
Hybrid Solutions to Instantaneous MIMO Blind Separation and Decoding

: Narrowband, QAM and Square Cases

Xu Zhao



A thesis submitted for the degree of Doctor of Philosophy.

The University of Edinburgh.

September 2009

Abstract

Future wireless communication systems are desired to support high data rates and high quality transmission when considering the growing multimedia applications. Increasing the channel throughput leads to the multiple input and multiple output and blind equalization techniques in recent years. Thereby blind MIMO equalization has attracted a great interest.

Both system performance and computational complexities play important roles in real time communications. Reducing the computational load and providing accurate performances are the main challenges in present systems. In this thesis, a hybrid method which can provide an affordable complexity with good performance for Blind Equalization in large constellation MIMO systems is proposed first. Saving computational cost happens both in the signal separation part and in signal detection part. First, based on Quadrature amplitude modulation signal characteristics, an efficient and simple nonlinear function for the Independent Component Analysis is introduced. Second, using the idea of the sphere decoding, we choose the soft information of channels in a sphere, and overcome the so-called curse of dimensionality of the Expectation Maximization (EM) algorithm and enhance the final results simultaneously. Mathematically, we demonstrate in the digital communication cases, the EM algorithm shows Newton-like convergence.

Despite the widespread use of forward-error coding (FEC), most multiple input multiple output (MIMO) blind channel estimation techniques ignore its presence, and instead make the simplifying assumption that the transmitted symbols are uncoded. However, FEC induces code structure in the transmitted sequence that can be exploited to improve blind MIMO channel estimates. In final part of this work, we exploit the iterative channel estimation and decoding performance for blind MIMO equalization. Experiments show the improvements achievable by exploiting the existence of coding structures and that it can access the performance of a BCJR equalizer with perfect channel information in a reasonable SNR range. All results are confirmed experimentally for the example of blind equalization in block fading MIMO systems.

Declaration of originality

I hereby declare that the research recorded in this thesis and the thesis itself was composed and originated entirely by myself in the Department of Electronics and Electrical Engineering at The University of Edinburgh.

XU ZHAO

Acknowledgements

Numerous faculty, friends, and family members have helped me to complete this dissertation. I would like to acknowledge with great appreciation the following individuals for their support and assistance.

My advisor, Professor Mike Davies: I would like to express my gratitude to Professor Mike Davies, my advisor and mentor, for his guidance, understanding and patience, and most importantly his friendship. He has been a strong and supportive advisor to me throughout my graduate studies. His mentorship provided me with a well-rounded research experience and formed the framework for my long-term research goal. His timely and valuable comments and evaluation at every stage of the process helped me complete this dissertation.

I would also like to thank the members in the institute of digital communications during my research study. Special thanks to Dr. John Thompson and Dr. Luis Barbero for their willingness to discuss with me the list sphere decoder. Thanks to Dr. Goertz Norbert and Miss Xiaoyan for the friendly interactive communications of the channel error correcting code. Also I owe a special debt to Dr. Mike Novey from University of Maryland, Baltimore County. I appreciate his valuable insights and help during the development of the fixed point algorithm. Dr. Mike Novey helped to make mathematical derivations an appealing subject area.

Lastly, I want to thank my parents, and my sister for financial and emotional support. Above all, I want to thank my affectionate wife for her love, sacrifices and kind indulgence. I also want to thank my new born son Li-an for inspiring and amazing me every day. Without them, I might never have finished this dissertation.

Contents

Declaration of originality	iii
Acknowledgements	iv
Contents	v
List of figures	vii
Acronyms and abbreviations	x
List of principal symbols	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Channel Equalization	2
1.3 Exploring Signal Properties	3
1.4 The Channel Model	4
1.5 Contributions and Thesis Outline	5
2 Review of Blind Equalization	7
2.1 Introduction	7
2.2 Notation and Statistical Prerequisites	7
2.3 Present Pilot Assisted Wireless Transmissions	10
2.4 Brief Survey of Blind Equalization	11
2.5 Implicit HOS – Bussgang Type Methods	14
2.5.1 Sato - Birth of BE	14
2.5.2 BGR Algorithm	15
2.5.3 Godard Algorithm	16
2.5.4 Constant Modulus Algorithm	16
2.5.5 Stop-and-Go Algorithm	19
2.5.6 Shalvi and Weinstein Algorithm	19
2.6 Multiple Input Multiple Output Systems	21
2.6.1 Second Order Statistics	24
2.6.2 HOS Methods	29
2.7 ICA for Blind Instantaneous MIMO Separation and Equalization	34
2.7.1 Contrast Functions and Basic Algorithms	35
2.7.2 Typical Complex ICA Algorithms for Blind Separations	39
2.8 Chapter Summary	45
3 An Efficient Nonlinearity for QAM signals	47
3.1 Introduction	47
3.2 Optimum Nonlinear Functions	47
3.3 PDF of QAM modulations	48
3.4 A Simple Nonlinear Functions Based On The QAM modulation	50
3.4.1 Split Complex Nonlinearities	51
3.4.2 Gradient Algorithm of Threshold Nonlinear ICA	54
3.4.3 Fixed-Point Algorithm for Threshold Nonlinear ICA	55

3.4.4	Local Stability Analysis	56
3.5	Nonlinear Threshold ICA With Bias Removal	59
3.6	Simulations	60
3.7	Chapter Summary	68
4	Blind MIMO Equalization and Decoding for Square QAM Modulation	70
4.1	Introduction	70
4.2	The EM Algorithm	71
4.2.1	The EM Principal	73
4.3	EM Algorithm for Noisy ICA	74
4.3.1	Convergence in Continuous Domains	75
4.3.2	Convergence in Discrete Domains	76
4.3.3	Limitations of EM	81
4.4	Sphere Decoder for MIMO Detection	82
4.4.1	List Fixed Sphere Decoder	85
4.5	SD EM algorithm	87
4.6	Simulations	89
4.7	Chapter Summary	94
5	Coding Assisted Blind MIMO Separation and Decoding	95
5.1	Introduction	95
5.2	The Proposed Iterative Procedure	99
5.2.1	Soft MIMO Likelihood and The BCJR Decoder	100
5.2.2	Soft Mapper and Demapper	103
5.2.3	EM Channel Estimation with a-Posteriori Probability	104
5.3	Codeword Over Multiple Blocks	106
5.4	Simulations	107
5.5	Chapter Summary	115
6	Conclusions and Future Work	116
6.1	Conclusions	116
6.2	Future Work	117
6.2.1	Convolutional Channel Estimate	117
6.2.2	Semi-Blind Methods	118
6.2.3	Complexity Reduction	118
A	Derivation of equation (3.42)	119
B	Proof of Equation (4.36)	122
C	Channels Used in Chapter V Simulations	124
	References	126

List of figures

1.1	Typical wireless communications.	2
2.1	Pilot placement patterns in existing wireless systems. Left: packet transmissions. Right: continuous transmissions. Shadow areas are pilot symbols.	11
2.2	Architecture of the SISO blind equalization.	13
2.3	CMA blind equalization.	17
2.4	Performance of MMA algorithm with different learning rates.	18
2.5	Architecture of Blind Equalization of MIMO system.	22
2.6	Architecture of a fractional sampling system.	25
2.7	Equalized signals by subspace blind equalization.	28
2.8	Received and separated signals by circular complex fast ICA in a 4×4 system with 16-QAM modulation.	42
2.9	BER and separability performances of circular complex fast ICA in a 4×4 system with 4-QAM and 16-QAM modulations.	42
2.10	Received and separated signals by JADE method in a 4×4 system with 16-QAM modulation.	44
2.11	BER and separability performances of the JADE algorithm in a 4×4 system with 4-QAM and 16-QAM modulations.	44
3.1	MoG PDF of 16-QAM with different variances.	49
3.2	Nonlinear approximation of MoG.	51
3.3	Nonlinear approximation of MoG.	53
3.4	Nonlinear threshold function and nonlinear selective function.	56
3.5	Separated signals. (a) : The complex fixed fast ICA algorithm [1]; (b): the JADE algorithm [2].	61
3.6	Separated signals. (a): the Douglas [3] algorithm [3]; (d): The proposed nonlinear threshold ICA algorithm.	61
3.7	Convergence Properties of the Threshold Nonlinearity.	62
3.8	Comparison of different threshold nonlinearities in noisy cases. The clipped version of pdf Vs. approximated version of pdf.	63
3.9	Separability of the Threshold Nonlinearity ICA vs. other complex ICA methods in a 4×4 system with 16-QAM modulation.	64
3.10	Separability of the Threshold Nonlinearity ICA vs. other complex ICA methods in a 9×9 system with 4-QAM,16-QAM and 64-QAM modulations simultaneously.	65
3.11	ICI vs. number of snapshots N for the different algorithms.	65
3.12	SNR = 20 dB; ICI vs. Number of sources	66
3.13	SNR = 15 dB; Separation performance of bias removal algorithm for QAM signals.	67
3.14	Separation performance of bias removal algorithm for QAM signals over various SNR with sample length of 256 symbols.	67

4.1	BER performance of different MIMO detectors.	72
4.2	EM likelihood surface.	81
4.3	Idea of choosing the admissible set.	83
4.4	Illustration of tree search of SD in a 4×4 system with 4-QAM modulation, from [4].	85
4.5	Tree Structure for the List Fixed Sphere Decoder.	87
4.6	BER performance of fixed sphere decoder in a 4×4 system with 4-QAM and 16-QAM modulation.	88
4.7	EM Log likelihood VS. True likelihood.	89
4.8	Likelihood of EM iterations and standard Gradient method.	90
4.9	BER Performance of ICA + EM method in a 2×2 system with 4-QAM modulation.	91
4.10	Separation quality of different ICA algorithms.	92
4.11	BER performance of the ZF algorithm, Nonlinear Threshold ICA algorithm, the SD-EM algorithm and SD with perfect CSI in a 4×4 MIMO system with 16-QAM modulation.	92
4.12	BER performance of the ZF algorithm, Nonlinear Threshold ICA algorithm, the SD-EM algorithm and SD with perfect CSI in a 4×4 MIMO system with 64-QAM modulation.	93
4.13	BER performance of the ICA-SD algorithm and the SD-EM algorithm in a 4×4 MIMO system with 16-QAM modulation. Left : 256 symbols in each block; right : 512 symbols in each block.	94
5.1	Joint channel estimation and symbol detection: Uncoded diagram vs. FEC diagram.	96
5.2	Receiver architecture of the proposed coding assisted system.	100
5.3	BER performance of a 2×2 uncoded MIMO system. An EM channel estimation with BPSK modulation. ML is obtained by the EM with perfect CSI.	108
5.4	Channel separability of the split threshold nonlinear ICA, the SD-EM and the coding assisted SD-EM algorithm with a rate $r = 1/2$ convolutional code over different SNR in 4×4 MIMO systems.	110
5.5	Channel separability of the split threshold nonlinear ICA, the SD-EM and the coding assisted SD-EM algorithm with a rate $r = 1/2$ convolutional code over different SNR in 8×8 MIMO systems.	110
5.6	BER performance of the ZF scheme, the split threshold nonlinear ICA, the SD-EM, the coding assisted SD-EM algorithm and known CSI SD with a rate $r = 1/2$ convolutional code in 4×4 MIMO systems.	111
5.7	BER performance of the ZF scheme, the split threshold nonlinear ICA, the SD-EM, the coding assisted SD-EM algorithm and known CSI SD with a rate $r = 1/2$ convolutional code in 8×8 MIMO systems.	112
5.8	BER improvements by utilizing time diversity and channel coding. Triangle line is the performance of the SD-EM algorithm following the BCJR algorithm and the square line is the performance of the coding assisted SD-EM algorithm. 2 blocks form a codeword.	112

5.9	BER improvements by utilizing time diversity and channel coding. Triangle line is the performance of the SD-EM algorithm following the BCJR algorithm and the square line is the performance of the coding assisted SD-EM algorithm. 4 blocks form a codeword.	113
5.10	BER improvements by utilizing time diversity and channel coding. Triangle line is the performance of the SD-EM algorithm following the BCJR algorithm and the square line is the performance of the coding assisted SD-EM algorithm. 8 blocks form a codeword.	113
5.11	Iterative BER improvements of the coding assisted SD-EM algorithm in a 4×4 system with 16-QAM modulation, eight iterations are used.	114
5.12	Iterative channel improvements of the coding assisted SD-EM algorithm in a 4×4 system with 16-QAM modulation.	115

Acronyms and abbreviations

APP	a posteriori probabilities
BCJR	Bahl-Cocke-Jelinek-Raviv
BE	blind equalization
BER	bit error rate
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BICM	bit interleaved coded modulation
BLAST	Bell laboratories Layered Space Time
BSS	blind source separation
CCI	cochannel interference
CMA	Constant Modulus Algorithm
CSI	channel state information
DAPE	differential amplitude phase encoded
DML	deterministic maximum likelihood
EASI	equivariant adaptive separation via independence
EM	expectation maximization
FEC	forward-error coding
FIR	finite impulse response
FOBI	Fourth-Order Blind Identification
FSD	fixed complexity sphere decoder
GSA	Generalized Sato Algorithm
HMM	hidden Markov model
HOS	high order statistics
I-MIMO	instantaneous MIMO
ICA	independent component analysis
ICI	Inter-Component Interference
IFC	inverse filter criterion
i.i.d	independent, identically distributed
ISI	intersymbol interference
JADE	joint approximate diagonalization of eigenmatrices

KL	Kullback Leibler
LFSD	list fixed-complexity sphere decoder
LLR	log-likelihood ratio
LOS	line-of-sight
LS	least square
LSCMA	Least squares constant modulus algorithm
LSD	List sphere decoding
LTl	linear time-invariant
MAI	multiple access interference
MCEM	Monte Carlo EM
MIMO	multiple input multiple output
ML	maximum likelihood
MMA	multimodulus algorithm
MMSE	minimum mean square error
MNC	maximum normalized cumulant
MoG	Mixture of Gaussian
MP	matrix pencil
MSE	mean square errors
non-MSE	non-minimum square error
NSC	non-systematic convolutional
OFDM	Orthogonal frequency division multiplexing
PAM	pulse amplitude modulation
PCA	principal components analysis
pdf	probability density function
pmf	probability mass function
PSAM	Pilot Symbol Assisted Modulation
PSP	Per-Survivor Processing
QAM	quadrature amplitude modulation
RCA	reduced constellation algorithm
SD	sphere decoding
SE	Schnorr-Euchner
SEA	super exponential algorithm
SIMO	single input multiple output

SISO	single input single output
SML	stochastic maximum likelihood
SOCS	second order cyclostationary statistics
SOS	second order statistics
SRM	subchannel response matching
STF	space-time-frequency
SVD	singular value decomposition
WSCS	wide sense cyclostationary
ZF	zero forcing

List of principal symbols

\sim	Distributed as
\otimes	Kronecker product
\odot	Element-wise multiplication
$\tilde{\nabla}^*$	conjugate gradient operation
$(\bullet)'$	First differential operation
$(\bullet)^{1/2}$	Square root matrix operation
$(\bullet)^T$	Transpose operation
$(\bullet)^H$	Conjugate transpose operation
$\Psi(\bullet)$	Cost function
ϵ_s	Distribution of nodes
$vec(\bullet)$	Stacked columns of the matrix \bullet
Σ	Noise covariance matrix
$CN(0, 1)$	Complex Gaussian distribution with zero mean and normalized convariance
$C_{p,q}(s)$	$p + q$ th-order cumulant of a random variable s
$d^2(\bullet)$	Second differential operation
$\exp(x)$	Exponential function e^x
$E\{s\}$	The expectation of a random variable s
$g(\bullet)$	Nonlinear function
$H(s)$	Differential entropy of a random vector s
$h_{i,j}$	The i th row j th column element
H	Channel matrix
H^{-1}	Inverse of channel matrix
\mathcal{H}	Complex Hessian
\mathbf{I}	Identity matrix
$I(s)$	Mutual information of a random vector s
$Im(\bullet)$	Imaginary part part
$J(\bullet)$	Mean cost function

$\kappa(\mathbf{s})$	The kurtosis of a random vector \mathbf{s}
$\kappa_4(\mathbf{s})$	The normalized kurtosis of a random vector \mathbf{s}
$L(\bullet)$	Log-likelihood function
N	Noise Matrix with covariance, Σ
N_T	Number of transmitted antennas
N_R	Number of receive antennas
$p(\bullet)$	probability density function
$Re(\bullet)$	Real part
$r[s]$	The autocorrelation function of \mathbf{s}
\mathbf{s}	Random variable
\mathbf{s}	Source vector
T	Number of symbols per block
\mathbf{u}	Equalized signal
\mathbf{w}	Equalizer coefficients
\mathbf{W}	Separation matrix or demixing matrix
\mathbf{x}	Whitening observed vector
$\mathbf{y}(k)$	Received signal at time index k
\mathbf{y}	Observed vector
\mathbf{Y}	Received signal Matrix

Chapter 1

Introduction

1.1 Motivation

For communications, the ultimate aim is to design a system which can carry information from one place to other places with reliability, effectivity, celerity, safety and high capacity. Different applying environments take people various communication systems. In wireless communication, the transmission is impaired by the channel due to the complicated surroundings as illustrated in the figure 1.1. Among them, three factors have an influence on final performance.

1. Path loss, which is a natural phenomenon of decreasing received power with increasing distance.
2. Slow fading, which is caused by obstruction of buildings, hills, trees and foliage.
3. Fast fading, which is caused by multipath reflection of a transmitted wave by objects.

Moreover, the mobility and flexibility of wireless transmitters and receivers leads to the received signals variable, the multiple access leads to new interferences and present a significant problem if the power level of the desired signal is significantly lower than the power level of the interfering. The trend of low power design increases all of these challenges above and makes realistic design more complex and intricate.

Even though people have to confront these difficulties in wireless media, it is still appealing. In comparison with wired communications, wireless communications has no cables and provides mobility, flexibility, scalability and many other environments that cable can not reach easily, such as in the rural, deep space and in emergency cases. Then, variety of commercial and government innovations and inventions had been invented, discovered and developed and such contributions led to an information based modern society. More demands for high quality service in such areas as mobile phone, the internet and multimedia is rapidly and constantly increasing. Consequently, this growth brings out more demands for channel capacity in wireless communications. In order to maintain high data rates over wireless channels, people have

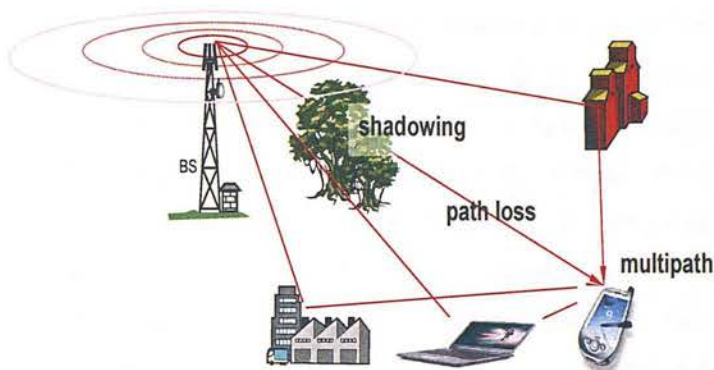


Figure 1.1: *Typical wireless communications.*

resorted to the equalization which is essentially a signal processing procedure to restore the distorted signals introduced by the channels. If the channel state information (CSI) is known, compensations for the intersymbol interference (ISI) or co-channel interference (CCI) is relatively easy to design at the receivers. However, in practice, channel information is usually not available in wireless systems. The receiver has to estimate the channels response, then equalize the received signal based on the estimated channel.

1.2 Channel Equalization

The conventional and most important equalization is the training based method and this kind of method is well developed in practice. Pilot signals are inserted periodically in the transmitted block along with data symbols. Receivers estimate the channel response based on these pilot sequences. In the presence of noise, more pilots are needed to improve the estimation accuracy.

As stated above, the wireless channel is a natural resource where the rate of communication is restricted by bandwidth and noise, with radio spectrum a scarce and expensive resource. To increase the data rate, one way is to explore methods approximating the channel capacity, such as two advanced channel coding techniques: turbo codes [5] and low density parity check codes [6]. Both of them make it possible to approach the Shannon capacity limit [7] for a single antenna link. Further advances are available through increasing the number of antennae at both the transmitter and the receiver. Recent research [8] in information theory has shown that large gains in capacity of communication over wireless channels are feasible in multiple-input multiple output (MIMO) systems, which grows approximately linearly with the number

of antennae. It has been demonstrated that the Bell laboratories Layered Space-Time (BLAST) coding technique [9]. It can attain the spatial efficiencies up to 42 bits/sec/Hz at a SNR of 20 dB in a 8×8 MIMO system. This represents an increase compared to current spectral efficiencies of 2-3 bits/sec/Hz, in cellular mobile and wireless LAN systems. These channel capacity limits highlight the bandwidth efficiency of MIMO channels. A typical way of improving this bandwidth efficiency is the blind channel estimation and decoding. This kind of methods allow us to calculate channel parameters and acquire operational conditions using only information symbols. Blind equalization (BE) became an interesting research area in digital signal processing because of particular features and it has received tremendous attention in the recent years. However, if any explicit prior knowledge is available about the transmitted sequences, such information should be exploited. Training data or pilot signals are generally available for other purposes, such as synchronization, in practical communications systems.

Using the knowledge of such signals, in conjunction with blind algorithms, leads to semi-blind methods. This hybrid method, using short training sequences and other mature methods to execute channel equalization, can achieve faster convergence and resolve some ambiguity problems that are typical in blind methods. Semi-blind methods are therefore potential solutions for near future practical wireless systems.

Another channel equalization technique uses the detected data to enhance the channel estimate giving rise to an iterative scheme for channel estimate and data detection. This kind of data aided channel equalization technique jointly estimates channel and detects data and provides more promising performance of achieving better spectral efficiency in fading channel in wireless communications. We will explore this in depth in the chapter 5.

1.3 Exploring Signal Properties

Blindness refers to the fact that there are no or little known pilot signals or training data. Only some statistical or structural properties of the transmitted and received signal are exploited in the process of adapting the equalizer. The developed algorithms depend on some statistical and structure properties of the transmitted signals. Temporal and spatial characteristics are often considered. Other information widely used in blind estimation include, finite alphabets, Bussgang statistics, cyclostationary, high order statistics (HOS), independence, shaping statistics (pre-coding) and matrix structure, such as Vandermonde or Toeplitz structure of the channel.

Blind receivers typically exploit one or a combination of these properties in order to acquire or track channel information for equalization and detection. Based on these characteristics, many algorithms for the solution of blind equalization have been proposed over the years.

1.4 The Channel Model

In the sense of wireless MIMO channel modeling, both the short-term effects caused by multipath propagation and the typical long-term effects, like path loss variations due to shadowing are considered in this thesis.

Multipath propagation causing spreading of the received signal in the angular and time domains are called frequency selective fading and time selective fading respectively. These components also are significant and meaningful in practice, but in this work, we only focus on the fundamental instantaneous narrow band MIMO systems. In narrow band environments, the bandwidth of the message does not significantly exceed the channel coherence bandwidth. Then the channel propagation does not cause significant spreading in the delay, angular and frequency domains. The wireless channel in this work concentrates on the instantaneous MIMO channel. This channel model is basic for MIMO systems and has attracted many studies [9][8].

In wireless communications, even if the channel changes frequently, it is assumed to be invariant during the whole block period. Then most algorithms in this thesis are based on a transmission block, where a group of symbols is transmitted as a unit and may be encoded for error control purposes. We have concentrated on the linear time-invariant (LTI) systems which are widely exploited in wireless community since the wireless propagation channel can often be modeled by linear systems. These LTI channel models actually are common flat fading (Rayleigh) channels. The entries in these channels are independent, identically distributed (i.i.d) circular symmetric complex Gaussian. Such a model is suitable for scenarios under the following conditions. First, received signals are a combination of multipath. These typically happens in wireless communications where the surrounding environments of both transmitter and receiver arrays possess some scatterers. Secondly, each antenna within both array needed to be widely separated which is usually greater than $\frac{1}{2}$ radio wavelength and then spatial correlations do not exist. Lastly, the line-of-sight (LOS) path is absent. Mathematically, these conditions guarantee that the elements within the channel matrix are zero mean complex Gaussian with unit magnitude variances. Then, the envelopes of the channel entries satisfy the well

known Rayleigh distribution. In wireless communications, this Rayleigh i.i.d channel is always used to model urban and indoor environments.

1.5 Contributions and Thesis Outline

This work focuses on blind equalization in large constellation MIMO systems with block transmission. The contribution of this study is mainly addressed in five aspects.

Firstly, based on the probability density function of modulation signals at digital transmitters, a simple and efficient nonlinear blind equalization method is proposed. Good separability and BER performance are obtained. The stability, convergence properties are also explored.

Second, for MIMO systems with high dimensional transmitter and receivers, eg. 4×4 and 8×8 systems, and dense constellation modulation such as 4-QAM and 16-QAM systems, an accurate blind detector are presented. This detector makes use of a maximum likelihood (ML) estimation to outperform the traditional linear MIMO detectors with affordable complexity.

Third, the Newton-like convergence property of this ML estimation are shown mathematically, the proof and its derivation are given in detail.

Furthermore, we explore the error correct code in this blind equalization system and design an iterative blind separation and decoding architecture for MIMO systems correspondingly. Simulations show that the coding structure in the transmitter can improve the final performance significantly.

In the last part of this study, with coding structure for multiple blocks, we can recover information effected by the deep fading or a very singular channel matrix.

The thesis is organized as follows:

Chapter 2 : A brief review of blind equalization is given. Major and popular algorithms are introduced, including single input single output and multiple input multiple output techniques. The performance of these algorithms are shown by explicit simulations. The independent component analysis (ICA) method is emphasized in detail for the foundation of this work and the suitability of ICA for MIMO systems is highlighted.

Chapter 3 : A simple nonlinear function based on QAM modulation is proposed. The gradient

form is developed first. The fixed point algorithm based upon this nonlinearity is also derived and the bias removal update is given as well. Stability analysis reveals this nonlinearity of the proposed algorithm is robust. Furthermore, this nonlinearity is flexible for many QAM distributions. We emphasize that this system can be easily implemented and integrated into real hardware systems.

Chapter 4 : A hybrid solution to MIMO blind separation and decoding is presented. This scheme provides good performance with an affordable complexity for large constellation MIMO systems. Computational savings happen both in the signal separation part and in the signal detection part. By utilizing a list sphere decoding and then obtaining the small set of symbol candidates, the so-called curse of dimensionality of the Expectation Maximization (EM) algorithm can be overcome. The Newton-like convergence of the EM algorithm is demonstrated mathematically. All these properties make a practical implementation feasible.

Chapter 5 : A coding assisted MIMO blind separation and decoding scenario is proposed. By using a-posteriori information, substantial gain over the uncoded system is shown. Moreover, the existence of coding structures appears to partly solve the problems of the EM getting trapped in a local minimum when the channel is close to singular or when the SNR is low. This new scheme appears to avoid local minimum and converge to the global minimum and then offer a good BER performance in a reasonable SNR range.

Chapter 6: Conclusions and future work base on this study are given. We point out main contributions of this thesis and the potential research such as, convolutive channel estimate, semi-blind methods and complexity reduction are emphasized.

Chapter 2

Review of Blind Equalization

2.1 Introduction

Blind equalization can be traced back to 30 years since the self-recovering equalization [10] appeared. A decision direct algorithm [11] is regarded as pioneers of the BE methods. Blind source separation (BSS) is a younger topic than BE since the seminal paper [12]. The corresponding separation technique can naturally applied into MIMO system in wireless communications. In this chapter, a history of BE and BSS are provided briefly. From the original single link to the multiple links. Both second order statistics and high order statistics methods are presented. Several simulations based on some popular BE and BSS approaches are given. First, we note the definitions used in this chapter at the beginning of this section.

2.2 Notation and Statistical Prerequisites

Statistical quantities play an important role in signal separation and equalization. It is necessary to introduce some related concepts used in this thesis.

Definition 1 (Expectation) *The expectation of a random variable s is given by*

$$E\{s\} = \int_{-\infty}^{\infty} sp(s)ds \quad (2.1)$$

Definition 2 (Wide-sense cyclostationary) *A random process $s[k]$ is defined as wide-sense cyclostationary (WSCS) with period M when it satisfies the following: The mean of s , $m_s[k] = E\{s[k]\}$ obeys*

$$m_s[k - \tau M] = m_s[k], \quad (2.2)$$

and the autocorrelation function $r_s[k, l] = E\{s[k]s^[l]\}$ satisfies,*

$$r_s[k - \tau M, k - l - \tau M] = r_s[k, k - l] \quad (2.3)$$

for all k and any integer τ .

Suppose a random process $s[k]$ is WSCS. It can be expanded as a Fourier series since its autocorrelation function $r_s[k, k-l]$ is periodic in k with period M , we can expand it as a Fourier series with the α th Fourier coefficient,

$$r_s^\alpha[l] = \frac{1}{M} \sum_{k=0}^{M-1} r_s[k, k-l] e^{-j2\pi\alpha \frac{k}{M}} \quad (2.4)$$

and its spectrum, the cyclic spectrum, is given by

$$S_s^\alpha(\omega) = \tilde{F}(r_s^\alpha[l]) = \sum_{l=-\infty}^{\infty} r_s^\alpha[l] e^{-j\omega l}, \quad \alpha = 0, 1, \dots, M-1. \quad (2.5)$$

Where $r_s^\alpha[l]$ is said to be the cyclic autocorrelation function indexed by the cycle frequency parameter α .

Definition 3 (Bussgang process) A stochastic process is said to have Bussgang statistics if its autocorrelation function equals the cross-correlation function between the process and the output of a zero-memory nonlinearity, g , i.e.

$$E\{s(k)s(k-i)\} = E\{s(k)g[s(k-i)]\}. \quad (2.6)$$

Let s_1, s_2, \dots, s_k be k complex random variables and $\mathbf{s} = (s_1, s_2, \dots, s_k)^T$, $s_l = s_{l,R} + js_{l,I}$, $l = 1, 2, \dots, k$.

Definition 4 (The first joint characteristic function) The first joint characteristic function of s_1, s_2, \dots, s_k is defined as

$$\Phi(\omega_{1,R}, \omega_{2,R}, \dots, \omega_{k,R}, \omega_{1,I}, \omega_{2,I}, \dots, \omega_{k,I}) \equiv \Phi(\omega) = E\left\{\exp\left(j \sum_{l=1}^k (\omega_{l,R} s_{l,R} + \omega_{l,I} s_{l,I})\right)\right\} \quad (2.7)$$

where $\omega_1, \omega_2, \dots, \omega_k$ are complex variables defined as $\omega_l = \omega_{l,R} + j\omega_{l,I}$, $l = 1, 2, \dots, k$, and $\omega = (\omega_1, \omega_2, \dots, \omega_k)^T$. It leads to the k th-order joint moment

Definition 5 (The k th-order joint moment)

$$E\{s_1 \cdots s_l s_{l+1}^* \cdots s_k^*\} = (-2j)^k \frac{\partial^k \Phi(\omega_1, \omega_2, \dots, \omega_k)}{\partial \omega_1^* \cdots \partial \omega_l^* \partial \omega_{l+1} \cdots \partial \omega_k} \Big|_{\omega=0}. \quad (2.8)$$

The second joint characteristic function is logarithmic function of the first joint characteristic function and is given as

Definition 6 (The second joint characteristic function)

$$\Psi_s(\omega) = \ln \Phi(\omega) \quad (2.9)$$

The k th-order joint cumulant of variable $\mathbf{s} = (s_1, s_2, \dots, s_k)^T$ is the partial derivative of the second joint characteristic function, this leads to the definition of *cumulant*,

Definition 7 (Cumulant)

$$\text{cum}\{s_1, s_2, \dots, s_k\} = (-2j)^k \frac{\partial^k \Psi^k(\omega_1, \omega_2, \dots, \omega_k)}{\partial \omega_1^* \cdots \partial \omega_l^* \partial \omega_{l+1} \cdots \partial \omega_k} \Big|_{\omega=0} \quad (2.10)$$

and the $p + q$ th-order cumulant of a random variable s is denoted by

$$C_{p,q}(s) = \text{cum}\left\{ \underbrace{s, \dots, s}_{p \text{ terms}}, \underbrace{s^*, \dots, s^*}_{q \text{ terms}} \right\}. \quad (2.11)$$

Definition 8 (Kurtosis)

$$\kappa(s) = E|s(k)|^4 - 2E^2\{|s(k)|^2\} - |E\{s(k)^2\}|^2 \quad (2.12)$$

Kurtosis is a fourth order cumulant and is a key statistical tool in blind signal processing area and a typical application is introduced in section 2.4.6.

Definition 9 (Circular) A complex random variable is defined as circular if for any real-valued number α , the probability density function (pdf) of $p(s)$ and $p(e^{j\alpha}s)$ are the same (i.e, $p(s)$ is phase invariant). $p(s) = p(e^{j\alpha}s)$.

Quadrature amplitude modulation (QAM) signals are not circular.

Definition 10 (Mutually Independent) Two complex random variables s_1 and s_2 are said to be mutually independent if $p(s_1, s_2) = p(s_1)p(s_2)$.

2.3 Present Pilot Assisted Wireless Transmissions

Traditional digital communication systems use training sequences or pilots for channel estimation. Generally, these pilot assisted transmissions multiplex known symbols with information data and then transmit these through antennae. Receivers know these pilot symbols according to the wireless protocols and employ this information to calculate the channel coefficients for processing. Some of the earliest studies on pilot assisted transmission focused on fast varying channels [13][14]. In paper [13], the author called the scheme Pilot Symbol Assisted Modulation (PSAM) and provided an analytical method for the design of pilot assisted transmissions. Thereafter, there has been continuing studies and research in the signal processing and digital communications communities. Some standardized schemes are illustrated in Figure 2.1 [14]. The GSM system [15], includes 26 pilot bits in the middle of each frame with some starting and tail bits. The North America TDMA standard [16] puts pilot symbols at the beginning of each frame. Third generation systems such as WCDMA [17] and CDMA2000 [18] transmit pilots and data simultaneously using separate spreading codes. Pilots are also used in broadband systems such as HiperLAN II [19], the IEEE 802.11 [20] and Long term evolution (LTE) [21]. LTE is considered as part of the fourth generation wireless systems. Wireless broadcast also has pilot sequences. The DVB-T [22] inserts different types of pilots in a periodic manner whereas the ATSC [23] sends pilots in the data stream with synchronization pilots. PSAM strategy is also successfully applied into wire communications, such as xDSL [24], cable modem [25] and voice band modems.

Generally speaking, equalization based on training sequences can estimate the channel state information with simple calculations and obtain desirable performances. Usually, the receiver applies zero forcing (ZF) or minimum mean-square error (MMSE) criteria to obtain channel information for further decoding operations. Receiver performance can be improved by increasing the length of pilot.

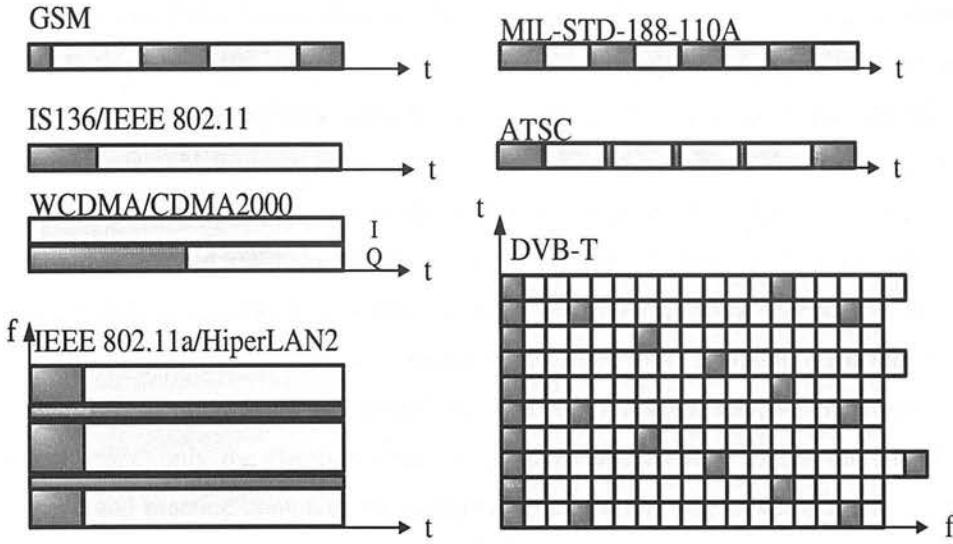


Figure 2.1: Pilot placement patterns in existing wireless systems. *Left: packet transmissions. Right: continuous transmissions. Shadow areas are pilot symbols.*

2.4 Brief Survey of Blind Equalization

The performance of equalizers with the assistance of PSAM is excellent so training based methods are widely applied in practical applications in present communication systems. With the aid of the pilot sequence, various approaches to the optimization and estimation of channel equalizers have been investigated and feasible strategies have been obtained in practice. Why bother with blind equalizations? Answers come from two sides. Firstly, in some cases, the channel response is unknown, time-varying, non-linear, and may also be not invertible, acquiring channel information is difficult in such environments. Secondly, in digital communications, bandwidth is scarce. The drawback of using training sequences (pilots) for channel estimation is obvious since pilots occupy an amount of channel resources. To struggle with limited bandwidth is an important part of communication systems. As we know, capacity is related with three parameters: numbers of channel, transmitted power and bandwidth. In consideration of human safety, the transmitted power is always constrained in wireless communications, thus saving bandwidth and multichannel transmission are important developing trends of further digital communications. Consequently, the multiple input multiple output (MIMO) system and BE have received lots of attention in this decade.

The current mobile network bears a lot of overhead for the equalization operation. Such as in GSM system, 18% of pilot bit in each slot. By eliminating this pilot sequence, BE would

present a bandwidth efficient solution. This is appealing since high data rate applications are needed in the future communications. In a broadcast network where a transmitter broadcasts to many receivers, if one receiver can not estimate its channel correctly, it must wait for the next pilot sequence, which means it loses all the data while it is waiting. An alternative strategy is that the transmitter send the current data again as the request. However, other receivers wait for this duration without any gain. If this happens too often, the data rate may be seriously reduced when the network capacity (subscriber) is large. Moreover, in some special applications, such as the radar signal detection or sonar detection, the only thing available is the received signals and blind equalization is the only applicable method. Generally, a BE scheme operates without channel input; only the channel output is available and as such makes the BE challenging in theory and practice compared to traditional equalization. BE has importance in broadcast channels or in any application where transmitter cooperation is impractical, e.g., satellite [26] and microwave radio [27][28]. A single chip demodulator for 64-QAM and 256-QAM digital television has been successfully implemented [29].

However, despite the fact that much work has been done in this area, applications of BE in commercial communication systems still are not widespread. Possibly one of the main reasons is that communication engineers are more inclined to use the mature techniques rather than new ones. Moreover, the application service providers and the carriers of wireless network do not want to expend the large amounts of capital required to change the existing wireless infrastructure. Another major reason for the insufficiency of broad uses of blind equalization to date is the fact that blind equalizers are, in general, more complicated than their counterparts such as training or pilot based systems. Usually, BE needs a larger sample size, lower convergence speed and higher complexity. Furthermore, ambiguities always remain. As a result, more effort is necessary to develop faster, simpler and more robust BE algorithms which can be widely applied to the practical applications.

In this chapter, we focus on various blind algorithms and architectures of single input single output (SISO), single input multiple output (SIMO) and MIMO systems. Many previous works [30] [31] [29] [32] have provided a thorough overview of the BE methods which focus on the single input and single output case. At the beginning, we briefly review the history of BE.

The basic blind SISO system is illustrated in Figure 2.2,

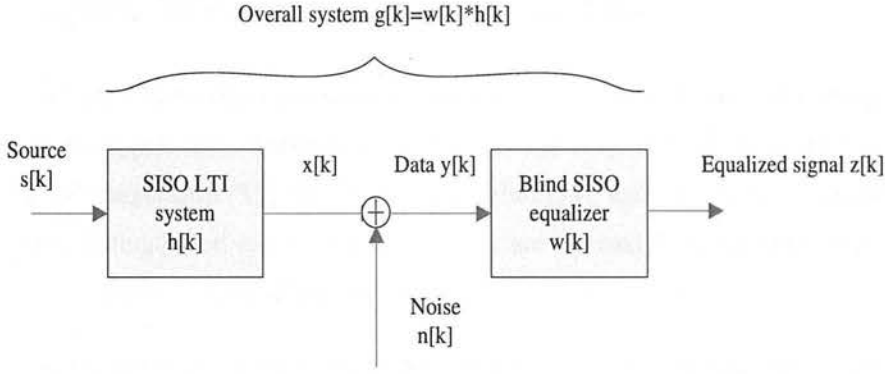


Figure 2.2: Architecture of the SISO blind equalization.

The system model is given as

$$y(k) = \sum_{t=0}^l h(k-t)s(t) + n(k) \quad (2.13)$$

$$u(k) = \mathbf{w}^T \mathbf{y}(k) \quad (2.14)$$

where $y(k)$ is the received signal at time index k and vector $\mathbf{y}(k)$ is the received signal sequence and is denoted as

$$\mathbf{y}(k) = [y_k, y_{k-1}, \dots, y_{k-l}]^T. \quad (2.15)$$

$s(t)$ is the digital source sequence and $n(k)$ is the additive white Gaussian noise. $u(k)$ is the output of the equalizer and \mathbf{w} is the equalizer coefficients and is denoted as

$$\mathbf{w} = [w_0, w_1, \dots, w_l]^T, \quad (2.16)$$

where l is the equalizer length and $l < \infty$. $h(k)$ is the linear time invariant system with bounded in bounded out stability.

The ultimate aim of blind equalization is to recover the input sequence without knowledge of the channel and any input sequence.

2.5 Implicit HOS – Bussgang Type Methods

The first BE algorithms proposed in the communication community were Bussgang type methods where the higher order statistics were utilized implicitly. Among them, the Sato algorithm [10], the BGR algorithm [33], the Godard algorithm [34], and the constant modulus (CM) algorithm are distinguished and well studied. They are referred to as Bussgang-type algorithms because the equalized signal is approximately a ‘Bussgang process’.

BE algorithms are often designed by minimizing non-minimum square error (non-MSE) cost functions. These show the level of intersymbol interference (ISI) in the equalizer output, but do not require the original inputs. Define the *mean cost function* as

$$J(\mathbf{w}) = E\{\Psi(u(k))\}, \quad (2.17)$$

where $\Psi(\bullet)$ is a scalar function of $u(k)$. $J(\mathbf{w})$, the mean cost function, is specified such that its minimum point \mathbf{w} corresponds to a minimum ISI or MSE condition. A stochastic gradient descent minimization algorithm is given as [35].

$$\begin{aligned} \mathbf{w}(k+1) &= \mathbf{w}(k) - \mu \frac{\partial}{\partial \mathbf{w}(k)} \Psi(u(k)) \\ &= \mathbf{w}(k) - \mu \Psi'[\mathbf{y}^H(k)\mathbf{w}(k)]\mathbf{y}^*(k) \end{aligned} \quad (2.18)$$

This equation is usually called the ‘standard form’ of the blind equalization algorithm. Then, the coefficients of a blind equalizer can be obtained by the mean cost function $\Psi(u)$. The design of the blind equalizer is converted into seeking a function Ψ such that the local minima of $J(\mathbf{w})$ is the equilibrium of the algorithm (2.18). In the following sections, some selections of Ψ are introduced.

2.5.1 Sato - Birth of BE

It is Sato who proposed the concept of blind equalization (self-recovering equalization) in his original work in 1975 [10]. It is one of the first widely used recursive identification schemes for discrete time system. This algorithm is suitable in a pulse amplitude modulation (PAM) system, in which the input symbol is uniformly distributed in M levels,

$$\{\pm(M-1)d, \pm(M-3)d, \dots, \pm 3d \pm d\} \quad (2.19)$$

PAM can be easily extended to QAM (complex value) straightforward. The corresponding error function is defined as,

$$\Psi_{sato}(u(k)) = u(k) - R_s \text{sgn}(u(k)). \quad (2.20)$$

In which R_s , the gain of the equalizer, is defined as,

$$R_s = \frac{E\{|s(k)|^2\}}{E\{|s(k)|\}} \quad (2.21)$$

The Sato algorithm replaces the channel input, $s(k - d)$, with $R_s \text{sgn}(u(k))$, called the slicer output, where d is the group delay. This algorithm takes the PAM signals as binary inputs and is based on an experimental formula instead of a theoretical one. Later researchers showed that local minima exist in the mean cost function of the Sato algorithm for M -ary PAM input and the size of the attractive region of these local minima decreases as the number of channel input levels increases [36].

2.5.2 BGR Algorithm

The first analytical study of blind equalization was develop by Benveniste, Goursat and Ruget in 1980 [33]. They extended the Sato algorithm by taking QAM signals as two mutually independent PAM signals. And then, applying the Sato algorithm to the two streams, the real and imaginary parts, respectively. The cost function is defined as,

$$\Psi_b(u(k)) = \tilde{\Psi}(\text{Re}[u(k)]) - R_b \text{sgn}(\text{Re}[u(k)]) + j\{\tilde{\Psi}(\text{Im}[u(k)]) - R_b \text{sgn}(\text{Im}[u(k)])\}. \quad (2.22)$$

$\tilde{\Psi}\{\bullet\}$ is an odd and twice differentiable function of, it also satisfies,

$$\Psi_b''(x) \geq 0 \quad \forall x \geq 0, \quad (2.23)$$

where $\Psi_b''(x)$ is the twice differential function of $\Psi_b(x)$. This algorithm has been shown to be convergent if two conditions are satisfied [37],

1. Signals have a sub-Gaussian.
2. The channel is a constant with no dynamics or for dynamical channels, a noncausal and infinitely parametrized equalizer is updated by the BGR algorithm.

If the odd function $\Psi_b(x)$ generalizes the linear function $\Psi_b(x) = x$, it is the same as the Sato algorithm. Generally, both Sato and BGR algorithms are call as Generalized Sato Algorithm (GSA) [37].

2.5.3 Godard Algorithm

In 1980, Godard proposed a new blind equalization algorithm using the cost function [34]

$$\Psi_g(u(k)) = \frac{1}{2}(|u(k)| - R)^2. \quad (2.24)$$

This cost function was subsequently generalized to,

$$\Psi(u(k)) = \frac{1}{2q}(|u(k)|^q - R_q)^2 \quad q = 1, 2, \dots, \quad (2.25)$$

where

$$R_q = \frac{E\{|s(k)|^{2q}\}}{E\{|s(k)|^q\}}. \quad (2.26)$$

This family of algorithms is distinguished by the integer q . Using the stochastic gradient descent, the update of the Godard algorithm is given by,

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu(|\mathbf{y}(k)^H \mathbf{w}(k)|^q - R_q)|\mathbf{y}(k)^H \mathbf{w}(k)|^{q-2} \mathbf{y}(k)^T \mathbf{w}(k) \mathbf{y}^*(k) \quad (2.27)$$

We will see, when $q = 1$, the Godard algorithm is essentially the Sato algorithm. When $q = 2$, the Godards algorithm is equivalent to the *constant modulus algorithm (CMA)* of Treichler[38] which will be introduced in details in the next section. The local convergence of this algorithm is accurately analyzed in [39]. There is no guarantee to converge to the desirable minima of the cost surface even for a noiseless channel.

2.5.4 Constant Modulus Algorithm

Constant Modulus Algorithm (CMA) was developed independently of the Sato and Godard algorithms and attracted a lot of attention in digital communications because its cost function is easy to analyze. For a channel input signal that has constant modulus, the CMA equalizer penalizes the equalized signals $z(k)$ that do not have the constant modulus characteristics. The

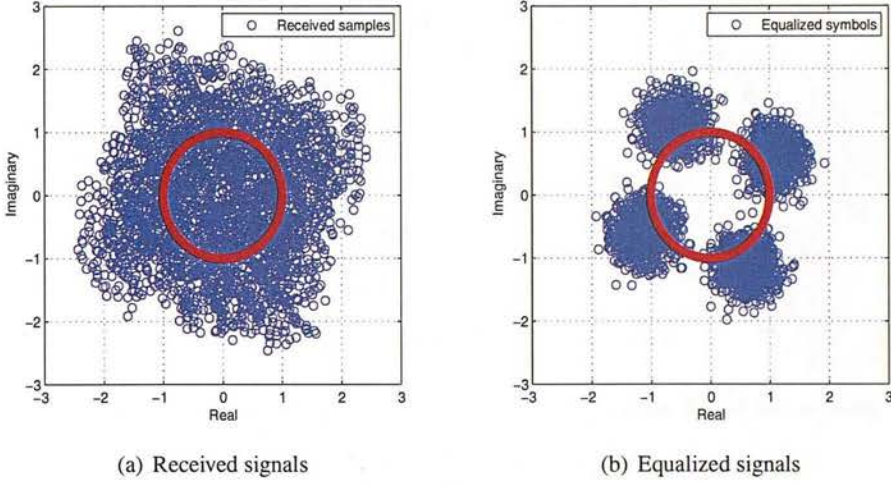


Figure 2.3: CMA blind equalization.

modulus error is simply

$$e(k) = |u(k)|^2 - R_2 \quad (2.28)$$

And its cost function is defined as,

$$\Psi_{CMA}(u(k)) = \frac{1}{4}(|u(k)|^2 - R_2)^2, \quad (2.29)$$

where

$$R_2 = \frac{E\{|s(k)|^4\}}{E\{|s(k)|^2\}}. \quad (2.30)$$

Obviously, R_2 is a constant when QAM modulations are employed in a digital system. In some sense, R_2 involves high order statistics (HOS). We can take it as the first expression using HOS tools to equalize the output of signal. In particular, CMA has been proven to be useful not only in blind equalizations but also in blind array signal processing systems [35]. For finite length baud rate sampled equalization, the constant-modulus criterion has been shown to have many local minima, thus, global convergence is not guaranteed [39]. Figure 2.3 illustrates the equalized signals after the CMA algorithm. This system is set up with a 4-QAM modulation, a single complex channel with 6 taps, the convergence step $\mu = 0.001$ and 3000 symbols were transmitted. As shown in the Figure 2.3(b), equalized signals converge to a constant circle, R_2 .

CMA and HOS techniques are not distinguished much since the CMA algorithm applies kurtosis implicitly. The CMA was developed and improved in the 90's last century. For improving stability and robustness, Lin and Lee proposed a method to avoid the gradient noise [40]. This

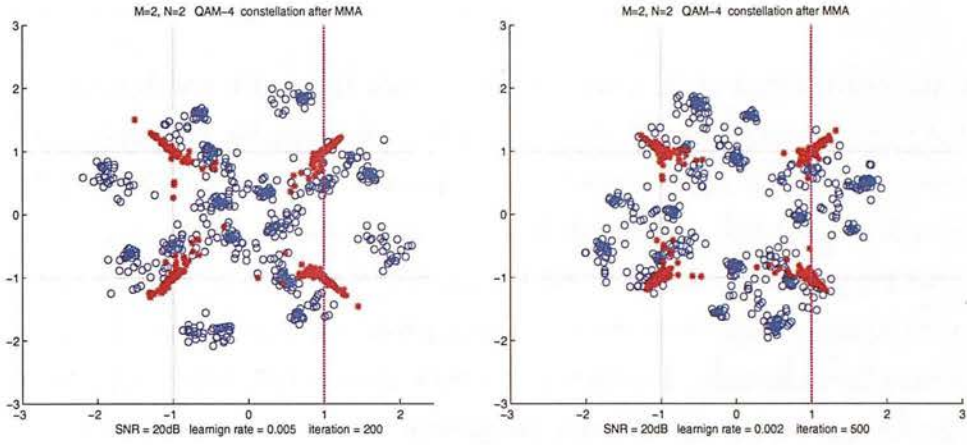


Figure 2.4: Performance of MMA algorithm with different learning rates.

method can not only accomplish blind equalization and carrier phase recovery simultaneously, but choose a nonlinear gain with a least squares algorithm.

The algorithm [41] combining CMA and the benefits of the reduced constellation algorithm (RCA) [42] was proposed and always was called multimodulus algorithm (MMA). The cost function RCA algorithm is simply defined as averaged distance between the received signals $y(k)$ and reference corner points, e.g., QAM points in the constellation. The MMA algorithm is suitable for certain types of signal constellations, such as non-square constellations and very dense constellations [41].

Figure 2.4 was implemented using the MMA. Experiments show that the equalized signal converges to some fixed point we desire. Simulation results were performed with $\text{SNR} = 20 \text{ dB}$, learning rates of 0.005 and 0.002 respectively, and the number of iterations were 200 and 500 respectively.

From the simulations above, both CMA and MMA might converge to the desirable point. These algorithms try to minimize the cost function defined by the CM criterion. The CM criterion penalizes deviations in the modulus of the equalized signal away from a fixed value and it can successfully equalize signals characterized by source those possessing a constant modulus.

2.5.5 Stop-and-Go Algorithm

From the standard form of the blind equalization algorithm 2.18, the convergence characteristics of these algorithms are mainly effected by the sign of the error signal. In order to force the coefficients of a blind equalizer to move to the optimum solution, the sign of its error signal should be associated with the sign of the LMS prediction error. The so-called stop-and-go method was proposed by G.Picchi [43] to improve the convergence of blind equalizers. The idea behind this algorithm is to stop updating the coefficients of equalizer when the error sign is not sufficiently reliable. This happens when the error function is less likely to have the true sign for the gradient descent direction. Updating the coefficients of the equalizer is determined by the error signs. Consider two algorithms with error functions $\Psi_{stop1}(u)$ and $\Psi_{stop2}(u)$. The stop-and-go algorithm can be written in the following forms:

$$\mathbf{w}(k+1) = \begin{cases} \mathbf{w}(k) + \mu \Psi_{stop}(u(k)) \mathbf{y}^*(k), & \text{if } \text{sgn}[\Psi_{stop1}\{u(k)\}] = \text{sgn}[\Psi_{stop2}\{u(k)\}] \\ \mathbf{w}(k), & \text{if } \text{sgn}[\Psi_{stop1}\{u(k)\}] \neq \text{sgn}[\Psi_{stop2}\{u(k)\}] \end{cases} \quad (2.31)$$

where the error function $\Psi_{stop}\{u(k)\}$ is given as

$$\Psi_{stop}\{u(k)\} = \frac{1}{2}(u(k) - \hat{s}(k)) + \frac{1}{2}|u(k) - \hat{s}(k)| \text{sgn}[u(k) - R_s \text{sgn}(u(k))]. \quad (2.32)$$

The $\hat{s}(k) = O(u(k))$ is the quantization of the equalized signal $z(k)$. This algorithm constructs the update by employing the difference between the quantizer output and equalizer output, $|u(k) - \hat{s}(k)|$. Obviously, this error function combines the Sato and the decision-directed algorithms to improve convergence speed. And the final performance depends on the accuracy of the estimate $\hat{s}(k)$ for the true symbol $s(k)$.

2.5.6 Shalvi and Weinstein Algorithm

Even though we try to partition HOS methodology with Bussgang methods clearly, it is not easy because of strong connections between them. The method of Shalvi-Weinstein [44] is based on higher order statistics of the equalizer output but this algorithm essentially keeps up with a form of Godard algorithms. It is the bridge between implicit HOS methods and the explicit kind of methods. The kurtosis of the equalizer output signal $y(k)$ is defined as equation (2.12),

$$\kappa(y) = E|y(k)|^4 - 2E^2\{|y(k)|^2\} - |E\{y(k)^2\}|^2 \quad (2.33)$$

The Shalvi-Weinstein algorithm maximizes κ subject to the constant power constraint,

$$E|u(k)|^2 = E|s(k)|^2 \quad (2.34)$$

Define \mathcal{B}_n as the combined channel equalizer impulse response given by,

$$\mathcal{B}_n = \sum_{t=0}^m h_k w_{k-t} \quad -\infty < n < \infty \quad (2.35)$$

If $s(k)$ is *i.i.d.*, it can be shown that,

$$E|u(k)|^2 = E|s(k)|^2 \sum_{i=1}^k |\mathcal{B}_i|^2 \quad (2.36)$$

So that

$$\kappa(y) = \kappa(s) \sum |\mathcal{B}_i|^4 \quad (2.37)$$

The Shalvi-Weinstein equalizer is therefore equivalent to,

$$\text{maximize } \sum_{i=-\infty}^{\infty} |\mathcal{B}_i|^4 \quad \text{subject to } \sum_{i=-\infty}^{\infty} |\mathcal{B}_i|^2 = 1 \quad (2.38)$$

It easy to be shown that Shalvi-Weinstein criterion reduces to the Godard cost function under special cases: a linear channel, no noise, an *i.i.d.* source.

We summarize the general properties of the algorithms based on Bussgang statistics. In digital communications, Bussgang type algorithms are popular for the blind equalization. Except for *Shalvi – Weinstein* criterion, all the algorithms proposed in the early stage actually belong to a generalized class, *Bussgang* blind equalization. They employ higher-order statistics implicitly, through zero-memory nonlinearities. They use a cost function whose corresponding minimum reflects the optimum solution of equalizer. The derivation of the cost function has the following general form,

$$\frac{\partial J(W(k))}{\partial W(k)} = -e(k)u(k) \quad (2.39)$$

Where $e(k)$ is the error between the output signal of equalizer and the estimated signal. $z(k)$ is output signal of equalizer. Early studies utilize a variety of combinations of purely blind methods and decision-directed approaches. Many local stability conditions and initialization selection have also been investigated [31] for the standard Bussgang-type algorithms.

Although there have been many extended equalization methods, The main drawback suffered by the mentioned blind equalizers is typically the slow convergence to the unknown CSI. Usually, several thousand observation samples are necessary to achieve channel identification. Moreover, the utilization of nonconvex cost functions can lead to estimates of the channel affected by high residual mean square errors (MSE) [45] [46].

2.6 Multiple Input Multiple Output Systems

Generally, MIMO systems are realized by multi-element array antennae. Channel capacity grows approximately linearly with $n = \min N_T, N_R$, where N_T and N_R are the number of transmitted antenna and receive antenna respectively. Thus MIMO systems have attracted the attention of scientists and engineers.

In this work, we are interested in the transmit and receive antennae connected by an independent flat Rayleigh fading process, as introduced in the first chapter. In channels with independent Rayleigh fading, a signal transmitted from each transmit antenna is uncorrelated at each of the receive antennae. As a result, the signal corresponding to every transmit antenna has a different spatial signature at a receive antenna. We note that this channel model does not exactly represent the real radio environment in all cases. In practical channels, the fades are not independent, and the channel capacity is reduced because of the spatial correlation among the receive antennae. However, this independent Rayleigh fading model is conventional model and provide fundamental research in academic society. It can be approximated in MIMO channels where antenna element spacing is larger than the carrier wavelength or the incoming wave incidence angle spread is relatively large, e.g. larger than 30 degrees [47]. An example of such a channel is the down link in cellular radio. In base stations placed high above the ground, the antenna signals get correlated due to a small angular spread of incoming waves and larger antenna separations are needed in order to obtain independent signals between adjacent antenna elements than if the incoming wave incidence angle spread is large. The displacement of receive antenna is needed to be widely separated which is usually greater than half wavelength to guarantee that spatial correlations do not exist.

For a channel matrix H in which each element represents the attenuation of a radio path. Define a time delay matrix T with the element of τ_{ij} representing the time delay of a particular path

from j th transmitted antenna to i th received antenna. The received signal vector \mathbf{y} is given,

$$\mathbf{y} = (H \odot \exp(-j\omega_{RF}\Gamma))\mathbf{s} + \mathbf{n} \quad (2.40)$$

where \odot is the element-wise multiplication of two matrices and ω_{RF} is the frequency operation. \mathbf{s} and \mathbf{n} are signal and noise vectors respectively. The amplitude and phase information can be combined into the matrix H with the element $h_{i,j}e^{-j\omega_{RF}\Gamma_{ij}}$. Then the channel matrix H can be considered as a matrix of complex scalars. In this work, the blind equalization and separation focus on instantaneous MIMO channels. Previous studies of BE have been concentrated on SISO systems, where people focused on the temporal distortion and inter-symbol interference cancellation. In the following, we pay more attention to MIMO systems.

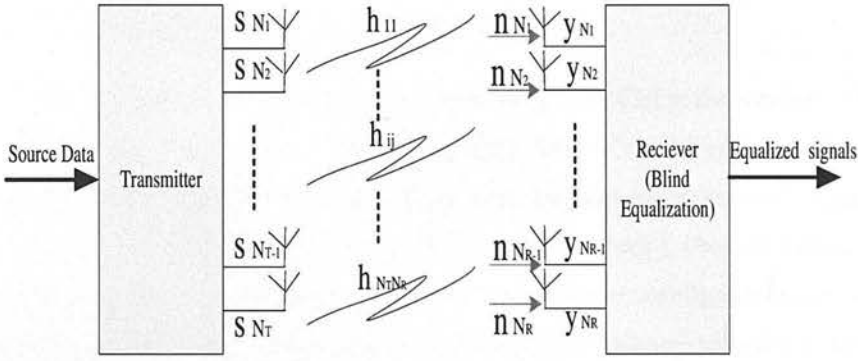


Figure 2.5: Architecture of Blind Equalization of MIMO system.

As illustrated in Figure 2.5, the MIMO system model is given by,

$$\mathbf{Y} = \mathbf{H}\mathbf{S} + \mathbf{N} \quad (2.41)$$

Where $\mathbf{Y} \in \mathbb{C}^{N_R \times T}$ is the observed matrix containing T observed signals from the N_R sensors, and $\mathbf{S} \in \mathbb{C}^{N_T \times T}$ is the complex discrete source signal matrix. T is the number of symbols in a block and \mathbf{H} is assumed constant for a coherence interval of a block, and change independently in the following block. $\mathbf{N} \in \mathbb{C}^{N_R \times T}$ is the matrix of the noise with covariance, Σ , which is uncorrelated with source signals. $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ is an unknown linear square matrix whose elements are drawn independently from a Rayleigh distribution and which satisfying circular symmetric complex Gaussian distribution. We assume that it is invertible. Note that, we do not guarantee \mathbf{H} is orthogonal. The square channel matrix can be expanded into the non-square overdetermined MIMO systems where the number of receive antennae is greater than the

number of transmitted antennae. The principal components analysis (PCA) or singular value decomposition (SVD) operations are needed as preprocessing at receivers to make $N_R = N_T$. This instantaneous MIMO model is very fundamental in communication systems, and can be extended to other popular models, such as MIMO OFDM systems. Since the number of transmit antennae and receive antennae are equal, we define the channel matrix as

$$\mathbf{H}(k) = \begin{bmatrix} h_{11}(k) & h_{12}(k) & \cdots & h_{1N_R}(k) \\ h_{21}(k) & h_{22}(k) & \cdots & h_{2N_R}(k) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R1}(k) & h_{N_R2}(k) & \cdots & h_{N_RN_R}(k) \end{bmatrix} \quad (2.42)$$

where $h_{i,j}(k), \sim CN(0, 1) 1 < i, j < n$ is the channel impulse response of the SISO system from the j th input $s_j[k]$ to the i th output $y_i[k]$, $1 < k < T$.

Most MIMO detection and decoding research assume perfect CSI at the receiver. The optimum MIMO equalization is studied in [48] when people know CSI information. It is a popular assumption for very slow fading channels with good channel estimation technique. However, this is not the case if the channel is fast fading. A pilot based channel estimation fails to fully exploit the channel information contained in the detected and decoded data symbols. The capacity is degraded when strong spatial correlation exists among antennae. Also, as stated in [49], imperfect CSI reduces the capacity gain and causes impairments of signal detection. Therefore, this makes it important to the blind channel estimation to MIMO detector studies.

When the channel parameters are not available, blind techniques have to be used to identify MIMO systems explicitly or implicitly to separate sources and equalize system distortion. Many BSS techniques for MIMO systems have been studied [50] [51][52][45][53]. Usually, for different applications with various source properties, there have been many different blind source separation and equalization algorithms which can be classified into four categories [54]:

1. HOS algorithms: The most popular approaches to solve BSS and BE problems if the sources are statistically independent without any temporal structure. Note that, in such a case, these methods do not allow more than one Gaussian source.
2. SOS algorithms: Utilizing the temporal correlation of sources, then second order statistics (SOS) are sufficient to estimate the mixing matrix and sources [55][56].
3. Non-stationarity algorithms: Approaches exploiting non-stationarity properties whose

joint probability distribution change when shifted in time or space, i.e. the second-order non-stationarity where source variances vary in time. The non-stationarity information methods allow the separation of coloured Gaussian sources with identical power spectra shapes [57][58]. This property attracts a lot research in blind speech separation since frequencies vary over time depending on pronouncing words.

4. Diversity algorithms: Approaches exploiting the various diversities of the sources, typically, time, frequency, and the time-frequency diversity. More generally, they make use of the joint space-time-frequency (STF) diversity [59] [60].

More advanced methods use combinations of all properties mentioned above, but further discussion is out of the scope of this thesis. Interested readers can look at the book [52] for details.

In following sections in this chapter, some popular MIMO blind separation and equalization methods are reviewed. From SIMO to MIMO systems, both SOS and HOS techniques are introduced. Since the significance of separating MIMO signals in this thesis, the independent component analysis (ICA) is emphasized in details.

2.6.1 Second Order Statistics

The methods and algorithms introduced previously focus on single input and single output systems. In this section, we concentrate on another important part of BE, single input and multiple output systems. SIMO channels arise in wireless communications from radio propagation between a single transmitter and multiple receivers or using oversampling technique when the outputs are sampled faster than the symbol rate at receivers. Such methods can be extended to this sort of MIMO system where the number of receive antennae is greater than that of the transmitted antennae. Second-order statistics (SOS) is an important statistics and is widespread used in digital signal processing. In a symbol rate sampling, SOS of a wide-sense-stationary (WSS) process do not retain phase information, and cannot distinguish between minimum-phase and non-minimum-phase systems. As a consequence, the SOS can not be applied to equalization of non-minimum phase systems in this case. Hence, people have to exploit other statistical properties to extract the phase information, such as high order statistics.

2.6.1.1 Fractionally Spaced Equalization

In 1991, Tong, Xu and Kailath [61] proposed the blind identifiability of SIMO LTI systems in which they utilized only the second order statistic (SOS) of system outputs with oversampling techniques. This work led to a number of SOS based SIMO blind system estimation and equalization algorithms such as the subspace and least-squares (LS) estimation approaches reported in [62].

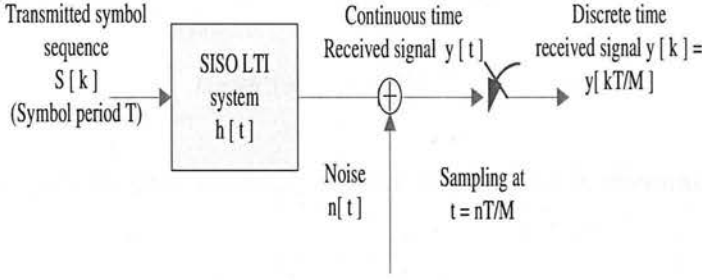


Figure 2.6: Architecture of a fractional sampling system.

In general, the autocorrelation function $r_s^\alpha[l]$ in equation (2.4) and its spectrum $S_s^\alpha(\omega)$ in equation (2.5) like second-order statistics are phase blind as α is set to zero. But this is not true when $\alpha > 1$. Consider a digital communication system illustrated Figure 2.6, the source $s[k]$ drawn from a finite alphabets is transmitted at time kT where T is the symbol period, and the discrete-time received signal is given below.

$$y[k] = y(k\tilde{T}) = \sum_{t=-\infty}^{\infty} s[t]h([k - tM]\tilde{T}) + n(k\tilde{T}), \quad (2.43)$$

where $1/T$ is called the symbol rate and \tilde{T} is called the sampling rate. $\tilde{T} = T/M$ is the sampling period and $M \geq 2$ is an integer referred to the oversampling factor. $h[k]$ is the discrete-time SISO linear time-invariant channel. We can express the noise-free received signal as $y[k] = \sum_{t=-\infty}^{\infty} s[t]h([k - tM])$ and its autocorrelation function is given by

$$r_y[k, k-l] = \sum_{t=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h[k - tM]h^*([k - l - mM])r_s[t - m] \quad (2.44)$$

The source signal variance of $s[k]$ is equal to σ_s^2 and its autocorrelation function $r_s[t - m] = \sigma_s^2$

when $t = m$. Then the equation 2.44 can be simply expressed as

$$r_y[k, k-l] = \sigma_s^2 \sum_{t=-\infty}^{\infty} h[k-tM]h^*[k-l-tM]. \quad (2.45)$$

Substituting (2.45) into (2.4) and using similar simplifications above, we get the α th Fourier coefficients

$$\begin{aligned} r_y^\alpha[l] &= \frac{\sigma_s^2}{M} \sum_{t=-\infty}^{\infty} \sum_{k=0}^{M-1} h[k-tM]h^*([k-l-tM])e^{-j2\pi\alpha k/M} \\ &= \frac{\sigma_s^2}{M} \sum_{m=-\infty}^{\infty} h[m]h^*[m-l]e^{-j2\pi\alpha m/M}, \quad \alpha = 0, 1, \dots, M-1. \end{aligned} \quad (2.46)$$

And the cyclic spectrum correspondingly is given by taking the Fourier transform with respect to l :

$$\begin{aligned} S_s^\alpha(\omega) &= \frac{\sigma_s^2}{M} \sum_{k=-\infty}^{\infty} h[k]e^{-j2\pi\alpha k/M} \sum_{l=-\infty}^{\infty} h^*[l]e^{-j\omega l} \\ &= \frac{\sigma_s^2}{M} H(\omega + \frac{2\pi\alpha}{M})H^*(\omega), \quad \alpha = 0, 1, \dots, M-1. \end{aligned} \quad (2.47)$$

Obviously, we can obtain phase information from the equation (2.47), in which $\alpha \neq 0$, while $S_s^0(\omega) = S_s(\omega)/M$ is phase blind, in which $\alpha = 0$. In digital communication system, such equalization is referred to a fractionally spaced equalization when sampling period \tilde{T} is less than the symbol rate. This means that the SISO multirate channel model is converted to an equivalent SIMO channel model. From (2.47), one can see that both the magnitude and phase information about the channel $h[k]$ are contained in this equation. This SOS technique led to a number of second order cyclostationary statistics (SOCS) based SISO blind system estimation and equalization algorithms. The most important advantage of such algorithms is that they are insensitive to additive noise as $\alpha \geq 1$. These SIMO second-order methods have been extended to other families, such as linear prediction techniques [63], subchannel response matching (SRM) approach [64] and subspace methods which we will introduce in the next section. In comparison with SISO methods where people conventionally resort to HOS techniques, SIMO only used the SOS and assume that the channels do not share common zero. In this sense, it is considered to be easier. On the other hand, blind equalization of SIMO systems using HOS was exploited by Treichler [65] as an extension of the CM algorithm.

2.6.1.2 Subspace Methods

Many recent blind channel estimation techniques have exploited subspace structures of the observation signals since the seminal paper [66]. The key idea behind this type of algorithm is that the channel vector is in a subspace of the observation statistics and is orthogonal to the noise subspace. People can decompose the observation matrix to the signal subspace and noise subspace and these two spaces are orthogonal. Taking advantage of this orthogonality allows us to identify the channel-coefficient vector. These methods have the attractive property that the channel estimates can often be obtained in a closed form.

Let us look at a SIMO system model $n > m = 1$ at time slot k ,

$$\mathbf{y}(k) = \mathbf{h}s(k) + \mathbf{n}(k) \quad (2.48)$$

Where $\mathbf{y}(k)$ is the $n \times 1$ vector, $\mathbf{h} = (h_1, h_2, \dots, h_n)^T$ is an $n \times 1$ channel of order equal to 0 for simplification. $\mathbf{n}(k)$ is a $n \times 1$ vector noise and it is the white Gaussian noise which is independent of $s(k)$ and its covariance matrix is $\Sigma_{\mathbf{n}} = \sigma^2 I$. After stacking T received signals, we have the following,

$$\begin{aligned} Y &= (\mathbf{y}(1)^T, \mathbf{y}(2)^T, \dots, \mathbf{y}(T)^T)^T \\ \mathbf{s} &= (s(1), s(2), \dots, s(T))^T \\ N &= (\mathbf{n}(1), \mathbf{n}(2), \dots, \mathbf{n}(T))^T. \end{aligned} \quad (2.49)$$

So, we obtain the system matrix form

$$Y = \mathcal{H}\mathbf{s} + N = X + N \quad (2.50)$$

where $X = \mathcal{H}\mathbf{s}$ and \mathcal{H} is the $MT \times T$ channel matrix, given as,

$$\mathcal{H} = \begin{bmatrix} \mathbf{h} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{h} & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{h} \end{bmatrix}$$

Obviously, this channel matrix has block Toeplitz structure. And the second order statistics of \mathbf{y}_k are

$$E\{YY^H\} = \mathbf{R}_y = \mathcal{H}\mathbf{R}_s\mathcal{H}^H + \Sigma_{\mathbf{n}}, \quad (2.51)$$

where $\mathbf{R}_y = E\{\mathbf{ss}^H\}$ is the source signal covariance matrix. By eigendecomposition, we get

$$\mathbf{R}_y = \mathbf{Q}_y \mathbf{\Lambda}_y \mathbf{Q}_y^H \quad (2.52)$$

The eigenvectors corresponding to T largest eigenvalues span the signal subspace, \mathbf{Q}_y , which is the same space as the channel vector, \mathcal{H} , spans. The other eigenvectors, noise subspace vectors \mathbf{Q}_n , span the noise subspace which is orthogonal to the signal subspace. So we can obtain the following equation,

$$\mathbf{Q}_n \mathcal{H} = \mathbf{0}. \quad (2.53)$$

By solving equation (2.53), the channel matrix \mathcal{H} [66] can be uniquely identified. The orthogonality of the signal and noise subspace provides the identification of the channel state information. Figure 2.7 illustrates the equalized 16-QAM signal resulting from the subspace method.

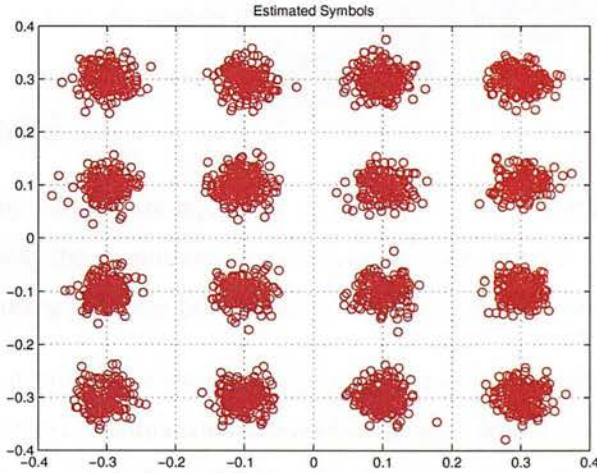


Figure 2.7: *Equalized signals by subspace blind equalization.*

As stated before, subspace methods are attractive because they can be identified by a closed form solution. From another point of view, since they depend on the property that the channel lies in a unique direction, they may not be robust against noise and modeling errors, especially when the channel matrix is close to being singular [67]. Then in practice, channel model mismatch should be considered carefully.

The other disadvantage of subspace methods is that they have an expensive computational load

when calculating a closed form solution. Fractionally spaced equalization and subspace methods utilize the SIMO model and the SOS signal property to blindly identify channels. They show two distinct features [68][45]: first, SOS can blindly identify non-minimum phase channels; second, finite impulse response (FIR) channels can be perfectly equalized in the noiseless case.

Some other methods based on the SOS characteristics assume that system inputs are temporally coloured such as AMUSE algorithm [55] and SOBI algorithm [56]. By exploring different power spectra, the identifiability of an instantaneous MIMO system using SOS of the system outputs has been proven by Hua and Tugnait [69]; meanwhile, some SOS based blind equalization methods have been reported such as minimum noise subspace method [53], the matrix pencil (MP) method [70], the blind identification via decorrelating subchannels approach [71], so on and so forth. However, these methods make the further assumption that the power spectra of the driving inputs are sufficiently diverse [71]. Such assumption are invalid in wireless communications. Therefore, we neglect this kind of SOS methods in this section.

2.6.2 HOS Methods

Most present methods use HOS for separating signals[72], because the HOS of the non-Gaussian signal contains not only the magnitude but also phase information of the unknown channel. BE of MIMO channels using HOS has been reported broadly in recent years [73][74][75].

In this section, considerations are focused on the problem of blind equalization of Linear Time Invariant MIMO systems as formulated in equation (2.41). MIMO systems are expected to equalize the observed signals both in spatial and temporal domains. In detail, MIMO equalization aims at eliminating inter-symbol interference due to possible delays introduced by multipaths propagation, and co-channel interference due to the possible presence of simultaneous users in the same band. A particular instance of the problem is the static mixture, often referred to as the source separation problem in which no inter-symbol interference is considered but only co-channel interference. The latter problem is relevant and formulated in narrow band communication system when the time delays are smaller than the symbol period, for example.

Methods based on SOS require the number of received antennae to be greater than the number of transmitted antennae or the source have distinct power spectra. When the number of receive antennae and transmitted antennae are equal and the sources are temporally i.i.d inputs, BE of

MIMO system usually use HOS technique as powerful tools. Some well-known blind MIMO methods are direct extensions of SIMO methods. Some MIMO methods employing higher-order statistics, cyclostationarity, subspace estimation, special matrix structures, and a constant modulus cost have been proposed. Blind source separation methods, assuming statistical independence among the transmitted sources, may also be used for solving the equalization problem in instantaneous MIMO channels.

Signals for digital communications are typically non-Gaussian and the higher-order statistics of these signals are therefore usually non-zero. This property can be applied in the construction of blind equalizers. Since the 1980s, the problem of SISO blind equalization has been tackled using HOS using the fact that cumulants and polyspectra contain both system magnitude information and system phase information. Algorithms using HOS exploit the non-Gaussian nature of transmitted signals. Gaussian signals are completely defined by their first and second moments. Consequently, higher-order cumulants and moments can be take as an measurement of being Gaussian. However, HOS and spectra have a large variance and, consequently, large sample sets are needed in order to obtain reliable estimates.

First, we make some assumptions for explicitly HOS algorithms.

1. The source signal $s[k]$ is statistically independent of the noise $n[k]$.
2. The SISO LTI system $h[k]$ is stable and the z-transform of its inverse system $h^{-1}[z]$ is also stable.
3. The source signal $s[k]$ is a zero-mean, i.i.d., stationary non-Gaussian process with $(p + q)th$ -order cumulant $C_{p,q}\{s[k]\} \neq 0$, where $C_{p,q}\{s[k]\}$ is defined as Eq. 2.11.
4. The noise $n[k]$ is a zero-mean, white or coloured, WSS Gaussian process with autocorrelation function $r_n[l]$.

In the following, we briefly introduce two basic and popular HOS based equalization algorithms utilizing HOS explicitly, the maximum normalized cumulant (MNC) equalization algorithm and the superexponential (SE) equalization algorithm. Both MNC and SE algorithm use cumulants as their criteria since cumulant can measure the independence and they are popular for their simplicity.

2.6.2.1 Maximum Normalized Cumulant Equalization Algorithm

As the name of Maximum Normalized Cumulant algorithm, MNC is designed through maximizing objective function [75],

$$J_{p,q}(\mathbf{w}_{MNC}) = \frac{|C_{p,q}\{u[k]\}|}{(\sigma_u^2)^{(p+q)/2}} \quad (2.54)$$

where $z[k] = \mathbf{w}_{MNC}^T \mathbf{y}[k]$ is the equalized signals and $\mathbf{y}[k]$ is the received signal vector. $C_{p,q}\{u[k]\}$ is $(p + q)th$ - order cumulant of $z[k]$ as defined as equation (2.11) and p, q are nonnegative integers, $p + q \geq 3$. This objective function is, in fact, the normalized cumulant of equalized signals. In MIMO case, the optimum $z[k]$ can be an estimate of any one of inputs and there are n stable local maxima [75] when $(p + q) \geq 2$. There proof is lengthy and omitted here. Equation (2.54) is highly nonlinear function, leading to the difficulty of derivation of a closed form solution. Therefore, people resort to gradient-type optimizations to deal with the MNC equalization algorithm for finding the local maximum of $J_{p,q}(\mathbf{w}_{MNC})$. The following table summarizes the steps of the MIMO-MNC algorithm.

Parameter choose	$p + q \geq 3$ and a convergence tolerance $\varsigma > 0$.
Initial setting	The iteration number i is set to 0. And set the initial value of $\mathbf{w}^{[0]}$.
Iteration $i = 1, 2, \dots$	Calculate the i - th approximation to the parameter $\mathbf{w}^{[i]}$ using the efficient Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [76] such that $J_{p,q}(\mathbf{w}^{[i+1]}) > J_{p,q}(\mathbf{w}^{[i]})$.
Convergence	If $\frac{ J_{p,q}(\mathbf{w}^{[i+1]}) - J_{p,q}(\mathbf{w}^{[i]}) }{J_{p,q}(\mathbf{w}^{[i]})} \geq \varsigma$, step into the next iteration; otherwise, obtain the optimum $\mathbf{w}_{MNC} = \mathbf{w}^i$.

The MNC method works well and shows good performance in the case of high SNR and the equalized coefficient \mathbf{w}_{MNC} is equivalent to the inverse system except for a scale and a time delay. For this reason, this criterion is referred to as the inverse filter criteria (IFC) [77] [78].

2.6.2.2 Super-Exponential Equalization Algorithm

Another algorithm which uses the cumulant technique for finding the blind equalizer $\mathbf{w}[k]$ is the super-exponential equalization algorithm. This iterative algorithm was proposed by Shalvi and Weinstein [79]. Consider the overall system $g = \mathbf{w} * \mathbf{h}$ in a SISO case, at iteration i , the

update of the overall system is given by [79],

$$\tilde{g}^{i+1}[k] = (g^i[k])^p (g^i[k]^*)^{q-1} \quad (2.55)$$

$$g^{i+1}[k] = \tilde{g}^{i+1}[k] / \|\tilde{g}^{i+1}[k]\| \quad (2.56)$$

$$(2.57)$$

where p and $q - 1$ are nonnegative, $p + q \geq 3$, and

$$\tilde{g}^{i+1}[k] = (\dots \tilde{g}^{[i+1]}[-1], \tilde{g}^{[i+1]}[0], \tilde{g}^{[i+1]}[1] \dots) \quad (2.58)$$

Consequently,

$$\left| \frac{g^{i+1}[k]}{g^{i+1}[m]} \right| = \left| \frac{g^i[k]}{g^i[m]} \right|^{p+q-1} = \dots = \left| \frac{g^0[k]}{g^0[m]} \right|^{(p+q-1)^{i+1}} \quad (2.59)$$

The coefficients of $g^{i+1}[k]$ decrease more rapidly when its has smaller amplitudes. This implies $g^{i+1}[k]$ approaches a delta function as $i \rightarrow \infty$. This indicates that $g^{i+1}[k]$ approaches a delta function more closely than $g^i[k]$. Then, the $i + 1$ th interference, $g^{i+1}[k]$ is less than i th interference $g^i[k]$.

The SE algorithm was first extended to the MIMO case [80], where authors assumed that all the system inputs have the same cumulant. Further research based on SE algorithm can be found in [81]. The following table summarizes the steps of the MIMO-SE algorithm [79].

Parameter choose	$p + q \geq 3$, and a tolerance $\zeta > 0$.
Initial setting	The iteration number i is set to 0. Set the initial value of $\mathbf{w}^{[0]}$ and calculate the inverse correlation matrix of the received signals $R_y = (E\{\mathbf{y}\mathbf{y}^H\})^{-1}$
Iteration $i = 1, 2, \dots$	Update the equalized coefficients \mathbf{w} by $\mathbf{w}^{i+1} = \frac{(R_y^*)^{-1} \mathbf{d}_{SE}^{[i]}}{\ (R_y^*)^{-1} \mathbf{d}_{SE}^{[i]}\ }$ where $\mathbf{d}_{SE}^{[i]}$ is the corresponding search direction and is defined as $\mathbf{d}_{SE}^{[i]} = \text{cum}\{u^i[k] : p, (u^i[k])^* : q - 1, \mathbf{y}^*[k]\}$ and $\mathbf{u}[k] = \mathbf{w}^T \mathbf{y}[k]$ is the received signals.
Convergence	If $ \langle \mathbf{w}^{i+1} \rangle^H \mathbf{w}^i \leq 1 - \frac{\zeta}{2}$, step into the next iteration; otherwise, obtain the optimum $\mathbf{w}_{SE} = \mathbf{w}^i$.

The major advantages of the SE are its faster convergence, a super-exponential rate. It need

smaller computational complexity over gradient type algorithms, such as the MNC algorithm. However, the SE equalization algorithm sometimes converge to an unfeasible solution if the initial $\mathbf{w}_{SE}[0]$ is inappropriate, or when the data length N is not sufficient large and the SNR is low. Some improvements of the convergence rate and final reliability can be obtained if we use whitening preprocessing. The full details are not expanded here, the interesting reader can make reference to the book [51].

Some other methods are also use the HOS technique directly in MIMO cases. Extensions of the CMA algorithm to a general MIMO case were investigated [74]. The Least-squares constant modulus algorithm (LSCMA) was first proposed in [82] for a two-sensor array. The essence of the LSCMA is to combine the well known least-squares estimator and constant modulus properties to blindly extract communication signals. This algorithm extends the CM algorithm to the MIMO system with fast convergence speed and good stability performance for any linearly independent set of input signals [82].

One of HOS measurement, kurtosis, also can be applied to blind MIMO separations directly, both in instantaneous mixtures and convolutive mixtures. To prevent the same source to be extracted at different outputs, the kurtosis based blind source separation algorithm (MUK) was proposed [83]. Papadias used a Gram-Schmidt orthogonalization approach to prevent this situation and he also proved the global convergence of the MUK algorithm in the noiseless case. The Gram-Schmidt orthogonalization was also developed in [84], where the authors apply idea of MMA algorithm to update real and imaginary parts of the equalizer matrix individually and then use the Gram-Schmidt orthogonalization to each part. Simulations show better performance than the MUK method. If the sources have distinct kurtosis, Cardoso proposed a simple and elegant algorithm, called Fourth-Order Blind Identification (FOBI) [85]. This method identifies the unknown matrix by eigen-decomposing the kurtosis matrix of observed data. Tong extend this method to the noisy case [86] and developed the EFOBI algorithm with good performance in noisy environments. In order to separate sources with identical kurtosis, FOBI was later generalized to JADE algorithm [87] which is based on joint diagonalization of several cumulant matrices. All these belong to the subject of ICA and will be introduced in detail in the following section.

2.7 ICA for Blind Instantaneous MIMO Separation and Equalization

The concept of BE is related to the problem of BSS in the signal processing community. Knowledge of the statistical and structural properties of the sources or mixing parameters is exploited in the blind separation process. Recent texts on the principles of source separation can be found in [52] [88]. Typically, source signals are assumed to be statistically independent, and the mixing system is assumed to be linear even though the mixing may be convolutive. Convolutive mixing with FIR filters may be described using FIR-MIMO model. Many ICA [89] algorithms [52][54][90][91] have developed for blind separation and equalization since it is implicitly connected with blind equalization and identification.

ICA has received a lot of attention and plays a very a important role in the signal processing community. Utilizing only the independence of the original signals, ICA identifies an unknown channel or mixing matrix and then estimates source signals typically by using some a priori information about the system and applying a suitable optimization procedure. It is essentially a statistical technique and can usually estimate the source signals up to certain indeterminacies, *e.g.* arbitrary scaling, permutation.

In a noisy ICA model, the observed data are typically expressed as a linear combination of sources such that

$$Y = AS + N \tag{2.60}$$

where S is the source matrix, N is noise matrix, Y is the observations and A is the mixing matrix. Obviously, this form is equivalent to the instantaneous MIMO system (2.41). For this ICA model, the following assumptions are adopted[52]:

1. Sources have to be statistically independent and Non-Gaussian.
2. Mixing matrix A is of full rank with the number of received sensors equal to or greater than the number of sources.
3. Each source has zero mean with unit variance.

All the above assumptions are very realistic and are almost automatically fulfilled in digital wireless telecommunication systems.

In digital wireless communications, assumptions (1) and (3) are very realistic and almost automatically fulfilled. Assumptions (2) can be satisfied when people design a MIMO system. Then, ICA drew the eyes of researchers and engineers, especially with the appearance of the fastICA algorithm [92]. This popular algorithm based on a fixed point iteration save a lot of computational loads, then make practical applications of ICA feasible. More and more applications combining ICA with other techniques have appeared recent years. Operations of the complex fastICA has been implemented by FPGA [93] in MIMO communication systems. ICA is a potentially useful tool in signal estimation and extraction, in particular in such fields as : Interference (ISI, CCI, MAI) cancellation [94] [95] [96]; signal separation in CDMA mobile communications [52]; Multiuser detection [94] [97] [98]; Radar detection [99]; Orthogonal frequency division multiplexing (OFDM) receiver design [100][91]; Blind separation of convolutive mixtures [101][102][103]; Refinement and improvement of the channel estimation [104].

2.7.1 Contrast Functions and Basic Algorithms

ICA usually operates by formulating a criteria (contrast) function and then optimizing it. The performance of separateness depends on both the contrast function and the algorithm optimizing it. The former determines the statistical properties of ICA, e.g., consistency, robustness and asymptotic variance. The latter determines the algorithm properties. e.g., convergence speed, numerical stability and memory requirements. In the following subsection, some popular contrast functions and basic optimizing algorithms are introduced.

2.7.1.1 Contrast Function

ICA separate mixtures by exploring the statistical independence of source signals. So the natural question is which tools should be used to measure the independence? Some basic contrast functions are listed below [52],

Normalized Kurtosis	$\kappa_4(\mathbf{s}) = E\{\mathbf{s}^4\} - 3(E\{\mathbf{s}^2\})^2 - 3$	
Differential entropy	$H(\mathbf{s}) = - \int p(\mathbf{s}) \log p(\mathbf{s}) d\mathbf{s}$	
Mutual information	$I(\mathbf{s}) = J(\mathbf{s}) - \sum_{i=1}^n J(s_i)$	
Negentropy	$J(\mathbf{s}) = H(\mathbf{s}_{gauss}) - H(\mathbf{s})$	(2.61)

The normalized kurtosis $\kappa_4(\mathbf{s})$ is commonly defined for a random vector \mathbf{s} and is simply the normalized fourth order cumulant. The second part at the end of kurtosis formula is explained as a correction to make the kurtosis of the normal distribution equal to zero. Then, for \mathbf{s} with unit variance the kurtosis becomes a normalized fourth moment $E\{\mathbf{s}^4\}$. \mathbf{s} is called sub-Gaussian and super-Gaussian for $\kappa_4(\mathbf{s}) < 0$ and $\kappa_4(\mathbf{s}) > 0$, respectively. In digital communication, sources are discrete, thus the probability mass function (pmf) is equals a discrete value. For binary sequences, \mathbf{s} is sub-Gaussian, the kurtosis, $\kappa_4(\mathbf{s}) = -2$. For complex sources, the normalized kurtosis of a zero-mean random variable is given as equation [44],

$$\kappa_4(s) = E\{|s|^4\} - 2(E\{|s|^2\})^2 - |E\{s^2\}|^2 - 3 \quad (2.62)$$

A high kurtosis shows a sharp peakedness in pdf or pmf.

The differential entropy $H_{en}(\mathbf{s})$ is defined as $H_{en}(\mathbf{s}) = - \int p(\mathbf{s}) \log p(\mathbf{s}) d\mathbf{s}$ for a given random vector $\mathbf{s} = [s_1 \ s_2 \ \dots \ s_n]^T$ with probability density function $p(\mathbf{s})$ [52].

The mutual information $I(\mathbf{s})$ between the random variables $s_i, i = 1, \dots, n$ is an information theoretic measure for the independence of random variables. It can be defined using the Kullback Leibler (KL) divergence between the joint density $p(\mathbf{s})$ and the factorized density $p_f(\mathbf{s}) = p(s_1) \cdots p(s_n)$ according to

$$- \int p(\mathbf{s}) \log \frac{p(\mathbf{s})}{p_f(\mathbf{s})} d\mathbf{s}. \quad (2.63)$$

The negentropy $J(\mathbf{s})$ is defined by normalizing the differential entropy $H_{en}(\mathbf{s})$, where \mathbf{s}_{gauss} is a Gaussian random vector with the same covariance matrix as \mathbf{s} .

To reduce the computational complexity, approximations of non-Gaussianity measures are

often used in most cases. As an example, in the case of the negentropy, approaches use higher-order moments [52] and the maximum-entropy principle [90] to approximate the non-Gaussianity. Note that, the intuitive interpretation of the cost functions is that they are measures of non-Gaussianity [89]. A family of such measures can be constructed by using practically any non-quadratic function G and then consider the difference between the expectation of G for the actual data and expectation of G for Gaussian data with the same variance as the actual data. Usually, people can measure the non-Gaussianity of a zero-mean random variable, u , using any even, non-quadratic, sufficiently smooth function G as follows [105].

$$J(u) = |E_u\{G(u)\} - E_v\{G(u)\}|^p, \quad (2.64)$$

where v is a standardized Gaussian *r.v.* u is assumed to be normalized to unit variance and the exponent $p = 1, 2$ typically. The subscripts denote expectation with respect to z and v .

Obviously, $J(u)$ is a generalization of kurtosis-based approximation. The point is that by choosing G , one want to obtain the approximation of negentropy. Some commonly used functions G are [52]:

$$\begin{aligned} G_1(u) &= \log \cosh(a_1 u), \\ G_2(u) &= \exp(-a_2 u^2/2) \\ G_3(u) &= \frac{1}{2} u^2. \end{aligned} \quad (2.65)$$

where $a_1, a_2 \geq 1$ are some suitable constants, often taken equal to one. In particular, one can obtain more robust estimators by choosing a G that does not grow too fast and to prevent overflows. G_3 is motivated by kurtosis. G_1 and G_2 grow more slowly than G_3 and thus they give more robust estimators.

These approximations of negentropy offer a very good compromise between the properties of the two classic non-Gaussianity measures given by kurtosis and negentropy. When we use the simple kurtosis function $G(u) = \frac{1}{2} u^2$ to measure the non-Gaussianity, the cost function is defined as

$$J_{ICA}(\mathbf{w}) = E\{|\mathbf{w}^H \mathbf{x}|^4\} - 2. \quad (2.66)$$

Recall the CMA cost function equation (2.29), which can be written as,

$$J_{CM}(\mathbf{w}) = E\{(|\mathbf{w}^H \mathbf{x}|^2 - R_2)^2\} \quad (2.67)$$

where R_2 is constant dependent on the signal constellation and can be normalized $R_2 = 1$. Assume the signal was first prewhitened then the CMA cost function is

$$J_{CM}(\mathbf{w}) = E\{(|\mathbf{w}^H \mathbf{x}|^4 - 1)\}. \quad (2.68)$$

As we see, it has the same form to the kurtosis-based ICA function.

2.7.1.2 Basic Algorithms

Before various ICA algorithms, people transform the observed vector, \mathbf{y} , to a new white vector, \mathbf{x} , which components are uncorrelated and their variances equal unity.

$$E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I} \quad (2.69)$$

This common data preprocessing method is generally called pre-whitening or data whitening. The principal component analysis processing simplifies the following processing and is one popular method for whitening. PCA finds an orthogonal basis for given data and can be implemented by singular value decomposition. The components of the received signals are uncorrelated and have unit variances after the process of data whitening. In digital communication, whitening can improve the separation performance and the convergence speed [52] since it reduces the dimension and randomness of the data, then prevents overlearning introduced by noise.

There are many algorithms to deal with ICA. It depends on variations of the basic problem and domains in which the algorithms are derived and their corresponding problems are posed. The problem variations include instantaneous and convolutive mixtures. The problem domains include the time domain and the frequency domain. Furthermore, algorithms based on adaptive or block methods can be used. ICA can be carried out with the help of on-line iterative algorithms, some of them extracting one source at a time (deflation). However, methods based on blocks (batch) become more attractive when computer power no longer appears to be an impediment. The use of block algorithms presents a number of advantages, such as improvements on the convergence speed (both estimation of statistics and optimization), and increased facility to deal with spurious local extrema. In the following, we list some important algorithms.

Four popular time domain algorithms are described: the Infomax [106], the natural gradient

[107], the equivariant adaptive separation via independence (EASI) [88] and the FastICA algorithm [90]. The iterative determination of the separating matrix W according to $\Delta W = W^+ - W$ that defines each algorithm is listed below,

$$\text{Infomax} \quad \Delta W = (W^T)^{-1} - g(\mathbf{u})\mathbf{x}^T \quad (2.70)$$

$$\text{Natural gradient} \quad \Delta W = \mu[\mathbf{I} - g(\mathbf{u})\mathbf{u}^T]W \quad (2.71)$$

$$\text{EASI} \quad \Delta W = \mu[\mathbf{I} - \mathbf{u}\mathbf{u}^T + g(\mathbf{u})\mathbf{u}^T]W \quad (2.72)$$

$$\text{FastICA} \quad \Delta W = \text{diag}(\alpha_i)[\text{diag}(\beta_i + E\{g(\mathbf{u})\mathbf{u}^T\})]W \quad (2.73)$$

where μ is the learning rate and \mathbf{I} is the identity matrix, g is a given nonlinearity and $g(\mathbf{u})$ applies the given scalar nonlinearity g to each component of the vector $\mathbf{u}, \mathbf{u} = W\mathbf{x}$.

The principle of the EASI algorithm is to cancel nonlinear cross correlations in output signals. Compared to the famous Jutten-Herault method [12], it reduces the computational load and improves the stability.

Using the Infomax principle which tries to maximize the mutual information between input and output, the Infomax algorithm [106] is a stochastic gradient algorithms intended to maximize log likelihood. It is the simplest method but it converges very slowly. The convergence can be improved by using the natural gradient or relative gradient methods, as given in Eq. 2.71. Natural gradient is based on the geometrical structure of the parameter space to correct the direction of stochastic gradient and then speed up the convergence speed. However, its convergence speed is still dependent on the choice of learning rate, μ . To overcome this problem, FastICA [90] uses a block algorithm with a fixed-point iteration. These two matrices $\text{diag}(\alpha_i), \text{diag}(\beta_i)$ given in 2.73 provide an optimal step size. This algorithm is a fast and reliable maximization method which originally is designed to maximize the measures of nongaussianity. Also, the FastICA algorithm can be directly applied to maximization of the likelihood [52]. We will describe its complex version in section 2.7.2.1.

2.7.2 Typical Complex ICA Algorithms for Blind Separations

Since a common modulation scheme in wireless communications is 4-QAM or 16-QAM which represent information in a complex domain, the introduction of ICA for separating complex

valued signals is needed. A complex variable s is defined in terms of two real variables s_R and s_I as $s = s_R + js_I$, where $j = \sqrt{-1}$. The statistics of a complex random vector $\mathbf{s} = \mathbf{s}_R + j\mathbf{s}_I$ are defined by the joint pdf or pmf $p_s(s_R, s_I)$ provided that it exists. The expectation of a complex random vector \mathbf{s} is then given with respect to this pdf and is written as $E\{\mathbf{s}\} = E\{\mathbf{s}_R\} + jE\{\mathbf{s}_I\}$. The covariance matrix is written as $R_{ss} = \text{cov}\{\mathbf{s}\} = E\{\mathbf{s} - E\{\mathbf{s}\}\}E\{\mathbf{s} - E\{\mathbf{s}\}\}^H$, where H denotes Hermitian transpose. The pseudo-covariance matrix is defined as $P_{ss} = E\{\mathbf{s} - E\{\mathbf{s}\}\}E\{\mathbf{s} - E\{\mathbf{s}\}\}^T$ [108], where T denotes transpose. These two quantities together define a complex random vector.

In noiseless cases, the closed-form solutions to ICA in the complex-mixture scheme can be found in [109]. For other approaches to ICA, for example when using cumulants as in joint approximate diagonalization of eigenmatrices (JADE) [2] or full second order structure as in strong-correlating transform [108], the extension to the complex case is straightforward. However, in approaches that use nonlinear functions to implicitly generate higher order statistics, such as Infomax [106] and FastICA, this is not the case. Methods using nonlinear updates provide simple and efficient solutions to the ICA problem and thus it is desirable to extend them to complex-valued data as well.

Usually, ICA has its limitations of blind separations in the complex domain, . There are three types of indeterminacies which can arise:

1. Permutation indeterminacy;
2. Sign and scaling indeterminacy;
3. Phase indeterminacy.

The second and third indeterminacies, when combined together, are referred to as the complex scale ambiguity. In consideration of noise, some popular complex ICA scenarios for digital communications are investigated below.

2.7.2.1 Complex Fast Fixed-Point ICA

The complex fast fixed-point ICA algorithm is the direct extension of Fast ICA [92]. Bingham *et al.* [1] focus on sources having circular distributions in order to simplify their derivations. Its contrast function is

$$J_G(\mathbf{w}) = E\{G(|\mathbf{w}^H \mathbf{x}|^2)\} \quad (2.74)$$

where $G : \mathbf{R}^+ \cup \{0\} \rightarrow \mathbf{R}$ is a smooth even function, \mathbf{w} is an n -dimensional complex weight vector and

$$E\{|\mathbf{w}^H \mathbf{x}|^2\} = 1 \quad (2.75)$$

where \mathbf{x} is the received signal after pre-whitening processing. Notice that finding the local or global maximum of a contrast function is a well defined problem only if the function is real. Consequently, people prefer to operate contrast functions on absolute values rather than on complex values. According to the Kuhn-Tucker conditions, the optima of (2.74) under the constraint (2.75) are obtained at points [1] where:

$$\nabla E\{G(|\mathbf{w}^H \mathbf{x}|^2)\} - \beta \nabla E\{|\mathbf{w}^H \mathbf{x}|^2\} = 0 \quad (2.76)$$

where β is a constant and the gradient is computed with respect to real and imaginary part of \mathbf{w} separately. Using a Newton method to solve (2.76),

one obtains the following approximating Newton iteration:

$$\begin{aligned} \mathbf{w}^+ &= E\{\mathbf{x}(\mathbf{w}^H \mathbf{x})^* g(|\mathbf{w}^H \mathbf{x}|^2)\} \\ &\quad - E\{g(|\mathbf{w}^H \mathbf{x}|^2) + g'(|\mathbf{w}^H \mathbf{x}|^2)\} \mathbf{w} \\ \mathbf{w} &= \mathbf{w}^+ / \|\mathbf{w}^+\| \end{aligned}$$

of which the non-linear function $g(u)$ is suggested [1]:

$$g_1(u) = \frac{1}{2\sqrt{\alpha_1 + u}}; g_2(u) = \frac{1}{\alpha_2 + u}; g_3(u) = u,$$

where g is the first derivative of the G and α_1, α_2 are some constants. Maximizing $J_G(\mathbf{w}_i)$ can estimate one independent source. n independent sources can be estimate by using a sum of n contrast functions, and a constraint of orthogonality. Thus one obtains the following optimization problem

$$\begin{aligned} \text{Maximize: } & \sum_{i=1}^n J_G(\mathbf{w}_i) \text{ with respect to } \mathbf{w}_i \\ & j = 1, \dots, n \end{aligned}$$

$$\text{Subject to: } E\{(\mathbf{w}_i^H \mathbf{x})(\mathbf{w}_k^H \mathbf{x})^*\} = \delta_{ik}$$

where $\delta_{ik} = 1$ for $i = k$ and $\delta_{ik} = 0$ otherwise.

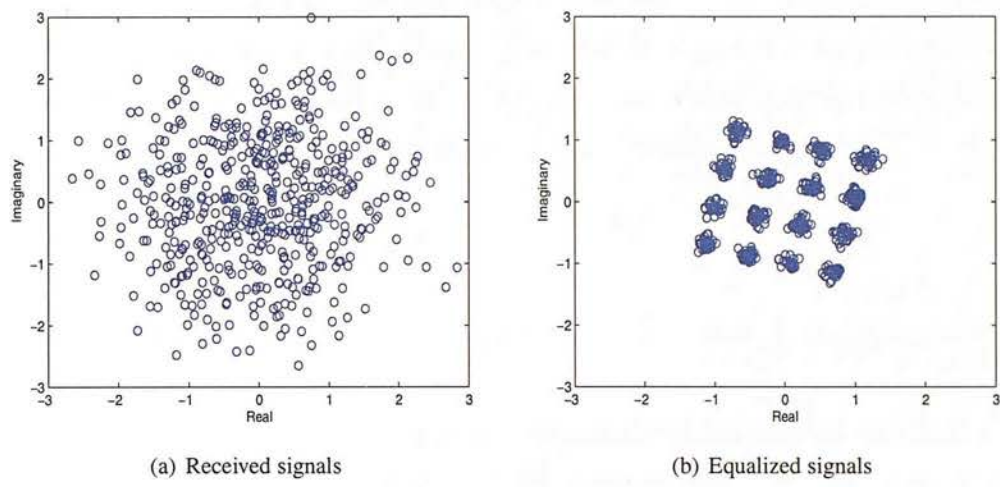


Figure 2.8: Received and separated signals by circular complex fast ICA in a 4×4 system with 16-QAM modulation.

Note that the contrast function defined in (2.74) does not preserve phase information and specifically uses the signal’s modulus for source separation. Because of this property, complex FastICA assumes source distributions to be circular and its performance significantly suffers when the sources are not circular. Despite the circularity assumption, this form of complex FastICA

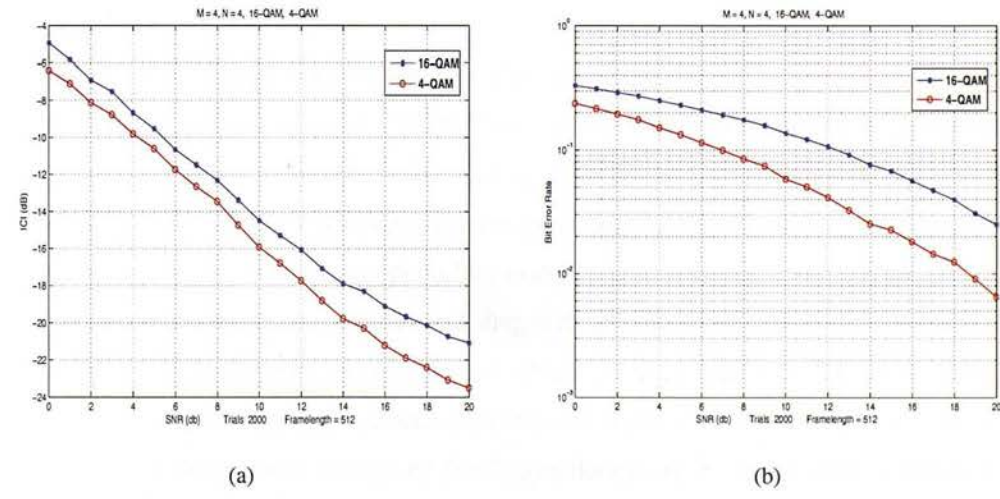


Figure 2.9: BER and separability performances of circular complex fast ICA in a 4×4 system with 4-QAM and 16-QAM modulations.

is a fast fixed point algorithm that works well when the sources are primarily circular because there is an additional ambiguity. Figure 2.8 shows the rotation of the estimated signals by this Complex Fast Fixed-Point algorithm. Where the source signals are 16-QAM modulation.

Clearly, it indicates the phase rotation. In this simulations, 1 symbol is inserted at the beginning of transmitted burst to indicate a phase information. Figure 2.9 illustrates the BER and separability performance which is defined (3.44) in detail. This circular fast ICA is set up in a 4×4 system with 4-QAM and 16-QAM modulation over variant SNR.

2.7.2.2 JADE

JADE algorithm [87] is another popular optimization method for ICA. The contrast function of JADE measures the mutual information by the fourth order statistics, cross-cumulant. Cardoso realizes that making the source independent can be translated as making the cumulants matrices of equalized signals as diagonal as possible. The cumulant matrix with elements, $[Q^{(x)}(M)]_{ij}$, is defined in [110] as:

$$[Q^{(x)}(M)]_{ij} = \sum_{k=1}^n \text{cum}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l) M_{kl} \quad (2.77)$$

where M is an $n \times n$ matrix and \mathbf{x} is the received signal after pre-whitening processing. The contrast function is the sum of squared cross cumulants form equation (2.77),

$$\phi_{jade}(\mathbf{x}) = \sum_{ijkl \neq iikl} (Q_{ijkl})^2 \quad (2.78)$$

Instead of optimizing directly with respect to all dimensions of data, one optimizes the function iteratively in a pairwise manner. JADE finds out the optimum of two dimensional data and transforms the data accordingly. This finally leads to the global optimum by iterating this several times to each pair. The optimization procedure is carried out by orthogonally jointly diagonalizing a set of forth-order cumulant matrices and trying to minimize the off-diagonals in a mean square sense. However, the off-diagonals of different matrices do not usually vanish simultaneously for any parameter value. An advantage of operating on cross-cumulants is that JADE does not require gradient descent and thus avoid problems of converge to local minimum. However, JADE requires storage of $O(n^4)$ cumulant matrices to calculate a complete set of fourth order cumulants.

Figures 2.10 and 2.11 show an estimation of the received signal by the JADE algorithm and illustrate the BER and separability performance of the JADE algorithm in a 4×4 system with 4-QAM and 16-QAM modulation over variant SNR. Again, 1 symbol is inserted at the beginning of transmitted burst to indicate the phase information for the decoding procedure.

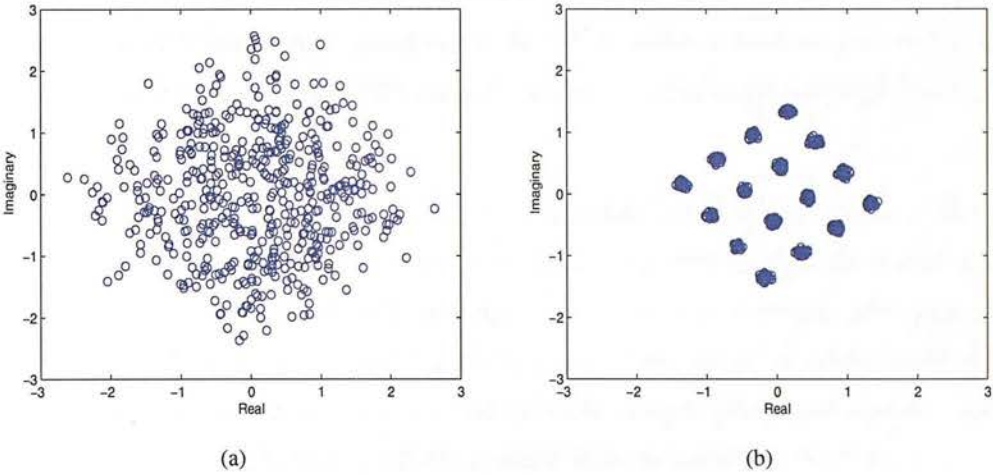


Figure 2.10: Received and separated signals by JADE method in a 4x4 system with 16-QAM modulation.

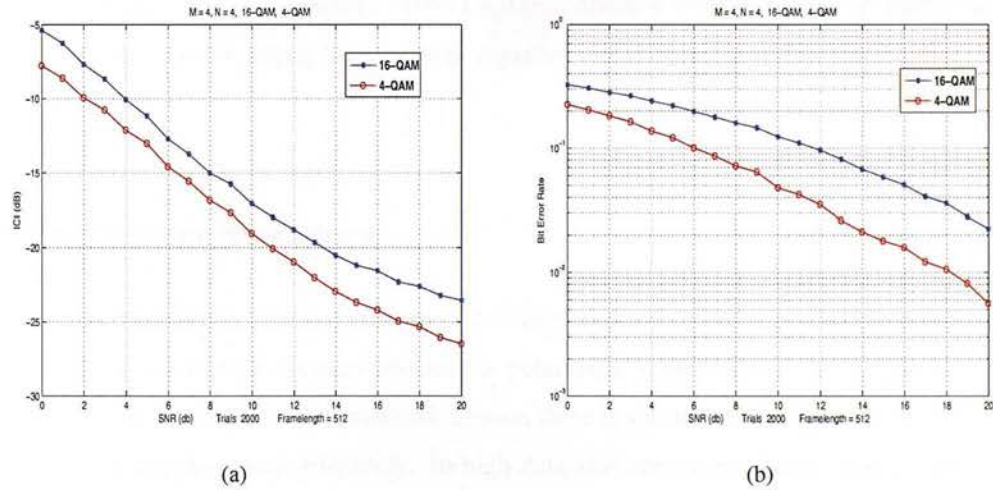


Figure 2.11: BER and separability performances of the JADE algorithm in a 4x4 system with 4-QAM and 16-QAM modulations.

Either complex FastICA or JADE algorithm can separate the QAM signals successfully. When SNR is greater than 20dB, the BER performance is nicely below 10^{-2} for 4-QAM modulation and a little higher SNR requirement is needed for 16-QAM signals. Both algorithm show reliable performance of separating QAM signals. However, they also have disadvantages. JADE estimates a complete set of fourth order cross-cumulants and requires storage of $O(n^4)$ cumulant matrices. This is not problematic and efficient in low dimensional problems. But its

convergence speed becomes increasingly slower as the number of sources increases [111]. The complex FastICA has no such question but the cost function is designed for circular sources [1]. It produces a biased result [112] when the mixture contains some real valued sources, such as BPSK signals.

As we have shown above, the mutual statistical independence of source streams in transmitted antennae exist in a MIMO system, thus this can be exploited through the use of ICA [89] which essentially is a specific HOS technique. As a statistical technique, ICA uses only the independence of the original signals to identify an unknown channel or mixing matrix first and then estimates source signals. As such independence among transmitted signals exists, it is very suitable for the separation of MIMO systems. ICA can usually estimate the source signals up to certain indeterminacies: arbitrary scaling, permutation. One of the advantages of an ICA aided approach is that the signal can be recovered regardless the constant modulus requirement and the spectral characteristics [113]. The ICA based estimation has also proved useful in refining linear detection [94]. This refinement comes from the alleviation of the error estimation of the channel. In more general MIMO models, there is evidence that the performance of ICA based Minimum Mean Square Error equalization is better than the conventional MMSE receivers[104].

2.8 Chapter Summary

As is well known, the radio spectrum is a scarce resource. The use of blind equalization in wireless communication system provides a potentially powerful tool for improving spectral efficiency, such as 18% in GSM systems. Hence, there is a need for intelligent receiver structures that use limited resources efficiently. In high data rate communications, high quality channel estimates are necessary. Blind and semi-blind methods can estimate channel and facilitate the mitigation of ISI with none or limited training data, so they are important and meaningful. Blind receivers further improve the performance by cancelling interference caused by the other users using the same frequency band. In addition, the use of multiple transmitters and receivers allows the systems to take advantage of properties such as spatial diversity in order to exploit existing the channel capacity better.

A brief review on blind equalization and separation is introduced. Both SISO, MIMO systems and SOS, HOS techniques were presented. For conventional blind equalization in SISO links,

Busgang family were exploited, this techniques is iterative equalization that usually employ stochastic gradient descent algorithm to minimize an non-convex cost function of the equalizer output. In the MIMO systems, techniques based on SOS and HOS are introduced. The popular fractionally spaced equalization and subspace method are presented in detail. Methods use explicit HOS for separating signals, such as MNC and SE are also given. They directly employs the minimization of higher order cumulant based nonlinear cost function.

ICA was exhibited mathematically. The corresponding cost function and some basic algorithms were shown. Simulations using the Fast ICA and JADE were illustrated. They all show the ability and feasibility of blind equalizations and separations. The underlying cost functions of BE are multimodal, then the convergence behaviour of these BE algorithms is difficult to converge to the global minimum. In order to avoid convergence to the local minima, a good initial value should be provided. However the good initialization is not always possible or successful. Theory analysis of these convergence properties concern a lot of mathematics, they are not summarized here for simplifications.

Chapter 3

An Efficient Nonlinearity for QAM signals

3.1 Introduction

From the beginning of blind equalization, nonlinearity has played an important part in digital communications. Approaches developed in the early stage utilized Bussgang nonlinearities, where the nonlinearities are memoryless. The statistical properties of Bussgang type algorithms used is that for the separated signal the expectation of the product of input and output of the Bussgang nonlinearity $g(\cdot)$ is equal to the squared input to the Bussgang nonlinearity, $E\{g(s)s\} = E\{s^2\}$. The role of nonlinearities is defined by contrast functions, which can be viewed as a probabilistic descriptor, i.e. likelihood, entropy or mutual information. Since then, blind separation and equalization methods based on nonlinearity attracted research interest. Such kind of ICA methods has two advantages: First, they provide simpler solutions to the ICA problem than joint approximate diagonalization of eigenmatrices (JADE)[2] or full second order structure as in the strong-correlating transform [114]; Second, these nonlinearities can reveal the high order correlation among the signals. Such high order correlations indicate mutual dependence, which then forms an error signal to drive the output signals to a state of higher independence. These high order statistics are produced by the nonlinear functions implicitly. An important point used in this chapter is that good nonlinear functions, in terms of separability performance, are essentially defined by the probability density function of the original source signals. In this chapter, a simple nonlinearity for QAM signals is proposed.

3.2 Optimum Nonlinear Functions

For blind channel identification that use nonlinearities to generate the higher-order statistics, such as complex infomax, maximum likelihood, and negentropy, the optimal choice of nonlinearity is based on the source distribution either through the score function in a likelihood framework, or through entropy in maximization of negentropy [115]. So, when using entropy/non-

Gaussianity as the cost function in ICA, a measurement of non-Gaussianity is negentropy defined for a vector \mathbf{y} as,

$$J_{neg}(\mathbf{y}) = H(\mathbf{y}_{Gauss}) - H(\mathbf{y}) \quad (3.1)$$

where $H(\mathbf{y}) = -E\{\log(p(\mathbf{y}))\}$ is the differential entropy, $p(\mathbf{y})$ is the pdf. Since \mathbf{y}_{Gauss} is a Gaussian random variable with the same covariance matrix as \mathbf{y} , $H(\mathbf{y}_{Gauss})$ is constant, this cost function then is determined by $E\{\log(p(\mathbf{y}))\}$. Also, this can be connected to ML estimation. Observe a noiseless ICA model,

$$\mathbf{y} = H\mathbf{s} \quad (3.2)$$

The pdf of the received signal $p(\mathbf{y})$ can be written

$$p_y(\mathbf{y}) = |\det W| p_s(\mathbf{s}) = |\det W| \prod_i p_i(s_i) \quad (3.3)$$

where $W = H^{-1}$ is the demixing matrix and $p_i(s_i)$ is pdf of i th source signal. Given $W = \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$, equation (3.3) can be written as

$$\begin{aligned} p_y(\mathbf{y}) &= |\det W| p_s(\mathbf{s}) \\ &= |\det W| \prod_i p_{s_i}(\mathbf{w}_i^T \mathbf{y}_i) \end{aligned}$$

The likelihood can be considered as the probability density of the received signal \mathbf{y} . The log-likelihood function $L(W)$ is then given by

$$L(W) = \log |\det W| + \log \prod_i p_{s_i}(\mathbf{w}_i^T \mathbf{y}_i) = \log |\det W| + \sum_{i=1}^n \log p_{s_i}(\mathbf{w}_i^T \mathbf{y}_i) \quad (3.4)$$

The equation above expresses the likelihood as a function of the source pdf, p_{s_i} . To find the maximum likelihood, the knowledge of source density is necessary.

To achieve ICA, both negentropy and ML require the nonlinearity to match the source density. Many source density adaptation methods have been proposed for performing ICA [116][117].

3.3 PDF of QAM modulations

M-ary quadrature amplitude modulation is an important digital modulation systems. It can be regarded as the extension of pulse amplitude modulation to the complex baseband representa-

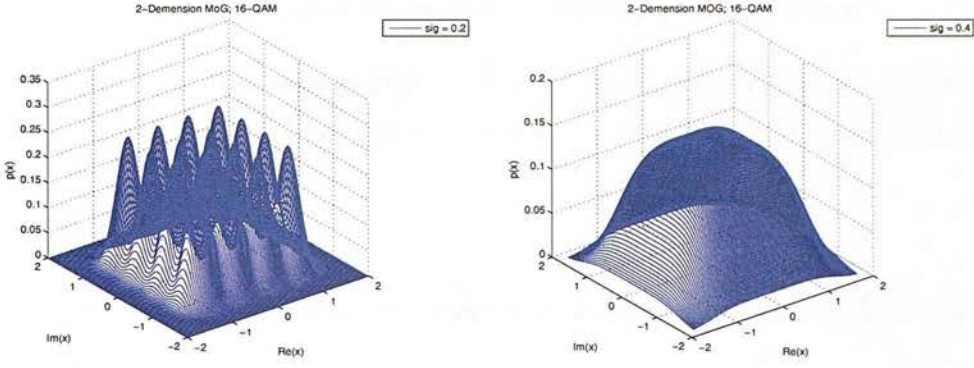


Figure 3.1: MoG PDF of 16-QAM with different variances.

tion. For the common case, in terms of I-Q balance, people always take M as an even power of two, e.g., $M = 4, 16$, the I and Q parts are independent of each other. Generally, QAM distribution is sub-Gaussian. For noiseless complex valued M -QAM sources, the probability mass function is

$$p(s) = \begin{cases} \frac{1}{M} & \text{if } s \in A \\ 0 & \text{otherwise} \end{cases} \quad (3.5)$$

where A is the set of complex points in the constellation, e.g. $A = [1 + j, 1 - j, -1 + j, -1 - j]^T$ for the 4-QAM source.

A mixture of Gaussian (MoG) model can provide a flexible alternative to source density matching. In reality, the system assumes the existence of noise, hence with the addition of complex white Gaussian noise to the sources, a MoG kernels is appropriate for representing the discrete source. Then we model the pdf in equation (3.7) for an M -QAM source with the Gaussian mixture model:

$$p_{QAM}(s) = \frac{1}{M2\pi\sigma^2} \sum_{i=1}^M e^{\left(\frac{-1}{2\sigma^2}((s^R - A_i^R)^2 + (s^I - A_i^I)^2)\right)} \quad (3.6)$$

where σ^2 is the variance of the Gaussian mixture. Figure 3.1 give us an illustration of MoG pdf with different variances. The noise covariance in Figure 3.1(a) is set to 0.2 and in Figure 3.1(b) is set to 0.4. With the increase of covariance, the source pdf change form a discrete distribution to a continuous probability distribution that approximates a pdf with a uniform distribution in the middle part and Gaussian distribution both sides. Different noise variances cause various separability. In following simulations, the noise variance, σ^2 , is fixed to .16 for 4-QAM and 16-QAM since this provides the good performance empirically. Note that for convenience we are treating the noise as being isotropic in the source domain as was also done in [118]. We

note, however a more realistic model is isotropic noise in the sensing domain.

This kind of density modeling method is widely used throughout the fields of machine learning and statistics [119] [120]. It also enable us to maintain the classical ICA model when using the MoG pdf.

3.4 A Simple Nonlinear Functions Based On The QAM modulation

Consider MoG pdf of QAM modulations, when the size of A increases, eg. 16-QAM or 64-QAM, solving the equation (3.6) need heavy calculations. The system computation load then becomes huge and it is prohibitive for real time operation. To reduce this unaffordable complexity, a simple nonlinear function to approximate this MoG pdf is desirable.

As emphasized above, the optimum nonlinear function for blind source separation is directly related to the source pdf. Furthermore, in terms of the Fisher information of a density, the optimum nonlinearity for given source distributions, $p'(s)/p(s)$, were found by applying ML methods. Further research shows that the exact curve of the nonlinearity did not matter too much [121]. These conclusions inspire the simple nonlinearity of the original optimum nonlinear function.

According to equation (3.1), the cost function for QAM signals is

$$J(\mathbf{w}) = E\{\log p_{QAM}(\mathbf{w}^H \mathbf{x})\} \quad (3.7)$$

where \mathbf{w} is the row of the unmixing matrix W and \mathbf{x} is the whitened received signal, which satisfies $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$, the identity matrix.

Figure 3.2 illustrates the real part MoG 16-QAM pdf, its approximation and the corresponding nonlinear function based on this approximation. Here, for simplicity, we give the 1 dimensional illustration in Figure 3.2. The top figure is obtained by the real part of the combination of discrete QAM pdf and 4 AWGN noise with covariance 0.25. The middle figure flare-out the curve of the top figure. And the bottom one is the score function with a pdf of middle figure.

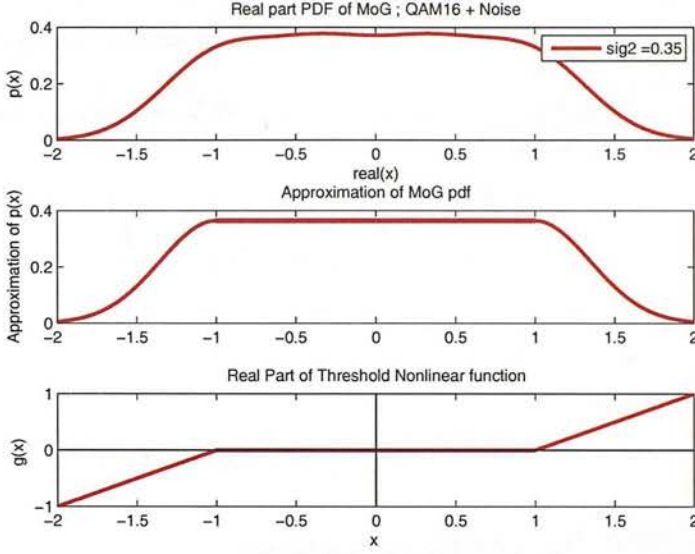


Figure 3.2: Nonlinear approximation of MoG.

3.4.1 Split Complex Nonlinearities

In square QAM systems, the real part and the imaginary parts of the signal are statistically independent. So one can apply the same nonlinearity independently to the real and the imaginary parts of the signal $s = s_R + js_I$. Such a nonlinearity is strongly motivated by the interpretation of the score function as the negative log likelihood of a distribution whose real and imaginary parts are independent. For QAM components, the joint pdf can be written as the product of marginal pdf of real and imaginary parts. And it can be written as

$$p_{QAM}(s) = p_{s_R}(s_R)p_{s_I}(s_I) \quad (3.8)$$

$p_{QAM} : R \times R \rightarrow R$ is the joint pdf of the QAM source. Hence, the nonlinearity becomes

$$\begin{aligned} g(s) &= -\frac{\frac{\partial p_{s_R}(s_R)p_{s_I}(s_I)}{\partial s_R} + j\frac{\partial p_{s_R}(s_R)p_{s_I}(s_I)}{\partial s_I}}{p_{s_R}(s_R)p_{s_I}(s_I)} \\ &= -\frac{p'_{s_R}(s_R)p_{s_I}(s_I) + jp_{s_R}(s_R)p'_{s_I}(s_I)}{p_{s_R}(s_R)p_{s_I}(s_I)} \\ &= -\left(\frac{p'_{s_R}(s_R)}{p_{s_R}(s_R)} + j\frac{p'_{s_I}(s_I)}{p_{s_I}(s_I)}\right) \\ &= g(s_R) + jg(s_I) \end{aligned} \quad (3.9)$$

Nonlinearities of the form given by (3.9) rotate the constellation of a complex signal such that the real and imaginary parts become independent of each other.

Let us look at the equilibrium of the separation point,

$$\begin{aligned}
 & E\{1 - g(s)s^*\} \\
 &= E\{1 - (g(s_R) + jg(s_I))(s_R - js_I)\} \\
 &= E\{1 - [g(s_R)s_R + g(s_I)s_I] + j[g(s_R)s_I - g(s_I)s_R]\} \\
 &= 0
 \end{aligned} \tag{3.10}$$

Two equations can be built from this,

$$E\{g(s_R)s_R + g(s_I)s_I\} = 1 \tag{3.11}$$

$$E\{g(s_R)s_I - g(s_I)s_R\} = 0 \tag{3.12}$$

Equation (3.11) shows that the signals are separated by using both real and imaginary parts. Equation (3.12) enforces the constellation rotation. So, this IQ independent nonlinearity could correct the phase ambiguity up to $k\pi/2$. It can therefore partly solve the question of *phase ambiguity*. We will call such nonlinear function, *Split Complex Nonlinearities*.

Obviously, we can describe this kind of Nonlinear function and the corresponding *Split Non-linear Threshold Function* is given by

$$g(u) = \begin{cases} 0, & |u_R| < \nu; |u_I| < \nu \\ \alpha[(u_R - \text{sgn}(u_R)\nu) + j(u_I - \text{sgn}(u_I)\nu)], & |u_R| \geq \nu; |u_I| \geq \nu \\ \alpha(u_R - \text{sgn}(u_R)\nu), & |u_R| \geq \nu; |u_I| < \nu \\ j\alpha(u_I - \text{sgn}(u_I)\nu), & |u_R| < \nu; |u_I| \geq \nu \end{cases} \tag{3.13}$$

where ν is the threshold and α is the slope of the nonlinearity and it affects the convergence capability and stability. The proposed I/Q independent nonlinearities, apply the same nonlinearity independently to the real and the imaginary parts of the signal $u = u_R + ju_I$.

The key advantages of this nonlinearity over the MoG kernels are:

- Simplicity - The nonlinearity only requires a small number of simple bit-level sign oper-

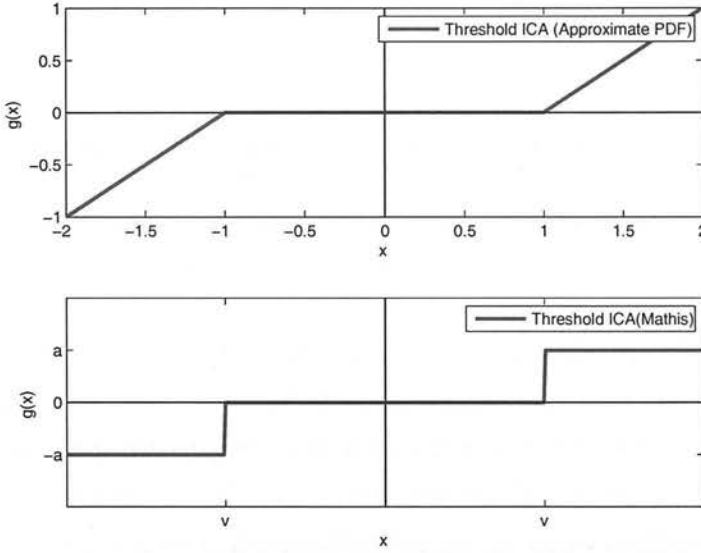


Figure 3.3: *Nonlinear approximation of MoG.*

ations which can take the place of a large amount of multiply and add operations.

- Unimodel density model - It avoids potential problems of introducing spurious local minima.
- Flexible - The single nonlinear activation function is applicable to all QAM modulation schemes irrespective of constellation size.

Another very attractive nonlinear function in the form of a threshold nonlinearity is developed by Mathis in [122] [123]. This universal nonlinearity introduces a parameter into three level quantizer and can separate any non-Gaussian distributions, symmetrically distributed signals. As illustrated in Figure 3.3, it is a clipped version of the score function for the source with uniform distribution with two parameters, the threshold v and the gain a . This threshold nonlinearity works directly for sub-Gaussian signals. Moreover, it can separate any non-Gaussian signal by adjusting the threshold v . The stability criteria for continuous and discrete distributions are also provided. It has the advantageous feature of low computational complexity. On the contrary, our proposed nonlinearity is not a simplification of polynomial function but an approximation of the true score function for Gaussian noise contained discrete sources. As illustrated in Figure 3.3, the slope in nonlinearity is coming from the sidelobe of Gaussian distribution. This property is more suitable for digital modulation with AWGN, as we will see in

Figures 3.8 and 3.9.

3.4.2 Gradient Algorithm of Threshold Nonlinear ICA

Generally, in order to identify a channel or mixing matrix, it is popular to resort to a stochastic gradient algorithm, where the update equation is given as

$$W^{n+1} = W^n + \mu[(W^n)^{-H} - \mathbf{g}(\mathbf{u})\mathbf{y}^H], \quad (3.14)$$

where $\mathbf{u} = \mathbf{w}^H \mathbf{y}$ is equalized signals. μ is the learning step and $\mathbf{g}(\mathbf{u}) = [g_1(u_1), \dots, g_N(u_N)]^T$ is the vector of score functions. Here $g_i(u_i)$ is the threshold nonlinearity mentioned above. To accelerate the convergence, the update equation can use the natural gradient method [107]. The separation matrix W can be formulated as

$$W^{n+1} = W^n + \mu(I - \mathbf{g}(\mathbf{u})\mathbf{u}^H)W^n \quad (3.15)$$

Computer simulation Figure 3.7 will illustrate the separated performance of this split complex nonlinearity based on the gradient algorithm.

3.4.2.1 Steady State Error

Generally speaking, the convergence behavior is tradeoff against convergence speed by the stepsize μ if the useful observed data is fixed. [124] give a measurement of steady state error. Here if we define

$$\begin{aligned} \kappa_+ &= E\{g'(u)E\{u^2\} + E\{g'(u)u\} \\ \kappa_- &= E\{g'(u)E\{u^2\} - E\{g'(u)u\} \\ \gamma_+ &= E\{g^2(u)E\{u^2\} + (E\{g'(u)u\})^2 \\ \gamma_- &= E\{g^2(u)E\{u^2\} - (E\{g'(u)u\})^2, \end{aligned}$$

where g is defined as 3.13. Then, the steady state error is proportional to

$$\xi = \frac{1}{2} \left(\frac{\gamma_+}{\kappa_+} + \frac{\gamma_-}{\kappa_-} \right) \quad (3.16)$$

3.4.3 Fixed-Point Algorithm for Threshold Nonlinear ICA

The gradient method can suffer from slow convergence speed and introduces the difficulty of selecting of the learning step size. So, a Newton-like approach is preferred to overcome these difficulties if an appropriate cost function can be defined for which the second derivative exists.

Given the cost function (3.7), and additionally under the unit power constraint $E\{u_i u_i^*\} = 1$, and constraining the weights to $\|\mathbf{w}\|^2 = 1$ due to the whitening transform, the optimal weight is [118]

$$\mathbf{w}_{opt} = \arg \max_{\|\mathbf{w}\|^2=1} E\{\log p_{QAM}(\mathbf{w}^H \mathbf{x})\}, \quad (3.17)$$

where \mathbf{x} is whitened the received signal. The fixed point algorithm [92] for the constrained optimization problem defined in (3.17) uses a Newton method based on the Lagrangian function

$$L(\mathbf{w}, \lambda) = J(\mathbf{w}) + \lambda(\mathbf{w}^H \mathbf{w} - 1) \quad (3.18)$$

where λ is the Lagrange multiplier. Using the complex gradient and Hessian in [125], the Newton update is defined as,

$$\Delta \tilde{\mathbf{w}} = - \left(\frac{\partial^2 J}{\partial \tilde{\mathbf{w}}^* \partial \tilde{\mathbf{w}}^T} \right)^{-1} \frac{\partial J}{\partial \tilde{\mathbf{w}}^*} \quad (3.19)$$

$$= -\mathcal{H}_J^{-1} \tilde{\nabla}_J^* \quad (3.20)$$

where \mathcal{H} is the complex Hessian, $\tilde{\nabla}^*$ is the conjugate gradient to the Lagrangian function and is of the form $\tilde{\nabla} \in \mathbb{C}^{2N} = [w_1, w_1^*, \dots, w_N, w_N^*]$. We follow the style of Novey and Adali [126], the update of (3.18) is written as

$$\Delta \tilde{\mathbf{w}} = -(\mathcal{H}_J + \lambda I)^{-1} (\tilde{\nabla}_J^* + \lambda \tilde{\mathbf{w}}^n). \quad (3.21)$$

After expanding,

$$(\mathcal{H}_J + \lambda I)^{-1} \tilde{\mathbf{w}}^{n+1} = -\tilde{\nabla}_J^* + \mathcal{H}_J \tilde{\mathbf{w}}^n \quad (3.22)$$

where $\Delta \tilde{\mathbf{w}} = \tilde{\mathbf{w}}^{n+1} - \tilde{\mathbf{w}}^n$. This equation is updated as [126][118]

$$\mathbf{w}^{n+1} = -\frac{1}{2} E\{\mathbf{x} g^*(u)\} + E\{g'_a(u)\} \mathbf{w}^n + E\{\mathbf{x} \mathbf{x}^T g'_b(u)\} (\mathbf{w}^n)^*. \quad (3.23)$$

Further details can be found in [126]. The above equation considers the noncircular nature of

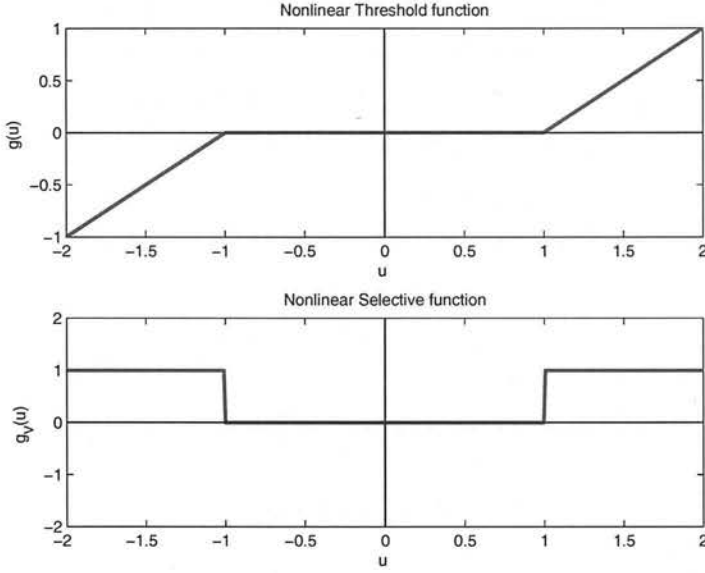


Figure 3.4: Nonlinear threshold function and nonlinear selective function.

the sources through $E\{\mathbf{x}\mathbf{x}^T\}$. In the light of (3.23) and equations in [126], the *split threshold nonlinear* update equations are

$$\begin{aligned} g(u) &= \frac{1}{2}\alpha \left\{ g_v(u_R)[u_R - \text{sgn}(u_R)v] + \right. \\ &\quad \left. j g_v(u_I)[u_I - \text{sgn}(u_I)v] \right\} \\ g'_a(u) &= \frac{1}{4}\alpha [g_v(u_R) + g_v(u_I)] \\ g'_b(u) &= \frac{1}{4}\alpha [g_v(u_R) - g_v(u_I)]. \end{aligned} \quad (3.24)$$

Where we define the *nonlinear selective function*, $g_v(s)$, as

$$g_v(u) = \begin{cases} 0, & |u| < v \\ \alpha, & |u| \geq v. \end{cases} \quad (3.25)$$

Fig. 3.4 illustrated the nonlinear threshold function and the corresponding nonlinear selective function, where $v = 1$ and $\alpha = 1$.

3.4.4 Local Stability Analysis

The stability property of ICA was well studied for the real case by Amari [127], by analyzing the second order statistics of ICA for real symbols. Also, Cardoso subsequently proposed a similar

solution in [128]. In this section, similar to Amari's analysis, we explore the local stability of the proposed complex nonlinearity for complex valued signals. We study the stability of $J(\mathbf{w})$ under the constraint that $\|\mathbf{w}\|^2 = 1$. We apply an orthogonal change of coordinates $\mathbf{q} = \mathbf{w}^H H$ and assume an optimum solution is $\mathbf{q}_i = [0, \dots, i, \dots, 0]^T$ without consideration of permutation and scalar ambiguity problems. Also, without loss of generality, we consider $\mathbf{q}_1 = [1, 0, \dots, 0]^T$. Therefore, the Lagrangian function of the constrained optimization is,

$$L(\mathbf{q}, \lambda) = J(\mathbf{q}) + \lambda(\mathbf{q}^H \mathbf{q} - 1) \quad (3.26)$$

The second order necessary and sufficient conditions for a local minimum are

$$\nabla_{\mathbf{q}} L(\mathbf{q}_1) = 0, \quad \nabla_{\lambda} L(\lambda) = 0 \quad (3.27)$$

and

$$\mathbf{z}^H \nabla_{\mathbf{q}}^2 L(\mathbf{q}_1) \mathbf{z} \geq 0 \quad (3.28)$$

where \mathbf{z} is a feasible direction $\{\mathbf{z} | (\nabla(\mathbf{q}^H \mathbf{q} - 1)^H \mathbf{z} = 0)\}$ [129].

The first condition (3.27) can be obtained by equation (13) in [118] at \mathbf{q}_1 . Since the sources are mutually independent,

$$\nabla_{\mathbf{q}} L(\mathbf{q}_1) = \frac{1}{2} \mathbb{E} \begin{pmatrix} s_1 g^*(s_1) \\ s_1^* g(s_1) \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3.29)$$

Setting either of the top two rows of equation 3.29 equal to zero and leads to

$$\lambda_1 = -\frac{1}{2} \mathbb{E}\{s_1 g^*(s_1)\}. \quad (3.30)$$

The second condition (3.28) can be obtained from paper [126], we find

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial q_i^* \partial q_j} &= \mathbb{E}\{s_i s_j^* g_a'\} & \frac{\partial^2 J(\mathbf{w})}{\partial q_i^* \partial q_j^*} &= \mathbb{E}\{s_i s_j g_b'\} \\ \frac{\partial^2 J(\mathbf{w})}{\partial q_i \partial q_j} &= \mathbb{E}\{s_i^* s_j^* g_b'\} & \frac{\partial^2 J(\mathbf{w})}{\partial q_i \partial q_j^*} &= \mathbb{E}\{s_i s_j^* g_a'\} \end{aligned} \quad (3.31)$$

In terms of our specific complex split nonlinearity, substituting (3.31) into the Hessian form of $\nabla_q^2 L(\mathbf{q}_1)$ defined in [118], we obtain

$$\nabla_q^2 L(\mathbf{q}_1) = \mathbb{E} \left[\begin{pmatrix} A & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & B_N \end{pmatrix} \right] + \lambda_1 \mathbf{I}. \quad (3.32)$$

This Hessian matrix is a block diagonal matrix with 2×2 submatrices A and B_i ,

$$A = \begin{pmatrix} s_1 s_1^* g'_a(s_1) & s_1^2 g'_b(s_1) \\ (s_1^*)^2 g'_b(s_1) & s_1 s_1^* g'_a(s_1) \end{pmatrix} \quad (3.33)$$

$$B_i = \begin{pmatrix} g'_a(s_1) & s_i^2 g'_b(s_1) \\ (s_i^*)^2 g'_b(s_1) & g'_a(s_1) \end{pmatrix}. \quad (3.34)$$

We now examine the behaviour through the submatrices A and B . The stability condition (3.28) requires $(A + \lambda_1 \mathbf{I})$ and $(B_i + \lambda_1 \mathbf{I})$ to be positive definite. Noting that the form of A and B_i is

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{12}^* & H_{11} \end{pmatrix} \quad (3.35)$$

The eigenvalues of H , κ_{\pm} , can be found,

$$\kappa_{\pm} = H_{11} \pm |H_{12}|. \quad (3.36)$$

For submatrices A and B_i , H_{11} is equal to A_{11} and B_{i11} respectively. Substituting (3.33) and (3.34) into the equation (3.36), the stability conditions can be obtained as:

$$\mathbb{E}\{s_1 s_1^* g'_a(s_1)\} + \lambda_1 \pm |\mathbb{E}\{s_1^2 g'_b(s_1)\}| > 0 \quad (3.37)$$

and

$$\mathbb{E}\{g'_a(s_1)\} + \lambda_1 \pm |\mathbb{E}\{s_i^2 g'_b(s_1)\}| > 0. \quad (3.38)$$

For QAM modulations, simplifications occur when the threshold value is set to the outermost symbol amplitude, either the real or imaginary part, *e.g.* $\nu = |Re(s_1)|$. Under this constraint,

$g'_b(s_1) = \frac{1}{4}[g^{RR} - g^{II}] = 0$, and

$$\begin{aligned}
 \lambda_1 &= -\frac{1}{2}E\{s_1 g^*(s_1)\} \\
 &= -\frac{1}{2}E\{g^R(s_1)s_1^R + g^I(s_1)s_1^I\} + jE\{g^R(s_1)s_1^I - g^I(s_1)s_1^R\} \\
 &= -\frac{1}{2}E\{g^R(s_1)s_1^R + g^I(s_1)s_1^I\} \\
 &= 0
 \end{aligned} \tag{3.39}$$

where $g^R(s_1) = g^I(s_1) = 0$. Then stability conditions (3.37) and (3.38) can be simplified as

$$E\{|s_1|^2 g'_a(s_1)\} > 0 \tag{3.40}$$

and

$$E\{g'_a(s_1)\} > 0. \tag{3.41}$$

A closer look at equation in [118] reveals that $E\{g'_a(s_1)\}$ is always greater than 0, So the stability conditions are set up when we select the threshold value appropriately, such as around the outermost symbol amplitude.

3.5 Nonlinear Threshold ICA With Bias Removal

An algorithm of the form given by the equation (3.23) will cause a biased estimate when AWGN is added at the receiver. The linear operations of ICA try to find the inverse of the channel matrix and multiplied by observed signals contaminated by AWGN. Therefore this operation introduces noise to the output. However, if the noise covariance of AWGN is known, an unbiased form can be designed.

This new complex bias removal for robust estimations extends the previous work [130] to complex domain. Using the known noise covariance, [130] take this into account to correct the second order statistics of the observed data. Then, he used the one-unit contrast functions in noise cases. These process can be considered as removing the asymptotic bias and are roust against noise. Generally, the data whitening, as mentioned in section 2.6.1.2, is not robust with respect to noise when the channel is near singular and then degrade the final separation performance.

Similar to data preprocessing operations in ICA, in the preliminary whitening, the effect of noise must be thought over, this is called quasiwhitening [130]. Then we apply the quasi-whitened data to Fast fixed ICA with little modification (refer appendix B). The unbiased form of the equation for fast fixed ICA in complex domain is given as

$$\begin{aligned} \mathbf{w}^+ &= \mathbf{E}\{\mathbf{z}^+(\mathbf{w}^H \mathbf{z}^+)^* g(|\mathbf{w}^H \mathbf{z}^+|^2)\} \\ &\quad -(I + \sigma^2(\Lambda - \Sigma)^{-1})\mathbf{w}\mathbf{E}\{g(|\mathbf{w}^H \mathbf{z}^+|^2) + |\mathbf{w}^H \mathbf{z}^+|^2 g'(|\mathbf{w}^H \mathbf{z}^+|^2)\} \end{aligned} \quad (3.42)$$

where \mathbf{z}^+ is the quasi-whitened data. A detailed derivation is given in appendix B. Λ is a diagonal matrix of eigenvalues from the eigendecomposition of the covariance matrix $\mathbf{E}\{\mathbf{Y}\mathbf{Y}^H\}$ and Σ is the noise covariance. This equation is the corresponding unbiased form of nonlinear complex fast ICA. Element $(I + \sigma^2(\Lambda - \Sigma)^{-1})$ can be considered as a correct factor which modifies the covariance matrix shape by the noise. Then, the corresponding fixed algorithm update equation (3.23) is given as,

$$\mathbf{w}^{n+1} = -\frac{1}{2}\mathbf{E}\{\mathbf{x}g^*(u)\} + \{I + \sigma^2(\Lambda - \Sigma)^{-1}\}\{\mathbf{E}\{g'_a(u)\}\mathbf{w}^n + \mathbf{E}\{\mathbf{x}\mathbf{x}^T g'_b(u)\}(\mathbf{w}^n)^*\}. \quad (3.43)$$

This bias removal technique inherit from the noise free fixed point algorithm the advantage of being simple computationally.

3.6 Simulations

In this section, we first provide an example to illustrate the $k\pi/2$ rotated signal recovered by the *Split Threshold nonlinearity*. The following figure shows the phase ambiguity improvement by our nonlinear function compared with other popular complex ICA methods [1][2][3] in a 4×4 MIMO system with independent Rayleigh channels.

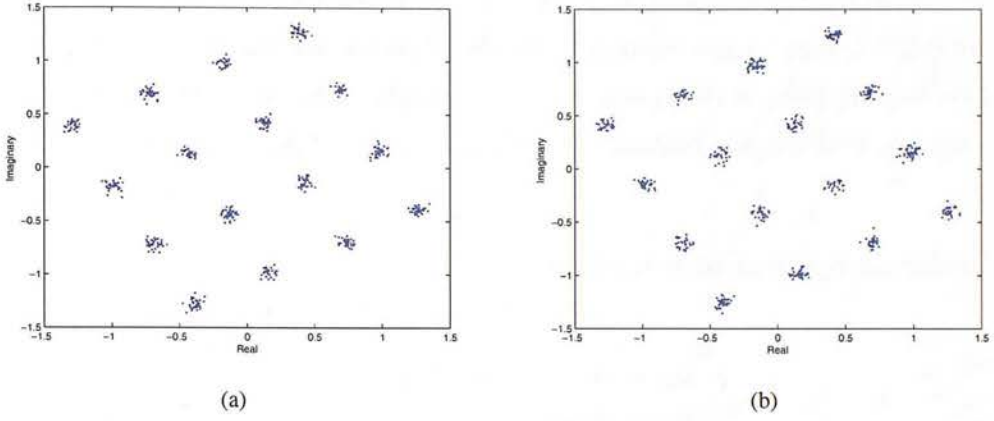


Figure 3.5: Separated signals. (a) : The complex fixed fast ICA algorithm [1]; (b): the JADE algorithm [2].

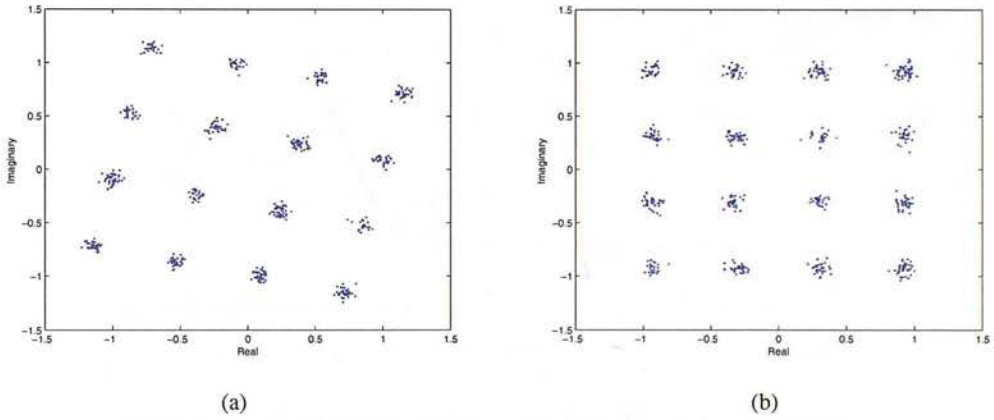


Figure 3.6: Separated signals. (a): the Douglas [3] algorithm [3]; (d): The proposed nonlinear threshold ICA algorithm.

We can see that in the figures 3.5 and 3.6, all algorithms have successfully separated the sources. There is full phase ambiguity in subfigures of 3.5 (a), (b) and 3.6 (a) while in subfigure 3.6 (b), the phase ambiguity is only integers of $\pi/2$. Consider cost functions of complex fast ICA, Eq. 2.74 and JADE, Eq.2.78, the argument of their cost functions are both operated on a absolute value. Thus the phase ambiguity then remains. On the contrary, our nonlinearity have not this limitation and can partly solve the phase ambiguity up to $k\pi/2$. The following $k\pi/2$ phase ambiguity can also be solved by further differential encoding techniques, as pointed out in [131].

Figure 3.7 shows the performance of gradient algorithm of the proposed threshold nonlinear

function and the threshold nonlinear function [122][123]. A 4 transmitter and 4 receiver MIMO system with 16-QAM modulation and flat Rayleigh channels was set up and 100 runs were taken. The optimum values according to Eqs. (5.39) and (5.40) in [122] are $\nu = 0.95$ and $a = 2.1$, and the learning step is fixed, $\mu = 0.001$. Our threshold and gain are $\nu = 1$ and $a = 1$ respectively.

Both nonlinearities convergence designed points with increasing the sample points. Our method is little better than the Mathis method in noiseless cases.

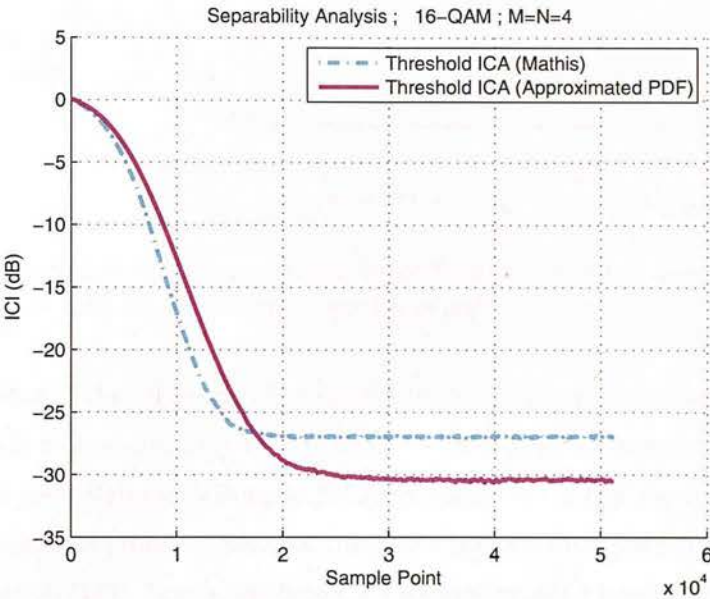


Figure 3.7: *Convergence Properties of the Threshold Nonlinearity.*

In this simulation, the SNR is set to 20dB, the sources are 64-QAM modulation. For Mathis nonlinearity, the optimum values according to Eqs. (5.39) and (5.40) in [122] are $\nu = 1.08$ and $a = 3.7$. Again, our threshold and gain are set to $\nu = 1$ and $a = 1$. The learning steps of both nonlinearities are equal to $\mu = 0.001$ and a total of 100 runs were taken. The separability performance is defined as Eq. 3.44. Figure 3.8 show that our method is better than Mathis threshold ICA. As we sated before, the threshold nonlinearity function [123] is a general function. Our proposed nonlinearity function is specifically designed for QAM signals with noise. It is an approximation of the true score function for Gaussian noise contained discrete sources. The slope in nonlinearity is coming from the sidelobe of Gaussian distribution. Our source model matches the noise model and then leads to good separability performances. Again, this explains that our accurate nonlinear function based on the score function can ameliorate separation qualities.

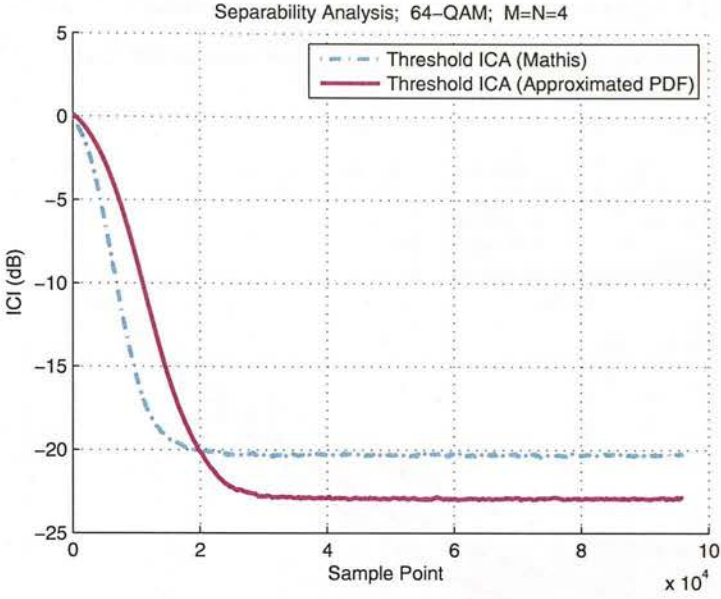


Figure 3.8: Comparison of different threshold nonlinearities in noisy cases. The clipped version of pdf Vs. approximated version of pdf.

The performance of our algorithm (Threshold ICA) is compared with complex FastICA (Circular FICA) [1] with nonlinearity $G(y) = \log(0.1 + |y|^2)$. Also we compare performance with a recent fixed-point algorithm (Douglas FICA) proposed in [3] that showed excellent performance for this kind of problem. Also, we compare it with the MoG pdf in [118] and the simple nonlinear function [123]. In this simulation, a 4 transmitter and 4 receiver MIMO system with 16-QAM modulation is employed and we fix the noise variance $\sigma^2 = 0.16$ which we found empirically provide the good performance in terms of separability. The channel H is a 4×4 complex instantaneous matrix, which is constant for each block interval (512 symbols), and it follows a Rayleigh Fading distribution. N is complex additive white and follows a Gaussian distribution. A total 1000 Monte Carlo runs were taken. We set $\nu = 1$ and $\alpha = 0.5$ in the following simulations. The separability performance of the *Split Threshold nonlinear* function can be further evaluated by measuring the distance of the estimated value from the true value. We define it as: $P = W^H H_{real}$

$$ICI(P) = \frac{1}{n} \sum_i \sum_j \left[\left(\frac{|P_{ij}|}{\max |P_{ij}|} \right)^2 - 1 \right], \quad (3.44)$$

where ICI stands for inter component interference.

Figure 3.9 shows that the *split threshold nonlinear* has the same capability as the true MoG pdf

for various SNR but the complexity is . Again, our method is much better than Mathis threshold ICA especially in low SNR region. And obviously, its separability performance is superior to all other methods.

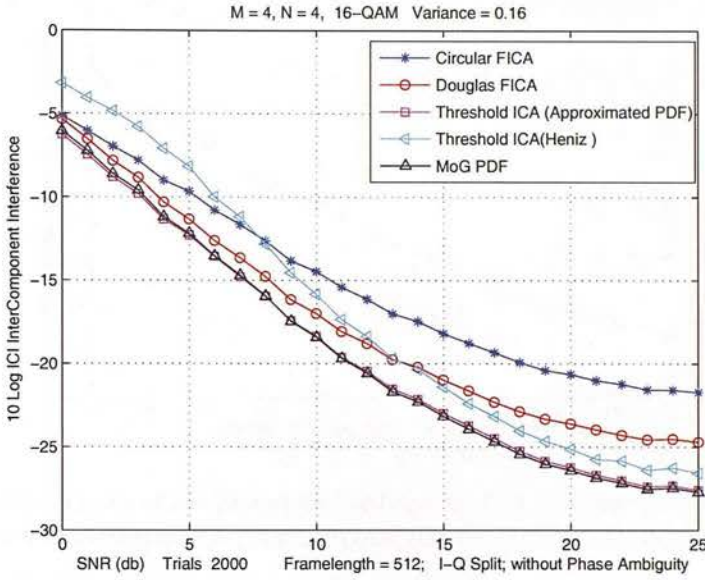


Figure 3.9: Separability of the Threshold Nonlinearity ICA vs. other complex ICA methods in a 4×4 system with 16-QAM modulation.

Such performance happens not only in low dimensional systems, but also with a large number of transmitters and receivers, such as in a 9×9 system. A test is set up by multi-modulation system with the sources: 3 4-QAM, 3 16-QAM and 3 64-QAM respectively. Figure 3.10 shows that the separability performance is always better than other methods. For fixed SNR = 15dB, Figure 3.11 shows the performance as a function of the number of samples and the *Nonlinear Threshold Function* achieves the better performance over all sample sizes. Figure 3.12 compares the performance for different numbers of transmitters and receivers. Note that although this measurement is unaffected by phase ambiguity, the separability performance based on our algorithm is still better than others.

The results indicate that implicitly matching the nonlinear function to the source distribution provides significant improvement over non-matching score functions. The updating equation (3.23) is essentially a complex version of FastICA [92] without phase ambiguities. So it should enjoy the properties of the FastICA, eg. convergence speed, stability.

Furthermore, the complexity is very important in real time wireless communication systems.

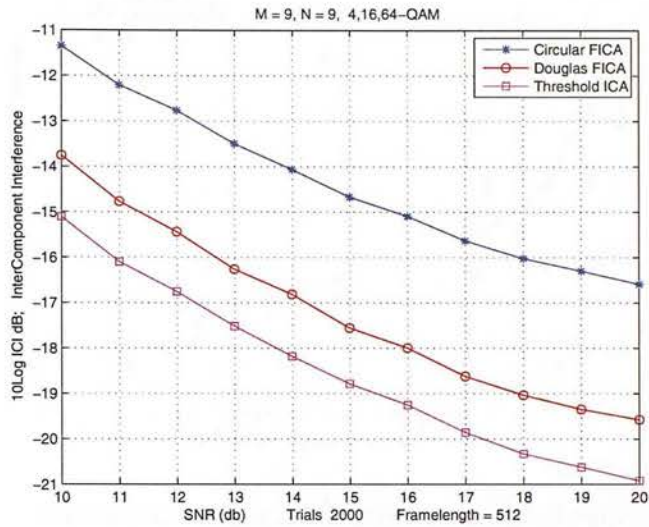


Figure 3.10: Separability of the Threshold Nonlinearity ICA vs. other complex ICA methods in a 9×9 system with 4-QAM,16-QAM and 64-QAM modulations simultaneously.

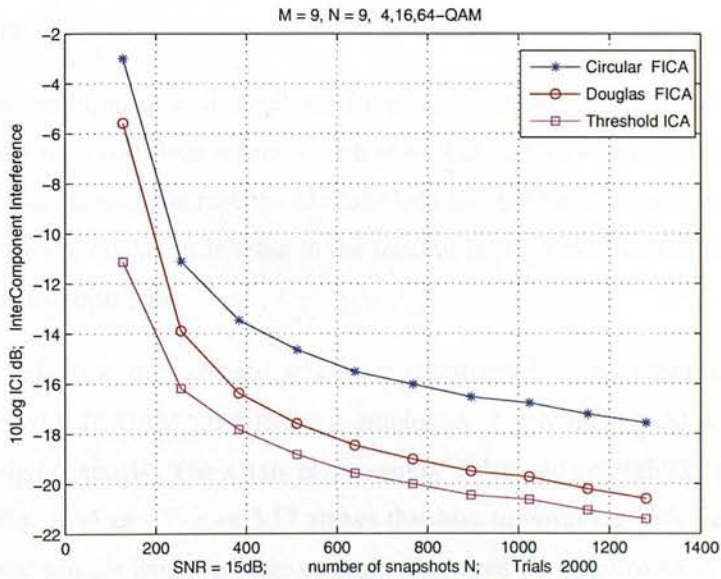


Figure 3.11: ICI vs. number of snapshots N for the different algorithms.

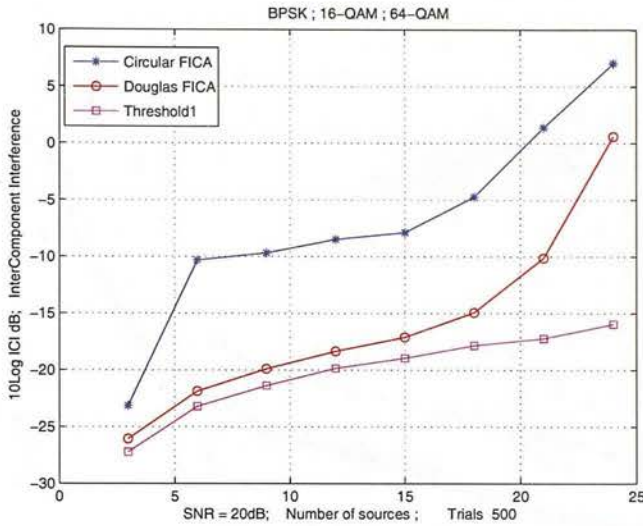


Figure 3.12: SNR = 20 dB; ICI vs. Number of sources

Compared with the nonlinear function based on the MoG pdf, the hyperbolic nonlinearities such as \tanh , \sinh , and the inverse circular nonlinearities such as \arctan , \arccos , this nonlinear threshold function requires only a small number of bit-level sign operations. As we know, sign operations always exist in real time processors such as digital signal processor(DSP), FPGA and avoids many multiply and add operations. Such a property is important for any real time implementation.

In addition, the nonlinearity is also suitable for adaptive QAM modulation schemes and combinations of various modulation schemes such as in WiMAX standards [132], where based on the received signal quality, the receiver can take into account the channel condition, and make the decision of the modulation scheme in the transmitter to maximize the throughput for the available signal-to-noise ratio.

In the final simulations, the unbiased scheme is investigated. A 4 transmitter and 4 receiver MIMO system with 16-QAM modulation is employed. The SNR is fixed at 15 dB and independent Rayleigh channels. The x axis is the sample index and separability performances are measured by Eq. 3.44 too. Figure 3.13 shows that bias removal algorithm can improve estimation when the sample length is large enough. However, the improvement is not significant when the data length is small as Figure 3.14 illustrated for a sample length of 256 symbols.

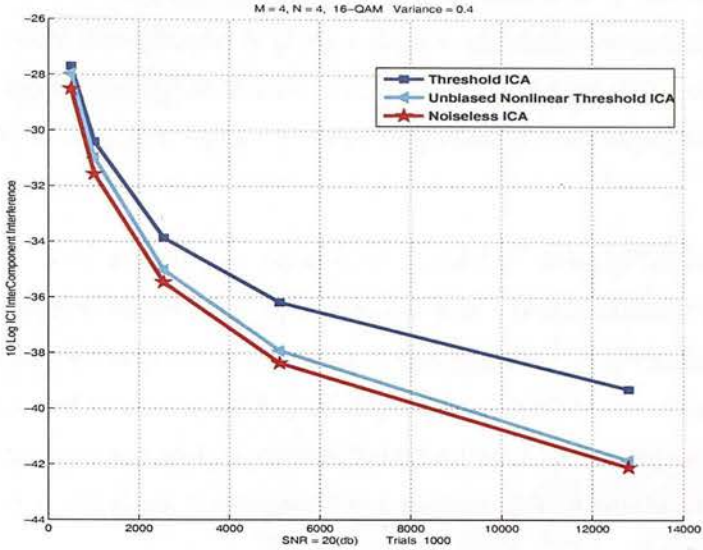


Figure 3.13: SNR = 15 dB; Separation performance of bias removal algorithm for QAM signals.

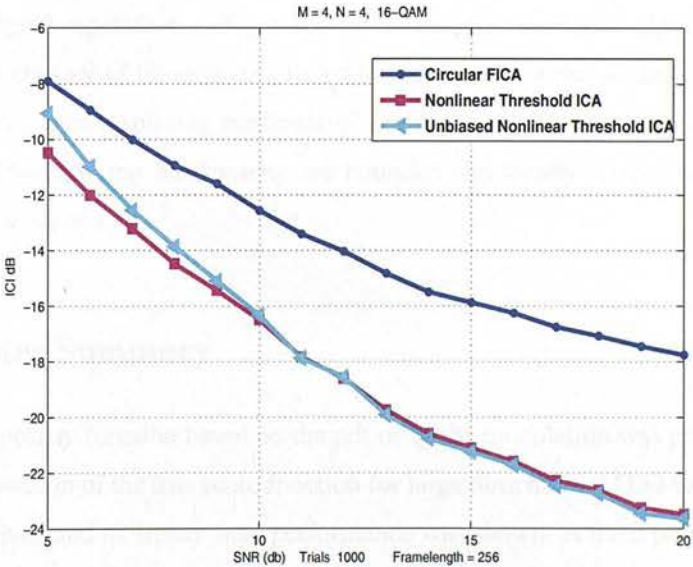


Figure 3.14: Separation performance of bias removal algorithm for QAM signals over various SNR with sample length of 256 symbols.

Although we proposed an efficient nonlinearity for QAM signals in this chapter, there is some previous research which has been designed for blind separation of QAM signals in a MIMO narrow band channel. For example, in [133], a discrete alphabet of sources is known. The authors take additional advantage of the alphabet knowledge, an intermediate contrast is designed and investigated. knowledge brings a significant improvement to the steady-state achievements of the adaptive separation system.

In another work [134], a maximum a-posteriori probability criterion for instantaneous BSS with a finite alphabet is implemented by minimization of a polynomial criterion. An efficient algorithm dedicated to this polynomial criteria is then proposed. In [135], for QAM pdf, each symbol is represented as a point and they construct a smoothed version of the ideal QAM pdf using the \tanh function and apply maximum likelihood blind source separation to QAM signals. By contrast, QAM signal is represented as a mixture of Gaussian (MoG) in our proposed scheme.

Based on the robust estimation and statistics technology, the authors introduced the Huber M-estimator cost function as a contrast function for use within the FastICA algorithm [136]. Similar to our method, this algorithm obtained from Huber M-estimator cost is particularly simple to implement and works well.

For complex signal separation, [137] proposed an adaptive nonlinear algorithm which maximized the joint entropy of the outputs. To avoid the restriction due to the Liouville's theorem [138], the authors used a splitting nonlinearity, one for the real and one for the imaginary part of the input. This splitting nonlinearity are bounded and locally analytic and works well in separating QAM signals.

3.7 Chapter Summary

A simple nonlinearity function based on the pdf of QAM modulation was proposed. This is an approximated version of the true score function for large dimensional QAM signals. The gradient form was given and its steady state performance was shown. A fixed point algorithm based upon this nonlinearity was also derived. The stability analysis is given and the bias removal update designed. Simulations show a good performance in terms of ICI both in large constellation QAM modulation system and few short block size wireless communication in various channels. Since the output of this proposed nonlinearity only has three values, the threshold

operation can be easily implemented in hardware by two comparators. Such a property makes implementation feasible for real time communication systems.

Chapter 4

Blind MIMO Equalization and Decoding for Square QAM Modulation

4.1 Introduction

In communication systems, system performance and computational complexity play key roles. Reducing the computational load and providing accurate performance are the main challenges in current systems. In this chapter, a hybrid method which can provide an affordable complexity with good performance for Blind Equalization in large constellation MIMO systems is proposed. Saving computational cost happens both in the signal separation part and in the signal detection part. Based on an efficient and simple nonlinear function for ICA introduced in the previous chapter, an initial channel estimate is obtained. Then the idea of the list sphere decoding (LSD) is used to acquire the channel soft information which is conditioned on the received data. This approach overcomes the so-called curse of dimensionality of the Expectation Maximization (EM) algorithm and enhances the final results simultaneously. Mathematically, we demonstrate for digital communication cases that the EM algorithm shows Newton-like convergence. All results are confirmed experimentally for the example of blind equalization in a block fading MIMO channel.

Although ICA can blindly equalize MIMO systems and has attracted enormous research interest, the performance is typically not good enough for communication applications. According to the common demands in digital communications, the BER performance should be below 10^{-2} in a reasonable range of power without any channel code. Therefore original ICA algorithms would need to be improved if it was to be used in realistic scenarios. Researchers and engineers always use Maximum likelihood estimator to make further improvements. However, to the authors knowledge, there is no efficient ML solution for high dimensional applications, particularly for large (greater than 16-QAM) constellation QAM modulation. The primary difficulty is the high complexity of the demanded computations. Such constraints make real

applications infeasible in practice. For this reason, we propose a solution to this problem with feasible complexity while maintaining good performance. In this chapter, this method is based on an ICA method as an initial estimation of channel and refined by the expectation maximization method, which can provide a significant improvement both in terms of mean square error and bit error rate (BER). More useful is that the signal detection based on the channel identification can improve the estimation performance compared with estimating signals directly [139][140]. However, we point out that the main drawback of this method is that the inaccuracies of the channel estimation by BE methods will effect the following signal detection stage.

In this chapter, our work also concentrates on the same narrow band system model, as we used in chapter 3.

$$Y = HS + N \quad (4.1)$$

This instantaneous ICA model is fundamental in communication systems [89][52], and can be extended to other popular models, such as MIMO OFDM systems[141][142]. The goal of this chapter is to develop a simple and efficient blind method for high dimensional QAM systems that can fully exploit the known QAM structure.

4.2 The EM Algorithm

As mentioned before, channel estimation by ICA does not produce a satisfactory performance in all circumstances. ICA, like most blind or semi-blind methods such as those based on Godard's method or the constant modulus algorithm, typically employs a linear equalization structure, even though the mechanism to adapt the equalizer coefficients is nonlinear. This restricts the ability to accurately estimate the signal in the presence of noise since the linear estimate amplifies the measurement noise and estimated signals maybe grossly inaccurate. For this reason, we resort to an efficient ML estimator.

Blind ML channel estimators and equalizers have been developed for deterministic signal and stochastic signal models, respectively. The noise is commonly assumed to be white Gaussian and the form of the pdf is assumed to be known.

The corresponding techniques are called deterministic maximum likelihood (DML) and stochastic maximum likelihood (SML) methods, respectively. When applying the stochastic maximum

likelihood method, the input is assumed to be a random i.i.d sequence and its distribution is also known. For example, the samples of a 16-QAM signal can take 16 different values, each one with probability 1/16. The channel provides the only unknown parameters. Our method in fact is a kind of SML method, the probability density is parameterized only by the channel state information, and the estimate is expressed as

$$[H_{ML}] = \arg \max L(y, H) \quad (4.2)$$

where $L(.)$ is the likelihood function of the observed signals. Although ML methods provide excellent performance, they suffer from a high complexity since the ML complexity increases exponentially with the number of transmitted antennae and the number of bits per modulation symbol. Thus finding the ML estimate with acceptable complexity is usually difficult. Typically, an iterative algorithm is applied to access a ML result, such as the EM [143] algorithm. The following figure given below is to illustrate the ML detector is much better than other linear and decision feedback detectors.

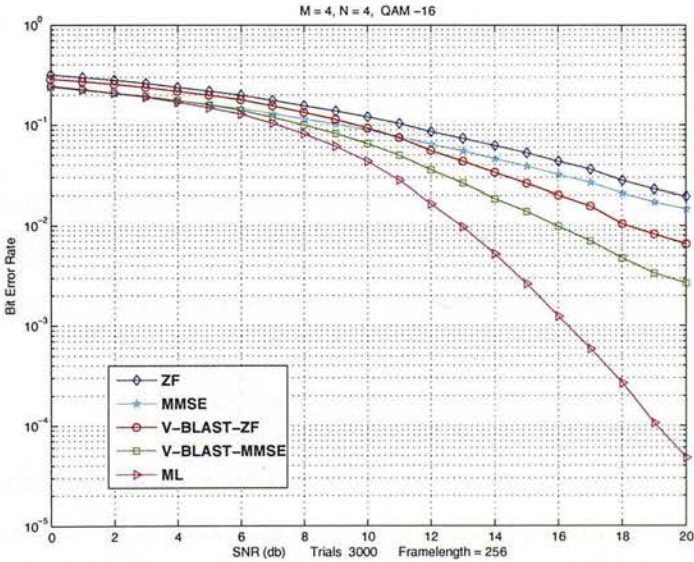


Figure 4.1: BER performance of different MIMO detectors.

Figure 4.1 shows the BER performance of the different MIMO detection algorithms in a 4×4 system with 16-QAM modulation. The linear ZF and MMSE detectors use ZF MMSE criteria for the decoding procedure. The V-BLAST-ZF and V-BLAST-MMSE [9] schemes use ZF and MMSE detections followed by nulling, cancellation and ordering operations. This kind of

equalizer is a decision feedback detection with the complexity lying between the linear and the ML detectors. The ML approach is implemented by an SD method [144] and has the best BER performance. Consequently, this BER improvement explains the interest in finding a counterpart to the blind MIMO detection while providing an approximation to ML performance. All detections mentioned here have perfect CSI at receivers. Obviously, from linear detection to the ML detection, the BER performance improves as the decoding complexity increases.

4.2.1 The EM Principal

In this section, a brief review of the EM algorithm is provided since it plays an important role in this dissertation. The contents presented can help the reader's understanding of subsequent material.

Denote $Y \in \mathbb{C}^{N_R \times T}$ as the observations from a system that relies on some unknown parameters which are denoted as $\Theta \in \Omega$. For our blind system, Y is the baud sampled signals from N_R receive antennae and T is the number of observations. Θ represents channel state information, H , and the noise covariance matrix Σ . Assume knowing Σ , the ML estimate is defined as:

$$\hat{H} = \arg \max_H p(Y|H) \quad (4.3)$$

where $p(Y|H)$ is the pdf of $p(Y)$ conditioned on H . Equation (4.3) looks brief but maximizing it may be intractable in some applications. In MIMO systems, the likelihood $p(Y|H)$ of the received signals is the marginal distribution since we do not know the transmitted signal matrix S .

$$p(Y|H) = \sum_S p(Y, S|H) = \sum_S p(Y|S, H)p(S) \quad (4.4)$$

where $p(Y, S|H)$ is the joint likelihood of S and Y . $p(S)$ is the prior probability of source signal S . With the assumption of the independence, the sum over $|S|^T$ terms reduce to $T|S|$ terms. $|S|$ denotes the number of signal configurations. The log likelihood function of received signals then is given as

$$L(H) = \log p(Y|H), \quad (4.5)$$

with the first derivative

$$\frac{\partial L}{\partial H} = \sum_{k=1}^T \sum_{i=1}^{|S|} p(s_i | y_k, H) \frac{\partial}{\partial H} \log p(y_k | s_i, H). \quad (4.6)$$

Then, the first part in the summation above can be calculated by

$$p(\mathbf{s}_i|\mathbf{y}_k, H) = \frac{p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i)}{p(\mathbf{y}_k, H)} = \frac{p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i)}{\sum_{j=1}^{|\mathcal{S}|} p(\mathbf{y}_k|\mathbf{s}_j, H)p(\mathbf{s}_j)}. \quad (4.7)$$

Then, the log likelihood function can be computed by $p(\mathbf{y}_k|\mathbf{s}_j, H)$ and its log derivative.

The EM algorithm maximizes this log likelihood function by the auxiliary function in the E-step[143]:

$$Q(H, H') = \sum_{|\mathcal{S}|^T} \log p(Y|S, H)p(S|Y, H'). \quad (4.8)$$

In a M-step, the EM iteratively computes

$$H^{i+1} = \arg \max_H Q(H, H^{i+1}). \quad (4.9)$$

It can be shown [143] that

$$Q(H, H') \geq Q(H', H'), \quad (4.10)$$

then, this indicates

$$L(H) \geq L(H'). \quad (4.11)$$

and the EM update produces a monotonic increase in the likelihood.

The EM algorithm has been applied to various applications, often with great success. However EM is very sensitive to the initial value and poor convergence properties have been reported in many papers. In this section, we investigate the convergence properties of the EM algorithm used for linear mixture models elaborated as defined by (4.1). We point out that different source domains give us very different convergence performances. Mathematically, we demonstrate in the discrete case, eg.digital modulation systems, the EM algorithm shows Newton-like convergence. The results are confirmed experimentally for the example of blind equalization in a digital communication system.

4.3 EM Algorithm for Noisy ICA

For a linear system,

$$Y = HS + N \quad (4.12)$$

When the sources, S , are continuous and distributed according to some prior $p(S)$. H is a linear matrix with full rank and the noise N is zero mean Gaussian, we get a Gaussian observation model as follows:

$$p(Y|S) = \frac{1}{(2\pi|\Sigma|)^N} \exp\{-Tr(Y - HS)^H \Sigma^{-1} (Y - HS)\} \quad (4.13)$$

We focus on estimating the parameter H . Using the EM algorithm with the source priors, we can write $p(Y)$ using the hidden variables S . The parameters are estimated as the optimal points of the log likelihood:

$$\frac{\partial \ln p(Y|H)}{\partial H^*} = 0 \quad (4.14)$$

and the solution to the M-step is:

$$H = \langle YS^H \rangle \langle SS^H \rangle^{-1} \quad (4.15)$$

where the angle bracket $\langle \cdot \rangle$ denotes average with respect to the source posterior $p(S|Y, H)$.

4.3.1 Convergence in Continuous Domains

In [145], the authors used a saddle point approximation to get the first and second moments of the sources with respect to the source posterior. The EM updating expressions for the mixing matrix H can be written as:

$$H^{i+1} = H^i + \Sigma \tilde{H}^i + O(\Sigma^2) \quad (4.16)$$

$$\tilde{H}^i = H^{-H} (H^H + \frac{1}{N} \frac{p'(S)}{p(S)} Y^H) H \quad (4.17)$$

Consider Eq.(4.16), to the zeroth order, the new estimated mixing matrix is identical to the previous one. The first order correction \tilde{H}^i is proportional to the gradient of the noiseless model's likelihood [106]. The second order is scaling with square of noise covariance, Σ^2 , which makes the change in the estimated noise extremely slow to converge when the noise level is small. We see that the update of the mixing matrix is 'frozen' in the low noise case. This conclusion is also mentioned in [146].

Even though many proposals for accelerating the EM algorithm exist, such as Aitken acceler-

ator [147], quasi-Newton approach [148] and conjugated gradient acceleration [149], they all require further analytical computations. In contrast, two new optimization alternatives were recently proposed which can accelerate EM in low noise. The first [150] is called adaptive over relaxed EM. This method appears to speed up the convergence but without further analytical computation and it introduces the increasing cost function. The second method is called UCMIF [151] and is a Quasi-Newton method with Broyden-Fletcher-Goldfarb-Shanno update to the inverse Hessian matrix and soft line search. However this needs extra work on calculation of the gradient.

4.3.2 Convergence in Discrete Domains

In the discrete domain, the convergence property of the EM algorithm is very different. It provides good convergence performance. Considering (5.1) again, we rewrite it in a vector form rather than matrix form,

$$\mathbf{y}_k = H\mathbf{s}_k + \mathbf{n} \quad (4.18)$$

where we express $\mathbf{y}_k = \mathbf{y}[k]$ and $\mathbf{s}_k = \mathbf{s}[k]$ for the sake of simplicity in the following derivations.

4.3.2.1 Gradient of Log likelihood

We estimate the mixing matrix H as the optimal points of the log likelihood:

$$\begin{aligned} & \frac{\partial \ln p(\mathbf{y}_k|H)}{\partial H^*} \\ &= \frac{\partial}{\partial H^*} \ln \sum_{i=1}^{|S|} p(\mathbf{y}_k, \mathbf{s}_i|H) \\ &= \frac{1}{p(\mathbf{y}_k|H)} \frac{\partial}{\partial H^*} \sum_{i=1}^{|S|} p(\mathbf{y}_k, \mathbf{s}_i|H) \\ &= \sum_{i=1}^{|S|} \frac{1}{p(\mathbf{y}_k|H)} \frac{\partial}{\partial H^*} p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i) \end{aligned} \quad (4.19)$$

where $|S|$ denotes the number of configurations in the source vector space, \mathbf{s}_k , e.g., for a 4×4 system with 16-QAM modulation, $|S| = 16^4$. \mathbf{s}_i is one instance of source vector at time index k , we omit time index k for simplicity. Since

$$\frac{1}{p(\mathbf{y}_k|H)} = \frac{p(\mathbf{s}_i|\mathbf{y}_k, H)}{p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i)}, \quad (4.20)$$

using the properties of the derivative of a logarithm, we get:

$$\frac{\partial \ln p(\mathbf{y}_k|H)}{\partial H^*} = \sum_{i=1}^{|\mathcal{S}|} p(\mathbf{s}_i|\mathbf{y}_k, H) \frac{\partial}{\partial H^*} [\ln p(\mathbf{y}_k|\mathbf{s}_i, H)].$$

Define log likelihood L

$$\begin{aligned} L &= \ln p(Y|H) \\ &= \ln p \prod_{k=1}^T p(\mathbf{y}_k|H) \\ &= \sum_{k=1}^T \ln p(\mathbf{y}_k|H). \end{aligned} \quad (4.21)$$

where T are the number of the observed signals, Take the first derivative of log likelihood, we get:

$$\begin{aligned} \nabla_{H^*} L &= \frac{\partial \ln p(Y|H)}{\partial H^*} \\ &= \sum_{k=1}^T \frac{\partial \ln p(\mathbf{y}_k|H)}{\partial H^*} \\ &= \sum_{k=1}^T \sum_{i=1}^{|\mathcal{S}|} p(\mathbf{s}_i|\mathbf{y}_k, H) \frac{\partial \ln p(\mathbf{y}_k|\mathbf{s}_i, H)}{\partial H^*} \end{aligned} \quad (4.22)$$

Given the system model in equation (2.41), $p(\mathbf{y}_k|\mathbf{s}_i, H)$ can be expressed as

$$p(\mathbf{y}_k|\mathbf{s}_i, H) = \frac{1}{\sqrt{2\pi}|\Sigma|} \exp\{-(\mathbf{y}_k - H\mathbf{s}_i)^H \Sigma^{-1} (\mathbf{y}_k - H\mathbf{s}_i)\} \quad (4.23)$$

After ignoring the effect of the constant $\sqrt{2\pi}$ since it is trivial in the following derivations, we then get,

$$\frac{\partial p(\mathbf{y}_k|\mathbf{s}_i, H)}{\partial H^*} = -\Sigma^{-1} (\mathbf{y}_k \mathbf{s}_i^H - H \mathbf{s}_i \mathbf{s}_i^H). \quad (4.24)$$

Substitute it into equation (4.22), we obtain:

$$\nabla_{H^*} L = \frac{\partial L}{\partial H^*} = \frac{\partial \ln p(Y|H)}{\partial H^*} = \Sigma^{-1} (\langle Y S^H \rangle - H \langle S S^H \rangle) \quad (4.25)$$

where $\langle \cdot \rangle$ denotes conditional sample average with respect to the source posterior $p(S|Y, H)$. Generally, the channel update is written as

$$H^{k+1} = H^k + \Delta H^k \quad (4.26)$$

In noisy ICA cases, from equation (4.15), the EM updating for channel H^k at k th iteration can be expressed

$$H^{k+1} = \langle Y S^H \rangle \langle S S^H \rangle^{-1}. \quad (4.27)$$

Then,

$$\begin{aligned} \Delta H^k &= H^{k+1} - H^k \\ &= \langle Y S^H \rangle \langle S S^H \rangle^{-1} - H^k \\ &= \Sigma \Sigma^{-1} (\langle Y S^H \rangle - H^k \langle S S^H \rangle) \langle S S^H \rangle^{-1}. \end{aligned} \quad (4.28)$$

$$(4.29)$$

The EM update is

$$H^{k+1} = H^k + \Sigma \Sigma^{-1} (\langle Y S^H \rangle - H^k \langle S S^H \rangle) \langle S S^H \rangle^{-1}. \quad (4.30)$$

Substitute Eq. 4.25 into Eq.4.30, the EM update for channel is written as

$$H^{k+1} = H^k + \Sigma \frac{\partial L}{\partial H^*} \langle S S^H \rangle^{-1}. \quad (4.31)$$

For clearly, we write this gradient form in vector expressions,

$$\text{vec}(H^{k+1}) = \text{vec}(H^k) + \text{vec}(\Sigma \frac{\partial L}{\partial H^*} \langle S S^H \rangle^{-1}), \quad (4.32)$$

where $\text{vec}[H]$ denotes the stacked columns of the matrix H . Since $\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$ [152], we get:

$$\text{vec}(H^{k+1}) = \text{vec}(H^k) + \langle S S^H \rangle^{-1} \otimes \Sigma \frac{\partial L}{\partial \text{vec}(H^*)}, \quad (4.33)$$

where \otimes denotes the Kronecker product [152].

In a concise form, the EM updating expressions for the mixing matrix H in gradient form is:

$$\text{vec}(H^{k+1}) = \text{vec}(H^k) + P(H^k) \frac{\partial L}{\partial \text{vec}(H^k)^*}, \quad (4.34)$$

where

$$P(H^k) = \langle S S^H \rangle^{-1} \otimes \Sigma. \quad (4.35)$$

4.3.2.2 Hessian matrix of Log likelihood

Let us consider the second statistical property of the log likelihood:

$$\frac{\partial^2 L}{\partial H^T \partial H^*} = \frac{\partial}{\partial H^T} \left(\frac{\partial L}{\partial H^*} \right)$$

Since

$$\frac{\partial \{f(H)g(H)\}}{\partial H} = g(H) \frac{\partial f(H)}{\partial H} + f(H) \frac{\partial g(H)}{\partial H}$$

from (4.19) and (4.21) we get: we get:

$$\begin{aligned} & \frac{\partial^2 L}{\partial H^T \partial H^*} \\ &= \sum_{k=1}^T \sum_{i=1}^{|S|} \frac{\partial}{\partial H^*} \ln p(\mathbf{y}_k | \mathbf{s}_i, H) \times \frac{\partial}{\partial H^T} p(\mathbf{s}_i | \mathbf{y}_k, H) \\ &+ \sum_{i=1}^{|S|} \sum_{k=1}^T p(\mathbf{s}_i | \mathbf{y}_k, H) \times \frac{\partial}{\partial H^T \partial H^*} \ln p(\mathbf{y}_k | \mathbf{s}_i, H) \end{aligned} \quad (4.36)$$

Consider the first term on the right hand side of (4.36). In the low noise case where H is sufficiently close to the local minimum, the posterior probability is dominated by its nearest point, therefore:

$$\frac{\partial}{\partial H^T} p(s = i | \mathbf{y}_i, H) \cong 0 \quad (4.37)$$

The mathematical derivation can be found in appendix C.

Consider the second term on right hand side of (4.36):

$$\sum_{k=1}^T \sum_{i=1}^{|S|} p(\mathbf{s}_i | \mathbf{y}_k, H) \frac{\partial}{\partial H^T \partial H^*} \ln p(\mathbf{y}_k | \mathbf{s}_i, H) \quad (4.38)$$

We denote

$$\begin{aligned} f(H) &= \ln \{p(\mathbf{y}_k | \mathbf{s}_i, H)\} \\ &= \ln \left(\frac{1}{2\pi|\Sigma|} \right) + [-(\mathbf{y}_k - H\mathbf{s}_i)^H \Sigma^{-1} (\mathbf{y}_k - H\mathbf{s}_i)] \end{aligned}$$

In trace form:

$$f(H) = \text{tr}\{c + [-(\mathbf{y}_k - H\mathbf{s}_i)^H \Sigma^{-1}(\mathbf{y}_k - H\mathbf{s}_i)]\}$$

Since $f(H)$ is twice differentiable, we take the second derivative with respect to H^* and H , respectively:

$$d^2 f(H) = d_H[d_H f(H)] = \text{tr}[\mathbf{s}_i \mathbf{s}_i^H d(H)^T \Sigma^{-1} dH], \quad (4.39)$$

where $d^2(\bullet)$ is the second differential operation. From [153], we know:

$$\begin{aligned} d^2 f(H) &= \text{tr}(B(dH)^T C dH) \iff \\ \mathcal{H}(f(H)) &= \frac{1}{2}(B^T \otimes C + B \otimes C^T) \end{aligned} \quad (4.40)$$

where \mathcal{H} is the Hessian operator. Thus,

$$\frac{\partial}{\partial H^T \partial H^*} \ln p(\mathbf{y}_k | \mathbf{s}_i, H) = (\mathbf{s}_i \mathbf{s}_i^H) \otimes \Sigma^{-1}$$

Substitute this equation into equation (4.38), we get

$$\mathcal{H}(H) = \frac{\partial^2 L}{\partial H^T \partial H^*} = \langle S S^H \rangle \otimes \Sigma^{-1} \quad (4.41)$$

According to $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, we can get:

$$\mathcal{H}(H^k)^{-1} = \langle S S^H \rangle^{-1} \otimes \Sigma. \quad (4.42)$$

Recall the EM update for channel matrix H^k in the noisy ICA, equations (4.34) and (4.35), we find that,

$$\mathcal{H}(H^k)^{-1} = P(H^k), \quad (4.43)$$

In discrete case, it is worthwhile to emphasize that the EM update for channel matrix in noisy ICA cases can be written as a Newton-Raphson method,

$$\text{vec}(H^{k+1}) = \text{vec}(H^k) + \mathcal{H}(H^k)^{-1} \frac{\partial L}{\partial \text{vec}(H^k)^*}. \quad (4.44)$$

This phenomenon can be explained as follows: In discrete cases, the log likelihood $L p(Y|S, H)$ is not continuous anymore when the mixing matrix H approximate the true value and noise is low. Then $\frac{\partial L}{\partial \text{vec}(H^k)}$ is very large and is determined by the inverse of noise covariance, Σ^{-1} , e.g. $\frac{\partial L}{\partial \text{vec}(H^k)}$ goes to infinity as the noise covariance approaches zero. This factor happens to compensate the freezing factor Σ which is the step size in the gradient form of the EM update.

Note that we do not use simplified summation operations for the EM algorithm. We stress again, in the discrete domain, the EM iteration approximates the Newton-Raphson method in low noise. This implies us that the EM algorithm guarantees fast convergence. The result is consistent with [154], which states that, when the missing information is small compared to the complete information, EM exhibits approximate Newton behaviour and enjoys fast, typically superlinear convergence in the neighbourhood of the local minima and suggests the use of the EM algorithm as a refinement of the previous ICA results for digital signal estimations.

4.3.3 Limitations of EM

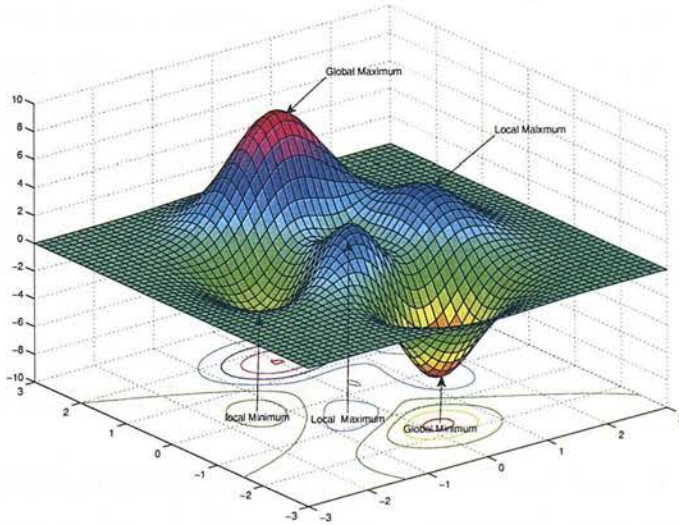


Figure 4.2: EM likelihood surface.

Although the EM algorithm has gained popularity in mixture analysis, it has limitations:

Limitation 1: The EM algorithm can be sensitive to its initial values.

Limitation 2: The computation cost is exponential in the number of sources.

The limitation 1 can be explained by the fact that the EM likelihood function can be multimodel and the EM is a local optimization method only [155]. Figure 4.2 show a sample of EM likelihood surface. Many local minimum and saddle points make the global convergence difficult. In wireless communications, mobile radio channels are complicated. This causes many local minimum in the likelihood surface, and EM is more liable to converges to a local optimum of the likelihood function. We will clarify this difficulty in detail later.

The limitation 2 is the curse of dimensionality which is a well known problem with the EM algorithm. The EM algorithm has computational cost that grows exponentially with the number of sources. Because the EM algorithm make uses of the full and whole hidden sources to build a functional whose maximization yields the ML estimate. When the number of sources increase, both the E-step and the M-step are difficult to solve. In digital MIMO systems, since the M-step is obtained from formula (4.15), the problem of high dimensions is due to the E-step, especially for large constellations. For example, in the case of 4 transmitters and 16-QAM signals, there are $16^4 = 65536$ different configurations for the source symbols at each sampling. Then, each iteration of the E-step requires the computation of $65536 \times T$ conditional probabilities. Where, T is the frame length of each transmitted block. Hence, for high-rate systems with large number of antennae, direct calculation proves to be infeasible. One solution is to use a stochastic algorithm and replace the summations over all the possible hidden source states by Monte Carlo integrations [156]. However, this is still not efficient enough for real time operations. We consider a deterministic approximation which we describe next.

4.4 Sphere Decoder for MIMO Detection

As illustrated in Figure 4.1, a optimum detector for a large constellation MIMO system is the ML detector. However, this sort of detector exhibits exponential complexity in the number of received antennae and this makes it impractical for real time communication systems. In order to overcome this difficulty, we propose the use of a soft sphere decoder. The principal spirit of the sphere decoder is to reduce the computational load of the maximum likelihood detector by only searching over the received signals that lie within a hypersphere of radius R around the received signal (Figure 4.3). This is reasonable since most of the constellation points located outside the sphere have vanishingly small probabilities since its probably is related to minus

exponentiation of distance. When the distance is relatively large, the contribution of points outside the sphere are negligible. The basic premise in the proposed algorithm is rather simple: In the case of digital modulations, it is reasonable to search over only the most probable hidden source points $s \in A_q^m$ that lie in a certain sphere of radius R around the received vector, where A_q^m denotes the n -dimensional hidden source points spanned by a q -QAM constellation in each dimension. Thereby reducing the search space and hence the required computational effort (see Figure 4.3). Clearly, the closest point inside the sphere, assuming there is one, will also be the closest point for the whole hidden source space. So this complexity reducing searching is rational.

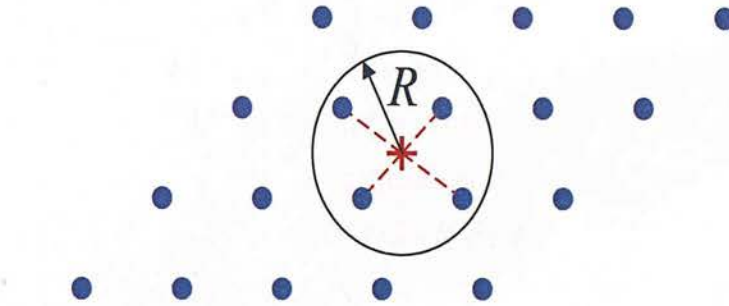


Figure 4.3: Idea of choosing the admissible set.

The summation over the points inside the circle. Figure 4.3 takes the place of the summation over the points lying in the entire source space. The points inside the sphere are good in a likelihood sense and their collection builds up a set called as the *admissible set* and its number is called the *size of the admissible set*. Closer scrutiny of this idea leads to two questions.

How should we choose the size of the admissible set? Obviously, it connects with the radius R directly. If R is too large, we may obtain too many points and the algorithm will remain exponential in size, whereas if the radius R is too small, we may not obtain enough points inside the sphere to accurately approximate the expectation sum. Choice of the radius R plays an important role in the final performance and the reduction of the complexity. Informally, when SNR is large, the size of the admissible set is selected to be small. Otherwise, the size of the admissible need to be large with increasing the sphere radius to guarantee that at least one point in the hypersphere. To assure the lattice point Hs lies inside the hypersphere, radius R is required to $R \geq \|y - Hs\|$, which is directly depends on noise covariance or SNR.

How can we determine which points are inside the sphere? Sphere decoding (SD) [157] pro-

vides a constructive answer this question. Figure 4.3 shows the basic aim of the SD algorithm. The dot points represent the noiseless constellation Hs and the cross stands for the actual received signal contaminated with noise. SD calculates the closest point by:

$$\hat{s}_{SD} = \arg \min_{s \in A_q^M} \|y - Hs\|^2 \leq R^2. \quad (4.45)$$

The main SD principals are revised here briefly for the reader. The SD optimization takes the place of ML solution and the source $s \in A_q^m$ is limited in the constellation. Define the unconstrained least square estimate, $\hat{s} = H^{-1}y$, via a mathematical matrix decomposition, equation (4.45) can be given as

$$\begin{aligned} & \|y - Hs\|^2 \\ &= \|H(\hat{s} - s)\|^2 \\ &= \|QU(\hat{s} - s)\|^2 \\ &= \|U(s - \hat{s})\|^2 \leq R^2, \end{aligned} \quad (4.46)$$

where Q is an orthonormal matrix and U is an $N \times N$ upper triangular matrix, $H = QU$ is the QR decomposition of H . Also U can be obtained by Cholesky decomposition of a Gram matrix $G = H^H H$. To solve the equation (4.46), we can use a depth-first constrained tree search over M levels. The search is recursive from top layer, $i = M$, down to the bottom layer, $i = 1$. Each layer in the search represents signal from one transmit antenna. Each node in a layer, except for the bottom layer, $i = 1$, has q child nodes lying in the next layer, and has a parent node on the tree, except for nodes on the top layer, $i = M$. The SD searches from the top layer of the tree, along the branch in which the child nodes satisfy

$$|s_i - z_i|^2 \leq \frac{T_i}{u_{ii}^2}, \quad (4.47)$$

where u_{ij}^2 is the element of decomposition matrix U and z_i

$$z_i = \hat{s}_i - \sum_{j=i+1}^M \frac{u_{ij}}{u_{ii}} (s_j \hat{s}_j), \quad (4.48)$$

and

$$T_i = R^2 - \sum_{j=i+1}^M u_{jj}^2 |s_j - z_j|^2, \quad (4.49)$$

with $T_M = R^2$.

At any level i , when $T_i \leq 0$, no nodes satisfy equation (4.46), which means this branch exceeds the radius constraint, the SD remove that part of the tree from further considerations. In comparison with exhaustive ML search, such operation reduce the complexity. At that point, the SD trace back to its parent node for further continuous search. The last path found corresponds the SD solution. Figure 4.4 illustrates a simplified diagram of a tree search of a 4×4 system with a 16-QAM modulation [4]. The dashed line is the paths discarded due to the constraint, equation (4.47).

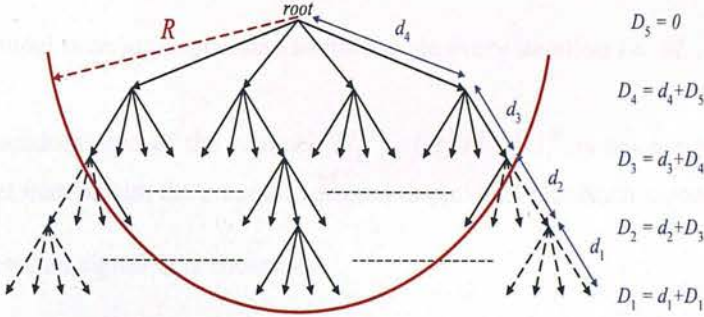


Figure 4.4: Illustration of tree search of SD in a 4×4 system with 4-QAM modulation, from [4].

4.4.1 List Fixed Sphere Decoder

Since the introduction of the SD for solving the ML question, many studies have been proposed to reduce its complexity. The most famous contribution is the Schnorr-Euchner (SE) [158] enumeration, in which the search visits the nodes in increasing distance to z_i , in equation 4.48. This kind of search can make the probability of finding the ML solution larger in the first layers search. The SE and its variants have attracted a lot of attention since they are widely considered as the most effective way of reducing SD complexity [159]. Another complexity reduction had been proposed by modifying the characteristics of the channel matrix [160] [159]. This lattice reduction approach can cut down the average computational load of the search stage. On other hand, some methods utilize channel statistics to improve the search stage [161].

Therefore, they need more calculation for limiting the threshold. All methods above suffer from requiring variable complexity depending on the noise level and the channel conditions and this has obstructed the integration into real time wireless communication systems.

Another approach of fixed complexity is the k -best lattice decoder [162]. It fixes the number of visiting nodes at each layer to guarantee the fixed complexity, but it neglects the statistical distribution of the visiting nodes. Furthermore, the computational burden is always higher than the computational burden of the SD.

To overcome the main drawback stated above, we resort to a fixed complexity sphere decoder (FSD) [144]. Figure 4.5 illustrates the idea of the fixed tree search following predetermined paths down the tree. This approach combines a novel channel matrix ordering technique and a search through a small subset of the received constellation.

The FSD channel ordering is operated as follows, in every iteration $i = M, \dots, 1$,

1. The pseudoinverse of the channel, $H_i^\dagger = (H_i^H H_i)^{-1} H_i^H$ is calculated, where H_i is the channel matrix with the columns selected in previous iterations zeroed.
2. The detected signal \hat{s}_k is chosen as

$$k = \begin{cases} \arg \max_j \|(H_i^\dagger)_j\|^2, & \text{if } n_i = P \\ \arg \min_j \|(H_i^\dagger)_j\|^2, & \text{if } n_i \neq P \end{cases} \quad (4.50)$$

where P is the number of points in QAM constellations and n_i is the number of visiting points at i th level, and $(H_i^\dagger)_j$ is the j th row of (H_i^\dagger) .

The purpose of the channel matrix ordering is to make sure that more points are visited in the first level while the number is reduced in the last levels. The channel matrix ordering proposed is such that makes it more probable to have a detection error in the levels where more points are searched. Generally known, the estimate of signals is more robust when more points are visited in a level. For the level where noise amplification is large, this scheme assigns more points at that level. Such assignment can guarantee the ML performance of the SD.

The other innovation is that the FSD exploits the distribution of the nodes. This distribution determines the final performance and the complexity reduction. Even though the distribution of nodes cannot be analyzed mathematically for any number of antennae and constellation sizes.

In [144], the author conjectured that for an uncoded spatially multiplexed $M \times N$ system with P constellation points per transmit antenna, there always exists a distribution of nodes ϵ_s in the form

$$\epsilon_s = (\underbrace{1, \dots, 1}_{l_1}, \underbrace{P, \dots, P}_{l_p})^T. \quad (4.51)$$

where l_1 is the number of levels where only one node is considered and l_p is the number of levels where the maximum number of nodes are considered. Such an assignment allows the FSD detector to approximate to ML performance with a fixed complexity [144]. At the bottom layer, P^{l_p} point remain and build up a list of candidate. This method is called list fixed-complexity sphere decoder(LFSD).

Figure 4.5 below shows a system in 4 transmit antenna, 4 receive antenna with 4-QAM constellations used at each transmit antenna. The number of points considered per level is $(\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4) = (1, 1, 2, 3)$.

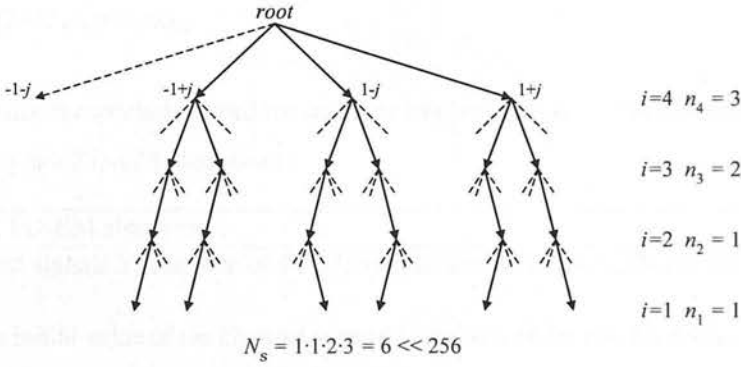


Figure 4.5: Tree Structure for the List Fixed Sphere Decoder.

BER performance was shown in Figure 4.6. A 4×4 system with 4-QAM and 16-QAM modulations. The node distributions are $(\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4) = (1, 1, 1, 4)$ and $(1, 1, 1, 16)$, respectively. Thus, there produce 4 and 16 candidates, a selection is on the basis of minimum distance used in this simulation.

4.5 SD EM algorithm

As stated above, the LFSD is integrated into our blind system. By searching a very small subset of the complete transmit constellation and a specific channel matrix ordering, the LFSD obtains a list of candidates that can be used to calculate symbol likelihood information. This

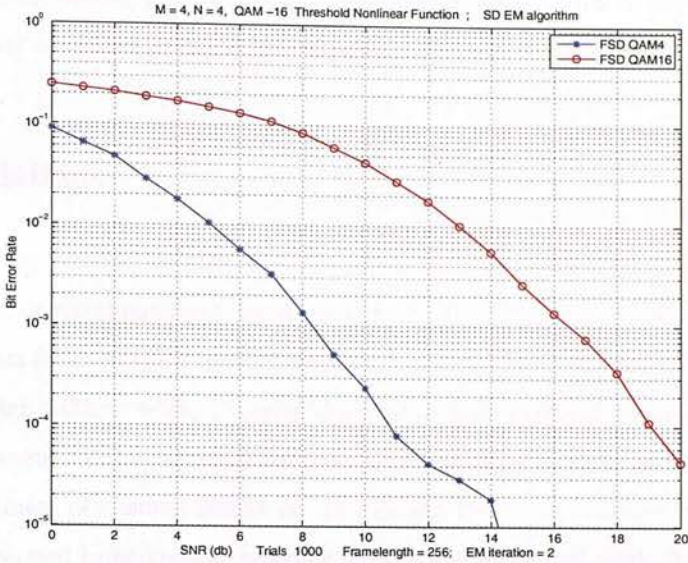


Figure 4.6: BER performance of fixed sphere decoder in a 4×4 system with 4-QAM and 16-QAM modulation.

fixed complexity is especially suited for real time implementation. With the introduction above, we can specify our SD-EM algorithm:

Algorithm 1 SD-EM algorithm

Input: received signals Y ; The size of the admissible set; Iteration number of the EM algorithm.

1. Get the initial value of the channel state information, \hat{H} , by our *Threshold Nonlinear ICA* algorithm or any standard ICA.
2. Using LFSD approach with \hat{H} to construct the admissible set A .
3. Updating the \hat{H} by the EM algorithm. The posterior is approximated by the points within the admissible set A .
4. Go back to step 3 with new \hat{H} until the iteration ends.

Output: Estimate signals \hat{S} ; \hat{H} .

Unlike the Monte Carlo EM (MCEM) algorithm, which approximates the conditional expectation in E-step by the Monte Carlo average, the essential idea behind this algorithm is that the SD-EM approximates the conditional expectation by the important samples which lie around

the received signal vector. This necessarily optimizes a lower bound to the likelihood at each iteration. Similar ideas can be found in [163] and [119].

4.6 Simulations

In this section we provide simulations to illustrate the convergence of the EM algorithm for 4-QAM signals. In the simulations, the mixing matrix H is 2×2 complex instantaneous matrix, which is constant for each block interval, and it follows a complex Gaussian distribution. N follows the complex additive white Gaussian distribution with diagonal covariance matrix. Here, the separation matrix W is initialized with the Threshold ICA method, and we take its inverse as a rough estimate of channel matrix H . In this and following chapters, the noise standard deviation is assumed to be known; however, with some additional work, it may be estimated using EM [164].

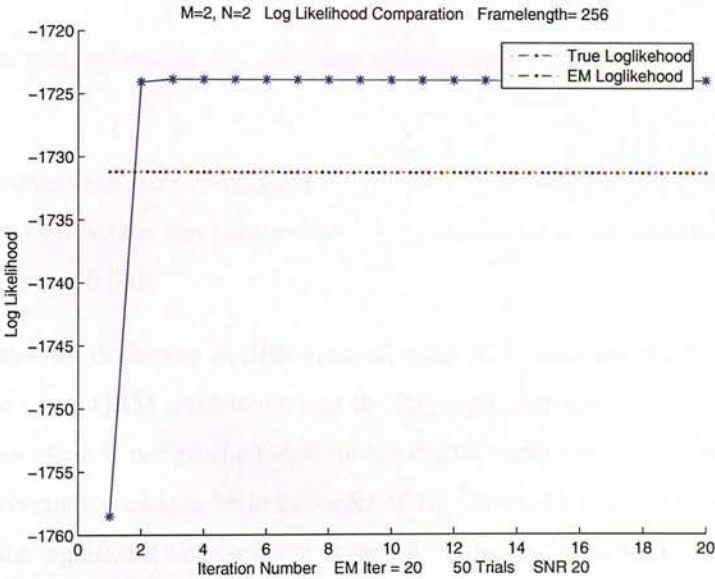


Figure 4.7: EM Log likelihood VS. True likelihood.

Figure 4.7 shows us that the log likelihood converges to the global minimum after 5 iterations of EM updating when SNR is 20 dB. For 50 experiments, the dotted line is the average true log likelihood curve with exact channel state information and the star line is the average EM log likelihood with the estimated channels. It exceeds the true value after only 2 iterations which shows us that the log likelihood converges to the local maximum after small iterations of EM

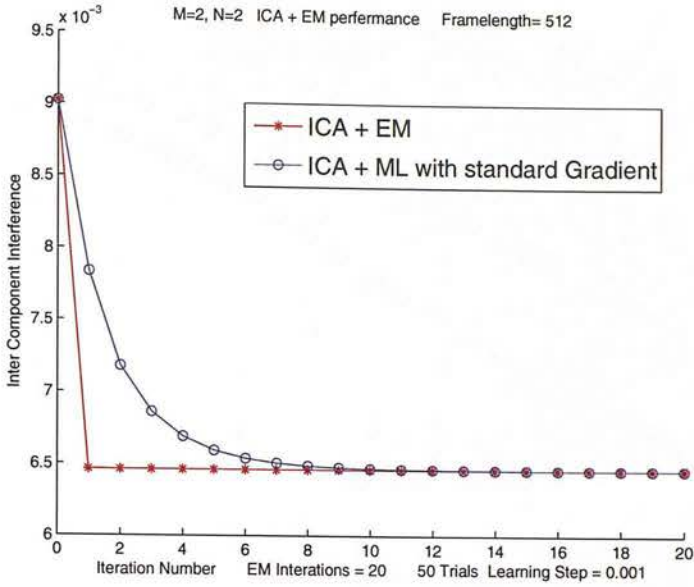


Figure 4.8: Likelihood of EM iterations and standard Gradient method.

updating. The margin between the true value and estimation is caused by the estimated bias of channels.

Figure 4.8 displays the inter component interference level, defined by (3.44). For an SNR of 20dB, we can observe the fast convergence speed compared to the standard gradient method with the step size = 0.001.

Figure 4.9 show the difference in BER between using ICA alone and the ICA+EM algorithm. Both of them use 4-QAM modulation and the Rayleigh channels. It indicates that using the ICA algorithm alone is not good enough for the digital communication system since, usually, the BER performance needs to be in the order of 10^{-2} when SNR is in the range of 10~20dB. We can see the significant improvement achieved by the EM algorithm. Here, we ignore the intrinsic permutation phase ambiguity problem in this simulation. The correct channel order by comparing the decoding sequence of EM with the true transmitted sequence. We choose the most likely order according to the minimum BER criteria. The phase ambiguity can be solved by further using a differential amplitude phase encoded (DAPE) QAM [165].

We next examine the SD-EM algorithm. The simulation setup is the same as the first simulation: 4x4 system, 16-QAM. 512 symbols in each block with slow Rayleigh fading channel and 1000 trials were run. First, we use the threshold nonlinear ICA to obtain rough estimations of

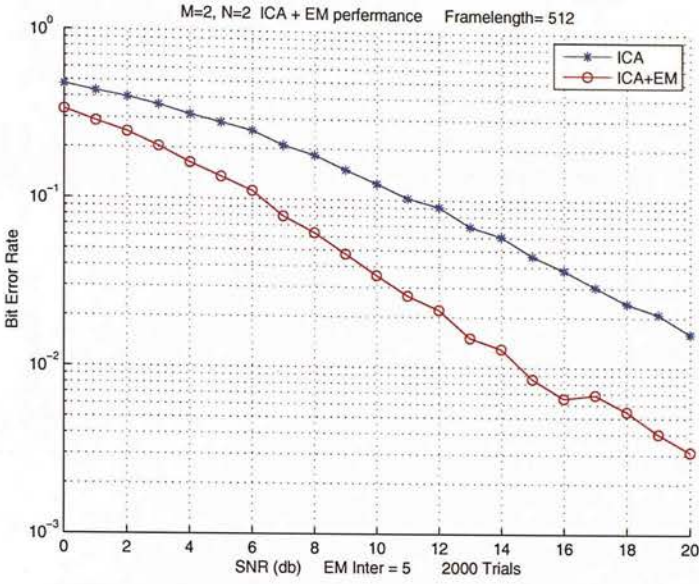


Figure 4.9: BER Performance of ICA + EM method in a 2x2 system with 4-QAM modulation.

channel, using this as initial values, we apply the SD-EM algorithm to further improve estimations. Here only 2 iterations of the EM algorithm are used. To maintain a low computational load, the size of the admissible set D is fixed at 16 which is much smaller than the entire 65536 configuration. The number of points considered in each level is (1, 1, 1, 16). Figure 4.10 shows substantial improvements by the SD-EM algorithm in terms of Inter-Component Interference (ICI). As we know, the performance can be improved by increasing the size of the admissible set. It is a tradeoff between the performance and complexity.

In digital communication transmission, the final performance is BER. In Figure 4.11, we compare our method with zero forcing detector with exact channel state information and the *threshold nonlinear* ICA using the ZF decoder. Also, we present the bound of the SD-EM method in which we apply the exact CSI to the SD algorithm. Using the same system configuration of the last simulation, we observe that the *threshold nonlinear* can reach similar performance of the ZF detector and the SD-EM algorithm improve it significantly over all SNR.

For more large QAM constellations, Figure 4.12 provides the BER performance of the SD-EM for 64QAM modulation. But the size of the admissible set D is fixed at 64 which is much smaller than the entire $64^4 = 16777216$ configuration. The number of points considered in each level is (1, 1, 1, 64). Other setups are the same as the previous 16-QAM system.

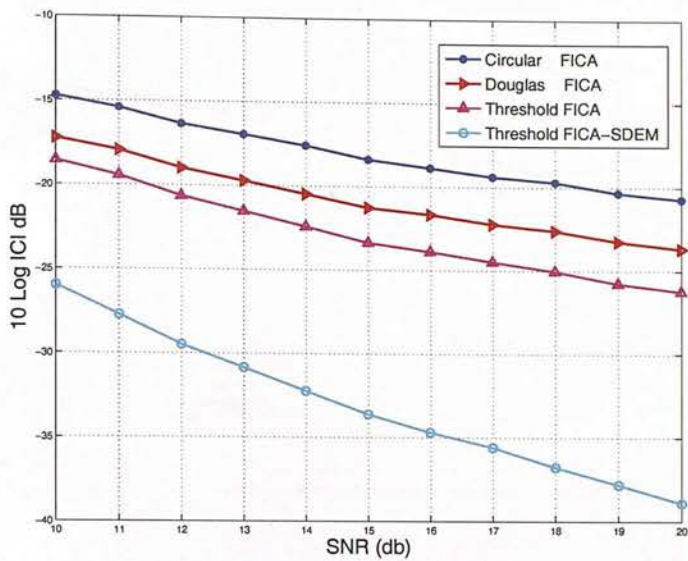


Figure 4.10: Separation quality of different ICA algorithms.

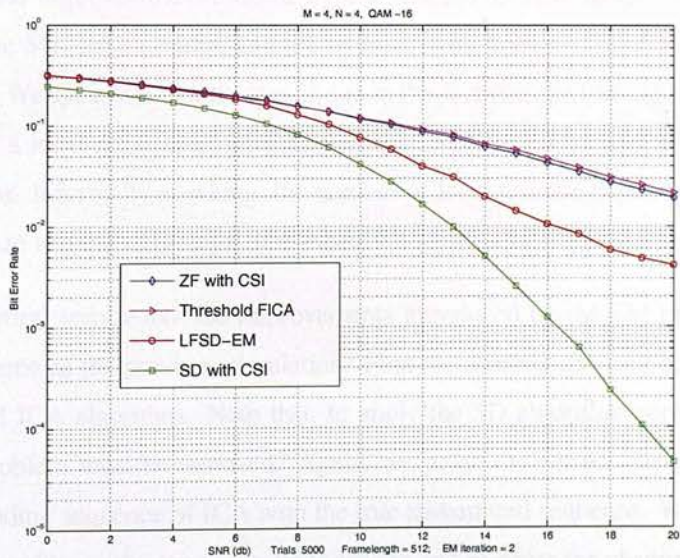


Figure 4.11: BER performance of the ZF algorithm, Nonlinear Threshold ICA algorithm, the SD-EM algorithm and SD with perfect CSI in a 4×4 MIMO system with 16-QAM modulation.

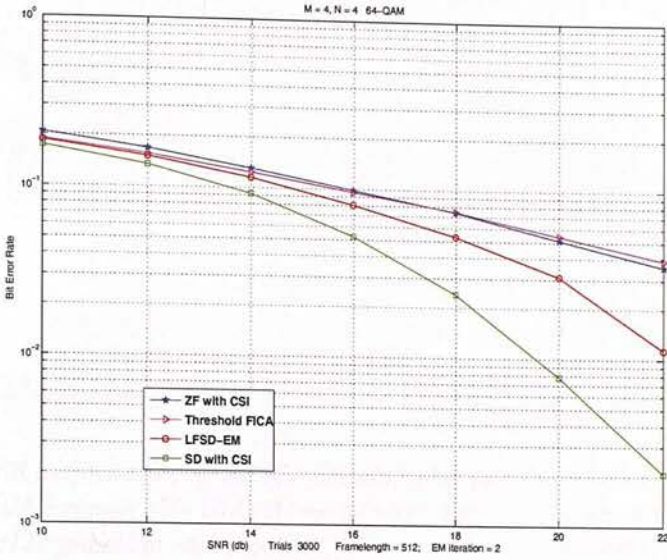


Figure 4.12: BER performance of the ZF algorithm, Nonlinear Threshold ICA algorithm, the SD-EM algorithm and SD with perfect CSI in a 4x4 MIMO system with 64-QAM modulation.

Although the proposed SD-EM algorithm makes the channel estimate better than other blind methods and thus improves the estimated signals in terms of BER, there is still a performance gap between the SD-EM algorithm and the optimal (with known CSI) SD solution, as shown in Figure 4.11. We speculate that the gap is due to the fact that the EM algorithm occasionally gets trapped in a local rather than global minimum, and that this occurs when the channel is close to singular. Informally speaking, the number of local minima depends on the number of mixtures, the size and the dimension of the data.

The last simulation emphasizes the improvements introduced by the EM update. The system setting is the same as the previous simulation. First the channel estimate blindly via the non-linear threshold ICA algorithm. Note that, to apply the SD algorithm correctly, the intrinsic permutation problem must be removed. Again, we judge the correct channel order by comparing the decoding sequence of ICA with the true transmitted sequence. We choose the most likely order according to the minimum BER criteria and re-order the channel row correspondingly. Then, we put it into the SD algorithm and the SD-EM respectively. The EM channel update can contribute to channel estimation with iteration going on and then improve the BER performance, as Figure 4.13 illustrated.

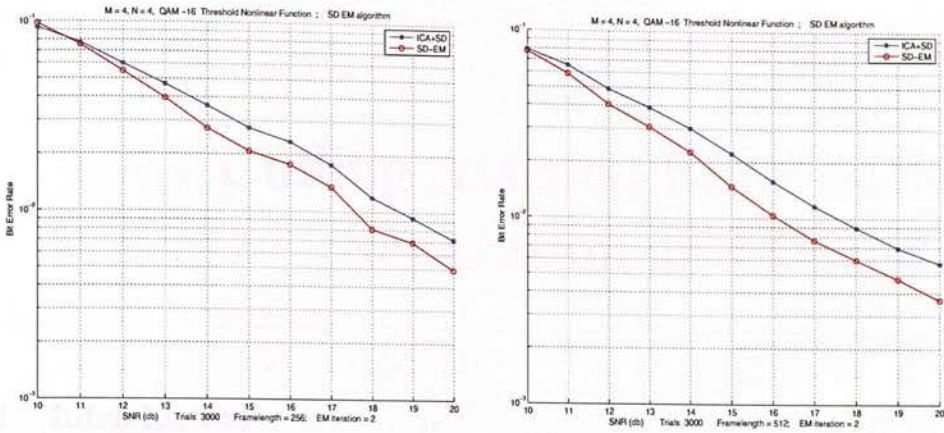


Figure 4.13: BER performance of the ICA-SD algorithm and the SD-EM algorithm in a 4x4 MIMO system with 16-QAM modulation. Left : 256 symbols in each block; right : 512 symbols in each block.

4.7 Chapter Summary

In this chapter, we first have presented the general mechanics of the EM principle and shown the convergence properties of EM exhibits dramatically different appearance. In digital communication cases, we point out that the EM updating algorithm has a fast convergence behaviour in the low noise limit. Moreover, to keep computational costs fixed, the E-step can be approximated by a fixed complexity sphere detector. Based on these findings, an efficient hybrid system for blind equalization of large constellation MIMO systems was presented. The proposed approaches for the signal separation and the signal detection reduce the computational complexity while maintaining the acceptable performance. Such an efficient combination makes this potentially feasible for a real time communication system. The numerical simulations demonstrate the effectiveness of this combined technique. Another benefit of such a hybrid system is that it utilizes soft decisions on the decoded symbols which could be further improved in an iterative error correcting decoder. This will be explored in the next chapter.

Chapter 5

Coding Assisted Blind MIMO Separation and Decoding

5.1 Introduction

Despite the widespread use of forward-error coding (FEC), most MIMO blind channel estimation techniques ignore its presence, and instead make the simplifying assumption that the transmitted sequences are uncoded and i.i.d. However, FEC induces code structure in the transmitted sequence that can be exploited to improve blind MIMO channel estimates. In this chapter, we exploit the iterative channel estimation and decoding performance for blind MIMO equalization. Experiments show the improvements achievable by exploiting the existence of coding structures and that it can approach the performance of a BCJR equalizer with perfect channel information in a reasonable SNR range.

All methods and algorithms referred and introduced in the previous chapters are based on estimates of the channel by a-priori soft information which is directly coming from the symbol detection techniques, even though some iterative procedures were introduced. However, if the channel is impaired by the deep fading, obtaining a precise estimation can be problematic. Consequently, the following symbol decisions are not so accurate and reliable. The natural question is how to improve the reliability of the channel estimation and symbol decoding in such a case.

In the last decade, joint channels estimation and symbol detection techniques have been proposed for the purpose of solving ISI problem in the SISO system. Channel codes, were originally designed for reliable transmission of digital communication in the noise environment. In this process, an information bearing sequence of length K , called the message, is encoded into a sequence of length $N > K$, called the codeword. Then, this codeword is modulated and transmitted through the communications channel. Such forward error coding, which restricts the transmitted sequence to a limited coding space so as to increase the minimum distance, can correct potentially wrong decoding due to noise contamination. Using the FEC code, the decoder at receiver can feedback a-posteriori information to the equalizer. The practical chal-

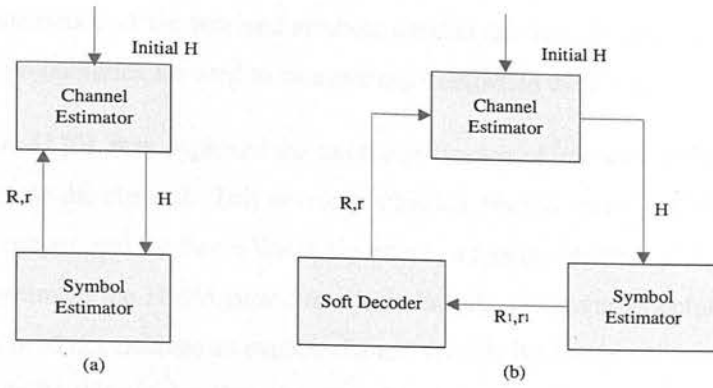


Figure 5.1: Joint channel estimation and symbol detection: Uncoded diagram vs. FEC diagram.

lenge, nonetheless, is the tremendous complexity demanded by this joint optimal ML decoding. To solve this problem, the iterative soft decoder has been studied and has been found to approach the optimal ML decoding performance at a practical and reasonable computational burden [166]. Furthermore, with powerful digital processors in the last decade, contributions of FEC to decoding with affordable complexity were explored in [167] [168]. Looking from a broader angle, we can take blind equalization as part of the decoding process. Thus, we can try to find a blind equalization and channel correcting scheme that together approximate the Shannon bound. Such methods combine the blind iterative channel estimation and turbo equalization. As illustrated in the block diagram of Figure 5.1(a), the equalizer uses the channel estimates to compute soft information of the transmitted symbols. The channel estimator then applies these soft symbols to improve the channel estimates, which in turn yields better symbol estimates, and so forth. In contrast, the FEC aware channel estimator based on soft symbol, a-priori information, feeds this information into the decoder in order to get more reliable soft bit information. Next, this posterior information is fed back to the channel estimator, and so on, as illustrated in the block diagram of Figure 5.1(b). Such a scheme utilizes FEC information in blind equalization. Nevertheless, there has been little work relating FEC to MIMO channel estimations. Although the independent, identically distributed (i.i.d.) assumption usually made in MIMO blind separation [51][52] no longer holds (due to the FEC coding), it has been shown that FEC does not impair the performance of some blind equalization techniques [169].

Some previous research has explored the FEC property on blind channel estimation. In [168] combined blind channel identification and turbo equalization, they used the EM algorithm to update channel state information. The covariance matrix was computed as a sample average

in which the likelihood of the received symbols weights the data. In contrast, in this chapter, pairwise joint probabilities are used to measure expectations in the E-step.

In another work [170], they exploited the turbo equalization of unknown ISI channels using a trellis to represent the channel. This develops a hidden Markov model (HMM) [171] for the noisy channel output, and the Baum Welch algorithm, a specific instance of the EM algorithm, is applied to estimate the HMM parameters including the observations before adding noise. This approach need not estimate an explicit channel directly but can be calculated using the estimated symbols. In this chapter, the soft-output BCJR equalizer depends on a channel estimate that is obtained from the EM algorithm iteratively.

In [172] [173] a blind iterative channel estimator is used that is also based on an EM algorithm. They applied a turbo equalizer loop with a decision feedback equalizer. Such schemes enjoy a low complexity. In the development of the EM channel estimator, they used a sample average to replace the ensemble average. In our work, the marginal and joint probabilities of each symbol are used to evaluate ensemble averages and this computation generates more accurate information.

In [174] [175], the author presented a generalized BCJR and LDPC algorithm to compute joint posteriori probabilities of symbols given noisy observations of those symbols to suppress at the output of an intersymbol interference channel. The pair-wise joint posterior probabilities are applied into the EM channel estimator. Both schemes are suitable for a single channel with a small number of channel taps since, in such a case, the surface of likelihood is simple and smooth enough to allow the EM algorithm will converge to a desirable point. Otherwise, the LDPC or the so call generalized BCJR may not converge to the correct state. In [175], for a single channel, 30 EM iterations were used to evaluate the system performance. This number of operations introduces a considerable computational burden.

Per-Survivor Processing (PSP) [176] is a seminal work in joint channel estimation and symbol detection. The PSP method embeds the data-aided channel estimation into the Viterbi algorithm. Each state has a separate channel estimate which is based on the survivor path leading to that state. PSP performs an ML estimation of the channel parameters. Then it estimated the candidate channel by applying a LMS algorithm or a Kalman filter to each survivor path. This kind of method was subsequently developed as an approximate minimum variance channel estimator. A similar but simplified approach was proposed by [177], where the author maximizes

the data log likelihood by weighting a quadratic function associated with each survivor path by the path probability. The authors then used an EM algorithm to update the channel estimate iteratively. To avoid initialization problem, they enumerated many initializations for the EM and select the most probable one in terms of the likelihood. Without an efficient proposal for initialization of the EM, this scheme is not feasible for high dimensional systems.

The methods described above were employed in single input single output channels and small constellations and these techniques simply do not extend to multi-antenna systems and large constellation QAM systems. In these cases, the increased dimensionality of the MIMO channel can make the convergence problematic.

In the last chapter, an efficient hybrid system for blind equalization of large constellation MIMO systems was proposed [178], in which the SD algorithm is integrated into the EM algorithm for the large multi-dimensional channel estimate. The initialization of the EM is provided by a fast and simple nonlinear ICA method which is specifically designed for QAM modulation. Such an efficient combination makes this feasible for a real time communication system. The numerical simulations demonstrate the effectiveness of this combined technique. However, it still suffered a loss of performance when the channel was close to singular or the noise was large.

In this chapter, a blind MIMO receiver that combines a soft channel estimation and a soft MIMO equalizer and decoder is proposed. A similar idea appears in [179], where the authors used a pilot sequence and a Wiener filter to initialize and update the channel respectively. This Minimum Mean Square Error based iterative channel estimator uses soft information from the output of the decoder to improve the mean square error of the channel estimates. However, taking the mean values of the data symbols calculated by the posteriori probabilities, is not an accurate way to improve the channel estimates. In our hybrid design, we improve the receiver's performance through efficiently incorporating the soft bit information from the decoder into the EM channel estimator. This enable us to exploit the a-priori knowledge imposed by the code. The system includes an efficient independent component analysis method which we mentioned in chapter 3, a selective sphere decoder process that computes the likelihood values (soft information) as stated in chapter 4 and a simple error correcting operation. This scheme has low complexity and improved convergence, being more likely to converge to the desirable stationary point. Moreover, through sending the bit interleaved coded modulation (BICM) bits on differently fading channels, we further make use of the temporal diversity. Combining this with the spatial diversity due to the statistical independence of transmitted sequences, blind

estimate and adjustment of the channel matrix can be performed simultaneously. Empirically this provides us with outstanding performance of MIMO blind equalization and decoding. This idea can also be easily extended to an MIMO-OFDM system, especially for the fast fading channel acquisition and tracking.

Again, we consider the $N_R \times N_R$ MIMO narrow band system model,

$$Y = HS + N, \quad (5.1)$$

where $Y \in \mathbb{C}^{N_R \times T}$ is the matrix containing observed signals from the receiver antennae, and $S \in \mathbb{C}^{N_R \times T}$ is the complex discrete source signal matrix. $N \in \mathbb{C}^{N_R \times T}$ is the noise matrix with covariance, Σ , which is assumed to be uncorrelated with the source signals and T is the sampled points of observations. $H \in \mathbb{C}^{N_R \times N_R}$, the Rayleigh channel, is the unknown linear square channel matrix whose elements are assumed to be drawn independently from a complex Gaussian distribution and we assume that it is invertible. Note that, H is instantaneous but we do not guarantee it is orthogonal. For the transmission of a frame of K_b bits, the transmitter encodes the K_u information bits using a convolutional code of rate r , where $K_u = K_b \times r$. The coded bits are interleaved and mapped into QAM symbols, forming a sequence of $K_s = K_b / \log_2 P$ symbols, where P is the number of possible symbols in the QAM constellation. Then the QAM sequence of symbols is split into N_R substreams corresponding to one Rayleigh fading channel, and is transmitted in parallel from each one of the N_R antennae. The problem above arises not only in MIMO systems, but also in multiuser DS/CDMA systems [180]. It further reduces to SIMO blind equalization when there is only one source signal or when fractionally spaced equalization is employed in single antenna communication systems.

5.2 The Proposed Iterative Procedure

In this work, a blind MIMO channel equalization algorithm is designed in which the BCJR and EM algorithms are iterated. Given initial estimates H^{ini} from the efficient nonlinear ICA method we introduced in chapter 3, the SD-BCJR algorithm computes the signal a-posteriori probabilities $p(s_k|Y, H^i, \Sigma)$ by utilizing the code structure and then feeds these to the EM algorithm. The EM algorithm uses these a-posteriori probabilities to evaluate the conditional expectations in (5.19) and (5.20). Thus we update new channel state information by (5.21) and pass this back to the SD-BCJR algorithm again. As the iterations proceed, estimates become

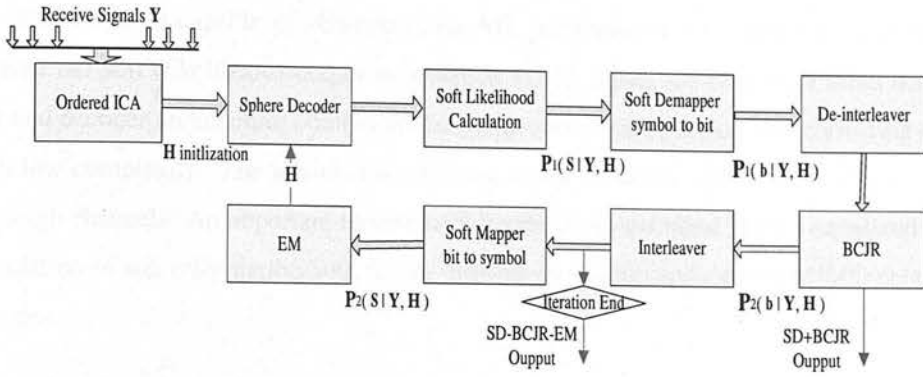


Figure 5.2: Receiver architecture of the proposed coding assisted system.

more accurate and the a-posteriori symbol probabilities become more precise.

Figure 5.2 shows the receiver structure using iterative equalization, whereby a soft equalizer interacts with a soft-input-soft-output error control decoder.

In MIMO channels, this soft decoding strategy for blind equalization consists of four stages:

- 1) Blindly estimate the channel state information from the statistics of the received signals. Here, we use an efficient nonlinear ICA approach to get the initial channel state information estimate.
- 2) Estimate the soft bits, i.e., the LLR of each transmitted bit, using the list version of the sphere decoder or its variants and the current channel state information estimate.
- 3) Make the soft bit information more reliable through a simple BCJR soft decoder.
- 4) Update the channel state information by the EM algorithm with the soft bit information input and feed it back to the channel estimator for further improvement.

Note that, this hybrid architecture does not calculate the full symbol probabilities in order to reduce the system complexity. It uses a number of approximations during the iterative procedure, *i.e.* in the list sphere decoder, the soft mapping and de-mapping parts and the details which are described in the following.

5.2.1 Soft MIMO Likelihood and The BCJR Decoder

The optimal ML receiver has exponential complexity with the signal modulation size and number of transmitted antennae, thus limiting its real time application. The sphere decoder, on

the other hand, is capable of achieving near ML performance [181] and can be designed to provide the soft (likelihood) output information [182]. Thus, we propose a blind soft equalizer and decoder architecture combining the SD decoder and a simple error correcting operator with low complexity. The low-complexity may enable iterative equalization for fast wireless Rayleigh channels. An important requirement for the proposed blind MIMO equalization is the calculation of soft information both for the channel estimation and the soft MIMO detector and decoder.

Channel coding has been extensively researched in the literature, we now introduce the techniques used in this chapter. The well known Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm [183] is used to compute a-posteriori probabilities (APP) of inputs to a finite state machine for received signals. It has been applied to many channel correcting codes such as, turbo code [5].

Define $\psi_k \in \{0, 1, \dots, Q - 1\}$ the state of the channel trellis at time index k , where $Q = |A|$ is the number of points in the constellation A . There is a one-to-one correspondence between symbols and the state value ψ_k . Let $s^{p,q}$ be the symbol input which indicates the state transition from p to q , then, the APP $P(s_k = i|\mathbf{y})$ is given as

$$P(s_k = i|\mathbf{y}) = \sum_{p,q:s^{p,q}=i} P(\psi_k = p, \psi_{k+1} = q|\mathbf{y}) \quad (5.2)$$

Note that the term in the summation in (5.2) can be factorized into three parts, as (5.2) shown. One, a *forward metric*, which relies on the previous channel outputs and which is composed of a vector, $\mathbf{y}_{l < k}$, and a *backward metric*, which relies only on future channel outputs and which is composed of a vector $\mathbf{y}_{l > k}$. The branch metric, depends on the current channel output \mathbf{y}_k . After some mathematical manipulations,

$$P(\psi_k = p, \psi_{k+1} = q|\mathbf{y}) = \alpha_k(p)\gamma_k(p, q)\beta_{k+1}(q)/p(\mathbf{y}) \quad (5.3)$$

where

$$\alpha_k(p) \equiv p(\psi_k = p, \mathbf{y}_{l < k}) \quad (5.4)$$

$$\beta_{k+1}(q) \equiv p(\mathbf{y}_{l > k}|\psi_{k+1} = q) \quad (5.5)$$

$$\gamma_k(p, q) \equiv p(\psi_{k+1} = q, \mathbf{y}_k|\psi_k = p) \quad (5.6)$$

Now, equation (5.4) can further be rewritten for the probability $\alpha_{k+1}(p)$,

$$\alpha_{k+1}(q) = \sum_{p \in \sigma_k} \gamma_k(p, q) \alpha_k(p) \quad (5.7)$$

where σ_k is the set of all states at time k . Thus, we can compute a forward metric $\alpha_{k+1}(q)$ for each state q at time $k + 1$ using the forward recursion. Similarly, equation (5.5) can be written for the probability $\beta_k(q)$ as

$$\beta_k(p) = \sum_{q \in \sigma_{k+1}} \gamma_k(p, q) \beta_{k+1}(q). \quad (5.8)$$

where σ_{k+1} is the set of all states at time $k + 1$. To avoid the underflow on finite precision computers during recursion in (5.7) and (5.8), people commonly normalize α_k and β_k at each time k , so that $\sum_{q \in \sigma_k} \alpha_k(q) = 1$ and $\sum_{p \in \sigma_k} \beta_k(p) = 1$. Finally, the branch metric is given as

$$\begin{aligned} \gamma_k(p, q) &= p(\psi_{k+1} = q, \mathbf{y}_k | \psi_k = p) \\ &= p(\mathbf{y}_k | \psi_k = p, \psi_{k+1} = q) p(\psi_{k+1} = q | \psi_k = p) \end{aligned} \quad (5.9)$$

If a state transition from state p to state q at time k . $p \rightarrow q$, is a valid state transition, for a continuous output AWGN channel, the first term in equation (5.9) can be written as,

$$p(\mathbf{y}_k | \psi_k = p, \psi_{k+1} = q) = \frac{1}{\sqrt{2\sigma^2}} \exp^{\frac{1}{2\sigma^2} \|\mathbf{y}_k - \mathbf{y}^{p,q}\|^2}, \quad (5.10)$$

where $\mathbf{y}^{p,q}$ is the noiseless channel output associated with the state transition from p to q .

While a full complexity BCJR soft decoder is used in this chapter, an efficient sliding window type scheme [184] can be applied in practical applications, which leads to suboptimal performance with a much lower complexity.

For a BPSK modulation scheme, the log-likelihood ratio (LLR) takes the form (LLR):

$$L(b_k) = \ln \frac{p(b_k = 1 | Y, \tilde{H})}{p(b_k = 0 | Y, \tilde{H})}, \quad (5.11)$$

where \tilde{H} is the channel estimate. This LLR value shows the reliability of the information bit. Given a convolutional code at the transmitter, the well known BCJR algorithm calculates the APP exactly if we know the true channel state information.

5.2.2 Soft Mapper and Demapper

The BCJR algorithm is designed for the convolutional bit sequence. Then, in a large constellation QAM modulation, the soft symbol information needs to transfer to the soft bit information for the following BCJR operations. In this section, a description of the QAM soft mapper and demapper employed in our proposed decoding scheme is given. Thanks to the bit interleaver in both the transmitters and receivers, the marginal posteriori probabilities of the coded symbols can be expressed as the product of the bit a-posteriori probabilities.

The soft demapper and mapper take the following three steps to compute the output symbol-wise APP to be fed back to the EM channel estimator.

1. Demapping with a-priori probabilities.

Define a coded symbol s , with m bits, as $s = \{b_0, \dots, b_j, \dots, b_{m-1}\}$. The demapper extracts a soft value of each coded bit for following decoding according to the encode scheme. The following gives a description of this demapper.

For a number of m coded bits, the L-value of bit b_j is given as

$$L(b_j) = L_a(b_j) \ln \frac{\sum_{s_k \in S_1^m} p(\mathbf{y}|s_k) e^{L(s_k)}}{\sum_{s_k \in S_0^m} p(\mathbf{y}|s_k) e^{L(s_k)}}, \quad (5.12)$$

where S_1^m and S_0^m define the subsets of S in which the bit b_j takes the values 1 and 0, correspondingly. s_k is the mapping symbol with the bit b_j taking the values 1 and 0. $L(s_k)$ is the likelihood of symbol point s_k at time index k . The a-priori L-values are ,

$$L_a(b_0) = \ln \frac{p(b_0 = 1)}{p(b_0 = 0)}, \quad L_a(b_1) = \ln \frac{p(b_1 = 1)}{p(b_1 = 0)}. \quad (5.13)$$

For $m = 2$, 4-QAM modulation, the channel decoder calculates LLR on the coded bits b_0, b_1 for each received symbol. The L-value of b_0 can be computed as

$$\begin{aligned} L(b_0|\mathbf{y}) &= \ln \frac{p(b_0 = 1|\mathbf{y})}{p(b_0 = 0|\mathbf{y})} \\ &= \ln \frac{p(b_0 = 1, b_1 = 0|\mathbf{y}) + p(b_0 = 1, b_1 = 1|\mathbf{y})}{p(b_0 = 0, b_1 = 0|\mathbf{y}) + p(b_0 = 0, b_1 = 1|\mathbf{y})}. \end{aligned} \quad (5.14)$$

Using Bayes's rule it further can be expressed as

$$L(b_0|\mathbf{y}) = L_a(b_0) + \ln \frac{p(\mathbf{y}|b_0 = 1, b_1 = 0) + p(\mathbf{y}|b_0 = 1, b_1 = 1)e^{L_a(b_1)}}{p(\mathbf{y}|b_0 = 0, b_1 = 0) + p(\mathbf{y}|b_0 = 0, b_1 = 1)e^{L_a(b_1)}}. \quad (5.15)$$

This probability of each mapping point s_k is computed from the equalizer output. From equation (5.12), this soft demapper is calculating marginal probabilities and we ignore the bit dependencies within the codeword.

2. Uses BCJR algorithm (or other FEC algorithms) to calculate bit-wise posterior probabilities based on the marginal probabilities obtained in last step.
3. Maps bit APPs to symbol APPs. For each symbol s_k , the channel joint symbol-wise posterior $p(s_k)$ can be written as the product of the input marginal bit-wise posterior $p(b_j)$. It is given as

$$p(s_k) = \prod_{j=1}^m p(b_j), \quad (5.16)$$

where

$$p(b_j) = \frac{e^{L(b_j)}}{1 + e^{L(b_j)}}, \quad (5.17)$$

denotes the input bit-wise priors offered by the channel decoder in last step. Generally, LLR clipping techniques [185][186] can be applied to reduce complexity. Here, we fix the LLR clipping level $L_{clip} = 3$ as used in [187].

Note that, a Gray mapping was employed in this work and the optimized symbol mappings for BICM with interleaved decoding were researched in high order constellations, the interested reader is referred to [188]. The soft mapper used in step 3 may lose some information since the joint symbol probability is set equal to the product of marginal bit probability. This assumption is only exact when the bit stream probabilities are strictly independent.

5.2.3 EM Channel Estimation with a-Posteriori Probability

Most prior work in blind iterative channel identification can be tied to the EM algorithm [143]. It is a general methodology for maximum likelihood or maximum a-posteriori estimation. The first use of EM with soft symbol estimates was proposed in [189]. An adaptive version of EM was applied in the identification problem in [190] and some modified EM algorithms were

proposed in [164][191]. The EM algorithm updates are analytically simple and numerically stable for distributions that belong to the exponential family. Similar to the EM application of the EM channel update in previous chapter, we explore EM Channel Estimation that exploits a-posteriori information.

Considering the system model described in equation (5.1), the EM algorithm estimates the channel H based on received signals $Y = \{y_k\}_1^T$. It maximizes the log likelihood, $\log P(Y|H)$ with an initial channel H^{ini} , by iteratively calculating,

$$H^{i+1} = \arg \min_H E[-\log P(Y|H^i, S^i)P(S^i|Y, H^i)], \quad (5.18)$$

where H^i is the i th estimate of the channel and S^i is the i th estimate of the symbols. As we know, the EM iteration in (5.18) only guarantees convergence to a local maximum of $P(Y|H)$ [192].

Again, as stated in the last chapter, the update of the equation (5.18) can be written in a closed-form solution [146] as follows,

$$r^j = \sum_{k=1}^T y_k E\{s_k^j | Y, H^k\} \quad (5.19)$$

$$R^i = \sum_{k=1}^T E\{s_k^i (s_k^i)^* | Y, H^i\} \quad (5.20)$$

$$H^{i+1} = (R^i)^{-1} r^i \quad (5.21)$$

Equations (5.19) and (5.20) depend on first-order statistics and the second-order statistics of the symbols respectively. Note that the computation of (5.19) and (5.20) also require the a-posteriori probabilities $P(s_k|Y, H)$ and $P(s_k s_k^* | Y, H)$, which are approximated in equation (5.16). It is generated by the BCJR algorithm or other error correcting code instead of from the symbol estimators directly when using an error correcting code. We emphasize that the EM algorithm can make full use of the soft a-priori information of the coded bits from the BCJR decoder and these posterior probabilities exploiting the coding structure can provide more accurate channel estimates.

An important problem in the performance of the EM algorithm is the appropriate selection of the initial estimate. In the case of low order constellation modulations and small number of received antennae, the EM algorithm may converge to the desirable point after several re-

initializations of iterative algorithm [193], we show this in the simulation section with 2×2 MIMO system with BPSK modulations. The experiments illustrate convergence to the global minimum in an uncoded system. However, if the likelihood surface is complicated, which happens in high order modulation and with a large number of receive antennae, the EM algorithm is liable to converge to a local minimum rather than a global minimum. Such convergence behaviour has been studied by many researchers, see [155] and references therein.

In the last chapter, the EM algorithm was shown to have a Newton-like convergence locally in digital communication systems. This potentially makes the EM algorithm suitable for real time applications in wireless communications. However, there is still the important but unsolved problem of whether the EM algorithm can converge to the correct solution, i.e., the consistent solution of the true channel parameters. An iterative joint channel estimation and symbol decoder with coding assist can partly solve such difficulty.

5.3 Codeword Over Multiple Blocks

In blind MIMO separation, poor estimates principally occur when the SNR is low or the channel is singular [178]. The latter is related to the channel matrix condition number, γ , which is the ratio of the largest singular value over the smallest singular value. It is a measure of how ill-conditioned the matrix is at receiver. When the channel is very singular, a precise blind estimate may be problematic and the information carrying capability is small with a reasonable SNR in wireless communications, *e.g.* 15dB. However, for a well-conditioned channel matrix, blind separation can usually provide good estimates. Here we utilize this reliable information to rescue the information in singular channel. The singular channel matrix appears with a small probability as we will see below, and the information received from good the channel estimates can be used to correct the bad information from singular channels.

One potential problem is that the long codeword will effect the system complexity and the decoding delay. Thus the next question is how many blocks are needed to form a codeword. Roughly speaking, this is a function of many parameters, such as : the SNR, the performance of blind separations and the frequency of the singular channels. For a fixed SNR and a particular blind separation method, the system performance is always affected by the condition number of channel matrices.

The condition number is determined by the size of channel and channel statistical properties.

For an $N_R \times N_R$ MIMO system, the Rayleigh fading channel assumption satisfies, $H \sim CN(0, 1)$ and the pdf of the normalized condition number $\tilde{\gamma} = \gamma/N_R$ is given by [194]

$$p_{\tilde{\gamma}}(x) = \frac{8}{x^3} \exp\left(-\frac{4}{x^2}\right). \quad (5.22)$$

With this equation, we can calculate the probability that the condition number will be greater than a specific value, γ_1 as:

$$P\{x > \gamma_1\} = \int_{\gamma_1}^{\infty} \frac{8}{x^3} \exp\left(-\frac{4}{x^2}\right) dx \quad (5.23)$$

which by the change of variable $y = 4/x^2$ can be simplified to:

$$\begin{aligned} P\{x > \gamma_1\} &= \int_{4/\gamma_1^2}^0 -\exp(-y) dy \\ &= 1 - \exp\left(-\frac{4}{\gamma_1^2}\right) \end{aligned} \quad (5.24)$$

This can help us to find a reasonable block length within a codeword. For example, in a 4×4 MIMO system, if the blind separation algorithm can not provide satisfactory performance when the channel condition number $\gamma > 20$ under practical SNRs, we can calculate its probability by

$$p_{\tilde{\gamma}}(x > 20) = 1 - \exp\left(-\frac{4}{(20/4)^2}\right) \approx 0.15 \quad (5.25)$$

Assume that singular channels appear uniformly, then, the expected frequency of a singular channel is 0.15 in a block. So, an appropriate block length takes the reciprocal of 0.15, approximated 7.

Note that, we just give a simple example to state that the long codeword with different fading factors can make use of the temporal diversity, then increase the error correction capability. Such an advance is applied to the blind channel estimation iteratively and improves the final performance both in the channel estimate and the BER significantly.

5.4 Simulations

As stated in [143], each local maximum has a domain of convergence. Near a local maximum, the EM algorithm converges linearly to that local maximum. If the likelihood is unimodal, any

EM converges to the global maximum starting anywhere in the parameter space [155]. In digital communications, we test the simple case where a 2×2 uncoded instantaneous MIMO system with BPSK modulation and Rayleigh channels. In Figure 5.3, a ZF detection with perfect CSI and a blind ICA approach [3] are run for comparison. The ML result come from a EM algorithm with exact CSI. The initial value of EM comes from the previous douglas ICA method. There is only a very small gap between the EM performance and the optimal ML solution. Through the empirical observations, we may speculate the EM may converge to the desired point if the likelihood surface is simple.

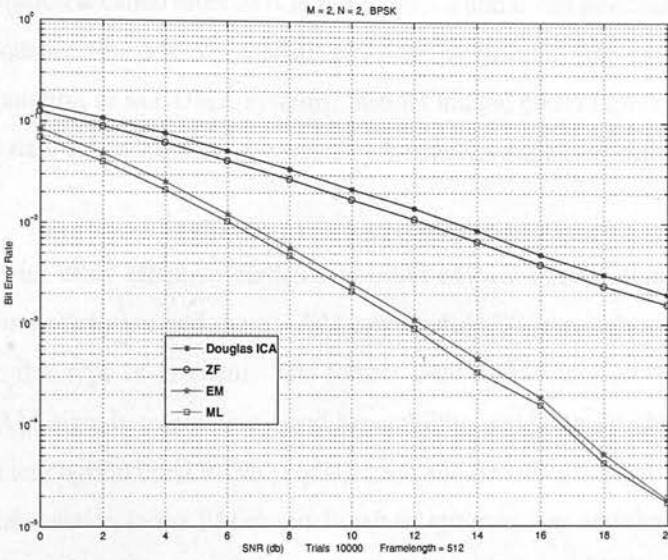


Figure 5.3: BER performance of a 2×2 uncoded MIMO system. An EM channel estimation with BPSK modulation. ML is obtained by the EM with perfect CSI.

Next, we consider high dimensional systems. A MIMO system with four transmitters, four receivers and 16-QAM modulation are employed. The Rayleigh fading channel H is a 4×4 or 8×8 complex instantaneous matrix, which is constant for each block interval (256 symbols). N follows the complex additive white Gaussian distribution. The results have been obtained for transmitting blocks of $K_b = 4096$ bits in a 4×4 system and $K_b = 812$ bits in a 8×8 system. For the error correcting system, a rate of $r = 1/2$ parallel concatenated convolutional code of memory 3 with two nonsystematic convolutional (NSC) code has been used. The generator polynomials are $G_1(D) = 1 + D + D3 + D4$, $G_2(D) = 1 + D3 + D4$ and the interleaver is set to pseudo random. The SD [195] is employed first to get soft information. Only two iterations of the channel estimation are employed in EM updates since the EM has a fast convergence

behaviour. In order to explore the full contribution provided by FEC, this work computes the symbol likelihood based on 16 points on each dimension. This calculation could be simplified by a list sphere decoder [185] or list-fixed-complexity sphere decoder [196] but with a potential performance degradation.

To explore the error correcting code correctly, the permutation problem of ICA should be overcome. The channel re-ordering technique is used in our simulations. Given initial estimates H_{ICA} , we try to calculate the channel permutation matrix, P , by making the $H_{ICA}^{-1}H = DP = I$, where D is a diagonal matrix. Then, the channel estimates $H^{ini} = H_{ICA}P$ is ordered with P . This kind of operation is called ordered ICA in Figure 5.2 and it can guarantee the correct order of the coded sequence. In practice, a small pilot can be inserted into data block to indicate the correct permutation or in CDMA systems, distinct unique codes (spread code) are applied into transmitted data stream in each antenna, then correct permutation can be identified at the receivers.

In comparison with other effective methods of blind MIMO equalization, such as the *Split Threshold nonlinear function* and the SD-EM approach [178], our scheme shows promising performance for this type of problem. The former used an efficient score function which is specified for QAM signals to obtain a good separability, as shown in chapter 3. The latter proposed an efficient hybrid blind MIMO equalization and decoding scheme as stated in chapter 4. It used soft information in the EM channel update, ignoring any additional information due to FEC. Figure 5.4 and 5.5 illustrate the separability improvements with the aid of channel code in 4×4 and 8×8 MIMO systems. This performance is measured by the distance of the estimated channel from the true value as defined by equation (3.44).

In digital communications, the ultimate goal is to obtain the optimum BER performance, these are shown in figures (5.6) and (5.7). Comparisons are set up with a ZF scheme and the two blind methods mentioned above. In order to illustrate the benefits from the channel improvement rather than the FEC, the BER is measured before the error correcting operations. Obviously, the coding assisted iterative structure improves the system performance significantly since it can help avoid the EM algorithm from becoming trapped in a local minimum. The performance of 8×8 system is better than those of 4×4 system since it can provide more useful channel estimates within a singular channel matrix, therefore more useful information can be fed to the FEC decoder. Although, the BER performance is improved, there is still a performance gap between the SD-BCJR-EM algorithm and the optimal (with known CSI) SD solution in

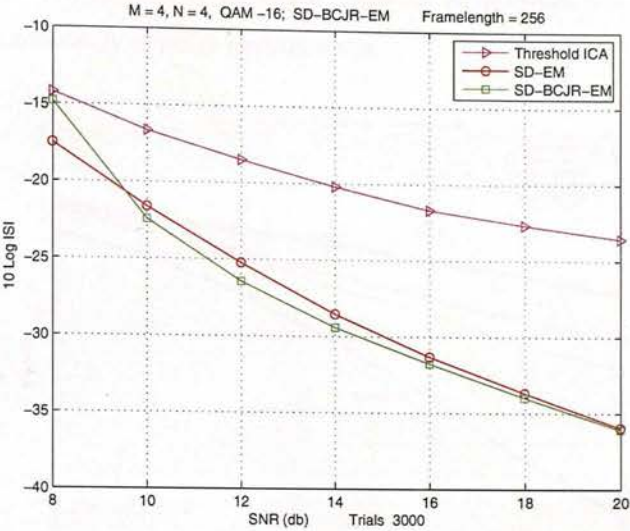


Figure 5.4: Channel separability of the split threshold nonlinear ICA, the SD-EM and the coding assisted SD-EM algorithm with a rate $r = 1/2$ convolutional code over different SNR in 4x4 MIMO systems.

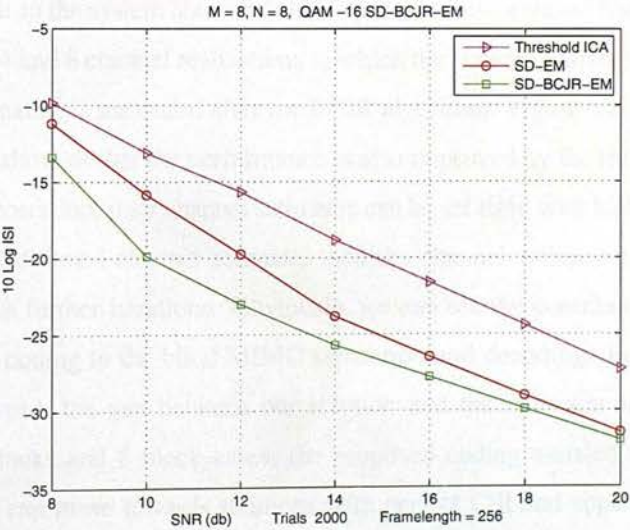


Figure 5.5: Channel separability of the split threshold nonlinear ICA, the SD-EM and the coding assisted SD-EM algorithm with a rate $r = 1/2$ convolutional code over different SNR in 8x8 MIMO systems.

both figures. Such phenomenon is introduced by some very singular channel estimates within a block, in which the amplitudes of one or two channels are so weak that blind estimation can not identify them accurately in noisy environments.

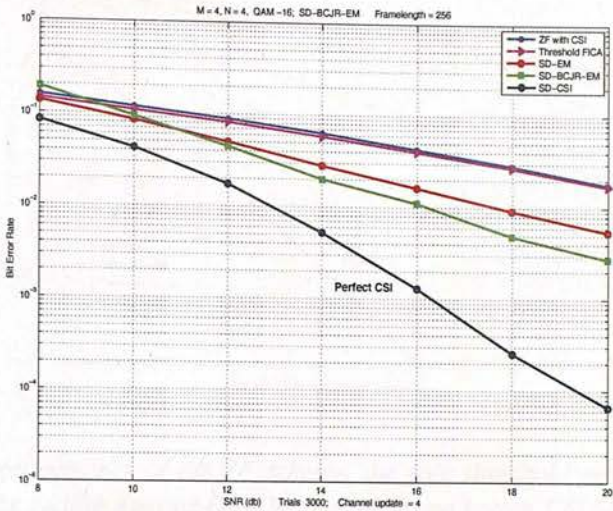


Figure 5.6: BER performance of the ZF scheme, the split threshold nonlinear ICA, the SD-EM, the coding assisted SD-EM algorithm and known CSI SD with a rate $r = 1/2$ convolutional code in 4×4 MIMO systems.

For a setup similar to the system above but which provides an increased time diversity, we apply a codeword to 2, 4 and 8 channel realizations in which the channel realizations are independent. The BER performance is measured after the BCJR algorithm. Figure 5.8, 5.9 and Figure 5.10, in the next page, show us that the performance is also improved by the time diversity. Through the interleaving operation, poor channel estimates can be set right with high quality information from the well conditioned channel estimate, then the channel estimate can be adjusted to be more reliable with further iterations. Obviously, we can see the contribution of time diversity with the channel coding to the blind MIMO separation and decoding. In the case of 2 blocks forming a codeword, the gap between our solution and the optimum solution is still large. However, in 4 blocks and 8 block cases, the proposed coding assisted blind separation and decoding system can move towards solutions with perfect CSI and approximate the optimum solutions at 20 dB and 12 dB, respectively. On the whole, a better performance can be obtained by further utilization of the time diversity. However, a codeword combining 4 blocks or more increases system computational loads and takes a long decoding delay, so there is a tradeoff between the complexity and BER performances.

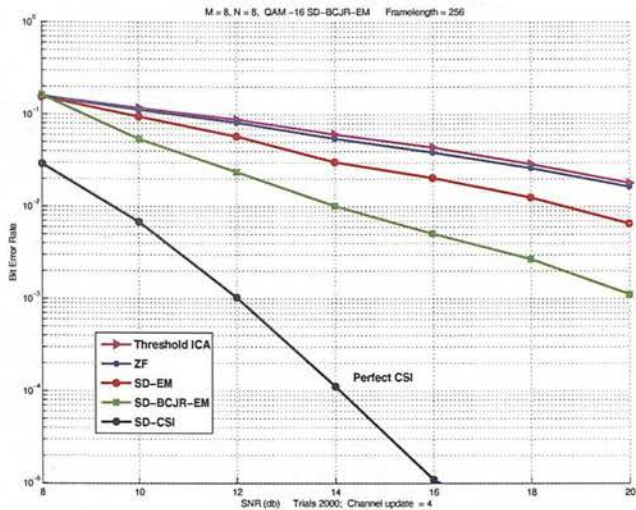


Figure 5.7: BER performance of the ZF scheme, the split threshold nonlinear ICA, the SD-EM, the coding assisted SD-EM algorithm and known CSI SD with a rate $r = 1/2$ convolutional code in 8x8 MIMO systems.

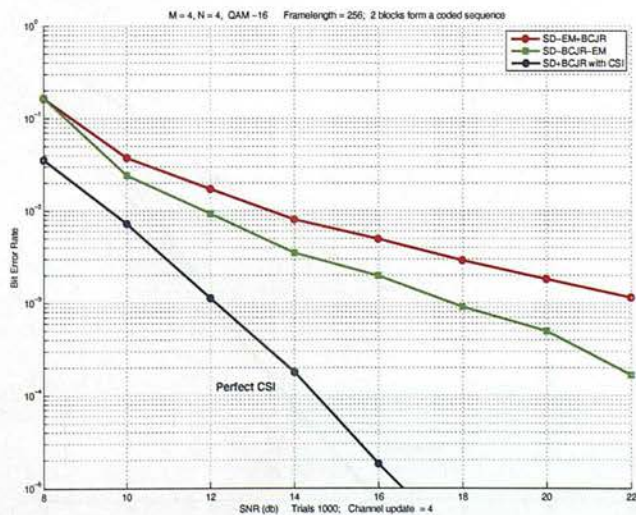


Figure 5.8: BER improvements by utilizing time diversity and channel coding. Triangle line is the performance of the SD-EM algorithm following the BCJR algorithm and the square line is the performance of the coding assisted SD-EM algorithm. 2 blocks form a codeword.

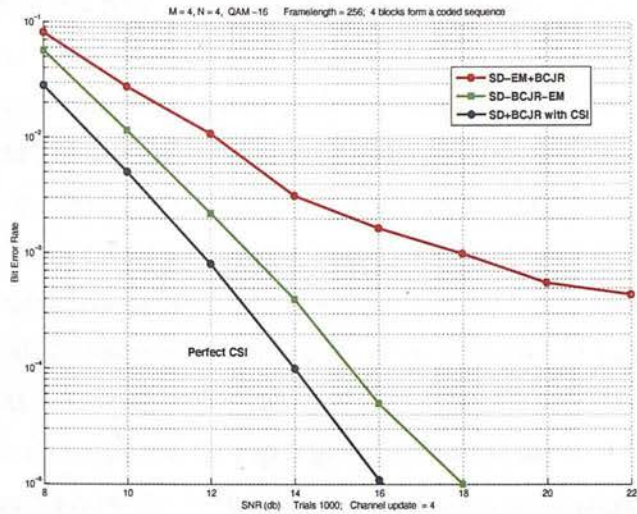


Figure 5.9: BER improvements by utilizing time diversity and channel coding. Triangle line is the performance of the SD-EM algorithm following the BCJR algorithm and the square line is the performance of the coding assisted SD-EM algorithm. 4 blocks form a codeword.

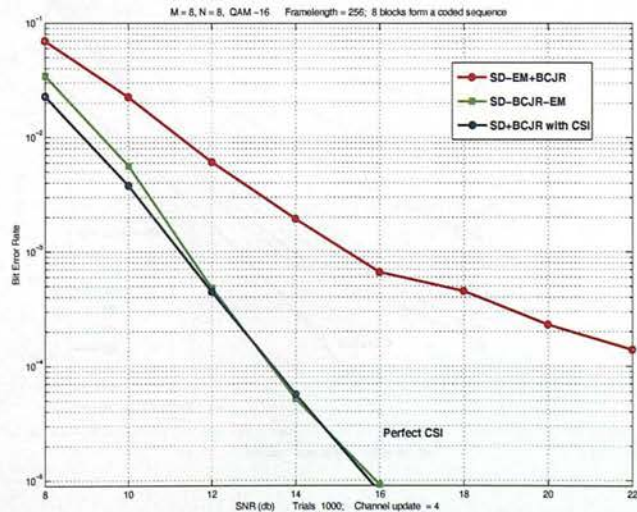


Figure 5.10: BER improvements by utilizing time diversity and channel coding. Triangle line is the performance of the SD-EM algorithm following the BCJR algorithm and the square line is the performance of the coding assisted SD-EM algorithm. 8 blocks form a codeword.

Our final simulation indicates the iterative gain of the coded soft channel estimation measured over 20,000 Monte Carlo runs. Similar to the setup above with 4 fixed channel realizations. One channel matrix is constrained to be very singular with a condition number over 25 so that it is difficult to estimate with a linear estimator blindly. The other channels are typically good in terms of the channel condition number and the parameters can be found in appendix D, eight iterations performance are studied. Clearly, from Figure 5.11, we can see that the performance progresses towards the optimal curve with CSI known at the receiver. The first singular channel matrix separability is illustrated in Figure 5.12 and the iterative improvements show the same trend of convergence. In this channel case, this is still a gap to the performance with the perfect CSI since there is only 256 symbol each block to estimate the channel through the EM update. We speculate that the gap is going to small as the block length going to large. As shown in Figure 5.11, for the 5–8th iterations, the convergence properties of the decoding process improves rapidly when SNR is greater than 18 dB.

The system complexity is high when we iteratively update the channel with the aid of a channel code. In practice, we can switch to only one or two iterations when the channel condition is good. A trade off can be built up flexibly between the performance and the complexity.

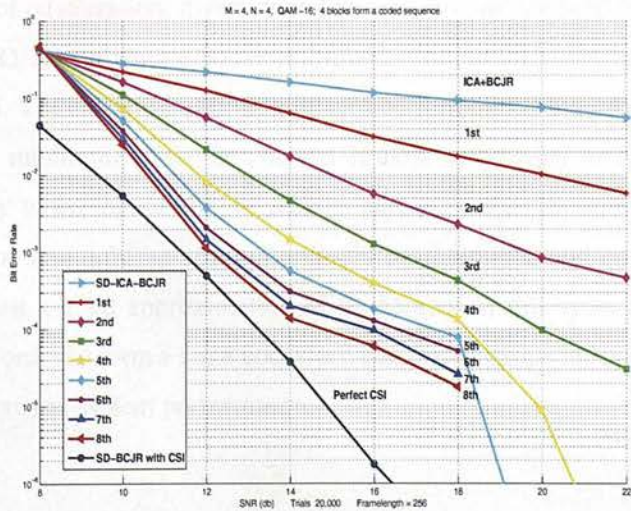


Figure 5.11: Iterative BER improvements of the coding assisted SD-EM algorithm in a 4×4 system with 16-QAM modulation, eight iterations are used.

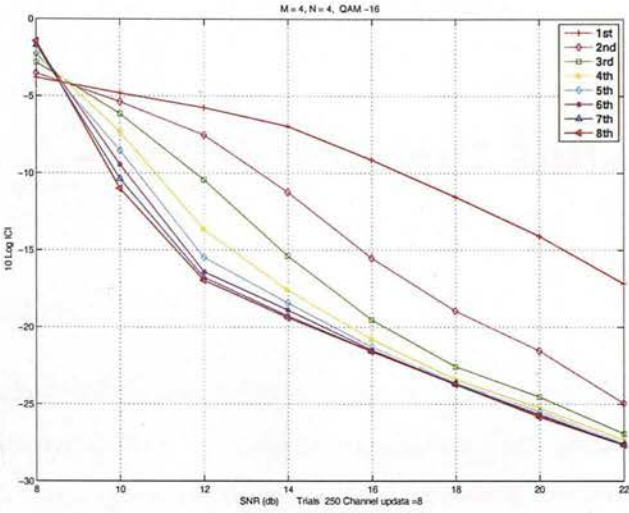


Figure 5.12: Iterative channel improvements of the coding assisted SD-EM algorithm in a 4x4 system with 16-QAM modulation.

5.5 Chapter Summary

In this section, a coding assisted MIMO blind equalization and decoding scheme is proposed. Three techniques of equalization, diversity and channel coding are used to improve fading link performance (BER). By utilizing a-posteriori information, a substantial gain over the uncoded system is provided. The existence of coding structures partly solves the problems of EM getting trapped in a local minimum when the channel is close to singular or the SNR is low. This happens frequently when the number of receive antennae, the size and the dimension of the data are large. The new scheme appears to avoid local minima and converges to the global minimum or at least a good approximation of it. Moreover, this system extends FEC to the multiple blocks in order to form a large codeword and then exploits the temporal diversity. This extension also improved system performance.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

Future wireless communications are desired to support high data rates transmission when considering the growing applications. Traditional equalization sends pilot on training sequences periodically, which make a poor use of the available bandwidth resources. Blind equalization can detect the transmitted symbols without the aid of a training sequence thus saving the limited bandwidth, therefore, blind equalization is an active research area in practice and can be verified by potential theory. In the recent two decades, with the appearance of MIMO systems, the blind MIMO channel estimation and equalization become an interesting and challenging research field.

For blind channel separation, ICA is very suitable for the identification of MIMO channels since the transmitted data on each antenna can be considered to be approximately independent. However, typical ICA algorithms can not provide accurate enough channel estimates to achieve good BER performance. In this work, based on the probability density function of modulation signals, we have proposed a simple and efficient nonlinear threshold function to separate the mixed MIMO signals. This simple but powerful function can provide good separability and BER performance as we have shown. The gradient update and the fixed point algorithm based upon this nonlinearity are also derived. The bias removal update is also presented and the corresponding stability and convergence properties of this nonlinear function are given. This nonlinear function is appropriate for large constellation QAM modulation systems in wireless communication as experiments have shown. A key advantage of this method is the threshold operation can be operated by two comparators. This easy implementation makes it feasible for application to a real time communication system.

For MIMO systems with a large number transmitter and receivers, a more accurate blind detector is presented. This detector makes use of EM estimation to update the channel estimates. We have demonstrated mathematically, in the digital communication cases, that the EM updating algorithm has a fast convergence behaviour in the low noise limit. Moreover, to save computa-

tional costs, the E-step can be approximated by a fixed complexity sphere detector. Based on these findings, an efficient hybrid system for blind equalization of large constellation MIMO systems was presented. Such a scheme outperforms the traditional linear MIMO detectors with affordable complexity.

A potential problem of the EM algorithm is that it is liable to get trapped in a local minima when the channel is close to singular or the SNR is low. This happens frequently when the number of receiver antennas, the dimension of the data are large. In order to overcome this problem, we explore the use of error correcting codes to enhance blind source separation. A coding assisted MIMO blind separation and decoding scheme has been proposed. By utilizing a-posteriori information, the channel update and soft symbol detection are operated iteratively. The existence of coding structures appears to avoid local minimum and converges to the global minimum. This design provides substantial gain over the uncoded system and improves the final performance significantly. In the last part of this study, the three techniques of equalization, diversity and channel coding are used to improve fading link performance (BER). Applying a error correcting codeword to multiple blocks, we can utilize the temporal diversity to compensate poor channel estimates introduced by the deep fading or very singular channel matrix and then recover information correspondingly.

6.2 Future Work

This thesis has focused on a general narrow band case where the non-time dispersive linear channel is modelled. For this popular wireless system, we have provided a simple and flexible blind receiver for large constellation modulation and there are several probabilities of future work based on this work.

6.2.1 Convolutional Channel Estimate

As an extension of linear channels, wireless channels can be represented as a linear time-dispersive system in broadband communications. The channel can be modelled as a convolutional mixing. This is always called blind wide band source separation in signal processing areas. Such problem is more complicated than instantaneous ICA when the various time delay in different wireless channels are considered. Generally, a popular solution is to reformulate it as an ordinary ICA problem by mapping it into the Fourier (frequency) domain. However, the

frequency permutations problem still exist [197]. Some methods directly extend instantaneous BSS methods to the convolutive case in the time domain, such as methods based on natural gradient [198], minimization of Kullback–Leibler divergence [199] and joint-diagonalization criterion [200]. In time domain, the permutation problem can be avoided and then these methods could be used to solve the convolutive channel estimate problem.

6.2.2 Semi-Blind Methods

Semi-blind approaches send a small amount of pilot in the transmitter and provide good initialization. Then by switching to a blind method, they use both the pilot and the information symbols to update the channel [201]. The initialization based on the pilot sequences allow for fast convergence of the channel estimates. Further, the training data could be designed to solve the phase ambiguities and permutation problems that always remain in blind approaches. Fast convergence is needed in fast time-varying channels, so, semi-blind methods will probably be more appropriate in practice. Further systematic research can be developed based on the our blind architecture of both coded and uncoded systems, *e.g.*, how best to integrate pilot sequences into the current blind estimate framework; what is the most appropriate length and location of pilot sequences across different SNR?

6.2.3 Complexity Reduction

For blind MIMO channel estimation and equalization, fast real time batch and on-line signal processing with low complexity is still challenging. A simple and efficient real time implementation is desirable for practical applications. Although the complexity of our blind method is reduced when compared to ML methods, it is still large if the coding assisted estimation and decoding are applied. Generally speaking, our SD-EM and SD-BCJR-EM algorithm are designed in order to correct the inaccurate estimation of ICA when the SNR is low and the channel is singular. The improvement is insignificant when the channel is good. We could therefore reconfigure the iteration number of channel update in coding assisted systems through the unbiased threshold ICA. We observe that it can obtain good condition number estimates even when the specific channel estimates is poor. Further research can be opened up for the purpose of complexity reduction.

Appendix A

Derivation of equation (3.42)

A brief review of the complex-valued FastICA algorithm is introduced here for completeness. The cost function of ICA is:

$$J_G(\mathbf{w}) = E\{G(|\mathbf{w}^H \mathbf{x}|^2)\} \quad (\text{A.1})$$

where $G : \mathbf{R}^+ \cup \{0\} \rightarrow \mathbf{R}$ is a smooth even function, W , de-mixing matrix, is an n -dimensional complex weight vector, \mathbf{w} is a column of W and

$$E\{|\mathbf{w}^H \mathbf{x}|^2\} = 1 \quad (\text{A.2})$$

where \mathbf{x} is the received signal after pre-whitening. Notice that the extreme of a contrast function is a well defined problem only if the function is real. For this reason our contrast functions operate on absolute values rather than on complex values. According to the Kuhn-Tucker conditions, the optima of equation (A.1) under the constraint A.2 are obtained at points where:

$$\nabla E\{G(|\mathbf{w}^H \mathbf{x}|^2)\} - \beta \nabla E\{|\mathbf{w}^H \mathbf{x}|^2\} = 0 \quad (\text{A.3})$$

where β is a constant and the gradient is computed with respect to real and imaginary part of \mathbf{w} separately. Using Newton method to solve (A.3), the following approximating Newton iteration can be obtained,

$$\begin{aligned} \mathbf{w}^+ &= E\{\mathbf{x}(\mathbf{w}^H \mathbf{x})^* g(|\mathbf{w}^H \mathbf{x}|^2)\} \\ &\quad - E\{g(|\mathbf{w}^H \mathbf{x}|^2) + |\mathbf{w}^H \mathbf{x}|^2 g'(|\mathbf{w}^H \mathbf{x}|^2)\} \mathbf{w} \\ \mathbf{w} &= \mathbf{w}^+ / \|\mathbf{w}^+\|. \end{aligned} \quad (\text{A.4})$$

Recall system model equation (2.41), the covariance of received signals is given as

$$C = E\{YY^H\} = E\{(HS + N)(HS + N)^H\} = HH^H + \Sigma \quad (\text{A.5})$$

The noise is independent of the sources, Then,

$$E\{SS^H\} = I \quad E\{SS^T\} = O \quad (A.6)$$

and the noise covariance is given as

$$E\{NN^H\} = \sigma^2 I = \Sigma. \quad (A.7)$$

After eigenvalue decompositions, the covariance of received signal C follows the form

$$C = U\Lambda U^H \quad (A.8)$$

and

$$\mathbf{z} = \Lambda^{-1/2} U^H Y \quad (A.9)$$

is the pre-whitened data. Similarly,

$$\mathbf{z}^+ = (\Lambda - \Sigma)^{-1/2} U^H Y \quad (A.10)$$

is quasi-whitened data. By defining $C^{-1/2} = (\Lambda - \Sigma)^{-1/2}$ [130], the pre-whitened data is given as

$$\mathbf{z}^+ = C^{-1/2} U^H Y. \quad (A.11)$$

The quasi-whitened data do not have the identity covariance matrix. The proof is given below,

$$\begin{aligned} E\{\mathbf{z}^+(\mathbf{z}^+)^H\} &= E\{C^{-1/2} U^H Y (C^{-1/2} U^H Y)^H\} \\ &= E\{C^{-1/2} U^H (HS + N) \{C^{-1/2} U^H (HS + N)\}^H\} \\ &= C^{-1/2} U^H H H^H U (C^{-1/2})^H + C^{-1/2} U^H \Sigma U (C^{-1/2})^H \end{aligned} \quad (A.12)$$

Since $C^{-1/2} = (\Lambda - \Sigma)^{-1/2}$ is a real diagonal matrix, then $(C^{-1/2})^H = C^{-1/2}$, and it follows:

$$E\{\mathbf{z}^+(\mathbf{z}^+)^H\} = C^{-1/2} U^H H H^H U C^{-1/2} + C^{-1/2} U^H \Sigma U C^{-1/2} \quad (A.13)$$

By defining $\tilde{\Sigma} = C^{-1/2} U^H \Sigma U C^{-1/2}$ and knowing $HH^H = C - \Sigma$, we achieve:

$$C^{-1/2} U^H HH^H U C^{-1/2} = \Lambda(\Lambda - \Sigma)^{-1} - C^{-1/2} U^H \Sigma U C^{-1/2} \quad (\text{A.14})$$

Taking the second part of equation (A.5) into this equation, we obtain:

$$\begin{aligned} & E\{\mathbf{z}^+(\mathbf{z}^+)^H\} \\ &= \Lambda(\Lambda - \Sigma)^{-1} - C^{-1/2} U^H \Sigma C^{-1/2} + C^{-1/2} U^H \Sigma U C^{-1/2} \\ &= I + \sigma^2(\Lambda - \Sigma)^{-1} \end{aligned} \quad (\text{A.15})$$

Obviously, the quasi-whitened data is not white. Then we modify the complex FastICA algorithm with quasi-whitening data input, the Newton method can be approximated by:

$$\begin{aligned} & \nabla^2 E\{G(|\mathbf{w}^H \mathbf{x}|^2)\} \\ & \approx (I + \sigma^2(\Lambda - \Sigma)^{-1})\{2E\{g(|\mathbf{w}^H \mathbf{z}^+|^2)\} + |\mathbf{w}^H \mathbf{z}^+|^2 g'(|\mathbf{w}^H \mathbf{z}^+|^2)\} \end{aligned}$$

and

$$\beta \nabla E\{|\mathbf{w}^H \mathbf{x}|^2\} = 2\beta(I + \sigma^2(\Lambda - \Sigma)^{-1}) \quad (\text{A.16})$$

Then the following new complex approximate Newton iteration can be achieved:

$$\begin{aligned} \mathbf{w}^+ &= E\{\mathbf{z}^+(\mathbf{w}^H \mathbf{z}^+)^* g(|\mathbf{w}^H \mathbf{z}^+|^2)\} \\ & - (I + \sigma^2(\Lambda - \Sigma)^{-1}) \mathbf{w} E\{g(|\mathbf{w}^H \mathbf{z}^+|^2) + |\mathbf{w}^H \mathbf{z}^+|^2 g'(|\mathbf{w}^H \mathbf{z}^+|^2)\} \end{aligned} \quad (\text{A.17})$$

Appendix B

Proof of Equation (4.36)

Taking the first term of equation (4.36). Since

$$\begin{aligned} & p(\mathbf{s}_i|\mathbf{y}_k, H) \\ = & \frac{p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i|H)}{p(\mathbf{y}_k|H)} \end{aligned} \quad (\text{B.1})$$

$$\begin{aligned} & = \frac{p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i|H)}{\sum_{j=1}^{|\mathcal{S}|} p(\mathbf{y}_k, \mathbf{s}_j|H)} \\ = & \frac{p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i|H)}{\sum_{j=1}^{|\mathcal{S}|} p(\mathbf{y}_k|\mathbf{s}_j, H)p(\mathbf{s}_j|H)} \end{aligned} \quad (\text{B.2})$$

From

$$\frac{\partial[f(H)/g(H)]}{\partial H} = \frac{1}{g^2(H)}[g(H)\frac{\partial f(H)}{\partial H} - f(H)\frac{\partial g(H)}{\partial H}] \quad (\text{B.3})$$

thus:

$$\begin{aligned} & \frac{\partial}{\partial H^T} p(\mathbf{s}_i|\mathbf{y}_k, H) \\ = & \frac{\partial}{\partial H^T} \frac{p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i|H)}{\sum_{j=1}^{|\mathcal{S}|} p(\mathbf{y}_k|\mathbf{s}_j, H)p(\mathbf{s}_j|H)} \\ = & \frac{1}{g^2(H)} \left\{ \sum_{j=1}^{|\mathcal{S}|} p(\mathbf{y}_k|\mathbf{s}_j, H)p(\mathbf{s}_j|H) \times \frac{\partial}{\partial H^T} [p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i|H)] \right. \\ & \left. - p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i|H) \times \frac{\partial}{\partial H^T} \left[\sum_{j=1}^{|\mathcal{S}|} p(\mathbf{y}_k|\mathbf{s}_j, H)p(\mathbf{s}_j|H) \right] \right\} \end{aligned} \quad (\text{B.4})$$

where $g(H) = \sum_{j=1}^{|\mathcal{S}|} p(\mathbf{y}_k|\mathbf{s}_j, H)p(\mathbf{s}_j|H)$ and $f(H) = p(\mathbf{y}_k|\mathbf{s}_i, H)p(\mathbf{s}_i|H)$.

Since $p(\mathbf{y}_k|\mathbf{s}_j, H) = c \exp(k(\mathbf{y}_k, \mathbf{s}_j, H))$ is an exponential function, we can obtain:

$$\begin{aligned} & \frac{\partial}{\partial H^T} p(\mathbf{y}_k|\mathbf{s}_j, H) \\ &= p(\mathbf{y}_k|\mathbf{s}_j, H) \frac{\partial}{\partial H^T} k(\mathbf{y}_k, \mathbf{s}_j, H) \end{aligned} \quad (\text{B.5})$$

Denote $\frac{\partial}{\partial H^T} k(\mathbf{y}_k, \mathbf{s}_j, H)$ as $k'_{k,j}(H)$ then

$$k'_{k,j}(H) = \frac{\partial}{\partial H^*} [-(\mathbf{y}_k - H\mathbf{s}_j)^H \Sigma^{-1} (\mathbf{y}_k - H\mathbf{s}_j)^H]$$

which is finite. Therefore in equation (B.4), the first part in the bracket is

$$\begin{aligned} & \sum_{j=1}^{|\mathcal{S}|} p(\mathbf{y}_k|\mathbf{s}_j, H) p(\mathbf{s}_j|H) \frac{\partial}{\partial H^T} [p(\mathbf{y}_k|\mathbf{s}_i, H) p(\mathbf{s}_i|H)] \\ &= \sum_{j=1}^{|\mathcal{S}|} p(\mathbf{y}_k|\mathbf{s}_j, H) p(\mathbf{s}_j|H) p(\mathbf{y}_k|\mathbf{s}_i, H) p(\mathbf{s}_i|H) k'_{k,j}(H). \end{aligned} \quad (\text{B.6})$$

When noise is low and k is sufficiently large,

$$p(\mathbf{y}_k|\mathbf{s}_j, H) = \begin{cases} 0 & i \neq j \\ p(\mathbf{y}_k|\mathbf{s}_i, H) & i=j \end{cases}$$

where \mathbf{s}_j is the local minimum. Then the equation (B.6) is equal to

$$p(\mathbf{y}_k|\mathbf{s}_i, H) p(\mathbf{s}_i|H) p(\mathbf{y}_k|\mathbf{s}_i, H) p(\mathbf{s}_i|H) k'_{k,i}(H). \quad (\text{B.7})$$

Consider the second part of equation (B.4) in the bracket and using the same idea, we obtain

$$p(\mathbf{y}_k|\mathbf{s}_i, H) p(\mathbf{s}_i|H) p(\mathbf{y}_k|\mathbf{s}_i, H) p(\mathbf{s}_i|H) k'_{k,i}(H). \quad (\text{B.8})$$

Thus

$$\frac{\partial}{\partial H^T} p(\mathbf{s}_i|\mathbf{y}_k, H) = 0, \quad (\text{B.9})$$

the equation (4.37) follows.

Appendix C

Channels Used in Chapter V Simulations

The linear square channel matrix, H , used in chapter 5. Its elements are assumed to be drawn independently from a Rayleigh distribution, where $h_{i,j}(k), \sim CN(0, 1) 1 < i, j < n$. The square channel matrix can be expanded into the non-square overdetermined MIMO systems where the number of receive antennae is greater than the number of transmitted antennae, $N_R > N_T$. Further principal components analysis or singular value decomposition operations are needed as preprocessing to make $N_R = N_T$.

We randomly draw 4 channel matrix and fixed them for 20,000 Monte Carlo runs. The condition numbers of 4 block channels in a codeword is 26.1580, 11.5966, 5.7314 and 5.9095, respectively.

$$H_1 = \begin{bmatrix} -0.2636 - 0.2704i & -0.9062 + 0.0505i & -1.5601 + 1.2145i & -0.4904 + 0.1816i \\ 0.9924 - 0.3147i & 0.2392 - 0.7332i & -0.0598 - 0.4830i & 0.3358 + 0.5572i \\ 1.1831 + 1.0356i & -0.9521 + 0.3823i & 0.2636 - 0.4605i & 0.2988 + 0.9400i \\ -0.1860 + 0.7397i & -0.0076 - 1.0483i & 1.4912 + 0.0384i & 0.1939 + 0.5999i \end{bmatrix}$$

$$H_2 = \begin{bmatrix} -0.3762 + 0.1472i & -0.3494 - 0.6691i & 0.3680 - 0.5868i & -0.4538 - 0.6661i \\ -0.1766 - 0.7189i & -0.6317 + 0.2621i & 0.2478 + 1.7522i & 0.0272 - 0.1648i \\ -0.1121 - 0.3579i & -0.5293 + 0.4152i & 0.0630 - 0.0100i & 0.3058 - 0.2477i \\ 1.2193 + 0.4520i & -1.2945 + 0.7046i & -0.4279 + 1.1214i & -0.0101 + 0.9134i \end{bmatrix}$$

$$H_3 = \begin{bmatrix} -0.0297 - 0.0818i & -0.8255 - 0.4025i & 0.5528 + 1.2157i & 0.1723 + 0.2118i \\ -0.7723 + 0.3704i & 0.0749 - 0.8959i & -0.5666 + 0.0317i & -1.3159 + 0.6531i \\ -0.0424 - 0.1636i & 0.4734 + 1.5803i & -0.1489 + 0.2011i & 0.8763 + 0.4389i \\ 0.1219 - 0.8974i & 0.0316 - 1.7762i & 0.4480 - 0.9119i & -0.0463 + 0.2618i \end{bmatrix}$$

$$H_4 = \begin{bmatrix} 0.1778 - 1.9344i & -1.5282 + 0.4596i & 0.5841 - 0.5065i & -0.3552 - 0.1945i \\ -0.5536 + 0.2296i & -1.2708 - 0.1256i & -0.5353 - 0.4438i & -0.3936 + 0.0086i \\ 0.2110 + 0.1353i & 0.4025 - 0.5441i & 0.4851 - 0.1228i & -0.5507 - 0.3385i \\ 0.7893 - 0.6173i & -0.1968 - 0.8766i & 0.4738 + 0.6765i & -0.4482 + 0.3673i \end{bmatrix}$$

References

- [1] E. Bingham and A. Hyvarinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *Neural Systems*, vol. 10, pp. 1–8, 2000.
- [2] S. A. Cardoso, J.F., "Blind beamforming for non-Gaussian signals," *IEE Proceedings F Radar and Signal Processing*, vol. 140, pp. 362–370, Dec. 1993.
- [3] S. Douglas, "Fixed-point fastica algorithms for the blind separation of complex-valued signal mixtures," in *Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, pp. 1320–1325, 2005.
- [4] L. G. Barbero and J. S. Thompson, "The fixed sphere decoder homepage." <http://www.see.ed.ac.uk/jst/sphere/index.html>.
- [5] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *IEEE International Conference on Communications ICC 93. Geneva. Technical Program, Conference Record*, vol. 2, pp. 1064–1070 vol.2, 1993.
- [6] R. Gallager, *Low Density Parity Check Codes*. MIT Press, Cambridge, Massachusetts, 1963.
- [7] C. Shannon, "A mathematical theory of communication," tech. rep., Bell Syst. Tech. J., vol. 27, pp. 379423 (Part one), pp. 623656 (Part two), Oct., 1948.
- [8] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. Vol. 10, pp. 585–595, 1999.
- [9] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2,, pp. 41–59, 1996.
- [10] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Transactions on Communications*, vol. 23, pp. 679–682, Jun 1975.
- [11] R. Lucky, "Automatic equalization for digital communications," *Bell System Technical Journal*, vol. 44, pp. 547–588, 1965.
- [12] C. Jutten and J. Hraut, "Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.
- [13] J. Cavers, "Pilot symbol assisted modulation and differential detection in fading and delay spread," *IEEE Transactions on Communications*, vol. 43, no. 7, pp. 2206–2212, 1995.
- [14] L. Tong, B. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Processing Magazine*, vol. 21, pp. 12–25, Nov. 2004.

- [15] GSM, "Group special mobile (gsm) recommendations." Series 01-12, 1990. GSM.
- [16] TIA/EIA/IS-136.1, "Tdma cellular/pcs - radio interface - mobile station - base station compatibility digital control channel addendum," No. 1 November 1996.
- [17] WCDMA, "Physical channels and mapping of transport channels onto physical channels (FDD)," 2001.
- [18] CDMA2000, "Physical layer standard for cdma2000 spread spectrum systems," 2001.
- [19] HIPERLAN-II, "Broadband radio access networks (bran); hipervlan type 2; physical (phy) layer," Feb 2001.
- [20] "Ansi/ieee std 802.11, part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications," 1999.
- [21] "Evolved universal terrestrial radio access (e-utra): Physical channels and modulation."
- [22] "Digital video broadcasting (dvb); framing structure, channel coding and modulation for digital terrestrial television," Jan 2001.
- [23] ATSC, "ATSC digital television standard (revision b)." Doc. A/53B, Aug 2001.
- [24] G.992.1, "International Telecommunication Union (ITU)."
- [25] C. Modem, "Ieee 802, subcommittee (802.14)."
- [26] S. D. Roberto Opez-Valcarce, Roberto Lpez-valcarce, "Blind equalization of nonlinear digital satellite links with psk modulation," in *Circuits and Systems, The 2001 IEEE International Symposium on*, 2001.
- [27] J. T. M Larimore, "Noise capture properties of the constant modulus algorithm," in *Acoustics, Speech, and Signal Processing, IEEE*, 1985.
- [28] R. G. Vincent Wolff and J. Treichler, "Specification and development of an equalizer-demodulator for wideband digital microwave radio signals," in *IEEE. Military. Communications Conference*, 1988.
- [29] J. Treichler, I. Fijalkow, and C. Johnson, "Fractionally-spaced equalizers: How long should they really be?," *IEEE Signal Processing Magazine*, vol. 13, pp. 65–81, 1996.
- [30] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 1996.
- [31] S. Haykin, ed., *Blind Deconvolution*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [32] J. Tugnait, L. Tong, and Z. Ding, "Single-user channel estimation and equalization," *IEEE Signal Processing Magazine*, vol. 17, pp. 16–28, May 2000.
- [33] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications," *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 385–399, 1980.

- [34] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications*, vol. 28, pp. 1867–1875, Nov. 1980.
- [35] Z. Ding, *Digital Signal Processing (DSP) Handbook*, ch. 17, pp. 1–24. CRC Press/IEEE Press, 1998.
- [36] Y.-M. Gu, B.-S. Seo, and C.-W. Lee, "Convergence analysis of sato blind equalizer with discrete channel input," in *Proc. IEEE International Symposium on Information Theory*, p. 355, 27 June–1 July 1994.
- [37] Z. Ding, R. Kennedy, B. Erson, and C. Johnson, "Local convergence of the sato blind equalizer and generalizations under practical constraints," *IEEE transactions on information theory*, vol. 39, pp. 129–144, 1993.
- [38] J. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 31, pp. 459–472, Apr. 1983.
- [39] Z. Ding, R. Kennedy, B. Anderson, and C. J. Johnson, "Ill-convergence of godard blind equalizers in data communication systems," *IEEE Transactions on Communications*, vol. 39, no. 9, pp. 1313–1327, 1991.
- [40] J.-C. Lin and L.-S. Lee, "A modified blind equalization technique based on a constant modulus algorithm," in *Conference Record.1998 IEEE International Conference on Communications ICC 98*, vol. 1, pp. 344–348 vol.1, 1998.
- [41] J. Yang, J.-J. Werner, and G. Dumont, "The multimodulus blind equalization and its generalized algorithms," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 997–1015, 2002.
- [42] S. Abrar, "Compact constellation algorithm for blind equalization of QAM signals," in *Proc. International Conference on Networking and Communication INCC 204*, pp. 170–174, 11–13 June 2004.
- [43] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a "stop-and-go" decision-directed algorithm," *IEEE Transactions on Communications*, vol. 35, no. 9, pp. 877–887, 1987.
- [44] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Transactions on Information Theory*, vol. 36, pp. 312–321, March 1990.
- [45] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Transactions Information Theory*, vol. 40, no. 2, pp. 340–349, 1994.
- [46] J. Tugnait and U. Gummadavelli, "Blind equalization and channel estimation with partial response input signals," *IEEE Transactions on Communications*, vol. 45, no. 9, pp. 1025–1031, 1997.
- [47] B. Vucetic and J. Yuan, *Space-Time Coding*. John Wiley & Sons Ltd, 2003.

- [48] B. R. Petersen and D. D. Falconer, "Suppression of adjacent-channel, cochannel, and intersymbol interference by equalizers and linear combiners," *IEEE Transactions on Communications*, vol. 42, pp. 3109–3118, Dec. 1994.
- [49] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2203–2214, 2006.
- [50] Z. Ding and Y. Li, *Blind Equalization and Identification*. CRC, 2001.
- [51] C.-Y. Chi, C.-C. Feng, C.-H. Chen, and C.-Y. Chen, *Blind Equalization and System Identification*. Springer, 2006.
- [52] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley & Sons, 2001.
- [53] K. Abed-Meraim and Y. Hua, "Blind identification of multi-input multi-output system using minimum noise subspace," *IEEE Transactions on Signal Processing*, vol. 45, pp. 254–258, Jan. 1997.
- [54] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. New York, John Wiley & Sons, 2002.
- [55] L. Tong, V. C. Soon, Y. F. Huang, and R. Liu, "Amuse: a new blind identification algorithm," in *Proc. IEEE International Symposium on Circuits and Systems*, pp. 1784–1787, 1–3 May 1990.
- [56] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, pp. 434–444, Feb. 1997.
- [57] D. T. Pham, "Exploiting source non stationary and coloration in blind source separation," in *Proc. 14th International Conference on Digital Signal Processing DSP 2002*, vol. 1, pp. 151–154, 2002.
- [58] A. Tanaka, H. Imai, and M. Miyakoshi, "Second-order-statistics-based blind source separation for non-stationary sources with stationary noise," in *Proc. IEEE International Symposium on Signal Processing and Information Technology*, pp. 647–650, Aug. 2006.
- [59] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Transactions on Signal Processing*, vol. 46, pp. 2888–2897, Nov. 1998.
- [60] M. P. Yannick Deville, "Temporal and time-frequency correlation-based blind source separation methods. part i: Determined and underdetermined linear instantaneous mixtures," *Signal Processing*, vol. 87, Issue 3, pp. 374–407, 2007.
- [61] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems and Computers*, pp. 856–860 vol.2, 1991.
- [62] Z. Ding and Y. Li, *Blind Equalization and Identification*. New York: Marcel Dekker, 2001.

- [63] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 694–705, 1997.
- [64] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, 1995.
- [65] J. Treichler and M. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 420–431, 1985.
- [66] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516–525, 1995.
- [67] L. Tong and S. Perreau, eds., *Multichannel Blind Identification: From Subspace to Maximum Likelihood Methods*, no. 1951–1968, Proceedings of the IEEE, October 1998 1998.
- [68] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-94*, pp. IV/585–IV/588, Volume iv,&9–22 April 1994.
- [69] Y. Hua and J. Tugnait, "Blind identifiability of FIR-MIMO systems with colored input using second order statistics," *IEEE Signal Processing Letters*, vol. 7, pp. 348–350, Dec. 2000.
- [70] C. T. Ma, Z. Ding, and S. F. Yau, "A two-stage algorithm for MIMO blind deconvolution of nonstationary colored signals," *IEEE Transactions on Signal Processing*, vol. 48, pp. 1187–1192, April 2000.
- [71] Y. Hua, S. An, and Y. Xiang, "Blind identification and equalization of FIR MIMO channels by BIDS," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 4, pp. 2157–2160, 7–11 May 2001.
- [72] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation," *IEEE Transactions on Signal Processing*, vol. 42, pp. 2158–2168, Aug. 1994.
- [73] Z. Ding and T. Nguyen, "Stationary points of a kurtosis maximization algorithm for blind signal separation and antenna beamforming," *IEEE Transactions on Signal Processing*, vol. 48, pp. 1587–1596, June 2000.
- [74] Y. Li and K. Liu, "Adaptive blind source separation and equalization for multiple-input/multiple-output systems," *IEEE Transactions on Information Theory*, vol. 44, pp. 2864–2876, Nov. 1998.
- [75] J. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria," *IEEE Transactions on Signal Processing*, vol. 45, pp. 658–672, March 1997.

- [76] M. Avriel, *Nonlinear Programming: Analysis and Methods*. Dover Publications Inc, 2003.
- [77] J. Tugnait, "Inverse filter criteria for estimation of linear parametric models using higher order statistics," in *Proc. International Conference on Acoustics, Speech, and Signal Processing ICASSP-91*, pp. 3101–3104 vol.5, 1991.
- [78] C.-Y. Chi, C.-Y. Chen, C.-H. Chen, and C.-C. Feng, "Batch processing algorithms for blind equalization using higher-order statistics," *IEEE Signal Processing Magazine*, vol. 20, pp. 25–49, Jan. 2003.
- [79] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 504–519, 1993.
- [80] M. Martone, "Non-Gaussian multivariate adaptive ar estimation using the super exponential algorithm," *IEEE Transactions on Signal Processing*, vol. 44, pp. 2640–2644, Oct. 1996.
- [81] K. L. Yeung and S. F. Yau, "A cumulant-based super-exponential algorithm for blind deconvolution of multi-input multi-output systems," *Signal Process*, vol. 67, no 2, pp. 141–162, 1998.
- [82] B. Agee, "The least-squares CMA: A new technique for rapid correction of constant modulus signals," in *Proc. IEEE International Conference on ICASSP '86. Acoustics, Speech, and Signal Processing*, vol. 11, pp. 953–956, 1986.
- [83] C. Papadias, "Globally convergent blind source separation based on a multiuser kurtosis maximization criterion," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3508–3519, 2000.
- [84] P. Sansrimahachai, D. Ward, and A. Constantinides, "Blind source separation for blast," in *Proc. 14th International Conference on Digital Signal Processing DSP 2002*, vol. 1, pp. 139–142 vol.1, 2002.
- [85] J. F. Cardoso, "Source separation using higher order moments," in *Proc. International Conference on Acoustics, Speech, and Signal Processing ICASSP-89*, pp. 2109–2112, 23–26 May 1989.
- [86] L. Tong, R. w. Liu, V. C. Soon, and Y. F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Transactions on In Circuits and Systems*, vol. 38, pp. 499–509, May 1991.
- [87] J. F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM Journal of Matrix Analysis and Applications*, vol. 17, pp. 161–164, 1996.
- [88] J. F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.
- [89] P. Comon, "Independent component analysis, a new concept ?," *Signal Processing, Elsevier*, vol. 36(3), pp. 287–314, 1994.

- [90] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13(4-5), pp. 411–430, 2000.
- [91] D. Obradovic, N. Madhu, A. Szabo, and C. S. Wong, "Independent component analysis for semi-blind signal separation in MIMO mobile frequency selective communication channels," in *Proc. IEEE International Joint Conference on Neural Networks*, vol. 1, 25–29 July 2004.
- [92] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, pp. 626–634, May 1999.
- [93] T. Inomoto, H. Shiomi, and Y. Okamura, "Implementation of the independent component analysis processing circuit for wireless communications by FPGA," in *Proc. IEEE Antennas and Propagation Society International Symposium AP-S 2008*, pp. 1–4, 5–11 July 2008.
- [94] T. Ristaniemi and J. Joutsensalo, "Advanced ICA-based receivers for DS-CDMA systems," in *Proc. 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2000*, vol. 1, pp. 276–281, 18–21 Sept. 2000.
- [95] R. Cristescu and J. K.-J. Ristaniemi, T. and Joutsensalo, "Delay estimation in CDMA communications using a fast ICA algorithm," in *International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, 2000.
- [96] I. Kostanic and W. Mikhael, "Blind source separation technique for reduction of co-channel interference," *Electronics Letters*, vol. 38, pp. 1210–1211, 26 Sept. 2002.
- [97] S. E. El-Khamy, "On the blind multi-user detection of dskdma signals using the independent component analysis," in *TWENTIEITH NATIONAL RADIO SCIENCE CONFERENCE*, March 18-20, 2003, Cairo, Egypt.
- [98] W. Leong and J. Homer, "Implementing ICA in blind multiuser detection," in *Proc. IEEE International Symposium on Communications and Information Technology ISCIT 2004*, vol. 2, pp. 947–952, 26–29 Oct. 2004.
- [99] S. Fiori, "Overview of independent component analysis technique with an application to synthetic aperture radar (sar) imagery processing," *Neural Networks*, vol. 16, p. 453467, 2003.
- [100] L. Sarperi, X. Zhu, and A. K. Nandi, "Blind OFDM receiver based on independent component analysis for multiple-input multiple-output systems," *IEEE Transactions on Wireless Personal Communications*, vol. 6, pp. 4079–4089, November 2007.
- [101] R. L. Comon P, "Blind separation of independent sources from convolutive mixtures," *IEICE Trans Fundam Electron Commun Comput Sci*, vol. E86-A, pp. 542–549, 2003.
- [102] C.-C. Feng and C.-Y. Chi, "Performance of cumulant based inverse filters for blind deconvolution," *IEEE Transactions on Signal Processing*, vol. 47, pp. 1922–1935, July 1999.

- [103] B. Chen and A. P. Petropulu, "Frequency domain blind MIMO system identification based on second and higher order statistics," *IEEE Transactions on Signal Processing*, vol. 49, pp. 1677–1688, Aug. 2001.
- [104] V. Zarzoso, "Exploiting independence for co-channel interference cancellation and symbol detection in multiuser digital communications," in *Proc. Seventh International Symposium on Signal Processing and Its Applications*, vol. 2, pp. 303–306, 1–4 July 2003.
- [105] A. Hyvarinen, "Survey on independent component analysis," tech. rep., Helsinki University of Technology, 1999.
- [106] A. Bell and T. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computing*, vol. 7, pp. 1129 – 1159, 1995.
- [107] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251–276, 1998.
- [108] J. Eriksson, A. Seppola, and V. Koivunen, "Complex ICA for circular and non- circular sources," in *MLSP, Sao Luis, Brazil*, 2004.
- [109] V. Zarzoso and A. K. Nandi, "Closed-form estimators for blind separation of sources - part ii: Complex mixtures," *Wireless Personal Communications*, vol. 21, pp. 29–48, 2002.
- [110] J. F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11. No.1, pp. 157–192, 1999.
- [111] G. A. S. C. D. Mantini, K. E. Hild, "Performance comparison of independent component analysis algorithms for fetal cardiac signal reconstruction: a study on synthetic fmcg data," *PHYSICS IN MEDICINE AND BIOLOGY*, vol. 51, pp. 1033–1046, 2006.
- [112] J. E. S. Douglas and V. Koivunen, "Fixed point complex ICA algorithms for the blind separation of sources using their real or imaginary components," in *6th International Conference on ICA and BSS*, 2006.
- [113] V. Zarzoso and A. Nandi, "Exploiting non-Gaussianity in blind identification and equalisation of MIMO FIR channels," *IEE Proceedings -Vision, Image and Signal Processing*, vol. 151, pp. 69–75, 5 Feb. 2004.
- [114] J. Eriksson, A.-M. Seppola, and V. Koivunen, "Complex ICA for circular and non-circular sources," in *Proc. 13th Eur. Signal Process Conf. (EUSIPCO)*, 2005.
- [115] A. Hyvarinen, "One-unit contrast functions for independent component analysis: a statistical analysis," in *Proc. IEEE Workshop Neural Networks for Signal Processing [1997] VII*, pp. 388–397, 1997.
- [116] L. Zhang, A. Cichocki, and S. Amari, "Self-adaptive blind source separation based on activation functions adaptation," *IEEE Transactions on Neural Network*, vol. 15, pp. 233–244, March 2004.
- [117] N. Vlassis and Y. Motomura, "Efficient source adaptivity in independent component analysis," *IEEE Transactions on Neural Network*, vol. 12, pp. 559–566, May 2001.

- [118] M. Novey and T. Adali, "Complex fixed-point ICA algorithm for separation of QAM sources using Gaussian mixture model," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2007*, vol. 2, pp. II-445-II-448, 15-20 April 2007.
- [119] H. Attias, "Independent factor analysis," *Neural Computation*, vol. 11, pp. 803-851, 1999.
- [120] E. Moulines, J. F. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-97*, vol. 5, pp. 3617-3620, 21-24 April 1997.
- [121] A. Hyvarinen and E. Oja, "Independent component analysis by general nonlinear hebbian-like learning rules," *Signal Processing*, vol. 64, no. 3, pp. 301-313, 1998.
- [122] H. Mathis, *Nonlinear functions for blind separation and equalization*. PhD thesis, Technische Wissenschaften ETH Zurich, 2001.
- [123] H. Mathis and S. Douglas, "On the existence of universal nonlinearities for blind source separation," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1007-1016, May 2002.
- [124] A. G. L. T. P. von Hoff and A. N. Kaelin, "Transpose properties in the stability and performance of the classic adaptive algorithms for blind source separation and deconvolution," *Signal Processing*, vol. 80, no. 9, pp. 1807-1822, 2000.
- [125] A. van den Bos, "Complex gradient and hessian," *IEE Proceedings -Vision, Image and Signal Processing*, vol. 141, pp. 380-383, Dec. 1994.
- [126] M. Novey and T. Adali, "Adaptable nonlinearity for complex maximization of nongaussianity and a fixed-point algorithm," in *Proc. 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, pp. 79-84, Sept. 2006.
- [127] S.-I. Amari, T. ping Chen, and A. Cichocki, "Stability analysis of adaptive blind source separation," *Neural Networks*, vol. 10, pp. 1345-1351, 1997.
- [128] J. Cardoso, "On the stability of source separation algorithms," in *Proc. IEEE Signal Processing Society Workshop Neural Networks for Signal Processing VIII*, pp. 13-22, 31 Aug.-2 Sept. 1998.
- [129] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1995.
- [130] A. Hyvarinen, "Gaussian moments for noisy independent component analysis," *IEEE Singal Processing Letter*, vol. 6, no. 6, pp. 145-147, 1999.
- [131] E. Tse and S. Pasupathy, "Blind equalization with differential detection for channels with isi and fading," *Wireless Personal Communications*, vol. Volume 10, No 2, pp. 245-269, 1999.
- [132] I. . L. S. Committee, "The ieee 802.16 working group on broadband wireless access standards." <http://grouper.ieee.org/groups/802/16/>.

- [133] O. Macchia and E. Moreau, "Adaptive unsupervised separation of discrete sources," *Signal Processing*, vol. 73, pp. 49–66, 1999.
- [134] O. Grellier and P. Comon, "Blind separation of discrete sources," *Signal Processing Letters, IEEE*, vol. 5, pp. 212–214, 1998.
- [135] K. Torkkola, "Blind signal separation in communications: making use of known signal distributions," in *Proc. IEEE DSP Workshop*, 1998.
- [136] J. C. S.C. Douglas, "Simple, robust, and memory-efficient fastica algorithms using the huber m-estimator cost function," *Journal of VLSI Signal Processing*, vol. 48, pp. 143–159, 2007.
- [137] R. P. M. Scarpiniti, D. Vigliano and A. Uncini, "Generalized splitting functions for blind separation of complex signals," *Neurocomputing, Elsevier*, vol. 71, pp. 2245–2270, 2008.
- [138] T. Clarke, "Generalization of neural networks to the complex plane," in *Proceedings of IJCNN*, vol. 2, pp. 435–440, 1990.
- [139] D. Gesbert, P. Duhamel, and S. Mayrargue, "On-line blind multichannel equalization based on mutually referenced filters," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2307–2317, Sept. 1997.
- [140] G. Giannakis and S. Halford, "Blind fractionally spaced equalization of noisy FIR channels: direct and adaptive solutions," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2277–2292, Sept. 1997.
- [141] S. Curnew and J. How, "Blind signal separation in MIMO OFDM systems using ica and fractional sampling," in *Proc. International Symposium on Signals, Systems and Electronics ISSSE '07*, pp. 67–70, July 30 2007–Aug. 2 2007.
- [142] L. Sarperi, X. Zhu, and A. K. Nandi, "Low-complexity ICA based blind multiple-input multiple-output OFDM receivers," *Neurocomputing*, vol. 69, no. 13-15, pp. 1529–1539, 2006.
- [143] A. D. N. Laird, and D. Rubink, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. Series B, 39(1), pp. 1–38, 1977.
- [144] L. Barbero and J. Thompson, "Fixing the complexity of the sphere decoder for MIMO detection," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2131–2142, 2008.
- [145] K. W. E. Peterson, "Explaining slow convergence of EM in low noise linear mixtures," tech. rep., Informatics and Mathematical Modelling, Technical University of Denmark, 2005.
- [146] O. Bermond and J. Cardoso, "Approximate likelihood for noisy mixtures," in *First International Workshop on Independent Component Analysis*, pp. 325–330, 1999.
- [147] T. C. Hesterberg, "Staggered aiken acceleration for EM," in *the American Statistical Association, Statistical Computing Section*, 2005.

- [148] K. Lange, "A quasi-newton acceleration of the EM algorithm," *Statistica Sinica*, vol. 5, pp. 1–18, 1995.
- [149] R. I. J. Mortaza Jamshidian, "Conjugate gradient acceleration of the EM algorithm," *Journal of the American Statistical Association*, vol. 88, pp. 221–228, 1993.
- [150] R. Salakhutdinov and S. Roweis, "Adaptive overrelaxed bound optimization methods," in *In Proceedings of International Conference on Machine Learning, ICML. International Conference on Machine Learning, ICML*, 2003.
- [151] H. B. Nielsen, "Ucminf: an algorithm for unconstrained nonlinear optimization," tech. rep., Technical University of Denmark, 2000.
- [152] J. R. Schott, *Matrix Analysis for Statistics*. John Wiley & Sons Inc., 1997.
- [153] Y. Miao and Y. Hua, "Fast subspace tracking and neural network learning by a novel information criterion," *IEEE Transactions on Signal Processing*, vol. 46, no. 7, pp. 1967–1979, 1998.
- [154] Z. R. Salakhutdinov, S. Roweis, "Optimization with EM and expectation-conjugate-gradient," in *Proc. 20th International Conf. On Machine Learning*, 2003.
- [155] C. F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11(1), p. 95103, 1983.
- [156] A. Belochrani and J. F. Cardoso, "Maximum likelihood source separation by the EM technique: Deterministic and stochastic implementation," in *In Proc NOLTA*, 1995.
- [157] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Mathematics of Computation*, vol. 44, pp. 463–471, 1985.
- [158] C. P. Schnorr and M. Euchner, "Lattice basis reduction: improved practical algorithms and solving subset sum problems," *Mathematical Programming*, vol. 66, pp. 181–199, 1994.
- [159] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2201–2214, 2002.
- [160] M. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2389–2402, 2003.
- [161] W. Zhao and G. Giannakis, "Sphere decoding algorithms with improved radius search," in *Proc. WCNC Wireless Communications and Networking Conference 2004 IEEE*, vol. 4, pp. 2290–2294 Vol.4, 2004.
- [162] Z. Guo and P. Nilsson, "VLSI implementation issues of lattice decoders for MIMO systems," in *Proc. International Symposium on Circuits and Systems ISCAS '04*, vol. 4, pp. IV-477–80 Vol.4, 2004.
- [163] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*, 1998.

- [164] S. Min and C. Nikias, "An ML/MMSE estimation approach to blind equalization," in *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP*, vol. 4, pp. 569–572, April 1994.
- [165] C.-D. Chung, "Differentially amplitude and phase-encoded QAM for the correlated rayleigh-fading channel with diversity reception," *IEEE Transactions on Communications*, vol. 45, no. 3, pp. 309–321, 1997.
- [166] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: principles and new results," *IEEE Transactions on Communications*, vol. 50, pp. 754–767, May 2002.
- [167] J. Garcia-Frias and J. Villaseñor, "Blind turbo decoding and equalization," in *Proc. IEEE 49th Vehicular Technology Conference*, vol. 3, pp. 1881–1885, 16–20 May 1999.
- [168] P. Ha and B. Honary, "Improved blind turbo detector," in *Proc. IEEE 51st VTC 2000-Spring Tokyo Vehicular Technology*, vol. 2, pp. 1196–1199, 15–18 May 2000.
- [169] J. Mannerkoski and V. Koivunen, "Autocorrelation properties of channel encoded sequences-applicability to blind equalization," *IEEE Transactions on Signal Processing*, vol. 48, pp. 3501–3507, Dec. 2000.
- [170] J. Garcia-Frias and J. Villaseñor, "Combined turbo detection and decoding for unknown ISI channels," *IEEE Transactions on Communications*, vol. 51, no. 1, pp. 79–85, 2003.
- [171] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *IEEE Proceedings*, vol. 77, no. 2, pp. 257–286, 1989.
- [172] R. Lopes and J. Barry, "Exploiting error-control coding in blind channel estimation," in *Proc. IEEE Global Telecommunications Conference GLOBECOM '01*, vol. 2, pp. 1317–1321 vol.2, 2001.
- [173] R. Lopes and J. Barry, "Blind iterative channel identification and equalization," in *Proc. IEEE International Conference on Communications ICC 2001*, vol. 7, pp. 2256–2260 vol.7, 2001.
- [174] J. Gunther, D. Keller, and T. Moon, "A generalized BCJR algorithm and its use in iterative blind channel identification," *IEEE Signal Processing Letters*, vol. 14, pp. 661–664, Oct. 2007.
- [175] J. Gunther, M. Annapure, and T. Moon, "A generalized LDPC decoder for blind turbo equalization," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3847–3856, 2005.
- [176] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing: a general approach to MLSE in uncertain environments," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 354–364, 1995.
- [177] H. N. B. Levy, "Blind and semi-blind equalization of CPM signals with the EMV algorithm," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2650–2664, Oct. 2003.
- [178] X. Z. M. Davies, "A feasible blind equalization scheme in large constellation MIMO systems," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2008*, pp. 1845–1848, March 31 2008–April 4 2008.

- [179] F. Sanzi, S. Jelting, and J. Speidel, "A comparative study of iterative channel estimators for mobile OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 5, pp. 849–859, 2003.
- [180] H. Liu, *Signal Processing Applications in CDMA Communications*. Norwood, MA: Artech House, 2000.
- [181] M. El-Khamy, H. Vikalo, and B. Hassibi, "Bounds on the performance of sphere decoding of linear block codes," in *Proc. IEEE Information Theory Workshop*, p. 5pp., 29 Aug.–1 Sept. 2005.
- [182] H. Vikalo and B. Hassibi, "On joint detection and decoding of linear block codes on Gaussian vector channels," *IEEE Transactions on Signal Processing*, vol. 54, pp. 3330–3342, Sept. 2006.
- [183] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (corresp.)," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, Mar 1974.
- [184] K. Sooyoung, H.-M. Kim, and J. Gwak, "Reduced complexity sliding window BCJR decoding algorithms for turbo codesldpc," in *Lecture Notes in Computer Science Proceedings of the 7th IMA International Conference on Cryptography and Coding*, 1999.
- [185] B. Hochwald and S. Ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 389–399, 2003.
- [186] Y. de Jong, T. J. Willink, "Iterative tree search detection for MIMO wireless systems," *IEEE Transactions on Communications*, vol. 53, pp. 930–935, June 2005.
- [187] D. L. Milliner, E. Zimmermann, J. R. Barry, and G. Fettweis, "Channel state information based LLR clipping in list MIMO detection," in *Proc. IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2008*, pp. 1–5, 15–18 Sept. 2008.
- [188] F. Schreckenbach, N. Gortz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Communications Letters*, vol. 7, no. 12, pp. 593–595, 2003.
- [189] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Transactions on Communications*, vol. 42, pp. 2406–2413, July 1994.
- [190] L. White, S. Perreau, and P. Duhamel, "Reduced computation blind equalization for FIR channel input Markov models," in *Proc. IEEE International Conference on Communications ICC 95 Seattle, Gateway to Globalization*, vol. 2, pp. 993–997, 18–22 June 1995.
- [191] H. Cirpan and M. Tsatsanis, "Stochastic maximum likelihood methods for semi-blind channel equalization," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Computers*, vol. 2, pp. 1629–1632, 2–5 Nov. 1997.
- [192] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, pp. 47–60, Nov. 1996.

- [193] G. A.-R. T. A.-N. A. Bahai, and J. Cioffi, "Exploiting error-control coding and cyclic-prefix in channel estimation for coded OFDM systems," *IEEE Communications Letters*, vol. 7, no. 8, pp. 388–390, 2003.
- [194] A. Edelman, *Eigenvalues and Condition Numbers of Random Matrices*. PhD thesis, Massachusetts Institute of Technology, Cambridge (MA), 1989.
- [195] K. S. I. B. I. Wassell, and X. Wang, "Efficient maximum-likelihood decoding of spherical lattice space-time codes," in *IEEE International Conference on Communications ICC '06*, vol. 7, pp. 3008–3013, 2006.
- [196] L. G. Barbero and J. S. Thompson, "Extending a fixed-complexity sphere decoder to obtain likelihood information for turbo-MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 57s, pp. 2804 – 2814, 2008.
- [197] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Transactions on Speech Audio Process*, vol. 12, pp. 530–538, Sept. 2004.
- [198] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, pp. 101–104, 16–18 April 1997.
- [199] H. Attias and C. E. Schreiner, "Blind source separation and deconvolution by dynamic component analysis," in *Proc. IEEE Workshop Neural Networks for Signal Processing [1997] VII*, pp. 456–465, 24–26 Sept. 1997.
- [200] B. H. Tadeusz A. Wysocki and B. J. Wysocki, *Time Domain Blind Separation of Non-stationary Convolutively Mixed Signals*, vol. 27 of *Multimedia Systems and Applications Series*. Springer US, 2005.
- [201] Z. Ding, T. Ratnarajah, and C. F. N. Cowan, "Hos-based semi-blind spatial equalization for MIMO rayleigh fading channels," *IEEE Transactions on Signal processing*, vol. 56, pp. 248–255, Jan. 2008.