ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΜΗ ΓΡΑΜΜΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΙΣ ΤΗΛΕΠΙΚΟΙΝΩΝΙΕΣ

Διαμαντής Κωτούλας

ΑΘΗΝΑ

ΦΕΒΡΟΥΑΡΙΟΣ 2012

NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCE

DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

PROGRAM OF POSTGRADUATE STUDIES

Ph.D. DISSERTATION

NONLINEAR SIGNAL PROCESSING APPLIED TO TELECOMMUNICATIONS

Diamantis Kotoulas

ATHENS

FEBRUARY 2012

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Μη γραμμική επεξεργασία σήματος με εφαρμογές στις τηλεπικοινωνίες

Διαμαντής Κωτούλας

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: Νικόλαος Καλουπτσίδης, Καθηγητής ΕΚΠΑ

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ:

Νικόλαος Καλουπτσίδης, (Καθηγητής ΕΚΠΑ)

Σέργιος Θεοδωρίδης, (Καθηγητής ΕΚΠΑ)

Σεραφείμ Καραμπογιάς, (Επίκουρος Καθηγητής ΕΚΠΑ)

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Νικόλαος Καλουπτσίδης
Καθηγητής ΕΚΠΑ

Σέργιος Θεοδωρίδης
Καθηγητής ΕΚΠΑ

Σεραφείμ Καραμπογιάς
Επίκουρος Καθηγητής ΕΚΠΑ

Εμμανουήλ Σαγκριώτης
Αναπλ. Καθηγητής ΕΚΠΑ

Κων/νος Μπερμπερίδης
Καθηγητής Παν/μίου Πατρών

Γεώργιος-Όθωνας Γλεντής
Αναπλ. Καθηγητής Πανεπιστημίου Πελοποννήσου

Ελευθέριος Κοφίδης
Επικ. Καθηγητής Πανεπιστημίου Πειραιά

Ημερομηνία Εξέτασης: 20/2/2012

Ph.D. DISSERTATION

Nonlinear signal processing applied to telecommunications

Diamantis Kotoulas

ADVISOR: Nicholas Kalouptsidis, Professor NKUA

THREE-MEMBER ADVISING COMMITTEE:

Nicholas Kalouptsidis, (Professor NKUA)

Sergios Theodoridis, (Professor NKUA)

Serafim Karambogias, (Assistant Professor NKUA)

SEVEN-MEMBER EXAMINATION COMMITTEE

Nicholas Kalouptsidis
Professor NKUA

Sergios Theodoridis
Professor NKUA

Serafim Karambogias
Assistant Professor NKUA

Emmanouil Sagriotis
Assoc. Professor NKUA

Konstantinos Berberidis
Professor University of Patras

George-Othon Glentis
Assoc. Professor University of Peloponnese

Eleutherios Kofidis
Assist. Professor University of Pireaus

Examination Date: 20/2/2012

Περίληψη

Μη γραμμικές συμπεριφορές παρουσιάζονται σχεδόν σε όλα τα τηλεπικοινωνιακά συστήματα, όπως δορυφορικά συστήματα, τηλεφωνικά κανάλια, κυψελωτά δίκτυα κινητής τηλεφωνίας, ασύρματα τοπικά δίκτυα, ραδιοτηλεοπτικά κανάλια, ψηφιακά μαγνητικά συστήματα κλπ. Γραμμικές προσεγγίσεις αυτών των φαινομένων που δεν λαμβάνουν υπόψη μη γραμμικές συμπεριφορές μπορούν να μειώσουν την απόδοση των συστημάτων και να οδηγήσουν στην απώλεια κυκλοφορίας. Για το λόγο αυτό, κατάλληλα μη γραμμικά πρότυπα πρέπει να αναπτυχθούν. Σημαντικό θέμα στη μελέτη γραμμικών αλλά και μη γραμμικών συστημάτων είναι αυτό του παραθύρου μνήμης (τάξης) των εμπλεκομένων υποσυστημάτων. Είναι κρίσιμο, καθώς η ακριβής γνώση του μπορεί να οδηγήσει σε αποτελεσματική αναγνώρισης συστήματος και ταυτοποίηση των εμπλεκομένων καναλιών. Ο κύριος στόχος αυτής της διατριβής είναι η λύση του γνωστού προβλήματος του προσδιορισμού της δομής ενός συστήματος, όπως αυτό σχετίζεται με την ισοστάθμιση και ταυτοποίησή του. Ειδικότερα, επιτυγχάνονται η ανάπτυξη μεθόδων και αλγορίθμων για τον προσδιορισμό της τάξης, τόσο σε γραμμικά όσο και μη γραμμικά συστήματα, καθώς και η ανάπτυξη μεθόδων και αλγορίθμων για την ταυτοποίηση των εμπλεκόμενων συστημάτων και την ισοστάθμιση των καναλιών τους.

Η εκτίμηση της σωστής τάξης για ένα σύστημα είναι δύσκολη, τόσο σε γραμμικό όσο και σε μη γραμμικό περιβάλλον. Για την περίπτωση των γραμμικών συστημάτων μιάς εισόδου και πολλών εξόδων έχουν αναπτυχθεί στο παρελθόν διάφορες μέθοδοι οι οποίες παρουσιάζουν πλεονεκτήματα αλλά και μειονεκτήματα. Η απόδοσή τους εξαρτάται σημαντικά τόσο από τις μεταβολές του SNR, όσο και από τον αριθμό των συμβόλων εξόδου που χρησιμοποιούνται για τους υπολογισμούς. Για τα συστήματα πολλαπλών εισόδων-εξόδων δεν έχουν έως τώρα αναπτυχθεί αντίστοιχοι αλγόριθμοι.

Η μεθοδολογία που αναπτύσσεται στην εργασία αυτή χρησιμοποιεί υπόχωρους που παράγονται από την προβολή των συμβόλων εξόδου τόσο σε παρελθούσες όσο και σε μελλοντικές τιμές. Η μεθοδολογία επίλυσης του προβλήματος βασίζεται σε τεχνικές βέλτιστης εξομάλυνσης. Συγκεκριμένα, επιλέγεται ένα παράθυρο μνήμης και όλες οι τιμές της εξόδου που εμπίπτουν σ' αυτό προβάλλονται ορθογώνια τόσο σε παραλθούσες όσο και σε μελλοντικές τιμές. Στη συνέχεια υπολογίζεται ο πίνακας λάθους των προβολών αυτών και, με βάση το βαθμό του, προσδιορίζονται οι τάξεις των υποσυστημάτων που αποτελούν το αρχικό σύστημα. Αποδεικνύεται ότι ο βαθμός του πίνακα προβολών είναι μία βηματική συνάρτηση της οποίας τα βήματα αύξησης εξαρτώνται γραμμικά από τον αριθμό των υποσυστημάτων που εμφανίζουν την ίδια τάξη αλλά και από την τιμή της τάξης αυτής. Με τον τρόπο αυτό απομονώνονται τα υποσυστήματα που εμφανίζουν το ίδιο παράθυρο μνήμης, τόσο γραμμικά

όσο και μη γραμμικά, αρκεί τα τελευταία να περιγράφονται από πεπερασμένες σειρές Volterra. Στη συνέχεια, έχοντας απομονώσει τα υποσυστήματα με την ίδια τάξη, λύνουμε την εξίσωση υπολογισμού του πίνακα λάθους προβολής, για κάθε ομάδα υποσυστημάτων με την ίδια μνήμη και προχωρούμε σε ταυτοποίηση καναλιών στη γραμμική περίπτωση, ή σε ισοστάθμισή τους στη μη γραμμική περίπτωση.

Θεματική περιοχή: Προσδιορισμός τάξης, Εκτίμηση γραμμικών καναλιών, Εκτίμηση μη γραμμικών καναλιών

Λέξεις κλειδιά: Σειρές Volterra, Τυφλή ταυτοποίηση, Τυφλή ισοστάθμιση

Abstract

Nonlinear behaviour appears in almost all digital communication systems, such as satellite systems, telephone channels, mobile cellular communications, wireless LAN devices, radio and TV channels, digital magnetic systems, etc. Linear approximations that do not take into account this type of behaviour, may lead to system performance degradation as well as loss of information. Therefore, appropriate models should be developed that tackle nonlinear system characteristics. Another important issue in studying both linear and nonlinear systems is that of the order (memory length) of the associated subsystems. It is critical, because knowing the exact subsystem orders may lead to accurate system identification and channel equalization. The primary objective of this dissertation is the solution of the structure determination problem in system identification. We develop methods and algorithms for the order determination of both linear and nonlinear systems and also for system identification and equalization.

Estimating the correct channel order is difficult in both SIMO and MIMO cases. For the case of SIMO systems several order estimation algorithms have been developed with several advantages and disadvantages. Their success depends highly on variations of SNR as well as the number of output data samples used for computations. The approaches addressing the problem of channel order estimation for MIMO systems are limited so far.

The methodology developed in this thesis builts upon subspaces induced by projecting system output data into past and future values. The concept is based in smoothing techniques. Specifically, a window of spesific length is chosen and data values that fall within window are orthogonally projected into past and future values. Then, the error projection matrix is developed and rank detection techniques are employed to identify subsystem orders. It turns out that, the rank of the error projection matrix is a step function, with increments depending linearly on the number of subsystems that attain the same order as well as the specific order value. This property helps identify and isolate the group of subsystems that attain the same order. This approach proves efficient for nonlinear systems when finite Volterra series is used for system modeling. Having identified discrete subsystem orders,we solve the projection error matrix equation for each group of subsystems and either identify a linear channel or equalize non-linear channels.

Subject Area: Order Determination, Nonlinear Channels Estimation, Linear Channels Estimation

Keywords: Volterra Series, Blind Identification, Blind Equalization

Στη γυναίκα μου Ειρήνη και το γιό μας Στέφανο

To my wife Irini and our son Stephanos

Acknowledgements

First and foremost, I would like to express my sincere appreciation and gratitude to my advisor Prof. Nicholas Kalouptsidis, who offered me the opportunity to pursue my Ph.D. degree and grow as a researcher at the University of Athens. I am grateful to him, for his patience, his continuous guidance, encouragement, enthusiasm and advice throughout my disseration efforts. His deep academic knowledge and kindness give him a unique talent to inspire his students with enthusiasm for their work and give them the right directions to reach their research targets. At difficulties, when research problems or pressure at work seemed too much to overcome, Prof. Kalouptsidis was always there to support me and give me a way out of the deadlocks. Having been his student from my early undergraduate years, having performed my graduation thesis at the Department of Computer Engineering and Information Science, at the Polytechnic School at the University of Patras with him as one of my supervisors, the least I can say is that people like him are examples to be followed. I am grateful and proud to be his student for all these years and I wish him health, family hapiness and academic success in years to come.

In addition, I would like to express my appreciation to my co-advisors Prof. Sergios Theodoridis and Assistant Prof. Serafeim Karambogias for their encourangement, help and support. A deep debt of gratidute is owed to Prof. Sergios Theodoridis, also one of my inspirers into the field from the early undergraduate years, for always being there to advise and support me. Besides my advisors, I would like to thank the members of my dissertation commitee Prof. Kostas Berberidis, Associate Prof. George-Othon Glentis, Associate Prof. Emmanouil Sagriotis and Assistant Prof. Eleftherios Kofidis for accepting the membership and reviewing the work on a short notice.

Moreover, I want to express my warmest thanks to Dr. Panos Koukoulas. Without his support at critical times of my research efforts it would be impossible to finish the requirements of the disseration. Dr. Gerasimos Mileounis, Dr. Kostis Xenoulis and Dr. Alexandros Katsiotis should also be thanked for helpful technical discussions as well their support and friendship.

Special thanks should also go to colleagues at work, Dr. Christos Vasilopoulos for his continuous encouragement and understanding, Petros Vouddas, Charalabos Skoufis, Abraham Asimakopoulos, Konstantinos Klotsotiras as well as Dr. Dimitrios Xenikos, for reading and commenting on drafts of my thesis, encouraging and pressing me to complete it.

Finally, I dedicate this dissertation to my wife Irini and our son Stephanos, apologizing for the hours I took away from them, as well as my parents and my brother for their endless love and support.

Table of Contents

List of Tables

1. Introduction

This dissertation addresses the problems of order determination and identification/equalization for both linear and nonlinear communication channels. In this introductory chapter we present the motivation and objectives of this work and we outline its contents.

1.1 Motivation

Equalization or deconvolution is essentially a signal processing procedure to restore a set of source signals which were distorted by an unknown linear or nonlinear system, whereas system identification is a signal processing procedure to identify and estimate the unknown linear or nonlinear system. The two problems arise in a variety of areas such as digital communications, speech signal processing, image signal processing, biomedical signal processing, seismic exploration, ultrasonic nondestructive evaluation, underwater acoustics, radio astronomy, sonar, radar, and so on. As the source signals are known a priori, the design of identification algorithms will be straightforward and effective. Certainly, the design of equalization algorithms will also be straightforward and effective when the system is completely known in advance. When both the source signals and the system are unknown, blind identification or equalization can be performed, but in this case, the equalization and identification problems are far from trivial. Obviously, the two problems are closely related to each other, and therefore similar design philosophies may frequently apply to the design of both equalization algorithms and system identification algorithms.

In this dissertation, we study blind equalization and identification methods, not only for linear systems but also for nonlinear systems modelled by finite Volterra series. The Volterra series is a power series with memory, first studied by the mathematician Vito Volterra in 1887 [1]. It was first used in nonlinear system theory by N. Wiener in early 1940s. In 1954 the group ”Statistical Theory of Communication” was formed at the Massachussets Institute of Technology (MIT) by Y.W.Lee, a student of N. Wiener. From 1950 till the late 1960s many students of Lee contributed to the theory of nonlinear systems [2]. A first attempt on the applications of functionals on the study of nonlinear systems was made by J.F. Barrett [3] Two fundamental works on the theory of Volterra series were the works by M. Schetzen [4] and W.J.Rugh [5].

The approach we take towards blind identification/equalization is the following. We develop a

method that identifies the orders (memory lengths) of the discrete subsystems that comprise the total system (either linear or nonlinear). This method is implemented by a computationally efficient algorithm. The algorithm detects the different subsystem orders as well as the number of subsystems that attain the same order. This is done both for Single Input Multiple Output (SIMO) and Multiple Input Multiple Output (MIMO) Linear Time Invariant (LTI) Finite Impulse Response (FIR) systems as well as for nonlinear SIMO Volterra systems.

Once the orders have been determined, subsystems are clustered in groups of the same memory length. For linear systems, Blind Source Separation techniques (BSS) are used to identify the system kernels. For nonlinear Volterra systems equalization is performed.

The complexity of the algorithm that is introduced depends on the number of subsystems that comprise the total system, the number of output channels, the degree of nonlinearity, the number of output samples and finally the statistical properties of the input. Our effort is concentrated on developing an algorithm that minimizes complexity with respect to all the above parameters.

1.2 Objectives

The objective of this dissertation is to propose new blind identification/equalization techniques and apply them to linear and nonlinear communication channels.

The research efforts focus on the following targets:

- To develop methods and algorithms for the order determination of both linear and nonlinear systems.
- To develop methods and algorithms for system identification and equalization.

For LTI FIR systems the identification problem is related to that of correct order determination. As it will be descibed in detail in the following chapter, order selection is a fundamental task in time-series analysis and signal processing. A common assumption of blind channel identification methods, is that the order of the true channel is known. However, this information is not available in most practice applications.

Therefore, most of the existing algorithms try to estimate the channel order by applying a rank detection procedure to an overmodeled data covariance matrix. Two commonly used approaches for this task are the Akaike Information Criterion (AIC) and the Minimum Description Length (MDL)

algorithms , which were originally proposed in [6] and popularized in the signal processing community by Wax and Kailath [7]. However, it is proved [6] that these criteria are very sensitive to SNR variations and number of data samples. This fact prohibits their successful application for channel order estimation in all cases.

The problem has been alleviated in [8] by channel approximation methods that attempt to model only the significant part of the channel that is composed of the large impulse response terms because efforts toward modeling small leading and/or trailing terms lead to effective overmodeling, which is generically ill-conditioned. These methods, perform blind identification techniques that assume model order equal to the order of the significant part of the true channel, called the effective channel order.

While the algorithm developed there applies to SIMO LTI FIR systems there is no work done addressing the problem of order determination for MIMO LTI FIR systems and done for nonlinear systems. In this dissertation, we present an algorithm for order detection that is applicable to both SIMO and MIMO linear systems as well as nonlinear systems modelled by finite Volterra series. Based on correct order detection, we next develop a procedure for kernel identification and equalization.

1.3 Outline

The dissertation is organized as follows:

- In *Chapter 2* a review of the literature concerning order detection methods as well as blind identification/equalization methods for linear and nonlinear systems is presented.
- In *Chapter 3* we develop a new order detection method for both linear as well as nonlinear Volterra systems. A computational efficient algorithm is then derived. QPSK inputs are used. For the linear case, the method is applied to several microwave channels and comparisons made to existing algorithms are discussed. The performance of the algorithm in a nonlinear context is checked for various systems and SNR values.
- *Chapter 4* we deals with a new method for blind identification of linear systems. In addition, it addresses equalization of Volterra systems.
- The main conclusions of this work along with directions for future research are presented in *Chapter 5*.

2. Background

In this chapter background is introduced. We present the issues of order determination and channel estimation. Next, we provide an overview of blind identification and equalization methods for Linear Time Invariant (LTI) systems with Finite Impulse Response (FIR). Next, we provide an overview of blind methods for nonlinear systems identification and equalization. The systems considered are modeled by finite Volterra series. We give an exposition of methods that can be used to estimate a Volterra system. Finally, we present the Blind Source Separation (BSS) concept and algorithms in this category, as BSS techniques are used to develop our algorithm.

2.1 Order Detection Methods

Following the work of Tong et al. [9], many methods have been recently proposed that claim blind single input/ multi-output (SIMO) channel identification under the so-called length and zero conditions [10]. A common assumption in all these works is that the order of the true channel is known. Of course, such information is not available in practice, and we are thus obliged to estimate the channel order by applying a rank detection procedure to an overmodeled data covariance matrix. The use of information theoretic criteria, as proposed in [11], has become the standard first step of many methods that treat the blind channel identification problem.

The development of information theoretic criteria is based on the assumptions that successive data vectors are i.i.d. zero mean Gaussian and that the noise is white Gaussian, uncorrelated to the signal. These assumptions seem realistic in some applications, but in many other, such as blind channel identification, they are not the most appropriate. First, in blind channel identification, the data covariance matrix is built from vectors that exhibit the so-called □shift property;□ thus, successive data vectors are not statistically independent. Furthermore, existence of □colored noise□ due to the influence of long tails of □small□ leading or trailing impulse response terms, is practically inevitable. These terms should not be modeled because the quality of the estimate degrades dramatically and thus, their influence on the data covariance matrix can be considered as □colored noise.□ Hence, the assumptions on which information theoretic criteria are based do not hold true in the blind channel identification context and do not provide reliable effective channel order estimates.

In addition to the above shortcomings, research has shown that information theoretic criteria are

very sensitive to variations in the SNR and the number of data samples. This is because these methods depend on the estimation of the effective rank of a perturbed matrix using its singular values. This problem has been addressed in [11] where several criteria have been introduced.

However, most of the above criteria, use threshold values that either do not appear to be based on any explicit analytical expressions but are selected on an ad hoc basis, or lower and upper bounds for them can be derived analytically assuming known noise statistics, which is not true for blind identification. All the above reasons prohibit the successful application of information theoretic criteria for effective channel order estimation. If used, they create, arguably, some ambiguity concerning the classification of channels into under- or over-modeled. This, in turn, creates confusion regarding the robustness and applicability of blind channel identification methods under realistic conditions.

In order to overcome the shortcomings of information theoretic criteria, Liavas at all [6] proposed a new approach based on numerical analysis arguments. Using the concept of canonical angles between subspaces and invariant subspace perturbation results, they developed a criterion that provides a stable decomposition of the range space of an overmodeled data covariance matrix into signal and noise subspaces. When used for the effective channel order determination of measured channels, the proposed criterion proves to be insensitive to variations in the SNR and the number of data samples.

The algorithm permits a classification into stable or well-conditioned and unstable or ill-conditioned cases. In the stable cases, it provides useful effective channel order estimates, leading to sufficiently good blind channel approximation/equalization considerably improving information theoretic criteria. In the unstable cases, however, sufficiently good blind channel order determination seems difficult.

Even though the above algorithm performs well in the case of SIMO systems, no extension of it is available for MIMO systems. Thus, the problem of correct order determination has not been fully addressed in recent years, even though a broad category of blind identification algorithms could benefit from its exact knowledge [12, 13].

2.2 Blind Methods for LTI Systems Identification and Equalization

Blind methods of system identification/equalization aim to identify the characteristics of a system and its input by using measurements available only at the system's output. They are desired in all applications where the measurements at the input of the system are either not possible or too expensive to obtain. Such applications include speech reverberation cancellation, seismic signal analysis, brain

signal detection, deblurring of distorted images as well as blind channel equalization for wireless communications [14].

Generally, the system under consideration is assumed Linear Time Invariant and causal, described as

$$\mathbf{y}(n) = \mathbf{H}(n) * \mathbf{x}(n) + \mathbf{w}(n) = \sum_{l=0}^{L} \mathbf{H}(l) + \mathbf{w}(n) \tag{2.1}$$

where * denotes convolution, $\mathbf{x}(n)$ is the sequence of the input vectors of dimension $K \times 1$, $\mathbf{y}(n)$ is the sequence of the output vectors of dimension $M \times 1$, $\mathbf{H}(n)$ is the sequence of the system's impulse response of dimension $M \times K$, L is an upper bound of the length of the system's finite impulse response (FIR) and $\mathbf{w}(n)$ is the noise vector. Noise is assumed to be additive, white, guassian.

An equivalent form to describe the same system is

$$\mathbf{y}(n) = \mathbf{H}(z) * \mathbf{x}(n) + \mathbf{w}(n) \tag{2.2}$$

where $\mathbf{H}(z)$ is the system's transfer function defined as $\mathbf{H}(z) = \sum_{l=0}^{L} \mathbf{H}(l) z^{-l}$. The main purpose of a blind method is to estimate $\mathbf{x}(n)$ (blind equalization) and/or $\mathbf{H}(z)$ (blind identification) using $\mathbf{y}(n)$.

Given $\mathbf{y}(n)$ alone, the system cannot in general be either equalized or identified. Further information is needed. Specifically, even in the absence of noise, certain identifiability conditions are required so that $\mathbf{y}(n)$ can be used to compute $\mathbf{x}(n)$ and/or $\mathbf{H}(z)$ to some extent of practical interest. In the following subsections, we shall provide an overview of the blind identification methods and the identifiability conditions. We shall describe the most representative algorithms for three categories of systems: single-input single-output (SISO) systems, single-input multiple-output (SIMO) systems and multiple-input multiple-output (MIMO) systems.

2.2.1 SISO Systems

A SISO system results from the above general model when $M = K = 1$. It is described by the equation:

$$y(n) = h(z)x(n) + w(n) \tag{2.3}$$

Given the above description, one can always write another SISO system as:

$$y(n) = h_1(z)x_1(n) + w(n) \tag{2.4}$$

where $h_1(z) = \frac{h(z)}{h_0(z)}$ and $x_1(n) = h_0(z)x(n)$. For this reason, the SISO system is not identifiable by simply knowing the output $y(n)$, without knowing anything about the input. The identifiability conditions that generally required to blindly identify the system are the following:

- The second-order statistics of the input is known.
- The second-order statistics of the noise is known.
- Both the input and the noise are white sequences.
- All zeros of $h(z)$ lie inside (or outside) the unit circle

The latter condition is known as the minimum (or maximum respectively) phase condition. If the above conditions hold, $h(z)$ can be constructed uniquely up to a scalar [15].

2.2.2 SIMO Systems

A SIMO system occurs when $M > K = 1$. It is described by the equation:

$$y(n) = \mathbf{h}(z)x(n) + \mathbf{w}(n) \tag{2.5}$$

It arises when data at the receiver are collected either from multiple sensors or using oversampling. When $\mathbf{h}(z)$ is a constant vector (independent of z), the SIMO model is memoryless; otherwise it is an FIR SIMO system. The two types of systems are discussed separately in the following.

The memoryless SIMO system

As already mentioned, this system is described by the equation

$$y(n) = \mathbf{h}x(n) + \mathbf{w}(n) \tag{2.6}$$

One can estimate $\mathbf{h}$ and $x(n)$ using the least squares criterion, with N samples. Specifically, using this approach an estimate,

$$\hat{x}(n) = (\hat{\mathbf{h}}^H\hat{\mathbf{h}})^{-1}\hat{\mathbf{h}}^H\mathbf{y}(n) \tag{2.7}$$

is computed, where $\hat{\mathbf{h}}$ is the first principal eigenvector of $\mathbf{C}_{\mathbf{yy}}(0) = \frac{1}{N}\sum_0^{N-1}\mathbf{y}(n)\hat{\mathbf{y}}(n)^H = \sigma_{xx}^2\mathbf{h}\mathbf{h}^H + \mathbf{C}_{\mathbf{ww}}(0)$. Here, $\mathbf{C}_{\mathbf{yy}}(0)$ and $\mathbf{C}_{\mathbf{ww}}(0)$ are the covariance matrices of the output and the noise. If the noise is

spatially white, then its covariance matrix is asymptotically proportional to the identity matrix, leading to estimates that are asymptotically exact up to a scalar. In the absence of noise, the above estimate is exact, even when a short sequence of the system's output is used [15].

The SIMO system with memory

As with SISO blind equalization, some statistical assumptions about the system inputs and some conditions about the unknown system are required by most algorithms that use either SOS or HOS methods. In 1991, Tong, Xu and Kailath [9] proposed the blind identifiability and equalizibility of SIMO LTI systems, using only SOS of the system outputs. Their work led to a number of SOS based blind system estimation and equalization algorithms. All of them have the attractive characteristic that system estimates can always be obtained by optimizing a quadratic cost function.

Among the most widely used approaches used for SIMO LTI systems are the *Noise Subspace Approach* and the *Single-Stage Linear Prediction Approach*.

The Noise Subspace Approach

This approach exploits the separability of noise and signal subspaces for both system estimation and equalization, provided that the system order is known in advance. The following assumptions are used in general [18]:

- The system order is known
- The system is BIBO stable.
- The polynomial columns of the system transfer function have no common factors, i.e. they are mutually prime.
- The system input $x(n)$ is a stationary zero-mean random process with full rank covariance matrix $R_x(0)$. This condition holds for temporally white inputs. Not all subspace methods however, require that the input is temporally white.
- The noise vector sequence is a zero-mean spatially uncorrelated and temporally white vector random process with diagonal covariance matrix.
- The system input is statistically independent of the noise.

Under the above conditions the impulse response of the system can be uniquely identified, up to a nonzero scalar ambiguity. Having estimated the system impulse response, either a Zero Force (ZF) or a Minimum Mean Squared Error (MMSE) equalizer can be built. The basic advantage of the noise subspace approach is that it gives a closed form solution for system estimation. However, it has some shortcomings. First of all, it requires that the system order is either given or estimated correctly in advance. Actually, its performance is very sensitive to system order mismatch. Additionally, it tends to fail when the channel disparity condition is nearly violated [19, 20]. The method was modified [15] so that it remains robust in the presence of common subchannel zeros and system order overestimation errors. This was achieved by exploiting transmitter redundancy using a trailing zero precoder. The use of the precoding procedure limits applications of the modified subspace method. Another disadvantage of the subspace based approach, is its computational complexity, due to the eigenvalue decomposition of data correlation matrices of large dimension.

The Single-Stage Linear Prediction Approach

The fundamental concept of linear prediction approaches for system estimation arises from the observation that a SIMO linear system can be modeled as a vector autoregressive process under certain conditions. This allows for blind multichannel equalization and estimation using linear least squares estimation, producing a closed-form solution [15].

The assumptions used by linear prediction approaches, in general, are the following:

- The system is BIBO stable.
- The polynomial columns of the system transfer function have no common factors, i.e. they are mutually prime.
- The system input is stationary zero-mean temporally white random process with variance σ_x^2.

Under the above assumptions the Linear Prediction method [22] produces an estimate of the system input $x(n)$. With this estimate available, channel estimation can be performed using input-output system identification methods, such as for example, the least-squares method. The method does not need knowledge of the system order and is therefore robust against system order misestimation error. However, its performance depends critically on the accuracy of the estimation of the leading coefficient channel vector of the SIMO system. If this coefficient is small then the system input estimate computed by the algorithm may not be accurate. To improve this performance degradation some other prediction

error based approaches have been developed such the outer product decomposition algorithm and the multi-step linear prediction algorithm [23].

2.2.3 MIMO LTI Systems

The SOS based blind equalization algorithms presented in the previous sections for the SIMO case could be extendable for the MIMO case. Unfortunately, this is not possible when the system inputs are spatially uncorrelated and temporally white, due to inherent ambiguities of the SOS of the MIMO system outputs. Some subspace methods can estimate the MIMO system up to an upper triangular unimodular matrix ambiguity ([24, 25, 26] and references therein). To resolve this ambiguity further information about the system structure is needed. This however may not be available in practice. The assumptions used by subspace algorithms in this category are, in general, the following:

- The system orders are known, for all the diferrent subsystems that comprise the total system.
- There are more outputs than inputs.
- The source sequences $\mathbf{x}(n)$ are mutually independent and persistently exciting.
- The channel matrix is irreducible and column reduced.

Tugnait, in [27], presented a multistep linear prediction based approach that removed the columned reduced requirement.

To design less restrictive algorithms for spatially uncorrelated and temporally white system inputs, HOS methods should be used, that provide some advantages. These include provision of system phase information without requiring channel diversity, ability to resolve matrix ambiguity to pure scaling and permutation indeterminacies and asymptotic insensitivity to additive Gaussian noise. Blind MIMO system identification using HOS has received much research attention over the recent years [15, 28, 29, 30, 31, 32, 33, 34, 35, 36] and references there in.

Among the algorithms that have been developed, a broad class is that of equation-error based approaches [28, 29, 30, 31]. Giannakis et al. [37] proposed an algorithm that equalizes the system provided that only the lower order and the higher order channel coefficient matrix have full column rank. Later on, Swami and his colleagues [15] presented a unified Kronecker product formulation to define cumulants of vector processes of arbitrary orders. Based on these works, parameter estimation algorithms for causal and noncausal multichannel AR,MA and ARMA models were developed. Channel

estimation algorithms have been proposed which can handle MIMO systems with fewer outputs than inputs [15]. However, the algorithms' implementations still needs improvement to avoid a number of matrix rank tests that are not practical for estimated cumulant matrices.

Another category is that of HOS algorithms that are based on statistics matching or inverse filtering. These cumulant matching techniques [38, 39, 40] usually involve nonlinear optimization problems that require accurate initial conditions to avoid local convergence. Some of these algorithms adopt an iterative procedure that successively recovers each active source signal using a MIMO equalizer and then estimates the corresponding subchannels based on the recovered source signal and the observed channel outputs. Adaptive equalization approaches such as the constant modulus algorithm [41],[42] and the Shalvi-Weinstein algorithm [15] have been also utilized. Both algorithms rely on the premise of accurately converging MIMO equalizers; however they tend to suffer from error propagation error when the number of users increases.

Recently, J. Liang and Z. Ding [44] proposed a HOS based algorithm that utilizes the common null space of a set of fourth-order cumulant matrices to identify an uknown MIMO channel impulse response up to a constant monomial matrix. The assumptions made by this algorithm are given below:

- The system orders are known, for all the diferrent subsystems that comprise the total system.
- There are more outputs than inputs.
- Channel noises are zero mean, Gaussian stationary processes. They are mutually independent and also independent of the channel input signals.
- The source sequences $\mathbf{x_i}(n)$ are mutually independent and persistently exciting.
- The lower order channel matrix $H[0]$ consists of non zero columns and has full column rank.

For the case of spatially uncorrelated and temporally colored system inputs with distinct power spectra the identifiability of an irreducible MIMO system using SOS of the system output vector has been proven by Hua and Tugnait [45]. In addition, some SOS based blind system identification and equalization methods have been reported. Examples are the Gorokhov and Loubaton's subspace method [46], Abed-Meraim and Hua's minimum noise subspace method [47], and, as explained in [49]and references there in, the matrix pencil method proposed by Ma et al. , the blind identification via decorrelating subchannels (BIVDS) approach proposed by Hua et al, An and Hua's blind identification via decorrelating the whole channel (BIVDW) approach and so forth. All of these SOS based methods

require that the number of outputs is greater than the number of inputs. Moreover, both the BIVDS and BIVDW methods require that the power spectra of the driving inputs are sufficiently diverse, an assumption that may be invalid in some practical applications such as wireless communications. This is avoided by the Matrix Pencil method, that is described below:

The Matrix Pencil Method

The method makes the following assumptions:

- The system orders are known, for all the diferrent subsystems that comprise the total system.
- The system transfer function is irreducible and column reduced.
- The channel inputs are persistently exciting, and mutually independent with distinct nonwhite power spectra.

In summary, the matrix pencil approach is a two-stage approach. In the first stage a matrix-pencil is utilized between the two chosen system output correlation matrices at different lags for extracting filtered inputs. For the group of the extracted filtered inputs, SIMO blind equalization/identification algorithms are used to estimate true inputs.

2.3 Blind Methods for Nonlinear Identification and Equalization

2.3.1 Volterra Modelling of Nonlinear Systems

Modern high-speed communication systems are frequently operated over nonlinear channels with memory. For example, most transmitters are equipped with Power Amplifiers (PAs) operating close to saturation to achieve power efficiency. To properly analyze a communication system, the nonlinear effects caused by the presence of PAs must be combined with the transmitting, receiving and channel filters [50].

One of the most popular models that are applied for the description of nonlinear phenomena are Volterra series. Volterra series allow for capturing combined effects of PAs, transmitting, receiving and channel filters. A Volterra series is described by the equation:

$$\mathbf{y}(n) = \sum_{p=1}^{\infty} \sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_p=-\infty}^{\infty} \mathbf{h}_p(\tau_1, \cdots, \tau_p) \prod_{i=1}^{p} x(n-\tau_i) \tag{2.8}$$

Real physical systems are causal, of finite duration. Therefore, the Volterra kernels $\mathbf{h}_p(\tau_1, \cdots, \tau_p)$ are typically taken to have have finite memory length M_p, meaning that $\mathbf{h}_p(\tau_1, \cdots, \tau_p) = 0, \forall i : \tau_i > M_p$. Accordingly, the infinite series described above reduces to a finite summation if we choose a suitable upper bound P , such that all the important nonlinear characteristics of the system are captured. The above methodology leads to causal discrete Volterra series of finite-order that have the following form (also referred as passband Volterra systems):

$$\mathbf{y}(n) = \sum_{p=1}^{P} \sum_{\tau_1=0}^{M_p} \cdots \sum_{\tau_p=0}^{M_p} \mathbf{h}_p(\tau_1, \cdots, \tau_p) \prod_{i=1}^{p} x(n-\tau_i) \tag{2.9}$$

The finite Volterra series has many important properties that are useful in the analysis of nonlinear systems [52]. The most important among them are the following:

- **Linearity in the parameters.** The key feature of the Volterra series is that nonlinearities occur as multiple products of delayed input values. On the other hand, kernel coefficients appear linearly in the output and Volterra series can be modelled by Kronecker products. Define $\mathbf{x}_{M_i}^{(i)}(n) = [x(n), x(n-1), \cdots, x(n-M_i)]^T$ and the pth-order Kronecker power $\mathbf{x}_{M_p}^{(p)}(n) = \mathbf{x}_{M_1}^{(p)}(n)$. The Kronecker power contains all the pth-order products of the input. Likewise $\mathbf{h} = [\mathbf{h}_1(.), \cdots, \mathbf{h}_p(.)]^T$

is obtained by treating the $p-$dimensional kernel as a $(\mathbf{M}_p)^p$ column vector. Then the output $\mathbf{y}(n)$ can be rewritten as

$$\mathbf{y}(n) = \mathbf{x}^T(n)\mathbf{h} \tag{2.10}$$

The property of linearity allows linear estimation techniques to be exploited in the identification of Volterra coefficients.

- **Multidimensional convolution.** The p-th order Volterra kernel can be considered as the impulse response of a $p-$dimensional linear system, whose $p-$dimensional input is assumed to be the product of p one-dimensional products. Volterra series can therefore be described as multidimensional convolutions. To give an example, we describe the convolution model for a 2nd-order Volterra system in the following table:

Regressor vectors	**Coefficient vectors**
$x(n)\cdots x(n-M)$	$h_1(0)\cdots h_1(M)$
$x^2(n)\cdots x^2(n-M)$	$h_2(0,0)\cdots h_2(M,M)$
$x(n)x(n-1)\cdots x(n-M)x(n-M-1)$	$h_2(0,1)\cdots h_2(M-1,M)$
$\vdots$	$\vdots$
$x(n)x(n-M)$	$h_2(0,M)$

Table 1: Convolution structure of a 2nd-order Volterra system

The above representation can be helpful for designing Volterra systems as multidimensional linear systems. Moreover, using the above representation a pth-order homogeneous Volterra system may be realized via a p-dimensional convolution along the diagonal slice.

- **MISO interpretation.** The Single Input Single Output Volterra model can be reformulated as a Multiple Input Single Output (MISO) linear system. The regressor vector formulation of the input presented abnove, allows us to express the $p-$th order regressor as a sum of multichannel linear regressors. The above representation does not imply the existence of a multiple input single output physical system but is only an equivalent way of modelling, useful for identification and equalization purposes as it will be shown in the following chapters.

- **Symmetric, triangular kernels.** A Volterra system can be described by an equivalent one with symmetric or triangular kernels. The p-th order Volterra kernel h_p contains M^p Volterra kernels.

This representation assumes that each permutation of the indices $\tau_1, \tau_2, \cdots, \tau_p$ gives a different kernel. Without loss of generality the Volterra kernels can be considered to be symmetric. Thus, to obtain the equivalent, unique, symmetric kernel representation we add together the kernels for all the distinct permutation of the indices $\tau_1, \tau_2, \cdots, \tau_p$ and divide the sum by $p!$, that is the number of terms in the summation. The symmetric kernel $\mathbf{h}_p^{(s)}$ is, therefore, given by the equation:

$$\mathbf{h}_p^{(s)}(\tau_1, \tau_2, \cdots, \tau_p) = \frac{1}{p!} \sum_{all permutations} \mathbf{h}_p(\tau_1, \tau_2, \cdots, \tau_p) \tag{2.11}$$

The second type of kernels that are of interest are triangular kernels. A kernel, $\mathbf{h}_p(\tau_1, \tau_2, \cdots, \tau_p)$ is called triangular if $\mathbf{h}_p(\tau_1, \tau_2, \cdots, \tau_p) = 0$ for $\tau_1 < \tau_2 < \cdots < \tau_p$. Again, it should be noticed that any Volterra series can be described by an equivalent one with triangular kernels, without any loss of generality.

Interconnections. Many signal processing systems consist of building blocks properly interconnected. The three basic interconnections are the additive, multiplicative and cascade connections.

The additive connection of two Volterra homogeneous systems $\mathcal{H}_1$ and $\mathcal{H}_2$ is established by applying the same inputs to both systems and summing their outputs, that is

$$\mathbf{y}(n) = \mathcal{H}_1(x(n)) + \mathcal{H}_2(x(n)) = (\mathcal{H}_1 + \mathcal{H}_2)(x(n)) \tag{2.12}$$

Similarly, the multiplicative connection of two Volterra homogeneous systems is described by

$$\mathbf{y}(n) = \mathcal{H}_1(x(n))\mathcal{H}_2(x(n)) = (\mathcal{H}_1\mathcal{H}_2)(x(n)) \tag{2.13}$$

Both additive and multiplicative connections are commutative and associative.

The cascade connection of two systems is defined by applying the input to the first system $\mathcal{H}_1$ and taking its output sequence as input sequence for the second system $\mathcal{H}_2$, that is

$$\mathbf{y}(n) = \mathcal{H}_1(\mathcal{H}_2(x(n)) = (\mathcal{H}_1 * \mathcal{H}_2)(x(n)) \tag{2.14}$$

The cascade operation is not commutative., but distributive with respect to addition, multiplication in particular orderings. It can be shown that systems resulting from any of the three interconnections are again Volterra systems.

In most digital communication systems signals are assumed band-limited, either because of channel

bandwidth limitations (ADSL case) or regulatory constraints as in wireless communications. In such cases the signal bandwidth and the receiver filter should be defined very carefully depending on the application. Thus, the output signal only contains spectral components near the carrier frequency. To model this type of applications the *baseband* Volterra system is used. It is defined as:

$$y(n) = \mathbf{h}_0 + \sum_{p=1}^{P} \sum_{m_i=0}^{N-1} \cdots \sum_{m_p=0}^{N-1} \mathbf{h}_p(m_1, \cdots, m_{2k+1}) x(n-m_1) \cdots x(n-m_{k+1}) x^*(n-m_{k+2}) \cdots x^*(n-m_{2k+1}) \tag{2.15}$$

The above representation only considers odd order powers with one unconjugated input more than conjugated input. This way the output does not create spectral components outside the frequency band of interest [53].

In general, Volterra series based models are suitable to systems with moderate nonlinearities. Their applicability is limited to weakly nonlinear systems, of low order. This is because of the non-convergence of the series for strong non-linearities and the fact that the number of coefficients increases exponentially as a function of the memory length and the nonlinear order.

2.3.2 Higher Order Moments and

Higher order ststistics (HOS) are important for signal as well as image processing applications [51], as they can be used for the description of non-Gaussian processes, identification of non mimimum phase systems and identification of nonlinear systems. HOS can be defined in terms of moments and cumulants. Let $\mathbf{x}(n) = [x(1), x(2), \cdots, x(n)]$ be a real random vector. The n-th order moment of $\mathbf{x}(n)$ is defined as

$$E(\mathbf{x}(n)) = \int_{-\infty}^{\infty} \mathbf{x}(n) p(\mathbf{x}(n)) d\mathbf{x}(n)$$

where $E(.)$ denotes the expectation operator and $p(\mathbf{x}(n))$ is the probability density function of $\mathbf{x}(n)$.

The function $M_x(\omega(1), \omega(2), \cdots, \omega(n)) = M_x(\omega(n)) = E(exp(j\omega^T(n)\mathbf{x}(n)))$ is called the *moment generating function* of $\mathbf{x}(n)$, since the n-th order moment of $\mathbf{x}(n)$ can be obtained by its n-th partial derivative around zero multiplied by $(-j)^n$.

Having defined the function $M_x(\omega(n))$ we define the function $K_x(\omega(n)) = ln(exp(j\omega^T(n)\mathbf{x}(n)))$, which is called the *cumulant generating function*. In accordance to the n-th order moment of $\mathbf{x}(n)$, the n-th order cumulant of $\mathbf{x}(n)$ is defined as the n-th partial derivative of the cumulant generating function around zero multiplied by $(-j)^n$.

The n-th order cumulant of a vector $\mathbf{x}(n) = [x(1), x(2), \cdots, x(n)]$ will be denoted in the following as $cum\{x(1), x(2), \cdots, x(n)\}$.

Higher order moments and cumulants are statistical measures of correlation designed to go to zero whenever any one or more quantities under study become statistically independent of the rest. Therefore, cumulants generalize the concept of a correlation measure.

Cumulants have some interest algebraic properties, that are listed below:

- Cumulants are scalar. If $a_{i_{i=1}}^{n}$ are scalars and $x(i)_{i=1}^{n}$ are random variables, then

$$cum\{a_1x(1), a_2x(2), \cdots, a_nx(n)\} = \prod a_{i_{i=1}}^{n} cum\{x(1), x(2), \cdots, x(n)\} \tag{2.16}$$

- Cumulants are symmetric functions in their arguments. That is

$$cum\{x(1), x(2), \cdots, x(n)\} = cum\{x(i_1), x(i_2), \cdots, x(i_n)\} \tag{2.17}$$

  where $(i_1, i_2, \cdots, i_n)$ is any linear permutation of $(1, 2, \cdots, n)$

- Cumulants are multi linear mappings. That is, if a and are constants:

$$\begin{aligned} cum\{ax(1) + y(1), ax(2) + y(2), \cdots, ax(n) + y(n)\} \\ = acum\{x(1), x(2), \cdots, x(n)\} \\ +cum\{y(1), y(2), \cdots, y(n)\} \end{aligned} \tag{2.18}$$

- Suppose that $x(i)_{i=1}^{n}$ and $y(i)_{i=1}^{n}$ are statistically independent random variables. Then,

$$\begin{aligned} cum\{x(1) + y(1), \cdots, x(n) + y(n)\} = cum\{x(1), x(2), \cdots, x(n)\} + \\ cum\{y(1), y(2), \cdots, y(n)\} \end{aligned} \tag{2.19}$$

- Cumulants are blind to additive constant. That is, if a is a constant, then:

$$cum\{a + x(1), x(2), \cdots, x(n)\} = cum\{x(1), x(2), \cdots, x(n)\}$$

- Suppose that a subset of the random variables $x(i)_{i=1}^{n}$ are statistically independent. Then,

$$cum\{x(1), x(2), \cdots, x(n)\} = 0$$

- If $x(n)$ is a k-th order stationary process then, the k-th order cumulants depend only on the time lags between the variables and not the specific variable values.

2.3.3 Identification and Equalization Algorithms for Nonlinear Systems

Most of the existing work in the field of nonlinear system modeling involves identification methods based on the availability of the input as well as the output signals. Blind identification and equalization of nonlinear systems is therefore an emerging area of advanced research. As it has been mentioned in the previous chapter, blind equalization approaches are of interest in digital communications since no training input and no interruption of the transmission are necessary to equalize the channel. Therefore, for channels exhibiting multipath phenomena, time-varying nonlinearitiies, or high data rates, blind methods are attractive [52, 64].

Data clustering techniques ([53] and references there in) used for blind equalization or identification of nonlinear channels require no explicit channel modeling. Blind decision feedback equalization is one of these methods ([54] and references there in). A drawback of the method is that it assumes that all the past detected symbols are correct, therefore it is not robust to error propagation. Significant work in the field has also been done in [65, 66, 67, 68]

Canonical picewise linear models were also proposed in for the blind equalization of nonlinear channels with memory, as a generalization of Donoho's minimum entropy deconvolution approach. An Maximum Likelihood algorithm for blind identification using expectation maximization was also proposed. However, the resulting method is suboptimal. Although ML provides good performance they suffer from computational complexity and the existence of local minima ([55] and references there in. Blind equalization of satellite links is dealt in [56] through Bayesian methods using Markov Chain Monte Carlo techniques. The disadvantage of this type of methods is that they require large data

samples. Oversampling is used to overcome this problem.

An approach for blind identification of nonlinear single-input, multiple-output (SIMO) truncated Volterra filters has been presented in [57]. It is shown that while impossible with a single output, several outputs make it possible to blindly deconvolve several Volterra channels simultaneously. This approach requires that the input sequence satisfies a persistent excitation condition and the channel matrix has a full row rank. The input could be deterministic or random with unknown distribution. The method, unlike previous algorithms, does not require knowledge of the higher-order statistics of the input, for equalization. The nonlinear channels are equalized using linear FIR filters for modelling. This can be justified intuitevely since the vector equalizer can be seen as a beamformer that, thanks to its diversity, is capable of nulling the nonlinearities and equalizing the linear part. The results, involving simulated magnetic recording channels as well as real and complex valued communication channels look promising. However, a number of questions, about the potential capability of the method are still open research issues. These involve analytic performance evaluation, the selection of optimum equalizer delay, order determination of the Volterra data model as well as the effect of high SNR values to the algorithm. The complexity is relatively high, and therefore, additional work on efficient on-line versions are required before the method may be used in practice.

In [64] a blind equalization algorithm for SIMO nonlinear systems was introduced that builts zero forcing (ZF) linear equalizers provided certain conditions are satisfied. Compared to the work in [57] it relaxes the assumption that the associated channel matrix is full rank and square. Eventhough squareness can always be achieved, if necessary, by decreasing the number of channels and increasing the equalizer length, a longer equalizer would increase the computational complexity; further, if some channels are to be dropped, it is not possible to ensure a priori that the surviving channels satisfy the corresponding full-rank condition, even if the original set did. Thus, the selection of the channels to drop is a difficult problem. Another advantage of the algorithm is that equalizers of any delay can be computed as opposed to the work in [57] that computes equalizers of zero and maximum delays. However, both the above algorithms assume that channel orders are known, an assumption that does not hold true in practice, especially for nonlinear systems.

2.4 Blind Source Separation (BSS)

When sensors are used to collect information or oversampling is used, the signals provided are linear mixtures of the signals of interest. We then have the problem of recovering the signals of interest, named as sources, from the observed signals. This problem, commonly known as Blind Source Separation (BSS), is aggravated by the fact that in many practical applications we cannot directly observe the sources nor the way that they are mixed. The BSS problem plays an important role in signal processing. It can be solved using a minimum amount of prior information, namely that the mixing system is invertible and that the sources are mutually independent. This model robustness makes BSS attractive to a large number of extremely diverse applications such as array processing, multiuser communications, signal restoration andbiomed ical engineering.

Specifically, we assume that we are given at the receiver some linear mixtures $\mathbf{x}_i(n)$ of a number of source signals $s_j(n), j = 1, 2, \cdots, P$ that obey the equation:

$$x_i(n) = \sum_{j=1}^{P} a_{ij} s_j(n) \tag{2.20}$$

or in matrix form,

$$\mathbf{X} = \mathbf{AS} \tag{2.21}$$

The nth column of $\mathbf{X}$ is $\mathbf{x}^t(n) = [x_1(n), \cdots, x_M(n)]$. Matrix $\mathbf{A}$ with elements a_{ij} has size $M \times P$, while the matrix of source signals $\mathbf{S}$ has column vectors $\mathbf{s}(n) = [s_1(n), \cdots, s_P(n)]$. The goal of BSS is to recover $\mathbf{A}$ from the observed mixtures $\mathbf{X}$.

As it is explained in [49] and references there in, since the pioneering work of Jutten and Herault, many efficient and robust algorithms for BSS have been proposed and their properties investigated. These algorithms have been developed from different points of view such as contrast functions , maximum likelihood estimation, information transfer maximization , Kullback□Leibler divergence minimization using the natural gradient approach , and non-linear principal component analysis (PCA) [69]. Despite their disparity, all these rules have a common characteristic: the utilization of non-linear functions of their outputs. This non-linearity is required because second-order statistics are not sufficient to solve the BSS problem. However, their convergence properties are extremely dependent on the sources distribution and the non-linearities used. Moreover, these properties are considerably

different when there is Gaussian noise in the mixture.

In addition to the above algorithms, other algorithms and methods such as Joint Diagonalization and the JADE algorithm, the Extended- Infomax algorithm, the Fast-ICA algorithm and the Zarzoso□Nandi algorithm , have been shown to be able to separate sources from a broadclass of distributions ([69] and references there in).

Joint Diagonalization is a BSS technique. Among the methods that perform joint diagonalization the JADE algorithm mentioned above is a popular approach for determining matrix **A** using cumulants. At first JADE, estimates a data whitening matrix **W** and sets $\mathbf{Z} = \mathbf{WX}$. Then, using fourth order cumulants of the input signals, it computes a maximal set of cumulant matrices to be jointly diagonilized. At a third step, it computes the rotation matrix **R** that makes the cumulant matrices as diagonal as possible and estimates the matrix **A** as $\mathbf{A} = \mathbf{RW}^{-1}$.

3. Order Determination

In this chapter a method for the solution of the order determination is presented. First, an overview of the necessary algebraic background and related definitions is provided. Next, we introduce for both LTI FIR MIMO systems as well nonlinear systems described by Volterra series. A set of theorems and lemmas is presented and finally the algorithm is introduced. Simulation results presented at the end of the chapter, validate the performace of the method compared to existing schemes.

3.1 Objectives

Given an LTI FIR system or a nonlinear system described by a finite Volterra series, our objective is to establish that

(a) The number of discrete inputs can be computed.

(b) The orders (memory lengths) of the subsystems that comprise the total system are determined.

(c) Computations can be carried out by an efficient algorithm. Specifically, if $L_1, L_2, \cdots, L_P$ are the different subsystem orders, they may be grouped into r distinct numbers J_1, J_2, $\cdots$, J_r such that $J_1 < J_2 < \cdots < J_r$. Then for all i, $1 \leq i \leq r$, the number m_i of subsystems that have order J_i can be computed by the proposed algorithm.

3.2 Prerequisites

The notation employed is standard. Signals are discrete-time and complex in general. Upper- and lower-case bold letters denote matrices and vectors respectively. $(\cdot)^t$ and $(\cdot)'$ are transpose and Hermitian operations. $\mathbf{0}_{m\times n}$ stands for the $m \times n$ zero matrix. Given a matrix $\mathbf{A}$, $\mathcal{R}(\mathbf{A})$ and $\mathcal{C}(\mathbf{A})$ stand for the row and column space of matrix $\mathbf{A}$ and $||\mathbf{A}||_\infty$ [1] denotes the ∞ norm of the matrix. For a given matrix $\mathbf{X}$ having the same number of columns as $\mathbf{A}$, $\mathcal{P}_{\mathbf{A}}\{\mathbf{X}\}$ denotes the projection of $\mathbf{X}$ onto the row space of $\mathbf{A}$. For a set of vectors $\mathbf{x}_1, \cdots, \mathbf{x}_n$, $sp\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is the linear subspace spanned by $\mathbf{x}_1, \cdots, \mathbf{x}_n$. If A, B are vector spaces over the same field F, $A \oplus B$ is their direct sum and $A \doteq B$ denotes that they are isomorphic, that is there exists an one to one and onto linear map from A to B. If $\mathbf{x}$, $\mathbf{X}$ are a vector and a matrix respectively and A is a vector space, then $\mathbf{x}_{|A}$ is the orthogonal projection of $\mathbf{x}$ onto A and $\mathbf{X}_{|A}$

[1] If $\mathbf{A}$ is an m rows by n columns matrix then $||\mathbf{A}||_\infty$ is defined as $||\mathbf{A}||_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|$

is the orthogonal projection of the rows of $\mathbf{X}$ onto A. Finally, $\mathcal{N}$ denotes the set of positive integers.

3.3 Model Description

3.3.1 LTI FIR MIMO Systems

We shall be concerned with FIR MIMO systems of the form:

$$\mathbf{x}(k) = \sum_{i=1}^{P} \sum_{j=0}^{L_i} \mathbf{h}_i(j) s_i(k-j) \tag{3.1}$$

The system has P inputs and M outputs. Thus the output signal $\mathbf{x}(k)$ is an $M \times 1$ dimensional vector. The input sequences consist of the signals $s_1(k)$, $s_2(k)$, $\cdots$, $s_P(k)$. The orders of the P subsystems are given by the integers $L_1, L_2, \cdots, L_P$. For each $1 \leq i \leq P$, and $0 \leq j \leq L_i$, $\mathbf{h}_i(j)$ is the corresponding $M \times 1$ kernel tap.

Equivalently, the above system can be described as

$$\mathbf{x}(k) = [\mathbf{H}(z)]\mathbf{s}(k) \tag{3.2}$$

where $[\mathbf{H}(z)]$ is the system transfer function and $\mathbf{s}(k) = [s_1(k) \cdots s_P(k)]^t$.

If L_q, $1 \leq q \leq P$, denotes the maximum of $L_1, L_2, \cdots, L_P$, then the channel polynomial matrix $\mathbf{H}(z)$ can be written as

$$\mathbf{H}(z) = \sum_{i=0}^{L_q} \mathbf{H}(i) z^{-i} \tag{3.3}$$

with

$$\mathbf{H}(i) = [\mathbf{h}_1(i)\mathbf{h}_2(i) \cdots \mathbf{h}_P(i)] \tag{3.4}$$

The memory lengths of the P subsystems are given by the integers $L_1, L_2, \cdots, L_P$. In the general case, some of the above integers may be equal to each other. Therefore, the r distint integer values appearing in the set $L_1, L_2, \cdots, L_P$ that denote the different subsystem orders, will in the following

be denoted by $J_1, J_2, \cdots, J_r$ and we shall assume, without loss of generality, that $J_1 < J_2 < \cdots < J_r$. In addition, for all i, $1 \leq i \leq r$, m_i will denote the number of subsystems attaining the same order J_i.

The following assumptions are made:

A1) The input sequences $s_1(k), s_2(k), \cdots, s_P(k)$ are stationary Independent Identically Distributed (I.I.D.) zero mean signals of finite variance that are mutually independent with each other.

A2) An upper bound L of the subsystems' orders is known.

A3) The number of inputs P is strictly less than the number of outputs M. Furthermore, the channel polynomial matrix **H**(z) is irreducible and column reduced.

As a result [70], for a smallest $w_0 > 0$, $w_0 \in \mathcal{N}$ so that the matrix $\mathbf{H}_{w_0}(\mathbf{h}) = [\mathbf{F}_{w_0}(\mathbf{h}_1) \cdots \mathbf{F}_{w_0}(\mathbf{h}_P)]$ where

$$\mathbf{F}_{w_0}(\mathbf{h}_i) = \begin{pmatrix} \mathbf{h}_i(0) & \cdots & \mathbf{h}_i(L_i) & \cdots & \mathbf{0}_{M\times 1} \\ \vdots & \ddots & \cdots & \ddots & \vdots \\ \mathbf{0}_{M\times 1} & \cdots & \mathbf{h}_i(0) & \cdots & \mathbf{h}_i(L_i) \end{pmatrix}$$

has full column rank. $\mathbf{F}_{w_0}(\mathbf{h}_i)$ is an $(Mw_0) \times (L_i + w_0)$ matrix. As stated in [70], to guarantee the full column rank of $\mathbf{H}_{w_0}(\mathbf{h})$ it suffices to select $w_0 \geq \sum_{i=1}^{P} L_i - 1$.

3.3.2 Discrete Volterra Systems

The nonlinear systems under consideration are assumed to be SIMO, discrete-time, time-invariant, causal and of finite memory. Furthermore, we assume that any small changes to the system's input $s(n)$ result in small changes in the system's output. Any such system [54] can be approximated over a uniformly bounded set of input signals by a truncated Volterra series expansion of finite order P.

The output $y(n)$ of a real valued SIMO Volterra system, as stated in [43], can be described as:

$$x(n) = \mathbf{h}_0 + \sum_{p=1}^{P} \sum_{m_i=0}^{N-1} \cdots \sum_{m_p=0}^{N-1} \mathbf{h}_p(m_1, \cdots, m_p) s(n-m_1) \cdots s(n-m_p) + \eta(n) \tag{3.5}$$

As stated in [53], for a narrowband communication system input-output relationship is described

by the equation:

$$x(n) = \mathbf{h}_0 + \sum_{p=1}^{P} \sum_{m_i=0}^{N-1} \cdots \sum_{m_p=0}^{N-1} \mathbf{h}_p(m_1, \cdots, m_{2k+1}) s(n-m_1) \cdots s(n-m_{k+1}) s^*(n-m_{k+2}) \cdots s^*(n-m_{2k+1}) + \eta(n) \tag{3.6}$$

We shall assume that $\eta(n) = 0$ (noiseless case) and examine the effect of noise in simulations. In the above equations , $s(n)$ is the system's input, $\forall 1 \leq p \leq P, \mathbf{h}_p$ are the discrete $p-th$ order Volterra kernels and P is the order of nonlinearity. Furthermore, without loss of generality, we may choose the range $m_1, \cdots, m_p$ so that $0 \leq m_1 \leq m_2 \cdots \leq m_p$, that is the Volterra kernels are upper triangular.

Using the aproach followed in [57] we may cast any SIMO Volterra system as an equivalent MIMO linear system, by appropriate redefinition of the vector kernels and the system's input.

Let M denote the number of output channels, then the vector kernel $\mathbf{h}_p(m_1, \cdots, m_p)$ can be indexed as:

$$\mathbf{h}_p(m_1, \cdots, m_p) = \mathbf{h}_{p,i_1:i_{p-1}}(m) = [h_p^1(m, m+i_1, \cdots, m+i_{p-1}) \cdots h_p^M(m, m+i_1, \cdots, m+i_{m-1})] \tag{3.7}$$

where $0 \leq i_1 \leq \cdots \leq i_{k-1} \leq N-1$ and $m = 0, 1, \cdots N-1-i_{p-1}$. To compactify the notation we use $i_1 : i_{p-1}$ to denote the set $(i_1, \cdots, i_{p-1})$. For $p = 2$, the range notation $(i_1 : i_1)$ denotes just the index i_1.

Similarly, we define the signals:

$$s_{p,i_1:i_{p-1}}(m) = s(m)s(m-i_1)\cdots s(m-i_{p-1}) \tag{3.8}$$

with $0 = i_0 \leq i_1 \leq \cdots \leq i_{p-1} \leq N-1$. We denote $\mathbf{h}_1 = \mathbf{h}_{1,i_0:i_0}$ and $s_1 = s_{1,i_0:i_0}$. Using the change of variables $m_p = m + i_{p-1}$, for $m = 1, 2, \cdots, P$ we have that:

$$\mathbf{x}(n) = \sum_{p=1}^{P} \sum_{0 \leq i_1 \leq \cdots \leq i_{p-1} \leq N-1} \sum_{m=0}^{N_p - i_{p-1}} \mathbf{h}_{p,i_1:i_{p-1}}(n) s_{p,i_1:i_{p-1}}(n-l) \tag{3.9}$$

Equation (4.5) describes a MIMO channel. Finally, for an appropriate B, we introduce the variables $u_i, 1 \leq i \leq B$ and use lexigographical ordering to name the sequences $u_i(n) = s_{p,i_1:i_{p-1}}(n)$. Then, applying the same technique and introducing variables $\mathbf{g}_i(m)$ we rename the kernels $\mathbf{g}_i(m) = \mathbf{h}_{p,i_1:i_{p-1}}(m)$. Taking all the forementioned steps, the above equation takes the form

$$\mathbf{x}(n) = \sum_{i=1}^{B} \sum_{j=0}^{K_i} \mathbf{g}_i(j) u_i(n-j) = [\mathbf{G}(z)]\mathbf{u}(n) \tag{3.10}$$

where $[\mathbf{G}(z)]$ is the system transfer function and $\mathbf{u}(k) = [u_1(k) \cdots u_B(k)]^t$. An example of how a SIMO Volterra system is cast as a MIMO system is given in [57] for $P = 2$.

In general, $B \neq P$, so the number of the subsystems appearing in the MIMO representation is not equal to the order of the system's non-linearity. The memory lengths of the B subsystems are given by the integers $K_1, K_2, \cdots, K_B$. In the general case, some of the above integers may be equal to each other. Therefore, the r distint integer values appearing in the set $K_1, K_2, \cdots, K_B$ that denote the different subsystem orders, will in the following be denoted by J_1, J_2, $\cdots$, J_r and we shall assume, without loss of generality, that $J_1 < J_2 < \cdots < J_r$. In addition, for all i, $1 \leq i \leq r$, m_i will denote the number of subsystems attaining the same order J_i.

We shall make the following assumptions:

B1) The input sequence $s(n)$ is zero mean, i.i.d., with values in a finite alphabet of at least $P + 1$ complex numbers. Examples include PSK or QAM signals.

B2) The system transfer matrix **G**(z) is irreducible. This guarantees that there are no common zeros in the (FIR) transfer functions of every pair of the subchannels involved. It is a common assumption in all methods based on Second Order Statistics.

B3) The memory of the linear kernel is strictly greater than the memory of any nonlinear term. Moreover, the first element of the zero-th tap of this kernel is equal to unity.

B4) An upper bound L of B is known.

3.3.3 Model Equivalence

As shown above, we may cast a nonlinear system described by a finite Volterra series as a linear MIMO system. It should be noticed that there is no physical equivalence of the two types of systems since the inputs of the casted Volterra system are products of the original input. However, the casting is useful for algebraic manipulations. Due to the casting, in the analysis presented in the following

chapters, we shall use the model as described for LTI FIR MIMO systems. It is obvious that this is without any loss of generality. Thus, the system under consideration is described as:

$$\mathbf{x}(k) = \sum_{i=1}^{B} \sum_{j=0}^{L_i} \mathbf{h}_i(j) s_i(k-j) \tag{3.11}$$

or equivalently

$$\mathbf{x}(k) = [\mathbf{H}(z)]\mathbf{s}(k) \tag{3.12}$$

where $[\mathbf{H}(z)]$ is the system transfer function and $\mathbf{s}(k) = [s_1(k) \cdots s_P(k)]^t$.

If L_q, $1 \leq q \leq P$, denotes the maximum of $L_1, L_2, \cdots, L_P$, then the channel polynomial matrix $\mathbf{H}(z)$ can be written as

$$\mathbf{H}(z) = \sum_{i=0}^{L_q} \mathbf{H}(i) z^{-i} \tag{3.13}$$

with

$$\mathbf{H}(i) = [\mathbf{h}_1(i)\mathbf{h}_2(i) \cdots \mathbf{h}_P(i)] \tag{3.14}$$

Again, the r distint integer values that denote the different subsystem orders, will in the following be denoted by $J_1, J_2, \cdots, J_r$ and we shall assume, without loss of generality, that $J_1 < J_2 < \cdots < J_r$. In addition, for all i, $1 \leq i \leq r$, m_i will denote the number of subsystems attaining the same order J_i.

3.4 Variables, Projections and Isomorphic Relations

3.4.1 Definition of Data Structures

The proposed approach extends the work by Tong et al, described in [58]. Linear prediction and smoothing techniques that exploit the isomorphic relationship between input and output subspaces are employed. Precise definitions of relevant variables, spaces and projections are given next.

Next we proceed with some definitions. Assuming that $k, Q \in \mathcal{N}$ and $Q > k$. We collect successive output vectors in the matrix

$$\mathbf{T}(k) = [\mathbf{x}(k) \quad \mathbf{x}(k+1) \quad \cdots \mathbf{x}(Q)] \tag{3.15}$$

Likewise, we combine successive input values in row vector form as follows

$$\begin{aligned} \mathbf{s}_1(k) &= [s_1(k) \quad s_1(k+1) \quad \cdots s_1(Q)] \\ \vdots \quad &\quad \vdots \\ \mathbf{s}_B(k) &= [s_B(k) \quad s_P(k+1) \quad \cdots s_B(Q)] \end{aligned} \tag{3.16}$$

Using a window of length w, we form the following array of output samples:

$$\mathbf{X}_{k,w} = \begin{pmatrix} \mathbf{T}(k) \\ \mathbf{T}(k-1) \\ \vdots \\ \mathbf{T}(k-w+1) \end{pmatrix} \tag{3.17}$$

$\mathbf{X}_{k,w}$ is the data matrix defined by stacking w consecutive such observations, starting with $\mathbf{T}(k)$ and going back to $\mathbf{T}(k-w+1)$. Similarly, for all $j, 1 \leq j \leq B$, we have

$$\mathbf{S}^j_{k,w} = \begin{pmatrix} \mathbf{s}_j(k) \\ \mathbf{s}_j(k-1) \\ \vdots \\ \mathbf{s}_j(k-w+1) \end{pmatrix} \tag{3.18}$$

Next, we consider the row spaces

$$\mathcal{X}_{k,w} = \mathcal{R}\left(\mathbf{X}_{k,w}\right) \tag{3.19}$$

$$\mathcal{S}^j_{k,w} = \mathcal{R}\left(\mathbf{S}^j_{k,w}\right) \tag{3.20}$$

Finally, for any $l \in \mathcal{N}$ we define the vector space

$$\begin{aligned} \dot{\mathcal{S}}_{k,l} &= \left(\mathcal{S}^1_{k-1,L_1+w} \oplus \cdots \oplus \mathcal{S}^B_{k-1,L_B+w}\right) \\ &\bigcup \left(\mathcal{S}^1_{k+l+w,L_1+w} \oplus \cdots \oplus \mathcal{S}^B_{k+l+w,L_B+w}\right) \end{aligned} \tag{3.21}$$

If we think of l as a smoothing window, we see that $\dot{\mathcal{S}}_{k,l}$ is a subspace constructed from both past and future data observations.

3.4.2 Lemmas and Theorems

In this subsection, we state and prove lemmas and theorems that establish the validity of the algorithm. First, we shall use the following lemma, the validity of which is established in [70]:

Lemma 1: For all $w \geq w_0$, $\mathbf{H}_w(\mathbf{h})$ has full column rank.

Based on the above lemma we present Theorem 1, which establishes the isomorphic relationship between input and output subspaces.

Theorem 1: For all $w \geq w_0$, $\mathcal{X}_{k,w} \doteq \mathcal{S}^1_{k,L_1+w} \oplus \cdots \oplus \mathcal{S}^B_{k,L_B+w}$.

The definitions preceding the theorem lead to

$$\mathbf{X}_{k,w} = \mathbf{H}_w(\mathbf{h}) \begin{pmatrix} \mathbf{S}^1_{k,L_1+w} \\ \mathbf{S}^2_{k,L_2+w} \\ \vdots \\ \mathbf{S}^B_{k,L_B+w} \end{pmatrix}$$

Due to Lemma 1, $\mathbf{H}_w(\mathbf{h})$ has full column rank, therefore

$$\mathcal{X}_{k,w} \doteq \mathcal{R}\left(\begin{pmatrix} \mathbf{S}^1_{k,L_1+w} \\ \mathbf{S}^2_{k,L_2+w} \\ \vdots \\ \mathbf{S}^B_{k,L_B+w} \end{pmatrix} \right) \tag{3.22}$$

When the system under consideration is LTI FIR MIMO the rows of the matrix

$$\begin{pmatrix} \mathbf{S}^1_{k,L_1+w} \\ \mathbf{S}^2_{k,L_2+w} \\ \vdots \\ \mathbf{S}^B_{k,L_B+w} \end{pmatrix} \tag{3.23}$$

are orthogonal with respect to the inner product induced by correlation in the Hilbert space of zero mean random vectors with finite second moments. This is due to the I.I.D. and mutual independence assumptions stated in A1). Therefore, the rows of the above matrix are linearly independent with probability one.

For SIMO Volterra systems, it suffices to notice that the above matrix equals the transpose of matrix $\mathbf{S}^{(0,Q+k)}$ as defined in [57]. There, it is proved that its rows are linearly independent with probability one, provided that assumption B1 holds. This, proves the theorem.

Theorem 1 imlies that

$$\mathcal{X}_{k-1,w} \doteq \left(\mathcal{S}^1_{k-1,L_1+w} \oplus \cdots \oplus \mathcal{S}^B_{k-1,L_B+w}\right) \tag{3.24}$$

and

$$\mathcal{X}_{k+l+w,w} \doteq \left(\mathcal{S}^1_{k+l+w,L_1+w} \oplus \cdots \oplus \mathcal{S}^B_{k+l+w,L_B+w}\right) \tag{3.25}$$

Using (3.21), (3.24) and (3.25) we conclude that,

$$\dot{\mathcal{S}}_{k,l} \doteq \mathcal{X}_{k-1,w} \oplus \mathcal{X}_{k+l+w,w} \tag{3.26}$$

Before proceeding, we define the projection error matrix and the kernel-input product matrices that are necessary to the development of the algorithm.

The projection error matrix $\mathbf{E}_{k,l}$ is defined as

$$\mathbf{E}_{k,l} = \begin{pmatrix} \mathbf{T}(k+l) - \mathbf{T}(k+l)_{|\dot{\mathcal{S}}_{k,l}} \\ \mathbf{T}(k+l-1) - \mathbf{T}(k+l-1)_{|\dot{\mathcal{S}}_{k,l}} \\ \vdots \\ \mathbf{T}(k) - \mathbf{T}(k)_{|\dot{\mathcal{S}}_{k,l}} \end{pmatrix}$$

$\mathbf{E}_{k,l}$ is a $(M(l+1)) \times (Q-k+1)$ matrix. Each block entry $\mathbf{T}(k+m) - \mathbf{T}(k+m)_{|\dot{\mathcal{S}}_{k,l}}$ of $\mathbf{E}_{k,l}$ is formed by the error resulting from the projection of $\mathbf{T}(k+m)$ on the space $\dot{\mathcal{S}}_{k,l}$ generated by past and future input values.

In the following analysis,we shall fix k, without loss of generality since the carried out analysis is valid for any choice of it. For all $i, 1 \leq i \leq r$, let $\mathbf{h}_{i_1}, \mathbf{h}_{i_2}, \cdots, \mathbf{h}_{i_{m_i}}$, be the m_i subsystems that attain order J_i. Denote by $s_{i_1}, s_{i_2}, \cdots, s_{i_{m_i}}$ the inputs to the above subsystems.

Then, for a given subsystem $\mathbf{h}_{i_s}$, for all $l, J_1 \leq l \leq L$ and $t, 1 \leq t \leq m_i$, we define the matrix:

$$\mathbf{D}_{l,J_i}(\mathbf{h}_{i_t}) = \begin{pmatrix} \mathbf{h}_{i_t}(0) & \cdots & \mathbf{0}_{M\times 1} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{i_t}(J_i) & \vdots & \mathbf{h}_{i_t}(0) \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{M\times 1} & \cdots & \mathbf{h}_{i_t}(J_i) \end{pmatrix},$$

Next, for all subsystems attaining the same order J_i, we form the matrix:

$$\mathbf{G}_{l,J_i} = \begin{pmatrix} \mathbf{D}_{l,J_i}(\mathbf{h}_{i_1}) & \cdots & \mathbf{D}_{l,J_i}(\mathbf{h}_{i_{m_i}}) \end{pmatrix},$$

Collect input values of the above subsystems, to form the matrix:

$$\tilde{\mathbf{S}}_{l,J_i}=\begin{pmatrix} \mathbf{s}_{i_1}(k+l-J_i) \\ \vdots \\ \mathbf{s}_{i_1}(k) \\ \vdots \\ \mathbf{s}_{i_{m_i}}(k+l-J_i) \\ \vdots \\ \mathbf{s}_{i_{m_i}}(k) \end{pmatrix}$$

Finally, for all $i, 1 \leq i \leq r$ and for all l, $J_i \leq l \leq L$ we define the kernel-input matrices

$$\mathbf{\Psi}_{l,J_i} = \mathbf{G}_{l,J_i}\tilde{\mathbf{S}}_{l,J_i} \tag{3.27}$$

Having defined the projection error matrix and the kernel-input product matrices we proceed to Theorem 2.

Theorem 2

(i) For any $l : l < J_1$, $\mathbf{E}_l = 0$

(ii) For $l : J_1 \leq l < J_2$,

$$\mathbf{E}_l = \mathbf{\Psi}_{l,J_1} \tag{3.28}$$

(iii) For $n : 3 \leq n \leq r$ and $l : J_{n-1} \leq l < J_n$,

$$\mathbf{E}_l - \sum_{m=1}^{n-2} \mathbf{\Psi}_{l,J_m} \quad = \quad \mathbf{\Psi}_{l,J_{n-1}} \tag{3.29}$$

(iv) For $l : J_r \leq l \leq L$,

$$\mathbf{E}_l - \sum_{m=1}^{r-1} \mathbf{\Psi}_{l,J_m} \quad = \quad \mathbf{\Psi}_{l,J_r} \tag{3.30}$$

We first observe that the dimension of $\mathbf{D}_{l,J_i}(\mathbf{h}_{(i_s)})$ is $M(l+1) \times (l - J_i + 1)$.

We have that:

(i) $l < J_1$: By the definition of $\dot{\mathcal{S}}_{k,l}$ (see (3.21)) we see that for all $i, j, \quad 1 \leq i \leq P, \quad 0 \leq j \leq L_i$: $\mathbf{s}_i(k+l-j) \in \dot{\mathcal{S}}_{k,l}$. Therefore, $\mathbf{T}(k+l)_{|\dot{\mathcal{S}}_{k,l}} = \mathbf{T}(k+l)$ and $\mathbf{E}_{k,l} = 0$.

(ii) $J_1 \leq l < J_2$: Let $0 \leq q \leq l$. All vectors of the form $\mathbf{s}_{i_s}(k+q-j)$ where $1 \leq s \leq m_1$ and $q-l+J_1 \leq j \leq q$ do not belong to $\dot{\mathcal{S}}_{k,l}$. Therefore, we have that

$$\mathbf{T}(k+q) - \mathbf{T}(k+q)_{|\dot{\mathcal{S}}_{k,l}} = \sum_{s=1}^{m_1} \sum_{j=q-l+J_1}^{q} \mathbf{h}_{i_s}(j)(\mathbf{s}_{i_s}(k+q-j) - \mathbf{s}_{i_s}(k+q-j)_{|\dot{\mathcal{S}}_{k,l}}) \tag{3.31}$$

The assertion follows if we take into account that, by definition, $\mathbf{h}_{i_s}(j) = 0$ for $j \leq 0$ or $j > J_1$ and write the previous equation in matrix form.

(iii) For all $n, 3 \leq n \leq r$, let $J_{n-1} \leq l < J_n$ and $0 \leq q \leq l$. Similar arguments lead to

$$\begin{aligned}\mathbf{T}(k+q) - \mathbf{T}(k+q)_{|\dot{\mathcal{S}}_{k,l}} = &\sum_{s=1}^{m_1} \sum_{j=q-l+J_1}^{q} \mathbf{h}_{i_s}(j)(\mathbf{s}_{i_s}(k+q-j) - \mathbf{s}_{i_s}(k+q-j)_{|\dot{\mathcal{S}}_{k,l}}) + \cdots \\ &+ \sum_{s=1}^{m_{n-1}} \sum_{j=q-l+J_{n-1}}^{q} \mathbf{h}_{i_s}(j)(\mathbf{s}_{i_s}(k+q-j) - \mathbf{s}_{i_s}(k+q-j)_{|\dot{\mathcal{S}}_{k,l}})\end{aligned} \tag{3.32}$$

Note that by definition for all $i, s, j : 1 \leq i \leq r, 1 \leq s \leq m_i$, and $j \leq 0$ or $j > J_i$, $\mathbf{h}_{i_s}(j) = 0$. If we write (3.32) in matrix form the equation is proved.

(iv) Proceeding in a similar fashion we prove the equation.

Next we establish the ranks of the projection error matrices $\mathbf{E}_l$.

Theorem 3: The following statements are true:

(i) For any $l : l < J_1$, $\qquad rank(\mathbf{E}_l) = 0$

(ii) For $l : J_1 \leq l < J_2$,

$$rank(\mathbf{E}_l) = (l - J_1 + 1)m_1 \tag{3.33}$$

(iii) For $n : 3 \leq n \leq r$ and $l : J_{n-1} \leq l < J_n$,

$$rank(\mathbf{E}_l) = \sum_{i=1}^{n-1}(l - J_i + 1)m_i \tag{3.34}$$

(iv) For $l : J_r \leq l \leq L$,

$$rank(\mathbf{E}_l) = \sum_{i=1}^{r}(l - J_i + 1)m_i \tag{3.35}$$

For each $n : 1 \leq n \leq r$ we define the matrix $\mathbf{\Theta}_n$ as

$$\mathbf{\Theta}_n = \left(\mathbf{H}_{l,J_1} | \cdots | \mathbf{H}_{l,J_n} \right)$$

Using induction on n we easily prove that $\mathbf{\Theta}_n$ has full column rank for each $n : 1 \leq n \leq r$.

For notational purposes set $J_{r+1} = L+1$. According to Theorem 2, for all $n : 2 \leq n \leq r+1$ and all $l : J_{n-1} \leq l < J_n$,

$$\mathbf{E}_l \;=\; \mathbf{\Psi}_{l,J_1} + \cdots \mathbf{\Psi}_{l,J_{n-1}} \tag{3.36}$$

Equation (3.36) is written compactly as

$$\mathbf{E}_l = \mathbf{\Theta}_{n-1} \begin{pmatrix} \tilde{\mathbf{S}}_{l,J_1} \\ \vdots \\ \tilde{\mathbf{S}}_{l,J_{n-1}} \end{pmatrix} \tag{3.37}$$

We know that $\mathbf{\Theta}_{n-1}$ has full column rank. Therefore,

$$rank(\mathbf{E}_l) = rank(\begin{pmatrix} \tilde{\mathbf{S}}_{l,J_1} \\ \vdots \\ \tilde{\mathbf{S}}_{l,J_{n-1}} \end{pmatrix}) \tag{3.38}$$

Note that each block entry $\tilde{\mathbf{S}}_{l,J_i}$ is a projection error. More precisely it has the form

$$\tilde{\mathbf{S}}_{l,J_i} = \begin{pmatrix} \mathbf{s}_{i_1}(k+l-J_i) - \mathbf{s}_{i_1}(k+l-J_i)_{|\dot{\mathcal{S}}_{k,l}} \\ \vdots \\ \mathbf{s}_{i_1}(k) - \mathbf{s}_{i_1}(k)_{|\dot{\mathcal{S}}_{k,l}} \\ \vdots \\ \mathbf{s}_{i_{m_i}}(k+l-J_i) - \mathbf{s}_{i_{m_i}}(k+l-J_i)_{|\dot{\mathcal{S}}_{k,l}} \\ \vdots \\ \mathbf{s}_{i_{m_i}}(k) - \mathbf{s}_{i_{m_i}}(k)_{|\dot{\mathcal{S}}_{k,l}} \end{pmatrix} \tag{3.39}$$

Using arguments similar to the ones used in the proof of Theorem 1, we conclude that for all $i,t,j : 1 \leq i \leq r, 1 \leq t \leq m_i, 0 \leq j \leq l - J_i$, $\mathbf{s}_{i_t}(k+j)$ is orthogonal to the vectors belonging to $\dot{\mathcal{S}}_{k,l}$. Therefore, $\mathbf{s}_{i_t}(k+j)_{|\dot{\mathcal{S}}_{k,l}} = 0$.

As a consequence, (for any k)

$$\tilde{\mathbf{S}}_{l,J_i}=\begin{pmatrix} \mathbf{s}_{i_1}(k+l-J_i) \\ \vdots \\ \mathbf{s}_{i_1}(k) \\ \vdots \\ \mathbf{s}_{i_{m_i}}(k+l-J_i) \\ \vdots \\ \mathbf{s}_{i_{m_i}}(k) \end{pmatrix} \tag{3.40}$$

Using similar arguments as the ones used in Theorem 1, we notice that the rows of the matrix

$$\begin{pmatrix} \tilde{\mathbf{S}}_{l,J_1} \\ \vdots \\ \tilde{\mathbf{S}}_{l,J_{n-1}} \end{pmatrix} \tag{3.41}$$

the rows of the above matrix are linearly independent with probability one.

The theorem then follows from (3.38) and (3.40) by a straightforward counting argument.[2]

The definition of the p-norm of a matrix and the previous theorem imply the following corollary:

Corollary 1: The following statements are true:

(i) For any $l : l < J_1$, $\quad ||\mathbf{E}_{k,l}||_p = 0$

(ii) For any $l : J_1 \leq l \leq L$, $\quad ||\mathbf{E}_{k,l}||_p > 0$

The proof is obvious for case (i).

We shall prove case (ii) by contradiction. Assume that there exists an $l : J_1 \leq l \leq L$ such that $||\mathbf{E}_l||_p = 0$. This implies that $\mathbf{E}_{k,l} = 0$. The latter as well as (3.37) imply that the rows of the matrix

$$\begin{pmatrix} \tilde{\mathbf{S}}_{l,J_1} \\ \vdots \\ \tilde{\mathbf{S}}_{l,J_{n-1}} \end{pmatrix} \tag{3.42}$$

are linearly dependent, something that cannot be true according to the proof of the previous theorem.

[2]The I.I.D. and mutual independence assumption on the input sequences can be replaced by a suitable deterministic condition. An example is a multichannel adaptation of the linear complexity condition stated in [58], guaranteeing that the input is generated by a multichannel AR model of sufficiently high order.

In the noiseless case, Theorem 3 indicates how to compute the system's different orders, as well as the number of subsystems that attain it in a straightforward manner.

Indeed, starting with $l = 0$ we compute $rank(\mathbf{E}_0)$ and we increase l by one until $rank(\mathbf{E}_l) > 0$. Equation (3.33) suggests that this value of l equals the smaller of the orders J_1. Moreover, it also gives the number m_1 of the subsystems that attain it.

Having determined m_1, we increase l in steps of one. As long as $l < J_2$, $rank(\mathbf{E}_l)$ remains a multiple of m_1. When this stops to hold (3.34) suggests that $l = J_2$. At this point having computed m_1, J_1 we use (3.34) to determine the number m_2 of the subsystems that attain the order J_2. We continue increasing l. As long as $l < J_3$, $rank(\mathbf{E}_l)$ increases by $m_1 + m_2$ each time l increases by one. Again, when this stops to hold, (3.34) suggests that $l = J_3$. At this point, having computed J_1, m_1, J_2, m_2 we use (3.34) to determine the number m_3 of the subsystems that have order equal to J_3.

We keep increasing l by one until we reach L and proceed in the same way, using (3.34) and (3.35) to determine J_i, m_i for all $i : 1 \leq i \leq r$. When $l = L$, we compute the number P of input signals as:

$$P = \sum_{i=1}^{r} m_i \tag{3.43}$$

The above approach is effective in a noiseless setting, where the rank of a matrix can be computed in a straightforward manner simply by counting the number of its nonzero singular values. In the presence of noise, however, the problem of rank determination can be proved difficult to solve because singular values that should be zero in theory could become small, but not necessarily zero. We address this issue in the next section, where we describe the implementation of our algorithm, using the concept of effective (numerical) matrix rank (*Criterion 1*) and a property implied by Corollary 1 (*Criterion 2*). Moreover, we provide a simpler version of the algorithm that applies when it is apriori known that the system is SIMO.

3.5 Algorithm Implementation

Two versions of the algorithm are supplied. The first applies in the general SIMO/MIMO and Volterra case. It can be used without any information regarding the application type. The second version is preferable in cases where it is apriori known that the system is SIMO, as it reduces significantly computation time by avoiding unnecessary SVDs.

Noise can be managed via the use of two criteria. The first one is employed to determine the effective (numerical) rank of a matrix [11].

Criterion 1: Let B be an $m \times n$ matrix perturbed by noise and denote $\beta_1, \beta_2, \cdots, \beta_r$, $r = min(m, n)$, its singular values with $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_r$. The effective rank of B is given by the integer t in the range $1 \leq t \leq r-1$ for which the ratio $\gamma_t = \beta_t/\beta_{t+1}$ is maximized.

The second criterion is suggested by Corollary 1.

Criterion 2: For fixed $k \in \mathcal{N}$, the ratio $\delta_{k,l} = ||\mathbf{E}_{k,l}||_p/||\mathbf{E}_{k,l+1}||_p$ is minimized for $l = J_1 - 1$, where J_1 is the smallest of the system orders.

Next, we present the two versions of our algorithm. We denote by L and Q an upper bound of the system orders and the number of output data vectors used, respectively.

3.5.1 SIMO, MIMO and Volterra Systems Order Determination

The proposed algorithm consists of the following steps.

1. For a fixed $w \geq w_0$ and for all l, $0 \leq l \leq L$ define the overall data matrices $\mathbf{Z}_{w,l}$ as

$$\mathbf{Z}_{w,l} = \left(\begin{array}{ccc} \mathbf{x}(2w+l+1) & \cdots & \mathbf{x}(Q) \\ \vdots & & \mathbf{F}_{w,l} \\ \mathbf{x}(w+l+2) & & \\ \hline \mathbf{x}(w+l+1) & \cdots & \\ \vdots & & \mathbf{Y}_{w,l} \\ \mathbf{x}(w+1) & & \\ \hline \mathbf{x}(w) & \cdots & \\ \vdots & & \mathbf{P}_{w,l} \\ \mathbf{x}(1) & & \end{array} \right)$$

2. For all l, $0 \leq l \leq L$, use SVD and *Criterion 1* to compute the effective rank of $\mathbf{Z}_{w,l}$. Reconstruct $\mathbf{Z}_{w,l}$, using the singular values that correspond to its effective rank and the associated left and right singular vectors to compute the matrix $\hat{\mathbf{Z}}_{w,l}$. From $\hat{\mathbf{Z}}_{w,l}$, compute the matrices $\hat{\mathbf{Y}}_{w,l}$ and $\hat{\mathbf{D}}_{w,l} = \begin{pmatrix} \hat{\mathbf{F}}_{w,l} \\ \hat{\mathbf{P}}_{w,l} \end{pmatrix}$, that correspond to the matrices $\mathbf{Y}_{w,l}$ and $\mathbf{D}_{w,l} = \begin{pmatrix} \mathbf{F}_{w,l} \\ \mathbf{P}_{w,l} \end{pmatrix}$ defined in the previous step. Step 2 will be referred as the *denoising* step.

3. For all l, $0 \leq l \leq L$, compute an orthogonal base of $\mathcal{R}(\hat{\mathbf{D}}_{w,l})$, $\{\mathbf{v}_1, \cdots, \mathbf{v}_f\}$. Use Theorems 1 and 2 to compute $\mathbf{E}_{w,l}$ in a way analogous to that given in [58], that is

$$\mathbf{E}_{w,l} = \hat{\mathbf{Y}}_{w,l} - \hat{\mathbf{Y}}_{w,l}\mathbf{V}'\mathbf{V}, \quad \mathbf{V} = \begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_f \end{pmatrix}$$

4. For all $0 \leq l \leq L$ compute $\delta_{w,l}$ and use *Criterion 2* to find J_1. Then use *Criterion 1* to compute the effective rank of $\mathbf{E}_{w,J_1}$. Conclude that this is the number m_1 of subsystems that have order J_1.
5. For all $J_1 < l \leq L$, use *Criterion 1* to find the effective rank of the matrices $\mathbf{E}_{w,l}$. Having determined J_1, m_1, use Theorem 3 to compute the pairs J_n, m_n, for all $1 < n \leq r$.
6. Determine the number P of system inputs from (3.43).

3.5.2 SIMO Systems

If it is apriori known that the system is SIMO the algorithm is simplified as follows.
1-3. Steps 1-3 remain the same as with the version described above.
4. For all $0 \leq l \leq L$ compute $\delta_{w,l}$ and use *Criterion 2* to find the order J_1 of the system.

Remark: In [58] it is proved that for a SIMO system, in the noiseless case, rank($\mathbf{Z}_{w,l}$)=$2w+l+L_\omega+1$, where L_ω is the true order of the system. Simulations showed that for some cases of low SNR values the estimated effective rank of $\mathbf{Z}_{w,l}$ could be either too low or two high introducing significant error when $\hat{\mathbf{Z}}_{w,l}$ is computed. Therefore, in our simulations we slightly modified Step 2 for SIMO systems, so that if for a given l the effective rank t_l of $\mathbf{Z}_{w,l}$ is less than $2w$ or greater than $2(w+L)$ the following actions are taken:

- If $l > 0$, t_l is set to $t_{l-1} + 1$, where t_{l-1} is the effective rank of $\mathbf{Z}_{w,l-1}$.
- If $l = 0$, t_l is set to $2w + 1$.

3.6 Simulation Examples

In this section we present the simulation experiments that we have performed for the LTI FIR SIMO and MIMO case as well as the nonlinear Volterra case.

3.6.1 LTI FIR MIMO and SIMO Systems

We assume that the system's output is given by the equation

$$\mathbf{y}(k) = \mathbf{x}(k) + \eta(k) \tag{3.44}$$

where $\mathbf{x}(k)$ is given by eq. (3.11) and $\eta(k) = [\eta_1(k) \cdots \eta_M(k)]^t$ is an $M \times 1$ spatial-temporal white noise vector sequence. Furthermore, the additive noise $\eta(k)$ is a zero mean stationary sequence uncorrelated to the inputs. For all i, $1 \leq i \leq M$, σ^2 is the finite variance of $\eta_i(k)$. Let Q denote the number of output vectors used. The signal-to-noise ratio is defined as

$$SNR = 10log_{10}\frac{\frac{1}{Q}\sum_{k=1}^{Q}||\mathbf{y}(k)||^2 - M\sigma^2}{M\sigma^2} \tag{3.45}$$

Even though Corollary 1 is true for any matrix norm, $||\mathbf{E}_{k,l}||_\infty$ was used in simulations, because the experiments showed that it achieves better performance over other known matrix norms. Finally, we mention that whenever the term *denoising* is used in the following, it refers to Step 2 of versions A and B of the proposed algorithm, as given in the previous section. Moreover, we shall use the terms output vector and output sample interchangeably to refer to the $M \times 1$ output data vector collected at the receiver.

A. MIMO Systems

Simulations entailing three different systems have been performed. In all cases the input sequences are QPSK, while 100 independent Monte-Carlo runs were obtained per system, number of output samples and SNR value. We depict the percentage of successful order detection related to SNR and the number of output vectors used. In all examples the predictor's size w equals to the upper bound of the system's orders $L = 10$.

Example 1. In this case we have simulated a two-input six-output MIMO system. The two subsystems $\mathbf{h}_1, \mathbf{h}_2$, with orders $L_1 = 3$ and $L_2 = 5$ respectively, are shown in Tables 3.1 and 3.2. According to *Criterion 2* and Theorem 3 the minimum order J_1 of the system should be equal to 3, while for $l = 3, 4, \cdots 10$ the corresponding ranks of $\mathbf{E}_{w,l}$ should be detected as $1, 2, 4, 6, 8, 10, 12, 14$.

Simulation results are presented in Table 3.3. The algorithm's success rate exceeds 90% using 1000 output vectors for SNR$\geq$ 20 dB. For SNR values between 20 and 25dB, increasing the amount of output samples from 400 to 800 or 1000 improves significantly the algorithm's performance. For

SNR values greater or equal to 27 dB high success rates are achieved even using 400 output samples.

Example 2. In this example, another two-input six-output MIMO system was tested. Subsystems $\mathbf{h}_1, \mathbf{h}_2$ comprising the total system are shown in Tables 3.4 and 3.5 respectively. Both their orders L_1, L_2 are equal to 4. According to *Criterion 2* and Theorem 3 the minimum order J_1 of the system should be equal to 4, while for $l = 4, 5, \cdots 10$ the corresponding ranks of $\mathbf{E}_{w,l}$ should be detected as $2, 4, 6, 8, 10, 12, 14$.

Results are shown in Table 3.6. The method's success rate exceeds 90% using 800 output vectors for SNR≥ 28 dB. As with the previous example we notice that for SNR values greater than 30dB high success rates are achieved even using 400 output samples.

Example 3. In this example a three-input six-output MIMO system was tested. The two subsystems of Example 1 were maintained and an extra subsystem $\mathbf{h}_3$ of order $L_3 = 3$ shown in Table 3.7 was added. According to *Criterion 2* and Theorem 3 the minimum order J_1 of the system should be equal to 3, while for $l = 3, 4, \cdots 10$ the corresponding ranks of $\mathbf{E}_{w,l}$ should be $2, 4, 7, 10, 13, 16, 19, 22$. The results are shown in Table 3.8.

In comparison to the previous examples we observe a performance degradation that can be explained by the extra amount of information that has to be detected by the algorithm due to the introduction of the new subsystem.

B. SIMO Systems

Simulations of SIMO systems are performed. The proposed method is compared with existing SIMO system order estimation techniques, namely those presented in [58], [6] and the MDL algorithm [59]. In order to study the effect of denoising, both J-LSS and the proposed algorithm were tested using denoised as well as non denoised data matrices. To make fair comparisons, all the algorithms were tested against the same data sequence. That is, in each Monte Carlo run, all the algorithms worked on the same received data. A success rate threshold of 90% was employed for comparison.

Four systems were simulated the first two of which are artificial. The remaining are real microwave channels given in http://spib.rice.edu/spib/microwave.html.

In all cases the input sequences were assumed to be QPSK and 200 output samples were used. One hundred independent Monte-Carlo runs were run per system and SNR value. Finally, in all the experiments the upper bound of system order L was set equal to 15 and the predictor's size w was set equal to L.

Example 1. A two-output SIMO system is considered. The subchannel coefficients are shown in

Table 3.9, while simulation results are presented in Table 3.10. The first and the last taps of the system concentrate 11% and 9, 6% approximately of the total power. Therefore, the estimated system order should equal 7.

Simulations showed that the proposed algorithm achieves a success rate greater than 90% for SNR$\geq$ 20dB. The corresponding SNR level for Liavas et al. method is 26dB, for MDL 21dB and for J-LSS's 35dB. Denoising improved J-LSS, giving an over 90% success rate for an SNR values equal to or greater 28dB. Applying no denoising slightly improves the proposed method giving a success rate of over 90% for an SNR value of 19dB.

Example 2. A two-output SIMO system is considered. The subchannel coefficients are shown in Table 3.11. It is easily verified that even though the last tap of the second subchannel is zero, the power of the corresponding tap of the first subchannel approximates 24% of the total channel energy. Therefore, blind order estimation should detect the maximum order of the subchannels, which equals 4. The results are shown in Table 3.12.

Simulations showed that the proposed algorithm achieves a success rate greater or equal to 90% for SNR$\geq$ 12dB. The corresponding SNR level for Liavas et al. method is 20dB, for MDL 15dB and for J-LSS is 26dB. Denoising improves J-LSS, giving an over 90% success rate for SNR values greater than 20dB. Applying no denoising improves the proposed method giving a success rate of over 90% for SNR values equal to or greater than 8dB.

Examples 3 and 4. Two FIR microwave radio channels chosen from the library above were tested, namely channel 10 and channel 13. Channels were oversampled by a factor of 2. Their characteristic is that their true impulse response is very long, but they can be partitioned into a "significant part" that contains the "large" channel terms and concentrates the most of the channel power and a "non-significant" part that comprises of "small" leading and/or trailing terms. The order of the significant part is the "effective channel order" as defined in [6]. Intuitevely, the satisfactory effective order estimate of these channels is two, giving three taps per subchannel. Simulation results are shown in Table 3.14 and Table 3.15.

Interpreting the results for channel 10, we should notice first that the 14th, 15th and 16th taps concentrate approximately 99% of the total channel power, while all the remaining leading and trailing taps account only for approximately 1%. Therefore the effective order of channel 10 should equal 2.

Simulations showed that the proposed method's success percentage exceeds 90% when SNR$\geq$ 14dB, while the corresponding SNR level for Liavas et al. method is 22dB and for MDL 15dB. J-LSS

success rate was not satisfactory. However, applying denoising in J-LSS dramatically improved its performance, giving a success rate exceeding 90% for SNR$\geq$ 22dB. MDL's performance proved poor for SNR$\geq$ 35dB, verifying that it tends to overmodel systems at high SNR values, as opposed to the proposed algorithm that kept a 100% success rate even for high SNR levels. For SNR values between 10 and 20 dB denoising improves significantly the performance of the proposed method and results in a 5dB gain in achieving the 90% success rate. For the same range of SNR values, the proposed algorithm proved superior to the Liavas et al. method. Finally, we notice that when the proposed algorithm failed to determine the channel's effective order it undermodelled it by one, giving one as the estimated order.

In case of channel 13 we mention that the 22nd, 23rd and 24th taps concentrate approximately 99% of the total channel power, while all the remaining leading and trailing taps concentrate approximately the remaining 1%. Therefore the effective order of channel 13 should equal 2.

Simulations showed that the proposed algorithm achieves a success rate greater than 90% for SNR$\geq$ 17dB. The corresponding SNR level for Liavas et al. method is 27dB and for MDL 16dB. As it was the case with channel 10, J-LSS's performance was not satisfactory, but again it was improved by denoising. For SNR$\geq$ 30dB MDL showed poor performance, due to overmodelling. On the contrary, the proposed method kept the 100% success rate even for high SNR levels. Similar overmodelling behaviour proved true for J-LSS, that is for SNR$\geq$ 40dB its performance dropped significantly even though denoising was used. For SNR values ranging between 10 and 20 dB the proposed method achieved superior performance compared to Liavas et al. algorithm. Furthermore, for SNR values between 10 and 20 dB denoising improves the performance of the new algorithm, gaining 4dB in achieving a success rate greater than or equal to 90%. As with the previous case, when the algorithm failed to determine the channel's effective order it undermodelled it by one, computing it as 1 rather than 2.

3.6.2 Volterra Systems

In simulation example 1 uniformly distributed, mutually independent, temporally i.i.d. symbols drawn from a QPSK constellation $\{1+j, 1-j, -1+j, -1-j\}$ with variance equal to 2 were used as input. In simulation example 2 (real valued SIMO VOLTERRA system), uniformly distributed, mutually independent, temporally i.i.d., input symbols drawn from a BPSK constellation with variance equal to 1 were used. QPSK as well as BPSK input sequences were randomly generated, using MATLAB functions. Simulations were performed according to the following methodology. For each system,

1000 independent Monte-Carlo experiments were executed. For each experiment, an input sequence consisting of 1500 symbols was generated, as described above. Having fixed the SNR value, 1000 runs of the algorithm per input sequence were performed. Then, the Symbol Error Rate (SER) was computed per experiment and SNR value, executing 1000 runs. The average was then calculated, per SNR value, for all different input sequences, executing 1000 Monte-Carlo runs, giving the SER per system and SNR value. The above procedure was repeated for all SNR values.

Example 1 We applied the proposed algorithm to the SIMO Volterra system described in *Example 1* of [57]. It is a three-output system, described by the following equation:

$$\mathbf{y}(n) = \sum_{l=0}^{2} \mathbf{h}_1(l)s(n-l) + \sum_{l=0}^{1} \mathbf{h}_{2,1}(l)s(n-l)s(n-1-l) + \eta(n)$$

where $\eta(n)$ is additive white Gaussian noise. As already mentioned, BPSK input was used. Kernel values as well as order determination results are presented in the following paragraphs.

Example 2 We applied the proposed algorithm to a three-output system as given in [64], described by the following equation:

$$\mathbf{y}(n) = \sum_{l=0}^{3} \mathbf{h}_1(l)s(n-l) + \sum_{l=0}^{1} \mathbf{h}_{3,12}(l)s(n-l)s(n-1-l)s^*(n-l-2) + \eta(n)$$

where $\eta(n)$ is additive white Gaussian noise. As already mentioned, QPSK input was used. Kernel values as well as order determination results are presented in the following paragraphs.

Table 3.1. MIMO Systems,example 1: Subchannel coefficients of subsystem 1

$\mathbf{h}_1(0)$	$\mathbf{h}_1(1)$	$\mathbf{h}_1(2)$	$\mathbf{h}_1(3)$
$0.15 + 0.0187i$	$0.0747 + 0.0825i$	$0.2241 + 0.2469i$	$-0.0148 - 0.0151i$
$0.124 + 0.0261i$	$0.0934 + 0.0844i$	$0.1905 + 0.2099i$	$-0.0160 - 0.0175i$
$0.199 + 0.0336i$	$0.1120 + 0.1236i$	$0.1569 + 0.1692i$	$-0.0147 - 0.1199i$
$0.173 + 0.0411i$	$0.1307 + 0.1255i$	$0.1233 + 0.1285i$	$-0.0134 - 0.1423i$
$0.15 + 0.1187i$	$0.2747 + 0.2825i$	$0.2241 + 0.2469i$	$-0.0448 - 0.0751i$
$0.124 + 0.1261i$	$0.2934 + 0.2844i$	$0.2905 + 0.2099i$	$-0.0560 - 0.0975i$

Table 3.2. MIMO Systems,example 1: Subchannel coefficients of subsystem 2

$\mathbf{h}_2(0)$	$\mathbf{h}_2(1)$	$\mathbf{h}_2(2)$	$\mathbf{h}_2(3)$	$\mathbf{h}_2(4)$	$\mathbf{h}_2(5)$
$0.04 + 0.017i$	$0.056 + 0.007i$	$0.126 + 0.020i$	$-0.047 + 0.010i$	$-0.013 - 0.041i$	$0.069 - 0.100i$
$0.10 + 0.02i$	$0.065 + 0.008i$	$0.104 + 0.019i$	$-0.043 + 0.006i$	$0.010 - 0.062i$	$0.004 - 0.034i$
$0.07 + 0.03i$	$0.074 + 0.010i$	$0.082 + 0.018i$	$-0.039 + 0.001i$	$0.034 - 0.084i$	$0.039 - 0.100i$
$0.04 + 0.03i$	$0.082 + 0.012i$	$0.061 + 0.017i$	$-0.034 - 0.019i$	$0.058 - 0.067i$	$-0.061 - 0.065i$
$0.04 + 0.027i$	$0.026 + 0.027i$	$0.226 + 0.220i$	$-0.247 + 0.210i$	$-0.213 - 0.241i$	$0.069 - 0.100i$
$0.10 + 0.22i$	$0.265 + 0.208i$	$0.204 + 0.219i$	$-0.243 + 0.206i$	$0.210 - 0.262i$	$0.004 - 0.034i$

Table 3.3. MIMO Systems,example 1: Percentage of successful order detection.

SYMBOLS	17 dB	20 dB	22 dB	25 dB	27 dB	30 dB	32 dB	40 dB	50 dB
400	0%	2%	17%	64%	85%	96%	100%	100%	100%
800	1%	79%	95%	98%	100%	100%	100%	100%	100%
1000	6%	96%	99%	100%	100%	100%	100%	100%	100%

Table 3.4. MIMO Systems, example 2: Subchannel coefficients of subsystem 1

$\mathbf{h}_1(0)$	$\mathbf{h}_1(1)$	$\mathbf{h}_1(2)$	$\mathbf{h}_1(3)$	$\mathbf{h}_1(4)$
$0.15 + 0.0187i$	$0.1747 + 0.1825i$	$0.1241 + 0.1469i$	$-0.1148 - 0.1151i$	$-0.1148 - 0.0151i$
$0.124 + 0.0261i$	$0.1934 + 0.1844i$	$0.2905 + 0.1099i$	$-0.1160 - 0.1175i$	$-0.1160 - 0.0175i$
$0.199 + 0.0336i$	$0.2120 + 0.2236i$	$0.1569 + 0.1692i$	$-0.1147 - 0.1199i$	$-0.1147 - 0.0199i$
$0.173 + 0.0411i$	$0.2307 + 0.2255i$	$0.2233 + 0.2285i$	$-0.1134 - 0.1423i$	$-0.1134 - 0.0423i$
$0.05 + 0.1187i$	$0.0747 + 0.0825i$	$0.0241 + 0.0469i$	$-0.0448 - 0.0751i$	$-0.0448 - 0.0751i$
$0.224 + 0.0461i$	$0.2934 + 0.2844i$	$0.1905 + 0.1099i$	$-0.1560 - 0.1975i$	$-0.1560 - 0.0975i$

Table 3.5. MIMO Systems, example 2: Subchannel coefficients of subsystem 2

$\mathbf{h}_2(0)$	$\mathbf{h}_2(1)$	$\mathbf{h}_2(2)$	$\mathbf{h}_2(3)$	$\mathbf{h}_2(4)$
$0.04 + 0.017i$	$0.256 + 0.107i$	$0.226 + 0.120i$	$-0.147 + 0.110i$	$-0.113 - 0.141i$
$0.10 + 0.12i$	$0.265 + 0.208i$	$0.204 + 0.119i$	$-0.143 + 0.106i$	$0.110 - 0.162i$
$0.07 + 0.13i$	$0.274 + 0.210i$	$0.182 + 0.118i$	$-0.139 + 0.101i$	$0.134 - 0.184i$
$0.04 + 0.13i$	$0.282 + 0.212i$	$0.161 + 0.117i$	$-0.134 - 0.119i$	$0.258 - 0.267i$
$0.04 + 0.127i$	$0.126 + 0.127i$	$0.226 + 0.220i$	$-0.247 + 0.210i$	$-0.213 - 0.241i$
$0.20 + 0.12i$	$0.365 + 0.308i$	$0.104 + 0.119i$	$-0.243 + 0.206i$	$0.210 - 0.262i$

Table 3.6. MIMO Systems, example 2: Percentage of successful order detection.

SYMBOLS	**25 dB**	**27 dB**	**28 dB**	**30 dB**	**35 dB**	**40 dB**	**50 dB**
400	0%	8%	17%	88%	100%	100%	100%
800	12%	62%	98%	100%	100%	100%	100%
1000	20%	87%	100%	100%	100%	100%	100%

Table 3.7. MIMO Systems, example 3: Subchannel coefficients of subsystem 3

$\mathbf{h}_3(0)$	$\mathbf{h}_3(1)$	$\mathbf{h}_3(2)$	$\mathbf{h}_3(3)$
$0.15 + 0.1187i$	$0.1747 + 0.1825i$	$0.1241 + 0.1469i$	$-0.1148 - 0.1151i$
$0.124 + 0.1261i$	$0.1934 + 0.1844i$	$0.1905 + 0.1099i$	$-0.1160 - 0.1175i$
$0.199 + 0.1336i$	$0.1120 + 0.1136i$	$0.1169 + 0.1192i$	$-0.1147 - 0.1099i$
$0.173 + 0.1411i$	$0.0307 + 0.0255i$	$0.0233 + 0.0285i$	$-0.1134 - 0.0423i$
$0.15 + 0.0187i$	$0.1747 + 0.1825i$	$0.1241 + 0.1469i$	$-0.1448 - 0.1751i$
$0.124 + 0.0261i$	$0.1934 + 0.1844i$	$0.0905 + 0.0099i$	$-0.1560 - 0.1975i$

Table 3.8. MIMO Systems, example 3: Percentage of successful order detection.

SYMBOLS	**25 dB**	**27 dB**	**30 dB**	**33 dB**	**35 dB**	**40 dB**	**50 dB**
400	0%	0%	0%	16%	73%	100%	100%
800	0%	0%	60%	97%	100%	100%	100%
1000	0%	4%	73%	99%	100%	100%	100%

Table 3.9. SIMO Systems, example 1: subchannel coefficients

$\mathbf{h}(0)$	$\mathbf{h}(1)$	$\mathbf{h}(2)$	$\mathbf{h}(3)$	$\mathbf{h}(4)$	$\mathbf{h}(5)$	$\mathbf{h}(6)$	$\mathbf{h}(7)$
0.4	0.18	0.22	0.5	0.4	0.16	0.43	0.38
0.25	0.31	0.16	0.41	0.20	0.61	0.43	0.22

Table 3.10. SIMO Systems, example 1: LIAVAS et al, MDL, J-LSS,PROPOSED METHOD comparison

SNR (dB)	**LIAVAS et al**	**MDL**	**J-LSS**	**J-LSS,D**	**PROP METHOD**	**PROP METHOD,D**
10	0%	0%	0%	0%	36%	16%
15	0%	0%	0%	0%	65%	50%
19	0%	59%	0%	12%	95%	84%
20	0%	82%	0%	15%	98%	92%
21	0%	94%	0%	31%	100%	93%
25	74%	98%	1%	71%	100%	100%
26	91%	98%	3%	76%	100%	100%
28	99%	99%	11%	92%	100%	100%
30	100%	99%	31%	95%	100%	100%
35	100%	99%	93%	97%	100%	100%
40	100%	99%	100%	100%	100%	100%
50	100%	99%	100%	100%	100%	100%

Table 3.11. SIMO Systems, example 2: subchannel coefficients

$\mathbf{h}(0)$	$\mathbf{h}(1)$	$\mathbf{h}(2)$	$\mathbf{h}(3)$	$\mathbf{h}(4)$
-0.215	0.086	0.688	0	-0.688
0.868	0	-0.496	0	0

Table 3.12. SIMO Systems, example 2: Liavas et al, MDL, J-LSS and the PROPOSED METHOD success rates

SNR (dB)	LIAVAS ET AL.	MDL	J-LSS	J-LSS,D	PROP METHOD	PROP METHOD,D
8	0%	0%	0%	0%	92%	56%
10	0%	4%	0%	1%	98%	82%
12	0%	31%	0%	3%	100%	90%
15	1%	90%	0%	22%	100%	98%
20	95%	99%	28%	93%	100%	100%
25	100%	100%	88%	100%	100%	100%
26	100%	100%	92%	100%	100%	100%
30	100%	100%	100%	100%	100%	100%
40	100%	99%	100%	100%	100%	100%
50	100%	99%	100%	100%	100%	100%

Table 3.13. SIMO Systems, Example 3 (Channel 10): Liavas et al, MDL, J-LSS and the PROPOSED METHOD success rates

SNR (dB)	LIAVAS ET AL.	MDL	J-LSS	J-LSS,D	PROP METHOD	PROP METHOD,D
10	0%	0%	0%	0%	2%	62%
14	0%	69%	0%	11%	13%	91%
15	0%	96%	0%	16%	23%	97%
19	15%	99%	0%	50%	90%	99%
22	99%	100%	0%	93%	100%	100%
30	100%	90%	9%	99%	100%	100%
40	100%	0%	31%	78%	100%	100%
50	100%	0%	46%	56%	100%	100%

Table 3.14. SIMO Systems, example 4 (Channel 13). Liavas et al, MDL, J-LSS and the PROPOSED METHOD success rates

SNR (dB)	LIAVAS ET AL.	MDL	J-LSS	J-LSS,D	PROP METHOD	PROP METHOD,D
10	0%	0%	0%	0%	0%	47%
15	0%	80%	0%	14%	1%	82%
16	0%	99%	0%	23%	2%	85%
17	0%	100%	1%	30%	6%	90%
20	0%	99%	3%	59%	71%	95%
21	0%	100%	7%	60%	92%	97%
25	39%	100%	18%	88%	100%	100%
27	97%	94%	22%	98%	100%	100%
30	100%	4%	26%	94%	100%	100%
40	100%	0%	0%	14%	100%	100%
50	100%	0%	0%	6%	100%	100%

Table 3.15. Example 3: SIMO Volterra System Kernels

$\mathbf{h}_1(0)$	$\mathbf{h}_1(1)$	$\mathbf{h}_1(2)$	$\mathbf{h}_{2,1}(0)$	$\mathbf{h}_{2,1}(1)$
1	-2.5	1	2	0.7
0.5	3	5	0.3	1.2
2	0	2	-0.7	3

Table 3.16. Example 3: Percentage of successful order detection

NUMBER OF SYMBOLS	**10 dB**	**12 dB**	**15 dB**	**18 dB**	**20 dB**
500	10%	32%	77%	94%	100%
1000	68%	80%	95%	98%	100%
1500	90%	96%	99%	100%	100%

Table 3.17. Example 4: SIMO Volterra System Kernels

$\mathbf{h}_1(0)$	$\mathbf{h}_1(1)$	$\mathbf{h}_1(2)$	$\mathbf{h}_1(3)$	$\mathbf{h}_{3,12}(0)$	$\mathbf{h}_{3,12}(1)$
1	$-2.5+2i$	$1+i$	$4+0.3i$	2	$0.7-0.8i$
$0.5+0.4i$	$3+2i$	$-1+i$	$5+i$	$0.3+0.2i$	$1.2+i$
$-1+i$	$1-2i$	$2+1.3i$	$-3+1.3i$	$-0.7+0.7i$	$3+0.1i$

Table 3.18. Example 4: Percentage of successful order detection

NUMBER OF SYMBOLS	**10 dB**	**12 dB**	**15 dB**	**18 dB**	**20 dB**
500	10%	30%	72%	93%	100%
1000	60%	85%	94%	99%	100%
1500	90%	95%	99%	100%	100%

4. LTI FIR MIMO Systems Kernel Identification and Volterra Systems Equalization

This chapter presents the algorithm for LTI FIR MIMO kernel identification and Volterra systems equalization. Definitions and algebraic notations remain the same as described in the previous chapter. Next, we describe briefly the system model for both LTI FIR MIMO systems as well nonlinear systems described by Volterra series and the assumptions used. A set of theorems and lemmas is presented and finally the algorithm is introduced. Simulation results presented at the end of the chapter, validate the performace of the method compared to existing schemes.

4.1 Objectives

Given an LTI FIR system or a nonlinear system described by a finite Volterra series the objective of the algorithm is

(a) To identify the kernels of the system in case it is LTI FIR MIMO.

(b) To equalize input symbols in case of SIMO Volterra systems.

4.2 Model Description

The model remains the same as described in the previous chapter for both finite memory SIMO Volterra systems as well as LTI FIR MIMO systems.

4.2.1 Discrete Volterra Systems

The nonlinear systems under consideration are assumed SIMO, discrete-time, time-invariant, causal and of finite memory. Furthermore, we assume that any small changes to the system's input $s(n)$ result in small changes to the system's output. As already mentioned, any such system can be approximated over a uniformly bounded set of input signals by a truncated Volterra series expansion of finite order P.

The output $y(n)$ of a real valued SIMO Volterra system, as stated in the previous chapter, can be described as:

$$y(n) = \mathbf{h}_0 + \sum_{p=1}^{P} \sum_{m_i=0}^{N-1} \cdots \sum_{m_p=0}^{N-1} \mathbf{h}_p(m_1, \cdots, m_p) s(n-m_1) \cdots s(n-m_p) + \eta(n) \tag{4.1}$$

On the other hand, for a narrowband communication system input-output relationship is described by the equation:

$$y(n) = \mathbf{h}_0 + \sum_{p=1}^{P} \sum_{m_i=0}^{N-1} \cdots \sum_{m_p=0}^{N-1} \mathbf{h}_p(m_1, \cdots, m_{2k+1}) s(n-m_1) \cdots s(n-m_{k+1}) s^*(n-m_{k+2}) \cdots s^*(n-m_{2k+1}) + \eta(n) \tag{4.2}$$

We shall assume that $\eta(n) = 0$ (noiseless case) and examine the effect of noise in simulations. In the above equations , $s(n)$ is the system's input, $\forall 1 \leq p \leq P$, $\mathbf{h}_p$ are the discrete $p-th$ order Volterra kernels and P is the order of nonlinearity. Furthermore, without loss of generality, we may choose the range $m_1, \cdots, m_p$ so that $0 \leq m_1 \leq m_2 \cdots \leq m_p$, that is, the Volterra kernels are upper triangular.

The SIMO Volterra system is then casted as an equivalent MIMO linear system, by appropriate redefinition of the vector kernels and the system's input. Let M denote the number of output channels, then the vector kernel $\mathbf{h}_p(m_1, \cdots, m_p)$ can be indexed as:

$$\mathbf{h}_p(m_1, \cdots, m_p) = \mathbf{h}_{p,i_1:i_{p-1}}(m) = [h_p^1(m, m+i_1, \cdots, m+i_{p-1}) \cdots h_p^M(m, m+i_1, \cdots, m+i_{m-1})] \tag{4.3}$$

where $0 \leq i_1 \leq \cdots \leq i_{k-1} \leq N-1$ and $m = 0, 1, \cdots N-1-i_{p-1}$. To make the notation simpler, we use $i_1 : i_{p-1}$ to denote the set $(i_1, \cdots, i_{p-1})$. For $p = 2$, the range notation $(i_1 : i_1)$ denotes just the index i_1.

Similarly, we define the signals:

$$s_{p,i_1:i_{p-1}}(m) = s(m)s(m-i_1) \cdots s(m-i_{p-1}) \tag{4.4}$$

with $0 = i_0 \leq i_1 \leq \cdots \leq i_{p-1} \leq N-1$. We denote $\mathbf{h}_1 = \mathbf{h}_{1,i_0:i_0}$ and $s_1 = s_{1,i_0:i_0}$. Using the change of variables $m_p = m + i_{p-1}$, for $m = 1, 2, \cdots, P$ we have that:

$$\mathbf{y}(n) = \sum_{p=1}^{P} \sum_{0 \leq i_1 \leq \cdots \leq i_{p-1} \leq N-1} \sum_{m=0}^{N_p - i_{p-1}} \mathbf{h}_{p,i_1:i_{p-1}}(n) s_{p,i_1:i_{p-1}}(n-l) \tag{4.5}$$

Equation (4.5) describes a MIMO system. Finally, for an appropriate B, we introduce the variables $u_i, 1 \leq i \leq B$ and use lexicographical ordering to name the sequences $u_i(n) = s_{p,i_1:i_{p-1}}(n)$. Then, applying the same technique and introducing variables $\mathbf{g}_i(m)$ we rename the kernels $\mathbf{g}_i(m) = \mathbf{h}_{p,i_1:i_{p-1}}(m)$. Taking all the aforementioned steps, the above equation takes the form

$$\mathbf{y}(n) = \sum_{i=1}^{B} \sum_{j=0}^{K_i} \mathbf{g}_i(j) u_i(n-j) = [\mathbf{G}(z)]\mathbf{u}(n) \tag{4.6}$$

where $[\mathbf{G}(z)]$ is the system transfer function and $\mathbf{u}(k) = [u_1(k) \cdots u_B(k)]^t$. An example of how a SIMO Volterra system is cast as a MIMO system is given in [10] for $P = 2$.

In general, $B \neq P$, meaning that the number of the subsystems appearing in the MIMO representation is not equal to the order of the system's nonlinearity. The memory lengths of the B subsystems are given by the integers $K_1, K_2, \cdots, K_B$. In general, some of the above integers may be equal to each other. Therefore, the r distinct integer values appearing in the set $K_1, K_2, \cdots, K_B$ and denote the different subsystem orders, will in the following be denoted by $J_1, J_2, \cdots, J_r$ and we shall assume, without loss of generality, that $J_1 < J_2 < \cdots < J_r$. In addition, for all i, $1 \leq i \leq r$, m_i will denote the number of subsystems attaining the same order J_i.

We shall make the following assumptions:

A1) The input sequence $s(n)$ is zero mean, i.i.d., with values from a finite alphabet of at least $P+1$ complex numbers. Examples include PSK or QAM signals.

A2) The system transfer matrix **G**(z) is irreducible. This guarantees that there are no common zeros in the (FIR) transfer functions of every pair of the subchannels involved. It is a common assumption in all methods based on Second Order Statistics.

A3) The memory of the linear kernel is strictly greater than the memory of any nonlinear term. The first element of the zero-th tap of this kernel is equal to unity.

A4) An upper bound L of B is known.

4.2.2 Discrete LTI FIR MIMO Systems

A discrete-time Linear Time Invariant FIR MIMO system can be described as:

$$\mathbf{y}(n) = \sum_{i=1}^{B} \sum_{j=0}^{K_i} \mathbf{g}_i(j) u_i(n-j) = [\mathbf{G}(z)]\mathbf{u}(n) \tag{4.7}$$

where $[\mathbf{G}(z)]$ is the system transfer function and $\mathbf{u}(n) = [u_1(n) \cdots u_B(n)]^t$. The number of system inputs is B, while the number of output channels is M. The following assumptions hold:

B1) The input sequences u_i are zero mean, complex i.i.d. processes. Moreover, forth-order kurtosis of all source signals have the same sign. For a broad class of digital communication signals, such as uniformly distributed QAM and PSK constellations their fourth-order kurtosis are negative. This assumption guarantees the validity of applying the JADE algorithm, as it will be described in the following section.

B2) $M > B$.

B3) **G**(z) is irreducible. As already mentioned above, this guarantees that there are no common zeros in the (FIR) transfer functions of every pair of the subchannels involved.

B4) An upper bound L of B is known.

4.3 Kernel Identification and Input Equalization

Data structures required, have already been defined in the previous chapter. In this section we state and prove a number of lemmas and theorems that establish the validity of the algorithm. In addition, the steps of the algorithm are presented.

4.3.1 Lemmas and Theorems

In this subsection, we state and prove lemmas and theorems that establish the validity of the algorithm. Assuming k is fixed, we shall denote $\mathbf{E}_{k,l}$ by $\mathbf{E}_l$, to simplify notation.

We state theorem 1 that is essential to the development of the proposed algorithm:

Theorem 1

For all $i, \quad 1 \leq i \leq r$ and each $l, \quad J_i \leq l \leq L$, the elements of each one of the matrices $\mathbf{\Psi}_{l,J_i}$ can be

computed from the elements of the matrices $\mathbf{E}_{J_1}, \mathbf{E}_{J_2}, \cdots, \mathbf{E}_{J_i}$.

Proof:

Assume that $l \geq L_i$ and that memory length J_i is attained by subsystem kernels $\mathbf{g}_{i_1}, \mathbf{g}_{i_2}, \cdots, \mathbf{g}_{i_{m_i}}$. By definition ${}_{l,J_i} = \mathbf{G}_{l,J_i}\tilde{\mathbf{U}}_{l,J_i}$, so each element p, t of ${}_{l,J_i}$ is the vector inner product of the $p-th$ row of $\mathbf{G}_{l,J_i}$ and the $t-th$ column of $\tilde{\mathbf{U}}_{l,J_i}$. Denote the $p-th$ row of $\tilde{\mathbf{U}}_{l,J_i}$ by $\boldsymbol{\alpha}_p$ and the $t-th$ column by $\boldsymbol{\beta}_t$. Thus, $\boldsymbol{\alpha}_p$ is an $1 \times (m_i \times (l - J_i + 1))$ row vector and $\boldsymbol{\beta}_t$ is an $(m_i \times (l - J_i + 1)) \times 1$ column vector. Denote

$$\boldsymbol{\alpha}_p = [a(1), a(2), \cdots, a(m_i \times (l - J_i + 1))] \tag{4.8}$$

$$\boldsymbol{\beta}_t = [b(1), b(2), \cdots, b(m_i \times (l - J_i + 1))]^t \tag{4.9}$$

Due to the structure of $\mathbf{G}_{l,J_i}$ we see that all the elements of $\boldsymbol{\alpha}_p$ that are $l - J_i$ columns apart are either zero or they belong to the same row of $\mathbf{G}_{J_i,J_i}$. Similarly, due to the structure of $\tilde{\mathbf{U}}_{l,J_i}$, all the elements of $\boldsymbol{\beta}_t$ that are $l - J_i$ rows apart belong to the same column of $\tilde{\mathbf{U}}_{J_i,J_i}$.

Therefore, the row by column inner product $\boldsymbol{\alpha}_p \times \boldsymbol{\beta}_t$ is equal to:

$$\alpha_p \times \beta_t = \sum_{j=1}^{l-J_i+1} \sum_{k=0}^{m_i-1} a(j + k \times (l - J_i))b(j + k \times (l - J_i)) \tag{4.10}$$

From the analysis above it occurs that the sum

$$\sum_{k=0}^{m_i-1} a(j + k \times (l - J_i))b(j + k \times (l - J_i)) \tag{4.11}$$

it equals zero or it is a product of a row of $\mathbf{G}_{J_i,J_i}$ by a column of $\tilde{\mathbf{U}}_{J_i,J_i}$. By definition, this product belongs to $\boldsymbol{\Psi}_{J_i,J_i}$, proving that all elements of $\boldsymbol{\Psi}_{l,J_i}$ can be computed from elements of $\boldsymbol{\Psi}_{J_i,J_i}$. Therefore, to prove the theorem it suffices to show that all elements of $\boldsymbol{\Psi}_{J_i,J_i}$ can be computed from elements of $\mathbf{E}_{J_1}, \mathbf{E}_{J_2}, \cdots, \mathbf{E}_{J_i}$. For that, we use theorem 1 stated in the previous chapter. Start, with J_1. According to theorem 1, stated in the previous chapter, $\boldsymbol{\Psi}_{J_1,J_1} = \mathbf{E}_{J_1}$, so the elements of all matrices $\boldsymbol{\Psi}_{l,J_1}$ can be computed from the elements of $\mathbf{E}_{J_1}$.Then, we use equation (3.29) of theorem 1 to compute $\boldsymbol{\Psi}_{J_2,J_2}$ from the elements of the matrix $\mathbf{E}_{J_2}$. Having computed $\boldsymbol{\Psi}_{J_2,J_2}$ we can compute all matrices $\boldsymbol{\Psi}_{l,J_2}$. Then, we use (3.30) to compute $\boldsymbol{\Psi}_{J_3,J_3}$ from $\mathbf{E}_{J_3}$ and continue the same way until all the corresponding matrices have been evaluated, proving the validity of the theorem.

4.3.2 Algorithm Description

A batch algorithm that performs system identification/equalization is developed. For LTI FIR MIMO systems the algorithm uses the Joint Diagonilization Principle through the JADE algorithm as described in [61]. The outline of the algorithm is the following.

A. Order Estimation

- Step A_1: At the receiver, group the collected symbols in blocks of Q symbols per block.
- Step A_2: Using the first block, for all l, $J_1 \leq l \leq L$ compute the matrices $\mathbf{E}_l$, using the method described in Appendix A. Store the matrices.
- Step A_3: Compuite the orders $J_1, J_2, \cdots J_r$ of the different subsystems that comprise the overall system following the algorithm established in [12]. Store $J_1, J_2, \cdots J_r$.
- Step A_4: For FIR LTI MIMO systems execute part B of the algorithm, for SIMO Volterra systems execute part C.

B. LTI FIR MIMO Systems Identification

- Step 1: Use matrices $\mathbf{E}_l$ computed in part **A**, to compute the matrices $\mathbf{\Psi}_{l,J_i}$, according to theorem 1.
- Step 2: Set $l = J_1$. Using the equation

$$\mathbf{E}_{J_1} = \mathbf{\Psi}_{J_1,J_1} \tag{4.12}$$

  and the JADE algorithm identify the kernels corresponding to the subsystems of order J_1 as columns of the matrix $\mathbf{G}_{J_1,J_1}$.

- Step 3: Set $l = J_2$. Using the equation

$$\mathbf{E}_{J_2} - \mathbf{\Psi}_{J_2,J_1} = \mathbf{\Psi}_{J_2,J_2} \tag{4.13}$$

  and the JADE algorithm identify the kernels that correspond to subsystems of order J_2 as columns of the matrix $\mathbf{G}_{J_2,J_2}$.

$\vdots$

- Step $r+1$: Set $l = J_r$. Use equation

$$\mathbf{E}_{J_r} - \sum_{i=1}^{r-1} \mathbf{\Psi}_{J_r,J_i} = \mathbf{\Psi}_{J_r,J_r} \tag{4.14}$$

  and the JADE algorithm to identify the kernels corresponding to subsystems of order J_r as columns of the matrix $\mathbf{G}_{J_r,J_r}$.

- **End of Kernel Identification**

C**. Volterra Systems Equalization**

- Step 1: Based on theorem 3, compute the matrices $\mathbf{\Psi}_{J_r,J_m}$, for all $m, 1 \leq m \leq r-1$.

- Step 2: Set $l = J_r$ and compute the matrix

$$\mathbf{E}_{J_r} - \sum_{m=1}^{r-1} \mathbf{\Psi}_{J_r,J_m} \quad = \quad \mathbf{\Psi}_{J_r,J_r} \tag{4.15}$$

- Step 3: Use the first row of the computed matrix $\mathbf{\Psi}_{J_r,J_r}$ to equalize the input vector. Due to assumption A3 and the definition of $\mathbf{\Psi}_{J_r,J_r}$, the first row of $\mathbf{\Psi}_{J_r,J_r}$ equals the vector $[s(k) \quad s(k+1) \cdots \quad s(k+Q-1)]$.

- Step 4: Use the next block of Q received symbols, to compute for all $i, 1 \leq i \leq r$ the matrices $\mathbf{E}_{J_i}$, using the method described in Appendix A and theorem 1. Store the matrices.

- Step 5:Repeat steps $1-4$ to equalize next block of Q symbols. Do so, until all received symbols have been processed.

- Step 6:**End of Equalization**

For Volterra systems, after equalization has been performed, the vector of the linear kernel taps can be evaluated as the first column of $\mathbf{\Psi}_{J_r,J_r}$ divided by $s(k)$. Furthermore, in order to equalize input symbols using equation(4.15) we only need to know the values of the first row of each of the matrices $\mathbf{\Psi}_{J_r,J_m}$. It is also important to point out the difference in executing the algorithm in the case of LTI FIR MIMO systems, as opposed to that of SIMO Volterra systems. In the first case, having computed the matrices $E_{J_1}, E_{J_2}, \cdots, E_{J_r}$, using the first block of Q symbols (or any block of Q symbols) at the receiver, we use theorem 3 to compute the matrices $\mathbf{\Psi}_{J_1,J_1}, \mathbf{\Psi}_{J_2,J_2}, \cdots, \mathbf{\Psi}_{J_r,J_r}$ and identify the

corresponding subsystem kernels. Once this is done and the algorithm stops. In case of SIMO Volterra systems, for each block of Q received symbols, we compute the matrices $E_{J_1}, E_{J_2}, \cdots, E_{J_r}$ and use theorem 3 to compute the matrices $\mathbf{\Psi}_{J_r,J_r}$ and equalize the corresponding input symbols. This is done repetitively until all symbol blocks of length Q are processed. At the j-th repetition, symbols $[s((j-1)Q+k) \quad s(k+1)\cdots \quad s(k+jQ-1)]$ are equalized.

4.4 Simulation Examples

In this section, we present the simulation experiments conducted for the LTI FIR MIMO case, as well as the SIMO Volterra case. To be able to correlate order estimation and kernel identification/symbol equalization performance we simulate both parts of the algorithm.

4.4.1 Simulation Methodology

In simulation examples 1,2 and 4, uniformly distributed, mutually independent, temporally i.i.d. symbols drawn from a QPSK constellation $\{1+j, 1-j, -1+j, -1-j\}$ with variance equal to 2 were used as input. In simulation example 3 (real valued SIMO VOLTERRA system), uniformly distributed, mutually independent, temporally i.i.d., input symbols drawn from a BPSK constellation with variance equal to 1 were used. QPSK as well as BPSK input sequences were randomly generated, using appropriate MATLAB functions. Simulations were performed as follows. For each system, 1000 independent Monte-Carlo experiments were executed. For each experiment, an input sequence of 1500 symbols was generated, as described above. Having fixed the SNR value, 1000 runs of the algorithm per input sequence were performed. Then,

- For LTI FIR MIMO systems: The normalized mean-squared error per experiment and SNR value was computed, for each subchannel $\mathbf{h}_i$ of the system. It was computed as

$$10log_{10}(\frac{\sum_{j=1}^{1000}(|\mathbf{h}_i - \hat{\mathbf{h}}_i^j|^2/|\mathbf{h}_i|^2)}{1000})$$

  where $\hat{\mathbf{h}}_i^j$ is the subchannel estimate computed by the algorithm for the specific SNR value at the $j-th$ run. The procedure was repeated over all 1000 independent Monte-Carlo experiments.

The resulting value of $NMSE$ per subchannel and SNR value was then computed as

$$NMSE = \frac{\sum_1^{1000}(10log_{10}(\frac{\sum_{j=1}^{1000}(|\mathbf{h}_i - \hat{\mathbf{h}}_i^j|^2/|\mathbf{h}_i|^2)}{1000}))}{1000}$$

The above procedure was repeated for all SNR values and 95% confidence bounds, based on the 1000 Monte-Carlo experiments conducted, were computed.

- For SIMO VOLTERRA systems: The Symbol Error Rate (SER) was computed per experiment and SNR value, executing 1000 runs. The average was then calculated, per SNR value, for all different input sequences, executing 1000 Monte-Carlo runs, giving the SER per system and SNR value. The above procedure was repeated for all SNR values.

4.4.2 LTI FIR MIMO Systems

The results of simulations entailing kernel identification for the two first out of total three different LTI FIR MIMO systems presented in the previous chapter are presented. As already mentioned, QPSK input was used in both cases.

Example 1. In this example, we have simulated a two-input six-output MIMO system, described by the following equation:

$$\mathbf{y}(n) = \sum_{l=0}^{3} \mathbf{h}_1(l)s_1(n-l) + \sum_{l=0}^{5} \mathbf{h}_2(l)s_2(n-l) + \eta(n)$$

where $\eta(n)$ is additive white Gaussian noise. Order detection and kernel identification were performed. The performance of the proposed method was checked against the algorithm presented in [61]. Order detection results are shown in Table I. Kernel identification results are presented in Tables II and III. The proposed method achieves a slight performance gain.

Example 2. In this example, the second two-input six-output MIMO system was tested. The system is described by the following equation:

$$\mathbf{y}(n) = \sum_{l=0}^{4} \mathbf{h}_1(l)s_1(n-l) + \sum_{l=0}^{4} \mathbf{h}_2(l)s_2(n-l) + \eta(n)$$

where $\eta(n)$ is additive white Gaussian noise. Both subsystem orders are equal to 4. Order detection and kernel identification were performed. The performance of the proposed method was checked against the algorithm presented in [62]. Order detection results are shown in Table IV. Kernel identification results are presented in Tables V and VI. The proposed method achieves significant performance gain.

4.4.3 SIMO Volterra Systems

Simulations for SIMO Volterra systems were performed. Specifically, the proposed method is compared to the methods presented in [57, 64].

Example 3. We applied the proposed algorithm to the SIMO Volterra system described in *Example 1* of [64]. It is a three-output system, described by the following equation:

$$\mathbf{y}(n) = \sum_{l=0}^{2} \mathbf{h}_1(l)s(n-l) + \sum_{l=0}^{1} \mathbf{h}_{2,1}(l)s(n-l)s(n-1-l) + \eta(n)$$

where $\eta(n)$ is additive white Gaussian noise. As already mentioned, BPSK input was used. Kernel coefficients are presented in Table VII. Order determination results are presented in Table VIII.

Comparison was made against the methods presented in [57, 64]. Results are presented in Table IX.

Example 4. We applied the proposed algorithm to a three-output system, described by the following equation:

$$\mathbf{y}(n) = \sum_{l=0}^{3} \mathbf{h}_1(l)s(n-l) + \sum_{l=0}^{1} \mathbf{h}_{3,12}(l)s(n-l)s(n-1-l)s^*(n-l-2) + \eta(n)$$

where $\eta(n)$ is additive white Gaussian noise. As already mentioned, QPSK input was used. Kernel coefficients of the simulated system are presented in Table X. Order determination results are presented in Table XI. The proposed algorithm was checked against the methods presented in [57, 64]. Results are presented in Table XII.

4.5 Remarks

In this chapter we developed an algorithm for kernel identification of LTI FIR MIMO systems as well as channel equalization for SIMO Volterra systems. For this type of systems, simulations

showed that accurate order detection and channel equalization can be achieved for SNR values as low as 10dB. Theorem 1 and Appendix A allows for lower complexity computation of the data matrices $E_{J_1}, E_{J_2}, \cdots, E_{J_r}$ and $\mathbf{\Psi}_{J_1,J_1}, \mathbf{\Psi}_{J_2,J_2}, \cdots, \mathbf{\Psi}_{J_r,J_r}$ that are needed to perform either kernel identification or symbol equalization. Moreover, based on assumption A3, a more efficient approach is developed for equalization of SIMO Volterra systems.

For LTI FIR MIMO systems the algorithm allows kernel identification in steps, based on the JADE algorithm. Kernels are identified in groups, depending on the memory length of the subsystems they belong to. Simulations showed that it achieves a slight performance improvement when compared to the method presented in [61]. When compared to the algorithm presented in [62], the method showed superior performance. Performance improvement emanates from the fact that kernel identification is performed in steps, starting with the subsystems attaining the lower memory and progressing to the subsystems with the higher memory. The systems used in simulation examples were two out of the three presented in [60]. Their kernels were generated randomly. Extensive simulation experiments performed using randomly generated LTI FIR MIMO systems, led to similar conclusions. The computational complexity of the proposed method is $O(Q^4 + (ML\ Q \log_2(Q)))$, where Q is the size of symbols block, This is because, even though the method is SOS based, the involvement of the JADE algorithm increases the complexity.

For Volterra systems, the performance of the algorithm is quite satisfactory, compared to the methods presented in [57, 64]. Simulations showed that the new method performs better than the one given in [57]. Compared to the algorithm presented in [64], the performance of the two methods is similar. An exception is made in case of Example 3, for SNR equal to 20dB. In this case, the proposed algorithm needs twice the number of input samples (that is 3000 samples) to achieve the performance of that in [64]. We should, however, notice that, compared to the method given in [57], the new method has several advantages. At first, it does not require that the input covariance matrix be positive definite. Instead, it can be applied when input constellation is strictly greater than the order of the system under consideration, an assumption easily achieved in practice. Next, the memory length of the linear channel part should not be known in advance. It is computed by the algorithm. Furthermore, it should not be greater than the sum of the memory lengths of the system's nonlinearities. It suffices that it is greater than the memory length of each individual nonlinear kernel. On the other hand, the proposed algorithm does not built equalizers, as it is done in [57, 64]. Instead, as explained in Section 4.3, equalization is performed by processing output data in blocks of Q samples per block. Since the processing is done

in a batch mode, according to Appendix A, the per block computational complexity of the algorithm, increases as high as $O(ML\,Q\log 2(Q))$, where Q is the block size. The overall algorithm complexity is therefore, $(\frac{N}{Q})O(ML\,Q\log_2(Q))$.

For both LTI FIR MIMO systems and SIMO Volterra systems, selecting the appropriate size Q of symbol blocks affects the correct order determination and by that the matrices E_l based on which identification/equalization are performed. Selection of Q depends on SNR as well as the complexity of the system examined, that is the number of subsystems it comprises of and their orders. Extensive simulation experiments, involving both LTI FIR MIMO and Volterra systems triggered by different types of PSK input, suggested that $Q \geq 1500$.

Appendix A. Computation of the Error Projection Matrices

In the following, we provide an efficient way of computing residual error matrices E_l denoted for simplicity purposes as $\mathbf{E}(\ell)$. Assume that Q is the symbol block length.

$$\mathbf{V}_\ell = \left[\begin{array}{cc} \mathbf{y}(2w+L+1) & \dots \quad \mathbf{y}(Q) \\ \vdots & \mathbf{Z}_f(\ell) \\ \mathbf{y}(w+\ell+1) & \\ \hline \mathbf{y}(w+\ell) & \dots \\ \vdots & \mathbf{D}(\ell) \\ \mathbf{y}(w+1) & \\ \hline \mathbf{y}(w) & \dots \\ \vdots & \mathbf{Z}_p \\ \mathbf{y}(w+1) & \end{array}\right] \tag{A-1}$$

Define

$$\mathbf{Z}(\ell) = \begin{bmatrix} \mathbf{Z}_f(\ell) \\ \mathbf{Z}_p \end{bmatrix} \tag{A-2}$$

$\mathbf{D}(\ell)$ is a $\ell M \times (Q-L-2w)$ matrix.

$\mathbf{Z}(\ell)$ is a $(2w+L-1)M \times (Q-L-2w)$ matrix.

We define the projection matrix $\mathcal{P}_{\mathbf{Z}(\ell)}$ onto the subspace spanned by the columns of $\mathbf{Z}(\ell)$,

$$\mathcal{P}_{\mathbf{Z}(\ell)} = \mathbf{Z}^t(\ell)(\mathbf{Z}(\ell)\mathbf{Z}^t(\ell))^{-1}\mathbf{Z}(\ell) \tag{A-3}$$

The projection matrix onto the orthogonal complement of the column space of $\mathbf{Z}(\ell)$ is expressed as

$$\mathcal{P}^{\perp}_{\mathbf{Z}(\ell)} = \mathbf{I} - \mathcal{P}_{\mathbf{Z}(\ell)} \tag{A-4}$$

The residual error is given by

$$\mathbf{E}(\ell) = \mathcal{P}^{\perp}_{\mathbf{Z}(\ell)}\mathbf{D}(\ell) = (\mathbf{I} - \mathcal{P}_{\mathbf{Z}(\ell)})\mathbf{D}(\ell) \tag{A-5}$$

or

$$\mathbf{E}(\ell) = \mathbf{D}(\ell) - \mathbf{Z}^t(\ell)(\mathbf{Z}(\ell)\mathbf{Z}^t(\ell))^{-1}\mathbf{Z}^t(\ell)\mathbf{D}(\ell) = \mathbf{D}(\ell) - \mathbf{Z}(\ell)\mathbf{C}^t(\ell) \tag{A-6}$$

The smoother is defined as the solution of

$$\mathbf{Z}(\ell)\mathbf{Z}^t(\ell)\mathbf{C}(\ell) = \mathbf{Z}(\ell)\mathbf{D}^t(\ell) \tag{A-7}$$

Define

$$\mathbf{R}(\ell) = \mathbf{Z}(\ell)\mathbf{Z}^t(\ell), \quad \mathbf{d}(\ell) = \mathbf{Z}(\ell)\mathbf{D}^t(\ell) \tag{A-8}$$

Then

$$\mathbf{R}(\ell)\mathbf{C}(\ell) = \mathbf{d}(\ell) \tag{A-9}$$

The minimum squared error attained is given by

$$A^s(\ell) = \mathbf{D}(\ell)\mathbf{D}^t(\ell) - \mathbf{C}^t\mathbf{d}(\ell) \tag{A-10}$$

Notice that

1) Matrix $\mathbf{Z}(\ell)$ is block Toeplitz and corresponds to a multichannel FIR system having $2M$ entry channels.

2) Matrix $\mathbf{D}(\ell)$ corresponds to a multichannel FIR system having ℓM output channels.

3) The optimal smoother $\mathbf{C}(\ell)$ can be obtained as the solution of the normal equations at a cost of $O(A^2\ell M^2) + O(M\ell\, Q \log 2(Q))$ where $A = (2w + L - 1)M$.

4) Residual error $\mathbf{E}(\ell)$ can be obtained at a cost of $O(M\ell\, Q \log 2(Q))$.

NUMBER OF SYMBOLS	17 dB	20 dB	22 dB	25 dB	27 dB	30 dB
400	0%	2%	17%	64%	85%	96%
800	1%	79%	95%	98%	100%	100%
1500	6%	96%	99%	100%	100%	100%

Table 4.1. MIMO Systems example 1: Percentage of successful order detection

SNR (dB)	NMSE (dB) Proposed Method	95% Confidence Bound Proposed Method	NMSE (dB) Ding's Algorithm
20	-12.1	$(-12.6, -11.6)$	-10.7
25	-13.3	$(-13.7, -12.9)$	-11.2
30	-22.6	$(-22.9, -22.3)$	-20.9
35	-30.7	$(-30.9, -30.5)$	-28.5

Table 4.2. MIMO Systems example 1: Channel 1 estimation

SNR (dB)	NMSE (dB) Proposed Method	95% Confidence Bound Proposed Method	NMSE (dB) Ding's Algorithm
20	-12.9	$(-13.6, -12.2)$	-10.8
25	-14.5	$(-15.1, -13.9)$	-12.3
30	-22.7	$(-23.2, -22.2)$	-19.6
35	-30.4	$(-30.6, -30.2)$	-27.3

Table 4.3. MIMO Systems example 1: Channel 2 estimation

NUMBER OF SYMBOLS	20 dB	25 dB	28 dB	30 dB
400	5%	8%	17%	88%
800	42%	62%	98%	100%
1500	90%	93%	100%	100%

Table 4.4. MIMO Systems example 2: Percentage of successful order detection

SNR (dB)	NMSE (dB) Proposed Method	95% Confidence Bound Proposed Method	NMSE (dB) SSUB
20	-20.1	$(-20.6, -19.6)$	-15.4
25	-23.8	$(-24.3, -23.3)$	-17.6
30	-30.5	$(-30.8, -30.2)$	-25.9
35	-32.6	$(-32.8, -32.4)$	-26.7

Table 4.5. MIMO Systems, example 2: Channel 1 Estimation

SNR (dB)	NMSE (dB) Proposed Method	95% Confidence Bound Proposed Method	NMSE (dB) SSUB
20	-21.4	$(-22.1, -20.7)$	-14.3
25	-24.7	$(-25.2, -24.2)$	-19.2
30	-27.5	$(-27.8, -27.2)$	-26.7
35	-31.3	$(-31.5, -31.1)$	-28.5

Table 4.6. MIMO Systems, example 2. Channel 2 Estimation

$\mathbf{h}_1(0)$	$\mathbf{h}_1(1)$	$\mathbf{h}_1(2)$	$\mathbf{h}_{2,1}(0)$	$\mathbf{h}_{2,1}(1)$
1	-2.5	1	2	0.7
0.5	3	5	0.3	1.2
2	0	2	-0.7	3

Table 4.7. Example 3. Real valued SIMO Volterra system kernel taps

NUMBER OF SYMBOLS	10 dB	12 dB	15 dB	18 dB	20 dB
500	10%	32%	77%	94%	100%
1000	68%	80%	95%	98%	100%
1500	90%	96%	99%	100%	100%

Table 4.8. Example 3: Percentage of successful order detection

SNR dB	Proposed Algorithm	Valcarce et al	Giannakis et al
10	$O(10^{-1})$	$O(10^{-1})$	$O(10^{0})$
15	$O(10^{-2})$	$O(10^{-2})$	$O(10^{-1})$
20	$O(10^{-4})$	$O(10^{-6})$	$O(10^{-4})$

Table 4.9. Example 3: SIMO Volterra Systems, SER Calculation

$\mathbf{h}_1(0)$	$\mathbf{h}_1(1)$	$\mathbf{h}_1(2)$	$\mathbf{h}_1(3)$	$\mathbf{h}_{3,12}(0)$	$\mathbf{h}_{3,12}(1)$
1	$-2.5+2i$	$1+i$	$4+0.3i$	2	$0.7-0.8i$
$0.5+0.4i$	$3+2i$	$-1+i$	$5+i$	$0.3+0.2i$	$1.2+i$
$-1+i$	$1-2i$	$2+1.3i$	$-3+1.3i$	$-0.7+0.7i$	$3+0.1i$

Table 4.10. Example 4: Complex valued SIMO Volterra system kernel taps

NUMBER OF SYMBOLS	10 dB	12 dB	15 dB	18 dB	20 dB
500	10%	30%	72%	93%	100%
1000	60%	85%	94%	99%	100%
1500	90%	95%	99%	100%	100%

Table 4.11. Example 4: Percentage of successful order detection

SNR dB	Proposed Algorithm	Valcarce et al	Giannakis et al
10	$O(10^{-1})$	$O(10^{-1})$	$O(10^{0})$
15	$O(10^{-2})$	$O(10^{-2})$	$O(10^{-1})$
20	$O(10^{-3})$	$O(10^{-3})$	$O(10^{-1})$

Table 4.12. Example 4: SIMO Volterra Systems, SER Calculation

5. Conclusions

We have presented a uniform algorithm, which, knowing the type of system under examination, can perform order detection and kernel equalization or symbol equalization, by solving the problem in steps. Order detection computations produce data structures that can be used for identification and/or equalization. This chapter presents the conclusions of our work and points out the direction for future work

5.1 Order Determination

The performance of the proposed algorithm in the presence of noise relies heavily on the validity of *Criteria* 1 and 2. Among them, the first one, that is the determination of the effective rank of a matrix perturbed by noise, is the most sensitive. In [11] various criteria are given that estimate the effective rank of a perturbed matrix using its singular values. All but one out of these, use threshold values that either do not appear to be based on any explicit analytical expressions but are selected on an ad hoc basis, or lower and upper bounds for them can be derived analytically assuming known noise statistics. The only criterion among those presented in [11] that does not use either an ad hoc threshold value or a threshold with upper and lower bounds based on known noise statistics, determines the numerical rank t of a perturbed matrix **B** from its singular values $_1 \geq _2 \geq \cdots \geq _r$ as the index t for which $_t >> _{t+1}$. It is straightforward to see that it is equivalent to Criterion 1, used by the proposed method. That is, it is equivalent to determining the effective rank t of a perturbed matrix **B** as the index t for which the ratio $_t/_{t+1}$ is maximized. This is the reason it is also referred as the "ratio test".

It is well known that the squares of the singular values of a matrix **B** are the eigenvalues of the matrix $\mathbf{BB}'$. Therefore, maximizing the ratio of successive singular values of **B** is equivalent to maximizing the ratio of the corresponding eigenvalues of the matrix $\mathbf{BB}'$.

It is also equivalent to the one given in [6] by Liavas et al, provided that there exists at least one i for which the successive eigenvalue ratio $\lambda_i/\lambda_{i+1} \geq 3$.

Simulation experiments showed the following:

- For MIMO systems, the validity of Criterion 2 was established even for low SNR values. On the other hand, the rank pattern suggested by Theorem 3 is harder to detect in the presence of noise.
- For SIMO systems the proposed method outperforms Liavas et al algorithm in all cases at least

by 6dB. MDL achieves the 90% success rate at about the same SNR levels with the proposed method, with an exception occurring in Example 2, where it proved worse by 3dB. For the microwave channels tested in Examples 3 and 4 MDL's performance proved poor for high SNR values due to overmodelling, while the proposed algorithm achieved a 100% success rate. The new method outperforms the J-LSS algorithm by at least 14dB, while J-LSS performed poorly in the case of true microwave channels.

- Numerous simulations conducted for MIMO systems showed that increasing the number of outputs improves the algorithm's performance. This implies additional hardware complexity and increases computational cost. On the contrary, two outputs proved sufficient for SIMO systems.

- Denoising improved dramatically the algorithm's performance for MIMO systems. In Examples 2 and 3, using 1000 output vectors and no denoising, the method achieves a success rate greater or equal to 90% for SNR levels of 37-38 and 41-42 dB respectively. The corresponding levels when denoising is applied drop down to 27-28 and 32-33 dB respectively.

  For SIMO systems, denoising had a mixed effect. In case of Examples 1 and 2, where all taps of the system have significant power, the performance was reduced because finding the effective rank of $\mathbf{Z}_{w,l}$ took away significant data information. In the case of Examples 3 and 4, that are real microwave channels where the system's energy is concentrated in a few taps but there also exist non zero head and tail taps, denoising improved the method. Finally, we should mention the performance boost that was achieved when denoising was applied at the J-LSS method.

- When the algorithm misses the correct value of the system order for the case of SIMO systems this results in system undermodelling in the vast majority of cases. Misses in the MIMO case can be interpreted as the false detection either of systems attaining non existing orders or of more systems that attain a specific order than they really are.

- In all our simulation experiments w was chosen equal to L the upper bound of system's orders. This is because, as noticed in [60], for a fixed data length, large values of w imply fewer columns in the algorithm's data matrices that correspond to smaller sample size when projections on subspaces are evaluated. Therefore, we tried to keep w as small as possible.

- We have conducted extensive simulation experiments on the microwave channels given in *http://spib.rice.edu/spib/microwave.html*. Most of these channels, (12 out of a total of 15), are

"good shaped". This means that their significant part (the part that contains most of the channel power) consists of a small number of successive taps. Applying the proposed method to all these channels, we saw that the estimated effective channel length varied, depending on the channel, from two to four taps that correspond to over 95% of the total channel power. As it was the case for channels 10 and 13, simulations verified the superiority of the proposed method's performance against all the other methods.

However, there are a few channels in *http://spib.rice.edu/spib/microwave.html* such as 3, 14 and 15, (3 out of a total of 15), that are "bad shaped". This means that most of the channel's power is not contained in a number of successive taps as it was the case with "good shaped" channels. Rather, there exist groups of successive taps that concentrate a significant amount of the channel's power as well as sporadic taps of significant power. An example of a "bad shaped" channel is channel 3. For this channel, kernel taps 0 to 9 are approximately zero, tap 10 accounts for 29% of the total channel power, tap 11 for 58.5%, tap 12 for 2.6%, tap 13 for 1%, while taps 13 to 37 are "small". However, tap 38 is a significant, sporadic tap, accounting for approximately 5% of the total channel power while the remaining terms of the channel impulse response are again "small" tail taps. Simulation experiments that were run using the proposed as well as all the other methods indicated one as the estimated effective channel order. The two taps found (taps 10 and 11) account for 87.5% of the total channel power, while the remaining 12.5% missed makes equalization poor for this channel, when SNR$\leq$40dB, as it is stated in [6].

The algorithm gives analogous results for order detection of Volterra Systems. It is important to notice that using 1500 symbols the algorithm achieves high rates of successful order detection even for low SNR values such as 10dB. This is due to the special structure of the Volterra systems, that is due to linearization of the system, where the new inputs were redefined as subproducts of delayed input values.

- The proposed algorithm performs $L+1$ SVD operations in step 2, $L+1$ SVD operations in step 3 and $L-J_1+1$ SVD operations in steps 4 and 5 in case of MIMO systems. For SIMO systems it performs $L+1$ SVD operations in step 2 and $L+1$ SVD operations in step 3, since SVD calculations associated with steps 4 and 5 are not performed. To calculate its complexity we take into account the cost of an SVD decomposition in flops, as well as the number of flops required for matrix multiplications in steps 2 (due to the reconstruction of $\mathbf{Z}_{w,l}$) and 3. Thus, we

conclude that the computational complexity of the proposed algorithm is $O(M^3Q^2(L+1)^4 + M^3QL^3(L+1) + M^3Q^2L^3(L+1) + M^3Q(L+1)^3(L-J_1+1))$ for MIMO systems. This reduces to $O(M^3Q^2(L+1)^4 + M^3QL^3(L+1) + M^3Q^2L^3(L+1))$ for SIMO systems. In the above M is the number of outputs and Q is the number of output data samples used. This suggests that the oversampling rate or the number of diverse sensors M used at the receiver should be kept small to avoid heavy computations. In addition, the less output samples used and the tighter the estimate of the system orders' upper bound L is, the less expensive computations become.

5.2 LTI FIR MIMO Kernel Identification and SIMO Volterra Systems Equalization

For LTI FIR MIMO systems the algorithm allows kernel identification in steps, based on the JADE algorithm. Kernels are identified in groups, depending on the memory length of the subsystems they belong to. Performance improvement emanates from the fact that kernel identification is performed in steps, starting with the subsystems attaining the lower memory and progressing to the subsystems with the higher memory.

The computational complexity of the proposed method is $O(Q^4 + (ML\,Q\log_2(Q)))$, where Q is the size of symbols block, This is because, even though the method is SOS based, the involvement of the JADE algorithm increases the complexity, as it requires the computation of fourth-order cumulants.

For Volterra systems, the performance of the algorithm is quite satisfactory. It has several advantages as it does not require that the input covariance matrix be positive definite. Instead, it can be applied when input constellation is strictly greater than the order of the system under consideration, an assumption easily achieved in practice. The method computes the memory lengths of the subsystems involved, without requiring any explicit assumption about them. On the other hand, the proposed algorithm does not built equalizers. Instead, as explained in Section 4.3, equalization is performed by processing output data in blocks of Q samples per block. Since the processing is done in a batch mode, according to Appendix A, the per block computational complexity of the algorithm, increases to $O(ML\,Q\log 2(Q))$, where Q is the block size. The overall algorithm complexity is therefore, $(\frac{N}{Q})O(ML\,Q\log_2(Q))$.

For both LTI FIR MIMO systems and SIMO Volterra systems, selecting the appropriate size Q of symbol blocks affects the correct order determination and by that the matrices E_l based on which

identification/equalization are performed. Selection of Q depends on SNR as well as the complexity of the system examined, that is the number of subsystems it comprises of and their orders. Extensive simulation experiments, involving both LTI FIR MIMO and Volterra systems triggered by different types of PSK input, suggested that $Q \geq 1500$.

5.3 Future Work

As it has already been mentioned the algorithm algorithm's complexity is $O(M^3Q^2(L+1)^4 + M^3QL^3(L+1) + M^3Q^2L^3(L+1) + M^3Q(L+1)^3(L-J_1+1))$ for the order identification part of MIMO systems and $O(M^3Q^2(L+1)^4 + M^3QL^3(L+1) + M^3Q^2L^3(L+1))$ for SIMO systems. Moreover, we pay $O(Q^4 + (ML\,Q\log_2(Q)))$ operations for identification or $(\frac{N}{Q})O(ML\,Q\log_2(Q))$ for Volterra systems equalization. It is obvious that there should be some future work to improve computational complexity. This can be achieved either by ausing a recursive implementation scheme or by introducing other type of BSS techniques, with less complexity than the JADE algorithm.

Bibliography

[1] V. Volterra, “Theory of functions and of Integral and Integro-Differential Equations ,” Dover 1959.

[2] C.W. Therrien, “The Lee-Wiener legacy (statistical theory of communication),” *IEEE Signal Processing Magazine,”* vol. 19(6), pp. 33-34, 2002.

[3] J.F. Barret, “The use of functionals in the analysis of non-linear physical systems,” *Journal of Electronic Control,* vol. 15, pp. 567-615, 1963.

[4] M. Schetzen, “The Volterra and Wiener Theories of Nonlinear Systems,” Wiley, 1980.

[5] W.J. Rugh, “Nonlinear System Theory: The Volterra/Wiener Approach,” John Hopkins University Press, 1981.

[6] A. P. Liavas, P. A. Regalia and J.-P. Delmas “Blind channel approximation: effective channel order determination,” *IEEE Trans. Signal Processing,* vol. 47, pp. 3336-3344, Dec. 1999.

[7] L. Tong, G. Xu and T. Kailath, “Blind identification and equalization based on second-order statistics: a time domain approach,” *IEEE Trans. Inform. Theory*, vol. 40, pp. 340-349, Mar. 1994.

[8] A. P. Liavas, P. A. Regalia and J.-P. Delmas “Robustness of least-squares and subspace methods for blind channel identification/equalization with respect to effective channel undermodelling/overmodelling,” *IEEE Trans. Signal Processing,* vol. 47, pp. 1636-1645, June 1999.

[9] L. Tong, G. Xu and T. Kailath, “A new approach to blind identification and equalization of multipath channels,” in *Proc. 25th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1991, pp. 856-860.

[10] E. Moulines, P. Duhamel, J. F. Cardoso and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing,* vol. 43, pp. 516-525, Feb. 1995.

[11] K. Konstantinides and K. Yao, "Statistical analysis of effective singular values in matrix rank determination," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 36, pp. 757-763, May 1988.

[12] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria," *IEEE Trans. Signal Processing*, vol. 45, pp. 658-672, Mar. 1997.

[13] Z. Ding and L. Qiu, "Blind MIMO channel identification from second order statistics using rank deficient channel convolution matrix," *IEEE Trans. Signal Processing*, vol. 51, pp. 535-544, Feb. 2003.

[14] N. Kalouptsidis, S. Theodoridis, "Adaptive System Identification and Signal Processing Algorithms," Pearson Education Limited.

[15] C-Y Chi et al, Blind Equalization and System Identification. NewYork:Marcel Dekker, 2001.

[16] A. Gorokhov, P. Loubaton and E. Moulines, "Second order blind equalization in multiple input multiple output FIR systems: A weighted least squares approach," in *Proc. ICASSP*, Atlanta, GA, May 1996, pp. 2415-2418.

[17] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 43, pp. 2714-2723, Nov. 1995.

[18] E. Moulines, P. Duhamel, J. F. Cardoso and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing,* vol. 43, pp. 516-525, Feb. 1995.

[19] Ding and Y. Li, ed., Blind Equalization and Identification. NewYork:Marcel Dekker, 2001.

[20] S. Haykin, ed., Blind Deconvolution. New Jersey: Prentice-Hall, 1994.

[21] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, ed., Signal Processing Advances in Wireless and Mobile Communications. Upper Saddle River: Prentice-Hall, 2001.

[22] K. K. Paliwal and M. M. Sondhi, ”Recognition of noisy speech using cumulantbased linear prediction analysis,” in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, Toronto, Ontario, Apr. 14-17, 1991, pp. 429-432.

[23] Z. Ding, “Matrix outer-product decomposition method for blind multiple channel identification,” *IEEE Trans. Signal Processing*, vol. 45, pp. 3053-3061, Dec. 1997.

[24] Tong and S. Perreau, ”Multichannel blind identification: From subspace to maximum likelihood methods,” Proceedings of the IEEE, vol. 86, no. 10, pp. 1951-1968, Oct. 1998

[25] T. Endres, B. D. O. Anderson, C. R. Johnson, and L. Tong, ”On the robustness of FIR channel identification from second-order statistics,” IEEE Signal Processing Letters, vol. 3, no. 5, pp. 153-155, May 1996.

[26] H. Ali, J. H. Manton, and Y. Hua, ”Modified channel subspace method for identification of SIMO FIR channels driven by a trailing zero filter band precoder,” in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, USA, May 7-11, 2001, pp. 2053-2056. 15. T. Kailath, Linear Systems. New Jersey: Prentice-Hall, 1980.

[27] J. K. Tugnait, “Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria,” *IEEE Trans. Signal Processing*, vol. 45, pp. 658-672, Mar. 1997.

[28] Y. Hua, S. An, and Y. Xiang, ”Blind identification and equalization of FIR MIMO channels by BIDS,” in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, USA, May 7-11, 2001, pp. 2157-2160.

[29] S. An and Y. Hua, ”Blind signal separation and blind system identification of irreducible MIMO channels,” in Proc. Sixth IEEE International Symposium on Signal Processing and its Applications, Kuala Lumpur, Malaysia, Aug. 13-16, 2001, pp. 276-279.

[30] D. T. M. Slock, ”Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction,” in Proc. IEEE International References 333 Conference on Acoustics, Speech, and Signal Processing, Adelaide, SA, Australia, Apr. 19-22 1994, pp. 585-588.

[31] R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis. New York: John Wiley, 1973.

[32] S. Choi, A. Cichocki, and A. Belouchrani, ”Second order nonstationary source separation,” Journal of VLSI Signal Processing, vol. 32, no. 1-2, pp. 93-104, Aug. 2002.

[33] 27. J. K. Tugnait, ”Adaptive blind separation of convolutive mixtures of independent linear signals,” Signal Processing, vol. 73, no. 1-2, pp. 139-152, Feb. 1999.

[34] Y. Inouye and T. Sato, ”Iterative algorithms based on multistage criteria for multichannel blind deconvolution,” IEEE Trans. Signal Processing, vol. 47, no. 6, pp. 1759-1764, June 1999.

[35] 29. Y. Inouye, ”Criteria for blind deconvolution of multichannel linear timeinvariant systems,” IEEE Trans. Signal Processing, vol. 46, no. 12, pp. 3432- 3436, Dec. 1998.

[36] 30. K. L. Yeung and S. F. Yau, ”A cumulant-based super-exponential algorithm for blind deconvolution of multi-input multi-output systems,” Signal Processing, vol. 67, no. 2, pp. 141-162, 1998.

[37] M. K. Tsatsanis and G. B. Giannakis, ”Blind estimation of direct sequence spread spectrum signals in multipath,” IEEE Trans. Signal Processing, vol. 45, no. 5, pp. 1241-1251, May 1997.

[38] Y. Li and Z. Ding, ”Global convergence of fractionally spaced Godard (CMA) adaptive equalizers,” IEEE Trans. Signal Processing, vol. 44, no. 4, pp. 818-826, Apr. 1996.

[39] Z. Ding and T. Nguyen, ”Stationary points of a kurtosis maximization algorithm for blind signal separation and antenna beamforming,” IEEE Trans. Signal Processing, vol. 48, no. 6, pp. 1587-1596, June 2000.

[40] Y. Li and K. J. R. Liu, ”Adaptive blind source separation and equalization for multiple-input/multiple-output systems,” IEEE Trans. Information Theory, vol. 44, no. 7, pp. 2864-2876, Nov. 1998.

[41] J. Gomes and V. Barroso, ”A super-exponential algorithm for blind fractionally spaced equalization,” IEEE Signal Processing Letters, vol. 3, no. 10, pp. 283- 285, Oct. 1996.

[42] C. B. Papadias and A. J. Paulraj, ”A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays,” IEEE Signal Processing Letters, vol. 4, no. 6, pp. 178-181, June 1997.

[43] O. Shalvi and E. Weinstein, ”Super-exponential methods for blind deconvolution,” IEEE Trans. Information Theory, vol. 39, no. 2, pp. 504-519, Mar. 1993.

[44] J. Liang and Z. Ding, “Blind MIMO system identification based on cumulant subspace decomposition,” *IEEE Trans. Signal Processing,* vol. 51, pp. 1457-1468, June 2003.

[45] J. K. Tugnait, “Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria,” *IEEE Trans. Signal Processing*, vol. 45, pp. 658-672, Mar. 1997.

[46] A. Gorokhov, P. Loubaton and E. Moulines, “Second order blind equalization in multiple input multiple output FIR systems: A weighted least squares approach,” in *Proc. ICASSP*, Atlanta, GA, May 1996, pp. 2415-2418.

[47] K. Abed-Meraim, P. Loubaton and E. Moulines, “A subspace algorithm for certain blind identification problems,” *IEEE Trans. Inform. Theory*, vol. 43, pp. 499-511, Mar. 1997.

[48] H. Liu and G. Xu, “Closed-form blind symbol estimation in digital communications,” *IEEE Trans. Signal Processing*, vol. 43, pp. 2714-2723, Nov. 1995.

[49] J-F, Cardoso, ”Higher Order Contrasts for Independent Component Analysis”, *Neural Computation*, vol. 11, pp. 157-192, 1999.

[50] V. J. Mathews, G.L. Sicuranza, Polynomial Signal Processing, Wiley Series in Telecommunications and Signal Processing, John Wiley, 2000.

[51] G. Mileounis, ”Nonlinear Signal Processing and its applications to communications”, Ph.D. Thesis, National Kapodistrian University of Athens, June 2010.

[52] K. Abed-Meraim and Y. Hua, “Blind identification of multi-input multi-output system using minimum noise subspace,” *IEEE Trans. Signal Processing*, vol. 45, pp. 254-258, Jan. 1997.

[53] S. Benedetto, E. Biglieri, ”Nonlinear Equalization of Digital Satellite Channels”, *IEEE Journal Selected Areas of Communicatios*, vol. SAC-1, pp. 57-62, Jan. 1983.

[54] A. Swami, G.B. Giannakis and G. Zhou, ”Bibliography on Higher Order Statistics,”, *Signal Processing*, vol. 60, no 1, pp. 65-126, 1997.

[55] R. Lopez-Valcarce, Z. Ding and S. Dasgupta, “Equalization and interference cancellation in linear multiuser systems based on second-order statistics,” *IEEE Trans. Signal Processing*, vol. 49, pp. 2042-2049, Sept. 2001.

[56] C.-H. Tseng, E. J. Powers, “Nonlinear channel equalization in digital satellite systems,” *Proc. GLOBECOM,*, Houston, TX, Nov./Dec 29-2, 1993, pp. 1639-1643.

[57] G.B. Giannakis, E. Serpedin, Linear Multichannel Blind Equalizers of Nonlinear FIR Volterra Channels, IEEE Transactions on Signal Processing, vol. 45, no. 1, pp. 67-81, January 1997. vol. 49, pp. 2042-2049, Sept. 2001.

[58] L. Tong and Q. Zhao, “Joint order detection and blind channel estimation by least squares smoothing,” *IEEE Trans. Signal Processing,* vol. 47, pp. 2345-2355, Sept. 1999.

[59] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 387-392, Apr. 1985.

[60] D. Kotoulas, P. Koukoulas, N. Kalouptsidis, Subspace Projection Based Blind Channel Order Estimation of MIMO Systems, IEEE Transactions on Signal Processing, vol. SP-54, no. 4, pp. 1351-1363, April 2006. Sept. 2003, pp. 409-412.

[61] J. Liang and Z. Ding, “Blind MIMO system identification based on cumulant subspace decomposition,” *IEEE Trans. Signal Processing,* vol. 51, pp. 1457-1468, June 2003.

[62] S. An, J. H. Manton, Y. Hua, ”A Sequential Subspace Method for Blind Identification of General FIR MIMO Channels,” *IEEE Transactions on Signal Processing*, vol. 53, pp. 3906-3910, October 2005.

[63] Serafeim Karambogias, ”Signals and Systems”, (book in greek)

[64] R. Lopez-Valcarce and S. Dasgupta, “Blind Equalization of Nonlinear Channels From Second-Order Statistics,” *IEEE Trans. Signal Processing*, vol. 49, pp. 3084-3097, Sept. 2001.

[65] G.M. Raz and B. D. Van Veen, ”Blind Equalization and Identification of Nonlinear and IIR systems - A Least Squares approach”, *IEEE Trans. Signal Processing*, vol. 48, pp. 192-200, Jan. 2000.

[66] V. Z. Marmarelis, ”Advanced Models of Physiological System Modeling”, New York, Plenum 1994.

[67] M. Tsatsanis and H. Cirpan, ”Blind Identification of Nonlinear Channels Excited by Discrete Alphabet Inputs”, *IEEE Signal Processing Workshop Statistical Signal Array Processing*, vol. 1, Corfu,Greece, pp. 176-179.

[68] S. Benedetto, E. Biglieri, and V. Castellani, ”Digital Transmission Theory”, Englewood Cliffs, NJ: Prentice-Hall, 1987

[69] J-F Cardoso, ”High-Order Contrasts for Independent Component Analysis”, *Neural Computation*, vol. 11, pp. 157-192, 1999.

[70] A. Gorokhov and P. Loubaton, ”Subspace-based techniques for blind separation of convolutive mixtures with temporally correlated sources”, *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 44, no 9, pp. 813-820, Sep. 1997.