# MA Parameter Estimation Using Higher-Order Cumulant Statistics 

Achilleas G Stogioglou



A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
-May 1996-


# Abstract 

The ability of higher-order statistics to preserve phase information makes them particularly useful in the study of non-Gaussian stationary linear processes amongst other things. This thesis derives some new results in the estimation of the parameters of MA models from the cumulants of the output processes.

New general relationships between the output cumulants and the system parameters are derived. These relationships involve different cumulant slices of the same order and the system parameters, and are used to develop new system identification methods which use only thirdorder or only fourth-order cumulants. Both least squares and recursive versions of the system identification algorithms are proposed. The identifiability of the algorithms is formally proved and asymptotic performance expressions are derived. Previous techniques of the same type required the use of second order statistics in order to ensure identifiability, sacrificing in this way the advantage of HOC -based methods in the presence of additive coloured Gaussian noise. The important issue of model order selection is also addressed and a practical method based on the minimisation of a cumulant-error function is proposed. It is also demonstrated that the MA parameter estimation methods are useful for the estimation of ARMA model parameters through double-MA modeling.

In many applications the primary objective is the estimation of the inverse filter coefficients. New general relationships are derived which involve the output cumulants and the inverse filter coefficients. Based on these relationships, a unified description of existing deconvolution methods is proposed and new deconvolution methods based on fourth-order cumulants or on a combination of second- and fourth-order cumulants are developed.

Finally, this thesis investigates properties that characterise sets of numbers as being the cumulants of some MA model. This problem is easier to analyse if the numbers are organised in a matrix form and the properties are expressed using matrix theoretic notions such as the rank of a matrix and the features of linear structured matrices. Because of estimation errors, sets of sample cumulants are not real cumulants of some MA model. Based on the characteristic properties of sets of cumulants, this thesis presents an iterative composite property mapping algorithm which maps the sample cumulants to a set of enhanced cumulants. If convergence is achieved, the enhanced cumulants are true cumulants of some MA model. If convergence has not been achieved, the enhanced cumulants are "nearer" to a set of true MA cumulants than the original set of sample cumulants was. It is shown that when the enhanced cumulants are used for parameter estimation, they can improve the performance of parameter estimation algorithms.

## Declaration of Originality

I hereby declare that this thesis and the work reported herein was composed and originated entirely by myself in the Department of Electrical Engineering at The University of Edinburgh.

## Acknowledgements

I would like to thank my supervisors Dr Steve McLaughlin and Prof. Anton Ziolkowski for their continuous support and guidance, and for reading and checking this thesis.

I would also like to thank the other members of the Signals and Systems Group for their support throughout my PhD. Special thanks go to Justin Fackrell who took the time to review the thesis.

Thanks are also due to the Leventis Foundation and NERC for providing financial support.

Finally, I would like to thank my family for their support and my fiancee, Marina, for her patience and support.

## Contents

List of Abbreviations ..... viii
List of Symbols ..... ix
List of Figures ..... xi
List of Tables ..... xii
1 Introduction ..... 1
1.1 Introduction ..... 1
1.2 Stochastic Modeling ..... 1
1.2.1 Modelling Assumptions ..... 2
1.3 Estimation Using Higher-Order Statistics ..... 5
1.4 Thesis Organisation ..... 6
2 Higher-Order Statistics ..... 8
2.1 Introduction ..... 8
2.2 Moments and Cumulants ..... 9
2.3 Properties of Moments and Cumulants ..... 11
2.4 Cumulants and Polyspectra ..... 12
2.4.1 Cumulants of Stationary Random Processes ..... 12
2.4.2 Polyspectra of Stationary Random Processes ..... 14
2.4.3 Cumulants and Polyspectra of $k^{t h}$-order White Non-Gaussian Noise ..... 16
2.5 Cumulants and Polyspectra of Non-Gaussian Linear Processes ..... 17
2.6 Cumulants of MA, AR and ARMA Processes ..... 19
2.6.1 Phase Properties of Systems with Rational Transfer Functions ..... 23
2.7 Estimation of Cumulants from Finite Samples ..... 27
2.7.1 Alternative Cumulant Estimation Methods ..... 30
2.8 Summary ..... 31
3 Moving Average Parameter Estimation ..... 32
3.1 Introduction ..... 32
3.2 Problem Definition ..... 33
3.2.1 The Method of Moments applied to MA parameter estimation ..... 34
3.3 Linear Methods for MA Parameter Estimation ..... 35
3.3.1 Relationships between cumulants of different order ..... 36
3.3.2 Relationships between different cumulant slices of the same order ..... 37
3.4 A Unified Framework for the Description of Linear Methods for MA identification ..... 38
3.5 MA Parameter Estimation Using Only Third-Order Cumulants ..... 41
3.5.1 Relationships between third-order cumulant slices ..... 41
3.5.2 Least-Squares Method using only Third-Order Cumulants ..... 42
3.5.3 Compensating for the over-parameterisation of the system of equa- tions ..... 43
3.5.4 A Closed-Form Solution and Uniqueness of the Least Squares Method ..... 45
3.6 MA Parameter Estimation Using Only Fourth-Order Cumulants ..... 47
3.6.1 Relationships between fourth order cumulants ..... 48
3.6.2 Least Squares Solution Using only Fourth-Order Cumulants ..... 51
3.6.3 A Recursive Solution Based on Fourth-Order Cumulants ..... 51
3.7 Asymptotic Performance Analysis ..... 52
3.8 MA Model Order Determination ..... 54
3.8.1 Model Order Selection Using a Criterion Based on the Confidence Factor ..... 55
3.8.2 Order selection through minimisation of a cumulant-error measure ..... 56
3.9 Applications to ARMA Parameter Estimation ..... 57
3.10 Numerical Simulations ..... 58
3.10.1 Third-order cumulant based estimation ..... 59
3.10.2 Fourth-order cumulant based estimation ..... 66
3.10.3 Results on model order selection ..... 68
3.11 Conclusion ..... 69
4 Blind Deconvolution of MA Models ..... 70
4.1 Introduction ..... 70
4.2 Problem Definition ..... 70
4.3 Direct Deconvolution ..... 72
4.3.1 Fundamental relationships for the inverse filter parameters ..... 73
4.3.2 Deconvolution scheme using only third-order cumulants ..... 75
4.3.3 Extension to fourth order cumulants ..... 80
4.3.4 Deconvolution schemes involving second- and third-order cumulants ..... 80
4.3.5 General Comments ..... 83
4.4 Indirect Deconvolution ..... 84
4.5 System Identification Through Deconvolution ..... 85
4.6 Asymptotic Performance Analysis ..... 86
4.7 Numerical Simulations ..... 88
4.7.1 Third-order cumulant-based deconvolution ..... 89
4.7.2 Fourth-order cumulant based deconvolution ..... 95
4.8 Conclusions ..... 102
5 MA Cumulant Enhancement ..... 103
5.1 Introduction ..... 103
5.2 The Linear Structure of Third-Order MA Cumulant Slices ..... 104
5.2.1 Geometric interpretation ..... 105
5.2.2 Characteristic properties of third-order MA cumulants ..... 107
5.3 Joint Structure of Second- and Third-Order Cumulants ..... 112
5.4 The Linear Structure of Fourth-Order MA Cumulant Slices ..... 116
5.4.1 Geometric interpretation ..... 116
5.4.2 Properties of a subset of the fourth order cumulants ..... 119
5.5 MA Cumulant Enhancement ..... 122
5.5.1 Mapping corresponding to the rank property ..... 123
5.5.2 Mapping corresponding to the linear structure property ..... 124
5.5.3 An iterative algorithm based on the composite mapping ..... 126
5.5.4 Asymptotic behavior of the cumulant enhancement algorithm ..... 127
5.6 Numerical Results ..... 128
5.7 Conclusions ..... 132
6 Summary and Conclusions ..... 137
6.1 Introduction ..... 137
6.2 Achievements of the Work ..... 137
6.3 Future Work ..... 139
Appendix A ..... 140
Appendix B ..... 142
List of Publications ..... 146
Bibliography ..... 147

# List of Abbreviations 

## AR

ARMA
CPMA
FIR
HOC
HOS
IID
IIR
LS
LTI
MA
SVD
TLS

## Autoregressive

Autoregressive Moving Average
Composite Property Mapping Algorithm
Finite Impulse Response
Higher-Order Cumulant
Higher-Order Statistics
Independent Identically Distributed
Infinite Impulse Response
Least Squares
Linear Time Invariant
Moving Average
Singular Value Decomposistion
Total Least Squares

## List of Symbols

()$^{\top}$
$\Phi_{X}(\xi)$
$M_{X}(\lambda)$
$K_{X}(\lambda)$
$\kappa_{n}(X)=\kappa_{n}\left(X_{1}, \ldots, X_{n}\right)$
$\mu_{n}(X)=\mu_{n}\left(X_{1}, \ldots, X_{n}\right)$
$c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)$
$m_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)$
$\hat{c}_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)$
$\hat{m}_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)$
$S_{k, x}\left(\omega_{1}, \ldots, \omega_{k-1}\right)$
$x(t) * y(t)$
$\otimes$
$H(z)$
$h(t)$
$\mathbf{0}_{\mathbf{m}, \mathbf{n}}$
$\|\cdot\|_{F}$
$\|\cdot\|_{E}$
$z^{*}$

Matrix/vector transpose.
The characteristics function of the random vector $X$.
Moment generating function of random vector $X$.
Cumulant generating function of random vector $X$.
$n^{\text {th }}$-order cumulant of the random vector $X=\left[X_{1}, \ldots, X_{n}\right]^{\top}$.
$n^{t h}$-order moment of the random vector $X=\left[X_{1}, \ldots, X_{n}\right]^{\top}$.
$k^{\text {th }}$-order cumulant of the stationary process $\{x(t)\}$.
$k^{t h}$-order moment of the stationary process $\{x(t)\}$. sample $k^{\text {th }}$-order cumulant of the stationary process $\{x(t)\}$.
sample $k^{t h}$-order moment of the stationary process $\{x(t)\}$.
$k^{t h}$ order polyspectrum of the ranndom process $\{x(t)\}$. convolution of the sequences $x(t)$ and $y(t)$.
Kronecker product operator.
transfer function of a linear time invariant system. impulse response of a linear time invariant system. ( $m \times n$ ) matrix whose elements are all zero. Frobenius matrix norm.
Euclidean vector norm.
the complex conjugate of $z$.

## List of Figures

2.1 (a) Symmetry regions of third-order cumulants $c_{3, x}\left(\tau_{1}, \tau_{2}\right)$. (b) Symme- try regions of the bispectrum $S_{3, x}\left(\omega_{1}, \omega_{2}\right)$. ..... 16
2.2 (a) The finite domain of support for $3^{\text {rd }}$ order cumulants of an MA(q) process. ..... 21
2.3 (a) The finite domain of support for $4^{\text {th }}$ order cumulants of an MA(4) process. ..... 22
2.4 A zero $z_{i}$ and its reciprocal conjugate $\frac{1}{z_{i}^{*}}$. ..... 24
2.5 The autocorrelation sequence is the same for all systems ..... 25
2.6 Third-order cumulants are different for the spectrally equivalently sys- tems ..... 26
3.1 Problem definition. ..... 33
3.2 2-D domain of support for cumulant slice $c_{4}(n, \alpha, \beta) n \geq 0$. ..... 50
3.3 Identification results for signal model 2. The number of output samples is 5000 . The graphs display the performance of system identification methods for $\mathrm{SNR}=50 \mathrm{~dB}$ and $\mathrm{SNR}=10 \mathrm{~dB}$. ..... 61
3.4 Estimated locations of zeros of signal model 2 after 50 Monte Carlo runs. The number of output samples is 5000 . ..... 62
3.5 Identification results for signal model 1. The number of output samples is 2000. The graphs display the performance of system identification methods for $\mathrm{SNR}=50 \mathrm{~dB}$ and $\mathrm{SNR}=0 \mathrm{~dB}$. ..... 63
3.6 Estimated locations of zeros of signal model 1 after 50 Monte Carlo runs. The number of output samples is 2000 . ..... 64
3.7 Identification results for signal model 4. The number of output samples is 2000 . The graphs display the performance of system identification methods for $\mathrm{SNR}=50 \mathrm{~dB}$ and $\mathrm{SNR}=0 \mathrm{~dB}$ ..... 65
3.8 Estimated locations of zeros of signal model 4 after 50 Monte Carlo runs. The number of output samples is 2000 . ..... 65
3.9 Fourth-order cumulants based identification results for signal model 3. The number of output samples is 5000 . The graphs display the perfor- mance of system identification methods for $\mathrm{SNR}=50 \mathrm{~dB}$. ..... 67
3.10 Estimated locations of zeros of signal model 3 after 50 Monte Carlo runs. The number of output samples is 5000 . ..... 67
3.11 Fourth-order cumulants based identification results for signal model 1. The number of output samples is 2000 . The graphs display the perfor- mance of system identification methods for $\mathrm{SNR}=50 \mathrm{~dB}$. ..... 68
4.1 Single channel system and deconvolution filter. ..... 71
4.2 Domain of support for equation 4.20. ..... 76
4.3 Geometric interpretation of $j$-diagonal slices. ..... 79
4.4 Deconvolution results for signal model 1 after 50 Monte Carlo runs. The number of output samples is 1000 samples and the SNR is 50 dB . ..... 91
4.5 Deconvolution results for signal model 1 after 50 Monte Carlo runs. The number of output samples is 1000 samples and the SNR is 10 dB . ..... 92
4.6 Standard deviation of estimated inverse impulse response and MSE of deconvolution for signal model 1 after 50 Monte Carlo runs. The number of output samples is 1000 samples and the SNR is 50 dB . ..... 93
4.7 Standard deviation of estimated inverse impulse response and MSE of deconvolution for signal model 1 after 50 Monte Carlo runs. The number of output samples is 1000 samples and the SNR is 10 dB . ..... 93
4.8 Deconvolution results for signal model 3 after 50 Monte Carlo runs. The SNR is 10 dB and the model order is overestimated to be 4 instead of 2 . Output sample size 1000. ..... 94
4.9 Deconvolution results for signal model 2 after 50 Monte Carlo runs. The number of output samples is 5000 samples and the SNR is 50 dB . ..... 96
4.10 Deconvolution results for signal model 2 after 50 Monte Carlo runs. The number of output samples is 5000 samples and the SNR is 10 dB ..... 97
4.11 Standard deviation of estimated inverse impulse response and MSE of deconvolution for signal model 2 after 50 Monte Carlo runs. The number of output samples is 5000 samples and the SNR is 50 dB . ..... 98
4.12 Standard deviation of estimated inverse impulse response and MSE of deconvolution for signal model 2 after 50 Monte Carlo runs. The number of output samples is 5000 samples and the SNR is 10 dB . ..... 98
4.13 Deconvolution results for signal model 3 after 50 Monte Carlo runs. The SNR is 10 dB and the model order is overestimated to be 4 instead of 2. Output sample size 2000. ..... 99
4.14 Deconvolution results for signal model 2 after 50 Monte Carlo runs. The number of output samples is 7000 samples and the SNR is 50 dB . ..... 100
4.15 Deconvolution results for signal model 2 after 50 Monte Carlo runs. The number of output samples is 7000 samples and the SNR is 10 dB ..... 101
5.1 The one-dimensional cumulant slices and their corresponding vectors. ..... 105
5.2 The detailed structure of $\mathbf{C}_{\mathbf{3}, \mathbf{q}} \cdot\left(c_{\tau_{1}, \tau_{2}}\right.$ is used to denote $c_{3, x}\left(\tau_{1}, \tau_{2}\right)$ ). ..... 106
5.3 The detailed structure of $\mathbf{C}_{2, \mathbf{3}, \mathbf{q}} \cdot\left(c_{\tau_{1}, \tau_{2}}\right.$ is used to denote $c_{3, x}\left(\tau_{1}, \tau_{2}\right)$ and $c_{\tau}$ is used to denote $c_{2, x}(\tau)$ ) ..... 113
5.4 The domain of support of the 2-D cumulant slice $c_{4, x}\left(m, \tau_{1}, \tau_{2}\right), 0 \leq m \leq$ $q$. ..... 117
5.5 Identification results for signal model 1. The number of output samples is 2500 and SNR is 50 dB . The results obtained without enhancement are shown in green and those obtained after enhancement are shown in blue. 129
5.6 The first column graphs show the singular value ratio, the second column graphs show the error of the enhanced cumulants and the third column graphs show the error of system identification of the closed formula [1] in blue, and the error of the LS methods of chapter 3 in green. The horizontal axis in all graphs represents the number of iterations. ..... 130
5.7 Identification results for signal model 2. The number of output samples is 5000 and SNR is 50 dB . The results obtained without enhancement are shown in green and those obtained after enhancement are shown in blue. 131
5.8 Deconvolution results for signal model 3. The number of output samples is 1000 and the SNR is 50 dB . The results are summarised after 50 Monte Carlo runs. ..... 133
5.9 Deconvolution results for signal model 2. The number of output samples is 4000 and the SNR is 50 dB . The results are summarised after 50 Monte Carlo runs. ..... 134
5.10 Mean Square Error of deconvolution after 50 Monte Carlo runs for signal model 3. The results obtained without enhancement are shown in green and those obtained after enhancement are shown in blue. ..... 135
5.11 Mean Square Error of deconvolution after 50 Monte Carlo runs for signal model 2. The results obtained without enhancement are shown in green and those obtained after enhancement are shown in blue. ..... 135

## List of Tables

3.1 System order selection for Signal Model 1: Successful selections in 100Monte Carlo trials with SNR=10dB69
3.2 System order selection for Signal Model 2: Successful selections in 100 Monte Carlo trials with SNR=20dB ..... 69

## Chapter 1

## Introduction

### 1.1 Introduction

Data acquisition is one of the most important stages in the study of both natural and artificial processes. The data quite often are in the form of a sequence or equivalently in the form of a discrete signal. Discrete signals can be classified in two broad categories: deterministic signals whose exact form can be reproduced at will using some well defined algorithm, and random signals which are sequences of numbers obeying a certain probabilistic law. The majority of discrete signals encountered in practice, and usually the most interesting ones, are random signals ${ }^{1}$. In most cases the underlying probabilistic law is not known, or at best is only partially known. In order to be able to process a random signal in a useful way it is necessary to make some statistical assumptions regarding its underlying probabilistic law. Such assumptions which aim at explaining or describing the mechanisms underlying the generation of random signals, are called random or stochastic models. The use of models for the study of random signals was first proposed by Yule in [2].

Within this thesis the problem of estimating the parameters of a particular class of stochastic models is considered. The purpose of this chapter is to give an informal introduction to the work undertaken in this project. The chapter begins with a brief overview of stochastic modeling which helps put the contributions of this thesis in the appropriate context. Following this, there is a discussion of the principal motivations for this work. Finally the organisation of the thesis is described.

### 1.2 Stochastic Modeling

Modeling is an essential step in practical problem solving. As Mendel points out in $[3,4]$ the modeling process can be decomposed into four stages: Representation, measurement, estimation and validation. The representation stage involves identifying the

[^0]class of models which is more appropriate for the problem under consideration. For example in problems involving random signals, the representation stage involves decisions as to whether the stochastic model should be linear or nonlinear, time-varying or static; in the time domain or frequency domain etc.

The measurement stage involves deciding which physical quantities related to the problem under consideration need to be measured and how they should be measured.

The estimation stage involves using the measurements for the estimation of non-measurable physical quantities related to the problem under consideration. These physical quantities are usually parameters characterizing the underlying model.

Finally the validation stage involves the design of statistical tests involving confidence limits for assessing the success of the modeling process.

The work in this thesis deals with the problem of estimation for problems where the measurements are random signals. Before looking at the estimation methods investigated in this thesis, it is necessary to examine the statistical mechanism which is assumed to generate the observed random signals. It is interesting to note here that the more we assume about the underlying mechanism the easier the estimation stage becomes. On the other hand, this restricts the range of applications of the model. The assumptions about the generating mechanism of the signals considered in this work are summarised in the next section.

### 1.2.1 Modelling Assumptions

Measured random signals are assumed to be realisations of discrete stationary random processes i.e. processes whose probabilistic law does not change over time [5]. Discrete stationary random processes offer two main advantages: First, discrete stationary processes are quite often ergodic processes [5] i.e. a single infinite length realisation of the process is sufficient to characterise the underlying probabilistic law. Lack of ergodicity prohibits the practical application of estimation methods in problems where only a single realisation of the random process is available. Second, their theoretical foundations have been studied extensively over the past 50 years, starting with the influential work of Wiener [6] and Wold [7]. Wold's decomposition theorem states that every discrete stationary process can be decomposed to a general linear process and a predictable process with these two processes being uncorrelated to each other.

## Linear Models

Wold's decomposition theorem is one of the most important results in modern statistical signal processing. It provides the theoretical justification for the construction of linear models for the study of discrete stationary processes. In this thesis it is assumed that the random signals under consideration are realisations of stationary processes whose predictable component is zero. The idea is that the measured random signal is the output of an unknown linear system whose input although un-measurable, is known to be a random independent identically distributed (IID) sequence. Before discussing the probability distribution of the input sequence, it is useful to make some further assumptions on the structure of the linear filter. To facilitate practical implementations it is easier to assume that the linear model is described by a finite number of parameters. The most popular approach to achieve this is to restrict attention to models whose corresponding filter transfer function is a rational polynomial. This translates in the time domain to an input-output relation for the stochastic model according to which the present output value of the model is a finite linear combination of past values of the model output (i.e. feedback) and present and past values of the model input [8]. The coefficients of the linear combination are the parameters of the model. Models satisfying the general input-output relation involving past and present values of both the input and output are called Autoregressive Moving Average models (ARMA models). Quite common are the following two special cases of ARMA models:

- Autoregressive models (AR models) which involve only the present value of the model input.
- Moving Average models (MA models) which involve no past values of the model output.

This thesis investigates batch estimation methods for MA models. From the discussion so far it is obvious that the proposed modeling of random signals involves three physical quantities: the input signal, the finite set of model parameters and the output signal. Here it is assumed that from these three quantities only the output signal is measurable. The other two can only be estimated using the measured output signal.

If the primary objective is the estimation of the model parameters, then the estimation problem is referred to as system identification and has received extensive attention in system theory, time-series analysis and control theory $[9,10,11,12]$. Novel methods for system identification of MA models are presented in this thesis. These methods can also be used in ARMA parameter estimation for the estimation of the MA part of the ARMA process or even for the estimation of the AR part as will be seen in chapter 3.

On the other hand, if the primary objective is the estimation of the input sequence, then the estimation problem is referred to as blind deconvolution or simply as deconvolution.

The term blind deconvolution first appeared in Stockham et al's paper [13]. Deconvolution involves estimating a linear system whose transfer function is reciprocal to the transfer function of the system corresponding to the linear stochastic model. This is also referred to as estimating the inverse system. Algorithms for blind deconvolution are also investigated in this thesis.

## Non-Gaussian linear processes

Another important issue in the modeling approach adopted in this thesis is that of the probability distribution characteristics of the input process of the linear system. The most common assumption usually made in the statistical signal processing literature is that the input process is a white Gaussian process with zero mean and constant variance. Parameter estimation theory for Gaussian processes is well established and it is relatively easy to derive objective performance benchmarks for estimators like the Cramer-Rao Bound [5]. In addition Maximum Likelihood estimation [5, 14] is relatively straightforward for Gaussian processes. Modeling using Gaussian processes also has some theoretical justification because of the implications of the Central Limit Theorem. Interestingly enough, there is a considerable number of practical applications where non-Gaussian signals have been identified. To name a few, radar returns, seismic reflectivity sequences and passive sonar are modeled as Rayleigh, mixture of Laplace, or log-normal distribution processes respectively. Communication applications also involve non-Gaussian processes. Motivated by evidence cited above, we depart in this thesis from the Gaussian model and consider specifically non-Gaussian linear processes and estimation methods for such processes.

The departure from the Gaussian model has some important implications for the statistical tools used in this thesis. Traditionally correlation-based methods [15, 16, 11, 9] have been employed for the study of Gaussian processes. This was because zero-mean Gaussian processes can be explicitly characterised by their second-order statistics and at the same time first- and second-order statistics do not carry any information on the phase properties of the underlying linear system.

For non-Gaussian processes, first- and second-order statistics are not sufficient for their complete characterisation. For the study of non-Gaussian processes, statistics of order higher than two, have been used in statistical signal processing during the past twenty years. In most practical applications involving non-Gaussian processes, third- and fourth-order moments and cumulants have been used, as well as their Fourier transforms known as bispectra and trispectra. The utilisation of higher than second-order statistics for the study of non-Gaussian linear processes allows the recovery of both amplitude and phase information of the underlying linear system. Consequently higher-order statisticsbased methods can identify non-minimum phase systems. The earlier techniques [17, $18,19]$ which are dated before the emergence of higher-order statistical techniques,
could achieve the same result but only under restrictive assumptions (first-order MA models, limited class of input distributions). Because of their theoretical advantages in the study of non-Gaussian linear processes, higher-order statistical techniques are good candidates for application to a wide range of practical problems. For example, areas of application include estimation of source signature in seismic data processing [ $21,22,23,24]$, construction of equalisers for communication systems [25, 26, 27, 28], biomedical signal processing [29] and other areas. However, it should be pointed out, that the type of stochastic models discussed in this section do not always describe accurately the natural processes which underlye practical problems. The most common problems are violations of the assumptions of stationarity and whiteness of the driving sequence [30].

### 1.3 Estimation Using Higher-Order Statistics

Dealing with non-Gaussian processes almost always implies abandoning Maximum Likelihood estimations methods. This is because the likelihood function is usually analytically intractable in the non-Gaussian case. This is reflected in the existing higher-order statistics-based estimation methods for non-Gaussian linear processes, which are all instances of the general Method of Moments [14]. The estimation methods considered in this thesis are also based on the Method of Moments. Although the results presented here are relevant to general ARMA processes, the investigation is mainly oriented towards MA processes.

The area of HOC-based (higher-order cumulant-based) MA parameter estimation is a mature area of research with many papers produced over the last ten years [31, 32, 33, $34,35,36,37,38,39,40]$. The majority of the existing work is concerned with the development of linear algebraic methods i.e. methods involving the solution of linear systems of equations with respect to expressions of the system parameters. Despite the maturity of the subject, some areas have been identified where there is scope for further development. More specifically, there is no unified description of methods based on the inter-relations of different cumulant slices of the same or different order. Such an analysis is undertaken in this thesis. In particular, I develop new general equations which inter-relate cumulants slices of both the same and different orders. I consider two versions of these equations, one involving the system parameters and a second one involving the inverse system parameters.

From my theoretical development, I have derive new methods for system identification and deconvolution which fill the gaps of existing techniques. For example I designed the new system identification methods to deal with signals contaminated with additive coloured Gaussian noise in a more efficient manner than earlier techniques. Furthermore I develop new deconvolution methods based on fourth-order cumulants and present a
first systematic study of the identifiability of HOC-based deconvolution methods.

Finally, an interesting issue concerning the higher-order cumulants of MA processes that has not been investigated before, is that of deriving properties that characterise real cumulants of MA processes. In this thesis I derive characteristic properties of cumulants which I then translate into properties of matrices consisting of cumulants. I use these results to enhance the quality of sets of sample cumulants and to develop novel iterative methods for MA parameter estimation.

### 1.4 Thesis Organisation

After this introduction the theory of higher-order statistics is reviewed. In Chapter 2 the cumulants of stationary processes are defined, and their symmetry properties resulting from the stationarity assumption are investigated. The cumulants of linear processes are examined later and the fundamental relationship between the output cumulants and the impulse response of linear system is presented. The phase properties of linear systems with a rational transfer function are reviewed and the ability of higher-order cumulants to resolve phase ambiguity is demonstrated with examples.

In chapter 3 the problem of MA system identification is considered. General equations relating cumulant slices of the same or different orders are developed. Existing system identification methods are reviewed, and expressed within a unified framework using the aforementioned equations. New linear system identification methods are developed and both least squares and recursive solutions are proposed. Applications of the new methods to model order selection and ARMA parameter estimation are also presented.

In chapter 4 the problem of blind deconvolution of MA models is considered. General equations are developed relating MA cumulants to the inverse filter coefficients. These equations are then used to develop linear methods for blind deconvolution for MA models. The structure of the matrices involved in these methods is studied in order to demonstrate the identifiability of the deconvolution methods. Finally, expressions for the asymptotic variance of the inverse filter coefficients are derived.

Chapter 5 deals with the subject of MA cumulant enhancement. Second- third- and fourth-order cumulants of MA processes are used to build matrices for which certain properties are derived. These properties are shown to be characteristic of matrices consisting of true MA cumulants. Property mappings are developed corresponding to the cumulant matrix properties. The property mappings are then used to build iterative procedures to map sample cumulant matrices to matrices possessing the prescribed properties. This procedure can be viewed as an alternative to nonlinear cumulant matching.

Chapter 6 summarises the conclusions of this work and suggests areas for useful future work.

## Chapter 2

## Higher-Order Statistics

### 2.1 Introduction

The traditional approach to the solution of almost every signal processing and system theory problem has been the application of second-order (correlation-based) methods. The mean and the covariances (first and second order statistics respectively) are sufficient for the characterisation of the probabilistic law of stationary Gaussian processes. In real world applications though, it is quite common to encounter non-Gaussian processes. For non-Gaussian processes apart from the first and second order moments, moments of order higher than 2 provide additional information about the probabilistic law of the processes. Utilisation of high-order moments in parameter estimation problems involving non-Gaussian processes can lead to improved accuracy. During the past 15 years there has been increasing interest in this direction.

Although moments provide all the available information for higher-order analysis of a random process, it is usually preferable to work with cumulants which are quantities closely related to moments. According to [46] cumulants were first defined by Thiele in about 1889. Most of Thiele's work is in Danish and his most accessible English translation appeared in 1931 [47]. The next important contribution to the statistical theory of cumulants was due to Fisher in 1929 [48]. For the signal processing community the most influential work was that of Leonov and Shiryaev [49], Brillinger [50] and Brillinger and Rosenblatt [51]. More recently hundreds of articles have appeared on different aspects of higher-order statistical methods in signal processing.

This chapter sets out the background theory of cumulants in general and more specially the theory of cumulants within the context of linear system theory. The chapter contains the most important theoretical results required for the development that follows in the subsequent chapters. The chapter begins with the formal definition of moments and cumulants for multivariate random variables. The properties of cumulants and moments are also discussed here. The discussion is then specialised on the cumulants of random stationary processes and more particularly on the cumulants of stationary linear random processes. The relation between cumulants and the impulse response of linear systems is established in section 2.5. In section 2.6, linear systems with rational
transfer functions are considered, and the ability of higher-order statistics to characterise their correct phase properties is discussed. Finally, the chapter ends with a discussion of sample cumulant and moment estimators and their asymptotic properties.

### 2.2 Moments and Cumulants

Cumulants are the main tool used in this thesis. Because cumulants are better understood as functions of the moments, this section begins with the definition of moments of vector random variables. Let $X$ be an $n$-dimensional random vector defined as $X=\left[X_{1}, \ldots, X_{n}\right]^{\top}$. The $n$ th-order moment $\mu_{n}(X)=\mu_{n}\left(X_{1}, \ldots, X_{n}\right)$ is defined as [5]

$$
\begin{equation*}
\mu_{n}\left(X_{1}, \ldots, X_{n}\right)=E\left\{X_{1} \cdots X_{n}\right\} . \tag{2.1}
\end{equation*}
$$

The moments are related to the characteristic function of the random vector $X$ which is defined as follows [5] :

$$
\begin{equation*}
\Phi_{X}(\xi)=E\left\{\exp \left(j \xi^{\top} X\right)\right\} \tag{2.2}
\end{equation*}
$$

where $\xi$ is an $n$-dimensional complex vector defined as $\xi=\left[\xi_{1}, \ldots, \xi_{n}\right]^{\top}$. It can easily be shown that

$$
\begin{equation*}
\mu_{n}\left(X_{1}, \ldots, X_{n}\right)=\left.(-j)^{n} \frac{\partial \Phi_{X}(\xi)}{\partial \xi_{1} \cdots \partial \xi_{2}}\right|_{\xi=0} \tag{2.3}
\end{equation*}
$$

$j \xi^{\top}$ in equation (2.2) is changed to $\lambda^{\top}$, the resulting function

$$
\begin{equation*}
M_{X}(\lambda)=E\left\{\exp \left(\lambda^{\top} X\right)\right\} \tag{2.4}
\end{equation*}
$$

is the moment-generating function. The $n$ th-order moment of the random vector $X$ is equal to the coefficient of the $\prod_{i=1}^{n} \lambda_{i}$ in the Taylor expansion around zero (provided it exists) of the moment-generating function. Equation (2.3) can now be written as

$$
\begin{equation*}
\mu_{n}\left(X_{1}, \ldots, X_{n}\right)=\left.\frac{\partial M_{X}(\lambda)}{\partial \lambda_{1} \cdots \partial \lambda_{n}}\right|_{\lambda=0} . \tag{2.5}
\end{equation*}
$$

The logarithm of the characteristic function is the second characteristic function of the random vector $X$ [5]:

$$
\begin{equation*}
\Psi_{X}(\xi)=\ln \Phi_{X}(\xi)=\ln \left(E\left\{\exp \left(j \xi^{\top} X\right)\right\}\right) . \tag{2.6}
\end{equation*}
$$

$j \xi^{\top}$ in equation (2.6) is changed to $\lambda^{\top}$, the resulting function

$$
\begin{equation*}
K_{X}(\lambda)=\ln \left(E\left\{\exp \left(\lambda^{\top} X\right)\right\}\right) \tag{2.7}
\end{equation*}
$$

is the second moment generating function or cumulant generating function. In analogy with the moments being the coefficients of the. Taylor expansion of the momentgenerating function, the cumulants are defined as the coefficients of the Taylor expansion around zero ${ }^{1}$ of the cumulant-generating function. In particular, the cumulant of the random vector $X$ is defined as the coefficient of the $\prod_{i=1}^{n} \lambda_{i}$ in the Taylor expansion. The $n$ th-order cumulant of the random vector $X$ will be denoted as $\kappa_{n}(X)=\kappa_{n}\left(X_{1}, \ldots, X_{n}\right)$. The above definition of cumulants clearly implies that,

$$
\begin{equation*}
\kappa_{n}\left(X_{1}, \ldots, X_{n}\right)=\left.(-j)^{n} \frac{\partial \Psi_{X}(\xi)}{\partial \xi_{1} \cdots \partial \xi_{2}}\right|_{\xi=\mathbf{0}}=\left.\frac{\partial K_{X}(\lambda)}{\partial \lambda_{1} \cdots \partial \lambda_{n}}\right|_{\lambda=\mathbf{0}} \tag{2.8}
\end{equation*}
$$

Both moments and cumulants are invariant to permutations of the components of $X$. Assuming that the moment-generating (or the cumulant-generating) function is analytic at the origin, the infinite set of moments (or cumulants) is sufficient to determine the joint distribution uniquely $[46,52]$.

From the definition of the moments and cumulants, it is evident that they are closely related. The exact formulas giving the relationship between cumulants and moments have been derived in [49, 50, 53, 14]. Before presenting these formulas it is necessary to introduce some notation on partition sets of integers. We follow the notation of [14] in which boldface lowercase letters denote sets of integers. Let $\mathbf{n}$ be the set $\{1,2, \ldots, n\}$, and let $\mathbf{m}$ be a subset of $\mathbf{n}$. The notation $X_{\mathrm{m}}$ denotes the vector consisting of the elements of $X$ with indeces in $\mathbf{m}$. The ordering of the components of $X_{\mathrm{m}}$ is irrelevant for the following discussion. A partition of $\mathbf{n}$ is a collection of disjoint subsets of $\mathbf{n}$ whose union is $\mathbf{n}$. For example $\{\{1,3\},\{2\},\{4\}\}$ is a partition of $\{1,2,3,4\}$. A generic partition of $\mathbf{n}$ will be denoted by $P(n)$ and the set of all partitions of $\mathbf{n}$ is denoted by $\mathcal{P}(n)$. The following two theorems provide us with the explicit relationships for the moments with respect to cumulants and the reverse.

Theorem 2.1 The nth-order moment of $X$ is related to the cumulants of the subvectors of $X$ via

$$
\begin{equation*}
\mu(X)=\sum_{P(n) \in \mathcal{P}(n)} \prod_{\mathbf{m} \in P(n)} \kappa\left(X_{\mathbf{m}}\right) \tag{2.9}
\end{equation*}
$$

Theorem 2.2 The nth-order cumulant of $X$ is related to the moments of the subvectors of $X$ via

$$
\begin{equation*}
\kappa(X)=\sum_{P(n) \in \mathcal{P}(n)}(-1)^{(r-1)}(r-1)!\prod_{\mathbf{m} \in P(n)} \mu\left(X_{\mathbf{m}}\right) \tag{2.10}
\end{equation*}
$$

Detailed proofs of the above theorems can be found in [14]. It is instructive to see

[^1]special cases of the moments to cumulants Theorem (2.2) for $n=1,2,3$ and 4.
$n=1:$
\[

$$
\begin{equation*}
\kappa_{1}\left(X_{1}\right)=\mu_{1}\left(X_{1}\right)=E\left\{X_{1}\right\} . \tag{2.11}
\end{equation*}
$$

\]

$n=2:$

$$
\begin{equation*}
\kappa_{2}\left(X_{1}, X_{2}\right)=\mu_{2}\left(X_{1}, X_{2}\right)-\mu_{1}\left(X_{1}\right) \mu_{1}\left(X_{2}\right) \tag{2.12}
\end{equation*}
$$

which shows that $\kappa_{2}\left(X_{1}, X_{2}\right)$ is the covariance of $X_{1}$ and $X_{2}$.
$n=3:$

$$
\begin{aligned}
& \kappa_{3}\left(X_{1}, X_{2}, X_{3}\right)=\mu_{3}\left(X_{1}, X_{2}, X_{3}\right)-\mu_{1}\left(X_{1}\right) \mu_{2}\left(X_{2} X_{3}\right)-\mu_{1}\left(X_{1}\right) \mu_{2}\left(X_{2} X_{3}\right)- \\
& \mu_{1}\left(X_{1}\right) \mu_{2}\left(X_{2} X_{3}\right)+2 \mu_{1}\left(X_{1}\right) \mu_{1}\left(X_{2}\right) \mu_{1}\left(X_{3} \nmid 2.13\right)
\end{aligned}
$$

$n=4$ :

$$
\begin{array}{r}
\kappa_{4}\left(X_{1}, X_{2}, X_{3}, X_{4}\right)=\mu_{4}\left(X_{1}, X_{2}, X_{3}, X_{4}\right)-\mu_{2}\left(X_{1}, X_{2}\right) \mu_{2}\left(X_{3} X_{4}\right) \\
-\mu_{2}\left(X_{1}, X_{3}\right) \mu_{2}\left(X_{2} X_{4}\right)-\mu_{2}\left(X_{1}, X_{4}\right) \mu_{2}\left(X_{2} X_{3}\right) \\
-\mu_{1}\left(X_{1}\right) \mu_{3}\left(X_{2} X_{3} X_{4}\right)-\mu_{1}\left(X_{2}\right) \mu_{3}\left(X_{1} X_{3} X_{4}\right) \\
-\mu_{1}\left(X_{3}\right) \mu_{3}\left(X_{1} X_{2} X_{4}\right)-\mu_{1}\left(X_{4}\right) \mu_{3}\left(X_{1} X_{2} X_{3}\right) \\
+2 \mu_{2}\left(X_{1} X_{2}\right) \mu_{1}\left(X_{3}\right) \mu_{1}\left(X_{4}\right)+2 \mu_{2}\left(X_{1} X_{3}\right) \mu_{1}\left(X_{2}\right) \mu_{1}\left(X_{4}\right) \\
+2 \mu_{2}\left(X_{1} X_{4}\right) \mu_{1}\left(X_{2}\right) \mu_{1}\left(X_{3}\right)+2 \mu_{2}\left(X_{2} X_{4}\right) \mu_{1}\left(X_{1}\right) \mu_{1}\left(X_{3}\right) \\
+2 \mu_{2}\left(X_{3} X_{4}\right) \mu_{1}\left(X_{1}\right) \mu_{1}\left(X_{2}\right)+2 \mu_{2}\left(X_{2} X_{3}\right) \mu_{1}\left(X_{1}\right) \mu_{1}\left(X_{4}\right) \\
-6 \mu_{1}\left(X_{1}\right) \mu_{1}\left(X_{2}\right) \mu_{1}\left(X_{3}\right) \mu_{1}\left(X_{4}\right) \tag{2.14}
\end{array}
$$

The above expressions become much simpler if one assumes that the random variables $X_{1}, X_{2}, X_{3}$ and $X_{4}$ are zero-mean.

### 2.3 Properties of Moments and Cumulants

The following properties of cumulants have been reported in the literature [51, 12, 54, 46, 49]:

Property 1: Cumulants and moments are symmetric in their arguments i.e.

$$
\mu_{n}\left(\tau_{1}, \ldots, \tau_{n}\right)=\mu_{n}\left(\tau_{i_{1}}, \ldots, \tau_{i_{n}}\right), \quad \text { and } \quad \kappa_{n}\left(\tau_{1}, \ldots, \tau_{n}\right)=\kappa_{n}\left(\tau_{i_{1}}, \ldots, \tau_{i_{n}}\right),
$$

where $\left(i_{1}, \ldots, i_{n}\right)$ is a permutation of $(1, \ldots, n)$.

Property 2: Suppose $X^{1}, X^{2}, \ldots, X^{m}$ are $m$ independent vector-valued random variables where $X^{r}$ has components $X_{1}^{r}, \ldots, X_{n}^{r}$. Then

$$
\kappa_{n}\left(\sum_{r=1}^{m} X_{1}^{r}, \ldots, \sum_{r=1}^{m} X_{n}^{r}\right)=\sum_{r=1}^{m} \kappa_{n}\left(X_{1}^{r}, \ldots, X_{n}^{r}\right)
$$

The moments do not have such a property.
Property 3: Suppose that $X$ consists of two independent sub-vectors $X_{a}$ and $X_{b}$. Then $\kappa_{n}(X)=0$ while in general $\mu_{n}(X) \neq 0$.

Property 4: For constants $a_{1}, \ldots, a_{n}$ we have that

$$
\mu_{n}\left(a_{1} X_{1}, \ldots, a_{n} X_{n}\right)=\mu_{n}\left(X_{1}, \ldots, X_{n}\right) \prod_{i=1}^{n} a_{i}
$$

and

$$
\kappa_{n}\left(a_{1} X_{1}, \ldots, a_{n} X_{n}\right)=\kappa_{n}\left(X_{1}, \ldots, X_{n}\right) \prod_{i=1}^{n} a_{i}
$$

Property 5: Both moments and cumulants are additive in their arguments i.e.

$$
\begin{aligned}
& \mu_{n}\left(X_{1}+Y_{1}, Z_{2}, \ldots, Z_{n}\right)=\mu_{n}\left(X_{1}, Z_{2}, \ldots, Z_{n}\right)+\mu_{n}\left(Y_{1}, Z_{2}, \ldots, Z_{n}\right) \\
& \kappa_{n}\left(X_{1}+Y_{1}, Z_{2}, \ldots, Z_{n}\right)=\kappa_{n}\left(X_{1}, Z_{2}, \ldots, Z_{n}\right)+\kappa_{n}\left(Y_{1}, Z_{2}, \ldots, Z_{n}\right)
\end{aligned}
$$

Property 6: If $\left\{X_{i}, i=1, \ldots, k\right\}$ are jointly Gaussian random variables then all cumulants of order higher than second are zero. This follows from the fact that the characteristic function of jointly Gaussian random variables with mean vector $\mu$ and covariance matrix $\mathbf{R}$ is given by

$$
\phi(\xi)=\exp \left(j \xi \mu-\frac{1}{2} \xi^{\top} \mathbf{R} \xi\right)
$$

The first and second order cumulants are just the means and covariances. Higherorder moments are not zero in this case and this is one of the disadvantages of moments compared to cumulants.

### 2.4 Cumulants and Polyspectra

### 2.4.1 Cumulants of Stationary Random Processes

The main subject of interest in this dissertation is the study of random stationary processes using high-order cumulants. In this section the cumulants of random stationary processes are defined and we examine the properties resulting from the stationarity assumption are examined. Let $\{x(t)\}$ be a zero-mean $k^{t h}$-order strictly stationary [8]
random process. The $k^{t h}$-order cumulant of this process, denoted as $c_{k, x}\left(\tau_{1}, \tau_{2}, \ldots, \tau_{k-1}\right)$, is defined as the joint $k^{t h}$-order cumulant of the random variables $x(t), x\left(t+\tau_{1}\right), x(t+$ $\left.\tau_{1}\right), \ldots, x\left(t+\tau_{k-1}\right)$ i.e.

$$
\begin{equation*}
c_{k, x}\left(\tau_{1}, \tau_{2}, \ldots, \tau_{k-1}\right)=\kappa_{k}\left(x(t), x\left(t+\tau_{1}\right), x\left(t+\tau_{1}\right), \ldots, x\left(t+\tau_{k-1}\right)\right) \tag{2.15}
\end{equation*}
$$

The $k^{t h}$-order moment is defined in a similar way:

$$
\begin{equation*}
m_{k, x}\left(\tau_{1}, \tau_{2}, \ldots, \tau_{k-1}\right)=\mu_{k}\left(x(t), x\left(t+\tau_{1}\right), x\left(t+\tau_{1}\right), \ldots, x\left(t+\tau_{k-1}\right)\right) \tag{2.16}
\end{equation*}
$$

The second-, third- and fourth-order cumulants of zero-mean stationary random processes are of particular interest in applications. In the following the equations (2.12) to (2.14) are modified for the zero-mean stationary random process $\{x(t)\}$ :

$$
\begin{gather*}
c_{2, x}(\tau)=m_{2, x}(\tau),  \tag{2.17}\\
c_{3, x}\left(\tau_{1}, \tau_{2}\right)=m_{3, x}\left(\tau_{1}, \tau_{2}\right),  \tag{2.18}\\
c_{4, x}\left(\tau_{1}, \tau_{2}, \tau_{3}\right)=m_{4, x}\left(\tau_{1}, \tau_{2}, \tau_{3}\right)-c_{2, x}\left(\tau_{1}\right) c_{2, x}\left(\tau_{2}-\tau_{3}\right)-c_{2, x}\left(\tau_{2}\right) c_{2, x}\left(\tau_{3}-\tau_{1}\right) \\
-c_{2, x}\left(\tau_{3}\right) c_{2, x}\left(\tau_{1}-\tau_{2}\right) . \tag{2.19}
\end{gather*}
$$

$c_{2, x}(\tau)$ is just the autocorrelation of $x(t)$.
As a direct sequence of the $k^{t h}$-order stationarity of the random process $\{x(t)\}$, the $k^{t h}{ }_{-}$ order cumulant is independent of $t$ and is only a function of the $k-1$ lags $\tau_{1}, \tau_{2}, \ldots, \tau_{k-1}$. More specifically, because of the stationarity assumption the following relationship holds:

$$
\begin{equation*}
\kappa_{k}\left(x\left(t_{1}\right), x\left(t_{2}\right), \ldots, x\left(t_{k}\right)\right)=c_{k, x}\left(t_{2}-t_{1}, t_{3}-t_{1}, \ldots, t_{k}-t_{1}\right) . \tag{2.20}
\end{equation*}
$$

Let

$$
\begin{equation*}
\tau_{n-1}=t_{n}-t_{1} \quad \text { for } \quad n=2, \ldots, k \tag{2.21}
\end{equation*}
$$

Equation (2.20) becomes

$$
\begin{equation*}
\kappa_{k}\left(x\left(t_{1}\right), x\left(t_{2}\right), \ldots, x\left(t_{k}\right)\right)=c_{k, x}\left(\tau_{1}, \ldots, \tau_{i}, \ldots, \tau_{k-1}\right) \text { where } 1 \leq i \leq k-1 \tag{2.22}
\end{equation*}
$$

Substituting the $\tau$ 's from (2.21) into $c_{k, x}\left(\tau_{1}-\tau_{i}, \ldots,-\tau_{i}, \ldots, \tau_{k-1}-\tau_{i}\right)(2.21)$ yields

$$
\begin{array}{r}
c_{k, x}\left(\tau_{1}-\tau_{i}, \ldots,-\tau_{i}, \ldots, \tau_{k-1}-\tau_{i}\right)= \\
c_{k, x}\left(t_{2}-t_{i+1}, \ldots, t_{i}-t_{i+1}, t_{1}-t_{i+1}, t_{i+2}-t_{i+1}, \ldots, t_{k}-t_{i+1}\right) . \tag{2.23}
\end{array}
$$

Using equation (2.20) and Property 1 of section 2.3, equation (2.23) becomes

$$
\begin{array}{r}
c_{k, x}\left(\tau_{1}-\tau_{i}, \ldots,-\tau_{i}, \ldots, \tau_{k-1}-\tau_{i}\right)= \\
\kappa_{k}\left(x\left(t_{i+1}\right), x\left(t_{2}\right), \ldots, x\left(t_{i}\right), x\left(t_{1}\right), x\left(t_{i+2}\right), \ldots, x\left(t_{k}\right)\right)= \\
\kappa_{k}\left(x\left(t_{1}\right), x\left(t_{2}\right), \ldots, x\left(t_{k}\right)\right) . \tag{2.24}
\end{array}
$$

Combining equations (2.22) and (2.24) yields the following formula for the transformation of lags of cumulants of stationary processes:

$$
\begin{equation*}
c_{k, x}\left(\tau_{1}-\tau_{i}, \ldots,-\tau_{i}, \ldots, \tau_{k-1}-\tau_{i}\right)=c_{k, x}\left(\tau_{1}, \ldots, \tau_{i}, \ldots, \tau_{k-1}\right) \tag{2.25}
\end{equation*}
$$

where $1 \leq i \leq k-1$. The above equation and property 1 of section 2.3 , show that the cumulants of a stationary process possess many symmetries, which are very important from a computational point of view because they make the computation of cumulants manageable. They divide the set of $k^{t h}$-order cumulants of stationary process into $k$ ! regions. Knowing the cumulants in any of these sets, enables the cumulants in the other $k!-1$ sets to be calculated. The set of cumulants corresponding to the lags $\tau_{1} \geq \ldots \geq \tau_{k-1} \geq 0$ is a minimal sufficient set [14]. Suppose that we need to calculate a cumulant $c_{k, x}\left(t_{1}, \ldots, t_{k-1}\right)$ which does not belong to the minimal sufficient set. We can use the following procedure to reflect this cumulant back to a cumulant in the minimal sufficient set:

Step 1: Order the lags in decreasing order $t_{i_{1}} \geq \ldots \geq t_{i_{k-1}}$. Then according to property 1 of section 2.3, we have that $c_{k, x}\left(t_{1}, \ldots, t_{k-1}\right)=c_{k, x}\left(t_{i_{1}}, \ldots, t_{i_{k-1}}\right)$. If $t_{i_{k-1}} \geq 0$ then the cumulant $c_{k, x}\left(t_{i_{1}}, \ldots, t_{i_{k-1}}\right)$ belongs to the minimal sufficient set, else proceed to Step 2.

Step 2: Perform the lag transformation described by equation (2.25) so that $c_{k, x}\left(t_{i_{1}}, \ldots, t_{i_{k-1}}\right)=c_{k, x}\left(t_{i_{1}}-t_{i_{k-1}}, t_{i_{2}}-t_{i_{k-1}}, \ldots,-t_{i_{k-1}}\right)$. Now all the lags are positive or zero and a final ordering in decreasing order results in lags corresponding to a cumulant in the minimal sufficient set.

In figure (2.1)(a) the 6 regions of symmetry of the third-order cumulants are shown.

### 2.4.2 Polyspectra of Stationary Random Processes

It is well known that the power spectral density of a stationary process is defined as the Fourier transform of the covariance sequence (whenever the transform exists). The same notion is extended for high order cumulants [50, 51, 54]. Assume that the
cumulant sequence is absolutely summable i.e.

$$
\begin{equation*}
\sum_{\tau_{1}=-\infty}^{+\infty} \ldots \sum_{\tau_{k-1}=-\infty}^{+\infty}\left|c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)\right|<\infty \tag{2.26}
\end{equation*}
$$

Then the $k^{\text {th }}$ order polyspectrum is defined as the ( $k-1$ )-dimensional discrete-time Fourier transform of the $k^{\text {th }}$-order cumulant i.e.,

$$
\begin{equation*}
S_{k, x}\left(\omega_{1}, \ldots, \omega_{k-1}\right)=\sum_{\tau_{k-1}=-\infty}^{+\infty} c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right) \exp \left(-j \sum_{i=1}^{k-1} \omega_{i} \tau_{i}\right), \tag{2.27}
\end{equation*}
$$

where $\left|\omega_{i}\right| \leq \pi$ for $i=1,2, \ldots, k-1$ and $\left|\omega_{1}+\omega_{2}+\cdots+\omega_{k-1}\right| \leq \pi$. It is obvious from equation (2.27) that the polyspectrum is periodic with period $2 \pi$ i.e.,

$$
\begin{equation*}
S_{k, x}\left(\omega_{1}, \ldots, \omega_{k-1}\right)=S_{k, x}\left(\omega_{1}+2 \pi, \ldots, \omega_{k-1}+2 \pi\right) \tag{2.28}
\end{equation*}
$$

The cumulant sequence can then be obtained from the polyspectra using the inverse Fourier transform formula:
$c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\frac{1}{(2 \pi)^{k-1}} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} S_{k, x}\left(\omega_{1}, \ldots, \omega_{k-1}\right) \exp \left(j \sum_{i=1}^{k-1} \omega_{i} \tau_{i}\right) d \omega_{1} \cdots d \omega_{k-1}$.

When $k=3, S_{3, x}\left(\omega_{1}, \omega_{2}\right)$ is called the bispectrum and when $k=4, S_{4, x}\left(\omega_{1}, \omega_{2}, \omega_{2}\right)$ is called the trispectrum. The symmetries of the cumulants analysed in the previous paragraph imply certain symmetries in the polyspectrum. There are also certain conjugate symmetries provided that the cumulants are real, and consequently the polyspectrum possesses a richer symmetry structure than the cumulants. We now take a closer look at the symmetries of the bispectrum. In figure (2.1)(b) we see the symmetry regions of the bispectrum. Knowledge of the bispectrum in the triangular region $A O B$ is enough for a complete description of the bispectrum. This can be achieved according to the following relations:

$$
\begin{array}{r}
S_{3, x}\left(\omega_{1}, \omega_{2}\right) \in A O B=S_{3, x}\left(\omega_{2}, \omega_{1}\right) \in B O C=S_{3, x}^{*}\left(-\omega_{2}, \omega_{1}+\omega_{2}\right) \in C O D \\
=S_{3, x}^{*}\left(-\omega_{1}, \omega_{1}+\omega_{2}\right) \in D O E=S_{3, x}\left(-\omega_{1}-\omega_{2}, \omega_{1}\right) \in E O F \\
=S_{3, x}\left(-\omega_{1}-\omega_{2}, \omega_{2}\right) \in F O G=S_{3, x}^{*}\left(-\omega_{1},-\omega_{2}\right) \in G O H=S_{3, x}^{*}\left(-\omega_{2},-\omega_{1}\right) \in H O I \\
=S_{3, x}^{*}\left(\omega_{1}+\omega_{2},-\omega_{2}\right) \in I O J=S_{3, x}^{*}\left(\omega_{1}+\omega_{2},-\omega_{1}\right) \in J O K \\
=S_{3, x}\left(\omega_{1},-\omega_{1}-\omega_{2}\right) \in K O L=S_{3, x}\left(\omega_{2},-\omega_{1}-\omega_{2}\right) \in K O A
\end{array}
$$

The symmetries of the trispectrum are far more complicated. In [55] Pflug et al. point out that the trispectrum of real processes possesses 96 symmetry regions.


Figure 2.1: (a) Symmetry regions of third-order cumulants $c_{3, x}\left(\tau_{1}, \tau_{2}\right)$. (b) Symmetry regions of the bispectrum $S_{3, x}\left(\omega_{1}, \omega_{2}\right)$.

### 2.4.3 Cumulants and Polyspectra of $k^{t h}$-order White Non-Gaussian Noise

A stationary zero-mean non-Gaussian process $\{w(n)\}$ is said to be white of order $k$, if

$$
\begin{equation*}
c_{k, w}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\gamma_{k, w} \delta\left(\tau_{1}, \ldots, \tau_{k-1}\right) \tag{2.30}
\end{equation*}
$$

where $\gamma_{k, w} \neq 0$ and $\delta\left(\tau_{1}, \ldots, \tau_{k-1}\right)$ is the ( $k-1$ )-dimensional Kronecker delta function. $\gamma_{2, w}, \gamma_{3, w}$ and $\gamma_{4, w}$, are the variance, skewness and kurtosis of $\{w(n)\}$ respectively. Since we have assumed that $\{w(n)\}$ is non-Gaussian, $\gamma_{k, w}$ cannot all be zero for $k \geq 3$. Combining equation (2.27) with (2.30) we obtain

$$
\begin{equation*}
S_{k, w}\left(\omega_{1}, \ldots, \omega_{k-1}\right)=\gamma_{k, w} \tag{2.31}
\end{equation*}
$$

Equation (2.31) shows that the polyspectrum is flat for all instances of ( $\omega_{1}, \ldots, \omega_{k-1}$ ). Quite often the term IID sequence (Independent Identically Distributed) sequence appears in the HOS literature. The term IID sequence simply implies whiteness of all orders. The generation of high-order white noise sequences is very important for the numerical simulation of linear non-Gaussian random processes.

### 2.5 Cumulants and Polyspectra of Non-Gaussian Linear Processes

Let $\{w(n)\}$ be a strictly stationary processes, all moments of which exist, and are absolutely summable, i.e.

$$
\begin{equation*}
\sum_{\tau_{1}, \ldots, \tau_{k-1}=-\infty}^{\infty}\left|c_{k, w}\left(\tau_{1}, \ldots, \tau_{k-1}\right)\right|<\infty \quad \forall k \geq 2 \tag{2.32}
\end{equation*}
$$

Let $\{x(n)\}$ be a random process, related to $\{w(n)\}$ according to the convolutional equation,

$$
\begin{equation*}
x(n)=\sum_{j=-\infty}^{\infty} h(n-j) w(j) \tag{2.33}
\end{equation*}
$$

where $h(j)$ is an absolutely summable sequence

$$
\begin{equation*}
\sum_{j=-\infty}^{\infty}|h(j)|<\infty \tag{2.34}
\end{equation*}
$$

The sequence $h(j)$ can be regarded as the impulse response of a linear time invariant system. The transfer function of the linear ${ }^{2}$ system is denoted by $H(z)$. The absolute summability of $h(j)$ ensures the BIBO (Bounded Input Bounded Output) stability of the linear system in equation (2.33) and ensures the strict stationarity of the output process $\{x(n)\}$ provided the input is strictly stationary. Since stationarity has been established, the $k^{t h}$-order cumulant of $\{x(n)\}$ is a function of $k-1$ lags $\left(\tau_{1}, \ldots, \tau_{k-1}\right)$. The basic relationship between the cumulant of the output process and the impulse response parameters is given by the following theorem:

Theorem 2.3 Let $\{x(n)\}$ be a random process satisfying equation (2.33). Then its cumulants are given by

$$
\begin{equation*}
c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\sum_{j_{1}, \ldots, j_{k-1}=-\infty}^{\infty} h_{k}\left(j_{1}, \ldots, j_{k-1}\right) c_{k, w}\left(\tau_{1}-j_{1}, \ldots, \tau_{k-1}-j_{k-1}\right) \tag{2.35}
\end{equation*}
$$

where

$$
\begin{equation*}
h_{k}\left(j_{1}, \ldots, j_{k-1}\right)=\sum_{i=-\infty}^{\infty} h(i) h\left(i+j_{1}\right) \cdots h\left(i+j_{k-1}\right) . \tag{2.36}
\end{equation*}
$$

From the above theorem it is straightforward to see that the absolute summability of the impulse response implies the absolute summability of the cumulants (and moments).

[^2]Formula 2.35 is known as the Brillinger Rosenblatt formula. The absolute summability of the cumulants is a necessary condition for the existence of the polyspectra. If the Fourier transform of equation (2.35) is taken, the equivalent relationship between the input and output $k^{\text {th }}$ order polyspectra is obtained [51]:

$$
\begin{equation*}
S_{k, x}\left(\omega_{1}, \ldots, \omega_{k-1}\right)=H_{k}\left(\omega_{1}, \ldots, \omega_{k-1}\right) S_{k, w}\left(\omega_{1}, \ldots, \omega_{k-1}\right) \tag{2.37}
\end{equation*}
$$

where

$$
\begin{equation*}
H_{k}\left(\omega_{1}, \ldots \omega_{k-1}\right)=H\left(\omega_{1}\right) \cdots H\left(\omega_{k-1}\right) H^{*}\left(\sum_{i=1}^{k-1} \omega_{i}\right) \tag{2.38}
\end{equation*}
$$

with $H(\omega)$ being the Fourier transform of the impulse response

$$
H(\omega)=\sum_{i=-\infty}^{\infty} h(i) \exp \left(-j \omega_{i}\right) .
$$

If $\{w(i)\}$ is a $k^{t h}$-order white noise process, then according to the discussion in section 2.4.3, $c_{k, w}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\gamma_{k, w} \delta\left(\tau_{1}, \ldots, \tau_{k-1}\right)$ and $S_{k, w}=\gamma_{k, w}$. When $\{w(i)\}$ is an IID process, $\{x(i)\}$ is called a linear random process, and equations (2.35) and (2.37) become respectively,

$$
\begin{equation*}
c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\gamma_{k, w} h_{k}\left(\tau_{1}, \ldots, \tau_{k-1}\right) \tag{2.39}
\end{equation*}
$$

and

$$
\begin{equation*}
S_{k, x}\left(\omega_{1}, \ldots, \omega_{k-1}\right)=\gamma_{k, w} H_{k}\left(\omega_{1}, \ldots, \omega_{k-1}\right) \tag{2.40}
\end{equation*}
$$

For $k=2$ equation (2.40) reduces to the well known spectral density:

$$
\begin{equation*}
S_{2, x}=\gamma_{2, x} H(\omega) H^{*}(\omega)=\gamma_{2, x}|H(\omega)|^{2} \tag{2.41}
\end{equation*}
$$

It is clear that the power spectrum $S_{2, x}(\omega)$ (or equivalently the covariance sequence), depends only on the magnitude of $H(\omega)$ and is independent of its phase. Because of this, it is often said that second-order statistics (covariances and the corresponding spectra) are phase - blind. There is no way to recover the phase characteristics of the transfer function $H(\omega)$ using only second-order statistical information. However, if the magnitude and phase spectra of the higher-order spectra are considered (equation 2.40 ), it can be seen that for the amplitude

$$
\begin{equation*}
\left|S_{k, z}\left(\omega_{1}, \ldots, \omega_{k-1}\right)\right|=\left|H\left(\omega_{1}\right)\right| \cdots\left|H\left(\omega_{k-1}\right)\right|\left|H\left(-\sum_{i=1}^{k-1} \omega_{i}\right)\right|\left|S_{k, v}\left(\omega_{1}, \ldots, \omega_{k-1}\right)\right| \tag{2.42}
\end{equation*}
$$

and for the phase

$$
\begin{equation*}
\psi_{k, z}\left(\omega_{1}, \ldots, \omega_{k-1}\right)=\phi_{h}\left(\omega_{1}\right)+\ldots+\phi_{h}\left(\omega_{k-1}\right)-\phi_{h}\left(\sum_{i=1}^{k-1} \omega_{i}\right)+\psi_{k, v}\left(\omega_{1}, \ldots, \omega_{k-1}\right) \tag{2.43}
\end{equation*}
$$

so that phase information is preserved.

The inability of second-order statistics to preserve phase information has motivated the research into the use of higher-order statistical information for the study of linear random processes. An important theoretical result in this direction was that of Lii and Rosenblatt [31], which states that for non-Gaussian processes it is possible to recover both the magnitude and the phase from the polyspectra. More specifically they have proved the following result [31]:

Theorem 2.4 Assume that the following properties hold for the Fourier transform $H(\omega)$ of the impulse response $h(t)$ :
(i).

$$
H(\omega) \neq 0 \quad \forall \omega
$$

(ii).

$$
\sum_{i=-\infty}^{\infty}|i h(i)|<\infty
$$

Assume also that $0<\gamma_{k, w}<\infty$ for some $k>2$. Then $H(\omega)$ can be computed from $S_{k, x}\left(\omega_{1}, \ldots, \omega_{k-1}\right)$ up to an unknown complex constant scale factor $A \exp (j \omega m)$ where $A$ is real and $m$ is an integer.

The above theorem has theoretical rather than practical value, since it requires knowledge of infinite polyspectral or equivalently cumulant lags. In the next section we restrict our attention to the class of linear processes whose corresponding system impulse responses have rational transfer functions.

### 2.6 Cumulants of MA, AR and ARMA Processes

Let $H(z)$ be the $z$-transform of the impulse response $h(n)$. We examine the case where $H(z)$ is a rational polynomial. In particular we look at the three basic types of linear models namely MA (Moving Average), AR (Autoregressive) and ARMA (Autoregressive Moving Average) models.

The MA model is defined as

$$
\begin{equation*}
x(n)=\sum_{i=0}^{q} b(i) w(n-i) \quad b(0), b(q) \neq 0 \tag{2.44}
\end{equation*}
$$

where $\{w(n)\}$ is a non-Gaussian IID processes with zero mean and finite cumulants $\gamma_{i, w}$ of up to $k$-order. The corresponding transfer function has only a nontrivial numerator polynomial:

$$
\begin{equation*}
H(z)=B(z) \tag{2.45}
\end{equation*}
$$

where $B(z)$ is the $z$-transform of $b(i)$. This model is also known as an all-zero model ${ }^{3}$ [56]. The cumulants of MA processes are given as special cases of equation (2.39) for finite impulse response:

$$
\begin{equation*}
c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\gamma_{k, w} \sum_{i=0}^{q-\tau_{1}} b(i) b\left(i+\tau_{1}\right) \cdots b\left(i+\tau_{k-1}\right) \tag{2.46}
\end{equation*}
$$

where $q \geq \tau_{1} \geq \cdots \geq \tau_{k-1} \geq 0$. We observe that the cumulants of MA processes have a finite domain of support. The domain of support for the third-order cumulants of MA processes is depicted in figure 2.2. The dark area depicts the minimal sufficient domain of support corresponding to $q \geq \tau_{1} \geq \tau_{2} \geq 0$. The domain of support for fourth-order cumulants is depicted in figure 2.3. The solid area depicts the minimal sufficient domain of support corresponding to $q \geq \tau_{1} \geq \tau_{2} \geq \tau_{3} \geq 0$.

The filters in the AR and ARMA models are recursive (IIR or Infinite Impulse Response). The AR process $\{x(n)\}$ satisfies the following convolution equation ${ }^{4}$ :

$$
\begin{equation*}
\sum_{i=0}^{p} a(i) x(n-i)=w(n) \quad a(0), a(p) \neq 0 \tag{2.47}
\end{equation*}
$$

where $\{w(n)\}$ is a non-Gaussian IID processes with zero mean and finite cumulants $\gamma_{i, w}$ of up to $k$-order. In the case of AR models the transfer function has only a non-trivial denominator polynomial.

$$
\begin{equation*}
H(z)=\frac{1}{A(z)}, \tag{2.48}
\end{equation*}
$$

where $A(z)$ is the $z$-transform of $a(i)$. AR models are also known and as all-zero models because their transfer function has no poles. The computation of the cumulants of AR processes is a special case of the computation of the cumulants of ARMA processes which we now introduce. The ARMA process $\{x(n)\}$ satisfies the following difference

[^3]

Figure 2.2: (a) The finite domain of support for $3^{r d}$ order cumulants of an MA(q) process.
equation:

$$
\begin{equation*}
\sum_{i=0}^{p} a(i) x(n-i)=\sum_{i=0}^{q} b(i) w(n-1) \quad a(0), a(p), b(0), b(q) \neq 0 \tag{2.49}
\end{equation*}
$$

where $\{w(n)\}$ is a non-Gaussian IID processes with zero mean and finite cumulants $\gamma_{i, w}$ of up to $k$-order. The transfer function has both numerator and denominator polynomials which are assumed to be co-prime ${ }^{5}$ :

$$
\begin{equation*}
H(z)=\frac{B(z)}{A(z)} . \tag{2.50}
\end{equation*}
$$

The cumulants of AR and ARMA processes can be calculated using equation (2.35), which involves the impulse response of the system. It should be pointed out though, that equation (2.35) contains an infinite number of terms and consequently in practice some truncation error is inevitable. In order to overcome this problem, a new method has been developed independently by Porat and Friendlander [57] and by Swami and Mendel [58], which calculates the cumulants with respect to AR or ARMA model parameters. This method is based on a state-space realisation of ARMA models. For completeness we summarise the main results here. In the following we assume that $a(0)=b(0)=1$ with no loss in generality. For the ARMA model of equation (2.49) let

[^4]

Figure 2.3: (a) The finite domain of support for $4^{\text {th }}$ order cumulants of an MA(4) process.
$m=\max (p, q+1)$ and define

$$
\begin{gather*}
\mathbf{A}=\left[\begin{array}{cccc}
-a(1) & \cdots & -a(m-1) & -a(m) \\
1 & & 0 & 0 \\
& \ddots & & \vdots \\
0 & & 1 & 0
\end{array}\right], \mathbf{B}=\left[\begin{array}{c}
1 \\
0 \\
\vdots \\
0
\end{array}\right] \text { and } \\
\mathbf{C}=\left[\begin{array}{cccc}
1 & b(1) & \cdots & b(m-1)
\end{array}\right], \tag{2.51}
\end{gather*}
$$

where $a(m)=0$ for $m>p$ and $b(m)=0$ for $m>q$. $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ is a state-space realisation of the ARMA process (2.49) which implies that

$$
\begin{equation*}
h_{i}(n)=h\left(n+\tau_{i}\right)=\mathbf{C A}^{n+\tau_{i}} \mathbf{B}=\mathbf{C A}^{\tau_{i}} \mathbf{A}^{n} \mathbf{B} . \tag{2.52}
\end{equation*}
$$

Equation (2.52) shows that $\left\{\mathbf{A}, \mathbf{B}, \mathbf{C A}^{\tau_{i}}\right\}$ is a state-space realisation of $h_{i}(n)$. Equation (2.39) can now be rewritten as

$$
\begin{equation*}
c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\sum_{j_{1}, \ldots, j_{k-1}=0}^{\infty} \prod_{i=0}^{k-1} h_{i}(n) \text { where } \tau_{0}=0 \tag{2.53}
\end{equation*}
$$

Combining equation (2.52) and (2.53) we can obtain an expression for the cumulants with respect to the state-space matrices [57, 58]. The derivation makes use of the Kronecker product operator $\otimes$, and is presented in Appendix A. Finally we obtain the following expression :

$$
\begin{equation*}
c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\gamma_{k, x} \mathbf{C}^{\otimes(k)}\left(\mathbf{I}-A^{\otimes(k)}\right)^{-1}\left(I \otimes\left(\bigotimes_{j=0}^{k-1} \mathbf{A}^{\tau_{j}}\right)\right)\left(\mathbf{B}^{\otimes(k)}\right) \tag{2.54}
\end{equation*}
$$

### 2.6.1 Phase Properties of Systems with Rational Transfer Functions

In this section we review minimum- and maximum-phase systems since they will play an important role in the sequel.

Distribution of zeros and phase properties A minimum-phase polynomial is one that has all of its zeros strictly inside the unit circle. A maximum-phase polynomial is one that has all its zeros strictly outside the unit circle.

Minimum-phase systems A minimum-phase system is a causal linear shift-invariant system with rational transfer function

$$
H(z)=\frac{B(z)}{A(z)}
$$

where both $B(z)$ and $A(z)$ are minimum phase polynomials. Thus a minimum-phase system is a causal stable system with a causal stable inverse.

Maximum-phase Systems A maximum-phase system is one where its transfer function is represented as the ratio of two maximum-phase polynomials. Thus it is an anticausal stable system with an anticausal and stable inverse.

Mixed-phase Systems A stable and casual system is called mixed-phase system when some of its zeros are located strictly inside the unit circle and the rest strictly outside the unit circle.

Maximum- and mixed-phase systems are jointly referred as non-minimum-phase systems. We should note here that in none of the above definitions are zeros allowed on the unit circle. For every non-minimum-phase system we can find a spectrally equivalent minimum-phase system by moving zeros (or poles) which are outside the unit circle to their conjugate reciprocal locations inside the unit circle (figure 2.4). This operation can be described formally as follows: Let $H(z)$ be the transfer function of


Figure 2.4: A zero $z_{i}$ and its reciprocal conjugate $\frac{1}{z_{i}^{*}}$.
a non-minimum-phase system. Suppose that $H(z)$ has a zero $z_{0}$ which is outside the unit circle. Then the system with transfer function $H_{\text {new }}(z)$ where

$$
\begin{equation*}
H_{n e w}(z)=H(z) \frac{z^{-1}-z_{0}^{*}}{1-z_{0} z^{-1}} \tag{2.55}
\end{equation*}
$$

has the zero $z_{0}$ replaced with the $\frac{1}{z_{0}^{*}}$. With a similar transformation we can replace the poles of a transfer function. Complex zeros or poles are always transformed in pairs.

It is easy to check that the magnitude of the frequency response is invariant under the previous transformations. In fact the phase of the frequency response is altered. Suppose that the transfer function of a stable casual system has $M$ zeros from which


Figure 2.5: The autocorrelation sequence is the same for all systems
$2 M_{C}$ are complex and $M_{R}$ are real so that $M=2 M_{C}+M_{R}$. Then, given that any real zero can be replaced with its reciprocal and that any complex conjugate pair of zeros can be replaced by its reciprocal conjugate pair without any change to its magnitude frequency response, there are $2^{M_{C}+M_{R}}$ possible systems with different combinations of zeros. One of these $2^{M_{C}+M_{R}}$ systems is minimum-phase, one is maximum-phase and the rest $2^{M_{C}+M_{R}-1}$ are mixed-phase. The following important conclusions can thus be drawn:

- Any non-minimum-phase system can be represented by the cascade of minimumphase system and an all-pass system that serves to move the zeros from the inside to the outside of the unit circle.
- Since correlation-based techniques are phase-blind, all the $2^{M}$ systems look identical if analysed with correlation-based techniques.

The latter is demonstrated in the following example. Suppose that we have a MA model whose corresponding system has the following zeros: $r_{1}=0.5, \quad r_{2}, r_{3}=0.4 \pm j 0.3$ where $j=\sqrt{(-1)}$. There are four possible configurations of the three transfer function zeros: one corresponding to a minimum-phase system, two corresponding to mixedphase systems and one corresponding to a maximum-phase system. The autocorrelation function for all systems is the same and is depicted in figure (2.5). The third order

Third order cumulants

Minimum Phase System


Mixed Phase
System I


Maximum Phase
System



Figure 2.6: Third-order cumulants are different for the spectrally equivalently systems
cumulants for the different systems are given in figure (2.6). The third-order cumulants can obviously distinguish the different zero configurations.

Now we take a look at a different criterion for the characterisation of the phase properties of systems.

Partial energy criterion: Let the unit impulse response of system $H(z)$ be $\{h(k) \mid k=$ $0,1, \ldots, \infty\}$. Then the total energy of the system can be defined as (using Parseval's equality)

$$
\begin{equation*}
E_{H}=\sum_{k=0}^{\infty}|h(k)|^{2}=\int_{-\pi}^{\pi}|H(\omega)|^{2} \frac{d \omega}{2 \pi} \tag{2.56}
\end{equation*}
$$

We see that all the $2^{M}$ systems have the same energy. The difference lies in the distribution of this energy. In the minimum-phase system the energy is concentrated at the early time. So defining the partial energy as

$$
E_{H}(n)=\sum_{k=0}^{n}|h(k)|^{2}
$$

then the minimum-phase system is the one of the $2^{M}$ systems for which $E_{h}(n)$ becomes minimum.

### 2.7 Estimation of Cumulants from Finite Samples

In most practical situations we need to estimate the cumulants from a finite sample of noisy data. Suppose we are given a data sequence of length $N$

$$
\mathbf{z}_{N}=[z(1), \ldots, z(N)]^{\top}
$$

which can be made zero-mean by subtracting the mean from the data. The estimation of cumulants is based on the estimation of moments and the use of the moments to cumulants formula (2.10). In the following we assume without loss of generality that $\tau_{1} \geq \tau_{2} \geq \cdots \geq \tau_{k-1}$. The sample estimator of the $k^{t h}$-order moment $m_{k, z}\left(\tau_{1}, \ldots, \tau_{k-1}\right)$ is given by the following expression:

$$
\begin{equation*}
\hat{m}_{k, z}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\frac{1}{N} \sum_{i=1}^{N-\tau_{1}} \prod_{j=0}^{k-1} z\left(i+\tau_{j}\right) \text { where } \tau_{0}=0 \tag{2.57}
\end{equation*}
$$

If we replace the moments with their sample estimators in equations (2.17), (2.18) and (2.19), we obtain the following expressions for the sample estimators of second-, third-,
and fourth-order cumulants:

$$
\begin{align*}
\hat{c}_{2, z}\left(\tau_{1}\right)= & \frac{1}{N} \sum_{i=1}^{N-\tau_{1}} z(i) z\left(i+\tau_{1}\right)  \tag{2.58}\\
\hat{c}_{3, z}\left(\tau_{1}, \tau_{2}\right)= & \frac{1}{N} \sum_{i=1}^{N-\tau_{1}} z(i) z\left(i+\tau_{1}\right) z\left(i+\tau_{2}\right)  \tag{2.59}\\
\hat{c}_{4, z}\left(\tau_{1}, \tau_{2}, \tau_{3}\right)= & \frac{1}{N} \sum_{i=1}^{N-\tau_{1}} z(i) z\left(i+\tau_{1}\right) z\left(i+\tau_{2}\right) z\left(i+\tau_{3}\right) \\
& -\hat{c}_{2, z}(m) \hat{c}_{2, z}\left(\tau_{2}-\tau_{3}\right) \\
& -\hat{c}_{2, z}(n) \hat{c}_{2, z}\left(\tau_{3}-\tau_{1}\right) \\
& -\hat{c}_{2, z}(l) \hat{c}_{2, z}\left(\tau_{2}-\tau_{1}\right) \tag{2.60}
\end{align*}
$$

Large Sample Properties of Cumulant Estimators: The large sample (or asymptotic) properties of sample cumulant estimators are crucial for the derivation of the asymptotic properties of practical methods based on higher-order cumulants. The asymptotic properties of cumulant estimators were first investigated in [59] and later in $[60,61]$. The following theorem summarises the main result:

## Theorem 2.5 Let

$$
\begin{equation*}
y(n)=\sum_{i=-\infty}^{\infty} h(i) w(n-i) \tag{2.61}
\end{equation*}
$$

be a non-causal linear convolutional process, whose impulse response is exponentially stable. Input $w(i)$ is stationary zero mean and IID, with $\gamma_{i, w} \neq 0$ for $i>1$ and finite moments ${ }^{\mathbf{6}}$ up to order $2 k$; i.e. $m_{i, w}<\infty, i=1, \ldots, 2 k$ where $m_{i, w}=E\left\{w^{i}\right\}$. Also let $z(n)=y(n)+e(n)$, where $e(n)$ is an additive noise process, independent of $w(n)$ with finite moments up to order $2 k$, and $\gamma_{i, e}=0$ for $i \geq 3$. Finally let

$$
\begin{equation*}
s_{k, y}(i)=y(i) \prod_{j=1}^{k-1} y\left(i+\tau_{j}\right) \quad \tau_{1} \geq \tau_{2} \geq \cdots \geq \tau_{k-1} \tag{2.62}
\end{equation*}
$$

and define the $k^{t h}$-order sample moment estimator as

$$
\begin{equation*}
\hat{m}_{k, z}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\frac{1}{N} \sum_{i=1}^{N-\tau_{1}} s_{k, y}(i) . \tag{2.63}
\end{equation*}
$$

The sample cumulants estimators $\hat{c}_{k, y}\left(\tau_{1}, \ldots, \tau_{k-1}\right)$ are obtained from equation 2.10 if we replace the moments with their sample estimators 2.63. Then

$$
\begin{equation*}
\hat{c}_{k, y}\left(\tau_{1}, \ldots, \tau_{k-1}\right) \xrightarrow{\substack{N \rightarrow \infty \\ w . p .1}} c_{k, y}\left(\tau_{1}, \ldots, \tau_{k-1}\right) \tag{2.64}
\end{equation*}
$$

[^5]and
\[

$$
\begin{equation*}
\hat{c}_{k, z}\left(\tau_{1}, \ldots, \tau_{k-1}\right) \xrightarrow{\substack{N \rightarrow \infty \\ w, p .1}} c_{k, z}\left(\tau_{1}, \ldots, \tau_{k-1}\right) . \tag{2.65}
\end{equation*}
$$

\]

Expression (2.64) establishes the strong consistency of the sample cumulant estimators. Expression (2.65) establishes the strong consistency of the sample cumulant estimators even in the presence of noise (asymptotic noise insensitivity of cumulants).

In [59, 61] it is also proved that if the assumptions of theorem (2.5) are valid, then the sample cumulant estimator is asymptotically normally distributed. Since cumulants are the main statistics used in the rest of this thesis, it is essential to be able to calculate the asymptotic covariances of the sample cumulants. The derivation of such expression is due to Porat and Friendlander [57]. They first derived expressions for the asymptotic covariance of sample moments. We summarise the main result here:
Suppose we want to calculate the asymptotic covariance of the following sample moments :

$$
\begin{equation*}
\hat{m}_{n, y}\left(\tau_{1}, \ldots, \tau_{n-1}\right)=\frac{1}{N} \sum_{t=0}^{N-\max \left\{\tau_{i}\right\}} y(t) y\left(t+\tau_{1}\right) \cdots y\left(t+\tau_{n-1}\right) \tag{2.66}
\end{equation*}
$$

and

$$
\begin{equation*}
\hat{m}_{k, y}\left(\sigma_{1}, \ldots, \sigma_{k-1}\right)=\frac{1}{N} \sum_{t=0}^{N-\max \left\{\sigma_{i}\right\}} y(t) y\left(t+\sigma_{1}\right) \cdots y\left(t+\sigma_{k-1}\right) . \tag{2.67}
\end{equation*}
$$

Then the asymptotic covariance of the above sample moments is given by the following expression:

$$
\begin{array}{r}
\lim _{N \rightarrow \infty} N \operatorname{cov}\left\{\hat{m}_{n, y}\left(\tau_{1}, \ldots, \tau_{n-1}\right), \hat{m}_{k, y}\left(\sigma_{1}, \ldots, \sigma_{k-1}\right)\right\}= \\
\sum_{t=-\infty}^{\infty}\left[E\left\{y(0) y\left(\tau_{1}\right) \cdots y\left(\tau_{n-1}\right) y(t) y\left(t+\sigma_{1}\right) \cdots y\left(t+\sigma_{k-1}\right)\right\}\right. \\
\left.-m_{n, y}\left(\tau_{1}, \ldots, \tau_{n-1}\right) m_{k, y}\left(\sigma_{1}, \ldots, \sigma_{k-1}\right)\right] \tag{2.68}
\end{array}
$$

Equation (2.68), can be regarded as a generalisation of Bartlett's asymptotic formula (for more details the reader is referred to Appendix B). The asymptotic covariance of sample cumulants can be obtained from the asymptotic covariance of sample moments and the Jacobian of the moments to cumulants transformation given by equation (2.10). The calculation of the covariance of sample cumulants will be examined more thoroughly in the following chapters, where derive asymptotic expressions for the performance of the proposed parameter estimation methods will be derived.

### 2.7.1 Alternative Cumulant Estimation Methods

One of the most serious problems when working with higher-order statistics is that the variance of the estimation of the higher-order cumulants is high and, in general, very long data sequences are needed to reduce the error of estimation. In [62] alternative cumulants estimates are proposed that result in lower variance of estimation at least for symmetrically-distributed data. A different approach ${ }^{7}$ is the following scheme which is based on segmenting and averaging the data sequence and usually results in smoothed higher-order cumulant estimates:
(i). Segment the data into $K$ records of $M$ samples each $(N=K M)$.
(ii). Subtract the average value of each record from the data.
(iii). Assuming that $\left\{z^{(i)}(k), k=0,1, \ldots, M-1\right\}$ is the data set per record $\mathrm{i}=1,2, \ldots, \mathrm{~K}$, obtain estimates of the higher-order moments

$$
m_{n}^{(i)}\left(\tau_{1}, \ldots, \tau_{n-1}\right)=\frac{1}{M} \sum_{k=S_{1}}^{S_{2}} z^{(i)}\left(k+\tau_{1}\right) \cdots z^{(i)}\left(k+\tau_{n-1}\right)
$$

where $n=2,3, \ldots, N i=1,2, \ldots, K$ and

$$
\begin{gathered}
S_{1}=\max \left(0,-\tau_{1}, \ldots, \tau_{n-1}\right) \\
S_{2}=\min \left(M-1, M-1-\tau_{1}, \ldots, M-1-\tau_{n-1}\right)
\end{gathered}
$$

(iv). Average over all segments

$$
\hat{m}_{n, z}\left(\tau_{1}, \ldots, \tau_{n-1}\right)=\frac{1}{K} \sum_{i=1}^{K} m_{n}^{(i)}\left(\tau_{1}, \ldots, \tau_{n-1}\right)
$$

(v). Finally generate the $n^{\text {th }}$-order cumulant sequence $\hat{C}_{n, z}\left(\tau_{1}, \ldots, \tau_{n-1}\right)$, which as we show in section 2.1.1 is a function of moments from second to $n$th order.

In the case of very short data records, the records of length $M$ can be overlapped. If the data sequence is not long enough we can use overlapping records for the calculation of the estimates.

[^6]
### 2.8 Summary

This chapter has described the main theory of cumulants and cumulant-related issues from linear systems theory that are required for the theoretical development of the following chapters. Many significant applications of higher-order statistics in signal processing have been excluded from this review since they were not closely related to the material that follows.

In particular in this chapter we have seen how cumulants of stationary linear processes are related to the impulse response of the corresponding linear system. This relationship has been used extensively in the HOS literature to derive new parameter estimation methods. The theoretical justification for using higher-order statistics for the characterisation of both the magnitude and phase response of linear systems is provided by the theorem 2.4 of Lii and Rosenblatt. Finally we have discussed the issue of estimating cumulants and moments from a finite number of samples. The asymptotic properties of cumulant estimators are instrumental in the asymptotic study of the methods developed in the following chapters.

## Chapter 3

## Moving Average Parameter Estimation

### 3.1 Introduction

In this chapter the problem of estimating the parameters of a MA model from the cumulant statistics of the noisy observations of the system output is considered. The system is driven by an IID non-Gaussian sequence that is not observed. The noise is additive and can be coloured and even non-Gaussian under certain conditions.

The chapter presents some new general equations that relate cumulant slices of the same order as well as cumulant slices of different orders with the system parameters. It will be shown that special cases of these equations have been used in existing MA parameter estimation techniques and thus the new equations allow a unified description of some of the most important linear algebraic system identification methods. More importantly, the new equation relating cumulants of the same order with the system parameters, is used to develop new parameter estimation algorithms based on only third-order or only fourth-order cumulants. Previous techniques of the same type required the use of second order statistics in order to ensure identifiability, sacrificing in this way the advantage of HOC-based methods in the presence of additive coloured Gaussian noise. The new methods can utilise the whole set of third- or fourth-order cumulants resulting in improved performance in comparison with methods utilising only a partial set.

This chapter also presents expressions for the asymptotic variance of the estimated parameters and application to MA model order selection and ARMA parameter estimation.

The performance of the proposed methods is demonstrated and compared with that of existing techniques with the use of Monte Carlo simulations.

### 3.2 Problem Definition

Consider the single-input single-output system depicted in Figure (3.1). The output process $\{y(n)\}$ is generated according to the following linear convolutional model:

$$
\begin{equation*}
x(n)=\sum_{k=0}^{q} h(k) w(n-k), \tag{3.1}
\end{equation*}
$$

and

$$
\begin{equation*}
y(n)=x(n)+u(n) . \tag{3.2}
\end{equation*}
$$

In equation (3.1), $\{x(n)\}$ is an MA process and $H(z)$ is the transfer function of the FIR filter $h(n)$. There are no restrictions on the phase characteristics of $H(z)$ which can possibly be non-minimum phase. $\{w(n)\}$ is a zero-mean non-Gaussian stationary process whose moments of order up to eight are finite. ${ }^{1}$


Figure 3.1: Problem definition.

The process $\{u(n)\}$ in equation (3.2) is an additive noise process which is independent of $\{x(n)\}$. We assume that $\{u(n)\}$ is Gaussian process which is not necessarily white.

The problem under consideration in this chapter is that of estimating the impulse response of $H(z)$ (or equivalently the parameters of the MA process) from noisy observations of the output process $\{y(n)\}$. From a statistical perspective, since the distribution of the output data is in general unknown, it is not possible to apply Maximum Likelihood Estimation [5, 14] methods in order to obtain the unknown MA parameters. All the existing methods for the solution of the parameter estimation problem formulated in this section belong to the general framework of the Method of Moments [14].

[^7]
### 3.2.1 The Method of Moments applied to MA parameter estimation

The Method of Moments allows the description of all HOC-based methods for MA parameter estimation within the same statistical framework. In summary the Method of Moments is applied as follows: We denote with $\boldsymbol{\theta}$ the unknown MA parameters i.e.

$$
\begin{equation*}
\boldsymbol{\theta}=\{h(0), \ldots, h(q)\} \tag{3.3}
\end{equation*}
$$

Suppose that we observe a data sequence $\mathbf{y}(N)=[y(1), \ldots, y(N)]$. From the data sequence $\mathbf{y}(N)$, we calculate a finite set of statistics

$$
\begin{equation*}
\mathbf{s}_{N}=\left\{s_{N}(i), i=1, \ldots, M\right\} \tag{3.4}
\end{equation*}
$$

which in our case consists of cumulants of various orders. The number of statistics $M$ is usually much smaller than the number of observations $N$, but hopefully still convey enough information about the unknown parameters. This operation of replacing the data with a statistic of lower dimensionality, is often referred to as data reduction. We assume that the vector statistic $\mathbf{s}_{N}$ asymptotically converges ${ }^{2}$ to the true vector statistic, which is denoted as $\mathbf{s}(\boldsymbol{\theta})$ in order to make explicit its dependence on the unknown parameter vector. The final step for parameter estimation is to devise a mapping $\mathcal{G}$, which maps the vector statistic $\mathbf{s}$, to the true parameter vector $\theta$ i.e.

$$
\begin{equation*}
\boldsymbol{\theta}=\mathcal{G}(\mathbf{s}(\boldsymbol{\theta})) . \tag{3.5}
\end{equation*}
$$

The performance of different parameter estimation methods depends on the selection of the mapping $\mathcal{G}$, as well as on the amount of information contained in the selected vector statistic $\mathbf{s}(\boldsymbol{\theta})$.

According to Mendel in [63], there are three basic types of algorithms for the identification of the parameters of MA models using cumulants:

Closed-Form Solutions: Giannakis in [1] derived some very simple formulas that give the MA parameters with respect to third or fourth order cumulants. In [32] he also developed a method for the recursive calculation of the parameters of an MA model, using the autocorrelation and diagonal third-order cumulants. The same method was later reformulated by Swami and Mendel in [58]. These methods were later extended by Tugnait in [33]. The closed-form solutions are not particularly useful as practical estimation procedures, since they do not smooth out the effects of errors in the estimation of cumulants.

Linear Algebraic Solutions: This type of method involves the solution of over-

[^8]determined systems of equations which are linear with respect to quantities related to the unknown parameters. The solution of the linear system is then transformed in an appropriate way so that it results in the unknown system parameters. Examples of these methods can be found in [32, 33, 34, 35, 38, 37]. The equations involved in these methods are all based on interrelationships between cumulant slices, and will be examined within this context later in this chapter. Another interesting method which differs from the rest of linear algebraic methods, although based on linear algebraic principles is that of Fonollosa [36]. It is based on an expression of cumulants as linear combination of cumulant slices. In contrast to the other linear methods which involve Least Squares solutions of over-determined systems, the method of [36] is based on minimum norm solutions of under-determined linear systems.

Non-linear Solutions: Non-linear methods usually involve the minimisation of quadratic cumulant-matching measures [57, 64, 63]. Nonlinear methods are computationally expensive and may converge to a local minimum. Good initial conditions, usually provided by linear methods, can help reduce computational complexity and avoid convergence to a local minimum. When a nonlinear method converges to a global minimum it is generally more accurate than both closed-form and linear algebraic solutions.

In the rest of this chapter we concentrate on linear algebraic methods for parameter estimation

### 3.3 Linear Methods for MA Parameter Estimation

The starting point in all linear algebraic methods for MA parameter estimation is the Brillinger and Rosenblatt [51] formula which was introduced in chapter 2. It is repeated here for convenience:

$$
\begin{equation*}
c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\gamma_{k, w} \sum_{i=0}^{q} h(i) h\left(i+\tau_{1}\right) \cdots h\left(i+\tau_{k-1}\right) \tag{3.6}
\end{equation*}
$$

where $\tau_{j} j=1, \ldots, k-1$ are integer lags. Equation (3.6), gives the expression of cumulants of an MA process with respect to the parameters of the process, but it does not show what is the direct relation between two different order cumulants of the same MA process, or between different cumulant slices of the same order. Equation (3.6) can be used to obtain such formulas which can then be used for MA parameter estimation. In the following two sections we examine separately the derivation in the time domain of formulas involving cumulants of different orders and formulas involving different cumulant slices of the same order.

### 3.3.1 Relationships between cumulants of different order

In this section we examine the relationship between $m^{t h}$ and $n^{t h}$ order cumulants of LTI processes. Suppose that the input sequence $\{w(t)\}$ has $m^{t h}$-order cumulant given by

$$
c_{m, w}\left(\tau_{1}, \ldots, \tau_{m-1}\right)=\gamma_{m, w} \delta\left(\tau_{1}, \ldots, \tau_{m-1}\right)
$$

and $n^{t h}$-order cumulant given by

$$
c_{n, w}\left(\tau_{1}, \ldots, \tau_{n-1}\right)=\gamma_{n, w} \delta\left(\tau_{1}, \ldots, \tau_{n-1}\right)
$$

where $\gamma_{m, w}, \gamma_{n, w} \neq 0$. In the following we assume that $n>m$. The analysis that follows is in the time domain. Let

$$
\begin{equation*}
P_{m, n}=\sum_{i, j} h(i) h(j)\left[\prod_{k=1}^{m-1} h\left(i+j+\tau_{k}\right)\right]\left[\prod_{k=m}^{n-1} h\left(i+\tau_{k}\right)\right] . \tag{3.7}
\end{equation*}
$$

Changing the order of the summation in equation (3.7) we can obtain different expressions for $P_{m, n}$. If we sum first with respect to $i$ and then with respect to $j$ we have,

$$
\begin{equation*}
P_{m, n}=\sum_{j} h(j) \sum_{i} h(i)\left[\prod_{k=1}^{m-1} h\left(i+j+\tau_{k}\right)\right]\left[\prod_{k=m}^{n-1} h\left(i+\tau_{k}\right)\right] . \tag{3.8}
\end{equation*}
$$

If we multiply both sides of equation (3.8) by $\gamma_{n, w}$, and make use of the Barlett, Brillinger Rosenblatt formula (3.6) we have,

$$
\begin{equation*}
\gamma_{n, w} P_{m, n}=\sum_{j} h(j) c_{n, x}\left(j+\tau_{1}, \ldots, j+\tau_{m-1}, \tau_{m}, \ldots, \tau_{n-1}\right) . \tag{3.9}
\end{equation*}
$$

If we now sum first with respect to $j$ and then with respect to $i$ we have,

$$
\begin{equation*}
P_{m, n}=\sum_{i} h(i)\left[\prod_{k=m}^{n-1} h\left(i+\tau_{k}\right)\right] \sum_{j} h(j)\left[\prod_{k=1}^{m-1} h\left(i+j+\tau_{k}\right)\right] . \tag{3.10}
\end{equation*}
$$

If we multiply both sides of equation (3.10) by $\gamma_{m, w}$ and using (3.6) again we obtain,

$$
\begin{equation*}
\gamma_{m, w} P_{m, n}=\sum_{i} h(i)\left[\prod_{k=m}^{n-1} h\left(i+\tau_{k}\right)\right] c_{m, x}\left(i+\tau_{1}, \ldots, i+\tau_{m-1}\right) . \tag{3.11}
\end{equation*}
$$

Combining equations (3.9) and (3.11), we obtain the following relationship between the $m^{\text {th }}$ and $n^{\text {th }}$ order cumulants of MA processes:

$$
\begin{gather*}
\sum_{j} h(j) c_{n, x}\left(j+\tau_{1}, \ldots, j+\tau_{m-1}, \tau_{m}, \ldots, \tau_{n-1}\right)= \\
\epsilon_{n, m} \sum_{i} h(i)\left[\prod_{k=m}^{n-1} h\left(i+\tau_{k}\right)\right] c_{m, x}\left(i+\tau_{1}, \ldots, i+\tau_{m-1}\right), \tag{3.12}
\end{gather*}
$$

where $\epsilon_{n, m}=\frac{\gamma_{n, w}}{\gamma_{m, w}}$. If the ranges of the summations are taken from $-\infty$ to $+\infty$ then equation (3.12) is valid for general non-causal stable LTI systems. In practice equation (3.12) is useful for the identification of FIR systems only. Special forms of equation (3.12) have been used in many linear methods for MA identification which will be examined later in this chapter.

### 3.3.2 Relationships between different cumulant slices of the same order

Similarly to the development of the previous section one can obtain relations between different slices of the cumulants with the same order. Assume that the input sequence $\{w(t)\}$ has $k^{t h}$-order cumulant given by

$$
c_{k, w}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\gamma_{k, w} \delta\left(\tau_{1}, \ldots, \tau_{k-1}\right)
$$

Let

$$
\begin{equation*}
Q_{l, k}=\gamma_{k, w} \sum_{i, j} h(i) h(j)\left[\prod_{t=1}^{l} h\left(i+m_{t}\right)\right]\left[\prod_{t=1}^{l} h\left(j+n_{t}\right)\right]\left[\prod_{t=l+1}^{k} h\left(i+j+\tau_{t}\right)\right] \tag{3.13}
\end{equation*}
$$

where $2 \leq l \leq k-1$. Changing the order of summations we obtain the following two expressions for $Q_{l, k}$ :

$$
\begin{align*}
Q_{l, k} & =\sum_{i} h(i)\left[\prod_{t=1}^{l} h\left(i+m_{t}\right)\right] \gamma_{k, w} \sum_{j} h(j)\left[\prod_{t=1}^{l} h\left(j+n_{t}\right)\right]\left[\prod_{t=l+1}^{k} h\left(i+j+\tau_{t}\right)\right] \\
& =\sum_{i} h(i)\left[\prod_{t=1}^{l} h\left(i+m_{t}\right)\right] c_{k, x}\left(n_{1}, \ldots, n_{l}, i+\tau_{l+1}, \ldots, i+\tau_{k}\right) \tag{3.14}
\end{align*}
$$

and

$$
\begin{align*}
Q_{l, k} & =\sum_{j} h(j)\left[\prod_{t=1}^{l} h\left(j+n_{t}\right)\right] \gamma_{k, w} \sum_{i} h(i)\left[\prod_{t=1}^{l} h\left(i+m_{t}\right)\right]\left[\prod_{t=l+1}^{k} h\left(i+j+\tau_{t}\right)\right] \\
& =\sum_{j} h(j)\left[\prod_{t=1}^{l} h\left(j+n_{t}\right)\right] c_{k, x}\left(m_{1}, \ldots, m_{l}, i+\tau_{l+1}, \ldots, i+\tau_{k}\right) \tag{3.15}
\end{align*}
$$

From equations (3.14) and (3.15) we obtain,

$$
\begin{align*}
& \sum_{i} h(i)\left[\prod_{t=1}^{l} h\left(i+m_{t}\right)\right] c_{k, x}\left(n_{1}, \ldots, n_{l}, i+\tau_{l+1}, \ldots, i+\tau_{k}\right)= \\
& \sum_{j} h(j)\left[\prod_{t=1}^{l} h\left(j+n_{t}\right)\right] c_{k, x}\left(m_{1}, \ldots, m_{l}, i+\tau_{l+1}, \ldots, i+\tau_{k}\right) . \tag{3.16}
\end{align*}
$$

If the lags are chosen so that $\left\{n_{1}, \ldots, n_{l}\right\} \neq\left\{m_{1}, \ldots, m_{l}\right\}$ then the equality in (3.16) is non-trivial. Equation (3.16) has not been used directly for parameter estimation. As we will see in the next section, special cases of equation (3.16) can be used to construct linear methods for the estimation of parameters of MA models.

### 3.4 A Unified Framework for the Description of Linear Methods for MA identification

In this section we briefly review the most important linear algebraic methods for MA parameter estimation. These methods will be examined in the context of the theory developed in the previous sections. The equations involved can be derived as special cases of either equation (3.12) which relates cumulants of different orders, or equation (3.16) which relates cumulants of the same order.

Giannakis - Mendel 1989 [32]: This is one of the earliest methods for MA parameter estimation. It is based on special instances of equation (3.12) for $2^{\text {nd }}$ and $3^{\text {rd }}$ order cumulants. Equation (3.12) for $n=3, m=2$ and $\tau_{1}=-\tau$ and $\tau_{2}=0$ becomes

$$
\begin{equation*}
\sum_{j=0}^{q} \epsilon_{3,2} h(j) c_{3,2}(\tau-j, \tau-j)=\sum_{j=0}^{q} h^{2}(j) c_{2, x}(\tau-j) \tag{3.17}
\end{equation*}
$$

Equation (3.17) holds for the noise free case. For $-q \leq \tau \leq 2 q$ we can construct a system of $(3 q+1)$ equations with $(2 q+1)$ unknowns. The unknowns in this case are $\epsilon_{3,2}, \epsilon_{3,2} h(1), \ldots, \epsilon_{3,2} h(q), h^{2}(1), \ldots, h^{2}(q)$. This method does not warrant the consistency of the obtained estimates since there exist values of the $h(k)$ 's such that the matrix of the linear system is not full rank.

Tugnait 1990 [33]: The approach of Giannakis and Mendel described in the previous paragraph has been modified by Tugnait in [33]. The equations (3.17) are augmented by special cases of equation (3.16) for third-order cumulants. The new equations can be obtained from (3.16) for $k=3, l=1, m_{1}=0, n_{1}=q$ and $\tau_{1}=\tau$ :

$$
\begin{equation*}
\left.h(q) c_{3, x}(-\tau, 0)-\sum_{j=1}^{q} h^{2}(j) c\right) 3, x(j-\tau, q)=c_{3, x}(-\tau, q) \tag{3.18}
\end{equation*}
$$

Taking equation (3.17) for $-q \leq \tau \leq-1$ and equation (3.18) for $q+1 \leq \tau \leq 2 q$, we obtain a system of $4 q$ equations with the following $2 q+2$ unknowns:

$$
\epsilon_{3,2}, h(q), \epsilon_{3,2} h(1), \ldots, \epsilon_{3,2} h(q), h^{2}(1), \ldots, h^{2}(q)
$$

The above set of equations is still valid in the presence of IID noise.
Tugnait 1991 [34]: In order to avoid problems of numerical ill-conditioning, Tugnait in 1991, reformulated the algorithm of the previous paragraph as follows:
Multiply both sides of equation (3.17) with $\epsilon^{\prime}=1 / \epsilon_{3,2}$ to obtain

$$
\begin{equation*}
\sum_{j=0}^{q} h(j) c_{3, x}(\tau-j, \tau-j)=\sum_{j=0}^{q} \epsilon^{\prime} h^{2}(j) c_{2, x}(\tau-j) \tag{3.19}
\end{equation*}
$$

Using $c_{3, x}(-\tau, 0)=\epsilon^{\prime} h(q) c_{2, x}(\tau)$ in equation (3.18) we obtain

$$
\begin{equation*}
\left.\epsilon^{\prime} h(q) c_{2, x}(\tau)-\sum_{j=1}^{q} h^{2}(j) c\right) 3, x(j-\tau, q)=c_{3, x}(-\tau, q) \tag{3.20}
\end{equation*}
$$

Taking equation (3.19) for $-q \leq \tau \leq 2 q$ and equation (3.20) for $-q \leq \tau \leq q$ we obtain $5 q+2$ equations of the $2 q+2$ unknowns,

$$
\epsilon^{\prime}, \epsilon^{\prime} h(q), h(1), \ldots, h(q), \epsilon^{\prime} h^{2}(1), \ldots, \epsilon^{\prime} h^{2}(q)
$$

Alshebeili, Venetsanopoulos, Cetin 1993 [35]: In [35] a new linear algebraic method is developed which is based on relations between second and third-order cumulants. For $n=3, m=2, \tau_{1}=-t_{1}$ and $\tau_{2}=t_{2}-t_{1}$, equation (3.12) becomes

$$
\begin{equation*}
\sum_{i=0}^{q} h(i) c_{3, x}\left(t_{1}-i, t_{2}-i\right)=\sum_{i=0}^{q} \epsilon_{2,3} h(i) h\left(t_{2}-t_{1}+i\right) c_{2, x}\left(t_{1}-i\right) . \tag{3.21}
\end{equation*}
$$

Taking the set of equations for $t_{1}, t_{2}$ in the region shown in figure (1), we obtain a linear system of $\left(5 q^{2}+4 q+1\right)$ equations with respect to the following $\left(q^{2}+5 q+2\right) / 2$ unknowns:

$$
h(1), \ldots, h(q), \epsilon_{2,3}, \epsilon_{2,3} h(1), \ldots, \epsilon_{2,3}, h(q), \epsilon_{2,3} h^{2}(1), \ldots, \epsilon_{2,3} h(1) h(q), \epsilon_{2,3} h^{2}(q)
$$

The solution of the linear system is used to form a matrix which is known to have rank 1. In practice due to estimation errors the rank of this matrix is larger than 1 , and so it is reduced to 1 using SVD-based rank reduction. The rank reduction method was used in [35] in order to overcome the over-parameterisation of the unknown vector.

## Robustness to additive noise

In practical situations, the received signal is usually a noise-corrupted version of the original one. The signal model is then expressed as

$$
y(n)=x(n)+u(n)
$$

For Gaussian processes only, cumulants of order greater than two are identically zero. As we have seen in chapter 2 , under the assumption that the additive noise $u(n)$ is Gaussian and independent of the signal $x(n)$ then the third-order cumulants of $y(n)$ are equal to the third-order cumulants of $x(n)$.

$$
c_{3, x}\left(\tau_{1}, \tau_{2}\right)=c_{3, y}\left(\tau_{1}, \tau_{2}\right)
$$

In practice though, the variance of $\hat{c}_{3, y}\left(\tau_{1}, \tau_{2}\right)$ is larger than the variance of $\hat{c}_{3, x}\left(\tau_{1}, \tau_{2}\right)$. On the other hand, second order cumulants are affected by presence of noise because ,

$$
c_{2, y}(\tau)=c_{2, x}(\tau)+c_{2, u}(\tau)
$$

The methods reviewed on the previous paragraphs all depend on equations involving second order cumulants with lags from $-q$ to $q$. Excluding equations ${ }^{3}$ containing $c_{2, y}(0)$ , can make them robust to additive white noise. The method of [35] can be modified to deal with additive noise whose second-order cumulants are non-zero only for lags in the range $|\tau|<\bar{q}$ where,

$$
\bar{q}= \begin{cases}(q / 2)-1 & \text { if } q \text { is even } \\ (q-1) / 2 & \text { if } q \text { is odd }\end{cases}
$$

In general, the need to make assumptions about the second order cumulants of the additive noise limits the range of practical applications of system identification methods, like the ones described previously in this section. The way to overcome this limitation is to derive linear algebraic methods which rely only on third- (or fourth-) order cumulants. Until recently no such linear methods were known. The first linear methods for FIR identification using only third-order cumulants appeared in [37, 38, 65]. Extensions of the method described in [65] will be the subject of the next sections. First we take a brief look at the methods of $[37,38]$ which have many similarities.

In [38], the following equation is developed, based on relationships between the $c_{3}(q, \cdot)$ and $c_{3}(0, \cdot)$ cumulant slices:

$$
\begin{equation*}
\sum_{i=0}^{q} h^{2}(i) c_{3}(q, i+\tau)=c_{3}(\tau, 0) c_{3}(q, q) / c_{3}(q, 0) \tag{3.22}
\end{equation*}
$$

[^9]For $-q \leq \tau \leq q$ we obtain a system of $2 q+1$ equations with respect to $q+1$ unknowns.
In [37], the following equation is developed, again based on relationships between the $c_{3}(q, \cdot)$ and $c_{3}(0, \cdot)$ cumulant slices:

$$
\begin{equation*}
\sum_{i=1}^{q} \frac{h(i)}{h(0)} c_{3}^{2}(q, k+i)-\gamma_{3} h(0) h^{2}(q) c_{3}(k, k)=-c_{3}^{2}(q, k) \tag{3.23}
\end{equation*}
$$

For $-q \leq k \leq q$ we obtain once more, a system of $2 q+1$ equations with respect to $q+1$ unknowns.

Both methods described above utilise only two of the cumulant slices and ignore the rest of the statistics which may contain useful information. They also differ from other linear algebraic methods for system identification since they are not linear with respect to cumulants. These terms involve products of cumulants which in most cases have higher variance than cumulants themselves. In the next section new methods for the estimation of the parameters of MA models are developed.

### 3.5 MA Parameter Estimation Using Only Third-Order Cumulants

From the discussion in the previous section, it is obvious that the existing HOC-based linear algebraic methods for MA parameter estimation do not account for the effects of additive Gaussian noise in an efficient manner. Methods involving correlations are affected by the bias of the correlation statistics in the presence of additive Gaussian noise, while the two methods which are based only on third-order cumulants are relatively primitive because they ignore relevant information contained in the unused cumulant slices, and because they are not linear with respect to the cumulants. In this section, a new method is presented which attempts to overcome some of the these problems.

### 3.5.1 Relationships between third-order cumulant slices

Consider equation (3.16), for $k=3, m_{1}=m, n_{1}=n$ and $\tau_{1}=\tau$ :

$$
\begin{equation*}
\sum_{i}^{q} h(i) h(i+n) c_{3, x}(m, i+\tau)=\sum_{i}^{q} h(i) h(i+m) c_{3, x}(n, i+\tau) . \tag{3.24}
\end{equation*}
$$

Equation (3.24) was first derived by Tugnait (equation (18) in [33]) but it has never been used directly for parameter estimation. Equation (3.24) depends on the three parameters ( $m, n, \tau$ ). It is important to examine the ranges of these parameters, where equation (3.24) is nontrivial. The following restrictions apply to the values of $m$ and $n$.

- $-q \leq m, n \leq q$. (Otherwise equation (3.24) becomes an identity $0=0$ ).
- $m>n$ because the equation parameterized by ( $m, n, \tau$ ), is identical to the equation parameterised by ( $n, m, \tau$ ).
- $n \geq 0$ because the equation parameterised by ( $m, n, \tau$ ), is identical to the equation parameterised by $(m,-n, \tau+n)$. This is obvious, if we observe that for $0 \leq n \leq$ $m \leq q$ and for a suitable $\tau$,

$$
\sum_{i=0}^{q} h(i) h(i+m) c_{3, x}(-n, i+\tau)=\sum_{i=0}^{q} h(i) h(i+m) c_{3, x}(n, i+\tau+n) .
$$

Finally from the previous points, it is clear that all the nontrivial equations in (3.24) can be described by the parameters $(m, n, \tau)$ for $0 \leq n<m \leq q$. Given $m$ and $n$, we want to find the range of values of $\tau$ such that the cumulants $c(m, i+\tau)$ are not all zero for $i=0, \ldots, q-n$ and the cumulants $c(n, i+\tau)$ are not all zero for $i=0, \ldots, q-m$ respectively. In both cases the range of $\tau$ is the same, depends on both $m$ and $n$ and is given by $m+n-2 q \leq \tau \leq q$. We define the set containing all the possible triplets ( $m, n, \tau$ ) as follows:

$$
\begin{equation*}
\mathcal{T}=\{(m, n, \tau): 0 \leq n<m \leq q, \quad m+n-2 q \leq \tau \leq q\} \tag{3.25}
\end{equation*}
$$

The set $\mathcal{T}$ contains $\sum_{n=0}^{q-1} \sum_{m=n+1}^{q}(3 q-m-n+1)=\frac{1}{2} q(q+1)(2 q+1)$ elements. In the parameter identification scenario as it was described in section 3.2, our theoretical objective is to obtain expressions for the $h(i)$ 's. In the next section we examine how it is possible to derive such expressions from equation (3.24).

### 3.5.2 Least-Squares Method using only Third-Order Cumulants

In this section a least squares approach is presented to the solution of equations (3.24). In the following we assume without loss of generality that $h(0)=1$. Equation (3.24) is treated as a system of linear equations with respect to the unknowns $h(1), \ldots, h(q), h^{2}(1), \ldots, h^{2}(q)$, and $h(i) h(i+l)$ for $1 \leq i \leq q-1$ and $1 \leq l \leq q-i$ and the number of unknowns is $q(q+3) / 2$. At this stage, one can consider a minimum of $M_{\text {min }}=q(q+3) / 2^{4}$ equations and a maximum of $M_{\max }=\frac{1}{2} q(q+1)(2 q+1)$ equations can be used to form a linear system of equations with respect to the $q(q+3) / 2$ unknowns. The equations can have the following two forms depending on the value of parameter n :

$$
\begin{equation*}
\sum_{k=0}^{q} h(k) h(k+m) c_{3, x}(k+\tau, n)-\sum_{k=0}^{q} h(k) h(k+n) c_{3, x}(k+\tau, m)=0, \quad \text { for } n \neq 0 \tag{3.26}
\end{equation*}
$$

[^10]and
\[

$$
\begin{array}{r}
\sum_{k=0}^{q} h(k) h(k+m) c_{3, x}(k+\tau, n)-\sum_{k=1}^{q} h^{2}(k) c_{3, x}(k+\tau, m)= \\
c_{3, x}(\tau, m), \text { for } n=0 \tag{3.27}
\end{array}
$$
\]

which can be expressed in a matrix form as follows:

$$
\begin{equation*}
\mathrm{Bg}=\mathrm{d} \tag{3.28}
\end{equation*}
$$

Where $\mathrm{g}=\left[h(1), \ldots, h(q), h^{2}(1), \ldots, h^{2}(q), h(1) h(2), \ldots, h(1) h(q), \ldots \ldots, h(q-1) h(q)\right]^{\top}$ is a $\frac{q^{2}+3 q}{2}$ element vector, $\mathbf{d}$ is a $q^{2}(q+1)$ element vector, and $\mathbf{B}$ is a $\left(M \times \frac{q^{2}+3 q}{2}\right)$ matrix. The contents of $\mathbf{d}$ and $\mathbf{B}$ are determined according to (3.26) and (3.27). The least squares solution of this over-determined system of linear equations is

$$
\begin{equation*}
\mathbf{g}=\left(\mathbf{B}^{\top} \mathbf{B}\right)^{-1} \mathbf{B}^{\top} \mathbf{d} \tag{3.29}
\end{equation*}
$$

In the selection of equations for the construction of the linear system, it is advisable to try to avoid some of the equations, which make use of cumulants with relatively "large" ${ }^{5}$ lags. The over-determined system of equation (3.28) can also be solved using the Total Least Squares (TLS) method [66]. The TLS method assumes that there are estimation errors in the elements of both $\mathbf{B}$ and $\mathbf{d}$. The $\mathbf{B}$ and $\mathbf{d}$ are then modified so that the rank of the extended matrix $[\mathbf{B} \mid \mathbf{d}]$ equals the rank of $\mathbf{B}$.

### 3.5.3 Compensating for the over-parameterisation of the system of equations

It is obvious from the construction of the vector of unknowns $\mathbf{g}$, that its elements are not independent of each other. The Least Squares solution on the other hand, assumes that the unknowns, i.e. the elements of vector $g$ are independent and so the resulting solution is sub-optimum in this respect. In fact, due to errors in the estimation of the cumulants, the elements of the solution vector obtained from (3.29) will not comply with the theoretical structure of $\mathbf{g}$. In order to compensate for this, and to exploit all the available information hidden in $\mathbf{g}$, two alternative solutions are proposed in the next two sections.

[^11]
## An SVD-based rank reduction solution

We can form the following matrix $\mathbf{R}$, in a similar fashion to [35]:

$$
\mathbf{R}=\left[\begin{array}{cccc}
1 & h(1) & \cdots & h(q)  \tag{3.30}\\
h(1) & h^{2}(1) & \cdots & h(1) h(q) \\
h(2) & h(2) h(1) & \cdots & h(2) h(q) \\
\vdots & \vdots & \ddots & \vdots \\
h(q-1) & h(q-1) h(1) & \cdots & h(q-1) h(q) \\
h(q) & h(q) h(1) & \cdots & h^{2}(q)
\end{array}\right] .
$$

It is clear from the structure of $\mathbf{R}$ that its rank is one. $\mathbf{R}$ may then be written in the following form:

$$
\mathbf{R}=\mathbf{h h}^{\top}=\left[\begin{array}{c}
1  \tag{3.31}\\
h(1) \\
\vdots \\
h(q)
\end{array}\right]\left[\begin{array}{llll}
1 & h(1) & \cdots & h(q)
\end{array}\right]
$$

In practice however, again due to estimation errors, its rank will be greater than 1. Now if the SVD of $\mathbf{R}$ is

$$
\mathbf{R}=\sum_{k=1}^{q+1} \sigma_{k} \mathbf{u}_{k} \mathbf{v}_{k}^{\top}
$$

where $\sigma_{1}>\sigma_{2} \geq \cdots \geq \sigma_{q} \geq \sigma_{q+1} \geq 0$. The system parameters can be found as

$$
\begin{equation*}
h(n-1)=\sigma_{1} u_{1,1} v_{n, 1}, \tag{3.32}
\end{equation*}
$$

where $u_{1,1}$ is the first element of $\mathbf{u}_{1}$ and $v_{n, 1}$ is the $n^{\text {th }}$ element of $\mathbf{v}_{1}$. In many practical situations, it is very useful to have a measure of "confidence" for the obtained least squares solution. From the previous discussion we can see that such a measure can be devised as follows:

$$
\begin{equation*}
\lambda=\frac{\sigma_{1}}{\sum_{i=1}^{q+1} \sigma_{i}} \tag{3.33}
\end{equation*}
$$

where $0<\lambda \leq 1$. The nearer $\lambda$ is to 1 , the more confident we can be of our solution.

## A second stage LS solution

Given that in practical situations the real cumulants are not known, sample estimates of the cumulants are used in equation 3.28. After solving equation 3.28 we obtain the
following $\frac{q^{2}+3 q}{2}$-element vector:
$\hat{\mathrm{g}}=\left[\widehat{h(1)}, \ldots, \widehat{h(q)}, \widehat{h^{2}(1)}, \ldots, \widehat{h^{2}(q)}, \widehat{h(1) h(2)}, \ldots, \widehat{h(1) h(q)}, \ldots \ldots, h(\underline{(q-1) h}(q)]^{\top}(3\right.$
Consequently, the result is an estimate $\hat{\mathbf{g}}$ of the vector $\mathbf{g}$. The rank reduction method presented in the previous section, removes the redundancy present in vector $\hat{g}$ and maps this vector to a $q$-element vector that is supposed to represent the true system parameters. Here, the same objective is achieved by forming and solving a system of linear equations. Suppose we want to obtain the system parameters which we denote as

$$
\begin{equation*}
\mathbf{h}=[h(1), \ldots, h(q)]^{\top} . \tag{3.35}
\end{equation*}
$$

The system consists of the following equations:

$$
\begin{equation*}
h(i) \widehat{h(j)}=\widehat{h(i) h(j)} \text { for } i, j=1, \ldots, q \text { and } h(i)=\widehat{h(i)} \text { for } i=1, \ldots, q \tag{3.36}
\end{equation*}
$$

In total we have $q^{2}+q$ equations with $q$ unknowns. In matrix form we have:

$$
\begin{equation*}
\mathbf{H} \cdot \mathbf{h}=\mathrm{g}_{\mathrm{e}} \tag{3.37}
\end{equation*}
$$

where,
$\mathbf{H}=\left[\begin{array}{c}\mathbf{H}_{\mathbf{1}} \\ \mathbf{H}_{\mathbf{2}} \\ \vdots \\ \mathbf{H}_{\mathbf{q}} \\ \mathbf{H}_{\mathbf{0}}\end{array}\right]$, where $\mathbf{H}_{\mathbf{i}}=\left[\begin{array}{cccc}\widehat{h(i)} & 0 & \ldots & 0 \\ 0 & \widehat{h(i)} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & \widehat{h(i)}\end{array}\right] \quad$ for $i=0, \ldots, q \quad(\widehat{h(0)}=1)$,
and
$\mathbf{g}_{\mathrm{e}}=\left[\widehat{h^{2}(1)}, \widehat{h(1) h(2)}, \ldots, h \widehat{h(1) h(q)}, \widehat{h(2) h(1)}, \ldots, \widehat{h(2) h(q)}, \ldots \ldots, \widehat{h^{2}(q)}, \widehat{h(1)}, \ldots, \widehat{h(q)}\right]$

### 3.5.4 A Closed-Form Solution and Uniqueness of the Least Squares Method

In order to claim the uniqueness of the least squares solution of the system of linear equations (3.28), it is necessary to ensure that the collected equations result in a fullrank matrix $\mathbf{B}$. In the following section, a closed-form recursive method is developed which will be used later to claim the uniqueness of the least squares solution.

A Recursive Algorithm In this section it is shown that the unknown parameters $h(1), \ldots, h(q), h^{2}(1), \ldots, h^{2}(q)$, and $h(i) h(i+l)$ for $1 \leq i \leq q-1$ and $1 \leq l \leq q-i$ can
be determined from (3.24), using a closed-form recursive algorithm. Taking (3.24) for $m=q, n=0$ and $\tau=q$ we have,

$$
\begin{equation*}
h(q)=\frac{c_{3, x}(q, q)}{c_{3, x}(0, q)} . \tag{3.38}
\end{equation*}
$$

Again for $m=q$, if we take $\tau=q-j$ for $j=1, \ldots, q$ we have the following recursive equation:

$$
\begin{equation*}
h^{2}(j)=\frac{h(q) c_{3, x}(q-j, 0)-\sum_{k=0}^{j-1} h^{2}(k) c_{3, x}(k+q-j, q)}{c_{3}(q, q)} . \tag{3.39}
\end{equation*}
$$

For $m<q$ and $\tau=m-2 q$ we also have,

$$
\begin{equation*}
h(q-m) h(q)=\frac{h^{2}(q) c_{3, x}(q-m, q)}{c_{3, x}(q, q)} . \tag{3.40}
\end{equation*}
$$

Finally for $m<q$ and $\tau=m+j-2 q, j=1, \ldots, q-m$,

$$
\begin{align*}
& h(q-j-m) h(q-j)=\left[\sum_{k=q-j}^{q} h^{2}(k) c_{3, x}(k-2 q-m-j, m)\right. \\
& \left.-\sum_{k=q-m-j+1}^{q-m} h(k) h(k+m) c_{3, x}(k-2 q+m+j, 0)\right] / c_{3, x}(q, q) . \tag{3.41}
\end{align*}
$$

All the divisions are well conditioned since $c_{3, x}(q, q) \neq 0$. The above recursive algorithm, shows that using only a subset of equations from (3.24) and assuming that $h(0)=1$, we can uniquely recover the unknown parameters $h(1), \ldots, h(q), h^{2}(1), \ldots, h^{2}(q)$, and $h(i) h(i+l)$ for $1 \leq i \leq q-1$ and $1 \leq l \leq q-i$.

We collect the triplets ( $m, n, \tau$ ) corresponding to the equations used in the previous recursive method, in the following set:

$$
\left.\begin{array}{rl}
\mathcal{R} & =\{(q, 0, q-j): 0 \leq j \leq q\} \\
\{(m, 0, m+j-2 q): & 0 \tag{3.42}
\end{array}\right)
$$

The set $\mathcal{R}$ contains $q(q+3) / 2$ elements.
It is worth noting that there is one more set of triplets $(m, n, \tau)$, which can be used to construct a different recursive solution. In this second set, the value of $m$ is fixed to $q$.

Suppose we construct a system of equations, corresponding to the triplets ( $m, n, \tau) \in \mathcal{M}$ where $\mathcal{M}$ is a superset of $\mathcal{R}$. The resulting system of equations in matrix form, is written as follows:

$$
\begin{equation*}
\mathbf{B}_{\mathcal{M}} \mathbf{g}=\mathbf{d}_{\mathcal{M}} . \tag{3.43}
\end{equation*}
$$

The following theorem ensures the uniqueness of the Least Squares solution of equation

Theorem 3.1 For the signal model (3.1) and assuming that we know the model order $q$, and that we are given the correct third-order cumulant statistics $c_{3, y}\left(\tau_{1}, \tau_{2}\right)$ for all $\tau_{1}$ and $\tau_{2}$ in the principal domain of support, the matrix $\mathbf{B}_{\mathcal{M}}$ in (3.43), is of full rank;

$$
\operatorname{rank}\left(\mathbf{B}_{\mathcal{M}}\right)=\frac{q^{2}+3 q}{2}
$$

Proof: The theorem can be proved by contradiction. Suppose that $\operatorname{rank}\left(\mathbf{B}_{\mathcal{M}}\right)<$ $\frac{q^{2}+3 q}{2}$. Then there are more than one different solutions to the system of equations (3.43). Since the equations corresponding to the triplets $(m, n, \tau) \in \mathcal{R}$ are included in the system of equations, all the solutions should satisfy equations (3.38) to (3.41); hence all the solutions must be identical. This a contradiction hence the rank of $\mathbf{B}_{\mathcal{M}}$ is $\frac{q^{2}+3 q}{2}$.

In practice, only a sample sequence of the noisy data is known. Then the true cumulants are replaced by their sample averages. Since it is known [59] that the sample cumulants converge (with probability one) to the true cumulants as the sample size goes to infinity, the new parameter estimation method is asymptotically consistent. It should be noted here that if the correct model order is not known, the solution of (3.43) gives totally erroneous results. The problem of model order selection will be considered later in this chapter.

### 3.6 MA Parameter Estimation Using Only Fourth-Order Cumulants

As we have seen in Chapter 2, third-order cumulants of non-skewed processes are identically zero. In many practical situations, when the signals under consideration are non-skewed, third-order cumulant-based methods are inappropriate. In such cases, one has to employ fourth-order cumulant based methods.

The linear methods reviewed in section 3.4, can be modified so that the third-order cumulant statistics are replaced by fourth-order cumulant statistics. Once again these existing linear methods have some disadvantages. Some methods require the use of correlations as well as fourth-order cumulants and so they are not suitable in applications where there is additive non-skewed coloured noise. The methods of [37, 38] can be extended to fourth order cumulants, but once again they ignore most of the information carried by the fourth order cumulants.

In the next section, we consider methods based on inter-relations between different
fourth order cumulant slices.

### 3.6.1 Relationships between fourth order cumulants

If we consider equation (3.16) for $k=4$, then we have two alternatives for the parameter $l$ :

Equation 1: For $k=4$ and $l=2$ equation (3.16) becomes,

$$
\begin{gather*}
\sum_{i=0}^{q} h(i) h\left(i+m_{1}\right) c_{4, x}\left(n_{1}, i+\tau_{1}, i+\tau_{2}\right)= \\
\sum_{i=0}^{q} h(i) h\left(i+n_{1}\right) c_{4, x}\left(m_{1}, i+\tau_{1}, i+\tau_{2}\right) . \tag{3.44}
\end{gather*}
$$

Equation 2: For $k=4$ and $l=3$ equation (3.16) becomes,

$$
\begin{gather*}
\sum_{i=0}^{q} h(i) h\left(i+m_{1}\right) h\left(i+m_{2}\right) c_{4, x}\left(n_{1}, n_{2}, i+\tau_{1}\right)= \\
\sum_{i=0}^{q} h(i) h\left(i+n_{1}\right) h\left(i+n_{2}\right) c_{4, x}\left(m_{1}, m_{2}, i+\tau_{1}\right) \tag{3.45}
\end{gather*}
$$

While equation (3.45) has been also derived by Tugnait in [33] equation (3.44) would not appear to have been reported before. Novel linear methods for MA parameter estimation using only fourth order cumulants can be developed using both of the above equations. However, equation (3.45) requires more extensive over-parameterisation than equation (3.44) since it involves triple products of the system parameters instead of double products. Because of this in the following we concentrate our attention on equation (3.44). Before using equation (3.44) for MA parameter estimation, it is important to obtain the ranges of the parameters ( $m, n, \tau_{1}, \tau_{2}$ ) that result in non-trivial equations. By setting $\tau_{1}=\tau_{2}=\tau$ in (3.44), the number of parameters is reduced to three ( $m, n, \tau$ ) and the fourth-order case can be treated in a similar manner to the third-order case. In this case the range of the parameters ( $m, n, \tau$ ) is the same as that in section 3 , and the recursive algorithm of section 3.1 can be easily extended to the fourth-order cumulant case by replacing $c_{3, x}(k+\tau, m)$ with $c_{4, x}(k+\tau, k+\tau, m)$. However if ones wishes to consider the possibility of $\tau_{1} \neq \tau_{2}$, then the analysis is less straightforward. The following section deals with this problem.

## Analysis of the parameterisation of the fourth order equation

Using the same arguments with the third-order cumulant equation of section (3), we conclude that we can form the non-redundant set containing all non-trivial equations
(3.44) for $m$ and $n$ satisfying the condition

$$
0 \leq n \leq m \leq q .
$$

However, in order to generate this set of equations in practice, we also need the ranges of the parameters $\tau_{1}$ and $\tau_{2}$. Once we have the ranges of all parameters, we will be able to calculate the maximum number of different equations generated by (3.44). Suppose that $n \geq 0$ in $c_{4}\left(n, i+\tau_{1}, i+\tau_{2}\right)$. For convenience we make the following substitutions:

$$
\alpha=i+\tau_{1} \quad \text { and } \beta=i+\tau_{2}
$$

The cumulant $c_{4}(n, \alpha, \beta)$ can be non-zero only if

$$
\begin{equation*}
n-q \leq \alpha \leq q \tag{3.46}
\end{equation*}
$$

For different orderings of the parameters $n$ and $\alpha$ we have the following ranges for the parameter $\beta$ :

- If $n \leq \alpha \leq q$ then $\alpha-q \leq \beta \leq q$.
- If $0 \leq \alpha \leq n \leq q$ then $n-q \leq \beta \leq q$.
- If $n-q \leq \alpha \leq 0$ then $n-q \leq \beta \leq q+\alpha$.

The above inequalities define the 2-D domain of support for the slice $c_{4}(n, \alpha, \beta)$ for $n \geq 0$ which is described by the polygon ABCDEF in Figure (3.2). Since we have now obtained the region which defines the possible values of $\alpha$ and $\beta$ we can now obtain the corresponding region for $\tau_{1}$ and $\tau_{2}$.

From the right hand side of equation (3.44), we observe that $i$ ranges from 0 to $q-m$. Consequently, a pair ( $\tau_{1}, \tau_{2}$ ) produces a non-trivial equation, only if there exists a value of $i$ for $0 \leq i \leq q-m$, such that the point with coordinates $\left(\tau_{1}+i, \tau_{2}+i\right)$ belongs to the polygon ABCDEF. If we look again at Figure (3.2), we can see that the set of points with coordinates ( $\tau_{1}, \tau_{2}$ ) satisfying the above condition, define the polygon ABGHKF. The number of different pairs ( $\tau_{1}, \tau_{2}$ ) which correspond to points in the polygon ABGHKF, can be calculated as follows:

$$
\begin{gather*}
\sum_{i=0}^{-m-n+2 q}(3 q-i-m-n+1)+\sum_{i=1}^{m+n-q+1}(3 q-m-n)+ \\
\sum_{i=m+n-q+1}^{q}(2 q-i+1)=5 q^{2}+4 q-2 q n-2 q m-m-n+1 \tag{3.47}
\end{gather*}
$$

The above equation means that, given a pair of values for ( $m, n$ ), we can use formula (3.44) to build $5 q^{2}+4 q-2 q n-2 q m-m-n+1$ non-trivial equations, which correspond to different selections of $\tau_{1}$ and $\tau_{2}$. Most of these equations appear twice. In order to create


Figure 3.2: 2-D domain of support for cumulant slice $c_{4}(n, \alpha, \beta) n \geq 0$.
a non-redundant set of equations, we consider ( $\tau_{1}, \tau_{2}$ ) such that the corresponding points belong to either polygon AFKH or ABGH. The number of equations then becomes,

$$
\begin{equation*}
\frac{1}{2}(q+1)(5 q+2-2 n-2 m) \tag{3.48}
\end{equation*}
$$

We sum over the allowed values of $m$ and $n$, we obtain the total number of different equations produced by formula (3.44):

$$
\begin{equation*}
\sum_{m=1}^{q} \sum_{n=0}^{m-1} \frac{1}{2}(q+1)(5 q+2-2 n-2 m)=\frac{1}{4} q(3 q+2)(q+1)^{2} . \tag{3.49}
\end{equation*}
$$

In summary the parameters $m, n, \tau_{1}$ and $\tau_{2}$ must satisfy the following conditions:

- $0 \leq n<m \leq q$,
- $\left(m+n-q \leq \tau_{1} \leq q\right.$ and $\left.\tau_{1}-q \leq \tau_{2} \leq \tau_{1}\right)$ or $\left(m+n-2 q \leq \tau_{1}<m+n-q\right.$ and $\left.m+n-2 q \leq \tau_{2} \leq \tau_{1}\right)$.


### 3.6.2 Least Squares Solution Using only Fourth-Order Cumulants

In a similar fashion with the third-order cumulant case of section 3.5.2, a system of linear equations can be constructed with the following equations:

$$
\begin{array}{r}
\sum_{k=0}^{q} h(k) h(k+m) c_{4, x}\left(k+\tau_{1}, k+\tau_{2}, n\right)- \\
\sum_{k=0}^{q} h(k) h(k+n) c_{4, x}\left(k+\tau_{1}, k+\tau_{2}, m\right)=0 \text { for } n \neq 0 \tag{3.50}
\end{array}
$$

and

$$
\begin{array}{r}
\sum_{k=0}^{q} h(k) h(k+m) c_{4, x}\left(k+\tau_{1}, k+\tau_{2}, n\right)- \\
\sum_{k=1}^{q} h^{2}(k) c_{4, x}\left(k+\tau_{1}, k+\tau_{2}, m\right)=c_{4, x}\left(\tau_{1}, \tau_{2}, m\right), \text { for } n=0 . \tag{3.51}
\end{array}
$$

The post-processing methods presented in section 3.5.3 can be applied to the solution of the system of linear equations in order to compensate for the effects of overparameterisation.

The total number of equations, given by equation (3.49), grows very quickly with the model order $q$ and so it becomes impractical to use all the equations. In practical situations, we need a tractable number of equations which warrant identifiability and relatively small variance of the estimated parameters. Equations which warrant identifiability can be deduced from the recursive solution which is developed in the next section.

### 3.6.3 A Recursive Solution Based on Fourth-Order Cumulants

Taking (3.44) for $m=q, n=0$ and $\tau_{1}=q$ we have

$$
\begin{equation*}
h(q)=\frac{c_{4, x}\left(q, q, \tau_{2}\right)}{c_{4, x}\left(0, q, \tau_{2}\right)} \quad 0 \leq \tau_{2} \leq q . \tag{3.52}
\end{equation*}
$$

The division is always well conditioned if $\tau_{2}=q$ or $\tau_{2}=0$. Again for $m=q$ and $n=0$, if we take $\tau_{1}=q-j$ for $j=1, \ldots, q$ we have the following recursive equation:
$h^{2}(j)=\frac{h(q) c_{4, x}\left(0, q-j, \tau_{2}\right)-\sum_{k=0}^{j-1} h^{2}(k) c_{4, x}\left(q, k+q-j, \tau_{2}\right)}{c_{4}\left(q, q, \tau_{2}+k\right)} \quad-j \leq \tau_{2} \leq q-j$.


Here all the divisions are always well conditioned, if $\tau_{2}=0$. For $m<q, n=0$ and $\tau_{2}=m-2 q$ we also have,

$$
h(q-m) h(q)=\frac{h^{2}(q) c_{4, x}\left(m, q+\tau_{1}, m-q\right)}{c_{4, x}\left(0, q-m+\tau_{1},-q\right)} \quad m-2 q \leq \tau_{1} \leq m-q .
$$

The above division is always well conditioned if $\tau_{1}=m-q$, or if $\tau_{1}=m-2 q$. Finally for $m<q, n=0$ and $\tau_{2}=m+j-2 q, j=1, \ldots, q-m$,

$$
\begin{gathered}
h(q-j-m) h(q-j)=\left[\sum_{k=q-j}^{q} h^{2}(k) c_{4, x}\left(m, k+\tau_{1}, k-2 q+m+j\right)\right. \\
\left.-\sum_{k=q-m-j+1}^{q-m} h(k) h(k+m) c_{4, x}\left(0, k+\tau_{1}, k-2 q+m+j\right)\right] / c_{4, x}\left(0, q-m-j+\tau_{1},-q\right) .
\end{gathered}
$$

In the above equation $m+j-2 q \leq \tau_{1} \leq m+j-q$. We can ensure that the division is well conditioned, if we select $\tau_{1}$ to be either $m+j-2 q$ or $m+j-q$.

The above recursive algorithm, shows that using only a subset of equations from (3.44) and assuming that $h(0)=1$, we can recover uniquely the unknown parameters $h(1), \ldots, h(q), h^{2}(1), \ldots, h^{2}(q)$, and $h(i) h(i+l)$ for $1 \leq i \leq q-1$ and $1 \leq l \leq q-i$.

The recursive algorithm presented here, is based on equations which have the value of the parameter $n$ fixed to 0 . In a similar way, a recursive solution based on equations with the value of the parameter $m$ fixed to $q$ can be developed.

The identifiability theorem 3.1 of section 3.5 .4 can easily be extended to linear systems constructed from fourth-order cumulant based equations.

### 3.7 Asymptotic Performance Analysis

In this section, we derive expressions which can be used to obtain the asymptotic performance of the MA parameter estimation methods developed in this chapter. The asymptotic performance is given as a function of the system parameters and the statistics of the input sequence. The following theorem (Theorem 3.16 in [14]) is required is order to explore the asymptotic distribution of the estimates of MA models.

Theorem 3.2 Let $d(N)$ be a positive sequence that tends to infinity with $N$, and assume that $d(N)\left(\mathbf{s}_{N}-\mathbf{s}(\theta)\right)$ converges in distribution to a Gaussian random vector with zero mean and positive definite covariance $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Assume that $\mathcal{G}(\mathbf{s})$ is continuously differentiable and its Jacobian $\mathbf{G}(\mathbf{s})$ is nonsingular for all $\mathbf{s}(\boldsymbol{\theta})$. Then $d(N)\left(\mathbf{s}_{N}-\mathbf{s}(\boldsymbol{\theta})\right)$ is asymptotically normal with zero mean and covariance $\mathbf{G}(\mathbf{s}) \boldsymbol{\Sigma}(\boldsymbol{\theta}) \mathbf{G}^{\top}(\mathbf{s})$.

Let us now see how the Theorem 3.2 applies to the method developed in section 3.5.2:

- The model order $q$ is assumed to be known so for the vector $\boldsymbol{\theta}$ we have

$$
\begin{equation*}
\boldsymbol{\theta}=\{h(1), \ldots, h(q)\} \tag{3.53}
\end{equation*}
$$

- the sample statistic vector $s_{N}$ consists of the sample third-order cumulants defined over the minimal domain of support i.e.

$$
\begin{align*}
\mathbf{s}_{N} & =\left\{s_{1}(N), s_{2}(N), \ldots, s_{q(q+1) / 2}(N)\right\} \\
& =\left\{\hat{c}_{3, x}(0,0), \hat{c}_{3, x}(0,1), \ldots, \hat{c}_{3, x}(q, q)\right\} . \tag{3.54}
\end{align*}
$$

$\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the asymptotic covariance matrix of the sample cumulant vector $\mathbf{s}_{N}$. In section 2.7 we described a method to calculate the asymptotic covariances of sample moments. Since for zero-mean processes third-order cumulants are equal to third-order moments, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the same as the asymptotic covariance matrix of the sample moment vector which corresponds to $s_{N}$. For fourth-order cumulants the asymptotic covariance matrix of the sample cumulants is obtained by transforming the asymptotic covariance matrix of second- and fourth-order cumulants with the Jacobian of the moment-to-cumulant transformation given by theorem 2.2.

- The mapping $\mathcal{G}(\mathbf{s})$ is a composite mapping given by

$$
\begin{equation*}
\mathcal{G}(\mathbf{s})=\left(\mathcal{G}_{2} \circ \mathcal{G}_{1}\right)(\mathbf{s}) \tag{3.55}
\end{equation*}
$$

where $\mathcal{G}_{1}$ corresponds to the least squares solution of equation (3.28) while $\mathcal{G}_{1}$ corresponds to the least squares solution of equation (3.37). For the Jacobian $\mathbf{G}(\mathbf{s})$ of $\mathcal{G}(\mathbf{s})$ we have that

$$
\begin{equation*}
\mathbf{G}(\mathbf{s})=\mathbf{G}_{2} \mathbf{G}_{1}, \tag{3.56}
\end{equation*}
$$

where $\mathbf{G}_{1}$ and $\mathbf{G}_{2}$ are the Jacobians of $\mathcal{G}_{1}$ and $\mathcal{G}_{2}$ respectively.

The exact form of the Jacobians $\mathbf{G}_{1}$ and $\mathbf{G}_{2}$ can be obtained according to the following theorem [57]:

Theorem 3.3 Assume that the parameter vector $\boldsymbol{\theta}$ and the statistic $\mathbf{s}(\boldsymbol{\theta})$ satisfy a linear constraint

$$
\begin{equation*}
\mathbf{A}(\mathrm{s}) \boldsymbol{\theta}=\mathbf{b}(\mathrm{s}) . \tag{3.57}
\end{equation*}
$$

Denote with $\mathcal{F}(\mathbf{s})$ the transformation of the statistics vector corresponding to the Least Squares solution of equation 3.57 i.e.

$$
\begin{equation*}
\hat{\boldsymbol{\theta}}=\left[\mathbf{A}^{\top}\left(\mathbf{s}_{N}\right) \mathbf{A}\left(\mathbf{s}_{N}\right)\right]^{-1} \mathbf{A}^{\top}\left(\mathbf{s}_{N}\right) \mathbf{b}\left(\mathbf{s}_{N}\right) \tag{3.58}
\end{equation*}
$$

The Jacobian of $\mathcal{F}(\mathbf{s})$ is then given by the following expression

$$
\begin{equation*}
\mathbf{F}=\left[\mathbf{A}^{\top}\left(\mathbf{s}_{N}\right) \mathbf{A}\left(\mathbf{s}_{N}\right)\right]^{-1} \mathbf{A}^{\top}\left(\mathbf{s}_{N}\right) \mathbf{D}(\mathbf{s}) \tag{3.59}
\end{equation*}
$$

where $\mathbf{D}(\mathrm{s})$ is a matrix whose $i^{\text {th }}$ column is

$$
\begin{equation*}
\mathbf{D}_{i}=\frac{\partial \mathbf{b}}{\partial s_{i}}-\left.\frac{\partial \mathbf{A}}{\partial s_{i}} \boldsymbol{\theta}\right|_{\mathrm{s}=\mathrm{s}(\boldsymbol{\theta})} \tag{3.60}
\end{equation*}
$$

where $s_{i}$ is the $i^{\text {th }}$ element of the vector statistic $\mathbf{s}$.

According to theorem 3.3 and the notation of section 3.5.2, the Jacobian $\mathbf{G}_{1}$ is given by:

$$
\begin{equation*}
\mathbf{G}_{1}=\left[\mathbf{B}^{\top} \mathbf{B}\right]^{-1} \mathbf{B}^{\top} \mathbf{D}_{\mathbf{1}} \tag{3.61}
\end{equation*}
$$

The $i^{\text {th }}$ column of $\mathbf{D}_{1}$ is given by the following expression:

$$
\begin{equation*}
\mathbf{D}_{\mathbf{1}_{i}}=\frac{\partial \mathbf{d}}{\partial s_{i}(N)}-\left.\frac{\partial \mathbf{B}}{\partial s_{i}(N)} \boldsymbol{\theta}\right|_{\mathbf{s}=\mathbf{s}(\boldsymbol{\theta})} \tag{3.62}
\end{equation*}
$$

where $s_{i}(N)$ is the $i^{\text {th }}$ element of the vector $\mathbf{s}(N)$ defined in equation 3.54. Similarly for the Jacobian $\mathbf{G}_{2}$ we have the following expression:

$$
\begin{equation*}
\mathbf{G}_{2}=\left[\mathbf{H}^{\top} \mathbf{H}\right]^{-1} \mathbf{H}^{\top} \mathbf{D}_{2} \tag{3.63}
\end{equation*}
$$

In equation 3.63 the $i^{\text {th }}$ column of $\mathbf{D}_{\mathbf{2}}$ is given by

$$
\begin{equation*}
\mathbf{D}_{\mathbf{2}_{i}}=\frac{\partial \mathbf{g}_{\mathrm{e}}}{\partial \hat{r_{i}}}-\left.\frac{\partial \mathbf{H}}{\partial \hat{r_{i}}} \boldsymbol{\theta}\right|_{\mathbf{r}=\mathbf{r}(\boldsymbol{\theta})} \tag{3.64}
\end{equation*}
$$

where $\hat{r_{i}}$ is the $i^{\text {th }}$ element of the vector $\hat{\mathbf{r}}$ which is defined as

$$
\begin{array}{r}
\hat{\mathbf{r}}=\left[\hat{h(1)}, \ldots, \hat{h(q)}, \widehat{h^{2}(1)}, \widehat{h(1) h(2)}, \ldots, \widehat{h(1) h(q)}, \widehat{h(2) h(1)}, \ldots, h \widehat{h(2) h(q)}, \ldots \ldots, \widehat{h^{2}(q)},\right. \\
\widehat{h(1)}, \ldots, \widehat{h(q)}] 3.65)
\end{array}
$$

### 3.8 MA Model Order Determination

In this section, the problem of MA order determination using higher order cumulants is addressed. The problem of order determination is crucial since most of the existing

HOC based system identification methods are very sensitive to incorrect model order. The main idea behind all the MA model order determination methods, is that the third-order cumulants of a MA model are identically zero for lags outside the region defined by $-q \leq \tau_{1}, \tau_{2} \leq q 2.2$. In [67] two methods were suggested for the order determination of FIR systems that are based on visual inspection and statistical testing. In [68] it is pointed out that although SVD is a numerically robust tool for AR order determination, it is not commonly used for MA order determination. They provide the first SVD criteria for MA model determination. In [35] the order is determined through the minimisation of a cumulant error measure. This method is particularly attractive since it can be implemented as a completely automated procedure without the involvement of visual inspection of singular values or subjective thresholding.

In this section, novel approaches to model order selection are considered which are based on the system identification methods developed in the previous sections. More specifically, two methods based on the optimisation of performance criteria are developed. The first involves the maximisation of the confidence factor defined in equation (3.33), and the second the minimisation of a cumulant matching error measure in the fashion of [35].

Both the new methods rely exclusively on third or fourth-order cumulants and consequently, they are of use in situations when the output sequence is contaminated by additive coloured Gaussian noise with unknown statistics.

### 3.8.1 Model Order Selection Using a Criterion Based on the Confidence Factor

The idea behind the proposed system order selection method is to select the order that results in the largest confidence measure $\lambda$ defined by equation (3.33). The proposed method can be summarised as follows:
(i). Assume that the system order is less than $p$. (In practice it is usually possible to make such an assumption.)
(ii). Assume that the model order is $q$, where
$q=1, \ldots, p$. For each value of $q$ do the following:
(a) From the given data calculate the sampled third-order cumulants assuming that the system order is $q$.
(b) Use the least-squares method of section 3.5.2 to calculate the parameter vector $\mathrm{g}_{q}$.
(c) Form the matrix $\mathbf{R}_{q}$ as shown in equation 3.30.
(d) Perform SVD on $\mathbf{R}_{q}$ and calculate the confidence measure $\lambda_{q}$ from equation (3.33).
(iii). Select the order $q$ that yields the maximum value of $\lambda_{q}$.

The described order selection algorithm offers the advantage of not requiring a visual inspection or subjective thresholding as in the approaches in [68, 67].

### 3.8.2 Order selection through minimisation of a cumulant-error measure

The model order selection method of the previous section is based on the principle that the confidence measure is maximised for the correct model order. The confidence measure corresponding to a model order q , is used as a criterion of how well the underlying model fits the data. In this section a different criterion of goodness of fit is used. This can be achieved using a cumulant-matching error measure. Such a criterion has been used before in [35]. In order to be implemented, the cumulant-matching error measure requires a good parameter estimation method. The MA parameter estimation method developed in this chapter is proposed for this purpose. The third-order cumulant based parameter estimation method used in [35] is a variation of the closed-formula solution for parameter estimation, and is prone to errors due to numerical ill-conditioning.

The model order selection algorithm for third-order cumulants is summarised in the following. The extension to fourth-order cumulants is straightforward. For the unknown model order q is assumed that $q \leq q_{\max }$. Suppose that the order is $q^{\prime} \leq q_{\max }$. Under this assumption use the methods of section 3.5.2, to obtain a vector of parameters

$$
\begin{equation*}
\mathbf{h}_{q^{\prime}}=\left[h_{q^{\prime}}(0), h_{q^{\prime}}(1), \ldots, h_{q^{\prime}}\left(q^{\prime}\right)\right]^{\top} \tag{3.66}
\end{equation*}
$$

where $h_{q^{\prime}}(0)=1$. Then using [51] we calculate third-order cumulants corresponding to the parameter vector $h_{q^{\prime}}$ :

$$
\begin{equation*}
c_{q^{\prime}}\left(\tau_{1}, \tau_{2}\right)=\sum_{i=0}^{q^{\prime}} h_{q^{\prime}}(i) h_{q^{\prime}}\left(i+\tau_{1}\right) h_{q^{\prime}}\left(i+\tau_{2}\right) . \tag{3.67}
\end{equation*}
$$

The cumulant-matching error function corresponding to $q^{\prime}$ is then given as

$$
\begin{equation*}
E_{q_{\max }}\left(q^{\prime}\right)=\sum_{\tau_{1}=0}^{q_{\max }} \sum_{\tau_{2}=0}^{\tau_{1}}\left(\frac{c_{3, x}\left(\tau_{1}, \tau_{2}\right)}{c_{3, x}(0,0)}-\frac{c_{q^{\prime}}\left(\tau_{1}, \tau_{2}\right)}{c_{q^{\prime}}(0,0)}\right)^{2} \tag{3.68}
\end{equation*}
$$

The selected model order $\hat{q}$ is obtained as follows:

$$
\begin{equation*}
\hat{q}=\min _{q^{\prime}} E_{q_{\max }}\left(q^{\prime}\right) \tag{3.69}
\end{equation*}
$$

The above algorithm is not affected by the presence of additive Gaussian noise since it relies on a purely third-order cumulant based estimation method.

### 3.9 Applications to ARMA Parameter Estimation

This section addresses the problem of estimating the parameters of non-Gaussian ARMA processes using only third-order cumulants of the observations. Assume that $\{y(n)\}$ is an ARMA $(p, q)$ process satisfying the following difference equation:

$$
\begin{equation*}
\sum_{i=0}^{p} \alpha(i) y(k-i)=\sum_{i=0}^{q} b(i) w(k-i) \tag{3.70}
\end{equation*}
$$

where $\{w(n)\}$ is an unobservable, stationary, zero-mean, iid, non-Gaussian process. The transfer function corresponding to model (3.70) is given by

$$
\begin{equation*}
H(z)=\frac{B(z)}{A(z)}=\frac{\sum_{i=0}^{q} b(i) z^{-i}}{\sum_{i=0}^{p} \alpha(i) z^{-i}} \tag{3.71}
\end{equation*}
$$

The $z$-transform of the third-order cumulants of $\{y(n)\}$ are given by the following equation [51]:

$$
\begin{equation*}
C_{3, y}\left(z_{1}, z_{2}\right)=\gamma_{3, x} H\left(z_{1}\right) H\left(z_{2}\right) H\left(z_{1}^{-1} z_{2}^{-2}\right)=\gamma_{3, x} \frac{B\left(z_{1}\right) B\left(z_{2}\right) B\left(z_{1}^{-1} z_{2}^{-1}\right)}{A\left(z_{1}\right) A\left(z_{2}\right) A\left(z_{1}^{-1} z_{2}^{-1}\right)} \tag{3.72}
\end{equation*}
$$

In [60], it was shown that the problem of estimating the $\operatorname{ARMA}(p, q)$ parameters can be reduced to two MA estimation problems. According to [60], equation (3.72) can be written as

$$
\begin{equation*}
C_{3, y}\left(z_{1}, z_{2}\right) A\left(z_{1}\right) A\left(z_{2}\right) A\left(z_{1}^{-1} z_{2}^{-1}\right)=\gamma_{3, x} B\left(z_{1}\right) B\left(z_{2}\right) B\left(z_{1}^{-1} z_{2}^{-1}\right) \tag{3.73}
\end{equation*}
$$

In the time domain equation (3.73) becomes

$$
\sum_{i, j=-p}^{p} \alpha_{3}(i, j) c_{3, y}(m-i, n-j)= \begin{cases}0 & m, n \notin S(q, p)  \tag{3.74}\\ \gamma_{3, x} b_{3}(m, n) \quad m, n \in S(q, p)\end{cases}
$$

where $S(q, p)=\left\{(m, n): q<\left|\tau_{n}\right| \leq q+2 p, n=1,2\right\}$. From equation (3.74), one can estimate the coefficients $\alpha_{3}(i, j)$ and $b_{3}(i, j)$ which are then considered as thirdorder cumulants of MA models corresponding to the AR part and the MA part of the ARMA model respectively. In [60] the $c(q, k)$-algorithm is used to obtain the system parameters. In general any MA parameter estimation method can be used for the same purpose.

### 3.10 Numerical Simulations

In this section numerical experiments are performed to demonstrate the performance of the methods developed in this chapter. Random signals are generated according to the following signal models:

## Model 1

$$
\begin{gathered}
x(n)=w(n)+0.9 w(n-1)+1.385 w(n-2)-0.771 w(n-3) \\
y(n)=x(n)+v(n)
\end{gathered}
$$

The zeros of the system transfer function $H(z)$ are located at $0.403,-0.651 \pm j 1.219$.

## Model 2

$$
\begin{array}{r}
x(n)=w(n)+0.1 w(n-1)-1.87 w(n-2)+3.02 w(n-3)-1.435 w(n-4) \\
+0.49 w(n-5)
\end{array}
$$

$$
y(n)=x(n)+v(n)
$$

The zeros of the system transfer function $H(z)$ are located at $-2,0.7 \pm j 0.7$ and $0.25 \pm j 0.433$. This model has also been used in $[5,6,9]$.

## Model 3

$$
\begin{gather*}
x(n)=w(n)+0.1 w(n-1)-1.87 w(n-2)+3.02 w(n-3) \\
-1.435 w(n-4)+1.49 w(n-5)  \tag{3.75}\\
y(n)=x(n)+v(n)
\end{gather*}
$$

The zeros of the system transfer function $H(z)$ are located at $-2.02,0.933 \pm j 0.7158$ and $0.0287 \pm j 0.729$.

## Model 4

$$
\begin{gathered}
x(n)=w(n)-1.13 w(n-1)+0.6 w(n-2) \\
y(n)=x(n)+v(n)
\end{gathered}
$$

The zeros of the system transfer function $H(z)$ are located at $0.565 \pm j 0.529882$. In all models the input signal $w(n)$ used in the simulations involving third-order cumulants is a zero-mean exponentially distributed IID noise sequence with $\sigma_{w}^{2}=1$ and $\gamma_{3, w}=2$. In simulations involving fourth-order cumulants the input signal $w(n)$ is an IID noise sequence distributed according to a Laplace distribution with parameter $l=1$, and $\sigma_{w}^{2}=2, \gamma_{3, w}=0$ and $\gamma_{4, w}=24$. Additive coloured noise is created as the output of
the following MA(4) model:

$$
\begin{equation*}
v(n)=0.5 u(n)-0.25 u(n-1)-0.5 u(n-2)+0.25 u(n-3)-0.25 u(n-4) \tag{3.76}
\end{equation*}
$$

where the input sequence is an IID Gaussian sequence. We define the signal-to-noise ratio as $S N R(d B)=10 \log \left(P_{x} / P_{v}\right)$ where $P_{x}$ denotes the power of the signal. The accuracy of system identification is assessed by calculating the Mean Square Error (MSE):

$$
M S E=\frac{\sum_{i=0}^{q}(h(n)-\hat{h}(m))^{2}}{\sum_{i=0}^{q} h^{2}(n)}
$$

where $\hat{h}(m)$ is the estimated system parameter.

### 3.10.1 Third-order cumulant based estimation

The system identification methods used in the simulation are marked as follows:

M1 A system of equations is constructed according to section 3.5.2. The system consists of equations of the type of 3.26 and 3.27 corresponding to the following set of triplets:

$$
\begin{gathered}
\{(0, q, \tau): \tau=-q, \ldots, 0\} \cup\{(0, q-1, \tau): \tau=q, q+1\} \cup \\
\{(0, m, \tau): m=1, \ldots, q-2 \text { and } \tau=-q, \ldots, 2 q-m\} \cup \\
\{(1, m, \tau): m=2, \ldots, q-1 \text { and } \tau=-q, \ldots, q-m\} \\
\cup\{(2, m, \tau): m=3, \ldots, q-1 \text { and } \tau=0, \ldots, q-m\} .]
\end{gathered}
$$

So for $q>2$ there are $\frac{11}{2} q(q-1)-1$ equations. The system is solved using TLS and the over-parameterisation of the resulting vector is reduced using the LS methods of section 3.5.3.

M2 Is the same as M1 with the only difference being that the over-parameterisation is reduced using the method SVD-based method described in section 3.5.3.

M3 this is the same as M1 with the only difference being that the linear system is solved using LS instead of TLS.

M4 the method of Alshebeili et al in [35]
M5 the method of Fonollosa et al [69]
M6 the method of Tugnait in [34]
M7 this is the same as M3 but TLS is used to solve the linear system.

M8 This is a nonlinear optimisation method based on the minimisation of a cumulantmatching error criterion [31]. The nonlinear method is initialised using method M1.

Other methods for MA parameter estimation like [ $1,32,37,38$ ] have not been considered in the simulations since they perform very poorly in the examples considered here. In the first example sequences were generated according to the signal model 2. To reduce the realisation dependency of our simulations, the parameter estimates are averaged over 50 Monte Carlo simulations. Graphs (a) to (e) in figure 3.3 represent identification results for the individual parameters of signal model 2 . The midpoints in the vertical bars represent the average value of the estimate after 50 Monte Carlo runs. The length of the vertical bar is twice the standard deviation of the estimated values. Graph (f) in figure 3.3 shows the MSE of the different identification methods. All the graphs show the results for $\mathrm{SNR}=50 \mathrm{~dB}$ (coloured green) and for $\mathrm{SNR}=10 \mathrm{~dB}$ (coloured blue). The results produced here show that the solution of the linear systems proposed in section 3.5.2 with TLS (method M1) produces better results than plain LS (method M3). Furthermore, when the linear systems are solved using TLS, the use of the method of section 3.5.3 (method M1) to overcome the effects of over-parameterisation produce better results that those of the methods of section 3.5 .3 (method M2). The results of figure 3.3 also show that the method M1 proposed in this chapter outperforms all the other linear methods M2 to M7. For high SNR the performance of Alshebeili et al [35] method M4 is comparable to that of M1. For lower SNR the involvement of second order statistics in M4 and M6 has a considerable effect on the accuracy of the estimates. Replacing the LS solution in M4 with TLS, results in method M7 which has significantly worse performance than M4. From figure 3.3 it is clear that the estimates of M4 and M5 are highly biased. As expected the nonlinear method M8 performs better than the linear methods. However the improvement over M1, which was used to initialise M8, is not significant. Figure 3.4 compares the location of zeros models estimated using M1 with the location of models estimated using M4, M5 and M6. It is seen from figure 3.4 that the effect of decreasing the SNR is to increase the dispersion of the estimated zeros. This is more prominent for the zeros estimated by M4 and M6 since they are more sensitive to additive coloured noise. The good performance of method M1 is evident from the small dispersion of the corresponding zeros and their relative robustness to additive coloured noise.

The next example involves signal model 1. System identification results for $\mathrm{SNR}=50 \mathrm{~dB}$ and $\mathrm{SNR}=0 \mathrm{~dB}$ are presented in figure 3.5. Once again the results are averaged over 50 Monte Carlo runs. For $\mathrm{SNR}=50 \mathrm{~dB}$ all methods perform very well for signal model 1. As expected the results for the methods involving second-order statistics are highly biased for $\mathrm{SNR}=0 \mathrm{~dB}$. The method M1 proposed in this section performs very well in both cases. It achieves results which are comparable only with the nonlinear methods M8, which itself uses M1 to provide the initial solution. Figure 3.6 compares the


Figure 3.3: Identification results for signal model 2. The number of output samples is 5000 . The graphs display the performance of system identification methods for $\mathrm{SNR}=50 \mathrm{~dB}$ and $\mathrm{SNR}=10 \mathrm{~dB}$.


Figure 3.4: Estimated locations of zeros of signal model 2 after 50 Monte Carlo runs. The number of output samples is 5000 .


Figure 3.5: Identification results for signal model 1. The number of output samples is 2000 . The graphs display the performance of system identification methods for $\mathrm{SNR}=50 \mathrm{~dB}$ and $\mathrm{SNR}=0 \mathrm{~dB}$.
estimated locations of zeros for signal model 1, obtained from M1 with those obtained with the rest of the methods. It is important to note that the results in figure 3.6 are



Figure 3.6: Estimated locations of zeros of signal model 1 after 50 Monte Carlo runs. The number of output samples is 2000 .
obtained a very low SNR of 0 dB . The results obtained from methods M4 and M6 are practically useless.

The final example involves signal model 4. This is a minimum-phase system. The system identification results summarised in figure 3.7 were averaged over 50 Monte Carlo runs. The length of the output sequence is 2000 samples. Once again the results demonstrate the significant advantage achieved by system identification methods based




Figure 3.7: Identification results for signal model 4. The number of output samples is 2000 . The graphs display the performance of system identification methods for $\mathrm{SNR}=50 \mathrm{~dB}$ and $\mathrm{SNR}=0 \mathrm{~dB}$.
$\mathrm{SNR}=0 \mathrm{db}$


Figure 3.8: Estimated locations of zeros of signal model 4 after 50 Monte Carlo runs. The number of output samples is 2000 .
only on third-order cumulants, in low SNR environments. The estimated locations of zeros for signal model 4 obtained from method M1, are compared with the locations obtained from M4 and M5, in figure 3.8. The SNR in 3.8 is 0 dB . Method M4 fails to locate the zeros accurately. Both M1 and M5 provide estimates relatively near the true zeros, but the estimates obtained from M1 are dispersed in a smaller region around the true zeros.

In summary, from the simulations examined in this section, the third-order cumulant method proposed in this section performs better than existing linear methods especially for very low SNR. The use of TLS for the solution of the linear system of section 3.5.2 improves significantly the accuracy in system identification.

### 3.10.2 Fourth-order cumulant based estimation

Fourth-order cumulant based methods have not been analysed extensively in the existing higher-order statistics literature. The second- and third-order cumulant method of Alshebeili et al can be extended to form a second- and fourth-order cumulant based method. However the information provided in the Appendix of [35] is not sufficient to allow an implementation of the second- and fourth-order cumulant based method. In theory, the extension of [35] has an important disadvantage compared to the method proposed in 3.6.2. The fourth-order extension in [35] is considerably more complicated than the third-order version and requires more extensive over-parameterisation. In contrast, the method proposed in 3.6 .2 can very simply be obtained from the thirdorder method of section 3.5 .2 by replacing the third-order cumulants with appropriate fourth-order cumulants.

The fourth-order cumulant-based system identification methods considered in this section are the following:

N1 This is the method proposed in section 3.6.2.
N2 This is the fourth-order method presented in [36]. It is the extension of M5 considered in the simulations of the previous section.

In the first example, signals consisting of 5000 samples were generated according to signal model 3. The SNR is 50 dB . The results of 50 Monte Carlo runs are summarised in figure 3.9. It can be seen that both in terms of bias of the individual parameter estimates and in terms of standard deviation, the method proposed in this section performs better than method N2. In fact method N2 practically fails to identify the correct system parameters. The locations of the estimated zeros for all Monte Carlo runs are depicted in figure 3.10.


Figure 3.9: Fourth-order cumulants based identification results for signal model 3. The number of output samples is 5000 . The graphs display the performance of system identification methods for $\mathrm{SNR}=50 \mathrm{~dB}$.


Figure 3.10: Estimated locations of zeros of signal model 3 after 50 Monte Carlo runs. The number of output samples is 5000 .

The last example involves signal model 1. The number of output samples is 2000 and the SNR is 50 dB . The results of 50 Monte Carlo simulations are summarised in figure 3.11. For this particular system both methods work quite well with slightly better results achieved with method N2. The zeros of the system are also located accurately with both methods.


Figure 3.11: Fourth-order cumulants based identification results for signal model 1. The number of output samples is 2000 . The graphs display the performance of system identification methods for $\mathrm{SNR}=50 \mathrm{~dB}$.

Summarising we can say that the fourth-order method proposed in this section performs very well in the simulations presented in this section. The method of Fonollosa et al although it performs well for signal model 1 , it practically fails to identify the parameters of signal model 3 when 5000 output samples are available.

### 3.10.3 Results on model order selection

In this section the performance of various model order selection methods is tested. In particular the methods of sections 3.8.1 and 3.8.2 are compared with the method based on equation (73) of [35]. In table 3.1 simulation results are presented for signal
model 1 which has model order 3. In order to illustrate the effectiveness of the algorithms, 100 Monte Carlo runs for determining the order of the system were performed at $\mathrm{SNR}=10 \mathrm{~dB}$ and output sample size $N=500,1000,2000$ and 4000 . The best performance is achieved with the algorithm of section 3.8 .2 followed by that of 3.8.1. A

|  | Number of Successful Selections |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Method | $\mathrm{N}=500$ | $\mathrm{~N}=1000$ | $\mathrm{~N}=2000$ | $\mathrm{~N}=4000$ |
| Sec 3.8.2 | 93 | 97 | 98 | 99 |
| Sec 3.8.1 | 89 | 95 | 97 | 97 |
| $[35]$ | 84 | 90 | 93 | 99 |

Table 3.1: System order selection for Signal Model 1: Successful selections in 100 Monte Carlo trials with SNR=10dB
similar experiment is performed for signal model 2 which has model order 5. The SNR in this case is 20 dB and the experiment is performed for output sample sizes $N=4000$, 5000 and 6000 . Once again the method of section 3.8 .2 performs better than the others followed by the method of section 3.8.1.

|  | Number of Successful Selections |  |  |
| :---: | :---: | :---: | :---: |
| Method | $\mathrm{N}=4000$ | $\mathrm{~N}=5000$ | $\mathrm{~N}=6000$ |
| Sec 3.8.2 | 74 | 84 | 91 |
| Sec 3.8.1 | 71 | 83 | 89 |
| $[35]$ | 40 | 57 | 69 |

Table 3.2: System order selection for Signal Model 2: Successful selections in 100 Monte Carlo trials with SNR=20dB

### 3.11 Conclusion

This chapter has considered the problem of estimating the parameters of an MA model using only third- or fourth-order cumulants. New algorithms have been presented which are based on equations relating different cumulant slices of the same in terms of the system parameters. The identifiability of the algorithms has been established through a recursive solution of these new equations. The simulation results demonstrate that the new methods perform better than existing HOC-based linear methods for MA parameter estimation.

## Chapter 4

## Blind Deconvolution of MA Models

### 4.1 Introduction

In this chapter the problem of estimating the inverse parameters of an MA model from the cumulant statistics of the noisy observations of the system output is considered. The system is driven by an IID non-Gaussian sequence that is not observed. The noise is additive and can be coloured and even non-Gaussian under certain conditions.

The chapter derives general equations relating the inverse system parameters with the output cumulants. These new equations are linear with respect to the inverse system parameters and they are used to develop linear methods for blind deconvolution. Existing deconvolution techniques are given a unified description with the use of the new equations. The generality of the approach adopted in this chapter allows the development of new fourth-order cumulant based deconvolution methods, which are not restricted to the use of only a single one-dimensional slice of the output cumulants.

This chapter also derives expressions for the asymptotic variance of the estimated inverse filter parameters.

Finally the performance of the deconvolution is demonstrated with the use of Monte Carlo simulations.

### 4.2 Problem Definition

The problem of blind deconvolution can be defined as follows. Consider the single-input single-output system depicted in Figure (3.1). The output process $\{y(n)\}$ is generated according to the following convolutional model:

$$
\begin{equation*}
x(n)=\sum_{k=0}^{q} h(k) w(n-k), \tag{4.1}
\end{equation*}
$$

where $\{w(n)\}$ is a zero-mean non-Gaussian stationary process whose moments of order up to eight ${ }^{1}$ is related to are finite, and

$$
\begin{equation*}
y(n)=x(n)+v(n) \tag{4.2}
\end{equation*}
$$

where $\{v(n)\}$ is a Gaussian additive noise process which is independent of $\{x(n)\}$. We assume that $\{v(n)\}$ is spectrally white but this assumption can be relaxed later for some cases. We assume without loss of generality that $h(0), h(q) \neq 0$.


Figure 4.1: Single channel system and deconvolution filter.

The objective of blind deconvolution ${ }^{2}$ is to find an inverse filter with transfer function

$$
\begin{equation*}
\Theta(z)=\sum_{i=r_{1}}^{r_{2}} \theta_{i} z^{-i} \tag{4.3}
\end{equation*}
$$

so that

$$
\begin{equation*}
\hat{w}(n)=\sum_{j=r_{1}}^{r_{2}} \theta_{j} x(n-j) \tag{4.4}
\end{equation*}
$$

where $\hat{w}(n) \approx w(n)$ and $r_{1}$ and $r_{2}$ are the orders of the anti-causal and the causal part respectively. The values that $r_{1}$ and $r_{2}$ can take depend on the phase characteristics of the generating system $H(z)$ :
$H(z)$ is minimum phase: The inverse system $\Theta(z)$ is a purely causal system i.e. $r_{1}=$ 0 and $r_{2} \geq 1$.
$H(z)$ is mixed phase: The inverse system $\Theta(z)$ has both causal and anti-causal parts i.e. $r_{1} \leq-1$ and $r_{2} \geq 1$.

[^12]$H(z)$ is maximum phase: The inverse system $\Theta(z)$ is a purely anticausal system i.e. $r_{1} \leq-1$ and $r_{2}=1$.

In this chapter we assume that $H(z)$ has no zeros on the unit circle. We must note here that since $H(z)$ is the transfer function of an FIR system its inverse $\frac{1}{H(z)}$ is the transfer function of an IIR system. Because the inverse system is assumed to be stable it can be approximated arbitrarily closely by $\Theta(z)$ as defined by equation (4.3). Hence, for sufficiently large values of $-r_{1}$ and $r_{2}$, the following equation holds:

$$
\begin{equation*}
\Theta(z)=\frac{1}{H(z)} . \tag{4.5}
\end{equation*}
$$

Next the two general approaches to blind deconvolution will be examined and the close relationship between deconvolution and system identification will be explained.

Direct and Indirect Deconvolution The inverse system $\Theta(z)$ can be obtained either directly or indirectly. As will be shown later in this section, it is possible to derive relationships between the inverse impulse response $\theta_{j}, \quad j=r_{1}, \ldots, r_{2}$ and the cumulants of the output process, and to use these expressions to obtain the inverse impulse response directly. We refer to this type of method as direct deconvolution. Alternatively, one can use equation (4.5) to obtain the inverse system $\Theta(z)$. This approach involves the estimation of the system transfer function $H(z)$, which is then inverted to obtain $\Theta(z)$. This type of method is referred to as indirect deconvolution and will also be examined more thoroughly later in this chapter.

Both direct and indirect deconvolution methods developed in this chapter belong to the general category of the Method of Moments [14]. In a similar manner to section 3.2.1, a vector of cumulant statistics is calculated from the observed sequence $\{y(n)\}$, which is then transformed so that it results in the unknown parameter vector $\boldsymbol{\theta}$, which in the case of deconvolution is given by

$$
\begin{equation*}
\boldsymbol{\theta}=\left\{\theta_{r_{1}}, \theta_{r_{1}+1}, \ldots, \theta_{r_{2}-1}, \theta_{r 2}\right\} \tag{4.6}
\end{equation*}
$$

In the following section HOC-based methods for direct deconvolution are examined.

### 4.3 Direct Deconvolution

In section 3.3 equations involving cumulants of the output process and the parameters of the generating system have been examined. In order to develop direct deconvolution methods it is required to develop expressions relating the cumulants of the output process to the parameters of the inverse filter.

### 4.3.1 Fundamental relationships for the inverse filter parameters

As shown in section 2.6 the cumulants of MA processes are given by the following equation:

$$
\begin{equation*}
c_{k, x}\left(\tau_{1}, \ldots, \tau_{k-1}\right)=\gamma_{k, w} \sum_{i=0}^{q-\tau_{1}} h(i) h\left(i+\tau_{1}\right) \cdots h\left(i+\tau_{k-1}\right) \tag{4.7}
\end{equation*}
$$

Let $\tau_{i}=\tau_{1}+m_{i}$ for $i>1$ in 4.7. Then after $z$-transforming with respect to $\tau_{1}$ we obtain :

$$
\begin{equation*}
C_{k, x}\left(z ; m_{2}, \ldots, m_{k-1}\right)=\gamma_{k, w} H\left(z^{-1}\right)\left[H(z) *\left[z^{m_{1}} H(z)\right] * \cdots *\left[z^{m_{k-1}} H(z)\right]\right] . \tag{4.8}
\end{equation*}
$$

In order to introduce the inverse filter transfer function in equation 4.8 we divide both sides of 4.8 with $H\left(z^{-1}\right)$ and, after making use of equation 4.5 , we obtain

$$
\begin{equation*}
\Theta\left(z^{-1}\right) C_{k, x}\left(z ; m_{2}, \ldots, m_{k-1}\right)=\gamma_{k, w}\left[H(z) *\left[z^{m_{1}} H(z)\right] * \cdots *\left[z^{m_{k-1}} H(z)\right]\right] . \tag{4.9}
\end{equation*}
$$

In the time domain, and assuming that the impulse response $\theta_{j}$ of the inverse system vanishes for $j<r_{1}$ or $j>r_{2}$, equation 4.9 is equivalent to,

$$
\begin{align*}
& \frac{1}{\gamma_{k, w}} \sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{k, x}\left(\tau+j, \tau+m_{1}+j, \ldots, \tau+m_{k-2}+j\right) \\
= & \begin{cases}h(\tau) h\left(m_{1}\right) \cdots h\left(m_{k-2}\right) & \tau, m_{1}, \ldots, m_{k-2} \in[0, q] \\
0 & \text { otherwise }\end{cases} \tag{4.10}
\end{align*}
$$

Of particular interest in practical situations are the special cases of 4.10 for $k=2,3$ and 4. The corresponding equations are:

$$
\begin{gather*}
\frac{1}{\sigma_{w}^{2}} \sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{2, x}(m+j)= \begin{cases}h(m) & m \in[0, q] \\
0 & \text { otherwise }\end{cases}  \tag{4.11}\\
\frac{1}{\gamma_{3, w}} \sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{3, x}(m+j, n+j)= \begin{cases}h(m) h(n) & m, n \in[0, q] \\
0 & \text { otherwise }\end{cases}  \tag{4.12}\\
\frac{1}{\gamma_{4, w}} \sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{4, x}\left(m_{1}+j, m_{2}+j, n+j\right)= \begin{cases}h\left(m_{1}\right) h\left(m_{2}\right) h(n) & m_{1}, m_{2}, n \in[0, q] . \\
0 & \text { otherwise }\end{cases} \tag{4.13}
\end{gather*}
$$

Equation 4.12 was first reported in [88] and a different derivation of 4.10 for the special case of diagonal cumulants appeared in [26]. The derivation in [26] is inelegant and leads to the false claim that equation 4.10 is a direct consequence of the Giannakis-Mendel equation which was derived in [32]. For suitably selected parameters $\tau, m_{1}, \ldots, m_{k-2}$
the right side of equation 4.10 becomes zero and then 4.10 involves only inverse system parameters and $k^{t h}$ order cumulants of the output process $\{x(n)\}$.

The equations developed so far in this section along with the equation 3.12 can be used to derive new families of equations involving the inverse filter coefficients, the secondorder cumulants and slices of higher-order cumulants. Consider equation 4.11 for $m=i$ $0 \leq i \leq q$ :

$$
\begin{equation*}
\sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{2, x}(i+j)=\sigma_{w}^{2} h(i) \tag{4.14}
\end{equation*}
$$

where $\gamma_{2, w}=\sigma_{w}^{2}$. Equation 4.10 can be rewritten as

$$
\begin{equation*}
\sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{k, x}\left(i+j, i+\tau_{1}+j, \ldots, i+\tau_{k-2}+j\right)=\gamma_{k, w} h(i) h\left(i+\tau_{2}\right) \cdots h\left(i+\tau_{k-2}\right) \tag{4.15}
\end{equation*}
$$

where $i, i+\tau_{1}, i+\tau_{2}, \ldots, i+\tau_{k-2} \in[0, q]$. Finally 3.12 for $n=k>2$ and $m=2$

$$
\begin{equation*}
\sum_{i=0}^{q} h(i) c_{k, x}\left(i+\tau, \tau_{1}, \tau_{2}, \ldots, \tau_{k-2}\right)=\epsilon_{k, 2} \sum_{i=0}^{q} h(i)\left[\prod_{k=1}^{k-2} h\left(i+\tau_{k}\right)\right] c_{2, x}(i+\tau) \tag{4.16}
\end{equation*}
$$

where $\epsilon_{k, 2}=\frac{\gamma_{k}, w}{\sigma_{w}^{2}}$. Equations 4.14 and 4.15 can be substituted in the left and right side of equation 4.16 respectively, in order to replace the system parameters $h(\cdot)$ 's with the inverse filter parameters:

$$
\begin{array}{r}
\sum_{i=0}^{q} \sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{2, x}(i+j) c_{k, x}\left(i+\tau, \tau_{1}, \tau_{2}, \ldots, \tau_{k-2}\right)= \\
\sum_{i=0}^{q} \sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{k, x}\left(i+j, \tau_{1}+i+j, \ldots, \tau_{k-2}+i+j\right) c_{2, x}(i+\tau) \tag{4.17}
\end{array}
$$

where $\tau_{1}, \tau_{2}, \ldots, \tau_{k-2} \in[0, q]$. Equation 4.17 for third and fourth order cumulants becomes

$$
\begin{equation*}
\sum_{i=0}^{q} \sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{2, x}(i+j) c_{3, x}\left(i+\tau, \tau_{1}\right)=\sum_{i=0}^{q} \sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{3, x}\left(i+j, \tau_{1}+i+j\right) c_{2, x}(i+\tau) \tag{4.18}
\end{equation*}
$$

and

$$
\begin{array}{r}
\sum_{i=0}^{q} \sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{2, x}(i+j) c_{4, x}\left(i+\tau, \tau_{1}, \tau_{2}\right)= \\
\sum_{i=0}^{q} \sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{4, x}\left(i+j, \tau_{1}+i+j, \tau_{2}+i+j\right) c_{2, x}(i+\tau) \tag{4.19}
\end{array}
$$

Equations 4.17 and consequently equations 4.17 and 4.19 are linear with respect to inverse system parameters and, as will be shown in the next section, they can be used
for direct deconvolution.

The equations developed in this section can be used to construct linear systems of equations with respect to the inverse filter coefficients. Different combinations of equations can be used for the construction of the linear system, resulting in different deconvolution schemes. The first scheme presented in the next section is based on equation 4.12.

### 4.3.2 Deconvolution scheme using only third-order cumulants

According to the problem definition of section 4.2, the only known quantities in equation 4.12 are the third-order cumulants. In order to have equations involving only inverse filter coefficients and cumulants and not the system's parameters, it is appropriate to select the equations where the right-hand side is zero, i.e. equations corresponding to $m, n$ in which at least one is outside $[0, q]$. Given a pair of parameters ( $m, n$ ), which at least one is outside $[0, q]$ the corresponding equation is

$$
\begin{equation*}
\sum_{j=r_{1}}^{r_{2}} \theta_{j} c_{3, x}(m+j, n+j)=0 \tag{4.20}
\end{equation*}
$$

This equation has been used in a similar deconvolution method developed in [39]. An adaptive version of that method has been reported in [89].

The equation corresponding to the parameter pair ( $m, n$ ) is the same as the equation corresponding to the parameter pair ( $n, m$ ) since $c_{3, x}(m+j, n+j)=c_{3, x}(n+j, m+j)$. In order to avoid duplicating equations, consider only the equations corresponding to parameter pairs ( $m, n$ ) with $m \leq n$ are considered. The set of cumulant lags involved in the equation corresponding to a parameter pair ( $m, n$ ) is the following:
$\mathcal{L}_{(m, n)}=\left\{\left(m+r_{1}, n+r_{1}\right),\left(m+r_{1}+1, n+r_{1}+1\right), \ldots,\left(m+r_{2}-1, n+r_{2}-1\right),\left(m+r_{2}, n+r_{2}\right)\right\}$
Denote with $\mathcal{L}_{M A(q)}$ the set of all third-order cumulant lags belonging to the domain of support of a MA $(q)$ process (the domain of support of MA processes has been discussed in section 2.6). Then, the equation corresponding to the parameter pair ( $m, n$ ) is nontrivial only if the intersection $\mathcal{L}_{(m, n)} \cap \mathcal{L}_{M A(q)}$ is non-empty. Consequently, the set of parameter pairs ( $m, n$ ), $m \leq n$ corresponding to non-trivial equations is defined as follows:

$$
\begin{array}{r}
\mathcal{D}=\left\{(-1-i, q-j-i): i=0, \ldots, r_{2}+q-1 \text { and } j=1, \ldots, q+1\right\} \\
\cup\left\{(i+j, q+1+i): i=0, \ldots,-r_{1} \text { and } j=1, \ldots, q+1\right\} . \tag{4.21}
\end{array}
$$

The set of non-trivial equations for $m \leq n$, is depicted in figure 4.2. The first set in the definition of $\mathcal{D}$, is represented by the rectangle DOKL excluding the edge DO. The


Figure 4.2: Domain of support for equation 4.20.
second set in the definition of $\mathcal{D}$, is represented by the rectangle ABCD excluding the the edge CD. Individual equations are represented by points with integers coordinates. The total number of equations is $\left(r_{2}+q\right)(q+1)-r_{1}(q+1)$. In the following it is assumed that we are dealing with mixed- or minimum-phase systems and that the inverse impulse response is normalised so that $\theta_{0}=1$. Equation 4.20 then becomes,

$$
\begin{equation*}
\sum_{j=r_{1}, j \neq 0}^{r_{2}} \theta_{j} c_{3, x}(m+j, n+j)=-c_{3, x}(m, n) \tag{4.22}
\end{equation*}
$$

Collecting the equations defined in 4.21, a system of linear equations with respect to the unknowns $\theta_{r_{1}}, \theta_{r_{1}+1}, \ldots, \theta_{-1}, \theta_{1}, \ldots, \theta_{r_{2}}$ can be formed. The detailed structure of the linear system is analysed in the following. Suppose that the equations corresponding to the parameters pairs belonging to the set

$$
\begin{equation*}
\left\{(-1-i, q-j-i): i=0, \ldots, r_{2}+q-1\right\} \tag{4.23}
\end{equation*}
$$

are collected for a given $j$ satisfying $1 \leq j \leq q+1$. A matrix with the coefficients of the inverse impulse response can then be formed. The coefficients of $\theta_{r_{1}}$ are located in column 1 and the coefficients of $\theta_{r_{2}}$ are located in the last column i.e. column $-r_{1}+r_{2}$. The matrix consisting of the coefficients of the selected equations has the following block structure:

$$
\mathbf{A}(j)=\left[\begin{array}{c|c}
\mathbf{A}_{\mathbf{1}, \mathbf{1}}(j) & \mathbf{A}_{\mathbf{1}, \mathbf{2}}(j)  \tag{4.24}\\
\hline \mathbf{0}_{\mathbf{r}_{\mathbf{2}}+\mathbf{1},-\mathbf{r}_{\mathbf{1}}} & \mathbf{A}_{\mathbf{2 , \mathbf { 2 }}}(j)
\end{array}\right] .
$$

$\mathbf{A}(j)$ is an $\left(r_{2}+q\right) \times\left(r_{2}-r_{1}\right)$ matrix. The matrices $\mathbf{A}_{\mathbf{1}, \mathbf{1}}(j)$ and $\mathbf{0}_{\mathbf{r}_{\mathbf{2}}+\mathbf{1},-\mathbf{r}_{\mathbf{1}}}$ contain the coefficients of the non-causal part of the inverse impulse response. The matrix $\mathbf{0}_{\mathbf{r}_{2}+1,-\mathbf{r}_{1}}$ is a $\left(r_{2}+1\right) \times\left(-r_{1}\right)$ matrix whose elements are all zero. For convenience of notation, in the following we assume that $c_{\tau_{1}, \tau_{2}}=c_{3, x}\left(\tau_{1}, \tau_{2}\right)$. The matrix $\mathbf{A}_{1,1}(j)$ is a ( $q-1$ ) $\times\left(-r_{1}\right)$ matrix and has the following structure:

$$
\mathbf{A}_{\mathbf{1}, 1}(j)=\left[\begin{array}{ccccccc}
0 & \cdots & 0 & c_{-q,-j+1} & c_{-q+1,-j+2} & \cdots & c_{-2, q-j-1}  \tag{4.25}\\
0 & \cdots & 0 & 0 & c_{-q,-j+1} & \cdots & c_{-3, q-j-2} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & 0 & 0 & c_{-q,-j+1}
\end{array}\right]
$$

$\mathbf{A}_{\mathbf{1}, \mathbf{2}}(j)$ and $\mathbf{A}_{\mathbf{2}, \mathbf{2}}(j)$ contain the coefficients of the causal part of the inverse impulse response ( $\theta_{j}$ for $0<j \leq r_{2}$ ). $\mathbf{A}_{1,2}(j)$ is a $(q-1) \times r_{2}$ matrix and has the following structure:

$$
\mathbf{A}_{\mathbf{1}, 2}(j)=\left[\begin{array}{cccccccc}
c_{0, q-j+1} & \cdots & c_{j-1, q} & 0 & 0 & \cdots & \cdots & 0  \tag{4.26}\\
c_{-1, q-j} & \cdots & c_{j-2, q-1} & c_{j-1, q} & 0 & 0 & \cdots & 0 \\
\vdots & \cdots & \cdots & \cdots & \ddots & \ddots & \ddots & \vdots \\
c_{-q+2,-j+3} & \cdots & \cdots & c_{j-2, q-1} & c_{j-1, q} & 0 & \cdots & 0
\end{array}\right]
$$

$\mathbf{A}_{\mathbf{2 , 2}}(j)$ is an $\left(\left(r_{2}+1\right) \times r_{2}\right)$ matrix and is defined as $\mathbf{A}_{\mathbf{2}, \mathbf{2}}(j)=$
$\left[\begin{array}{cccccccc}c_{-q+1,-j+2} & \cdots & c_{j-1, q} & 0 & 0 & 0 & \cdots & 0 \\ c_{-q,-j+1} & c_{-q+1,-j+2} & \cdots & c_{j-1, q} & 0 & 0 & \cdots & 0 \\ 0 & c_{-q,-j+1} & \cdots & c_{j-2, q-1} & c_{j-1, q} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & c_{-q,-j+1} & \cdots & c_{j-2, q-1} & c_{j-1, q} & 0 \\ 0 & 0 & \cdots & 0 & c_{-q,-j+1} & \cdots & c_{j-2, q-1} & c_{j-1, q} \\ 0 & 0 & \cdots & 0 & 0 & c_{-q,-j+1} & \cdots & c_{j-2, q-1} \\ \vdots & \cdots & \cdots & \cdots & \cdots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & c_{-q,-j+1}\end{array}\right]$.

The constant terms of the right-hand side of 4.22 , are collected in the following vector:

$$
\begin{equation*}
\mathbf{a}(j)=[\underbrace{-c_{-1, q-j}, \ldots,-c_{-q+1,-j+2}}_{(q-1) \text {-elements }}, \underbrace{-c_{-q,-j+1}, 0, \ldots, 0}_{\left(r_{2}+1\right)-\text { elements }}]^{\top} \tag{4.27}
\end{equation*}
$$

Let us now consider equations corresponding the parameter pairs belonging to the set

$$
\begin{equation*}
\left\{(i+j, q+1+i): i=0, \ldots,-r_{1}-1\right\} \tag{4.28}
\end{equation*}
$$

for a given $j$ satisfying $1 \leq j \leq q+1$. The matrix containing the coefficients of these equations has the following block structure:

$$
\begin{equation*}
\mathbf{B}(j)=\left[\mathbf{B}_{\mathbf{1}}(j) \mid \mathbf{0}_{-\mathbf{r}_{1}, \mathbf{r}_{\mathbf{2}}}\right] . \tag{4.29}
\end{equation*}
$$

$\mathbf{B}(j)$ is an $\left(\left(-r_{1}\right) \times\left(r_{2}-r_{1}\right)\right)$ matrix. $\mathbf{0}_{-\mathbf{r}_{1}, \mathbf{r}_{2}}$ is an $\left(-r_{1} \times r_{2}\right)$ matrix with all its elements equal to zero. $\mathbf{0}_{-\mathbf{r}_{1}, \mathbf{r}_{2}}$ contains the coefficients of the causal part of the inverse impulse response. $\mathbf{B}_{1}(j)$ is an $\left(\left(-r_{1}\right) \times\left(r_{1}\right)\right)$ matrix containing the coefficients of the anticausal part of the impulse response:
$\mathbf{B}_{\mathbf{1}}(j)=\left[\begin{array}{cccccccc}0 & \cdots & 0 & 0 & c_{-q,-j+1} & \cdots & c_{j-2, q-1} & c_{j-1, q} \\ 0 & \cdots & 0 & c_{-q,-j+1} & c_{-q+1,-j+2} & \cdots & c_{j-1, q} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{-q,-j+1} & c_{-q+1,-j+2} & \cdots & c_{j-1, q} & 0 & 0 & \cdots & 0 \\ c_{-q+1,-j+2} & \cdots & c_{j-1, q} & 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{j-1, q} & 0 & 0 & \cdots & 0 & 0 & 0 & 0\end{array}\right]$.
The vector containing the constant terms of the equations has all its elements equal to zero:

$$
\begin{equation*}
\mathbf{b}(j)=\mathbf{0}_{-\mathbf{r}_{1}} \tag{4.30}
\end{equation*}
$$

It should be noted here that the matrices $\mathbf{A}(j)$ and $\mathbf{B}(j)$ as well as the vectors $\mathbf{a}(j)$ and $\mathbf{b}(j)$, all contain cumulants from only one diagonal slice. More specifically it is the


Figure 4.3: Geometric interpretation of $j$-diagonal slices.
diagonal slice

$$
\begin{equation*}
c_{3, x}(-q,-j+1), \ldots, c_{3, x}(j-1, q) \tag{4.31}
\end{equation*}
$$

We call this the $j$-diagonal slice. The $j$-diagonal slices are depicted in figure 4.3. It is possible to build a system of equations involving all the ( $q+1$ ) diagonal slices depicted in figure 4.3. Let $\boldsymbol{\theta}$ be the vector containing the unknown inverse filter coefficients:

$$
\begin{equation*}
\boldsymbol{\theta}=\left[\theta_{r_{1}}, \theta_{r_{1}+1}, \ldots, \theta_{-1}, \theta_{1}, \ldots, \theta_{r_{2}}\right]^{\top} \tag{4.32}
\end{equation*}
$$

Then the system of linear equations can be written in matrix form as follows:

$$
\left[\begin{array}{c}
\mathbf{A}(1)  \tag{4.33}\\
\mathbf{B}(1) \\
\mathbf{A}(2) \\
\mathbf{B}(2) \\
\vdots \\
\mathbf{A}(q+1) \\
\mathbf{B}(q+1)
\end{array}\right] \boldsymbol{\theta}=\left[\begin{array}{c}
\mathbf{a}(1) \\
\mathbf{b}(1) \\
\mathbf{a}(2) \\
\mathbf{b}(2) \\
\vdots \\
\mathbf{a}(q+1) \\
\mathbf{b}(q+1)
\end{array}\right]
$$

Equation 4.33 can now be written as

$$
\begin{equation*}
\mathbf{C}_{1} \theta=\mathbf{c}_{1} \tag{4.34}
\end{equation*}
$$

The matrix $\mathbf{C}_{\mathbf{1}}$ has $\left(r_{2}-r_{1}+q\right)(q+1)$ rows and $\left(r_{2}-r_{1}\right)$ columns. In section 4.2 we
have assumed that $h(0), h(q) \neq 0$. This implies that

$$
\begin{equation*}
c_{3, x}(-q, 0)=\gamma_{3, w} h(0) h^{2}(q) \neq 0 \quad \text { and } \quad c_{3, x}(0, q)=\gamma_{3, w} h^{2}(0) h(q) \neq 0 \tag{4.35}
\end{equation*}
$$

As a consequence of property 4.35 and its structure, the matrix of coefficients generated by the 1-diagonal slice is full rank.

The results presented here can be extended very easily to the fourth-order case. In the next paragraph a new fourth-order cumulant-based method for direct blind deconvolution is presented.

### 4.3.3 Extension to fourth order cumulants

For symmetrically-distributed signals third-order cumulants are zero, and the use of fourth-order cumulants is required. Take equation 4.13 for $m_{1}=m_{2}=m, m, n$, such that at least one is outside $[0, q]$, and $\theta_{0}=1$ :

$$
\begin{equation*}
\sum_{j=r_{1}, j \neq 0}^{r_{2}} \theta_{j} c_{4, x}(m+j, m+j, n+j)=-c_{4, x}(m, m, n) . \tag{4.36}
\end{equation*}
$$

This equation is new and has not been used for blind deconvolution before. The parameterisation in 4.13 has been reduced, since we have taken $m_{1}=m_{2}=m$. Equation 4.36 is now parameterised in a way similar to the third order cumulant equation 4.22. This is very convenient, since all the analysis developed in section 4.3 .2 can be extended directly to fourth-order cumulants if equation $c_{\tau_{1}, \tau_{2}}=c_{3, x}\left(\tau_{1}, \tau_{2}\right)$ is replaced with $c_{\tau_{1}, \tau_{2}}=c_{4, x}\left(\tau_{1}, \tau_{1}, \tau_{2}\right)$. We denote the resulting matrix equation as follows:

$$
\begin{equation*}
\mathrm{D}_{1} \theta=\mathrm{d}_{1} \tag{4.37}
\end{equation*}
$$

The matrix $\mathrm{D}_{1}$ has $\left(r_{2}-r_{1}+q\right)(q+1)$ rows and $\left(r_{2}-r_{1}\right)$ columns.

### 4.3.4 Deconvolution schemes involving second- and third-order cumulants

In many cases it is desirable to involve second-order statistical information along with higher-order statistics in order to recover the inverse filter coefficients. Second-order statistics have lower variance than higher-order statistics so their inclusion can have positive effects on the estimated parameters.

One possibility of introducing second-order statistical information is to use equation 4.11. Zheng et al in [45] have used equation 4.11 along with the equation involving
the ( $\mathrm{q}+1$ )-diagonal third-order cumulant slice. Equation 4.22 allows us to extend the method of [45] to include all the available third-order statistics of the output process. Assuming that $\theta_{0}=1$, equation 4.11 can be written as follows:

$$
\begin{equation*}
\sum_{j=r_{1}, j \neq 0}^{r_{2}} \theta_{j} c_{2, x}(m+j)=-c_{2, x}(m) \tag{4.38}
\end{equation*}
$$

where $m \notin[0, q]$. Equation 4.38 is parametrised only with respect to $m$. After collecting the equations corresponding to $m=-1, \ldots,-r_{2}-q$ and $m=q+1, \ldots,-r_{1}+q$, we obtain a coefficient matrix with the following block structure:

$$
\mathbf{C}^{(2)}=\left[\begin{array}{cc}
\mathbf{A}_{1,1}^{(2)} & \mathbf{A}_{1,2}^{(2)}  \tag{4.39}\\
\mathbf{0}_{\mathbf{r}_{2}+1,-\mathbf{r}_{1}} & \mathbf{A}_{2,2}^{(2)} \\
\mathbf{B}_{1}^{(2)} & \mathbf{0}_{-\mathbf{r}_{1}, \mathbf{r}_{2}}
\end{array}\right] .
$$

The matrices $\mathbf{A}_{\mathbf{1}, \mathbf{1}}^{(\mathbf{2})}, \mathbf{A}_{\mathbf{1}, \mathbf{2}}^{(\mathbf{2})}, \mathbf{A}_{\mathbf{2}, \mathbf{2}}^{(\mathbf{2})}$, and $\mathbf{B}_{\mathbf{1}}^{(\mathbf{2})}$ have the same structure with the matrices $\mathbf{A}_{\mathbf{1}, \mathbf{1}}(q+1), \mathbf{A}_{\mathbf{1 , 2}}(q+1), \mathbf{A}_{\mathbf{2}, \mathbf{2}}(q+1)$, and $\mathbf{B}_{1}(q+1)$ respectively, provided we replace $c_{\tau, \tau}$ with $c_{2, x}(\tau)$. Using the same substitution, we can get the vector of the constant terms of equation 4.38 , which can be written as $\mathbf{c}^{(2)}=\left[\mathbf{a}^{(2)}, \mathbf{b}^{(2)}\right]^{\top}$. The new equations can now be added to the system of third-order cumulant based equations 4.34:

$$
\left[\begin{array}{c}
C_{1}  \tag{4.40}\\
\mathbf{C}^{(2)}
\end{array}\right] \theta=\left[\begin{array}{c}
c_{1} \\
\mathbf{c}^{(2)}
\end{array}\right]
$$

In exactly the same way we can add the second-order cumulant-based equations to system 4.37 which is based on fourth-order cumulants.

There is another way of adding second-order statistical information to the deconvolution schemes developed in the previous section. This can be achieved with equation 4.18 for the third-order statistics case, and with equation 4.19 for the fourth-order statistics case. Let us examine the third-order cumulant case first.

After collecting the coefficients of $\theta_{j}$ 's in equation 4.18 and assuming $\theta_{j}=1$ we obtain the following equation (We also rename the parameters ( $\tau, \tau_{1}$ ) of equation 4.18 as $\left.\left(\tau_{1}, \tau_{2}\right)\right):$

$$
\begin{equation*}
\sum_{j=r_{1}(j \neq 0)}^{r_{2}} \theta_{j} f_{j}\left(\tau_{1}, \tau_{2}\right)=f_{0}\left(\tau_{1}, \tau_{2}\right) \tag{4.41}
\end{equation*}
$$

where the $f_{j}\left(\tau_{1}, \tau_{2}\right)$ is given as follows:

$$
\begin{equation*}
f_{j}\left(\tau_{1}, \tau_{2}\right)=\sum_{i=0}^{q}\left(c_{3, x}\left(i+j, i+j+\tau_{1}\right) c_{2, x}\left(\tau_{2}+i\right)-c_{2, x}(i+j) c_{3, x}\left(\tau_{2}+i, \tau_{1}\right)\right) \tag{4.42}
\end{equation*}
$$

In equation (4.42), $\tau_{1}=-q, \cdots, q$ and $\tau_{2}=-2 q, \cdots, q$. For $\tau_{1}=0$ equation 4.42
involves only ( $q+1$ )-diagonal cumulants and it has been used for deconvolution in [45]. Equation 4.42 allows us to generalise the approach of [45] to include all the available third-order cumulants. It is easy to observe that for $\tau_{2}=-2 q$ and for $\tau_{1}=1, \ldots, q$ equation 4.42 is the same as equation 4.22 for $(m, n)=\left(q+\tau_{1}, q\right)$. Similarly, equation 4.42 for $\tau_{2}=q$ and $\tau_{1}=-q, \ldots,-1$ is the same as equation 4.22 for $(m, n)=\left(0, \tau_{1}\right)$. Consequently, there are $(2 q+1)(3 q+1)-2 q$ new equations resulting from the general equation 4.42. These equations in matrix form can be written as

$$
\begin{equation*}
\mathbf{C}^{(2,3)} \theta=\mathbf{c}^{(2,3)} \tag{4.43}
\end{equation*}
$$

It is interesting to note that the inverse filter parameters involved in equation 4.42 are the parameters $\theta_{-2 q}, \ldots, \theta_{-1}, \theta_{1}, \ldots, \theta_{q}$. The rest of the inverse filter coefficients are obtained using a linear system represented by equation 4.34 , so we can combine 4.34 and 4.43 and obtain the following:

$$
\left[\begin{array}{c}
\mathbf{C}_{1}  \tag{4.44}\\
\mathbf{C}^{(2,3)}
\end{array}\right] \theta=\left[\begin{array}{c}
\mathbf{c}_{1} \\
\mathbf{c}^{(2,3)}
\end{array}\right]
$$

For symmetrically-distributed signals, we use equation 4.19 which, after collecting the coefficients of the $\theta_{j}$ 's and assuming $\theta_{0}=1$, becomes

$$
\begin{equation*}
\sum_{j=r_{1}(j \neq 0)}^{r_{2}} \theta_{j} g_{j}\left(\tau_{1}, \tau_{2}, \tau_{3}\right)=g_{0}\left(\tau_{1}, \tau_{2}, \tau_{3}\right), \tag{4.45}
\end{equation*}
$$

where the $g_{j}\left(\tau_{1}, \tau_{2}, \tau_{3}\right)$ is given as follows:

$$
\begin{array}{r}
g_{j}\left(\tau_{1}, \tau_{2}, \tau_{3}\right)=\sum_{i=0}^{q}\left(c_{2, x}(i+j) c_{4, x}\left(i+\tau_{1}, \tau_{2}, \tau_{3}\right)-\right. \\
\left.c_{4, x}\left(i+j, \tau_{2}+i+j, \tau_{3}+i+j\right) c_{2, x}\left(i+\tau_{1}\right)\right) \tag{4.46}
\end{array}
$$

Once again we have renamed the parameters of 4.19 from $\left(\tau, \tau_{1}, \tau_{2}\right)$ to $\left(\tau_{1}, \tau_{2}, \tau_{3}\right)$. Equation 4.46 is not trivial for the following values of $\tau_{1}, \tau_{2}$ and $\tau_{3}$ :

$$
\begin{equation*}
-2 q \leq \tau_{1} \leq q, \quad-q \leq \tau_{2}, \tau_{3} \leq q \tag{4.47}
\end{equation*}
$$

If we set $\tau_{2}=\tau_{3}$, equation 4.45 can be used to form a set of linear equations in a similar way to equation 4.42 . Thus we can use equation 4.45 to obtain a system of $(2 q+1)(3 q+1)-2 q$ linear equations which in matrix form can be written as

$$
\begin{equation*}
\mathbf{D}^{(2,4)} \boldsymbol{\theta}=\mathbf{d}^{(2,4)} . \tag{4.48}
\end{equation*}
$$

The new equations 4.48 can be used together with equations 4.37 to obtain the following
combined system:

$$
\left[\begin{array}{c}
D_{1}  \tag{4.49}\\
D^{(2,4)}
\end{array}\right] \theta=\left[\begin{array}{c}
d_{1} \\
d^{(2,4)}
\end{array}\right]
$$

### 4.3.5 General Comments

Uniqueness of least-squares solutions: As we have seen at the end of section 4.3.2, the inclusion of the linear equations whose coefficients come from the 1-diagonal cumulant slice ensures that matrix $\mathbf{C}_{\mathbf{1}}$ is full rank. Consequently the linear systems $4.33,4.40$ and 4.44 all involve full rank matrices. The uniqueness of the least-squares solutions of the above systems is guaranteed. In practice, the cumulants that are used to construct the linear systems of the previous sections are not known and they have to be estimated. Assuming that the sample estimates are asymptotically consistent, the corresponding estimates of the inverse filter coefficients will also be asymptotically consistent. The uniqueness and consistency of the least-squares solutions of the systems involving fourth-order cumulants (linear systems $4.37,4.49$ ) is also guaranteed by the same arguments. The fourth-order cumulant slice that ensures that the matrices are full rank is the one-dimensional slice $c_{4, x}(-q, 0,0), \ldots, c_{4, x}(0, q, q)$.

Order of the inverse filter: Assuming that $\rho_{1}$ is the maximum modulus of the roots of $H(z)$ located inside the unit circle, and $\rho_{2}$ is the minimum modulus of the roots of $H(z)$ outside the unit circle, then $r_{2}$ should be selected to be proportional to $1 / \log \left(\rho_{1}\right)$ and $r_{1}$ should be proportional to $-1 / \log \left(\rho_{2}\right)[14,56]$. The algorithms studied in the previous sections are relatively insensitive to the choice of $r_{1}$ and $r_{2}$ as long as they are taken to be large (in absolute value) enough. In fact equations 4.22 for example, are completely insensitive to the selection of $r_{1}$ and $r_{2}$, as long as for the parameters ( $m, n$ ) the following condition holds: supposing that the points ( $x_{1}, y_{1}$ ) and ( $x_{2}, y_{2}$ ) define the segment that results from the intersection of the segment defined by the points ( $m+r_{1}, n+r_{1}$ ) and ( $m+r_{2}, n+r_{2}$ ) with the domain of support of third-order cumulants of the MA(q) process, then

$$
\begin{align*}
m+r_{1} & <x_{1}<x_{2}<m+r_{2} \\
n+r_{1} & <y_{1}<y_{2}<n+r_{2} . \tag{4.50}
\end{align*}
$$

Similar conditions hold for the second- and fourth-order equations 4.38 and 4.36 respectively. Equations 4.42 and 4.45 are completely insensitive to the selection of $r_{1}$ and $r_{2}$ as long $r_{1}$ is selected to be smaller than $-2 q$ and $r_{2}$ bigger that $q$. When $-r_{1}$ and $r_{2}$ are not sufficiently large, there will be a relatively significant residual in the equations which does not satisfy the above conditions.

A common assumption for all the algorithms of this chapter is that $\theta_{0}=1$. This assumption implies that the underlying system is mixed or minimum phase. For maximumphase systems, as we have seen in section 4.2, the inverse impulse response is purely anti-causal. All the algorithms can easily be converted to deal with maximum-phase systems if the assumption $\theta_{0}=1$ is replaced with $\theta_{-1}=1$.

Model order $q$ : In this chapter we have assumed that we are given the true model order $q$. MA model order selection techniques have been examined in chapter 3. For maximum phase systems, the model order determines the exact structure of the linear systems and affects the estimated inverse impulse response. It is interesting to note though, that for minimum phase systems the structure of the systems does not depend closely on the true model order and consequently in these cases the estimation of the inverse impulse response is insensitive to model order selection.

Implementation issues: As discussed earlier, if the underlying system has zeros very near the unit circle, then the inverse model orders $-r_{1}$ and/or $r_{2}$ can get very big. In such a case, the linear systems examined in the previous sections, can become forbiddingly large. Fortunately when this happens these linear systems become very sparse and can be implemented using sparse matrices techniques.

### 4.4 Indirect Deconvolution

Indirect deconvolution is a two-stage procedure. The first stage involves the estimation of the system parameters $h(i) i=0, \ldots, q$. For this purpose one can use the methods developed in chapter 3 or other methods available in the literature [ $63,90,68,37$ ]. The second stage involves the calculation of the inverse filter coefficients $\theta_{j}$ for $j=$ $r_{1}, \ldots, r_{2}$. The inverse filter coefficients can be obtained as the least-squares solution of the following system of equations:

$$
\begin{equation*}
\sum_{l=\max \{0, k-q\}}^{\min \left\{k, r_{2}-r_{1}\right\}} \theta_{l} h(k-l)=\delta\left(k+r_{1}\right) \quad k=0, \ldots, r_{2}-r_{1}+q . \tag{4.51}
\end{equation*}
$$

In matrix notation we have

$$
\begin{equation*}
\mathbf{B}_{\mathrm{inv}} \boldsymbol{\theta}_{\boldsymbol{*}}=\mathbf{b}_{\mathrm{inv}}, \tag{4.52}
\end{equation*}
$$

where $\boldsymbol{\theta}_{*}$ is the vector $\left[\theta_{r_{1}}, \ldots, \theta_{-1}, \theta_{0}, \theta_{1}, \ldots, \theta_{r_{2}}\right]$. Finally we normalise with respect to $\theta_{0}$.

### 4.5 System Identification Through Deconvolution

The equations developed in section 4.3.1 can also be used to estimate the filter parameters $h(i) \quad i=0, \ldots, q$. More specifically, assuming we have estimated the inverse filter parameters $\theta_{j}$ for $j=-2 q, \ldots, q$ we can use equation 4.11 to obtain directly the system impulse response $h(i)$ for $i=0, \ldots, q$. However this approach is very sensitive to additive coloured noise and so for low SNR a method based on equation 4.12 is now developed. Consider equation (4.12) for all the $m$ and $n$ in the range between 0 and $q$. Equation 4.12 can now be rewritten as follows:

$$
\begin{equation*}
\sum_{j=-2 q}^{q} \theta_{j} c_{3, x}(m+j, n+j)=\gamma_{3, w} h(m) h(n) \quad m, n \in[0, q] \tag{4.53}
\end{equation*}
$$

We can now form a matrix $\mathbf{D}$ with the products $\gamma_{3, x} b_{m} b_{n}$ calculated from equation (4.53):

$$
\mathbf{H}=\left[\begin{array}{cccc}
\gamma_{3, w} h^{2}(0) & \gamma_{3, w} h(0) h(1) & \cdots & \gamma_{3, w} h(0) h(q)  \tag{4.54}\\
\gamma_{3, w} h(0) h(1) & \gamma_{3, w} h^{2}(1) & \cdots & \gamma_{3, w} h(1) h(q) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{3, w} h(0) h(q) & \gamma_{3, w} h(1) h(q) & \cdots & \gamma_{3, w} h^{2}(q)
\end{array}\right]
$$

It is obvious from the structure of $\mathbf{H}$ that $\operatorname{rank}(\mathbf{H})=1$. In fact we have

$$
\mathbf{H}=\gamma_{3, w}\left[\begin{array}{c}
h(0)  \tag{4.55}\\
h(1) \\
\vdots \\
h(q)
\end{array}\right] \cdot\left[\begin{array}{llll}
h(0) & h(1) & \cdots & h(q)
\end{array}\right]
$$

In theory taking the SVD of $\mathbf{H}$ would result in only one nonzero singular value. In practice though, there will be more than one nonzero singular value. In this case we take the most dominant singular value and we use the SVD to perform rank reduction on $\mathbf{H}$. Say $\mathbf{H}=\mathbf{U} \Sigma \mathbf{V}^{\top}$. Then if $\mathbf{u}_{\mathbf{1}}$ is the first column of $\mathbf{U}, \mathbf{v}_{\mathbf{1}}$ is the first of $\mathbf{V}$ and $\sigma_{1}$ is the dominant singular value of $H$ we have

$$
\left[\begin{array}{llll}
h(0) & h(1) & \cdots & h(q) \tag{4.56}
\end{array}\right]=r \cdot \mathbf{v}_{\mathbf{1}}^{\top}
$$

where $r$ is a constant. After normalising so that $h(0)=1$ we find the system parameters (we must note here that sometimes this normalisation may lead to numerical instability due to division with a very small number). From the previous we see that in order to find the system parameters we need only to know the inverse system parameters $\theta_{j}$ for $j=-2 q, \ldots, q$. This means that the estimation of the $h(i)$ 's is immune to errors in the estimation of the $\theta_{j}$ 's for $j=r_{1}, \ldots,-2 q-1$ and for $j=q, \ldots, r_{2}$.

The above procedure can easily be extended to the fourth-order cumulant case. Con-
sider equation (4.13) for $m_{1}=0$ and $^{3}$ for all $m_{2}$ and $n$ in the range between 0 and $q$.

$$
\begin{equation*}
\sum_{j=-2 q}^{q} \theta_{j} c_{4, x}\left(j, m_{2}+j, n+j\right)=\gamma_{4, w} h(0) h\left(m_{2}\right) h(n) \quad m_{2}, n \in[0, q] . \tag{4.57}
\end{equation*}
$$

Assuming that we have already estimated $\theta_{-2 q}, \ldots, \theta_{q}$, we can use the right-hand side of equation 4.57 to form the following matrix:

$$
\mathbf{H}_{(4)}=\left[\begin{array}{cccc}
\gamma_{4, w} h^{3}(0) & \gamma_{4, w} h^{2}(0) h(1) & \cdots & \gamma_{4, w} h^{2}(0) h(q)  \tag{4.58}\\
\gamma_{4, w} h^{2}(0) h(1) & \gamma_{4, w} h(0) h^{2}(1) & \cdots & \gamma_{4, w} h(0) h(1) h(q) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{4, w} h^{2}(0) h(q) & \gamma_{4, w} h(0) h(1) h(q) & \cdots & \gamma_{4, w} h(0) h^{2}(q)
\end{array}\right]
$$

Once again $\operatorname{rank}\left(\mathbf{H}_{(4)}\right)=1 . \mathbf{H}_{(4)}$ can now be written as

$$
\mathbf{H}_{(4)}=\gamma_{4, w} h(0)\left[\begin{array}{c}
h(0)  \tag{4.59}\\
h(1) \\
\vdots \\
h(q)
\end{array}\right] \cdot\left[\begin{array}{llll}
h(0) & h(1) & \cdots & h(q)
\end{array}\right] .
$$

The system parameters are then obtained using exactly the same procedure as the third-order cumulant case (equation 4.56 ).

The standard method to obtain the system parameters from the inverse system is to perform a straightforward inversion procedure in the frequency or time domain (similar to the method of section 4.4). The advantage of the present method is that the application of the SVD can potentially smooth out the effects of errors in the inverse impulse response. Furthermore the present method involves only a portion of the inverse impulse response from $\theta_{-2 q}, \ldots, \theta_{q}$ which usually has smaller length than the full inverse impulse response. This can reduce the computational requirements of having to deal with a very long impulse response in time domain calculations.

### 4.6 Asymptotic Performance Analysis

The asymptotic performance of the deconvolution methods developed in this section can be analysed using the theorems 3.2 and 3.3 of section 3.7.

Asymptotic covariance expressions for the estimates of direct deconvolution: The deconvolution methods developed in section 4.3 use the output cumulants to build

[^13]a system of linear equations which is then solved using the least-squares method. Here we develop analytic expressions for the asymptotic covariance matrix of the random vector $\hat{\boldsymbol{\theta}}$ obtained from equation 4.34. The corresponding expressions for the rest of the direct deconvolution methods can be derived using a similar procedure. The method resulting in equation 4.34 involves only third-order cumulants. We collect the thirdorder cumulants corresponding to the minimally sufficient set of lags in the following vector:
\[

$$
\begin{equation*}
\mathbf{s}_{3}=\left[c_{3, x}(0,0), c_{3, x}(0,1), \ldots, c_{3, x}(q, q)\right]^{\top} \tag{4.60}
\end{equation*}
$$

\]

The least-squares solution of 4.34 is expressed as

$$
\begin{equation*}
\hat{\boldsymbol{\theta}}=\left(\mathbf{C}_{\mathbf{1}}^{\top} \mathbf{C}_{\mathbf{1}}\right)^{-1} \mathbf{C}_{\mathbf{1}}^{\top} \mathbf{c}_{\mathbf{1}} \tag{4.61}
\end{equation*}
$$

According to theorem 3.3, the Jacobian of the transformation 4.61 is given by

$$
\begin{equation*}
\mathbf{G}_{\mathbf{C}_{1}}=\left[\mathbf{C}_{\mathbf{1}}^{\top} \mathbf{C}_{\mathbf{1}}\right]^{-1} \mathbf{C}_{\mathbf{1}}^{\top} \mathbf{D}_{\mathbf{C}_{\mathbf{1}}} \tag{4.62}
\end{equation*}
$$

where the $i^{\text {th }}$ column of $\mathbf{D}_{\mathbf{C}_{1}}$ is given by

$$
\begin{equation*}
\mathbf{D}_{\mathbf{C}_{1}, i}=\frac{\partial \mathbf{c}_{\mathbf{1}}}{\partial \mathbf{s}_{\mathbf{3}}(i)}-\frac{\partial \mathbf{C}_{\mathbf{1}}}{\partial \mathbf{s}_{\mathbf{3}}(i)} . \tag{4.63}
\end{equation*}
$$

In particular, if we assume that $\mathrm{s}_{3}(i)=c_{3, x}\left(\tau_{1}, \tau_{2}\right)$ and that $\mathcal{L}_{3, q}(a, b)=(c, d)$ is a mapping such that $c_{3, x}(a, b)=c_{3, x}(c, d)$ and $0 \leq c \leq d \leq q$, we have the following expression for the $k^{\text {th }}$ element of $\mathbf{D}_{\mathbf{C}_{1, i}}$ :

$$
\begin{align*}
& \mathbf{D}_{\mathbf{C}_{1}(k, i)}=\frac{\partial c_{3, x}(m, n)}{\partial c_{3, x}\left(\tau_{1}, \tau_{2}\right)}-\sum_{j=r_{1}(j \neq 0)}^{r_{2}} \theta_{j} \frac{\partial c_{3, x}(m+j, n+j)}{\partial c_{3, x}\left(\tau_{1}, \tau_{2}\right)}= \\
& \begin{cases}0 & \left(\tau_{1}, \tau_{2}\right) \notin\left\{\mathcal{L}_{3, q}(m+j, n+j): j=r_{1}, \ldots, r_{2}\right\} \\
1 & \left(\tau_{1}, \tau_{2}\right)=\mathcal{L}_{3, q}(m, n) \\
-\theta_{i} & \left(\tau_{1}, \tau_{2}\right)=\mathcal{L}_{3, q}(m+i, n+i) \quad i=r_{1}, \ldots, r_{2}, i \neq 0 .\end{cases} \tag{4.64}
\end{align*}
$$

Finally, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is given by

$$
\begin{equation*}
\lim _{N \rightarrow \infty} N \operatorname{cov}\left\{\hat{\boldsymbol{\theta}}^{\top} \hat{\boldsymbol{\theta}}\right\}=\mathbf{G}_{\mathbf{C}_{\mathbf{1}}} \boldsymbol{\Sigma}\left(\hat{\mathbf{s}_{\mathbf{3}}}\right) \mathbf{G}_{\mathrm{C}_{1}}{ }^{\top} . \tag{4.65}
\end{equation*}
$$

Asymptotic covariance expressions for the estimates of indirect deconvolution: The asymptotic expressions of the estimates of indirect deconvolution are derived in a similar way as those of direct deconvolution. The multi-stage nature of indirect deconvolution translates to a cascade of Jacobian transformations of the asymptotic covariance matrix of the sample cumulants.

As we have seen in section 4.4, the solution of equation 4.52 is normalised with respect
to $\theta_{0}$. The Jacobian of this transformation is defined as

$$
\mathbf{G}_{n o r m}=\frac{1}{\theta_{0}}\left[\begin{array}{ccccccccc}
1 & 0 & 0 & \cdots & 0 & -\frac{\theta_{r_{1}}}{\theta_{0}} & 0 & \cdots & 0  \tag{4.66}\\
0 & 1 & 0 & \cdots & 0 & -\frac{\theta_{r_{1}+1}}{\theta_{0}} & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & 1 & -\frac{\theta_{-1}}{\theta_{0}} & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 & -\frac{\theta_{1}}{\theta_{0}} & 1 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & 0 & -\frac{\theta_{r_{2}}}{\theta_{0}} & 0 & \cdots & 1
\end{array}\right] .
$$

$\mathbf{G}_{\text {norm }}$ is a $\left(r_{2}-r_{1}\right) \times\left(r_{2}-r_{1}+1\right)$ matrix.

The Jacobian corresponding to the solution of equation 4.52 is given by the following expression

$$
\begin{equation*}
\mathbf{G}_{\mathbf{i n v}}=\left[\mathbf{B}_{\mathbf{i n v}}{ }^{\top} \mathbf{B}_{\mathbf{i n v}}\right]^{-1} \mathbf{B}_{\mathbf{i n v}}{ }^{\top} \mathbf{D}_{\mathbf{i n v}}, \tag{4.67}
\end{equation*}
$$

where the ( $i, j$ ) element of $\mathbf{D}_{\text {inv }}$, is

$$
\begin{gather*}
\mathbf{D}_{\mathbf{i n v}(i, j)}=\frac{\partial \delta\left(i-1+r_{1}\right)}{\partial h(j)}-\sum_{l=\max (0, i-1-q)}^{\min \left(i-1, r_{2}-r_{1}\right)} \frac{\partial \theta_{l} h(i-1-l)}{\partial h(j)}= \\
\begin{cases}-\theta_{i-j-1} & (i \leq q+1, \quad j \leq i-1) \quad \text { or } \quad\left(q+2 \leq i \leq r_{2}-r_{1}\right) \\
0 & \text { or }\left(r_{2}-r_{1}+1 \leq i \leq r_{2}-r_{1}+q, \quad j \geq r_{2}-r_{1}-i+1\right) \\
0 & \text { otherwise. }\end{cases} \tag{4.68}
\end{gather*}
$$

Finally, the asymptotic covariance of the inverse filter coefficients obtained using indirect deconvolution is given by

$$
\begin{equation*}
\lim _{N \rightarrow \infty} N \operatorname{cov}\left\{\boldsymbol{\theta}^{\top} \boldsymbol{\theta}\right\}=\mathbf{G}_{n o r m} \mathbf{G}_{i n v} \mathbf{G}_{s y s I D} \boldsymbol{\Sigma}\left(\hat{\mathbf{s}_{3}}\right) \mathbf{G}_{s y s I D}^{\top} \mathbf{G}_{i n v}^{\top} \mathbf{G}_{n o r m}^{\top}, \tag{4.69}
\end{equation*}
$$

where $\mathbf{G}_{\text {sysID }}$ is the Jacobian matrix corresponding to the system identification method used to derive the filter parameters $h(0), \ldots, h(q)$.

### 4.7 Numerical Simulations

In this section numerical experiments are performed to demonstrate the performance of the methods developed in this chapter. Random signals are generated according to the following signal models:

## Model 1

$$
\begin{gathered}
x(n)=w(n)-2.4 w(n-1)+0.8 w(n-2) \\
y(n)=x(n)+v(n)
\end{gathered}
$$

The zeros of the system transfer function $H(z)$ are located at 2 and 0.4 . The inverse filter model order is set to $r_{1}=-10$ and $r_{2}=8$.

## Model 2

$$
\begin{gathered}
x(n)=w(n)-0.8 w(n-1)-0.86 w(n-2)+0.768 w(n-3)+1.0205 w(n-4) \\
y(n)=x(n)+v(n)
\end{gathered}
$$

The zeros of the system transfer function $H(z)$ are located at $1.1 \pm j 0.6$ and $-0.7 \pm j 0.4$. The inverse filter model order is set to $r_{1}=-30$ and $r_{2}=25$.

## Model 3

$$
\begin{gathered}
x(n)=w(n)+0.3 w(n-1)-0.4 w(n-2) \\
y(n)=x(n)+v(n)
\end{gathered}
$$

The zeros of the system transfer function $H(z)$ are located at 0.8 and 0.5 . The inverse filter model order is set to $r_{1}=0$ and $r_{2}=20$.

Similarly to the previous section, the input signal $w(n)$ used in the simulations involving third-order cumulants is a zero-mean exponentially distributed IID noise sequence with $\sigma_{w}^{2}=1$ and $\gamma_{3, w}=2$. In simulations involving fourth-order cumulants the input signal $w(n)$ is an IID noise sequence distributed according to a Laplace distribution with parameter $l=1$, and $\sigma_{w}^{2}=2, \gamma_{3, w}=0$ and $\gamma_{4, w}=24$. Additive coloured noise is created as the output of the following MA(4) model:

$$
\begin{equation*}
v(n)=0.5 u(n)-0.25 u(n-1)-0.5 u(n-2)+0.25 u(n-3)-0.25 u(n-4) \tag{4.70}
\end{equation*}
$$

where the input sequence is an IID Gaussian sequence. We define the signal-to-noise ratio as $S N R(d B)=10 \log \left(P_{x} / P_{v}\right)$ where $P_{x}$ denotes the power of the signal. The accuracy of deconvolution is assessed by calculating the Mean Square Error (MSE):

$$
M S E=\frac{\sum_{i=0}^{q}\left(\theta_{n}-\hat{\theta}_{m}\right)^{2}}{\sum_{i=0}^{q} \theta_{n}^{2}}
$$

where $\hat{\theta}_{m}$ is the estimated impulse response corresponding to time $m$. The MSE can also be expressed in dB's as $10 \log _{10}(M S E) \mathrm{dB}$ 's.

### 4.7.1 Third-order cumulant-based deconvolution

The deconvolution methods used in this section are defined as follows:

Method 1: This method involves system identification using method M1 of chapter 3. The inverse system parameters are obtained according to the method described in section 4.4.

Method 2: This is the third-order cumulant based method described in section 4.3.2. This method is similar to that described in [39].

Method 3: This is the second- and third-order cumulant method described by equation 4.40.

Method 4: This is the second- and third-order cumulant method described by equation 4.44.

Method 5: This is a second- and third-order cumulant method described by the special case of equation 4.44 which invoves only main diagonal cumulants. This is the the same as the method first presented in [45].

Method 6: This method is also similar to method 5 but it includes the equations 4.41 for $\tau_{1}=-1, \cdots, 1$ and $\tau_{2}=0$.

The first example involves signal model 1. The number of output samples is 1000 and two experiment were performed, each one involving 50 Monte Carlo runs. The SNR during the first experiment is 50 dB while in the second experiment is 10 dB . Figure 4.4 depicts the results of individual deconvolution methods for $\mathrm{SNR}=50 \mathrm{~dB}$. The graphs in 4.4 show the true inverse impulse response, the average of the estimated impulse response and the average plus/minus the standard deviation of the estimates. To facilitate comparison the standard deviations of the estimated impulse response obtained from all methods and the MSE are shown in the graphs of figure 4.6. All methods perform well in this experiment but the best results are achieved with the indirect deconvolution method which is marked as Method 1. The results for $\mathrm{SNR}=10 \mathrm{~dB}$ are shown in figures 4.5 and 4.7. The variance of the estimates is higher for all deconvolution methods because of the effects of the noise. The method of [45] (Method 5) is more seriously affected by the lower SNR. It is also interesting to note that under $S N R=10 \mathrm{~dB}$ the performance of Method 1 is no longer better than that of the other methods.

This example involves channel model 3 . This is a minimum phase system and the purpose of this example is to demonstrate the robustness of deconvolution methods to overestimation of the model order. The model order in this example is assumed to be 4 instead of 2 . The results are shown in graph 4.8. 50 output sequences of 5000 samples are generated with $\mathrm{SNR}=10 \mathrm{~dB}$. Despite the overestimation of the model order and the low SNR, both methods perform well in the estimation of the inverse impulse response as predicted by the theory.


Figure 4.4: Deconvolution results for signal model 1 after 50 Monte Carlo runs. The number of output samples is 1000 samples and the SNR is 50 dB .


Figure 4.5: Deconvolution results for signal model 1 after 50 Monte Carlo runs. The number of output samples is 1000 samples and the SNR is 10 dB .


Figure 4.6: Standard deviation of estimated inverse impulse response and MSE of deconvolution for signal model 1 after 50 Monte Carlo runs. The number of output samples is 1000 samples and the SNR is 50 dB .


Figure 4.7: Standard deviation of estimated inverse impulse response and MSE of deconvolution for signal model 1 after 50 Monte Carlo runs. The number of output samples is 1000 samples and the SNR is 10 dB .


Figure 4.8: Deconvolution results for signal model 3 after 50 Monte Carlo runs. The SNR is 10 dB and the model order is overestimated to be 4 instead of 2 . Output sample size 1000 .

The final example involves signal model 2. The deconvolution of this model requires longer data sequences than signal model 1 , so output sequences of 5000 samples are used in this example. Identification results for $\mathrm{SNR}=50 \mathrm{~dB}$ are shown in the graphs of figure 4.9 and results for $\mathrm{SNR}=10 \mathrm{~dB}$ are shown in figure 4.10 . Methods 2 and 3 which are based on equations 4.20 and 4.38 perform better than methods 4,5 and 6 which are mainly based on equation 4.41. Furthermore, it is clear from figure 4.10 that methods 4,5 and 6 are much more sensitive to additive noise.

### 4.7.2 Fourth-order cumulant based deconvolution

In this section fourth-order cumulant-based methods are examined.

Method 1(4): This the new fourth-order method described in section 4.3.3. It is based solely on the fourth-order cumulant equation 4.36 .

Method 2(4): This is the fourth-order equivalent of Method 3 in the previous section. It combines the fourth-order equation 4.36 with the second-order equation 4.38 .

Method 3(4): This method involves system identification with the method of the previous section 3.6.2 and then inversion to obtain the inverse system parameters.

Method 4(4): This methods involves all equations 4.36, 4.38 and 4.13.

In the first example sequences were generated according to signal model 3. The SNR is set to 10 dB and the model order is overestimated to be 4 . The results are averaged after 50 Monte Carlo runs, each run involving output sequences of 2000 samples. Methods $1(4)$ and 4(4) are used for deconvolution. The results are reported in figure 4.13. As we can see both methods manage to estimate the inverse impulse response quite accurately.

Finally, sequences were generated according to signal model 2 . The number of output samples is 7000 . It is bigger than that of third-order cumulants, in order to compensate for the increase in the variance of the fourth-order cumulants. The results of deconvolution for $\mathrm{SNR}=50 \mathrm{~dB}$ and $\mathrm{SNR}=10 \mathrm{~dB}$ averaged after 50 Monte Carlo runs are reported in figures 4.14 and 4.15 respectively. The performance of Methods $1(4), 2(4)$ and $3(4)$ are the same in terms of their MSE, for both low and high SNR. The estimates obtained from Method 3(4) have higher variance than those obtained from Methods 1(4) and 2(4), but they are less biased. The inclusion of equations 4.36 has a negative effect on the performance of Method 4(4).


Figure 4.9: Deconvolution results for signal model 2 after 50 Monte Carlo runs. The number of output samples is 5000 samples and the SNR is 50 dB .


Figure 4.10: Deconvolution results for signal model 2 after 50 Monte Carlo runs. The number of output samples is 5000 samples and the SNR is 10 dB .


Figure 4.11: Standard deviation of estimated inverse impulse response and MSE of deconvolution for signal model 2 after 50 Monte Carlo runs. The number of output samples is 5000 samples and the SNR is 50 dB .


Figure 4.12: Standard deviation of estimated inverse impulse response and MSE of deconvolution for signal model 2 after 50 Monte Carlo runs. The number of output samples is 5000 samples and the SNR is 10 dB .


Figure 4.13: Deconvolution results for signal model 3 after 50 Monte Carlo runs. The SNR is 10 dB and the model order is overestimated to be 4 instead of 2 . Output sample size 2000 .


Figure 4.14: Deconvolution results for signal model 2 after 50 Monte Carlo runs. The number of output samples is 7000 samples and the SNR is 50 dB .


Figure 4.15: Deconvolution results for signal model 2 after 50 Monte Carlo runs. The number of output samples is 7000 samples and the SNR is 10 dB .

### 4.8 Conclusions

This chapter has considered the problem of blind deconvolution of FIR systems. In order to build methods for blind deconvolution, some new general equations have been derived, which relate the inverse response to the output cumulants. Previously, only special cases of these equations had been derived. The new equations can be combined in various ways to produce deconvolution algorithms. These algorithms involve the solution of linear systems of equations. The structure of the matrices involved in these systems was studied in order to facilitate implementation and, most importantly, to formulate for the first time conditions that ensure the identifiability of the inverse system parameters. Expressions for the asymptotic variance of the inverse system parameters obtained from HOC-based deconvolution have also been derived. Simulation results have been presented which demonstrate the performance of the blind deconvolution algorithms.

## Chapter 5

## MA Cumulant Enhancement

### 5.1 Introduction

One of the great concerns when we apply higher-order statistical techniques to real signals is the higher variance of estimates of the higher-order statistics relative to the variance of estimated second-order statistics. This disadvantage of higher-order statistics is compensated, to some degree, by the fact that the higher the order of the statistics is, the larger the number of available statistics becomes. For example, suppose that we observe a stationary non-Gaussian MA(q) process. The non-redundant set of information bearing second-order statistics (correlation statistics) contains $q+1$ elements. On the other hand, the non-redundant set of information bearing third-order statistics (third-order cumulants) contains $(q+1)(q+2) / 2$ elements. Intuitively, one could deduce that there is some form of internal structure in a set of the third-order cumulants of an MA $(q)$ process. Indeed not all sets of $(q+1)(q+2) / 2$ numbers can be considered to be true third-order cumulants of some MA $(q)$ process. Consequently, there must be some structural properties that characterise MA $(q)$ cumulants. In fact, HOC-based MA parameter estimation methods in the literature, as well as those developed earlier in this thesis, implicitly rely on the internal structure of cumulants. This chapter explicitly focuses on the characteristic properties of higher-order cumulants and study has two objectives:
(i). To gain an understanding of the kind of properties which are sufficient to characterise sets of cumulants. Such a result, apart from its theoretical value would be useful for the realisation of the second objective.
(ii). To be able to "enhance" sets of sample estimates of cumulant statistics.

Sets of sample estimates of cumulant statistics, possess the characteristic properties of sets of true cumulants only approximately. The accuracy of this approximation depends on how accurate are the sample estimates of the cumulants. If we know the characteristic properties of sets of cumulants (objective 1), then we can build a mapping that maps, according to some optimality criterion, the set of sample estimates of cumulants to a set of enhanced cumulants which possess the desired properties.

The enhanced cumulants can then be used for parameter estimation using any of the existing methods.

The chapter starts with the study of the properties of third-order cumulants. This is achieved through a geometric (or vectorial) interpretation of equation 3.24 of chapter 3. The cumulants are then used to form a matrix and the properties of cumulants can be conveniently expressed as properties of a matrix. The sufficiency of the derived properties to characterise third-order cumulants is formally proved. This method is then extended to account for the joint structure of second- and third-order cumulants.

The next subject of this chapter is to study the properties of fourth-order cumulants. For fourth-order cumulants there are two alternative approaches. The first involves the study of the full set of the available fourth-order cumulants. The large number of statistics makes this study difficult and limits its practical value for relatively large model order $q$. Furthermore, the fourth-order cumulant-based methods developed in the previous chapters use only a subset of the available statistics. For these reasons the second approach considers only a subset of the fourth-order cumulant statistics. The properties of this set are derived and expressed in matrix notation.

Finally, the chapter considers the development of mappings corresponding to the properties of the cumulant matrices. The composition of these mappings is then used to develop an iterative algorithm that given a set of sample cumulants and tries to map this set to an enhanced set possessing the theoretical properties. The convergence properties of the iterative algorithm are also analysed. The chapter concludes with the presentation of results obtained from numerical experiments.

### 5.2 The Linear Structure of Third-Order MA Cumulant Slices

The starting point of the analysis of the structure of third-order cumulants is equation 3.24 of chapter 3 which is repeated here for convenience:

$$
\begin{equation*}
\sum_{i}^{q} h(i) h(i+n) c_{3, x}(m, i+\tau)=\sum_{i}^{q} h(i) h(i+m) c_{3, x}(n, i+\tau) . \tag{5.1}
\end{equation*}
$$

Consider equation 5.1 for $m=q$ and then divide both sides with $h(0) h(q)$ we obtain the following:

$$
\begin{equation*}
c_{3, x}(n, \tau)=\sum_{i=0}^{q} \frac{h(i) h(i+n)}{h(0) h(q)} c_{3, x}(q, i+\tau) . \tag{5.2}
\end{equation*}
$$

Equation 5.2 shows how the third-order cumulants of an MA(q) processes, can be expressed as functions of the system parameters and of cumulants one of whose lags equals the model order $q$.

### 5.2.1 Geometric interpretation

Equation (5.2) has an interesting geometric interpretation which reveals the linear structure inherent in the third-order cumulants of MA processes. This becomes more evident after constructing a vector equation with the aid of equation 5.2. Let us define the following vectors of cumulants ${ }^{1}$ :

$$
\begin{equation*}
\mathbf{c}_{\mathbf{n}}=[\underbrace{c_{3, x}(n-q, n), c_{3, x}(n-q+1, n), \ldots, c_{3, x}(q-1, n), c_{3, x}(q, n)}_{2 q-n+1}, \underbrace{0, \ldots, 0}_{n}]^{\top} . \tag{5.3}
\end{equation*}
$$

Also, for $0 \leq d \leq q$ we define as $\mathbf{C}_{\mathbf{q}}^{\mathbf{d}}$ the following vector:

$$
\begin{equation*}
\mathbf{c}_{\mathbf{q}}^{\mathbf{d}}=[\underbrace{0, \ldots, 0}_{d}, \underbrace{c_{3, x}(0, q), c_{3, x}(1, q), c_{3, x}(2, q), \ldots, c_{3, x}(q, q)}_{q+1}, \underbrace{0, \ldots, 0}_{q-d}]^{\top} . \tag{5.4}
\end{equation*}
$$

The vectors $\mathbf{c}_{\mathbf{q}}^{\mathbf{d}}$ and $\mathbf{c}_{\mathbf{n}}$ have $2 q+1$ elements each. Equation (5.2) can be written in


Figure 5.1: The one-dimensional cumulant slices and their corresponding vectors.

[^14]vector form as,
\[

$$
\begin{equation*}
\mathbf{c}_{\mathbf{n}}=\sum_{i=0}^{q} \frac{h(i) h(i+n)}{h(0) h(q)} \mathbf{c}_{\mathbf{q}}^{\mathbf{q}-\mathbf{n - i} \mathbf{i}} . \tag{5.5}
\end{equation*}
$$

\]

We observe, that

$$
\begin{equation*}
\mathcal{B}=\left\{\mathbf{c}_{\mathbf{q}}^{\mathbf{0}}, \mathbf{c}_{\mathbf{q}}^{\mathbf{1}}, \ldots, \mathbf{c}_{\mathbf{q}}^{\mathbf{q}}\right\} \tag{5.6}
\end{equation*}
$$

is a set of linearly independent vectors. Equation (5.5) also shows that

$$
\begin{equation*}
\mathbf{c}_{\mathbf{n}} \in \operatorname{span}(\mathcal{B}) \quad n=0, \ldots, q . \tag{5.7}
\end{equation*}
$$

Due to the symmetries of third-order cumulants, we need only consider the vectors $\mathbf{c}_{\mathbf{n}}$ for $n=q, \ldots, 0$, each one containing cumulants belonging to a one-dimensional thirdorder cumulant slice. These cumulant slices are depicted in figure 5.1. Expression 5.7 shows that the vectors corresponding to different one-dimensional cumulant slices belong to the same $(q+1)$-dimensional subspace and that a basis of that subspace contains vectors of cumulants taken from a single one-dimensional cumulant slice. This subspace interpretation of equation 5.5 provides the motivation to organize the cumulants in a matrix form as follows:

$$
\begin{equation*}
\mathbf{C}_{3, \mathbf{q}}=\left[\mathbf{c}_{\mathbf{q}}, \mathbf{c}_{\mathbf{q}}^{1}, \mathbf{c}_{\mathbf{q}}^{2}, \ldots, \mathbf{c}_{\mathbf{q}}^{\mathbf{q}}, \mathbf{c}_{\mathbf{q}-1}, \mathbf{c}_{\mathbf{q}-1}, \ldots, \mathbf{c}_{1}, \mathbf{c}_{\mathbf{0}}\right]^{\top} \tag{5.8}
\end{equation*}
$$

The detailed structure of $\mathbf{C}_{\mathbf{3 , q}}$ is shown in figure 5.2: The matrix $\mathbf{C}_{\mathbf{3}, \mathrm{q}}$ has dimension

$$
\left[\begin{array}{ccccccccc}
c_{0, q} & c_{1, q} & \cdots & c_{q-1, q} & c_{q, q} & 0 & 0 & \cdots & 0 \\
0 & c_{0, q} & \cdots & c_{q-2, q} & c_{q-1, q} & c_{q, q} & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & c_{0, q} & c_{1, q} & \cdots & c_{q-1, q} & c_{q, q} & 0 \\
0 & \cdots & 0 & 0 & c_{0, q} & c_{1, q} & \cdots & c_{q-1, q} & c_{q, q} \\
c_{-1, q-1} & c_{0, q-1} & c_{1, q-1} & \cdots & c_{q-1, q-1} & c_{q, q-1} & 0 & \cdots & 0 \\
c_{-2, q-2} & c_{-1, q-2} & c_{0, q-2} & \cdots & c_{q-2, q-2} & c_{q-1, q-2} & c_{q, q-2} & \cdots & 0 \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
c_{q-1,0} & c_{q-1,1} & c_{q-1,2} & \cdots & c_{q-1,2 q-4} & c_{q-1,2 q-3} & c_{q-1,2 q-2} & c_{q-1,2 q-1} & 0 \\
c_{-q, 0} & c_{-q+1,0} & c_{-q+2,0} & \cdots & c_{q-4,0} & c_{q-3,0} & c_{q-2,0} & c_{q-1,0} & c_{q, 0}
\end{array}\right]
$$

Figure 5.2: The detailed structure of $\mathbf{C}_{3, \mathbf{q}} \cdot\left(c_{\tau_{1}, \tau_{2}}\right.$ is used to denote $\left.c_{3, x}\left(\tau_{1}, \tau_{2}\right)\right)$.
$(2 q+1) \times(2 q+1)$. The matrix $\mathbf{C}_{3, \mathbf{q}}$ contains all the third-order cumulants in the principle domain $\mathcal{D}$, (where $\mathcal{D}=\left\{c_{3, x}\left(\tau_{1}, \tau_{2}\right): 0 \leq \tau_{1} \leq \tau_{2} \leq q\right\}$ ). There are $(q+1)(q+$ 2)/ 2 cumulants in $\mathcal{D}$. Clearly, matrix $\mathbf{C}_{3, q}$, possesses the following two properties:

- The rank of $\mathbf{C}_{\mathbf{3 , q}}$ is $q+1$. This is a direct consequence of expression 5.7
- C has some particular structural characteristics. For example, as we can see from 5.8 , there are certain elements of $\mathbf{C}_{3, \mathbf{q}}$ which are zero. In addition, the non-zero elements of $\mathbf{C}_{\mathbf{3}, \mathbf{q}}$ are interrelated in a manner dictated by the symmetries of the third-order cumulants. Because of these symmetries, different one-dimensional cumulant slices share some of their elements.

The structural characteristics of the matrix $\mathbf{C}_{3, \mathbf{q}}$ can be put now in a more formal context. The following definition is required (taken from [91]):

Definition 5.1 Let $\alpha_{i, j}\left(\xi_{1}, \xi_{2}, \ldots, \xi_{p}\right)$ for $1 \leq i \leq m$ and $1 \leq j \leq n$ be a given set of functions which depend on the parameters $\left(\xi_{1}, \xi_{2}, \ldots, \xi_{p}\right)$ in which $p \leq m n$. Furthermore, let the class $\mathcal{A}$ consist of all $m \times n$ matrices whose components are governed by the functional relationship

$$
\begin{align*}
X(i, j) & =\alpha_{i, j}\left(\xi_{1}, \xi_{2}, \ldots, \xi_{p}\right) \\
& =\alpha_{i, j}(\xi) \tag{5.9}
\end{align*}
$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. The matrix class $\mathcal{A}$ is said to have a structure induced by the functions $\alpha_{i, j}(\boldsymbol{\xi})$ and to have " $p$ " degrees of freedom. If the functions $\alpha_{i, j}(\boldsymbol{\xi})$ are linear, the matrix set $\mathcal{A}$ is said to have a linear structure.

The vector $\boldsymbol{\xi}$ is the generating or principal parameter vector of the linear structured matrix. The vector $\boldsymbol{\xi}_{\mathbf{3}}$ containing the third-order cumulants, corresponding to the minimally sufficient set of lags, plays the role of vector $\boldsymbol{\xi}$ in the construction of matrix $C_{3, q}$ i.e.

$$
\xi_{3}=\left[c_{3, x}(0,0), c_{3, x}(0,1), \ldots, c_{3, x}(q, q)\right]^{\top}
$$

The functions $\alpha_{i, j}\left(\boldsymbol{\xi}_{3}\right)$, corresponding to matrix $\mathbf{C}_{3, \mathrm{q}}$ are all linear since for some of the $(i, j)$ the functions are constant $\alpha_{i, j}\left(\xi_{3}\right)=0$, and for the rest $(i, j)$ s the functions return an element of the vector $\boldsymbol{\xi}_{\mathbf{3}}$ i.e. $\alpha_{i, j}\left(\boldsymbol{\xi}_{\mathbf{3}}\right)=\xi_{k}$. From the previous discussion, it is obvious that the matrix $\mathbf{C}_{\mathbf{3}, \mathbf{q}}$ has a linear structure induced by the functions $\boldsymbol{a}_{i, j}\left(\boldsymbol{\xi}_{3}\right)$, with " $(q+1)(q+2) / 2$ " degrees of freedom.

The next section deals with the question of whether the rank and structure properties are sufficient to characterise a matrix as being constructed by true third-order cumulants of some MA $(q)$ model.

### 5.2.2 Characteristic properties of third-order MA cumulants

Suppose the matrix $\mathbf{S}$ has both the desired structure and rank properties defined in the previous section. It is interesting to examine whether this matrix consists of real
cumulants of some MA model. The approach adopted in the following is to construct a MA(q) model whose third-order cumulants are the elements of the matrix $\mathbf{S}$.

Since the matrix $\mathbf{S}$ has the same structural characteristics as those of a matrix constructed of real cumulants, then, if $s_{i, j}$ is a non-zero element of $\mathbf{S}$, we denote $s_{i, j}$ as $s\left(\tau_{1}, \tau_{2}\right)$, where ( $\tau_{1}, \tau_{2}$ ) are the lags we associate with the $i, j$-element of a structurallyequivalent matrix which is constructed from real cumulants. Then, because of the structure property, the same symmetries that apply to lags of cumulants will apply to these associated lags of $s\left(\tau_{1}, \tau_{2}\right)$. In the following it is assumed that $s(0, q), s(q, q) \neq 0$. The following Lemma is required:

Lemma 1 Suppose that we are given a $(2 q+1) \times(2 q+1)$ matrix $\mathbf{S}$, which has the two prescribed properties (structure and rank). Then the following equation holds for $s\left(\tau_{1}, \tau_{2}\right)$ :

$$
\begin{equation*}
s(j, n)=\sum_{i=0}^{q} \frac{s(i, q) s(i+n, q)}{s(0, q) s(q, q)} s(j+i, q) \tag{5.10}
\end{equation*}
$$

for $n=0, . . q-1$ and $j=n-q, \ldots, q$.

Proof: The vectors corresponding to the first $q+1$ rows of the matrix are denoted as $\mathbf{s}_{\mathbf{q}}^{\mathbf{d}}$ where $d=0, \ldots, q$ and the vectors corresponding to the last $q$ rows of the matrix are denoted by $\mathbf{s}_{\mathbf{q}-\mathbf{1}}, \ldots, \mathbf{s}_{\mathbf{0}}$. We assume that $s(0, q), s(q, q) \neq 0$, then it is obvious from their structure, that the $q+1$ vectors $\mathbf{s}_{\mathbf{q}}^{\mathbf{d}}, d=0, \ldots, q$ are linearly independent. Given that the rank of the matrix is $q+1$, we can conclude that the vectors corresponding to the last $q$ rows of the matrix, belong to the space spanned by the first $q+1$ rows. In particular, since the last $n$ elements of the vector $\mathbf{s}_{\mathbf{n}}(n=q-1, \ldots, 0)$ are zero, it can easily be seen that they belong to the space spanned only by $\mathbf{s}_{\mathbf{q}}^{\mathbf{d}}$ for $d=0, \ldots, q-n$. We can write this as follows:

$$
\begin{equation*}
\mathbf{s}_{\mathbf{n}} \in \operatorname{span}\left\{\mathbf{s}_{\mathbf{q}}^{\mathbf{0}}, \ldots, \mathbf{s}_{\mathbf{q}}^{\mathbf{q}-\mathbf{n}}\right\} \quad n=q-1, \ldots, 0 \tag{5.11}
\end{equation*}
$$

Now we take $n=q-1$, it is straight forward to prove that,

$$
\begin{equation*}
\mathbf{s}_{\mathbf{q}-\mathbf{1}}=\frac{s(1, q)}{s(0, q)} \mathbf{s}_{\mathbf{q}}^{\mathbf{0}}+\frac{s(q-1, q)}{s(q, q)} \mathbf{s}_{\mathbf{q}}^{\mathbf{1}} \tag{5.12}
\end{equation*}
$$

In scalar form, this translates to ,

$$
\begin{equation*}
s(j, q-1)=\sum_{i=0}^{1} \frac{s(i, q) s(i+q-1)}{s(q, q) s(0, q)} s(j+i, q) \quad j=-1, \ldots, q \tag{5.13}
\end{equation*}
$$

so equation (5.10) holds for $n=q-1$.

Assumption 1: Let us suppose that equation (5.10) holds for every $n$ such that $q-1 \geq$ $n \geq k$, for some $k>0$.
We want to prove that it also holds for $n=k-1$,

$$
\begin{equation*}
\mathbf{s}_{k-1}=\sum_{i=0}^{q-k+1} \lambda_{k-1, i} \mathbf{s}_{q}^{i} \tag{5.14}
\end{equation*}
$$

Because of the cumulant-like symmetries in the lags of $s(k-q-1, k-1)$, the first element of $\mathbf{s}_{k-1}$ is $s(k-q-1, k-1)=s(q-k+1, q)$. It is related with $s(0, q)$ as follows:

$$
\begin{equation*}
s(q-k+1, q)=\lambda_{k-1,0} s(0, q) \tag{5.15}
\end{equation*}
$$

From equation (5.15) we can obtain the value of $\lambda_{k-1,0}=s(q-k+1, q) / s(0, q)$. Since $s(k-q-1, k-1)=(s(q-k+1, q) / s(0, q)) s(0, q)$, equation (5.10) holds for $n=k-1$ and $j=k-q-1$.

Assumption 2: Suppose that equation (5.10) holds for $n=k-1$ and $k-q-1 \leq j \leq m$ where $m \leq-2$.
In other words, we assume that we know that
$s(m, k-1)=\sum_{i=-m}^{q-k+1} \frac{s(i, q) s(i+k-1, q)}{s(0, q) s(q, q)} s(i+m, q)=\sum_{i=-m}^{q-k+1} \lambda_{k-1, q-k+1-i} s(i+m, q)$.

We want to obtain the value of $\lambda_{k-1, m-k+q+2}$ and use this to show that equation (5.10) is valid for $n=k-1$ and $j=m+1$. So,
$s(m+1, k-1)=\sum_{i=-m}^{q-k+1} \frac{s(i, q) s(i+k-1, q)}{s(0, q) s(q, q)} s(i+m+1, q)+\left(\lambda_{k-1, m-k+q+2}\right) s(0, q)$,
but $s(m+1, k-1)=s(-m-1, k-m-2)$, where $k-m-2 \geq k$. Then according to Assumption 1 we have,

$$
\begin{align*}
s(m+1, k-1) & =s(-m-1, k-m-2) \\
& =\sum_{j=0}^{q-k+m+2} \frac{s(j, q) s(j+k-m-2, q)}{s(0, q) s(q, q)} s(j-m-1, q) . \tag{5.18}
\end{align*}
$$

Consequently, the previous equation can now be rewritten as follows,

$$
\begin{array}{r}
s(m+1, k-1)=\sum_{j=1}^{q-k+m+2} \frac{s(j, q) s(j+k-m-2, q)}{s(0, q) s(q, q)} s(j-m-1, q) \\
+\frac{s(0, q) s(k-m-2, q)}{s(0, q) s(q, q)} s(-m-1, q) \tag{5.19}
\end{array}
$$

If we make the transformation $j=i+m+1$ in equation (5.19) we obtain,

$$
\begin{array}{r}
s(m+1, k-1)=\sum_{i=-m}^{q-k+1} \frac{s(i+m+1, q) s(i+k-1, q)}{s(0, q) s(q, q)} s(i, q) \\
+\frac{s(0, q) s(k-m-2, q)}{s(0, q) s(q, q)} s(-m-1, q) \tag{5.20}
\end{array}
$$

Now observe that the summations in equations (5.20) and (5.17) are equal, thus we can deduce that

$$
\lambda_{k-1, m-k+q+2}=\frac{s(k-m-2, q) s(-m-1, q)}{s(0, q) s(q, q)}
$$

and, consequently, equation (5.17) can be rewritten as

$$
\begin{equation*}
s(m+1, k-1)=\sum_{i=-m-1}^{q-k+1} \frac{s(i, q) s(i+k-1, q)}{s(0, q) s(q, q)} s(i+m+1, q) . \tag{5.21}
\end{equation*}
$$

Equation (5.21) demonstrates that equation (5.10) is valid for $j=m+1$ and $n=k-1$. Now, knowing that the initial equation corresponding to $n=k-1$ and $j=k-q-1$ holds, we have demonstrated that we can prove equation (5.10) to be valid for $n=k-1$ and $k-q-1 \leq j \leq-1$. From expression (5.11) we know that,

$$
\begin{equation*}
\mathbf{s}_{k-1}=\sum_{i=0}^{q-k+1} \lambda_{k-1, i} \mathbf{s}_{q}^{i} \tag{5.22}
\end{equation*}
$$

We have already obtained the values of $\lambda_{k-1,0}$ to $\lambda_{k-1, q-k}$, but we still need to find $\lambda_{k-1, q-k+1}$. This is easily obtained if we consider the following expression for the last non-identically zero element of $\mathbf{s}_{k-1}$ :

$$
\begin{gather*}
s(q, k-1)=\lambda_{k-1, q-k+1} s(q, q)  \tag{5.23}\\
\lambda_{k-1, q-k+1}=\frac{s(q, k-1)}{s(q, q)} \tag{5.24}
\end{gather*}
$$

Since we know all the $\lambda$ 's in (5.22), we can now write (5.22) in scalar form:

$$
\begin{equation*}
s(j, k-1)=\sum_{i=0}^{q-k+1} \frac{s(i, q) s(i+k-1, q)}{s(0, q) s(q, q)} s(i+j, q) \quad j=k-1-q, \ldots, q . \tag{5.25}
\end{equation*}
$$

Given that equation (5.10) is valid for $n=q-1$ we have shown that it is valid for every $n$ such that $q-1 \geq n \geq 0$.
Since we know that $s(0, q), s(q, q) \neq 0$, we can find a $\gamma_{3} \neq 0$ such that

$$
\begin{equation*}
s(i, q)=\gamma_{3} \frac{s(0, q)}{s(0, q)} \frac{s(q, q)}{s(0, q)} \frac{s(i, q)}{s(0, q)} \tag{5.26}
\end{equation*}
$$

We combine equations (5.10) and (5.26), we obtain the following:

$$
\begin{array}{r}
s(j, n)=\sum_{i=0}^{q} \frac{s(i, q) s(i+n, q)}{s(0, q) s(q, q)} \gamma_{3} \frac{s(0, q)}{s(0, q)} \frac{s(q, q)}{s(0, q)} \frac{s(i+j, q)}{s(0, q)} \\
=\gamma_{3} \sum_{i=0}^{q} \frac{s(i, q)}{s(0, q)} \frac{s(i+n, q)}{s(0, q)} \frac{s(i+j, q)}{s(0, q)} . \tag{5.27}
\end{array}
$$

Equation (5.27) shows that $s(j, n)$ is the third order cumulant of an MA model with parameters $h(i)=s(i, q) / s(0, q)$. Thus the following theorem holds:

Theorem 5.1 Every $(2 q+1) \times(2 q+1)$ matrix $\mathbf{S}$ possessing the structure and rank properties defined in the previous section, consists of real cumulants of some MA(q) model.

In the previous section it was stated that the matrix $\mathbf{C}_{3, \mathbf{q}}$, whose elements are cumulants of an MA $(q)$ model, possesses the rank and linear structure properties. The above theorem states that the inverse is also true. The above two sentences provide the necessary and sufficient conditions for a set of $(q+1)(q+2) / 2$ numbers ${ }^{2}$ to be the set of third-order cumulants of some MA $(q)$ model. It is interesting to notice here, how the construction of matrices of cumulants allows the expression of cumulant properties in terms of matrix properties which are easier to manipulate and analyse using standard linear algebra. The utility of this matrix representation is exploited later in this chapter, when mappings corresponding to matrix properties are developed.

The next section extends the work on the structure of third-order cumulants and studies second- and third-order cumulants in the same framework.

[^15]
### 5.3 Joint Structure of Second- and Third-Order Cumulants

The joint structure of second- and third-order cumulants can be studied using equation 3.12 for $m=2$ and $n=3$ :

$$
\begin{equation*}
\sum_{j=0}^{q} h(j) c_{3, x}\left(j+\tau_{1}, \tau_{2}\right)=\epsilon_{3,2} \sum_{i=0}^{q} h(i) h\left(i+\tau_{2}\right) c_{2, x}\left(i+\tau_{1}\right) \tag{5.28}
\end{equation*}
$$

where $\epsilon_{3,2}=\gamma_{3, x} / \sigma_{x}^{2}$. After taking $\tau_{2}=q$ and dividing both sides by $h(0) h(q)$, equation 5.28 becomes

$$
\begin{equation*}
c_{2, x}\left(\tau_{1}\right)=\epsilon_{2,3} \sum_{j=0}^{q} \frac{h(j)}{h(0) h(q)} c_{3, x}\left(j+\tau_{1}, q\right) \tag{5.29}
\end{equation*}
$$

where $\epsilon_{2,3}=1 / \epsilon_{3,2}$. Equation 5.29 can be interpreted in a similar way to equation 5.2. The second-order cumulants are used to form the following vector:

$$
\begin{equation*}
\mathbf{c}_{(2)}=\left[c_{2, x}(-q), \ldots, c_{2, x}(0), \ldots, c_{2, x}(q)\right]^{\top} \tag{5.30}
\end{equation*}
$$

Equation 5.29 can now be written in vector form:

$$
\begin{equation*}
\mathbf{c}_{(2)}=\epsilon_{2,3} \sum_{j=0}^{q} \frac{h(j)}{h(0) h(q)} \mathbf{c}_{\mathbf{q}}^{\mathbf{q}-\mathbf{j}}, \tag{5.31}
\end{equation*}
$$

where the vectors $\mathbf{c}_{\mathbf{q}}^{\mathbf{d}}$ are defined by equation 5.4. Equation 5.31 clearly shows that the vector containing the second-order cumulants belongs to the same subspace as the vectors $\mathbf{c}_{\mathbf{n}}, n=0, \ldots, q$ which contain one-dimensional third-order cumulants slices:

$$
\begin{equation*}
\mathbf{c}_{(2)} \in \operatorname{span}(\mathcal{B}) \tag{5.32}
\end{equation*}
$$

where $\mathcal{B}$ is defined by expression 5.6. The matrix $\mathbf{C}_{3, q}$ defined by equation 5.8 can now be extended by including the second-order cumulant vector:

$$
\begin{equation*}
\mathbf{C}_{2,3, \mathbf{q}}=\left[\mathbf{c}_{(2)}, \mathbf{c}_{\mathbf{q}}, \mathbf{c}_{\mathbf{q}}^{\mathbf{1}}, \mathbf{c}_{\mathbf{q}}^{2}, \ldots, \mathbf{c}_{\mathbf{q}}^{\mathbf{q}}, \mathbf{c}_{\mathbf{q}-1}, \mathbf{c}_{\mathbf{q}-1}, \ldots, \mathbf{c}_{1}, \mathbf{c}_{0}\right]^{\top} \tag{5.33}
\end{equation*}
$$

The detailed structure of $\mathbf{C}_{\mathbf{2}, \mathbf{3}, \mathrm{q}}$ is shown in figure 5.3.
The matrix $\mathbf{C}_{\mathbf{2}, \mathbf{3}, \mathbf{q}}$ has dimension $(2 q+2) \times(2 q+1)$ and contains all the second-order statistics $c_{2, x}(\tau), \tau=0, \ldots q$ as well as all the third-order cumulants in the principal domain $\mathcal{D}$, (where $\mathcal{D}=\left\{c_{3, x}\left(\tau_{1}, \tau_{2}\right): 0 \leq \tau_{1} \leq \tau_{2} \leq q\right\}$ ). Similarly with matrix $\mathbf{C}_{3, \mathbf{q}}$ defined in 5.2 , the matrix $\mathbf{C}_{\mathbf{2}, \mathbf{3}, \mathbf{q}}$ possesses the following two properties:

- The rank of $\mathbf{C}_{\mathbf{2}, \mathbf{3}, \mathbf{q}}$ is $q+1$. This is a direct consequence of expressions 5.7 and 5.32.
- $\mathbf{C}_{2, \mathbf{3 , q}}$ has a linear structure with $(q+1)(q+4) / 2$ degrees of freedom. The generating parameter vector is defined as

$$
\begin{equation*}
\boldsymbol{\xi}_{\mathbf{2}}=\left[c_{2, x}(0), \ldots, c_{2, x}(q), c_{3, x}(0,0), c_{3, x}(0,1), \ldots, c_{3, x}(q, q)\right]^{\top} \tag{5.34}
\end{equation*}
$$

The linear functional relationships between the parameter vector $\boldsymbol{\xi}_{2}$ and the elements of matrix $\mathbf{C}_{\mathbf{2}, \mathbf{3}, \mathbf{q}}$ are determined by the symmetries of third-order cumulant lags and the symmetry of negative and positive lags of second-order cumulants.

$$
\left[\begin{array}{ccccccccc}
c_{-q} & c_{-q+1} & \cdots & c_{-1} & c_{0} & c_{1} & \cdots & c_{q-1} & c_{q} \\
c_{0, q} & c_{1, q} & \cdots & c_{q-1, q} & c_{q, q} & 0 & 0 & \cdots & 0 \\
0 & c_{0, q} & \cdots & c_{q-2, q} & c_{q-1, q} & c_{q, q} & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & c_{0, q} & c_{1, q} & \cdots & c_{q-1, q} & c_{q, q} & 0 \\
0 & \cdots & 0 & 0 & c_{0, q} & c_{1, q} & \cdots & c_{q-1, q} & c_{q, q} \\
c_{-1, q-1} & c_{0, q-1} & c_{1, q-1} & \cdots & c_{q-1, q-1} & c_{q, q-1} & 0 & \cdots & 0 \\
c_{-2, q-2} & c_{-1, q-2} & c_{0, q-2} & \cdots & c_{q-2, q-2} & c_{q-1, q-2} & c_{q, q-2} & \cdots & 0 \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
c_{q-1,0} & c_{q-1,1} & c_{q-1,2} & \cdots & c_{q-1,2 q-4} & c_{q-1,2 q-3} & c_{q-1,2 q-2} & c_{q-1,2 q-1} & 0 \\
c_{-q, 0} & c_{-q+1,0} & c_{-q+2,0} & \cdots & c_{q-4,0} & c_{q-3,0} & c_{q-2,0} & c_{q-1,0} & c_{q, 0}
\end{array}\right]
$$

Figure 5.3: The detailed structure of $\mathbf{C}_{\mathbf{2 , 3 , \mathbf { q }}} \cdot\left(c_{\tau_{1}, \tau_{2}}\right.$ is used to denote $c_{3, x}\left(\tau_{1}, \tau_{2}\right)$ and $c_{\tau}$ is used to denote $c_{2, x}(\tau)$ )

Naturally the following question arises: Given a matrix with the same rank and structure properties as those of matrix $\mathbf{C}_{\mathbf{2}, \mathbf{3}, \mathbf{q}}$ is it possible to say that the elements of the given matrix are true second- and third-order cumulants of some MA(q) model? This question has been partially answered in the previous section. Excluding the first row, all the rows of $\mathbf{C}_{\mathbf{2 , 3 , q}}$ were shown to contain true one-dimensional third-order cumulant slices. So the question now reduces to whether the first row can be considered to contain second-order cumulants of the same MA $(q)$ model. The rest of this section deals with this question.

Suppose that a matrix $\mathbf{S}^{\prime}$ has the same rank and structure properties with the matrix $\mathbf{C}_{2,3, \mathbf{q}}$. The first row of $\mathbf{S}^{\prime}$ is denoted as

$$
\begin{equation*}
\mathbf{s}^{\prime}=[s(-q), s(-q+1), \cdots, s(-1), s(0), s(1), \cdots, s(q-1), s(q)] . \tag{5.35}
\end{equation*}
$$

The notation used for the rest of the elements of the matrix is the same as that used used in section 5.2.2. In the following it is assumed that $s(0, q), s(q, q), s(q) \neq 0$. Because the rank of $\mathbf{S}^{\prime}$ is $q+1$, all its rows can be expressed as linear combinations of the
linearly-independent rows $\mathbf{s}_{\mathbf{q}}^{\mathbf{d}}$ where $d=0, \ldots, q$ (rows 2 to $q+2$ ):

$$
\begin{equation*}
\mathbf{s}^{\prime}=\sum_{i=0}^{q} \alpha_{i} \mathbf{s}_{\mathbf{q}}^{\mathbf{i}} \tag{5.36}
\end{equation*}
$$

In scalar form equation 5.36 becomes

$$
\begin{equation*}
s(\tau)=\sum_{i=0}^{q} \alpha_{i} s(i+\tau, q) \tag{5.37}
\end{equation*}
$$

The objective here is find the values of $\alpha_{i} \mathrm{~s}$ so that $s(\tau)$ can be decomposed into its constituent parts. It will then be possible to decide whether $s(\tau)$ for $\tau=0, \ldots, q$ are second-order cumulants of some MA model. By setting $\tau=q$ in equation 5.37 it is possible to obtain the value of $\alpha_{0}$ :

$$
\begin{equation*}
s(q)=\alpha_{0} s(q, q) \Rightarrow \alpha_{0}=\frac{s(q)}{s(q, q)} . \tag{5.38}
\end{equation*}
$$

Similarly, by setting $\tau=-q$ in 5.37 , it is possible to obtain the value of $\alpha_{q}$ :

$$
\begin{equation*}
s(-q)=\alpha_{q} s(0, q) \Rightarrow \alpha_{q}=\frac{s(-q)}{s(0, q)} \Rightarrow \alpha_{q}=\frac{s(q)}{s(0, q)} . \tag{5.39}
\end{equation*}
$$

The rest of the coefficients $\alpha_{i}$ can be calculated recursively. For example the value of $\alpha_{1}$ can be calculated as a function of $\alpha_{0}$ after setting $\tau=q-1$ in 5.37:

$$
\begin{equation*}
s(q-1)=\alpha_{0} s(q-1, q)+\alpha_{1} s(q, q) \Rightarrow \alpha_{1}=s(q-1)-\alpha_{0} s(q-1, q) . \tag{5.40}
\end{equation*}
$$

Unfortunately equation 5.40 does not provide any information on the nature of $s(q-1)$. Alternatively, the following system of equations is formed, based on the fact that $s(\tau)=$ $s(-\tau)$ :

$$
\begin{equation*}
\sum_{i=0}^{q-\tau} \alpha_{i} s(\tau+i, q)=\sum_{i=\tau}^{q} \alpha_{i} s(-\tau+i, q) \quad 0<\tau<q . \tag{5.41}
\end{equation*}
$$

After changing the variable of summation on the right-hand side, equation 5.41 becomes

$$
\begin{equation*}
\sum_{i=0}^{q-\tau} \alpha_{i} s(\tau+i, q)=\sum_{i=0}^{q-\tau} \alpha_{\tau+i} s(i, q) \quad 0<\tau<q . \tag{5.42}
\end{equation*}
$$

The values of $\alpha_{0}$ and $\alpha_{q}$ are known from equations 5.38 and 5.39 and so expression 5.42 can be regarded as a system of $q-1$ linear equations with respect to the $q-1$ unknowns $\alpha_{1}, \ldots, \alpha_{q-1}$. It is easy to observe that

$$
\begin{equation*}
\alpha_{i}=\frac{s(q) s(i, q)}{s(q, 0)} \tag{5.43}
\end{equation*}
$$

is a solution of the system 5.42. This solution extended to $i=0$ and $i=q$ is consistent
with the values of $\alpha_{0}$ and $\alpha_{q}$ given by equations 5.38 and 5.39 . In order for the solution 5.43 to be also a solution of 5.37 , it is required that 5.43 is a unique solution of the linear system 5.42 . For $q \geq 3$, this can only happen when the matrix of the coefficients of the unknowns of the linear system 5.42 is full rank. Suppose this matrix is called $\mathbf{P}$. Then $\mathbf{P}$ can be expressed as

$$
\begin{equation*}
\mathbf{P}=\mathbf{P}_{l e f t}-\mathbf{P}_{r i g h t} \tag{5.44}
\end{equation*}
$$

where $\mathbf{P}_{\text {left }}$ is defined as

$$
\mathbf{P}_{l e f t}=\left[\begin{array}{cccccc}
s(2, q) & \cdots & \cdots & \cdots & \cdots & s(q, q)  \tag{5.45}\\
s(3, q) & \cdots & \cdots & \cdots & s(q, q) & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
s(q-1, q-1) & s(q, q) & 0 & \cdots & \cdots & 0 \\
s(q, q) & 0 & 0 \cdots & \cdots & 0 & 0
\end{array}\right]
$$

and $\mathbf{P}_{\text {right }}$ is defined as

$$
\mathbf{P}_{\text {right }}=\left[\begin{array}{cccccc}
s(0, q) & \ldots & \ldots & \ldots & \cdots & s(q-2, q)  \tag{5.46}\\
0 & s(0, q) & \cdots & \cdots & \cdots & s(q-3, q) \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 0 & s(0, q) & s(1, q) \\
0 & 0 & \cdots & 0 & 0 & s(1, q)
\end{array}\right]
$$

$\mathbf{P}_{l e f t}$ contains the coefficients of $\alpha_{i} \quad i=1, \ldots, q-1$ in the left-hand side of equation 5.42 and $\mathbf{P}_{\text {right }}$ the corresponding coefficients in the right-hand side of 5.42. Although both $\mathbf{P}_{l e f t}$ and $\mathbf{P}_{\text {right }}$ are always ${ }^{3}$ full-rank, $\mathbf{P}$ is not guaranteed to be full-rank. $\mathbf{P}$ can fail to be full rank for a few degenerative cases like for example $s(0, q)=s(1, q)=\cdots=s(q, q)$. It is important to note that it is always to possible to check if $\mathbf{P}$ is full-rank. Assuming that $\mathbf{P}$ is full-rank, then 5.43 is the solution of 5.37 , which can then be re-written as

$$
\begin{equation*}
s(\tau)=\sum_{i=0}^{q} \frac{s(q) s(i, q)}{s(q, 0)} s(i+\tau, q) . \tag{5.47}
\end{equation*}
$$

Taking equation 5.26 into account, equation 5.47 becomes

$$
\begin{align*}
s(\tau) & =\sum_{i=0}^{q} \frac{s(q) s(i, q)}{s(q, 0)} \gamma_{3} \frac{s(0, q)}{s(0, q)} \frac{s(q, q)}{s(0, q)} \frac{s(i+\tau, q)}{s(0, q)} \\
& =\frac{\gamma_{3} s(q) s(q, q)}{s(q, 0)} \sum_{i=0}^{q} \frac{s(i, q)}{s(0, q)} \frac{s(i+\tau, q)}{s(q, 0)} \tag{5.48}
\end{align*}
$$

Equation (5.48) shows that $s(\tau)$ is the second-order cumulant of an MA model with parameters $h(i)=s(i, q) / s(0, q)$.

[^16]In summary, the following theorem holds:

Theorem 5.2 Suppose that a $(2 q+2) \times(2 q+1)$ matrix $\mathbf{S}^{\prime}$ possesses the structure and rank properties of matrix 5.33. Then, if one of the following conditions holds,

- $q \leq 2$,
- $q \geq 3$ and the determinant of the square matrix $\mathbf{P}$ defined by equation 5.44 is non-zero,
then the matrix $\mathbf{S}^{\prime}$ consists of real second- and third-order cumulants of some $M A(q)$ model.

The structure of fourth-order cumulants is now considered in the next section.

### 5.4 The Linear Structure of Fourth-Order MA Cumulant Slices

The structure of fourth-order cumulants of MA models is much richer than that of third-order cumulants and consequently its analysis is more complicated than that of third-order cumulants. In a similar manner to the development of the previous sections, an obvious starting point for the analysis of the structure of the fourth-order cumulants, would be an expression of the inter-relationships between different slices of fourth-order cumulants. Two such expressions have been developed in section 3.6.1. Equation 3.44 is considered first. For $n_{1}=q$ and $m_{1}=m$ equation 3.44 becomes

$$
\begin{equation*}
c_{4, x}\left(m, \tau_{1}, \tau_{2}\right)=\sum_{i=0}^{q} \frac{h(i) h(i+m)}{h(0) h(q)} c_{4, x}\left(q, i+\tau_{1}, i+\tau_{2}\right) . \tag{5.49}
\end{equation*}
$$

Equation 5.49 shows how all fourth-order cumulants can be expressed as a linear combination of a single 2-D fourth-order cumulant slice.

### 5.4.1 Geometric interpretation

In section 5.2.1, the third-order equation involving 1-D cumulant slices was rewritten in a vectorial form. Extending that approach to the fourth-order equation 5.49 which involves 2-D cumulant slices, means rewriting equation 5.49 in a matrix form. The domain of support of the 2-D fourth-order cumulant slice $c_{4, x}\left(m, \tau_{1}, \tau_{2}\right)$, for some $m$ such that $0 \leq m \leq q$, is depicted in figure 5.4. The polygon ABCDEF in figure


Figure 5.4: The domain of support of the 2-D cumulant slice $c_{4, x}\left(m, \tau_{1}, \tau_{2}\right), 0 \leq m \leq q$.
5.4 can be seen as a horizontal slice ${ }^{4}$ from the 3-D domain of support of fourth-order cumulants depicted in figure 2.3. The matrices $\mathbf{D}_{\mathbf{m}}$ for $0 \leq m \leq q$ are constructed with the fourth-order cumulants whose first lag is fixed to $m$. They are defined as follows:

$$
\mathbf{D}_{\mathbf{m}}=\left[\begin{array}{ccc}
c_{4, x}(m, m-q, q) & \cdots & c_{4, x}(m, q, q)  \tag{5.50}\\
\vdots & \cdots & \vdots \\
c_{4, x}(m, m-q, m-q) & \cdots & c_{4, x}(m, q, m-q)
\end{array}\right] \cdot 0 \leq m \leq q
$$

The dimensions of $\mathbf{D}_{\mathbf{m}}$ are $(2 q-m+1) \times(2 q-m+1)$. In the following $\mathbf{0}_{\mathbf{i}, \mathbf{j}}$ denotes a ( $i \times j$ ) matrix whose elements are all zero. The matrix $\mathbf{D}_{\mathbf{q}}$ is used to define the following set of block matrices:

$$
\mathbf{C}_{\mathbf{q}}^{\mathbf{d}}=\left[\begin{array}{c|c|c}
\mathbf{0}_{\mathbf{q}-\mathbf{d}, \mathbf{d}} & \mathbf{0}_{\mathbf{q}-\mathbf{d}, \mathbf{q}+\mathbf{1}} & \mathbf{0}_{\mathbf{q}-\mathbf{d}, \mathbf{q}-\mathbf{d}}  \tag{5.51}\\
\hline \mathbf{0}_{\mathbf{q}+\mathbf{1}, \mathbf{d}} & \mathbf{D}_{\mathbf{q}} & \mathbf{0}_{\mathbf{q}+\mathbf{1}, \mathbf{q}-\mathbf{d}} \\
\hline \mathbf{0}_{\mathbf{d}, \mathbf{d}} & \mathbf{0}_{\mathbf{d}, \mathbf{q}+\mathbf{1}} & \mathbf{0}_{\mathbf{d}, \mathbf{q}-\mathbf{d}}
\end{array}\right] \cdot 0 \leq d \leq q .
$$

The set of matrices $\left\{\mathbf{C}_{\mathbf{q}}^{\mathbf{d}}\right.$ : $\left.0 \leq d \leq q\right\}$ is a set of linearly independent matrices, whose dimension is $(2 q+1) \times(2 q+1)$.

The matrices $\mathbf{D}_{\mathbf{m}}$ for $0 \leq m \leq q-1$ are used to construct the following set of block

[^17]matrices:
\[

\mathbf{C}_{\mathbf{m}}=\left[$$
\begin{array}{c|c}
\mathbf{0}_{\mathbf{m}, \mathbf{2 q - m}+\mathbf{1}} & \mathbf{0}_{\mathbf{m}, \mathbf{m}}  \tag{5.52}\\
\hline \mathbf{D}_{\mathbf{m}} & \mathbf{0}_{\mathbf{2 q}-\mathbf{m}+\mathbf{1}, \mathbf{m}}
\end{array}
$$\right] \quad 0 \leq m \leq q-1 .
\]

The dimension of these matrices is also $(2 q+1) \times(2 q+1)$.
Equation 5.49 can now be written in matrix form as follows:

$$
\begin{equation*}
\mathbf{C}_{\mathbf{m}}=\sum_{i=0}^{q} \frac{h(i) h(i+m)}{h(0) h(q)} \mathbf{C}_{\mathbf{q}}^{\mathbf{q}-\mathbf{m}-\mathbf{i}} \quad 0 \leq m \leq q-1 \tag{5.53}
\end{equation*}
$$

Equation 5.53 shows that the set of matrices $\left\{\mathbf{D}_{\mathbf{m}}: 0 \leq m \leq q-1\right\}$ belongs to the linear space spanned by the base matrices $\left\{\mathbf{C}_{\mathbf{q}}^{\mathbf{d}}: 0 \leq d \leq q\right\}$ :

$$
\begin{equation*}
\mathbf{C}_{\mathbf{m}} \subset \operatorname{span}\left(\left\{\mathbf{D}_{\mathbf{q}}^{\mathbf{d}}: 0 \leq d \leq q\right\}\right) \quad m=0, \ldots, q-1 \tag{5.54}
\end{equation*}
$$

The matrix $\mathbf{D}_{\mathbf{q}}$, which is used to construct the matrices $\mathbf{C}_{\mathbf{q}}^{\mathbf{d}}$, has rank 1. This is easily seen if $\tau_{1}$ is set to $q$ in 5.49:

$$
\begin{equation*}
c_{4, x}\left(m, q, \tau_{2}\right)=\frac{h(m)}{h(q)} c_{4, x}\left(q, q, \tau_{2}\right) . \tag{5.55}
\end{equation*}
$$

Equation 5.55 shows that the rows (and columns) of $\mathbf{D}_{\mathbf{q}}$ and consequently the rows and columns of $\mathbf{C}_{\mathbf{q}}^{\mathbf{d}}$, are collinear.

Expressions 5.54 and 5.55 provide important information on the theoretical structure of fourth-order cumulants of MA $(q)$ models. That information can be used to construct a matrix of fourth-order cumulants, so that the properties of the matrix reflect the properties of the set of cumulants. Before defining the final cumulant matrix, it is helpful to define some additional auxiliary matrices denoted by $\mathbf{D}_{\mathbf{q}}^{\mathrm{d}}$ for $0 \leq d \leq q$ and $\mathbf{D}_{\mathbf{m}}^{\prime}$ for $0 \leq m \leq q-1$ :

$$
\begin{gather*}
\mathbf{D}_{\mathbf{q}}^{\mathbf{d}}=\left[\mathbf{0}_{\mathbf{q}+\mathbf{1}, \mathbf{d}}\left|\mathbf{D}_{\mathbf{q}}\right| \mathbf{0}_{\mathbf{q}+\mathbf{1}, \mathbf{q}-\mathbf{d}}\right], \quad 0 \leq d \leq q  \tag{5.56}\\
\mathbf{D}_{\mathbf{m}}^{\prime}=\left[\mathbf{D}_{\mathbf{m}} \mid \mathbf{0}_{\mathbf{2 q - \mathbf { m } + \mathbf { 1 } , \mathbf { m }}}\right] \quad 0 \leq m \leq q-1 \tag{5.57}
\end{gather*}
$$

Finally, the following matrix is defined:

$$
\mathbf{C}_{\mathbf{4}, \mathbf{q}}=\left[\begin{array}{c}
\frac{\mathbf{D}_{\mathbf{q}}^{0}}{\vdots}  \tag{5.58}\\
\hline \frac{\mathbf{D}_{\mathbf{q}}^{\mathbf{q}}}{\mathbf{D}_{\mathbf{q}-1}^{\prime}} \\
\hline \vdots \\
\hline \mathbf{D}_{0}^{\prime}
\end{array}\right] .
$$

The dimension of matrix $\mathbf{C}_{4, \mathbf{q}}$ is $\left(\frac{(q+1)(5 q+4)}{2}\right) \times(2 q+1)$ and contains all the available fourth-order cumulant statistics. Clearly, from the discussion earlier in this section, $\mathrm{C}_{\mathbf{4}, \mathrm{q}}$ possesses the following properties:

- The rank of $\mathbf{C}_{4, \mathbf{q}}$ is $q+1$. This is a direct consequence of expressions 5.54 and 5.55.
- $\mathbf{C}_{4, \mathbf{q}}$ has a linear structure with $\frac{1}{6}(q+3)(q+2)(q+1)$ degrees of freedom.

The principal parameter vector of the linear structured matrix is the vector containing all the fourth-order cumulants. The structure of matrix $\mathbf{C}_{4, q}$ is much more complicated than the third-order equivalent $\mathbf{C}_{\mathbf{3}, \mathbf{q}}$, mainly because of the large number of fourthorder statistics and the large number of regions of symmetry in the lags of fourth-order cumulants. Consequently, it is very difficult to prove that the properties of $\mathbf{C}_{4, \mathbf{q}}$ are sufficient to characterise fourth-order cumulants of some MA $(q)$ model. Consequently only a reduced set of fourth-order cumulants is considered in the next section.

### 5.4.2 Properties of a subset of the fourth order cumulants

The parameter estimation methods developed in the previous chapters, when extended to the fourth-order cumulant case, involve only the fourth-order cumulants whose two of the lags are equal. It is therefore sensible to examine the properties of this particular subset of cumulants. When examining the properties of a subset of fourth-order cumulants, it is important to ensure that the subset contains a sufficient amount of structure to characterise the selected cumulants. In the subset considered here, as it will become obvious later in this section, it is necessary to also include all the cumulants contained in the 2 D -slice $c_{4, x}\left(q, \tau_{1}, \tau_{2}\right)$. Suppose that the all the selected fourth-order cumulants are collected in a vector $\boldsymbol{\xi}_{4}$. The following paragraph gives a formal definition of the vector $\boldsymbol{\xi}_{4}$.

Definition of $\boldsymbol{\xi}_{4}$ As we have seen in the previous paragraph the vector $\boldsymbol{\xi}_{4}$ contains the fourth-order cumulants which have two equal lags and the cumulants which have one lag fixed to $q$. In order to formally define $\boldsymbol{\xi}_{4}$, it is necessary to define the cumulant lags that result in a minimal description of the set. There are three separate cases for the lags of $c_{4, x}(\tau, \tau, m)$ :
(i). $\tau, m$ satisfy the condition $0 \leq \tau \leq m \leq q$. This results in $(q+2)(q+1) / 2$ different cumulants.
(ii). $\tau, m$ satisfy the condition $0 \leq m<\tau \leq q$. This results in $q(q+1) / 2$ different cumulants.
(iii). $\tau, m$ satisfy the conditions $m-q \leq \tau<0 \leq m \leq q$. This results in $(q+2) q / 2$ different cumulants.

The above add up to $\left(3 q^{2}+6 q+1\right) / 2$ cumulants. Apart from these cumulants, vector $\boldsymbol{\xi}_{4}$ also contains the cumulants $c_{4, x}\left(q, \tau_{1}, \tau_{2}\right)$ for $0<\tau_{2}<\tau_{1}<q$; that is $q(q-3) / 2$ additional cumulants. Consequently $\boldsymbol{\xi}_{4}$ contains $\left(4 \boldsymbol{q}^{2}+3 q+1\right) / 2$ cumulants.

## Linear relationships for the cumulants of $\boldsymbol{\xi}_{4}$

In the following we make use of equation 3.45 which is repeated here using more convenient notation for the cumulant lags:

$$
\begin{gather*}
\sum_{i=0}^{q} h(i) h\left(i+m_{1}\right) h\left(i+m_{2}\right) c_{4, x}\left(\tau_{1}, \tau_{2}, i+m\right)= \\
\sum_{i=0}^{q} h(i) h\left(i+\tau_{1}\right) h\left(i+\tau_{2}\right) c_{4, x}\left(m_{1}, m_{2}, i+m\right) \tag{5.59}
\end{gather*}
$$

Equation 5.59 for $m_{1}=q, m_{2}=0$ and $\tau_{1}=\tau_{2}=\tau$ becomes

$$
\begin{equation*}
c_{4, x}(\tau, \tau, m)=\sum_{i=0}^{q} \frac{h(i) h^{2}(i+\tau)}{h(q) h^{2}(0)} c_{4, x}(q, 0, i+m) \tag{5.60}
\end{equation*}
$$

Equation 5.60 can be given a vectorial interpretation in a similar way to the third-order equation 5.2. Let us define the following vectors ${ }^{5}$ :
$\mathbf{c}_{\mathbf{q}-\mathbf{n}}=[\underbrace{c_{4, x}(q-n, q-n,-n), c_{4, x}(q-n, q-n,-n+1), \ldots, c_{4, x}(q-n, q-n, q)}_{q+n+1}, \underbrace{0, \ldots, 0}_{q-n}]^{\top}$,
where $n=1, \ldots, 2 q-1$, and

$$
\begin{equation*}
\mathbf{c}_{\mathbf{q}}^{\mathbf{d}}=[\underbrace{0, \ldots, 0}_{d}, \underbrace{c_{4, x}(q, 0,0), c_{4, x}(q, 0,1), c_{4, x}(q, 0,2), \ldots, c_{4, x}(q, 0, q)}_{q+1}, \underbrace{0, \ldots, 0}_{q-d}]^{\top}, \tag{5.62}
\end{equation*}
$$

for $d=0, \ldots, q$. The vectors defined by $5.61,5.62$ have $2 q+1$ elements. Equation 5.60 can now be written in vector form as follows:

$$
\begin{equation*}
\mathbf{c}_{\mathbf{q}-\mathbf{n}}=\sum_{i=0}^{q} \frac{h(i) h^{2}(i+q-n)}{h(q) h^{2}(0)} \mathbf{c}_{\mathbf{q}}^{\mathbf{q}-\mathbf{n}-\mathbf{i}} \tag{5.63}
\end{equation*}
$$

[^18]A different kind of relationship, exists between the cumulants $c_{4, x}\left(q, \tau, m_{1}\right)$ and $c_{4, x}\left(q, \tau, m_{2}\right)$ :

$$
\begin{equation*}
c_{4, x}\left(q, \tau, m_{1}\right)=\frac{h\left(m_{1}\right)}{h\left(m_{2}\right)} c_{4, x}\left(q, \tau, m_{2}\right) \tag{5.64}
\end{equation*}
$$

where $h\left(m_{2}\right)$ is assumed to be non-zero. This equation can also be given a useful vectorial interpretation: Define the following vectors

$$
\begin{equation*}
\mathbf{c}_{\mathbf{q}, \mathbf{l}}=[\underbrace{c_{4, x}(q, l, 0), c_{4, x}(q, l, 1), c_{4, x}(q, l, 2), \ldots, c_{4, x}(q, l, q)}_{q+1}, \underbrace{0, \ldots, 0}_{q}]^{\top}, \tag{5.65}
\end{equation*}
$$

for $l=0, \ldots, q$. Then 5.64 in vector form is written as

$$
\begin{equation*}
\mathbf{c}_{\mathbf{q}, \mathbf{m}_{1}}=\frac{h\left(m_{1}\right)}{h\left(m_{2}\right)} \mathbf{c}_{\mathbf{q}, \mathbf{m}_{2}} \tag{5.66}
\end{equation*}
$$

for $h\left(m_{2}\right) \neq 0$. The vectors $\mathbf{c}_{\mathbf{q}, \mathbf{m}_{1}}$ also have $2 q+1$ elements.
From equations 5.63 and 5.66 it is easy to see that

$$
\begin{equation*}
\mathbf{c}_{\mathbf{q}-\mathbf{n}}, \mathbf{c}_{\mathbf{q}, \mathbf{l}} \in \operatorname{span}\left(\left\{\mathbf{c}_{\mathbf{q}}^{\mathbf{d}}: d=0, \ldots, q\right\}\right) \tag{5.67}
\end{equation*}
$$

where $n=0, \ldots, q$ and $l=0, \ldots, q$. In a similar manner to the previous section, these vectors can be used to form a matrix with specific rank and structure properties:

$$
\begin{equation*}
\mathbf{C}_{4, \mathbf{q}}^{\prime}=\left[\mathbf{c}_{\mathbf{q}, \mathbf{q}}, \ldots, \mathbf{c}_{\mathbf{q}, \mathbf{1}}, \mathbf{c}_{\mathbf{q}}^{\mathbf{0}}, \ldots, \mathbf{c}_{\mathbf{q}}^{\mathbf{q}}, \mathbf{c}_{\mathbf{q}-1}, \ldots, \mathbf{c}_{0}, \ldots, \mathbf{c}_{1-\mathbf{q}}\right]^{\top} \tag{5.68}
\end{equation*}
$$

$\mathrm{C}_{4, \mathrm{q}}^{\prime}$ is a $(4 q) \times(2 q+1)$ matrix containing cumulants belonging to the vector $\boldsymbol{\xi}_{4}$ defined earlier. Considering only the subset of cumulants which belong in $\boldsymbol{\xi}_{4}$, results in a significant reduction in the size of $\mathbf{C}_{4, \mathrm{q}}^{\prime}$ when this is compared with the size of $\mathbf{C}_{\mathbf{4}, \mathbf{q}}$ defined in equation 5.58. Furthermore $\mathbf{C}_{4, \mathrm{q}}^{\prime}$ possesses similar properties to those of $\mathbf{C}_{\mathbf{4}, \mathbf{q}}$. More specifically $\mathbf{C}_{\mathbf{4}, \mathbf{q}}^{\prime}$ possesses the following properties:

- The rank of $\mathbf{C}_{\mathbf{4}, \mathbf{q}}^{\prime}$ is $q+1$. This is a direct consequence of expression 5.67.
- $\mathbf{C}_{4, \mathbf{q}}^{\prime}$ has a linear structure with $\left(4 q^{2}+3 q+1\right) / 2$ degrees of freedom.

The simple structure of $\mathbf{C}_{4, \mathbf{q}}^{\prime}$ makes it possible to prove the following theorem:

Theorem 5.3 Every ( $4 q) \times(2 q+1)$ matrix S possessing the above structure and rank properties, consists of real fourth-order cumulants of some MA(q) model.

The proof is an extension of the proof of theorem 5.1 and is given in Appendix B.

### 5.5 MA Cumulant Enhancement

So far this chapter has been devoted to the investigation of the structure of MA(q) cumulants. Cumulant properties have been identified, and they have been shown to characterise cumulant matrices. In this section sample cumulants are considered. In order to make the discussion as general as possible, the order of the statistics used is not stated explicitly in most of the material that follows.

Suppose that $\hat{\mathbf{x}}=[x(0), \ldots, x(N)]$ is a single realisation of an MA $(q)$ process. This observation vector is used to calculate a vector of sample cumulant estimates $\hat{\boldsymbol{\xi}}$. The vector $\hat{\boldsymbol{\xi}}$ represents the sample estimate version of either $\hat{\xi_{3}}$ or $\hat{\xi}_{2}$ or $\hat{\xi}_{4}$. The number of sample cumulants in $\hat{\boldsymbol{\xi}}$ is $\nu_{\xi}$, so $\hat{\boldsymbol{\xi}} \in \mathbb{R}^{\nu_{\xi}}$. The corresponding vector consisting of true cumulants of the MA $(q)$ process is denoted by $\boldsymbol{\xi}_{\boldsymbol{h}}$. The set of all true MA $(q)$ cumulant vectors $\boldsymbol{\xi}_{\boldsymbol{h}}$ is denoted by $\mathcal{K}_{q} \subset \mathbb{R}^{\nu_{\xi}}$. In practical situations, due to the inaccuracies in the estimation of the sample cumulants, the inter-relationships between the elements of $\hat{\boldsymbol{\xi}}$ are not the same as those that exist between true cumulants of an MA $(q)$ process.

Problem definition: The problem under consideration here is that of finding a function that can be used to map the sample cumulant vector $\hat{\boldsymbol{\xi}}$ to another vector $\dot{\boldsymbol{\xi}}$ whose elements are true cumulants of some MA $(q)$ model i.e. to a vector in $\mathcal{K}_{q}$. This function should also satisfy some optimality criteria, for example $\check{\boldsymbol{\xi}}$ should be the "nearest" vector to $\hat{\boldsymbol{\xi}}$ possessing the prescribed structure. We denote such a matrix as $\mathcal{F}_{\mathcal{K}}: \mathbb{R}^{\nu_{\xi}} \rightarrow \mathcal{K}_{q}$.

In the previous sections it was shown that the properties of cumulants can be translated to matrix properties. Consequently, any mapping corresponding to cumulant properties can be translated to a matrix mapping. Suppose that $\mathbf{C}_{\mathbf{q}}$ is an ( $m \times n$ ) matrix ${ }^{6}$ representing one of the matrices $\mathbf{C}_{3, \mathbf{q}}, \mathbf{C}_{2,3, \mathbf{q}}$ or $\mathbf{C}_{4, \mathbf{q}}^{\prime}$. The rank of $\mathbf{C}_{\mathbf{q}}$ is $q+1$ and in addition has a characteristic linear structure which depends on the order of cumulants used. The set of all such $\mathbf{C}_{\mathbf{q}}$ matrices corresponding to all possible MA $(q)$ models, is denoted by $\mathcal{M}_{q} \subset \mathcal{H}$, where $\mathcal{H}$ is the Hilbert space of ( $m \times n$ ) matrices. The inner product in $\mathcal{H}$ is defined as

$$
\begin{equation*}
\left\langle\mathbf{X}_{1} \mathbf{X}_{2}\right\rangle=\operatorname{trace}\left(\mathbf{X}_{1} \mathbf{X}_{2}^{\top}\right), \quad \mathbf{X}_{1} \mathbf{X}_{2} \in \mathcal{H} \tag{5.69}
\end{equation*}
$$

The corresponding norm is the Frobenius norm (denoted as $\|\ldots\|_{F}$ ). Let $\hat{\mathbf{C}}_{\mathbf{q}}$ denote the matrix obtained from $\mathbf{C}_{\mathbf{q}}$ after replacing its elements with their corresponding finite sample estimates. Since the symmetries in the lags of sample cumulants are the same as those of true cumulants and the linear structure depends only on the symmetries of the cumulants involved, the sample cumulant matrix $\hat{\mathbf{C}}_{\mathbf{q}}$ has the same linear structure as $\mathbf{C}_{\mathbf{q}}$. The estimation errors present in the elements of $\hat{\mathbf{C}}_{\mathbf{q}}$, force $\hat{\mathbf{C}}_{\mathbf{q}}$ to be a full-rank

[^19]matrix.
Let $\mathcal{S}_{l s}$ denote the set of ( $m \times n$ ) matrices possessing the linear structure induced by the cumulant symmetries. Also let $\mathcal{S}_{q+1}$ denote the set of ( $m \times n$ ) matrices whose rank is $q+1$. Because of theorems 5.1,5.2 and 5.3, it follows ${ }^{7}$ that $\mathcal{M}_{q}=\mathcal{S}_{l s} \cap \mathcal{S}_{q+1}$. The problem of cumulant enhancement defined earlier, can now be decomposed as follows:
(i). Map the sample cumulant vector $\hat{\boldsymbol{\xi}}$ to the corresponding linear structured matrix $\hat{\mathbf{C}}_{\mathbf{q}}$ which belongs in $\mathcal{S}_{l s}$. Denote this mapping as
\[

$$
\begin{equation*}
\mathcal{F}_{l s}: \mathbb{R}^{\nu_{\xi}} \rightarrow \mathcal{S}_{l s} \tag{5.70}
\end{equation*}
$$

\]

(ii). Given a matrix $\hat{\mathbf{C}}_{\mathbf{q}}$ in $\mathcal{S}_{l s}$, find a mapping $\mathcal{G}: \mathcal{S}_{l s} \rightarrow \mathcal{S}_{l s} \cap \mathcal{S}_{q+1}$, such that $\mathcal{G}\left(\hat{\mathbf{C}}_{\mathbf{q}}\right)$ is the "nearest" matrix to $\hat{\mathbf{C}}_{\mathbf{q}}$ among the matrices of $\mathcal{S}_{l s} \cap \mathcal{S}_{q+1}$.
(iii). Finally, the inverse mapping $\mathcal{F}_{l s}^{-1}: \mathcal{S}_{l s} \rightarrow \mathbb{R}^{\nu_{\xi}}$ is used to obtain the enhanced cumulant vector $\check{\boldsymbol{\xi}}$. This is because given that $\mathcal{G}\left(\hat{\mathbf{C}}_{\mathbf{q}}\right) \in \mathcal{S}_{l s} \cap \mathcal{S}_{q+1}$, it follows that $\mathcal{F}_{l s}^{-1}\left(\mathcal{G}\left(\hat{\mathbf{C}}_{\mathbf{q}}\right)\right) \in \mathcal{K}_{q}$.

The second step is the most difficult to implement and requires closer examination. Step (ii) can be recast as an optimisation problem: Find a matrix $\dot{\mathbf{C}} \in \mathcal{S}_{l s} \cap \mathcal{S}_{q+1}$ which solves the optimisation problem

$$
\begin{equation*}
\inf _{\mathbf{x} \in \mathcal{S}_{l s} \cap \mathcal{S}_{q+1}}\left\|\mathbf{X}-\hat{\mathbf{C}}_{q}\right\|_{F} \tag{5.71}
\end{equation*}
$$

At this point the Composite Property Mapping Algorithm (CPMA) [91] is used to solve the optimisation problem (5.71). According to [91], in order to render a tractable solution procedure to the optimisation problem (5.71), it is beneficial to decompose the original problem into two subproblems relating to the two individual matrix properties, namely the structure and rank properties. The general theory of property mappings has been developed first by Zangwill in [93]. In [94] Cardoso used the CPMA for fourth-order cumulant "structure forcing" in a blind array processing problem.

Mappings corresponding to the structure and rank properties are discussed in the following sections.

### 5.5.1 Mapping corresponding to the rank property

The objective here is to find a mapping $\mathcal{G}_{q+1}: \mathcal{H} \rightarrow \mathcal{S}_{q+1}$ such that, given a full rank matrix $\mathbf{X} \in \mathcal{H}, \mathcal{G}_{q+1}(\mathbf{X})$ is the $(q+1)$-rank matrix that lies closest to $\mathbf{X}$ in

[^20]the minimum Frobenious norm sense. The widely used Singular Value Decomposition (SVD) representation $[66,91]$ can be used to devise such a mapping. Suppose that the SVD representation of $\mathbf{X}$ is given by
\[

$$
\begin{equation*}
\mathbf{X}=\sum_{k=1}^{2 q+1} \sigma_{k} \mathbf{u}_{k} \mathbf{v}_{k}^{\prime} \tag{5.72}
\end{equation*}
$$

\]

where $\sigma_{k}, k=1, \ldots, 2 q+1$ are the singular values of $\mathbf{X}$ order in decreasing order, and $\mathbf{u}_{k}, \mathbf{v}, k=1, \ldots, 2 q+1$ are the corresponding orthonormal left and right singular vectors. Let $\mathcal{S}_{q+1}^{\prime}$ be the subset of $\mathcal{H}$ whose elements have rank smaller than or equal to $q+1$. Provided that $0<s_{q+1} \neq s_{q+2}$, the desired mapping can be implemented by discarding the $q$ smallest singular values of $\mathbf{X}$ :

$$
\begin{equation*}
\mathcal{G}_{q+1}(\mathrm{X})=\sum_{k=1}^{q+1} \sigma_{k} \mathbf{u}_{k} \mathbf{v}_{k}^{\prime} . \tag{5.73}
\end{equation*}
$$

If $\sigma_{q+1} \geq \sigma_{q+2}$, then the mapping is from $\mathcal{H}$ to $\mathcal{S}_{q+1}^{\prime}$. In practice, it is important always to check that $\sigma_{q+1}$ is sufficiently bigger than $\sigma_{q+2}$, otherwise the rank of $\mathcal{G}_{q+1}(\mathbf{X})$ might be smaller that $q+1$. Since it is safe to assume that $\operatorname{rank}(\mathbf{X})>\operatorname{rank}\left(\mathcal{G}_{q+1}(\mathbf{X})\right)$, the Pythagoras' Theorem holds:

$$
\begin{equation*}
\|\mathbf{X}\|^{2}=\left\|\mathcal{G}_{q+1}(\mathbf{X})\right\|^{2}+\left\|\mathbf{X}-\mathcal{G}_{q+1}(\mathbf{X})\right\|^{2} \tag{5.74}
\end{equation*}
$$

Equation 5.74 is easily obtained using $\|\mathbf{X}\|^{2}=\sum_{i=1}^{2 q+1} \sigma_{i}^{2}$. It is important to note here that, in the general case where $s_{q+1} \geq s_{q+2}, \mathcal{G}_{q+1}() \subset \mathcal{H}$ is a projection on a non-convex cone ${ }^{8}$ [95].

### 5.5.2 Mapping corresponding to the linear structure property

The objective here is to find a mapping $\mathcal{G}_{l s}: \mathcal{H} \rightarrow \mathcal{S}_{l s}$ such that it maps a given matrix $\mathbf{X}$ to the "nearest" matrix that has the same linear structure as the matrix $\mathbf{C}_{\mathbf{q}}$. It would be useful here to characterise the matrix in a more formal way. Suppose $\mathcal{T}: \mathcal{H} \rightarrow \mathbb{R}^{m n}$ denotes a linear transformation such that if $\mathbf{x}=\mathcal{T}(\mathbf{X})$ then $\mathbf{x}$ is the concatenation of column vectors ${ }^{9}$ of $\mathbf{X}$. Furthermore we have that $\mathbf{X}=\mathcal{T}^{-1}(\mathbf{x})(\mathcal{T}$ is an isomorphism). Another useful property of this transform is that it preserves norms in the sense that

$$
\begin{equation*}
\|\mathcal{T}(\mathbf{X})\|_{\mathbf{E}}=\|(\mathbf{X})\|_{\mathbf{F}} \tag{5.75}
\end{equation*}
$$

[^21]where $\|\cdot\|_{E}$ denote the Euclidean norm. $\mathbf{C}_{\mathbf{q}}$ has a linear structure with $\boldsymbol{\xi}$ as the principal parameter vector. The elements of $\mathbf{C}_{\mathbf{q}}$ are either zero or equal to some element of $\boldsymbol{\xi}$. Obviously the same holds for the elements of the vector $\mathcal{T}\left(\mathbf{C}_{\mathbf{q}}\right)$. Then there exists an $m n \times \nu_{\xi}$ matrix $\mathbf{A}_{\mathbf{q}}$ such that
\[

$$
\begin{equation*}
\mathbf{A}_{\mathbf{q}} \boldsymbol{\xi}=\mathcal{T}\left(\mathbf{C}_{\mathbf{q}}\right) \text { or equivalently } \mathcal{T}^{-1}\left(\mathbf{A}_{\mathbf{q}} \boldsymbol{\xi}\right)=\mathbf{C}_{\mathbf{q}} \tag{5.76}
\end{equation*}
$$

\]

Each row of $\mathbf{A}_{\mathbf{q}}$ corresponds to an element of $\mathbf{C}_{\mathbf{q}}$. If this element is zero, then the whole row is zero, while if it is equal to the $i^{\text {th }}$ element of $\boldsymbol{\xi}$ then the $i^{\text {th }}$ element of the row is equal to one and the rest are zero. The matrix $\mathbf{A}$ is called the characteristic matrix of the linear structure [91]. In general, the matrix $\mathbf{A}$ is another way of expressing the functions $\alpha_{i, j}(\cdots)$ in the definition of linear structured matrices 5.1. From the above discussion it is easy to see that the mapping $\mathcal{F}_{l s}(\cdot)$ in 5.70 is implemented as

$$
\begin{equation*}
\mathcal{F}_{l s}(\xi)=\mathcal{T}^{-1}\left(\mathbf{A}_{\mathbf{q}} \boldsymbol{\xi}\right) \tag{5.77}
\end{equation*}
$$

The mapping $\mathcal{G}_{l s}$ can now be expressed in terms of the characteristic matrix: Given a matrix $\mathbf{X}, \mathcal{G}_{l s}(\mathbf{X})$ should satisfy the following conditions:

- It should have the prescribed linear structure, which means that there exist a principal parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{\nu_{\xi}}$, such that

$$
\begin{equation*}
\mathcal{G}_{l s}(\mathbf{X})=\mathcal{T}^{-1}\left(\mathbf{A}_{\mathbf{q}} \boldsymbol{\theta}\right) \tag{5.78}
\end{equation*}
$$

- The $\mathcal{G}_{l s}(\mathbf{X})$ should be selected so that it minimises $\left\|\mathbf{X}-\mathcal{G}_{l s}(\mathbf{X})\right\|_{F}$, which, according to the first condition, becomes

$$
\begin{equation*}
\left\|\mathbf{X}-\mathcal{T}^{-1}\left(\mathbf{A}_{\mathbf{q}} \boldsymbol{\theta}\right)\right\|_{F}=\left\|\mathcal{T}(\mathbf{X})-\mathbf{A}_{\mathbf{q}} \boldsymbol{\theta}\right\|_{E} . \tag{5.79}
\end{equation*}
$$

It is now easy to see that $\mathcal{G}_{l s}(\mathbf{X})$ should be defined as the least squares solution of $\mathbf{A}_{\mathbf{q}} \boldsymbol{\theta}=\mathcal{T}(\mathbf{X}):$

$$
\begin{equation*}
\mathcal{G}_{l s}(\mathbf{X})=T^{-1}\left(\mathbf{A}_{\mathbf{q}}\left[\mathbf{A}_{\mathbf{q}}^{\top} \mathbf{A}_{\mathbf{q}}\right]^{-1} \mathbf{A}_{\mathbf{q}}^{\top} \mathcal{T}(\mathbf{X})\right) \tag{5.80}
\end{equation*}
$$

The set $\mathcal{S}_{l s}$ of $m \times n$ linear-structured matrices characterised by the matrix $\mathbf{A}_{\mathbf{q}}$ is a convex subset of $\mathcal{H}$. The mapping $\mathcal{G}_{l s}(\cdot)$ is an orthogonal projection on that convex subset. The Pythagoras' relation applied to this projection gives

$$
\begin{equation*}
\|\mathbf{X}\|_{F}^{2}=\left\|\mathcal{G}_{l s}(\mathbf{X})\right\|_{F}^{2}+\left\|\mathbf{X}-\mathcal{G}_{l s}(\mathbf{X})\right\|_{F}^{2} \tag{5.81}
\end{equation*}
$$

In the next section the mappings corresponding to the rank and structure properties are combined to form an iterative method for cumulant enhancement.

### 5.5.3 An iterative algorithm based on the composite mapping

Earlier it was shown that the sample cumulant matrix $\hat{\mathbf{C}}_{\mathbf{q}}$ has the same linear structure as a matrix of true cumulants, but it does not have the same rank. Applying the rank reduction mapping 5.73 to $\hat{\mathbf{C}}_{\mathbf{q}}$ produces a matrix with the right rank but destroys the linear structure. The linear structure mapping can be applied now to fix the structure, and then this procedure is repeated until convergence to a matrix possessing both properties is achieved. The composite property mapping is defined as

$$
\begin{equation*}
\mathcal{G}=\mathcal{G}_{q} \mathcal{G}_{q+1}, \tag{5.82}
\end{equation*}
$$

so, given a matrix $\mathbf{X}$, applying the mapping 5.82 gives $\mathcal{G}(\mathbf{X})=\mathcal{G}_{q}\left(\mathcal{G}_{q+1}(\mathbf{X})\right)$. It is possible to get an idea of how close a matrix $\mathbf{X}$ is to a $q+1$-rank matrix, by examining how close the following quantity is to 1 , provided that $\sigma_{q+1}>\sigma_{q+2}$ :

$$
\begin{equation*}
\lambda_{q+1}(X)=\frac{\sum_{k=1}^{q+1} \sigma_{k}}{\sum_{k=1}^{2 q+1} \sigma_{k}} \tag{5.83}
\end{equation*}
$$

The proposed iterative procedure for cumulant enhancement is summarised in the following:
(i). From the available data obtain a sample estimate $\hat{\boldsymbol{\theta}}$ of the Principal Parameter Vector (PPV). (The model order $q$ is assumed to be known.)
(ii). Calculate the characteristic matrix $\mathbf{A}_{\mathbf{q}}$ which corresponds to the known model order $q$.
(iii). Using the PPV $\hat{\boldsymbol{\theta}}$, form the sample estimate $\hat{\mathbf{C}}_{\mathbf{q}}=\mathbf{A}_{\mathbf{q}} \hat{\boldsymbol{\theta}}$.
(iv). $\operatorname{Set} \hat{\mathbf{C}}^{(0)}=\hat{\mathbf{C}}_{\mathbf{q}}$.

Repeat

$$
\begin{equation*}
\hat{\mathbf{C}}^{(k+1)}=\mathcal{G}\left(\hat{\mathbf{C}}^{(k)}\right)=\mathcal{G}_{q}\left(\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)\right) \tag{5.84}
\end{equation*}
$$

until $1-\lambda_{q+1}\left(\hat{\mathbf{C}}^{(k+1)}\right) \leq \epsilon$, where $\epsilon$ is a predefined small positive number.
(v). Finally use the components of the final matrix $\check{\mathbf{C}}_{\mathbf{q}}$ to create the enhanced PPV given by $\check{\boldsymbol{\xi}}=\left[\mathbf{A}_{\mathbf{q}}{ }^{\top} \mathbf{A}_{\mathbf{q}}\right]^{-1} \mathbf{A}_{\mathbf{q}}{ }^{\top} \check{\mathbf{C}}_{\mathbf{q}}$.

The asymptotic behaviour of the proposed algorithm is discussed in the following section.

### 5.5.4 Asymptotic behavior of the cumulant enhancement algorithm

The algorithm developed in the previous section is similar to the method of Successive Projection On Convex Sets presented in [96]. The difference here is that, as it was pointed out earlier, the set $\mathcal{S}_{q+1}$ is not convex and consequently the convergence results in [96] cannot be applied in this case.

A different kind of analysis is adopted here which is similar to the asymptotic study of the multichannel modeling algorithm of [95]. The analysis is based on the Pythagoras equations 5.745 .81 for the two property mappings. Suppose that the iterative algorithm generates an infinite sequence of matrices $\mathbf{C}^{(k)}$ for $k=0, \ldots, \infty$. A single iterative step from $\hat{\mathbf{C}}^{(\mathbf{k})}$ to $\hat{\mathbf{C}}^{(\mathbf{k}+1)}$ is given by:

$$
\begin{equation*}
\hat{\mathbf{C}}^{(k+1)}=\mathcal{G}_{l s}\left(\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)\right) \tag{5.85}
\end{equation*}
$$

The Pythagoras equation for the $\left(\mathcal{G}_{q+1}(\cdot)\right.$ projection is

$$
\begin{equation*}
\left\|\mathbf{C}^{(k)}\right\|_{F}^{2}=\left\|\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)\right\|_{F}^{2}+\left\|\hat{\mathbf{C}}^{(k)}-\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)\right\|_{F}^{2} \tag{5.86}
\end{equation*}
$$

The Pythagoras equation for the $\mathcal{G}_{l s}(\cdot)$ projection is,

$$
\begin{equation*}
\left\|\mathcal{G}_{q+\mathbf{1}}\left(\hat{\mathbf{C}}^{(k)}\right)\right\|_{F}^{2}=\left\|\hat{\mathbf{C}}^{(k+1)}\right\|_{F}^{2}+\left\|\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)-\hat{\mathbf{C}}^{(k+1)}\right\|_{F}^{2} \tag{5.87}
\end{equation*}
$$

From equations 5.86 and 5.87 it is obvious that

$$
\begin{equation*}
\left\|\hat{\mathbf{C}}^{(k+1)}\right\|_{F}^{2} \leq\left\|\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)\right\|_{F}^{2} \leq\left\|\hat{\mathbf{C}}^{(k)}\right\|_{F}^{2} \tag{5.88}
\end{equation*}
$$

Inequality 5.88 shows that the composite property mapping $\mathcal{G}(\cdot)$ is a norm-reducing mapping and consequently the sequence of norms $\left\|\hat{\mathbf{C}}^{(k)}\right\|, k=0, \ldots, \infty$ is decaying. The latter directly implies that the sequence $\left\|\hat{\mathbf{C}}^{(k)}\right\|, k=0, \ldots, \infty$ converges. Furthermore, the matrix sequence $\hat{\mathbf{C}}^{(k)}, k=0, \ldots, \infty$ is a bounded sequence of the Hilbert space $\mathcal{H}$.

Equations 5.86 and 5.87 combined, give

$$
\begin{equation*}
\left\|\hat{\mathbf{C}}^{(k)}-\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)\right\|_{F}^{2}+\left\|\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)-\hat{\mathbf{C}}^{(k+1)}\right\|_{F}^{2}=\left\|\hat{\mathbf{C}}^{(k)}\right\|_{F}^{2}-\left\|\hat{\mathbf{C}}^{(k+1)}\right\|_{F}^{2} \tag{5.89}
\end{equation*}
$$

which, because of the convergence of $\left\|\hat{\mathbf{C}}^{(k)}\right\|, k=0, \ldots, \infty$, results in

$$
\begin{equation*}
\lim _{k \rightarrow \infty}\left\|\hat{\mathbf{C}}^{(k)}-\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)\right\|_{F}^{2}+\left\|\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)-\hat{\mathbf{C}}^{(k+1)}\right\|_{F}^{2}=0 . \tag{5.90}
\end{equation*}
$$

Because of the triangular inequality the following equation holds:

$$
\begin{equation*}
\left\|\hat{\mathbf{C}}^{(k)}-\hat{\mathbf{C}}^{(k+1)}\right\| \leq\left\|\hat{\mathbf{C}}^{(k)}-\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)\right\|_{F}^{2}+\left\|\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)-\hat{\mathbf{C}}^{(k+1)}\right\|_{F}^{2} \tag{5.91}
\end{equation*}
$$

Equations 5.91 and 5.90 combined, give

$$
\begin{equation*}
\lim _{k \rightarrow \infty}\left[\hat{\mathbf{C}}^{(k)}-\hat{\mathbf{C}}^{(k+1)}\right]=\mathbf{0} \tag{5.92}
\end{equation*}
$$

Equation 5.92 shows that the composite property mapping is a distance-reducing mapping. After every application of the composite property mapping in 5.85 the resulting matrix has the correct structure property, but not necessarily the correct rank. However, the minimum distance of $\hat{\mathbf{C}}^{(k)}$ from the set ${ }^{10} \mathcal{S}_{q+1}^{\prime}$ is given by

$$
\begin{equation*}
\mathrm{d}\left(\hat{\mathbf{C}}^{(k)}, \mathcal{S}_{q+1}^{\prime}\right)=\left\|\hat{\mathbf{C}}^{(k)}-\mathcal{G}_{q+1}\left(\hat{\mathbf{C}}^{(k)}\right)\right\|_{F} \tag{5.93}
\end{equation*}
$$

The square of the right-hand of 5.93 is the same as the first term of 5.90 and consequently converges to zero i.e. $\lim _{k \rightarrow \infty} \mathrm{~d}\left(\hat{\mathbf{C}}^{(k)}, \mathcal{S}_{q+1}^{\prime}\right)=0$. This result shows that the sequence of matrices resulting from the iterative algorithm tends asymptotically to the set of matrices of $\mathcal{H}$ whose rank is less that or equal to $q+1$. If during the evolution of the iteration the singular value $\sigma_{q+1}$ remains non-zero, then the sequence tends asymptotically to $\mathcal{S}_{q+1}$, i.e. the subset of $\mathcal{H}$ that contains all the matrices which have rank $q+1$.

### 5.6 Numerical Results

In this section numerical simulations are presented which demonstrate the effect of cumulant enhancement on system identification and blind deconvolution. Random signals are generated according to the following models: Model 1

$$
\begin{gathered}
x(n)=w(n)+1.5 w(n-1)-2.6 w(n-2)-0.89 w(n-3) \\
y(n)=x(n)+v(n) .
\end{gathered}
$$

The zeros of the system transfer function $H(z)$ are located at $0.560 \pm j 0.158,-2.621$.

## Model 2

$$
\begin{align*}
& x(n)=w(n)+0.1 w(n-1)-1.87 w(n-2)+3.02 w(n-3) \\
&-1.435 w(n-4)+1.49 w(n-5)  \tag{5.94}\\
& y(n)=x(n)+v(n)
\end{align*}
$$

The zeros of the system transfer function $H(z)$ are located at $-2.02,0.933 \pm j 0.7158$ and $0.0287 \pm j 0.729$. The inverse filter model order is set to $r_{1}=-35$ and $r_{2}=15$.

[^22]
## Model 3

$$
\begin{gathered}
x(n)=w(n)-0.8 w(n-1)-0.86 w(n-2)+0.768 w(n-3)+1.0205 w(n-4) \\
y(n)=x(n)+v(n)
\end{gathered}
$$

The zeros of the system transfer function $H(z)$ are located at $1.1 \pm j 0.6$ and $-0.7 \pm j 0.4$. The inverse filter model order is set to $r_{1}=-30$ and $r_{2}=25$. The following methods


Figure 5.5: Identification results for signal model 1. The number of output samples is 2500 and SNR is 50 dB . The results obtained without enhancement are shown in green and those obtained after enhancement are shown in blue.
are used for system identification:

M1 This the TLS method of chapter 3 (M1 of section 3.10).
M2 Is the same linear system as M1 but solved with LS instead of TLS.
M3 the method of Alshebeili et al in [35]
M4 the method of Fonollosa et al [69]
M5 the method of Tugnait in [34]


Figure 5.6: The first column graphs show the singular value ratio, the second column graphs show the error of the enhanced cumulants and the third column graphs show the error of system identification of the closed formula [1] in blue, and the error of the LS methods of chapter 3 in green. The horizontal axis in all graphs represents the number of iterations.


Figure 5.7: Identification results for signal model 2. The number of output samples is 5000 and SNR is 50 dB . The results obtained without enhancement are shown in green and those obtained after enhancement are shown in blue.

System identification results for signal models 1 and 2 are reported in figures 5.5 and 5.7 respectively. The maximum number of iterations allowed in the CPMA algorithm is 50 . Convergence may or may not take place after 50 iterations. The results obtained after cumulant enhancement are improved both in terms of bias and variance for all algorithms except M1. The methods M1, M2 and M4 are based only on third-order cumulants. The fact that the results obtained from these three methods are not identical shows that for the signalmodels tested here the enhanced cumulants after 50 iterations are not true cumulants.

The following methods are used for deconvolution:

Method 1: This is the third-order cumulant based method described in section 4.3.2.
Method 2: This is the second- and third-order cumulant method described by equation 4.40.

Method 3: This is a second- and third-order cumulant method described by the special case of equation 4.44 which invoves only main diagonal cumulants.

The results of blind deconvolution for signal model 3 are summarised in figure 5.8. The MSE of deconvolution for the same signal model is shown in figure 5.10 . The number of output samples is 1000 and the SNR is 50 dB . The use of enhanced cumulants has a positive effect on the results obtained by all deconvolution algorithms. The effects are more significant for method 3 which uses only diagonal third-order cumulants as well as second-order cumulants. The enhanced diagonal cumulants have been enriched with information from non-diagonal cumualant slices, and this has resulted in a significant improvement on the deconvolution results.

The same experiment is performed for signal model 2. The number of output samples is 4000 and the SNR is 50 dB . The results of deconvolution are summarised in figure 5.9. The MSE after 50 Monte Carlo runs is shown on figure 5.11. The estimated impulse response obtained from non-enhanced cumulants is severely biased. The preprocessing of the cumulants using 50 iteration of the CPMA for cumulant enhancement has reduced the bias significantly and thus resulted in lower MSE.

### 5.7 Conclusions

This chapter has considered theoretical properties of sets of MA cumulant statistics. The higher-order cumulants of MA models were organised in suitably constructed matrices and it is shown that such matrices have a rank depending on the order of the MA model, and a linear structure dictated by the symmetries of the cumulants. These two properties of the cumulant matrices are referred to as the rank and structure properties.


Figure 5.8: Deconvolution results for signal model 3. The number of output samples is 1000 and the SNR is 50 dB . The results are summarised after 50 Monte Carlo runs.


Figure 5.9: Deconvolution results for signal model 2. The number of output samples is 4000 and the SNR is 50 dB . The results are summarised after 50 Monte Carlo runs.


Figure 5.10: Mean Square Error of deconvolution after 50 Monte Carlo runs for signal model 3. The results obtained without enhancement are shown in green and those obtained after enhancement are shown in blue.


Figure 5.11: Mean Square Error of deconvolution after 50 Monte Carlo runs for signal model 2. The results obtained without enhancement are shown in green and those obtained after enhancement are shown in blue.

It is also shown that any matrix possessing these two properties can be considered as a matrix consisting of true MA cumulants. Based on these results a CPMA was devised with the objective of mapping matrices of sample cumulants, which do not possess the two properties, to matrices of true MA cumulants. The CPMA algorithm is an iterative algorithm, and when it converges, it performs a function similar to nonlinear cumulantmatching methods. The problem of convergence is also discussed and it is shown that provided the rank reduction mapping involved in the CPMA are one-to-one mappings, the distance of the generated matrices to the desired set, converges asymptotically to zero. In practice it is observed that convergence may require thousands of iterations. It is also demonstrated that even if convergence is not achieved within a prespecified number of iterations, the enhanced cumulants can improve the results obtained from system identification and deconvolution algorithms.

## Summary and Conclusions

### 6.1 Introduction

The work described in this thesis has been primarily concerned with the development and analysis of higher-order cumulant-based techniques for the estimation of the parameters involved in MA modeling. New estimation methods for both the problems of systems identification and blind deconvolution have been proposed in this thesis. Within this chapter the main conclusions of the work are highlighted and pointers towards future work are presented.

### 6.2 Achievements of the Work

For the problem of system identification, theoretical work has been performed which results in new general relationships between cumulant slices of the same order and cumulant slices of two different orders. These relationships also involve the system parameters and, consequently, they can be used for system identification. In fact, some of the most important existing MA parameter estimation methods are based on special cases of the equation relating cumulants of different orders, and so this equation has been used in chapter 3 to give a unified description of existing estimation methods. The main contribution of chapter 3 was to use the equation relating third-order cumulant slices, and the one relating fourth-order cumulant slices, to develop a new third-order cumulant-based method and a new fourth-order cumulant-based method for system identification in MA modeling. The new methods are based on the solution of systems of equations which are linear with respect to double products of the system impulse response. Both LS/TLS and recursive methods are proposed for the solution of the systems of equations. In addition, SVD and LS methods are proposed for the efficient recovery of the system parameters from the double products of the impulse response, which result from the solution of the linear systems. The important issue of identifiability was also addressed in chapter 3. Expressions for the asymptotic variance of the parameters estimated with the new methods were derived. The system identification algorithms proposed in chapter 3 do not work if the wrong model-order is used. It is thus evident that the robustness of the system identification algorithms, depends on
the availability of good model order selection methods. For this purpose methods for MA order were also proposed in chapter 3. The simulations presented at the end of chapter 3 demonstrate the ability of the new methods to provide accurate estimates of the system parameters even in the presence of high-levels of additive coloured Gaussian noise. The MA models used in the simulations are similar to MA models used to test HOC-based identification methods in the statistical signal processing literature and are characterised by relatively low model-orders. In some practical problems the underlying systems have longer impulse responses, and an AR or ARMA model could result in more efficient modeling. The MA identification methods of chapter 3 can also be used for the identification of the paramaters of AR or ARMA models.

For the problem of blind deconvolution, new general expressions were derived which relate the inverse impulse response with the output cumulants. Expressions involving cumulants of the same order and expressions involving cumulants of two different orders were derived. The only previously reported HOC-based methods for blind deconvolution are those of Nikias and Chiang [39] and Zheng and McLaughlin [45]. In [39], a purely third-order cumulant-based method is presented, while in [45], methods based on second-order and diagonal third-order cumulants as well as methods based on second-order and diagonal fourth-order cumulants are reported. The general equations derived in chapter 4, allow for a unified description of these existing methods and, most importantly, allow for the development of new methods combining second-order cumulants with all third-order cumulant slices and new methods based on fourth-order cumulants. An important part of the analysis of chapter 4, is the study of the structure of the matrices involved in the deconvolution methods. As a result of this analysis criteria for the identifiability of the inverse filter parameters were formulated in chapter 4. Another result reported in chapter 4 is that of system identification through the inverse filter parameters without the need of inversion. Additionaly new asymptotic performance expressions have been derived for the deconvolution methods.

Finally in chapter 5 , the theory of MA cumulant enhancement was presented. This is a novel concept, which has not been considered before in the HOC-based MA parameter estimation. Chapter 5 , using the expressions developed in chapter 3, presented some characteristic properties of cumulants of MA models, and formulated them using matrix theoretic properties. The use of matrix concepts to express cumulant properties, facilitated the formulation of property mappings corresponding to the cumulant properties. Using the theory of CPMA (Composite Property Mappings Algorithm), an iterative algorithm was proposed in chapter 5 , which maps sets of sample cumulants to sets of real cumulants of some MA model. The convergence properties of the iterative algorithm were studied and it was pointed out that even if convergence can sometimes takes many iterations to be realised, only a few iterations are enough to "enhance" the properties of the original set of sample cumulants. The enhanced set of sample cumulants can then be used for system identification or deconvolution using any of the available methods. It has been seen through numerical simulations, that applying
cumulant enhancement to the sample cumulants can sometimes improve significantly the performance of parameter estimation methods.

### 6.3 Future Work

To conclude the thesis, I suggest some pointers to further areas of development, and some alternative applications.

The work presented in this thesis was exclusively concerned with batch estimation problems, in which the entire data set is available to the estimator. In many real situations, for example in communication systems, we frequently encounter situations in which the data stream is continuous and it is required to update the estimate continuously. Adaptive versions of the parameter estimation algorithms presented in this thesis can be derived in the spirit of the theory presented in [8].

Another issue that requires further investigation is that of parameter estimation for Multiple Input Multiple Output (MIMO) linear models. The work presented in this thesis concentrated on the problem of parameter estimation of Single Input Single Output systems. Many practical problems have a multi-dimensional nature. In [97], the formula of Brillinger Rosenblatt 2.35 which holds for SISO models is extended to the MIMO case. The extended equation is based on the Kronecker product and its structure is remarkably similar to that of the SISO equation. It will be interesting to examine whether the similarity at the level of the Barlett Brillinger formula can lead to an extension of the results presented in this thesis, to similar results for MIMO systems.

Finally, in this thesis the problem of cumulant enhancement was studied only for the case of MA models. An interesting extension would be to consider models involving IIR systems. Such models are the AR and ARMA models. Trying to establish properties that characterise sets of cumulants as being the cumulants of some ARMA or AR model is less straightforward than the MA case, because cumulants of AR or ARMA models do not have a finite domain of support. A possible way to overcome this problem is to use identifiability results for AR and ARMA models such as those presented in [61]. These results define subsets of the infinite set of ARMA or AR cumulants which are sufficient for the characterisation of the underlying models. One can then concentrate in these subsets to try establish properties which characterise true cumulants of some AR or ARMA model.

## Appendix A

## $z$-Transform

The bilateral $z$-transform of the sequence $y(k), k=-\infty, \ldots, \infty$ is defined as

$$
\begin{equation*}
Y(z)=\sum_{k=-\infty}^{\infty} y(k) z^{-k} \tag{B.1}
\end{equation*}
$$

More information on the properties of the $z$-transform can be found in [56].

## The Kronecker Product

The Kronecker product of $\mathbf{A}(p \times q)$ and $\mathbf{B}(m \times n)$ is denoted $\mathbf{A} \otimes \mathbf{B}$ and is $p m \times q n$ matrix defined by

$$
\mathbf{A} \otimes \mathbf{B}=\left[\begin{array}{c|c|c|c}
\alpha_{11} \mathbf{B} & \alpha_{12} \mathbf{B} & \cdots & \alpha_{1 q} \mathbf{B}  \tag{B.2}\\
\hline \alpha_{21} \mathbf{B} & & & \\
\hline \vdots & & & \\
\alpha_{p 1} \mathbf{B} & & & \alpha_{p q} \mathbf{B}
\end{array}\right]
$$

More information on the properties of the Kronecker product can be found in [98].

## Barlett's asymptotic formula and equation 2.68

Equation 2.68 for $n=k=2$ is written as

$$
\begin{array}{r}
\lim _{N \rightarrow \infty} N \operatorname{cov}\left\{\hat{m}_{2, y}(\tau), \hat{m}_{2, y}(\sigma)\right\}= \\
\sum_{t=-\infty}^{\infty}\left[E\{y(0) y(\tau) y(t) y(t+\sigma)\}-m_{2, y}(\tau) m_{2, y}(\sigma)\right]= \\
\sum_{t=-\infty}^{\infty}\left[m_{4, y}(\tau, t, t+\sigma)-m_{2, y}(\tau) m_{2, y}(\sigma)\right] \tag{B.3}
\end{array}
$$

From equation 2.10 we know that

$$
\begin{array}{r}
m_{4, y}(\tau, t, t+\sigma)=c_{4, y}(\tau, t, t+\sigma)+m_{2, y}(\tau) m_{2, y}(\sigma)+ \\
m_{2, y}(t) m_{2, y}(t+\sigma-\tau)+m_{2, y}(t+\sigma) m_{2, y}(t-\tau) . \tag{B.4}
\end{array}
$$

Finally combining equations B. 3 and B. 4 we obtain Barlett's asymptotic formula

$$
\begin{array}{r}
\lim _{N \rightarrow \infty} N \operatorname{cov}\left\{\hat{m}_{2, y}(\tau), \hat{m}_{2, y}(\sigma)\right\}= \\
\sum_{t=-\infty}^{\infty}\left(m_{2, y}(t) m_{2, y}(t+\sigma-\tau)+m_{2, y}(t+\sigma) m_{2, y}(t-\tau)+c_{4, y}(\tau, t, t+\sigma)\right) \tag{B.5}
\end{array}
$$

## Appendix B

## Proof of Theorem 5.3

Assume that the $(4 q) \times(2 q+1)$ matrix $\mathbf{S}$, posseses the desired rank and structure properties defined in section 5.4.2. Since $\mathbf{S}$ has the same structural characteristics as a matrix constructed from real fourth-order cumulants, then if the $(i, j)$-element of $\mathbf{S}$ is not identically zero, it can be expressed as $s\left(\tau_{1}, \tau_{2}, \tau_{3}\right)$ where $\left(\tau_{1}, \tau_{2}, \tau_{3}\right)$ are the lags we associate with ( $i, j$ )-element of a structurally equivalent matrix consisting of real fourth-order cumulants. S can be written as follows:

$$
\begin{equation*}
\mathbf{S}=\left[\mathbf{s}_{\mathbf{q}, \mathbf{q}}, \ldots, \mathrm{s}_{\mathbf{q}, \mathbf{1}}, \mathbf{s}_{\mathbf{q}}^{\mathbf{0}}, \ldots, \mathbf{s}_{\mathbf{q}}^{\mathbf{q}}, \mathbf{s}_{\mathbf{q}-\mathbf{1}}, \ldots, \mathbf{s}_{\mathbf{0}}, \ldots, \mathrm{s}_{1-\mathbf{q}}\right]^{\top} \tag{B.1}
\end{equation*}
$$

where

$$
\begin{equation*}
\mathbf{s}_{\mathbf{q}, \mathrm{l}}=[\underbrace{s(q, l, 0), s(q, l, 1), s(q, l, 2), \ldots, s(q, l, q)}_{q+1}, \underbrace{0, \ldots, 0}_{q}]^{\top}, \tag{B.2}
\end{equation*}
$$

for $l=0, \ldots, q$,

$$
\begin{equation*}
\mathbf{s}_{\mathbf{q}}^{\mathbf{d}}=[\underbrace{0, \ldots, 0}_{d}, \underbrace{s(q, 0,0), s(q, 0,1), s(q, 0,2), \ldots, s(q, 0, q)}_{q+1}, \underbrace{0, \ldots, 0}_{q-d}]^{\top}, \tag{B.3}
\end{equation*}
$$

for $d=0, \ldots, q$ and

$$
\begin{equation*}
\mathbf{s}_{\mathbf{q}-\mathbf{n}}=[\underbrace{s(q-\dot{n}, q-n,-n), s(q-n, q-n,-n+1), \ldots, s(q-n, q-n, q)}_{q+n+1}, \underbrace{0, \ldots, 0}_{q-n}]^{\top}, \tag{B.4}
\end{equation*}
$$

for $n=1, \ldots, 2 q-1$. It is further asssumed that $s(q, q, 0), s(q, q, q), s(q, 0,0) \neq 0$. As a result of the way they are constructed the $q+1$ vectors $\mathbf{s}_{\mathbf{q}}^{\mathbf{d}}$ for $d=0, \ldots, q$ are linearly independent. Since the rank of $\mathbf{S}$ is $q+1$, each of the vectors $\mathbf{s}_{\mathbf{q}, 1}$ for $l=0, \ldots, q$ is a scaled version of $\mathbf{s}_{\mathbf{q}}^{0}$ and consequently the following eqations hold:

$$
\begin{equation*}
s(q, q, \tau)=\frac{s(q, 0, q) s(q, 0, \tau)}{s(q, 0,0)} \tag{B.5}
\end{equation*}
$$

and

$$
\begin{equation*}
s(\tau, \tau, q)=\frac{s(q, 0, \tau) s(q, 0, \tau)}{s(q, 0,0)} \tag{B.6}
\end{equation*}
$$

It is necessary to prove that

$$
\begin{equation*}
s(q-k, q-k, m)=\sum_{i=0}^{q} \frac{s(q, 0, i) s^{2}(q, 0, i+q-k)}{s^{2}(q, 0,0) s(q, 0, q)} s(q, 0, i+m) \tag{B.7}
\end{equation*}
$$

for $k=1, \ldots, 2 q-1$. The vectors $\mathbf{s}_{\mathbf{q}-\mathbf{k}}$ for $k=1, \ldots, q$ can be expressed as a linear combination of the vectors $\mathrm{s}_{\mathrm{q}}^{\mathrm{d}}$ for $d=0, \ldots, q$ awccording to the following equation:

$$
\begin{equation*}
\mathbf{s}_{\mathbf{q}-\mathbf{k}}=\sum_{i=0}^{k} \lambda_{i}^{(q-k)} \mathbf{s}_{\mathbf{q}}^{\mathbf{k}-\mathbf{i}} . \tag{B.8}
\end{equation*}
$$

Equation B. 7 will be derived as a by-product of the solution of equation B. 8 for $k=$ $1, \ldots, q$ with respect to the $\lambda$ 's. More specifically it is desired to prove that

$$
\begin{equation*}
\lambda_{i}^{(q-k)}=\frac{s(q, 0, i) s^{2}(q, 0, i+q-k)}{s^{2}(q, 0,0) s(q, 0, q)} \tag{B.9}
\end{equation*}
$$

for $k=1, \ldots, q$ and $i=0, \ldots, k$. Equation B. 8 for $k=1$ becomes:

$$
\begin{equation*}
\mathbf{s}_{\mathbf{q}-\mathbf{1}}=\lambda_{0}^{(q-1)} \mathbf{s}_{\mathbf{q}}^{\mathbf{1}}+\lambda_{1}^{(q-1)} \mathbf{s}_{\mathbf{q}}^{\mathbf{0}} \tag{B.10}
\end{equation*}
$$

Then the first element of $\mathbf{s}_{\mathbf{q - 1}}$ satisfies the following:

$$
\begin{array}{r}
s(q-1, q-1,-1)=\lambda_{1}^{(q-1)} s(q, 0,0) \Rightarrow \lambda_{1}^{(q-1)}=\frac{s(q-1, q-1,-1)}{s(q, 0,0)} \Rightarrow \\
\lambda_{1}^{(q-1)}=\frac{s^{2}(q, 0, q) s(q, 0,1)}{s^{2}(q, 0,0) s(q, 0, q)} . \tag{B.11}
\end{array}
$$

The last non-zero element of $\mathbf{s}_{\mathbf{q}-\mathbf{1}}$ satisfies the following:

$$
\begin{equation*}
s(q-1, q-1, q)=\lambda_{0}^{(q-1)} s(q, 0, q) \Rightarrow \lambda_{0}^{(q-1)}=\frac{s^{2}(q, 0, q-1) s(q, 0,0)}{s^{2}(q, 0,0) s(q, 0, q)} . \tag{B.12}
\end{equation*}
$$

In deriving equations B. 11 and B. 12 equation B. 6 has been used. Equations B. 11 and B. 12 show that equation B. 9 is valid for $k=1$.

Assume that B. 9 is valid for $k \leq K$ where $K<q$. It is now necessary to prove that is also holds for $k=K+1$. As a result of the rank property the following equation holds:

$$
\begin{equation*}
\mathbf{s}_{\mathbf{q}-(\mathbf{K}+\mathbf{1})}=\sum_{i=0}^{K+\mathbf{1}} \lambda_{i}^{(q-(K+1))} \mathbf{s}_{\mathbf{q}}^{\mathbf{K}+\mathbf{1 - i}} . \tag{B.13}
\end{equation*}
$$

The first element of $\mathbf{s}_{\mathbf{q}-(\mathbf{K + 1})}$ satisfies the following equation:

$$
\begin{equation*}
s(q-(K+1), q-(K+1),-K-1)=\lambda_{K+1}^{(q-(K+1))} s(q, 0,0) . \tag{B.14}
\end{equation*}
$$

From the lag symmetries it is known that

$$
\begin{equation*}
s(q-(K+1), q-(K+1),-K-1)=s(q, q, K+1) \tag{B.15}
\end{equation*}
$$

Taking equation B. 5 for $\tau=K+1$ results in

$$
\begin{equation*}
s(q, q, K+1)=\frac{s(q, 0, q) s(q, 0, K+1)}{s(q, 0,0)} \tag{B.16}
\end{equation*}
$$

Finally, combining equations B.14,B. 15 and B. 16 results in the following expression for $\lambda_{K+1}^{(q-(K+1))}$ :

$$
\begin{equation*}
\lambda_{K+1}^{(q-(K+1))}=\frac{s(q, 0, K+1) s^{2}(q, 0, q)}{s^{2}(q, 0,0) s(q, 0, q)} . \tag{B.17}
\end{equation*}
$$

Equation B. 17 shows that $\lambda_{K+1}^{(q-(K+1))}$ has the form expected by equation B.9.
Assume that $\lambda_{i}^{(q-(K+1))}$ has the correct form for $i=M, \ldots, K+1$ where $1<M \leq K+1$. Now it is necessary to prove that $\lambda_{M-1}^{(q-(K+1))}$ also has the correct form. The element $s(q-(K+1), q-(K+1),-M+1)$ can be written as follows:

$$
\begin{array}{r}
s(q-(K+1), q-(K+1),-M+1)= \\
\lambda_{M-1}^{(q-(K+1))} s(q, 0,0)+\sum_{i=M}^{K+1} \lambda_{i}^{(q-(K+1))} s(q, 0, i-M+1) . \tag{B.18}
\end{array}
$$

From the symmetries of cumulant lags it is known that

$$
\begin{array}{r}
s(q-(K+1), q-(K+1),-M+1)= \\
s(q-K+M-2, q-K+M-2, M-1) \tag{B.19}
\end{array}
$$

Since $M \geq 2$ it follows that $K-M+2 \leq K$. Consequently $s(q-K+M-2, q-K+$ $M-2, M-1$ ) is given by the following expression:

$$
\begin{array}{r}
s(q-K+M-2, q-K+M-2, M-1)= \\
\sum_{i=0}^{K-M+2} \lambda_{i}^{(q-(K-M+2))} s(q, 0, i+M-1) \tag{B.20}
\end{array}
$$

where the $\lambda_{i}^{(q-(K-M+2))}$ can be obtained from equation B. 9 after replacing $k$ with ( $K-M+2$ ). To aid comparison with equation B.18, equation B. 20 can be written as
follows:

$$
\begin{array}{r}
s(q-K+M-2, q-K+M-2, M-1)=\lambda_{0}^{(q-(K-M+2))} s(q, 0,0+M-1)+ \\
\sum_{i=1}^{K-M+2} \lambda_{i}^{(q-(K-M+2))} s(q, 0, i+M-1)(1 \tag{B.21}
\end{array}
$$

It is easy to check by substituting the known expressions for the $\lambda$ 's in equation B. 18 and B. 21 such that

$$
\begin{equation*}
\sum_{i=M}^{K+1} \lambda_{i}^{(q-(K+1))} s(q, 0, i-M+1)=\sum_{i=1}^{K-M+2} \lambda_{i}^{(q-(K-M+2))} s(q, 0, i+M-1) \tag{B.22}
\end{equation*}
$$

Finally combining B. 19 and B. 22 results in the following:

$$
\begin{array}{r}
\lambda_{M-1}^{(q-(K+1))} s(q, 0,0)=\lambda_{0}^{(q-(K-M+2))} s(q, 0,0+M-1) \Rightarrow \\
\lambda_{M-1}^{(q-(K+1))} s(q, 0,0)=\frac{s(q, 0,0) s^{2}(q, 0, q-(K-M+2))}{s^{2}(q, 0,0) s(q, 0, q)} s(q, 0, M-1) \Rightarrow \\
\lambda_{M-1}^{(q-(K+1))}=\frac{s(q, 0, M-1) s^{2}(q, 0, q-(K-M+2))}{s^{2}(q, 0,0) s(q, 0, q)} \tag{B.23}
\end{array}
$$

Equation B. 23 shows that $\lambda_{M-1}^{(q-(K+1))}$ has the form expected from equation B. 9 and this concludes the proof of B.7.

Since $s(q, 0,0)$ is assumed to be non-zero then it is possible to find $\gamma_{4} \neq 0$ such that

$$
\begin{equation*}
s(q, 0, i+m)=\gamma_{4} \frac{s^{2}(q, 0,0) s(q, 0, q) s(q, 0, i+m)}{s^{4}(q, 0,0)} \tag{B.24}
\end{equation*}
$$

Substituting B. 24 into B. 7 results in
$s(q-k, q-k, m)=\sum_{i=0}^{q} \frac{s(q, 0, i) s^{2}(q, 0, i+q-k)}{s^{2}(q, 0,0) s(q, 0, q)} \gamma_{4} \frac{s^{2}(q, 0,0) s(q, 0, q) s(q, 0, i+m)}{s^{4}(q, 0,0)} \Rightarrow$
$s(q-k, q-k, m)=\gamma_{4} \sum_{i=0}^{q} \frac{s(q, 0, i) s^{2}(q, 0, i+q-k) s(q, 0, i+m)}{s^{4}(q, 0,0)}$
Equation B. 25 shows that $s(q-k, q-k, m)$ can be considered as the fourth-order cumulant with lags ( $q-k, q-k, m$ ), of an MA $(q)$ model with parameters $h(i)=$ $s(q, 0, i) / s(q, 0,0)$ for $i=0, \ldots, q$.

# List of Publications 

- A G Stogioglou, S McLaughlin "MA Parameter Estimation and Cumulant Enhancement", IEEE Transactions on Signal Processing, 44, No 7, p, July 1996.
- A G Stogioglou, S McLaughlin "ARMA Parameter Estimation Through Enhanced Double MA Modelling", $7^{\text {th }}$ European Signal Processing Conference (EUSIPCO), Trieste, September 1996.
- A G Stogioglou, S McLaughlin "Composite Cumulant-Property Mapping for MA Cumulant Matching" $8^{\text {th }}$ IEEE Signal Processing Workshop on Statistical Signal and Array Processing, Greece, June 1996.
- A G Stogioglou, S McLaughlin "Third Order Cumulant Enhancement for MA Models" Proceedings of the $4^{\text {th }}$ IEEE Workshop on Higher Order Statistics", Spain, June 1995.
- A G Stogioglou, S McLaughlin "Overcoming the effects of sampling for bandlimited stationary analogue signals using HOC-based MA modelling", Proceedings of the IEE Colloquium on Higher Order Statistics:Are they of Any use, Digest no 1995/138, pp4/1-4/8, London, May 1995.
- A G Stogioglou, S McLaughlin "New Results in Blind Deconvolution and Systems Identification of FIR Systems", Proceedings of ATHOS Workshop on Higher Order Statistics, September 1994.
- A G Stogioglou, S McLaughlin "A Robust Approach to Least Squares Solutions in HOC Based System Identification", Proceedings of the $7^{\text {th }}$ European Signal Processing Conference (EUSIPCO), pp208-212, Edinburgh 1994.
- A G Stogioglou, S McLaughlin "New Results in System Identification of FIR Systems", Proceedings of the 28th Asilomar Workshop on Circuits and Systems, pp430-435, California, November 1994.


## References

[1] G Giannakis. Cumulants: A powerful Tool in Signal Processing. Proceedings of the IEEE, 75(9):1333-1334, September 1987.
[2] G U Yule. On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. Philosophical Transactions of the Royal Society of London, A226:267-298, 1927.
[3] J M Mendel. Discrete Techniques of Parameter Estimation: the Equation Error Formulation. Marcel-Dekker, 1973.
[4] J M Mendel. Lessons in Digital Estimation Theory. Prentice-Hall, 1987.
[5] A Papoulis. Probability, Random Variables and Stochastic Processes. McGraw Hill International Editions, 3nd edition, 1991.
[6] N Wiener. Extrapolation, Interpolation, and Smoothing of Stationay Time Series. MIT Press, 1949.
[7] H Wold. The Analysis of Stationary Time Series. Almquist and Wicksell, Uppsala, Sweden, 2nd edition, 1954.
[8] S Haykin. Adaptive Filter Theory. Prentice-Hall International Editions, 2nd edition, 1991.
[9] G E P Box and G M Jenkins. Time Series Analysis, Forecasting and Control. Holden Day, 1970.
[10] L Ljung and T Soderstrom. Theory and Practice of Recursive Identification. MIT Press, Cambridge MA, 1983.
[11] P Faurre. Stochastic realisation algorithms. In R K Mehra and D G Lainiotis, editors, System Identification: Advances and Case Studies. Academic, New York, 1976.
[12] M Rosenblatt. Stationary Sequences and Random Fields. Birkhauser, 1985.
[13] T G Stockham, T M Cannon and R B Ingebretsen. Blind deconvolution through digital signal processing. Proceedings of the IEEE, 63(4):678-692, April 1975.
[14] B Porat. Digital Processing of Random Signals: Theory and Methods. Prentice Hall Information and System Sciences Series, 1994.
[15] H Akaike. Canonical correlation analysis of time series and the use of an onformation criterion. In R K Mehra and D G Lainiotis, editors, System Identification: Advances and Case Studies. Academic, 1976.
[16] J A Cadzow. Spectral estimation: An overdetermined rational model equation approach. Proceedings of the IEEE, 70:907-939, 1982.
[17] A Benveniste, M Goursat and G Roget. Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communications. IEEE Transactions on Automatic Control, AC-25:385-399, 1985.
[18] J J Kormylo and J M Mendel. Identifiability of nonminimum phase linear stochastic systems. IEEE Transaction on Automatic Control, AC-28:1081-1090, 1983.
[19] J D Scargle. Absolute value optimisation to estimate phase properties of stochastic time series. IEEE Transaction on Information Theory, IT-23:140-143, 1977.
[20] A V Oppenheim and Schafer. Digital Signal Processing. Prentice Hall, Englewood Cliffs NJ, 1975.
[21] M Boumahdi and S Mansar. Blind deconvolution of field seismic data using higherorder statistics. In IEEE Signal Processing ATHOS workshop on Higher-Order Statistics, pages 454-458, Girona, Spain, June 1995.
[22] G Giannakis and J Mendel. Tomographic wavelet estimation via higher-order statistics. In Proc. of the 56th Int. Conf. Soc. Explor. Geophys., pages 512-514, Houston, 1986.
[23] G Giannakis. Wavelet Parameter and Phase Estimation Using Cumulant Slices. IEEE Transactions on Geoscience and Remote Sensing, 27(4):452-455, 1989.
[24] T Deeming. Fourth Order Statistics for Seismic Deconvolution. In C Nikias and J Mendel, editors, Workshop on Higher-Order Spectral Analysis at Vail, Colorado, pages 907-938, June 1989.
[25] D Hatzinakos and C L Nikias. Blind equalisation using a trispectrum based approach. IEEE Transactions on Communications, 39:669-682, May 1991.
[26] M Dogan and J Mendel. On the applications of the gm equation to blind deconvolution. In Proceeding of the IEEE Signal Processing/ATHOS Workshop on HOS, pages 459-463, Parador Aguiablava, Spain, 1995.
[27] J Tugnait. Blind channel equalization and adaptive equalizer initialization. In Proc. of 1991 Int. Conf. on Communications, Denver CO, June 1991.
[28] A G Bessios and F M Caimi. Reverberation canceling for data communications in a multimodal acoustic ocean waveguide. In IEEE Signal Processing ATHOS workshop on Higher-Order Statistics, Girona, Spain, June 1995.
[29] H Mino K Yana, H Marushima and N Takeuchi. Bispectral analysis of filtered impulse process with applications to the analysis of bioelectric phenomena. In Proc. of Workshop on Higher Order Spectral Analysis, pages 140-145, Vail, Colorado, June 1989.
[30] A Ziolkowski. Why don't we measure seismic signatures? Geophysics, 56(2):190201, February 1991.
[31] K Lii and M Rosenblatt. Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes. Ann. Statist, 10:1195-1208, 1982.
[32] G Giannakis and J Mendel. Identification of Non-Minimum Phase Systems Using Higher-order Statistics. IEEE Transactions on Acoustics, Speech and Signal Processing, 37(3):360-377, March 1989.
[33] J Tugnait. Approaches to FIR System Identification with Noisy Data Using Higher Order Statistics. IEEE Transactions on Acoustics, Speech and Signal Preocessing, 38(7):1307-1317, 1990.
[34] J Tugnait. New results on FIR System Identification Using Higher Order Statistics. IEEE Transactions on Acoustics, Speech and Signal Preocessing, 39(10):2216-2221, 1991.
[35] S Alshebeili, A Venetsanopoulos and E Cetin. Cumulant Based Identification Approaches for Nonminimum Phase FIR Systems. IEEE Transactions on Signal Processing, 41(4):1576-1588, April 1993.
[36] J A R Fonollosa and J Vidal. System identification using a linear combination of cumulant slices. IEEE Transactions on Signal Procesing, 1993.
[37] A Nandi. Blind Identification of FIR Systems Using Third Order Cumulants. Signal Processing, 39(3), September 1994.
[38] X Zhang and Y Zhang. FIR System Identification Using Higher Order Statistics Alone. IEEE Transactions on Signal Processing, 42(10):2854-2858, October 1994.
[39] C Nikias and H Chiang. Higher-Order Spectrum Estimation Via Noncausal Autoregressive Modeling and Deconvolution. IEEE Transactions on Acoustics, Speech and Signal Processing, 36(12):1911-1913, December 1988.
[40] F-C Zheng. Blind Deconvolution: Techniques and Applications. PhD thesis, Department of Electrical Engineering, The University of Edinburgh, 1992.
[41] D R Brillinger. Time Series, Data Analysis and Theory. Oakland, CA: Holden Day, 1981.
[42] K L Chung. A Course in Probability Theory. Academic, 2nd edition, 1974.
[43] G Clarke. Time-Varying Deconvolution Filters. Geophysics, 33(6):936-944, 1968.
[44] B Friedlander and B Porat. Asymptotically Optimal Estimation of MA and ARMA Parameters of Non-Gaussian Processes from High-Order Moments. IEEE Transactions on Automatic Control, 35(1):27-35, January 1990.
[45] S McLaughlin F-C Zheng and B Mulgrew. Cumulant-based deconvolution and identification: Several new families of linear equations. Signal Processing, (30):199219, 1993.
[46] P McCullagh. Tensor Methods in Statistics. Chapman and Hall, 1987.
[47] T N Theile. Theory of Observations. The Annals of Mathematical Statistics, 2:165-308, 1931.
[48] R A. Fisher. Moments and Product Moments of Sampling Distributions. Proc. Lond. Math. Soc, 30(Series 2):199-238, 1931.
[49] V C Leonov and A N Shiryaev. On the Technique of Computing Semi-Invariants. Theory of Probability Applications, 38(4):319-329, 1959.
[50] D Brillinger. An introduction to polyspectra. Annals of Mathematical Statistics, 36:1351-1374, 1965.
[51] D Brillinger and M Rosenblatt. Computation and Iterpretation of k-th Order Spectra. In B Harris, editor, Spectral Analysis of Time Signals, pages 907-938. Wiley, 1967.
[52] W Feller. An Introduction to Probability Theory and Its Applications. New York, Wiley, 1971.
[53] M. Rosenblatt. Stationary Sequences and Random Fields. Birkhauser, 1985, Boston MA.
[54] C L Nikias and A P Petropoulou. Higher-Order Spectral Analysis: A Nonlinear Signal Processing Framework. Prentice-Hall Signal Processing Series, 1st edition, 1993.
[55] A L Pflug, G E Ioup, J W Ioup and R L Fiels. Properties of Higher-Order Correlations and Spectra for Band Limited Deterministic Transients. J. Acoust. Soc. Am., 91(2):975-988, February 1992.
[56] J G Proakis and D G Manolakis. Digital Signal Processing: Principles, Algorithms and Applications. Macmillan, 1992.
[57] B Porat and B Friedlander. Performance Analysis of Parameter Estimation Algorithms Based on Higher-Order Moments. International Journal of Adaptive Control and signal Processing, 3:191-229, 1989.
[58] A Swami and J Mendel. Closed-Form Recursive Estimation of MA Coefficients Using Autocorrelations and Third-Order Cumulants. IEEE Transactions on Acoustics, Speech and Signal Processing, 37(11):1794-1795, November 1989.
[59] M Rosenblatt and J Van Ness. Estimation of the Bispectrum. Annals of Mathematical Statistics, 36:1120-1136, 1965.
[60] G B Giannakis and A Swami. On estimating noncausal nonminimum phase arma models of non-gaussian processes. IEEE Transactions on ASSP, 38(3):478-495, March 1990.
[61] G B Giannakis and M K Tsatsanis. A Unifying Maximum-Likelihood View of Cumulant and Polyspectral Measures for Non-Gaussian Signal Classification and Estimation. IEEE Transactions on Information Theory, 38(2):386-406, 1992.
[62] A Nandi. Robust estimation of third-order cumulants in applications of higherorder statistics. IEE PROCEEDINGS F. Radar and Signal Processing, 6, December 1993.
[63] J M Mendel. Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications. Proceedings of the IEEE, 79:278-305, March 1991.
[64] J K Tugnait. Identification of linear stochastic systems via second- and fourthorder cumulant matching. IEEE Transactions on Information Theory, 1987.
[65] A G Stogioglou, S McLaughlin. New Results in System Identification of FIR Systems. In Proceedings of ASILOMAR Conference, November 1994.
[66] G Golub and C Van Loan. Matrix Computations. The Johns Hopkins University Press, 1983.
[67] G Giannakis and J Mendel. Cumulant-Based Order Determination of NonGaussian ARMA Models. IEEE Transactions on Acoustics, Speech and Signal Processing, 38(8):1411-1423, August 1990.
[68] X Zhang and Y Zhang. Singular Value Decomposition-Based MA Order Determination of Non-Gaussian ARMA Models. IEEE Transactions on Signal Processing, 41(8):2657-2873, August 1993.
[69] J A R Fonollosa and J Vidal. System Identification Using a Linear Combination of Cumulant Slices. IEEE Transactions on Signal Processing, 41(7):2405-2412, July 1993.
[70] G Giannakis. On the Identifiability of Non-Gaussian ARMA Models Using Cumulants. IEEE Transactions on Automatic Control, 35(1):18-26, January 1990.
[71] M J Hinich and M A Wolinsky. A test for aliasing using bispectral components. J. Amer. Stat. Assoc. Theory and Methods, 83(402):499-502, 1993.
[72] P Huber. Robust Statistics. Wiley, 1981.
[73] J L Peyre, J L LeRoux, D Rossille and G Favier. MA HOS Modelization from Samples of Band-Limited Third Order Stationary Signals. In Proceedings of ATHOS workshop on HOS, September 1994.
[74] K Konstantinides and K Yao. Statistical Analysis of Effective Singular Values in Matrix Rank Determination. IEEE Transactions on Acoustics, Speech and Signal Preocessing, 36(5):757-763, 1988.
[75] C Coroyer J LeRoux and D Rossille. Illustration of the Effects of Sampling on Higher-Order Spectra. Signal Processing, (36):375-390, 1994.
[76] C Mullins. Multiple Window Cumulant Estimation. In IEEE Workshop on HigherOrder Spectral Analysis at Lake Tahoe, California, June 1993.
[77] H Nguyen and G Rogers. Fundamentals of Mathematical Statistics, volume 1 and 2. Springer-Verlag Texts in Statistics, 1989.
[78] C Nikias and M Raghuveer. Bispectrum Estimation: A Digital Signal Processing Framework. Proceedings of the IEEE, 75(7):869-891, 1987.
[79] J G Proakis C M Rader F Ling and C L Nikias. Advanced Digital Signal Processing. Maxwell Macmillan.
[80] S Orfanides. Optimum Signal Processing: An Introduction. McGraw Hill International Editions, Electrical Engineering Series, 1988.
[81] A T Parsons and M L Williams. Limitations on the Use of Discrete Linear Models of Continuous Random Processes. Circuits, Systems and Signal Processing, 13(4):403-410, 1994.
[82] M Raghuveer and C Nikias. Bispectrum Estimation: A Parametric Approach. IEEE Transactions on Acoustics, Speech and Signal Processing, 33(4):1213-1230, October 1985.
[83] A Swami and J Mendel. ARMA Parameter Estimation Using Only Output Cumulants. IEEE Transactions on Acoustics, Speech, and Signal Processing, 38(7):12571265, July 1990.
[84] C Therrien. Discrete Random Signals and Statistical Signal Processing. PrebticeHall International Editions, 1992.
[85] A Walden and J Hosken. An Investigation of the Spectral Properties of Primary Reflection Coefficients. Geophysical Prospecting, 33:400-435, 1985.
[86] A Walden. Non-Gaussian Reflectivity, Entropy, and Deconvolution. Geophysics, 50(12):2862-2888, December 1985.
[87] A Walden and J Hosken. The Nature of the Non-Gaussianity of Primary Reflection Coefficients and its Significance for Deconvolution. Geophysical Prospecting, 34:1038-1066, 1986.
[88] A Stogioglou and S McLaughlin. Blind Deconvolution of FIR Systems: New Results. ATHOS Workshop on Higher Order Statistics in Signal Processing, September 1994.
[89] H Chiang and C Nikias. Adaptive Deconvolution and Identification of Nonminimum Phase FIR Systems Based on Cumulants. IEEE Transactions on Automatic Control, 35(1):36-47, January 1990.
[90] A G Stogioglou, S McLaughlin. MA Cumulant Enhancement and Parameter Estimation. IEEE Transaction on Signal Processing, July 1996.
[91] J Cadzow. Signal Enhancement- A Composite Property Mapping Algorithm. IEEE Transactions on Acoustics, Speech and Signal Preocessing, 36(1):49-62, 1988.
[92] J A Cadzow and D M Wilkes. Enhanced rational signal modeling. Signal Processing, 25:171-188, 1991.
[93] W Zangwill. Nonlinear Programming: A Unified Approach. Prentice-Hall, Englewood Cliffs, NJ, 1969.
[94] J-F Cardoso. Fourth-Order Cumulant Structure Forcing. Application to Blind Array Processing. In Proceedings IEEE SP Workshop on SSAP-92, 1992.
[95] I Dologlou, J-C Pesquet and J Skowronski. Projection Based Rank Reduction Algorithms for Multichannel Modelling and Image Compression. Subitted to Signal Processing, 1995.
[96] P L Combettes. The Foundations of set Theoretic Estimation. Proceedings of the IEEE, 81(2):181208, February 1993.
[97] A Swami and J Mendel. Time and Lag Recursive Computation of Cumulants from a State-Space Model. IEEE Transactions on Automatic Control, 35(1):4-17, November 1989.
[98] J W Brewer. Kronecker products and matrix calculus in system theory. IEEE Transactions on Circuits and Systems, CAS-25(9), 1978.


[^0]:    ${ }^{1}$ It is interesting to point out that the definition of random signals used here, includes deterministic signals observed in random noise.

[^1]:    ${ }^{1}$ Provided that such expansion exists.

[^2]:    ${ }^{2}$ The transfer function is defined as the $z$-transform of the impulse response. The $z$-transform is defined in Appendix A.

[^3]:    ${ }^{3}$ The term all-zero is misleading since it implies that the transfer function has only zeros and no poles, while in reality the transfer function has $z^{q}$ as a denominator and 0 is a pole with multiplicity $q$.
    ${ }^{4}$ Since this is a causal AR model it must be minimum phase.

[^4]:    ${ }^{5}$ Which means that there are no pole-zero cancelations

[^5]:    ${ }^{6}$ In case $w(i)$ is not IID we require that its moments are absolutely summable.

[^6]:    ${ }^{7}$ which is widely used in practice.

[^7]:    ${ }^{1}$ The asymptotic variance of sample fourth-order cumulants depend on up to eighth-order cumulants.

[^8]:    ${ }^{2}$ In the case where the statistics are sample cumulants, the convergence is in probability.

[^9]:    ${ }^{3}$ Excluding equations is not always possible. We should always make sure that the exclusion of the selected equations does not affect the consistency and uniqueness of the least squares solution.

[^10]:    ${ }^{4}$ The minimal set of equations that ensures identifiability is discussed later in section (3.3)

[^11]:    ${ }^{5}$ Here "large" means lags near to the model order $q$.

[^12]:    ${ }^{1}$ The expression for the variance of the fourth-order sample cumulants involves up to eighth-order cumulants.
    ${ }^{2}$ The term blind refers to the fact that we perform deconvolution without prior knowledge of the input sequence $\{w(n)\}$.

[^13]:    ${ }^{3}$ Another option is to use $m_{1}=q$.

[^14]:    ${ }^{1}$ The symbol ()$^{\top}$ is used to denote transpose.

[^15]:    ${ }^{2}$ The set has $(q+1)(q+2) / 2$ elements since this is the degrees of freedom of the linear structure.

[^16]:    ${ }^{3}$ Assuming that $s(0, q), s(q, q) \neq 0$.

[^17]:    ${ }^{4}$ The slice is taken parallel and above the plane of $\tau_{1}, \tau_{2}$

[^18]:    ${ }^{5}$ The notation of the fourth-order cumulant vectors in this section is the same as the notation of the third-order cumulant vectors of section 5.2.1. This should not cause any confusion since the the present section and section 5.2.1 are independent of each other.

[^19]:    ${ }^{6}(n=2 q+1, m>n)$

[^20]:    ${ }^{7}$ Theorem 5.2 requires some extra conditions which are assumed to be satisfied here.

[^21]:    ${ }^{8}$ The set $\mathcal{S}_{q+1}^{\prime}$ of ( $m \times n$ ) matrices with rank less than or equal to $q+1$, is a cone since if $\mathbf{X} \in \boldsymbol{S}_{q+1}^{\prime}$ then $\forall \alpha \in \mathbb{R}, \alpha \mathbf{X} \in \mathcal{S}_{q+1}^{\prime}$. On the other hand $\mathcal{S}_{q+1}^{\prime}$ is non-convex adding two matrices from $\mathcal{S}_{q+1}^{\prime}$ results in a matrix with rank in general larger than $q+1$.
    ${ }^{9}$ If $\mathbf{X}$ is an $(m \times n)$ matrix, then the $(i, j)$-element of $\mathbf{X}$, equals the ( $i-j m-m$ )-element of $\mathcal{T}(\mathbf{X})$.

[^22]:    ${ }^{10}$ This is the set of matrices in $\mathcal{H}$ with rank less than or equal to $q+1$.

