar{\Psi}_{2k+1}(\mathbf{s}(i)),\end{aligned} \tag{3.26}$$

where $\bar{\Psi}_{2k+1}(\mathbf{s}(i)) \in \mathbb{C}^{N\times 1}$ is the frequency domain version of $\Psi_{2k+1}(\mathbf{s}(i))$, i.e.

$$\bar{\Psi}_{2k+1}(\mathbf{a}) \equiv \mathbf{V}\Psi_{2k+1}(\mathbf{a}). \tag{3.27}$$

Substituting (3.26) into (3.24), we get:

$$\bar{\mathbf{x}}(i) = \Lambda \sum_{k=0}^{K} f_{2k+1}\bar{\Psi}_{2k+1}(\mathbf{s}(i)). \tag{3.28}$$

Equation (3.28) can be written in a scalar form as:

$$\bar{x}_{i,n} = \sum_{k=0}^{K} \lambda_n f_{2k+1}\bar{\psi}_{2k+1}(s_{i,n}), \tag{3.29}$$

for $1 \leq n \leq N$, where $\lambda_n = [\Lambda]_{n,n}$, $\bar{x}_{i,n} = [\bar{\mathbf{x}}(i)]_n$ is the frequency domain received signal and $\bar{\psi}_{2k+1}(s_{i,n}) = [\bar{\Psi}_{2k+1}(\mathbf{s}(i))]_n$ is the frequency domain version of $\psi_{2k+1}(s_{i,n})$. A global channel representation including the PA and the wireless channel coefficients can be defined from (3.29) as:

$$h_{n,k+1} = \lambda_n f_{2k+1}, \tag{3.30}$$

for $1 \leq n \leq N$ and $0 \leq k \leq K$, which gives

$$\bar{x}_{i,n} = \sum_{k=0}^{K} h_{n,k+1}\bar{\psi}_{2k+1}(s_{i,n}). \tag{3.31}$$

Equation (3.31) can be expressed as:

$$\bar{x}_{i,n} = h_{n,1}\bar{s}_{i,n} + \sum_{k=1}^{K} h_{n,k+1}\bar{\psi}_{2k+1}(s_{i,n}), \tag{3.32}$$

showing that the frequency domain received signal $\bar{x}_{i,n}$ is a scaled version of the

information signal $\bar{s}_{i,n}$ plus the nonlinear ICI. It should be remarked that the term $\bar{\psi}_{2k+1}(s_{i,n})$ depends on information symbols of other subcarriers, which means that a given subcarrier interferes in other subcarriers. This phenomenon can be viewed as consequence of the spectral broadening provided by the nonlinear PA.

Note also that $\bar{x}_{i,n}$ is not corrupted with interferences from other information symbols $\bar{s}_{i',n}$, for $i' \neq i$. In this case, an equalizer should be used to eliminate the nonlinear ICI terms $\sum_{k=1}^{K} h_{n,k+1}\bar{\psi}_{2k+1}(s_{i,n})$ and to remove the scalar factor $h_{n,1}$.

Let us consider a set of I_B transmission blocks, i.e. I_B information symbols per subcarrier, and by $\bar{\mathbf{x}}_n = [\bar{x}_{1,n} \cdots \bar{x}_{I_B,n}] \in \mathbb{C}^{1\times I_B}$ the row-vector containing the I_B frequency domain signals received at the n^{th} subcarrier. Equation (3.31) can then be rewritten as:

$$\bar{\mathbf{x}}_n = \mathbf{h}_n^T \bar{\mathbf{S}}_n, \tag{3.33}$$

where

$$\mathbf{h}_n = [h_{n,1}\ h_{n,3} \cdots h_{n,2K+1}]^T \in \mathbb{C}^{(K+1)\times 1} \tag{3.34}$$

is a vector containing the global channel coefficients of the n^{th} subcarrier and $\bar{\mathbf{S}}_n \in \mathbb{C}^{(K+1)\times I_B}$ is defined as:

$$\bar{\mathbf{S}}_n = \begin{bmatrix} \bar{s}_{1,n} & \cdots & \bar{s}_{I_B,n} \\ \bar{\psi}_3(s_{1,n}) & \cdots & \bar{\psi}_3(s_{I_B,n}) \\ \vdots & \ddots & \vdots \\ \bar{\psi}_{2K+1}(s_{1,n}) & \cdots & \bar{\psi}_{2K+1}(s_{I_B,n}) \end{bmatrix}. \tag{3.35}$$

As the equalizers proposed in this chapter assume that the global channel coefficients are known, a channel estimation technique is then needed. Equations (3.31) and (3.33) show the linear dependence between the received signals and the global channel coefficients, which can then be estimated during a training period by means of a LS approach. In the sequel, channel estimation of the global channel coefficients is addressed.

3.2.2 Channel Estimation

Channel estimation of OFDM channels is usually done by using one of two different approaches [67]. The first one assumes that pilot symbols are allocated on all subcarriers and the second one assumes that pilot symbols are allocated on a

reduced set of subcarriers. In the second case, the channel coefficients on all the subcarriers are estimated using interpolation. The channel estimation techniques proposed in this chapter consider the second case, the number of pilot subcarriers being denoted by N_P, with $N_P \leq N$. In the simulation results of this chapter, the channel $h_{n,k+1}$ is interpolated using truncated FFT matrices. See [67, 66] for further details about interpolation using the FFT method.

The channel estimation technique presented in the sequel is based on the global channel representation (3.33) that includes the coefficients of the PA and the wireless channel. The LS estimate of these global channel coefficients can be done by using I_P pilot symbols per pilot subcarrier $\{\bar{s}_{1,n}^{(p)}, \cdots, \bar{s}_{I_P,n}^{(p)}\}$, for $1 \leq n \leq N_P$. Let

$$\bar{\mathbf{x}}_n^{(p)} = [\bar{x}_{1,n}^{(p)} \cdots \bar{x}_{I_P,n}^{(p)}] \in \mathbb{C}^{1\times I_P} \tag{3.36}$$

be the row-vector containing the I_P frequency domain signals received during the training period at the n^{th} pilot subcarrier. From (3.33), the LS estimate of the global channel vector $\mathbf{h}_n$ can be done as:

$$\hat{\mathbf{h}}_n = \left[\bar{\mathbf{x}}_n^{(p)} \left[\bar{\mathbf{S}}_n^{(p)}\right]^{\dagger}\right]^T, \tag{3.37}$$

where $(\cdot)^{\dagger}$ denotes the matrix pseudo-inverse and $\bar{\mathbf{S}}_n^{(p)} \in \mathbb{C}^{(K+1)\times I_P}$ is defined as:

$$\bar{\mathbf{S}}_n^{(p)} = \begin{bmatrix} \bar{s}_{1,n}^{(p)} & \cdots & \bar{s}_{I_P,n}^{(p)} \\ \bar{\psi}_3(s_{1,n}^{(p)}) & \cdots & \bar{\psi}_3(s_{I_P,n}^{(p)}) \\ \vdots & \ddots & \vdots \\ \bar{\psi}_{2K+1}(s_{1,n}^{(p)}) & \cdots & \bar{\psi}_{2K+1}(s_{I_P,n}^{(p)}) \end{bmatrix}, \tag{3.38}$$

with $s_{i,n}^{(p)}$ being the time domain version of $\bar{s}_{i,n}^{(p)}$.

Note that, a necessary condition for (3.37) is $I_P \geq K+1$. However, based on a realistic assumption that the PA can be modeled using a third-order polynomial ($K+1=2$) [15, 16, 39, 95, 66, 38, 119, 3], we can use $I_P = 2$ pilot symbols per subcarrier.

3.2.3 Power Diversity-based Receivers (PDRs)

In this section, we propose two receivers for the nonlinear OFDM channel described in Section 3.2.1 assuming that channel coefficients $h_{n,k+1}$ are known, for $1 \leq n \leq N$

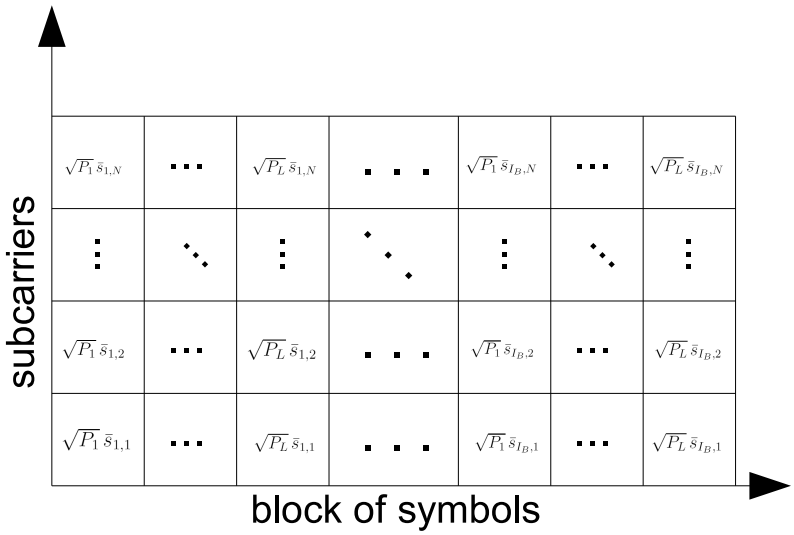

Figure 3.2: PDR transmission scheme.

and $0 \leq k \leq K+1$. However, as we will see, of the proposed techniques use only the linear coefficients $h_{n,1}$ of the channel.

Transmission scheme

The PDRs are based on the transmission scheme shown in Fig. 3.2. For a given subcarrier n $(1 \leq n \leq N)$ and transmission block i $(1 \leq i \leq I_B)$, we have:

$$\bar{s}^{(pd)}_{((i-1)L+l),n} = \sqrt{P_l}\,\bar{s}_{i,n}, \quad \text{for} \quad 1 \leq l \leq L, \tag{3.39}$$

where $\bar{s}^{(pd)}_{k,n}$ is the k^{th} frequency domain transmitted signal associated with the n^{th} subcarrier, L is the repetition factor and $P_1, ..., P_L$ are the transmission powers. Equation (3.39) can be written in a vector form as:

$$\bar{\mathbf{s}}^{(pd)}((i-1)L+l) = \bar{\mathbf{s}}(i)\sqrt{P_l}, \quad \text{for} \quad 1 \leq l \leq L, \tag{3.40}$$

where $\bar{\mathbf{s}}^{(pd)}(k) = [\bar{s}^{(pd)}_{k,1} \ \cdots \ \bar{s}^{(pd)}_{k,N}]^T \in \mathbb{C}^{N\times 1}$ is the k^{th} vector of frequency domain transmitted symbols. This means that each block $\bar{\mathbf{s}}(i) \in \mathbb{C}^{N\times 1}$ of frequency domain QAM-modulated information symbols is repeated L times using transmission powers equal to $P_1, ..., P_L$.

Note that, for each subcarrier, one frequency domain information symbol $\bar{s}_{i,n}$ ge-

nerates a set of L frequency domain transmitted symbols $\bar{s}^{(pd)}_{((i-1)L+l),n}$ $(1 \leq l \leq L)$ and, hence, a set of L frequency domain received signals, denoted by $\bar{x}^{(pd)}_{((i-1)L+l),n}$. Let us define

$$\bar{\mathbf{x}}^{(pd)}_{i,n} = [\bar{x}^{(pd)}_{((i-1)L+1),n} \cdots \bar{x}^{(pd)}_{iL,n}]^T \in \mathbb{C}^{L\times 1} \tag{3.41}$$

as the column-vector containing the L frequency domain received signals at the n^{th} subcarrier associated with the frequency domain information symbol $\bar{s}_{i,n}$. From (3.33), we have:

$$\bar{\mathbf{x}}^{(pd)}_{i,n} = \left[\bar{\mathbf{S}}^{(pd)}_{i,n}\right]^T \mathbf{h}_n, \tag{3.42}$$

where $\bar{\mathbf{S}}^{(pd)}_{i,n} \in \mathbb{C}^{(K+1)\times L}$ is defined as:

$$\bar{\mathbf{S}}^{(pd)}_{i,n} = \begin{bmatrix} \bar{s}^{(pd)}_{((i-1)L+1),n} & \bar{s}^{(pd)}_{((i-1)L+2),n} & \cdots & \bar{s}^{(pd)}_{iL,n} \\ \bar{\psi}_3(s^{(pd)}_{((i-1)L+1),n}) & \bar{\psi}_3(s^{(pd)}_{((i-1)L+2),n}) & \cdots & \bar{\psi}_3(s^{(pd)}_{iL,n}) \\ \vdots & & \ddots & \vdots \\ \bar{\psi}_{2K+1}(s^{(pd)}_{((i-1)L+1),n}) & \bar{\psi}_{2K+1}(s^{(pd)}_{((i-1)L+2),n}) & \cdots & \bar{\psi}_{2K+1}(s^{(pd)}_{iL,n}) \end{bmatrix}. \tag{3.43}$$

By using (3.39), the matrix $\bar{\mathbf{S}}^{(pd)}_{i,n}$ can be rewritten as:

$$\bar{\mathbf{S}}^{(pd)}_{i,n} = \text{diag}\left[\bar{\mathbf{s}}_{i,n}\right] \mathbf{P}^T, \tag{3.44}$$

where diag$[\cdot]$ denotes the diagonal matrix formed from the vector argument, $\bar{\mathbf{s}}_{i,n} = [\bar{s}_{i,n}, \bar{\psi}_3(s_{i,n}), \cdots, \bar{\psi}_{2K+1}(s_{i,n})]^T \in \mathbb{C}^{(K+1)\times 1}$ and

$$\mathbf{P} = \begin{bmatrix} P_1^{\frac{1}{2}} & \cdots & P_1^{\frac{2K+1}{2}} \\ \vdots & \ddots & \vdots \\ P_L^{\frac{1}{2}} & \cdots & P_L^{\frac{2K+1}{2}} \end{bmatrix} \in \mathbb{C}^{L\times(K+1)}. \tag{3.45}$$

Substituting (3.44) into (3.42), we get:

$$\begin{aligned} \bar{\mathbf{x}}^{(pd)}_{i,n} &= \mathbf{P}\,\text{diag}\left[\bar{\mathbf{s}}_{i,n}\right] \mathbf{h}_n \\ &= \mathbf{P}\,\text{diag}\left[\mathbf{h}_n\right] \bar{\mathbf{s}}_{i,n}. \end{aligned} \tag{3.46}$$

By defining

$$\bar{\mathbf{X}}_n^{(pd)} = [\bar{\mathbf{x}}_{1,n}^{(pd)} \cdots \bar{\mathbf{x}}_{I_B,n}^{(pd)}] \in \mathbb{C}^{L \times I_B}, \tag{3.47}$$

we can deduce from (3.46):

$$\bar{\mathbf{X}}_n^{(pd)} = \mathbf{P}\,\mathrm{diag}\,[\mathbf{h}_n]\,[\bar{\mathbf{s}}_{1,n} \cdots \bar{\mathbf{s}}_{I_B,n}]\,, \tag{3.48}$$

or, equivalently,

$$\bar{\mathbf{X}}_n^{(pd)} = \mathbf{P}\,\mathrm{diag}\,[\mathbf{h}_n]\,\bar{\mathbf{S}}_n, \tag{3.49}$$

where $\bar{\mathbf{S}}_n \in \mathbb{C}^{(K+1)\times I_B}$ is defined in (3.35). By comparing (3.33) with (3.49), it can be seen that re-transmitting the symbols $\bar{s}_{i,n}$ L times with different amplitudes has the effect of creating L sub-channels for the nonlinear OFDM system, the equivalent channel matrix given by $(\mathbf{P}\,\mathrm{diag}\,[\mathbf{h}_n])$. It can then be concluded that the power variation can be viewed as diversity for the received signals.

Equalization

Due to the multi-channel representation above described, the frequency domain information signals $\bar{s}_{i,n}$ can be recovered from the matrix of frequency domain received signals $\bar{\mathbf{X}}_n^{(pd)}$ by means of a multi-channel processing technique. Several techniques can be used for this purpose as, for instance, the ZF and MMSE receivers. In this case, the MMSE and ZF receivers are respectively given by:

$$\mathbf{W}_n = \mathbf{R}_{\bar{\mathbf{s}}}\,\mathrm{diag}\,[\mathbf{h}_n^*]\;\mathbf{P}^H\left(\mathbf{P}\,\mathrm{diag}\,[\mathbf{h}_n]\,\mathbf{R}_{\bar{\mathbf{s}}}\,\mathrm{diag}\,[\mathbf{h}_n^*]\;\mathbf{P}^H + \mathbf{I}_L\sigma^2\right)^{-1} \in \mathbb{C}^{(K+1)\times L}, \tag{3.50}$$

and

$$\mathbf{W}_n = (\mathbf{P}\,\mathrm{diag}\,[\mathbf{h}_n])^{\dagger} \in \mathbb{C}^{(K+1)\times L}, \tag{3.51}$$

where σ^2 is the noise variance and $\mathbf{R}_{\bar{\mathbf{s}}} = \mathbb{E}[\bar{\mathbf{s}}_{i,n}\bar{\mathbf{s}}_{i,n}^H] \in \mathbb{C}^{(K+1)\times(K+1)}$ is the covariance matrix of the vector $\bar{\mathbf{s}}_{i,n}$, the matrix $\bar{\mathbf{S}}_n$ being estimated as:

$$\hat{\bar{\mathbf{S}}}_n \;=\; \mathbf{W}_n\bar{\mathbf{X}}_n^{(pd)}. \tag{3.52}$$

As we can see in (3.35), the first row of $\hat{\bar{\mathbf{S}}}_n$ corresponds to the linear terms, which means that frequency domain symbols $\bar{s}_{i,n}$ can be estimated from the first row of $\hat{\bar{\mathbf{S}}}_n$. Moreover, it can be shown after some manipulations that expressions (3.50) and (3.51) are equivalent when $\sigma^2 = 0$.

The main advantage of the MMSE receiver is that it is more robust to noise than the ZF receiver. However, in order to use the MMSE receiver, we have to assume that the noise variance is known, contrarily to the ZF receiver. Besides, as it is shown in the sequel, the use of a ZF receiver allows a great simplification in the estimation of the information signals $\bar{s}_{i,n}$. Indeed, by replacing (3.51) into (3.52), we get:

$$\hat{\bar{\mathbf{S}}}_n = \mathrm{diag}\left[\mathcal{I}(\mathbf{h}_n)\right] \mathbf{P}^{\dagger} \bar{\mathbf{X}}_n^{(pd)} \quad \in \mathbb{C}^{(K+1)\times I_B}, \tag{3.53}$$

where $\mathcal{I}(\mathbf{a})$ is an operation that inverses each element of the vector $\mathbf{a}$ in the argument (element-wise inverse), i.e. $\left[\mathcal{I}(\mathbf{a})\right]_k = 1/\left[\mathbf{a}\right]_k$. In fact, denoting by $\tilde{\mathbf{p}} \in \mathbb{C}^{1\times L}$ the first row of $\mathbf{P}^{\dagger}$, the first row of $\hat{\bar{\mathbf{S}}}_n$ can be expressed as:

$$[\hat{\bar{\mathbf{S}}}_n]_{1,\cdot} = \frac{\tilde{\mathbf{p}}}{h_{n,1}} \bar{\mathbf{X}}_n^{(pd)} \quad \in \mathbb{C}^{1\times I_B}, \tag{3.54}$$

The recovery of the information signals of the N subcarriers and I_B blocks can be done by defining

$$\bar{\mathbf{X}}^{(pd)} = \left[\bar{\mathbf{X}}_1^{(pd)} \cdots \bar{\mathbf{X}}_N^{(pd)}\right] \in \mathbb{C}^{L\times I_B N} \tag{3.55}$$

and

$$\hat{\bar{\mathbf{s}}} = \left[[\hat{\bar{\mathbf{S}}}_1]_{1,\cdot} \cdots [\hat{\bar{\mathbf{S}}}_N]_{1,\cdot}\right] \in \mathbb{C}^{1\times I_B N}. \tag{3.56}$$

A compact form of (3.54) is then given by:

$$\begin{aligned} \hat{\bar{\mathbf{s}}} &= \tilde{\mathbf{p}} \left[h_{1,1}^{-1} \bar{\mathbf{X}}_1^{(pd)} \cdots h_{N,1}^{-1} \bar{\mathbf{X}}_N^{(pd)}\right] \\ &= \tilde{\mathbf{p}} \bar{\mathbf{X}}^{(pd)} \,\mathrm{diag}\left(\mathcal{I}(\mathbf{h}^{(lin)}) \otimes \mathbf{1}_{I_B}\right), \end{aligned} \tag{3.57}$$

where $\otimes$ denotes the Kronecker product, $\mathbf{1}_{I_B} \in \mathbb{R}^{I_B\times 1}$ is the all ones vector of dimension I_B and

$$\mathbf{h}^{(lin)} = [h_{1,1} \cdots h_{N,1}]^T \in \mathbb{C}^{N\times 1}. \tag{3.58}$$

The ZF- and MMSE-PDR expressions are summarized in Tables 3.1 and 3.2, respectively. Note that, a necessary condition for these techniques is that $\mathbf{P}$ be full column-rank, which means that the repetition factor must satisfy $L \geq K+1$. An important drawback of these equalization methods is the fact that the transmission

Table 3.1: Minimum Mean Square Error-Power Diversity-based Receiver (MMSE-PDR)

Transmission scheme

For $1 \leq i \leq I_B$ and $1 \leq l \leq L$:

$\bar{\mathbf{s}}^{(pd)}((i-1)L+l) = \bar{\mathbf{s}}(i)\sqrt{P_l} \quad \in \mathbb{C}^{N\times 1}$

Equalization:

For $1 \leq n \leq N$:

1) Construct $\bar{\mathbf{X}}_n^{(pd)} = [\bar{\mathbf{x}}_{1,n}^{(pd)} \cdots \bar{\mathbf{x}}_{I_B,n}^{(pd)}] \in \mathbb{C}^{L\times I_B}$, where $\bar{\mathbf{x}}_{i,n}^{(pd)} = [\bar{x}_{((i-1)L+1),n}^{(pd)} \cdots \bar{x}_{iL,n}^{(pd)}]^T \in \mathbb{C}^{L\times 1}$.

2) Calculate $\hat{\bar{\mathbf{S}}}_n = \mathbf{R}_{\bar{\mathbf{s}}}\,\mathrm{diag}\,[\mathbf{h}_n^*]\;\mathbf{P}^H\left(\mathbf{P}\,\mathrm{diag}\,[\mathbf{h}_n]\,\mathbf{R}_{\bar{\mathbf{s}}}\,\mathrm{diag}\,[\mathbf{h}_n^*]\;\mathbf{P}^H + \mathbf{I}_L\sigma^2\right)^{-1}\bar{\mathbf{X}}_n^{(pd)}$ $\in \mathbb{C}^{(K+1)\times I_B}$.

3) The estimate of the I_B frequency domain symbols associated with the n^{th} subcarrier is the first row of $\hat{\bar{\mathbf{S}}}_n$.

rate is divided by L. However, note that when $K+1=2$ (third-order polynomial nonlinearity) [15, 16, 39, 95, 66, 38, 119, 3], we can use $L=2$.

An interesting characteristic of the proposed ZF equalization method is that it uses only the coefficients $h_{n,1}$ associated with the linear terms. This means that we do not need to know the coefficients of the PA, excepting linear one f_1. Moreover, when $K+1=L=2$, the ZF-PDR technique needs the computation of only one inverse matrix of dimensions 2×2 for all $1 \leq n \leq N$ and $1 \leq i \leq I_B$, which means that the ZF receiver has a smaller computational cost than the MMSE receiver.

In fact, it should be highlighted that the above proposed channel estimation and equalization techniques can also be applied when the PA is modeled as a Hammerstein system. That is due to the fact that the impulse response of the FIR filter corresponding to the PA can be incorporated with the impulse response of the wireless channel.

Table 3.2: Zero-Forcing Power Diversity-based Receiver (ZF-PDR)

Transmission scheme

For $1 \leq i \leq I_B$ and $1 \leq l \leq L$:

$\bar{\mathbf{s}}^{(pd)}((i-1)L+l) = \bar{\mathbf{s}}(i)\sqrt{P_l} \ \in \mathbb{C}^{N\times 1}$

Equalization:

1) Construct $\mathbf{P}$ from (3.45) and calculate $\tilde{\mathbf{p}}$ as the fitst row of $\mathbf{P}^{\dagger}$.

2) Construct $\bar{\mathbf{X}}^{(pd)} = \left[\bar{\mathbf{X}}_1^{(pd)} \cdots \bar{\mathbf{X}}_N^{(pd)}\right] \in \mathbb{C}^{L\times I_B N}$, where $\bar{\mathbf{X}}_n^{(pd)} = [\bar{\mathbf{x}}_{1,n}^{(pd)} \cdots \bar{\mathbf{x}}_{I_B,n}^{(pd)}] \in \mathbb{C}^{L\times I_B}$ and $\bar{\mathbf{x}}_{i,n}^{(pd)} = [\bar{x}_{((i-1)L+1),n}^{(pd)} \cdots \bar{x}_{iL,n}^{(pd)}]^T \in \mathbb{C}^{L\times 1}$.

3) The estimate of the information signals of the N subcarriers and I_B blocks is given by: $\hat{\bar{\mathbf{s}}} = \tilde{\mathbf{p}}\bar{\mathbf{X}}^{(pd)} \operatorname{diag}\left(\mathcal{I}(\mathbf{h}^{(lin)}) \otimes \mathbf{1}_{I_B}\right) \in \mathbb{C}^{1\times I_B N}$.

3.3 MIMO-OFDM channel with memoryless polynomial PAs

This section extends the above channel estimation and equalization techniques for the case of a MIMO-OFDM system assuming memoryless polynomial PAs. As well as in Section 3.2, the equalizer proposed in this section should be placed after the FFT stage. The channel characterization, estimation and equalization are treated respectively in Sections 3.3.1, 3.3.2 and 3.3.3.

3.3.1 Channel characterization

Let us consider that the PA of each source is nonlinear and represented by a memoryless polynomial model of order $2K+1$, with equivalent baseband coefficients denoted by $f_{2k+1,t}$, for $0 \leq k \leq K$ and $1 \leq t \leq T$, where T is the number of sources. Besides, let us denote respectively by $\bar{\mathbf{s}}(i,t) \in \mathbb{C}^{N\times 1}$ and $\mathbf{s}(i,t) \in \mathbb{C}^{N\times 1}$ the i^{th} vector of frequency and time domain symbols of the t^{th} user, with components $\bar{s}_{i,n,t} = [\bar{\mathbf{s}}(i,t)]_n$ and $s_{i,n,t} = [\mathbf{s}(i,t)]_n$.

As the signals from the multiple source are linearly mixed, from (3.24), the i^{th}

vector $\bar{\mathbf{x}}(i,r) \in \mathbb{C}^{N\times 1}$ of frequency domain signals received by the r^{th} antenna, for $1 \leq r \leq R$, can be written as:

$$\bar{\mathbf{x}}(i,r) = \sum_{t=1}^{T} \Lambda(r,t)\bar{\mathbf{u}}(i,t), \tag{3.59}$$

where R is the number of receive antennas,

$$\bar{\mathbf{u}}(i,t) = \mathbf{V}\mathbf{u}(i,t) \in \mathbb{C}^{N\times 1} \tag{3.60}$$

is a vector containing the frequency domain version of t^{th} PA output, with

$$\mathbf{u}(i,t) = [u_{i,(M_{cp}+1),t} \cdots u_{i,(M_{cp}+N),t}]^T \in \mathbb{C}^{N\times 1}, \tag{3.61}$$

and

$$\Lambda(r,t) = \mathbf{V}\mathbf{W}(r,t)\mathbf{V}^H \in \mathbb{C}^{N\times N} \tag{3.62}$$

is a diagonal matrix containing the eigenvalues of circulant channel matrix $\mathbf{W}(r,t)$, constructed as in (3.14), using the impulse response of the wireless channel $w_m(r,t)$ $(0 \leq m \leq M)$ associated with the r^{th} receive antenna and the t^{th} source, for $1 \leq r \leq R$ and $1 \leq t \leq T$. The main diagonal of $\Lambda(r,t)$ contains N samples of the frequency response of the wireless channel between source t and receive antenna r.

From (3.26), we have:

$$\bar{\mathbf{u}}(i,t) = \sum_{k=0}^{K} f_{2k+1,t}\bar{\Psi}_{2k+1}(\mathbf{s}(i,t)), \tag{3.63}$$

where

$$\bar{\Psi}_{2k+1}(\mathbf{s}(i,t)) = \mathbf{V}\Psi_{2k+1}(\mathbf{s}(i,t)) \in \mathbb{C}^{N\times 1}. \tag{3.64}$$

Thus substituting (3.63) into (3.59), we get:

$$\bar{\mathbf{x}}(i,r) = \sum_{t=1}^{T} \Lambda(r,t) \sum_{k=0}^{K} f_{2k+1,t}\bar{\Psi}_{2k+1}(\mathbf{s}(i,t)). \tag{3.65}$$

Equation (3.65) can be expressed in a scalar form as:

$$\bar{x}_{i,n,r} = \sum_{t=1}^{T} \sum_{k=0}^{K} \lambda_{n,r,t} f_{2k+1,t} \bar{\psi}_{2k+1}(s_{i,n,t}), \tag{3.66}$$

for $0 \leq i \leq I_B$, $1 \leq n \leq N$ and $1 \leq r \leq R$, where $\lambda_{n,r,t} = [\Lambda(r,t)]_{n,n}$, $\bar{x}_{i,n,r} = [\bar{\mathbf{x}}(i,r)]_n$ and $\bar{\psi}_{2k+1}(s_{i,n,t}) = [\bar{\Psi}_{2k+1}(\mathbf{s}(i,t))]_n$. By defining:

$$h_{n,k+1,r,t} = \lambda_{n,r,t}\, f_{2k+1,t}, \tag{3.67}$$

for $1 \leq n \leq N$, $0 \leq k \leq K$, $1 \leq r \leq R$ and $1 \leq t \leq T$, a global channel representation of (3.66) can be obtained:

$$\bar{x}_{i,n,r} = \sum_{t=1}^{T} \sum_{k=0}^{K} h_{n,k+1,r,t}\ \bar{\psi}_{2k+1}(s_{i,n,t}). \tag{3.68}$$

Equation (3.68) can be expressed as:

$$\bar{x}_{i,n,r} = h_{n,1,r,1} \bar{s}_{i,n,1} + \sum_{k=1}^{K} h_{n,k+1,r,1}\ \bar{\psi}_{2k+1}(s_{i,n,1}) + \sum_{t=2}^{T} \sum_{k=0}^{K} h_{n,k+1,r,t}\ \bar{\psi}_{2k+1}(s_{i,n,t}), \tag{3.69}$$

showing that the frequency domain received signal $\bar{x}_{i,n}$ can be viewed as a scaled version of the information signal of the first user $\bar{s}_{i,n,1}$ corrupted with MAI and nonlinear ICI from itself.

Similarly as in (3.33), denoting by $\bar{\mathbf{X}}_n \in \mathbb{C}^{R \times I_B}$, with $[\bar{\mathbf{X}}_n]_{r,i} = \bar{x}_{i,n,r}$, for $1 \leq i \leq I_B$ and $1 \leq r \leq R$, the matrix containing the I_B frequency domain signals received by all the R antennas at the n^{th} subcarrier, equation (3.68) may be written in a matrix form as:

$$\bar{\mathbf{X}}_n = \mathbf{H}_n\, \bar{\mathbf{S}}_n, \tag{3.70}$$

where $\bar{\mathbf{S}}_n \in \mathbb{C}^{(K+1)T \times I_B}$ is defined as:

$$\bar{\mathbf{S}}_n = \begin{bmatrix} \bar{s}_{1,n,1} & \cdots & \bar{s}_{I_B,n,1} \\ \bar{\psi}_3(s_{1,n,1}) & \cdots & \bar{\psi}_3(s_{I_B,n,1}) \\ \vdots & \ddots & \vdots \\ \bar{\psi}_{2K+1}(s_{1,n,1}) & \cdots & \bar{\psi}_{2K+1}(s_{I_B,n,1}) \\ & \vdots & \\ \bar{s}_{1,n,T} & \cdots & \bar{s}_{I_B,n,T} \\ \bar{\psi}_3(s_{1,n,T}) & \cdots & \bar{\psi}_3(s_{I_B,n,T}) \\ \vdots & \ddots & \vdots \\ \bar{\psi}_{2K+1}(s_{1,n,T}) & \cdots & \bar{\psi}_{2K+1}(s_{I_B,n,T}) \end{bmatrix}, \tag{3.71}$$

and $\mathbf{H}_n \in \mathbb{C}^{R \times (K+1)T}$ contains the global channel coefficients of the n^{th} subcarrier, with $[\mathbf{H}_n]_{r,((t-1)(K+1)+k+1)} = h_{n,k+1,r,t}$, for $0 \leq k \leq K$, $1 \leq r \leq R$ and $1 \leq t \leq T$.

Note that the matrix $\mathbf{H}_n$ can be expressed as:

$$\mathbf{H}_n = \Lambda_n \, \mathbf{F}, \tag{3.72}$$

with

$$\mathbf{F} = \begin{bmatrix} \mathbf{f}_1^T & \cdots & \mathbf{0}_{K+1,1}^T \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{K+1,1}^T & \cdots & \mathbf{f}_T^T \end{bmatrix} \in \mathbb{C}^{T \times (K+1)T}, \tag{3.73}$$

where $\Lambda_n \in \mathbb{C}^{R \times T}$ is the linear channel matrix, with $[\Lambda_n]_{r,t} = \lambda_{n,r,t}$, for $1 \leq r \leq R$ and $1 \leq t \leq T$, and $\mathbf{f}_t = [f_{1,t} \; f_{3,t} \cdots f_{2K+1,t}]^T \in \mathbb{C}^{(K+1)\times 1}$ contains the PA coefficients associated with the t^{th} source. As the rank of $\mathbf{F}$ is equal to T, it can be concluded that the rank of the channel matrix $\mathbf{H}_n$ can not be higher than T. Thus, the matrix $\bar{\mathbf{S}}_n$ can not be estimated by using the pseudo-inverse of $\mathbf{H}_n$, which means that the spatial diversity provided by the antenna array can not be exploited to recover the information signals $\bar{s}_{i,n,t}$.

However, substituting (3.72) into (3.70), we get:

$$\bar{\mathbf{X}}_n = \Lambda_n \, \bar{\mathbf{U}}_n, \tag{3.74}$$

where

$$\bar{\mathbf{U}}_n = \mathbf{F} \, \bar{\mathbf{S}}_n \in \mathbb{C}^{T \times I_B}, \tag{3.75}$$

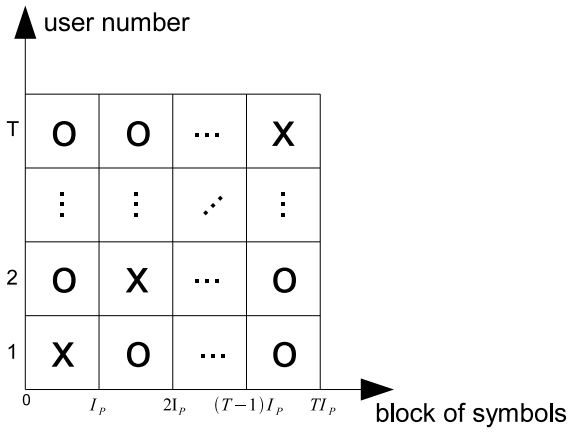

Figure 3.3: Transmission scheme for MIMO channel estimation.

which allows us to conclude that the Fourier transform of the output of the t^{th} PA $[\bar{\mathbf{U}}_n]_{t,i} = \bar{u}_{i,n,t}$ (see (3.18)) can be estimated from $\bar{\mathbf{X}}_n$ by means of a pseudo-inverse of Λ_n. This means that the PA outputs can be separated by exploiting the spatial diversity.

3.3.2 Channel Estimation

As well as in Section 3.2.2, we assume that pilot symbols are allocated on N_P subcarriers ($N_P \leq N$), the channel coefficients on all the subcarriers being estimated using the FFT interpolation method. Moreover, we assume that, when a given source is transmitting pilot symbols, the other ones do not transmit any information. This transmission scheme for MIMO channel estimation is shown in Fig. 3.3, with X representing "transmitting" and O "not-transmitting". Thus, the channel estimation is carried out as in a single-user system, the total number of transmission blocks being equal to TI_P.

Let us denote by $\{\bar{s}^{(p)}_{1,n,t}, \cdots, \bar{s}^{(p)}_{I_P,n,t}\}$ the I_P pilot symbols of the t^{th} user at n^{th} pilot subcarrier and by $\bar{\mathbf{X}}^{(p)}_{n,t} \in \mathbb{C}^{R\times I_P}$ the matrix containing the frequency domain signals received by all the R antennas during the training period of the t^{th} user at the n^{th} pilot subcarrier, i.e. for a given t ($1 \leq t \leq T$) and n ($1 \leq n \leq N_B$), we have $[\bar{\mathbf{X}}^{(p)}_{n,t}]_{r,i} = \bar{x}_{i,n,r}$, for $1 \leq r \leq R$ and $(t-1)I_P + 1 \leq i \leq tI_P$. Thus, from (3.70), the LS estimate of the channel matrix $\mathbf{H}_n$ is obtained as follows:

$$\hat{\mathbf{H}}_n = [\hat{\mathbf{H}}_{n,1} \;\; \hat{\mathbf{H}}_{n,2} \cdots \hat{\mathbf{H}}_{n,T}] \in \mathbb{C}^{R\times(K+1)T}, \tag{3.76}$$

with

$$\hat{\mathbf{H}}_{n,t} = \bar{\mathbf{X}}_{n,t}^{(p)} \left[\bar{\mathbf{S}}_{n,t}^{(p)}\right]^{\dagger} \in \mathbb{C}^{R\times(K+1)}, \tag{3.77}$$

for $1 \leq t \leq T$, where $(\cdot)^{\dagger}$ denotes the matrix pseudo-inverse and $\bar{\mathbf{S}}_{n,t}^{(p)} \in \mathbb{C}^{(K+1)\times I_P}$ is defined as:

$$\bar{\mathbf{S}}_{n,t}^{(p)} = \begin{bmatrix} \bar{s}_{1,n,t}^{(p)} & \cdots & \bar{s}_{I_P,n,t}^{(p)} \\ \bar{\psi}_3(s_{1,n,t}^{(p)}) & \cdots & \bar{\psi}_3(s_{I_P,n,t}^{(p)}) \\ \vdots & \ddots & \vdots \\ \bar{\psi}_{2K+1}(s_{1,n,t}^{(p)}) & \cdots & \bar{\psi}_{2K+1}(s_{I_P,n,t}^{(p)}) \end{bmatrix}. \tag{3.78}$$

Note that a necessary condition for (3.77) is $I_P \geq (K+1)$.

3.3.3 MIMO Power Diversity-based Receivers (MIMO-PDRs)

As pointed out earlier, it is not possible to recover the frequency domain information signals $\bar{s}_{i,n,t}$ from the matrix $\bar{\mathbf{X}}_n$ by exploiting the spatial diversity, as the channel matrix $\mathbf{H}_n$ can never be full-column rank. However, it can be viewed from (3.74) that if Λ_n is full column-rank, the spatial diversity can be used to separate the outputs of the PAs of the different sources, without canceling the nonlinear distortions. Thus, the receivers proposed in this section use an antenna array to cancel the MAI and the power diversity transmission scheme to cancel the nonlinear ICI. Indeed, as we will see in the sequel, the joint use of R receive antennas and a power diversity transmission scheme with a repetition factor equal to L provides RL sub-channels for the nonlinear OFDM system. The information symbols will then be recovered exploiting jointly the space and power transmission diversities.

Transmission Scheme

The proposed power diversity transmission scheme in a MIMO-OFDM channel is described in the following. For a given subcarrier n ($1 \leq n \leq N$), a transmission block i ($1 \leq i \leq I_B$) and user t ($1 \leq t \leq T$), we have:

$$\bar{s}_{((i-1)L+l),n,t}^{(pd)} = \sqrt{P_{l,t}}\, \bar{s}_{i,n,t}, \quad \text{for} \quad 1 \leq l \leq L, \tag{3.79}$$

where $\bar{s}^{(pd)}_{k,n,t}$ are the frequency domain transmitted signals of the t^{th} user, L is the repetition factor and $P_{1,t}, ..., P_{L,t}$ are the transmission powers of the t^{th} user.

Let us denote by $\bar{x}^{(pd)}_{((i-1)L+l),n,r}$ the frequency domain signal received by the r^{th} antenna at the n^{th} subcarrier and associated with the i^{th} transmission block and l^{th} symbol repetition. Moreover, let $\bar{\mathbf{X}}^{(pd)}_{i,n} \in \mathbb{C}^{R\times L}$ be the matrix containing the L frequency domain signals received by all the R antennas at the n^{th} subcarrier and associated with the i^{th} transmission block, with

$$[\bar{\mathbf{X}}^{(pd)}_{i,n}]_{r,l} = \bar{x}^{(pd)}_{((i-1)L+l),n,r}, \tag{3.80}$$

for $1 \leq l \leq L$ and $1 \leq r \leq R$. Thus, using (3.70) and the following relationship: $\text{vec}\,[\mathbf{ABC}] = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \text{vec}\,[\mathbf{B}]$, we may write:

$$\text{vec}\left[\bar{\mathbf{X}}^{(pd)}_{i,n}\right] \;=\; (\mathbf{I}_L \otimes \mathbf{H}_n)\,\text{vec}\left[\bar{\mathbf{S}}^{(pd)}_{i,n}\right] \in \mathbb{C}^{RL\times 1}, \tag{3.81}$$

where $\bar{\mathbf{S}}^{(pd)}_{i,n} \in \mathbb{C}^{(K+1)T\times L}$ defined as:

$$\bar{\mathbf{S}}^{(pd)}_{i,n} = \begin{bmatrix} \bar{s}^{(pd)}_{((i-1)L+1),n,1} & \cdots & \bar{s}^{(pd)}_{iL,n,1} \\ \vdots & \ddots & \vdots \\ |\bar{s}^{(pd)}_{((i-1)L+1),n,1}|^{2K}\,\bar{s}^{(pd)}_{((i-1)L+1),n,1} & \cdots & |\bar{s}^{(pd)}_{iL,n,1}|^{2K}\,\bar{s}^{(pd)}_{iL,n,1} \\ & \vdots & \\ \bar{s}^{(pd)}_{((i-1)L+1),n,T} & \cdots & \bar{s}^{(pd)}_{iL,n,T} \\ \vdots & \ddots & \vdots \\ |\bar{s}^{(pd)}_{((i-1)L+1),n,T}|^{2K}\,\bar{s}^{(pd)}_{((i-1)L+1),n,T} & \cdots & |\bar{s}^{(pd)}_{iL,n,T}|^{2K}\,\bar{s}^{(pd)}_{iL,n,T} \end{bmatrix}, \tag{3.82}$$

which can be rewritten as:

$$\bar{\mathbf{S}}^{(pd)}_{i,n} = \text{diag}\left[\bar{\mathbf{s}}^{(pd)}_{i,n}\right]\,\mathbf{P}^T, \tag{3.83}$$

where $\bar{\mathbf{s}}^{(pd)}_{i,n} = [\bar{s}_{i,n,1}, \ldots, |\bar{s}_{i,n,1}|^{2K}\bar{s}_{i,n,1},\ \cdots\ \bar{s}_{i,n,T}, \ldots, |\bar{s}_{i,n,T}|^{2K}\bar{s}_{i,n,T}]^T \in \mathbb{C}^{(K+1)T\times 1}$ and $\mathbf{P} = [\mathbf{P}_1 \cdots \mathbf{P}_T] \in \mathbb{C}^{L\times(K+1)T}$, with

$$\mathbf{P}_t = \begin{bmatrix} P_{1,t}^{\frac{1}{2}} & \cdots & P_{1,t}^{\frac{2K+1}{2}} \\ \vdots & \ddots & \vdots \\ P_{L,t}^{\frac{1}{2}} & \cdots & P_{L,t}^{\frac{2K+1}{2}} \end{bmatrix} \in \mathbb{C}^{L\times(K+1)}. \tag{3.84}$$

Substituting (3.83) into (3.81), we get:

$$\begin{aligned} \text{vec}\left[\bar{\mathbf{X}}_{i,n}^{(pd)}\right] &= (\mathbf{I}_L \otimes \mathbf{H}_n)\left(\mathbf{I}_L \otimes \text{diag}\left[\bar{\mathbf{S}}_{i,n}^{(pd)}\right]\right)\text{vec}\left[\mathbf{P}^T\right] \\ &= (\mathbf{P} \diamond \mathbf{H}_n)\,\bar{\mathbf{s}}_{i,n}^{(pd)} \in \mathbb{C}^{RL\times 1}, \end{aligned} \tag{3.85}$$

where $\diamond$ denotes the Khatri-Rao product.

In order to take all the transmission blocks into account, we can define:

$$\bar{\mathbf{X}}_n^{(pd)} = \left[\text{vec}\left[\bar{\mathbf{X}}_{1,n}^{(pd)}\right]\cdots\text{vec}\left[\bar{\mathbf{X}}_{I_B,n}^{(pd)}\right]\right] \in \mathbb{C}^{RL\times I_B}, \tag{3.86}$$

which can be expressed as:

$$\bar{\mathbf{X}}_n^{(pd)} = (\mathbf{P} \diamond \mathbf{H}_n)\,\bar{\mathbf{S}}_n, \tag{3.87}$$

where $\bar{\mathbf{S}}_n \in \mathbb{C}^{(K+1)T\times I_B}$ is defined in (3.71). Note that each row of the matrix $\bar{\mathbf{X}}_n^{(pd)}$ corresponds to one of the RL sub-channels provided by the proposed MIMO transmission scheme, the corresponding channel matrix given by $(\mathbf{P} \diamond \mathbf{H}_n) \in \mathbb{C}^{RL\times(K+1)T}$.

Furthermore, substituting (3.72) into (3.87) and using (A.10), we get:

$$\bar{\mathbf{X}}_n^{(pd)} = (\mathbf{I}_L \otimes \Lambda_n)\,(\mathbf{P} \diamond \mathbf{F})\,\bar{\mathbf{S}}_n. \tag{3.88}$$

Denoting by $\bar{u}_{((i-1)L+l),n,t}^{(pd)}$ the FFT of the t^{th} PA output at the n^{th} subcarrier and associated with the i^{th} transmission block and l^{th} symbol repetition, we define:

$$\bar{\mathbf{U}}_n^{(pd)} = \begin{bmatrix} \bar{u}_{1,n,1}^{(pd)} & \bar{u}_{L+1,n,1}^{(pd)} & \cdots & \bar{u}_{((I_B-1)L+1),n,1}^{(pd)} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{u}_{1,n,T}^{(pd)} & \bar{u}_{L+1,n,T}^{(pd)} & \cdots & \bar{u}_{((I_B-1)L+1),n,T}^{(pd)} \\ & & \vdots & \\ \bar{u}_{L,n,1}^{(pd)} & \bar{u}_{2L,n,1}^{(pd)} & \cdots & \bar{u}_{I_BL,n,1}^{(pd)} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{u}_{L,n,T}^{(pd)} & \bar{u}_{2L,n,T}^{(pd)} & \cdots & \bar{u}_{I_BL,n,T}^{(pd)} \end{bmatrix} \in \mathbb{C}^{TL\times I_B}. \tag{3.89}$$

Using (3.75), the matrix $\bar{\mathbf{U}}_n^{(pd)}$ can then be expressed as:

$$\begin{aligned}\bar{\mathbf{U}}_n^{(pd)} &= \begin{bmatrix} \mathbf{F}\mathrm{diag}\left[\mathbf{P}_{1,\cdot}\right] \\ \vdots \\ \mathbf{F}\mathrm{diag}\left[\mathbf{P}_{L,\cdot}\right] \end{bmatrix} \begin{bmatrix} \bar{s}_{1,n,1} & \cdots & \bar{s}_{I_B,n,1} \\ \vdots & \ddots & \vdots \\ |\bar{s}_{1,n,1}|^{2K}\bar{s}_{1,n,1} & \cdots & |\bar{s}_{I_B,n,1}|^{2K}\bar{s}_{I_B,n,1} \\ & \vdots & \\ \bar{s}_{1,n,T} & \cdots & \bar{s}_{I_B,n,T} \\ \vdots & \ddots & \vdots \\ |\bar{s}_{1,n,T}|^{2K}\bar{s}_{1,n,T} & \cdots & |\bar{s}_{I_B,n,T}|^{2K}\bar{s}_{I_B,n,T} \end{bmatrix} \\ &= (\mathbf{P} \diamond \mathbf{F})\,\bar{\mathbf{S}}_n. \end{aligned} \tag{3.90}$$

Thus, substituting (3.90) into (3.88), we get:

$$\bar{\mathbf{X}}_n^{(pd)} = (\mathbf{I}_L \otimes \Lambda_n)\,\bar{\mathbf{U}}_n^{(pd)}. \tag{3.91}$$

It can be noted from (3.89) that the informations relative to the t^{th} source are in rows t, $(T+t)$,...,$((L-1)T+t)$ of $\bar{\mathbf{U}}_n^{(pd)}$, for $1 \leq t \leq T$. Thus, by defining

$$\bar{\mathbf{U}}_{n,t}^{(pd)} = \left[\left[\bar{\mathbf{U}}_n^{(pd)}\right]_{t,\cdot}^T \quad \left[\bar{\mathbf{U}}_n^{(pd)}\right]_{(T+t),\cdot}^T \cdots \left[\bar{\mathbf{U}}_n^{(pd)}\right]_{((L-1)T+t),\cdot}^T \right] \in \mathbb{C}^{L \times I_B}. \tag{3.92}$$

we get from (3.90):

$$\bar{\mathbf{U}}_{n,t}^{(pd)} = \mathbf{P}_t \mathrm{diag}\left[\mathbf{f}_t\right] \bar{\mathbf{S}}_{n,t}, \tag{3.93}$$

where $\mathbf{P}_t$ is defined in (3.84) and

$$\bar{\mathbf{S}}_{n,t} = \begin{bmatrix} \bar{s}_{1,n,t} & \cdots & \bar{s}_{I_B,n,t} \\ \vdots & \ddots & \vdots \\ |\bar{s}_{1,n,t}|^{2K}\bar{s}_{1,n,t} & \cdots & |\bar{s}_{I_B,n,t}|^{2K}\bar{s}_{I_B,n,t} \end{bmatrix}. \tag{3.94}$$

Thus, it is possible to develop a method for recovering the information symbols $\bar{s}_{i,n,t}$ based on (3.91) and (3.93), which is done in the sequel.

Table 3.3: MMSE MIMO-Power Diversity-based Receiver (MMSE MIMO-PDR)

Transmission scheme

For $1 \leq i \leq I_B$, $1 \leq l \leq L$ and $1 \leq t \leq T$:

$\bar{\mathbf{s}}^{(pd)}((i-1)L+l,t) = \bar{\mathbf{s}}(i,t)\sqrt{P_{l,t}} \quad \in \mathbb{C}^{N\times 1}$

Equalization:

For $1 \leq n \leq N$:

1) Construct: $\bar{\mathbf{X}}_n^{(pd)} = \left[\text{vec}\left[\bar{\mathbf{X}}_{1,n}^{(pd)}\right] \cdots \text{vec}\left[\bar{\mathbf{X}}_{I_B,n}^{(pd)}\right]\right] \in \mathbb{C}^{RL\times I_B}$, where $[\bar{\mathbf{X}}_{i,n}^{(pd)}]_{r,l} = \bar{x}_{((i-1)L+l),n,r}^{(pd)} \in \mathbb{C}^{R\times L}$, for $1 \leq l \leq L$ and $1 \leq r \leq R$.

2) Calculate: $\hat{\bar{\mathbf{S}}}_n = \mathbf{R}_{\bar{\mathbf{S}}}\ (\mathbf{P} \diamond \mathbf{H}_n)^H \left[(\mathbf{P} \diamond \mathbf{H}_n)\,\mathbf{R}_{\bar{\mathbf{S}}}\ (\mathbf{P} \diamond \mathbf{H}_n)^H + \mathbf{I}_{RL}\sigma^2\right]^{-1} \bar{\mathbf{X}}_n^{(pd)} \in \mathbb{C}^{(K+1)T\times I_B}$

3) The estimate of the I_B frequency domain symbols associated with the n^{th} subcarrier and t^{th} source is the $[(t-1)(2K+1)+1]^{th}$ row of $\hat{\bar{\mathbf{S}}}_n$.

Equalization

Using the multi-channel representation (3.87), the MMSE estimate of the matrix $\bar{\mathbf{S}}_n$, defined in (3.71), is given by:

$$\hat{\bar{\mathbf{S}}}_n = \mathbf{R}_{\bar{\mathbf{S}}}\ (\mathbf{P} \diamond \mathbf{H}_n)^H \left[(\mathbf{P} \diamond \mathbf{H}_n)\,\mathbf{R}_{\bar{\mathbf{S}}}\ (\mathbf{P} \diamond \mathbf{H}_n)^H + \mathbf{I}_{RL}\sigma^2\right]^{-1} \bar{\mathbf{X}}_n^{(pd)} \quad \in \mathbb{C}^{(K+1)T\times I_B}, \tag{3.95}$$

where $\mathbf{R}_{\bar{\mathbf{S}}} = \mathbb{E}[\bar{\mathbf{S}}_n\bar{\mathbf{S}}_n^H] \in \mathbb{C}^{(K+1)T\times(K+1)T}$ is covariance matrix of $\bar{\mathbf{S}}_n$. The expressions of the MMSE MIMO-PDR are summarized in Table 3.3.

The ZF PDR for the considered MIMO-ODFM systems is based on two steps: the first one consists in separating the signals of the sources using the spacial diversity and the second one in separating information signals from the nonlinear distortions using the power diversity. Using the multi-channel representation (3.91), the ZF

estimate of the matrix $\bar{\mathbf{U}}_n^{(pd)}$ is given by:

$$\hat{\bar{\mathbf{U}}}_n^{(pd)} = (\mathbf{I}_L \otimes \Lambda_n)^\dagger \bar{\mathbf{X}}_n^{(pd)} \in \mathbb{C}^{TL \times I_B}. \tag{3.96}$$

Thus, using (3.92) and (3.93), as in Section 3.2.3, the first row of $\bar{\mathbf{S}}_{n,t}$, containing the frequency domain symbols $\bar{s}_{i,n,t}$ of the t^{th} user, can be estimated as:

$$[\hat{\bar{\mathbf{S}}}_{n,t}]_{1,\cdot} = \frac{\tilde{\mathbf{p}}_t}{f_{1,t}} \hat{\bar{\mathbf{U}}}_{n,t}^{(pd)} \quad \in \mathbb{C}^{1 \times I_B}, \tag{3.97}$$

where $\tilde{\mathbf{p}}_t \in \mathbb{C}^{1 \times L}$ is the first row of $\mathbf{P}_t^\dagger$.

Taking all the subcarriers into account, we get:

$$\hat{\bar{\mathbf{U}}}^{(pd)} = \left[(\mathbf{I}_L \otimes \Lambda_1)^\dagger \bar{\mathbf{X}}_1^{(pd)} \quad \cdots \quad (\mathbf{I}_L \otimes \Lambda_n)^\dagger \bar{\mathbf{X}}_n^{(pd)} \right], \tag{3.98}$$

where

$$\hat{\bar{\mathbf{U}}}^{(pd)} = \left[\hat{\bar{\mathbf{U}}}_1^{(pd)} \cdots \hat{\bar{\mathbf{U}}}_N^{(pd)} \right] \in \mathbb{C}^{TL \times I_B N}. \tag{3.99}$$

Thus, defining:

$$\hat{\bar{\mathbf{U}}}_t^{(pd)} = \left[\left[\hat{\bar{\mathbf{U}}}^{(pd)} \right]_{t,\cdot}^T \quad \left[\hat{\bar{\mathbf{U}}}^{(pd)} \right]_{(T+t),\cdot}^T \cdots \left[\hat{\bar{\mathbf{U}}}^{(pd)} \right]_{((L-1)T+t),\cdot}^T \right] \in \mathbb{C}^{L \times N I_B}, \tag{3.100}$$

the t^{th} user's information signal on the N subcarriers and I_B blocks can be estimated as:

$$\hat{\bar{\mathbf{s}}}_t = \frac{\tilde{\mathbf{p}}_t}{f_{1,t}} \hat{\bar{\mathbf{U}}}_t^{(pd)} \quad \in \mathbb{C}^{1 \times N I_B}. \tag{3.101}$$

The expressions of the ZF MIMO-PDR are summarized in Table 3.4, where the matrix $\mathbf{H}_n^{(lin)} \in \mathbb{C}^{R \times T}$ is defined as

$$\mathbf{H}_n^{(lin)} = \Lambda_n \text{diag}\left(\mathbf{f}^{(lin)} \right), \tag{3.102}$$

with $\mathbf{f}^{(lin)} = [f_{1,1} \cdots f_{1,T}]^T \in \mathbb{C}^{T \times 1}$ being the vector containing the linear PA coefficients of all the sources. Recalling that

$$[\mathbf{H}_n]_{r,((t-1)(K+1)+k+1)} = h_{n,k+1,t,r}, \tag{3.103}$$

Table 3.4: ZF MIMO-Power Diversity-based Receiver (ZF MIMO-PDR)

Transmission scheme

For $1 \leq i \leq I_B$, $1 \leq l \leq L$ and $1 \leq t \leq T$:

$\bar{\mathbf{s}}^{(pd)}((i-1)L+l,t) = \bar{\mathbf{s}}(i,t)\sqrt{P_{l,t}} \ \in \mathbb{C}^{N\times 1}$

Equalization:

1) For $1 \leq n \leq N$, construct: $\bar{\mathbf{X}}_n^{(pd)} = \left[\text{vec}\left[\bar{\mathbf{X}}_{1,n}^{(pd)}\right] \cdots \text{vec}\left[\bar{\mathbf{X}}_{I_B,n}^{(pd)}\right]\right] \in \mathbb{C}^{RL\times I_B}$, where $[\bar{\mathbf{X}}_{i,n}^{(pd)}]_{r,l} = \bar{x}_{((i-1)L+l),n,r}^{(pd)} \in \mathbb{C}^{R\times L}$, for $1 \leq l \leq L$ and $1 \leq r \leq R$.

2) Calculate: $\hat{\bar{\mathbf{U}}}^{(pd)} = \left[\left(\mathbf{I}_L \otimes \mathbf{H}_1^{(lin)}\right)^{\dagger} \bar{\mathbf{X}}_1^{(pd)} \ \cdots \ \left(\mathbf{I}_L \otimes \mathbf{H}_n^{(lin)}\right)^{\dagger} \bar{\mathbf{X}}_n^{(pd)}\right] \in \mathbb{C}^{TL\times I_B N}$, where $\mathbf{H}_n^{(lin)}$ is given by (3.104).

For $1 \leq t \leq T$:

3) Construct $\mathbf{P}_t$ from (3.84) and calculate $\tilde{\mathbf{p}}_t$ as the first row of $\mathbf{P}_t^{\dagger}$

4) Construct $\hat{\bar{\mathbf{U}}}_t^{(pd)} = \left[\left[\hat{\bar{\mathbf{U}}}^{(pd)}\right]_{t,\cdot}^{T} \quad \left[\hat{\bar{\mathbf{U}}}^{(pd)}\right]_{(T+t),\cdot}^{T} \cdots \left[\hat{\bar{\mathbf{U}}}^{(pd)}\right]_{((L-1)T+t),\cdot}^{T}\right] \in \mathbb{C}^{L\times I_B N}$

5) The information signal of t^{th} source on the N subcarriers and I_B blocks is estimated as: $\hat{\bar{\mathbf{s}}}_t = \tilde{\mathbf{p}}_t \hat{\bar{\mathbf{U}}}_t^{(pd)} \in \mathbb{C}^{1\times I_B N}$

the matrix $\mathbf{H}_n^{(lin)}$ can be estimated directly from the global channel matrix $\mathbf{H}_n$ as:

$$\mathbf{H}_n^{(lin)} = \left[[\mathbf{H}_n]_{\cdot,1}^{T} \quad [\mathbf{H}_n]_{\cdot,(K+1)+1}^{T} \cdots [\mathbf{H}_n]_{\cdot,(T-1)(K+1)+1}^{T}\right]. \tag{3.104}$$

A necessary condition for the ZF and MMSE MIMO-PDRs is that Λ_n and $\mathbf{P}_t$ $(1 \leq t \leq T)$ be full column-rank, which implies respectively $R \geq T$ and $L \geq K+1$. It should be highlighted that the matrices $\mathbf{P}_t$ $(1 \leq t \leq T)$ can be the same for all the users. Moreover, as well as in the SISO case, the proposed ZF equalization method uses only the channel coefficients associated with the linear terms.

3.4 MIMO-OFDM system with memory polynomial PAs

In this section, we deal with a MIMO-OFDM channel assuming that each PA is represented by a memory polynomial model (2.81). The main result of this section is that the channel estimation and equalization techniques proposed respectively in Sections 3.3.2 and 3.3.3 can be directly applied for the case of memory polynomial PAs. This is due a theorem stating that a memory polynomial PA model can be expressed as a memoryless polynomial model with coefficients that depend on the subcarrier.

In what follows, we assume that the length of the cyclic prefix satisfy: $M_{cp} \geq M + M_{pa}$, where M_{pa} is the memory of the PAs, assumed to be the same for all the sources. Thus, denoting by $u_{i,n,t}$ $(1 \leq n \leq N + M_{cp})$ the output of the t^{th} memory polynomial PA at the i^{th} transmission block, we may write for $1 \leq n \leq N$:

$$u_{i,(n+M_{cp}),t} = \sum_{m=0}^{M_{pa}} \sum_{k=0}^{K} f_{2k+1,t}(m)\ \psi_{2k+1}(s^{(cir)}_{i,(n-m),t}), \tag{3.105}$$

where $f_{2k+1,t}(m)$ are the coefficients of the t^{th} PA and

$$s^{(cir)}_{i,n,t} = \begin{cases} s_{i,n,t}, & \text{for} \quad 1 \leq n \leq N, \\ s_{i,N+n,t}, & \text{for} \quad 1 - N \leq n \leq 0. \end{cases} \tag{3.106}$$

Note that, due to the fact that $M_{cp} \geq M + M_{pa}$, the signal $u_{i,n,t}$ contains a cyclic block of $(M_{cp} - M_{pa})$ symbols: $\{u_{i,(M_{pa}+1,t)}, \cdots, u_{i,M_{cp},t}\}$ is equal to $\{u_{i,(N-M_{cp}+M_{pa}+1),t}, \cdots, u_{i,N,t}\}$. As a consequence, equation (3.59) is still valid in this case.

From (3.59), it can be concluded that frequency domain received signals are linear mixtures of the T frequency domain PA outputs $\bar{u}_{i,n,t}$, with $\bar{u}_{i,n,t} = [\bar{\mathbf{u}}(i,t)]_n$. In what follows, Theorem 3.1 demonstrates that, when then PA is represented by a memory polynomial model, the signal $\bar{u}_{i,n,t}$ can be written as the frequency domain of the output of a memoryless polynomial PA, the coefficients of which varying from one subcarrier to another. Before that, Lemma 3.1 demonstrates an important result used in the proof of Theorem 3.1.

Lemma 3.1: Let $\mathbf{V} \in \mathbb{C}^{N \times N}$ be the FFT matrix of order N and $\mathbf{U} \in \mathbb{C}^{N \times N}$ be a

circulant lower shift matrix of order N defined as:

$$\mathbf{U} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{3.107}$$

Then, we have:

$$\mathbf{V}\mathbf{U}^i = \sqrt{N}\ \mathrm{diag}_{(i+1)}[\,\mathbf{V}\,]\mathbf{V}, \tag{3.108}$$

for $0 \leq i \leq N-1$, where $\mathrm{diag}_i[\mathbf{A}]$ is the diagonal matrix formed from the i^{th} row of $\mathbf{A}$.

Proof:

Post-multiplying a matrix by $\mathbf{U}^i$ is equivalent to circularly shifting its columns to the left i times, that is, by defining $\mathbf{V}(i) = \mathbf{V}\mathbf{U}^i$, the n^{th} column of $\mathbf{V}(i)$ is given by:

$$\mathbf{V}(i)_{\cdot,n} = \mathbf{V}_{\cdot,(\mathrm{mod}(n+i-1,N)+1)\cdot} \tag{3.109}$$

As the k^{th} column of $\mathbf{V}$, defined in (3.2), is given by $\mathbf{V}_{\cdot,k} = \frac{1}{\sqrt{N}}[1\ \omega^{(k-1)}\ \omega^{2(k-1)}\ \cdots\ \omega^{(N-1)(k-1)}]^T$, with $\omega = e^{-\mathbf{j}2\pi/N}$, we have:

$$\mathbf{V}(i) = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \omega^i & \omega^{(i+1)} & \cdots & \omega^{(i+N-1)} \\ \omega^{2i} & \omega^{2(i+1)} & \cdots & \omega^{2(i+N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{(N-1)i} & \omega^{(N-1)(i+1)} & \cdots & \omega^{(N-1)(i+N-1)} \end{pmatrix}. \tag{3.110}$$

Equation (3.110) can be reexpressed as:

$$\mathbf{V}(i) = \mathrm{diag} \begin{bmatrix} 1 \\ \omega^i \\ \omega^{2i} \\ \vdots \\ \omega^{(N-1)i} \end{bmatrix} \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{(n-1)} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{(n-1)} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)(N-1)} \end{pmatrix}, \tag{3.111}$$

which is equivalent to (3.108) .

■

Theorem 3.1: Let $\bar{u}_{i,n,t}$ $(1 \leq n \leq N)$ be the frequency domain output of a memory polynomial PA with coefficients denoted by $f_{2k+1,t}(m)$, for $0 \leq k \leq K$ and $0 \leq m \leq M_{pa}$. Then, the signal $\bar{u}_{i,n,t}$ can expressed as the frequency domain output of a memoryless polynomial PA:

$$\bar{u}_{i,n,t} = \sum_{k=0}^{K} f'_{2k+1,t,n} \; \bar{\psi}_{2k+1}(s_{i,n,t}), \tag{3.112}$$

with subcarrier dependent coefficients given by:

$$f'_{2k+1,t,n} = \sum_{m=0}^{M_{pa}} f_{2k+1,t}(m) \; e^{-\jmath 2\pi m(n-1)/N}. \tag{3.113}$$

Proof:

Let us express (3.105) in a vector form as:

$$\mathbf{u}(i,t) = \sum_{m=0}^{M_{pa}} \sum_{k=0}^{K} f_{2k+1,t}(m) \; \Psi_{2k+1}(\mathbf{s}_m^{(cir)}(i,t)) \in \mathbb{C}^{N\times 1} \tag{3.114}$$

where $[\mathbf{u}(i,t)]_n = u_{i,(n+M_{cp}),t}$ for $1 \leq n \leq N$, and

$$\mathbf{s}_m^{(cir)}(i,t) = [s_{i,(1-m),t}^{(cir)} \; \cdots \; s_{i,(N-m),t}^{(cir)}]^T \in \mathbb{C}^{N\times 1}. \tag{3.115}$$

Note that, due to its circularity property, the vector $\mathbf{s}_m^{(cir)}(i,t)$ can be expressed as:

$$\begin{aligned} \mathbf{s}_m^{(cir)}(i,t) &= \mathbf{U}\mathbf{s}_{m-1}^{(cir)}(i,t) \\ &= \mathbf{U}^m \mathbf{s}_0^{(cir)}(i,t) = \mathbf{U}^m \mathbf{s}(i,t), \end{aligned} \tag{3.116}$$

with $\mathbf{U}$ defined in (3.107) and $[\mathbf{s}(i,t)]_n = s_{i,n,t}$.

Thus, substituting (3.116) into (3.114), we get:

$$\begin{aligned} \mathbf{u}(i,t) &= \sum_{m=0}^{M_{pa}} \sum_{k=0}^{K} f_{2k+1,t}(m) \; \Psi_{2k+1}(\mathbf{U}^m \mathbf{s}(i,t)) \\ &= \sum_{m=0}^{M_{pa}} \sum_{k=0}^{K} f_{2k+1,t}(m) \; \mathbf{U}^m \Psi_{2k+1}(\mathbf{s}(i,t)) \end{aligned} \tag{3.117}$$

Calculating the FFT of both sides of (3.117), we get:

$$\bar{\mathbf{u}}(i,t) = \sum_{m=0}^{M_{pa}} \sum_{k=0}^{K} f_{2k+1,t}(m) \mathbf{V}\mathbf{U}^m \Psi_{2k+1}(\mathbf{s}(i,t)). \tag{3.118}$$

Using Lemma 3.1, we have:

$$\begin{aligned} \bar{\mathbf{u}}(i,t) &= \sum_{m=0}^{M_{pa}} \sum_{k=0}^{K} f_{2k+1,t}(m)\ \sqrt{N}\ \mathrm{diag}_{(m+1)}[\mathbf{V}]\mathbf{V}\Psi_{2k+1}(\mathbf{s}(i,t)), \\ &= \sum_{m=0}^{M_{pa}} \sum_{k=0}^{K} f_{2k+1,t}(m)\ \sqrt{N}\ \mathrm{diag}_{(m+1)}[\mathbf{V}]\bar{\Psi}_{2k+1}(\mathbf{s}(i,t)). \end{aligned} \tag{3.119}$$

Thus, by defining:

$$f'_{2k+1,t,n} = \sum_{m=0}^{M_{pa}} f_{2k+1,t}(m)\ e^{-\jmath\, 2\pi\, m\, (n-1)/N} \tag{3.120}$$

or, using the vector from $\mathbf{f}'_{2k+1,t} = [f'_{2k+1,t,1} \cdots f'_{2k+1,t,N}]^T \in \mathbb{C}^{N\times 1}$:

$$\mathrm{diag}[\mathbf{f}'_{2k+1,t}] = \sqrt{N}\ \sum_{m=0}^{M_{pa}} f_{2k+1,t}(m)\ \mathrm{diag}_{(m+1)}[\mathbf{V}], \tag{3.121}$$

we obtain from (3.119):

$$\bar{\mathbf{u}}(i,t) = \sum_{k=0}^{K} \mathrm{diag}[\mathbf{f}'_{2k+1,t}]\bar{\Psi}_{2k+1}(\mathbf{s}(i,t)), \tag{3.122}$$

which gives the desired result. ■

Theorem 3.1 is of great importance as, by comparing the expressions for the frequency domain output of the PA (3.63) and (3.122), it can be concluded that, with respect to the signal $\bar{u}_{i,n,t}$, the memory polynomial model is equivalent to a subcarrier dependent memoryless polynomial model, the relationship between the parameters of these two models being given by (3.113).

Thus, using (3.59), the i^{th} vector of frequency domain signals received by the r^{th}

antenna is given by:

$$\bar{\mathbf{x}}(i,r) = \sum_{t=1}^{T} \Lambda(r,t) \sum_{k=0}^{K} \mathrm{diag}[\mathbf{f}'_{2k+1,t}] \bar{\Psi}_{2k+1}(\mathbf{s}(i,t)) \in \mathbb{C}^{N\times 1}, \tag{3.123}$$

or, equivalently,

$$\begin{aligned} \bar{x}_{i,n,r} &= \sum_{t=1}^{T} \sum_{k=0}^{K} \lambda_{n,r,t} f'_{2k+1,t,n}\ \bar{\psi}_{2k+1}(s_{i,n,t}) \\ &= \sum_{t=1}^{T} \sum_{k=0}^{K} h'_{n,k+1,r,t}\ \bar{\psi}_{2k+1}(s_{i,n,t}), \end{aligned} \tag{3.124}$$

with the global channel coefficients defined as:

$$h'_{n,k+1,r,t} = \lambda_{n,r,t} f'_{2k+1,t,n}. \tag{3.125}$$

Comparing (3.124) and (3.68), it can be viewed that the frequency domain received signals $\bar{x}_{i,n,r}$ have the same expression for memoryless and memory polynomial PAs, the global channel coefficients given by (3.67) and (3.125), respectively. The only difference between these two cases is the fact that the PA coefficients $f_{2k+1,t}$ in (3.67) are the same for all the subcarriers, while the PA coefficients $f'_{2k+1,t,n}$ in (3.125) vary from one subcarrier to another. Thus, similarly as in (3.70), the matrix $\bar{\mathbf{X}}_n \in \mathbb{C}^{R\times I_B}$, containing the I_B frequency domain signals received by all the R antennas at the n^{th} subcarrier, can be expresses as:

$$\bar{\mathbf{X}}_n = \mathbf{H}'_n\, \bar{\mathbf{S}}_n, \tag{3.126}$$

where $\bar{\mathbf{S}}_n \in \mathbb{C}^{(K+1)T\times I_B}$ is given in (3.71) and $\mathbf{H}'_n \in \mathbb{C}^{R\times(K+1)T}$ contains the global channel coefficients of the n^{th} subcarrier, with $[\mathbf{H}'_n]_{r,((t-1)(K+1)+k+1)} = h'_{n,k+1,r,t}$, for $0 \le k \le K$, $1 \le r \le R$ and $1 \le t \le T$.

In this case, the matrix $\mathbf{H}'_n$ can be expressed as $\mathbf{H}'_n = \Lambda_n\, \mathbf{F}_n$, with:

$$\mathbf{F}_n = \begin{bmatrix} \mathbf{f}_{1,n}^T & \cdots & \mathbf{0}_{K+1,1}^T \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{K+1,1}^T & \cdots & \mathbf{f}_{T,n}^T \end{bmatrix} \in \mathbb{C}^{T\times(K+1)T}, \tag{3.127}$$

where $\mathbf{f}_{t,n} = [f_{1,t,n}\ f_{3,t,n} \cdots f_{2K+1,t,n}]^T \in \mathbb{C}^{(K+1)\times 1}$. Similarly as in Section 3.3.1, the matrix $\bar{\mathbf{S}}_n$ can not be estimated by using the pseudo-inverse of $\mathbf{H}'_n$ as the matrix $\mathbf{H}'_n$ can not be full column rank. However, the matrix $\bar{\mathbf{X}}_n$ can still be expressed

as $\bar{\mathbf{X}}_n = \Lambda_n \bar{\mathbf{U}}_n$, with $\bar{\mathbf{U}}_n = \mathbf{F}_n \bar{\mathbf{S}}_n$.

In summary, owing to Theorem 3.1, it was possible to demonstrate that, when the PAs are represented by memory polynomial models, the expressions for the frequency domain received signals in terms of the global channel parameters are the same than the ones developed in Section 3.3 for memoryless polynomial PAs. Indeed, in both cases the global channel parameters have four indices: subcarrier, nonlinearity order, antenna number and source number. The main result of this section is then the fact that, as the channel estimation and equalization techniques proposed in Sections 3.3.2 and 3.3.3 are based on the use of the global channel parameters, these techniques can be directly applied in the case of memory polynomial PAs. In other words, we have shown that a more general case (memory polynomial PA) can be treated as particular case (memoryless polynomial PA). This means that the matrix $\mathbf{H}_n^{'}$ can be estimated as in (3.77)-(3.76), i.e.:

$$\hat{\mathbf{H}}_n^{'} = \bar{\mathbf{X}}_n^{(p)} \left[\bar{\mathbf{S}}_n^{(p)}\right]^{\dagger}, \tag{3.128}$$

and the information signals of the t^{th} source on the N subcarriers and I_B blocks can be estimated by the MIMO-PDRs, summarized in Tables 3.3 and 3.4, using respectively the channel matrix $\hat{\mathbf{H}}_n^{'}$ and the linear part of the channel matrix $\hat{\mathbf{H}}_n^{'}$:

$$\mathbf{H}_n^{'(lin)} = \left[\left[\mathbf{H}_n^{'}\right]_{\cdot,1}^{T} \quad \left[\mathbf{H}_n^{'}\right]_{\cdot,(K+1)+1}^{T} \cdots \left[\mathbf{H}_n^{'}\right]_{\cdot,(T-1)(K+1)+1}^{T}\right]. \tag{3.129}$$

3.5 Simulation Results

In this section, the proposed channel estimation and equalization techniques are evaluated by means of simulations. MIMO-OFDM systems with memoryless and memory third-order polynomial PAs have been considered for the simulations, with the PA coefficients shown in Table 3.5 for $t = 1, \ldots, T$. The $R \times T$ wireless link, corresponding to R receive and T transmit antennas, has a frequency selective fading due to multipath propagation, with the wireless link memory equal to 4 ($M = 4$). The length of the cyclic prefix L_{cp} is equal to 4 in the memoryless case and to 7 in the memory case. The results were obtained with $N = 64$ subcarriers and 64-QAM transmitted signals, via Monte Carlo simulations using at least $N_R = 100$ independent data realizations. In all the simulations, the PDRs use a repetition factor $L = 2$, with $P_{1,t} = 1$ and $P_{2,t} = 0.8$, for $t = 1, ..., T$.

Table 3.5: Memoryless and memory polynomial PA coefficients

	PA coefficients
memoryless polynomial [15]	$f_{1,t} = 0.9798 - 0.2887\jmath,\ f_{3,t} = -0.2901 + 0.4350\jmath$
memory polynomial [94]	$f_{1,t}(0) = 1.9702 + 0.1931\jmath,\ f_{3,t}(0) = -0.5934 - 0.1174\jmath$
	$f_{1,t}(1) = -0.9606 + 0.0036\jmath,\ f_{3,t}(1) = 0.2300 + 0.0560\jmath$
	$f_{1,t}(2) = 0,\ f_{3,t}(2) = 0$
	$f_{1,t}(3) = 0.1591 - 0.0132\jmath,\ f_{3,t}(3) = -0.0112 - 0.0094\jmath$

3.5.1 Memoryless polynomial PA - channel estimation

The next three figures contain simulation results corresponding to channel estimation in the case of memoryless polynomial PAs. The proposed channel estimation method is evaluated by means of the Normalized Mean Squared Error (NMSE) of the estimated global channel parameters, defined as:

$$NMSE_H = \frac{1}{N_R} \sum_{l=1}^{N_R} \frac{\| \mathbf{H} - \hat{\mathbf{H}}_l \|_F^2}{\| \mathbf{H} \|_F^2}, \tag{3.130}$$

where $\hat{\mathbf{H}}_l$ represents the channel matrix estimated at the l^{th} Monte Carlo simulation and $\| \cdot \|_F$ the *Frobenius norm*.

Fig. 3.4 shows the NMSE versus signal-to-noise-ratio (SNR) provided by proposed channel estimation technique for various values of N_P (number of pilot subcarriers), with $R = T = 1$ and $I_P = 2$ pilot symbols per subcarrier. It can be viewed that the channel estimates obtained with the tested values of N_P are very close, except for $N_P = 8$ and high SNRs. This figure shows that, in this case, it is interesting to use $N_P = 8$ or $N_P = 16$ pilot subcarriers.

To have a performance comparison for our technique, we have simulated the channel estimation method proposed in [67]. This technique assumes that the PA coefficients of all the users are known at the receiver. In practice, the PA parameters have to be estimated at the transmitter and this information has to be sent to the receiver. In the case of a memory PA, all coefficients of the memory polynomial model are sent to the receiver. The transmission of these parameters must be included in the system initialization process before the channel estima-

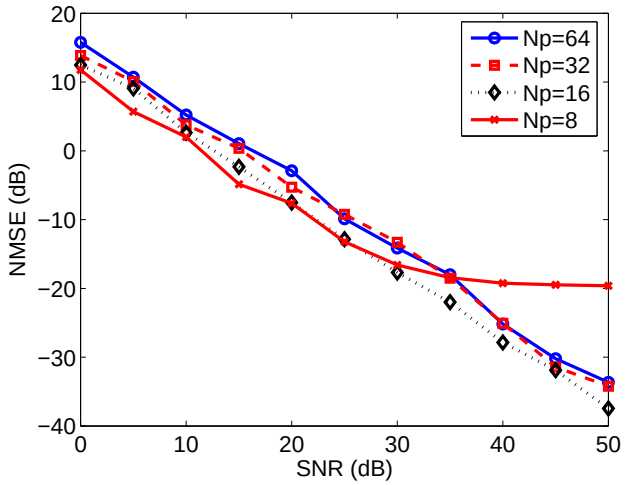

Figure 3.4: NMSE versus SNR for various values of N_P - $R = T = 1$ with memoryless PA

tion. Thus, this method performs the estimation of the wireless channel coefficients only. In this case, the estimation method is evaluated by means of the NMSE of the estimated wireless channel coefficients, defined as:

$$NMSE_L = \frac{1}{N_R} \sum_{l=1}^{N_R} \frac{\sum_{n=1}^{N} \parallel \Lambda_n - \hat{\Lambda}_n^{(l)} \parallel_F^2}{\sum_{n=1}^{N} \parallel \Lambda_n \parallel_F^2}, \tag{3.131}$$

where $\hat{\Lambda}_n^{(l)} \in \mathbb{C}^{R \times T}$ represents the estimate of Λ_n at the l^{th} Monte Carlo simulation.

However, note that the $NMSE_L$ can not be computed for the proposed estimation technique, as it does not provide an estimate of the wireless channel coefficients. Thus, as $\mathbf{H}_n^{(lin)} = \Lambda_n f_1$, where $f_1 = f_{1,t}$ for $t = 1, ..., T$, a figure of merit similar to (3.131) can be defined as:

$$NMSE_{Hlin} = \frac{1}{N_R} \sum_{l=1}^{N_R} \frac{\sum_{n=1}^{N} \parallel \mathbf{H}_n^{(lin)} - \hat{\mathbf{H}}_n^{(lin)}(l) \parallel_F^2}{\sum_{n=1}^{N} \parallel \mathbf{H}_n^{(lin)} \parallel_F^2}, \tag{3.132}$$

where $\hat{\mathbf{H}}_n^{(lin)}(l) \in \mathbb{C}^{R \times T}$ represents the estimate of $\mathbf{H}_n^{(lin)}$ at the l^{th} Monte Carlo simulation.

Fig. 3.5 shows the $NMSE_H$ and $NMSE_{Hlin}$ provided by proposed channel estimation technique, for $R = T = 1$, $N_P = 16$ and $I_P = 2$, and the $NMSE_L$

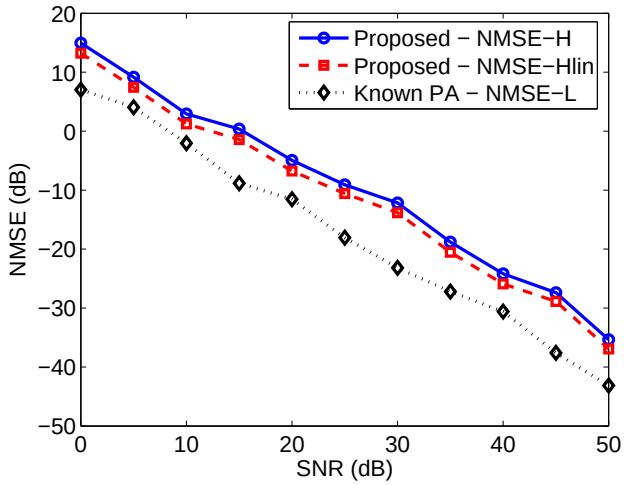

Figure 3.5: NMSE versus SNR for proposed and Known PA techniques - $R = T = 1$ with memoryless PA

provided by the method proposed in [67], henceforth referred to as Known PA channel estimation technique. In this case, as this technique uses only one pilot symbol per user per subcarrier, we have used $N_P = 32$ and $I_P = 1$, so that both techniques use 32 pilot symbols. Note that the Known PA technique provides a SNR gain of approximatively 6dB with respect to the proposed technique. The advantage of the proposed technique is that it does not require the knowledge of the PA coefficients. Indeed, the transmission of the PA coefficients leads to additional information to be transmitted, implying in a loss of transmission rate. Furthermore, as the transmission of these parameters is done before the channel estimation, a linear channel with good quality must be used to avoid errors on the values of the PA coefficients.

Similar results were obtained for the MIMO case. Fig. 3.6 shows the $NMSE_H$ and $NMSE_{Hlin}$ provided by proposed channel estimation technique, for $R = T = 2$, $N_P = 8$ and $I_P = 2$, and the $NMSE_L$ provided by the Known PA estimation technique. In this case, we have used $N_P = 16$ and $I_P = 1$, so that both technique use 16 pilot symbols per user. The conclusions that we can draw from Fig. 3.6 are similar to those of Fig. 3.5.

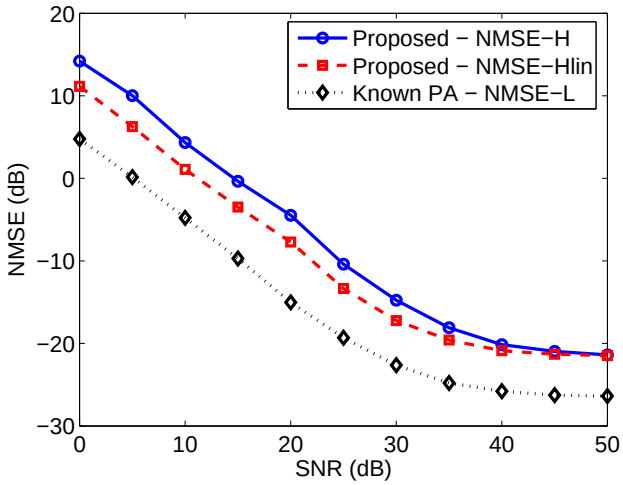

Figure 3.6: NMSE versus SNR for proposed and Known PA techniques - $R = T = 2$ with memoryless PA

3.5.2 Memoryless polynomial PA - channel equalization

The next figures contain simulation results corresponding to channel equalization in the case of memoryless polynomial PAs. The channel equalization techniques are evaluated by means of the Bit-Error-Rate (BER).

Fig. 3.7 shows the BER versus SNR provided by the proposed ZF and MMSE PDRs using channel estimates provided by the proposed channel estimation method, for $R = T = 1$, $N_P = 16$ and $I_P = 2$. In the case of a linear PA, the equalization of an OFDM channel can be carried out by a 1-tap equalizer that de-rotate the received signals. Thus, for comparison purposes, Fig. 3.7 also shows the BER provided by a 1-tap equalizer that simply divides the received signal $\bar{x}_{i,n}$ by the channel coefficient $g_{n,1}$. Note that the proposed receivers provide BERs much lower than the 1-tap equalizer. Moreover, as expected, the ZF and MMSE receivers have similar performances when the SNR is high.

Fig. 3.8 shows the BER versus SNR provided by the proposed ZF and MMSE PDRs using channel estimates provided by the proposed channel estimation method, and assuming that the global channel coefficients are known, for $R = T = 1$, $N_P = 16$ and $I_P = 2$. It is also shown the BER provided by the proposed MMSE receiver for a linear PA, assuming that the channel coefficients are known. Once again, the ZF and MMSE receivers have similar performances when the SNR is high. Besides, BER obtained with the known and estimated channel are similar

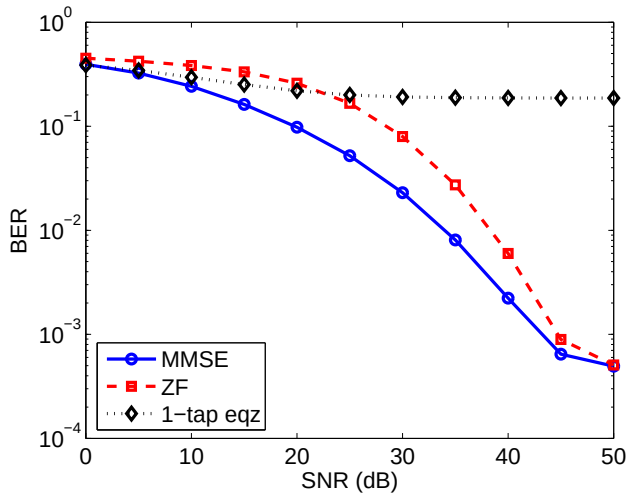

Figure 3.7: BER versus SNR provided by a single-tap equalizer and by the proposed ZF and MMSE PDRs - $R = T = 1$ with memoryless PA

for SNRs up to 35dB.

Fig. 3.9 shows the BER versus SNR provided by the proposed ZF and MMSE MIMO-PDRs using channel estimates provided by the proposed channel estimation method, for $R = T = 2$, $N_P = 16$ and $I_P = 2$, and the BER provided by the PA nonlinearity cancelation (PANC) [67], for $N_P = 32$ and $I_P = 1$. Note that the proposed PDRs provide a BER of 3.10^{-4} for high SNRs, while the PANC provides a BER of 6.10^{-3}. The drawback of the PDR receivers is that their transmission rate is two times smaller than the one of the PANC.

Fig. 3.10 shows the BER versus SNR provided by the proposed ZF and MMSE MIMO-PDRs using channel estimates provided by the proposed channel estimation method, and assuming that the global channel coefficients are known, for $R = T = 2$, $N_P = 16$ and $I_P = 2$. It is also shown the BER provided by the proposed MMSE receiver for $R = 3$ and $T = 2$. Once again, the BER obtained with the known and estimated channel are similar for SNRs up to 35dB. Besides, it can be seen that the BER provided by the use of an additional antenna is significant

Moreover, concerning Figs. 3.7-3.10, it can viewed that, in some cases, the BERs are not equal to zero even for high SNRs. This is due to the fact that the ICI and/or MAI are not perfectly canceled.

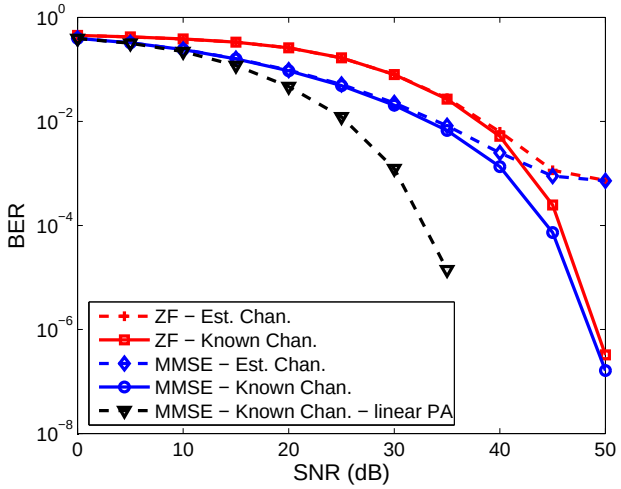

Figure 3.8: BER versus SNR provided by the proposed ZF and MMSE PDRs with known and estimated channels - $R = T = 1$ with memoryless PA

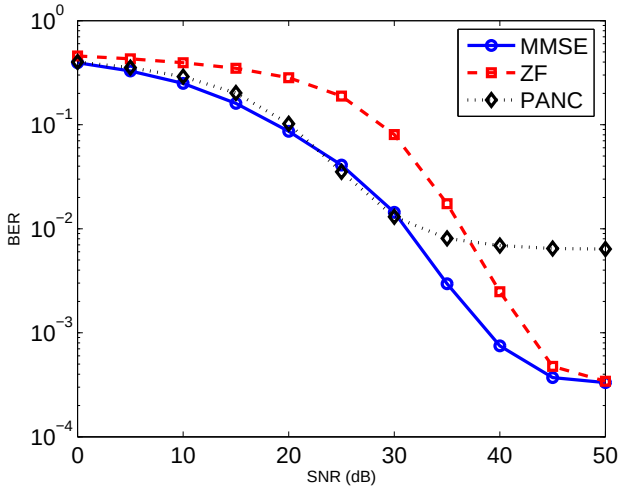

Figure 3.9: BER versus SNR provided by the PANC and by the proposed ZF and MMSE MIMO-PDRs - $R = T = 2$ with memoryless PA

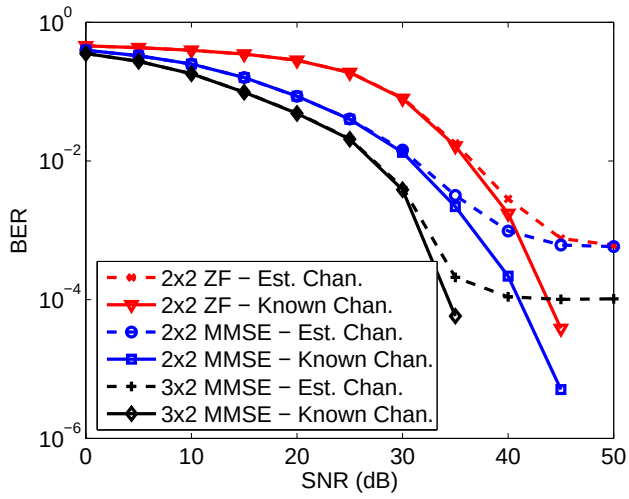

Figure 3.10: BER versus SNR provided by the proposed ZF and MMSE MIMO-PDRs with known and estimated channels - $R = T = 2$ and $R = 3$, $T = 2$ with memoryless PA

3.5.3 Memory polynomial PA - channel estimation

This section provides simulation results corresponding to channel estimation in the case of memory polynomial PAs. As mentioned earlier, in this case, the length of the cyclic prefix is equal to $M + M_{pa} = 7$. Fig. 3.11 shows the $NMSE_H$ and $NMSE_{Hlin}$ provided by proposed channel estimation technique, for $R = T = 1$, $N_P = 16$ and $I_P = 2$, and the $NMSE_L$ provided by the Known PA estimation technique, for $N_P = 32$ and $I_P = 1$. In this case, the SNR gain provided by the Known PA technique with respect to the proposed one is approximatively equal to 10dB. However, it should be highlighted that, if the Known PA technique is used, the transmitter has to estimate and transmit the 8 complex-valued coefficients that characterizes the memory polynomial PA. Similar results were obtained for the MIMO case. Fig. 3.12 shows the $NMSE_H$ and $NMSE_{Hlin}$ provided by proposed channel estimation technique, for $R = T = 2$, $N_P = 8$ and $I_P = 2$, and the $NMSE_L$ provided by the Known PA estimation technique, for $N_P = 16$ and $I_P = 1$.

The next figure shows the influence of the number of pilot subcarriers N_P and the number I_P of pilot symbols per subcarrier per user on the estimation of MIMO global channel coefficients in the case of a memory polynomial PA. Fig. 3.13 shows the NMSE versus SNR provided by proposed channel estimation technique for various values of N_P and I_P, for $R = T = 2$. It can be viewed that the use

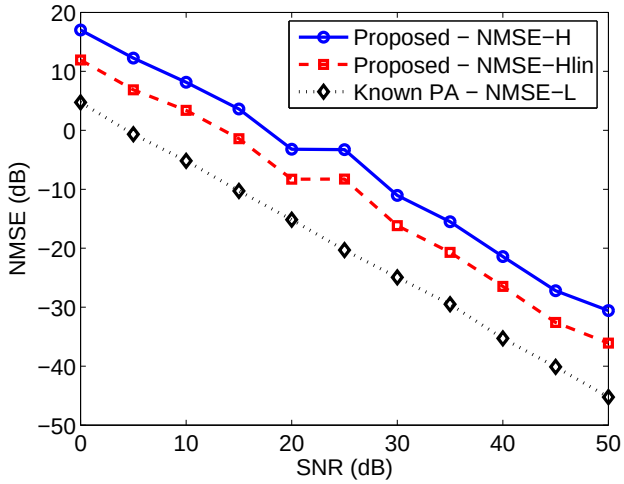

Figure 3.11: NMSE versus SNR for proposed and Known PA techniques - $R = T = 1$ with memory polynomial PA

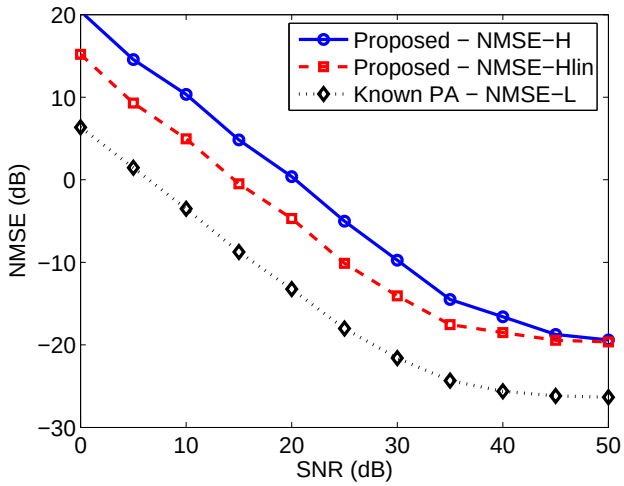

Figure 3.12: NMSE versus SNR for proposed and Known PA techniques - $R = T = 2$ with memory polynomial PA

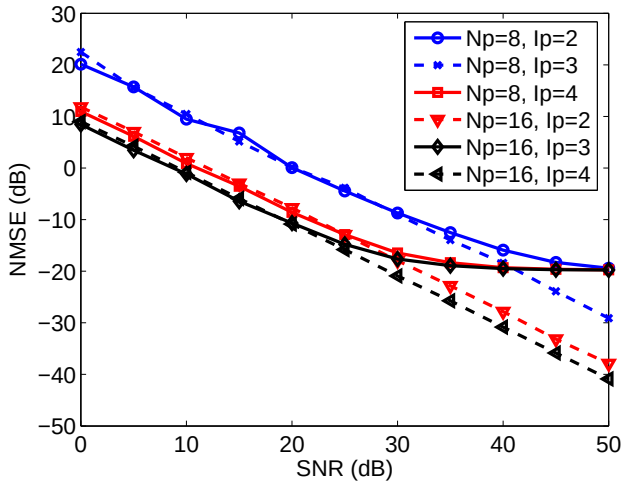

Figure 3.13: NMSE versus SNR for various values of N_p and I_P - $R = T = 2$ with memory polynomial PA

of $N_P = 16$ pilot subcarriers improves the channel estimation with respect to the case $N_P = 8$ when the SNR is higher than 30dB. Moreover, it can be noted that the NMSE is significantly improved if 3 or 4 pilot symbols are used per subcarrier per user, instead of 2. This result shows that, if we increase the number of pilot symbols, the proposed channel estimation technique may provide results similar to those obtained by the Known PA technique, with the advantage of not assuming that the PA coefficients are known.

3.5.4 Memory polynomial PA - channel equalization

The next two figures contain simulation results concerning channel equalization in the case of memory polynomial PAs. Fig. 3.14 shows the BER versus SNR provided by the proposed ZF and MMSE MIMO-PDRs using channel estimates provided by the proposed channel estimation method, for $R = T = 2$, $N_P = 16$ and $I_P = 2$, and the BER provided by the PANC technique, for $N_P = 32$ and $I_P = 1$. Note that the PANC does not perform well in the case of a memory polynomial PA, while the PDRs provide a BER of 7.10^{-4} for high SNRs.

Fig. 3.15 shows the BER versus SNR provided by the proposed ZF and MMSE MIMO-PDRs using channel estimates provided by the proposed channel estimation method, and assuming that the global channel coefficients are known, for $R = T = 2$, $N_P = 16$ and $I_P = 2$. It should be highlighted that the BER obtained with the

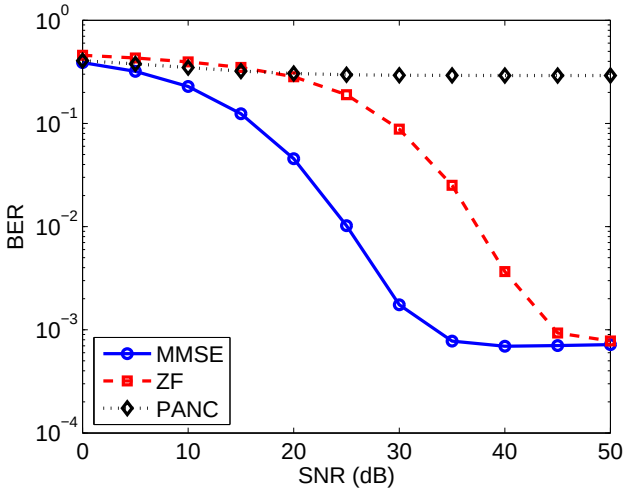

Figure 3.14: BER versus SNR provided by the PANC and by the proposed ZF and MMSE MIMO-PDRs - $R = T = 2$ with memory polynomial PA

known and estimated channel are similar for SNRs up to 25dB in the case of the MMSE MIMO-PDR and up to 35dB in the case of the ZF MIMO-PDR.

The main conclusion that can be drawn from the last five figures is that, when PDRs are used, the simulation results obtained with memory polynomial PAs are very similar to those obtained with memoryless polynomial PAs. That reinforces the results obtained in Section 3.4, i.e. if channel estimation and equalization methods based on the use of the global channel parameters are used, the cases of memoryless and memory polynomial PAs can be treated in the same way.

3.6 Conclusion

In this chapter, we have proposed techniques for channel estimation and equalization of MIMO-OFDM channels considering nonlinear PAs. These techniques are based on the use of global channel coefficients that characterize the cascade of the nonlinear PA and the linear wireless link. Initially, we have considered OFDM systems with memoryless PAs. In this case, we have developed a supervised LS channel estimation technique whose main advantage is that it estimates the linear and nonlinear coefficients jointly, do not requiring the knowledge of the PA parameters. Also in the case of memoryless PAs, two channel equalization techniques were proposed based on a transmission scheme that re-transmits all the symbols

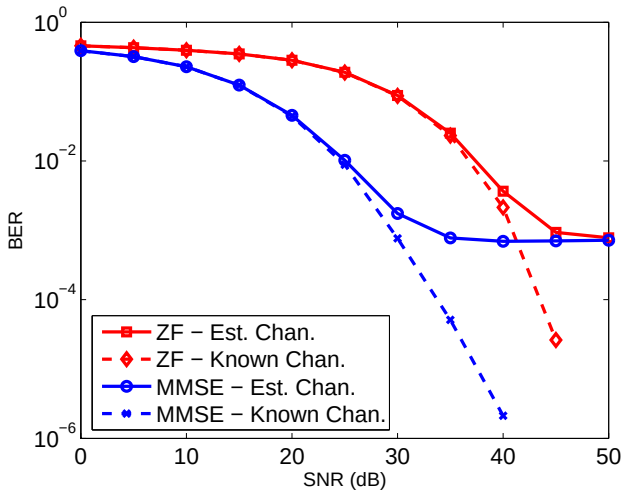

Figure 3.15: BER versus SNR provided by the proposed ZF and MMSE MIMO-PDRs with known and estimated channels - $R = T = 2$ with memory polynomial PA

several times with a different transmission power each time. These techniques have provided good simulation results, their main drawback being the fact that the transmission rate is divided by the repetition factor.

Moreover, we have demonstrated that, in a OFDM system, a memory polynomial PA can be expressed as a memoryless polynomial PA with coefficients that vary from one subcarrier to another. We have then shown that, when the global channel parameters are used, memoryless and memory polynomial PAs lead to identical expressions for the frequency domain received signals. Thus, the channel estimation and equalization techniques proposed for memoryless PAs can be directly applied in the case of memory polynomial PAs. As expected, the simulation results concerning memory polynomial PAs were very similar to those concerning memoryless PAs.

In the next chapter, we develop techniques for supervised estimation of MIMO Volterra channels. However, in Chapter 4, the proposed techniques are designed for time and space division multiple access (TDMA-SDMA) systems, instead of OFDM systems. In the next chapter, we deal with MIMO Volterra channels more general than the ones used in the present chapter, the channel model being the most general type of MIMO Volterra system used in this thesis. A well-known and efficient approach for estimating SISO Volterra models is the use of orthonormal polynomials [14, 108, 105, 137, 120, 134, 59]. In Chapter 4, we extend the procedure

of construction and the use of orthonormal polynomials to the case of MIMO Volterra systems. Moreover, the proposed techniques are applied to the estimation of a multiuser nonlinear ROF channel.

CHAPTER 4

Supervised Estimation of MIMO Volterra Channels Using Orthonormal Polynomials

THIS chapter deals with the estimation of discrete-time equivalent baseband MIMO Volterra communication channels in a supervised scenario, i.e. the transmitted signals are assumed to be known. The channel model considered in this chapter is the general MIMO Volterra system (2.57) or, equivalently, (2.56). This channel model is the most general MIMO Volterra system used in this thesis, with the channel possibly having different memories with respect to the sources. Few authors have proposed channel estimation or equalization techniques for MIMO Volterra communication channels and, to the best of our knowledge, none of them correspond to the general MIMO Volterra model (2.56)-(2.57). A receiver for MIMO ultra-wideband (UWB) systems was proposed in [113], based on a MIMO Volterra model of order two with binary PSK (BPSK) transmitted signals. Besides, channel estimation and equalization of MIMO Volterra channels that do not correspond to a nonlinear mixture of the sources were proposed in [67, 133] and [127], respectively for OFDM and CDMA systems. Some works dealing with MIMO Volterra systems were also proposed considering other kinds of applications, however, to the best of our knowledge, they also use less generical models than the one considered in this chapter [123, 1, 124, 136, 2].

The main drawback of Volterra models is their high number of coefficients, which may significantly difficult the estimation of these coefficients. For instance, in the case of adaptive schemes, although the estimation of the Volterra parameters can

be done by a number of algorithms, it is often carried out by the Least Mean Square (LMS) algorithm [163] due to its simplicity [14, 108]. However, the convergence speed of the LMS algorithm is slow if the condition number (or eigenvalue spread) of the covariance matrix of the regression vector is high [74]. The condition number of a matrix is defined as the ratio between its highest and smallest eigenvalues. For Volterra systems, this covariance matrix has, in general, a high eigenvalue spread [108], specially when the input signals are QAM modulated. Hence, the estimation of Volterra channels by means of the LMS algorithm is, in general, very slow if the transmitted signals are QAM modulated. To overcome this problem, we develop a set of orthonormal polynomials to improve the conditioning of the covariance matrix of the nonlinear input vector associated with a MIMO Volterra system with QAM inputs. Thus, the use of this set of orthonormal polynomials accelerates the convergence of the LMS algorithm when performing adaptive estimation of MIMO Volterra channels.

Several works have used similar orthogonalization approaches for SISO Volterra systems [14, 108, 137, 134, 59, 84]. However, these methods developed for SISO systems can not be applied to a MIMO Volterra system like (2.57), since the received signals are nonlinear mixtures of the sources that can have different probability density functions (PDFs) and different delay spreads (channel memory). To the best of our knowledge, there is no similar orthogonalization technique for MIMO Volterra systems. In this chapter, we propose a procedure of construction of orthonormal polynomials in the case of MIMO Volterra systems, allowing different PDFs for the input signals and different memories with respect to the inputs, so that the orthonormal polynomials can be applied to the general equivalent baseband MIMO Volterra channel (2.57).

The orthonormal polynomials can also be exploited in the calculus of the Minimum Mean Squared Error (MMSE) estimation of MIMO Volterra systems in a block processing scheme. In this case, the use of orthonormal polynomials makes possible a decoupled estimation of each coefficient of the MIMO Volterra system, leading to a reduction of the computational complexity associated with the calculus of the MMSE estimate of the system coefficients.

Other contribution of this chapter is the application of the proposed techniques for the estimation of uplink nonlinear MIMO Radio-Over-Fiber (ROF) channels, considering multiple users and multiple receive antennas. Some works have been done in the last years on the estimation and equalization of nonlinear ROF channels, specially by X. N. Fernando, S. Z. Pinter and A. B. Sesay. In the SISO case, a Hammerstein-type decision feedback equalizer for a Wiener ROF uplink channel was proposed in [57]. References [55, 54] developed a pre-distorter for the downlink, while a post-compensation equalization technique was proposed for the

uplink, assuming an Additive White Gaussian Noise (AWGN) wireless channel. A pre-distortion technique for ROF links was also developed in [82].

In the MISO case, i.e. considering multiple users and a single antenna at the reception, an estimation technique for a ROF uplink channel was presented in [116, 114] for a CDMA system using pseudo-noise spreading codes. The same authors have proposed an equalizer for the ROF downlink channel in a CDMA environment using Walsh codes [117]. Although these estimation and equalization techniques proposed for the MISO case could be applied in the MIMO case, they are designed for CDMA systems, contrarily to the ones developed in this chapter. The development of techniques for ROF channels in CDMA systems will be treated in Chapter 6 of this thesis.

Other approaches have been considered for the mitigation of nonlinear distortions on ROF links, like equivalent circuit model based solutions [29, 156], "post nonlinearity recovery block" [125] and electronic pre-distortion schemes [58, 70]. However, as pointed out in [55], these solutions have many drawbacks, as the fact that they are device dependent and that their accuracy depend on the knowledge of physical parameters. In addition, these approaches can not take into account nonlinear distortions due to other possible nonlinear devices, such as PAs.

This chapter is organized as follows. Section 4.1 presents the channel model considered in this chapter. In Section 4.2, the problem associated with the estimation of Volterra systems by the LMS algorithm is presented. In Section 4.3, the set of orthonormal polynomials is developed. In Section 4.4, MIMO Volterra channel estimation using orthonormal polynomials is addressed. Section 4.5 provides some simulation results to illustrate the performance of the proposed methods and Section 4.6 draws some conclusions about the work. Moreover, in Appendix B, the procedure of construction of the orthonormal monomials used by the orthonormal polynomials is presented. The results presented in this chapter have been published in [52, 46].

4.1 Channel Model

The channel model used in this chapter is a noisy version of the discrete-time equivalent baseband MIMO Volterra channel (2.57):

$$\begin{aligned} y_r(n) &= x_r(n) + v_r(n), \\ y_r(n) &= \sum_{k=0}^{K} \sum_{\bar{m}_1=1}^{\bar{M}} \cdots \sum_{\bar{m}_{k+1}=\bar{m}_k}^{\bar{M}} \sum_{\bar{m}_{k+2}=1}^{\bar{M}} \cdots \sum_{\bar{m}_{2k+1}=\bar{m}_{2k}}^{\bar{M}} h_{2k+1}^{(r)}(\bar{m}_1, \ldots, \bar{m}_{2k+1}) \\ &\quad \prod_{i=1}^{k+1} \bar{s}_{\bar{m}_i}(n) \prod_{i=k+2}^{2k+1} \bar{s}^*_{\bar{m}_i}(n) + v_r(n), \end{aligned} \tag{4.1}$$

where $y_r(n)$ and $x_r(n)$ $(1 \leq r \leq R)$ are respectively the noisy and noiseless versions of the signal received by antenna r at the time instant n, R is the number of receive antennas, $(2K+1)$ is the nonlinearity order of the model, $h_{2k+1}^{(r)}(\bar{m}_1, \ldots, \bar{m}_{2k+1})$ are the coefficients of the $(2k+1)^{th}$-order Volterra kernel of the r^{th} sub-channel, $\bar{s}_{\bar{m}_i}(n)$ corresponds to the $\bar{m}_i^{th}$ component of the linear input vector given by (2.58), $v_r(n)$ is the zero-mean AWGN component and $\bar{M} = M_1 + \cdots + M_T + T$, M_t being the channel memory with respect to the t^{th} source and T the number of users. In this chapter, we assume that the transmitted signals $s_t(n)$ are known and independent and identically-distributed (i.i.d.). Moreover, the noise components $v_r(n)$ are assumed to be zero mean, independent from each other and from the transmitted signals $s_t(n)$.

Note that, in Chapter 2, triangular MIMO Volterra filters were denoted by $\tilde{h}_{2k+1}^{(r)}(\bar{m}_1, \ldots, \bar{m}_{2k+1})$. However, for simplifying the notation, henceforth they will be denoted by $h_{2k+1}^{(r)}(\bar{m}_1, \ldots, \bar{m}_{2k+1})$.

As described in Chapter 2, the MIMO Volterra model (4.1) can be expressed in a compact form:

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n) = \mathbf{H}\mathbf{w}(n) + \mathbf{v}(n), \tag{4.2}$$

where $\mathbf{y}(n) = [y_1(n)\ y_2(n) \ldots y_R(n)]^T \in \mathbb{C}^{R\times 1}$, $\mathbf{v}(n) = [v_1(n)\ v_2(n) \ldots v_R(n)]^T \in \mathbb{C}^{R\times 1}$, $\mathbf{w}(n) \in \mathbb{C}^{Q\times 1}$ is given by (2.60), $\mathbf{x}(n) \in \mathbb{C}^{R\times 1}$ and $\mathbf{H} \in \mathbb{C}^{R\times Q}$ are defined as in (2.59), and Q is given by (2.61).

4.2 LMS Volterra Channel Estimation

As mentioned earlier, the adaptive estimation of Volterra systems is generally done by the LMS algorithm due to its simplicity. The LMS estimation of the channel matrix $\mathbf{H}$ is based on the minimization of the MMSE cost function:

$$J = \mathbb{E}\left[\|\mathbf{y}(n) - \mathbf{H}\mathbf{w}(n)\|_F^2)\right] = \mathbb{E}\left[\mathbf{e}^H(n)\mathbf{e}(n)\right], \tag{4.3}$$

where $\|\cdot\|_F$ denotes the *Frobenius norm* and $\mathbf{e}(n) = \mathbf{y}(n) - \mathbf{H}\mathbf{w}(n) \in \mathbb{C}^{R\times 1}$. Denoting by $\mathbf{h}_r \in \mathbb{C}^{Q\times 1}$ the transpose of the r^{th} row of $\mathbf{H}$, the gradient of (4.3) with respect to $\mathbf{h}_r$ is given by:

$$\nabla_{\mathbf{h}_r} J = \mathbb{E}\left[-2y_r(n)\mathbf{w}^*(n) + 2[\mathbf{h}_r^T\mathbf{w}(n)]\mathbf{w}^*(n)\right]. \tag{4.4}$$

The LMS algorithm tries to minimize the MMSE cost function based on the steepest descent algorithm (or gradient descent algorithm) [74]. The steepest descent algorithm is an iterative optimization technique that starts at an initial point and, at each iteration, takes a step proportional to the negative of the gradient of the cost function at the current point. The LMS algorithm can be viewed as a stochastic version of the steepest descent algorithm that uses instantaneous estimates of the gradient vector. The LMS estimation of the channel matrix $\mathbf{h}_r$ is then given by the following adaptation equation:

$$\hat{\mathbf{h}}_r(n+1) = \hat{\mathbf{h}}_r(n) + \mu\left(y_r(n) - \hat{\mathbf{h}}_r^T(n)\mathbf{w}(n)\right)\mathbf{w}^*(n), \tag{4.5}$$

where μ is the step-size parameter and the vector $\hat{\mathbf{h}}_r(n)$ represents the estimate value of $\mathbf{h}_r$ at the n^{th} time instant. Equivalently, we may write:

$$\hat{\mathbf{H}}(n+1) = \hat{\mathbf{H}}(n) + \mu\left(\mathbf{y}(n) - \hat{\mathbf{H}}(n)\mathbf{w}(n)\right)\mathbf{w}^H(n), \tag{4.6}$$

where the matrix $\hat{\mathbf{H}}(n)$ denotes the estimate value of $\mathbf{H}$ at the n^{th} time instant.

Let us define the weight error vector for the r^{th} received signal ($1 \le r \le R$) as: $\mathbf{e}_r^{(h)}(n) \equiv \hat{\mathbf{h}}_r(n) - \mathbf{h}_r \in \mathbb{C}^{Q\times 1}$, which leads to [108]:

$$\mathbb{E}[\mathbf{e}^{(h)}(n+1)] = (\mathbf{I}_Q - \mu\mathbf{R}_{\mathbf{w}})\,\mathbb{E}[\mathbf{e}^{(h)}(n)], \tag{4.7}$$

where $\mathbf{R}_{\mathbf{w}} = \mathbb{E}[\mathbf{w}(n)\mathbf{w}^H(n)]$ is the covariance matrix of the nonlinear input vector $\mathbf{w}(n)$. Let $\mathbf{Q} \in \mathbb{C}^{Q\times Q}$ be the matrix containing the eigenvectors of $\mathbf{R}_{\mathbf{w}}$. Thus, by

defining:

$$\mathbf{v}^{(h)}(n) = \mathbf{Q}^H \mathbf{e}^{(h)}(n), \tag{4.8}$$

we may write:

$$\mathbb{E}[\mathbf{v}^{(h)}(n+1)] = (\mathbf{I}_Q - \mu\mathbf{\Lambda})\,\mathbb{E}[\mathbf{v}^{(h)}(n)], \tag{4.9}$$

where $\mathbf{\Lambda} \in \mathbb{C}^{Q\times Q}$ is a diagonal matrix containing the eigenvalues λ_q $(1 \leq q \leq Q)$ of $\mathbf{R_w}$. Equation (4.9) can be rewritten in a scalar form as:

$$\mathbb{E}[v_q^{(h)}(n+1)] = (1 - \mu\lambda_q)\,\mathbb{E}[v_q^{(h)}(n)], \tag{4.10}$$

for $q = 1, ..., Q$, which leads to

$$\mathbb{E}[v_q^{(h)}(n)] = (1 - \mu\lambda_q)^n\,\mathbb{E}[v_q^{(h)}(0)]. \tag{4.11}$$

Hence, for convergence we must have:

$$0 < \mu < \frac{2}{|\lambda_{max}|}, \tag{4.12}$$

where λ_{max} is the eigenvalue of $\mathbf{R_w}$ with the highest magnitude. Moreover, the equation (4.11) can be approximated by [108]:

$$\mathbb{E}[v_q^{(h)}(n)] \cong e^{-\tau_q n}\mathbb{E}[v_q^{(h)}(0)], \tag{4.13}$$

with the time constants given by

$$\tau_q = \frac{1}{2\mu\lambda_q}, \tag{4.14}$$

for $q = 1, ..., Q$. It can then be concluded from (4.14) that the convergence speed of the LMS algorithm is limited by the smallest eigenvalue λ_{min}, the maximal time constant being given by $\tau_{min} = \frac{1}{2\mu\lambda_{min}}$.

On the other hand, a measure of the misadjustment of the LMS algorithm is provided by the Excess Mean Square Error (EMSE) of the r^{th} received signal, defined as:

$$\xi_{EMSE}^{(r)}(n) = \mathbb{E}[\hat{\xi}^{(r)}(n)] - \xi_{min}^{(r)}, \quad 1 \leq r \leq R, \tag{4.15}$$

where $\hat{\xi}^{(r)}(n) = \mathbb{E}[|y_r(n) - \mathbf{f}_r\mathbf{w}(n)|^2]$ is the mean square error of the r^{th} received signal provided by the LMS algorithm at iteration n and $\xi_{min}^{(r)}$ is the corresponding minimum mean square error provided by the Wiener solution. The EMSE provided

by the LMS algorithm in steady state can be approximated by [108]:

$$\xi_{EMSE}^{(r)}(\infty) \cong \mu\, \xi_{min}^{(r)} \sum_{q=1}^{Q} \lambda_q. \tag{4.16}$$

The ill-conditioning of the correlation matrix $\mathbf{R_w}$ means that its eigenvalues are widely spread, i.e. some eigenvalues are much higher than some others. In this case, it can be seen from (4.16) that the EMSE produced by the LMS algorithm is mainly determined by the largest eigenvalues. Thus, if there exist some eigenvalues with high values, the step-size parameter should have a small value in order to obtain a low EMSE. On the other hand, small eigenvalues and a small step-size lead to high time constants, which slows down the convergence of the LMS algorithm. On other words, for an ill-conditioned matrix $\mathbf{R_w}$, the smallest eigenvalues lead to a slow convergence, while the highest eigenvalues lead to a high EMSE.

For Volterra systems, the covariance matrix $\mathbf{R_w}$ of the nonlinear input vector has often a high eigenvalue spread [108], specially for QAM signals, as it can be viewed in Table 4.1. This figure shows the eigenvalue spread of the matrix $\mathbf{R_w}$ associated with a equivalent baseband MIMO Volterra channel for P-PSK, 16-QAM, 64-QAM and 256-QAM constellations, assuming an uniform and i.i.d. distribution. Three different cases are considered, according to the number of sources (T) and channel memory (M): (1) $T = 2$ and $M = 1$, (2) $T = 3$ and $M = 1$, (3) $T = 2$ and $M = 2$. In Table 4.1, the input signals have unit power and the channel nonlinearity order is three ($K = 1$), which means that these cases correspond respectively to $Q = 4, 12$ and 28 virtual sources for PSK sources and to $Q = 8, 21$ and 44 for QAM sources. It can be viewed that the eigenvalue spread is much higher for QAM constellations than for PSK constellations. In fact, the eigenvalue spread is equal to one for PSK signals. This is due to the fact that the matrix $\mathbf{R_w}$ is diagonal for this kind of signals [99].

Table 4.1: Eigenvalue spread of the covariance matrix of the nonlinear input vector - uniform i.i.d. signals.

Eigenvalue spread / Modulation	Case 1	Case 2	Case 3
P-PSK	1	1	1
16-QAM	63.27	94.41	131.76
64-QAM	51.46	77.55	108.90
256-QAM	49.48	74.67	104.93

In the next section, a set of orthonormal polynomials is developed to decrease the eigenvalue spread of the covariance matrix of the nonlinear input vector associated with an equivalent baseband MIMO Volterra system, assuming that all the sources transmit uniformly distributed i.i.d. QAM symbols. Besides, it is assumed that

$$P_t > 2K + 1, \quad \text{for} \quad t =, 1..., T, \tag{4.17}$$

where P_t is the number of points of the QAM constellation of the t^{th} source. Equation (4.17) corresponds to the well-known persistence of excitation condition for a Volterra system of order $2K + 1$ [107]. Moreover, for simplifying the developments, we also assume that P_t is even, for $t = 1, ..., T$, i.e. the signals transmitted by all the sources belong to square QAM constellations. In this case, the following property holds:

$$\mathbb{E}[s_t^i(n)s_y^{*^j}(n)] = 0, \quad \forall(i, j) \quad i \neq j, \quad i + j \leq 2K + 1, \tag{4.18}$$

where $s_t(n)$ is the signal transmitted by the t^{th} user, which means that the random variables $s_t(n)$ are circular of order $2K + 1$, for $t =, 1..., T$.

Using an orthonormal basis, the covariance matrix $\mathbf{R_w}$ is equal to the identity matrix and the eigenvalue spread is equal to 1, leading to a faster convergence of the LMS algorithm. Several works have used similar orthogonalization approaches for SISO Volterra systems [14, 108, 137, 134, 59, 84]. However, these techniques can not be applied to the MIMO Volterra channel (4.1) as the received signals are nonlinear mixture of the sources. The developments of the next sections extend the procedure of construction and the use of orthonormal polynomials to the case of MIMO Volterra systems, allowing different PDF's for the source signals and different channel memories with respect to the sources.

.

4.3 Orthonormal Polynomials

In this section, a set of orthonormal polynomials is developed for the discrete-time equivalent baseband MIMO Volterra channel (4.1). The orthonormalization problem considered here consists in finding a lower triangular matrix $\mathbf{T} \in \mathbb{C}^{Q\times Q}$ so that $\breve{\mathbf{w}}(n) = \mathbf{T}\mathbf{w}(n)$, satisfying $\mathbf{R}_{\breve{\mathbf{w}}} = \mathbf{T}\mathbf{R}_{\mathbf{w}}\mathbf{T}^H = \mathbf{I}_Q$, where $\breve{\mathbf{w}}(n)$ is a nonlinear input vector in an orthonormal basis, $\mathbf{R}_{\breve{\mathbf{w}}} = \mathbb{E}[\breve{\mathbf{w}}(n)\breve{\mathbf{w}}^H(n)]$ is the covariance matrix of $\breve{\mathbf{w}}(n)$ and $\mathbf{I}_Q$ is the identity matrix of order Q. As pointed out in Section 4.2, this orthonormalization procedure is not necessary in the case of uniformly distributed

PSK signals.

The elements of the nonlinear input vector $\mathbf{w}(n)$, given by (2.60), are multivariable functions of the delayed input signals that can be expressed as products of monomials in $\bar{s}_{\bar{m}}(n)$. We recall that $\bar{s}_{\bar{m}}(n)$, for $\bar{m} = 1, ..., \bar{M}$, correspond to the delayed signals transmitted by all the sources. The basic idea of the orthonormalization method is that by exploiting the hypothesis of independency between the inputs and their i.i.d. characteristic, an orthonormalization can be carried out by applying the Gram-Schmidt procedure to the set of monomials that composes $\mathbf{w}(n)$. The orthonormal polynomials are then obtained as products of orthonormal monomials. An advantage of this method is that the Gram-Schmidt orthonormalization is applied to calculate only few monomials, even if the system has a high number of nonlinear input terms.

Let $C_{2K+1}[s_1, \cdots, s_{\bar{M}}] = C_{2K+1}[S]$, with $S = \{s_1, \cdots, s_{\bar{M}}\}$, be the space of polynomials of order equal or smaller than $2K+1$ in the $\bar{M}$ random variables $s_1, s_2, ..., s_{\bar{m}}$, these variables being assumed to be complex-valued, persistently exciting of order $2K+1$ [107] and independent from each other. It it is important to highlight that the random variables $\{s_1, \cdots, s_{\bar{M}}\}$ does not need to have the same PDF. In addition, it is considered that the polynomials have only odd-order kernels with one more non-conjugated term than conjugated terms, i.e. polynomials have the form $\prod_{i=1}^{k+1} s_{\bar{m}_i} \prod_{i=k+2}^{2k+1} s^*_{\bar{m}_i}$, with $0 \leq k \leq K$, $1 \leq \bar{m}_1 \leq ... \leq \bar{m}_{k+1} \leq \bar{M}$ and $1 \leq \bar{m}_{k+2} \leq ... \leq \bar{m}_{2k+1} \leq \bar{M}$.

Let us associate the space $C_{2K+1}[S]$ with the following scalar product:

$$< A(S), B(S) > \equiv \mathbb{E}[A(S)B^*(S)], \tag{4.19}$$

where $A(S)$ and $B(S)$ are polynomials in $C_{2K+1}[S]$. The canonical basis of the space $C_{2K+1}[S]$ is constituted by the following components:

$$\begin{aligned} W^{(2k+1)}_{\bar{m}_1,...,\bar{m}_{2k+1}}(S) &\equiv \\ &\equiv W^{(2k+1)}_{\bar{m}_1,...,\bar{m}_{2k+1}}(s_{\bar{m}_1}, \ldots, s_{\bar{m}_{k+1}}, s^*_{\bar{m}_{k+2}}, \ldots, s^*_{\bar{m}_{2k+1}}), \qquad (4.20) \\ &= \prod_{i=1}^{\bar{M}} T_{\alpha_i,\beta_i}(s_i), \qquad (4.21) \end{aligned}$$

where α_i (resp. β_i) is the cardinality of s_i (resp. ${s_i}^*$) in the set $\{s_{\bar{m}_1}, \ldots, s_{\bar{m}_{k+1}}\}$ (resp. $\{s^*_{\bar{m}_{k+2}}, \ldots, s^*_{\bar{m}_{2k+1}}\}$) and $T_{\alpha_i,\beta_i}(s_i)$ are canonical monomials given by

$$T_{\alpha_i,\beta_i}(s_i) = (s_i)^{\alpha_i} ({s_i}^*)^{\beta_i}. \tag{4.22}$$

Note that $0 \leq \alpha_i \leq k+1$, $0 \leq \beta_i \leq k$, $\sum_{i=1}^{\bar{M}} \alpha_i = k+1$ and $\sum_{i=1}^{\bar{M}} \beta_i = k$. Note that the basis function $W^{(2k+1)}_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}(S)$ is equal to the term $\{s_1^{\alpha_1} s_1^{*\beta_1} \ldots s_T^{\alpha_T} s_T^{*\beta_T}\}$. For instance, for $K=1$ (linear and cubic terms), the canonical basis is given by:

$$W^{(1)}_{l}(S) = T_{1,0}(s_l) = s_l, \tag{4.23}$$

$$W^{(3)}_{l,l,l}(S) = T_{2,1}(s_l) = s_l^2 s_l^*, \tag{4.24}$$

$$W^{(3)}_{l,l,j}(S) = T_{2,0}(s_l)T_{0,1}(s_j) = s_l^2 s_j^*,\ l \neq j \tag{4.25}$$

$$W^{(3)}_{l,j,l}(S) = T_{1,1}(s_l)T_{1,0}(s_j) = s_l s_l^* s_j,\ l < j \tag{4.26}$$

$$W^{(3)}_{l,j,k}(S) = T_{1,0}(s_l)T_{1,0}(s_j)T_{0,1}(s_k) = s_l s_j s_k^*,\ l < j, \tag{4.27}$$

where $l, j, k = 1, ..., \bar{M}$.

The scalar product between two components of the canonical basis is equal to:

$$\begin{aligned}\left\langle W^{(2k+1)}_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}(S), W^{(2k'+1)}_{\bar{m}'_1,\cdots,\bar{m}'_{2k'+1}}(S) \right\rangle &= \mathbb{E}\left[\prod_{i=1}^{\bar{M}} T_{\alpha_i,\beta_i}(s_i) \prod_{i=1}^{\bar{M}} T^*_{\alpha'_i,\beta'_i}(s_i)\right] \\ &= \prod_{i=1}^{\bar{M}} \mathbb{E}\left[T_{\alpha_i,\beta_i}(s_i)\, T^*_{\alpha'_i,\beta'_i}(s_i)\right] = \prod_{i=1}^{\bar{M}} \mathbb{E}\left[(s_i)^{\alpha_i+\beta'_i} (s_i{}^*)^{\beta_i+\alpha'_i}\right]. \end{aligned} \tag{4.28}$$

Note that if $\alpha_i + \beta'_i = \beta_i + \alpha'_i$ for all i, the scalar product (4.28) is non-null. This means that even if the bases $W^{(2k+1)}_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}(S)$ and $W^{(2k'+1)}_{\bar{m}'_1,\cdots,\bar{m}'_{2k'+1}}(S)$ are different, their scalar product can be different of zero. That demonstrates that the canonical basis is not orthogonal.

The set of canonical monomials $\{T_{\alpha_i,\beta_i}(y_i)\}$ can be orthonormalized by using the Gram-Schmidt procedure, as described in Appendix B. That leads to a set of orthonormal monomials, denoted by $P_{\alpha_i,\beta_i}(s_i)$, $0 \leq \alpha_i \leq k+1$, $0 \leq \beta_i \leq k$, $i = 1, \cdots, \bar{M}$. For instance, for $K=1$ (linear-cubic nonlinearity), we have (see Appendix B):

$$P_{0,0}(s) = 1, \tag{4.29}$$

$$P_{1,0}(s) = \frac{s}{\sqrt{\rho_{s,1,1}}}, \tag{4.30}$$

$$P_{0,1}(s) = \frac{s^*}{\sqrt{\rho_{s,1,1}}}, \tag{4.31}$$

$$P_{2,0}(s) = \frac{s^2}{\sqrt{\rho_{s,2,2}}}, \tag{4.32}$$

$$P_{1,1}(s) = \frac{|s|^2 - \rho_{s,1,1}}{\sqrt{\rho_{s,2,2} - \rho_{s,1,1}^2}}, \tag{4.33}$$

$$P_{2,1}(s) = \frac{\rho_{s,1,1}|s|^2 s - \rho_{s,2,2} s}{\sqrt{\rho_{s,1,1}^2 \rho_{s,3,3} - \rho_{s,1,1}\rho_{s,2,2}^2}}. \tag{4.34}$$

where $\rho_{s,p,q} = \mathbb{E}[s^p s^{q^*}]$. Note that the orthonormal monomials $P_{\alpha_i,\beta_i}(s_i)$ depend on the statistics of the random variable s_i.

The multivariable orthonormal basis associated with the space $C_{2K+1}[S]$ is then given by the following product of the orthonormal monomials:

$$Q^{(2k+1)}_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}(S) = \prod_{i=1}^{\bar{M}} P_{\alpha_i,\beta_i}\left(s_i\right), \tag{4.35}$$

where α_i (resp. β_i) is the cardinality of s_i (resp. ${s_i}^*$) in the set $\{s_{\bar{m}_1}, \ldots, s_{\bar{m}_{k+1}}\}$ (resp. $\{s^*_{\bar{m}_{k+2}}, \ldots, s^*_{\bar{m}_{2k+1}}\}$). For instance, for $K = 1$, the orthonormal basis is given by:

$$\begin{aligned}
Q^{(1)}_l(S) &= P_{1,0}(s_l), && (4.36)\\
Q^{(3)}_{l,l,l}(S) &= P_{2,1}(s_l), && (4.37)\\
Q^{(3)}_{l,l,j}(S) &= P_{2,0}(s_l)P_{0,1}(s_j),\ l \neq j && (4.38)\\
Q^{(3)}_{l,j,l}(S) &= P_{1,1}(s_l)P_{1,0}(s_j),\ l < j && (4.39)\\
Q^{(3)}_{l,j,k}(S) &= P_{1,0}(s_l)P_{1,0}(s_j)P_{0,1}(s_k),\ l < j, && (4.40)
\end{aligned}$$

where $l, j, k = 1, ..., \bar{M}$. The above described polynomials are called the *Q-polynomials* by Schetzen in the case of real-valued Gaussian variables [134].

So, we have:

$$\left\langle Q^{(2k+1)}_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}(S), Q^{\left(2k'+1\right)}_{\bar{m}'_1,\cdots,\bar{m}'_{2k'+1}}(S) \right\rangle = \prod_{i=1}^{\bar{M}} \mathbb{E}\left[P_{\alpha_i,\beta_i}(s_i)\, P^{*}_{\alpha'_i,\beta'_i}(s_i) \right]$$
$$= \prod_{i=1}^{\bar{M}} \delta(\alpha_i - \alpha'_i)\delta(\beta_i - \beta'_i), \qquad (4.41)$$

where $\delta(\cdot)$ is the Kronecker symbol, i.e. $\delta(\alpha_i - \alpha'_i) = 1$ if $\alpha_i = \alpha'_i$ and $\delta(\alpha_i - \alpha'_i) = 0$ if $\alpha_i \neq \alpha'_i$. Equation (4.41) shows that the scalar product of two components of the basis is non-null if and only if $\alpha_i = \alpha'_i$ and $\beta_i = \beta'_i$, $\forall\, i$, $(l = 1, 2, ..., \bar{M})$, which means that $k = k'$, $\{\bar{m}'_1, \ldots, \bar{m}'_{k+1}\}$ is a permutation of $\{\bar{m}_1, \ldots, \bar{m}_{k+1}\}$ and $\{\bar{m}'_{k+2}, \ldots, \bar{m}'_{2k+1}\}$ is a permutation of $\{\bar{m}_{k+2}, \ldots, \bar{m}_{2k+1}\}$. As we have by definition $\bar{m}_1 \leq \bar{m}_2 \leq \cdots \leq \bar{m}_{k+1}$ and $\bar{m}_{k+2} \leq \bar{m}_{k+3} \leq \cdots \leq \bar{m}_{2k+1}$, we can conclude that $\{\bar{m}'_1, \ldots, \bar{m}'_{k+1}\} = \{\bar{m}_1, \ldots, \bar{m}_{k+1}\}$ and $\{\bar{m}'_{k+2}, \ldots, \bar{m}'_{2k+1}\} = \{\bar{m}_{k+2}, \ldots, \bar{m}_{2k+1}\}$. That demonstrates the orthonormality of the basis $\{Q^{(2k+1)}_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}(S)\}$.

4.4 MIMO Volterra Channel Estimation Using Orthonormal Polynomials

In this section, the set of orthonormal polynomials developed in the last section is applied to the estimation of the MIMO Volterra channel (4.1). From (4.1), it can be seen that the received signals $x_r(n)$ $(1 \leq r \leq R)$ are complex polynomial functions of $S = \{s_1(n), ..., s_1(n - M_1 + 1), \cdots, s_T(n), ..., s_T(n - M_T + 1)\} = \{\bar{s}_1(n), \bar{s}_2(n), ..., \bar{s}_{\bar{M}}(n)\}$, these polynomials having the form assumed in Section 4.3: $\prod_{i=1}^{k+1} s_{\bar{m}_i}(n) \prod_{i=k+2}^{2k+1} s^{*}_{\bar{m}_i}(n)$, with $0 \leq k \leq K$, $1 \leq \bar{m}_1 \leq ... \leq \bar{m}_{k+1} \leq \bar{M}$ and $1 \leq \bar{m}_{k+2} \leq ... \leq \bar{m}_{2k+1} \leq \bar{M}$. Moreover, assuming that the signals transmitted by the sources are uniformly distributed i.i.d. square QAM symbols, the signals $s_t(n)$ are circular of order $2K + 1$, i.e. (4.18) holds.

The received outputs can then be expressed in the orthonormal basis $Q^{(2k+1)}_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}(S)$ as:

$$x_r(n) = \sum_{k=0}^{K} \sum_{\bar{m}_1=1}^{\bar{M}} \cdots \sum_{\bar{m}_{k+1}=\bar{m}_k}^{\bar{M}} \sum_{\bar{m}_{k+2}=1}^{\bar{M}} \cdots \sum_{\bar{m}_{2k+1}=\bar{m}_{2k}}^{\bar{M}} f^{(r)}_{2k+1}(\bar{m}_1, \ldots, \bar{m}_{2k+1}) Q^{(2k+1)}_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}(S), \quad (4.42)$$

where

$$f_{2k+1}^{(r)}(\bar{m}_1,\ldots,\bar{m}_{2k+1}) = \left\langle x_r(n),\, Q_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}^{(2k+1)}(S)\right\rangle \tag{4.43}$$

are the channel coefficients in the orthonormal basis and $Q_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}^{(2k+1)}(S)$ is given by (4.35). It should be highlighted that the polynomials $P_{\alpha_i,\beta_i}(\bar{s}_i(n))$ and, consequently $Q_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}^{(2k+1)}(S)$, are functions of the input moments $\rho_{\bar{s}_i,p,q}$, which means that the input signals may have different PDF's.

Equation (4.42) can be written in a vector form as:

$$\mathbf{x}(n) = \mathbf{F}\breve{\mathbf{w}}(n), \tag{4.44}$$

where $\mathbf{F} \in \mathbb{C}^{R\times Q}$, with $\mathbf{F} = [\mathbf{f}_1\ \mathbf{f}_2\ \ldots\ \mathbf{f}_R]^T$, is the channel matrix in the orthonormal basis, the vector $\mathbf{f}_r \in \mathbb{C}^{Q\times 1}$ containing the channel parameters $f_{2k+1}^{(r)}(\bar{m}_1,\ldots,\bar{m}_{2k+1})$, and $\breve{\mathbf{w}}(n)$ is the nonlinear input vector in the orthonormal basis, containing the components $Q_{\bar{m}_1,\cdots,\bar{m}_{2k+1}}^{(2k+1)}(S)$. The noisy version of (4.44) can then be written as:

$$\mathbf{y}(n) = \mathbf{F}\breve{\mathbf{w}}(n) + \mathbf{v}(n). \tag{4.45}$$

Adaptive Estimation

As pointed out earlier, the convergence speed of LMS algorithm when estimating a MIMO Volterra channel can be very slow if the canonical basis is used, due to the high eigenvalue spread of $\mathbf{R}_{\mathbf{w}}$. On the other hand, with an orthonormal basis, the eigenvalue spread of $\mathbf{R}_{\breve{\mathbf{w}}}$ is equal to 1. The LMS adaptation equation for the estimation of the channel matrix in the orthonormal basis $\mathbf{F}$ is given by:

$$\hat{\mathbf{F}}(n+1) = \hat{\mathbf{F}}(n) + \mu\left(\mathbf{y}(n) - \hat{\mathbf{F}}(n)\breve{\mathbf{w}}(n)\right)\breve{\mathbf{w}}^H(n), \tag{4.46}$$

where the matrix $\hat{\mathbf{F}}(n)$ represents the estimate of $\mathbf{F}$ at the n^{th} iteration.

As $\mathbf{R}_{\breve{\mathbf{w}}\breve{\mathbf{w}}} = \mathbf{I}_Q$, it can be deduced from (4.12) that the convergence of (4.46) is assured if $0 < \mu < 1$. Moreover, the EMSE provided by the LMS algorithm using the orthonormal basis can be obtained from (4.16) :

$$\xi_{EMSE}^{(r)}(\infty) \cong \mu\,\xi_{min}^{(r)}Q. \tag{4.47}$$

MMSE Estimation

The orthonormal polynomials can also be used to provide a reduction of the computational complexity associated with the calculus of the MMSE estimation of the channel matrix in a block processing scheme. Indeed, the MMSE estimate of the system coefficient matrix $\mathbf{H}$ in the canonical basis can be calculated as: $\hat{\mathbf{H}} = \hat{\mathbf{R}}_{\mathbf{yw}}\hat{\mathbf{R}}_{\mathbf{w}}^{-1}$, where $\hat{\mathbf{R}}_{\mathbf{yw}}$ and $\hat{\mathbf{R}}_{\mathbf{w}}$ are respectively the sample estimates of $\mathbf{R}_{\mathbf{yw}} = \mathbb{E}\left[\mathbf{y}(n)\mathbf{w}^H(n)\right]$ and $\mathbf{R}_{\mathbf{w}} = \mathbb{E}\left[\mathbf{w}(n)\mathbf{w}^H(n)\right]$. In the orthonormal basis, the MMSE estimate becomes:

$$\hat{\mathbf{F}} = \hat{\mathbf{R}}_{\mathbf{y}\breve{\mathbf{w}}}\hat{\mathbf{R}}_{\breve{\mathbf{w}}}^{-1}, \tag{4.48}$$

where $\hat{\mathbf{R}}_{\mathbf{y}\breve{\mathbf{w}}}$ and $\hat{\mathbf{R}}_{\breve{\mathbf{w}}}$ are respectively the sample estimates of $\mathbf{R}_{\mathbf{y}\breve{\mathbf{w}}} = \mathbb{E}[\mathbf{y}(n)\breve{\mathbf{w}}^H(n)]$ and $\mathbf{R}_{\breve{\mathbf{w}}} = \mathbb{E}[\breve{\mathbf{w}}(n)\breve{\mathbf{w}}^H(n)]$.

A low-complexity estimate of the channel matrix in the orthonormal basis can be obtained by taking into account the fact that $\mathbf{R}_{\breve{\mathbf{w}}} = \mathbf{I}_Q$. From (4.48), $\hat{\mathbf{F}}$ can then be estimated as:

$$\hat{\mathbf{F}} = \hat{\mathbf{R}}_{\mathbf{y}\breve{\mathbf{w}}}. \tag{4.49}$$

Thus, the orthonormal polynomials make possible a decoupled estimation of the elements of $\mathbf{F}$, avoiding the calculus of the inverse of the covariance matrix of the nonlinear input vector. As Volterra system may have a large number of coefficients, this matrix inversion can be a hard computational task. For large values of Q and N, the block estimation in the canonical basis requires approximatively $(RQ+Q^2)N+(Q^3/3)$ multiplications, $(Q^2/2)$ divisions and $(RQ+Q^2)N+(Q^3/3)$ additions, considering the Gauss elimination method for matrix inversion. In the orthonormal basis, it requires only $[RQ + (Q^2/2)]N$ multiplications and $RQN + (Q^2/2)N$ additions.

Information Recovery

Once the channel coefficients in the orthonormal basis are estimated, the estimation of the transmitted signals can be carried out by a technique of maximum likelihood sequence estimation (MLSE) such as the Viterbi algorithm [118]. This means that the knowledge of the channel coefficients in the orthonormal basis is sufficient to recover the transmitted symbols.

Nevertheless, if we are interested in finding the parameters in the canonical basis,

they can be obtained from (4.49):

$$\begin{aligned}\hat{\mathbf{F}} &= \mathbb{E}[\mathbf{y}(n)\breve{\mathbf{w}}^H(n)] \\ &= \mathbb{E}[\hat{\mathbf{H}}\mathbf{w}(n)\breve{\mathbf{w}}^H(n)] + \mathbb{E}[\mathbf{v}(n)\breve{\mathbf{w}}^H(n)],\end{aligned} \tag{4.50}$$

which gives

$$\hat{\mathbf{H}} = \hat{\mathbf{F}}\mathbf{T}, \tag{4.51}$$

with $\mathbf{T} = \mathbb{E}[\mathbf{w}(n)\breve{\mathbf{w}}^H(n)]^{-1}$ being a lower triangular matrix. For uniformly distributed i.i.d. QAM signals, the matrix $\mathbf{T}$ has many zeroes below the main diagonal, which allows to develop simple relationships between the parameters in the two bases.

4.5 Simulation Results

In this section, the proposed estimation method is evaluated by means of simulations. A MIMO Wiener filter corresponding to the model of an uplink channel of a Radio Over Fiber (ROF) multiuser communication system [114, 44] has been considered for the simulations. The $R \times T$ wireless link, corresponding to an array of R half-wavelength spaced antennas and T users, has a frequency selective fading due to multipath propagation and is modeled as a convolutive $R \times T$ linear mixer of memory M. The electrical-optical (E/O) conversion in each antenna is modeled by the following third-order polynomial $f_1 x + f_3|x|^2 x$, with $f_1 = -0.2952$ and $f_3 = 1.078$. In a ROF channel, the received signals are subject to optical and wireless channel noise, however, it is assumed that only the wireless noise is significant. The results were obtained via Monte Carlo simulations using 100 independent data realizations and the amplitude of the signals transmitted by all the users is equal to 1. In order to accentuate the averaging effect, the Normalized Mean Squared Error (NMSE) curves of the adaptive algorithms were passed through a low-pass filter of order equal to 2 and with a cut-off frequency equal to 10^{-2}. For the adaptive algorithms, we adjusted the step-size parameter in such a way that the orthonormal and the canonical bases provide approximately the same steady-state error. The adaptive filters are initialized with zeroes in all the components.

The adaptive channel estimation techniques are initially evaluated by means of

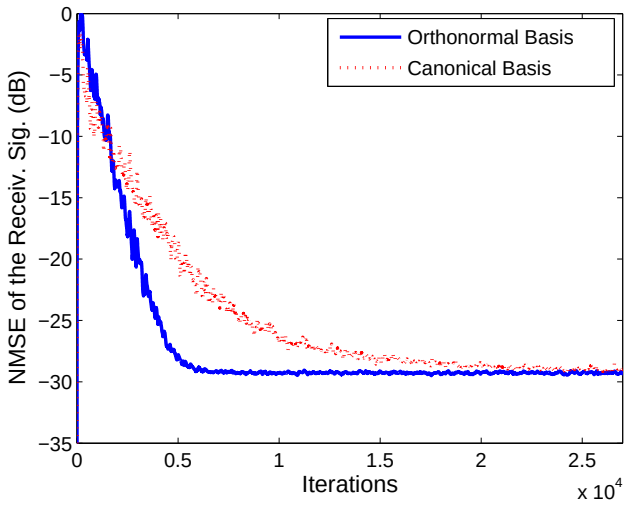

Figure 4.1: NMSE of the received signals using the LMS algorithm - $R = 4$, $T = 4$, $M_t = 2$ (t=1,2,3,4).

the NMSE of the estimated received signals parameters, defined as:

$$e_S(n) = \frac{1}{N_R} \sum_{i=1}^{N_R} \frac{\| \mathbf{y}_i(n) - \hat{\mathbf{y}}_i(n) \|_2^2}{\| \mathbf{y}_i(n) \|_2^2}, \tag{4.52}$$

where $\| \cdot \|_2$ denotes the l^2 norm and $\hat{\mathbf{y}}_i(n)$ represents the estimate of the received signal vector using the channel estimated at time instant n of the i^{th} Monte Carlo simulation, i.e. $\hat{\mathbf{y}}_i(n) = \hat{\mathbf{F}}_i(n)\breve{\mathbf{w}}(n)$ for the orthonormal basis and $\hat{\mathbf{y}}_i(n) = \hat{\mathbf{H}}_i(n)\mathbf{w}(n)$ for the canonical basis, with $\hat{\mathbf{F}}_i(n)$ and $\hat{\mathbf{H}}_i(n)$ denoting respectively the channel matrix in the orthonormal and canonical bases estimated at time instant n of i^{th} Monte Carlo simulation.

Fig. 4.1 shows the evolution of the NMSE of the received signals $e_S(n)$, using the LMS algorithm with the canonical and orthonormal bases, for $R = 4$, $T = 4$, $M_t = 2$, for $t = 1, 2, 3, 4$ ($Q = 296$), and a fixed Signal-to-Noise-Ratio (SNR) of 30dB. The four users transmit uniformly distributed P-QAM signals, with $P = 16, 16, 32$ and 64 symbols respectively. The step-size parameter was set to $4 \cdot 10^{-4}$ for the canonical basis and 10^{-3} for the orthonormal basis. The adaptation in the orthonormal basis converges approximately after 5000 iterations and in the canonical basis after 18000 iterations. The eigenvalue spread of the autocorrelation matrix estimated using the 27000 transmitted symbols is equal to 368.31 in the canonical basis and 2.03 in the orthonormal basis.

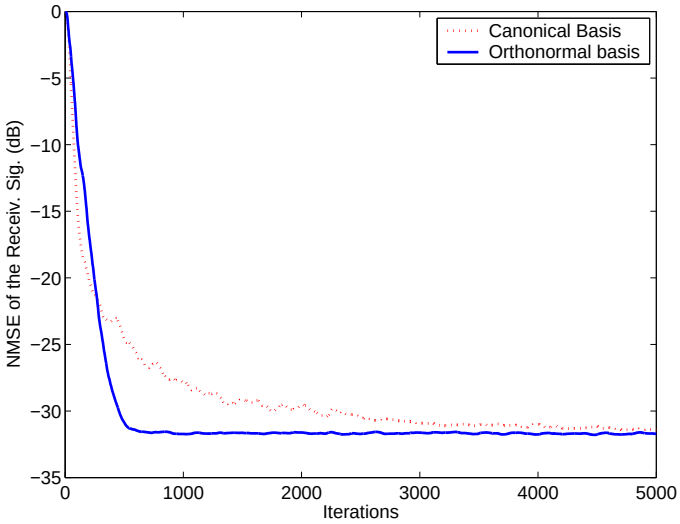

Figure 4.2: NMSE of the received signals using the LMS algorithm - $R = 3$, $T = 3$, $M_t = 1$ (t=1,2,3).

Similar results are found for 3 users and 3 antennae. Fig. 4.2 shows the evolution of the NMSE of the received signals $e_S(n)$, for $R = 3$, $T = 3$, $SNR = 30$dB, M_t=1 for $t = 1, 2, 3$ ($Q = 21$). The four users transmit uniformly distributed 16-QAM signals. In this case, the orthonormal basis provides a more significant gain in performance, the adaptation in the orthonormal and canonical bases converging approximately after 500 and 3000 iterations respectively. The step-size parameter was set to 10^{-2} for the two bases.

Fig. 4.3 shows the evolution of the NMSE of the estimated channel coefficients, defined as:

$$e_P(n) = \frac{1}{N_R} \sum_{i=1}^{N_R} \frac{\| \mathbf{H} - \hat{\mathbf{H}}_i(n) \|_F^2}{\| \mathbf{H} \|_F^2}, \tag{4.53}$$

where $\| \cdot \|_F$ denotes the *Frobenius norm*, using the LMS algorithm for $R = 2$, $T = 2$, $M_1 = M_2 = 1$ ($Q = 8$) and a SNR of 30dB. The two users transmit uniformly distributed 16-QAM signals. The step-size parameter was set to $4 \cdot 10^{-2}$ for the canonical basis and 10^{-2} for the orthonormal basis. Once again, the gain in the rate of convergence provided by the orthonormal basis is evident. The adaptation in the orthonormal basis converges approximately after 550 iterations and in the canonical basis after 2100 iterations. The eigenvalue spread of the autocorrelation matrix estimated using the 2200 transmitted symbols is equal to 64.02 in the canonical basis and 1.17 in the orthonormal basis.

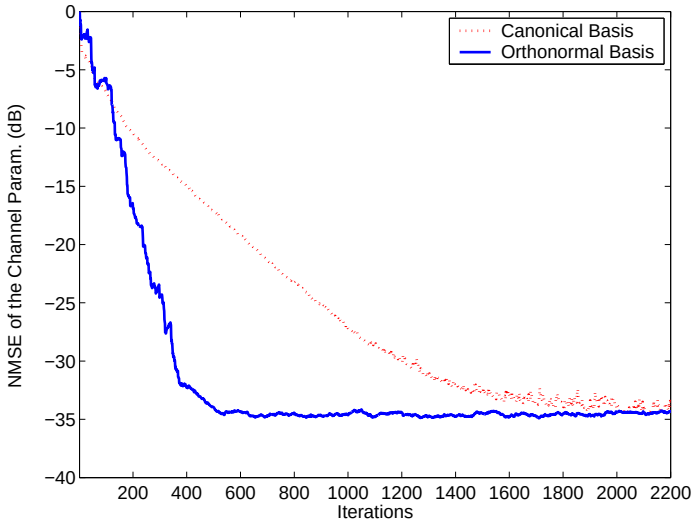

Figure 4.3: NMSE of the channel coefficients using the LMS algorithm - $R = 2$, $T = 2$, $M_1 = M_2 = 1$.

In what concerns the block processing, Fig. 4.4 shows the NMSE of the received signals versus SNR provided by the MMSE estimator in the canonical and orthonormal bases, for $R = 4$, $T = 4$, $M_1 = M_2 = 2$, $M_3 = 1$, $M_4 = 3$ ($Q = 296$) and $N = 5000$. In this figure, the MMSE estimate (4.48) is called "Orthonormal" and the MMSE estimate (4.49) is called "Orthonormal-RC" (Orthonormal with Reduced Complexity). The four users transmit uniformly distributed P-QAM signals, with $P = 16, 16, 32$ and 64 symbols respectively. Note that the NMSEs provided by MMSE estimator in the canonical basis and in the orthonormal basis (4.48) are quite similar. However, the performance of the MMSE estimator (4.49) is worst than the ones of other two techniques. It should be highlighted that the computational complexity of the orthonormal estimator (4.49) is much smaller than the ones of the other two estimators.

In order to have another performance reference, we have also simulated the MIMO Hammerstein system estimation technique proposed in [123], using a different simulation scenario. Although this technique concerns real-valued systems, it can also be applied to complex-valued systems. This technique estimates MIMO Hammerstein coefficients in the canonical basis using a block processing based on polyspectra [123]. The system considered in the next figure is composed of T memoryless nonlinearities given by $x + |x|^2 x$, followed by a complex-valued T by R linear convolutional mixer with a fixed memory M for all the inputs. The real and imaginary parts of the input signals are uniformly generated from the set $[-1, 1]$.

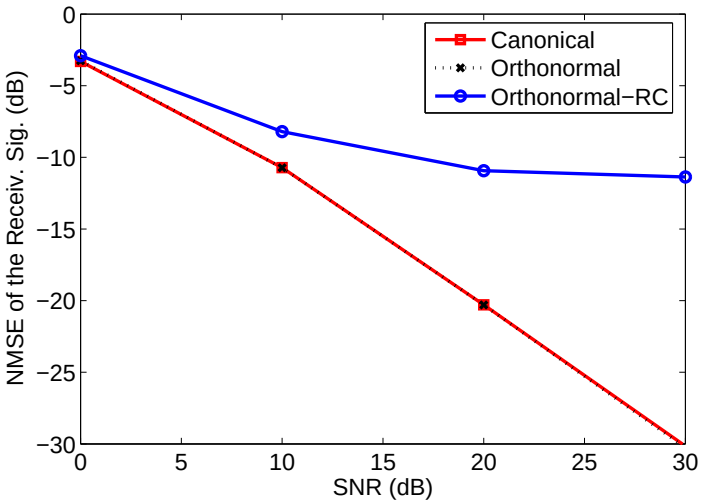

Figure 4.4: NMSE of the received signals versus SNR provided by the MMSE estimator in the orthonormal and canonical bases- $R = 4$, $T = 4$, $N = 5000$, $M_1 = M_2 = 2$, $M_3 = 1$ and $M_4 = 3$.

Note that this signal satisfies the circularity condition (4.18). We have not used a smooth function in the polyspectra estimations. Fig. 4.5 shows the NMSE of the received signals versus SNR provided by the MMSE estimator in the orthonormal basis (4.49) and the polyspectra based technique, for $M = 4$, $N = 2000$ and $T = R = 4, 6, 8, 10$. For a Hammerstein system, the number of quasi-sources is given by $Q = 2TM$. Note that the proposed method performs significantly better than the other one for high values of Q. Moreover, the proposed technique provides a better robustness to noise than the polyspectra based technique.

4.6 Conclusion

In this chapter, a method for supervised estimation of equivalent baseband MIMO Volterra channels is developed. This method is based on the utilization of orthonormal polynomials to improve the conditioning of the covariance matrix of the nonlinear input vector. The proposed orthonormalization technique can be viewed as an extension of orthonormal polynomials to the case of MIMO Volterra systems, allowing different PDFs for the input signals and different memories with respect to the inputs, so that the orthonormal polynomials can be applied to a general equivalent baseband MIMO Volterra channel. The proposed channel estimation

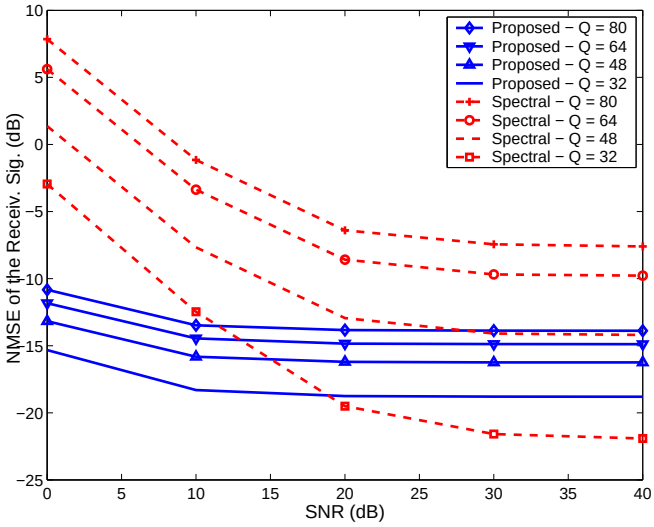

Figure 4.5: NMSE of the received signals versus SNR provided by the MMSE estimator in the orthonormal basis and the polyspectra based technique - $M = 4$, $N = 2000$ and $R = T = 4, 6, 8, 10$ (Q=32,48,64,80).

method has been applied for identifying an uplink channel in a multiuser ROF communication system, the main advantage of the orthonormalization approach being the fact that it improves significantly the convergence speed of the LMS algorithm.

In the next chapter, we also develop techniques for estimating MIMO Volterra communication channels in the context of TDMA-SDMA systems. However, the techniques proposed in Chapter 5 consider a blind scenario and memoryless MIMO Volterra models. As the approach used by the channel estimation method developed in this chapter can not be used in a blind scenario, the MIMO Volterra techniques developed in the next chapter make use of tensor decompositions, exploiting the fact that Volterra models are linear with respect to the system coefficients. As we will see, such decompositions are possible owing to a new precoding scheme developed for PSK signals modeled as Markov chains.

CHAPTER 5

Blind Estimation of Memoryless MIMO Volterra Channels Using Tensor Decomposition and Precoding

This chapter proposes two blind estimation methods for memoryless MIMO Volterra communication channels. As seen in Chapter 2, MIMO Volterra models have important applications in the field of telecommunications, e.g. to model multiuser nonlinear uplink channels in ROF communication systems. In fact, up to several Mbps, the ROF channel can be considered as a memoryless link [54]. This means that assuming that the nonlinearities are memoryless and that the transmitted signals are narrowband with respect to the wireless channel's coherence bandwidth, i.e. the wireless channel frequency response is flat, the ROF channel can be considered as memoryless MIMO Volterra model.

There are few works dealing with the problem of blind channel identification or source separation in the context of multiuser or MIMO nonlinear communication systems. Reference [127] proposes a blind zero forcing receiver for multiuser code division multiple access (CDMA) systems with nonlinear channels and [113] develops blind and semi-blind source separation algorithms for memoryless Volterra channels in ultra-wide-band systems.

The proposed channel identification methods rely on the Parallel Factor (PARAFAC) decomposition [73] of a tensor (multidimensional array) composed of spatio-temporal covariances of the signals received by an antenna array. A great advantage of using the PARAFAC decomposition is that it allows to work when

the number of receive antennas is smaller than the number of virtual sources, i.e. the number of nonlinear terms of the Volterra filter. This is particularly interesting since the Volterra filters may have a large number of parameters. Indeed, working with a number of receive antennas higher than or equal to the number of virtual sources imposes a strong constraint on the number of antennas to be used; see previous works [127, 43, 113].

In telecommunications, the transmitted signals are usually assumed to be white. Thus, if we intend to exploit the temporal correlation of the sources for estimating the channel, some strategy must be used to induce correlation on the transmitted signals. It is shown that the input signals must satisfy some orthogonality constraints associated with the channel nonlinearities in order to get the PARAFAC decomposition of the considered tensor. A precoding scheme is then proposed so that these constraints be satisfied. In this scheme, PSK transmitted signals are modeled as discrete time Markov chains (DTMC) inducing temporal correlation in a controlled way and some orthogonality properties. The proposed precoding scheme induces correlation by introducing temporal redundancy on the signals, which is carried out by imposing some constraints on the symbol transitions. In fact, the proposed transmission scheme can be viewed as a special case of differential encoding. The introduction of redundancy in the transmitted signals is sometimes used with bandwidth-constrained channels, where a performance gain can be achieved without expanding the channel bandwidth or the transmission power [118].

Some properties of nonlinearly distorted PSK signals established in [100] have motivated the use of PSK signals in the present chapter. As PSK signals have less envelope fluctuations than QAM signals, the PSK modulation provides less nonlinear distortions than the QAM. In fact, when the input signals are PSK-modulated, Volterra models can be rewritten with a smaller number of coefficients. The performance of PSK signals over nonlinear satellite channels was investigated by several authors, e.g. [9, 78].

Two algorithms are proposed to perform channel estimation: a two-step version of the Alternating Least Squares (ALS) algorithm [73, 144] and a joint diagonalization algorithm (JDA) [19, 7]. The second estimation method can be viewed as an extension of the second order blind identification (SOBI) algorithm [7] to nonlinear channels. The SOBI algorithm is a blind source separation and identification technique for linear memoryless mixtures based on the joint diagonalization of covariance matrices, and exploiting the temporal correlation of the sources.

Second-order statistics have been used for blind identification and equalization of nonlinear single-input-multiple-output (SIMO) channels [60, 43, 99]. PARAFAC-

based blind channel identification and source separation have also been addressed in the case of linear channels in the context of CDMA systems [144, 142, 35, 143, 33, 53]. In [130], a time-varying user power loading was proposed to enable the application of the PARAFAC analysis, in order to perform blind estimation of spatial signatures. Blind source separation using a PARAFAC tensor composed of covariance matrices was also proposed in [37]. In the case of nonlinear channels, a deterministic blind PARAFAC-based receiver was presented for SIMO channels in [87] and a blind identification method based on the PARAFAC decomposition of a channel output data tensor was recently proposed for Wiener-Hammerstein type channels [86].

The chapter is organized as follows. Section 5.1 presents the channel model used in this chapter. In Section 5.2, a tensor composed of channel output covariances is introduced. In Section 5.3, some orthogonality constraints are established to get a PARAFAC decomposition of this tensor. In Section 5.4, these constraints are rewritten in terms of the transition probability matrix (TPM) of a Markov chain and a procedure to design TPMs satisfying such constraints is described. Section 5.5 presents the proposed blind channel estimation algorithms. In Section 5.6, we evaluate the performance of these algorithms by means of simulation results. Finally, some conclusions and perspectives are drawn in Section 5.7. The results presented in this chapter have been published in [44, 45, 46, 47, 51].

5.1 The Channel Model

In this chapter, the discrete-time equivalent baseband model of the nonlinear communication channel is assumed to be expressed as a memoryless MIMO Volterra model:

$$y_r(n) = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{k+1}=t_k}^{T} \underbrace{\sum_{t_{k+2}=1}^{T} \cdots \sum_{t_{2k+1}=t_{2k}}^{T}}_{\{t_{k+2},\ldots,t_{2k+1}\}\cap\{t_1,\ldots,t_{k+1}\}=\emptyset} h_{2k+1}^{(r)}(t_1,\ldots,t_{2k+1}) \prod_{i=1}^{k+1} s_{t_i}(n) \prod_{i=k+2}^{2k+1} {s_{t_i}}^*(n) + \upsilon_r(n), \qquad (5.1)$$

where $y_r(n)$ $(1 \leq r \leq R)$ is the signal received by antenna r at the time instant n, R is the number of receive antennas, $(2K+1)$ is the nonlinearity order of the model, $s_t(n)$ $(1 \leq t \leq T)$ is the stationary signal transmitted by the t^{th} user at the time instant n, T is the number of users, $h_{2k+1}^{(r)}(t_1,\ldots,t_{2k+1})$ are the coefficients of the

$(2k+1)^{th}$-order Volterra kernel of the r^{th} sub-channel and $\upsilon_r(n)$ is the zero-mean additive white Gaussian noise (AWGN), with variance σ^2 for $r = 1, 2, ..., R$.

It is assumed that $s_t(n)$ and $s_{t'}(n)$, with $t \neq t^{'}$, are independent and that the noise component $\upsilon_r(n)$ is independent from $\upsilon_{r'}(n)$, with $r \neq r^{'}$, and from $s_t(n)$. It should also be highlighted that the transmitted signals $s_t(n)$ are colored, as we will see later. Moreover, they are assumed to be P-PSK modulated, i.e. $s_t(n)$ belongs to the set $\{a_p = A\ e^{\mathbf{j}2\pi(p-1)/P}, p = 1, 2, ..., P\}$, where $\mathbf{j}$ is the imaginary unit, A is the amplitude of the transmitted signals and P is the number of points of the PSK constellation. The parameters A and P are not necessarily the same for all the users; however, for simplifying the notation, the user index t is omitted from these parameters. Besides, it is assumed that

$$P > 2K + 1, \tag{5.2}$$

which corresponds to the well-known persistence of excitation condition for a Volterra system of order $(2K + 1)$ [107].

The nonlinear terms corresponding to $t_i = t_j$, for all $i \in \{1, ..., k + 1\}$ and $j \in \{k + 2, ..., 2k + 1\}$, are absent in (5.1) due to the fact that, for constant modulus signals, the term $|s_{t_i}(n)|^2$ reduces to a multiplicative constant that can be absorbed by the associated channel coefficient. As a consequence, some nonlinear terms degenerate in terms of smaller order.

From (2.59), equation (5.1) can be expressed in a compact form:

$$\mathbf{y}(n) = \mathbf{H}\mathbf{w}(n) + \mathbf{v}(n), \tag{5.3}$$

where $\mathbf{y}(n) = [y_1(n) \dots y_R(n)]^T \in \mathbb{C}^{R\times 1}$, $\mathbf{H} = [\mathbf{h}_1 \dots \mathbf{h}_R]^T \in \mathbb{C}^{R\times Q}$ and $\mathbf{v}(n) = [\upsilon_1(n) \dots \upsilon_R(n)]^T \in \mathbb{C}^{R\times 1}$, with Q representing the number of virtual sources , i.e. the number of linear and nonlinear terms in (5.1). For a linear-cubic channel $(2K + 1 = 3)$, we have $Q = T + \frac{T^2(T-1)}{2}$. In this case, the nonlinear input vector $\mathbf{w}(n) = [w_1(n) \dots w_Q(n)]^T \in \mathbb{C}^{Q\times 1}$ contains all the linear and nonlinear terms in $s_t(n)$ and $s_t^*(n)$ of (5.1), and is constructed as follows:

$$\mathbf{w}(n) = \Theta\, \tilde{\mathbf{w}}(n) \tag{5.4}$$

where

$$\tilde{\mathbf{w}}(n) = [\mathbf{s}^T(n) \quad \otimes_*^3 \mathbf{s}^T(n) \ \cdots \ \otimes_*^{2K+1}\mathbf{s}^T(n)]^T, \tag{5.5}$$

with $\mathbf{s}(n) = [s_1(n) \dots \ s_T(n)]^T \in \mathbb{C}^{T\times 1}$ and the operator $\otimes_*^{2k+1}$ defined as:

$$\otimes_*^{2k+1}\mathbf{s}(n) = \left[\otimes^{k+1}\mathbf{s}(n)\right] \otimes \left[\otimes^{k}\mathbf{s}^*(n)\right], \tag{5.6}$$

$\otimes$ denoting the Kronecker product and $\otimes^k\mathbf{s}(n) = \mathbf{s}(n) \otimes \cdots \otimes \ \mathbf{s}(n)$, with $k-1$ Kronecker products. The matrix Θ is a row-selection matrix that selects all the elements of $\tilde{\mathbf{w}}(n)$ corresponding to $\prod_{i=1}^{k+1} s_{t_i}(n) \prod_{i=k+2}^{2k+1} {s_{t_i}}^*(n)$ with $t_1 \leq \cdots \leq t_{k+1}$, $t_{k+2} \leq \cdots \leq t_{2k+1}$ and $\{t_{k+2}, \dots, t_{2k+1}\} \cap \{t_1, \dots, t_{k+1}\} = \emptyset$, for $k = 1, 2, ..., K$.

5.2 PARAFAC Decomposition of a Channel Output Covariance Tensor

The proposed identification methods rely on the PARAFAC decomposition of a tensor composed of spatio-temporal covariances of the received signals. Assuming that these signals are stationary and ergodic, we have:

$$\mathbf{R}_{\mathbf{y}}(d) = \mathbb{E}\left[\mathbf{y}(n+d)\mathbf{y}^H(n)\right] = \mathbf{H}\mathbf{R}_{\mathbf{w}}(d)\mathbf{H}^H + \sigma^2\mathbf{I}_R\delta(d) \ \in \mathbb{C}^{R\times R}, \tag{5.7}$$

with

$$\mathbf{R}_{\mathbf{w}}(d) = \mathbb{E}\left[\mathbf{w}(n+d)\mathbf{w}^H(n)\right] \ \in \mathbb{C}^{Q\times Q}, \tag{5.8}$$

where $0 \leq d \leq D-1$, D is the number of delays (time lags) taken into account, $\delta(\cdot)$ is the Kronecker symbol and $\mathbf{I}_R$ is the identity matrix of order R. In the sequel, it is assumed that the noise variance σ^2 is known, allowing the subtraction of the noise term in (5.7). Then, from now on, the noise term will be omitted. However, in practice, this noise variance has to be estimated [7, 27] or the proposed identification methods can be applied without using the zero-lag covariance matrix ($d = 0$).

A third-order tensor $\mathcal{R} \in \mathbb{C}^{D\times R\times R}$ can be defined from the matrices $\mathbf{R}_{\mathbf{y}}(d)$, with:

$$r_{d,r_1,r_2} = \mathbb{E}[y_{r_1}(n+d-1)y_{r_2}^*(n)], \tag{5.9}$$

as entries, for $1 \leq d \leq D$ and $1 \leq r_1, r_2 \leq R$. From (5.7), we get:

$$r_{d,r_1,r_2} = \sum_{q_1=1}^{Q}\sum_{q_2=1}^{Q} h_{r_1,q_1}h_{r_2,q_2}^*\tilde{r}_{d,q_1,q_2}, \tag{5.10}$$

where $h_{r,q} = [\mathbf{H}]_{r,q}$ and $\tilde{r}_{d,q_1,q_2} = \mathbb{E}[w_{q_1}(n+d-1)w_{q_2}^*(n)]$, $w_q(n)$ $(q = 1, ..., Q)$ being

the q^{th} component of the nonlinear input vector $\mathbf{w}(n)$. Note that equation (5.10) corresponds to the scalar writing of a Tucker2 model [155].

If the covariance matrices $\mathbf{R}_{\mathbf{w}}(d-1)$ of the nonlinear input vector are diagonal for $1 \leq d \leq D$, the scalar writing (5.10) of $\mathcal{R}$ becomes:

$$r_{d,r_1,r_2} = \sum_{q=1}^{Q} z_{d,q} h_{r_1,q} h^*_{r_2,q}, \tag{5.11}$$

which corresponds to the PARAFAC decomposition of $\mathcal{R}$ (see Appendix C) with factor matrices equal to $\mathbf{Z}$, $\mathbf{H}$ and $\mathbf{H}^*$, the matrix $\mathbf{Z} \in \mathbb{C}^{D \times Q}$ being formed with the diagonal elements of $\mathbf{R}_{\mathbf{w}}(d-1)$ for $1 \leq d \leq D$, i.e.

$$\mathbf{Z} = \begin{bmatrix} \tilde{r}_{1,1,1} & \cdots & \tilde{r}_{1,Q,Q} \\ \vdots & \ddots & \vdots \\ \tilde{r}_{D,1,1} & \cdots & \tilde{r}_{D,Q,Q} \end{bmatrix}, \tag{5.12}$$

or $z_{d,q} = \tilde{r}_{d,q,q}$. The main advantages of the PARAFAC model over the Tucker2 model are its simplicity and the essential uniqueness of its factors, assured if the Kruskal's condition is verified (see Appendix C):

$$2k_{\mathbf{H}} + k_{\mathbf{Z}} \geq 2Q + 2, \tag{5.13}$$

where $k_{\mathbf{A}}$ is the k-rank of matrix $\mathbf{A}$, i.e. the greatest integer $k_{\mathbf{A}}$ such that every set of $k_{\mathbf{A}}$ columns of $\mathbf{A}$ is linearly independent. The essential uniqueness property means that the matrices $\mathbf{H}$, $\mathbf{H}^*$ and $\mathbf{Z}$ are unique up to column scaling and permutation ambiguities, i.e. any matrices $\hat{\mathbf{H}}_a$, $\hat{\mathbf{H}}_b$ and $\hat{\mathbf{Z}}$ satisfying (5.11) are linked to $\mathbf{H}$, $\mathbf{H}^*$ and $\mathbf{Z}$ by: $\hat{\mathbf{H}}_a = \mathbf{H}\Pi\Lambda_a$, $\hat{\mathbf{H}}_b = \mathbf{H}^*\Pi\Lambda_b$ and $\hat{\mathbf{Z}} = \mathbf{Z}\Pi\Lambda_c$, where Λ_a, Λ_b and Λ_c are $Q \times Q$ diagonal matrices such that $\Lambda_a\Lambda_b\Lambda_c = \mathbf{I}_Q$ and Π is a $Q \times Q$ permutation matrix.

When $\mathbf{Z}$ is known, we have $\hat{\mathbf{Z}} = \mathbf{Z}$ and, hence, $\Pi = \mathbf{I}_Q$, $\Lambda_c = \mathbf{I}_Q$ and $\Lambda_b = \Lambda_a^{-1} = \Lambda^{-1}$. Thus, $\hat{\mathbf{H}}_a = \mathbf{H}\Lambda$ and $\hat{\mathbf{H}}_b = \mathbf{H}^*\Lambda^{-1}$. This means that the permutation ambiguity is eliminated. Moreover, the scaling ambiguity does not represent an effective problem, as it can be canceled by using a differential modulation [118]. Another possible solution consists in using a few pilot signals to estimate the scaling ambiguity matrix Λ.

Assuming that the matrices $\mathbf{H}$ and $\mathbf{Z}$ are full k-rank, i.e. $k_{\mathbf{H}} = \min(R, Q)$ and $k_{\mathbf{Z}} = \min(D, Q)$, the Kruskal's condition becomes: $2\min(R, Q) + \min(D, Q) \geq 2Q + 2$, which implies that the tensor approach allows working even if $R < Q$, contrarily to previous works that require $R \geq Q$ [127, 43, 113]. This is particularly interesting

for identifying Volterra systems characterized by a large number of parameters.

In the next section, we establish some conditions for ensuring that the covariance matrices $\mathbf{R}_{\mathbf{w}}(d)$ be diagonal for $0 \leq d \leq D-1$, in order to get a PARAFAC decomposition of the tensor $\mathcal{R}$.

5.3 Orthogonality Conditions

The following theorem states sufficient conditions to ensure that the covariance matrices of the nonlinear input vector $\mathbf{R}_{\mathbf{w}}(d)$ $(0 \leq d \leq D-1)$ be diagonal when the transmitted signals are PSK modulated.

Theorem 5.1 *Assuming that the transmitted signals $s_t(n)$ $(1 \leq t \leq T)$ are stationary and PSK modulated, with cardinality $P > 2K+1$, the covariance matrices $\boldsymbol{R}_{\boldsymbol{w}}(d)$ for $0 \leq d \leq D-1$, are diagonal if the following orthogonality conditions are satisfied for $(T-1)$ users:*

(i) $\mu_t^{(i,j)}(d) = 0$, *for all* $0 \leq i, j \leq K+1$ *with* $i \neq j$;

(ii) $\varrho_t^{(i,j)}(d) = 0$, *for all* $1 \leq i \leq K+1$, $1 \leq j \leq K$;

where

$$\mu_t^{(i,j)}(d) \equiv \mathbb{E}\left[s_t^i(n+d)\left[s_t^j(n)\right]^*\right] \tag{5.14}$$

and

$$\varrho_t^{(i,j)}(d) \equiv \mathbb{E}\left[s_t^i(n+d)s_t^j(n)\right]. \tag{5.15}$$

Proof:

The elements of $\mathbf{R}_{\mathbf{w}}(d-1)$ $(1 \leq d \leq D)$ are defined as:

$$\tilde{r}_{d,q_1,q_2} = \mathbb{E}[w_{q_1}(n+d-1)w_{q_2}^*(n)], \tag{5.16}$$

where $w_{q_1}(n)$ and $w_{q_2}(n)$ can be written respectively as:

$$w_{q_1}(n) = \prod_{t=1}^{T} s_t^{\alpha_t}(n) \left[s_t^{\beta_t}(n) \right]^*, \tag{5.17}$$

$$w_{q_2}(n) = \prod_{t=1}^{T} s_t^{\alpha_t'}(n) \left[s_t^{\beta_t'}(n) \right]^*, \tag{5.18}$$

for some non negative integers $\alpha_t, \beta_t, \alpha_t', \beta_t'$ satisfying:

$$\begin{aligned} &\sum_{t=1}^{T} \alpha_t = k+1, \qquad \sum_{t=1}^{T} \beta_t = k, \\ &\sum_{t=1}^{T} \alpha_t' = k'+1 \text{ and } \sum_{t=1}^{T} \beta_t' = k'. \end{aligned} \tag{5.19}$$

Note that, due to the circularity property of PSK signals, we have $s_t^P(n) = 1$ and, consequently, $0 \leq \alpha_t, \alpha_t' \leq \min(K+1, P-1)$ and $0 \leq \beta_t, \beta_t' \leq \min(K, P-1)$. However, from the persistence of excitation condition (6.27), we have $\min(K+1, P-1) = K+1$ and $\min(K, P-1) = K$. Moreover, from the constraints $t_{k+2}, ..., t_{2k+1} \neq t_1, ..., t_{k+1}$ in (5.1), it can be deduced that α_t or β_t (or both) equals zero, as well as α_t' or β_t' (or both), for all $t = 1, ..., T$. Hence, (5.17) and (5.18) can be rewritten respectively as:

$$w_{q_1}(n) = \prod_{t=1}^{T} \dot{s}_t^{\gamma_t}(n) \quad \text{and} \quad w_{q_2}(n) = \prod_{t=1}^{T} \ddot{s}_t^{\gamma_t'}(n), \tag{5.20}$$

where $\gamma_t = \max(\alpha_t, \beta_t)$, $\gamma_t' = \max(\alpha_t', \beta_t')$,

$$\dot{s}_t(n) = \begin{cases} s_t(n), & \text{if } \beta_t = 0, \gamma_t = \alpha_t, \\ s_t^*(n), & \text{if } \alpha_t = 0, \gamma_t = \beta_t, \end{cases} \qquad \ddot{s}_t(n) = \begin{cases} s_t(n), & \text{if } \beta_t' = 0, \gamma_t' = \alpha_t', \\ s_t^*(n), & \text{if } \alpha_t' = 0, \gamma_t' = \beta_t'. \end{cases} \tag{5.21}$$

Substituting (5.20) into (5.16), we get:

$$\tilde{r}_{d,q_1,q_2} = \prod_{t=1}^{T} \mathbb{E}\left[\dot{s}_t^{\gamma_t}(n+d-1) \left[\ddot{s}_t^{\gamma_t'}(n) \right]^* \right]. \tag{5.22}$$

If $q_1 \neq q_2$, there is at least two users t_1 and t_2 such that $(\alpha_{t_1}, \beta_{t_1}) \neq (\alpha'_{t_1}, \beta'_{t_1})$ and $(\alpha_{t_2}, \beta_{t_2}) \neq (\alpha'_{t_2}, \beta'_{t_2})$. Thus, (5.22) can be rewritten as:

$$\tilde{r}_{d,q_1,q_2} = \prod_{\substack{t=1 \\ t\neq t_1,\, t_2}}^{T} \mathbb{E}\left[\dot{s}_t^{\gamma_t}(n+d-1)\left[\ddot{s}_t^{\gamma'_t}(n)\right]^*\right] \mathbb{E}\left[\dot{s}_{t_1}^{\gamma_{t_1}}(n+d-1)\left[\ddot{s}_{t_1}^{\gamma'_{t_1}}(n)\right]^*\right] \mathbb{E}\left[\dot{s}_{t_2}^{\gamma_{t_2}}(n+d-1)\left[\ddot{s}_{t_2}^{\gamma'_{t_2}}(n)\right]^*\right]. \quad (5.23)$$

Depending on the different possible configurations of the couples $(\dot{s}_{t_1}(n), \ddot{s}_{t_1}(n))$ and $(\dot{s}_{t_2}(n), \ddot{s}_{t_2}(n))$, the two last factors of (5.23) can be expressed in terms of the following quantities:

- $\mu_t^{(i,j)}(d-1)$, with $0 \leq i, j \leq K+1$ and $i \neq j$;
- $\varrho_t^{(i,j)}(d-1)$, with $1 \leq i \leq K+1$, $1 \leq j \leq K$;

for $d = 1, ..., D$. Thus, for each couple of users, if at least one user satisfies conditions (i) and (ii) of Theorem 5.1, we have $\tilde{r}_{d,q_1,q_2} = 0$. Therefore, if at least $(T-1)$ users satisfy these conditions, the covariance matrices $\mathbf{R}_{\mathbf{w}}(d-1)$ $(1 \leq d \leq D)$ are diagonal. ■

To illustrate Theorem 5.1, let us consider the covariance matrix $\mathbf{R}_{\mathbf{w}}(d)$ for 2 users $(t = 1, 2)$ and $K = 1$, given by:

$$\mathbf{R}_{\mathbf{w}}(d) = \begin{pmatrix} \mu_1^{(1,1)}(d) & \mu_1^{(1,0)}(d)\mu_2^{(0,1)}(d) & \mu_1^{(1,2)}(d)\mu_2^{(1,0)}(d) & \varrho_1^{(1,1)}(d)\mu_2^{(0,2)}(d) \\ \mu_1^{(0,1)}(d)\mu_2^{(1,0)}(d) & \mu_2^{(1,1)}(d) & \mu_1^{(0,2)}(d)\varrho_2^{(1,1)}(d) & \mu_1^{(1,0)}(d)\mu_2^{(1,2)}(d) \\ \mu_1^{(2,1)}(d)\mu_2^{(0,1)}(d) & \mu_1^{(2,0)}(d)\left[\varrho_2^{(1,1)}(d)\right]^* & \mu_1^{(2,2)}(d)\left[\mu_2^{(1,1)}(d)\right]^* & \varrho_1^{(2,1)}(d)\left[\varrho_2^{(1,2)}(d)\right]^* \\ \left[\varrho_1^{(1,1)}(d)\right]^*\mu_2^{(2,0)}(d) & \mu_1^{(0,1)}(d)\mu_2^{(2,1)}(d) & \left[\varrho_1^{(1,2)}(d)\right]^*\varrho_2^{(2,1)}(d) & \left[\mu_1^{(1,1)}(d)\right]^*\mu_2^{(2,2)}(d) \end{pmatrix}$$

and associated with the following nonlinear input vector:

$$\mathbf{w}(n) = [s_1(n) \quad s_2(n) \quad s_1^2(n)s_2^*(n) \quad s_1^*(n)s_2^2(n)]. \quad (5.24)$$

Note that all the off-diagonal components of $\mathbf{R}_{\mathbf{w}}(d)$ are the product of two terms like (5.14) and (5.15), with $t = 1$ or 2. Then, if the conditions (i) and (ii) hold for at least one user, the matrix $\mathbf{R}_{\mathbf{w}}(d)$ is diagonal.

5.4 Transmitted Signal Design

In this section, a precoding scheme is proposed so that the transmitted signals satisfy the orthogonality constraints of Theorem 5.1. Each transmitted signal is modeled as a discrete time Markov chain (DTMC), the states of the DTMC being given by the P PSK symbols $a_p = A\, e^{\mathrm{j}2\pi(p-1)/P}, p = 1, 2, ..., P$. The coding induces time correlation by introducing redundancy on the signals, which is done by imposing some constraints on the transition probability matrix (TPM) associated with the DTMC. The correlation is introduced in a controlled way so that the constraints of Theorem 5.1 are satisfied, the TPM playing a key role in this scheme.

Let us denote by L_B the number of input bits of the encoder, assumed to be independent and identically distributed (i.i.d.) and uniformly distributed over the set $\{0, 1\}$. Moreover, we assume that $L = 2^{L_B} < P$, which imposes some restrictions on the symbol transitions. This means that, for each state, there are L equiprobable possible transitions and $(P - L)$ not assigned transitions. The code rate is therefore equal to $L_B / \log_2 P$.

Let us denote by $\mathbf{T} = \{T_{p_1,p_2}\}$, with p_1, $p_2 \in \{1, 2, ..., P\}$, the TPM for a given user, T_{p_1,p_2} being the probability of transition from the state a_{p_1} to the state a_{p_2}. Each user is associated with a different TPM. However, for simplifying the notation, the user index t will be omitted from $\mathbf{T}$. Note that $\sum_{p_2=1}^{P} T_{p_1,p_2} = 1$, for $1 \leq p_1 \leq P$. Hence, each row of the TPM has $L = 2^{L_B}$ elements equal to $1/L = 2^{-L_B}$ and $(P - L)$ elements equal to zero. For instance, Fig. 5.1 shows the state transition diagram of a DTMC corresponding to $P = 4$ and $L_B = 1$, with the following TPM:

$$\mathbf{T} = \frac{1}{2} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}. \tag{5.25}$$

5.4.1 Orthogonality Constraints in Terms of the TPM

The orthogonality constraints of Theorem 5.1 are now rewritten in terms of the TPM $\mathbf{T}$ of the DTMC associated with each user. Some important properties of DTMC are first recalled [71]. In what follows, T_{n,p_1,p_2} denotes the (p_1, p_2) element of $\mathbf{T}^n = \mathbf{T}\mathbf{T}\ldots\mathbf{T}$ (n times the matrix $\mathbf{T}$), which represents the probability of being in the state a_{p_2} after n transitions, supposing that the current state is a_{p_1}.

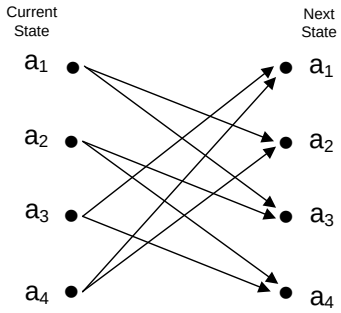

Figure 5.1: Example of state transition diagram for $P = 4$ and $L_B = 1$.

Definition 5.1 *A state a_p of a DTMC is said to be aperiodic if the great common divisor of the set of integers n such that $T_{n,p_1,p_2} > 0$ is equal to 1. If all the states are aperiodic, then the DTMC is also aperiodic.*

Definition 5.2 *A state a_{p_2} of a DTMC is said to be* accessible *from the state a_{p_1} if there exists some integer n such that $T_{n,p_1,p_2} > 0$.*

Definition 5.3 *A DTMC is said to be* irreducible *if all the states are accessible from each other.*

Definition 5.4 *The limiting probability θ_{p_2} $(1 \leq p_2 \leq P)$ of a given state of a DTMC is defined as:*

$$\theta_{p_2} = \lim_{n \to \infty} T_{n,p_1,p_2}, \ \forall\, p_1 \in \{1, 2, ..., P\}. \tag{5.26}$$

Definition 5.5 *A probability distribution $\Phi = [\phi_1 \ \phi_2 \ ... \ \phi_P]$, $0 \leq \phi_1, ..., \phi_P \leq 1$, is stationary if the following conditions are satisfied:*

$$\Phi\, \mathbf{T} = \Phi, \tag{5.27}$$

$$\sum_{p=1}^{P} \phi_p = 1. \tag{5.28}$$

Definition 5.6 *An irreducible and aperiodic DTMC is said to be stationary if the initial state is chosen according to the stationary distribution.*

In what follows, we reformulate the conditions of Theorem 5.1 in terms of the TPM of a given user. For that, we first establish in Lemma 5.1 the conditions to be satisfied by the TPM for generating a stationary and uniformly distributed signal. Then, Theorem 5.2 expresses the quantities $\mu_t^{(i,j)}(d)$ and $\varrho_t^{(i,j)}(d)$ in terms of the TPM. In what follows, the user index t will be omitted from these quantities to simplify the notation.

Lemma 5.1 *Let $\boldsymbol{T}$ be the TPM of a DTMC with P states. If the following conditions hold:*

(C1) *the DTMC is irreducible and aperiodic;*

(C2) $\sum_{p_2=1}^{P} T_{p_1,p_2} = 1$, *for* $1 \leq p_1 \leq P$;

(C3) $\sum_{p_1=1}^{P} T_{p_1,p_2} = 1$, *for* $1 \leq p_2 \leq P$;

then the corresponding signal is stationary and uniformly distributed.

Proof:

As already mentioned, Condition C2 must be satisfied by any TPM. Moreover, the aperiodicity and irreducibility properties (condition C1) assure that [71]: (i) all the limiting probabilities of the DTMC exist and are positive, (ii) the stationary distribution exists and is unique, and (iii) the limiting probabilities distribution is equal to the stationary distribution. Thus, the limiting probabilities can be obtained by finding the stationary distribution, i.e. by solving equations (5.27) and (5.28). It can be easily verified that if $\sum_{p_1=1}^{P} T_{p_1,p_2} = 1$ (condition C3), then

$$\Phi = [1/P \ ... \ 1/P] \tag{5.29}$$

is solution of (5.27) and (5.28), which shows that the limiting probabilities correspond to an uniform distribution. Hence, if the initial state is equiprobably drawn from the set of PSK symbols $\{a_1, ..., a_P\}$, the DTMC is stationary with an uniform distribution. ■

Theorem 5.2 *If conditions C1-C3 are satisfied, then the quantities (5.14) and (5.15) can be rewritten as:*

$$\mu^{(i,j)}(d) = \frac{1}{P} \left[\mathbf{a}^{\bullet j}\right]^H \mathbf{T}^d \mathbf{a}^{\bullet i} \tag{5.30}$$

and

$$\varrho^{(i,j)}(d) = \frac{1}{P}\left[\mathbf{a}^{\bullet j}\right]^T \mathbf{T}^d \mathbf{a}^{\bullet i} \tag{5.31}$$

where $\boldsymbol{a} = [a_1,\, a_2,\, \ldots, a_P]^T$ *and* $\boldsymbol{a}^{\bullet i} = [a_1^i,\, a_2^i,\, \ldots, a_P^i]^T$.

Proof:

From Lemma 5.1, conditions C1-C3 ensure that the DTMC is stationary with an uniform distribution. Thus, we get

$$\mu^{(i,j)}(d) = \mathbb{E}\left[s^i(n+d)\left[s^j(n)\right]^*\right] = \sum_{p_1=1}^{P}\sum_{p_2=1}^{P} \mathsf{p}(\alpha_n = a_{p_1})\left[a_{p_1}^j\right]^* \mathsf{p}(\alpha_{n+d} = a_{p_2}|\alpha_n = a_{p_1}) a_{p_2}^i, \tag{5.32}$$

where $\mathsf{p}(\alpha_n = a_{p_1})$ is the probability of being in the state a_{p_1} at the time instant n and $\mathsf{p}(\alpha_{n+d} = a_{p_2}|\alpha_n = a_{p_1})$ is the conditional probability of being in the state a_{p_2} at the time instant $(n+d)$, given the state a_{p_1} at the time instant n. Then, we have:

$$\begin{aligned}\mu^{(i,j)}(d) &= \sum_{p_1=1}^{P}\sum_{p_2=1}^{P} \frac{1}{P}\left[a_{p_1}^j\right]^* T_{d,p_1,p_2} a_{p_2}^i \\ &= \frac{1}{P}\left[\mathbf{a}^{\bullet^j}\right]^H \mathbf{T}^d \mathbf{a}^{\bullet i}.\end{aligned} \tag{5.33}$$

Expression (5.31) can be derived in a similar way. ■

Remark: Note that, when i or $j = 0$, Condition (i) of Theorem 5.1 becomes:

$$\mathbb{E}\left[s^i(n)\right] = 0, \tag{5.34}$$

for $1 \leq i \leq K+1$. On the other hand, for $d = 0$, Condition (i) becomes:

$$\mu^{(i,j)}(0) = \mathbb{E}\left[s^i(n)\left[s^j(n)\right]^*\right] = \begin{cases} A^{2j}\ \mathbb{E}\left[s^{i-j}(n)\right], & \text{if} \quad i > j, \\ A^{2i}\ \mathbb{E}\left[s^{j-i}(n)\right]^*, & \text{if} \quad i < j, \end{cases} \tag{5.35}$$

with $1 \leq |i-j| \leq K+1$, which is equivalent to (5.34). Since for a stationary and uniformly distributed P-PSK signal, we have:

$$\mathbb{E}\left[s^i(n)\right] = \frac{A^i}{P}\sum_{p=1}^{P} e^{\mathbf{j}2\pi(p-1)i/P} = \frac{A^i\left(e^{\mathbf{j}2\pi i}-1\right)}{P\left(e^{\mathbf{j}2\pi i/P}-1\right)} = 0, \tag{5.36}$$

for $1 \leq i \leq K+1 < P$, we can conclude that Condition (i) of Theorem 5.1 is satisfied for $d = 0$, and i or $j = 0$.

In summary, combining Lemma 5.1, Theorem 5.2 and the above remark, the conditions of Theorem 5.1 can be reformulated as follows:

(C1) the DTMC is irreducible and aperiodic;

(C2) $\sum_{p_2=1}^{P} T_{p_1,p_2} = 1$, for $1 \leq p_1 \leq P$;

(C3) $\sum_{p_1=1}^{P} T_{p_1,p_2} = 1$, for $1 \leq p_2 \leq P$;

(C4) $\left[\mathbf{a}^{\bullet j}\right]^H \mathbf{T}^d \mathbf{a}^{\bullet i} = 0$, for all i and j such that $1 \leq i, j \leq K+1$ with $i \neq j$;

(C5) $\left[\mathbf{a}^{\bullet j}\right]^T \mathbf{T}^d \mathbf{a}^{\bullet i} = 0$, for all i and j such that $1 \leq i \leq K+1$, $1 \leq j \leq K$;

for $1 \leq d \leq D-1$ and at least $(T-1)$ users.

5.4.2 Determination of the Transition Probability Matrices

For a given user, Conditions C2 and C3 can be written as the following set of linear equations:

$$\begin{bmatrix} \Theta_3 \\ \Theta_4 \end{bmatrix} \mathrm{vec}(\mathbf{T}) = \mathbf{1}_{2P-1}, \tag{5.37}$$

where $\mathrm{vec}(\cdot)$ is the vectorization operator that stacks the columns of its matrix argument, $\mathbf{1}_{2P-1} \in \mathbb{R}^{(2P-1)\times 1}$ is the all ones vector of dimension $2P-1$ and the matrices $\Theta_3 \in \mathbb{R}^{P\times P^2}$ and $\Theta_4 \in \mathbb{R}^{(P-1)\times P^2}$ are respectively given by:

$$\Theta_3 = \mathbf{1}_P^T \otimes \mathbf{I}_P \tag{5.38}$$

and

$$\Theta_4 = \left[\mathbf{I}_{P-1}\ \mathbf{0}_{P-1}\right]\left(\mathbf{I}_P \otimes \mathbf{1}_P^T\right), \tag{5.39}$$

where $\mathbf{0}_{P-1}$ is the zero vector of dimension $P-1$. Note that each row of the left side of (5.37) corresponds to the sum of the elements of one row or one column of $\mathbf{T}$. The sum of the elements of the last column of $\mathbf{T}$ is not included as it represents a redundant constraint.

Moreover, for all (i,j) such that $1 \leq i,j \leq K+1$, conditions C4 and C5 can be written in a matrix form respectively as:

$$\mathbf{A}^H \mathbf{T}^d \mathbf{A} = \mathbf{0}_{(K+1)\times(K+1)} \text{ and } \mathbf{A}^T \mathbf{T}^d \mathbf{A} = \mathbf{0}_{(K+1)\times(K+1)} \tag{5.40}$$

where

$$\mathbf{A} = \left[\mathbf{a} \ \mathbf{a}^{\bullet 2} \cdots \mathbf{a}^{\bullet K+1}\right] \ \in \mathbb{C}^{P\times(K+1)}. \tag{5.41}$$

Applying the vec$(\cdot)$ operator to the two members of equations (5.40) and using the following property: $\text{vec}\left(\mathbf{ABZ}\right) = \left(\mathbf{Z}^T \otimes \mathbf{A}\right)\text{vec}\left(\mathbf{B}\right)$, we get:

$$\left(\mathbf{A}^T \otimes \mathbf{A}^H\right)\text{vec}\left(\mathbf{T}^d\right) = \mathbf{0}_{(K+1)^2} \tag{5.42}$$

and

$$\left(\mathbf{A}^T \otimes \mathbf{A}^T\right)\text{vec}\left(\mathbf{T}^d\right) = \mathbf{0}_{(K+1)^2}. \tag{5.43}$$

By restricting the values of i and j as indicated in conditions C4 and C5, equations (5.42) and (5.43) become:

$$\left[\begin{array}{c} \Theta_1\left(\mathbf{A}^T \otimes \mathbf{A}^H\right) \\ \Theta_2\left(\mathbf{A}^T \otimes \mathbf{A}^T\right) \end{array}\right] \text{vec}(\mathbf{T}^d) = \left[\begin{array}{c} \mathbf{0}_{2(K^2+K)} \end{array}\right], \tag{5.44}$$

where $\Theta_1 \in \mathbb{R}^{(K^2+K)\times(K+1)^2}$ is a row selection matrix that eliminates the rows of $\left(\mathbf{A}^T \otimes \mathbf{A}^H\right)$ corresponding to $\left(\mathbf{a}_i^T \otimes \mathbf{a}_i^H\right)$, for $i = 1,2,...,K+1$, and $\Theta_2 \in \mathbb{R}^{(K^2+K)\times(K+1)^2}$ is a row selection matrix that eliminates the rows of $\left(\mathbf{A}^T \otimes \mathbf{A}^T\right)$ corresponding to $\left(\mathbf{a}_i^T \otimes \mathbf{a}_{K+1}^T\right)$, for $i = 1,2,...,K+1$.

Thus, the TPMs must satisfy (5.37),(5.44) and Condition C1. It should be highlighted that, once chosen the values of K, P and L_B, these constraints only depend on the matrix $\mathbf{T}$, which means that $\mathbf{T}$ can be a priori designed. By exploiting the fact that $T_{p_1,p_2} \in \{0, 1/L\}$, the next theorem proposes a procedure to determine TPMs that verify (5.37) and (5.44) for any values of K, P and L_B.

Definition 5.7 *The p^{th} circulant diagonal ($p = 1,...,P$) of a $P \times P$ matrix is the set of entries corresponding to the following indices: $(k, mod(p+k-2,P)+1)$, for $k = 1,...,P$, where $mod(\cdot,P)$ denotes the modulo operation, i.e. the remainder of the division of the argument by P.*

Definition 5.8 *Let us define* $\boldsymbol{T}_P(p_1, ..., p_L)$ *as the* $P \times P$ *matrix having entries equal to* $1/L$ *on the circulant diagonals* $p_1, ..., p_L$ *and to zero elsewhere, with* $L = 2^{L_B} < P$.

For instance, for $P = 4$ and $L_B = 1$ ($L = 2$), the TPM (5.25) is denoted by $\mathbf{T}_4(2, 3)$.

Theorem 5.3 *The matrices* $\boldsymbol{T}_P(p_1, ..., p_L)$ *satisfy (5.37) and (5.44) for all* $1 \leq p_1 < p_2 < ... < p_L \leq P$.

Proof:

Each row and column of $\mathbf{T}_P(p_1, ..., p_L)$ contains L elements equal to $1/L$ and $(P - L)$ elements equal to zero. Hence, Conditions C2 and C3, i.e. equation (5.37), are always satisfied. In the sequel, it is proved that $\mathbf{T}_P(p_1, ..., p_L)$ also satisfies Condition C4 for all $d \geq 1$.

For $d \geq 1$, defining $\mathbf{q} = \mathbf{T}\mathbf{a}^{\bullet i} \in \mathbb{C}^{P \times 1}$, Condition C4 can be rewritten as:

$$\left[\mathbf{a}^{\bullet j}\right]^H \mathbf{T}^d \mathbf{a}^{\bullet i} = \left[\mathbf{a}^{\bullet j}\right]^H \mathbf{T}^{d-1} \mathbf{q}, \tag{5.45}$$

The first element of the vector $\mathbf{q}$ can be developed as:

$$q_1 = \sum_{p=1}^{P} T_{1,p} a_p^i = \sum_{l=1}^{L} T_{1,p_l} a_{p_l}^i = \frac{1}{L} \sum_{l=1}^{L} a_{p_l}^i. \tag{5.46}$$

By using Definition 5.8, the k^{th} element ($k = 2, ..., P$) of $\mathbf{q}$ can be expressed as:

$$q_k = \sum_{p=1}^{P} T_{k,p} a_p^i = \sum_{l=1}^{L} T_{k,[\mathrm{mod}(p_l+k-2,P)+1]} a_{[\mathrm{mod}(p_l+k-2,P)+1]}^i. \tag{5.47}$$

For PSK modulated symbols, we have:

$$\begin{aligned} a_{[\mathrm{mod}(p_l+k-2,P)+1]}^i &= A^i \, e^{\mathbf{j}2\pi[\mathrm{mod}(p_l+k-2,P)]i/P} = A^i \, e^{\mathbf{j}2\pi(p_l+k-2)i/P} \\ &= \frac{a_{p_l}^i a_k^i}{A^i}. \end{aligned} \tag{5.48}$$

Substituting (6.16) into (6.17) gives:

$$q_k = \frac{1}{L \, A^i} \sum_{l=1}^{L} a_{p_l}^i a_k^i = \frac{a_k^i q_1}{A^i}. \tag{5.49}$$

Thus, the vector $\mathbf{q}$ can be written as:

$$\mathbf{q} = \mathbf{T}\mathbf{a}^{\bullet i} = \frac{q_1}{A^i}\mathbf{a}^{\bullet i}. \tag{5.50}$$

By substituting (6.55) into (5.45), we get the following recursive equation:

$$\left[\mathbf{a}^{\bullet j}\right]^H \mathbf{T}^d \mathbf{a}^{\bullet i} = \frac{q_1}{A^i}\left[\mathbf{a}^{\bullet j}\right]^H \mathbf{T}^{d-1}\mathbf{a}^{\bullet i}, \tag{5.51}$$

which leads to:

$$\begin{aligned}\left[\mathbf{a}^{\bullet j}\right]^H \mathbf{T}^d \mathbf{a}^{\bullet i} &= \left(\frac{q_1}{A^i}\right)^d \left[\mathbf{a}^{\bullet j}\right]^H \mathbf{a}^{\bullet i} \\ &= \left(\frac{q_1}{A^i}\right)^d A^{i+j} \sum_{p=1}^{P} e^{\mathbf{j}2\pi(p-1)(i-j)/P} \\ &= \left(\frac{q_1}{A^i}\right)^d A^{i+j} \frac{e^{\mathbf{j}2\pi(i-j)} - 1}{e^{\mathbf{j}2\pi(i-j)/P} - 1},\end{aligned} \tag{5.52}$$

which is equal to zero for $i \neq j$. That proves that the matrices $\mathbf{T}_P(p_1, ..., p_L)$ satisfy Condition C4. A similar proof can be made for Condition C5.

■

Theorem 5.3 provides a set of TPMs satisfying (5.37) and (5.44) for all K, P and L_B. However, due to the complexity of (5.37) and (5.44), it is very difficult to find an analytical expression for the TPMs. Note that (5.44) corresponds to a system of nonlinear equations with respect to T_{p_1,p_2} $(1 \leq p_1, p_2 \leq P)$. Moreover, in order to find an analytical solution, one has to take into account that $T_{i,j} \in \{0, 1/L\}$. Thus, it may exist other TPMs satisfying these equations than the matrices $\mathbf{T}_P(p_1, ..., p_L)$. For linear-cubic Volterra systems ($K = 1$) and 4-PSK input signals, due to the reduced dimension of the TPM, it is easy to verify by an exhaustive search that the matrices given in Theorem 5.3 are the only TPMs satisfying (5.37) and (5.44). Moreover, it can be verified that the matrices $\mathbf{T}_4(1, 3)$ and $\mathbf{T}_4(2, 4)$ correspond respectively to a reducible and a periodical DTMC. Thus, for 4-PSK signals, the matrices $\mathbf{T}_4(1, 2)$, $\mathbf{T}_4(2, 3)$, $\mathbf{T}_4(3, 4)$ and $\mathbf{T}_4(1, 4)$ are the only matrices satisfying the orthogonality conditions C1-C5.

5.4.3 Interpretation of the TPM

An interesting characteristic of the matrix $\mathbf{T}_P(p_1, ..., p_L)$ is that the corresponding precoding can be viewed as a differential coding. For a given row of $\mathbf{T}_P(p_1, ..., p_L)$,

Table 5.1: Bit mapping for the TPM $\mathbf{T}_4(2,3)$.

Current State \ Next State	a_1	a_2	a_3	a_4
a_1		$B_n = \{0\}$	$B_n = \{1\}$	
a_2			$B_n = \{0\}$	$B_n = \{1\}$
a_3	$B_n = \{1\}$			$B_n = \{0\}$
a_4	$B_n = \{0\}$	$B_n = \{1\}$		$B_n = \{0\}$

each non-zero element is associated with one of the L combinations of the L_B input bits of the encoder. From Definition 5.8, the difference between the row and the column indices of an element of the p^{th} circulant diagonal ($1 \leq p \leq P$) is given by $(\mathrm{mod}(p-2, P)+1) = p-1$, which means that all the P elements of the p^{th} circulant diagonal correspond to the same phase shift $2\pi(p-1)/P$. Thus, if we associate all the P elements of the p^{th} circulant diagonal to the same combination of the L_B input bits, this combination will be associated with the same phase shift, regardless of the input state. The symbols may then be decoded using only the difference of phase of two consecutive symbols, which is the principle of a differential coding. This characteristic simplifies the decoding process and makes it insensitive to scaling ambiguities.

The difference between the proposed coding and the conventional differential coding is that, in the proposed approach, there are some phase shifts that are not allowed. The allowed phase shifts are determined by the circulant diagonals of the TPM, the circulant diagonal p corresponding to a phase shift of $(2\pi(p-1)/P)$. For instance, let us consider the TPM $\mathbf{T}_4(2,3)$, given in (5.25), and corresponding to the state transition diagram shown in Fig 5.1. If the bit mapping defined in Table 5.1 is used, the symbols may then be decoded from the phase shift of two consecutive symbols: if this phase shift is equal to $\pi/2$ (resp. π), the input bit of the encoder is equal to 0 (resp. 1).

The choice of the circulant diagonals determines then an important characteristic of the coding: the distance between the possible phase shifts. With respect to this characteristic, it is desirable to choose the circulant diagonals so that the distance between the allowed phase shifts $2\pi(p-1)/P$ be high, due to the fact that close phase shifts are more difficult to recover in presence of noise and interference. For instance, for $P = 8$ and $L_B = 1$, it is easy to verify that the matrices $\mathbf{T}_8(i, i+3)$ ($1 \leq i \leq 8$) provide the maximal euclidean distance between the allowed phase shifts. Note that the matrices $\mathbf{T}_8(i, i+4)$ ($1 \leq i \leq 8$) correspond to reducible DTMCs. Moreover, for $P = 8$, $L_B = 2$ and considering only irreducible and aperiodic DTMCs, we found by an exhaustive search for all the values of p_1, p_2, p_3, p_4 such that $1 \leq p_1 < p_2 < p_3 < p_4 \leq 8$, that the TPMs maximizing the euclidean

distance between the allowed phase shifts are $\mathbf{T}_8(i, i+2, i+4, i+5)$ $(1 \leq i \leq 8)$.

5.5 Channel Estimation Algorithms

When conditions C1-C5 hold, i.e. when the tensor $\mathcal{R}$ admits the PARAFAC decomposition (5.11), the following expressions for the first and third-mode slices of $\mathcal{R}$ can be deduced from Appendix C:

$$\mathbf{R}_{d\cdot\cdot} = \mathbf{H}\text{diag}_d[\mathbf{Z}]\mathbf{H}^H \quad \text{and} \quad \mathbf{R}_{\cdot\cdot r} = \mathbf{Z}\,\text{diag}_r[\mathbf{H}^*]\mathbf{H}^T, \tag{5.53}$$

where $1 \leq d \leq D$, $1 \leq r \leq R$, $\text{diag}_i[\mathbf{A}]$ is the diagonal matrix formed from the i^{th} row of $\mathbf{A}$ and $\mathbf{R}_{d\cdot\cdot}$ (resp. $\mathbf{R}_{\cdot\cdot r}$) is the first- (resp. third-) mode matrix slice of $\mathcal{R}$, obtained by fixing the first (resp. third) index of $\mathcal{R}$ and varying the indices associated with the two other modes.

Let us denote respectively by $\mathbf{R_1} \in \mathbb{C}^{RD\times R}$ and $\mathbf{R_3} \in \mathbb{C}^{RD\times R}$ the first and third-mode unfolded matrices of the tensor $\mathcal{R}$, defined as:

$$\mathbf{R_1} \equiv \begin{bmatrix} \mathbf{R}_{1\cdot\cdot} \\ \vdots \\ \mathbf{R}_{D\cdot\cdot} \end{bmatrix}, \qquad \mathbf{R_3} \equiv \begin{bmatrix} \mathbf{R}_{\cdot\cdot 1} \\ \vdots \\ \mathbf{R}_{\cdot\cdot R} \end{bmatrix}. \tag{5.54}$$

These matrices are given by:

$$\mathbf{R_1} = (\mathbf{Z} \diamond \mathbf{H})\,\mathbf{H}^H \quad \text{and} \quad \mathbf{R_3} = (\mathbf{H}^* \diamond \mathbf{Z})\,\mathbf{H}^T, \tag{5.55}$$

where $\diamond$ denotes the Khatri-Rao (column-wise Kronecker) product.

In the sequel, two estimation methods are proposed for estimating the channel based on (5.55): a two-steps Alternate Least Squares (ALS) algorithm [44] and a joint diagonalization algorithm [45].

5.5.1 Alternating Least Squares (ALS) algorithm

The first proposed channel estimation method uses the ALS algorithm [73, 144], the principle of which is to estimate, in the least square sense, a subset of the parameters by using a previous estimation of other subsets of parameters. In fact, a two-steps version of the ALS algorithm is used due to the fact that the matrix

$\mathbf{Z}$ is assumed to be known, as it can be precomputed using the formula:

$$\mu^{(i,i)}(d) = \frac{1}{P}\left[\mathbf{a}^{\bullet^i}\right]^H \mathbf{T}^d \mathbf{a}^{\bullet i}, \tag{5.56}$$

for $i = 0, ..., K+1$ and $d = 1, ..., D$. The channel estimation problem is solved by minimizing the two following conditional least squares cost functions in an alternate way:

$$J_1 = \left\| \hat{\mathbf{R}}_{\mathbf{1}} - \left(\mathbf{Z} \diamond \hat{\mathbf{H}}_a^{(it-1)}\right) \hat{\mathbf{H}}_b^T \right\|_F^2, \quad J_2 = \left\| \hat{\mathbf{R}}_{\mathbf{3}} - \left(\hat{\mathbf{H}}_b^{(it)} \diamond \mathbf{Z}\right) \hat{\mathbf{H}}_a^T \right\|_F^2, \tag{5.57}$$

where $\hat{\mathbf{R}}_{\mathbf{1}}$ and $\hat{\mathbf{R}}_{\mathbf{3}}$ are respectively the sample estimates of the unfolded matrices $\mathbf{R}_{\mathbf{1}}$ and $\mathbf{R}_{\mathbf{3}}$, it and $\|\cdot\|_F$ denote respectively the iteration number and the Frobenius norm. The covariances of the received signals are estimated in the following way:

$$\hat{r}_{d,r_1,r_2} = \frac{1}{N}\sum_{n=1}^{N} y_{r_1}(n+d-1) y_{r_2}^*(n), \tag{5.58}$$

where N is the number of data symbols.

The ALS algorithm is summarized in Table 5.2, where $(\cdot)^\dagger$ denotes the matrix pseudo-inverse, ϵ is an arbitrary small positive constant and $\hat{\mathbf{H}}_{ab}^{(it)} = 0.5 \cdot [\hat{\mathbf{H}}_a^{(it)} + (\hat{\mathbf{H}}_b^{(it)})^*]$. At each iteration, two LS channel estimates, denoted by $\hat{\mathbf{H}}_a^{(it)}$ and $\hat{\mathbf{H}}_b^{(it)}$, corresponding respectively to estimates of $\mathbf{H}$ and $\mathbf{H}^*$, are calculated. This process continues until the convergence of the parameters is achieved. After convergence, three channel estimates can then be obtained: $\hat{\mathbf{H}}_a^{(it)}$, $(\hat{\mathbf{H}}_b^{(it)})^*$ and $\hat{\mathbf{H}}_{ab}^{(it)}$, the final channel estimate being chosen as the one that provides the smallest value of the cost function (5.57).

One of the drawbacks of the ALS algorithm is that it may exhibit a slow convergence speed if the number of factors is large relative to the tensor dimensions [18, 122]. The ALS algorithm also works if the matrix $\mathbf{Z}$ is unknown. In this case, three least squares estimates are calculated at each iteration.

Note that the ALS algorithm does not take the fact that $\hat{\mathbf{H}}_a$ is the complex conjugate of $\hat{\mathbf{H}}_b$ into account. In fact, we have tested by means of computer simulations a modified version of the ALS algorithm taking the constraint $\hat{\mathbf{H}}_b = \hat{\mathbf{H}}_a^*$ into account. However, this algorithm presents some convergence problems and it does not allow to improve the performance.

Table 5.2: ALS algorithm

Initialization: $\hat{\mathbf{H}}_a^{(0)} \rightarrow R \times Q$ Gaussian random matrix
Iterations $(it = it + 1)$ **:**   1) $\hat{\mathbf{H}}_b^{(it)} = \left[\left(\mathbf{Z} \diamond \hat{\mathbf{H}}_a^{(it-1)}\right)^{\dagger} \hat{\mathbf{R}}_{\mathbf{1}}\right]^T$ 2) $\hat{\mathbf{H}}_a^{(it)} = \left[\left(\hat{\mathbf{H}}_b^{(it)} \diamond \mathbf{Z}\right)^{\dagger} \hat{\mathbf{R}}_{\mathbf{3}}\right]^T$
Stop Criteria: $\frac{\left\|\hat{\mathbf{H}}_{ab}^{(it)} - \hat{\mathbf{H}}_{ab}^{(it-1)}\right\|_F^2}{\left\|\hat{\mathbf{H}}_{ab}^{(it-1)}\right\|_F^2} < \epsilon$

5.5.2 Joint Diagonalization Algorithm (JDA)

The channel matrix $\mathbf{H}$ can also be estimated from the set of covariance matrices $\mathbf{R}_{\mathbf{y}}(d)$ by using a joint diagonalization algorithm (JDA). The estimation algorithm is summarized in Table 5.3 (for further details, see [7]). In the simulations of the next section, the step 3 of this method is carried out by using the joint diagonalization algorithm of [19]. The resulting identification method can then be viewed as an extension of the SOBI algorithm [7] to nonlinear channels. Note that, unlike the ALS algorithm, the joint diagonalization estimator requires $R \geq Q$, i.e. it does not work in the underdetermined case. Besides, the JDA does not assume the knowledge of the source covariance matrix $\mathbf{R}_{\mathbf{w}}(d)$.

The uniqueness of the joint diagonalizer based estimator is given by the following theorem, due to [7]. Without loss of generality, this theorem assumes that the covariation matrix $\mathbf{Z}(0)$ is normalized, i.e. $\mathbf{Z}(0) = \mathbf{I}_Q$.

Theorem 5.4 *Let $\mathcal{B} = \{\boldsymbol{B}_1, ..., \boldsymbol{B}_D\}$ be a set of D matrices $Q \times Q$ such that $\boldsymbol{B}_d = \boldsymbol{M}\boldsymbol{A}_d\boldsymbol{M}^H$, for $d = 1, ..., D$, where $\boldsymbol{M} \in \mathbb{C}^{Q \times Q}$ is a unitary matrix and $\boldsymbol{A}_d \in \mathbb{C}^{Q \times Q}$, for $d = 1, ..., D$, are diagonal matrices, the elements of which are*

Table 5.3: JDA algorithm

1) Calculate the whitening matrix $\mathbf{U}$ as:

$$\mathbf{U} = \left[\lambda_1^{-\frac{1}{2}}\,\mathbf{u}_1 \cdots \lambda_Q^{-\frac{1}{2}}\,\mathbf{u}_Q\right]^H, \tag{5.59}$$

where $\{\lambda_q\}_{q=1}^{Q}$ are the Q largest eigenvalues of $\hat{\mathbf{R}}_{\mathbf{y}}(0)$ and $\{\mathbf{u}_q\}_{q=1}^{Q}$ are the corresponding eigenvectors, $\hat{\mathbf{R}}_{\mathbf{y}}(0)$ being the sample estimate of $\mathbf{R}_{\mathbf{y}}(0)$. It is considered that the estimated noise variance $\hat{\sigma}^2$ was subtracted from $\hat{\mathbf{R}}_{\mathbf{y}}(0)$, as mentioned earlier.

2) Calculate the following set of prewhitened matrices: $\hat{\mathbf{R}}_{\mathbf{p}}(d) = \mathbf{U}\hat{\mathbf{R}}_{\mathbf{y}}(d)\mathbf{U}^H$, for $0 \leq d \leq D-1$, where $\hat{\mathbf{R}}_{\mathbf{y}}(d)$ is the sample estimate of $\mathbf{R}_{\mathbf{y}}(d)$.

3) Determine an unitary matrix $\hat{\mathbf{M}}$ as the joint diagonalizer of the matrices $\hat{\mathbf{R}}_{\mathbf{p}}(d)$, for $0 \leq d \leq D-1$.

4) Estimate the channel matrix as $\hat{\mathbf{H}} = \mathbf{U}^{\dagger}\hat{\mathbf{M}}$.

denoted by $a_d(q) = [\boldsymbol{A}_d]_{q,q}$. *If*

$$\forall (q_1, q_2) \text{ such that } 1 \leq q_1 \neq q_2 \leq Q, \quad \exists\, d,\ 1 \leq d \leq D, \text{ such that } a_d(q_1) \neq a_d(q_2), \tag{5.60}$$

then any joint diagonalizer of $\mathcal{B}$ *is equal to* $\Pi\Lambda\boldsymbol{M}$, *where* Λ *is a diagonal matrix and* Π *a permutation matrix.*

Note that, if the channel matrix is full column rank, the identifiability condition (5.13) of the ALS algorithm becomes $k_{\mathbf{Z}} \geq 2$, which is equivalent to condition (5.60). Thus, for full column rank channel matrices, the identifiability conditions of the ALS and JDA algorithms become equivalent.

5.6 Simulation Results

In this section, the proposed channel estimation methods are evaluated by means of simulations. A memoryless linear-cubic MIMO Volterra system corresponding to a

MIMO Wiener of an uplink channel of a radio over fiber multiuser communication system [114, 44] has been considered for the simulations. The wireless link is modeled as a Rayleigh $R \times T$ linear channel, with an array of R half-wavelength spaced antennas and $T = 2$ or 3 users. The electrical-optical (E/O) conversion in each antenna is modeled by the following polynomial $f_1 x + f_3 |x|^2 x$, with $f_1 = 1$ and $f_3 = -0.35$, as in [114, 116]. In a ROF channel, the received signals are subject to optical and wireless channel noise, however, it is assumed that only the wireless noise is significant. The results were obtained with 8-PSK input signals ($P = 8$), via Monte Carlo simulations using at least 100 independent data realizations. The amplitude of the signals transmitted by all the users is equal to 1.

The proposed channel estimation methods are evaluated by means of the Normalized Mean Squared Error (NMSE) of the estimated channel parameters, defined as:

$$NMSE = \frac{1}{N_R} \sum_{l=1}^{N_R} \frac{\| \mathbf{H} - \hat{\mathbf{H}}_l \|_F^2}{\| \mathbf{H} \|_F^2}, \tag{5.61}$$

where $\hat{\mathbf{H}}_l$ represents the channel matrix estimated at the l^{th} Monte Carlo simulation after eliminating the ambiguities. As a performance reference for the proposed channel estimation techniques, we also show the NMSE obtained with the Wiener solution, given by:

$$\hat{\mathbf{H}} = \hat{\mathbf{R}}_{\mathbf{yw}} \hat{\mathbf{R}}_{\mathbf{ww}}^{-1}, \tag{5.62}$$

where $\hat{\mathbf{R}}_{\mathbf{yw}}$ and $\hat{\mathbf{R}}_{\mathbf{ww}}$ are the sample estimates of $\mathbf{R}_{\mathbf{yw}} = \mathbb{E}\left[\mathbf{y}(n)\mathbf{w}^H(n)\right]$ and $\mathbf{R}_{\mathbf{ww}} = \mathbb{E}\left[\mathbf{w}(n)\mathbf{w}^H(n)\right]$, respectively, and $\mathbf{w}(n)$ is the nonlinear input vector defined in (5.4). This non-blind solution needs to known the input signals.

Table 5.4 describes the various tested simulation configurations, the matrices $\mathbf{T}_8(p_1, ..., p_L)$ being constructed as in Definition 5.8. All the configurations of Table 5.4 provide matrices $\mathbf{Z}$ such that $k_{\mathbf{Z}} = \min(D, Q)$. Remark that Configurations A, B, E and F correspond to a code rate of 1/3 while Configurations C and D lead to a code rate of 2/3. In our simulations, we have remarked that, in general, the best channel estimates are provided by the TPMs that induce high a correlation on the transmitted signals, the correlation being calculated using the following formula:

$$\mu^{(1,1)}(d) = \frac{1}{P} \mathbf{a}^H \mathbf{T}_P^d(p_1, ..., p_L) \mathbf{a}. \tag{5.63}$$

Indeed, if the induced correlation is low, the transmitted and received signals are

Table 5.4: Simulation Configurations

Config.	T	Q	L_B	TPM of user 1	TPM of user 2	TPM of user 3
A	2	4	1	$\mathbf{T}_8(1,4)$	$\mathbf{T}_8(2,5)$	-
B	2	4	1	$\mathbf{T}_8(1,2)$	$\mathbf{T}_8(2,3)$	-
C	2	4	2	$\mathbf{T}_8(1,3,5,6)$	$\mathbf{T}_8(2,4,6,7)$	-
D	2	4	2	$\mathbf{T}_8(1,4,5,6)$	$\mathbf{T}_8(2,5,6,7)$	-
E	3	12	1	$\mathbf{T}_8(1,4)$	$\mathbf{T}_8(2,5)$	$\mathbf{T}_8(3,6)$
F	3	12	1	$\mathbf{T}_8(1,2)$	$\mathbf{T}_8(2,3)$	$\mathbf{T}_8(3,4)$

"almost blind", which means that a small value of D should be used due to an inaccurate estimation of the correlations $\hat{r}_{d,r_1,r_2}$. Thus, the circulant diagonals of the TPMs of Configurations B, D and F were chosen so that the correlation of the transmitted signals be maximized. By doing an exhaustive search for all the values of p_1, p_2 such that $1 \leq p_1 < p_2 \leq 8$, it was found that, for $P = 8$ and $L_B = 1$, the matrices $\mathbf{T}_8(i, i+1)$ $(1 \leq i \leq 8)$ provide the maximal time correlation, i.e. this choice of circulant diagonals maximizes $\sum_{d=0}^{D-1} |\mu^{(1,1)}(d)|^2$, for $D = 4$. Similarly, for $P = 8$, $L_B = 2$ and $D = 4$, the TPMs that maximize the time correlation are $\mathbf{T}_8(i, i+3, i+4, i+5)$ $(1 \leq i \leq 8)$.

5.6.1 Simulations with a code rate of 1/3

The next three figures compare the performance of the two proposed estimation algorithms using Configurations A and B of Table 5.4, i.e. for $T = 2$ users and a code rate of 1/3 ($L_B = 1$). Fig. 5.2 shows the NMSE versus signal-to-noise-ratio (SNR) provided by the ALS and JDA algorithms and by the Wiener solution, for $R = 5$, $D = 4$ and data blocks of $N = 1024$ symbols. It is also shown the NMSE obtained with the ALS algorithm in the case of Configuration B and an unknown noise variance (ALS-UNV), using covariances with delays $d = 1, 2, ..., 4$. The following conclusions can be drawn from Fig. 5.2:

- Configuration B provides better performance than Configuration A, for both ALS and JDA algorithms. As pointed out earlier, this is probably due to the fact that Configuration B is the one that induces the highest correlation to the transmitted signals.

- The performance of JDA is better than that of ALS, except when Configuration B is used and the SNR is lower than 15dB.
- The NMSE provided by the ALS-UNV algorithm is approximatively 3dB higher than the one obtained with the ALS algorithm.

Fig. 5.3 evaluates the performance of the proposed channel identification methods in terms of bit-error-rate (BER). It shows the BER versus SNR provided by the Minimum Mean Square Error (MMSE) receiver:

$$\hat{\mathbf{W}}_{MMSE} = \mathbf{R}_{\mathbf{ww}} \hat{\mathbf{H}}^H \left[\hat{\mathbf{H}} \mathbf{R}_{\mathbf{ww}} \hat{\mathbf{H}}^H + \sigma^2 \mathbf{I}_R \right]^{-1} \in \mathbb{C}^{Q \times R}, \tag{5.64}$$

using the ALS and JDA channel estimates, with Configurations A and B, $R = 5$, $D = 4$ and $N = 1024$. For comparison, it is also plotted the BER provided by the MMSE receiver assuming an exact knowledge of the channel, using Configuration A and Differential Binary PSK (DBPSK) input signals. The following conclusions can be drawn from this figure:

- When JDA is used, Configuration A provides a lower BER than Configuration B. As pointed out earlier, this is certainly due to the distance of the allowed phase shifts of these configurations.
- Using the ALS algorithm, Configuration A provides a lower BER than Configuration B when the SNR is smaller than 15dB. This is due to the poor channel estimation performed by the ALS algorithm when Configuration A is used and to the fact that the multiuser interference is the main source of degradation when the SNR is high.
- Moreover, for a BER of 10^{-2} and considering the case of a known channel, the SNR gap between Configuration A and DBPSK modulation is equal to 1.9dB. This result indicates the SNR lose provided by the proposed coding with respect to DBPSK signals, regardless of the channel estimation.

The advantage of ALS over JDA is illustrated by evaluating the influence of the antenna number R for a small value of SNR. Fig. 5.4 shows the NMSE versus R provided by JDA and ALS using Configurations A and B, for $D = 4$, $N = 1024$ and $SNR = 0$dB. The following remarks can be highlighted from this figure:

- The ALS algorithm allows to work with $R < Q = 4$, the JDA requiring at least $Q = 4$ antennas.

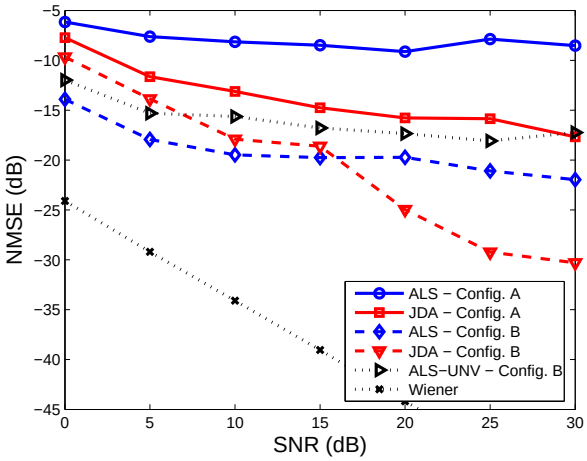

Figure 5.2: NMSE versus SNR provided by the JDA, ALS, ALS-UNV and Wiener solution for Configurations A and B.

- The ALS algorithm provides a good channel estimation even when the Kruskal's condition is not satisfied, i.e. with $R = 2$ and $D = 4$.
- Using Configuration B, ALS performs better than JDA for $SNR = 0$dB.

Fig. 5.5 considers the case of $T = 3$ users ($Q = 12$) and rate 1/3 codes ($L_B = 1$), using Configurations E and F of Table 5.4, for $R = 12$, $D = 8$ and $N = 1024$. It shows the NMSE versus SNR provided by the JDA and ALS. It is also shown the NMSE obtained the ALS algorithm with $R = 8$. The conclusions that we can draw from this figure are similar to those of Fig. 5.2:

- Configuration F provides better NMSE performances than Configuration E for both ALS and JDA algorithms.
- The performance of JDA is better than that of ALS.
- The ALS algorithm with Configuration F and $R = 8$ ($R < Q$) provides a NMSE approximatively 1.8dB higher than the one obtained with $R = 12$.

Moreover, we have carried out some simulations concerning the BER corresponding to Fig. 5.5, the results being similar to those of Fig. 5.3. This is expected as Configurations A and E correspond to similar TPMs, as well as Configurations B and F.

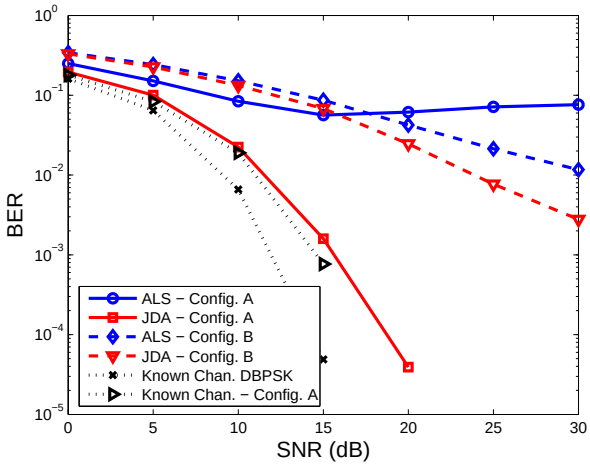

Figure 5.3: BER versus SNR provided by the MMSE receiver using the JDA and ALS channel estimates, and the exact channel, for Configurations A and B.

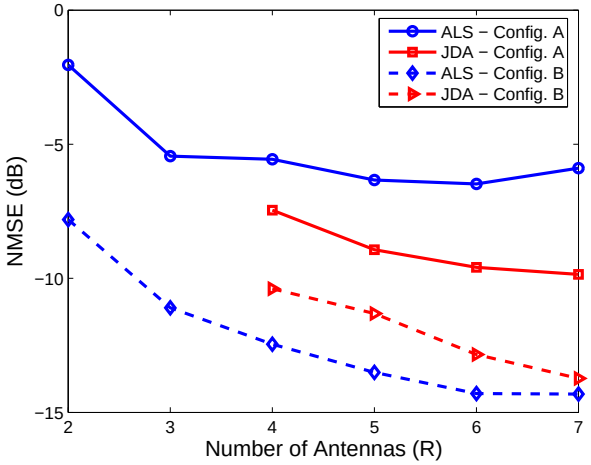

Figure 5.4: NMSE versus R provided by ALS and JDA for SNR=0dB.

5.6.2 Simulations with a code rate of 2/3

The two next figures show the performance of the proposed estimation algorithms using Configurations C and D of Table 5.4, i.e. for $T = 2$ users and a code rate of $2/3$ ($L_B = 2$). Fig. 5.6 and 5.7 plot respectively the NMSE and BER versus SNR, for $R = 5$, $D = 4$ and $N = 1024$. The following conclusions can be drawn from

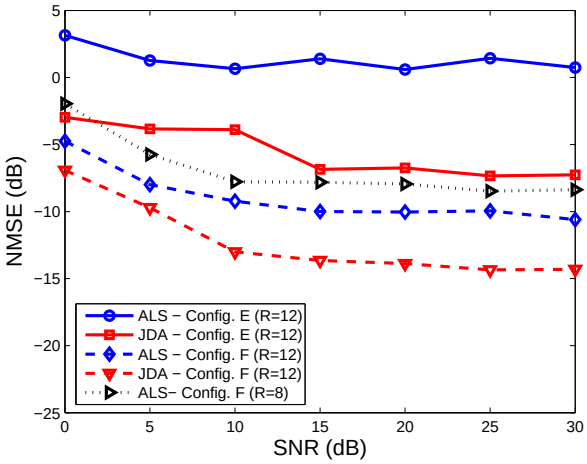

Figure 5.5: NMSE versus SNR provided by the JDA and ALS for Configurations E and F.

Fig. 5.6 :

- Configuration D provides lower NMSE than Configuration C, for both ALS and JDA algorithms. This is certainly due to higher correlation induced by Configuration D.
- The performance of JDA is always better than that of ALS.

From Fig. 5.7, we can conclude that:

- Configuration D provides lower BER than Configuration C. Although the euclidean distance between the possible transitions of Configuration C is higher than the one of Configuration D, this difference is not very significant. The sum of all the euclidean distances between the possible transitions is equal to 30.8 for Configuration C and to 28.0 for Configuration D. In this case, the better channel estimate provided by Configuration D becomes more relevant than the distance between the possible transitions.
- The MMSE receiver calculated with the JDA channel estimate gives better performance than the one calculated with the ALS channel estimate.

The simulation results presented in Sections 5.6.1 and 5.6.2 allow to put in evidence some interesting characteristics of the proposed tensor-based identification

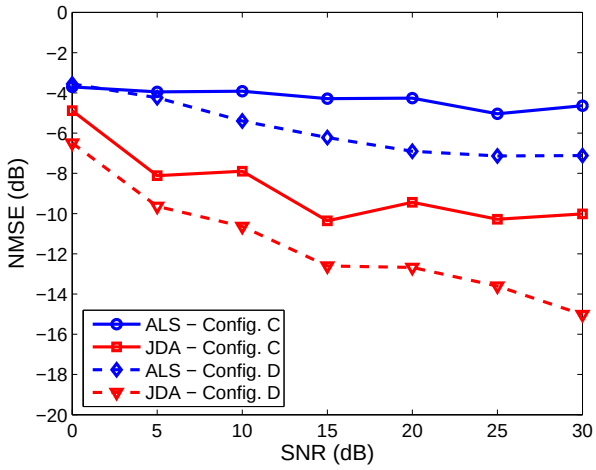

Figure 5.6: NMSE versus SNR provided by the JDA and ALS algorithms for Configurations C and D.

methods. Configurations B, D and F provide better channel estimates, as they induce a high correlation to the transmitted signals. On the other hand, Configurations A, C and E, corresponding to greater distances between the phase shifts, are more robust to channel noise and interference.

5.6.3 Semi-blind ALS algorithm

The objective of the next two figures is to illustrate the performance of a "semi-blind ALS" (SB-ALS) algorithm, i.e. the ALS algorithm initialized by means of the Wiener solution (5.62) calculated using 8 known symbols. In this case, we consider a small data block composed of $N = 256$ symbols. Fig. 5.8 shows the NMSE versus SNR provided by the ALS (blind) and SB-ALS algorithms for Configuration B, with $R = 5$ and $D = 4$. From this figure, we draw the following conclusions:

- For high SNRs, a NMSE performance improvement of about 2dB is obtained when the SB-ALS algorithm is used instead of the ALS algorithm.
- The SB-ALS algorithm needs a smaller number of iterations than the ALS algorithm to achieve the convergence, as shown in Fig. 5.9.

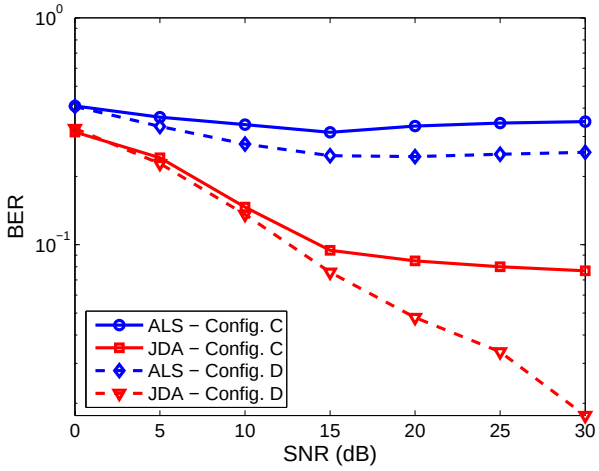

Figure 5.7: BER versus SNR provided by MMSE receiver using the JDA and ALS channel estimates for Configurations C and D.

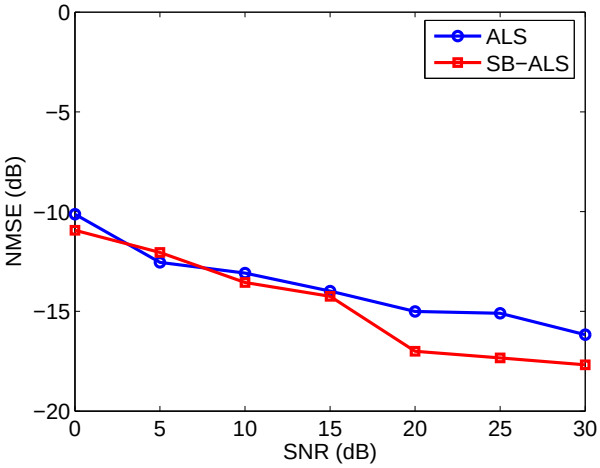

Figure 5.8: NMSE versus SNR provided by the ALS and SB-ALS algorithms for Configuration B.

5.7 Conclusion

In this chapter, two tensor-based methods for estimating memoryless MIMO Volterra channels have been proposed. These methods result from the PARAFAC decomposition of an output covariance tensor. To get this PARAFAC decomposi-

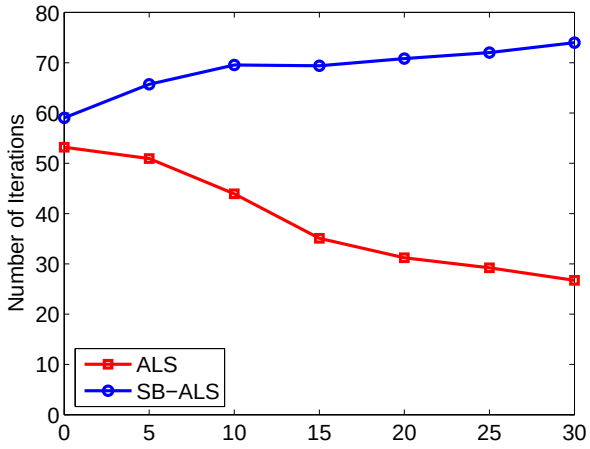

Figure 5.9: Number of iterations for convergence versus SNR for the ALS and SB-ALS algorithms for Configuration B.

tion, a new precoding scheme has been developed so that the transmitted signals satisfy some orthogonality constraints. In this scheme, the transmitted signals are PSK modulated and modeled as DTMCs. A method for designing TPMs that satisfy the orthogonality constraints has been proposed. The channel estimation was carried out using two different algorithms: ALS and JDA.

The proposed identification methods have been applied for identifying an uplink channel in a multiuser ROF communication system. The tested TPM configurations can be divided into two groups: the ones that induce high temporal correlation to the transmitted signals and the ones corresponding to spaced values of phase shifts. The configurations of the first group provide better channel estimates than the ones of the second group, whereas the configurations of the second group lead to better robustness to noise and interference, due to higher distances between the symbol phase shifts. A tradeoff between channel estimation accuracy and equalization robustness to noise and interference is to be taken into account in order to choose the best configurations in terms of BER performance.

In most of the cases, JDA outperforms ALS. However, the ALS algorithm is able to work when the number of antennas is smaller than the number of virtual sources, which is not the case for JDA. Besides, ALS has well performed using only 256 symbols to estimate the covariances. When a short training sequence is used for initializing ALS, its convergence can be significantly accelerated with an improvement of the channel estimation accuracy.

In the next chapter, we deal with MIMO Volterra channels in the context of a CDMA communication system. As well as the methods developed in the present chapter, the techniques presented in Chapter 6 consider a blind scenario and are based on a tensorial approach. The key aspect concerning these techniques is the fact that the use of spreading codes induces a new diversity on the received signals. The received signals can then be treated as as three-dimensional variables, with indices corresponding to symbol, chip and space. The tensor-based techniques exploit the diversity provided by the spreading codes by means of the PARAFAC decomposition.

CHAPTER 6

Estimation and Equalization of MIMO Volterra Channels in CDMA systems

In this chapter, techniques for estimation and equalization of discrete-time equivalent baseband MIMO Volterra channels in the context of a CDMA communication system are developed. As seen in Chapter 2, MIMO Volterra systems have many applications in communication systems as, for instance, in the modeling of uplink ROF multiuser channels. Several authors have considered ROF links in the context of CDMA systems [56, 117, 116, 114]. Moreover, concerning this application, a related work can be found in [112], with semiconductor laser's nonlinearity compensation being carried out for CDMA systems. It should be highlighted that the equalizers proposed in this chapter are developed to cancel Inter-Chip Interference and Multiple-Access Interference, instead of traditional equalizers that cancel inter-symbol interference (ISI).

As well as in Chapter 5, the techniques developed in this chapter are based on tensor decompositions. The key aspect of the proposed algorithms is the fact that the spreading codes induce a new diversity on the received signals. As a consequence, the signals received by an antenna array can be viewed as 3-D variables, with indices corresponding to symbol, chip and space. In this chapter, the terms deterministic and stochastic are related to techniques that use tensors composed of received signals and received signals statistics, respectively.

The main contributions of this chapter are divided in three parts. In the first part (Section 6.1), two deterministic receivers are proposed for memoryless MIMO Volterra channels in CDMA systems, one of them being blind and other one semi-

blind. These techniques are based on the PARAFAC decomposition of a third-order tensor composed of received signals with space, time and code diversities. This tensor-based approach allows joint estimation of the channel coefficients and transmitted signals. Two algorithms are considered for carrying out the PARAFAC decomposition: the ALS and a modified version of the ALS that uses a short training sequence and takes the structure of one of the factor matrices into account. Channel estimation and equalization based on deterministic tensors have been addressed by many authors in the case of linear CDMA channels [144, 143, 33, 34]. In the case of nonlinear channels, deterministic blind PARAFAC-based receivers for Volterra channels were developed in [87, 83], but for single-user TDMA channels. Moreover, a deterministic blind identification method based on the PARAFAC decomposition was recently proposed for single-user Wiener-Hammerstein channels [85], also in the context of a TDMA system.

In the second part (Section 6.2), blind techniques for memoryless MIMO Volterra channel estimation in CDMA systems are developed. As well as the methods developed in Chapter 5, these techniques are based on a stochastic tensorial approach considering that the transmitted signals are PSK modulated. They rely on the PARAFAC decomposition of a fifth-order tensor composed of covariances of the received signals. The PARAFAC model is estimated by three different algorithms, the first one being the ALS algorithm. As the ALS algorithm may need many iterations to achieve the convergence, two non-iterative estimation methods are proposed. The first one is based on Eigenvalue Decompositions (EVD) and the other one exploits a property of the Khatri-Rao product. These two algorithms lead to relaxed restrictions on the number of receive antennas.

The third part of the chapter (Section 6.3) is dedicated to the proposition of a blind estimation method for MIMO Volterra channels with short memory in a CDMA system. The approach considered in this section can be viewed as an extension of the one developed in Chapter 5 for channels with short memory. Indeed, this proposed method is also based on the PARAFAC decomposition of a third-order tensor composed of channel output covariances using PSK transmitted signals.

As mentioned earlier, a great advantage of these tensor-based approaches is that they allow working with weak uniqueness conditions compared with previous works [127, 43, 113], which require that the number of channel outputs be greater than the number of virtual sources. The proposed tensorial techniques developed for CDMA systems provide a great flexibility on the number of antennas and spreading factor, leading to an interesting tradeoff between complexity and bandwidth.

This chapter is organized as follows. Section 6.1 develops deterministic techniques for joint estimation and equalization of memoryless MIMO Volterra channels. In

Section 6.2, stochastic techniques for estimation of memoryless MIMO Volterra channels are proposed. In Section 6.3, a stochastic method for estimating MIMO Volterra channels with short memory is presented. The performance of the proposed techniques is evaluated by means of computer simulations in Section 6.4, and Section 6.5 draws some conclusions about the chapter. Some of the results presented in this chapter have been published in [48, 49, 50].

6.1 Deterministic approach for estimation and equalization of memoryless channels

In this section, after presenting the considered MIMO Volterra CDMA channel, two techniques for joint channel estimation and information recovery are developed based on the PARAFAC decomposition of a deterministic tensor. Indeed, these techniques use a third-order tensor composed of received signals considering space, time and code diversities. The PARAFAC decomposition of this tensor allows joint estimation of the channel, spreading codes and transmitted signals, using only one known pilot symbol. Besides, as we will see, the Kruskal condition for the PARAFAC uniqueness provides a great flexibility on the number of antennas, spreading factor and length of the data block.

The estimation of the PARAFAC factors is first carried out by means of the ALS algorithm, resulting in a technique that can be viewed as an extension of [144] to nonlinear channels. A modified version of the ALS algorithm is also proposed by taking the structure of one of the factor matrices into account and by using a short training sequence. It should be mentioned that the techniques developed in the present section can be used with PSK or QAM modulations.

6.1.1 The CDMA system with memoryless MIMO Volterra channel

The discrete-time equivalent baseband model of the communication channel is assumed to be expressed as a memoryless MIMO Volterra model:

$$y_{r,n,p} = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{k+1}=t_k}^{T} \sum_{t_{k+2}=1}^{T} \cdots \sum_{t_{2k+1}=t_{2k}}^{T} h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}) \prod_{i=1}^{k+1} u_{t_i,n,p} \prod_{i=k+2}^{2k+1} u_{t_i,n,p}^{*} + \upsilon_{r,n,p}, \tag{6.1}$$

with

$$y_{r,n,p} = y_r\left((n-1)P+p\right), \tag{6.2}$$
$$\upsilon_{r,n,p} = \upsilon_r\left((n-1)P+p\right), \tag{6.3}$$
$$u_{t,n,p} = u_t\left((n-1)P+p\right), \tag{6.4}$$

where $y_{r,n,p}$ $(1 \leq r \leq R,\ 1 \leq p \leq P)$ is the chip-rate sampled signal received by antenna r at the p^{th} chip period of the n^{th} symbol period, i.e. received at the $[(n-1)P+p]^{th}$ chip period, R is the number of receive antennas, P is the length of the spreading code (number of chips per symbol), $(2K+1)$ is the nonlinearity order of the model, $h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1})$ are the kernel coefficients of the r^{th} sub-channel, $u_{t,n,p}$ $(1 \leq t \leq T)$ is the spread signal transmitted by user t at the p^{th} chip period of the n^{th} symbol period, T is the number of users and $\upsilon_{r,n,p}$ is the Additive White Gaussian Noise (AWGN).

Equation (6.1) can be rewritten as:

$$y_{r,n,p} = \sum_{q=0}^{Q} h_{r,q}\tilde{u}_{n,p,q} + \upsilon_{r,n,p} = \mathbf{h}_r^T \tilde{\mathbf{u}}_{n,p} + \upsilon_{r,n,p} \ , \tag{6.5}$$

where the vector $\mathbf{h}_r = [h_{r,1}\, h_{r,2}\, \ldots\, h_{r,Q}]^T \in \mathbb{C}^{Q\times 1}$ contains the Volterra kernel coefficients $h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1})$ of the r^{th} sub-channel, the number of parameters of each sub-channel being given by $Q = \sum_{k=0}^{\bar{K}} C_{T,k}C_{T,k+1}$, with $C_{T,k} = \frac{(T+k-1)!}{(T-1)!k!}$. Moreover, $\tilde{\mathbf{u}}_{n,p} = [\tilde{u}_{n,p,1}\, \ldots\, \tilde{u}_{n,p,Q}]^T \in \mathbb{C}^{Q\times 1}$ is the nonlinear input vector containing the products of the spread input signals $u_{t_i,n,p}$ in (6.1), having the following

form:

$$\tilde{u}_{n,p,q} = \prod_{i=1}^{k+1} u_{t_i,n,p} \prod_{i=k+2}^{2k+1} u^*_{t_i,n,p}. \tag{6.6}$$

where the index q depends on the indices $t_1, t_2, ..., t_{2k+1}$. In fact, $\tilde{u}_{n,p,q}$ is defined so that the products $\prod_{i=1}^{k+1} u_{t_i,n,p} \prod_{i=k+2}^{2k+1} u^*_{t_i,n,p}$ are placed at the vector $\tilde{\mathbf{u}}_{n,p}$ according to the ordering defined by the sums of (6.1).

The spread signal $u_{t,n,p}$ is obtained by upsampling the information signal $s_t(n)$ and multiplying it by the spreading code $c_t(p)$ before transmission, leading to:

$$u_{t,n,p} = s_t(n)c_t(p), \tag{6.7}$$

for $p = 1, ..., P$, where $c_t(p)$ is the p^{th} element of the spreading code of the t^{th} user and $s_t(n)$ is the n^{th} information symbol of the t^{th} user. The signal $s_t(n)$ $(1 \leq t \leq T)$ is assumed to be stationary and independent from $s_{t'}(n)$, for $t \neq t'$. Substituting (6.7) into (6.6), we get:

$$\tilde{u}_{n,p,q} = \tilde{s}_{n,q}\tilde{c}_{p,q}, \tag{6.8}$$

where

$$\tilde{s}_{n,q} = \prod_{i=1}^{k+1} s_{t_i}(n) \prod_{i=k+2}^{2k+1} s^*_{t_i}(n) \tag{6.9}$$

and

$$\tilde{c}_{p,q} = \prod_{i=1}^{k+1} c_{t_i}(p) \prod_{i=k+2}^{2k+1} c^*_{t_i}(p). \tag{6.10}$$

Note that the signal $\tilde{s}_{n,q}$ is defined in a similar way to the q^{th} element $\mathbf{w}(n)$ of the nonlinear input vector (2.60). The signal $\tilde{s}_{n,q}$ $(1 \leq q \leq Q)$ denotes a product of information signals in a CDMA system and $w_q(n)$ $(1 \leq q \leq Q)$ denotes a product of transmitted signals in a TDMA system. Substituting (6.8) into (6.5), we get:

$$y_{r,n,p} = \sum_{q=0}^{Q} h_{r,q}\tilde{s}_{n,q}\tilde{c}_{p,q} + \upsilon_{r,n,p}. \tag{6.11}$$

If the information signals $s_t(n)$ are PSK modulated and the spreading codes $c_t(p)$ have an unitary modulus, then the transmitted signals $u_{t,n,p}$ are constant modulus.

In this case, as well as pointed out in Chapter 5, the nonlinear terms corresponding to $t_i = t_j$, for all $i \in \{1, ..., k+1\}$ and $j \in \{k+2, ..., 2k+1\}$, are absent in (6.1) due to the fact that the term $|u_{t_i,n,p}|^2$ reduces to a multiplicative constant that can be absorbed by the associated channel coefficient. In this case, the memoryless MIMO Volterra channel (6.1) becomes:

$$y_{r,n,p} = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{k+1}=t_k}^{T} \underbrace{\sum_{t_{k+2}=1}^{T} \cdots \sum_{t_{2k+1}=t_{2k}}^{T}}_{t_{k+2},\ldots,t_{2k+1} \neq t_1,\ldots,t_{k+1}} h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}) \prod_{i=1}^{k+1} u_{t_i,n,p} \prod_{i=k+2}^{2k+1} u_{t_i,n,p}^{*} + \upsilon_{r,n,p}. \tag{6.12}$$

Equation (6.12) can also be written as (6.11), with a smaller value of Q. For instance, for a linear-cubic channel ($2K + 1 = 3$), we have $Q = T + \frac{T^2(T-1)}{2}$.

6.1.2 Third-Order Received Signal Tensor

For simplifying the development of this section, we consider that there is no AWGN term in the received signals. Let $\mathcal{Y} \in \mathbb{C}^{R \times N \times P}$ be the third-order tensor composed of received signals $y_{r,n,p}$ for $1 \leq r \leq R$, $1 \leq n \leq N$ and $1 \leq p \leq P$, with $[\mathcal{Y}]_{r,n,p} = y_{r,n,p}$, where N is the number of data symbols. Equation (6.11) represents the scalar writing of the PARAFAC decomposition of the third order-tensor $\mathcal{Y}$ with rank $\leq Q$ and matrix components $\mathbf{H} \in \mathbb{C}^{R \times Q}$, $\tilde{\mathbf{C}} \in \mathbb{C}^{P \times Q}$ and $\tilde{\mathbf{S}} \in \mathbb{C}^{N \times Q}$, where

$$\mathbf{H} = [\mathbf{h}_1 \ \ldots \ \mathbf{h}_R]^T \in \mathbb{C}^{R \times Q} \tag{6.13}$$

is the channel matrix,

$$\tilde{\mathbf{S}} = [\tilde{\mathbf{s}}_1 \ \cdots \ \tilde{\mathbf{s}}_N]^T, \in \mathbb{C}^{N \times Q} \tag{6.14}$$

with $\tilde{\mathbf{s}}_n = [\tilde{s}_{n,1} \ \cdots \ \tilde{s}_{n,Q}]^T \in \mathbb{C}^{Q \times 1}$, is the matrix containing the products of the information signals $\tilde{s}_{n,q}$ given in (6.9), and

$$\tilde{\mathbf{C}} = [\tilde{\mathbf{c}}_1 \ \cdots \ \tilde{\mathbf{c}}_P]^T \in \mathbb{C}^{P \times Q}, \tag{6.15}$$

with $\tilde{\mathbf{c}}_p = [\tilde{c}_{p,1} \ \cdots \ \tilde{c}_{p,Q}]^T \in \mathbb{C}^{Q \times 1}$, is the nonlinear code matrix, i.e. the matrix containing the products of the spreading codes $\tilde{c}_{p,q}$ given in (6.10).

For instance, for $T = 2$, $K = 1$ and constant modulus transmitted signals, the

matrices $\mathbf{H}$, $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{C}}$ are respectively given by:

$$\mathbf{H} = \begin{pmatrix} h_1^{(1)}(1) & h_1^{(1)}(2) & h_3^{(1)}(1,1,2) & h_3^{(1)}(2,2,1) \\ \vdots & \vdots & \vdots & \vdots \\ h_1^{(R)}(1) & h_1^{(R)}(2) & h_3^{(R)}(1,1,2) & h_3^{(R)}(2,2,1) \end{pmatrix},$$

$$\tilde{\mathbf{S}} = \begin{pmatrix} s_1(1) & s_2(1) & s_1^2(1)s_2^*(1) & s_2^2(1)s_1^*(1) \\ \vdots & \vdots & \vdots & \vdots \\ s_1(N) & s_2(N) & s_1^2(N)s_2^*(N) & s_2^2(N)s_1^*(N) \end{pmatrix}$$

and

$$\tilde{\mathbf{C}} = \begin{pmatrix} c_1(1) & c_2(1) & c_1^2(1)c_2^*(1) & c_2^2(1)c_1^*(1) \\ \vdots & \vdots & \vdots & \vdots \\ c_1(P) & c_2(P) & c_1^2(P)c_2^*(P) & c_2^2(P)c_1^*(P) \end{pmatrix}.$$

The matrix slices of the tensor $\mathcal{Y}$ are then given by (see Appendix C):

$$\mathbf{Y}_{r,\cdot,\cdot} = \tilde{\mathbf{S}}\ \mathrm{diag}_r[\mathbf{H}]\ \tilde{\mathbf{C}}^T \in \mathbb{C}^{N\times P}, \tag{6.16}$$

$$\mathbf{Y}_{\cdot,n,\cdot} = \tilde{\mathbf{C}}\ \mathrm{diag}_n[\tilde{\mathbf{S}}]\ \mathbf{H}^T \in \mathbb{C}^{P\times R}, \tag{6.17}$$

$$\mathbf{Y}_{\cdot,\cdot,p} = \mathbf{H}\ \mathrm{diag}_p[\tilde{\mathbf{C}}]\ \tilde{\mathbf{S}}^T \in \mathbb{C}^{R\times N}, \tag{6.18}$$

where $\mathrm{diag}_i[\cdot]$ denotes the diagonal matrix formed from the i^{th} row of the matrix argument. The unfolded matrices of the tensor $\mathcal{Y}$ can be defined as:

$$\mathbf{Y}_{[1]} = \begin{bmatrix} \mathbf{Y}_{1,\cdot,\cdot} \\ \vdots \\ \mathbf{Y}_{R,\cdot,\cdot} \end{bmatrix} \in \mathbb{C}^{NR\times P}, \tag{6.19}$$

$$\mathbf{Y}_{[2]} = \begin{bmatrix} \mathbf{Y}_{\cdot,1,\cdot} \\ \vdots \\ \mathbf{Y}_{\cdot,N,\cdot} \end{bmatrix} \in \mathbb{C}^{PN\times R}, \tag{6.20}$$

$$\mathbf{Y}_{[3]} = \begin{bmatrix} \mathbf{Y}_{\cdot,\cdot,1} \\ \vdots \\ \mathbf{Y}_{\cdot,\cdot,P} \end{bmatrix} \in \mathbb{C}^{RP\times N}, \tag{6.21}$$

which leads to:

$$\mathbf{Y}_{[1]} = \left(\mathbf{H} \diamond \tilde{\mathbf{S}}\right) \tilde{\mathbf{C}}^T, \tag{6.22}$$

$$\mathbf{Y}_{[2]} = \left(\tilde{\mathbf{S}} \diamond \tilde{\mathbf{C}}\right) \mathbf{H}^T, \tag{6.23}$$

$$\mathbf{Y}_{[3]} = \left(\tilde{\mathbf{C}} \diamond \mathbf{H}\right) \tilde{\mathbf{S}}^T, \tag{6.24}$$

where $\diamond$ denotes the Khatri-Rao (column-wise Kronecker) product.

The essential uniqueness of the PARAFAC decomposition of $\mathcal{Y}$ is assured by the Kruskal condition:

$$k_{\mathbf{H}} + k_{\tilde{\mathbf{S}}} + k_{\tilde{\mathbf{C}}} \geq 2Q + 2, \tag{6.25}$$

where $k_{\mathbf{A}}$ denote the k-rank of the matrix $\mathbf{A}$. In this case, the essential uniqueness means that if any other set of matrices $\mathbf{H}^{'}$, $\tilde{\mathbf{S}}^{'}$ and $\tilde{\mathbf{C}}^{'}$ satisfies (6.22)-(6.24), then $\mathbf{H}^{'} = \mathbf{H}\Pi\Lambda_1$, $\tilde{\mathbf{S}}^{'} = \tilde{\mathbf{S}}\Pi\Lambda_2$ and $\tilde{\mathbf{C}}^{'} = \tilde{\mathbf{C}}\Pi\Lambda_3$, where Λ_1, Λ_2 and Λ_3 are diagonal matrices such that $\Lambda_1\Lambda_2\Lambda_3 = \mathbf{I}_Q$ and Π is a permutation matrix. Assuming that the matrices $\mathbf{H}$, $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{C}}$ are full k-rank, condition (6.25) becomes:

$$\min(R, Q) + \min(N, Q) + \min(P, Q) \geq 2Q + 2. \tag{6.26}$$

In particular, if we choose $N \geq Q$, we get:

$$\min(R, Q) + \min(P, Q) \geq Q + 2. \tag{6.27}$$

The flexibility on the choice of R, N and P provided by the Kruskal condition is one of the main advantages of using this tensor-based approach. It leads to an interesting tradeoff between complexity (number of receiver antennas R) and capacity (spreading factor P). In particular, note that it is possible to choose $R < Q$ and $P < Q$ i.e. the number of receive antennas and spreading gain are smaller than the number of virtual sources.

In the sequel, two algorithms are presented for carrying out the PARAFAC decomposition of the tensor $\mathcal{Y}$. In the following developments, the matrix $\tilde{\mathbf{C}}$ is assumed to be known, as it can be calculated if the codes are known. So, if the Kruskal condition (6.25) is satisfied, we have $\hat{\tilde{\mathbf{C}}} = \tilde{\mathbf{C}}$ and, hence, $\Pi = \Lambda_3 = \mathbf{I}_Q$ and $\Lambda_2 = \Lambda_1^{-1}$. Therefore, $\hat{\mathbf{H}} = \mathbf{H}\Lambda_1$ and $\hat{\tilde{\mathbf{S}}} = \tilde{\mathbf{S}}\Lambda_1^{-1}$. This means that the permutation ambiguity is eliminated. Moreover, due to the structure of the matrix $\tilde{\mathbf{S}}$, the scaling ambiguity matrix Λ_1 can be identified by using one known pilot symbol for

each user $(s_t(n), t = 1, .., T)$, i.e. by assuming that the first row of $\tilde{\mathbf{S}}$ is known, as:

$$\Lambda_1 = \text{diag}\left[\left[\frac{\tilde{s}_{1,1}}{\hat{\tilde{s}}_{1,1}} \cdots \frac{\tilde{s}_{1,Q}}{\hat{\tilde{s}}_{1,Q}}\right]^T\right], \tag{6.28}$$

where diag[·] denotes the diagonal matrix formed from the vector argument.

6.1.3 Estimation Algorithms

In this section, two algorithms are proposed for carrying out the PARAFAC decomposition of $\mathcal{Y}$ assuming that $\tilde{\mathbf{C}}$ is known. It is important to highlight that, if the matrix $\tilde{\mathbf{C}}$ is unknown, the following algorithms are able to jointly estimate the channel coefficients, transmitted signals and spreading codes. However, in this case, the performance of these estimation algorithms is worst than in the case where $\tilde{\mathbf{C}}$ is known.

Alternating Least Squares algorithm

The first presented channel estimation algorithm is a two-steps version of the ALS algorithm. In this case, the channel estimation problem is solved by minimizing the two following cost functions:

$$J_1 = \left\| \bar{\mathbf{Y}}_{[3]} - \left(\tilde{\mathbf{C}} \diamond \hat{\mathbf{H}}^{(it-1)}\right) \hat{\tilde{\mathbf{S}}}^T \right\|_F^2, \tag{6.29}$$

$$J_2 = \left\| \bar{\mathbf{Y}}_{[2]} - \left(\hat{\tilde{\mathbf{S}}}^{(it)} \diamond \tilde{\mathbf{C}}\right) \hat{\mathbf{H}}^T \right\|_F^2, \tag{6.30}$$

where $\bar{\mathbf{Y}}_{[3]}$ and $\bar{\mathbf{Y}}_{[2]}$ are noisy versions of $\mathbf{Y}_{[3]}$ and $\mathbf{Y}_{[2]}$, respectively. The ALS algorithm is summarized in Table 6.1, where ϵ is a small positive constant, and $\hat{\mathbf{H}}^{(it)}$ and $\hat{\tilde{\mathbf{S}}}^{(it)}$ denotes respectively the estimates of the matrices $\mathbf{H}$ and $\tilde{\mathbf{S}}$ at iteration it. The existence of the left inverse of the matrices $(\hat{\tilde{\mathbf{S}}}^{(it-1)} \diamond \tilde{\mathbf{C}})$ and $(\tilde{\mathbf{C}} \diamond \hat{\mathbf{H}}^{(it)})$ is asymptotically assured if the Kruskal condition (6.25) is satisfied [98].

This technique can be viewed as a generalization of [144] to nonlinear channels. Indeed, in [144], the factor matrices $\mathbf{H}$, $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{C}}$ contain only the elements corresponding to the linear kernel.

Table 6.1: ALS algorithm - deterministic tensor

Initialization:

$\hat{\tilde{\mathbf{H}}}^{(0)} \rightarrow R \times Q$ random matrix

Iterations $(it = it + 1)$ **:**

1) $\hat{\tilde{\mathbf{S}}}^{(it)} = \left[\left(\tilde{\mathbf{C}} \diamond \hat{\mathbf{H}}^{(it-1)}\right)^{\dagger} \bar{\mathbf{Y}}_{[3]}\right]^{T}$

2) $\hat{\mathbf{H}}^{(it)} = \left[\left(\hat{\tilde{\mathbf{S}}}^{(it)} \diamond \tilde{\mathbf{C}}\right)^{\dagger} \bar{\mathbf{Y}}_{[2]}\right]^{T}$

Stop Criteria:

$$\frac{\left\|\hat{\mathbf{H}}^{(it)} - \hat{\mathbf{H}}^{(it-1)}\right\|_F^2}{\left\|\hat{\mathbf{H}}^{(it-1)}\right\|_F^2} < \epsilon \quad \text{and} \quad \frac{\left\|\hat{\tilde{\mathbf{S}}}^{(it)} - \hat{\tilde{\mathbf{S}}}^{(it-1)}\right\|_F^2}{\left\|\hat{\tilde{\mathbf{S}}}^{(it-1)}\right\|_F^2} < \epsilon$$

ALS with Direct Decision and Block Initialization (ALS-DD-BI) algorithm

The performance of the ALS algorithm can be improved by taking into account the structure of matrix $\tilde{\mathbf{S}}$ and the fact that the transmitted symbols belong to a finite alphabet. The second proposed estimation algorithm consists in a modified version of the ALS algorithm where direct decisions are used to construct the matrix $\hat{\tilde{\mathbf{S}}}^{(it)}$ and some known pilot symbols are used to obtain an initial estimate for $\hat{\mathbf{H}}^{(0)}$.

Let us denote by $\hat{\tilde{\mathbf{S}}}_L^{(it)} \in \mathbb{C}^{N \times T}$ the matrix composed of the T first columns of $\hat{\tilde{\mathbf{S}}}^{(it)}$, i.e. the matrix containing the linear part of $\hat{\tilde{\mathbf{S}}}^{(it)}$, and by $\hat{\tilde{\mathbf{S}}}_{NL}^{(it)} \in \mathbb{C}^{N \times (Q-T)}$ the matrix composed of the $(Q-T)$ last columns of $\hat{\tilde{\mathbf{S}}}^{(it)}$, i.e. the matrix containing the nonlinear part of $\hat{\tilde{\mathbf{S}}}^{(it)}$. That gives $\hat{\tilde{\mathbf{S}}}^{(it)} = [\hat{\tilde{\mathbf{S}}}_L^{(it)} | \hat{\tilde{\mathbf{S}}}_{NL}^{(it)}]$.

Moreover, let us define $\hat{\tilde{\mathbf{S}}}_{L,DD}^{(it)} \in \mathbb{C}^{N \times T}$ as the matrix composed of the elements of $\hat{\tilde{\mathbf{S}}}_L^{(it)}$ after a decision device that projects its inputs into the alphabet of symbols. Finally, let $\hat{\tilde{\mathbf{S}}}_{NL,DD}^{(it)} \in \mathbb{C}^{N \times (Q-T)}$ be the nonlinear part of the information signal

Table 6.2: ALS-DD-BI algorithm - deterministic tensor

Initialization:

Calculate $\hat{\mathbf{H}}^{(0)}$ using (6.31).

Iterations $(it = it + 1)$ **:**

1) $\hat{\tilde{\mathbf{S}}}^{(it)} = \left[\left(\tilde{\mathbf{C}} \diamond \hat{\mathbf{H}}^{(it-1)}\right)^{\dagger} \bar{\mathbf{Y}}_{[3]}\right]^{T}$

2) Eliminate the scaling ambiguity from $\hat{\tilde{\mathbf{S}}}_{L}^{(it)}$ by using (6.28).

3) Construct $\hat{\tilde{\mathbf{S}}}_{L,DD}^{(it)}$ by projecting the elements of $\hat{\tilde{\mathbf{S}}}_{L}^{(it)}$ on the alphabet of user symbols.

4) Reconstruct the nonlinear part of the information signal matrix $\hat{\tilde{\mathbf{S}}}_{NL,DD}^{(it)}$ from $\hat{\tilde{\mathbf{S}}}_{L,DD}^{(it)}$.

5) $\hat{\mathbf{H}}^{(it)} = \left[\left(\hat{\tilde{\mathbf{S}}}_{DD}^{(it)} \diamond \tilde{\mathbf{C}}\right)^{\dagger} \bar{\mathbf{Y}}_{[2]}\right]^{T}$

Stop Criteria:

$$\frac{\left\|\hat{\mathbf{H}}^{(it)} - \hat{\mathbf{H}}^{(it-1)}\right\|_F^2}{\left\|\hat{\mathbf{H}}^{(it-1)}\right\|_F^2} < \epsilon \quad \text{and} \quad \frac{\left\|\hat{\tilde{\mathbf{S}}}^{(it)} - \hat{\tilde{\mathbf{S}}}^{(it-1)}\right\|_F^2}{\left\|\hat{\tilde{\mathbf{S}}}^{(it-1)}\right\|_F^2} < \epsilon$$

matrix reconstructed from $\hat{\tilde{\mathbf{S}}}_{L,DD}^{(it)}$.

The ALS-DD-BI algorithm is summarized in Table 6.2, where a short training sequence is used to obtain an initial estimate of the channel matrix $\hat{\mathbf{H}}^{(0)}$. Let us denote by $\tilde{\mathbf{S}}_0 \in \mathbb{C}^{N_t \times Q}$ the matrix composed of the N_t first rows of $\tilde{\mathbf{S}}$ and by $\bar{\mathbf{Y}}_{[2],0} \in \mathbb{C}^{PN_t \times R}$ the corresponding unfolded matrix of the tensor $\mathcal{Y}$, where N_t is the length of the training sequence. The initial estimation of $\mathbf{H}$ is obtained as:

$$\hat{\mathbf{H}}^{(0)} = \left[\left(\tilde{\mathbf{S}}_0 \diamond \tilde{\mathbf{C}}\right)^{\dagger} \bar{\mathbf{Y}}_{[2],0}\right]^{T}. \tag{6.31}$$

Note that a necessary identifiability condition for this initialization is $r_{(\tilde{\mathbf{S}}_0 \diamond \tilde{\mathbf{C}})} = Q$, which implies $N_t P \geq Q$.

6.2 Stochastic approach for estimation of memoryless channels

In this section, techniques for estimating MIMO Volterra channels are presented, considering the channel model presented in Section 6.1.1. However, instead of using tensor composed of channel outputs, the techniques developed in the present section use a tensor composed of channel output covariances, similarly as in Chapter 5. In the sequel, the spatio-temporal covariances of the received signals are characterized, which allows the construction of a fifth-order tensor composed of such covariances. The algorithms used to perform the PARAFAC decomposition of this tensor are treated at the end of the section, with three estimation algorithms being proposed.

It is assumed that the information signals $s_t(n)$ $(1 \leq t \leq T)$ belong to a PSK constellation and that the spreading codes $c_t(p)$ have an unitary modulus, which lead to transmitted signals $u_{t,n,p}$ with constant modulus.

6.2.1 Covariance Matrices of the Received Signals

Equation (6.11) can be expressed in a vector form as:

$$\mathbf{y}(n,p) = \mathbf{H}\text{diag}_p\left[\tilde{\mathbf{C}}\right]\tilde{\mathbf{s}}_n + \mathbf{v}(n,p). \tag{6.32}$$

where $\mathbf{y}(n,p) = [y_{1,n,p}\ldots y_{R,n,p}]^T \in \mathbb{C}^{R\times 1}$ and $\mathbf{v}(n,p) = [\upsilon_{1,n,p}\ \ldots \upsilon_{R,n,p}]^T \in \mathbb{C}^{R\times 1}$. Let us define the spatio-temporal covariance matrices of the chip-rate sampled received signals $\mathbf{y}(n,p)$ as:

$$\begin{aligned}\mathbf{R}_{\mathbf{y}}(d,p_1,p_2) &\equiv \mathbb{E}\left[\mathbf{y}(n+d,p_1)\mathbf{y}^H(n,p_2)\right] \in \mathbb{C}^{R\times R}\\ &= \mathbf{H}\,\text{diag}_{p_1}\left[\tilde{\mathbf{C}}\right]\mathbf{R}_{\tilde{\mathbf{s}}}(d)\,\text{diag}_{p_2}\left[\tilde{\mathbf{C}}^*\right]\mathbf{H}^H + \sigma^2\mathbf{I}_R\delta(d)\delta(p_1-p_2),\end{aligned} \tag{6.33}$$

where $0 \leq d \leq D-1$, D is the number of delays taken into account and

$$\mathbf{R}_{\tilde{\mathbf{s}}}(d) = \mathbb{E}\left[\tilde{\mathbf{s}}(n+d)\tilde{\mathbf{s}}^H(n)\right] \in \mathbb{C}^{Q\times Q}. \tag{6.34}$$

Assuming that the noise variance σ^2 is known, the noise covariance matrix can be ignored in (6.33) as it can be subtracted from $\mathbf{R}_y(0,p_1,p_2)$. If the noise variance is not known, the proposed methods can be carried out without using the covariance matrices $\mathbf{R}_{\mathbf{y}}(0,p,p)$, for $p = 1,...,P$.

When the transmitted signals are i.i.d. and have a PSK modulation, the covariance matrix $\mathbf{R}_{\tilde{\mathbf{s}}}(0)$ is non-null and diagonal, and the matrices $\mathbf{R}_{\tilde{\mathbf{s}}}(d)$ are null for $d > 0$. However, if the precoding scheme developed in Chapter 5 is used, the matrices $\mathbf{R}_{\tilde{\mathbf{s}}}(d)$ are non-null and diagonal for all d. Thus, if the matrices $\mathbf{R}_{\tilde{\mathbf{s}}}(d)$ ($d = 0, ..., D-1$) are diagonal, (6.33) can be rewritten as:

$$\mathbf{R}_{\mathbf{y}}(d, p_1, p_2) = \mathbf{H}\,\mathrm{diag}_{p_1}\left[\tilde{\mathbf{C}}\right]\mathrm{diag}_{d+1}[\mathbf{Z}]\mathrm{diag}_{p_2}\left[\tilde{\mathbf{C}}^*\right]\mathbf{H}^H, \tag{6.35}$$

where the rows of the matrix $\mathbf{Z} \in \mathbb{C}^{D\times Q}$ contain the diagonal elements of $\mathbf{R}_{\tilde{\mathbf{s}}}(d)$ for $0 \leq d \leq D-1$, i.e. $z_{d+1,q} = [\mathbf{Z}]_{d+1,q} = [\mathbf{R}_{\tilde{\mathbf{s}}}(d)]_{q,q}$.

As we will see later, the use of this precoding scheme is not mandatory for the techniques proposed in this section, as the channel estimation algorithms may work with $D = 1$, i.e. by using only the covariance matrices $\mathbf{R}_{\mathbf{y}}(0, p_1, p_2)$, for $1 \leq p_1, p_2 \leq P$. However, the precoding has the advantage of adding redundancy to the transmitted signals, which induces a supplementary dimension to the tensor.

6.2.2 Fifth-Order Tensor of Covariances

Let us define $\mathcal{R} \in \mathbb{C}^{D\times R\times R\times P\times P}$ as the fifth-order tensor composed of the covariance matrices $\mathbf{R}_y(d, p_1, p_2)$, for $0 \leq d \leq D-1$ and $1 \leq p_1, p_2 \leq P$, constructed in such a way that the $(r_1, r_2)^{th}$ element of the matrix $\mathbf{R}_y(d, p_1, p_2)$ corresponds to the $(d+1, r_1, r_2, p_1, p_2)^{th}$ element of $\mathcal{R}$, i.e. $[\mathbf{R}_y(d, p_1, p_2)]_{r_1,r_2} = [\mathcal{R}]_{d+1,r_1,r_2,p_1,p_2}$. From (6.35), a typical element of the tensor can be expressed by:

$$r_{d+1,r_1,r_2,p_1,p_2} = \sum_{q=1}^{Q} h_{r_1,q}\tilde{c}_{p_1,q}z_{d+1,q}\tilde{c}^*_{p_2,q}h^*_{r_2,q}. \tag{6.36}$$

where $r_{d+1,r_1,r_2,p_1,p_2} = [\mathcal{R}]_{d+1,r_1,r_2,p_1,p_2}$. Equation (6.36) corresponds to the PARAFAC decomposition (see Appendix C) of a fifth-order tensor with rank $\leq Q$ and matrix factors $\mathbf{H} \in \mathbb{C}^{R\times Q}$, $\tilde{\mathbf{C}} \in \mathbb{C}^{P\times Q}$, $\mathbf{Z} \in \mathbb{C}^{D\times Q}$, $\tilde{\mathbf{C}}^* \in \mathbb{C}^{P\times Q}$ and $\mathbf{H}^* \in \mathbb{C}^{R\times Q}$.

Its is possible to define the 10 types of matrix slices of the tensor $\mathcal{R}$, depending on which indexes are fixed. Using the notation introduced in Appendix C, some

of these matrix slices are given by:

$$\mathbf{R}_{d+1,r_1,\cdot,p_1,\cdot} = \mathbf{H}^* \operatorname{diag}_{d+1}[\mathbf{Z}] \operatorname{diag}_{r_1}[\mathbf{H}] \operatorname{diag}_{p_1}\left[\tilde{\mathbf{C}}\right] \tilde{\mathbf{C}}^H \in \mathbb{C}^{R\times P}, \quad (6.37)$$

$$\mathbf{R}_{d+1,\cdot,\cdot,p_1,p_2} = \mathbf{H} \operatorname{diag}_{d+1}[\mathbf{Z}] \operatorname{diag}_{p_1}\left[\tilde{\mathbf{C}}\right] \operatorname{diag}_{p_2}\left[\tilde{\mathbf{C}}^*\right] \mathbf{H}^H \in \mathbb{C}^{R\times R}, \quad (6.38)$$

$$\mathbf{R}_{d+1,\cdot,r_2,p_1,\cdot} = \mathbf{H} \operatorname{diag}_{d+1}[\mathbf{Z}] \operatorname{diag}_{r_2}[\mathbf{H}^*] \operatorname{diag}_{p_1}\left[\tilde{\mathbf{C}}\right] \tilde{\mathbf{C}}^H \in \mathbb{C}^{R\times P}, \quad (6.39)$$

for $0 \leq d \leq D-1$, $1 \leq p_1, p_2 \leq P$ and $1 \leq r_1, r_2 \leq R$.

The unfolded matrices of the tensor $\mathcal{R}$ are constructed by stacking all the matrix slices of a given type. The channel estimation algorithms presented in the next section are based on the following unfolded matrices of $\mathcal{R}$:

$$\mathbf{R}_{[1]} = \left(\mathbf{H} \diamond \tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^*\right) \mathbf{H}^H \quad \in \mathbb{C}^{RP^2D\times R}, \quad (6.40)$$

$$\mathbf{R}_{[2]} = \left(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^* \diamond \mathbf{H}^*\right) \mathbf{H}^T \quad \in \mathbb{C}^{RP^2D\times R}, \quad (6.41)$$

$$\mathbf{R}_{[3]} = \left(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^*\right) (\mathbf{H}^* \diamond \mathbf{H})^T \quad \in \mathbb{C}^{P^2D\times R^2}. \quad (6.42)$$

These unfolding matrices are constructed so that the element $[\mathcal{R}]_{(d+1),r_1,r_2,p_1,p_2}$ of the tensor is placed at the position $(i_{\text{lin}}, i_{\text{col}})$ of the unfolding matrix, with i_{lin} and i_{col} defined as:

$$i_{\text{lin}} = (r_1-1)P^2D + (p_1-1)PD + dP + p_2, \quad i_{\text{col}} = r_2, \quad (6.43)$$

$$i_{\text{lin}} = (p_1-1)RPD + dRP + (p_2-1)R + r_2, \quad i_{\text{col}} = r_1, \quad (6.44)$$

$$i_{\text{lin}} = (p_1-1)PD + dP + p_2, \quad i_{\text{col}} = (r_2-1)R + r_1, \quad (6.45)$$

for the matrices $\mathbf{R}_{[1]}$, $\mathbf{R}_{[2]}$ and $\mathbf{R}_{[3]}$, respectively.

The essential uniqueness of the PARAFAC decomposition of the tensor $\mathcal{R}$ is assured by the Kruskal theorem. In this case, the Kruskal theorem implies that if:

$$2k_{\mathbf{H}} + 2k_{\mathbf{C}} + k_{\mathbf{Z}} \geq 2Q + 4, \quad (6.46)$$

then the matrix factors $\mathbf{H}$, $\mathbf{H}^*$, $\tilde{\mathbf{C}}$, $\tilde{\mathbf{C}}^*$ and $\mathbf{Z}$ are unique up to column scaling and permutation ambiguities. That means that if any other set of matrices $\mathbf{H}^{'}$, $\mathbf{H}^{''}$, $\mathbf{C}^{'}$, $\mathbf{C}^{''}$ and $\mathbf{Z}^{'}$ satisfy (6.36), then $\mathbf{H}^{'} = \mathbf{H}\Pi\Lambda_a$, $\mathbf{H}^{''} = \mathbf{H}^*\Pi\Lambda_b$, $\mathbf{C}^{'} = \tilde{\mathbf{C}}\Pi\Lambda_c$,

$\mathbf{C}^{''} = \tilde{\mathbf{C}}^{*}\Pi\Lambda_d$ and $\mathbf{Z}^{'} = \mathbf{Z}\Pi\Lambda_e$, where Λ_a, Λ_b, Λ_c, Λ_d and Λ_e are diagonal matrices such that $\Lambda_a\Lambda_b\Lambda_c\Lambda_d\Lambda_e = \mathbf{I}_Q$ and Π is a permutation matrix. Assuming that the matrices $\mathbf{H}$, $\mathbf{C}$ and $\mathbf{Z}$ are full k-rank, equation (6.46) becomes:

$$\min(R, Q) + \min(P, Q) + \frac{\min(D, Q)}{2} \geq Q + 2. \tag{6.47}$$

The matrix $\mathbf{Z}$, containing the information about the time correlation introduced by the precoding scheme, can also be assumed to be known, as seen in Chapter 5. It should be mentioned that the configurations of transition probability matrices of Table 5.4 provide full k-rank matrices $\mathbf{Z}$. Moreover, as mentioned in Section 6.1, the matrix $\tilde{\mathbf{C}}$ containing the code products is assumed to be known, as it can be calculated from the spreading codes $c_t(p)$. Thus, the spreading codes should be chosen such that $\mathbf{C}$ is full k-rank. Thus, if the Kruskal condition (6.47) is verified, we have $\mathbf{C}^{'} = \tilde{\mathbf{C}}$, $\mathbf{C}^{''} = \tilde{\mathbf{C}}^{*}$, $\mathbf{Z}^{'} = \mathbf{Z}$ and, hence, $\Pi = \mathbf{I}_Q$, $\Lambda_c = \Lambda_d = \Lambda_e = \mathbf{I}_Q$ and $\Lambda_b = \Lambda_a^{-1} = \Lambda^{-1}$, where Λ is a $Q \times Q$ diagonal matrix. That gives $\mathbf{H}^{'} = \mathbf{H}\Lambda$ and $\mathbf{H}^{''} = \mathbf{H}^{*}\Lambda^{-1}$. The scaling ambiguity does not represent an effective problem, as its effects can be canceled by using a differential modulation. Another possible solution is to use a few pilot signals to estimate this ambiguity.

The uniqueness condition (6.47) is weaker than that associated with other estimation methods [127, 43, 113] and that obtained in Chapter 5. The flexibility on the choice of R and P provided by the Kruskal condition is one of the main advantages of using a tensor-based approach, allowing an interesting tradeoff between complexity and bandwidth. As well as in Section 6.1, it is possible to choose $R < Q$ and $P < Q$.

Moreover, the sufficient condition (6.46) is not necessary for the uniqueness of the tensor decomposition. In the next section, we state alternative sufficient conditions by taking into account the fact that some matrix factors are known.

6.2.3 Channel Estimation Algorithms

This section presents three blind algorithms to carry out the PARAFAC decomposition. The first one is a two-steps ALS algorithm. As the ALS algorithm may need many iterations to achieve the convergence, two non-iterative estimation methods are also proposed. It is worth mentioning that these algorithms do not require the use of the precoding if the number of used covariance matrices is set to one ($D = 1$).

Two-steps ALS algorithm

The first channel estimation method is a two-steps version of the ALS algorithm. In the case of the fifth-order tensor $\mathcal{R}$, the channel estimation problem is solved by minimizing the two following least squares cost functions in an alternate way:

$$\begin{aligned} J_1 &= \left\| \hat{\mathbf{R}}_{[1]} - \left(\hat{\mathbf{H}}_a^{(it-1)} \diamond \tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^* \right) \mathbf{H}_b^T \right\|_F^2 \\ J_2 &= \left\| \hat{\mathbf{R}}_{[2]} - \left(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^* \diamond \hat{\mathbf{H}}_b^{(it)} \right) \mathbf{H}_a^T \right\|_F^2, \end{aligned} \tag{6.48}$$

where the matrices $\hat{\mathbf{H}}_a^{(it)}$ and $\hat{\mathbf{H}}_b^{(it)}$ denote respectively the estimates of $\mathbf{H}$ and $\mathbf{H}^*$ at the it^{th} iteration, $\hat{\mathbf{R}}_{[1]}$ and $\hat{\mathbf{R}}_{[2]}$ are respectively the sample estimates of $\mathbf{R}_{[1]}$ and $\mathbf{R}_{[2]}$, the covariances of the received signals being estimated in the following way:

$$\hat{r}_{d+1,r_1,r_2,p_1,p_2} = \frac{1}{N} \sum_{n=1}^{N} y_{r_1,(n+d-1),p_1} y^*_{r_2,n,p_2}, \tag{6.49}$$

where N is the number of data symbols. The ALS algorithm is described in Table 6.3, where ϵ is an arbitrary small positive constant and $\hat{\mathbf{H}}_{ab}^{(it)} = 0.5 \cdot [\hat{\mathbf{H}}_a^{(it)} + (\hat{\mathbf{H}}_b^{(it)})^*]$. Three channel estimates are obtained: $\hat{\mathbf{H}}_a^{(it)}$, $(\hat{\mathbf{H}}_b^{(it)})^*$ and $\hat{\mathbf{H}}_{ab}^{(it)}$; the final channel estimate being chosen as the one providing the smallest value of the cost function (6.48).

The existence of the left inverse of the matrices $(\hat{\mathbf{H}}_a^{(it)} \diamond \tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^*)$ and $(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^* \diamond \hat{\mathbf{H}}_b^{(it)})$ is asymptotically assured if the if the Kruskal condition (6.46) is verified. As well as in the deterministic approach, if the matrix $\tilde{\mathbf{C}}$ is unknown, the ALS algorithm is able to jointly estimate the matrices $\mathbf{H}$ and $\tilde{\mathbf{C}}$. However, in our simulations, the performance of the estimation algorithms for $\tilde{\mathbf{C}}$ unknown is worst than in the case where $\tilde{\mathbf{C}}$ is known.

EVD-LS algorithm

In order to avoid possible convergence problems associated with the ALS algorithm, we propose a non-iterative method to estimate the channel matrix $\mathbf{H}$ from the tensor $\mathcal{R}$, by using the unfolded matrix (6.42). This technique, called Eigenvalue Decomposition-based Least Squares (EVD-LS) algorithm, takes into account the fact that the matrices $\tilde{\mathbf{C}}$ and $\mathbf{Z}$ are known and allows the development of an alternative identifiability condition. In this case, the channel estimation algorithm

Table 6.3: ALS algorithm - stochastic tensor

Initialization:

$\hat{\mathbf{H}}_a^{(0)} \rightarrow R \times Q$ random matrix

Iterations $(it = it + 1)$ **:**

$$1)\ \hat{\mathbf{H}}_b^{(it)} = \left[\left(\hat{\mathbf{H}}_a^{(it-1)} \diamond \tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^*\right)^{\dagger} \hat{\mathbf{R}}_{[1]}\right]^T$$

$$2)\ \hat{\mathbf{H}}_a^{(it)} = \left[\left(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^* \diamond \hat{\mathbf{H}}_b^{(it)}\right)^{\dagger} \hat{\mathbf{R}}_{[2]}\right]^T$$

Stop Criteria:

$$\frac{\left\|\hat{\mathbf{H}}_{ab}^{(it)} - \hat{\mathbf{H}}_{ab}^{(it-1)}\right\|_F^2}{\left\|\hat{\mathbf{H}}_{ab}^{(it-1)}\right\|_F^2} < \epsilon$$

is based on the minimization of the following cost function:

$$J = \left\|\hat{\mathbf{R}}_{[3]} - \left(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^*\right)\left(\hat{\mathbf{H}}^* \diamond \hat{\mathbf{H}}\right)^T\right\|_F^2, \tag{6.50}$$

where $\hat{\mathbf{R}}_{[3]}$ is the sample estimate of $\mathbf{R}_{[3]}$ calculated using N data symbols.

The EVD-LS algorithm is summarized in Table 6.4. The first step consists in the LS estimation of the matrix $\mathbf{W} = (\mathbf{H}^* \diamond \mathbf{H}) \in \mathbb{C}^{R^2 \times Q}$ from the unfolding matrix $\hat{\mathbf{R}}_{[3]}$. In the second and third steps, the q^{th} column of the channel matrix $\mathbf{H}$ is estimated from a matrix formed from the q^{th} column of $\hat{\mathbf{W}}$ by means of the EVD, as in [86]

Remark that the EVD-LS method requires that the following identifiability condition be satisfied: $r_{(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^*)} = Q$, i.e. the matrix $(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^*)$ must be full column rank. A sufficient condition for assuring that the Khatri-Rao product of two matrices is full column rank was established in [142]. This condition can be generalized

Table 6.4: EVD-LS algorithm - stochastic tensor

1) $\hat{\mathbf{W}} = \left[\left(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^{*}\right)^{\dagger} \hat{\mathbf{R}}_{[3]}\right]^{T}$, where $\mathbf{W} = \mathbf{H}^{*} \diamond \mathbf{H} \in \mathbb{C}^{R^2 \times Q}$.

2) For $q = 1, ..., Q$: Construct

$$\hat{\mathbf{W}}(q) = \mathbf{unvec}(\hat{\mathbf{W}}_{\cdot\, q}) = \mathbf{unvec}(\hat{\mathbf{H}}^{*}_{\cdot\, q} \diamond \hat{\mathbf{H}}_{\cdot\, q}) = \hat{\mathbf{H}}_{\cdot\, q}\hat{\mathbf{H}}^{H}_{\cdot\, q},$$

where $\hat{\mathbf{W}}_{\cdot\, q}$ and $\hat{\mathbf{H}}_{\cdot\, q}$ denote the q^{th} column of $\hat{\mathbf{W}}$ and $\hat{\mathbf{H}}$ respectively, and the operator $\mathbf{unvec}\,(\cdot)$ forms a $R \times R$ matrix from its vector argument.

3) Calculate $\hat{\mathbf{H}}_{\cdot q}$ as the eigenvector associated with the largest eigenvalue of $\hat{\mathbf{W}}(q)$.

to multi Khatri-Rao products by using the following result due to [153]:

$$k_{(\mathbf{A}_1 \diamond \mathbf{A}_2)} \geq \min\left(k_{\mathbf{A}_1} + k_{\mathbf{A}_2} - 1, Q\right), \tag{6.51}$$

where $\mathbf{A}_1 \in \mathbb{C}^{L_1 \times Q}$ and $\mathbf{A}_2 \in \mathbb{C}^{L_2 \times Q}$ do not contain an all-zero column. By applying this result $F - 1$ times, we get [168]:

$$k_{(\mathbf{A}_1 \diamond \cdots \diamond \mathbf{A}_F)} \geq \min\left(\sum_{f=1}^{F} k_{\mathbf{A}_f} - (F-1), Q\right), \tag{6.52}$$

where $\mathbf{A}_f \in \mathbb{C}^{L_f \times Q}$, for $f = 1, ..., F$. Thus, if

$$\sum_{f=1}^{F} k_{\mathbf{A}_f} \geq Q + (F-1), \tag{6.53}$$

then $k_{(\mathbf{A}_1 \diamond \cdots \diamond \mathbf{A}_F)} = Q$ and, hence, $r_{(\mathbf{A}_1 \diamond \cdots \diamond \mathbf{A}_F)} = Q$. A sufficient identifiability condition for the EVD-LS algorithm is therefore given by:

$$2k_{\tilde{\mathbf{C}}} + k_{\mathbf{Z}} \geq Q + 2. \tag{6.54}$$

Assuming that the matrices $\tilde{\mathbf{C}}$ and $\mathbf{Z}$ are full k-rank, this condition becomes:

$$2\min(P, Q) + \min(D, Q) \geq Q + 2. \tag{6.55}$$

Thus, an advantage of this approach is that it does not impose constraints on the number R of antennas, contrarily to the ALS algorithm. This is due to the fact that the uniqueness condition (6.55) takes into account the fact that the matrices $\mathbf{Z}$ are $\tilde{\mathbf{C}}$ are known.

Single-LS algorithm

The channel estimation technique proposed in this section, called Single-LS algorithm, can be viewed as a simplified version of the EVD-LS algorithm. In fact, the Single-LS algorithm uses the first step of the EVD-LS algorithm to estimate the matrix $\mathbf{W} = (\mathbf{H}^* \diamond \mathbf{H})$. Then, it exploits a property of the Khatri-Rao product to estimate the channel matrix $\mathbf{H}$.

Indeed, the matrix $\mathbf{W}$ can be expressed as:

$$\mathbf{W} = (\mathbf{H}^* \diamond \mathbf{H}) = \begin{pmatrix} \mathbf{H}\text{diag}_1\left[\mathbf{H}^*\right] \\ \vdots \\ \mathbf{H}\text{diag}_R\left[\mathbf{H}^*\right] \end{pmatrix}. \tag{6.56}$$

Thus, by defining $\hat{\mathbf{W}}^{(r)} \in \mathbb{C}^{R \times Q}$ $(1 \leq r \leq R)$ as the matrix formed from the rows $[(r-1)R+1]$ up to (rR) of $\hat{\mathbf{W}}$, the channel matrix $\hat{\mathbf{H}}$ can then be estimated up to a diagonal matrix as the mean of $\hat{\mathbf{W}}^{(r)}$ for $r = 1, ..., R$. The Single-LS algorithm is summarized in Table 6.5. Note that the computational complexity of the Single-LS algorithm is smaller than that of the EVD-LS algorithm, as it computes only one LS operation and no EVD operation. Moreover, the identifiability condition of the Single-LS algorithm is the same as the one of the EVD-LS algorithm, i.e. $r_{(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^*)} = Q$, which leads to the sufficient identifiability condition (6.54).

6.2.4 Comparison between the uniqueness conditions of the stochastic and deterministic approaches

This section provides a brief comparison between the uniqueness conditions of the proposed stochastic and deterministic tensor-based techniques. Table 6.6 shows the uniqueness condition of the techniques developed in Sections 6.1 and 6.2, for $K = 1$ (third-order nonlinearity) with $T = 2$ and $T = 3$ sources, which corresponds to $Q = 4$ and $Q = 12$ virtual sources respectively. It is assumed that the factor matrices are full k-rank. Moreover, for simplifying the comparison, we have made some realistic assumptions about the number of delays D, symbols N, receive antennas R and

Table 6.5: Single-LS algorithm - stochastic tensor

1) $\hat{\mathbf{W}} = \left[\left(\tilde{\mathbf{C}} \diamond \mathbf{Z} \diamond \tilde{\mathbf{C}}^{*}\right)^{\dagger} \hat{\mathbf{R}}_{[3]}\right]^{T}$, where $\mathbf{W} = \mathbf{H}^{*} \diamond \mathbf{H} \in \mathbb{C}^{R^2 \times Q}$.

2) For $r = 1, ..., R$: Construct $\hat{\mathbf{W}}^{(r)} = \begin{bmatrix} \hat{\mathbf{W}}_{(r-1)R+1,\cdot} \\ \vdots \\ \hat{\mathbf{W}}_{rR,\cdot} \end{bmatrix}$, where

$\hat{\mathbf{W}}_{r,\cdot}$ denotes the r^{rh} row of $\hat{\mathbf{W}}$.

3) $\hat{\mathbf{H}} = \frac{1}{R} \sum_{r=1}^{R} \hat{\mathbf{W}}^{(r)}$.

Table 6.6: Uniqueness Conditions of the Proposed Techniques

Algorithm	2 users	3 users
ALS (deterministic)	$R + P \geq 6$	$R + P \geq 14$
ALS-DD-BI (deterministic)	$R + P \geq 6$ $N_t P \geq 4$	$R + P \geq 14$ $N_t P \geq 12$
ALS (stochastic)	$R + P \geq 4$	$R + P \geq 10$
EVD-LS and Single-LS (stochastic)	$P \geq 1$	$P \geq 3$

spreading factor P. We considered that $D, N \geq 4$ and $R, P \leq 4$ in the case of $T = 2$. For $T = 3$, we considered that $D \geq 8$, $N \geq 12$ and $R, P \leq 12$. Note that the EVD-LS and Single-LS algorithms provide the most relaxed uniqueness condition, while the deterministic techniques provide the strongest uniqueness conditions. This is due to the fact that these techniques doe not impose constraints on the number of receive antennas.

6.3 Stochastic approach for estimation of short memory channels

In this section, a method for blind estimation of MIMO Volterra CDMA channels with short memory is proposed. In fact, the zero-memory assumption considered in Sections 6.1 and 6.1 represents a strong physical constraint if the spreading gain P is large. The developed method can be viewed as an extension of the techniques developed in Chapter 5 for channels with short memory. The information signals $s_t(n)$ $(1 \leq t \leq T)$ are assumed to be PSK modulated and generated using the precoding scheme developed in Chapter 5. The key aspect of the approach presented in this section is the use of spreading codes containing "guard-chips" [144] that allow an equivalent memoryless writing of the MIMO Volterra channel. As we will see, the joint use of PSK modulation and guard-chips leads to a great simplification of the Volterra model.

6.3.1 The CDMA system with short memory MIMO Volterra channel

The communication channel considered in this section is modeled as an equivalent baseband MIMO Volterra system with memory:

$$y_{r,n,p} = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \sum_{t_3=1}^{T} \cdots \sum_{t_{2k+1}=1}^{T} \sum_{m_1=0}^{M} \sum_{m_3=0}^{M} \cdots \sum_{m_{2k+1}=0}^{M}$$
$$h_{2k+1}^{(r)}(t_1, t_3, \ldots, t_{2k+1}, m_1, m_3, \ldots, m_{2k+1}) \prod_{i=1}^{k+1} u_{t_i}(\bar{n} - m_i) \prod_{i=k+2}^{2k+1} u_{t_i}^*(\bar{n} - m_i)$$
$$+ \upsilon_{r,n,p}, \qquad (6.57)$$

where $\bar{n} = (n-1)P + p$ represents the p^{th} chip period of the n^{th} symbol period and the spread signals $u_t(\bar{n}) = u_{t,n,p}$ are generated using (6.7).

In the sequel, we show that, if the channel memory M corresponds to a few chips, i.e. $M < P$, nonlinear Inter-Symbol Interference (ISI) can be avoided by considering that the spreading codes contain guard-chips [144]. However, as we will see in the sequel, the use of guard-chips does not avoid nonlinear Inter-Chip Interference. In this case, the M last elements of the spreading codes are equal to

zero, i.e.

$$c_t(p) = 0, \quad \text{for} \quad P - M + 1 \leq p \leq P \quad \text{and} \quad 1 \leq t \leq T. \tag{6.58}$$

In the following development, we also consider that $c_t(p) = 0$, for $p \leq 0$. From (6.7), we may write:

$$\begin{aligned} u_t(\bar{n} - m) &= u_t((n-1)P + p - m) \\ &= \begin{cases} c_t\,(p-m)\,s_t(n), & \text{if} \quad 1 \leq p - m \leq P, \\ c_t\,(P+p-m)\,s_t(n-1), & \text{if} \quad p - m \leq 0. \end{cases} \end{aligned} \tag{6.59}$$

Note that if $p - m \leq 0$, then $P - M + 1 \leq P + p - m \leq P$. Thus, from (6.58), we have $c_t(P + p - m) = 0$, leading to $u_t(\bar{n} - m) = 0$, which can be expressed as $u_t(\bar{n} - m) = c_t(p - m)s_t(n) = 0$.

We can therefore replace $u_t(\bar{n} - m)$ by $c_t(p - m)s_t(n)$ in (6.57), leading to:

$$y_{r,n,p} = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{2k+1}=1}^{T} \bar{g}_{2k+1}^{(r,p)}(t_1, \ldots, t_{2k+1}) \prod_{i=1}^{k+1} s_{t_i}(n) \prod_{i=k+2}^{2k+1} s_{t_i}^*(n) + \upsilon_{r,n,p}, \tag{6.60}$$

where

$$\begin{aligned} \bar{g}_{2k+1}^{(r,p)}(t_1, \ldots, t_{2k+1}) = \sum_{m_1=0}^{M} \cdots \sum_{m_{2k+1}=0}^{M} & h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, m_1, \ldots, m_{2k+1}) \\ & \prod_{i=1}^{k+1} c_{t_i}(p - m_i) \prod_{i=k+2}^{2k+1} c_{t_i}^*(p - m_i). \end{aligned} \tag{6.61}$$

Note that the use of guard-chips leads to an equivalent memoryless Volterra representation of the channel, the kernel coefficients $\bar{g}_{2k+1}^{(r,p)}(t_1, \ldots, t_{2k+1})$, given by (6.61), depending on the spreading codes $c_t(p)$ and the original kernel coefficients $h_{2k+1}^{(r)}(t_1, \ldots, t_{2k+1}, m_1, \ldots, m_{2k+1})$. Note also that the linear kernel $\bar{g}_1^{(r,p)}(t)$ is given by the convolution of the linear kernel $h_1^{(r)}(t, m)$ with the spreading code $c_t(p)$. It should be also highlighted that each value of p generates a new output for the equivalent channel $\bar{g}_{2k+1}^{(r,p)}(t_1, \ldots, t_{2k+1})$, leading to a Volterra system with RP outputs at each symbol period n.

As the information signals $s_t(n)$ are PSK modulated, the nonlinear terms corresponding to $t_i = t_j$, for all $i \in \{1, ..., k+1\}$ and $j \in \{k+2, ..., 2k+1\}$, can be

eliminated from (6.60), leading to the following triangular form:

$$y_{r,n,p} = \sum_{k=0}^{K} \sum_{t_1=1}^{T} \cdots \sum_{t_{k+1}=t_k}^{T} \underbrace{\sum_{t_{k+2}=1}^{T} \cdots \sum_{t_{2k+1}=t_{2k}}^{T}}_{t_{k+2},\ldots,t_{2k+1} \neq t_1,\ldots,t_{k+1}} g_{2k+1}^{(r,p)}(t_1,\ldots,t_{2k+1}) \prod_{i=1}^{k+1} s_{t_i}(n) \prod_{i=k+2}^{2k+1} s_{t_i}^{*}(n) + \upsilon_{r,n,p}. \tag{6.62}$$

The RP output signals at the n^{th} symbol period can be expressed in the following form:

$$\check{\mathbf{y}}(n) = \mathbf{G}\,\tilde{\mathbf{s}}_n + \check{\mathbf{v}}(n), \tag{6.63}$$

where $\check{\mathbf{y}}(n) = [y_{1,n,1}\, y_{2,n,1} \ldots y_{R,n,1} \;\cdots\; y_{1,n,P}\, y_{2,n,P} \ldots y_{R,n,P}]^T \in \mathbb{C}^{RP\times 1}$ is the vector composed of the signals received by the R antennas and P chips of the n^{th} symbol period, $\mathbf{G} = [\mathbf{g}^{(1,1)}\, \mathbf{g}^{(2,1)} \ldots \mathbf{g}^{(R,1)} \;\cdots\; \mathbf{g}^{(1,P)}\, \mathbf{g}^{(2,P)} \ldots \mathbf{g}^{(R,P)}]^T \in \mathbb{C}^{RP\times Q}$ is the channel matrix, with $\mathbf{g}^{(r,p)} = [g_1^{(r,p)}\, g_2^{(r,p)} \ldots g_Q^{(r,p)}]^T \in \mathbb{C}^{Q\times 1}$ containing the Volterra kernel coefficients $g_{2k+1}^{(r,p)}(t_1,\ldots,t_{2k+1})$ of the $((p-1)R+r)^{th}$ sub-channel and $\check{\mathbf{v}}(n) = [v_{1,n,1}\, v_{2,n,1} \ldots v_{R,n,1} \;\cdots\; v_{1,n,P}\, v_{2,n,P} \ldots v_{R,n,P}]^T \in \mathbb{C}^{RP\times 1}$, with Q being defined as in (6.12).

6.3.2 Third-Order Tensor of Covariances

The proposed tensor-based channel estimation method is based on the fact that (6.63) is equivalent to the memoryless MIMO Volterra channel (5.3). However, in (6.63), the channel $g_{2k+1}^{(r,p)}(t_1,\ldots,t_{2k+1})$ has RP outputs, while in (5.3), the channel $h_{2k+1}^{(r)}(t_1,\ldots,t_{2k+1})$ has R outputs. The covariance matrix of $\check{\mathbf{y}}(n)$ can then be expressed as:

$$\mathbf{R}_{\check{\mathbf{y}}}(d) = \mathbb{E}\left[\check{\mathbf{y}}(n+d)\check{\mathbf{y}}^H(n)\right] = \mathbf{G}\mathbf{R}_{\tilde{\mathbf{s}}}(d)\mathbf{G}^H \in \mathbb{C}^{RP\times RP}, \tag{6.64}$$

for $0 \leq d \leq D-1$, where $\mathbf{R}_{\tilde{\mathbf{s}}}(d)$ is given by (6.34).

A third-order tensor $\mathcal{R} \in \mathbb{C}^{D\times RP\times RP}$ composed of received signal covariances can be defined similarly as in Chapter 5. Let $[\mathcal{R}]_{d+1,i_1,i_2} = [\mathbf{R}_y(d)]_{i_1,i_2}$, for $0 \leq d \leq D-1$ and $1 \leq i_1, i_2 \leq RP$. Assuming that the information signals are generated using the precoding scheme developed in Chapter 5, a typical element of $\mathcal{R}$ can be

expressed by:

$$r_{(d+1),i_1,i_2} = \sum_{q=1}^{Q} g_{i_1,q} g^*_{i_2,q} z_{d,q}, \tag{6.65}$$

where $r_{(d+1),i_1,i_2} = [\mathcal{R}]_{d+1,i_1,i_2}$, $z_{d+1,q} = [\mathbf{Z}]_{d+1,q} = [\mathbf{R}_{\tilde{\mathbf{s}}}(d)]_{q,q}$ and $g_{i,q} = [\mathbf{G}]_{i,q} = g_q^{(r,p)}$, with $i = ((p-1)R + r)$. Note that (6.65) corresponds to the PARAFAC decomposition of the tensor $\mathcal{R}$ with rank $\leq Q$ and factor matrices equal to $\mathbf{G}$, $\mathbf{G}^*$ and $\mathbf{Z}$. Hence, the slice matrices of the tensor $\mathcal{R}$ are given by:

$$\mathbf{R}_{(d+1)\,\cdot\cdot} = \mathbf{G}\,\mathrm{diag}_{d+1}[\mathbf{Z}\,]\mathbf{G}^H \in \mathbb{C}^{RP\times RP}, \tag{6.66}$$

$$\mathbf{R}_{\cdot r_1 \cdot} = \mathbf{G}^* \mathrm{diag}_{r_1}[\mathbf{G}]\mathbf{Z}^T \in \mathbb{C}^{RP\times D}, \tag{6.67}$$

$$\mathbf{R}_{\cdot\,\cdot\,r_2} = \mathbf{Z}\,\mathrm{diag}_{r_2}[\mathbf{G}^*]\mathbf{G}^T \in \mathbb{C}^{D\times RP}. \tag{6.68}$$

Thus, defining the unfolding matrices of $\mathcal{R}$ as:

$$\mathbf{R}_{[1]} \equiv \begin{bmatrix} \mathbf{R}_{1\cdot\cdot} \\ \vdots \\ \mathbf{R}_{D\cdot\cdot} \end{bmatrix}, \qquad \mathbf{R}_{[2]} \equiv \begin{bmatrix} \mathbf{R}_{\cdot 1\cdot} \\ \vdots \\ \mathbf{R}_{\cdot RP\cdot} \end{bmatrix}, \qquad \mathbf{R}_{[3]} \equiv \begin{bmatrix} \mathbf{R}_{\cdot\cdot 1} \\ \vdots \\ \mathbf{R}_{\cdot\cdot RP} \end{bmatrix}, \tag{6.69}$$

we get:

$$\mathbf{R}_{[1]} = (\mathbf{Z} \diamond \mathbf{G})\,\mathbf{G}^H \in \mathbb{C}^{RPD\times RP}, \tag{6.70}$$

$$\mathbf{R}_{[2]} = (\mathbf{G} \diamond \mathbf{G}^*)\,\mathbf{Z}^T \in \mathbb{C}^{R^2P^2\times D}, \tag{6.71}$$

$$\mathbf{R}_{[3]} = (\mathbf{G}^* \diamond \mathbf{Z})\,\mathbf{G}^T \in \mathbb{C}^{RPD\times RP}. \tag{6.72}$$

It is important to note that, in the case of a memoryless channel ($M = 0$), equation (6.61) becomes:

$$\bar{g}^{(r,p)}_{2k+1}(t_1,\ldots,t_{2k+1}) = h^{(r)}_{2k+1}(t_1,\ldots,t_{2k+1},0,\ldots,0)\prod_{i=1}^{k+1} c_{t_i}(p)\prod_{i=k+2}^{2k+1} c^*_{t_i}(p). \tag{6.73}$$

As expected, the contributions of the channel coefficients $h^{(r)}_{2k+1}(\cdot)$ and spreading codes $c_t(p)$ in the Volterra kernel coefficients $\bar{g}^{(r,p)}_{2k+1}(\cdot)$ can be decoupled, which allowed us to construct a fifth-order PARAFAC tensor of channel output covariances in Section 6.2.

6.3.3 Channel Estimation

The estimation algorithms used for identifying the channel matrix $\mathbf{G}$ are the same as in Section 6.2, that is: (i) the two-steps ALS algorithm, (ii) the EVD-LS algorithm and (iii) the Single-LS algorithm. They are summarized respectively in Tables 6.7, 6.8 and 6.9, where $\hat{\mathbf{G}}_{ab}^{(it)} = 0.5[\hat{\mathbf{G}}_{a}^{(it)} + (\hat{\mathbf{G}}_{b}^{(it)})^*]$ and $\mathbf{W} = (\mathbf{G} \diamond \mathbf{G}^*) \in \mathbb{C}^{R^2P^2 \times Q}$. These algorithms assume that the matrix $\mathbf{Z}$ is known. Moreover, all these algorithms provide an estimate of the channel matrix $\mathbf{G}$ up to a diagonal matrix. As well as in Section 6.2, this scaling ambiguity does not represent an effective problem, as it can be removed by a gain control at the receiver or using a differential modulation.

The Kruskal sufficient condition for the essential uniqueness of the PARAFAC decomposition of $\mathcal{R}$ is given by:

$$2k_{\mathbf{G}} + k_{\mathbf{Z}} \geq 2Q + 2. \tag{6.74}$$

Assuming that the matrix factors are full k-rank, that leads to:

$$2\min(RP, Q) + \min(D, Q) \geq 2Q + 2. \tag{6.75}$$

Note that this uniqueness condition allows working with R, P and $D < Q$.

Similarly as in Section 6.2, a sufficient identifiability condition for EVD-LS and Single-LS algorithms is $r_{\mathbf{Z}} = Q$, i.e. the matrix $\mathbf{Z}$ is full column rank or, equivalently, $\mathbf{Z}$ has a left inverse. That implies $D \geq Q$, which means that these two algorithms do not impose constraints on R and P.

6.4 Simulation Results

In this section, the proposed channel estimation and equalization methods are evaluated by means of simulations. A linear-cubic MIMO Volterra system corresponding to a MIMO Wiener channel of an uplink radio over fiber multiuser communication system [114, 44] is considered for the simulations. In Sections 6.4.1, 6.4.2 and 6.4.3, the $R \times T$ wireless link, corresponding to R receive antennas and T users, has a frequency flat fading. In this case, the wireless link is modeled as a memoryless $R \times T$ linear mixer. In Section 6.4.4, the wireless link has a frequency selective fading and is modeled as a convolutive $R \times T$ linear mixer with memory $M = 1$ chip period and spreading gain $P = 3$. The electrical-optical (E/O) conversion in each antenna is modeled by the following polynomial $f_1 x + f_3 |x|^2 x$, with

Table 6.7: ALS algorithm - short memory channel

Initialization: $\hat{\mathbf{G}}_a^{(0)} \rightarrow RP \times Q$ random matrix
Iterations $(it = it + 1)$ **:** 1) $\hat{\mathbf{G}}_b^{(it)} = \left[\left(\mathbf{Z} \diamond \hat{\mathbf{G}}_a^{(it-1)}\right)^{\dagger} \hat{\mathbf{R}}_{[1]}\right]^T$ 2) $\hat{\mathbf{G}}_a^{(it)} = \left[\left(\hat{\mathbf{G}}_b^{(it)} \diamond \mathbf{Z}\right)^{\dagger} \hat{\mathbf{R}}_{[3]}\right]^T$
Stop Criteria: $\frac{\left\|\hat{\mathbf{G}}_{ab}^{(it)} - \hat{\mathbf{G}}_{ab}^{(it-1)}\right\|_F^2}{\left\|\hat{\mathbf{G}}_{ab}^{(it-1)}\right\|_F^2} < \epsilon$

$f_1 = 1$ and $f_3 = -0.35$ [114, 116]. The results were obtained via Monte Carlo simulations using at least 100 independent data realizations and complex-valued Walsh-Hadamard spreading codes [62].

6.4.1 Memoryless channels: deterministic approach

In this section, the performance of the proposed deterministic tensor-based techniques for joint channel estimation and equalization are evaluated by means of simulation results. The channel equalization is evaluated by means of the Bit-Error-Rate (BER) and the channel estimation by means of the Normalized Mean Square Error (NMSE) of the estimated channel parameters, defined as:

$$NMSE = \frac{1}{N_R} \sum_{l=1}^{N_R} \frac{\| \mathbf{H} - \hat{\mathbf{H}}_l \|_F^2}{\| \mathbf{H} \|_F^2}, \tag{6.76}$$

where $\hat{\mathbf{H}}_l$ represents the channel matrix estimated at the l^{th} Monte Carlo simulation. All the simulations concerning deterministic tensor-based techniques are

Table 6.8: EVD-LS algorithm - short memory channel

1) $\hat{\mathbf{W}} = \hat{\mathbf{R}}_{[2]} \left(\mathbf{Z}^T\right)^{\dagger}$, where $\mathbf{W} = (\mathbf{G} \diamond \mathbf{G}^*) \in \mathbb{C}^{R^2P^2 \times Q}$.

2) For $q = 1, ..., Q$: Construct

$$\hat{\mathbf{W}}(q) = \mathbf{unvec}(\hat{\mathbf{W}}_{\cdot\, q}) = \mathbf{unvec}(\hat{\mathbf{G}}_{\cdot\, q} \diamond \hat{\mathbf{G}}^*_{\cdot\, q}) = \hat{\mathbf{G}}^*_{\cdot\, q} \hat{\mathbf{G}}^T_{\cdot\, q},$$

where $\hat{\mathbf{W}}_{\cdot\, q}$ and $\hat{\mathbf{G}}_{\cdot\, q}$ denote the q^{th} column of $\hat{\mathbf{W}}$ and $\hat{\mathbf{G}}$ respectively, and the operator $\mathbf{unvec}\,(\cdot)$ forms a $RP \times RP$ matrix from its vector argument.

3) Calculate $\hat{\mathbf{G}}_{\cdot q}$ as the conjugate of the eigenvector associated with the largest eigenvalue of $\hat{\mathbf{W}}(q)$.

Table 6.9: Single-LS algorithm - short memory channel

1) $\hat{\mathbf{W}} = \hat{\mathbf{R}}_{[2]} \left(\mathbf{Z}^T\right)^{\dagger}$, where $\mathbf{W} = (\mathbf{G} \diamond \mathbf{G}^*) \in \mathbb{C}^{R^2P^2 \times Q}$.

2) For $i = 1, ..., RP$: Construct $\hat{\mathbf{W}}^{(i)} = \begin{bmatrix} \hat{\mathbf{W}}_{(i-1)RP+1,\cdot} \\ \vdots \\ \hat{\mathbf{W}}_{iRP,\cdot} \end{bmatrix}$, where

$\hat{\mathbf{W}}_{i,\cdot}$ denotes the i^{rh} line of $\hat{\mathbf{W}}$.

3) $\hat{\mathbf{G}} = \frac{1}{RP} \sum_{i=1}^{RP} \left[\hat{\mathbf{W}}^{(i)}\right]^*$.

obtained with 4-PSK transmitted signals.

Fig. 6.1 shows the NMSE versus Signal to Noise Ratio (SNR) provided by the ALS and ALS-DD-BI techniques for $N = 32$, $N_t = 4$, $P = 3$ $R = 3$ and $T = 2$. For comparison, it is also shown the NMSE provided by the ALS algorithm with $N = 8$ and by the ALS-DD algorithm with $N = 32$. This algorithm corresponds to the ALS-DD-BI without the block-initialization (a random initialization is used in this case). From this figure, it can be concluded that the NMSE provided by the

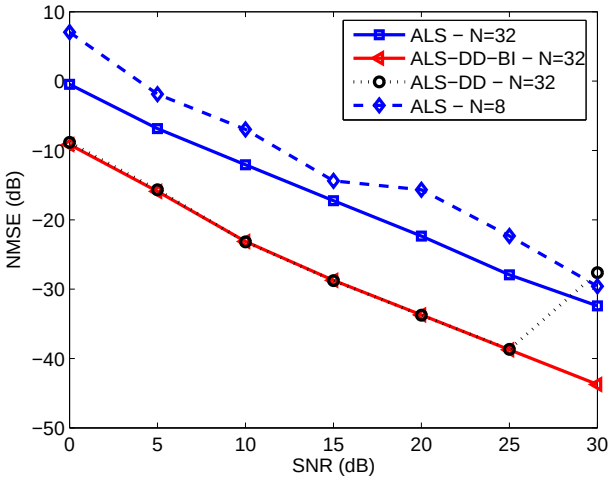

Figure 6.1: NMSE versus SNR provided by the deterministic tensor-based techniques.

ALS-DD-BI algorithm is significantly smaller than the one obtained with the ALS algorithm. Moreover, it can be remarked that the proposed ALS method is able to blindly estimate the channel using only 8 symbols. In fact, only 1 known pilot symbol is used to remove the scaling ambiguity. Note also that the performance of the ALS-DD technique is different to that of the ALS-DD-BI when the SNR is equal to 30dB. This is certainly due to the fact that the ALS-DD-BI is not monotonically convergent, contrarily to the ALS algorithm [18]. This means that the use of direct decisions may deteriorate the convergence of the ALS algorithm. However, as it can be viewed in this figure, the use of a block-initialization seems to overcome this problem.

Fig. 6.2 shows the number of iterations needed to achieve the convergence versus SNR for the ALS and ALS-DD-BI algorithms with $N = 32$, $N_t = 4$, $P = 3$, $R = 3$ and $T = 2$, and for the ALS algorithm with $N = 8$. It can be remarked that the ALS-DD-BI technique converges more quickly than the ALS algorithm in most of the cases. Note also that the ALS-DD-BI algorithm converges after approximately 2 iterations when the SNR is higher that 15dB.

Fig. 6.3 shows the BER versus SNR provided by the ALS and ALS-DD-BI algorithms with $N = 32$, $N_t = 4$, $P = 3$, $R = 3$ and $T = 2$, and by the ALS algorithm with $N = 8$. It is also shown the BER provided by the zero forcing (ZF) receiver

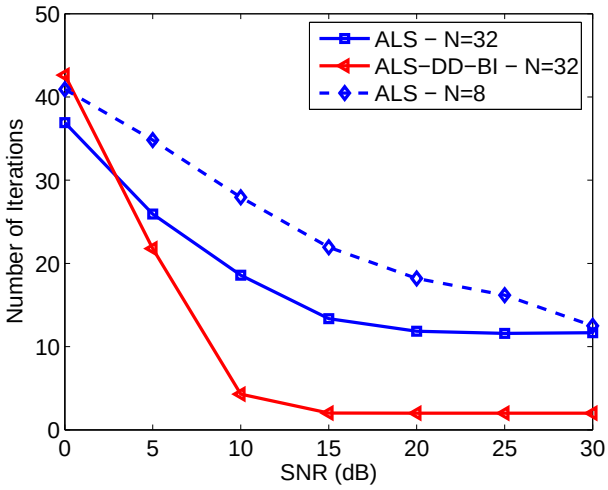

Figure 6.2: Number of iterations needed to achieve the convergence versus SNR for the deterministic tensor-based techniques.

assuming the channel knowledge:

$$\hat{\tilde{\mathbf{S}}} = \left[\left(\tilde{\mathbf{C}} \diamond \mathbf{H}\right)^{\dagger} \bar{\mathbf{Y}}_{[3]}\right]^{T}. \tag{6.77}$$

Note that (6.77) corresponds to the first step of the ALS algorithm with a known channel. The conclusions that we can draw from Fig. 6.3 are similar to those of Fig. 6.1, the ALS-DD-BI algorithm providing a BER smaller than that of the ALS algorithm and close to that of the ZF receiver with a known channel.

6.4.2 Memoryless channels: stochastic approach

In this section, the proposed stochastic tensor-based techniques for estimating memoryless MIMO Volterra channels are evaluated by means of simulations. Fig. 6.4 shows the NMSE versus SNR provided by the ALS, EVD-LS and Single-LS algorithms for $N = 256$, $D = 4$, $P = 3$, $R = 3$ and $T = 2$, using Configurations A and B of Table 5.4. It can be concluded from this figure that Configuration B provides a better performance than Configuration A for the ALS and EVD-LS algorithms, as well as in Chapter 5. It can also be viewed that the NMSE provided by the EVD-LS is a little smaller than the one obtained with the ALS and much smaller than the one obtained with the Single-LS. However, it should be highlighted that the Single-LS algorithm has a computational cost significantly smaller than the other two techniques. For instance, in Fig. 6.4, when Configuration B is

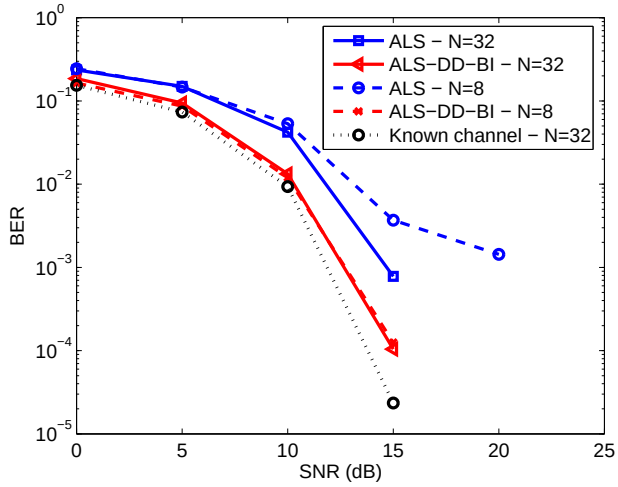

Figure 6.3: BER versus SNR provided by the deterministic tensor-based techniques.

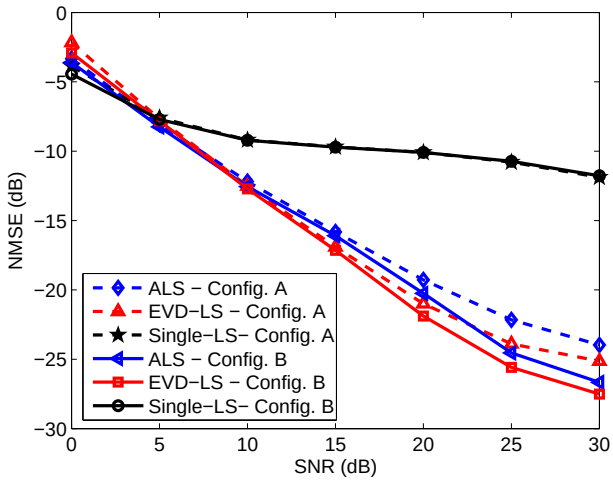

Figure 6.4: NMSE versus SNR provided by the stochastic tensor-based techniques.

used and the SNR is equal to 0dB, the ALS algorithm needs approximatively 15 iterations to converge, with two LS estimate computations per step, while the EVD-LS and Single-LS algorithms computes respectively 5 and 1 LS estimates.

The next figure evaluates the influence of spreading gain P and number of covariance delays D on the channel estimation accuracy. Fig. 6.5 shows the NMSE versus D provided by the ALS algorithm for $P = 1, 2$ and 3, with $N = 256$,

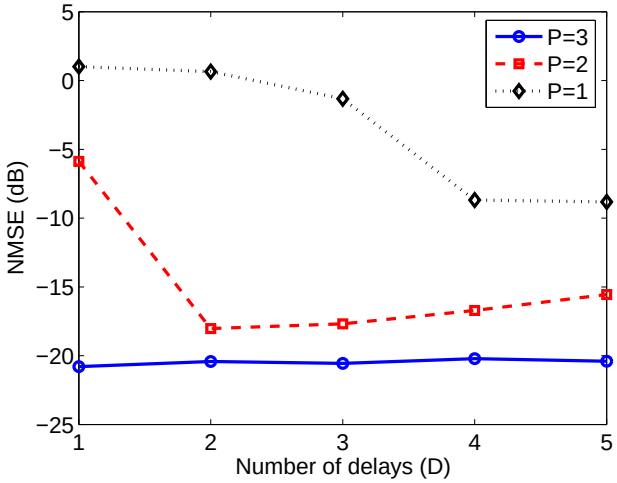

Figure 6.5: NMSE versus the number of covariance delays D provided by the ALS algorithm.

$SNR = 20dB$, $R = 3$ and $T = 2$. Note that the case $P = 1$ corresponds to the ALS estimation technique developed in Chapter 5. When $P = 3$, we can see that the accuracy of the ALS is quite similar for all the tested values of D. Indeed, for $P = R = 3$, the uniqueness condition (6.47) becomes $D \geq 1$. In this case, the use of the precoding scheme is not very attractive. However, it can be viewed in Fig. 6.5 that, for $P = 2$ and $P = 1$, the ALS does not work well for $D = 1$. Indeed, in these cases, condition (6.47) becomes $D \geq 2$ and $D \geq 4$, respectively. It can then be concluded that the use of the precoding scheme is mandatory in such cases.

Fig. 6.6 shows the BER versus SNR provided by the following Minimum Mean Square Error (MMSE) receiver based on (6.24):

$$\hat{\mathbf{W}}_{MMSE} = \mathbf{R}_{\tilde{\mathbf{s}}}(0)\left(\tilde{\mathbf{C}} \diamond \hat{\mathbf{H}}\right)^H \left[\left(\tilde{\mathbf{C}} \diamond \hat{\mathbf{H}}\right)\mathbf{R}_{\tilde{\mathbf{s}}}(0)\left(\tilde{\mathbf{C}} \diamond \hat{\mathbf{H}}\right)^H + \sigma^2\mathbf{I}_{RP}\right]^{-1} \in \mathbb{C}^{Q\times RP}, \quad (6.78)$$

using ALS, EVD-LS and Single-LS channel estimates obtained with Configuration A, and ALS channel estimates obtained with Configuration B, for $N = 256$, $D = 4$, $P = 3$, $R = 3$ and $T = 2$. For comparison, it is also plotted the BER provided by the MMSE receiver assuming an exact knowledge of the channel, using Configuration A. Two main remarks should be highlighted from this figure. The first one is that, by comparing the BER curves obtained with the ALS, one can see that Configuration A performs better than Configuration B, as well as in Chapter 5. The second one is that the BERs provided by the ALS and EVD-LS are very close to that of the MMSE receiver with the known channel.

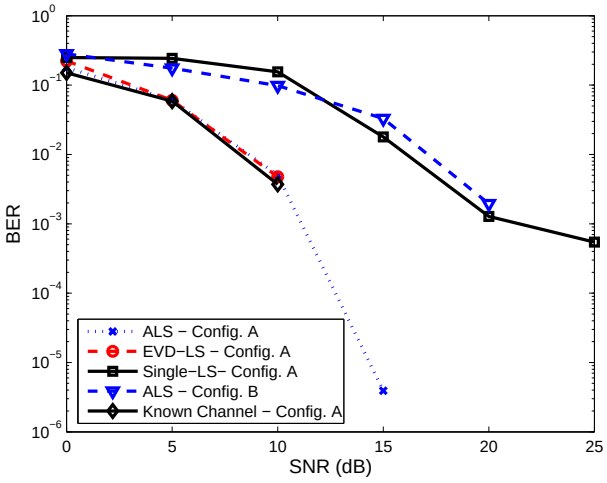

Figure 6.6: BER versus SNR provided by the stochastic tensor-based techniques.

6.4.3 Comparison between deterministic and stochastic approaches

This section provides a comparison between the proposed deterministic and stochastic tensor-based methods by means of simulations. Three techniques are compared: the ALS with the deterministic tensor of received signals using 4-PSK transmitted signals and the ALS with the stochastic tensor of received signal covariances using Configurations A and C of Table 5.4. We recall that code rates associated with Configurations A and C are respectively 1/3 and 2/3. Thus, as these configurations use 8-PSK signals, the transmission rate provided by Configuration C is the same as the one of a 4-PSK signal, while the transmission rate provided by Configuration A is the half of the one associated with a 4-PSK signal. Figs. 6.7, 6.8 and 6.9 show respectively the NMSE, number of iterations needed to achieve the convergence and BER versus SNR provided by these techniques, for $N = 128$, $D = 4$, $P = 3$, $R = 3$ and $T = 2$. The BERs associated with the stochastic tensor-based methods were calculated using the MMSE receiver (6.78).

It can be viewed from Fig. 6.7 that, for low SNRs, the NMSE provided by the stochastic techniques is a little better than the one obtained with the deterministic technique. However, for high SNRs, the deterministic technique provides smaller NMSEs than other methods. That is due to the fact that the noise is main source of performance degradation of the deterministic technique. On the other hand, in the case of the stochastic techniques, the errors on the estimation of the covariances

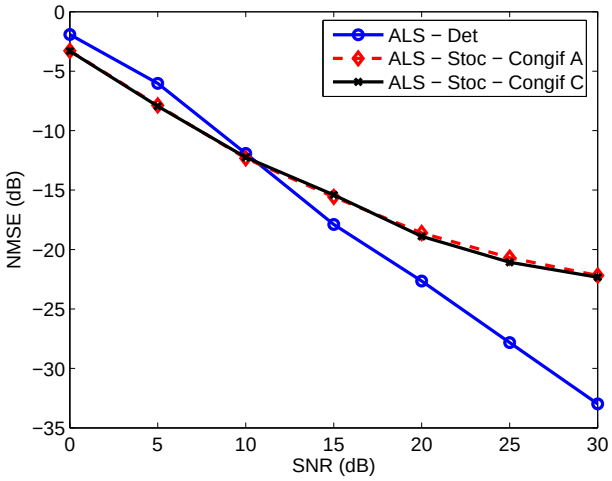

Figure 6.7: NMSE versus SNR provided by the deterministic and stochastic tensor-based techniques.

are main source of performance degradation. Indeed, in our simulations, we found that the performance of the stochastic techniques can be significantly improved if a higher number of symbols N is used, which is not the case for the deterministic technique.

From Fig. 6.8, it can be concluded that the stochastic techniques need a smaller number of iterations to converge than the deterministic one. Moreover, it should be highlighted that, in this case, each iteration of the deterministic ALS algorithm computes the pseudo-inverse of 9×4 and 384×4 matrices, while the stochastic ALS algorithm computes, in each iteration, the pseudo-inverse of two 108×4 matrices. Thus, it can be concluded that, for the case treated in Fig. 6.8, the computational cost of the stochastic techniques is significantly smaller than the one of the deterministic technique.

From Fig. 6.9, we can see that, when Configuration A is used, the stochastic technique performs better than the deterministic one. However, as mentioned earlier, the transmission rate provided by Configuration A is twice smaller than the one of a 4-PSK signal. On the other hand, when Configuration C is used, the deterministic technique performs better than the stochastic one. In this case, both techniques have the same transmission rate.

Thus, we can conclude that the use of deterministic techniques is more interesting when we have to use small blocks of data (small N). However, if a high value of N can be used, then the stochastic techniques may be an interesting choice.

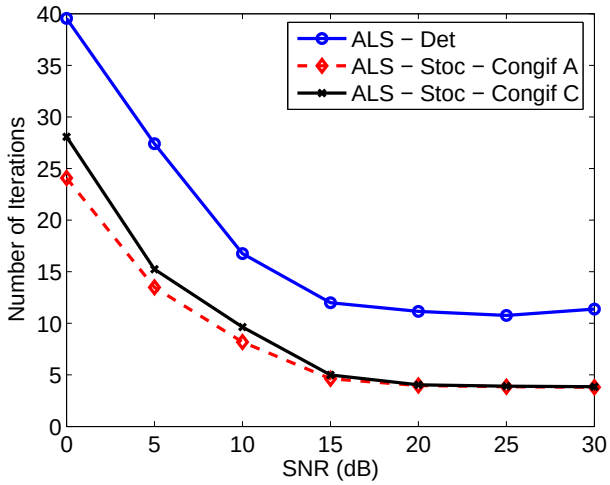

Figure 6.8: Number of iterations needed to achieve the convergence versus SNR provided by the deterministic and stochastic tensor-based techniques.

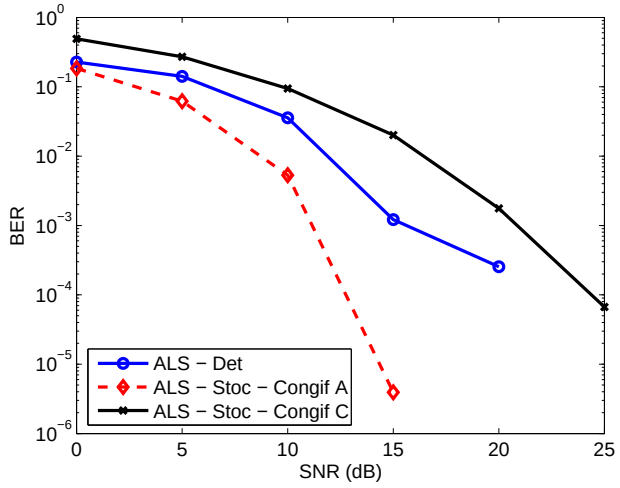

Figure 6.9: BER versus SNR provided by the deterministic and stochastic tensor-based techniques.

6.4.4 Short memory channels: stochastic approach

In this section, the proposed stochastic tensor-based techniques for estimating MIMO Volterra channels with short memory are evaluated by means of simulations. The channel equalization is evaluated by means of the NMSE of the

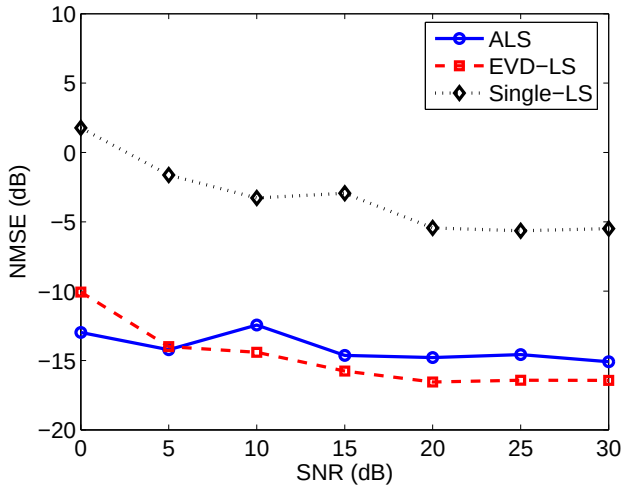

Figure 6.10: NMSE versus SNR provided by the stochastic tensor-based techniques - channel with short memory

estimated equivalent memoryless channel parameters, defined as:

$$NMSE = \frac{1}{N_R}\sum_{l=1}^{N_R}\frac{\parallel \mathbf{G} - \hat{\mathbf{G}}_l \parallel_F^2}{\parallel \mathbf{H} \parallel_F^2}, \tag{6.79}$$

where $\hat{\mathbf{G}}_l$ represents the channel matrix estimated at the l^{th} Monte Carlo simulation. Fig. 6.10 shows the NMSE versus SNR provided by the ALS, EVD-LS and Single-LS algorithms for $N = 256$, $D = 4$, $P = 3$, $R = 3$ and $T = 2$, using Configuration B of Table 5.4. As well as in Fig. 6.4, the performance of the EVD-LS and ALS algorithms are close, the Single-LS providing worse NMSEs. Once again, it should be highlighted that the Single-LS has a computational cost significantly smaller than the other two algorithms. In this case, when the SNR is equal to 0dB, the ALS algorithm needs approximatively 15 iterations to converge.

Fig. 6.11 evaluates the influence of number of receive antennas R and length of the data block N on the BER. It shows the BER versus SNR provided by the MMSE receiver:

$$\hat{\mathbf{W}}_{MMSE} = \mathbf{R}_{\tilde{\mathbf{s}}}(0)\hat{\mathbf{G}}^H\left[\hat{\mathbf{G}}\mathbf{R}_{\tilde{\mathbf{s}}}(0)\hat{\mathbf{G}}^H + \sigma^2\mathbf{I}_{RP}\right]^{-1} \in \mathbb{C}^{Q\times RP}, \tag{6.80}$$

using ALS channel estimates obtained with Configuration B, for $N = 256$ and $N = 1024$, $R = 2$ and $R = 3$, with $D = 4$, $P = 3$ and $T = 2$. For comparison, it is also shown the BER provided by the MMSE receiver assuming that the channel

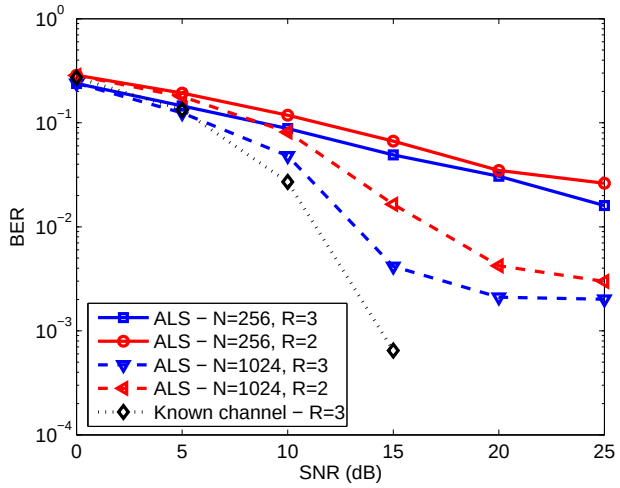

Figure 6.11: NMSE versus SNR provided by the deterministic and stochastic tensor-based techniques - channel with short memory.

is known, with $R = 3$. It can be concluded from this figure that the BER can be significantly improve when N is increased. In particular, when $N = 1024$ and the SNR is low, the BER provided by the proposed ALS method is relatively close to that of the MMSE receiver with the known channel. Note also that the cases corresponding to $R = 3$ provide little performance gains with respect to the cases where $R = 2$.

6.5 Conclusion

This chapter has addressed the problem of blind estimation and equalization of Volterra communication channels in a multiuser CDMA environment. The proposed techniques are based on tensor decompositions, the key aspect of them being the fact that the spreading codes induce a new diversity on the received signals. The main advantage of these tensor-based approaches is that they allow for a great flexibility on the number of antennas and spreading factor.

The theoretical contributions of the chapter are divided in three parts. In the first part, a method for joint channel estimation and equalization is developed based on the PARAFAC decomposition of a third-order tensor composed of received signals, with two algorithms being considered for carrying out the tensor decomposition. The first one is the classical ALS algorithm, while the second one, called the ALS-DD-BI algorithm, includes a block initialization and a decision device, which

implies a significant performance improvement in terms of channel estimation and BER. In fact, the approach developed in this section can be viewed as an extension of [144] to nonlinear channels.

In the second part of the chapter, MIMO Volterra channel estimation techniques are developed based on the PARAFAC decomposition of a fifth-order tensor composed of covariances of the received signals, assuming that the transmitted signals have a PSK modulation. In particular, we should highlight that the EVD-LS algorithm has performed similarly as the ALS algorithm, with a smaller computational cost. In the third part of the chapter, we have proposed a new method for estimating MIMO Volterra communication channels with short memory, based on the PARAFAC decomposition of a third-order tensor composed of channel output covariances. In this case, the spreading codes are assumed to contain guard chips and transmitted signals are generated using the precoding scheme developed in Chapter 5. In fact, this approach can be considered as an extension of the one developed in Chapter 5 for channels with short memory.

The proposed techniques were applied to an uplink channel of a nonlinear ROF-CDMA multiuser communication system. Some simulation results have illustrated the good performance of these algorithms. In general, the main advantage of the deterministic approach is that it allowed joint blind channel estimation and equalization with a small number of symbols. On the other hand, the main advantage of the stochastic approaches is that they have provided weaker uniqueness conditions and smaller computational cost.

CHAPTER 7

Conclusion

THE main objective of this thesis is to study and develop techniques for channel estimation and information recovery in nonlinear MIMO communication systems based on the use of Volterra models. Many devices in communication systems are potential sources of nonlinearities. In this thesis, we were particularly concerned with nonlinear distortions in MIMO systems due to power amplifiers (PAs) and electrical-optical (E/O) conversion devices. Indeed, we have made use of MIMO Volterra systems to model the channel behavior of three kinds of communication systems: OFDM, ROF-TDMA and ROF-CDMA. The digital signal processing techniques developed through this thesis are designed to eliminate or reduce the effect of such nonlinear distortions. In the sequel, a brief conclusion of each chapter is given:

- In **Chapter 2**, we have provided an overview of MIMO Volterra communication channels. The main system models used through this work are described and relationships between several block structured nonlinear models and the MIMO Volterra system are developed. In fact, these relationships are original contributions of this chapter. Another contribution is the development of general expressions for equivalent baseband discrete-time MIMO Volterra channels. This chapter also presents the applications of MIMO Volterra models in communication systems considered in this thesis.

- In **Chapter 3**, we have proposed techniques for estimation and equalization of MIMO-OFDM channels with nonlinear PAs. The proposed techniques are based on a global channel representation that characterizes the cascade

of the PA and the wireless channel, with two PA models being considered: the memoryless polynomial model and the memory polynomial model. For the case of memoryless PAs, we have developed a supervised technique for estimating the global channel coefficients, its main advantage being that it does not require the knowledge of the PA parameters, contrarily to previous methods. Then, still in the case of memoryless PAs, two channel equalization techniques were developed based on the use of an antenna array at the reception and on a proposed transmission scheme. Another important contribution of this chapter is the demonstration that memoryless and memory polynomial PAs provide identical expressions for the frequency domain received signals in terms of the global channel parameters. Therefore, the techniques developed for memoryless PAs can be applied to the case of memory polynomial PAs. This was possible due to a theorem demonstrated in this chapter, stating that a memory polynomial PA in a OFDM system can be expressed as a memoryless polynomial PA with coefficients that vary from one subcarrier to another.

- **Chapter 4** deals with supervised estimation of MIMO Volterra channels in the case of TDMA-SDMA systems. This chapter considers the most general type of MIMO Volterra systems used in this thesis. The proposed estimation method uses orthonormal polynomials to improve the eigenvalue spread of the covariance matrix of the nonlinear input vector. The developed orthonormalization technique is an extension of existing methods to the case of MIMO Volterra systems, with the property of allowing different PDFs for the input signals and different memories with respect to the inputs. The proposed supervised channel estimation method was applied to the estimation of a multiuser nonlinear ROF channel, its main advantage being the significant improvement of the convergence speed of the LMS algorithm with respect to the case where canonical polynomials are used.

- In **Chapter 5**, two techniques for blind estimation of memoryless MIMO Volterra channels have been proposed for TDMA-SDMA systems. These methods are based on the PARAFAC decomposition of a tensor composed of covariances of the received signals. Such a decomposition is possible owing to a new precoding scheme developed so that the transmitted signals are temporally correlated and satisfy some orthogonality constraints. A great advantage of using the PARAFAC decomposition is that it provides relaxed uniqueness conditions. In fact, the proposed estimation methods exploit the redundancy introduced on the transmitted signal by the precoding. These methods have been applied for estimating an uplink channel in a multiuser ROF communication system, providing good and promising results. Some of

the tested precoding configurations have provided good channel estimates, while some others have shown good robustness to noise and interference. A tradeoff between channel estimation accuracy and robustness to noise and interference must then be taken into account in order to choose the best configuration in terms of signal detection.

- **Chapter 6** proposes estimation and equalization techniques for MIMO Volterra channels in a CDMA communication system. The developed techniques are based on the PARAFAC decomposition, the spreading codes being used to induce a new diversity on the received signals. As well as in Chapter 5, the main advantage of these tensor-based approaches is that they provide relaxed uniqueness conditions. This chapter treated memoryless and short memory channels. In the first case, two PARAFAC-based approaches were developed. The first one allows joint blind channel estimation and information recovery, considering a third-order tensor composed of received signals. The second one allows blind estimation of the considered MIMO Volterra channel by using a fifth-order tensor composed of covariances of the received signals. In the case of short memory channels, an estimation method was developed based on the PARAFAC decomposition of a third-order tensor composed of channel output covariances, the spreading codes containing guard chips and transmitted signals being generated using the precoding scheme developed in Chapter 5. This approach can be viewed as an extension of the one developed in Chapter 5 for channels with short memory. These techniques were applied to an uplink channel of a nonlinear ROF-CDMA multiuser communication system. The principal conclusion drawn from the simulations is that, in the case of memoryless channels, the main advantage of the first approach is that it works with a small number of data symbols. On the other hand, the main advantage of the second approach is that it provides weaker uniqueness conditions and smaller computational cost than the first approach.

It should be highlighted that the proposed techniques use different kinds of approaches, according to the considered application and kind of MIMO Volterra system model. Moreover, it should be remarked that this thesis considers supervised transmission schemes (Chapters 3 and 4), as well as blind and semi-blind scenarios (Chapters 5 and 6). A common point linking the channel estimation techniques developed in Chapters 4 and 5 is that they make use of covariances and are based on methods for orthogonalization of virtual sources. In Chapter 4, the diagonalization of the zero-delay covariance matrix of the nonlinear input vector is carried out in order to improve the convergence speed of the LMS algorithm. In Chapter 5, the diagonalization of covariance matrices of the nonlinear input vector for various delays is performed for satisfying some identifiability conditions.

In fact, the stochastic channel estimation techniques of Chapter 6 also exploit the orthogonality of the virtual sources.

Another common point linking different chapters of this thesis is the use of tensors in Chapters 5 and 6. In these cases, tensor decompositions are used to exploit some kind of redundancy of the received signals. It should also be highlighted that the ZF and MMSE receivers used in Chapter 3 are also considered in Chapters 5 and 6. In all these chapters, once the channel is estimated, ZF and MMSE receivers are used to separate the virtual sources, providing an estimate of the transmitted signals. Furthermore, we should also mention that the transmission schemes of Chapters 3 and 5 are based on the same principle of introducing redundancy on the transmitted signals in order to induce a new diversity on the received signals.

Finally, we remark that, although there is a lack of works dealing with signal processing techniques for nonlinear MIMO communication systems, the techniques developed in the thesis have shown promising theoretical and simulation results.

Perspectives

In what follows, we provide a list of the main perspectives and future works related to this thesis:

Chapter 3

- An interesting topic for a future work is the extension of the results of Section 3.4 to more general PA models. For instance, if the PA is represented by a Volterra model, we believe that the PA model can be rewritten with a much smaller number of virtual sources, the coefficients of which varying from one subcarrier to another, as well as in Theorem 3.1. In this case, a global channel representation would also be possible.

- Another perspective concerning Chapter 3 is the blind estimation and equalization of nonlinear MIMO-OFDM systems using tensor decompositions. For instance, by using (3.66), one could define a tensor composed of received signals with space, frequency and time diversities, i.e. with indices corresponding to receive antenna, sub-carrier and transmission block. In this case, tensor decompositions more complex than the PARAFAC [34] could be used to jointly estimate the channel and transmitted signals.

- A deeper comparative study between the proposed channel estimation and equalization techniques and other methods is considered for a future work.

Chapter 4

- It has been shown that all the coefficients of a Volterra system corresponding to Wiener or Wiener-Hammerstein systems can be calculated uniquely from the diagonal coefficients of the Volterra system [85]. This result was demonstrated for the SISO case, but it can also be demonstrated for the MIMO case in a similar way. In this case, one could develop a technique that uses an orthonormal basis to estimate only the diagonal coefficients of a MIMO Volterra system and, then, the non-diagonal coefficients are calculated using the results of [85]. The resulting technique would have a reduced computational cost as it does not need to estimate all the MIMO Volterra coefficients from the data.

- Concerning the ROF system considered in Chapter 4, an interesting perspective is to assume that the received signals are subject to optical and wireless channel noise. ROF systems with nonlinearities due to electrical-optical (E/O) conversion and PA will be also considered in a future work.

Chapter 5

- The main perspective concerning Chapter 5 is a deeper study about the optimal choice of the transition probability matrices (TPMs). We believe that a criterion based on the entropy of the precoded signals can be a good idea to find the optimal TPMs and, as a consequence, minimize the bit-error-rate (BER) provided by the proposed method.

- The application of the proposed tensor-based blind channel estimation methods to the case of ROF-OFDM systems is to be considered in a future work. Two cases will be considered: ROF systems with nonlinearities due to E/O conversion only and due to E/O conversion and PA.

Chapter 6

- We believe that the approaches developed in Chapter 6 can be extended to the case of CDMA systems with large memory. In this case, based on more general channel propagation models, one could use general tensor decompositions [34] to provide channel estimation and equalization techniques for nonlinear MIMO CDMA systems.

Appendices

APPENDIX A

The Kronecker, truncated Kronecker and Khatri-Rao products

A.1 The Kronecker product

The *Kronecker product* of the matrices $\mathbf{A} \in \mathbb{C}^{L_1 \times L_2}$ and $\mathbf{B} \in \mathbb{C}^{L_3 \times L_4}$ is defined as:

$$\mathbf{A} \otimes \mathbf{B} \equiv \begin{pmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,L_2}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,L_2}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{L_1,1}\mathbf{B} & a_{L_1,2}\mathbf{B} & \cdots & a_{L_1,L_2}\mathbf{B} \end{pmatrix} \in \mathbb{C}^{L_1 L_3 \times L_2 L_4}, \tag{A.1}$$

where $a_{i,j}$ are the elements of $\mathbf{A}$.

In particular, the Kronecker product of a vector $\mathbf{a} \in \mathbb{C}^{L \times 1}$ by itself is given by:

$$\otimes^2\mathbf{a} \equiv \mathbf{a} \otimes \mathbf{a} = \begin{pmatrix} a_1\mathbf{a} \\ a_2\mathbf{a} \\ \vdots \\ a_L\mathbf{a} \end{pmatrix} \in \mathbb{C}^{L^2 \times 1}, \tag{A.2}$$

where a_i are the elements of the vector $\mathbf{a}$. Note that the $[\otimes^2\mathbf{a}]_q = a_i a_j$, with $q = (i-1)L + j$ and $1 \leq i, j \leq L$, where $[\otimes^2\mathbf{a}]_q$ denotes the q^{th}entry of the vector $\otimes^2\mathbf{a}$. The N^{th}-order power of the Kronecker product of the vector $\mathbf{a}$ by itself is

defined in a similar way: $\otimes^N \mathbf{a} \equiv \mathbf{a} \otimes \cdots \otimes \mathbf{a} \quad \mathbb{C}^{L^N \times 1}$ ($N-1$ times the operator $\otimes$). In this case, we have $[\otimes^N \mathbf{a}]_q = a_{i_1} a_{i_2} \dots a_{i_N}$, with

$$q = i_N + (i_{N-1}-1)L + \cdots + (i_2-1)L^{N-2} + (i_1-1)L^{N-1} = 1 + \sum_{n=1}^{N} (i_n-1)L^{N-n} \quad \text{(A.3)}$$

and $1 \leq i_1, ..., i_N \leq L$.

A.2 The truncated Kronecker product

The *truncated Kronecker product* of the vector $\mathbf{a}$ by itself is defined in the following way:

$$\oslash^2 \mathbf{a} \equiv \mathbf{a} \oslash \mathbf{a} \equiv \begin{pmatrix} a_1 \bar{\mathbf{a}}_1 \\ a_2 \bar{\mathbf{a}}_2 \\ \vdots \\ a_{L-1} \bar{\mathbf{a}}_{L-1} \\ a_L^2 \end{pmatrix} \quad \in \mathbb{C}^{\frac{L(L+1)}{2} \times 1}, \quad \text{(A.4)}$$

where $\bar{\mathbf{a}}_i = [a_i \; a_{i+1} \dots a_L]^T$. The truncated Kronecker product does not consider the redundant terms that are present in the Kronecker product of a vector by itself, which means that the vector $\oslash^2 \mathbf{a}$ does not contain repeated components.

It is also possible to define the N^{th}-order power of the truncated Kronecker product of a vector $\mathbf{a}$, denoted $\oslash^N \mathbf{a} \equiv \mathbf{a} \oslash \cdots \oslash \mathbf{a}$ ($N-1$ times the operator $\oslash$), by means of the following recursion:

$$\oslash^N \mathbf{a} \equiv \begin{pmatrix} a_1 \oslash^{N-1} \bar{\mathbf{a}}_1 \\ a_2 \oslash^{N-1} \bar{\mathbf{a}}_2 \\ \vdots \\ a_{L-1} \oslash^{N-1} \bar{\mathbf{a}}_{L-1} \\ a_L^k \end{pmatrix}, \quad \text{(A.5)}$$

with $\oslash^1 \mathbf{a} = \mathbf{a}$. The vector $\oslash^N \mathbf{a}$ contains all the N^{th}-order products of the elements of $\mathbf{a}$, with no repeated terms. The dimension of the vector $\oslash^N \mathbf{a}$ is given by the number of subsets of cardinality N with elements taken from a set of cardinality L, i.e. the number of combinations with repetition of N elements drawn from a

set of cardinality L:

$$C_{L,N} = C_N^{L+N-1} = \frac{(L+N-1)!}{(L-1)!N!}, \tag{A.6}$$

where C_N^{L+N-1} denotes the number of combinations without repetition of N elements drawn from a set of cardinality $L+N-1$.

For instance, the vector $\oslash^3\mathbf{a}$, for $L=3$, is given by:

$$\oslash^3\mathbf{a} = \begin{pmatrix} a_1^3 \\ a_1^2 a_2 \\ a_1^2 a_3 \\ a_1 a_2^2 \\ a_1 a_2 a_3 \\ a_1 a_3^2 \\ a_2^3 \\ a_2^2 a_3 \\ a_2 a_3^2 \\ a_3^3 \end{pmatrix}, \tag{A.7}$$

with $C_{3,3} = 10$.

A.3 The Khatri-Rao product

The Khatri-Rao (column-wise Kronecker) product of the matrices $\mathbf{A} \in \mathbb{C}^{L_1 \times L}$ and $\mathbf{B} \in \mathbb{C}^{L_2 \times L}$ is defined as:

$$\mathbf{A} \diamond \mathbf{B} \equiv [(\mathbf{a}_1 \otimes \mathbf{b}_1)\ (\mathbf{a}_2 \otimes \mathbf{b}_2)\ \cdots\ (\mathbf{a}_L \otimes \mathbf{b}_L)] \quad \in \mathbb{C}^{L_1 L_2 \times L}, \tag{A.8}$$

where $\mathbf{a}_l$ and $\mathbf{b}_l$ $(1 \le l \le L)$ denote the l^{th} column of $\mathbf{A}$ and $\mathbf{B}$, respectively. The Khatri-Rao product of $\mathbf{A}$ and $\mathbf{B}$ can also be expressed as:

$$\mathbf{A} \diamond \mathbf{B} = \begin{pmatrix} \mathbf{B}\,\mathrm{diag}_1[\mathbf{A}\,] \\ \vdots \\ \mathbf{B}\,\mathrm{diag}_{L_1}[\mathbf{A}\,] \end{pmatrix}, \tag{A.9}$$

here $\mathrm{diag}_l[\mathbf{A}]$ denotes the diagonal matrix formed from the l^{th} row of $\mathbf{A}$.

An important relationship concerning the Khatri-Rao product is given by:

$$(\mathbf{A}_1\mathbf{A}_2) \diamond (\mathbf{B}_1\mathbf{B}_2) = (\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \diamond \mathbf{B}_2), \tag{A.10}$$

where $\mathbf{A}_1 \in \mathbb{C}^{L_1 \times L_2}$, $\mathbf{A}_2 \in \mathbb{C}^{L_2 \times L}$, $\mathbf{B}_1 \in \mathbb{C}^{L_3 \times L_4}$ and $\mathbf{B}_2 \in \mathbb{C}^{L_4 \times L}$.

APPENDIX B

Orthonormal Monomials

The set of orthonormal monomials $P_{\alpha,\beta}(y)$ used in Chapter 4 are constructed by applying the Gram-Schmidt formula to the set of canonical monomials $T_{\alpha,\beta}(y) = y^{\alpha}y^{*\beta}$, for $0 \leq k \leq K$, $0 \leq \alpha \leq k+1$ and $0 \leq \beta \leq k$, as in [84]. The Gram-Schmidt procedure assumes a pre-specified order of the canonical monomials. Let $\mathcal{T}$ be an ordered set of pairs (α, β) associated with canonical monomials $T_{\alpha,\beta}(y) = y^{\alpha}y^{*\beta}$, for $0 \leq k \leq K$, $0 \leq \alpha \leq k+1$ and $0 \leq \beta \leq k$. Let us denote the canonical monomial ordering as follows: $(\alpha_1, \beta_1) \prec (\alpha_2, \beta_2)$ means that (α_1, β_1) precedes (α_2, β_2) in the set $\mathcal{T}$. The canonical monomials are ordered according to the following criteria:

- $(\alpha_1, \beta_1) \prec (\alpha_2, \beta_2)$ if:
  $\alpha_1 + \beta_1 < \alpha_2 + \beta_2$, or
  $\alpha_1 + \beta_1 = \alpha_2 + \beta_2$ and $\alpha_1 > \alpha_2$,
  otherwise we have $(\alpha_2, \beta_2) \prec (\alpha_1, \beta_1)$.

The orthonormal monomials are then calculated by using the following recursive formula:

$$P_{\alpha,\beta}(y) = \frac{\tilde{P}_{\alpha,\beta}(y)}{\|\tilde{P}_{\alpha,\beta}(y)\|}, \tag{B.1}$$

with

$$\tilde{P}_{\alpha,\beta}(y) = T_{\alpha,\beta}(y) - \sum_{p,q \in \mathcal{T}} < T_{\alpha,\beta}(y), P_{p,q}(y) > P_{p,q}(y), \tag{B.2}$$

where $P_{\alpha,\beta}(y)$ is the orthonormal monomial associated with the term $y^{\alpha}{y^*}^{\beta}$,

$$< T_{\alpha,\beta}(y), P_{p,q}(y) >= \mathbb{E}\left[T_{\alpha,\beta}(y)P^*_{p,q}(y)\right], \tag{B.3}$$

$$\|\tilde{P}_{\alpha,\beta}(y)\| = \sqrt{\mathbb{E}\left[\left|\tilde{P}_{\alpha,\beta}(y)\right|^2\right]} \tag{B.4}$$

and the sum in (B.2) is carried out according to the ordering defined by $\mathcal{T}$

For instance, let us consider $K = 1$ (third-order monomials). The canonical monomials are given by:

$$\begin{aligned} T_{0,0}(y) &= 1 && \text{(B.5)}\\ T_{1,0}(y) &= y && \text{(B.6)}\\ T_{0,1}(y) &= y^* && \text{(B.7)}\\ T_{2,0}(y) &= y^2 && \text{(B.8)}\\ T_{1,1}(y) &= yy^* && \text{(B.9)}\\ T_{2,1}(y) &= y^2y^*. && \text{(B.10)} \end{aligned}$$

The ordered set of pairs (α, β), for $0 \leq k \leq 1$, $0 \leq \alpha \leq k+1$ and $0 \leq \beta \leq k$, is therefore given by:

$$\mathcal{T} = \{(0,0), (1,0), (0,1), (2,0), (1,1), (2,1)\}. \tag{B.11}$$

In this case, the Gram-Schmidt formulas (B.1) and (B.2) are applied to the set of monomials (B.5)-(B.10) using the ordered set $\mathcal{T}$ defined in (B.11). Thus, taking into account the fact that the random variable y is assumed to be circular of order $2K+1$ and defining $\rho_{y,i,j} = \mathbb{E}[y^i{y^*}^j]$ and $k = i + j$, we get:

- for k=0:

$$P_{0,0}(y) = 1. \tag{B.12}$$

- For k=1:

$$\dot{P}_{1,0}(y) = y - \mathbb{E}[y] = y,$$

$$P_{1,0}(y) = \frac{y}{\sqrt{\rho_{y,1,1}}}. \tag{B.13}$$

$$\dot{P}_{0,1}(y) = y^* - \mathbb{E}[y^*] - \mathbb{E}\left[y^* \frac{y^*}{\sqrt{\rho_{y,1,1}}}\right] \frac{y}{\sqrt{\rho_{y,1,1}}} = y^*,$$

$$P_{0,1}(y) = \frac{y^*}{\sqrt{\rho_{y,1,1}}}. \tag{B.14}$$

- For k=2:

$$\dot{P}_{2,0}(y) = y^2 - \mathbb{E}[y^2] - \mathbb{E}\left[\frac{y^2 y^*}{\sqrt{\rho_{y,1,1}}}\right] \frac{y}{\sqrt{\rho_{y,1,1}}} - \mathbb{E}\left[\frac{y^3}{\sqrt{\rho_{y,1,1}}}\right] \frac{y^*}{\sqrt{\rho_{y,1,1}}} = y^2,$$

$$P_{2,0}(y) = \frac{y^2}{\sqrt{\rho_{y,2,2}}}. \tag{B.15}$$

$$\begin{aligned}\dot{P}_{1,1}(y) &= yy^* - \mathbb{E}[yy^*] - \mathbb{E}\left[\frac{yy^{*^2}}{\sqrt{\rho_{y,1,1}}}\right] \frac{y}{\sqrt{\rho_{y,1,1}}} - \mathbb{E}\left[\frac{y^2 y^*}{\sqrt{\rho_{y,1,1}}}\right] \frac{y^*}{\sqrt{\rho_{y,1,1}}} \\ &- \mathbb{E}\left[\frac{yy^{*^3}}{\sqrt{\rho_{y,2,2}}}\right] \frac{y^2}{\sqrt{\rho_{y,2,2}}} = yy^* - \rho_{y,1,1},\end{aligned}$$

$$P_{1,1}(y) = \frac{yy^* - \rho_{y,1,1}}{\sqrt{\rho_{y,2,2} - \rho_{y,1,1}^2}}. \tag{B.16}$$

- For k=3:

$$\begin{aligned}\dot{P}_{2,1}(y) &= y^2 y^* - \mathbb{E}[y^2 y^*] - \mathbb{E}\left[\frac{y^2 y^{*^2}}{\sqrt{\rho_{y,1,1}}}\right] \frac{y}{\sqrt{\rho_{y,1,1}}} - \mathbb{E}\left[\frac{y^3 y^*}{\sqrt{\rho_{y,1,1}}}\right] \frac{y^*}{\sqrt{\rho_{y,1,1}}} \\ &- \mathbb{E}\left[\frac{y^2 y^{*^3}}{\sqrt{\rho_{y,2,2}}}\right] \frac{y^2}{\sqrt{\rho_{y,2,2}}} - \mathbb{E}\left[\frac{y^3 y^{*^2} - y^2 y^* \rho_{y,1,1}}{\sqrt{\rho_{y,2,2} - \rho_{y,1,1}^2}}\right] \frac{yy^* - \rho_{y,1,1}}{\sqrt{\rho_{y,2,2} - \rho_{y,1,1}^2}},\end{aligned}$$

$$\dot{P}_{2,1}(y) = y^2 y^* - y\frac{\rho_{y,2,2}}{\rho_{y,1,1}},$$

$$P_{2,1}(y) = \frac{\rho_{y,1,1}y^2y^* - \rho_{y,2,2}y}{\sqrt{\rho_{y,1,1}^2\rho_{y,3,3} - \rho_{y,1,1}\rho_{y,2,2}^2}}. \quad \text{(B.17)}$$

APPENDIX C

The PARAFAC Decomposition

Definition C.1 *Let $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ be a N^{th}-order tensor with entries $x_{i_1,i_2,\ldots,i_N}$, for $1 \leq i_n \leq I_n$, with $n \in [1,\, N]$. The Parallel Factor (PARAFAC) decomposition, also known as Canonical Decomposition (CANDECOMP), of the tensor $\mathcal{X}$ is given by:*

$$x_{i_1,i_2,\ldots,i_N} = \sum_{q=1}^{Q} a^{(1)}_{i_1,q} a^{(2)}_{i_2,q} \ldots a^{(N)}_{i_N,q}\,, \tag{C.1}$$

where $a^{(n)}_{i_n,q}$ denotes the $(i_n, q)^{th}$ element of the matrix factor $\boldsymbol{A}^{(n)} \in \mathbb{C}^{I_n \times Q}$, with $n \in [1,\, N]$, and Q is the rank of $\mathcal{X}$.

The PARAFAC decomposition expresses a tensor as a sum of rank-1 tensors, a N^{th}-order rank-1 tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ being defined as a tensor whose elements $a_{i_1,i_2,\ldots,i_N}$, for $1 \leq i_n \leq I_n$, with $n \in [1,\, N]$, can be written as:

$$a_{i_1,i_2,\ldots,i_N} = a^{(1)}_{i_1} a^{(2)}_{i_2} \ldots a^{(N)}_{i_N}\,. \tag{C.2}$$

The essential uniqueness of the PARAFAC decomposition of the tensor $\mathcal{X}$ is assured by the Kruskal theorem [93]. It states that if:

$$\sum_{n=1}^{N} k_{\mathbf{A}^{(n)}} \geq 2Q + N - 1, \tag{C.3}$$

then the matrix factors $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times Q}$, with $n \in [1, N]$, are unique up to column scaling and permutation ambiguities, with $k_{\mathbf{A}}$ denoting the k-rank of the matrix $\mathbf{A}$, i.e. the greatest integer $k_{\mathbf{A}}$ such that every set of $k_{\mathbf{A}}$ columns of $\mathbf{A}$ is linearly independent. That means that if any other set of matrices $\bar{\mathbf{A}}^{(n)} \in \mathbb{C}^{I_n \times Q}$, with $n \in [1, N]$, satisfy (C.1), then

$$\bar{\mathbf{A}}^{(n)} = \mathbf{A}^{(n)} \Pi \Lambda_n, \tag{C.4}$$

where Π is a permutation matrix and Λ_n, with $n \in [1, N]$, are diagonal matrices such that

$$\prod_{n=1}^{N} \Lambda_n = \mathbf{I}_Q. \tag{C.5}$$

The uniqueness of the PARAFAC decomposition was addressed by several authors [146, 97, 36].

A *matrix slice* or 2-D slice of the tensor $\mathcal{X}$ is obtained by fixing $N-2$ indices of $\mathcal{X}$ and varying the two other indices. There are many ways of defining a matrix slice of a tensor $\mathcal{X}$, depending on which indexes are fixed. For instance, the matrix slice obtained by varying the two first indices and fixing the n^{th} index i_n, for $n \in [3, N]$, is given by:

$$\mathbf{X}_{:,:,i_3,i_4,\ldots,i_N} = \mathbf{A}^{(1)} \, \mathsf{diag}_{i_3}\left[\mathbf{A}^{(3)}\right] \cdots \mathsf{diag}_{i_N}\left[\mathbf{A}^{(N)}\right] \, \mathbf{A}^{(2)^T} \in \mathbb{C}^{I_1 \times I_2}, \tag{C.6}$$

with $[\mathbf{X}_{:,:,i_3,i_4,\ldots,i_N}]_{i,j} = x_{i,j,i_3,\ldots,i_N}$. Similarly, the matrix slice obtained by varying the second and third indices, and fixing the n^{th} index i_n, for $n = 1, 4, 5, \ldots, N$, is given by:

$$\mathbf{X}_{i_1,:,:,i_4,i_5,\ldots,i_N} = \mathbf{A}^{(2)} \, \mathsf{diag}_{i_1}\left[\mathbf{A}^{(1)}\right] \mathsf{diag}_{i_4}\left[\mathbf{A}^{(4)}\right] \cdots \mathsf{diag}_{i_N}\left[\mathbf{A}^{(N)}\right] \, \mathbf{A}^{(3)^T} \in \mathbb{C}^{I_2 \times I_3}, \tag{C.7}$$

with $[\mathbf{X}_{i_1,:,:,i_4,i_5,\ldots,i_N}]_{i,j} = x_{i_1,i,j,i_4,\ldots,i_N}$.

An *unfolded matrix* of the tensor $\mathcal{X}$ is constructed by stacking all the matrix slices of a given type so that all the tensor elements are placed in a matrix, which can be done in many different ways. For instance, the most part of the tensor-based techniques presented in this thesis are based on the unfolded matrices of the form:

$$\mathbf{X}_{[1]} = \left(\mathbf{A}^{(1)} \diamond \mathbf{A}^{(2)} \diamond \cdots \diamond \mathbf{A}^{(N-1)}\right) \mathbf{A}^{(N)^T} \in \mathbb{C}^{I_1 I_2 \ldots I_{N-1} \times I_N}, \tag{C.8}$$

where $\diamond$ denotes the Khatri-Rao product, with $\mathbf{X}_{[1]}$ being constructed so that $x_{i_1,i_2,\ldots,i_N}$ is placed at the position (i_{lin}, i_N) of $\mathbf{X}_{[1]}$, with i_{lin} given by:

$$i_{\text{lin}} = (i_1 - 1)I_2 I_3 \ldots I_{N-1} + \cdots + (i_{N-3} - 1)I_{N-2}I_{N-1} + (i_{N-2} - 1)I_{N-1} + i_{N-1}. \quad \text{(C.9)}$$

Similarly, we may construct an unfolded matrix as:

$$\mathbf{X}_{[2]} = \left(\mathbf{A}^{(2)} \diamond \mathbf{A}^{(3)} \diamond \cdots \diamond \mathbf{A}^{(N)}\right) \mathbf{A}^{(1)^T} \in \mathbb{C}^{I_2 I_3 \ldots I_N \times I_1}, \quad \text{(C.10)}$$

with $x_{i_1,i_2,\ldots,i_N}$ being placed at the position (i_{lin}, i_1) of $\mathbf{X}_{[2]}$, with:

$$i_{\text{lin}} = (i_2 - 1)I_3 I_4 \ldots I_N + \cdots + (i_{N-2} - 1)I_{N-1}I_N + (i_{N-1} - 1)I_N + i_N. \quad \text{(C.11)}$$

Another kind of unfolded matrix used in this thesis has the following form:

$$\mathbf{X}_{[3]} = \left(\mathbf{A}^{(1)} \diamond \mathbf{A}^{(2)} \diamond \cdots \diamond \mathbf{A}^{(N-2)}\right) \left(\mathbf{A}^{(N-1)} \diamond \mathbf{A}^{(N)}\right)^T \in \mathbb{C}^{I_1 I_2 \ldots I_{N-2} \times I_{N-1} I_N}, \quad \text{(C.12)}$$

with $x_{i_1,i_2,\ldots,i_N}$ being placed at the position $(i_{\text{lin}}, i_{\text{col}})$ of $\mathbf{X}_{[3]}$, with:

$$i_{\text{lin}} = (i_1 - 1)I_2 I_3 \ldots I_{N-2} + \cdots + (i_{N-3} - 1)I_{N-2} + i_{N-2} \quad \text{(C.13)}$$

and

$$i_{\text{col}} = (i_{N-1} - 1)I_N + i_N. \quad \text{(C.14)}$$

Bibliography

[1] C. K. An, E. J.Powers, and C. P. Ritz. Frequency domain modeling of dual–input/multiple–output quadratic systems with general random inputs. In *IEEE International Symposium on Circuits and Systems*, volume 3, pages 2209–2212, Monterey, CA, USA, Jun. 1998.

[2] B. A. Anouar and M. Hassani. Reduced Volterra model of non linear MIMO systems with decoupled outputs. In *International Symposium on Communications, Control and Signal Processing*, pages 430–433, St. Julians, Malta, Mar. 2008.

[3] E. Aschbacher. *Digital Pre-distortion of Microwave Power Amplifiers*. PhD thesis, Vienna University of Technology, Austria, Sep. 2005.

[4] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. Separating convolutive post non-linear mixtures. In *Workshop on Independent Component Analysis and Signal Separation*, pages 138–143, San Diego, CA, USA, 2001.

[5] P. Banelli, G. Baruffa, and S. Cacopardi. Effects of HPA nonlinearity on frequency multiplexed OFDM signals. *IEEE Transactions on Broadcasting*, 47(2):123–136, Jun. 2001.

[6] P. Banelli and S. Cacopardi. Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels. *IEEE Transactions on Communications*, 48(3):430–441, Mar. 2000.

[7] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, Feb. 1997.

[8] D. A. Bendersky, J. W. Stokes, and H. S. Malvar. Nonlinear residual acoustic echo suppression for high levels of harmonic distortion. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 261–264, Las Vegas, NV, USA, Apr. 2008.

[9] S. Benedetto and E. Biglieri. Nonlinear equalization of digital satellite channels. *IEEE Journal on Selected Areas in Communications*, 1(1):57–62, Jan. 1983.

[10] S. Benedetto, E. Biglieri, and V. Castellani. *Digital Transmission Theory.* Prentice-Hall, 1987.

[11] S. Benedetto, E. Biglieri, and R. Daffara. Modeling and performance evaluation of nonlinear satellite links - A Volterra series approach. *IEEE Transactions on Aerospace Electronic Systems*, 15:494–507, Jul. 1979.

[12] E. Biglieri. High-level modulation and coding for nonlinear satellite channels. *IEEE Transactions on Communications*, 32(5):616–626, May 1984.

[13] E. Biglieri, E. Chiaberto, G. P. Maccone, and E. Viterbo. Compensation of nonlinearities in high-density magnetic recording channels. *IEEE Transactions on Magnetics*, 30(6):5079–5086, Nov. 1994.

[14] E. Biglieri, A. Gersho, R. Gitlin, and T. Lim. Adaptive cancellation of nonlinear intersymbol interference for voiceband data transmission. *IEEE Journal on Selected Areas in Communications*, 2(5):765–777, 1984.

[15] V. A. Bohara and S. H. Ting. Theoretical analysis of OFDM signals in nonlinear polynomial models. In *International Conference on Information, Communications and Signal Processing*, pages 10–13, Singapore City, Singapore, Dec. 2007.

[16] V. A. Bohara and S. H. Ting. Analysis of OFDM signals in nonlinear high power amplifier with memory. In *IEEE International Conference on Communications*, pages 3653–3657, Beijing, China, May 2008.

[17] S. Boyd and L. O. Chua. Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions on Circuits and Systems*, 32(11):1150–1161, Nov. 1985.

[18] R. Bro. *Multi-way analysis in the food industry: Models, algorithms and applications*. PhD thesis, University of Amsterdam, Netherlands, 1998.

[19] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 17(1):161–164, Jan. 1996.

[20] M. Castella. Exact inversion of MIMO nonlinear polynomial mixtures. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 3, pages III–1429–III–1432, Honolulu, HI, USA, Apr. 2007.

[21] M. Castella. Inversion of polynomial systems and separation of nonlinear mixtures of finite-alphabet sources. *IEEE Transactions on Signal Processing*, 56(8):3905–3917, Aug. 2008.

[22] H. Chen and A.M. Haimovich. Iterative estimation and cancellation of clipping noise for OFDM signals. *IEEE Communications Letters*, 7(7):305–307, Jul. 2003.

[23] S.-W. Chen, W. Panton, and R. Gilmore. Effects of nonlinear distortion on CDMA communication systems. *IEEE Transactions on Microwave Theory and Techniques*, 44(12):2743–2750, Dec. 1996.

[24] C.-H. Cheng and E.J. Powers. Optimal Volterra kernel estimation algorithms for a nonlinear communication system for PSK and QAM inputs. *IEEE Transactions on Signal Processing*, 49(1):147–163, 2001.

[25] C. J. Clark, G. Chrisikos, M. S. Muha, A. A. Moulthrop, and C. P. Silva. Time-domain envelope measurement technique with application to wideband power amplifier modeling. *IEEE Transactions on Microwave theory and techniques*, 46(12):2531–2540, Dec. 1998.

[26] A. Conti, D. Dardari, and V. Tralli. An analytical framework for CDMA systems with a nonlinear amplifier and AWGN. *IEEE Transactions on Communications*, 50(7):1110–1120, Jul. 2002.

[27] J. Coon, M. Sandell, M. Beach, and J. McGeehan. Channel and noise variance estimation and tracking algorithms for unique-word based single-carrier systems. *IEEE Transactions on Wireless Communications*, 5(6):1488–1496, Jun. 2006.

[28] E. Costa and S. Pupolin. M-QAM-OFDM system performance in the presence of a nonlinear amplifier and phase noise. *IEEE Transactions on Communications*, 50(3):462–472, Mar. 2002.

[29] C. Cox III, E. Ackerman, R. Helkey, and G. E. Betts. Techniques and performance of intensity-modulation direct detection analog optical links. *IEEE Transactions on Microwave theory and techniques*, 45(8):1375–1383, Aug. 1997.

[30] S. C. Cripps. *RF Power Amplifiers for Wireless Communications*. Artech House, 1999.

[31] A. N. D'Andrea, V. Lottici, and R. Reggiannini. Nonlinear predistortion of OFDM signals over frequency-selective fading channels. *IEEE Transactions on Communications*, 49(5):837–843, May 2001.

[32] D. Dardari, V. Tralli, and A. Vaccari. A theoretical characterization of nonlinear distortion effects in OFDM systems. *IEEE Transactions on Communications*, 48(10):1755–1764, Oct. 2000.

[33] A. L. F. de Almeida, G. Favier, and J. C. M. Mota. PARAFAC-based unified tensor modeling of wireless communication systems with application to blind multiuser equalization. *Signal Processing - Special Issue on Tensor Signal Processing*, 87(2):337–351, Feb. 2007.

[34] A. L. F. de Almeida, G. Favier, and J. C. M. Mota. Space-time spreading MIMO-CDMA downlink systems using constrained tensor modeling. *Signal Processing*, 88(10):2403–2416, Oct. 2008.

[35] A. de Baynast, L. De Lathauwer, and B. Aazhang. Blind PARAFAC receivers for multiple access-multiple antenna systems. In *IEEE Vehicular Technology Conference - Fall*, volume 2, pages 1128–1132, Orlando, FL, USA, Oct. 2003.

[36] L. De Lathauwer. A link between the Canonical Decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 28(3):642–666, Aug. 2006.

[37] L. De Lathauwer and J. Castaing. Blind identification of underdetermined mixtures by simultaneous matrix diagonalization. *IEEE Transactions on Signal Processing*, 56(3):1096–1105, Mar. 2008.

[38] L. Ding. *Digital Predistortion of Power Amplifiers for Wireless Applications*. PhD thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, USA, Mar. 2004.

[39] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina. A robust digital baseband predistorter constructed using memory polynomials. *IEEE Transactions on Communications*, 52(1):159–165, Jan. 2004.

[40] Y. Ding and A. Sano. Time-domain adaptive predistortion for nonlinear amplifiers. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 2, pages ii–865–ii–868, Montreal, Canada, May 2004.

[41] N.Y. Ermolova. OFDM equalization in nonlinear time-varying channels. In *International Symposium on Wireless Communication Systems*, pages 358–362, Valencia Spain, Sep. 2006.

[42] N.Y. Ermolova, N. Nefedov, and S. Haggman. An iterative method for nonlinear channel equalization in OFDM systems. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, volume 1, pages 484–488, Barcelona, Spain, Sep. 2004.

[43] J. Fang, A. R. Leyman, Y. H. Chew, and H. Duan. Some further results on blind identification of MIMO FIR channels via second-order statistics. *Signal Processing*, 87(6):1434–1447, Jun. 2007.

[44] C. A. R. Fernandes, G. Favier, and J. C. M. Mota. Blind source separation and identification of nonlinear multiuser channels using second order statistics and modulation codes. In *IEEE International Workshop Signal Processing Advances in Wireless Communications (SPAWC)*, Helsinki, Finland, Jun. 2007.

[45] C. A. R. Fernandes, G. Favier, and J. C. M. Mota. Blind tensor-based identification of memoryless multiuser Volterra channels using SOS and modulation codes. In *European Signal Processing Conference (EUSIPCO)*, Poznan, Poland, Sep. 2007.

[46] C. A. R. Fernandes, G. Favier, and J. C. M. Mota. Input orthogonalization methods for third-order MIMO Volterra channel identification. In *Colloque GRETSI*, Troyes, France, Sep. 2007.

[47] C. A. R. Fernandes, G. Favier, and J. C. M. Mota. A modulation code-based blind receiver for memoryless multiuser Volterra channels. In *ASILOMAR Conference on Siginal, Systems, and Computers*, Pacific Grove, CA, USA, Nov. 2007.

[48] C. A. R. Fernandes, G. Favier, and J. C. M. Mota. Blind estimation of nonlinear instantaneous channels in multiuser CDMA systems with PSK inputs. In *IEEE International Workshop Signal Processing Advances in Wireless Communications (SPAWC)*, Recife, Brazil, Jul. 2008.

[49] C. A. R. Fernandes, G. Favier, and J. C. M. Mota. Tensor-based blind identification of MIMO volterra channels in a multiuser CDMA environment. In *European Signal Processing Conference (EUSIPCO)*, Lausanne , Switzerland, Aug. 2008.

[50] C. A. R. Fernandes, G. Favier, and J. C. M. Mota. Tensor based receivers for nonlinear radio over fiber uplinks in multiuser CDMA systems. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Cannes, France, Sep. 2008.

[51] C. A. R. Fernandes, G. Favier, and J. C. M. Mota. Blind identification of multiuser nonlinear channels using tensor decomposition and precoding. *Signal Processing*, 2009. To Appear.

[52] C. A. R. Fernandes, A. Kibangou, G. Favier, and J. C. M. Mota. Identification of nonlinear MIMO radio over fiber uplink channels. In *International Telecommunications Symposium*, Fortaleza, Brazil, Sep. 2006.

[53] C. E. R. Fernandes, G. Favier, and J. C. M. Mota. Blind channel identification algorithms based on the parafac decomposition of cumulant tensors: The single and multiuser cases. *Signal Processing*, 88(6):1382–1401, Jun. 2008.

[54] X. N. Fernando and A. B. Sesay. Higher order adaptive filter based predistortion for nonlinear distortion compensation of radio over fiber links. In *IEEE International Conference on Communications*, volume 1/3, pages 367–371, New-Orleans, LA, USA, Jun. 2000.

[55] X. N. Fernando and A. B. Sesay. Adaptive asymmetric linearization of radio over fiber links for wireless access. *IEEE Transactions on Vehicular Technology*, 51(6):1576–1586, Nov. 2002.

[56] X. N. Fernando and A. B. Sesay. Characteristics of directly modulated ROF link for wireless access. In *IEEE Canadian Conference on Electrical and Computer Engineering*, volume 4, pages 2167–2170, Corfu, Greece, May 2004.

[57] X. N. Fernando and A. B. Sesay. A Hammerstein-type equalizer for concatenated fiber-wireless uplink. *IEEE Transactions on Vehicular Technology*, 54(6):1980–1991, 2005.

[58] L. S. Fock and R. S. Tucker. Simultaneous reduction of intensity noise and distortion in semiconductor laser by feed forward compensation. *Electronic Letters*, 27(14):1297–1298, Jul. 1991.

[59] P. Koukoulas G. Mileounis and N. Kalouptsidis. Input-output identification of nonlinear channels using PSK, QAM and OFDM inputs. *Signal Processing*, 89(7):1359–1369, Jan. 2009.

[60] G. B. Giannakis and E. Serpedin. Linear multichannel blind equalizers of nonlinear FIR Volterra channels. *IEEE Transactions on Signal Processing*, 45(1):67–81, Jan. 1997.

[61] G. B. Giannakis and E. Serpedin. A bibliography on nonlinear system identification. *Signal Processing*, 81(3):533–580, Mar. 2001.

[62] A. Goldsmith. *Wireless communications*. Cambridge University Press, 2005.

[63] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE Journal on Selected Areas in Communications*, 21(5):684–702, Jun. 2003.

[64] R. M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239, 2005.

[65] W. Greblicki. Nonparametric identification of Wiener systems by orthogonal series. *IEEE Transactions on Automatic Control*, 39(10):2077–2086, Oct. 1994.

[66] F. Gregorio. *Analysis and compensation of nonlinear power amplifiers effects in multi-antenna OFDM systems*. PhD thesis, Helsinki University of Technology, Finland, 2007.

[67] F. Gregorio, S. Werner, T. I. Laakso, and J. Cousseau. Receiver cancellation technique for nonlinear power amplifier distortion in SDMA–OFDM systems. *IEEE Transactions on Vehicular Technology*, 56(5):2499–2516, Sep. 2007.

[68] A. Gutierrez and W. E. Ryan. Performance of adaptive Volterra equalizers on nonlinear satellite channels. In *IEEE International Conference on Communications*, volume 1, pages 488–492, Seattle, WA, USA, Jun. 1995.

[69] A. Gutierrez and W. E. Ryan. Performance of Volterra and MLSD receivers for nonlinear band-limited satellite systems. *IEEE Transactions on Communications*, 48(7):1171–1177, Jul. 2000.

[70] H. Gysel and M. Ramachandran. Electrical pre-distortion to compensate for combined effect laser chirp and fiber dispersion. *Electronic Letters*, 27(5):421–423, Feb. 1991.

[71] O. Haggstrom. *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press, 2002.

[72] S. Harmeling, A. Ziehe, M. Kawanabe, and K.-R. Muller. Kernel-based nonlinear blind source separation. *Neural Computation*, 15(5):1089–1124, May 2003.

[73] R. A. Harshman. *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis*. UCLA Working Papers in Phonetics, 16th edition, Dec. 1970.

[74] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, 3rd edition, 1996.

[75] R. Hermann. Volterra modeling of digital magnetic saturation recording channels. *IEEE Transactions on Magnetics*, 26(5):2125–2127, Sep. 1990.

[76] X. Hu, M. Hong, J. Peng, and T. Chen. State-of-the-art in Volterra series modeling for ADC nonlinearity. In *Asia International Conference on Modelling & Simulation*, Kuala Lumpur, Malaysia, May 2008.

[77] D. Hummels and R. Gitchell. Equivalent low-pass representations for bandpass Volterra systems. *IEEE Transactions on Communications*, 28(1):140–142, Jan. 1980.

[78] I.-K. Hwang and L. Kurz. Digital data transmission over nonlinear satellite channels. *IEEE Transactions on Communications*, 41(11):1694–1702, Nov. 1993.

[79] M. Ibnkahla, N. J. Bershad, J. Sombrin, and F. Castanie. Neural network modeling and identification of nonlinear channels with memory: algorithms, applications, and analytic models. *IEEE Transactions on Communications*, 46(5):1208–1220, May 1998.

[80] C. Jutten, M. Babaie-Zadeh, and S. Hosseini. Three easy ways for separating nonlinear mixtures? *Signal Processing*, 84(2):217–229, Feb. 2004.

[81] G. Karam and H. Sari. A data predistortion technique with memory. *IEEE Transactions on Communications*, 39(2):336–344, Feb. 1991.

[82] A. Katz, W. Jemison, M. Kubak, and J. Dragone. Improved radio over fiber performance using predistortion linearization. In *IEEE MTT-S International Microwave Symposium Digest*, volume 2, pages 1403–1406, Philadelphia, PA, USA, Jun. 2003.

[83] A. Kibangou and G. Favier. Blind equalization of nonlinear channels using a tensor decomposition with code/space/time diversities. *Signal Processing*, 89(2):133–143, Feb. 2009.

[84] A. Y. Kibangou. *Reduced complexity Volterra models: Parametric estimation and application to channel equalization.* PhD thesis, University of Nice - Sophia Antipolis, France, 2005.

[85] A. Y. Kibangou and G. Favier. Wiener-Hammerstein systems modeling using diagonal Volterra kernels coefficients. *IEEE Signal Processing Letters*, 13(6):381–384, Jun. 2006.

[86] A. Y. Kibangou and G. Favier. Identification aveugle de canaux de communication non-linéaires basée sur la décomposition PARAFAC. In *Colloque GRETSI*, Troyes, France, Sep. 2007.

[87] A. Y. Kibangou, G. Favier, and M. M. Hassani. Blind receiver based on the PARAFAC decomposition for nonlinear communication channels. In *Colloque GRETSI*, pages 177–180, Louvain-la-neuve, Belgium, Sep. 2005.

[88] Y. Kim, S. Doucet, M. E. M. Pasandi, and S. LaRochelle. Optical multicarrier generator for radio-over-fiber systems. *Optics Express*, 16(2):1068–1076, Jan. 2008.

[89] P. R. King and S. Stavrou. Capacity improvement for a land mobile single satellite MIMO system. *IEEE Antennas and Wireless Propagation Letters*, 5(1):98–100, Dec. 2006.

[90] P. R. King and S. Stavrou. Characteristics of the land mobile satellite MIMO channel. In *IEEE Vehicular Technology Conference - Fall*, pages 1–4, Montreal, Canada, Sep. 2006.

[91] P. R. King and S. Stavrou. Low elevation wideband land mobile satellite MIMO channel characteristics. *IEEE Transactions on Wireless Communications*, 6(7):2712–2720, Jul. 2007.

[92] M. J. Korenberg and I. W. Hunter. The identification of nonlinear biological systems: LNL cascade models. *Biological Cybernetics*, 55(2–3):125–134, Nov. 1986.

[93] J. Kruskal. Three way arrays: Rank and uniqueness of trilinear decomposition with applications to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18:95–138, 1977.

[94] H. Ku. *Behavior modeling of nonlinear RF power amplifiers for digital wireless communication systems with implications for predistortion linearization systems*. PhD thesis, Georgia Institute of Technology, USA, Oct. 2003.

[95] H. Ku and J. S. Kenney. Behavioral modeling of nonlinear RF power amplifiers considering memory effects. *IEEE Transactions on Microwave Theory and Techniques*, 51(12):2495–2504, Dec. 2003.

[96] S. Litsyn. *Peak Power Control in Multicarrier Communications*. Cambridge University Press, 1st edition, 2007.

[97] X. Liu and N. D. Sidiropoulos. Cramer-Rao lower bounds for low-rank decomposition of multidimensional arrays. *IEEE Transactions on Signal Processing*, 49(9):2074–2086, 2001.

[98] X. Liu and N.D. Sidiropoulos. Cramer-Rao lower bounds for low-rank decomposition of multidimensional arrays. *IEEE Transactions on Signal Processing*, 49(9):2074–2086, Sep. 2001.

[99] R. Lopez-Valcarce and S. Dasgupta. Blind equalization of nonlinear channels from second-order statistics. *IEEE Transactions on Signal Processing*, 49(12):3084–3097, Dec. 2001.

[100] R. Lopez-Valcarce and S. Dasgupta. Second-order statistical properties of nonlinearly distorted phase-shift keyed (PSK) signals. *IEEE Communications Letters*, 7(7):323–325, Jul. 2003.

[101] P. Z. Marmarelis and V. Z. Marmarelis. *Analysis of Physiological Systems*. Plenum, New-York, USA, 1978.

[102] R. Marsalek. *Contributions to the power amplifier linearization using digital baseband adaptive predistortion*. PhD thesis, Université de Marne-la-Vallée, France, 2003.

[103] R. Marsalek, P. Jardin, and G. Baudoin. From post-distortion to predistortion for power amplifiers linearization. *IEEE Communications Letters*, 7(7):308–310, Jul. 2003.

[104] R. J. Martin. Volterra system identification and Kramer's sampling theorem. *IEEE Transactions on Signal Processing*, 47(11):3152–3155, Nov. 1999.

[105] V. J. Mathews. Adaptive Volterra filters using orthogonal structures. *IEEE Signal Processing Letters*, 3(12):307–309, Dec. 1996.

[106] J. E. Mitchell. Performance of OFDM at 5.8 GHz using radio over fibre link. *Electronics Letters*, 40(21):1353–1354, Oct. 2004.

[107] R. D. Nowak and B. D. Van Veen. Random and pseudorandom inputs for Volterra filter. *IEEE Transactions on Signal Processing*, 42(8):2124–2135, 1994.

[108] T. Ogunfunmi and S. L. Chang. Second-order adaptive Volterra system identification based on discrete nonlinear Wiener model. *IEE Proceedings - Vision, Image and Signal Processing*, 148(1):21–29, Feb. 2001.

[109] A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-Time Signal Processing*. Prentice-Hall, 2nd edition, 1998.

[110] A. Paraschiv-Ionescu, C. Jutten, and G. Bouvier. Source separation based processing for integrated Hall sensor arrays. *IEE Sensors Journal*, 2(6):663–673, Dec. 2002.

[111] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei. An overview of MIMO communications: A key to gigabit wireless. *Proc. of IEEE*, 92(2):198–218, Feb. 2004.

[112] R. Pervez and M. Nakagawa. Semiconductor laser's nonlinearity compensation for DS-CDMA optical transmission system by post nonlinearity recovery block. *IEICE Transactions on Communications*, E-79 B(3):1576–1586, Mar. 1996.

[113] N. Petrochilos and K. Witrisal. Semi-blind source separation for memoryless Volterra channels in UWB and its uniqueness. In *IEEE Workshop on Sensor Array and Multichannel Processing*, pages 566–570, Waltham, MA, USA, Jul. 2006.

[114] S. Z. Pinter and X. N. Fernando. Estimation of radio-over-fiber uplink in a multiuser CDMA environment using PN spreading codes. In *Canadian Conference on Electrical and Computer Engineering*, pages 1–4, Saskatoon, Canada, May 2005.

[115] S. Z. Pinter and X. N. Fernando. Fiber-Wireless solution for broadband multimedia access. *IEEE Canadian Review - Summer*, (50):6–9, 2005.

[116] S. Z. Pinter and X. N. Fernando. Concatenated fiber-wireless channel identification in a multiuser CDMA environment. *IET Communications*, 1(5):937–944, Oct. 2007.

[117] S. Z. Pinter and X. N. Fernando. Equalization of multiuser wireless CDMA downlink considering transmitter nonlinearity using Walsh codes. *EURASIP Journal on Wireless Communications and Networking*, 2007(1), Jan. 2007.

[118] J. G. Proakis. *Digital Communications*. McGraw-Hill, 4*rd* edition, 2001.

[119] R. Raich. *Nonlinear System Identification and Analysis with Applications to Power Amplifier Modeling and Power Amplifier Predistortion*. PhD thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, USA, Mar. 2004.

[120] R. Raich, H. Qian, and G. T. Zhou. Digital baseband predistortion of nonlinear power amplifiers using orthogonal polynomials. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 6, pages 689–692, Hong-Kong, Apr. 2003.

[121] R. Raich and G. T. Zhou. On the modeling of memory nonlinear effects of power amplifiers for communication applications. In *IEEE Digital Signal Processing Workshop*, pages 7–10, Pine Mountain, GA, USA, Oct. 2002.

[122] M. Rajih and P. Comon. Enhanced line search: A novel method to accelerate PARAFAC. In *European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey, Sep. 2005.

[123] J. C. Ralston and A. M. Zoubir. Identification of a class of multiple input–output nonlinear systems driven by stationary non-Gaussian processes. In *IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pages 379–382, Corfu, Greece, Jun. 1996.

[124] K. D. Rao and D. C. Reddy. Design of Multi-Input Multi-Output adaptive Volterra filters. In *Digital Signal Processing Workshop*, pages 8.11.1–8.11.2, Sep. 1992.

[125] P. Raziq and M. Nagakawa. Semiconductor laser's nonlinearity compensation for DS–CDMA optical transmission system by post nonlinearity recovery block. *IEICE Transactions on Communications*, E79-B(3):424–431, Mar. 1996.

[126] A. J. Redfern and G. T. Zhou. Nonlinear channel identification and equalization for OFDM systems. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 6, pages 3521–3524, Seattle, WA,USA, May 1998.

[127] A. J. Redfern and G. T. Zhou. Blind zero forcing equalization of multichannel nonlinear CDMA systems. *IEEE Transactions on Signal Processing*, 49(10):2363–2371, Oct. 2001.

[128] P. Ritosa, B. Batagelj, and M. Vidmar. Optically steerable antenna array for radio over fibre transmission. *Electronics Letters*, 41(16):47–48, Aug. 2005.

[129] N. Rodriguez, I. Soto, and R. Carrasco. Adaptive predistortion of COFDM signals for a mobile satellite channel. *International Journal of Communications Systems*, 16(2):137–150, 2003.

[130] Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos. Blind spatial signature estimation via time-varying user power loading and parallel factor analysis. *IEEE Transactions on Signal Processing*, 53(5):1697–1709, May 2005.

[131] A. A. M. Saleh. Frequency-independent and frequency dependent nonlinear models of TWT amplifiers. *IEEE Transactions on Communications*, COM-29:1715–1720, Nov. 1981.

[132] M. Sauer, A. Kobyakov, and J. George. Radio over fiber for picocellular network architectures. *Journal of Lightwave Technology*, 25(11):3301–3320, Nov. 2007.

[133] T. C. W. Schenk, C. Dehos, D. Morche, and E. R. Fledderus. Receiver-based compensation of transmitter-incurred nonlinear distortion in multiple-antenna OFDM systems. In *IEEE Vehicular Technology Conference - Fall*, pages 1346–1350, Baltimore, MD, USA, Oct. 2007.

[134] M. Schetzen. *The Volterra and Wiener theories of nonlinear systems*. John Wiley and Sons, Inc., New-York, USA, 1980.

[135] R. T. Schwarz, A. Knopp, D. Ogermann, C. A. Hofinann, and B. Lankl. Optimum-capacity MIMO satellite link for fixed and mobile services. In *International ITG Workshop on Smart Antennas*, pages 209–216, Darmstadt, Germany, Feb. 2008.

[136] C. Seretis and E. Zafiriou. Nonlinear dynamical system identification using reduced Volterra models with generalized orthonormal basis functions. In *American Control Conference*, volume 5, pages 3042–3046, Albuquerque, NM, USA, Jun. 1997.

[137] S. Serfaty, J.L. LoCicero, and G.E. Atkin. Cancellation of nonlinearities in bandpass QAM systems. *IEEE Transactions on Communications*, 38(10):1835–1843, 1990.

[138] A.R. Shah and B. Jalali. Adaptive equalisation for broadband predistortion linearisation of optical transmitters. *IEE Proceedings - Optoelectronics*, 152(1):16–32, Feb. 2005.

[139] K. Shi, X. Ma, and G. T. Zhou. Adaptive acoustic echo cancellation in the presence of multiple nonlinearities. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 3601–3604, Las Vegas, NV, USA, Apr. 2008.

[140] K. Shi, X. Ma, and G. T. Zhou. A residual echo suppression technique for systems with nonlinear acoustic echo paths. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 257–260, Las Vegas, NV, USA, Apr. 2008.

[141] J. J. Shynk and R. P. Gooch. The constant modulus array for cochannel signal copy and direction finding. *IEEE Transactions on Signal Processing*, 44(3):652–660, Mar. 1996.

[142] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis. Parallel factor analysis in sensor array processing. *IEEE Transactions on Signal Processing*, 48(8):2377–2388, Aug. 2000.

[143] N. D. Sidiropoulos and G. Z. Dimic. Blind multiuser detection in WCDMA systems with large delay spread. *IEEE Signal Processing Letters*, 8(3):87–89, Mar. 2001.

[144] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro. Blind PARAFAC receivers for DS-CDMA systems. *IEEE Transactions on Signal Processing*, 48(3):810–823, Mar. 2000.

[145] N. D. Sidiropoulos and X. Liu. Identifiability results for blind beamforming in incoherent multipath with small delay spread. *IEEE Transactions on Signal Processing*, 49(1):228–236, Jan. 2001.

[146] N.D. Sidiropoulos and R. Bro. On the uniqueness of multilinear decomposition of N-way arrays. *Journal of Chemometrics*, 14(3):229–239, 2000.

[147] Y. T. Su, M.-C. Chiu, and Y.-C. Chen. Turbo equalization of nonlinear TDMA satellite signals. In *IEEE Global Telecommunications Conference*, volume 3, pages 2860–2864, Taipei, Taiwan, Nov. 2002.

[148] A. I. Sulyman and M. Ibnkahla. Performance analysis of non-linearly amplified M-QAM signals in MIMO channels. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 4, pages iv–401–iv–404, Montreal, Canada, May 2004.

[149] A. Taleb and C. Jutten. Source separation in post-nonlinear mixtures. *IEEE Transactions on Signal Processing*, 47(10):2807–2820, Sep. 1999.

[150] S. Talwar, M. Viberg, and A. Paulraj. Blind separation of synchronous co-channel digital signals using an antenna array – Part I: Algorithms. *IEEE Transactions on Signal Processing*, 44(5):1184–1197, May 1996.

[151] I. E. Telatar. Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, Nov. 1999.

[152] J. Tellado, L. Hoo, and J. Cioffi. Maximum-likelihood detection of nonlinearly distorted multicarrier symbols by iterative decoding. *IEEE Transactions on Communications*, 51(2):218–228, Feb. 2003.

[153] J. M. F. ten Berge. The k-rank of a Khatri-Rao product. *Unpublished Note*, Heijmans Institute of Psychological Research, University of Groningen, The Netherlands, 2000.

[154] J. Tsimbinos and K. V. Lever. Input Nyquist sampling suffices to identify and compensate nonlinear systems. *IEEE Transactions on Signal Processing*, 46(10):2833–2837, Oct. 1998.

[155] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.

[156] R. S. Tucker. High speed modulation of semiconductor lasers. *IEEE Journal of Lightwave Technology*, 3(6):1180–1192, Dec. 1985.

[157] A. Uncini, L. Vecci, P. Campolucci, and F. Piazza. Complex-valued neural networks with adaptive spline activation function for digital radio links nonlinear equalization. *IEEE Transactions on Signal Processing*, 47(2):505–514, Feb. 1999.

[158] H. Valpola, E. Oja, A. Ilin, A. Honkela, and J. Karhunen. Nonlinear blind source separation by variational bayesian learning. *IEICE Transactions*, E86-A(3):532–541, Mar. 2003.

[159] A.-J. van der Veen. Algebraic methods for deterministic blind beamforming. *Proc. of IEEE*, 86(10):1987–2008, Oct. 1998.

[160] V. Volterra. *Theory of Functionals and of Integral and Integro-Differential Equations*. New York: Dover Publications, 1959.

[161] W. I. Way. Subcarrier multiplexed lightwave system design considerations for subscriber loop applications. *IEEE Journal of Lightwave Technology*, 7(11):1806–1818, Nov. 1989.

[162] W.I. Way. Optical fiber based microcellular systems. An overview. *IEICE Transactions on Communications*, E76-B(9):1091–1102, Sep. 1993.

[163] B. Widrow and M. E. Hoff. Adaptive switching circuits. In *IRE WESCON Convention Record*, pages 96–104, 1960.

[164] N. Wiener. *Nonlinear Problems in Random Theory*. MIT Press, Cambridge, MA, USA, 1958.

[165] K. Witrisal, G. Leus, M. Pausini, and C. Krall. Equivalent system model and equalization of differential impulse radio UWB systems. *IEEE Journal on Selected Areas in Communications*, 23(9):1851–1862, Sep. 2005.

[166] C. Xia and J. Ilow. Blind compensation of memoryless nonlinear effects in OFDM transmissions using CDF. In *Communication Networks and Services Research Conference*, Moncton, Canada, May 2003.

[167] F. Yamashita, K. Kobayashi, M. Ueba, and M. Umehira. Broadband multiple satellite MIMO system. In *IEEE Vehicular Technology Conference - Fall*, volume 4, pages 2632–2636, Dallas, TX, USA, Sep. 2005.

[168] Y. Yu and A. P. Petropulu. PARAFAC-based blind estimation of possibly underdetermined convolutive MIMO systems. *IEEE Transactions on Signal Processing*, 56(1):111–124, Jan. 2008.

[169] G. Zhou and R. Raich. Spectral analysis of polynomial nonlinearity with applications to RF power amplifiers. *EURASIP Journal on Applied Signal Processing*, 12:1831–1840, 2004.

[170] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Muller. Blind separation of post-nonlinear mixtures using linearizing transformations and temporal decorrelation. *Journal of Machine Learning Research*, 4(7-8):1319–1338, 2003.

[171] A. Ziehe, K.-R. Muller, G. Nolte, B.-M. Mackert, and G. Curio. Artifact reduction in magnetoneurography based on time-delayed second-order correlations. *IEEE Transactions on Biomedical Engineering*, 47(1):75–87, 2000.

[172] Q. Zou, M. Mikhemar, and A. H. Sayed. Digital compensation of RF nonlinearities in software-defined radios. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2921–2924, Las Vegas, NV,USA, Apr. 2008.