

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
MAY 31 2005		
347		

**Blind Source Recovery: Theoretical Formulations,
Implementations and Application to CDMA
Communication Systems**

By

Khurram Waheed

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering

2003

Blind

Blind Sou

approach t

include co

is to recov

environme

adopted b

constraint

adaptation

Th

framework

general u

interestin

laws for

have bee

algorithm

undercor

BSR alg

ABSTRACT

Blind Source Recovery: Theoretical Formulations, Implementations and Application to CDMA Communication Systems

By

Khurram Waheed

Blind Source Recovery (BSR) is an autonomous (or unsupervised) stochastic adaptation approach that denotes recovering signals from measurements in environments that may include convolution, transients, and even possible nonlinearity. The primary goal of BSR is to recover original source signals, as best as possible, even in the absence of precise environment identifiability. A discrete-time optimization framework for BSR has been adopted based on minimization of the Kullback-Lieblar Divergence, subject to the constraints of a state-space representation, using the Riemannian contra-variant gradient adaptation.

The modeling of the environment and its representation is vital in the proposed framework. The adoption of the state space framework allows for the derivation of more general update laws capable of catering to most known filtering paradigms. For the interesting case of multi-variable linear time-invariant dynamic BSR, parametric update laws for discrete-time canonical feedforward and feedback state space configurations have been derived. Higher Order Statistics have been explored to obtain adaptation algorithms that are purely blind of actual source distributions. The BSR cases of undercomplete and overcomplete mixtures have also been investigated. A new two-stage BSR algebraic algorithm for sparse static overcomplete mixtures has been developed.

Further. t

modern C

demonstr.

C

research t

scenarios.

and comm

this thesis

- Info-th

investi

phase

feedba

- An ex

of ada

source

- The c

algori

overc

of AI

- Deve

CDM

frame

Further, the BSR framework has been successfully formulated for multi-user detection in modern CDMA wireless communication networks. Promising new results clearly demonstrate the effectiveness and practicality of the formulated approach.

Computer Simulations have been widely conducted during the course of this research to evaluate the performance of all the developed algorithms for a variety of scenarios. In most cases, the performance verification has been done using actual speech and communication data in the presence of noise. In summary the main contributions of this thesis are

- Info-theoretic update laws for a multi-variable dynamic state space network were investigated. Natural gradient based BSR update laws for the cases of minimum phase and non-minimum phase mixing environments as well as both feedforward and feedback canonical state space configurations were developed and implemented.
- An exploration and extension of parametric source distribution models and derivation of adaptive score-functions (or non-linearities) for the BSR of sources with multiple source distributions.
- The development of the Algebraic ICA (AICA) Algorithm, which is a new ICA algorithm for blind, sparse, static mixing matrix recovery. It enables the BSR from overcomplete mixtures of speech and other sparse distributions using a combination of AICA and interior point linear programming (IP-LP) techniques.
- Development and simulation of new Blind Multi-user Detection algorithms for DS-CDMA and WCDMA wireless communication networks based on the BSR framework.

Dedicated to the most important people in my life

My parents, my wife and family

I would i

for his g.

his value

stimulatio

Sir

Professor

Professor

Departmen

my career

I w

encourage

individua

M

affection,

completi

enriched

L

sisters, in

course of

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to Professor Fathi M. Salem for his guidance, support and inspiration throughout the course of this research. Without his valuable, kind consideration, understanding, firm belief in my capabilities and stimulation of my mind this dissertation would not be an object of my pride.

Sincere gratitude is extended to the members of my guidance committee: Professor Habib Salehi from the Department of Statistics, Professor Hassan Khalil, Professor John R. Deller Jr., Professor Ning Xi and Professor Tongtong Li from the Department of Electrical and Computer Engineering for their valuable contributions to my career and service on my committee.

I would not be who I am today without the unconditional love, support and encouragement of my parents Abdul Waheed and Najma. They have shaped me into the individual I am and inspired me to pursue my dreams in life.

My sincere thanks and gratitude to my wife Ayesha for her patience, love affection, support, encouragement and sacrifice during every step that I made towards the completion of this research. My love and adoration to my kids Rafay and Iqra, who have enriched my life but have to spend most of their time without me.

Last, but not least, I would like to express my sincere thanks to my brother, sisters, in-laws and their family members for their spiritual and moral support during the course of this research.

LIST O

LIST O

Chapter

INTRO

1.1

1.2

1.3

1.4

1.4.1

1.5

Chapter

REVIEW

2.1

2.1.1

2.1.2

2.1.3

2.2

2.2.1

TABLE OF CONTENTS

LIST OF TABLES	xiv
LIST OF FIGURES	xv

Chapter 1

INTRODUCTION	1
1.1 Background	3
1.2 Importance of Modeling	4
1.3 The State Space Framework	4
1.4 Problem Definition: Adaptive Framework for Blind Source Recovery	6
1.4.1 Blind Source Separation: An Example	8
1.5 Organization of the dissertation	11

Chapter 2

REVIEW OF LITERATURE.....	14
2.1 The Natural Gradient Framework.....	18
2.1.1 Definition of the Natural Gradient.....	19
2.1.2 Theorem 2.1	20
2.1.3 The Natural Gradient Learning.....	21
2.2 Blind Source Separation.....	22
2.2.1 Theorem 2.2	23

2.3

2.3.

2.4

Chapter

**BLIND
SPACE**

3.1

3.1.1

3.1.2

3.2

3.2.1

3.3

3.3.1

3.3

3.3

3

3

3.3.2

3.3

3.3

3

2.3	Blind Source Deconvolution.....	24
2.3.1	Theorem 2.3	24
2.4	Equivariant Performance of the Natural Gradient	25

Chapter 3

BLIND SOURCE RECOVERY: A FRAMEWORK IN THE STATE SPACE

3.1	Formulation of the Optimization Structure	27
3.1.1	Algorithms for the Nonlinear Dynamic Case	28
3.1.2	Algorithm for Nonlinear Dynamic Canonical Neural Structure.....	31
3.2	Algorithms for the Linear Dynamic Case.....	34
3.2.1	Extensions to the Natural Gradient	37
3.3	Blind Source Recovery Algorithms for Minimum Phase Environments... 38	
3.3.1	Theorem 3.1: Feedforward Minimum Phase Demixing Network	38
3.3.1.1	Derivation of Update Laws for Feedforward BSR	39
3.3.1.2	Auxiliary Conditions for Convergence.....	41
3.3.1.2.1	Remark 3.1:.....	41
3.3.1.2.2	Remark 3.2:.....	42
3.3.2	Theorem 3.2: Feedback Minimum Phase Demixing	43
3.3.2.1	Derivation of Update Laws for Feedback BSR	43
3.3.2.2	Auxiliary Conditions for Convergence.....	46
3.3.2.2.1	Remark 3.3:.....	46

3.4

Enviro

3.4.

3.

B

3.5

3.5.1

3.5

3.5

3.5

Mi

3.6

Chapter

BLIND

FUNCTI

4.1

4.1.1

4.1.2

4.1.3

4.2

4.2.1

3.4	Blind Source Recovery Algorithms for Non-minimum Phase Environments	47
3.4.1	Theorem 3.3: Feedforward Non-minimum Phase Demixing	47
3.4.1.1	Derivation of Update Laws for Feedforward Non-minimum Phase BSR	48
3.4.1.1.1	Remark 3.4:.....	55
3.5	Simulation Results	56
3.5.1	BSR Simulations -- Minimum Phase Mixing Environments.....	58
3.5.1.1	Simulation I: Feedforward Recovery of Minimum Phase Mixing	61
3.5.1.2	Simulation II: Feedback Recovery of Minimum Phase Mixing	65
3.5.1.3	Simulation III: Feedforward Recovery of Non-minimum Phase Mixing	69
3.6	Remarks.....	74

Chapter 4

BLIND SOURCE RECOVERY USING ADAPTIVE SCORE FUNCTIONS.....75

4.1	Generalized Gaussian Distribution Model	77
4.1.1	Generalized Kurtosis Measures	79
4.1.2	Relationship between $\beta_i (\sigma_i, \alpha_i)$ and α_i	81
4.1.3	Derivation of Generalized Gaussian Adaptive Score Function (GGASF)	82
4.2	Hyperbolic Source Density Models	84
4.2.1	Background.....	84

4.2.1.1	Pearson Distribution Model	84
4.2.1.2	Generalized Cauchy Distribution.....	86
4.2.2	Proposed Hyperbolic Distribution Models	87
4.2.2.1	Hyperbolic Sub-Gaussian Model.....	87
4.2.2.2	Hyperbolic Super-Gaussian Model.....	89
4.2.3	Derivation of Hyperbolic Adaptive Score Function (HASF)	91
4.3	Adaptive Estimation of the Fisher (or normalized) Kurtosis	94
4.4	Parameter Regulation for the Proposed Adaptive Score Functions	95
4.4.1	Choice of parameters for Generalized Gaussian Adaptive Score Function	96
4.4.2	Choice of parameters for Hyperbolic Adaptive Score Function.....	96
4.5	Simulation Results	97
4.5.1	Environment Model	98
4.5.2	Demixing Network Structure.....	98
4.6	Remarks	103

Chapter 5

BLIND SOURCE RECOVERY OF UNDERCOMPLETE AND OVERCOMPLETE MIXTURES

5.1	BSR of Undercomplete Mixtures.....	107
5.1.1	Efficient reduction of Mixture Dimensions	108
5.1.1.1	Algebraic Principal Component Analysis.....	108
5.1.1.2	Adaptive Principal Component Analysis.....	109

5.2

5.2

5.3

5.3

5.3

5.3

5.3

5.3

5.3

5.3

5.3

5.3

5.3

5.3

5.4

Chapte

BLIND
CDMA

6.1

6.2

6.3

6.3

6.3

5.2	BSR of Overcomplete Mixtures	110
5.2.1	Two Stage Approach to Overcomplete ICA.....	110
5.3	Algebraic Independent Component Analysis	112
5.3.1	The Algebraic ICA Algorithm.....	116
5.3.2	Algebraic Matrix-Distance Index (AMDI).....	120
5.3.3	Resolution of AICA Local Minima and Ambiguities.....	121
5.3.4	Inferring Sources from Overcomplete Mixtures.....	123
5.3.5	Simulation Examples	124
5.3.5.1	Simulation I: 4×2 Laplacian Data	125
5.3.5.2	Simulation II: Overcomplete Speech Data (3×2 Case).....	127
5.3.5.3	Simulation III: 5×3 Speech Data.....	130
5.3.5.4	Simulation IV: Independent Identically Distributed (iid) Laplacian Data	134
5.4	Remarks	136

Chapter 6

	BLIND SOURCE RECOVERY APPROACHES FOR WIRELESS DS- CDMA AND WCDMA DOWNLINK SYSTEMS	138
6.1	User Detection in CDMA Systems	139
6.2	CDMA Downlink Communication Channel	143
6.3	Downlink Signal Models for CDMA Systems	146
6.3.1	DS-CDMA Signal Model	147
6.3.2	WCDMA (UMTS FDD) Signal Model	151

6.4	CDMA User Detection Structures.....	154
6.4.1	Conventional Detection Schemes	154
6.4.1.1	Matched Filter (MF) Detector.....	155
6.4.1.2	RAKE Detector.....	156
6.4.1.3	Linear Minimum Mean Squared Error (LMMSE) Detector.....	157
6.4.1.4	Channel Estimation Using The Common Pilot Channel (CPICH).....	158
6.4.1.4.1	Channel Estimation Procedure.....	159
6.4.2	Proposed Blind Source Recovery Detection Schemes.....	160
6.4.2.1	Natural Gradient Blind Multi-User Detection (BMUD) Algorithms.	160
6.4.2.1.1	Feedforward BMUD Configuration:.....	163
6.4.2.1.2	Feedback BMUD Configuration I:	164
6.4.2.1.3	Feedback BMUD Configuration II:	165
6.4.2.2	RAKE-Blind Source Recovery (RAKE-BSR) and RAKE-Principal Component Analysis (RAKE-PCA) Detectors.....	167
6.5	Simulation Results	169
6.5.1	Simulation I: DS-CDMA System	170
6.5.2	Simulation II: WCDMA (UMTS FDD) System.....	173
6.6	Remarks.....	176

Chapter 7

CONCLUSIONS.....	178
-------------------------	------------

APPENDIX

Appendix A

List of

Appendix B

**Kullback
Source**

Appendix C

State Space

C.1.

C.2.

Appendix D

Performance

Appendix D.1

D.1.1.

D.1.2.

D.1.3.

D.1.4.

D.1.5.

D.1.6.

APPENDICES

Appendix A

List of Abbreviations and Acronyms	184
---	------------

Appendix B

Kullback Lieblar Divergence: A Performance Functional for Blind Source Recovery.....	187
---	------------

Appendix C

State Space Network: Specialized Filtering Structures	190
C.1. IIR Filtering.....	190
C.2. FIR Filtering.....	193

Appendix D

Performance Benchmarks for Blind Source Recovery	195
Appendix D.1. Conventional Performance Benchmarks.....	195
D.1.1. Signal to Noise Ratio (SNR).....	195
D.1.2. Multichannel Intersymbol Interference (MISI)	196
D.1.3. Maximum signal to interference plus noise ratio (SINRM)	197
D.1.4. Mean Squared Error (MSE).....	198
D.1.5. Distortion Measure.....	198
D.1.6. Separation Measure.....	199

Appen

D.2

D.2

D

D

D.2

D.2

D.2

Append

Append

Multi-I

E.1.

BIBL

Appendix D.2. Quadratic Information Measure (QIM)	200
D.2.1. Practical Issues in Estimating Independence of Output Sequences.....	200
D.2.2. Independence Benchmark using Quadratic Mutual Information.....	202
D.2.2.1. Renyi’s Entropy	202
D.2.2.2. Parzen Window Estimator	203
D.2.3. Quadratic Independence Measure (QIM)	204
D.2.4. Simulations Results.....	209
D.2.5. Observations	211
Appendix D.3. Algebraic Matrix Distance Index (AMDI)	212
Appendix E	
Multi-Input Multi-Output (MIMO) Canonical Form	214
E.1. Transformations required for a MIMO Transfer Function	215
BIBLIOGRAPHY	219

LIST OF TABLES

Table 3.1. Quantitative Signal Recovery Comparison of the proposed algorithms..... 73

Figure 1

Figure 1

Figure 2
Function

Figure 3

Figure 3.1

Figure 3.2

Figure 3.3
and Poles

Figure 3.5
Environment

Figure 3.6

Figure 3.7
Demixing

Figure 3.8
Network T

Figure 3.9
Demixing

Figure 3.10
Demixing

Figure 3.11
Network T

Figure 3.12
Demixing

Figure 3.13
Zeros and P

LIST OF FIGURES

Figure 1.1. General framework for Blind Source Recovery.....	6
Figure 1.2. Blind Source Separation (a) Original Sources (b) Observed Mixtures	9
Figure 1.3. Blind Source Separation (a) Separated Signals (b) Global Transfer Function	10
Figure 3.1. State Space Mixing Environment.....	34
Figure 3.2. State Space Demixing Network.....	35
Figure 3.3. State Space Feedback Demixing Structure.....	43
Figure 3.4. Minimum Phase Mixing Environment (a) Environment Transmission Zeros and Poles, (B) Impulse Response of the Environment Model	59
Figure 3.5. Minimum Phase Mixing Environment, Impulse Response of (a) Theoretical Environment Inverse, (b) Estimated Demixing Network	60
Figure 3.6. Minimum Phase Mixing Environment, Final Global Transfer Function	61
Figure 3.7. Minimum Phase Feedforward BSR – All poles at the theoretical solution: (a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index	62
Figure 3.8. Minimum Phase Feedforward BSR – All poles set to zero: (a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index	63
Figure 3.9. Minimum Phase Feedforward BSR – All poles initialized arbitrarily: (a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index	64
Figure 3.10. Minimum Phase Feedback BSR – All poles at the theoretical solution: (a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index	66
Figure 3.11. Minimum Phase Feedback BSR – All poles set at zero: (a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index	67
Figure 3.12. Minimum Phase Feedback BSR – All poles initialized arbitrarily: (a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index	68
Figure 3.13. Non-minimum Phase Mixing Environment (a) Environment Transmission Zeros and Poles, (b) Impulse Response of the Environment Model.....	70

Figure
Theoret

Figure
Function

Figure
Performa

Figure 4.

Figure 4.

Figure 4
parameter

Figure 4
function

Figure 4
range of

Figure 4
range of

Figure 4
(b) hype

Figure 4
.....

Figure 4
Inverse.

Figure 4
Network

Figure
Batch K

Figure
Index, (

Figure
Functio

Figure 3.14. Non-minimum Phase Mixing Environment, Impulse Response of (a) Theoretical environment Inverse, (b) Estimated Demixing Network.....	71
Figure 3.15. Non-minimum Phase Mixing Environment, (a) Final Global Transfer Function, (b) Demixing Network Transmission Zeros and Poles.....	72
Figure 3.16. Non-minimum Phase Mixing Environment, Convergence of MISI Performance Index.....	73
Figure 4.1. Some members of Generalized Gaussian Distribution Family	78
Figure 4.2. Fisher kurtosis Characteristics for the Generalized Gaussian Family.....	81
Figure 4.3. Scaled Generalized Gaussian Score Function for different values of parameter α	83
Figure 4.4. Fisher Kurtosis Characteristics for the symmetric Pearson Model as a function of μ	85
Figure 4.5. (a) Proposed Symmetric Sub-gaussian Density Model, (b) Fisher Kurtosis range of the proposed model; for $0 \leq k \leq 3$ and $\lambda_i = 1$	88
Figure 4.6. (a) Proposed Symmetric Super-gaussian Density Model, (b) Fisher Kurtosis range of the proposed model; for $0 \leq k \leq 3$; $\lambda_i = 1$	90
Figure 4.7. Score Functions based on the proposed: (a) hyperbolic sub-gaussian model, (b) hyperbolic super-gaussian model; for $0 \leq k \leq 3$; $\lambda_i = 1$	92
Figure 4.8. Proposed Adaptive Hyperbolic Score Function: some possible combinations.	94
Figure 4.9. (a) Mixing Environment Transfer Function, (b) Theoretical Environment Inverse.....	99
Figure 4.10. Feedforward BSR using Adaptive Score Functions: (a) Estimated Demixing Network Transfer Function, (b) Convergence of MISI Index	100
Figure 4.11. Feedforward BSR using Adaptive Score Functions: (b) Convergence of Batch Kurtosis, (c) Final Global Transfer Function	101
Figure 4.12. Feedback BSR using Adaptive Score Functions: (a) Convergence of MISI Index, (b) Convergence of Batch Kurtosis	102
Figure 4.13. Feedback BSR using Adaptive Score Functions: Final Global Transfer Function	103

Figure 5
space.....

Figure 5
initial w

Figure 5
Converg

Figure 5
Mixed sp

Figure 5
Converge

Figure 5.6

Figure 5.7

Figure 5.8
.....

Figure 5.9
 \hat{H}^2

Figure 6.1

Figure 6.2

Figure 6.3

Figure 6.4

Figure 6.5

Figure 6.6

Figure 6.7.
G=63, (b) u

Figure 6.8.
with Perfect
Estimate.....

Figure 5.1. A scatter plot example of the projection of two sources onto the 2×2 mixture space.....	115
Figure 5.2. Convergence of the weights in the 2×2 mixture space; $w_i(0)$ denotes the initial weight locations while $w_i(\infty)$ denotes the final weight location.	119
Figure 5.3. BMMR for Laplacian Data using AICA: (a) Interim Results, (b) Final Convergence	126
Figure 5.4. Overcomplete Speech Mixture (3×2 Case): (a) Original speech sources, (b) Mixed speech	128
Figure 5.5. Overcomplete Speech Mixture (3×2 Case): (a) Recovered speech, (b) Convergence of Normalized AMDI over 100 experiments.....	129
Figure 5.6. Overcomplete Speech Mixture (5×3 Case): 3 Overcomplete Mixtures	131
Figure 5.7. Overcomplete Speech Mixture (5×3 Case): Original 5 Speech Utterances	132
Figure 5.8. Overcomplete Speech Mixture (5×3 Case): Recovered 5 Speech Utterances	133
Figure 5.9. Unambiguous AICA estimate using multiple runs: (a) using $\tilde{H}1$, (b) using $\tilde{H}2$	135
Figure 6.1. Signal Generation Model for a QPSK DS-CDMA system	147
Figure 6.2. Signal generation based on the proposed 3GPP UMTS FDD standard	152
Figure 6.3. Channel Estimation Using CPICH	160
Figure 6.4. Feedforward Demixing Structure	163
Figure 6.5. Feedback Demixing Structure I.....	164
Figure 6.6. Feedforward Demixing Structure II	165
Figure 6.7. DS-CDMA downlink performance for $K=20$, 50 users (a) using Gold Codes $G=63$, (b) using OVSF Codes $G=64$	172
Figure 6.8. WCDMA Downlink System for $K=20$, 50 users (a) Performance Comparison with Perfect Channel Estimation, (b) Performance Comparison with Imperfect Channel Estimate.....	174

Figure 4
with Im

Figure 1
(b) Qua

Figure 1
of QIM

Figure 6.9. WCDMA Downlink System for K=20, 50 users: Performance Comparison with Imperfect Channel Estimate and Inter-Cellular Interference.....	175
Figure D.1. Effect of length of observation sequence on (a) Cross Information Measure, (b) Quadratic Independence Measure (QIM).....	208
Figure D.2. (a) Convergence of MISI performance Index, (b) Corresponding Performance of QIM performance Index	210

Blind Sou

recover is

source pro

environme

info-theor

of the env

optimal au

Blind

in the linea

separation a

of recoverin

their *linear*

deconvolvin

a priori info

[33, 54]. Th

Chapter 1

INTRODUCTION

Blind Source Recovery (BSR) is an unsupervised adaptation technique to stochastically recover independent original source signals from corrupted measurements caused by source propagation through a noisy, linear or nonlinear, dynamic mixing and filtering environment. The measurements from this environment are the only inputs to an adaptive info-theoretic BSR network, which seeks to counteract, to the extent possible, the effects of the environment [34, 64, 139, 140]. Eventually, the purpose of BSR is to develop an optimal autonomous blind estimate of original signals [8, 9, 53, 110].

Blind Source Recovery (BSR) is a challenging signal processing formulation that in the linear case includes the well-known adaptive filtering sub-problems of blind source separation and blind source deconvolution. Blind Source Separation (BSS) is the process of recovering a number of original stochastically independent sources/signals when only their *linear static* mixtures are available. Blind Source Deconvolution (BSD) deals with deconvolving the effects of a temporal or a spatial mixing *linear* filter on signals without a priori information about the filtering medium/network or the original sources/signals [33, 54]. This dissertation presents a generalized framework based on the multi-variable

canonic

system.

from the

have als

F

independ

signals s

learning

world ap

T

analyses a

approach

77], we p

representa

The fram

minimum

Separate u

and decon

[112, 140]

derived fro

mixtures o

undercompl

solution, ca

canonical state space representation of both the mixing environment and the demixing system. Various filtering paradigms have been consequently derived as special cases from the proposed framework. Simulation results verifying the theoretical developments have also been included in this work [109, 110, 126].

For this unsupervised adaptive filtering task, only the property of signal independence is assumed. No additional a priori knowledge of the original signals/sources is assumed. BSR requires few assumptions and possesses the self-learning capability, which render such networks attractive from the viewpoint of real world applications where on-line training is often desired.

The challenges for the BSR reside in the development of sound mathematical analyses and a framework capable of handling a variety of diverse problems. While other approaches to BSR have used more conventional signal processing structures [33, 60, 77], we propose a comprehensive BSR framework based on multi-variable state-space representations, optimization theory, the calculus of variations and higher order statistics. The framework is utilized to develop natural gradient adaptation algorithms for both minimum phase and non-minimum phase environments [110, 112, 126, 129, 135, 139]. Separate update laws for the feedforward and feedback recovering (i.e., both demixing and deconvolving) network parameters have been derived and verified by simulations [112, 140]. Further, algorithms have been extended using adaptive score functions derived from comprehensive source distribution models for unmitigated recovery from mixtures of multiple source distributions [127, 130, 136, 138]. The problems of undercomplete and overcomplete BSR have also been extensively studied. A new solution, called Algebraic Independent Component Analysis (AICA) for the case of

static. C
propose
commu
problem
convent

1.1 B

Interest i
during re
separatio
parallel a
ability to
stimuli an
transmitted
informatio
informatio
further pro

BSI

systems, s
enhanceme
example, th
mixed sign
vantage poi
that usually

static, overcomplete mixtures of sparse sources has been proposed [132-134, 137]. The proposed adaptive BSR framework is also applied to the modern CDMA wireless communication networks (such as DS-CDMA, WCDMA etc.) to efficiently address the problem of multi-user detection in downlink channel with favorable results compared to conventional techniques [122-125].

1.1 Background

Interest in the field of blind source separation and deconvolution has grown dramatically during recent years, motivated to a large extent by its similarity to the mixed signal separation capability of the human brain. The brain makes use of unknown nonlinear parallel and complex dynamic signal processing with auto-learning and self-organization ability to perform such tasks. The peripheral nervous system integrates complex external stimuli and endogenous information into packets, which are transformed, filtered and transmitted in a manner that is yet to be completely understood. This complex mixture of information is received by the central nervous system (brain), split again into original information, and relevant information relayed to various sections of cerebral cortex for further processing and action [89].

BSR is valuable in numerous applications that include telecommunication systems, sonar, acoustics and radar systems, mining and oil drilling systems, image enhancement, feature extraction and biomedical signal processing. Consider, for example, the audio and sonar applications where the original signals are sounds, and the mixed signals are the output of several microphones or sensors placed at different vantage points. A network would receive, via each sensor, a mixture of original sounds that usually undergo multi-path delays. The network's role in this scenario will be to

dynam.

signals

applica

cardiac

engine c

1.2 In

Develop

multi-pat

practical

highly s

network

variation

propagat

even nor

1.3 TI

The state

variety

computa

This state

laws for

input sig

are sever

dynamically reproduce, as closely as possible, the original signals. These separated signals can subsequently be channeled for further processing or transmission. Similar application scenarios can be described in situations involving measurement of neural, cardiac or other vital biological parameters, communication in noisy environments, engine or plant diagnostics, and cellular mobile communications.

1.2 Importance of Modeling

Development of adequate models of the environment that include time delays or filtering, multi-path effects, time-varying parameters or sensor dynamics is important in desired practical applications. The choice of inadequate models of the environment may result in highly sensitive and un-robust processing by a network. Indeed, in order to render the network operable in real world scenarios, robust operations must account for parameter variations, dynamic influences and signal delays that often result in asynchronous signal propagation. The environment needs to be modeled as an appropriate dynamic linear (or even nonlinear) system [53, 86, 151].

1.3 The State Space Framework

The state space formulation provides a general framework capable of dealing with a variety of situations. The multi-variable state space provides a compact and computationally attractive representation of multi-input multi-output (MIMO) filters. This state space representation allows for the derivation of generalized iterative update laws for the BSR. The state notion abridges weighted past as well as filtered versions of input signals and can be easily extended to include non-linear networks [17, 103]. There are several reasons for choosing this framework.

- Tra
con
real
- Stat
choi
bein
- The
inver
case
inver
the cr
- Par
partic
finite
- All c
Gamr
- The l
non-l
gener

- Transfer function models, although equivalent to the state space models when initial conditions are zero, do not exploit any common features that may be present in the real dynamic systems.
- State space models give an efficient internal description of a system. Further, this choice allows various equivalent state space realizations for a system, more important being the canonical observable and controllable forms.
- The inverse for a state space representation is easily derived subject to the *invertibility* of the instantaneous relational mixing matrix between input-output – in case this matrix is not square; the condition reduces to the existence of the pseudo-inverse of this matrix. This feature ensures *recoverability* of original sources provided the environment model is invertible.
- Parameterization methods are well known for specific classes of models. In particular, the state space model enables much more general description than standard finite/infinite impulse response (FIR/IIR) convolutive filtering.
- All conventional (dynamic) filtering models, like AR, MA, ARMA, ARMAX and Gamma filters, can be considered as special cases of flexible state space models.
- The linear state space formulation of a problem is conveniently extendable to include non-linear and time-varying models. The state space models are ideal candidates for generalized non-linear model representations that include, e.g., neural networks.

In the r

unknow

outputs

about th

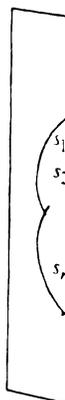
T

unknown

to comp

without a

of the un



1.4 Problem Definition: Adaptive Framework for Blind Source Recovery

In the most general setting, the mixing/convolving environment may be represented by an unknown dynamic process \bar{H} with inputs being the n -d independent sources \underline{s} and the outputs being the m -d measurements \underline{m} . In this extreme case, no structure is assumed about the model of the environment.

The environment may be also modeled as a dynamic system with fixed but unknown parameters. The processing network H must be constructed with the capability to compute the “inverse” (or the “closest to an inverse”) of the environment model without assuming any knowledge of the mixing environment or the distribution structure of the unknown sources.

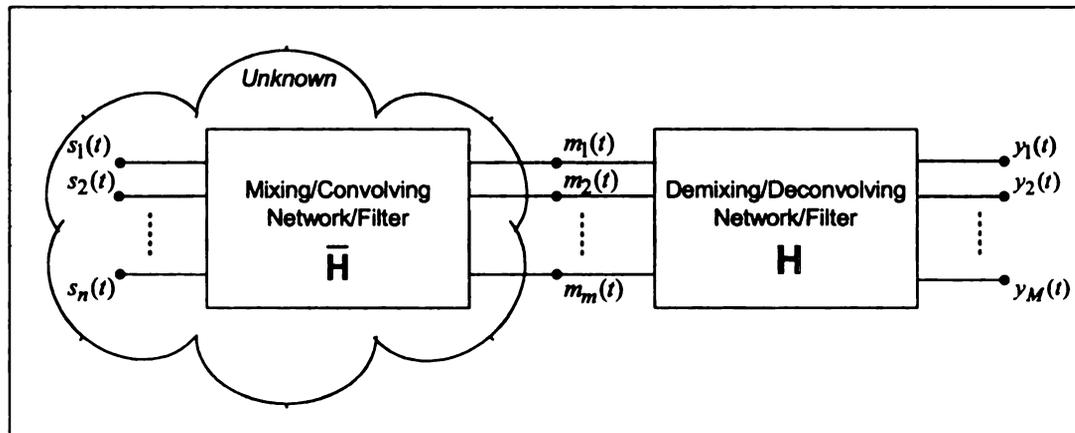


Figure 1.1. General framework for Blind Source Recovery

environ

environ

state sp

dynamic

equilibri

F

Y

where

P - is a g

D - is a d

* - repr

represents

Po

stem from

impossible

mixtures [4

- For

- For

dim

- For

in the

It is possible that an augmented network be constructed so that the inverse of the environment is merely a subsystem of the network with learning. In this case, even if the environment is unstable (e.g., due to existence of non-minimum phase zeros in a linear state space model), the overall augmented network may represent a nonlinear adaptive dynamic system, which may converge towards the desired parameters as a stable equilibrium point or set. Thus achieving the global task of blind identification [126, 156].

For the linear filtering environments, this problem may be presented as

$$\underline{Y} = H * \underline{m} = H * \bar{H} * \underline{s} = P * D * \underline{s} \quad (1.1)$$

where

P – is a generalized permutation matrix

D – is a diagonal matrix of filters with each diagonal filter having only one non-zero tap.

$*$ – represents ordinary matrix multiplication for the static mixing case, while it represents polynomial matrix multiplication for the multi-tap deconvolution problems.

Possible differences in the dimensions of the signal, mixture and output vector stem from the blindness of the problem. However, in most cases it is very difficult if not impossible to determine the number of unknown sources directly from the observed mixtures [4,6]. This results in three types of BSR problems, i.e.

- For $m = n$, the problem is referred to as simply *BSR*.
- For $m > n$, the problem is termed as *undercomplete BSR*, due to higher dimensionality of the projected mixture space
- For $m < n$, the problem is called *overcomplete BSR*, due to scarcity of dimension in the mixtures space.

BSR a

conven

for all

easily

measure

decompe

for $m <$

algorith

BSR, wh

1.4.1

In this

matrix .

simulat

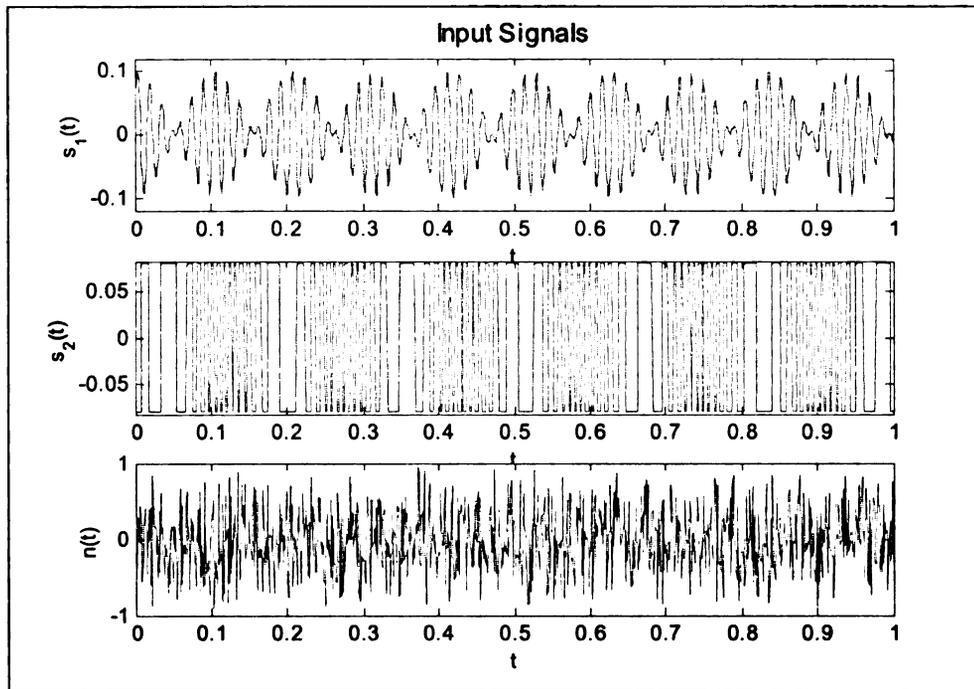
For notational and mathematical convenience, in Chapter 3 we will derive all BSR algorithms for the case where $n = m = M$, unless otherwise specified. This convenient choice allows for unambiguous mathematical manipulations in the derivation for all the algorithms. The under-complete mixture case of BSR, where $m > n$, can be easily reduced to the fore mentioned case by appropriately discarding some measurements using pre-processors, e.g., Principal Component Analysis (PCA), eigen decomposition etc. [33, 52, 53, 60]. On the other hand, the over-complete mixture case for $m < n$, in general, is not tractable and cannot be handled by the existing BSR algorithms in any straightforward fashion. A solution for a special case of overcomplete BSR, where the underlying source distributions are sparse, is discussed in chapter 5.

1.4.1 Blind Source Separation: An Example

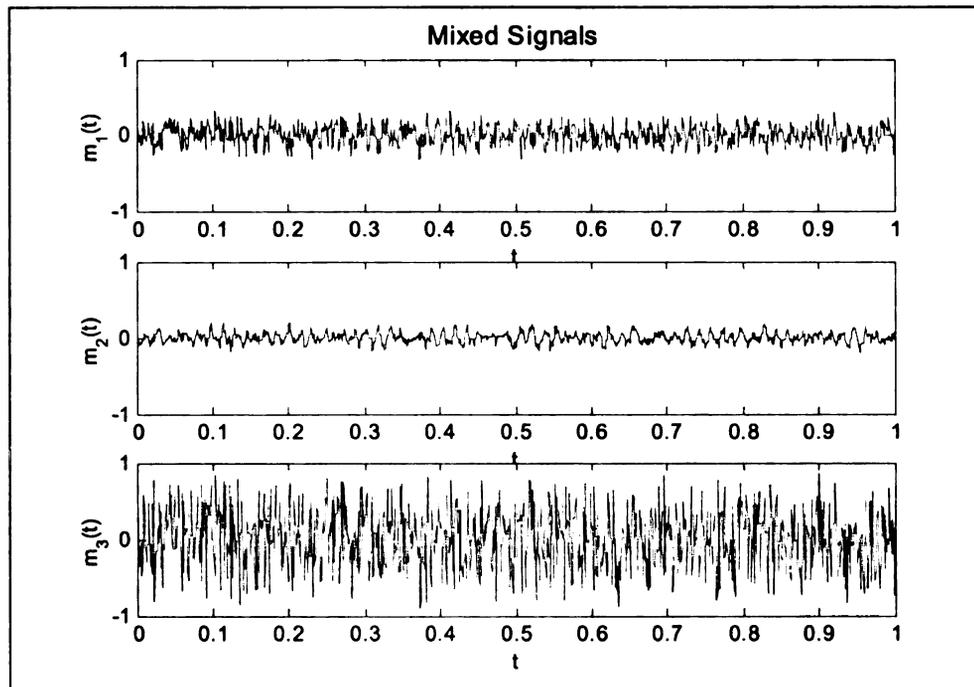
In this example, we mix three signals using an arbitrarily chosen normalized mixing matrix . The original sources are

$$s(t) = \begin{cases} 0.1 \sin(400t) \cos(30t) \\ 0.08 \text{sign}(\sin(500t + 9 \cos(40t))) \\ \text{noise, uniformly distributed} \end{cases}$$

The nonlinearity used in this case is based on Edgeworth expansion [8, 47]. The simulation results are presented below

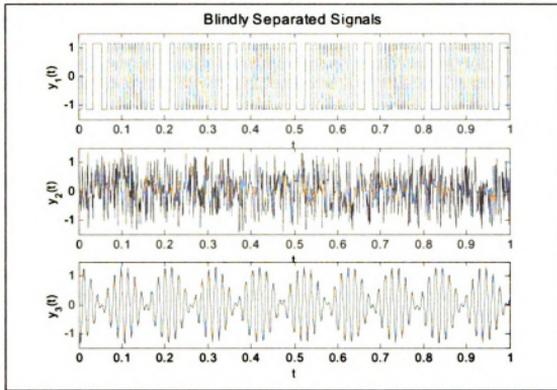


(a)

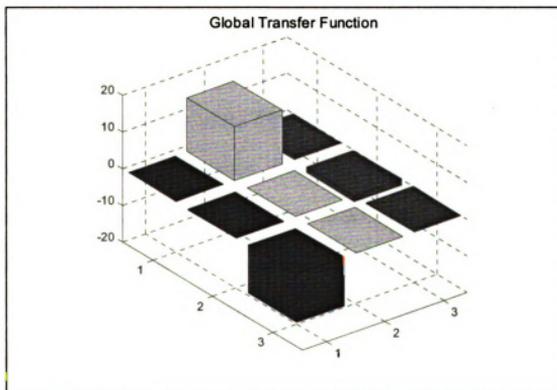


(b)

Figure 1.2. Blind Source Separation (a) Original Sources (b) Observed Mixtures



(a)



(b)

Figure 1.3. Blind Source Separation (a) Separated Signals (b) Global Transfer Function

1.5 O

This dis

motivati

overview

Chapter 2

approach

with a di

recovery p

general n

dynamical

for more e

are deriv

classificat

feedforwa

performan

4 discuss

chapter 3.

Generalize

models, re

parameter

source dist

are also pr

and overco

1.5 Organization of the dissertation

This dissertation is divided into seven chapters and five indices. Some basic introduction, motivating applications and the definition of the problem are presented in Chapter 1. An overview of the current literature on the problems of BSS and BSD is presented in Chapter 2. This is followed by a concise discussion on the natural gradient learning approach and its application to blind source recovery networks. In chapter 3, we start with a discussion on the multivariable optimization framework for the blind source recovery problem. We initially present BSR update laws previously derived in a more general nonlinear mixing setting, the framework is then specialized to the linear dynamical state space representation. The natural gradient based update laws are derived for more efficient estimation of the recovery network parameters. Separate update laws are derived for the minimum and non-minimum phase environments. Further classification is done on the basis of the BSR network structure, which can be either in a feedforward or a feedback formulation. Simulation results and a comparison of the performance for all the proposed BSR structures make up the rest of the chapter. Chapter 4 discusses the issue of score function adaptation for all the update laws proposed in chapter 3. Two separate adaptive score functions are proposed for BSR based on the Generalized Gaussian distribution model and the proposed Hyperbolic distribution models, respectively. A discussion on adaptive estimation of the batch output kurtosis, parameter regulation for the proposed score functions and their effectiveness for various source distributions is included, The simulation results using the proposed score functions are also presented. In chapter 5, we discuss the recovery issues related to undercomplete and overcomplete mixtures, which may arise due to excess and deficiency of dimension

in mix

sparse

Analys

Simula

are incl

to the a

The foc

CDMA

Compre

equival

derived.

schemes

the prop

also sug

this diss

S

dissertat

this wor

entropy

in the

simplific

(FIR) an

BSR are

in mixture space with respect to the original source space. For the case of overcomplete sparse distributions, a new ICA algorithm called Algebraic Independent Component Analysis (AICA) is presented and related implementation issues are discussed. Simulation examples using both independent identical distribution (iid) data and speech are included as a demonstration of the AICA recovery performance. Chapter 6 is devoted to the application of BSR approaches to the CDMA wireless communication networks. The focus is primarily placed on the forward link (or downlink) for both a generic DS-CDMA implementation, and a UMTS FDD (or WCDMA) wireless network. Comprehensive signal models are initially developed and then cast into convolutive equivalents. Three adaptive BSR algorithms for a user's signal detection are subsequently derived. These proposed detection schemes are then compared to conventional detection schemes for CDMA using extensive simulation results that outline the effectiveness of the proposed approaches. Chapter 7 summarizes the developments made in this work and also suggests various possible future extensions possible to the work being presented in this dissertation.

Several appendices have been included to supplement the main text of this dissertation. Appendix A enlists the various abbreviations and acronyms used throughout this work. In appendix B, we present the Kullback-Lieblar divergence [72] or the relative entropy measure, which is used as the objective performance functional to be minimized in the BSR optimization framework. Appendix C presents specific implementation simplifications in the state space BSR framework for the cases of finite impulse response (FIR) and infinite impulse response (IIR) filters. Various performance benchmarks for BSR are presented in Appendix D. Two of these benchmarks namely the Quadratic

Inform

develo

technic

formul

Information Measure (QIM) and the Algebraic Matrix distance Index (AMDI) have been developed during the course of this work. Lastly appendix E presents a transformation technique to transform a generic MIMO filtering structure to a canonical state space formulation.

Blind S

informa

applicat

telecom

is to rec

mixture

based t

decorre

depend

compon

sensors

blind s

algorith

independ

Chapter 2

REVIEW OF LITERATURE

Blind Source Separation (BSS) by Independent Component Analysis (ICA) using the information theoretic approaches has been a hot research topic because of its potential applications in signal processing such as in speech recognition systems, telecommunications, time series analysis and medical signal processing. The goal of ICA is to recover independent sources given only sensor observations that are unknown linear mixtures of the unobserved independent source signals. In contrast to the correlation-based transformations such as Principal Component Analysis (PCA), ICA not only decorrelates the signals (2^{nd} order statistics) but also reduces higher-order statistical dependencies, attempting to make the signals as independent as possible [53].

Two different research communities have considered the analysis of independent components. On one hand, the study of separating mixed sources observed in an array of sensors has been a classical and difficult signal processing problem. The seminal work on blind source separation was by Herault and Jutten [56] where they introduced an adaptive algorithm in a simple feedback architecture that was able to separate several unknown independent sources. Their approach was further developed by Jutten and Herault [63];

Karhunen and Joutsensalo [65]; Cichocki, Unbehauen and Rummert [36]; Comon [37] elaborated the concept of independent component analysis and proposed cost functions related to the approximate minimization of mutual information between the sensors.

In parallel to blind source separation studies, unsupervised learning rules based on information theory were proposed by Linsker [83]. The goal was to maximize the mutual information between the inputs and outputs of a neural network. This approach is related to the principle of redundancy reduction suggested by Barlow [20] as a coding strategy in neurons. Each neuron should encode features that are as statistically independent as possible from other neurons over a natural ensemble of inputs; decorrelation as a strategy for visual processing was explored by Atick [19]. Salam, et. al., [45, 106] first proposed the use of linear state space models for BSS and derived a novel adaptive algorithm based on multivariable optimization theory. Nadal and Parga [88] showed that in the low-noise case, the maximum of the mutual information between the input and output of a neural network implied that the output distribution was factorial; that is, the multivariate probability density function (pdf) can be factorized as a product of marginal pdfs. Roth and Baram [102] and Bell and Sejnowski [21] independently derived stochastic gradient learning rules for this maximization and applied them, respectively, to forecasting, time series analysis, and the blind separation of sources. Bell and Sejnowski [21] put the blind source separation problem into an information-theoretic framework and demonstrated the separation and deconvolution of mixed sources. Their adaptive methods are more plausible from a neural processing perspective than the cumulant-based cost functions proposed by Comon [37]. A similar adaptive method for source separation was proposed by Cardoso and Laheld [29].

viewp

propos

Girolan

Project

kurtosis

linear n

separati

similarit

process

develop

algorithm

the con

the natu

Cichock

and sou

[30]. G

higher d

index, c

Gaussian

intercon

of exte

compreh

Other algorithms for performing ICA have been proposed from different viewpoints. Maximum Likelihood Estimation (MLE) approaches to ICA were first proposed by Gaeta and Lacoume [43] and elaborated by Pearlmutter and Parra [94]. Girolami and Fyfe [51], motivated by information-theoretic indices for Exploratory Projection Pursuit (EPP) used marginal negentropy as a projection index and showed that kurtosis-seeking projection pursuit will extract one of the underlying sources from a linear mixture. A multiple output EPP network was eventually developed to allow full separation of all the underlying sources [50]. Bell and Sejnowski [21] have pointed out a similarity between their infomax algorithm and the Bussgang algorithm in signal processing. Torkkola [118] and Girolami [48] also made theoretical and implementation developments using infomax algorithm. Lambert [73, 74] proposed extensions in BSS algorithms using frequency domain polynomial matrix algebra, further he also elucidated the connection between three different Bussgang cost functions. Amari [7, 8] proposed the natural gradient based learning rules based on information geometry while Amari, Cichocki and Yang [13] proposed implementations based on use of higher order statistics and source pdf approximations using Gram Charlier and Edgeworth series expansions [30]. Gharbi and Salam [45, 47] extended their state space algorithms derived using higher order statistics. Girolami and Fyfe [51] chose negentropy as a projection pursuit index, developing a learning rule that is able to blindly separate mixed sub- and super-Gaussian. Lee, Girolami, Bell and Sejnowski [76, 77, 79] have shown the interconnection of all the above mentioned approaches, unifying them under the banner of extended infomax algorithm. Amari, Douglas, Cichocki and Yang [14] proposed a comprehensive algorithm using the natural gradient for the multichannel blind source

separa

repres

algorit

equaliz

Source

contra-

The use

47, 111

using m

[153, 15

space alg

filters fo

for the s

126, 129

dynamic

Source F

minimum

distributi

O

contravar

blind dec

and the o

separation problems. Constraining both the mixing and demixing environment to be represented by FIR filters. Several research groups have further worked to optimize these algorithms for various domains of applications especially for mixtures of speech and equalization of communication channels.

It has been proved that the algorithms for the Blind Source separation and Blind Source deconvolution give superior performance when optimized using the Riemannian contra-variant or the natural gradient [6, 9, 16], or equivalently the relative gradient [29]. The use of state space for BSS was first proposed by Salam [106]. Gharbi and Salam [45, 47, 111] extended their algorithms for continuous time static and dynamic environments using multivariable optimization theory and the natural gradient. Zhang, Cichocki et. al. [153, 155, 156] followed suit with several research papers, focused on derivation of state space algorithms based on Lie Group structures. They have also advocated use of Kalman filters for state estimation [157]. Salam and Erten. [109] proposed a general framework for the state space blind source recovery. More recently, Waheed and Salam [110, 112, 126, 129] have proposed extended discrete time state space algorithms, for static and dynamic environments, both in feedforward and feedback configurations for Blind Source Recovery (BSR). They have demonstrated the application of algorithms for both minimum and non-minimum phase systems, and also for a variety of unknown source distributions using adaptive non-linearities.

Our current research proposes extended discrete-time natural (or the Riemannian contravariant) gradient algorithms for the Blind Source recovery (BSR) or multichannel blind deconvolution (MCBD) problems fused with the use the state-space formulation, and the optimization theory. BSS in this approach can be viewed as a trivial static case.

Before providing details of the conducted research, the notion of the natural gradient and its application to the problems of BSS and BSD has been concisely discussed.

The choice of the natural gradient, for the problems of Blind Source Recovery (BSR) or Multichannel Blind Deconvolution (MCBD) has been made due its superior and equivariant performance, which is preserved in the state-space formulation for this research

2.1 The Natural Gradient Framework

For any general nonlinear optimization framework, the stochastic gradient method [5, 104, 119, 147] is a popular learning method. However, in many cases the parameter space is not Euclidean but has a Riemannian metric structure. In these cases, the ordinary gradient does not give the steepest direction of a target function; rather, the steepest direction is given by the natural (or the contravariant) gradient. The Riemannian metric structures were introduced by means of information geometry [6-9, 16]. The natural gradient is discussed explicitly in the space of matrices for blind source separation (BSS), and the space of linear dynamical systems for the blind source deconvolution (BSD) problem.

A performance comparison between the ordinary stochastic gradient and the natural gradient has been presented in [9]. The reference studies the asymptotic behavior of online natural gradient learning. For online adaptation, the training examples can be used only once as they appear during the learning/adaptation phase. Therefore, the asymptotic performance of online learning/adaptation cannot be better than the optimal batch procedure where all the examples can be reused again and again. This has been proved that natural gradient online learning gives the Fisher-efficient estimator in the

sense

asympt

loss fu

than b

by the

gradien

2.1.1

Let $S =$

Euclidean

incremen

where d

orthonor

W

length of

The $n \times n$

general o

It reduces

g.

sense of asymptotic statistics when the loss function is differentiable, so that it is asymptotically equivalent to the optimal batch procedure, see [7, 91]. However, when the loss function is non-differentiable, the accuracy of asymptotic online learning is worse than batch learning by a factor of 2, for example see [25]. In this section, what is meant by the natural gradient is briefly outlined, and its relationship to the ordinary stochastic gradient is established.

2.1.1 Definition of the Natural Gradient

Let $S = \{w \in R^n\}$ be a parameter space on which a function $L(w)$ is defined. When S is a Euclidean space with an orthonormal coordinate system w , the squared length of a small incremental vector connecting w and $w + dw$ is given by

$$|dw|^2 = \sum_{i=1}^n (dw_i)^2 \quad (2.1)$$

where dw_i are the components of dw . However, when the coordinate system is non-orthonormal, the squared length is given by the quadratic form

$$|dw|^2 = \sum_{i,j} g_{ij}(w) dw_i dw_j \quad (2.2)$$

When S is a curved manifold, there are no orthonormal linear coordinates, and the length of dw is always written as in equation (2.2). Such a space is a Riemannian space. The $n \times n$ matrix $G = (g_{ij})$ is called the Riemannian metric tensor, and it depends in general on w .

It reduces to

$$g_{ij}(w) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (2.3)$$

in the

that m

for a s

2.1.2

The ste

where

gradie

the su

When

is def

notati

is Euc

in the Euclidean orthonormal case, so that G is the unit matrix I in this case.

The steepest descent direction of a function $L(w)$ at w is defined by the vector dw that minimizes $L(w+dw)$ where $|dw|$ has a fixed length, that is, under the constraint

$$|dw|^2 = \varepsilon^2 \quad (2.4)$$

for a sufficiently small constant ε , see [7, 9, 16].

2.1.2 Theorem 2.1

The steepest descent direction of $L(w)$ in a Riemannian space is given by

$$-\tilde{\nabla}L(w) = -G^{-1}(w)\nabla L(w) \quad (2.5)$$

where $G^{-1} = (g^{ij})$ is the inverse of the metric $G = (g_{ij})$ and ∇L is the conventional gradient,

$$\nabla L(w) = \left(\frac{\partial}{\partial w_1} L(w), \dots, \frac{\partial}{\partial w_n} L(w) \right)^T$$

the superscript T denoting the transposition.

Where

$$\tilde{\nabla}L(w) = G^{-1}\nabla L(w) \quad (2.6)$$

is defined to be the *natural gradient* of L in the Riemannian space.

Thus, $-\tilde{\nabla}L$ represents the steepest descent direction of L . Using the tensorial notation, this is nothing but the contravariant form of $-\nabla L$. Observe that when the space is Euclidean and the coordinate system is orthonormal, this reduces to $\tilde{\nabla}L = \nabla L$

probl

presen

2.1.3

Let us

variab

signals

adjusta

process

where

stochas

where

definite

$C(w)$ e

where

The natural gradient learning has been briefly outlined and its use for the problems of BSS and BSR (or MCBBD) is shown below. For complete proofs of the presented theorems, please see [9, 16].

2.1.3 The Natural Gradient Learning

Let us consider an information source that generates a sequence of independent random variables $s_1, s_2, \dots, s_t, \dots$, subject to the same probability distribution $q(s)$. The random signals s_t are processed by an adaptive processor (e.g., a neural network) that has a set of adjustable parameters w . Let $l(s, w)$ be a loss function when signal s is processed by the processor whose parameter is w . Then the risk function or the average loss is

$$L(w) = E[l(s, w)] \quad (2.7)$$

where E denotes the expectation with respect to s .

Learning is a procedure to search for the optimal w^* that minimizes $L(w)$. The stochastic gradient descent learning method can be formulated in general as

$$w_{t+1} = w_t - \eta_t C(w_t) \nabla l(s_t, w_t) \quad (2.8)$$

where η_t is a learning rate that may depend on t and $C(w)$ is a suitably chosen positive definite matrix [5]. In the natural gradient online learning method, it is proposed to put $C(w)$ equal to $G^{-1}(w)$ when the Riemannian structure is defined as (2.6).

This suggests the natural gradient descent algorithm of the form

$$w_{t+1} = w_t - \eta_t \tilde{\nabla} L(w_t), \quad (2.9)$$

where η_t is the learning rate that determines the step size.

Two illustrative examples for the natural gradient learning are presented that include Blind Source Separation (BSS) and Blind Source Deconvolution (BSD). A brief introduction to the setup of each problem is followed by the derivation of update laws based on the natural gradient, this approach will serve to outline both problems and also present the motivation in the use of natural gradient while searching for an adaptive solution.

2.2 Blind Source Separation.

Let us consider n signal sources that produce n independent signals $s_i(t)$, $i = 1, \dots, n$, at discrete times $t = 1, 2, \dots$. Assuming that $s_i(t)$ are independent at different times and that the expectations of s_i are 0. Let $r(s)$ be the joint probability density function of s . Then it is written in the product form

$$r(s) = \prod_{i=1}^n r_i(s_i) \quad (2.10)$$

Consider the case where there exists no direct access to the source signals $s(t)$ but only their n instantaneous mixtures $m(t)$ can be observed,

$$m(t) = Hs(t), \text{ or} \\ m_i(t) = \sum_{j=1}^n H_{ij}s_j(t) \quad (2.11)$$

where $H = (H_{ij})$ is an $n \times n$ nonsingular mixing matrix that does not depend on t , and

$m = (m_1, \dots, m_n)^T$ is the observed mixtures.

1. 2. .

because

The "3

densiti

where

form (

follow

2.2.1

The n

there

when

The

poin

Blind source separation is the problem of recovering the original signals $s(t)$, $t = 1, 2, \dots$, from the observed signals $m(t)$, $t = 1, 2, \dots$, [63]. If H is known, this is trivial, because $s(t) = H^{-1}m(t)$

The “blind” implies that neither the mixing matrix H nor the probability distribution densities $r_i(s_i)$ are known.

A typical algorithm to solve the problem is to transform $m(t)$ into

$$y(t) = W_t m(t) \quad (2.12)$$

where W_t is an estimate of H^{-1} , and is modified by the following update equation of the form (2.9). The specific form of the update law for BSS can be derived using the following theorem by Amari [9].

2.2.1 Theorem 2.2

The natural gradient in the matrix space is given by

$$\tilde{\nabla} l = (\nabla l) W^T W \quad (2.13)$$

therefore, the natural gradient update equation for BSS is

$$\frac{dW}{dt} = \eta_t (I - \varphi(y)y^T) W \quad (2.14)$$

where $\varphi(y)$ is the column vector

$$\begin{aligned} \varphi(y) &= [\varphi_1(y_1), \dots, \varphi_m(y_m)]^T \\ \varphi_i(y_i) &= -\frac{d}{dy} \log f_i(y_i) \end{aligned} \quad (2.15)$$

The efficiency of this equation is studied from the statistical and information geometrical point of view [11, 15].

2.3

When

signal

introd

The m

where

To rec

of deg

where

where

2.3.1

The n

2.3 Blind Source Deconvolution.

When the original signals $s(t)$ are mixed not only instantaneously but also with past signals as well, the problem is called blind source deconvolution or equalization. By introducing the time delay operator z^{-1} ,

$$z^{-1}s(t) = s(t-1) \quad (2.16)$$

The mixing matrix filter H is denoted by

$$H(z) = \sum_{k=0}^{\infty} H_k z^{-k} \quad (2.17)$$

where A_k are $m \times m$ matrices. The observed mixtures are

$$m(t) = H(z)s(t) = \sum_k H_k s(t-k) \quad (2.18)$$

To recover the original independent sources, using the finite impulse response model

$$W(z) = \sum_{k=0}^d W_k z^{-k} \quad (2.19)$$

of degree d . The original signals are recovered by

$$y(t) = W_t(z)m(t) \quad (2.20)$$

where W_t is adaptively modified by

$$W_{t+1}(z) = W_t(z) - \eta_t \tilde{\nabla} l \{m_t, m_{t-1}, \dots, W_t\} \quad (2.21)$$

where the natural gradient in this case is given by the following theorem [9].

2.3.1 Theorem 2.3

The natural gradient of the manifold of systems is given by

$$\tilde{\nabla} l = \nabla l(z) W^T (z^{-1}) W(z), \quad (2.22)$$

where

2.4

An al

equiv

mixing

deconv

only d

where

Post r

where

that i

algor

where operator z^{-1} should be operated adequately.

2.4 Equivariant Performance of the Natural Gradient

An algorithm for source separation of instantaneous additive mixtures is defined to have equivariant performance if its behavior does not depend explicitly on the form of the mixing matrix H [29]. This definition can be extended for the case of source deconvolution [13] by stating that a deconvolution algorithm is equivariant if its behavior only depends on the combined filter $G(z,k)$

Using (2.21), the update algorithm can be expressed in the form

$$\Delta W(z,k) = -\eta_t \tilde{\nabla} l \{m_t, m_{t-1}, \dots, W_t\} = -\eta_t \frac{\partial l(z)}{\partial X(z)} W(z,k) \quad (2.23)$$

where $dX(z) = dW(z) \{W(z)\}^{-1}$

Post multiplying both sides of (2.23) by $H(z)$, the relation is given by

$$\Delta G(z,k) = -\eta_t \frac{\partial l(z)}{\partial X(z)} G(z,k) \quad (2.24)$$

where $G(z,k) = W(z,k) * H(z)$,

Note, that both (2.24) and the definition of gradient can be expressed in a form that is independent of $H(z)$, which proves the equivariant property of the natural gradient algorithms.

Chapter 3

BLIND SOURCE RECOVERY: A FRAMEWORK IN THE STATE SPACE

Blind Source Recovery (BSR) denotes recovery of original sources/signals from environments that may include convolution, temporal variation, and even nonlinearity. It also infers the recovery of sources even in the absence of precise environment identifiability. This chapter describes a generalized BSR formulation, see Salem, et. al. [109, 110], achieved by the application of stochastic optimization principles to the Kullback-Lieblar divergence as a performance functional subject to the constraints of the general (i.e., nonlinear and time-varying) state space representation. This technique was used to derive update laws for nonlinear time-varying dynamical systems, which are subsequently specialized to canonical neural and time-invariant linear systems. Some of this work has also been implemented in dedicated hardware/software platforms [40-42, 46].

The state space feedforward and feedback demixing network structures have been exploited to develop natural gradient based learning rules. These adaptation algorithms

are capable of handling most filtering paradigms and which can be conveniently extended to nonlinear models. In the special cases, distinct linear state-space algorithms are presented for the minimum phase and non-minimum phase mixing environment models. Conventional (FIR/IIR) filtering models can be subsequently derived from this general structure (see, Appendix C). Illustrative simulation examples are presented to demonstrate the *online* adaptation capabilities of the developed BSR algorithms.

3.1 Formulation of the Optimization Structure

This section formally describes the state space BSR optimization framework, see [109, 112] for more details. In this section, the BSR stochastic gradient update laws are derived via the proposed constrained multivariable optimization framework [109, 110] using Kullback-Lieblar divergence as the performance functional (see, Appendix B). This performance functional is a well known information-theoretic “distance” measure, appropriate for quantifying the mutual *dependence* of signals in a mixture. Using the calculus of variations and the optimization theory, the update laws are initially derived for a more general non-linear dynamic structure. Subsequently, the update algorithm is specialized to the case of a linear dynamical state space model.

For the case of discrete-time, independent, wide sense stationary output vector $y(k)$, the Kullback Lieblar divergence (or the relative entropy) is given by

$$L(y(k)) = -H(y(k)) + \sum_{i=1}^n H_i(y_i(k)) \quad (3.1)$$

where, $H(y(k))$: represents the entropy of the signal vector $y(k)$, given by

$$H(y) = -E \left[\ln |p_y(y)| \right] = \begin{cases} - \int_{y \in Y} p_y(y) \ln |p_y(y)| dy & \text{Continuous Case} \\ - \sum_{y \in Y} p_y(y) \ln |p_y(y)| & \text{Discrete Case} \end{cases}$$

and $H_i(y_i(k))$: represents the marginal entropy of a component signal $y_i(k)$

3.1.1 Algorithms for the Nonlinear Dynamic Case

Assume that the environment can be modeled as the following nonlinear discrete-time dynamic forward model

$$X_e(k+1) = f_e(X_e(k), s(k), h_1) \quad (3.2)$$

$$m(k) = g_e(X_e(k), s(k), h_2) \quad (3.3)$$

where

$s(k)$: n -d vector of original source signals

$m(k)$: m -d vector of measurements

$X_e(k)$: M_e -d state vector for the environment

h_1 : constant parameter vector (or matrix) of dynamic state equation

h_2 : constant parameter vector (or matrix) of output equation

$f_e(\cdot)$ and $g_e(\cdot)$: differentiable nonlinear functions [86] that specify the structure of the environment.

Further it is assumed that existence and uniqueness of solutions are satisfied for any given initial conditions $X_e(t_0)$ and sources $s(k)$ thus a Lipschitz condition on $f_e(\cdot)$ is satisfied [71].

feedf

netwo

where

$m(k)$:

$y(k)$:

$X(k)$:

w_1 : pa

w_2 : pa

$f(\cdot)$ a

netwo

T

equati

and m

L

so tha

funct.

optim

N

The processing demixing network model may be represented by a dynamic feedforward or feedback network. Focusing on a feedforward network model, assume the network to be represented by

$$X(k+1) = f(X(k), m(k), w_1) \quad (3.4)$$

$$y(k) = g(X(k), m(k), w_2) \quad (3.5)$$

where

$m(k)$: m -d vector of measurements

$y(k)$: N -d vector of network output

$X(k)$: M -d state vector for the processing network

w_1 : parameters of the network state equation

w_2 : parameters of the network output equation

$f(\cdot)$ and $g(\cdot)$: differentiable nonlinear functions defining the structure of the demixing network

The assumption of existence and uniqueness of solutions of the nonlinear difference equations is also assumed for the network model for any given initial conditions $X(t_0)$ and measurement vector $m(k)$.

In order to derive the update law, the notation is abused for the sake of convenience so that $y(k)$ in (3.1) is represented as y_k and $L(y(k))$ is generalized to $L^k(y_k)$ so that the functional may also be a function of the time index k . Thus, the following constrained optimization problem can be formulated [109, 112] to be

Minimize

subje

with t

compu

becom

where

Conse

and th

where

$$J_o(w_1, w_2) = \sum_{k=k_0}^{k_1} L^k(y_k) \quad (3.6)$$

subject to

$$X_{k+1} = f^k(X_k, m_k, w_1) \quad (3.7)$$

$$y_k = g^k(X_k, m_k, w_2) \quad (3.8)$$

with the initial condition X_{k_0} , where $[k_0, k_1)$ is the discrete-time frame used for the computation of the cost functional. Thus, the augmented cost functional to be optimized becomes

$$J(w_1, w_2) = \sum_{k=k_0}^{k_1} L^k(y_k) + \lambda_{k+1}^T (f^k(X_k, m_k, w_1) - X_{k+1}) \quad (3.9)$$

where λ_k is the Langrange variable [81]. Define the Hamiltonian as

$$H^k = L^k(y_k) + \lambda_{k+1}^T f^k(X_k, m_k, w_1) \quad (3.10)$$

Consequently, the *necessary* conditions for optimality are

$$X_{k+1} = \frac{\partial H^k}{\partial \lambda_{k+1}} = f^k(X_k, m_k, w_1) \quad (3.11)$$

$$\lambda_k = \frac{\partial H^k}{\partial X_k} = (f_{X_k}^k)^T \lambda_{k+1} + \frac{\partial L^k}{\partial X_k} \quad (3.12)$$

and the change in “weight” parameters become

$$\Delta w_1 = -\eta_k \frac{\partial H^k}{\partial w_1} = -\eta_k (f_{w_1}^k)^T \lambda_{k+1} \quad (3.13)$$

$$\Delta w_2 = -\eta_k \frac{\partial H^k}{\partial w_2} = -\eta_k \frac{\partial L^k}{\partial w_2} \quad (3.14)$$

where

$$f_a^k \triangleq$$

$$\eta_k : \text{re}$$

3.1.2

Most c

There

capacit

these r

[109,

neural

scenar

deriva

where

$$m(k)$$

$$y(k)$$

$$X_i(k)$$

$$w_i - 1$$

$f_a^k \triangleq \frac{\partial f^k}{\partial a}$: represents the partial derivative w.r.t the parameter a , in the limit.

η_k : represents a positive time-adaptive learning rate.

3.1.2 Algorithm for Nonlinear Dynamic Canonical Neural Structure

Most of the real life problems cannot be adequately modeled using linear representations. There is a requirement for rich non-linear demixing network models [151] that have the capacity to undo effects of an intrinsically non-linear environment. The non-linearities in these real-life environments can take a variety of forms. The nonlinear dynamic structure [109, 110] presented above can be specialized to a hierarchical dynamic feedforward neural structure. This neural structure has the capacity to deal with several practical scenarios. Since the canonical structure is more amenable to practical implementations, a derivation for such a neural structure is presented below

We assume the demixing network to be represented by the structure

$$\begin{aligned}
 X_1(k+1) &= f_1(X_1(k), m(k), w_1) \\
 X_2(k+1) &= f_2(X_1(k), X_2(k), m(k), w_2) \\
 &\vdots \\
 X_n(k+1) &= f_n(X_1(k), \dots, X_n(k), m(k), w_n) \\
 y(k) &= g(X_1(k), \dots, X_n(k), m(k), w_y)
 \end{aligned} \tag{3.15}$$

where

$m(k)$ - represents an m dimensional measurement vector

$y(k)$ - represents an N dimensional output vector

$X_i(k)$ - represents $m \times m$ dimensional i^{th} network state vector

w_i - represents parameters of the i^{th} network state equation, $i = 1, 2, \dots, n$.

$w_y - r$

$f_i(\cdot)$

differ

and m

recove

diverg

symbol

subject

and the

w_y - represents parameters of the network output equation

$f_i(\cdot)$ and $g(\cdot)$ - represent differentiable non-linear functions.

Further it is assumed that existence and uniqueness of the solutions of the differential equations are satisfied for any given initial conditions $X_i(t_0), i=1,2,\dots,n$ and measurement vector $m(k)$ [71].

In order to derive the update laws for this non-linear case of blind source recovery, we set up the following optimization problem using the Kullback-Liebler divergence [72] as the performance functional. Please note the convenient change in symbolic notation used for the derivation.

Minimize

$$J(w_1, \dots, w_n, w_y) = \sum_{k=k_0}^{k_1} L^k(y_k) \quad (3.16)$$

subject to

$$\begin{aligned} X_1^{k+1} &= f_1^k(X_1^k, m_k, w_1) \\ X_2^{k+1} &= f_2^k(X_1^k, X_2^k, m_k, w_2) \\ &\vdots \\ X_n^{k+1} &= f_n^k(X_1^k, \dots, X_n^k, m_k, w_n) \\ y_k &= g(X_1^k, \dots, X_n^k, m_k, w_y) \end{aligned} \quad (3.17)$$

and the initial conditions $X_1^{k_0}, X_2^{k_0}, \dots, X_n^{k_0}$

The augmented cost functional to be minimized becomes

Defin

and the

triangu

upper tr

$$J(w_1, \dots, w_n, w_y) = \sum_{k=k_0}^{k_1} L^k(y_k) + \begin{bmatrix} \lambda_1^{k+1} \\ \lambda_2^{k+1} \\ \vdots \\ \lambda_n^{k+1} \end{bmatrix}^T \begin{bmatrix} f_1^k(X_1^k, m_k, w_1) - X_1^{k+1} \\ f_2^k(X_1^k, X_2^k, m_k, w_2) - X_2^{k+1} \\ \vdots \\ f_n^k(X_1^k, \dots, X_n^k, m_k, w_n) - X_n^{k+1} \end{bmatrix} \quad (3.18)$$

Define the Hamiltonian as

$$H^k = L^k(y_k) + \begin{bmatrix} \lambda_1^{k+1} \\ \lambda_2^{k+1} \\ \vdots \\ \lambda_n^{k+1} \end{bmatrix}^T \begin{bmatrix} f_1^k(X_1^k, m_k, w_1) \\ f_2^k(X_1^k, X_2^k, m_k, w_2) \\ \vdots \\ f_n^k(X_1^k, \dots, X_n^k, m_k, w_n) \end{bmatrix} \quad (3.19)$$

Consequently the necessary conditions for optimality are

$$X_i^{k+1} = \frac{\partial H^k}{\partial \lambda_i^{k+1}} = f_i^k(X_1^k, \dots, X_i^k, m_k, w_i), i = 1, 2, \dots, n \quad (3.20)$$

$$\lambda_i^k = \frac{\partial H^k}{\partial X_i^k} = \frac{\partial f_i^k(\cdot)}{\partial X_i^k} \lambda_i^{k+1} + \dots + \frac{\partial f_n^k(\cdot)}{\partial X_i^k} \lambda_n^{k+1} + \frac{\partial L^k}{\partial X_i^k}, i = 1, 2, \dots, n \quad (3.21)$$

and the change in weighting parameters become

$$\Delta w_i = -\eta_k \frac{\partial H^k}{\partial w_i} = -\eta_k \begin{bmatrix} \frac{\partial f_i^k(\cdot)}{\partial w_i} \\ \vdots \\ \frac{\partial f_n^k(\cdot)}{\partial w_i} \end{bmatrix}^T \begin{bmatrix} \lambda_i^{k+1} \\ \vdots \\ \lambda_n^{k+1} \end{bmatrix}, i = 1, 2, \dots, n \quad (3.22)$$

$$\Delta w_y = -\eta_k \frac{\partial H^k}{\partial w_y} = -\eta_k \frac{\partial L^k}{\partial w_y} \quad (3.23)$$

In this proposed neural structure the update of the states takes the structural lower triangular form, while the update of co-states and the weighting parameters assumes upper triangular structures.

3.2

In the
form

The exi
existenc
environm
Spec
(Kullback

3.2 Algorithms for the Linear Dynamic Case

In the linear dynamic case, the environment model is assumed to be in the state space form

$$X_e(k+1) = A_e X_e(k) + B_e s(k) \quad (3.24)$$

$$m(k) = C_e X_e(k) + D_e s(k) \quad (3.25)$$

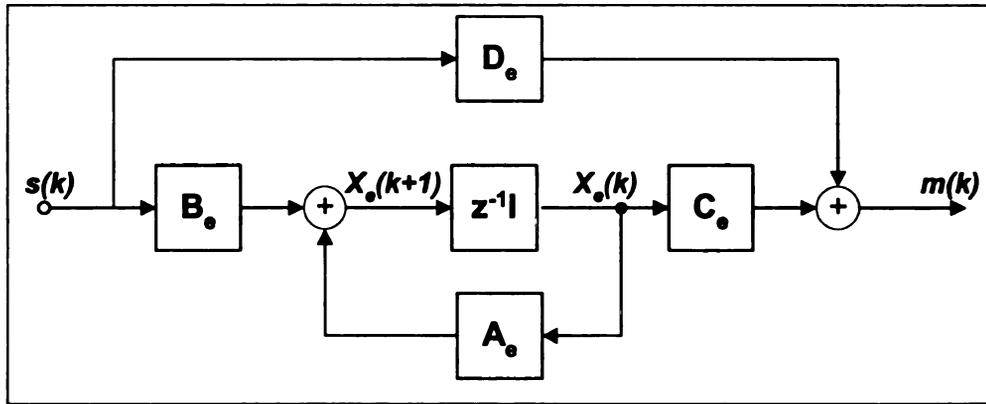


Figure 3.1. State Space Mixing Environment

In this case the feedforward separating network will attain the state space form

$$X(k+1) = A X(k) + B m(k) \quad (3.26)$$

$$y(k) = C X(k) + D m(k) \quad (3.27)$$

The existence of an explicit solution in this case has been shown by [106, 107]. This existence of solution ensures that the network has the capacity to compensate for the environment and consequently recover the original signals, see Figures 3.1 and 3.2.

Specializing the general BSR algorithm to the linear network case, one forms the (Kullback-Lieblar divergence) performance (“distance”) measure, see (3.6) as

subject

The a

Again

F

given

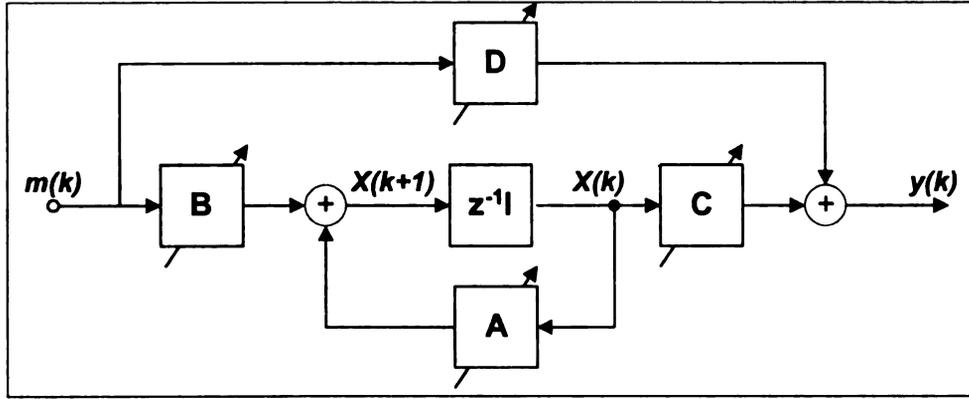


Figure 3.2. State Space Demixing Network

$$J_o(A, B, C, D) = \sum_{k=k_o}^{k_1} L^k(y_k) \quad (3.28)$$

subject to (3.26) and (3.27) with the initial conditions X_{k_o}

The augmented cost functional to be optimized becomes

$$J(A, B, C, D) = \sum_{k=k_o}^{k_1} L^k(y_k) + \lambda_{k+1}^T (A X_k + B m_k - X_{k+1}) \quad (3.29)$$

Again, define the Hamiltonian as

$$H^k = L^k(y_k) + \lambda_{k+1}^T (A X_k + B m_k) \quad (3.30)$$

For the linear time-invariant case the ordinary stochastic gradient update laws are given by [109, 110].

$$X_{k+1} = \frac{\partial H^k}{\partial \lambda_{k+1}} = A X_k + B m_k \quad (3.31)$$

$$\lambda_k = \frac{\partial H^k}{\partial X_k} = A_k^T \lambda_{k+1} + C_k^T \frac{\partial L^k}{\partial y_k} \quad (3.32)$$

provid

matric

guaran

η_k : po

$[D]^{-T}$

the tran

$\varphi(y)$:

which

say φ

where

$p(y_i)$:

136].

$$\Delta A = -\eta_k \frac{\partial H^k}{\partial A} = -\eta_k \lambda_{k+1} X_k^T \quad (3.33)$$

$$\Delta B = -\eta_k \frac{\partial H^k}{\partial B} = -\eta_k \lambda_{k+1} m_k^T \quad (3.34)$$

$$\Delta C = -\eta_k \frac{\partial H^k}{\partial C} = -\eta_k \frac{\partial L^k}{\partial C} = -\eta_k \varphi(y) X^T \quad (3.35)$$

$$\Delta D = -\eta_k \frac{\partial H^k}{\partial D} = -\eta_k \frac{\partial L^k}{\partial D} = \eta_k ([D]^{-T} - \varphi(y) m^T) \quad (3.36)$$

The above derived update laws form a comprehensive theoretical algorithm and provides the update laws for the states X_k , the co-states λ_k and all the parametric matrices in the state space. The invertibility of the state space as discussed in [45, 106] is guaranteed if the matrix D is invertible. In the above derived laws

η_k : positive learning rate of the algorithm, which may be adaptive

$[D]^{-T}$: represents the transpose of the inverse of the matrix D if it is a square matrix or the transpose of its pseudo-inverse in case D is not a square matrix.

$\varphi(y)$: represents a vector of the usual nonlinearity or score function [67, 127, 136, 138], which acts individually on each component of the output vector y , with each component, say $\varphi(y_i)$, is given as

$$\varphi(y_i) = -\frac{\partial \log p(y_i)}{\partial y_i} = -\frac{\partial p(y_i) / \partial y_i}{p(y_i)} \quad (3.37)$$

where

$p(y_i)$: is the (estimate) of the probability density function of each source [127, 130, 136].

result

update

and B

An in

and th

3.2.1

In thi

Recov

[6, 7,

enviro

[110,

system

unstal

the en

the u

demi

inver

[11,

filter

exhib

the r

The update law in (3.35) and (3.36) is similar to the standard gradient descent results [12, 13, 155], indicating its optimality for the Euclidean parametric structure. The update law provided above although non-causal for the update of parametric matrices A and B , can be easily implemented using frame-buffering and memory storage [52, 54]. An inherent delay or latency in the recovered signal are related to the length of the buffer and the dimension of the state space.

3.2.1 Extensions to the Natural Gradient

In this section, the linear state space algorithms for the problem of Blind Source Recovery (BSR) derived in the previous section are extended using the natural gradient [6, 7, 9, 14, 16]. Specialized algorithms for the class of minimum phase mixing environments are presented both in feedforward and feedback state space configuration [110, 112, 126]. For the non-minimum phase mixing environment, the requisite demixing system becomes unstable due to the presence of poles outside the unit circle. These unstable poles are required to cancel out the non-minimum phase transmission zeros of the environment. In order to avoid instability due to the existence of these poles outside the unit circle, the natural gradient algorithm may be derived with the constraint that the demixing system is a double sided FIR filter, i.e., instead of trying to determine the IIR inverse of the environment, the inverse is approximated using an all zero non-causal filter [11, 129]. The double-sided filters have been proven to adequately approximate IIR filters at least in the magnitude terms with a certain associated delay [22].

The minimum phase algorithms have lesser computational requirements and exhibit asymptotically better convergence characteristics, compared to the algorithm for the non-minimum phase mixing environments, when applied to the same class of

minimum phase systems. The minimum phase algorithm also exhibits more robust performance with respect to the choice of the order of demixing/deconvolving filter per channel, and the multi-source distribution composition of mixtures that may include at most one pure gaussian source [33]. The non-minimum phase algorithm gives good performance for the non-gaussian mixtures and/or the non-gaussian contents of a complex multi-source distribution mixture. The gaussian source although separated from other sources may still be auto-convolved (Note that the sum of co-centric copies of a gaussian distributions is also gaussian). Also notice that the presented algorithms in this chapter focus primarily on adaptive determination of zeros. The poles are kept fixed either at zero or at randomly selected stable locations. Otherwise, they are assumed known using other methods, e.g., adaptive state estimation techniques such as the recursive least squares (RLS), kalman filter [155, 157] etc.

3.3 Blind Source Recovery Algorithms for Minimum Phase Environments

3.3.1 Theorem 3.1: Feedforward Minimum Phase Demixing Network

Assume the (mixing) environment is modeled by a multiple-input, multiple-output (MIMO) minimum phase transfer function. Then, the update laws for the zeros of the feedforward state space demixing network using the natural gradient are given by

$$\Delta C(k) = \eta \left((I_N - \varphi(y(k))y^T(k))C(k) - \varphi(y(k))X^T(k) \right) \quad (3.38)$$

$$\Delta D(k) = \eta (I_N - \varphi(y(k))y^T(k))D(k) \quad (3.39)$$

where $\varphi(y)$ is a nonlinear score function given by (3.37)

33.1.1

For the fu

be as in

feedforw

gradient

Riemman

better co

(3.35) an

been dro

Def

so the a

Th

the sto

where

Con

pseu

3.3.1.1 Derivation of Update Laws for Feedforward BSR

For the feedforward demixing network, its linear state space representation is assumed to be as in (3.26)-(3.27). A formal formulation has been developed for deriving the feedforward update laws for the problem using the output equation (3.27) and the natural gradient learning. This leads to modified update laws for the matrices C and D using the Riemmanian manifold. Further, these new update laws based on the natural gradient have better convergence performance compared to those based on the stochastic gradient (3.35) and (3.36). Note that in the following derivation, instantaneous time index k has been dropped for convenience

Defining augmented vectors \tilde{y} and \tilde{x} , and the matrix \tilde{W} as

$$\tilde{y} = \begin{bmatrix} y \\ X \end{bmatrix} \quad \tilde{x} = \begin{bmatrix} m \\ X \end{bmatrix} \quad \tilde{W} = \begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix} \quad (3.40)$$

so the augmented output equation becomes

$$\tilde{y} = \tilde{W} \tilde{x} \quad (3.41)$$

The update law for this augmented parameter matrix \tilde{W} is similar in form to (3.36) or the stochastic gradient law for the static mixing case. Thus the update is

$$\Delta \tilde{W} = \eta \left[\tilde{W}^{-T} - \varphi(\tilde{y}) \tilde{x}^T \right] \quad (3.42)$$

where

$$\tilde{W}^T = \begin{bmatrix} D^T & 0 \\ C^T & I_{(L-1)m} \end{bmatrix}$$

Consequently for the general case where D may not be square, its inverse (assuming the pseudo-inverse for D to exist) is given by

$$\bar{W}^{-T}$$

Factoring out

$$\Delta \bar{W}$$

Post-multiplying

$$\Delta \bar{W}$$

Using (3.41)

$$\Delta \bar{W}$$

Writing in

$$\Delta \left[\begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \right]$$

Considering

$$[\Delta \bar{W}]$$

Therefore

$$\Delta C$$

$$\Delta D$$

Thus

exact agreement

and (3.48)

$$\tilde{W}^{-T} = \begin{bmatrix} D(D^T D)^{-1} & 0 \\ -C^T D(D^T D)^{-1} & I_{(L-1)m} \end{bmatrix}$$

Factoring out the augmented weight term \tilde{W}^{-T} , (3.42) can be written as

$$\Delta \tilde{W} = \eta \left[I - \varphi(\tilde{y}) \tilde{x}^T \tilde{W}^T \right] \tilde{W}^{-T} \quad (3.43)$$

Post-multiplying by the matrix $\tilde{W}^T \tilde{W}$, the update law becomes

$$\Delta \tilde{W} = \eta \left[I - \varphi(\tilde{y}) \tilde{x}^T \tilde{W}^T \right] \tilde{W} \quad (3.44)$$

Using (3.41), the above update rule can be written as

$$\Delta \tilde{W} = \eta \left[I - \varphi(\tilde{y}) \tilde{y}^T \right] \tilde{W} \quad (3.45)$$

Writing in terms of the original state space variables the update law (3.45) is given by

$$\Delta \begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix} = \eta \left[I - \varphi \begin{bmatrix} y \\ X \end{bmatrix} \begin{bmatrix} y^T & X^T \end{bmatrix} \right] \begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix} \quad (3.46)$$

Considering the update laws for matrices C and D only, the relation is

$$[\Delta D \quad \Delta C] = \eta [D \quad C] - \eta \begin{bmatrix} \varphi(y) y^T & \varphi(y) X^T \end{bmatrix} \begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix}$$

Therefore the final instantaneous update laws for the matrices C and D are

$$\Delta C(k) = \eta \left((I_N - \varphi(y(k)) y^T(k)) C(k) - \varphi(y(k)) X^T(k) \right) \quad (3.47)$$

$$\Delta D(k) = \eta (I_N - \varphi(y(k)) y^T(k)) D(k) \quad (3.48)$$

The resulting update laws for the natural gradient update derived in [155] are in exact agreement with the update laws derived above [110, 112]. The update law in (3.47) and (3.48) is related to the earlier derived update law (3.35) and (3.36) by the relation.

$\bar{\nabla} /$

where

$\nabla /$

denotes the

in (3.49) is s

3.3.1.2

Examining t

$$0 = e$$

Also

$\Delta /$

By a rearran

$\varphi(\lambda)$

Using the r

$\varphi(\lambda)$

3.3.1.2.1

The resulti

agreement

derivation a

$$\tilde{\nabla}l = \nabla l \begin{bmatrix} D^T D & D^T C \\ C^T D & I_{(L-1)m} + C^T C \end{bmatrix} = \nabla l \begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix}^T \begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix} \quad (3.49)$$

where

$$\nabla l = \begin{bmatrix} \frac{\partial L^k}{\partial D} & \frac{\partial L^k}{\partial C} \end{bmatrix} \quad (3.50)$$

denotes the update according to the ordinary stochastic gradient, the conditioning matrix in (3.49) is symmetric and positive definite.

3.3.1.2 Auxiliary Conditions for Convergence

Examining the update of the remaining terms in equation (3.46),

$$0 = \varphi(X)y^T D \Rightarrow \varphi(X)y^T = 0 \quad (3.51)$$

Also

$$\Delta I_{(L-1)m} = 0 = I_{(L-1)m} - \varphi(X)y^T C - \varphi(X)X^T$$

By a rearrangement of terms, the condition is

$$\varphi(X)y^T C + \varphi(X)X^T = I_{(L-1)m}$$

Using the relation (3.51), the condition reduces to

$$\varphi(X)X^T = I_{(L-1)m} \quad (3.52)$$

3.3.1.2.1 Remark 3.1:

The resulting update laws for the natural gradient update derived in [155] are in exact agreement with the update laws (3.47)-(3.48), also see [107, 110]. However, the derivation approaches and methodologies are different.

3.3.1.2.2

The new au
on the no
network. It
explicitly a
update law
feedback st

However, fo

(in a stocha

$E[$

$E[$

The first e

as defined

must be sa

the causa

condition

the states

special ca

measurem

measurem

the updat

algorithm

3.3.1.2.2 Remark 3.2:

The new auxiliary relations (3.51) and (3.52) define supplementary stochastic conditions on the nonlinear “correlation” between the outputs and the states of the demixing network. In this feedforward minimum phase structure case, these relations do not explicitly appear in the update laws and thus can not be used directly to simplify the update laws (3.47)-(3.48). (In contrast, similar auxiliary conditions derived for the feedback structure *do* simplify the resulting update laws, see proof of Theorem 3.2.) However, for effective convergence of the algorithm, these conditions need to be satisfied (in a stochastic sense), i.e.,

$$E\left[\varphi(X)y^T\right] = 0_{(L-1)m \times N}, \text{ and}$$

$$E\left[\varphi(X)X^T\right] = I_{(L-1)m}$$

The first equality signifies that the output and the states of the network are “uncorrelated” as defined. This condition supplements the condition $E\left[\varphi(y)X^T\right] \rightarrow 0_{(L-1)m \times N}$, which must be satisfied in (3.47) at convergence. Viewing the states as prior signal information, the causal algorithm necessarily includes only causal correlation while the auxiliary condition (3.51) signifies the noncausal correlation. The second condition signifies that the states of the demixing network need to be a white process. It is noted that, in the special case of FIR filtering where the states are basically delayed versions of previous measurements, i.e., $X_i(k) = m_{k-i}$, the second condition (3.52) implies that whitening the measurement signals prior to applying the algorithm would enhance the convergence of the update law. These conditions are in addition to the stability conditions for the algorithm [11, 27, 109].

*Assume the
system or
demixing*

 Δ Δ

where ϕ

3.3.2.1

*An alter
compris
compos
paramet
feedback*

The

3.3.2 Theorem 3.2: Feedback Minimum Phase Demixing

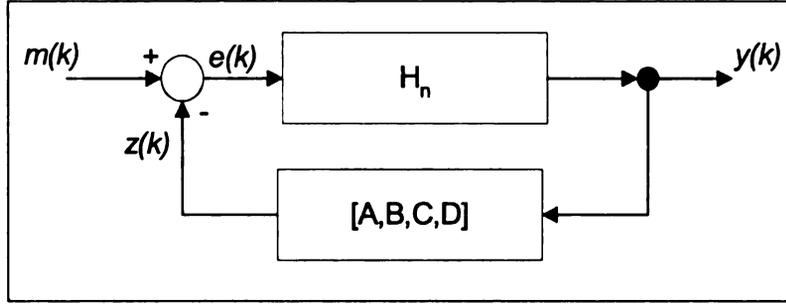


Figure 3.3. State Space Feedback Demixing Structure

Assume the MIMO (mixing) environment modeled by a minimum phase linear dynamic system or transfer function. Then the update laws for the zeros of the feedback state space demixing network using the natural gradient are given by

$$\Delta D = \eta \left[(I_N + D)(\varphi(y)y^T - I_N) \right] \quad (3.53)$$

$$\Delta C = \eta [(I_N + D)\varphi(y)X^T] \quad (3.54)$$

where $\varphi(y)$ is the appropriate score function by (3.37)

3.3.2.1 Derivation of Update Laws for Feedback BSR

An alternate BSR structure was presented in [112], where the demixing network comprises of a feedback transfer function, see Figure 3.3. This feedback path is composed of tunable poles and zeros while the feedforward path is assumed to have fixed parameters. Using only the update laws for adaptively determining the zeros of this feedback path can potentially alter the poles of the overall demixing structure.

The network equations for the feedback path are (see Figure 3.3)

$z(k)$

Defining

$e(k)$

the output

$y(k)$

For simplicity

$y(k)$

rearranging

$(I_N$

In matrix form

$\begin{bmatrix} I_N$

or

$\begin{bmatrix} y \\ x \end{bmatrix}$

Defining the

\tilde{X}

(3.59) can

$\tilde{Y} =$

Using the

$$z(k) = C X(k) + D y(k) \quad (3.55)$$

Defining

$$e(k) = m(k) - z(k) \quad (3.56)$$

the output of the feedback structure is given by

$$y(k) = H_n * e(k)$$

For simplicity, assume $H_n = I$, so

$$y(k) = I * (m(k) - z(k)) = m(k) - z(k) \quad (3.57)$$

rearranging terms, the output equation becomes

$$(I_N + D) y(k) + C X(k) = m(k) \quad (3.58)$$

In matrix form, (3.58) can be written as

$$\begin{bmatrix} I_N + D & C \\ 0 & I_{(L-1)m} \end{bmatrix} \begin{bmatrix} y \\ X \end{bmatrix} = \begin{bmatrix} m \\ X \end{bmatrix}$$

or

$$\begin{bmatrix} y \\ X \end{bmatrix} = \begin{bmatrix} I_N + D & C \\ 0 & I_{(L-1)m} \end{bmatrix}^{-1} \begin{bmatrix} m \\ X \end{bmatrix} \quad (3.59)$$

Defining the augmented vectors \tilde{X} and \tilde{Y} , and the matrix \tilde{W} as

$$\tilde{X} = \begin{bmatrix} m \\ X \end{bmatrix}, \quad \tilde{Y} = \begin{bmatrix} y \\ X \end{bmatrix}, \quad \text{and} \quad \tilde{W} = \begin{bmatrix} I_N + D & C \\ 0 & I_{(L-1)m} \end{bmatrix}$$

(3.59) can be expressed in the compact form

$$\tilde{Y} = \tilde{W}^{-1} \tilde{X} = \tilde{\tilde{W}} \tilde{X} \quad \text{where} \quad \tilde{\tilde{W}} = \tilde{W}^{-1} \quad (3.60)$$

Using the natural gradient, the update law for $\tilde{\tilde{W}}$ is

$$\begin{aligned}\Delta \tilde{\tilde{W}} &= \eta \left[\tilde{\tilde{W}}^{-T} - \varphi(\tilde{Y}) \tilde{X}^T \right] \tilde{\tilde{W}}^T \tilde{\tilde{W}}, \text{ or} \\ \Delta \tilde{\tilde{W}} &= \eta \left[I_{M+N} - \varphi(\tilde{Y}) \tilde{X}^T \tilde{\tilde{W}}^T \right] \tilde{\tilde{W}}\end{aligned}\quad (3.61)$$

Differentiating $\tilde{\tilde{W}}\tilde{\tilde{W}} = I$

$$\tilde{\tilde{W}}' \tilde{\tilde{W}} + \tilde{\tilde{W}} \tilde{\tilde{W}}' = 0$$

and by rearranging terms

$$\tilde{\tilde{W}}' = -\tilde{\tilde{W}} \tilde{\tilde{W}}' \tilde{\tilde{W}}^{-1} = -\tilde{\tilde{W}}^{-1} \tilde{\tilde{W}}' \tilde{\tilde{W}}^{-1}\quad (3.62)$$

Using 1st order Euler approximation for the derivatives in (3.62), the variation in $\tilde{\tilde{W}}$ can be expressed as

$$\Delta \tilde{\tilde{W}} = -\tilde{\tilde{W}}^{-1} (\Delta \tilde{\tilde{W}}) \tilde{\tilde{W}}^{-1}\quad (3.63)$$

Also note that from (3.60)

$$\tilde{Y}^T = \tilde{X}^T \tilde{\tilde{W}}^T\quad (3.64)$$

Using (3.62) and (3.64), the update law in (3.61) becomes

$$-\tilde{\tilde{W}}^{-1} (\Delta \tilde{\tilde{W}}) \tilde{\tilde{W}}^{-1} = \eta \left[I - \varphi(\tilde{Y}) \tilde{Y}^T \right] \tilde{\tilde{W}}, \text{ or}$$

$$-\tilde{\tilde{W}}^{-1} (\Delta \tilde{\tilde{W}}) = \eta \left[I - \varphi(\tilde{Y}) \tilde{Y}^T \right]$$

since $\tilde{\tilde{W}}\tilde{\tilde{W}} = I$.

Arranging terms

$$\Delta \tilde{\tilde{W}} = -\eta \tilde{\tilde{W}} \left[I - \varphi(\tilde{Y}) \tilde{Y}^T \right] = \eta \tilde{\tilde{W}} \left[\varphi(\tilde{Y}) \tilde{Y}^T - I \right]\quad (3.65)$$

Inserting the definition of $\tilde{\tilde{W}}$ and \tilde{Y} , (3.65) becomes

Therefo

rows bl

3.3.2.2

By equa

O

Δ

φ

3.3.2.2.1

Observe

suppleme

case, hov

network

feedback

signify th

decorrelate

$$\Delta \begin{bmatrix} I_N + D & C \\ 0 & I_{(L-1)m} \end{bmatrix} = \eta \begin{bmatrix} I_N + D & C \\ 0 & I_{(L-1)m} \end{bmatrix} \times \begin{bmatrix} \varphi(y)y^T - I_N & \varphi(y)X^T \\ \varphi(X)y^T & \varphi(X)X^T - I_{(L-1)m} \end{bmatrix}$$

Therefore the update laws for the matrices C and D are extracted by equating the first rows block wise to get

$$\Delta(I_N + D) = \Delta D = \eta \left[(I_N + D)(\varphi(y)y^T - I_N) + C\varphi(X)y^T \right] \quad (3.66)$$

$$\Delta C = \eta [(I_N + D)\varphi(y)X^T + C(\varphi(X)X^T - I_{(L-1)m})] \quad (3.67)$$

3.3.2.2 Auxiliary Conditions for Convergence

By equating the second row block wise, one gets

$$0_{(L-1)m \times N} = \varphi(X)_{(L-1)m \times 1} y_{1 \times N}^T \quad (3.68)$$

$$\Delta I_{(L-1)m} = 0 = \varphi(X)X^T - I_{(L-1)m}, \quad \text{or}$$

$$\varphi(X)X^T = I_{(L-1)m} \quad (3.69)$$

3.3.2.2.1 Remark 3.3:

Observe that (3.68) and (3.69) are identical to (3.51) and (3.52). These relations form *supplementary conditions* for the convergence of the proposed feedback algorithm. In this case, however, the states of the feedback structure are a “function” of the previous network outputs, i.e., $X_i(k) = \zeta(y_{k-i})$. Therefore, the auxiliary conditions in the feedback case define nonlinear correlation conditions on the demixing outputs that signify that the expectation of the demixing network outputs will be (nonlinearly) decorrelated leading to the components being stochastically independent.

Un
in the deri
conditions
but also in
final updat

$$\Delta D$$

$$\Delta C$$

The
earlier prop
(3.71) exhib

3.4 Blind

Pha

3.4.1 TH

*Assume the
function. TH
using the na*

$$\Delta C_i$$

$$\Delta D =$$

where the ste

$$C = [$$

Unlike the feedforward update laws, however, these conditions appear explicitly in the derived update laws (3.66) and (3.67). Therefore, enforcing the derived auxiliary conditions in the update laws (3.66) and (3.67) not only simplifies them computationally, but also improves the convergence dynamic properties of the proposed algorithm. The final update law for the feedback structure is given by

$$\Delta D = \eta \left[(I_N + D)(\varphi(y)y^T - I_N) \right] \quad (3.70)$$

$$\Delta C = \eta [(I_N + D)\varphi(y)X^T] \quad (3.71)$$

The above natural gradient update laws for feedback BSR are different from earlier proposed feedback (or recurrent) update laws in [35, 75]. The algorithm (3.70) and (3.71) exhibits better stability and faster convergence. See [123] for a comparison.

3.4 Blind Source Recovery Algorithms for Non-minimum Phase Environments

3.4.1 Theorem 3.3: Feedforward Non-minimum Phase Demixing

Assume the MIMO (mixing) environment modeled by a non-minimum phase transfer function. Then, the update law for the zeros of the state space FIR demixing network using the natural gradient is given by

$$\Delta C_i = \eta(k) \left[C_i - \varphi(y(k))u(k-i)^T \right], i = 1, 2, \dots, L-1 \quad (3.72)$$

$$\Delta D = \eta(k) \left[D - \varphi(y(k))u(k)^T \right] \quad (3.73)$$

where the state space matrix C is defined as

$$C = [C_1 \quad C_2 \quad \dots \quad C_{L-1}] \quad (3.74)$$

$C_j - b_j$

$u(k) -$

or eqn

when

given

3.4.

On

input

(3.7)

need

time

is

ini

the

the

(3

(3

C_i - being the MIMO FIR filter coefficients corresponding to delay z^{-i}

$u(k)$ – represents an information back-propagation filter, given as

$$u(k) = \sum_{i=1}^{L-1} C_i^T y(k+i) + D^T y(k)$$

or equivalently in the adjoint state space as

$$\lambda(k-1) = A^T \lambda(k) + C^T y(k)$$

$$u(k) = B^T \lambda(k) + D^T y(k)$$

where T represents the matrix Hermitian operator, $\varphi(y)$ is the appropriate score function given by (3.37) and A and B are given by (43), with $A_{1j} = 0$.

3.4.1.1 Derivation of Update Laws for Feedforward Non-minimum Phase BSR

On the outset, please note that the algorithm requires a frame of $3L-2$ samples of the input signal to initiate the computations of the update of the parameter matrices in (3.72)-(3.73). Note also that the time index of the parameter update (3.72)-(3.73) is not necessarily the same time index of the signals. In practice, it may be chosen so that the time index of the parameters is a delayed version of the signal time index so that no issues of non-causality would arise. Also in practice, see Simulation III in section 4.2, the initialization of the update law (3.72)-(3.73) uses an identity (or a nonsingular matrix) for the center tap matrix (specifically, $j = \lfloor (L-1)/2 \rfloor$) and small random or zero values for the remaining tap matrices.

Let the mixing environment to assume the same state space representation as in (3.24)-(3.25), while the demixing network assumes the state space representation (3.26)-(3.27), repeated below for convenience

network

intend

and B

where

I_m : 10

O_m : 2

frequ

where

$\bar{W}(z)$

i^{th} filt

matrix

$$X(k+1) = A X(k) + B m(k)$$

$$y(k) = C X(k) + D m(k)$$

As mentioned before, for the non-minimum phase mixing case, the demixing network is constrained to be a double-sided FIR filter [22, 156] so as to approximate the intended unstable IIR inverse. Using the MIMO controller canonical form, the matrices A and B take the form

$$A = \begin{bmatrix} O_m & O_m & \cdots & O_m & O_m \\ I_m & O_m & \cdots & O_m & O_m \\ O_m & I_m & \cdots & O_m & O_m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ O_m & O_m & \cdots & I_m & O_m \end{bmatrix}, B = \begin{bmatrix} I_m \\ O_m \\ O_m \\ \vdots \\ O_m \end{bmatrix}$$

where

I_m : Identity matrix of dimension $m \times m$

O_m : Zero matrix of dimension $m \times m$

For compactness, we can represent (3.26)-(3.27) using the compact (mixed time-frequency) transfer function notation of the MIMO FIR filter as

$$y_k = [\bar{W}(z)] m_k \quad (3.75)$$

where, $\bar{W}(z)$ is the transfer function of the demixing MIMO FIR filter of the form,

$$\bar{W}(z) = W_0 + W_1 z^{-1} + \dots + W_{L-1} z^{-L+1} \quad (3.76)$$

Further we can define a matrix \bar{W} , comprised of the coefficients of the filter $\bar{W}(z)$. Note that each constituent sub-matrix W_i contains co-efficients pertaining to the i^{th} filter lag acting on the time instant $k-i$. This matrix \bar{W} is related to the state space matrices C and D as given below.

wh

as

For

tran

i.e.

wh

coe

tim

cau

Usi

don

wh

the

$$\bar{W} \triangleq [W_0 \ W_1 \ \dots \ W_{L-1}] = [D \ C]$$

where the state space matrices C and D in terms of the sub-matrices W_i are represented

as

$$\begin{aligned} D &= W_0 \\ C &= [W_1 \ W_2 \ \dots \ W_{L-1}] \end{aligned} \quad (3.77)$$

For the MIMO transfer function $\bar{W}(z)$, the conjugate (or Hermitian) transpose of the transfer function can be defined in a fashion similar to the one proposed by Lambert [73],

i.e.,

$$\begin{aligned} [\bar{W}(z)]^T &= [W_0 + W_1 z^{-1} + \dots + W_{L-1} z^{-L+1}]^T \\ &= W_0^T + W_1^T z^1 + \dots + W_{L-1}^T z^{L-1} \\ &= \bar{W}^T(z^{-1}) \end{aligned} \quad (3.78)$$

where, T is the Hermitian transpose operator. Notice that again we can define a matrix of coefficients \bar{W}^T comprised of the transpose of the constituent sub-matrices acting in a time reversed fashion, i.e., the i^{th} lag of this time reversed filter acts on the future (non-causal) time instant $k+i$, i.e.,

$$\bar{W}^T \triangleq [W_0^T \ W_1^T \ \dots \ W_{L-1}^T] \quad (3.79)$$

Using the definition of \bar{W} , we can express the network output equation (3.75) in the time domain as

$$y_k = \bar{W} \underline{m}_k \equiv D m_k + \sum_{i=1}^{L-1} C_i m_{k-i} \quad (3.80)$$

where \underline{m}_k represents an $Lm \times 1$ matrix of observations at the k^{th} iteration, which includes the k^{th} and previous $L-1$ m -d measurement (or observation) vectors, i.e.,

n

y_k : repre

Fo

divergenc

maximiza

H

where q_j

$j \in [0, L -$

Usi

47, 105], th

ΔH

where δ_{ji}

It

minimum

$j = \lfloor (L -$

will be se

(3.91) or

will includ

filter matri

$$\underline{m}_k \triangleq \left[m_k^T \quad m_{k-1}^T \quad \cdots \quad m_{k-L+1}^T \right]^T$$

y_k : represents the output vector at the k^{th} iteration

For the derivation of the update laws, we minimize the Kullback-Liebler divergence with the constraints of an FIR representation as described above. Using the maximization of entropy approach, the Hamiltonian can be defined [9, 47, 105, 109] as

$$H^k = -\log \det |W_j| - \sum_i \log(q_i(y_i, W(z))) \quad (3.81)$$

where $q_i(y_i, W(z))$ is the online estimated marginal distribution of $y_i(k)$ and $j \in [0, L-1]$.

Using the ordinary stochastic gradient update laws (3.35) and (3.36), also see [9, 47, 105], the update law for the sub-matrices W_i is given by

$$\Delta W_i = -\eta_k \frac{\partial H^k}{\partial W_i} = \eta_k \left[\left(W_j^{-1} \right)^T \delta_{ji} - \varphi(y_k) m_{k-i}^T \right] \quad (3.82)$$

where δ_{ji} equals 1 only when $i = j$, and $j \in [0, L-1]$.

It should be noted that W_j is required here to be nonsingular. For the non-minimum phase case, j is appropriately chosen to be an intermediate tap, e.g. $j = \lfloor (L-1)/2 \rfloor$, in order to render an appropriate double-sided FIR filter. In any event, it will be seen below that using the natural gradient, the resulting parameter update laws (3.91) or (3.92) do not depend on the explicit nonsingularity of the j -th tap matrix and will include all tap matrices. The Stochastic Gradient (SG) law for the complete MIMO filter matrix is given by

modi

Riem

where

while

where

propa

$$[\Delta \bar{W}(z)]_{SG} = \sum_{i=0}^{L-1} \Delta W_i z^{-i} = \eta_k \left[\left(W_j^{-1} \right)^T \delta_{ji} - \sum_{i=0}^{L-1} \varphi(y_k) m_{k-i}^T z^{-i} \right] \quad (3.83)$$

Using the definition of the natural gradient in the systems space [9], we can modify (3.83) to have the update of the demixing filter parameters according to the Riemannian contra-variant gradient as

$$\begin{aligned} \Delta \bar{W}(z) &= [\Delta \bar{W}(z)]_{SG} \bar{W}^T(z^{-1}) \bar{W}(z) \\ &= \eta_k \left[\left(W_j^{-1} \right)^T \delta_{ji} - \sum_{i=0}^{L-1} \varphi(y_k) m_{k-i}^T z^{-i} \right] \bar{W}^T(z^{-1}) \bar{W}(z) \\ &= \eta_k \left[\left(W_j^{-1} \right)^T \delta_{ji} \bar{W}^T(z^{-1}) \bar{W}(z) - \sum_{i=0}^{L-1} \varphi(y_k) m_{k-i}^T z^{-i} \bar{W}^T(z^{-1}) \bar{W}(z) \right] \end{aligned} \quad (3.84)$$

where, the first term on the right hand side can be simplified as

$$\begin{aligned} \left(W_j^{-1} \right)^T \delta_{ji} \bar{W}^T(z^{-1}) \bar{W}(z) &= \left(W_j^{-1} \right)^T \left(\delta_{ji} \bar{W}^T(z) \right) \bar{W}(z) \\ &= \left(W_j^{-1} \right)^T W_j^T \bar{W}(z) = I^T \bar{W}(z) = \bar{W}(z) \end{aligned} \quad (3.85)$$

while the second term can be simplified as

$$\begin{aligned} \sum_{i=0}^{L-1} \varphi(y_k) m_{k-i}^T z^{-i} \bar{W}^T(z^{-1}) \bar{W}(z) &= \varphi(y_k) \sum_{i=0}^{L-1} \left([\bar{W}(z)] m_{k-i} \right)^T \bar{W}(z) z^{-i} \\ &= \varphi(y_k) \sum_{i=0}^{L-1} y_{k-i}^T \bar{W}(z) z^{-i} \\ &= \varphi(y_k) \sum_{i=0}^{L-1} \left([\bar{W}^T(z^{-1})] y_{k-i} \right)^T z^{-i} \\ &= \varphi(y_k) \sum_{i=0}^{L-1} u_{k-i}^T z^{-i} \end{aligned} \quad (3.86)$$

where, we used equality (3.75) and u_k^T is the transpose of the non-causal back-propagation filter output, defined as

the k

then

Usin

tran

req

imp

upc

am

pra

Wr

W_i

$$u_k \triangleq \left[\bar{W}^T(z^{-1}) \right] y_k \equiv D^T y_k + \sum_{i=1}^{L-1} C_i^T y_{k+i} \quad (3.87)$$

Let \underline{y}_k represent an $NL \times 1$ matrix of outputs at the k^{th} iteration, which includes the k^{th} and the future $L-1$ N -d output vectors, i.e.,

$$\underline{y}_k \triangleq \left[y_k^T \quad y_{k+1}^T \quad \cdots \quad y_{k+L-1}^T \right]^T$$

then the back-propagation filter can also be expressed in the compact form

$$u(k) = \bar{W}^T \underline{y}_k \quad (3.88)$$

Using (3.85) and (3.86) in (3.84), we get the following update laws for the MIMO FIR transfer Function

$$\Delta \bar{W}(z) = \eta_k \left[\bar{W}(z) - \sum_{i=0}^{L-1} \varphi(y_k) u_{k-i}^T z^{-i} \right] \quad (3.89)$$

The update law (3.89) is non-causal as computation of $u_{k-i}^T; i = 0, 1, 2, \dots, L-1$ requires up to $L-1$ future outputs to be available. Practically the update law can be implemented (or made causal) by introducing a delay of $L-1$ iterations between the update of the parameters and the computation of the output at the k^{th} iteration (this in fact amount of using a frame-buffering in order to initialize the update law as is used in practice). Thus, a causal version of (3.89) becomes

$$\left[\Delta \bar{W}(z) \right]_C = \eta_k \left[\bar{W}(z) - \sum_{i=0}^{L-1} \varphi(y_{k-L+1}) u_{k-L-i+1}^T z^{-i} \right] \quad (3.90)$$

Writing down the update laws in the time domain for the i^{th} component coefficient matrix W_i , we have

$$[\Delta W_i]_C = \eta_k \left[W_i - \varphi(y_{k-L+1}) u_{k-L-i+1}^T \right] \quad (3.91)$$

Defining the new time index $\bar{k} = k - L + 1$, we have

$$[\Delta W_i]_C = \eta_k \left[W_i - \varphi(y(\bar{k})) u^T(\bar{k} - i) \right] \quad (3.92)$$

where the computation of the network output $y(\bar{k})$ and the back-propagation filter output $u(\bar{k})$ is computed as in the following expressions

$$y(\bar{k}) \triangleq y(k - L + 1) = \bar{W} \underline{m}_{k-L+1} = \sum_{i=1}^{L-1} C_i m(k - L + 1 - i) + D m(k - L + 1) \quad (3.93)$$

$$u(\bar{k}) \triangleq u(k - L + 1) = \bar{W}^T \underline{y}_{k-L+1} = \sum_{i=1}^{L-1} C_i^T y(k - L + 1 + i) + D^T y(k - L + 1) \quad (3.94)$$

Therefore the explicit causal update laws for the state space matrices C and D , using the time index \bar{k} are

$$\Delta C_i = \eta_k \left[C_i - \varphi(y(\bar{k})) u(\bar{k} - i)^T \right], i = 1, 2, \dots, L - 1 \quad (3.95)$$

$$\Delta D = \eta_k \left[D - \varphi(y(\bar{k})) u(\bar{k})^T \right] \quad (3.96)$$

In the state space regime, the information back-propagation filter assumes the form

$$\lambda(\bar{k} - 1) = A^T \lambda(\bar{k}) + C^T y(\bar{k}) \quad (3.97)$$

$$u(\bar{k}) = B^T \lambda(\bar{k}) + D^T y(\bar{k}) \quad (3.98)$$

The non-minimum natural gradient algorithms [14, 126, 129, 130, 154-156] require both forward and backward in time propagation by the adaptive MIMO FIR filter. One would require a buffer of $3L-2$ input samples to initiate the adaptation according to the update laws.

encomp

conven

any los

been ig

where

minim

consti

proces

phase

to sho

3.4.1.

It sho

FIR S

In tha

simila

The non-minimum phase state space BSR update laws, being more general, also encompass the domain of the minimum phase mixing network case. This can be verified conveniently by using definition (3.87) in the update law for the matrix D . Note, without any loss of generality, the delay introduced in the update law (3.96) to make it causal has been ignored for clarity.

$$\begin{aligned}
\Delta D &= \eta_k \left[D - \varphi(y(k))u(k)^T \right] \\
&= \eta_k \left[D - \varphi(y(k)) \left[D^T y(k) + \sum_{i=1}^{L-1} C_i^T y(k+i) \right]^T \right] \\
&= \eta_k \left[D - \varphi(y(k)) \left(y(k)^T D + \sum_{i=1}^{L-1} y(k+i)^T C_i \right) \right] \\
&= \eta_k \left[\left(I - \varphi(y(k))y(k)^T \right) D - \varphi(y(k)) \sum_{i=1}^{L-1} y(k+i)^T C_i \right]
\end{aligned} \tag{3.99}$$

where the first term on the right hand side of equation (3.99) comprises the instantaneous minimum phase update of the matrix D (see equation (3.48)), while the second term constitutes the “non-causal” update term corresponding to the back-propagating processing required for the double sided FIR demixing filter in the case of non-minimum phase environments. Similarly, the update law (3.95) can be expanded using (3.80)-(3.87) to show the explicit relationship to the minimum phase feedforward update law (3.47).

3.4.1.1.1 *Remark 3.4:*

It should be noted that Equation (3.84) holds for the doubly infinite filter space. For an FIR Space domain, it does not satisfy a Lie group in the sense of the formulation in [154]. In that event, the matrix product in (3.84) is defined differently and consequently a similar, but different, update law results.

3.5

MAT

all thr

follow

when

A_i, E

aver

$v(k)$

m, n

dep

aux

pol

kn

rep

we

eff

13

ga

co

3.5 Simulation Results

MATLAB simulation results for the three linear BSR algorithms are being presented. In all three cases the environment comprises of a 3×3 IIR mixing convolving filter with the following model [110, 112, 155].

$$\sum_{j=0}^{m-1} A_j m(k-i) = \sum_{i=0}^{n-1} B_i s(k-i) + v(k) \quad (3.100)$$

where

A_i, B_i - are the co-efficient matrices for the i^{th} tap of the autoregressive and moving average sections of the filter respectively, and

$v(k)$ - is the additive gaussian measurement noise.

m, n - are the lengths of the auto-regressive and moving average lags in the filter.

The demixing network for each simulation can be initialized in a number of ways, depending on the available environment information, which may be collected via auxiliary means [155, 157]. This may include complete, partial, or no knowledge of the poles of the mixing environment [126, 135, 139]. The primary advantage of incorporating knowledge of poles in a BSR algorithm is to obtain a more compact demixing network representation. All the three cases have been verified via simulation and the algorithms were able to converge satisfactorily using *online* update algorithms. However, to make efficient use of space only typical results are being presented here, see [110, 112, 128-130, 135, 139, 140]. Unless otherwise specified, the original sources had sub-gaussian, gaussian and super-gaussian densities respectively.

The convergence performance of the BSR algorithms, for the linear and convolutive mixing, is measured using the multi-channel intersymbol interference (MISI)

benchma

permutati

A

$$G(z) = H$$

$$\bar{H}(z) = [$$

$$H(z) = [$$

A

depend o

converge

simulatio

on the b

(3.102)

outputs

where

$\kappa_4(y)$:

γ : repr

benchmark. MISI is a measure of the global transfer function diagonalization and permutation as achieved by the demixing network and is defined as

$$MISI_k = \sum_{i=1}^N \frac{\left| \sum_j \sum_p |G_{p ij}| - \max_{p,j} |G_{p ij}| \right|}{\max_{p,j} |G_{p ij}|} + \sum_{j=1}^N \frac{\left| \sum_i \sum_p |G_{p ij}| - \max_{p,i} |G_{p ij}| \right|}{\max_{p,i} |G_{p ij}|} \quad (3.101)$$

$G(z) = H(z) * \bar{H}(z)$ - Global Transfer Function,

$\bar{H}(z) = [A_e, B_e, C_e, D_e]$ - Transfer Function of Environment,

$H(z) = [A, B, C, D]$ - Transfer Function of Network.

As earlier discussed, the optimal score function for the derived update laws depend on the density function of the sources to be separated, which upon successful convergence of the algorithm is similar to the density of the separated outputs. For simulations, a hyperbolic score function [126, 127, 130, 136] has been used, which relies on the batch kurtosis of the output of the demixing system. This score function given by (3.102) converges to the optimal non-linearity for the demixing system as the network's outputs approach stochastic independence.

$$\varphi(y) = \begin{cases} y - \alpha \tanh(\beta y) & \kappa_4(y) \leq -\gamma \\ y & \text{for } |\kappa_4(y)| < \gamma \\ \alpha \tanh(\beta y) & \kappa_4(y) \geq \gamma \end{cases} \quad (3.102)$$

where

$\kappa_4(y)$: represents batch kurtosis of the output signals.

γ : represents a positive density classification threshold.

3.5.1 B

In this section

(3.100) are

A_0

B_0

In the

3 matrix with

scaled to a

map, the

environment

overall g

shown in

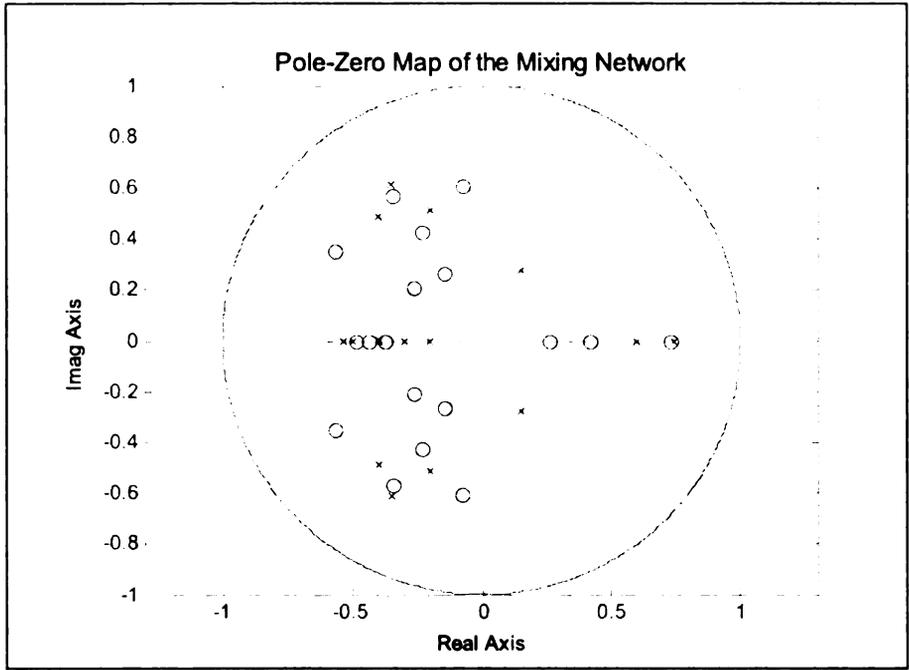
3.5.1 BSR Simulations – Minimum Phase Mixing Environments.

In this section, the BSR results for a minimum phase IIR mixing model of the form (3.100) are presented, where $m = n = 3$ and

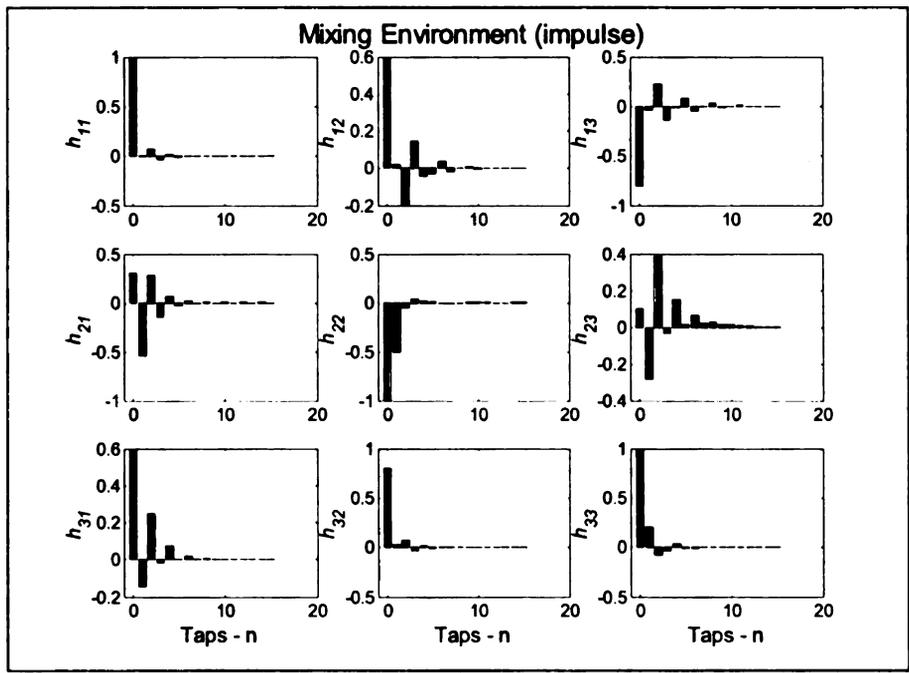
$$A_0 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, A_1 = \begin{bmatrix} 0.5 & 0.8 & -0.7 \\ 0.8 & 0.3 & -0.2 \\ -0.1 & -0.5 & 0.4 \end{bmatrix}, A_2 = \begin{bmatrix} 0.06 & 0.4 & -0.5 \\ 0.16 & -0.1 & -0.4 \\ -0.3 & -0.06 & 0.3 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.3 & 1 & 0.1 \\ 0.6 & -0.8 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} 0.5 & 0.5 & 0.6 \\ -0.3 & 0.2 & -0.3 \\ -0.2 & -0.43 & 0.6 \end{bmatrix}, B_2 = \begin{bmatrix} .125 & 0.06 & 0.2 \\ -0.1 & 0 & 0.4 \\ 0.08 & -0.13 & 0.3 \end{bmatrix}$$

In the transfer function domain, this MIMO transfer function will constitute a 3×3 matrix with each element being an order 18 IIR filter (provided each element has been scaled to account for all the environment transmission poles). The transmission pole-zero map, the impulse response of the environment transfer function, the theoretical environment inverse, the corresponding typical demixing network estimate, and the overall global transfer function using minimum phase BSR algorithms are respectively shown in Figure 3.4, 3.5 and 3.6.

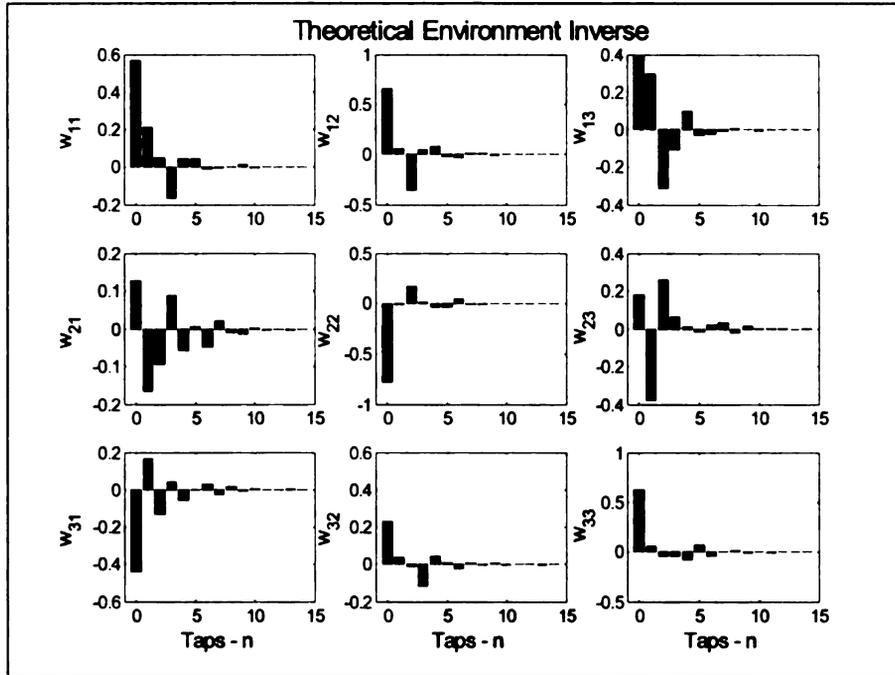


(a)

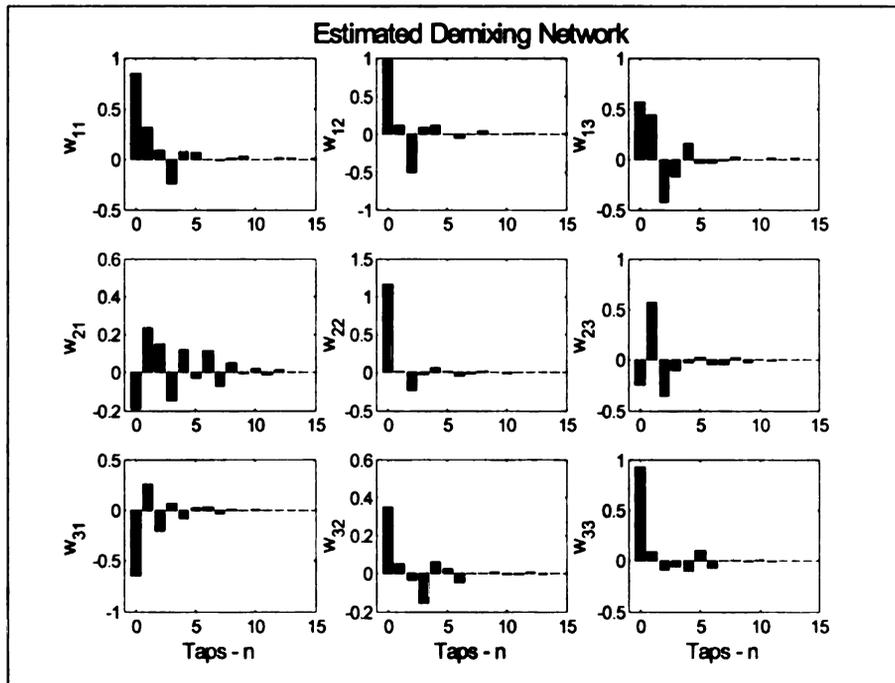


(b)

Figure 3.4. Minimum Phase Mixing Environment (a) Environment Transmission Zeros and Poles, (B) Impulse Response of the Environment Model



(a)



(b)

Figure 3.5. Minimum Phase Mixing Environment, Impulse Response of (a) Theoretical Environment Inverse, (b) Estimated Demixing Network

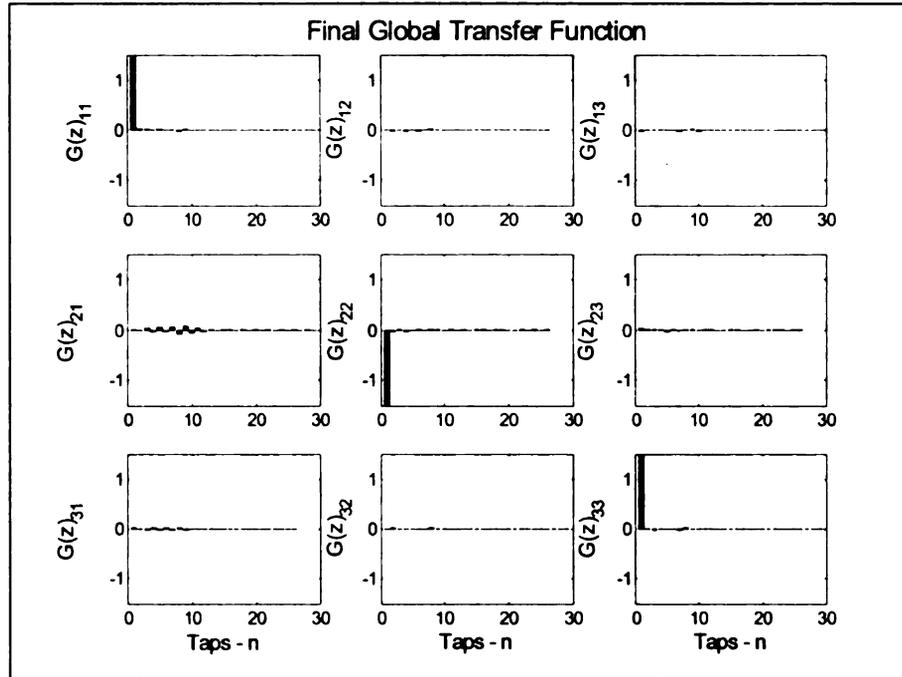
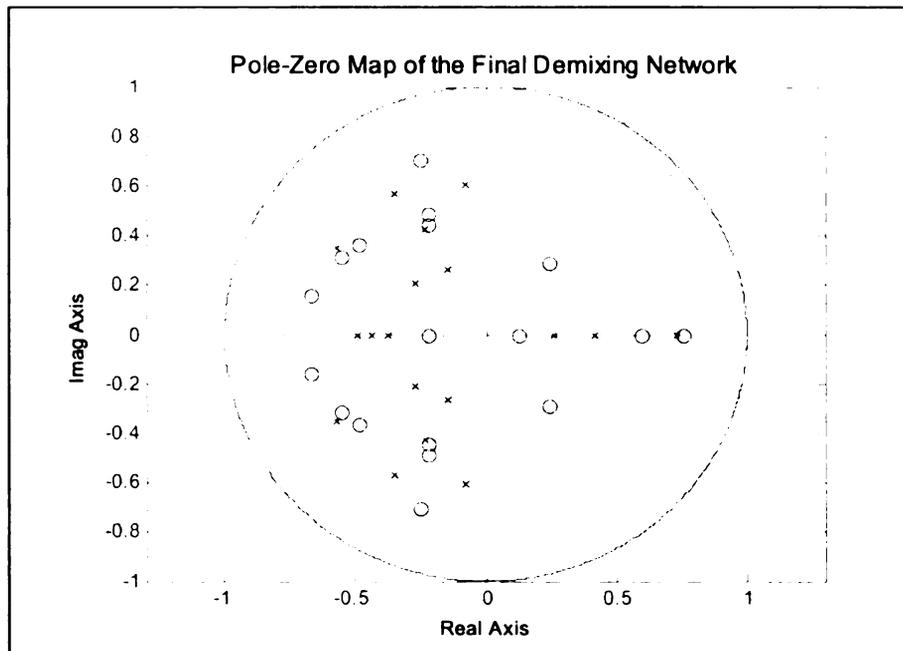


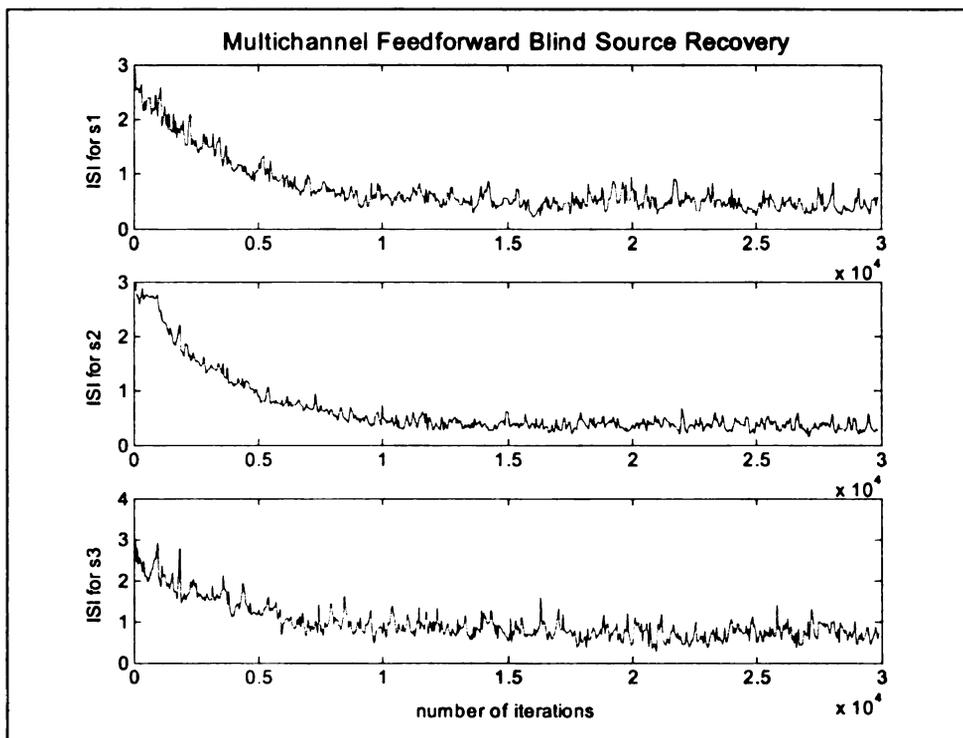
Figure 3.6. Minimum Phase Mixing Environment, Final Global Transfer Function

3.5.1.1 Simulation I: Feedforward Recovery of Minimum Phase Mixing

Using the minimum phase feedforward structure, three different simulation scenarios are being presented, where the poles of the demixing network have been fixed (a) at the solution, (b) all set to zero, or (c) set randomly but stable. In all cases, the matrix C is initialized to have small random elements, the matrix D is initialized to be the identity matrix, while matrices A and B are correspondingly set to be in Canonical Form I. The number of states in the demixing state space network was correspondingly chosen to be 18, 60, and 27. For all three cases, the pole-zero map of the final demixing network and the online MISI convergence characteristics are being presented in Figures 3.7-3.9 after 30,000 online iterations. It is observed that in all the three cases the global transfer function is diagonalized, although the demixing network have quite different pole-zero maps, indicating multiple possible solutions to the same problem. A quantitative comparative summary of recovered results has been summarized in Table 3.1.

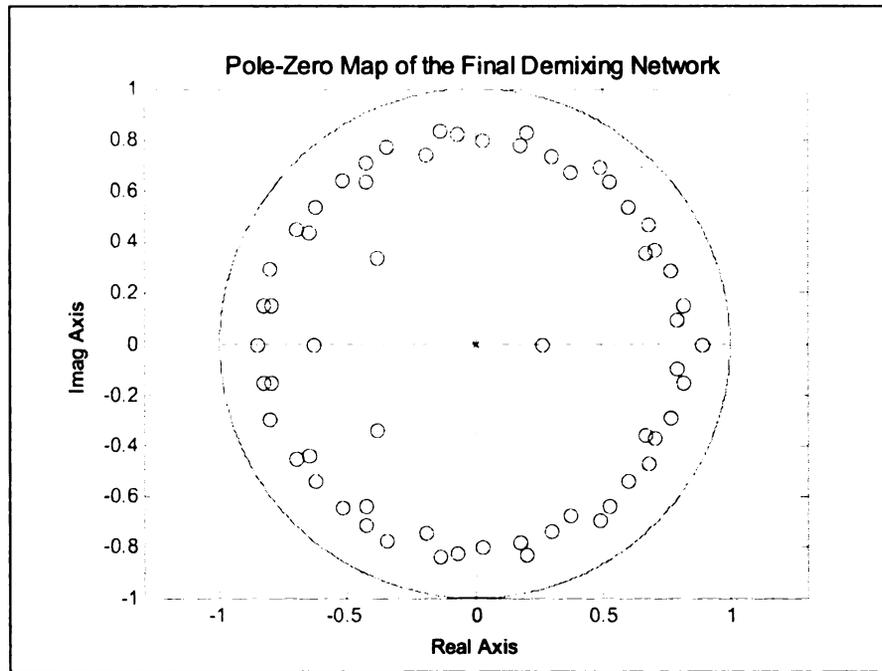


(a)

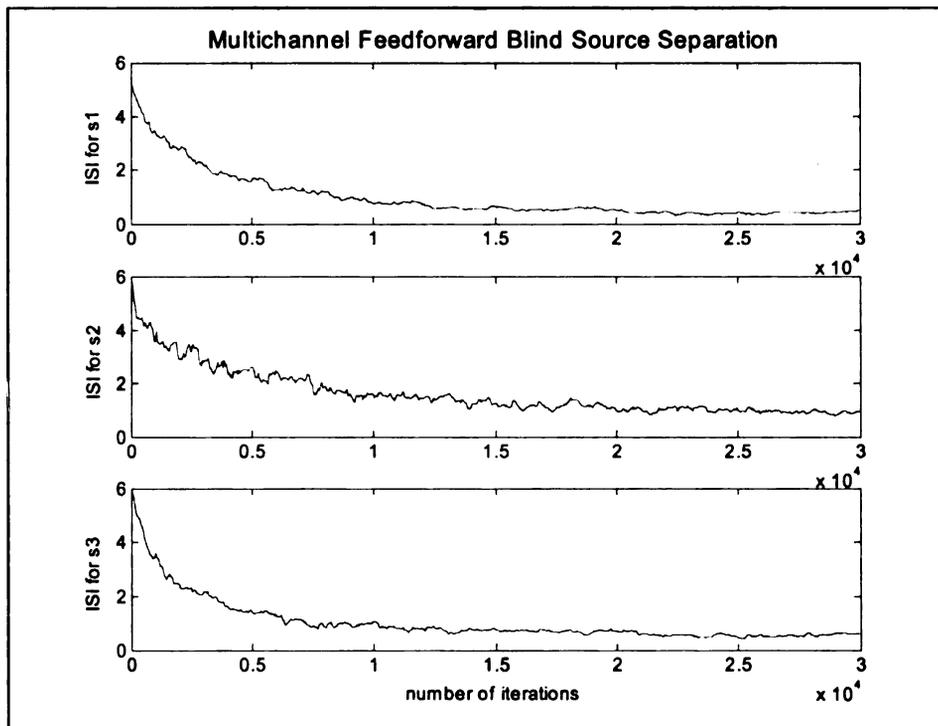


(b)

Figure 3.7. Minimum Phase Feedforward BSR – All poles at the theoretical solution: (a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index

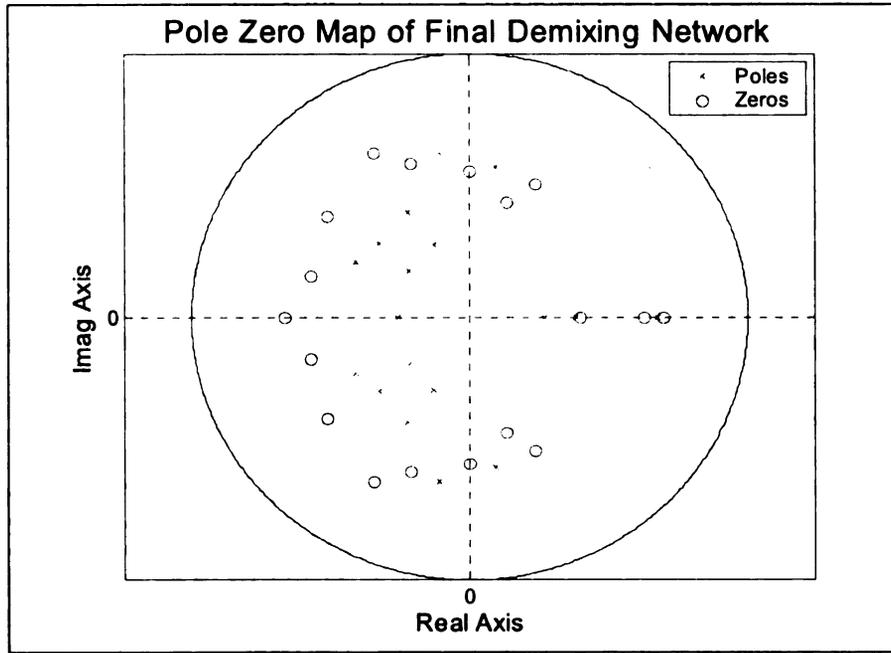


(a)

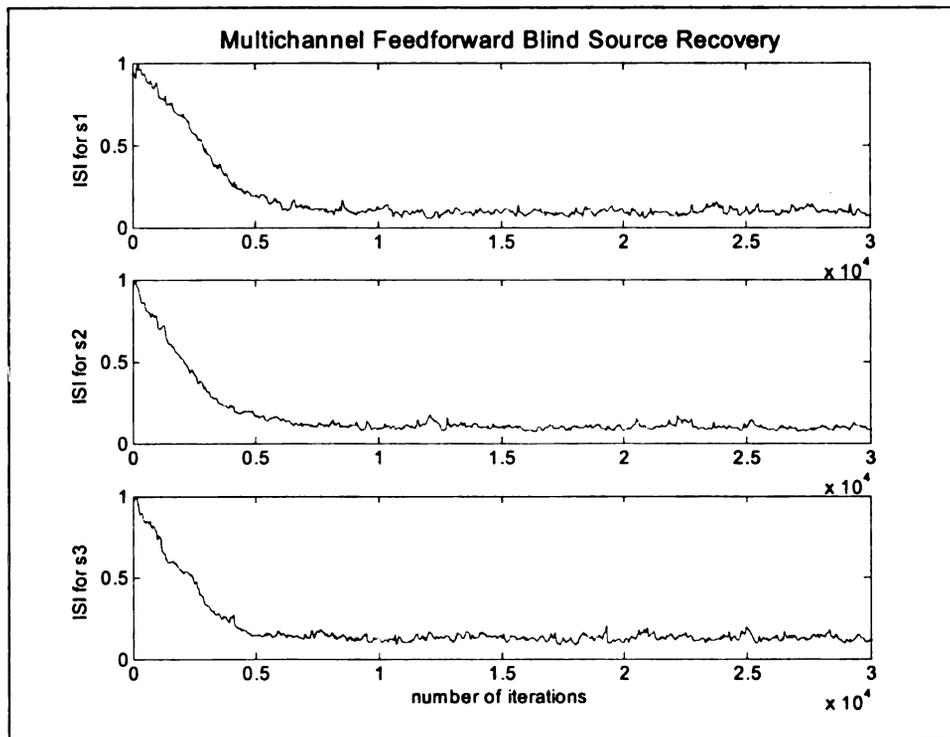


(b)

Figure 3.8. Minimum Phase Feedforward BSR – All poles set to zero: (a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index



(a)



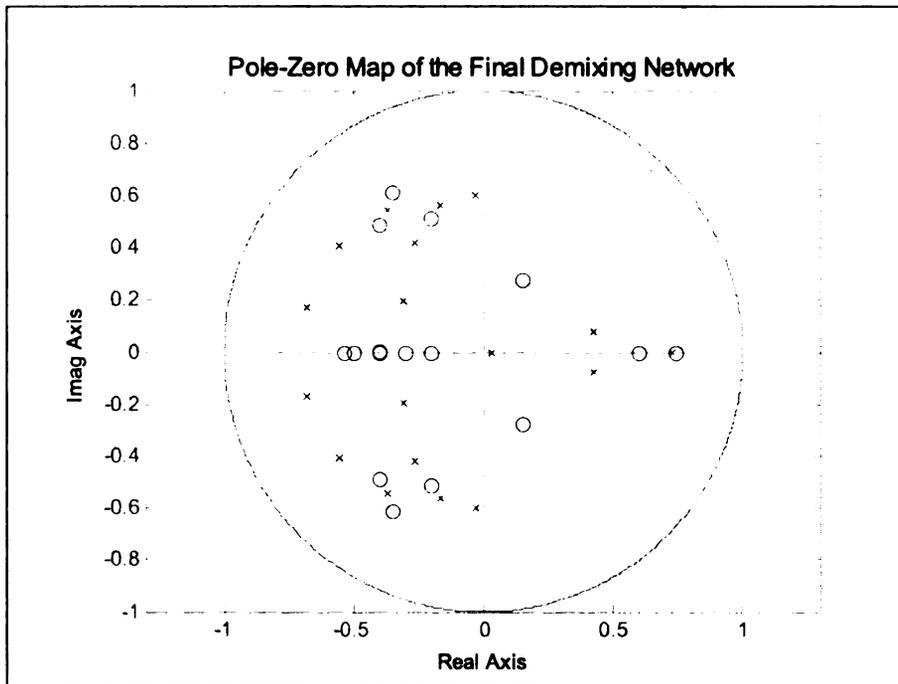
(b)

Figure 3.9. Minimum Phase Feedforward BSR – All poles initialized arbitrarily:

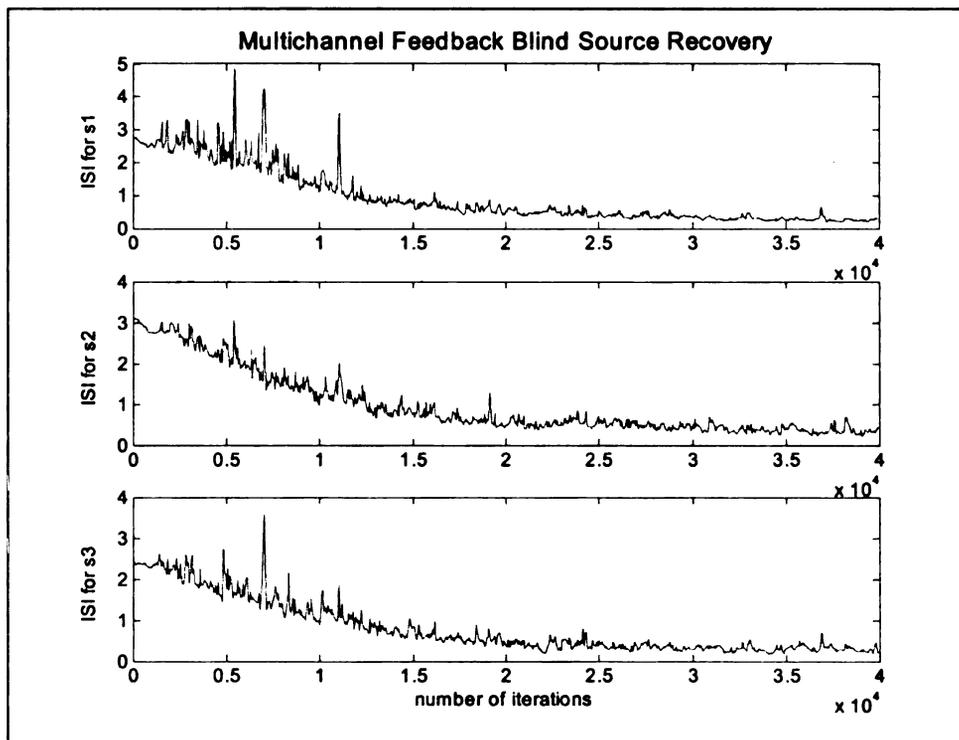
(a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index

3.5.1.2 Simulation II: Feedback Recovery of Minimum Phase Mixing

For the minimum phase feedback structure, similar three different simulation scenarios are being presented, where the demixing network poles have been (a) fixed at the solution, (b) all set to zero, or (c) set to be random but stable. The state space matrices and the number of states for the feedback network were chosen similar to the feedforward case. Note that again three different global diagonalizing demixing network pole-zero maps are obtained. The qualitative simulation result summary is presented in Figure 3.10-3.12. See Table 3.1 for a quantitative comparison summary.

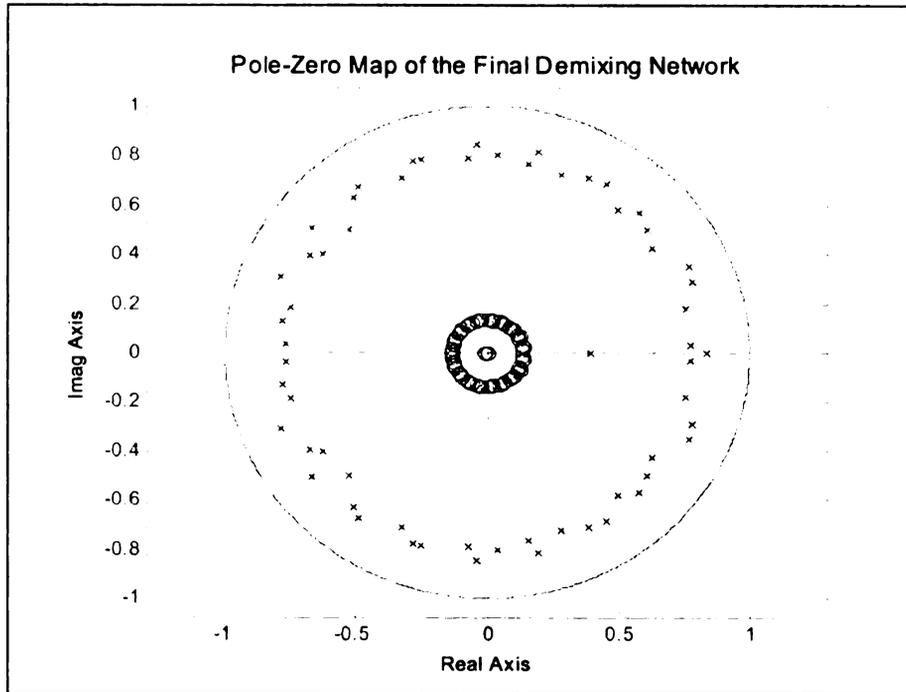


(a)

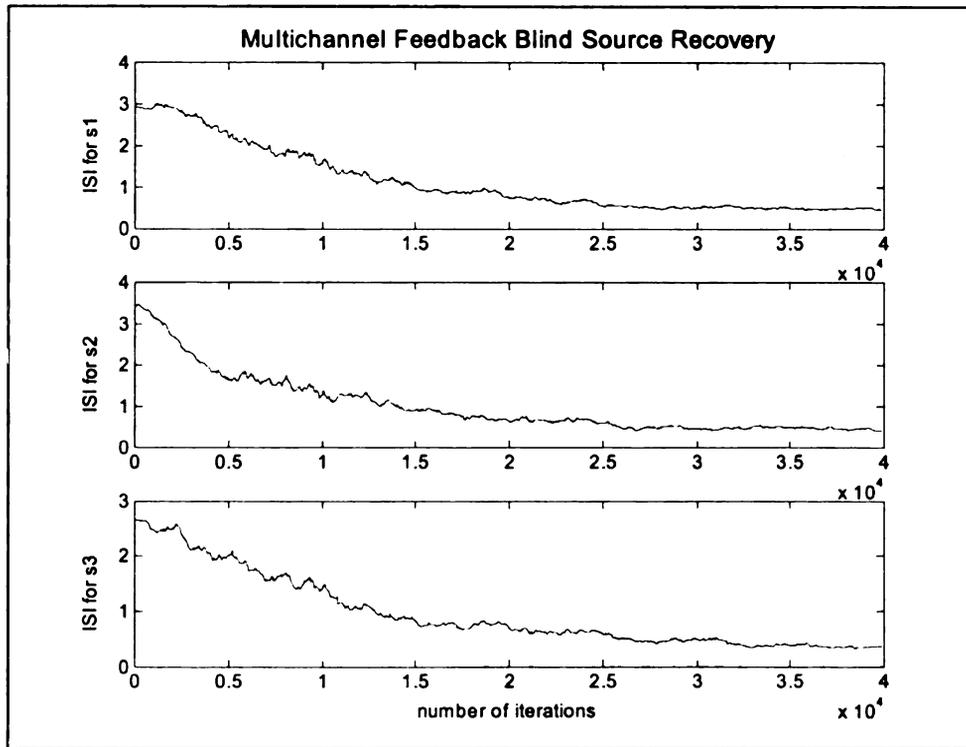


(b)

Figure 3.10. Minimum Phase Feedback BSR – All poles at the theoretical solution:
 (a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index

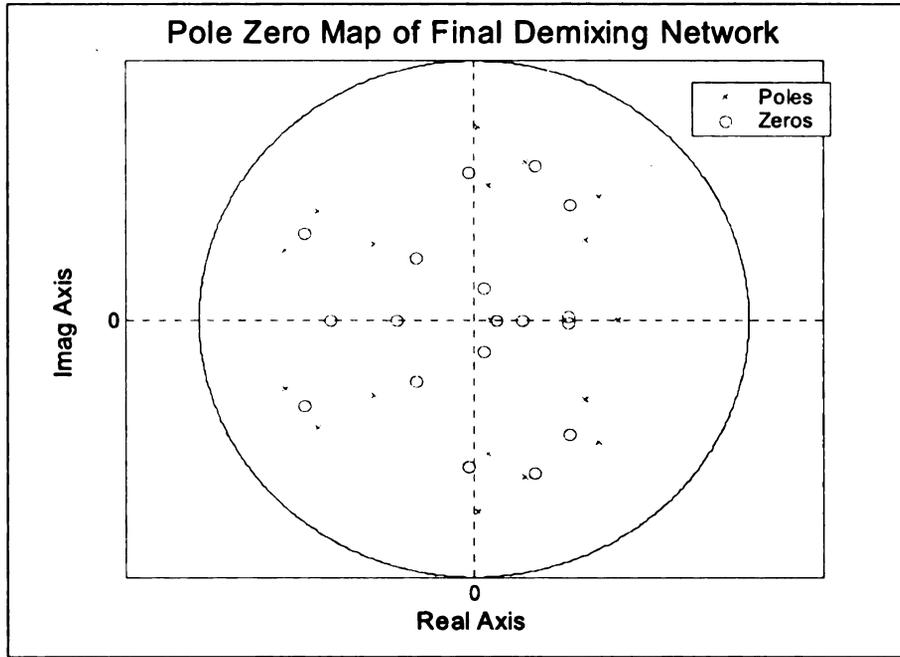


(a)

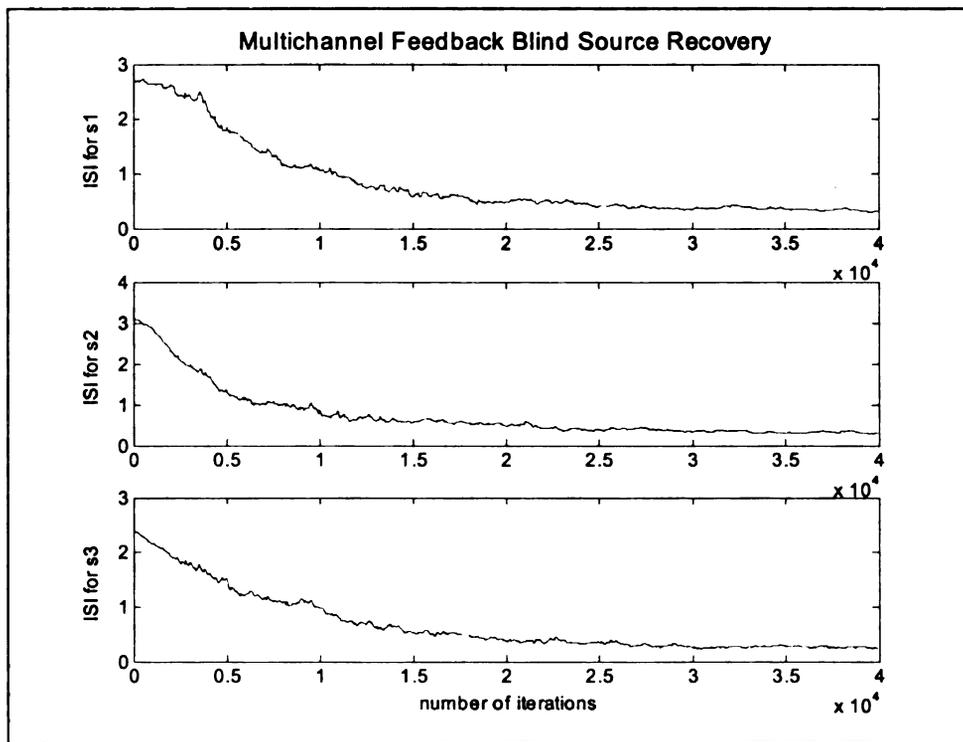


(b)

Figure 3.11. Minimum Phase Feedback BSR – All poles set at zero: (a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index



(a)



(b)

Figure 3.12. Minimum Phase Feedback BSR – All poles initialized arbitrarily:

(a) Demixing Network Transmission Pole-Zero Map, (b) Convergence of MISI Index

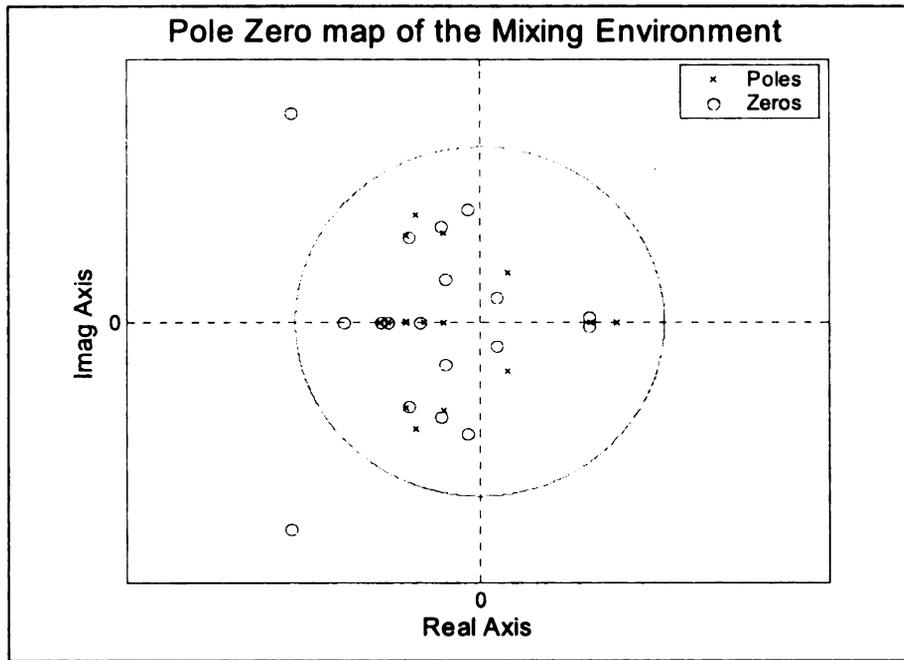
3.5.1.3 Simulation III: Feedforward Recovery of Non-minimum Phase Mixing

For the case of non-minimum phase, we again choose $m = n = 3$ for the IIR mixing environment model (3.100). In this case, the coefficient matrices are given by

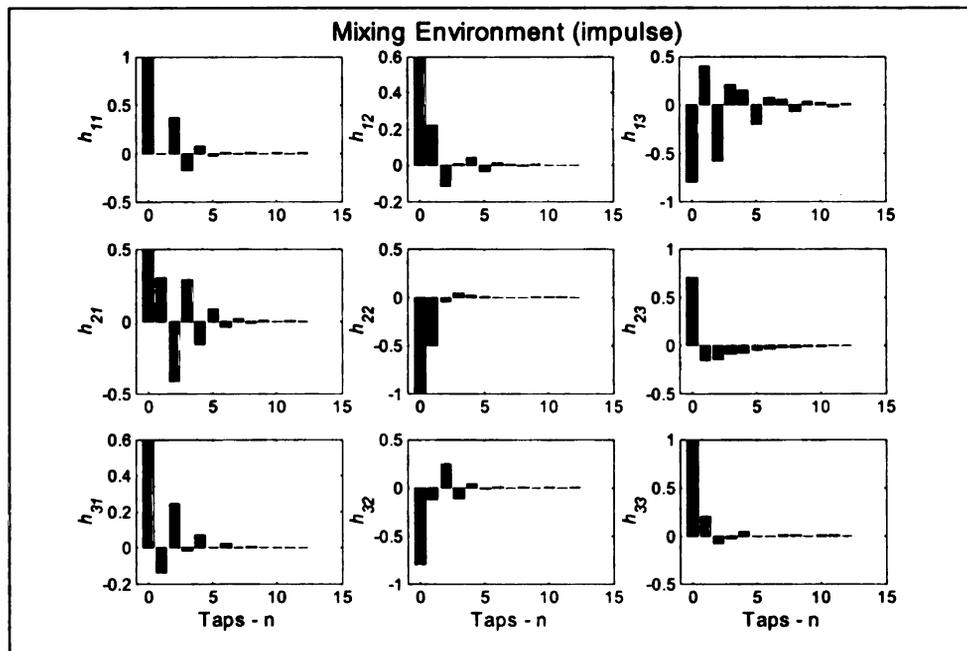
$$A_0 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, A_1 = \begin{bmatrix} 0.5 & 0.8 & -0.7 \\ 0.8 & 0.3 & -0.2 \\ -0.1 & -0.5 & 0.4 \end{bmatrix}, A_2 = \begin{bmatrix} 0.06 & 0.4 & -0.5 \\ 0.16 & -0.1 & -0.4 \\ -0.3 & -0.06 & 0.3 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.5 & 1 & 0.7 \\ 0.6 & 0.8 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} 0.5 & 0.7 & 0.16 \\ 0.7 & 0.2 & -0.3 \\ -0.2 & 0.53 & 0.6 \end{bmatrix}, B_2 = \begin{bmatrix} .425 & 0.3 & 0.7 \\ -0.1 & 0 & -0.4 \\ 0.08 & -0.13 & 0.3 \end{bmatrix}$$

The theoretical inverse of this environment model will be an unstable IIR filter with a minimum of 20 states. This instability stems from having two poles of the intended demixing network to be outside the unit circle. However, the demixing network is setup to be comprised of a doubly-finite 3×3 FIR demixing filter with 31 taps per filter (i.e., a total of 90 states) and supplied with mixtures of multiple source distributions. The polynomial matrix $\bar{W} = [D \ C]$ is initialized to be full rank, while A and B are in Canonical form I. Instantaneous update results after 40,000 iterations are presented in Figures 3.13-3.16. A close comparison of the theoretical environment inverse and the adaptively estimated inverse will reveal that the theoretical results form a part of the much larger (in taps), estimated inverse.

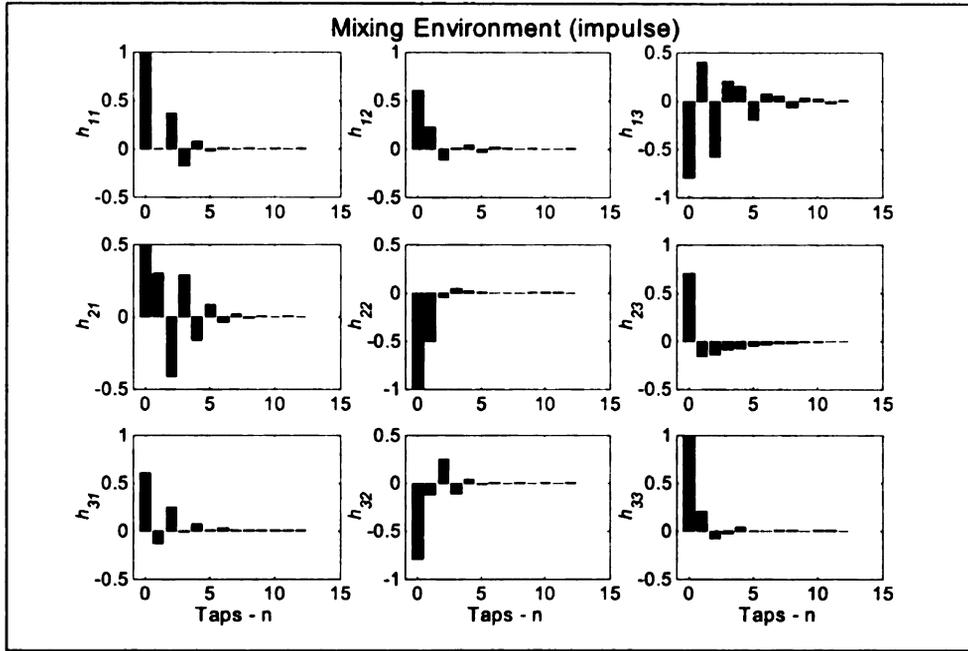


(a)

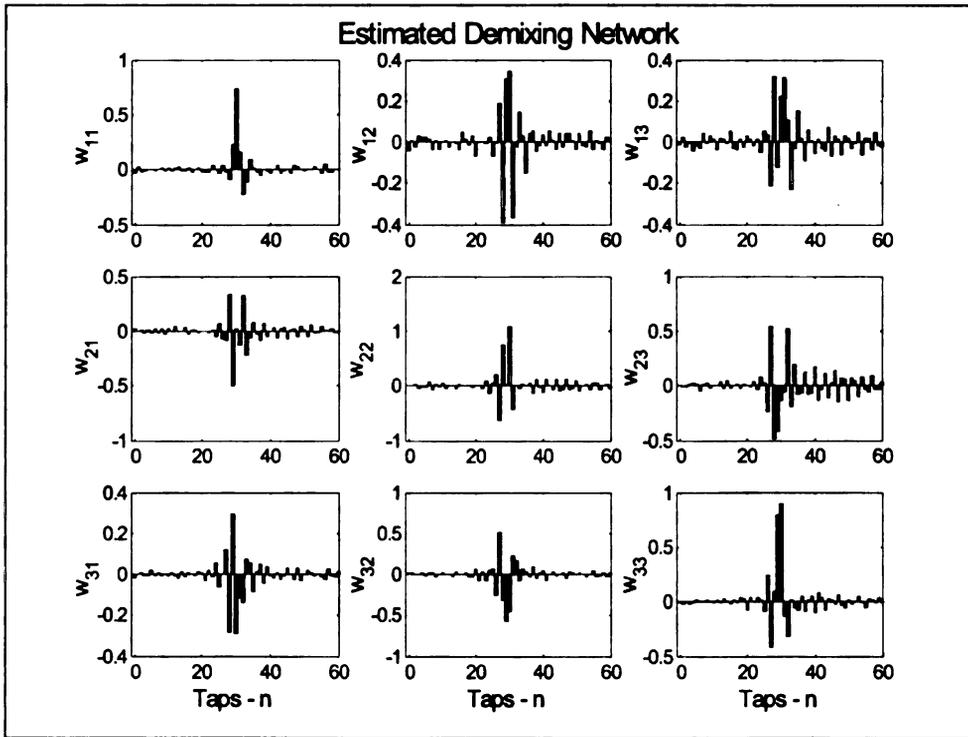


(b)

Figure 3.13. Non-minimum Phase Mixing Environment (a) Environment Transmission Zeros and Poles, (b) Impulse Response of the Environment Model



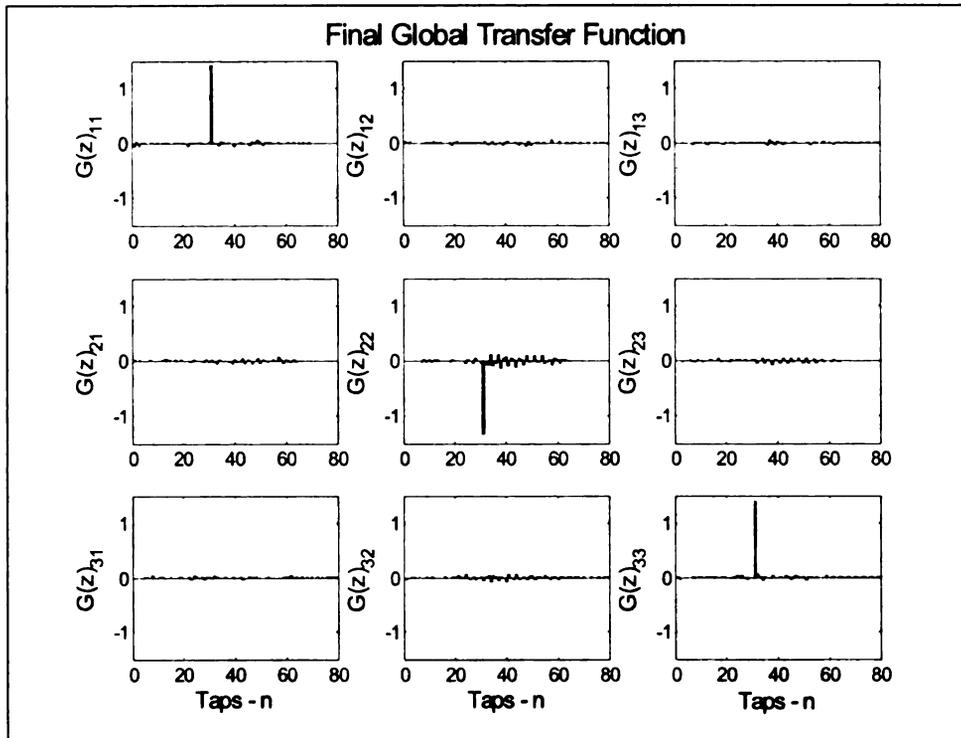
(a)



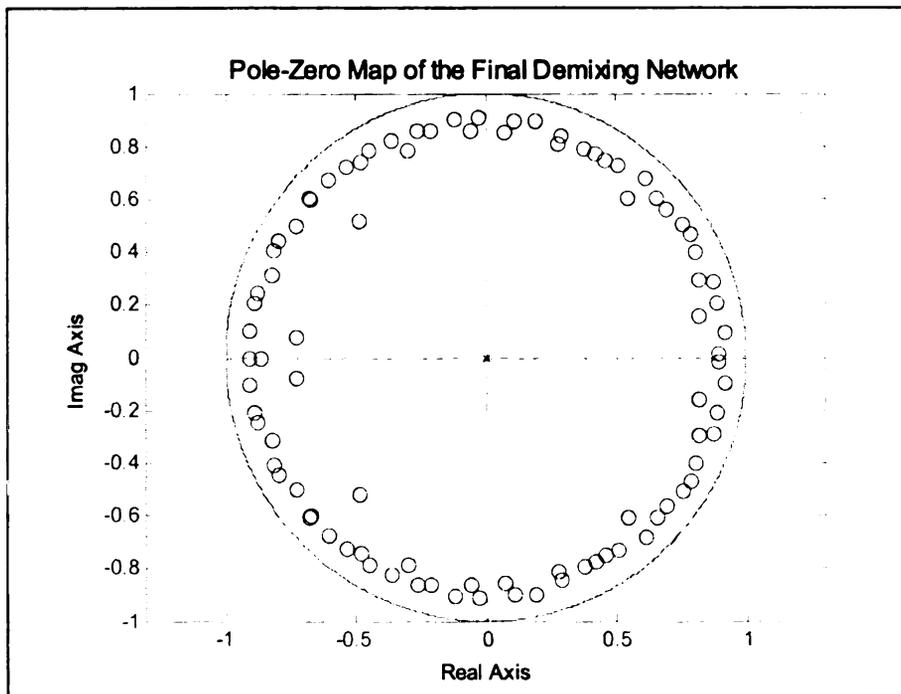
(b)

Figure 3.14. Non-minimum Phase Mixing Environment, Impulse Response of

(a) Theoretical environment Inverse, (b) Estimated Demixing Network.



(a)



(b)

Figure 3.15. Non-minimum Phase Mixing Environment, (a) Final Global Transfer Function, (b) Demixing Network Transmission Zeros and Poles.

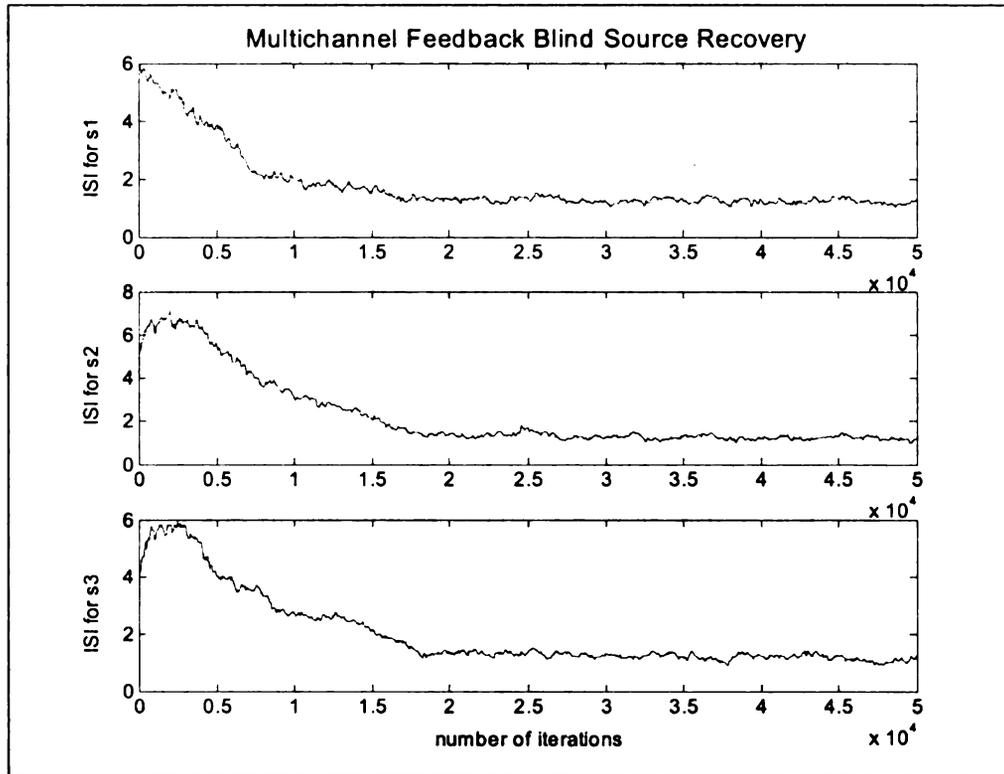


Figure 3.16. Non-minimum Phase Mixing Environment, Convergence of MISI Performance Index.

A quantitative performance comparison of the blindly recovered communication signals using the presented BSR algorithms is provided in Table 3.1.

		Minimum Phase Feedforward Demixing			Minimum Phase Feedback Demixing			Non-minimum Phase Demixing
		<i>Poles known</i>	<i>Poles fixed at zero</i>	<i>Random, stable fixed poles</i>	<i>Poles known</i>	<i>Poles fixed at zero</i>	<i>Random, stable fixed poles</i>	
ISI (in dB)	<i>AM Signal</i>	-22.55	-21.77	-26.02	-27.67	-16.47	-23.42	-17.67
	<i>PSK Signal</i>	-26.66	-18.48	-23.52	-23.52	-18.85	-24.85	-11.43
	<i>White Noise</i>	-21.09	-21.85	-21.57	-26.03	-18.36	-23.96	-11.39

Table 3.1. Quantitative Signal Recovery Comparison of the BSR algorithms

3.6 Remarks

This chapter presents a generalized Blind Source Recovery (BSR) framework based on the theory of optimization under constraints of a state space representation with the Kullback-Lieblar divergence as the performance functional. The framework is then specialized to the state space blind source recovery for both static and dynamic environments, which may possess either a minimum phase or a non-minimum phase structure. For the case of minimum phase environments, two possible state-space demixing networks in feedforward and feedback configuration are proposed. For non-minimum phase (convolutive) mixing, to avoid any instability, the demixing network is set as a non-causal all-zero feedforward state space network. The update laws for all the cases have been derived using the natural (or the Riemannian contra-variant) gradient.

The algorithms have been extensively simulated for both synthetic and physical data. The simulation results for all the cases have been provided using mixtures of multiple source densities, including one gaussian source. The algorithms exhibit success in recovery of sources in all cases. Furthermore, for the case of minimum phase environments, various possible initial conditions have been explored. A comparative table for a set of communication signals has also been provided as a guideline to the achievable separation/recovery results using the proposed online adaptive algorithms.

Chapter 4

BLIND SOURCE RECOVERY USING ADAPTIVE SCORE FUNCTIONS

The chapter discusses Blind Source Recovery (BSR) of minimum phase and non-minimum phase mixtures of multiple source distributions using adaptive score functions. For the linear static and the dynamic convolutive mixtures, the ability to adaptively determine discerning nonlinear score functions allows for recovery of original sources independent from (and even in the absence of) precise environment *identifiability*.

The performance of BSR algorithms strongly depends on the choice of an appropriate score function which appears as an element wise acting non-linearity on the output signals [1, 3, 14, 22, 24, 29, 30]. For a particular problem, the optimal score function depends on the distribution of the original source signals which are unknown in a blind scenario. In such cases, unless some assumptions about the distribution of the sources are made, any BSR algorithm will potentially be unable to deliver the desired performance. Therefore, adaptive estimation of appropriate score functions is very attractive from a practical implementation viewpoint. The proposed score functions have

been applied to BSR for both the cases of minimum phase and non-minimum phase mixing environment models [4, 10, 11, 16]. Several closed form entropy expansion based deterministic expressions for adaptive score functions have been earlier proposed, but they are typically suited for a limited range of distributions and converge to the true estimating score function for static mixtures only [2, 5, 6, 21].

Most of the mixtures encountered in practical BSR problems are from sources with a variety of non-gaussian distributions, e.g., speech and acoustics, communication signals, internet data, interstellar signals, bio-medical signals etc. On the other hand, most noise phenomena or unidentified sources are assumed to possess gaussian distributions. This results in practical situations, where one has to cope with multiple source distribution mixtures including gaussian distributions. This chapter describes the use of the proposed adaptive score functions for the linear convolutive class of BSR.

Two adaptive score functions are presented in this chapter. The first score function is derived from the generalized gaussian distribution model [7, 12, 26]. The second score function is based on proposed new hyperbolic source models, which extend current Pearson and Hyper-Cauchy distribution models [12]. The proposed score functions can be tuned online by certain control parameters. An adaptive algorithm to estimate these control parameters for the proposed tuning of score functions using the batch kurtosis of BSR output is also presented. The primary advantage of the proposed online parameter estimation is that it renders the adaptive estimation of the demixing network to be completely blind. No a priori information about the distribution structure of the original sources is required. Simulation examples verifying the proposed BSR

formulation (inclusive of the proposed score function estimation technique) are included as an illustration of the proposed framework.

4.1 Generalized Gaussian Distribution Model

The generalized gaussian distribution class was proposed by Miller et. al. [17] and used in the detection theory as a model for non-gaussian noise. This class has a symmetric *unimodal* density characteristic. Various classes of densities are generated by a parameterization of the exponential decay of this density function. The Generalized Gaussian Family has the following density relation

$$f(x_i, \alpha_i, \sigma_i) = \frac{\alpha_i \beta_i(\sigma_i, \alpha_i)}{2\Gamma\left(\frac{1}{\alpha_i}\right)} \exp\left\{-\left(\beta_i(\sigma_i, \alpha_i)|x_i|\right)^{\alpha_i}\right\} \quad (4.1)$$

where

$\alpha_i > 0$, is a positive parameter that controls the distribution's exponential rate of decay

σ_i^2 is the variance of the distribution

$\Gamma(\cdot)$ is the gamma function, given by

$$\Gamma(a) = \int_0^{\infty} \tau^{a-1} e^{-\tau} d\tau \quad (4.2)$$

$\beta_i(\sigma_i, \alpha_i)$ is a scaling function of σ_i and α_i , defined as

$$\beta_i(\sigma_i, \alpha_i) \triangleq \sigma_i^{-1} \left(\frac{\Gamma\left(\frac{3}{\alpha_i}\right)}{\Gamma\left(\frac{1}{\alpha_i}\right)} \right)^{1/2} \quad (4.3)$$

This generalized gaussian family can encompass a wide variety of distributions, e.g., for $\alpha_i = 1$, the model reduces to the laplacian probability density function

$$f(x_i, 1, \sigma_i) = \frac{\beta_i(\sigma_i, 1)}{2} \exp\{-(\beta_i(\sigma_i, 1)|x_i|)\} \quad (4.4)$$

For $\alpha_i = 2$, the end results is the gaussian probability density function

$$f(x_i, 2, \sigma_i) = \frac{\beta_i(\sigma_i, 1)}{\sqrt{\pi}} \exp\{-(\beta_i(\sigma_i, 2)|x_i|)^2\} \quad (4.5)$$

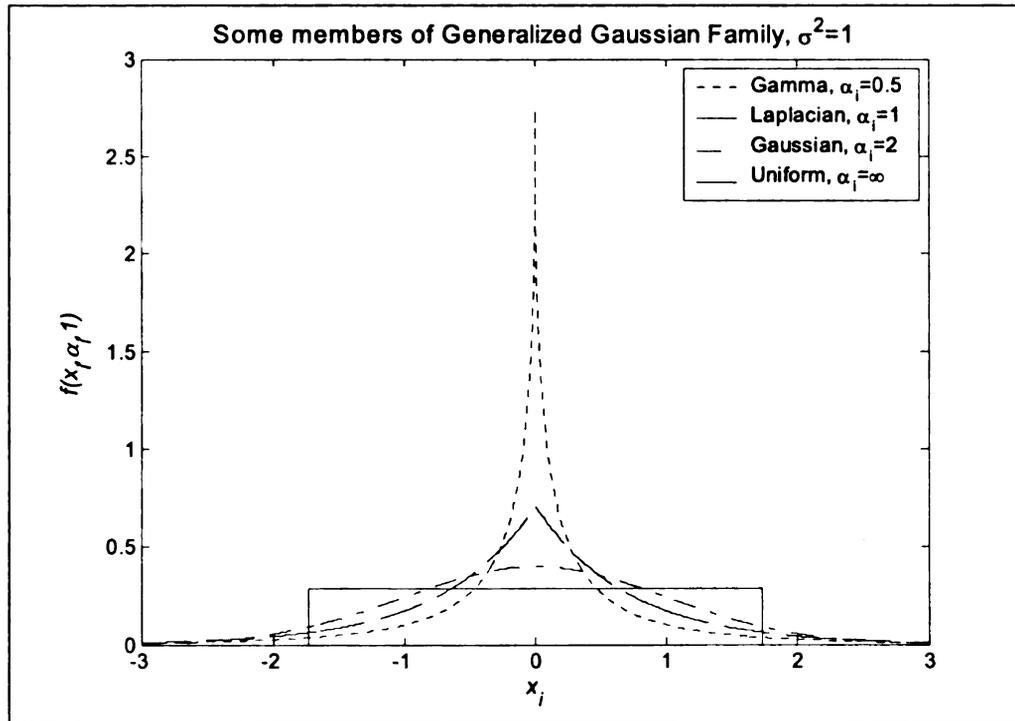


Figure 4.1. Some members of Generalized Gaussian Distribution Family

As α_i tends to infinity, the distribution approaches the uniform density function

$$f(x_i, \infty, \sigma_i) = \frac{\beta_i(\sigma_i, \infty)}{2}; \frac{-1}{\beta_i} \leq x_i \leq \frac{1}{\beta_i} \quad (4.6)$$

While as α_i tends to zero, the distribution converges to an impulse probability function

As a function of α_i and for $\sigma_i^2 = 1$, some members of the family are presented in

Figure 4.1.

Based on the generalized gaussian density model in (4.1), the k^{th} absolute moment can be characterized as

$$E\{|x_i|^k\} = \beta_i(\sigma_i, \alpha_i)^k \frac{\Gamma\left(\frac{k+1}{\alpha_i}\right)}{\Gamma\left(\frac{1}{\alpha_i}\right)} \quad (4.7)$$

4.1.1 Generalized Kurtosis Measures

The generalized kurtosis is a statistical measure to describe the “tail weight” for a distribution, it does so by estimating the deformation of the distribution from normality. Kurtotic measures have been extensively used in signal processing for classification of data distributions, and as a distance measure from gaussianity. In order to avoid any ambiguity, it is important to clarify that there are two types of kurtosis measures.

Pearson Kurtosis [19] is based on the mean statistics [12, 19] and in its generalized form it is defined as

$$K^P(\zeta, x_i) = \frac{E\{|x_i|^{2\zeta}\}}{E^2\{|x_i|^\zeta\}} \quad (4.8)$$

Fisher Kurtosis [12, 13] is based on the k-statistics [13] and is correspondingly defined to be the ratio of cumulants, i.e.,

$$K^F(\zeta, x_i) = \frac{\kappa_{2\zeta}(x_i)}{\kappa_\zeta^2(x_i)} \quad (4.9)$$

Typically the above-mentioned kurtosis measures are used for $\zeta = 2$, referred to as the standard kurtosis. For this standard case, the above two kurtosis measures are related by

$$K_4^F(x_i) = K_4^P(x_i) - 3 \quad (4.10)$$

where $K_4^i(x_i) = K^i(2, x_i)$ represents the computed kurtosis for the standard case of $\zeta = 2$

The primary advantage in using the Fisher Kurtosis is that a perfect gaussian source will have zero fisher kurtosis. A *leptokurtic* distribution (Fisher kurtosis > 0) represents the super-gaussian family of distributions, while a *platykurtic* (Fisher kurtosis < 0) distribution represents the family of sub-gaussian distributions [12, 13].

For the generalized gaussian family, the generalized Pearson kurtosis can be simplified to

$$K^P(\zeta, x_i) = \frac{\Gamma\left(\frac{2\zeta+1}{\alpha_i}\right)\Gamma\left(\frac{1}{\alpha_i}\right)}{\Gamma^2\left(\frac{\zeta+1}{\alpha_i}\right)} \quad (4.11)$$

Therefore, the standard Fisher kurtosis can be written as a function of the shaping parameter a_i only, i.e.,

$$K_4^F(a_i) = \frac{\Gamma\left(\frac{5}{a_i}\right)\Gamma\left(\frac{1}{a_i}\right)}{\Gamma^2\left(\frac{3}{a_i}\right)} - 3 \quad (4.12)$$

Note that for the parameter $a_i = 2$, the distribution reduces to the true gaussian with zero kurtosis. The complete kurtosis characteristics of the generalized gaussian family are given in Figure 4.2. As generalized gaussian family can only represent unimodal distributions. The Fisher kurtosis approaches approximately -1.2 as a_i approaches infinity (i.e., the Fisher Kurtosis for a flat or uniform distribution).

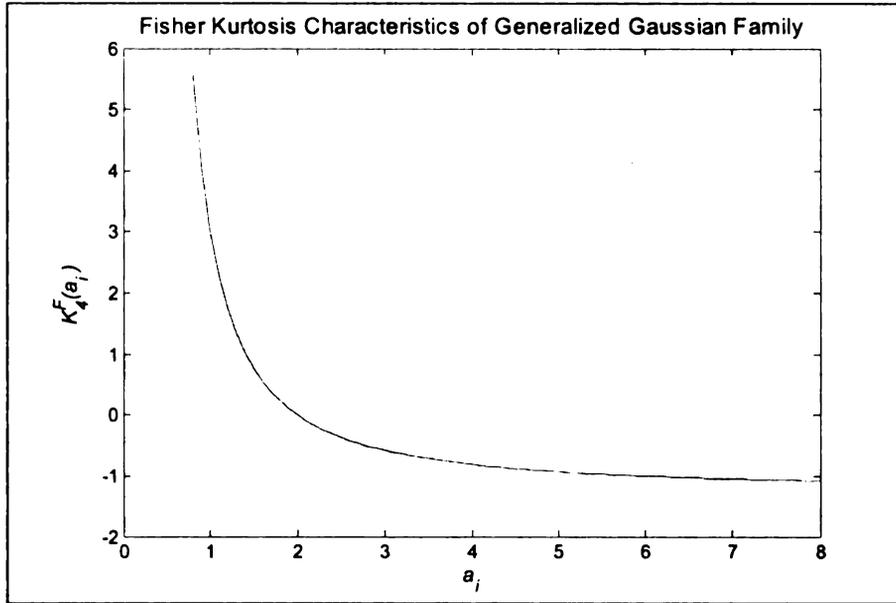


Figure 4.2. Fisher kurtosis Characteristics for the Generalized Gaussian Family

4.1.2 Relationship between $\beta_i(\sigma_i, \alpha_i)$ and α_i

Imposing the statistical constraint of distribution integral on the generalized Gaussian distribution family, i.e.,

$$\int_{-\infty}^{\infty} f(x_i, \alpha_i, \sigma_i) dx_i = 1 \quad (4.13)$$

where, the relation between $\beta_i(\sigma_i, \alpha_i)$ and α_i can be estimated to be

$$\beta_i(\sigma_i, \alpha_i)^{-\alpha_i} = \alpha_i E\{|x_i|^{\alpha_i}\} \quad (4.14)$$

This relationship is helpful in the determining the final form of the generalized gaussian score function.

4.1.3 Derivation of Generalized Gaussian Adaptive Score Function (GGASF)

Due to the very parameterized nature of the generalized gaussian family and its ability to model a variety of unimodal source densities of interest to blind source recovery, a parametric score function (nonlinearity) can be derived based on this family. This comprehensive score function inherits a nice parametric structure from its parameterized parent density. In an adaptive setting, the regulation parameter α_i can be adapted during the adaptation for the source recovery algorithm.

The element-wise nonlinear score function required for the BSR problem has the definition [1, 3, 22, 24, 29, 30]

$$\varphi(y) = -\frac{\partial \log p(y)}{\partial y} = [\varphi_1(y_1) \quad \varphi_2(y_2) \quad \cdots \quad \varphi_n(y_n)]^T \quad (4.15)$$

For the i^{th} output with an exponential density function, the equation can be simplified to

$$\varphi_i(y_i) = -\frac{\partial p_i(y_i) / \partial y_i}{p_i(y_i)} \quad (4.16)$$

where

y_i : represents the i^{th} output of the network

$p_i(y_i)$: represents the statistical probability density of the i^{th} output

Using the generalized gaussian family as the candidate density in relation (4.16) for the score function

$$\varphi_i(y_i) = - \frac{\frac{\partial}{\partial y_i} \left(\frac{\alpha_i \beta_i(\sigma_i, \alpha_i)}{2\Gamma(1/\alpha_i)} \exp\left\{ -(\beta_i(\sigma_i, \alpha_i)|x_i|)^{\alpha_i} \right\} \right)}{\frac{\alpha_i \beta_i(\sigma_i, \alpha_i)}{2\Gamma(1/\alpha_i)} \exp\left\{ -(\beta_i(\sigma_i, \alpha_i)|x_i|)^{\alpha_i} \right\}} \quad (4.17)$$

by simple calculus, the relation reduces to

$$\varphi_i(y_i) = \alpha_i \beta_i(\sigma_i, \alpha_i)^{\alpha_i} |y_i|^{\alpha_i-1} \frac{\partial}{\partial y_i} |y_i| \quad (4.18)$$

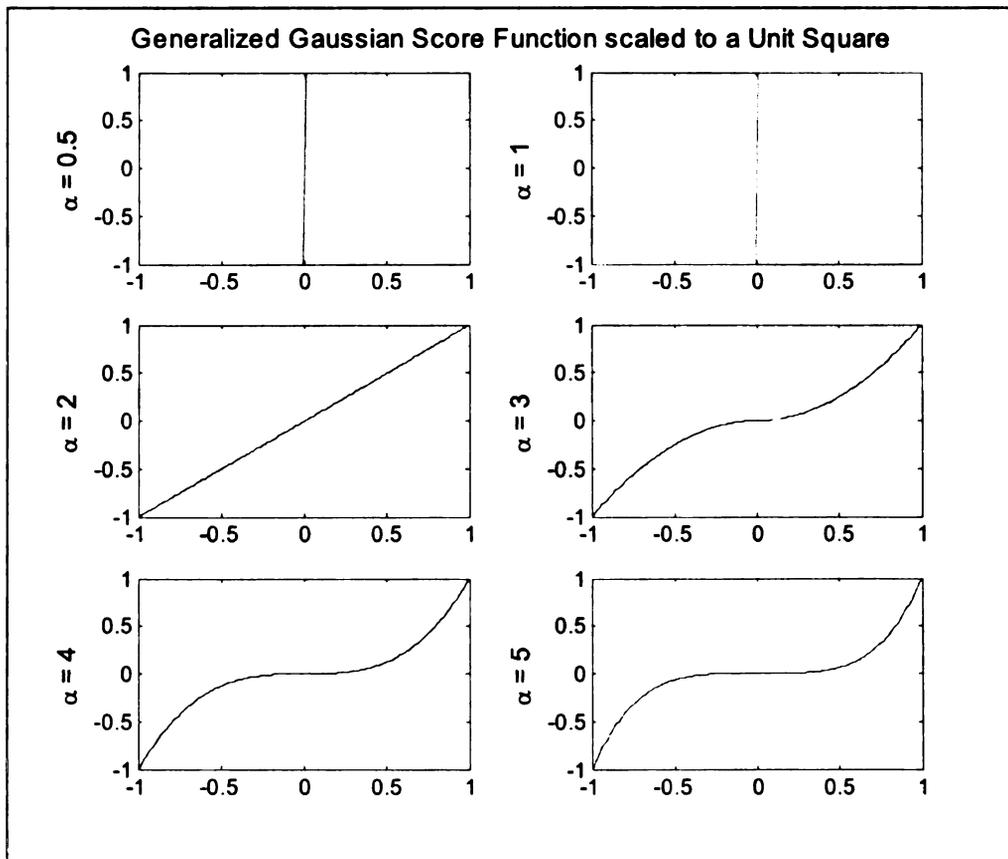


Figure 4.3. Scaled Generalized Gaussian Score Function for different values of parameter α

Using (4.14) and the definition $|y_i| = y_i \text{sign}(y_i)$, the score function is given by

$$\varphi_i(y_i) = \frac{|y_i|^{\alpha_i-1} \text{sign}(y_i)}{E\{|y_i|^{\alpha_i}\}} \quad (4.19)$$

where an appropriate choice of α_i makes the proposed score function (or nonlinearity) suitable to a particular source distribution. This characterization of the source distribution can be done adaptively using the Fisher kurtosis estimate for the batch-output of the BSR network. Figure 4.3 shows some members of the generalized gaussian score function family as a function of α_i .

4.2 Hyperbolic Source Density Models

In this section, first new probability density models are proposed and their relationship to existing well-known density models is established. Consequently, a score function for each of the proposed density models is derived. A combination of these two score functions results in the proposed Hyperbolic adaptive score function [30]. Interestingly the proposed score function is quite similar to our earlier proposed adaptive score function based on heuristic justifications, see Waheed, et al. [28].

4.2.1 Background

4.2.1.1 Pearson Distribution Model

A simple mixture of two gaussian densities gives the Pearson density model [18, 19], i.e.,

$$f(x_i, a_i) = (1 - a_i)f_1(x_i) + a_i f_2(x_i) \quad (4.20)$$

where

$f_i(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right)$ represents a gaussian pdf.

For the symmetric case of the Pearson model, i.e. $a_i = 1/2$, $|\mu_1| = |\mu_2| = |\mu|$ and $\sigma_1^2 = \sigma_2^2 = \sigma$; the Pearson model possesses two distinct modes for $\mu > 0$. However, for $\mu = 0$ the density reduces to the regular gaussian density. This model can effectively model gaussian and sub-gaussian densities including bimodal (binary) densities. For the symmetric Pearson density, the model can be expressed as

$$f(x_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{\mu_i^2}{2\sigma_i^2}\right) \exp\left(-\frac{x_i^2}{2\sigma_i^2}\right) \cosh\left(\frac{\mu_i x_i}{\sigma_i^2}\right) \quad (4.21)$$

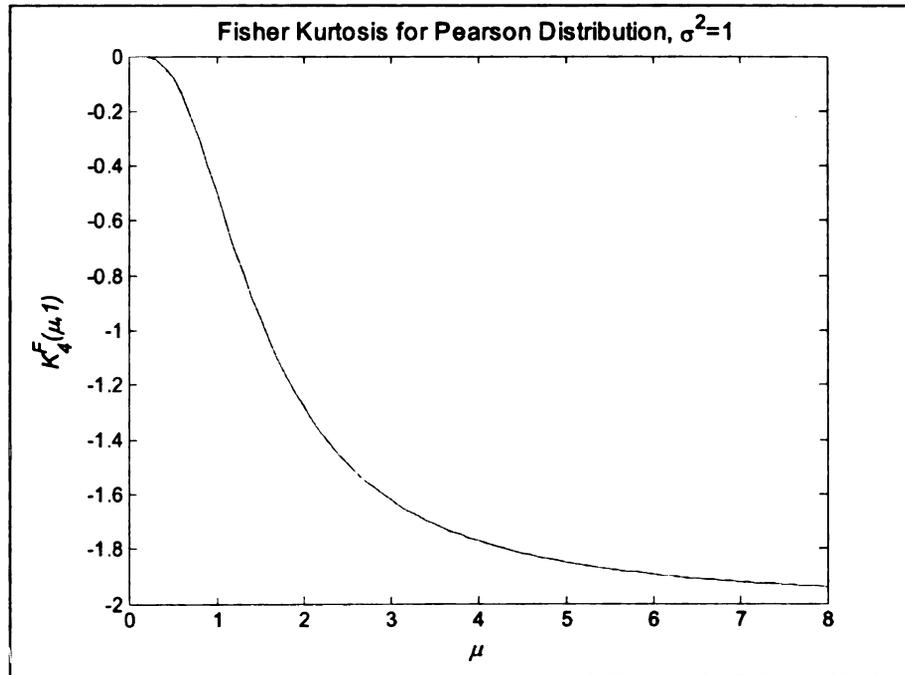


Figure 4.4. Fisher Kurtosis Characteristics for the symmetric Pearson Model as a function of μ

The Fisher kurtosis [12, 13] for the symmetric Pearson distribution model is given by

$$K_4^F(\mu, \sigma^2) = \frac{-2\mu^4}{(\mu^2 + \sigma^2)^2} \quad (4.22)$$

It is evident that for $\mu > 0$, the kurtosis of the distribution is strictly negative; see Figure 4.4, making it a suitable model for sub-gaussian to gaussian distributions only.

4.2.1.2 Generalized Cauchy Distribution

The symmetric super-gaussian density models are best represented by the Generalized Cauchy density. The generalized standard Cauchy density is given by [12, 18]

$$f(x_i) = \frac{k_i}{(1 + x^2)^{m_i}} \quad (4.23)$$

where

m_i : represents the order of the Cauchy model (i.e., determines the kurtosis of the distribution)

k_i : represents a scaling parameter to satisfy the constraints of a pdf, e.g., for $m_i = 1$,

$$k_i = 1/\pi.$$

The Cauchy distribution is a symmetric distribution with heavy tails and a single peak at the center of the distribution. Due to the heavier tail, the kurtosis is much larger than that for a normal distribution. In short, the kurtosis for a generalized Cauchy model ($m_i > 0$) is strictly positive and has an infinite range.

4.2.2 Proposed Hyperbolic Distribution Models

4.2.2.1 Hyperbolic Sub-Gaussian Model

An alternate model is proposed based on the simplified Pearson model in (4.21). Not only can this model represent all the kurtosis range of the original Pearson model; but also due to an increased degree of freedom in the hyperbolic component, it can model additional range of bimodal densities, which the original Pearson model cannot. The proposed model is given by

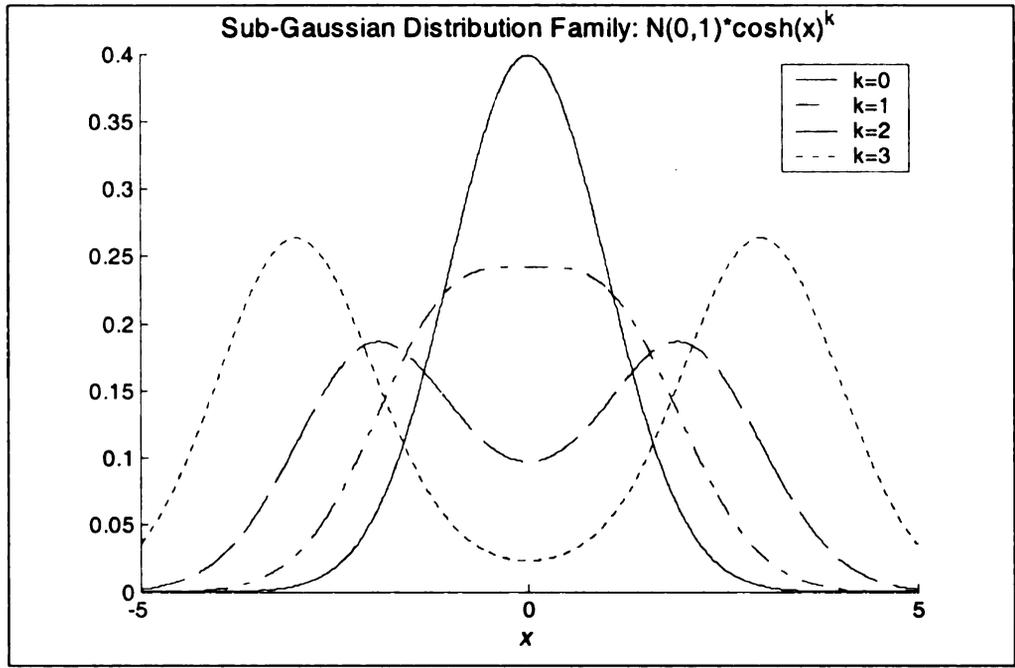
$$\hat{f}_{sub}(x_i) = N(\mu_i, \sigma_i^2) \cosh(\lambda_i(x_i - \mu_i))^k \quad (4.24)$$

where

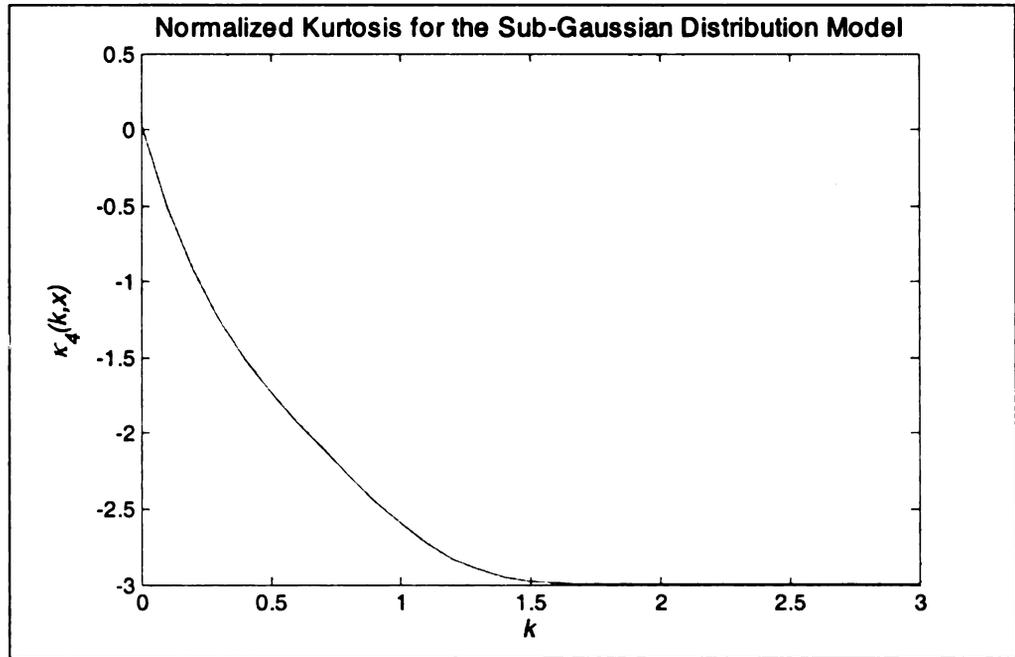
λ_i, k : represent positive shaping parameters

$N(\mu_i, \sigma_i^2)$: represents a gaussian pdf [32] with mean μ_i and variance σ_i^2 . μ_i is the mean for the symmetric distribution. Without loss of any generality, assume $\mu_i = 0$.

Just like the Pearson Model, this model is capable of representing density models from the gaussian down to the sub-gaussian bimodal densities. The family of density curves generated by this probability model for $\lambda_i = 1$ is shown in Figure 4.5.a. For $k = 0$, this model represents the gaussian density with a Fisher kurtosis equal to zero.



(a)



(b)

Figure 4.5. (a) Proposed Symmetric Sub-gaussian Density Model, (b) Fisher Kurtosis range of the proposed model; for $0 \leq k \leq 3$ and $\lambda_i = 1$.

The Fisher kurtosis [12, 18, 32] of this proposed model for the range $0 \leq k \leq 3$ and $\lambda_i = 1$ is presented in Figure 4.5.b. Note that unlike the original Pearson model, this new representation is capable of representing the bimodal distributions with a Fisher kurtosis < -2 . In fact, the model can model bimodal distributions with a Fisher kurtosis value of down to -3.

4.2.2.2 Hyperbolic Super-Gaussian Model

Similar to the above approach, the product of a gaussian and a hyperbolic function can be used to approximate the generalized Cauchy density model, i.e.,

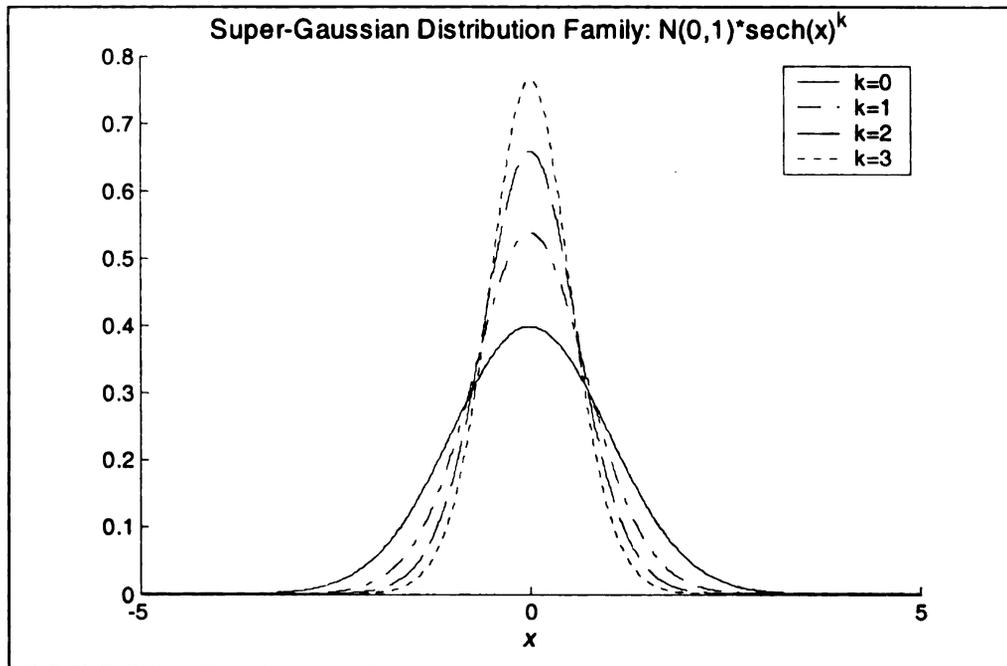
$$\hat{f}_{\text{sup}}(x_i) = N(0, \sigma_i^2) \text{sech}(\lambda_i x_i)^k \quad (4.25)$$

where

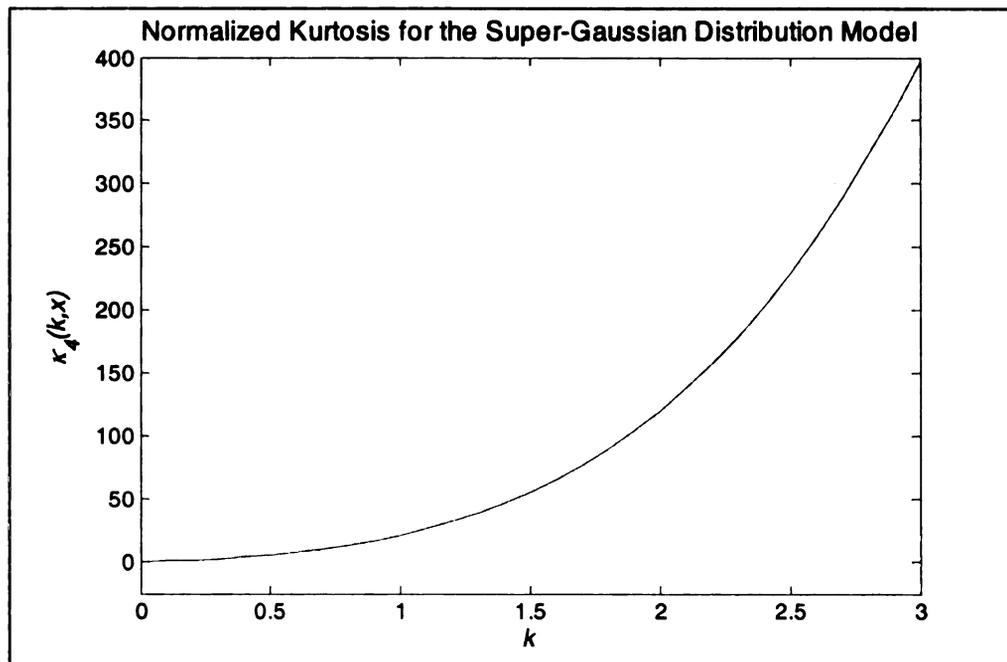
λ_i, k : represent positive shaping parameters

The proposed model can effectively model all possible generalized standard Cauchy densities. In addition, for $k = 0$, the proposed model reduces to the gaussian density. Figure 4.6.a presents some members of this family for $\lambda_i = 1$ and $0 \leq k \leq 3$.

As expected the normalized kurtosis for this family can range from zero to infinitely large positive values. For the range $0 \leq k \leq 3$ and $\lambda_i = 1$, the Fisher kurtosis for the family is shown in Figure 4.6.b.



(a)



(b)

Figure 4.6. (a) Proposed Symmetric Super-gaussian Density Model, (b) Fisher Kurtosis range of the proposed model; for $0 \leq k \leq 3$; $\lambda_i = 1$.

4.2.3 Derivation of Hyperbolic Adaptive Score Function (HASF)

The two proposed density models can represent a wide range of sources which are of particular interest to the source separation models. This representation covers the complete range from the binary distributed sources in communications to the very sparse distributions obtained during stellar and biological measurements. The score functions are initially derived for each individual proposed density model. Later both of these score functions are merged to formulate the proposed adaptive hyperbolic score function.

For the proposed sub-gaussian distribution model (4.24), the BSR score function can be computed as

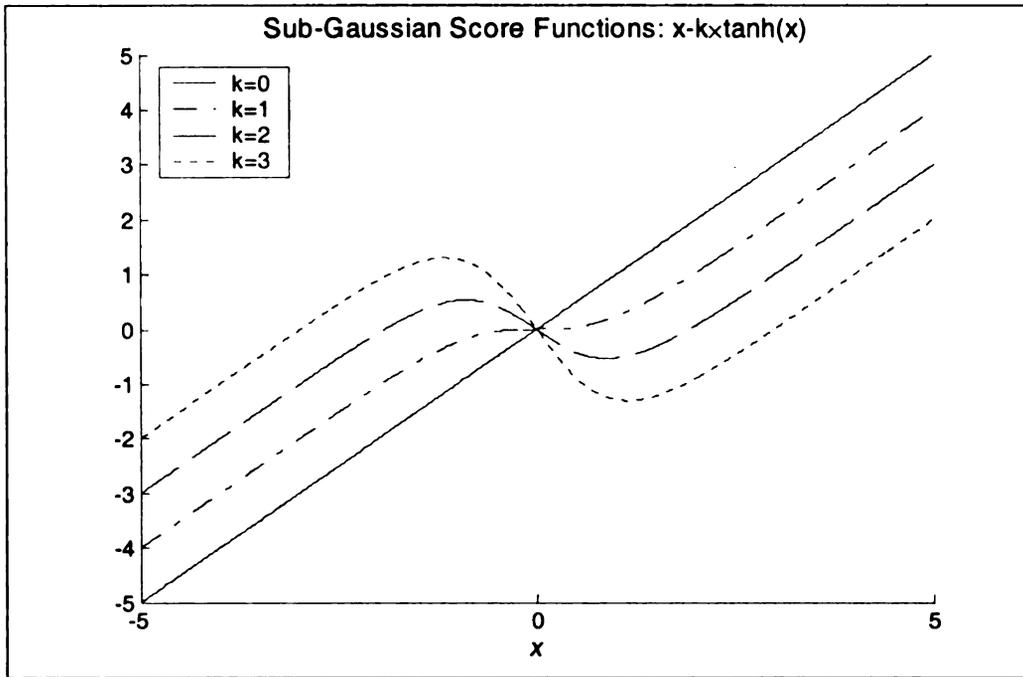
$$\varphi_i(y_i) = -\frac{\partial p_i(y_i)/\partial y_i}{p_i(y_i)} = \frac{y_i}{\sigma_i^2} - \lambda_i k_i \tanh(\lambda_i y_i) \quad (4.26)$$

The sub-gaussian family of score functions generated in the range $0 \leq k \leq 3$, $\sigma_i^2 = 1$ and $\lambda_i = 1$ is shown in Figure 4.7.a.

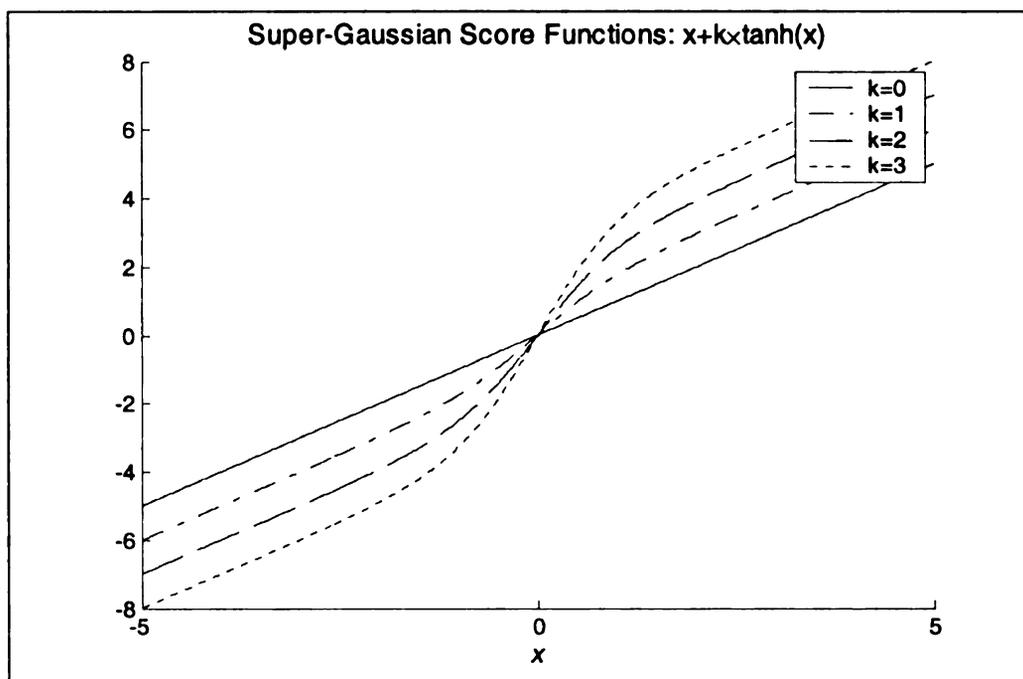
While for the proposed super-gaussian distribution model (4.25), the BSR score function is given by

$$\varphi_i(y_i) = -\frac{\partial p_i(y_i)/\partial y_i}{p_i(y_i)} = \frac{y_i}{\sigma_i^2} + \lambda_i k_i \tanh(\lambda_i y_i) \quad (4.27)$$

The super-gaussian family of score functions generated in the range $0 \leq k \leq 3$, $\sigma_i^2 = 1$ and $\lambda_i = 1$ is shown in Figure 4.7.b.



(a)



(b)

Figure 4.7. Score Functions based on the proposed: (a) hyperbolic sub-gaussian model,

(b) hyperbolic super-gaussian model; for $0 \leq k \leq 3$; $\lambda_i = 1$.

Comparing (4.26) and (4.27), it is evident that both score functions have a similar structure but differ by the sign of the hyperbolic tangent term only. The appropriate sign of this term depends on whether the BSR output distribution is sub-gaussian or super-gaussian. This hyperbolic term, however, should disappear for the gaussian constituents in a BSR problem. A good adaptive estimate of the Fisher (or normalized) kurtosis of the BSR network output can provide us with this discrimination. In an adaptive sense, the switch between these different possible score functions can be made based on the estimate of output kurtosis. Both fore-mentioned score functions can be thus combined into an adaptive hyperbolic score function given as:

$$\varphi_i(y_i) = \nu_i y_i + \text{sign}(K_4^F(y_i)) \alpha_i \tanh(\beta_i y_i) \quad (4.28)$$

where

ν_i : is a scaling parameter for the linear term. ν_i is inversely proportional to the variance of the reference gaussian (or the resulting non-gaussian) distribution. Hence, for the distributions with large variance, this term will approach zero.

α_i, β_i : are two scaling parameters of the hyperbolic term which control the magnitude and the slope of the score function near the origin. α_i is directly proportional to the shaping parameter k in the proposed distribution models. For the gaussian model, $k = 0$ and therefore $\alpha_i = 0$ and the hyperbolic term should vanish. Some possible score functions are presented in Figure 4.8.

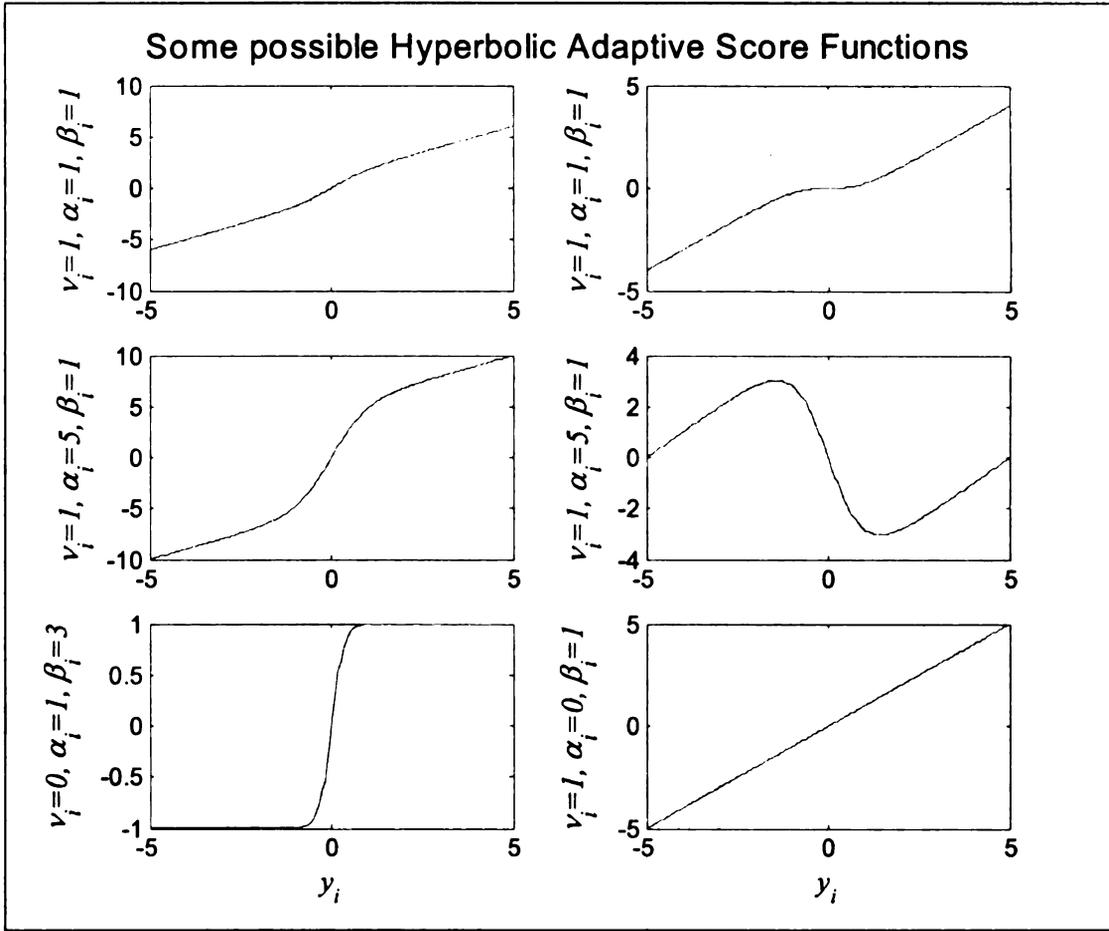


Figure 4.8. Proposed Adaptive Hyperbolic Score Function: some possible combinations.

4.3 Adaptive Estimation of the Fisher (or normalized)

Kurtosis

In order to perform an efficient adaptive estimate of the Fisher (or the normalized) kurtosis [12, 32] of the BSR output, a batch update algorithm is employed. Practically, one efficient simple formulation to adaptively estimate the output Fisher kurtosis is using a forgetting factor as follows:

$$K_i(n+1) = \gamma_i K_i(n) + (1 - \gamma_i) K_4^F(y_i(n)) \quad (4.29)$$

where

$K_i(n)$: represents the adaptively determined Fisher kurtosis for the i^{th} output at the n^{th} iteration

γ_i : represents the forgetting factor for adaptive kurtosis determination of the i^{th} output;

$$0 \leq \gamma_i \leq 1$$

$K_4^F(y_i(n))$: represents the normalized or Fisher kurtosis estimate for the n^{th} batch of the i^{th} output, which is computed as

$$K_4^F(y_i(n)) = \frac{\sum (\bar{y}_i - \bar{\mu}_i)^4}{N \bar{\sigma}_i^4} - 3 \quad (4.30)$$

where

\bar{y}_i : represents a batch of length N for the i^{th} BSR network output

$\bar{\mu}_i, \bar{\sigma}_i^2$: represent the mean and the variance of the batch of the i^{th} BSR network output, respectively.

4.4 Parameter Regulation for the Proposed Adaptive Score Functions

Using the estimated output Fischer kurtosis (4.29), discrimination can be done between the three possible cases of gaussian, sub-gaussian or super-gaussian distributions. This is done by appropriate selection of the score function shaping parameters for both the proposed adaptive score functions. This can be effectively done for the case of source recovery from linear convolutive mixtures, where only a discriminating score function is adequate to separate sources with different distributions [8, 15, 26, 28, 30, 31, 33]. On the other hand, for nonlinear BSR, a good match between the true source distribution and the

chosen score function is required for estimation of the correct separating network [7, 9, 24].

4.4.1 Choice of parameters for Generalized Gaussian Adaptive Score Function

In order to employ the proposed adaptive generalized gaussian score function for the linear convolutive BSR, one needs to primarily focus on only three cases of parameter α_i such that a classification between super-gaussian, sub-gaussian and gaussian distribution families can be made. Practically, a choice of α_i for the computation of the score function, see (4.19), is made using the adapted value of kurtosis; see (4.29), by the following simple scheme.

- if $\kappa_i(.) \geq \psi$, $1 \leq \alpha_i < 1.5$; where $\psi > 0$ is a constant threshold (for super-gaussian source distributions)
- elseif $\kappa_i(.) \leq -\xi$, $\alpha_i \geq 3$; where $\xi > 0$ a constant threshold (for sub-gaussian source distributions)
- else $\alpha_i = 2$ (for gaussian distribution)

Typical values chosen for α_i in the super-gaussian case is 1, and for the sub-gaussian case is 3.

4.4.2 Choice of parameters for Hyperbolic Adaptive Score Function

In order to employ the proposed hyperbolic adaptive score function for the linear convolutive BSR, the estimated normalized batch kurtosis of the BSR network output is used to discriminate between the three possible cases of gaussian, sub-gaussian or super-

gaussian distributions. This is done by appropriate selection of the score function shaping parameters ν_i , α_i and β_i . Outlined below is an algorithm to choose the salient feature of the proposed score function

- if $\kappa_i(\cdot) \geq \psi$, $0 \leq \nu_i \leq 1$; $\alpha_i \geq 1$ and $\beta_i \geq 1$ where $\psi > 0$ is a constant threshold (for super-gaussian source distributions)
- elseif $\kappa_i(\cdot) \leq -\xi$, $\nu_i \geq 1$; $\alpha_i \geq 1$ and $\beta_i \geq 1$ where $\xi > 0$ a constant threshold (for sub-gaussian source distributions)
- else $\alpha_i = 0$ (for gaussian distribution)

Note that for the case of sub-gaussian and super-gaussian distributions, α_i and β_i can be fine tuned to match the dynamic range and also further improve the signal to noise ratio (SNR) and intersymbol interference (ISI) of the recovered BSR sources.

4.5 Simulation Results

Simulation results using a 3×3 IIR mixing environment are being presented. The demixing system is formulated as a state-space network [20, 22, 29, 33]. For both simulations involving feedforward and feedback networks respectively, the structure for matrices A and B is kept fixed [23], whereas elements of matrices C and D are adapted. For an in depth discussion on the BSR algorithms see [22, 23, 25, 27-29]. In all presented simulations, the source signals are chosen to possess different distributions, i.e., super-gaussian, gaussian and sub-gaussian respectively; the score function is adapted online along with the BSR algorithm.

The convergence performance of the algorithm is measured using the multi-channel intersymbol interference (MISI) benchmark. MISI is a measure of the global

transfer function diagonalization and permutation as achieved by the demixing network, see Appendix D for details.

4.5.1 Environment Model

This environment model is assumed to be a 3×3 IIR filter

$$\sum_{j=0}^{m-1} A_j m(k-i) = \sum_{i=0}^{n-1} B_i s(k-i) + v(k) \quad (4.31)$$

where

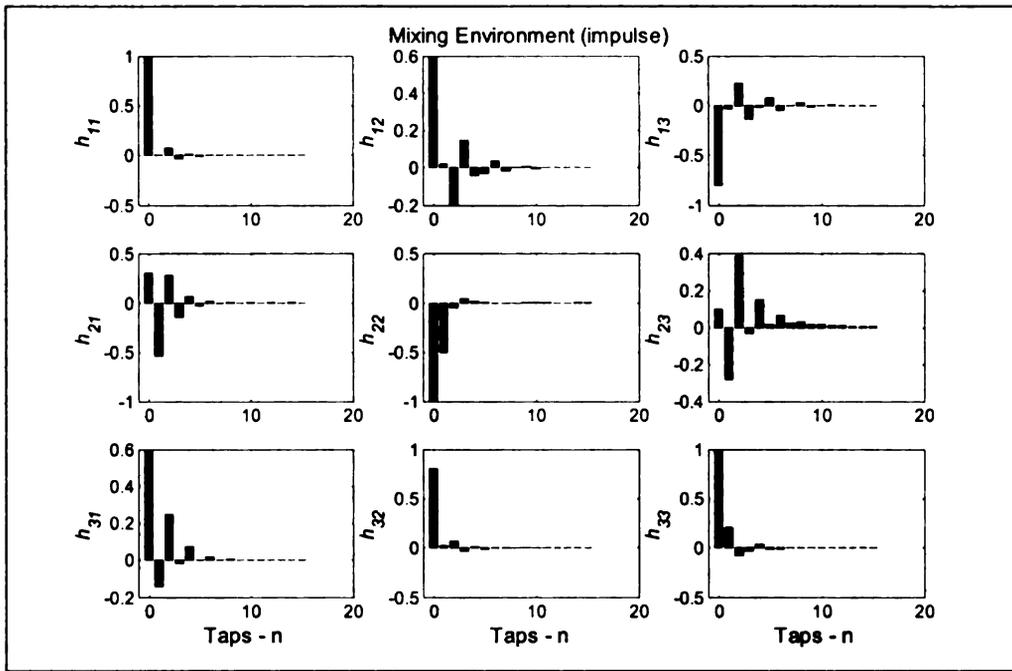
$$A_0 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, A_1 = \begin{bmatrix} 0.5 & 0.8 & -0.7 \\ 0.8 & 0.3 & -0.2 \\ -0.1 & -0.5 & 0.4 \end{bmatrix}, A_2 = \begin{bmatrix} 0.06 & 0.4 & -0.5 \\ 0.16 & -0.1 & -0.4 \\ -0.3 & -0.06 & 0.3 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.3 & 1 & 0.1 \\ 0.6 & -0.8 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} 0.5 & 0.5 & 0.6 \\ -0.3 & 0.2 & -0.3 \\ -0.2 & -0.43 & 0.6 \end{bmatrix}, B_2 = \begin{bmatrix} .125 & 0.06 & 0.2 \\ -0.1 & 0 & 0.4 \\ 0.08 & -0.13 & 0.3 \end{bmatrix}$$

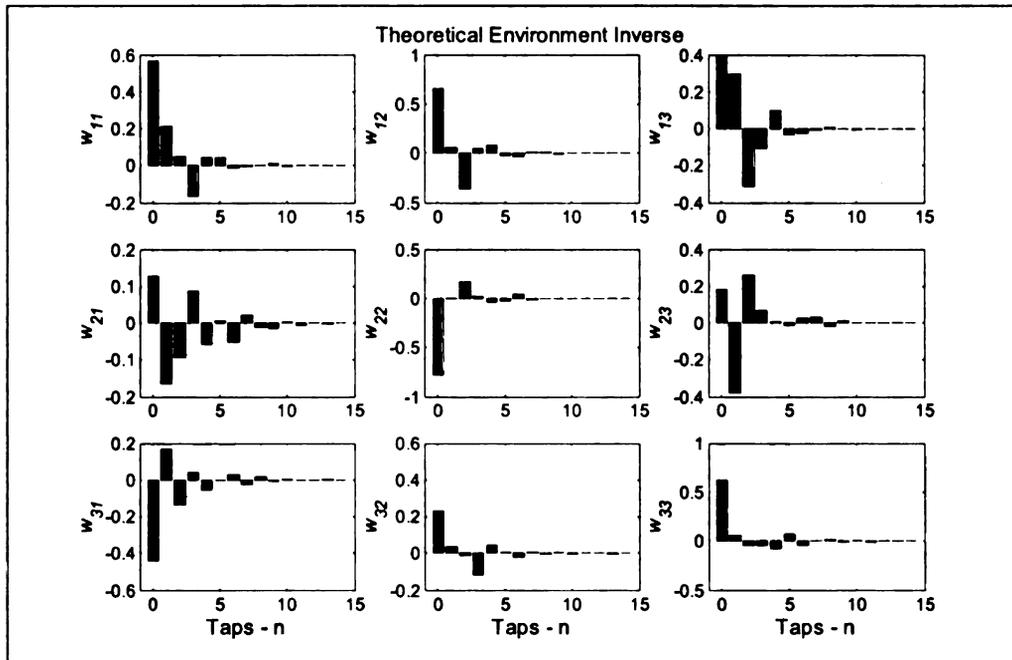
and $v(k)$ - represents additive gaussian noise

4.5.2 Demixing Network Structure

The theoretical inverse of this IIR mixing environment will be a 3×3 IIR matrix filter with each individual element being an IIR (or ARMA) filter of order 18. For both the feedforward and the feedback demixing cases, matrix C can be initialized with very small random numbers or all zero elements, while the matrix D is chosen to be an identity matrix. The number of taps per filter was chosen to be 21. Note for these equivariant minimum phase setups, any over-determination in the number of taps does not affect the performance of the algorithm.

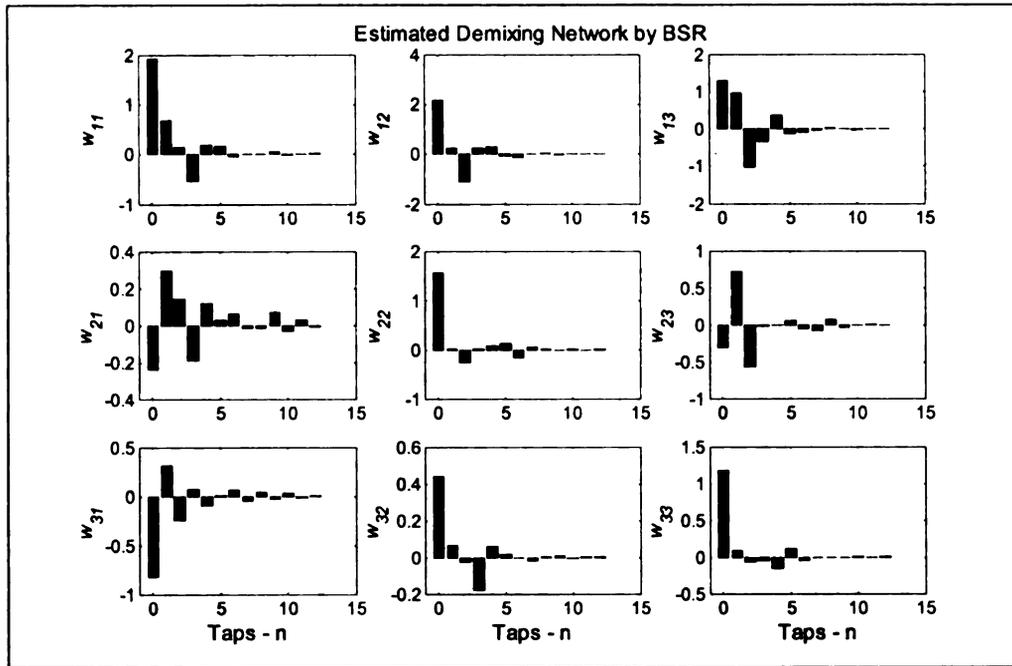


(a)

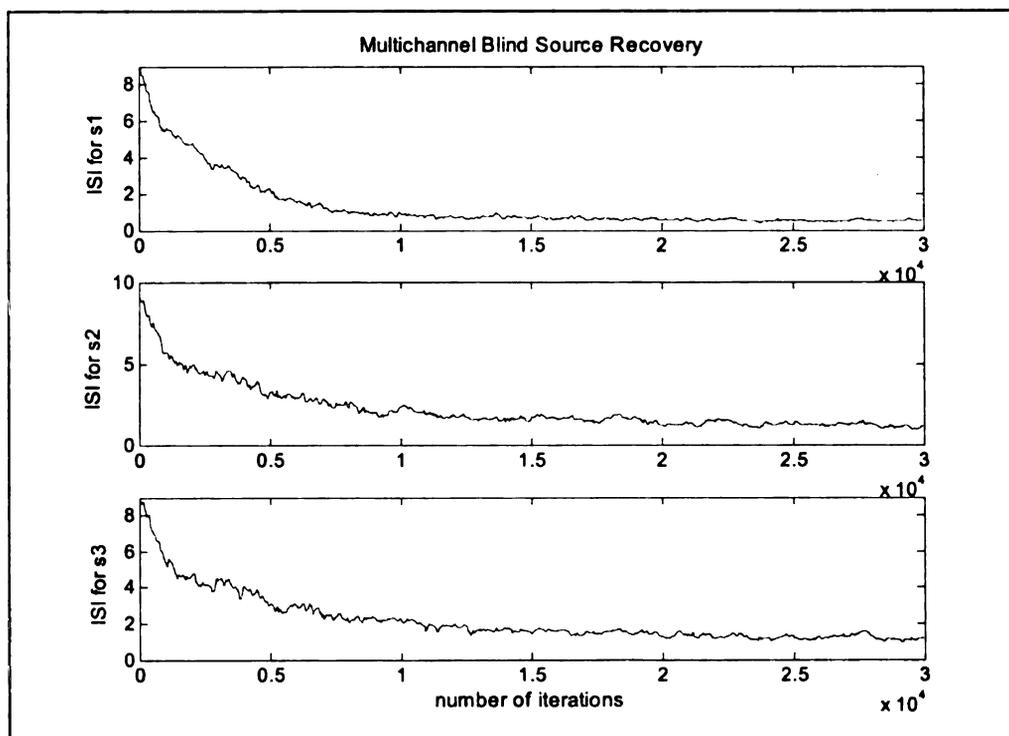


(b)

Figure 4.9. (a) Mixing Environment Transfer Function, (b) Theoretical Environment Inverse

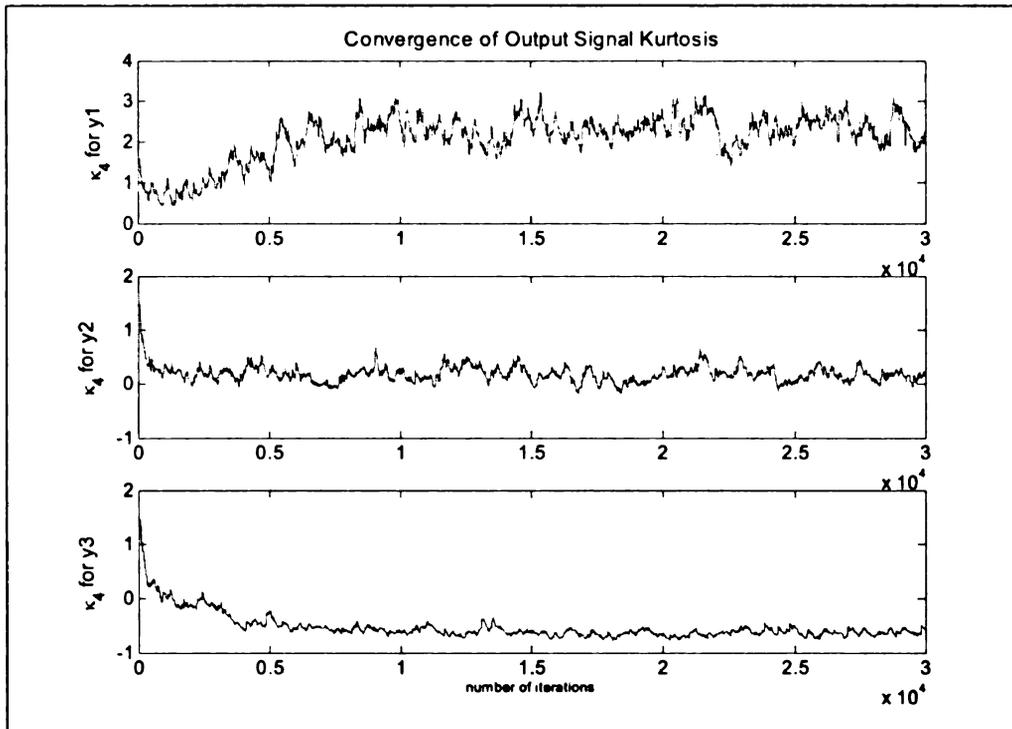


(a)

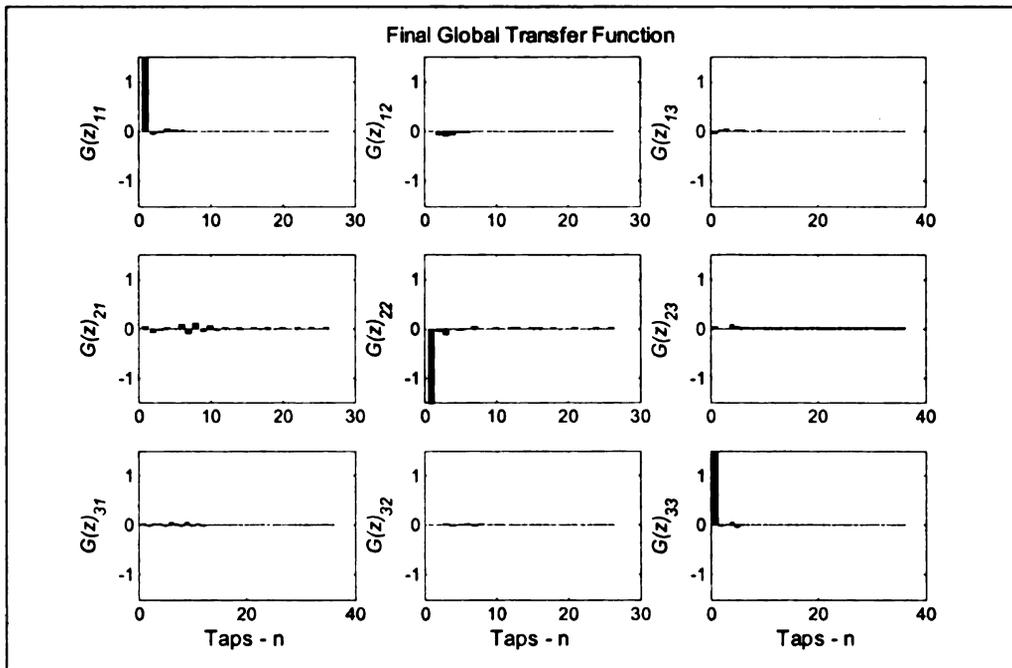


(b)

Figure 4.10. Feedforward BSR using Adaptive Score Functions: (a) Estimated Demixing Network Transfer Function, (b) Convergence of MISI Index



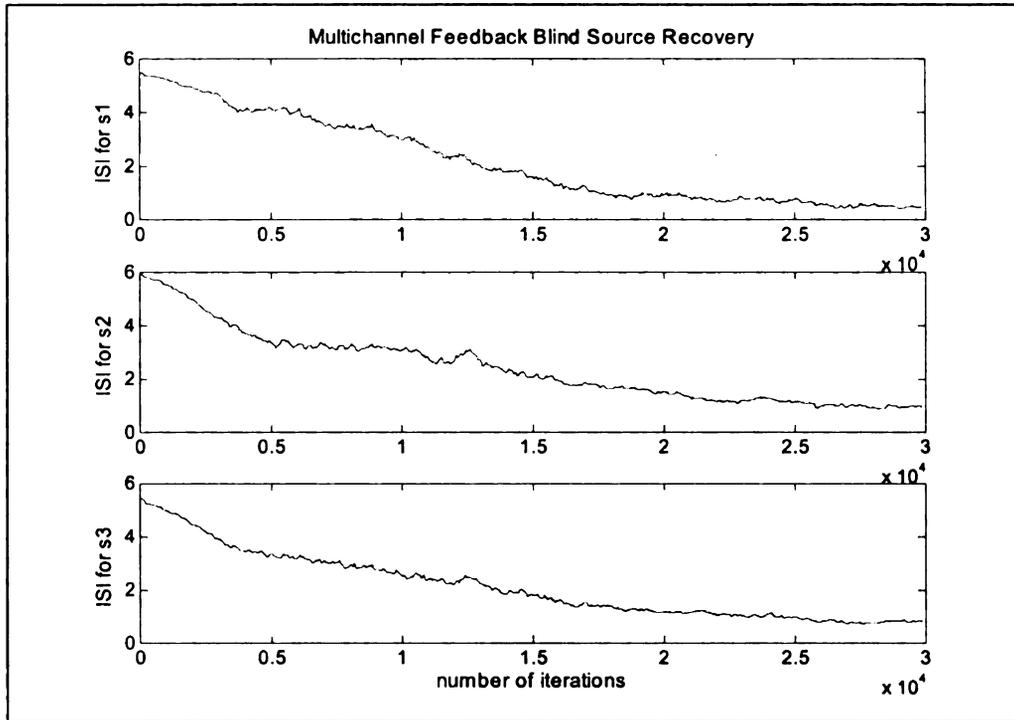
(a)



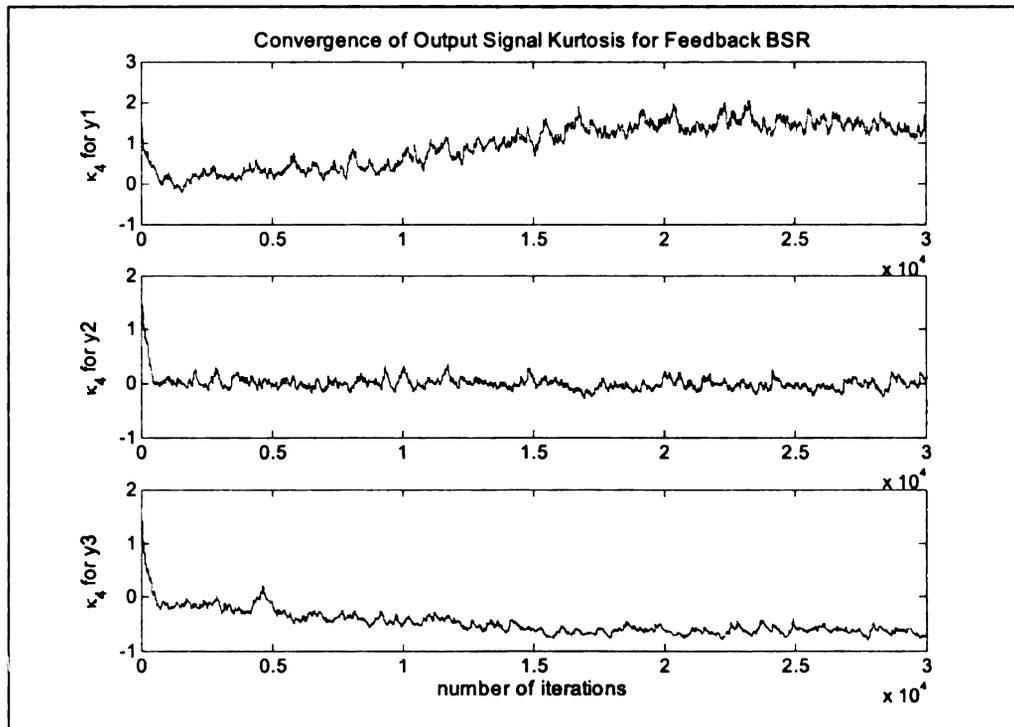
(b)

Figure 4.11. Feedforward BSR using Adaptive Score Functions: (b) Convergence of

Batch Kurtosis, (c) Final Global Transfer Function



(a)



(b)

Figure 4.12. Feedback BSR using Adaptive Score Functions: (a) Convergence of MISI Index, (b) Convergence of Batch Kurtosis

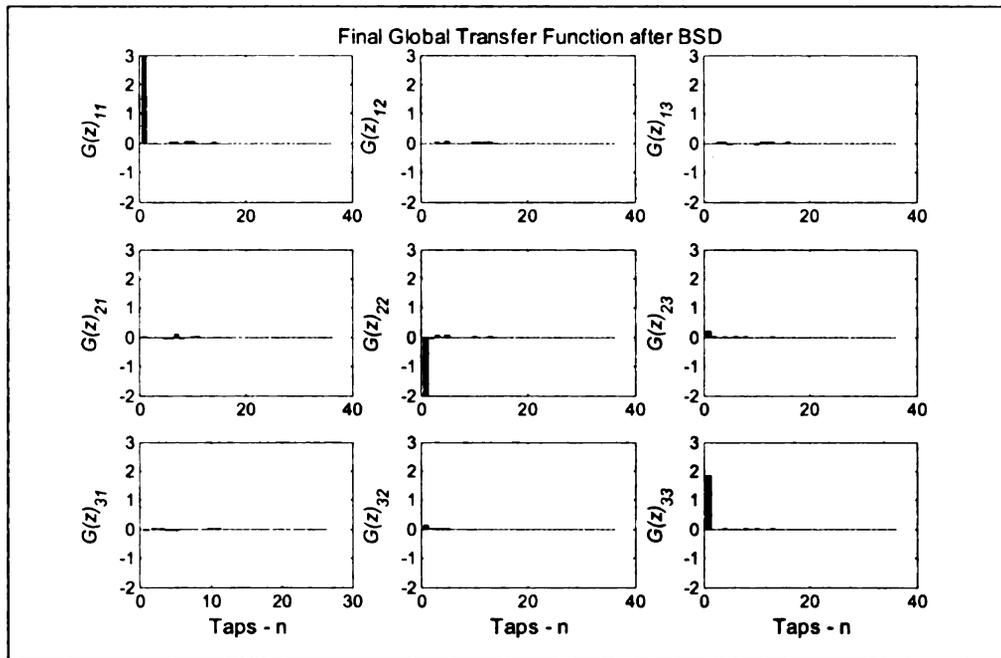


Figure 4.13. Feedback BSR using Adaptive Score Functions: Final Global Transfer Function

4.6 Remarks

In this chapter, two separate adaptive score functions for blind source recovery algorithms have been proposed. One of the score functions is based on the generalized gaussian density model, while the other is derived using proposed hyperbolic extensions to the Pearson and hyper-Cauchy models. Further we presented a method for online adaptation of the parameters for these score functions. This online adaptation scheme has been extensively tested for both minimum phase and non-minimum phase, instantaneous as well as convolutive mixing. For linear convolutive mixing, the results using both the proposed score functions are comparable. For simulations using bimodal distributions only, the latter score function outperforms the former that is best suited for unimodal

distributions. For nonlinear BSR, however, the slightly different kurtosis characteristics of the two proposed distribution models may be better exploited to estimate the correct demixing network. A couple of simulation examples using two different demixing architectures have also been presented. It is observed that the proposed adaptive framework can recover sources from a complex convolutive mixture efficiently even when all the sources have different probability density structure.

Chapter 5

BLIND SOURCE RECOVERY OF UNDERCOMPLETE AND OVERCOMPLETE MIXTURES

Blind Source Recovery is a framework for the estimation of unknown original source signals. Most of the existing BSR literature as well as most theoretical developments in this dissertation have been done under the assumption that the number of observed mixtures is equal to the number of original sources. In essence, this implies that the dimension of original sources to be estimated is known a priori and thus sufficient dimension in mixtures is recorded. In such a quadratic dimensional scenario, the recoverability of the sources is guaranteed provided the environment is *invertible* [13, 14]. In case the environment is linear in nature, this theoretical demixing network is unique and BSR adaptation results approach a permuted and scaled version of this

solution. However, in practical settings, the number of mixtures may or may not exactly equal the number of original sources. In case the dimension of observed mixture space exceeds the dimension of the original sources, the BSR problem is termed as *Undercomplete* BSR. Contrarily, if the dimension of the mixture space is smaller than the signal space, the recovery problem is called *Overcomplete* BSR [1, 4, 8, 27].

This variation in dimension may arise due to a number of reasons. Firstly, in a blind setting, the number of original stochastically independent sources may not be exactly known, e.g., consider stellar measurements, cocktail party problem, financial markets etc. Secondly, the dimension of the mixture or observation space may be determined by the structure of the problem [6, 18, 19], e.g., communication networks, number of installed recording devices etc. Thirdly, the additive noise in an environment may be dynamic and originate from a spurious number of sources causing the number of actual sources to vary over time, e.g., number of users in a wireless communication system, channel additive sources and multi-path effects in multiple access communication systems or in fact any measurements in dynamic noisy environments [5, 8, 11]. Due to these practical issues and difficulties in BSR problems, it is imperative to consider the cases of undercomplete and overcomplete BSR.

In this chapter, assuming the mixing to be linear in nature, a methodology is presented to use the BSR algorithms presented earlier in this work for the case of undercomplete mixtures. The case of overcomplete BSR is intractable in general. However, only possible solutions exist for sparse sources and are discussed in subsequent sections [1, 4, 10, 16, 24, 25, 27].

5.1 BSR of Undercomplete Mixtures

Undercomplete BSR is a more commonly encountered problem especially in the realm of communication systems, radar and sonar measurements or any other noisy BSR formulations, where adequate sensors are employed to collect the mixed signals and/or corrupted sources. For the case of undercomplete mixtures, the original sources occupy only a subspace of the mixture dimensions. In noisy undercomplete BSR, the update laws presented in earlier chapters can be used as is on the mixture dimensions to recover an equal number of sources. However, in such a recovered set of estimated signals, some of the sources may be recovered with multiple permutations and deflationary approaches may be needed to efficiently estimate the actual estimated sources [1, 5].

Alternately undercomplete BSR problem can be converted to the standard BSR problem by efficiently discarding some of the mixture data. This can be achieved by using signal pre-processors [4, 5, 7], such as PCA, projection pursuit etc. Efficient data reduction is useful in several ways; firstly, it reduces the dimension of the BSR problems and therefore results in a reduction of computational requirements. Secondly, data reduction techniques such as PCA based pre-whitening may also result in reduction of additive noise in the mixture space simplifying the BSR problem at hand. However, the performance of the data reduction stage is critical, as any data discarded at this stage is not used by the BSR stage and thus in-efficient data reduction may limit the performance of the overall recovery setup [5, 8]. See chapter 6, for the performance of BSR algorithms with a data-reducing pre-processor.

5.1.1 Efficient reduction of Mixture Dimensions

In undercomplete mixtures, efficient estimation of the desired signal subspace results in both reduction of overall computational load as well as it may serve as a pre-processor to reduce the additive measurement noise in the mixture space. Two best candidates for such reduction of data dimensions are basis pursuit and pre-whitening using principal component analysis (PCA). Under good implementation both techniques lead to similar performance [4, 5, 7], therefore only PCA is discussed in this work. PCA may be implemented either using algebraic or adaptive techniques.

5.1.1.1 Algebraic Principal Component Analysis

In algebraic PCA, an auto-correlation matrix of the received mixture data is computed or estimated online as [6].

$$R(m) = \alpha r(m)r^H(m) + (1-\alpha)R(m-1) \quad (5.1)$$

An efficient data reducing pre-whitening matrix in this case is then constructed as follows:

$$W = D_s^{-1/2} V_s^H \quad (5.2)$$

where \hat{D}_s and \hat{V}_s are the estimates of the n dominant eigen-values and eigen-vectors of the auto-correlation matrix R . These eigen values and vectors can also be computed adaptively, see [8, 9].

5.1.1.2 Adaptive Principal Component Analysis

Instead of trying to estimate the auto-correlation of the mixture data or its eigen decomposition, the pre-whitening matrix may be directly adapted from the mixture data [6, 7] by the following simple update law

$$W(k+1) = W(k) + \eta(k) \left[I - x(k)x^H(k) \right] W(k) \quad (5.3)$$

or its averaged version

$$\Delta W(k) = \eta(k) \left[I - \langle x(k)x^H(k) \rangle \right] W(k) \quad (5.4)$$

where $x(k) = W(k)m(k)$ and $\langle \cdot \rangle$: represents statistical averaging.

In case, $m(k)$ is noisy such that $m(k) = \hat{m}(k) + v(k)$, and $\hat{m}(k)$ is the corresponding noiseless mixture estimate. Let $\hat{x}(k) = W(k)\hat{m}(k)$ be the corresponding noiseless whitened estimate. The additive noise $v(k)$ within $m(k)$ introduces a bias in the estimated pre-whitening matrix W . Computing the covariance matrix for the whitened data as

$$R_{xx} = \langle x(k)x^H(k) \rangle = WR_{\hat{m}\hat{m}}W^H + WR_{vv}W^H \quad (5.5)$$

where $R_{\hat{m}\hat{m}} = \langle \hat{m}(k)\hat{m}^H(k) \rangle$ and $R_{vv} = \langle v(k)v^H(k) \rangle$.

Assuming the covariance matrix of the noise to be either known or estimated online [5-8], an unbiased PCA estimate can be determined by

$$W(k+1) = W(k) + \eta(k) \left[I - x(k)x^H(k) + \sigma_v^2 W(k)W^H(k) \right] W(k) \quad (5.6)$$

where, it is assumed that $R_{vv} = \sigma_v^2 I$. For more details, see [5].

5.2 BSR of Overcomplete Mixtures

The overcomplete BSR poses an intractable problem with no straightforward theoretical solution. In fact, even with the precise knowledge of the mixing structure for a simple static case, source estimation is not possible using any simple linear algebraic manipulation. A possible solution for the case of sparsely distributed (super-gaussian) statically mixed sources is inspired from *basis pursuit* [4], where the overcomplete mixtures is decomposed using a combination of adaptation for the mixing matrix and source inference using an interior point linear programming minimization of an L_1 norm based performance functional. Therefore, the discussion on overcomplete BSR will be restricted to the static case of ICA only.

Several possible solutions to the separation of independent components from their overcomplete mixtures have been proposed for super-gaussian sources [1, 2, 4, 10, 27, 28], but either the algorithms are computationally very demanding or the achieved separation quality is not very good. Recently published results have proposed efficient two stage approach to the recovery of sparse overcomplete mixtures [3, 16, 24]. This approach results in remarkably reduced computational requirements as well relevantly improved quality of received sources.

5.2.1 Two Stage Approach to Overcomplete ICA

In two stage overcomplete recovery approach, the overall computational burden of the adaptation is reduced by decoupling the two sub-problems of mixing matrix estimation and original source recovery. Data space projection methods [15, 22] are employed to first blindly estimate the underlying mixing matrix, this is called Blind Mixing Matrix

Recovery (BMMR). In the second Overcomplete Source Estimation (OSE) stage, source data is estimated directly from the mixture data minimizing an L_1 norm objective *distance* functional for interior point linear programming [4, 10]. As the two steps are carried out independently of each other, rather than recursively [1, 10, 27], the computational burden is greatly reduced.

Another technique for OSE using the shortest path approximation is explained in [3, 17]. This technique requires less computation than the aforementioned technique. However, in our experience the quality of sources recovered using the IP-LP approach is better than the latter approach, especially when the mixture space has a larger dimension.

In our work, the overcomplete ICA is solved in two stages, our proposed Algebraic Independent Component Analysis (AICA) Algorithm is used to estimate the rectangular mixing matrix, followed by optimal source inferencing using L_1 norm based interior point linear programming technique.

The proposed algebraic ICA algorithm is meant for the blind estimation of the mixing matrix and provides promising results for all cases of ICA. Under the assumption of sparse sources, AICA provides a blind estimate of the mixing matrix, termed as Blind Mixing Matrix Recovery (BMMR) [16]. BMMR directly constitutes a solution for quadratic ICA (and undercomplete ICA after discarding the appropriate undesired subspace), while for the case of overcomplete ICA, it is followed by a stage for the recovery of more sources than observations based on the already estimated mixing matrix [4, 10, 16, 25].

In the following sections first Algebraic ICA algorithm is presented followed by a permutation invariant algebraic performance benchmark for the evaluation of ICA

algorithms. The source recovery step has been summarized for the convenience of the reader. Further, a heuristic technique is discussed to enhance the algebraic resolution of the AICA solution for cases where some of the columns of the mixing matrix are algebraically “close” to each other. This is followed by simulation examples and a discussion on the algorithm itself.

5.3 Algebraic Independent Component Analysis

Independent Component Analysis can be best visualized as a projection method where the observed mixture data is projected onto a new basis, which minimizes the statistical dependence among the various data components. Such a representation is very appealing in a number of domains including statistics, signal processing, bio-medical, molecular, communications, audio, speech, sonar, astro-physics, textual-topics etc. The best-known application of ICA is termed as blind source separation (BSS). BSS provides the framework for the estimation of the original source signals directly from their measured static mixture with the only underlying assumption of mutual independence of the actual source signals. Blind Source Separation (BSS) is a special case of Blind Source Recovery (BSR), where the mixing is assumed to be static.

Algebraic Independent Component Analysis (AICA) is a new ICA algorithm that exploits algebraic operations and vector-distance measures to estimate the unknown mixing matrix in a scaled algebraic domain. AICA possesses stability and convergence properties similar to earlier proposed geometric ICA (geo-ICA) algorithms, however, the choice of the proposed algebraic measures in AICA has several advantages. Firstly, it leads to considerable reduction in the computational complexity of the AICA algorithm as compared to similar algorithms relying on geometric measures making AICA more

suitable for online implementations. Secondly, algebraic operations exhibit robustness against the inherent permutation and scaling issues in ICA, which simplifies the performance evaluation of the ICA algorithms using algebraic measures. Thirdly, the algebraic framework is directly extendable to any dimension of ICA problems exhibiting only a linear increase in the computational cost as a function of the mixing matrix dimension. The algorithm has been extensively tested for over-complete, under-complete and quadratic ICA using unimodal super-gaussian distributions. For other less peaked distributions, the algorithm can be applied either using a neighbourhood function, averaging of data [7].

The Algebraic ICA (AICA) algorithm is based on the linear algebraic notion of dot-product between two n -d data vectors as a distance measure. The dot product between two vectors \bar{a} and \bar{b} in the same basis $\Psi \in R^n$ is defined as

$$\bar{a}\bar{b} \triangleq \bar{a}^T\bar{b} = |\bar{a}||\bar{b}|\cos\theta_{ab} \quad (5.7)$$

where θ_{ab} is the n -d angle between the vectors \bar{a} and \bar{b} .

In case, the two vectors have been projected onto a unit hyper-sphere $\Psi^1 \in R^n$, then $|\bar{a}| = |\bar{b}| = 1$, and the dot product simply becomes a measure of the n -d. angle or difference in directions of the two vectors.

If the two vectors are co-incident (i.e., lie along the same n -d direction), the absolute value of the cosine of the n -d angle θ_{ab} is maximum (i.e., one), while if they are orthogonal to each other, the cosine of the angle θ_{ab} becomes zero. Therefore, after projection onto a unit hyper-sphere, the notion of minimum distance is equivalent to the notion of maximizing the absolute dot-product between the projected vectors. Using the

absolute value of the dot product produces a sign ambiguity, which is inherent in any ICA algorithm. This results in considerable computational savings during the implementation of the proposed algorithm, i.e.,

$$|\bar{a}\bar{b}| = |\hat{a}\bar{b}| \text{ where } \bar{a} = -\hat{a} \quad (5.8)$$

Another useful property for the dot-product is its insensitivity to the order of the product, i.e., $\bar{a}\bar{b} = \bar{b}\bar{a}$

In order to satisfy the stability and convergence requirements of AICA, the original source distributions should be unimodal, super-gaussian and preferably symmetric [3, 15, 23].

Consider the case where there are n original independent Lebesgue-continuous sources $s_k \in R^n; k = 1, 2, \dots, n$

$$s(i) = [s_1(i) \quad s_2(i) \quad \dots \quad s_n(i)]^T; i = 1, 2, \dots, N \quad (5.9)$$

that are mixed by a mixing matrix $H \in R^{m \times n}$, where m may or may not be equal to n , i.e.,

$$H = [h_1 \quad h_2 \quad \dots \quad h_n]; h_k \in R^m; k = 1, 2, \dots, n \quad (5.10)$$

where each column maps the k^{th} source signal s_k to the mixture space $m_k \in R^m$, i.e., $m = Hs$.

Without any loss of generality, it is assumed that the columns of the mixing matrix reside on the m -d unit hyper-sphere. In case the columns of H do not satisfy this constraint, they can be projected onto this hyper-sphere by

$$\tilde{H} = [\tilde{h}_1 \quad \tilde{h}_2 \quad \dots \quad \tilde{h}_n] \quad (5.11)$$

where

$$\tilde{h}_k = \frac{h_k}{\|h_k\|}; k = 1, 2 \dots n \quad (5.12)$$

This assumption is equivalent to having a scaling ambiguity in the mixing matrix estimates by the algorithm, which is again inherent in any ICA formulation.

In the mixture basis $m \in R^m$, the k^{th} source data is clustered along the m -d hyper-direction formed by the intersection of \tilde{h}_k and $-\tilde{h}_k$. The AICA algorithm therefore tries to estimate these hyper-directions (and hence the mixing matrix) using algebraic operations on the available mixture data.

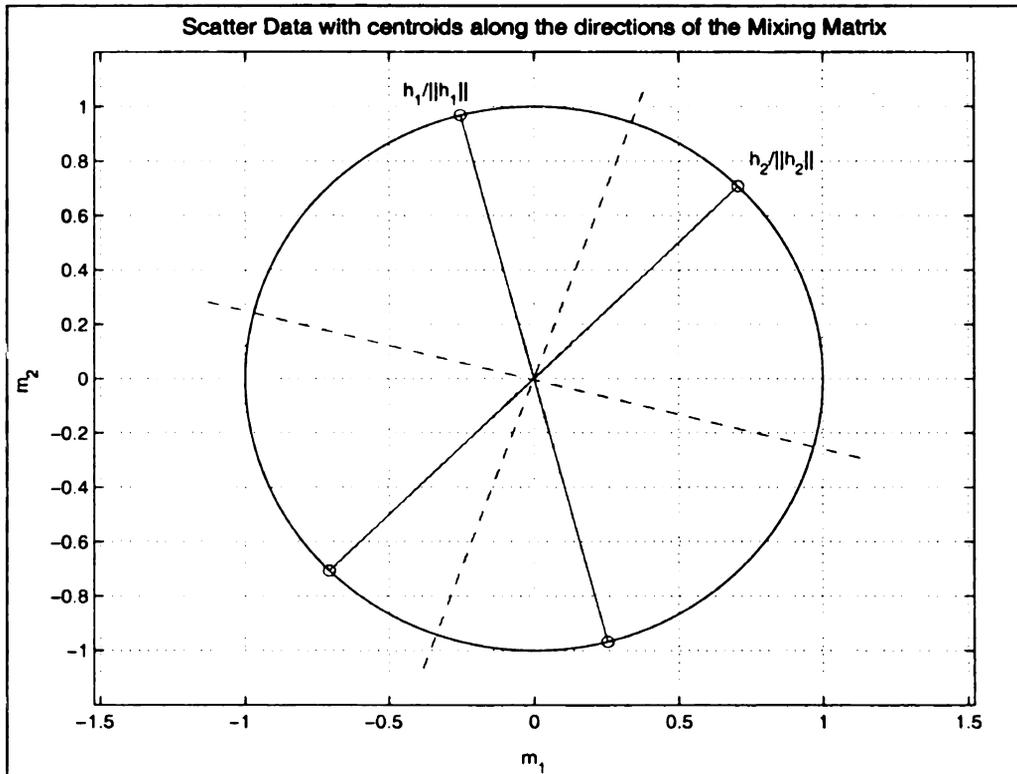


Figure 5.1. A scatter plot example of the projection of two sources onto the 2×2 mixture space.

For clarity of illustration, consider the case for $m = n = 2$ as shown in Figure 5.1. The columns of the mixing matrix \tilde{H} have been normalized so that they are projected onto the unit circle. In order to clearly indicate the direction of the plotted mixing matrix, lines are drawn between $h_i/\|h_i\|$ and $-h_i/\|h_i\|$. The mixture data has also been scaled to fit within the unit circle. The dashed lines indicate the basins of attraction for each of these directions.

5.3.1 The Algebraic ICA Algorithm

Assuming that the n -d mixture data has been accumulated over some period of time and there are N samples available that in essence can capture the distribution of the original sources and are also adequate for the algorithm to converge. Furthermore, the dimension $m \times n$ of the desired weight network is known. The algorithm can be implemented as follows:

1. Pick an n number of m -d starting elements for the weight matrix $W_0 = [w_1 \ w_2 \ \dots \ w_n]$ on the unit hyper-sphere. These initial starting points can be chosen in a number of ways, e.g., they can be either chosen randomly or evenly distributed around the unit hyper-sphere.
2. Choose an m -d mixture sample vector m_i taken at the i^{th} sampling instant, such that $m_i \neq 0, i = 1, 2, \dots, N$
3. Project this sample mixture vector onto the m -d unit hyper-sphere by

$$M_i = \frac{m_i}{\|m_i\|}, i = 1, 2, \dots, N \quad (5.13)$$

4. Find the dot-product of this projected mixture vector M_i with the current i^{th} estimates of the weight matrix $W_i; i = 0, 1 \dots N-1$

5. Determine the closest weight direction (i.e., the *winner column*) by

$$w_j^* = \arg \max_{w_j} \left| M_i^T * w_j \right|; j = 1, 2, \dots, n, i = 1, 2, \dots, N \quad (5.14)$$

6. Update only the winning weight direction by the following algorithm

$$w_j(k+1)^* = \begin{cases} \psi \left(w_j(k)^* + \eta(t) \operatorname{sgn}(M_i - w_j(k)^*) \right), & \text{if } M_i^T * w_j^* > 0 \\ \psi \left(w_j(k)^* + \eta(t) \operatorname{sgn}(-M_i - w_j(k)^*) \right), & \text{if } M_i^T * w_j^* < 0 \end{cases} \quad (5.15)$$

where, $\eta(t)$ - represents a small time adaptive learning rate. Often the learning rate is chosen to be an exponentially decreasing function with a decay rate that depends on the characteristics of the source distributions to be estimated.

$\psi(\cdot)$: is a nonlinear function that projects the updated weight vector again onto the

unit hyper-sphere, i.e., $\psi(\cdot) = \frac{(\cdot)}{\|(\cdot)\|}$

7. Determine the change in weight space

$$\Delta W_i = W_i - W_{i-1}; i = 1, 2, \dots, N \quad (5.16)$$

8. If $|\Delta W_i| < \varepsilon$ where $\varepsilon > 0$ is a small positive valued stopping criterion; the algorithm has converged to its final state.

9. Otherwise, update the iteration counter $i = i+1$, and repeat steps 2 to 9 till the algorithm converges.

For the proposed algorithm, the update is according to the rule: *one winner takes all*.

The algorithm reaches a *fixed-point* (or settles down to a solution), when

$E(\text{sign}(M_i - w_j(k)^*)) = 0$. The AICA algorithm at this stage has converged to the median of the distribution, which for a symmetric distribution also matches with the mean of the distribution. At convergence, there is no change in the weights and $E(W_{k+1}) = E(W_k)$. In [3, 15, 16] similar conditions have been derived for geo-ICA, termed as *Geometric Convergence Condition (GCC)*, but these conditions are in general satisfied by all self-organizing algorithms (SOM). Both AICA and geo-ICA algorithms, resemble the Kohonen's algorithm for self-organizing maps (SOM) with a trivial zero-neighborhood [7]. In a 2-sources 2-mixtures case, the algorithm possesses only two fixed points [15],

$$W^{(\infty)} = \begin{cases} X \begin{bmatrix} \tilde{h}_1 & \tilde{h}_2 \end{bmatrix} \\ X \begin{bmatrix} (\tilde{h}_1 + \tilde{h}_2)/2 & (\tilde{h}_1 - \tilde{h}_2)/2 \end{bmatrix} \end{cases} \quad (5.17)$$

where, X is a generalized permutation matrix.

The convergence of the algorithm to the true solution is therefore a function of initialization and distribution of the sources to be estimated. In case of super-gaussian sources, however, only the true solution is the stable fixed point. Another observation for lower dimension problems is that a weight vector may get trapped between two data clusters and may require re-initialization especially if the two directions are very close to each other. However, for larger dimension problems, the probability of this occurrence is negligible.

The proposed algorithm framework has similarities to the Geometric ICA algorithm [3, 12, 15, 17] but requires significantly less computation due to the choice of a different distance measure and a reduced number of parameters to update. The use of a neighborhood function in the proposed algorithm is currently being investigated for

source distributions other than the super-gaussian. There is interest in the ability to handle signals with multiple source distributions [20].

Figure 5.2 presents the results after convergence of the proposed algorithm for the 2×2 case with initialization as presented in Figure 5.1. The initial estimates of the weight vector were chosen randomly. It is evident from Figure 5.2 that the algorithm caused the initialized weights to converge to the nearest desired direction.

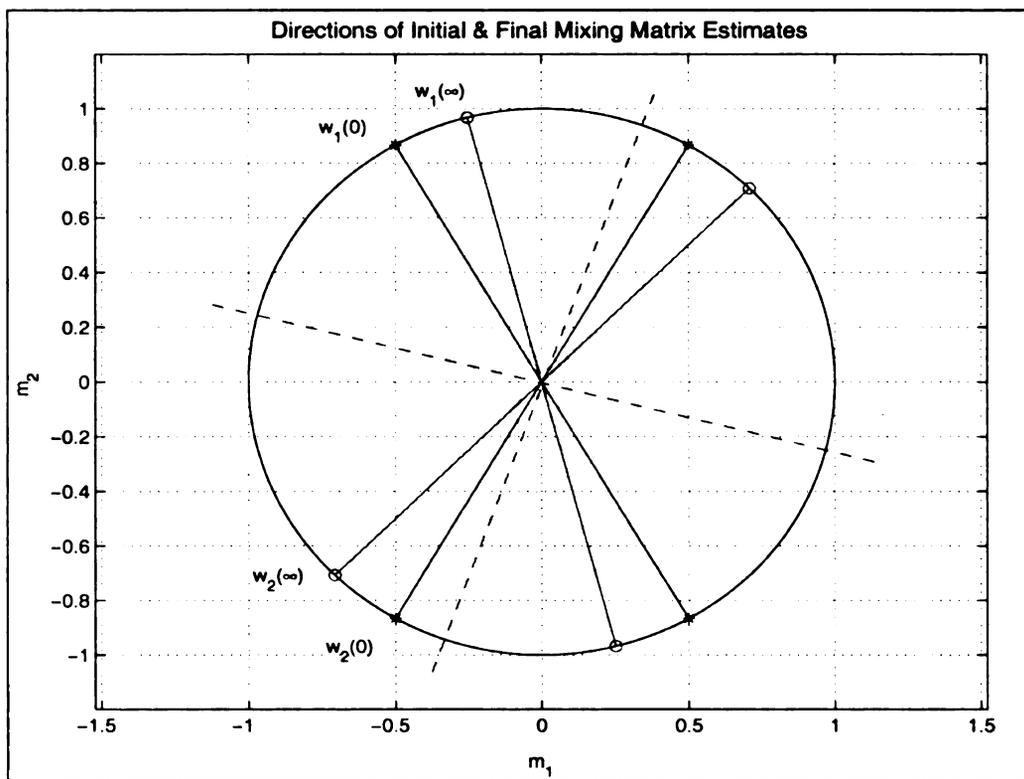


Figure 5.2. Convergence of the weights in the 2×2 mixture space; $w_i(0)$ denotes the initial weight locations while $w_i(\infty)$ denotes the final weight location.

5.3.2 Algebraic Matrix-Distance Index (AMDI)

Algebraic Matrix Distance Index (AMDI) is primarily a scaling, sign and permutation invariant symmetric numerical index to estimate the “similarity” between the column-bases of two “normalized matrices” with conformal dimensions. Unlike the Multichannel Intersymbol Intereference (MISI) measure (also called Amari Index by some) used in ICA to measure the diagonalization of the global transfer function, AMDI is a measure of similarity between two column-normalized matrices with similar dimensions.

These attributes also make AMDI suitable for comparison of results for algorithms such as AICA and geo-ICA especially for overcomplete ICA where computation of MISI is not possible. In order to compute this metric, the columns of the matrices are normalized by projecting them onto the conformal unit hyper-sphere by dividing all the elements of each column by the norm of the corresponding column. The AMDI is given by

$$AMDI(W, H) = \frac{n - \sum_{rows} \max(|W \cdot H|)}{n} + \frac{n - \sum_{cols} \max(|W \cdot H|)}{n} \quad (5.18)$$

where

W, H – are two column-normalized matrices of similar dimensions.

The Algebraic Matrix Distance Index (AMDI) has the following properties:

- $0 \leq AMDI(W, H) \leq 2$; $AMDI(W, H) = 2$ only if both matrices are orthogonal to each other in the n -d space, also
- $AMDI(W, H) = AMDI(H, W)$

5.3.3 Resolution of AICA Local Minima and Ambiguities

Algebraic ICA is a member of a class of data-dependant self-organizing algorithms. The performance of the algorithm is a function of the choice of the initial weights, learning rate and most importantly characteristics of the data. While AICA weight columns converge to the directions (or columns) of the actual mixing matrix in the limit, provided the data is unimodal with a high value of kurtosis, and the mixing matrix directions are sparse [3, 16, 22]. For data that does not meet these requirements either due to lower kurtosis, or due to the actual mixing matrix directions being “close” to each other (especially for the overcomplete cases), the algorithm does not give the desired estimate performance. For such cases, a better estimate of the mixing matrix may be derived using more than one trials of the algorithm. This technique is also useful for the cases where the available data is inadequate for the AICA algorithm to estimate the mixing matrix. A heuristic stepwise approach for such ambiguous cases can be followed as:

1. Initialize all the columns of the estimator matrix at the same location. For each successive trial, choose a different initial location, e.g., one way to do so is to initialize the weights at possible m -d binary combinations of 1's and zeros. For m mixtures there are possible $2^m - 1$ such initializations possible, i.e. leaving aside the trivial combination of all zeros.
2. Execute the AICA algorithm for each initial condition. For a sparse mixing matrix, the algorithm converges independent of the initialization provided adequate data is available [17, 24]. In most cases, the algorithm will converge during one of the trials.

3. In case, the algorithm fails to converge for any choice of initialization. Note, in a blind scenario this condition cannot be verified and multiple trials serve as a mean of estimating the mixing matrix with a higher confidence level. The AICA estimate for all cases can be post-processed to determine the high confidence candidate weight columns, see steps 4 to 10.
4. Determine the list of “close” columns in all trials. This can be algebraically realized by choosing the corresponding columns with dot product values above a certain threshold α . This threshold α can be chosen as a function of data sparseness and the desired resolution of the AICA algorithm.
5. Concatenate the column pairs with dot products better than a higher threshold β , where $\beta \geq \alpha$ and a common column.
6. Repeat Step 5 till all possible concatenations are made (this may take $\eta - 1$ iterations for η trials, if 1 combination is made each time).
7. Determine the independent z columns as the mixing matrix candidate columns.
8. The concatenated column cases are given a relatively higher rank. The candidate columns are arranged as a function of their dot-products values. At this instant, one option is to choose the M highest ranked columns as the final mixing matrix candidate columns.
9. In case the final column candidates making the cut are smaller than M , deflationary approaches can be used after estimation of the output using the mixing matrix estimate. The remaining columns can now be determined in a consecutive stage.

10. Alternately when the candidate number of columns is larger than M , instead of choosing the columns via ranking, all the columns may be used for the source recovery stage, then the selection of appropriate outputs is made by measuring the independence of recovered outputs, e.g., such an estimate can be made using Quadratic Information Measure [21].

Using the above-mentioned technique, for a 3-d mixture space, the mixing matrix with a resolution of better than a 15° (approx. 0.25 rad) hyper-cone in the weight space has been successfully estimated. See simulation III for some more details.

5.3.4 Inferring Sources from Overcomplete Mixtures

Once the mixing matrix has been estimated by the AICA algorithm, the original sources can be estimated. For quadratic ICA, the sources can be estimated by computing the inverse of the estimated mixing matrix W , i.e.

$$\hat{s} = W^{-1}m \quad (5.19)$$

For the case of overcomplete mixtures, the recovery of the sources is in general not possible. However, for the case of sparse super-gaussian sources, once the mixing matrix has been estimated, interior point linear programming (IP-LP) techniques [4] can be used to determine the best possible estimate of M -d sources from m -d mixture where $M > m$. The technique uses interior point LP with an L_1 norm based performance functional. Standard interior point L_2 norm based LP methods can be used for this technique according to the following procedure [4, 10]:

$$\begin{array}{ll} \min c^T x & \min \|\hat{s}\|_1 \\ \text{s.t.} & \Rightarrow \text{s.t.} \\ Ax = b, x \geq 0 & W\hat{s} = m \end{array} \quad (5.20)$$

where

$A = [W \quad -W]$: W is the estimated mixing matrix of dimension $m \times M$, note by choice $M = n$ in this case.

$x = [u \quad v]^T$: represents the M -d intermediate recovered sub-signals with the constraints $u \geq 0$ and $v \geq 0$

$c = [\underline{1} \quad \underline{1}]^T$: constrains the recovered M -d sub-signals u and v equally, $\underline{1}$ is an M -d vector of ones.

$b = m$: m are the m -d overcomplete mixtures.

The M -d outputs are then determined using the relation

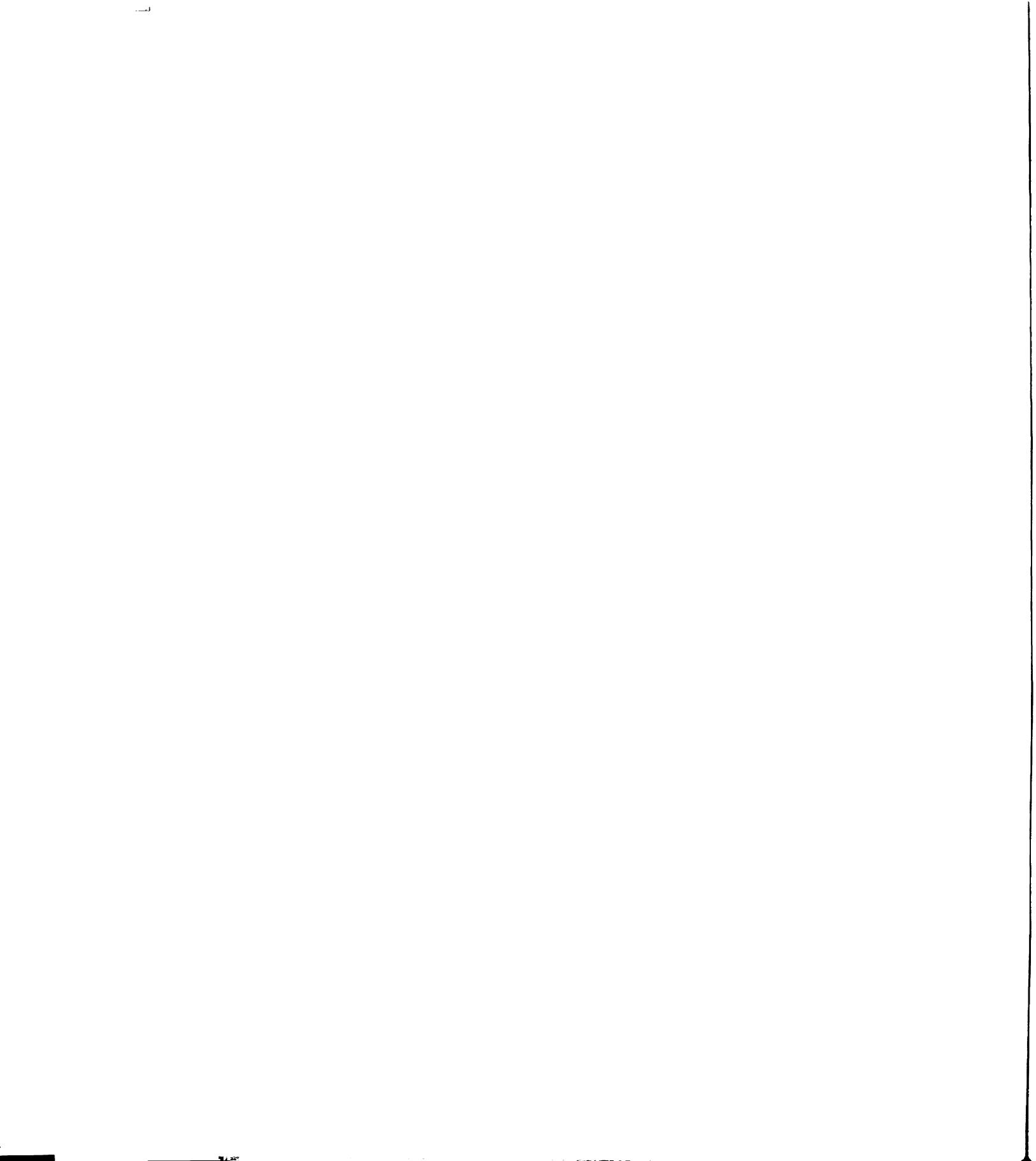
$$\hat{s} = u - v \tag{5.21}$$

Another technique using the shortest path approximation is explained in [3, 16].

This technique requires less computation than the aforementioned technique. However, in our experience the quality of sources recovered using the LP approach is better than the latter approach.

5.3.5 Simulation Examples

Four simulation examples are being presented: the first example demonstrates the issue of multiple fixed points (or presence of local minima) in the AICA algorithm using independent identically distributed (IID) data. Then there are two examples using overcomplete mixtures of 3 and 5 sparse speech segments, respectively. In these sparse speech samples, words are spoken with certain time intervals to increase the sparseness of the actual speech signals. The recovery of these speech signals from their 2 and 3 static mixtures, respectively, is presented. The AICA algorithm is first used to estimate the



mixing matrix, while the speech sources are inferred using the L_1 norm interior point linear programming.

The fourth example discusses BMMR for a 6×3 mixture using IID laplacian sources. The motivation for this example are cases where some of the mixing directions are “so close” that algebraic ICA algorithm cannot discriminate these algebraically adjacent columns due to the relatively high overlap of data aligned to them (resolution). For such cases, the proposed ambiguity resolution technique is applied, which allows for the exploration of the complete mixture bases in an orderly and organized manner.

5.3.5.1 Simulation I: 4×2 Laplacian Data

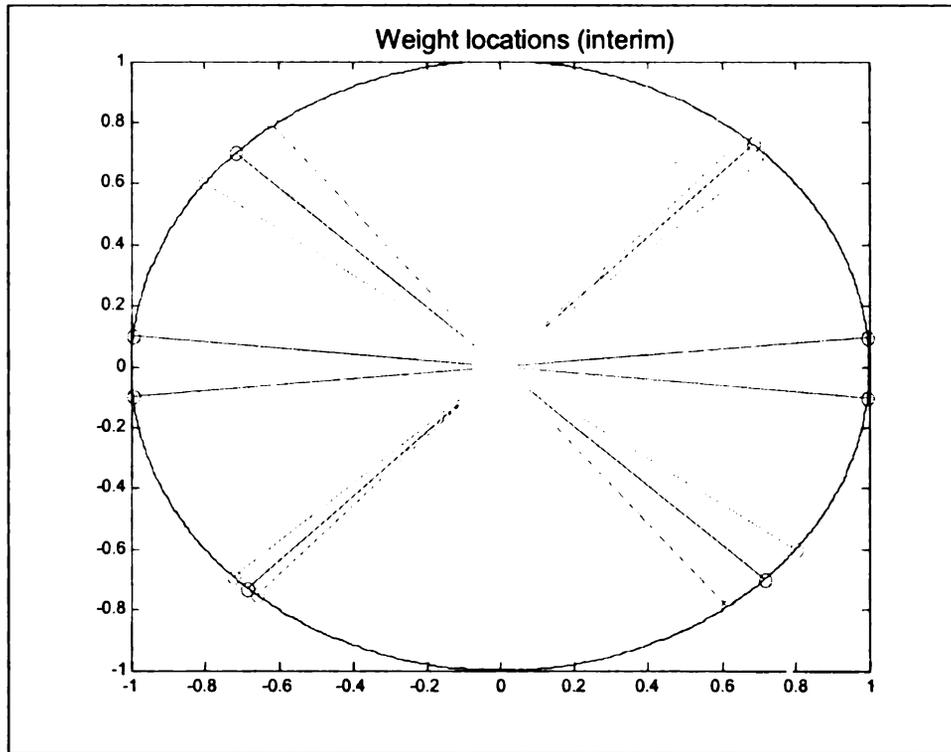
In this simulation, a synthetic random matrix mixes 3 independent identically distributed laplacian sources. For this simulation the mixing matrix is

$$\tilde{H} = \begin{bmatrix} 0.7231 & -0.6121 & -0.6587 & 0.8005 \\ 0.6908 & 0.7908 & 0.7524 & -0.5994 \end{bmatrix}$$

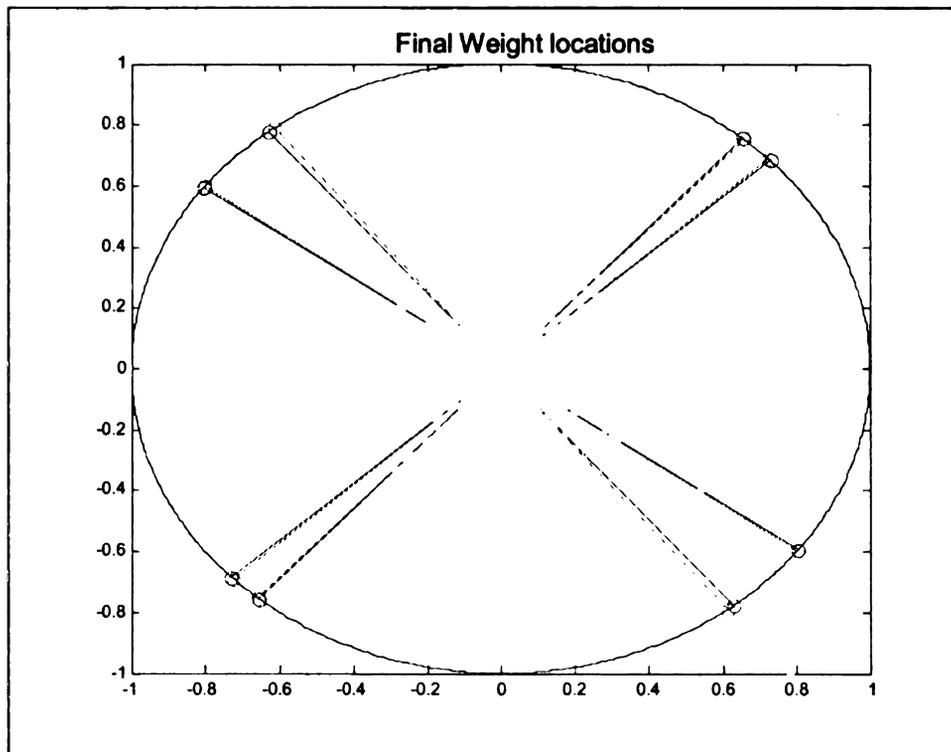
All the weights for the AICA algorithm in this case are initialized along the real axis. It is observed that the algorithm converges to the undesired fixed point. This is sensed by AICA observing that only two of the weights are updated, while the other two are not, see Figure 5.3.a. The algorithm in this case re-initializes the discarded weights very close to the weights which are currently being updated. This results in the convergence of the algorithm to the true solution, see Figure 5.3.b. The determined final weights for this simulation are

$$W = \begin{bmatrix} 0.7265 & 0.8039 & 0.6448 & 0.6075 \\ 0.6872 & -0.5948 & 0.7644 & -0.7943 \end{bmatrix}$$

and the achieved AMDI performance is less than 0.0003.



(a)



(b)

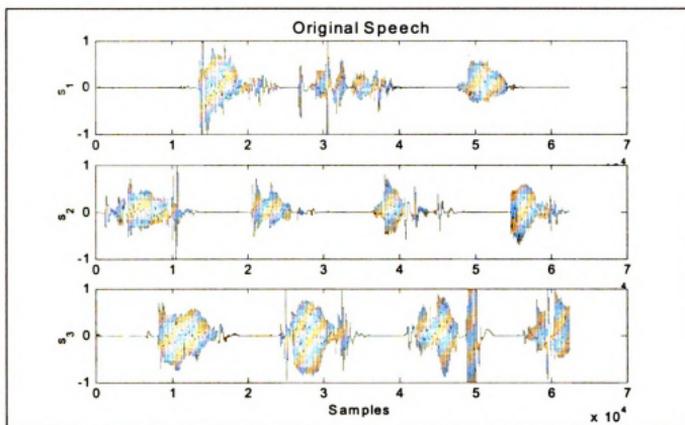
Figure 5.3. BMMR for Laplacian Data using AICA: (a) Interim Results,
(b) Final Convergence

5.3.5.2 Simulation II: Overcomplete Speech Data (3×2 Case)

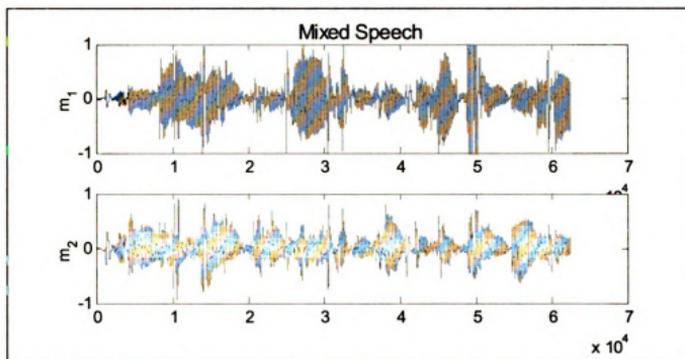
A simulation example for an overcomplete mixture which comprises of three speech signals mixed by a random mixing matrix into two mixtures is being presented. The AICA algorithm is first used to estimate the mixing matrix, while the speech sources are inferred using the L_1 norm interior point linear programming.

The AICA estimated mixing matrix was very close to the original mixing matrix used with a normalized AMDI of 0.001. The recovered sources show a permutation between the first and third sources. Also notice that the recovered signals from the overcomplete mixture in Figure 5.5.a are not without artifacts but very closely resemble the original sources in Figure 5.4.a.

Observe that the separated quality achieved using separate Blind Mixing Matrix Recovery (BMMR) and source recovery steps far exceeds other algorithms [1, 4, 10], provided a good mixing matrix estimate was found.



(a)



(b)

Figure 5.4. Overcomplete Speech Mixture (3×2 Case): (a) Original speech sources,

(b) Mixed speech

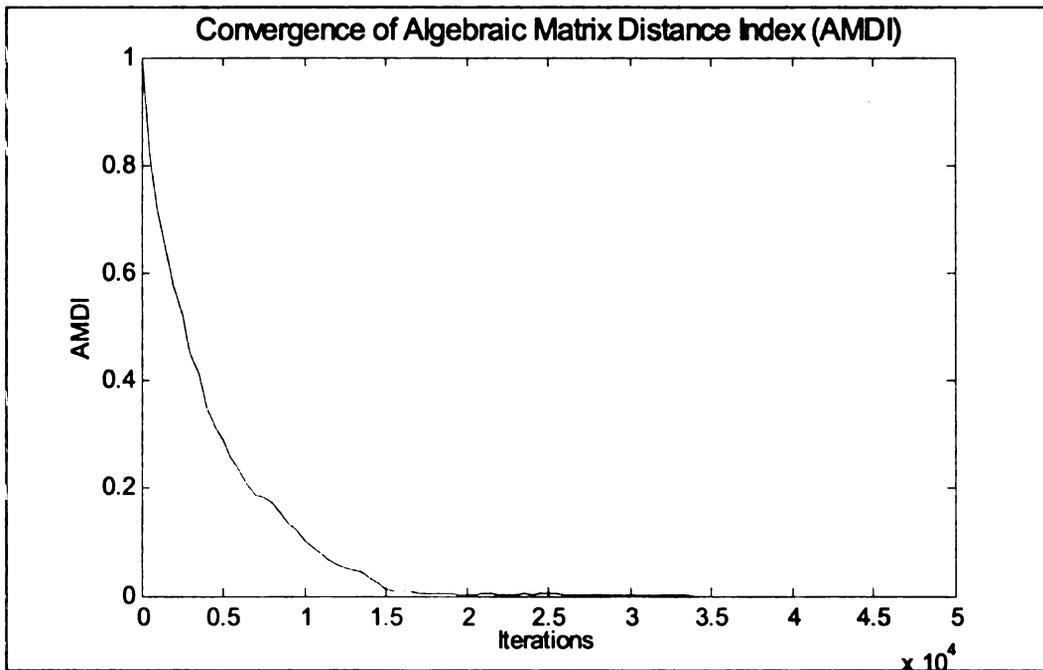
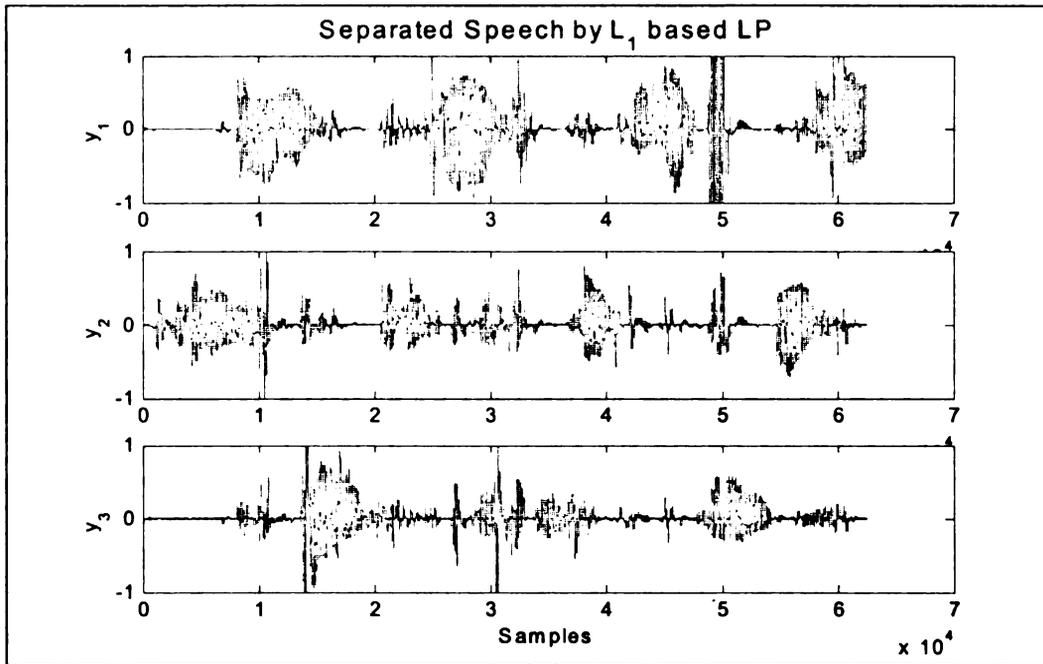


Figure 5.5. Overcomplete Speech Mixture (3×2 Case): (a) Recovered speech,
 (b) Convergence of Normalized AMDI over 100 experiments

5.3.5.3 Simulation III: 5 × 3 Speech Data

For this simulation, the simulation example comprises of an overcomplete speech mixture. Five utterance are mixed using a 5 × 3 mixing matrix, as given below.

$$\tilde{H} = \begin{bmatrix} 0.4872 & -0.9063 & 0.3285 & 0.6203 & -0.8837 \\ 0.2021 & -0.0247 & 0.2714 & -0.6752 & 0.2019 \\ -0.8489 & -0.4220 & 0.9047 & 0.3990 & 0.4223 \end{bmatrix}$$

The AICA algorithm is first used to estimate the mixing matrix W , while the speech sources are inferred using the L_1 norm interior point linear programming.

The AICA algorithm converges to the estimate in approximately 25000 iterations achieving an AMDI of 0.002. The estimated mixing matrix W exhibits a permutation, where the columns appear in the order 4, 3, 5, 1 and 2. Also the second and fourth columns of W have an inverted sign.

$$W = \begin{bmatrix} 0.6243 & -0.3298 & -0.8787 & -0.4861 & 0.9015 \\ -0.6724 & -0.2773 & 0.1993 & -0.2020 & 0.0228 \\ 0.3949 & -0.9157 & 0.4240 & 0.8445 & 0.4160 \end{bmatrix}$$

In Figures 5.6-5.8, the 3 speech mixtures, the original 5 speech utterances and the 5 inferred speech sources are shown. Notice that the recovered signals from the overcomplete mixture in Figure 5.8 are not without artifacts (introduced primarily by the source inferring stage) but closely resemble the original speech utterances. A listening test proves this. The speech simulation data can be downloaded at <http://www.egr.msu.edu/bsr/results>.

Observe that the two stage overcomplete mixture recovery approaches [16, 24, 25] are relatively simpler to implement as compared to other complex approaches [1, 4, 10]. Also combined BMMR and source inferencing provide satisfactory results as long as

the underlying probability distribution is sparse. The performance of the AICA algorithm depends on the quantity of preamble data available for estimation of the weight matrix. Although the algorithm may be implemented for any dimensions, the amount of data required may be immense. This can be dealt with by repeatedly training for weight estimation on a batch of data, provided the data batch size is adequate to represent the actual underlying source distributions.

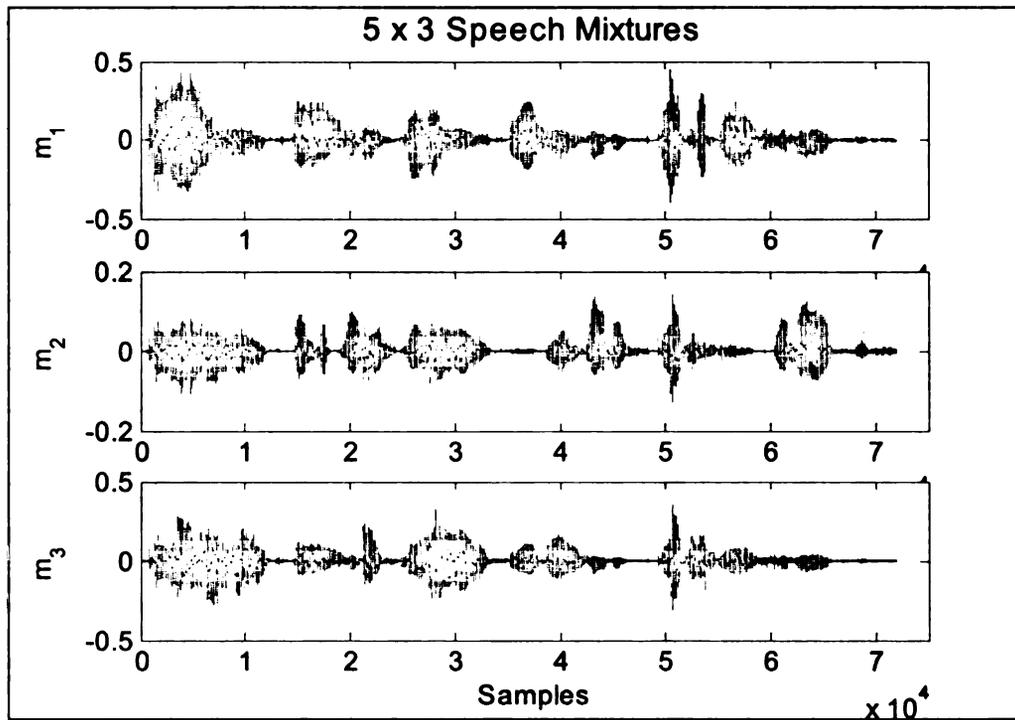


Figure 5.6. Overcomplete Speech Mixture (5×3 Case): 3 Overcomplete Mixtures

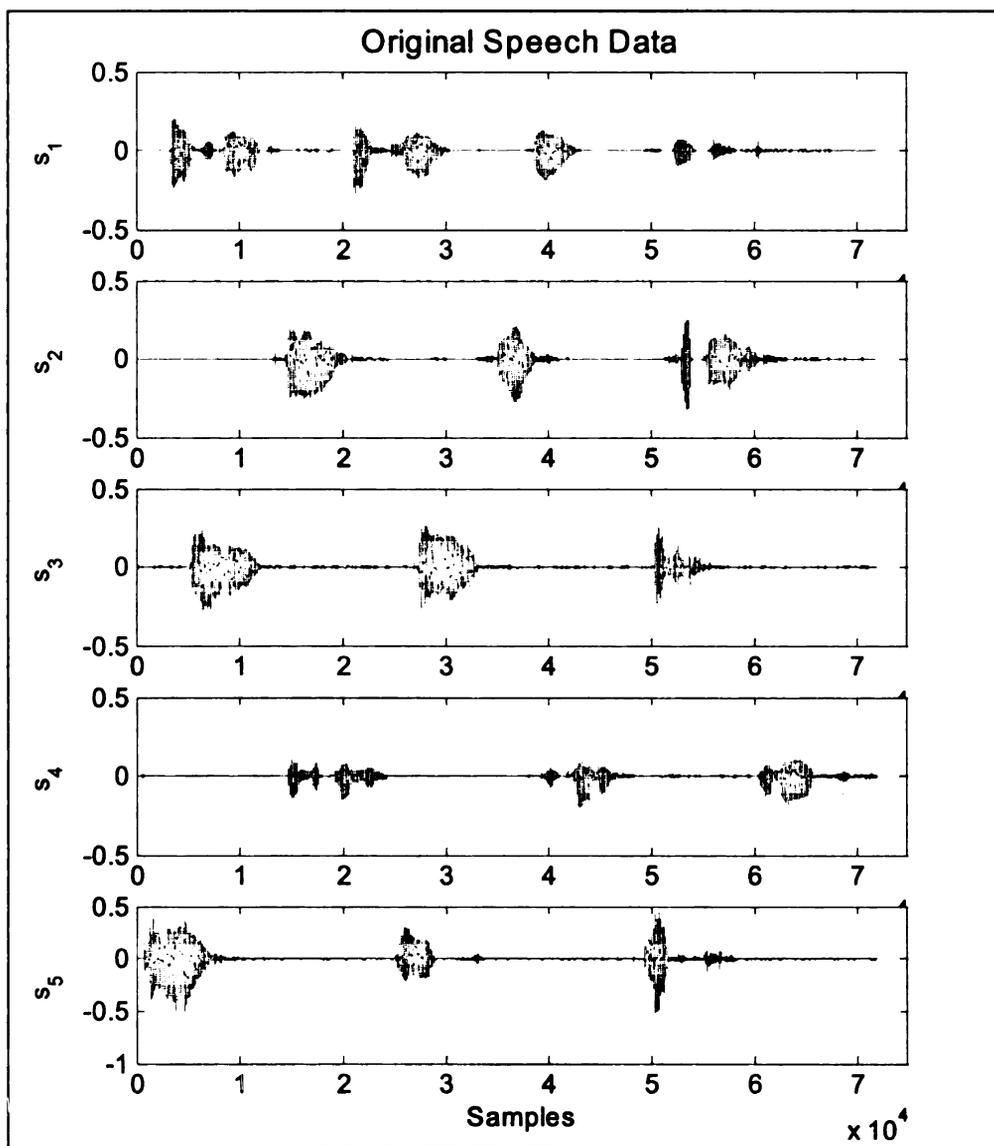


Figure 5.7. Overcomplete Speech Mixture (5×3 Case): Original 5 Speech Utterances

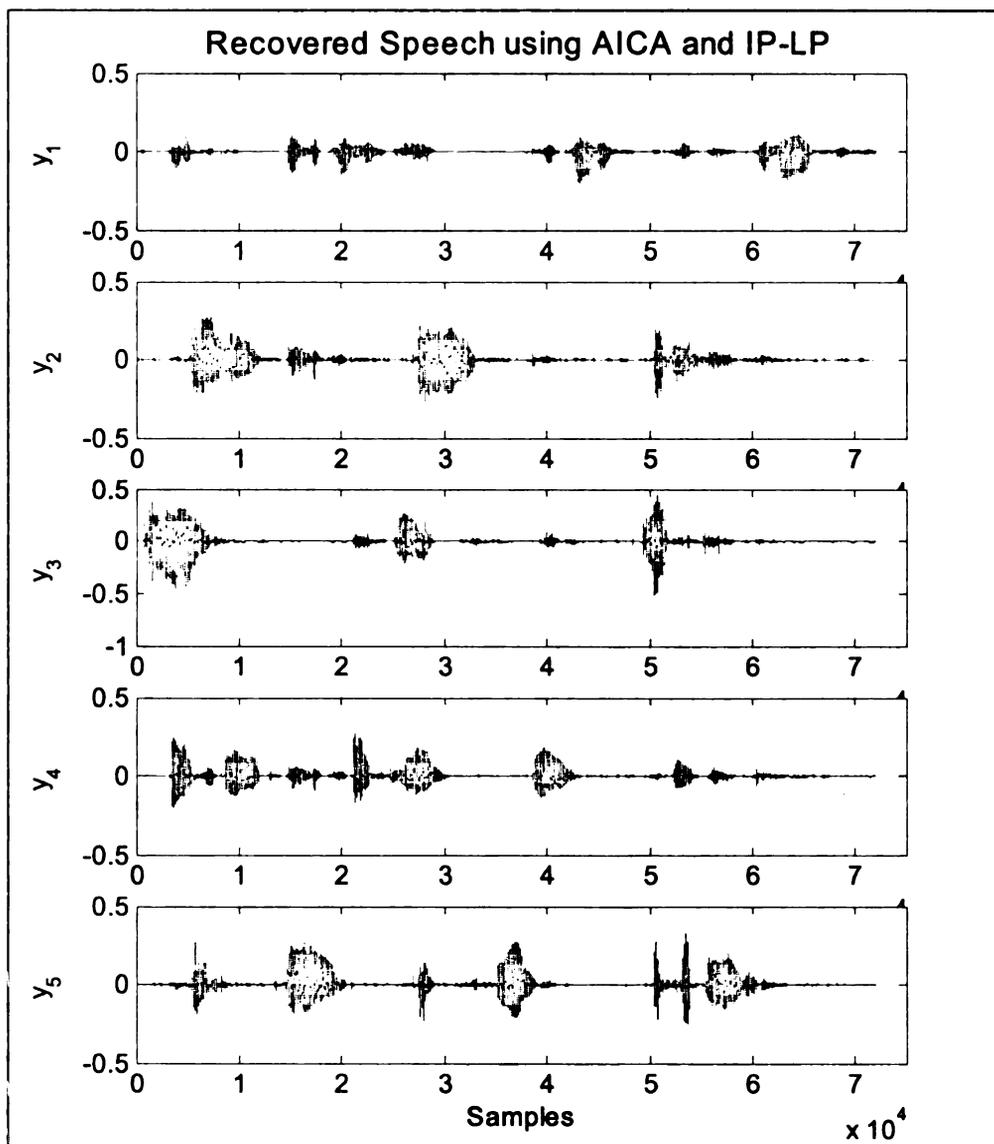


Figure 5.8. Overcomplete Speech Mixture (5×3 Case): Recovered 5 Speech Utterances

5.3.5.4 Simulation IV: Independent Identically Distributed (iid) Laplacian Data

For this simulation, six iid laplacian sources are mixed by a randomly generated matrix of dimension 6×3 into three mixtures. Also presented are some of the results where the columns of the mixing matrix are close and the matrix is estimated using the ambiguity resolution scheme is presented in section 1.3.4.

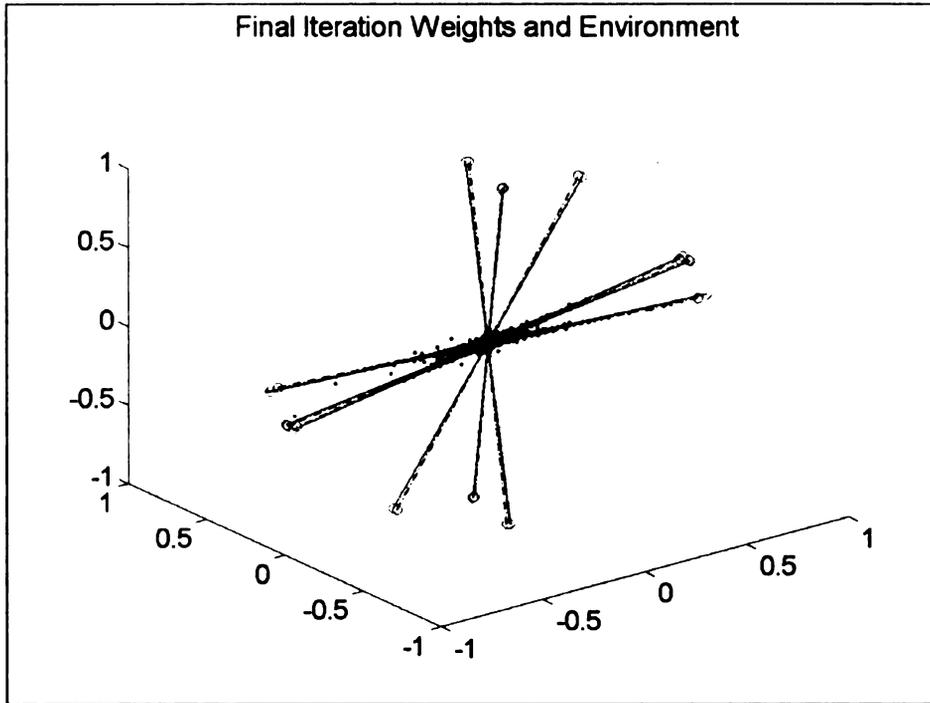
For the mixing matrix below the columns 5 and 6 are within 0.4 rad of each other. AICA is able to estimate the mixing matrix successfully using the data from 3 trials.

$$\tilde{H}1 = \begin{bmatrix} 0.0226 & -0.4698 & 0.5311 & -0.2640 & 0.9597 & -0.9498 \\ -0.0620 & -0.0327 & -0.5890 & -0.4616 & 0.0309 & 0.1029 \\ 0.9978 & -0.8822 & 0.6091 & -0.8469 & 0.1793 & 0.0021 \end{bmatrix}$$

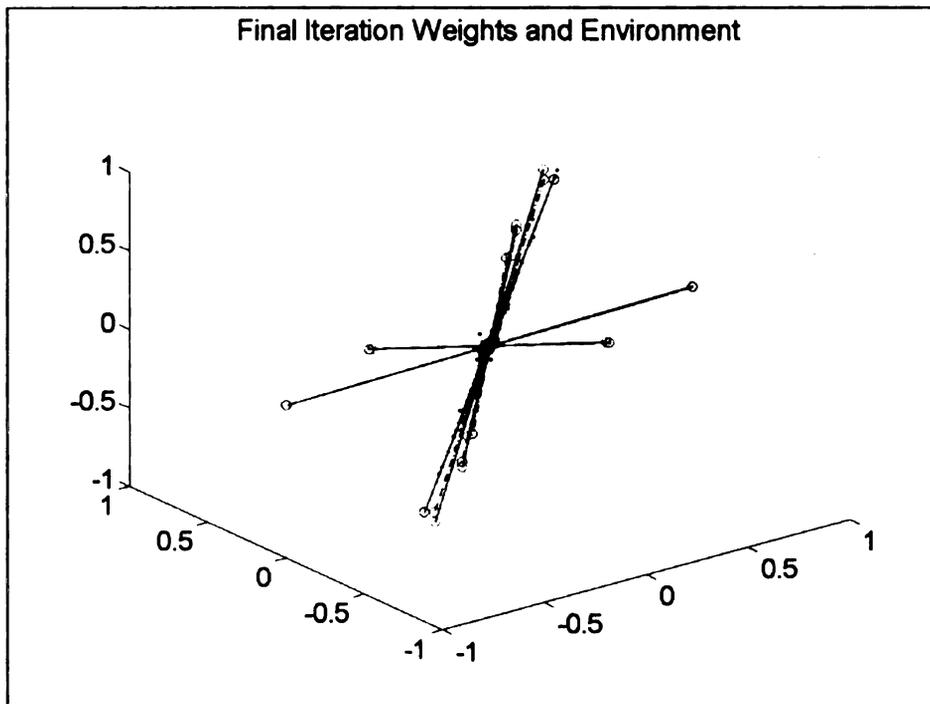
Figure 5.9.a shows the results using the proposed technique. The recovered weight are shown in red, while the blue dash-dot line shows the actual mixing matrix locations. In another simulation, the mixing matrix is

$$\tilde{H}2 = \begin{bmatrix} 0.0839 & 0.6345 & 0.6606 & 0.9997 & 0.3875 & 0.3164 \\ 0.8850 & 0.6525 & 0.7506 & 0.0029 & 0.1546 & -0.0014 \\ -0.4579 & 0.2144 & -0.0149 & 0.0248 & 0.8509 & 0.9486 \end{bmatrix}$$

where columns pairs 2 and 3; 5 and 6 are within 0.35 rad of each other. The AICA can resolve these columns successfully using results from 7 trials. The final result is shown in Figure 5.9.b.



(a)



(b)

Figure 5.9. Unambiguous AICA estimate using multiple runs: (a) using $\tilde{H}1$,
 (b) using $\tilde{H}2$

The performance of the AICA algorithm depends on the quantity of preamble data available for estimation of the weight matrix. Although the algorithm may be implemented for any dimensions, the amount of data required may be immense. In such cases, the same data batch or a randomly drawn batch may be repeated during the weight estimation process provided the amount of data in the batch is adequate to represent the actual source distributions [7]. The convergence of weights certainly depends on the initialization. Another observation for lower dimension problems is that there are more chances for a weight vector to be trapped between two algebraically close data clusters. However, for larger dimension ICA problems, the probability of this occurrence is minimum.

5.4 Remarks

In this chapter, the adaptation issues for the undercomplete and the overcomplete BSR have been discussed. A new Algebraic ICA (AICA) algorithm based dot-product has been presented. The algorithm estimates the static mixing matrix under the assumption that the original sources are unimodal and possess positive fisher kurtosis. It is demonstrated that the proposed AICA algorithm for the case of overcomplete mixing is used for BMMR. The original sources can then be estimated using interior point L_1 norm based LP technique. Also presented is AMDI, which is a matrix-distance measure to compare two matrices, which may be both scaled and permuted with respect to each other. Further, a heuristic technique to determine the BMMR solution when the mixing matrix columns are close to each other is also discussed. The algorithm has been successfully applied to the problems where the dimension of mixtures is half the number of original independent signal sources. Currently algorithm extension using a

neighborhood function is being investigated for gaussian and sub-gaussian distributed signals. The speech data for the examples can be accessed at the website <http://www.egr.msu.edu/bsr/results> [26].

Chapter 6

BLIND SOURCE RECOVERY APPROACHES FOR WIRELESS DS-CDMA AND WCDMA DOWNLINK SYSTEMS

Code Division Multiple Access (CDMA) is based on spread-spectrum technology and is a dominant air interface for the proposed modern 3G and 4G wireless networks [145]. The transmitted CDMA signals propagate through noisy multipath fading communication channels before arriving at the receiver of the user equipment (UE). In contrast to classical single-user detection (SUD) algorithms, which do not provide the requisite performance for modern high data rate applications, multi-user detection (MUD) approaches require a lot of a-priori information not available to the UE. In this chapter,

three promising adaptive natural-gradient based info-theoretic blind user detection approaches capable of handling the wireless dynamic environments are proposed. The first approach, Blind Multi User Detection (BMUD), is the process of simultaneously estimating multiple symbol sequences associated with all users in the downlink of a Code Division Multiple Access (CDMA) communication system using only the received wireless data. This approach is applicable to CDMA systems with short codes but becomes impractical for systems using long codes. Also proposed are two other adaptive approaches, namely RAKE-Blind Source Recovery (RAKE-BSR) and RAKE-Principal Component Analysis (RAKE-PCA) that fuse an info-theoretic stage into a generic RAKE receiver. This results in robust detection algorithms with performance exceeding the standard LMMSE detectors for both DS-CDMA and WCDMA systems under conditions of congestion, imprecise channel estimation and unmodeled multiple access interference (MAI).

6.1 User Detection in CDMA Systems

The conventional detection schemes for CDMA signals only exploit second order statistics among user codes. Practically, however, the underlying user data symbol sequences are in general mutually (near-) “independent”. This is a key assumption, which propels the application of info-theoretic learning approaches such as information maximization [21] and minimum mutual information [14] to the realm of CDMA. The use of these algorithms is justified since a wide sense stationary slowly fading multipath CDMA environment can be conveniently represented as a linear multi-channel convolution model (see section 6.3.1) [39, 123]. The received CDMA signal can be considered as a sum of several non-gaussian random variables generated by the linear

convolutive transformations of statistically (near-) independent component user variables [58, 141]. This linear transformation accounts for the user spreading codes, the cell scrambling codes (in case of a cellular architecture), multiple channel paths and slowly fading channel effects [141]. Our goal is to estimate a linear transformation to counteract, as “optimally” as possible, the effects of a first transformation-- resulting in the recovery of the original user signals under the constraint of knowing only the user’s signature code (and the corresponding cell scrambling code for a two stage implementation) [58, 62, 122, 123, 141].

For a generic DS-CDMA implementation using only a single stage of user spreading codes (e.g., using PN codes, gold codes or Walsh Hadamard codes etc.), info-theoretic blind source recovery techniques [62, 123, 124] can be employed to recover the user data sequences in the multi user environment. Blind Source Recovery (BSR) is the process of estimating the original “independent” user-specific symbol sequences independent of, and even in the absence of, precise system/channel identification [109, 110]. In typical downlink signal processing, where many of the system parameters are unknown, including the codes for co-existing users at any instant of time, one can use the blind techniques for better estimate of the user-specific signals [62, 123, 124]. Alternately, Blind Multi-User Detection (BMUD) algorithms based on the Natural Gradient Blind Source Recovery (BSR) techniques in both feedback and feedforward structures can be used [109, 110, 135, 139]. The “quasi-orthogonality” of the spreading codes and the inherent “independence” among the various transmitted user symbol sequences form the basis of the proposed BMUD methods. The proposed structures and algorithms demonstrate promising results as compared to the conventional techniques

comprising, e.g., Match Filter (MF), RAKE and block LMMSE [57, 98, 120, 121]. Our proposed algorithms can be implemented either using the batch or the more computationally efficient instantaneous update methods [110, 112]. Although batch implementations exhibit better performance [123], it is however accompanied by longer latency and more involved implementation structures not suitable for a UE/MS. Focus is placed on the instantaneous (or on-line) performance of the BSR algorithms only, which still exceeds the performance of other approaches [122, 124].

This (on-line) detection technique can be easily extended to other CDMA implementations, using relatively short scrambling codes, but becomes impractical in WCDMA downlink using long scrambling codes. In spite of the fact that very low bit error rates (BER) can be achieved with this technique and the detection process does not even require the knowledge of user's own signature code, the recovered signal stream is at the symbol level with no user identification [62, 123]. Further, inherent sign and permutation ambiguities exist in BSR (scaling is not relevant as the recovered streams are either BPSK or QPSK etc.); therefore user identification is not possible unless some preamble or pilot data is transmitted periodically [62, 123]. This periodic requirement stems from the dynamic nature of the wireless communication scenario where users may dynamically enter or exit the system. The environment structure also varies widely due to the mobility of the MS/UE [4].

With these practical constraints in mind, new algorithms are proposed by an infusion of info-theoretic learning algorithms such as static Blind Source Recovery (BSR) (or Independent Component Analysis, ICA) and Principal Component Analysis (PCA) in the existing structure of a RAKE receiver [122, 125]. The purpose of this

additional info-theoretic stage is to counter, as best as possible, the unmodeled multiple access interference (MAI) and the additive (white gaussian) noise in the channel [122]. Further, use of a simple info-theoretic stage does not make the receiver structure too complex (in fact, it is simpler than most other proposed adaptive LMMSE implementations [44, 121]). RAKE-PCA uses up to second order statistics, as compared to RAKE-BSR, which utilizes higher order statistics. This results in slightly simpler update structure for the RAKE-PCA, but the performance of the RAKE-BSR is found to be better than RAKE-PCA. Further, assuming the score-function for the ICA update law to be chosen properly, the resulting equalization matrix in case of RAKE-BSR has relatively smaller element values (energy) as compared to the corresponding matrix for RAKE-PCA, which can be translated to the need of fewer memory bits. Lastly, both RAKE-BSR and RAKE-PCA use all the available user information, so that there are no issues of user identification and sign ambiguities in this case [122, 125]. The main advantage of both the RAKE-BSR and RAKE-PCA algorithms is the improved BER performance for the UE/MS without the need of any additional information than what a standard RAKE receiver already has. The proposed algorithms can be applied to both generic DS-CDMA and modern 3G UMTS implementations.

The remaining chapter is organized as follows: the next section discusses the CDMA forward communication channel (or downlink) and the limitations of the conventional user detection approaches. In the third section, the signal models are derived for a generic single stage spreading DS-CDMA and the two stage WCDMA (or 3GPP UMTS FDD) downlink scenarios [143]. This is followed by a section on detection algorithms for UE/MS. Initially the structure of the well-known linear detectors such as

MF, RAKE, LMMSE etc. is briefly described. This is followed by a description of the proposed info-theoretic structures for DS-CDMA and WCDMA wireless networks. The fifth section presents simulation results under various test conditions and congestion levels for the systems. Simulation results using the BER performance as a performance measure support the superiority of the proposed info-theoretic formulations for various DS-CDMA and WCDMA situations. In the conclusion section, the findings of the work are enlisted and summarizing remarks are made on the effectiveness of the proposed algorithms.

6.2 CDMA Downlink Communication Channel

Code Division Multiple Access (CDMA) uses spread spectrum techniques in which all the users share the same temporal and spectral resources [58, 98, 121]. In the downlink signal processing, each user is identified by a unique signature code, which is chosen to be “orthogonal” (or quasi-orthogonal) to the signature codes allotted to other users in the system. The Direct-Sequence Code-Division Multiple Access (DS-CDMA) is a promising data transmission technique capable of high data rates and better immunity to channel impairments and noise [98]. In DS-CDMA, the signal energy is “spread” over a wide frequency range, which reduces the effects of fading channels. CDMA systems utilize the given spectral and temporal resources more effectively as compared to other multiple access schemes such as FDMA and TDMA [145]. This results in a larger user capacity per given resources [57, 98, 121], making CDMA systems a cost effective choice for modern and proposed future wireless networks [1-4]. Other advantageous features include soft capacity limit, cell frequency reuse, soft handover of users etc. [57, 98, 121]. Wide bandwidth CDMA will be a dominant technology for third generation

(3G) and future wireless communication systems and form an integral part of the CDMA2000, 3rd Generation Partnership Project (3GPP) Universal Mobile Telecommunication System (UMTS) Frequency Domain Duplex (FDD, also called WCDMA) and Time Domain Duplex (TDD) standards [1-4, 143]. Note that the UMTS TDD standard uses multi-carrier CDMA (MC-CDMA) instead of DS-SS-SS-SS. While indoor, personal, ad-hoc and local non-cellular implementations of CDMA use one stage of signature codes for user identification [27, 149, 152], the modern 2.5G and beyond wireless voice-data-multimedia networks such as IS-95, CDMA2000, UMTS FDD (or WCDMA) and beyond use both user-specific spreading codes and cell-specific long scrambling codes to permit cost-effective cellular implementations [1-4].

Unlike the uplink communication channel, where all user codes are known and the base station (BS) possesses superior signal processing capabilities and resources, the downlink path is constrained by restricted computational resources and limited knowledge of the user equipment (UE)/mobile station (MS) [3, 121]. In a CDMA system, the symbol train to a user may be detected using either a single-user or a multi-user detector. A single-user detector (SUD) such as Matched Filter (MF) detector, Zero Forcing (ZF) detector [121], RAKE etc., does not model the multiple access interference (MAI) due to the presence of other users and extraneous signals in its signal path, rather it treats all interfering users and disturbances as additive noise. This severely limits its detection performance and it fails to provide the performance levels vital to modern high data rate applications [4]. In contrast, a multi-user detector (MUD) [58, 120, 121, 146] includes all the users in the signal model. Significant improvement can be obtained with a multi-user receiver [121, 146]. However the optimal MUD [120] is computationally

intensive and requires several dynamic system parameters to be known precisely. Several linear multi-user detection (MUD) techniques such as Best linear unbiased estimator (BLUE) [146], linear minimum mean squared error (LMMSE) estimator [44, 59, 146] have been proposed for the wireless downlink based on the linear convolutive channel model. In practical situations most of these estimators require extensive knowledge of channel parameters and massive computations. Even if approximated using block level implementations, the system performance becomes sub-optimal, which further degrades quickly if the a-priori channel estimates are far from true system parameters. Adaptive approximations of these algorithms also require some a-priori information for correct initialization and may not match the performance of the algebraic counterparts under all conditions.

For the proposed 3GPP UMTS FDD [2, 3] and future tightly synchronous CDMA standards, the user specific spreading is done by Orthogonal Variable Spreading Factor (OVSF) codes, while the cell specific complex scrambling is done using a set of long gold sequences (and/or Kasami sequences) [66]. Long codes have their own advantages such as *code-hopping* which results in similar performance for all users in the system (i.e., better QoS). In addition long codes also result in improved power control for users with relatively smaller data rates in the system and better rejection of extra-cellular interference (i.e., lower effective BER) [58]. Contrarily, the use of long scrambling codes makes the implementation of the exact linear detection algorithms and BMUD impractical due to the excessive computational requirements [44, 146]. Further, the actual MUD model also does not capture effects such as inter-cellular interference of signals and interference from other services sharing the same temporal and spectral resources.

Even the performance of block implementations of linear detectors such as block LMMSE (B-LMMSE) is severely degraded due to imprecise channel estimation (and/or inaccurate estimate of the phase reference for the principal path), if possible at all [59, 121, 122].

The restrictive downlink performance of the conventional detectors in practical communication systems and the higher performance requirements of the modern multimedia and other dedicated user services, necessitate better detection schemes for the downlink CDMA channel.

6.3 Downlink Signal Models for CDMA Systems

In this section, the signal model for a generic DS-CDMA implementation using one layer of spreading codes only is first discussed. In the subsequent sub-section, this model is extended for the modern wireless networks such as IS-95, CDMA2000, UMTS etc. These modern CDMA implementations utilize another stage of codes for cell-specific scrambling [1-4].

6.3.1 DS-CDMA Signal Model

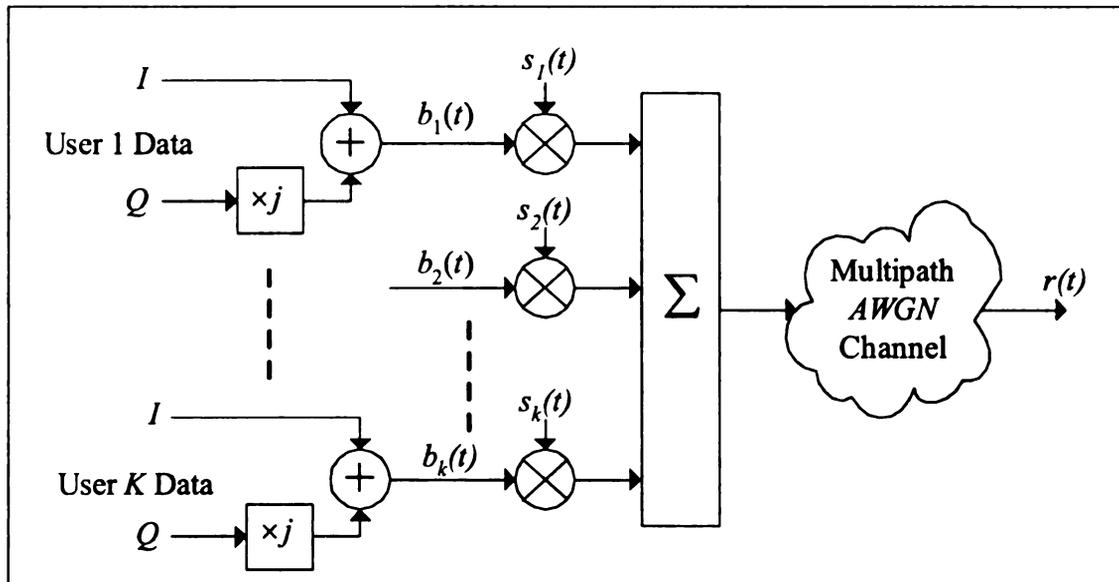


Figure 6.1. Signal Generation Model for a QPSK DS-CDMA system

In a typical downlink synchronous DS-CDMA system employed for indoor ATM and certain ad-hoc wireless networks [149, 152], each user's data is spread using an individual signature waveform (or spreading code), then the data for all users is combined and transmitted by the Base Station (BS). Each User Equipment (UE) or Mobile station (MS) synchronizes itself with the BS using the broadcast synchronization/pilot channels, once synchronized the BS and UE/MS can communicate on the traffic channel (comprising of both data and control streams). Assuming the data transmission to be QPSK, i.e., comprising of two composite data channels created by a serial-to-parallel (S/P) stage, which are constellated in quadrature. At the UE/MS receiver, the received signal is first passed through a chip-matched filter and sampled at chip rate.

Considering a total of K active users in an L multipath environment and N transmitted symbols during the observation frame T_F , the received signal is given by [82, 122, 124]

$$r(t) = \sum_{n=1}^N \sum_{k=1}^K \sum_{l=0}^{L-1} \sqrt{\varepsilon_{kn}(t)} b_k(n) h_l(t) s_k(t - nT - \tau_l) + n(t) \quad (6.1)$$

where ε_{kn} is the energy of the n^{th} symbol for the k^{th} user, $b_k(n) \in \{\pm 1 \pm i\}$ is the n^{th} complex symbol for the k^{th} user, h_l and τ_l are the l^{th} path's gain co-efficients and delay, respectively. $n(t)$ is the additive noise and $s_k(t)$ is the k^{th} user's signature code (or spreading sequence) generated by

$$s_k(t) = \sum_{m=0}^{G-1} \alpha_k(m) p(t - mT_c); \quad \alpha_k(m) \in \{-1, 1\}; \quad 0 \leq m \leq G-1 \quad (6.2)$$

$\alpha_k(m)$ is a real spreading sequence (i.e., any of the standard CDMA PN codes, such as the Gold, Walsh-Hadamard, Kasami etc. [66]) sequence for the k^{th} user containing G chips per symbol, i.e., $G = T_b/T_c$. $p(t)$ is a chipping pulse of duration T_c . T_b being the symbol period.

Under the assumption of time-invariance, the model (6.1) can be more compactly written in a vector-matrix format as

$$\bar{r} = H S \bar{b} + \bar{n} \quad (6.3)$$

where, H is a $(NG + L - 1) \times NG$ multipath propagation co-efficient matrix containing the channel coefficients. S is a $NG \times NG$ block diagonal matrix with the matrix of spreading codes forming the diagonal elements, \bar{b} is an NG -d vector containing the user symbols, while \bar{n} is the $(NG + L - 1)$ -d channel noise vector with covariance matrix Q .

The structure of the above defined matrices and vectors is given by

$$H = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ h_{L-1} & & \ddots & 0 \\ 0 & \ddots & & h_0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{L-1} \end{bmatrix}$$

$$S = \text{diag}[\bar{S} \ \bar{S} \ \cdots \ \bar{S}], \quad \bar{S} = [s_1 \ s_2 \ \cdots \ s_k]$$

$$\bar{b} = [b(1)^T \ b(2)^T \ \cdots \ b(N)^T]^T, \quad b(n) = [\sqrt{\varepsilon_{1n}}b_1(n) \ \sqrt{\varepsilon_{2n}}b_2(n) \ \cdots \ \sqrt{\varepsilon_{kn}}b_K(n)]^T$$

The compact linear model (6.3) is useful in deriving the linear detectors such as matched filter (MF), linear minimum mean squared error (LMMSE) etc. (see, section 6.4.1) for recovery of the transmitted symbol train for a desired user. However, the primary disadvantage of this model is the prohibitive dimensions of the constituent matrices, especially with longer frame durations and larger G , the matrices become excessively large, making this model unsuitable for any real-time implementation at UE/MS.

Alternately, the signal model can be represented as a linear convolutive model [57, 62, 122-124], i.e., during symbol time, the received chip data is constituted of the chips corresponding to the currently transmitted symbol and its delayed multipath components as well as delayed chips from some previously transmitted symbols and the channel added noise. In this formulation, G chips arriving at the UE/MS during the n^{th} symbol time are computed as the sum of the chips from L multipaths of the n^{th} transmitted symbol and the multipath components of the previous $J-1$ symbols ($n-1, \dots, n-J-1$), where

$$J = \left\lceil \frac{\tau_L}{G} \right\rceil + 1 \quad (6.4)$$

τ_L being the maximum chip delay in L multipaths. The n^{th} received symbol can be expressed as

$$r_n(t) = \sum_{k=1}^K b_k(n) \sqrt{\varepsilon_{kn}(t)} \sum_{l=0}^{L-1} h_l(t) s_k(t - nT - \tau_l) + n_n(t) + \sum_{j=1}^{J-1} \sum_{k=1}^K b_k(n-j) \sqrt{\varepsilon_{k(n-j)}(t)} \sum_{l=0}^{L-1} h_l(t) s_k(t - (n-j)T - \tau_l); nT \leq t \leq (n+1)T \quad (6.5)$$

Under the assumption $L \leq G$, the above model can be expressed just in terms of the current and the preceding symbol. That is, the multipaths with delay greater than symbol period either do not exist or are weak enough to be ignored. In this case, the output samples of the chip-matched filter can be written as:

$$r_n = \sum_{k=1}^K \left[b_{kn} \sqrt{\varepsilon_{kn}(t)} \sum_{l=0}^{L-1} h_l \bar{z}_{kl} + b_{k,n-1} \sqrt{\varepsilon_{k,n-1}(t)} \sum_{l=0}^{L-1} h_l z_{kl} \right] + n_n \quad (6.6)$$

where, \bar{z}_{kl} and z_{kl} are G -d early and late code vectors, i.e.,

$$\bar{z}_{kl} = \left[\underbrace{0 \ \cdots \ 0}_{\tau_l} \ s_k[1] \ \cdots \ s_k[G - \tau_l] \right]^T \quad (6.7)$$

$$z_{kl} = \left[s_k[G - \tau_l + 1] \ \cdots \ s_k[G] \ \underbrace{0 \ \cdots \ 0}_{G - \tau_l} \right]^T \quad (6.8)$$

τ_l is the discretized delay satisfying the constraint $0 \leq \tau_l \leq T_b$. Imposing time invariant constraints, the multipath slowly fading environment model (6.6) can be represented in the form

$$r_n = H_0 b_n + H_1 b_{n-1} + n_n \quad (6.9)$$

where b_n and b_{n-1} are the K -d vectors of current and previous symbol for all the K users.

H_0 and H_1 are $G \times K$ mixing matrices with the structure

$$H_0 = [H_{0,0} \quad H_{0,1} \quad \cdots \quad H_{0,K}], \quad H_1 = [H_{1,0} \quad H_{1,1} \quad \cdots \quad H_{1,K}]$$

Such that

$$H_{0,k} = \sqrt{\varepsilon_0} \sum_{l=0}^{L-1} h_l \bar{z}_{kl} \quad (6.10)$$

$$H_{1,k} = \sqrt{\varepsilon_1} \sum_{l=0}^{L-1} h_l z_{kl} \quad (6.11)$$

and $\varepsilon_0, \varepsilon_1$ represent the energy of the current and the previous symbol respectively.

6.3.2 WCDMA (UMTS FDD) Signal Model

Consider the forward link (or downlink) for a modern wireless 3G UMTS FDD (best known as WCDMA) cellular mobile communications network [1-4]. For WCDMA, a user's signal stream is first spread by a user-specific OVSF sequence (of length 4-512 chips) followed by a cell-specific scrambling using long gold codes of length 38400 chips [3]. This arrangement allows for the signal streams to be distinguishable both at the cell (BS) and the user (UE/MS) level (see Figure 6.2) [3, 4]. Due to the presence of long scrambling codes used in the WCDMA downlink, the signal model for the mobile unit differs from the one discussed in the previous section. The intra-cell multiple user signals sharing the same multipath environment are considered to be the main cause of the Multiple Access Interference (MAI) in the received signal for a desired user [57, 121].

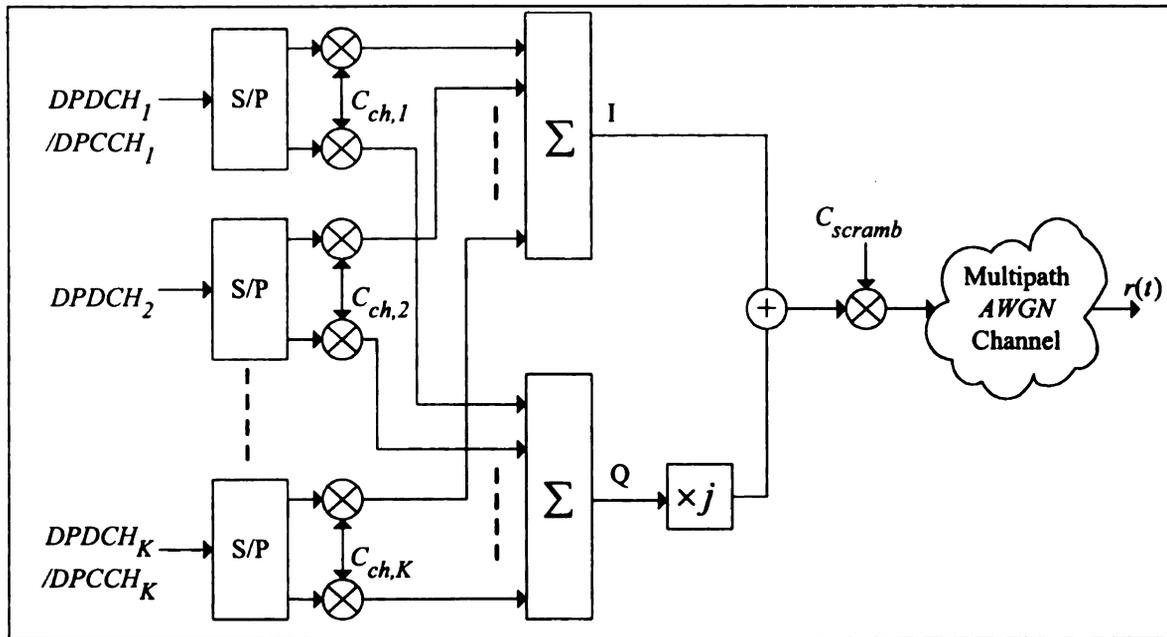
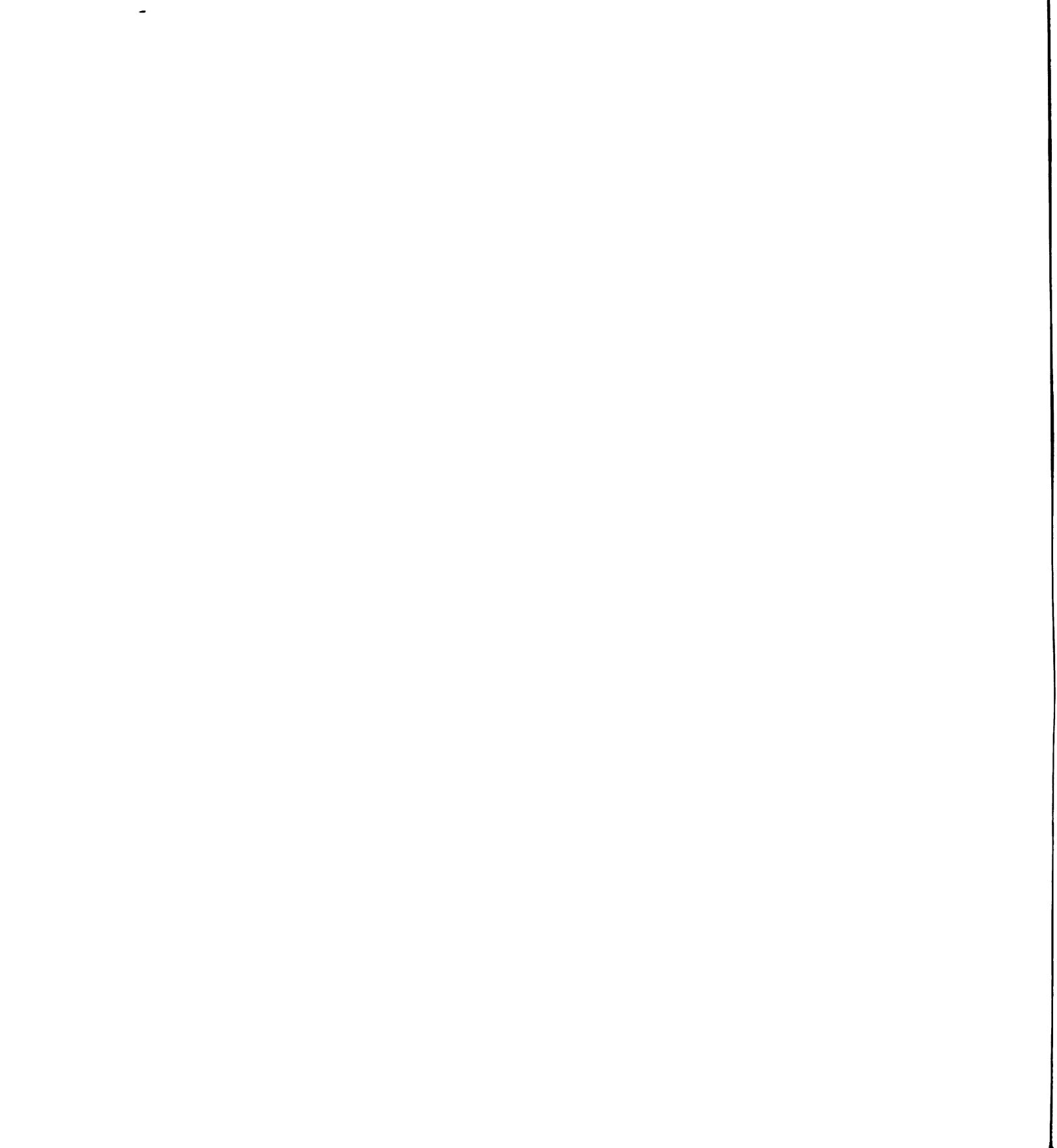


Figure 6.2. Signal generation based on the proposed 3GPP UMTS FDD standard

User specific data is delivered to the physical layer in blocks often termed as Transport Blocks (TB), which include bits for CRC, FEC etc. The transport channel multiplexer (or a parallel-to-serial converter P/S) converts all the data to a single $BPSK$ stream. After interleaving and addition of physical channel information for transmission power control (TPC), pilot, transport format combination indicator (TFCI) etc., the data stream comprises of two dedicated logical channels, one for data ($DPDCH$) and one for corresponding control ($DPCCH$) information, in multiplexed form they form one user specific dedicated physical data channel ($DPCH$) [2]. However, note that in case several data streams are associated to a single user, they can share the same logical control channel. Before transmission, the data goes through a serial-to-parallel (S/P) converter for QPSK modulation. Each user specific stream is first spread using a relatively short (4-512 chips) OVFS channelization code C_{ch} (or real user-specific spreading code, $s_k(t)$). The parallel channels are mapped to the I and the Q branches. At this stage the data



streams for all the users in a cell are summed together. This combined QPSK stream is subsequently scrambled by the long (38400 chips) scrambling code C_{scramb} (or cell-level complex gold code, $c(t)$). The collective data for all users is then modulated and synchronously transmitted via the transmission medium, which is assumed to be a wide sense stationary slowly fading multipath frequency selective channel. All physical channels transmitted from a BS are mutually orthogonal at the time of transmission as the spreading codes are the OVSF (or Walsh Hadamard codes), which are confined to the symbol period only. Different users can have different data-rates, which is accommodated by assigning different length OVSF spreading sequences $s_k(t)$ in a global prefix-free code assignment tree [3]. The complex gold scrambling code $c(t)$ has a length of 38400 chips, which lasts for a 10 ms frame [3, 66]. The downlink system is designed to be tightly synchronous at the frame level [2].

The received signal at the mobile receiver within a transmitted frame of length T_F can be represented by

$$r(t) = \sum_{n=1}^N \sum_{k=1}^K \sum_{l=0}^{L-1} b_k(n) \sqrt{\varepsilon_{kn}(t)} h_l(t) c(t - \tau_l) s_k(t - nT - \tau_l) + n(t) \quad (6.12)$$

where $c(t) \in \{\pm 1 \pm i\}$ is the complex cell-specific scrambling sequence. Other variables are defined in a fashion similar to the model (6.1).

The received signal at UE/MS is passed through a chip-matched filter and sampled at chip rate. The received vector \bar{r} in this case can be expressed as [122, 125]

$$\bar{r} = HCS\bar{b} + \bar{n} \quad (6.13)$$

where H , S , \bar{b} and \bar{n} are defined in a fashion similar to (6.3). C is the $NG \times NG$ complex diagonal scrambling matrix with $CC^H = 2 \times I_{NG \times NG}$. N being the number of

symbols within a frame time T_F . C is defined as: $C = \text{diag}[c_1 \ c_2 \ \cdots \ c_{NG}]$;

$$c_i \in \{\pm 1 \pm i\}; 1 \leq i \leq NG$$

Note that in a physical cellular environment, the unmodeled channel noise has a more intricate structure than simple additive white Gaussian noise. The additive noise may include inter-cellular signals especially near cell boundaries or in congested metropolitan areas, and cross-interference with other services sharing the frequency spectrum [57].

The size of the above model is excessively large due to the dimension of the long scrambling code. Even the convolutive model, in this case, is not practical due to the large dimension of the underlying code structure. A number of approaches will be discussed in the subsequent section to recover the desired user's symbol train at UE/MS while keeping in mind the constraints of performance and ease of implementation.

6.4 CDMA User Detection Structures

In this section, the linear detection schemes such as the matched filter (MF), multipath matched filter or RAKE receiver and the linear minimum mean squared error (LMMSE) detector are described. Other linear detector formulations such as the least squares (LS) or the zero forcing (ZF) detector, best linear unbiased estimator (BLUE) detector are not discussed as their performance is worse as compared to the LMMSE detector. For more details on these detectors, please see [57, 84, 121, 146].

6.4.1 Conventional Detection Schemes

In this section, we describe three conventional user detection schemes dominantly used in CDMA communication systems. These schemes are the matched filter (MF), RAKE

detector and the linear minimum mean squared error (LMMSE) detector. Also presented is a practical technique for CDMA channel estimation [57].

6.4.1.1 Matched Filter (MF) Detector

The standard matched filter is a single user detector which just utilizes the user's own signature sequence to make the best possible estimate of user's transmitted sequence from the raw chip data received at the UE/MS. The detection algorithm completely ignores the presence of convolutions due to multipaths in the in the receiver environment as well as the MAI due to additional users sharing the resources. For the DS-CDMA system (6.3), MF detector for the i^{th} user becomes

$$\hat{b}_{i,MF} = S_i^H \bar{r} \quad (6.14)$$

where $S_i = \text{diag}[\bar{S}_i \quad \bar{S}_i \quad \dots \quad \bar{S}_i]$, and $\bar{S}_i = [0 \quad \dots \quad s_i \quad \dots \quad 0]$. s_i being the i^{th} user's signature code, \bar{r} is the received chip train as defined in (6.3).

In case of WCDMA system, the user's symbol undergoes spreading as well as scrambling, therefore the MF detector becomes

$$\hat{b}_{i,MF} = S_i^H C^H \bar{r} \quad (6.15)$$

where $C = \text{diag}[c_1 \quad c_2 \quad \dots \quad c_{NG}]$; $c_i \in \{\pm 1 \pm i\}$; $1 \leq i \leq NG$ is the scrambling code matrix.

Practically, MF can be implemented as

$$\hat{b}_{i,MF}(n) = s_i^H \tilde{r}(n) \quad (6.16)$$

where

$$\tilde{r}(n) = \begin{cases} \bar{r}((n-1)G+1 : nG) & \text{for DS-CDMA} \\ C^H(n) \bar{r}((n-1)G+1 : nG) & \text{for WCDMA} \end{cases}$$

and $C^H(n) = \left\{ \text{diag} \left[c_{(n-1)G+1} \quad \cdots \quad c_{nG} \right] \right\}^H$

6.4.1.2 RAKE Detector

The RAKE detector in CDMA is created using multiple chip-delayed MF detector fingers in parallel. The RAKE receiver first identifies three or more (depending on the number of fingers) strongest multipath signals arriving at the receiving antenna using a maximal ratio path search algorithm [57, 121]. These multipath signals are combined after adjustment for their corresponding delays (and possibly phase and path attenuation if channel estimation is available) to produce a relatively stronger received signal chip stream. In this work, the RAKE receiver is assumed to be implemented with the knowledge of both channel multipath delays (maximal ratio path search) as well as the corresponding multipath attenuation co-efficients (channel estimation). The RAKE detector for the DS-CDMA system therefore can be mathematically expressed as

$$\hat{b}_{i,RAKE} = S_i^H \hat{H}^H \bar{r} \quad (6.17)$$

where \hat{H}^H is the hermitian of the estimated channel matrix, other variables are as defined earlier for MF detector. Similarly for the WCDMA system, RAKE detectors is given by

$$\hat{b}_{i,RAKE} = S_i^H C^H \hat{H}^H \bar{r} \quad (6.18)$$

Similar to (6.16), once the channel estimate is known, RAKE can also be efficiently implemented as

$$\hat{b}_{i,RAKE}(n) = s_i^H \tilde{r}(n) \quad (6.19)$$

where

$$\tilde{r}(n) = \begin{cases} \hat{H}^H(n) \bar{r}((n-1)G+1 : nG) & \text{for DS-CDMA} \\ C^H(n) \hat{H}^H(n) \bar{r}((n-1)G+1 : nG) & \text{for WCDMA} \end{cases},$$

$C^H(n)$ is as defined above, and $\hat{H}^H(n)$ is the hermitian of the channel estimate MA filter.

6.4.1.3 Linear Minimum Mean Squared Error (LMMSE) Detector.

The optimal LMMSE detector is considered to be the best linear detector for DS-CDMA reception. While other linear detectors such as LS, ZF, BLUE etc. do not provide good performance in the presence of excessive noise (especially colored noise). LMMSE detectors do a trade-off by not performing perfect orthogonalization of received signal stream at low SNRs by trying to minimize additive noise variance.

The LMMSE detector for single stage spreading DS-CDMA systems is given by

$$\hat{b}_{i,LMMSE} = S_i^H \hat{H}^H (\sigma^2 \hat{H} \hat{H}^H + \hat{Q})^{-1} \bar{r} \quad (6.20)$$

where $R = E[\bar{r} \bar{r}^H] = (\sigma^2 H H^H + Q)$ is the auto-correlation matrix for the signal chip train received at the UE/MS, σ^2 being the average power of transmitted user signals.

For the WCDMA system, the symbol recovery needs to be descrambled before despreading, and the LMMSE detector becomes

$$\hat{b}_{i,LMMSE} = S_i^H C^H \hat{H}^H (\sigma_x^2 \hat{H} \hat{H}^H + \hat{Q})^{-1} \bar{r} \quad (6.21)$$

There are several difficulties in the implementation of the LMMSE receiver. Practically, the dimension of H especially in case of WCDMA is impractical. Typically LMMSE is implemented on mG -sized blocks of received data and is called block LMMSE (B-LMMSE).

Observe that the terms on the left side of matrix inverse in (6.20) and (6.21) are similar to (6.17) and (6.18), respectively. Therefore LMMSE detector can typically be implemented by adding a pre-processing stage before RAKE receiver. The data

correlation matrix may be adaptively estimated resulting in adaptive LMMSE implementations [44, 84, 96, 146].

For DS-CDMA systems, since the number of users K is typically smaller than the spreading gain G , the user data occupies a sub-space only. In such cases, instead of inverting the autocorrelation matrix R , the data-dimensions may be reduced without any significant performance degradation [142] using eigen decomposition, i.e.

$$\hat{b}_{i,LMMSE} = S_i^H C^H \hat{H}^H (\hat{V}_s \hat{D}_s^{-1} \hat{V}_s^H) \bar{r} \quad (6.22)$$

where \hat{D}_s and \hat{V}_s are the estimates of the K dominant eigen-values and eigen-vectors of the auto-correlation matrix R . These eigen values and vectors can also be computed adaptively, see [52, 62].

6.4.1.4 Channel Estimation Using The Common Pilot Channel (CPICH)

The Common Pilot Channel (*CPICH*) is a fixed rate (30 kbps, $G=256$) downlink physical channel that carries a pre-defined bit sequence [1-4]. Each cell has a Primary Common Pilot Channel (*P-CPICH*) and possibly one or more Secondary Common Pilot Channels (*S-CPICH*). What configuration of these pilot channels is available and whether they can provide phase reference and possibly channel estimation for the dedicated data channels *DPCH* is communicated to the UE/MS by higher-level signaling. It is possible in a UMTS implementation that neither *the P-CPICH* nor any *S-CPICH* is a phase reference for any downlink *DPCH* [4].

If applicable, *CPICH* allows the UE/MS to equalize the channel in order to achieve a phase reference with the *SCH* (Synchronization Channel). *CPICH* also allows active estimation for power control. Primary *CPICH* always uses the same channelization code.

6.4.1.4.1 Channel Estimation Procedure

In order to outline a simple and practical channel estimation technique using *CPICH* [1, 4], we assume that the transmitted pilot comprises of a stream of a single symbol train of $1 - i$, and uses an all ones spreading code (i.e., the first OVSF code)

1. For each independent multipath, multiply the incoming symbol train with the corresponding delayed scrambling sequence.
2. Remove the modulation from CPICH by simply multiplying the CPICH data by its conjugate, i.e., $1 + i$. The resulting channel estimate is noisy because of AWGN and multiple-access interference.
3. Pass the noisy channel estimate through a smoothing filter to achieve better noise immunity. This filter can be either a moving average window of length $2M+1$, e.g.,

$$\hat{h}_i = \frac{1}{2M+1} \sum_{k=-M}^M p_{ik} \quad (6.23)$$

Since, a moving UE/MS represents a dynamic channel, the channel estimates may be processed using a relatively smaller factor M followed by a single-zero smoothing filter

$$\bar{h}_i = (1 - \rho)\bar{h}_i + \rho\hat{h}_i; \text{ where } 0 \leq \rho \leq 1 \quad (6.24)$$

4. Decimate or interpolate the filtered channel estimate obtained in step 2 to match the data rate of the CPICH to the data rate of the DPCCH/DPDCH. This simple technique works well in most cases because the channel is assumed to be stable for the symbol duration.

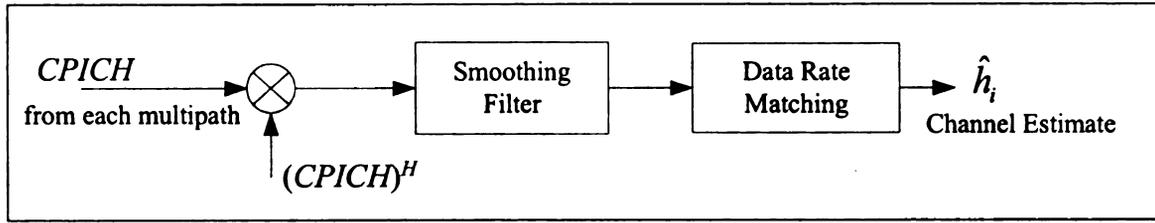


Figure 6.3. Channel Estimation Using CPICH

This simple technique works well in most cases. The channel is typically assumed to be stable for symbol duration.

6.4.2 Proposed Blind Source Recovery Detection Schemes

In this section, three new MUD detection algorithms using BSR techniques are presented.

6.4.2.1 Natural Gradient Blind Multi-User Detection (BMUD) Algorithms.

Blind Multi-user Detection (BMUD) is the process to blindly estimate all the user symbol sequences directly from the received composite CDMA signal using the Blind Source Recovery (BSR) techniques. BSR framework implies recovery of original signals from environments that may include transient, convolution and even non-linearity [109, 110, 112, 126]. The linear BSR algorithms [110, 112, 129] have been developed for linear convolutive mixing environments by the minimization of mutual information (Kullback Lieblar Divergence) using the natural gradient subject to the structural constraints of a recovery network. The natural gradient BMUD network can be either in the feedforward or the feedback configuration [112]. The proposed BMUD algorithm adaptively estimates a set of matrices to counter the linear convolutive environment model (6.9).

The justification for BMUD algorithms is based on the convenient convolutive signal model representation of DS-CDMA systems, see (6.9), and the reasonable

assumption that the various transmitted user symbol sequences are mutually “independent” as they are generated independently of each other. In this framework, both the transmitted sequence and the mixing matrices in the model (6.9) are unknown to the user. The only known entity to the user is the self-identification code. Other available prior information is the nature of transmitted data, which is assumed to be quaternary sub-gaussian distribution, e.g., QPSK data distorted by the multipath channel and additive noise. There exists enough information to apply the info-theoretic natural gradient Blind Source Recovery (BSR) algorithms for BMUD in this case [55, 62, 101, 123, 124].

Further assume that the DS-CDMA channel is not over-saturated and $K \leq G$. The proposed BMUD algorithms do not require any pre-whitening of received data. However, in most modern WCDMA, G is chosen to be very large and in general $K < G$. Therefore, it is computationally advantageous to pre-process the data for dimension reduction to K which is the actual number of principal independent symbol sequences (or users) in the received data. The process of pre-whitening will also remove the second order dependence among the received data samples and some of the additive noise [33, 39, 62, 101, 123-125]. The data pre-whitening can be achieved either online using adaptive principal component analysis (PCA) algorithms or it may be done algebraically [33, 52, 60, 98].

The pre-whitening matrix can be computed as

$$W = D_s^{-1/2} V_s^H \quad (6.25)$$

where D_s and V_s are as defined in section 6.4.1.3.

The whitened version of (6.9) is given by

$$r_n^w = W(H_0 b_n + H_1 b_{n-1} + n_n) \cong \bar{H}_0 b_n + \bar{H}_1 b_{n-1} \quad (6.26)$$

where

r_n^w represents the G -d whitened data received during the n^{th} symbol time and \bar{H}_0 , \bar{H}_1 represent the equivalent $G \times K$ convolutive mixing matrices for the current and the previous symbol.

BMUD algorithms blindly adapt a set of matrices to estimate the independent user symbols y_n at the n^{th} instant. This is followed by a decision stage to interpret, as best as possible, y_n and estimate the corresponding user symbol estimates \hat{b}_n also at the n^{th} instant.

$$\hat{b}_n = \psi(y_n) \quad (6.27)$$

where $\psi(\cdot)$: represents the (nonlinear) decision stage.

The update laws for both feedforward and feedback structures [110, 112] have been presented. In the next section, the performance of the proposed algorithms is discussed and compared with conventional user detection algorithms.

6.4.2.1.1 **Feedforward BMUD Configuration:**

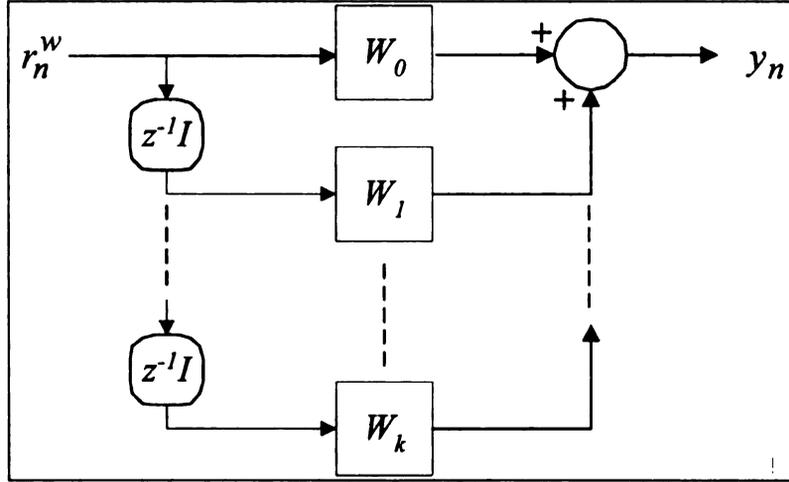


Figure 6.4. Feedforward Demixing Structure

For the feedforward configuration, the BMUD stage output is computed as

$$y_n = W_0 r_n^w + \sum_{k=1}^K W_k r_{n-k}^w \quad (6.28)$$

The update laws for this feedforward structure can be easily derived, using the techniques outlined in Chapter 3. For the feedforward parametric matrices W_0 and W_k , the update laws are

$$\Delta W_0 \propto (I - \varphi(y_n) y_n^H) W_0 \quad (6.29)$$

and

$$\Delta W_k \propto (I - \varphi(y_n) y_n^H) W_k - \varphi(y_n) (r_{n-k}^w)^H \quad (6.30)$$

where $\varphi(\cdot)$ is an element-wise acting score function [12, 27, 36, 37, 60, 114, 127, 130, 136], and I is a K -d identity matrix.

For the initialization of the algorithm, W_0 is chosen to be either an identity or a diagonally dominantly matrix, while all other matrices W_k are initialized to have either random elements with a very small variance or as a matrix of all zeros [110, 112, 155]. Note that no matrix inversion is required for the feedforward algorithm.

6.4.2.1.2 Feedback BMUD Configuration I:

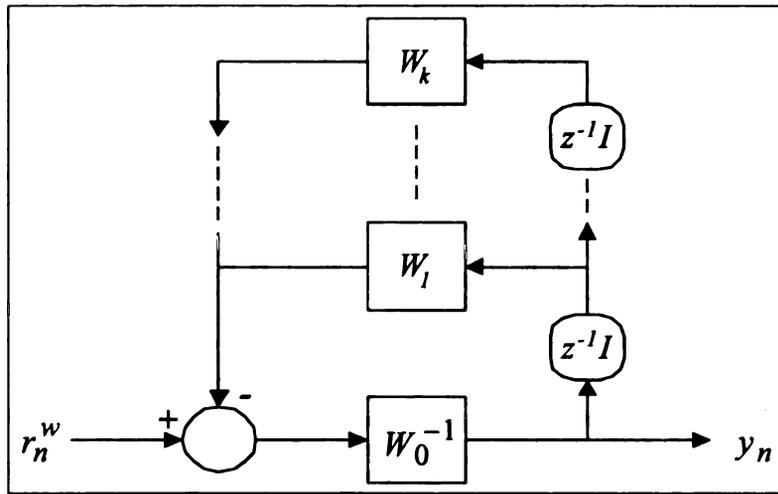


Figure 6.5. Feedback Demixing Structure I

In the feedback configuration I, the output is estimated by

$$y_n = W_0^{-1} \left(r_n^w - \sum_{k=1}^K W_k y_{n-k} \right) \quad (6.31)$$

The update laws for this structure using the natural gradient have been derived in [112]. The update law for the matrix W_0 is given by

$$\Delta W_0 \propto -W_0 \left(I - \varphi(y_n) y_n^H \right) \quad (6.32)$$

While for the feedback matrices W_k , the update law is

$$\Delta W_k \propto W_0 \left(\varphi(y_n) y_{n-k}^H \right) \quad (6.33)$$

The matrices in this case are also initialized in a fashion similar to the feedforward case. However, note that at least one matrix inversion is required in this formulation.

6.4.2.1.3 Feedback BMUD Configuration II:

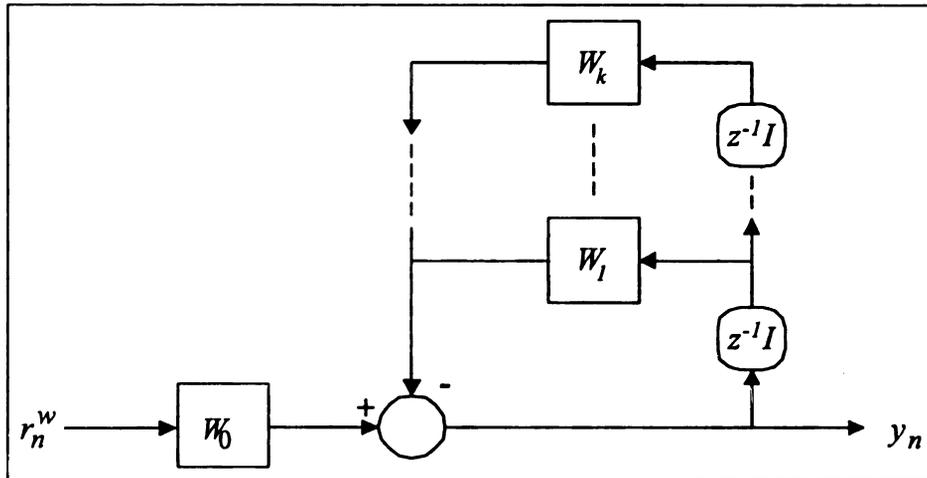


Figure 6.6. Feedforward Demixing Structure II

An alternate feedback configuration, see Waheed et al. {Waheed, 2003 #630}, implements the feedback structure without the need for any matrix inversion. For this feedback configuration II, the BMUD stage output is computed as

$$y_n = W_0 r_n^w - \sum_{k=1}^K W_k y_{n-k} \quad (6.34)$$

The update laws for this feedback structure have been derived in [110, 112, 155] and are given by

$$\Delta W_0 \propto \left(I - \varphi(y_n) y_n^H \right) W_0 \quad (6.35)$$

$$\Delta W_k \propto \left(I - \varphi(y_n) y_n^H \right) W_k + \varphi(y_n) y_{n-k}^H \quad (6.36)$$

In case the channel can be estimated [57], the performance of the proposed BMUD algorithms may be adjudged by the diagonalization of the absolute value of the global transfer function. The global transfer function presents the combined effect of the complex mixing and demixing transfer functions. For the two symbol convolutive models for the case of $L \leq G$, the global transfer function for the natural gradient algorithms in the z-domain are given by [123-125]:

$$G = G_0 + G_1 z^{-1} + G_2 z^{-2} \quad (6.37)$$

where, for the *feedforward* configuration

$$\begin{aligned} G_0 &= W_0 \bar{H}_0 = W_0 W H_0, \\ G_1 &= W_0 \bar{H}_1 + W_1 \bar{H}_0 = W_0 W H_1 + W_1 W H_0 \text{ and} \\ G_2 &= W_1 \bar{H}_1 = W_1 W H_1 \end{aligned} \quad (6.38)$$

while, for the *feedback* configuration I

$$\begin{aligned} G_0 &= W_0^{-1} \bar{H}_0 = W_0^{-1} W H_0, \\ G_1 &= W_0^{-1} (\bar{H}_1 - W_1) = W_0^{-1} (W H_1 - W_1) \text{ and} \\ G_2 &= 0 \end{aligned} \quad (6.39)$$

and for the for the *feedback* configuration II

$$\begin{aligned} G_0 &= W_0 \bar{H}_0 = W_0 W H_0, \\ G_1 &= W_0 \bar{H}_1 + W_1 = W_0 W H_1 + W_1 \text{ and} \\ G_2 &= 0 \end{aligned} \quad (6.40)$$

The proposed BMUD algorithms as formulated result in recovery of the user symbols directly. The algorithms can be conveniently applied to DS-CDMA systems

using only user-specific spreading sequences. They may also be extended to CDMA systems using relatively short scrambling codes, though the dimension of matrices may still become large for DSP implementations in a UE/MS. The WCDMA system uses long codes in the downlink, making the application of these algorithms impractical because of the requisite dimension of the demixing network matrices.

6.4.2.2 RAKE-Blind Source Recovery (RAKE-BSR) and RAKE-Principal Component Analysis (RAKE-PCA) Detectors

RAKE-BSR and RAKE-PCA are two new proposed adaptive detectors [122, 125], which utilize the same knowledge as a RAKE receiver. An info-theoretic adaptive weighting matrix of dimension $G \times G$ is introduced into the RAKE structure, which gives a big performance boost to the RAKE receiver. The performance of RAKE-BSR/RAKE-PCA exceeds the performance of LMMSE detectors under the conditions of high network congestion, imprecise channel estimation, and unmodeled inter-cellular interference etc [57, 122, 125]. The closed form structure of these proposed detectors is given by [122]

$$\hat{b}_{i,RAKE-ICA/PCA} = \begin{cases} S_i^H \tilde{W} \hat{H}^H \bar{r} & \text{for DS-CDMA Systems} \\ S_i^H C^H \tilde{W} \hat{H}^H \bar{r} & \text{for WCDMA Systems} \end{cases} \quad (6.41)$$

where $\tilde{W} = \text{diag}[A \ A \ \dots \ A]$, and A is the $G \times G$ matrix that is adaptively estimated either using static BSR (ICA) or PCA algorithms.

It is proposed to adapt the matrix A using the natural gradient update laws [7, 9, 122]. However, there exist several other methods for ICA/PCA and any other suitable method may be used for these adaptations [21, 60, 77]. This blind adaptation of the A matrix has several advantages and improves the performance of the overall equalization process in several ways. Firstly, it can counter artifacts in the estimated channel co-

efficients \hat{H} . Secondly, the channel estimation process (as in RAKE receivers) may be limited by the structure (such as number of fingers) and may estimate only a few of the dominant channel parameters. \tilde{W} stage tends to counteract this anomaly, as best as possible, and provides better performance than LMMSE in such cases [122]. Thirdly, this adaptive stage minimizes the effect of the additive channel noise, which may have an unmodeled intricate underlying structure. Lastly, the natural gradient ICA /PCA algorithms inherently reduce near-far problems by removing any ill conditioning in the signal space for all the users in the system. This results in all the mobile users in the system to have a BER performance similar to the average BER performance of the downlink channel [57, 122]. The matrix A is adaptively estimated using the update laws

$$A(k+1) = A(k) + \eta_k \Delta A(k) \quad (6.42)$$

where

$$\Delta A(k) = \begin{cases} (I - \varphi(y(k))y(k)^H)A(k) & \text{for static BSR (or ICA)} \\ (I - y(k)y(k)^H)A(k) & \text{for PCA} \end{cases} \quad (6.43)$$

and $\varphi(\cdot)$ is a nonlinear score function [7, 11, 12, 28, 36, 126, 130, 138] which depends on the underlying distribution structure of the signals involved. For QPSK signal, a suitable score-function [136] is

$$\varphi_i(y_i) = \nu_i y_i - \alpha_i \left(\tanh(\beta_i \operatorname{Re}\{y_i\}) + \tanh(\beta_i \operatorname{Im}\{y_i\}) \right) \quad (6.44)$$

Of these proposed RAKE-BSR/RAKE-PCA structures, RAKE-BSR exhibits relatively faster and more stable convergence [110, 155]. However, in CDMA systems the underlying code structure is chosen to be “orthogonal”, and RAKE-PCA may converge to a slightly lower BER solution if the channel impairments are linear in nature.

Note that in (6.41), if the channel estimate \hat{H} [57] is either not available or changes very

dynamically, the detector can be estimated without using the channel estimate and the structure reduces to Matched Filter BSR/PCA, i.e.,

$$\hat{b}_{i, MF-BSR/PCA} = \begin{cases} S_i^H \tilde{W}_F & \text{for DS-CDMA Systems} \\ S_i^H C^H \tilde{W}_F & \text{for WCDMA Systems} \end{cases} \quad (6.45)$$

The performance of this structure is better than MF alone, and approaches RAKE performance as the underlying matrix A converges. However, in this chapter, this structure will not be discussed any further.

6.5 Simulation Results

To verify the performance of the proposed BSR algorithms, a series of experiments for both DS-CDMA and WCDMA systems were conducted. Some selected results are being provided for the purpose of illustration. The proposed algorithms can be easily extended to multirate CDMA transmission. However, for a clear comparison of the proposed algorithms with the conventional approaches, the simulations have been restricted to the case where all the users have the same data rate. Assuming a constant spreading gain, $G = 64$. The channel is assumed to be a wide sense stationary slowly fading with several possible multiple paths to the UE/MS. The transmitted signal is also corrupted by additive white gaussian noise (AWGN) during the transmission. The simulated SNR range for all simulations is from -10 to 20 dB. The dominant multi-path delays and attenuation co-efficients for the signal received at the UE/MS are assumed fixed (i.e., the UE/MS is assumed to be static). Further all the multi-path attenuation co-efficients are assumed to be complex, i.e., each multipath applies both scaling and rotation to the propagated signal. For all the included simulations, the received CDMA signal comprises

of five multipaths with delays of 0, 1, 2, 3, 4 chips. The corresponding channel attenuation co-efficients are chosen to be.

$$h = [0.25 + 0.18i \quad 0.21 + 0.14i \quad 0.18 + 0.11i \quad 0.14 + 0.11i \quad 0.11 + 0.07i], \text{ respectively.}$$

The multipath channel estimate is either estimated online [57] or assumed to be known unless otherwise specified. The score function for both the BMUD and proposed RAKE-BSR algorithms is chosen to be $\varphi_i(y_i) = y_i - (\tanh(\text{Re}\{y_i\}) + \tanh(\text{Im}\{y_i\}))$. The UE/MS and BS are assumed to be in perfect synchronism. A decaying time-adaptive learning rate is chosen for all the adaptive algorithms. The B-LMMSE algorithm is applied on a block size of G chips, the auto-correlation matrix (for the B-LMMSE algorithm) is computed from the whole ensemble of the received data. The conjugate of the channel filter H , is applied recursively. These steps were done to ensure that the performance of the simulated B-LMMSE is not restricted due to implementation. In order to efficiently utilize the space, for each simulation, presentation is restricted to just two congestion scenarios of 20 users (approx. 30% congestion) and 50 users (approx 80% congestion).

The final symbol decision stage for all algorithms is given by

$$\psi(y_n) = \text{sign}(\text{Re}\{y_n\}) + \text{sign}(\text{Im}\{y_n\})i \quad (6.46)$$

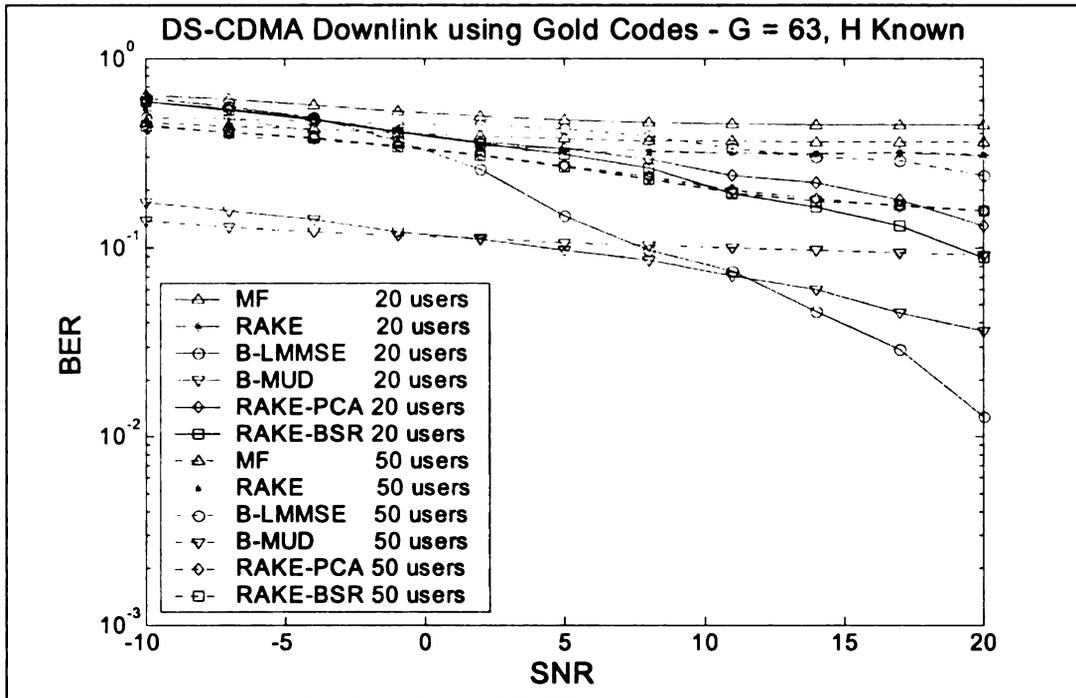
Any sign ambiguity in the recovered symbols is fixed using the pilot bits in each user's dedicated data channel.

6.5.1 Simulation I: DS-CDMA System

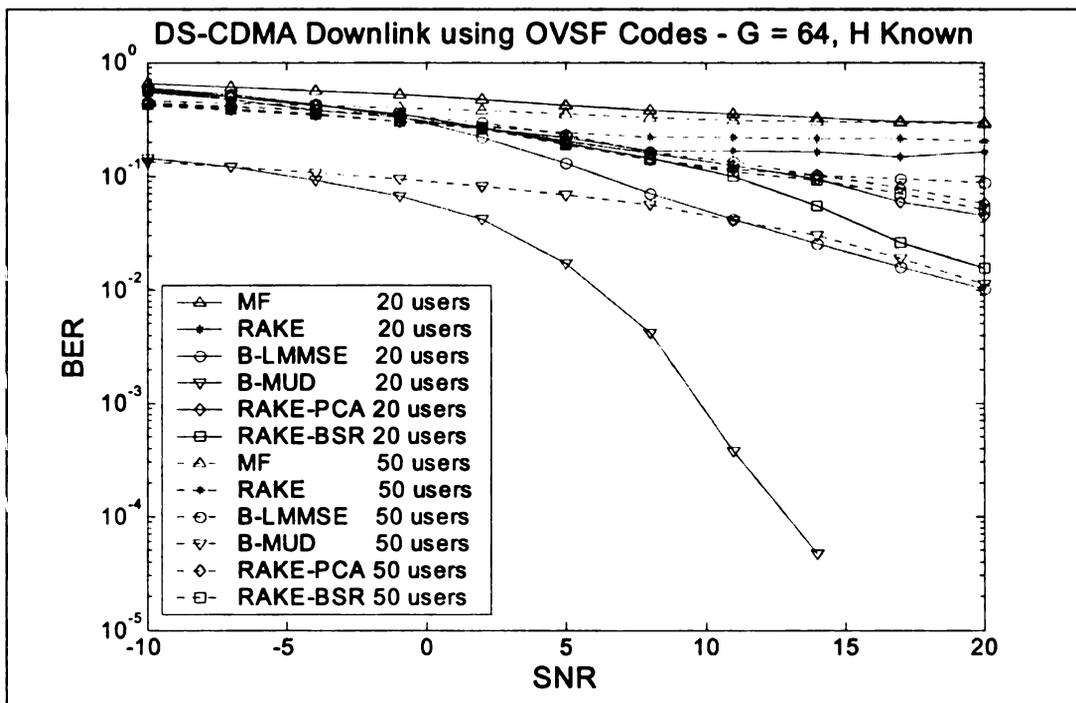
In this simulation, a DS-CDMA systems is simulated and modeled as in (6.9). All the 5 multi-path channel co-efficients and delays are assumed known. The system has been

simulated using two types of spreading codes, namely gold codes with a spreading gain $G = 63$, and OVSF (or Walsh Hadamard) codes with a spreading gain $G = 64$. The simulation results are presented below in Fig. 6.6.

It is evident from the presented results that the performance of proposed BMUD algorithm far exceeds any other user detection method. Under very good SNR conditions, the BER performance of B-LMMSE using gold codes improves over the proposed BMUD algorithm. This is because the channel estimate is assumed to be perfectly known for B-LMMSE and secondly, the BMUD algorithm is applied in instantaneous (on-line) mode with 5000 iterations only. The performance of the BMUD exceeds B-LMMSE if the algorithm is allowed to adapt longer.



(a)



(b)

Figure 6.7. DS-CDMA downlink performance for $K=20, 50$ users (a) using Gold Codes

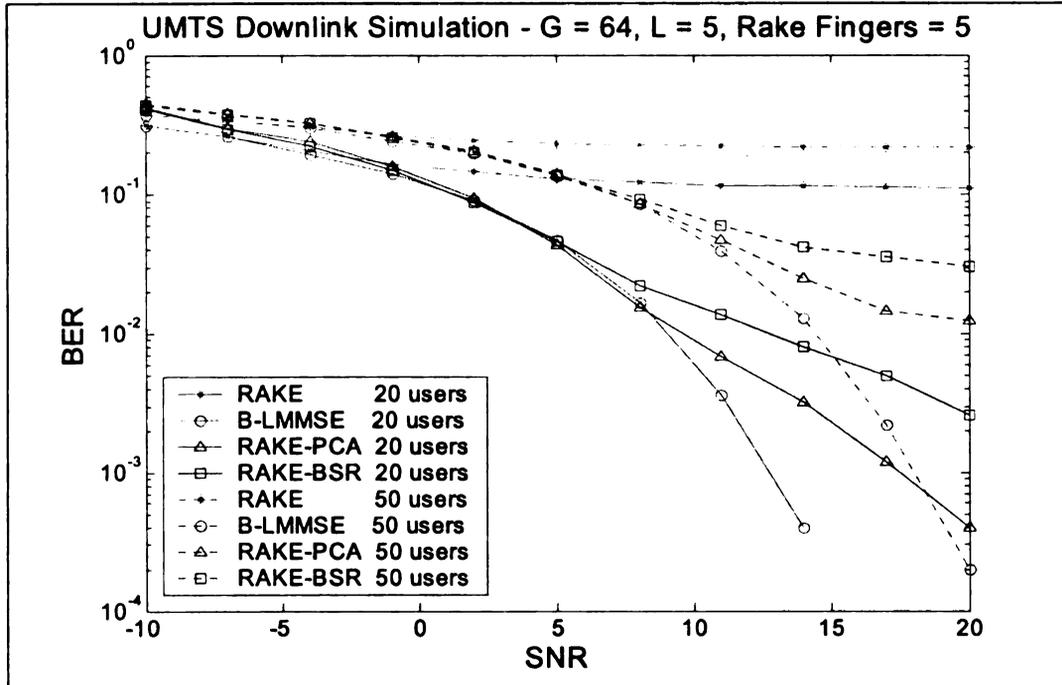
$G=63$, (b) using OVSF Codes $G=64$.

The performance of the proposed RAKE-PCA and RAKE-BSR algorithms is also better than B-LMMSE at lower SNRs and under congestion of the DS-CDMA system. However, for lower congestion of the system and high SNR, the B-LMMSE performance supercedes these algorithms if the channel estimate is perfect. In case, the channel estimate is imprecise [57], the proposed algorithms are always better than B-LMMSE.

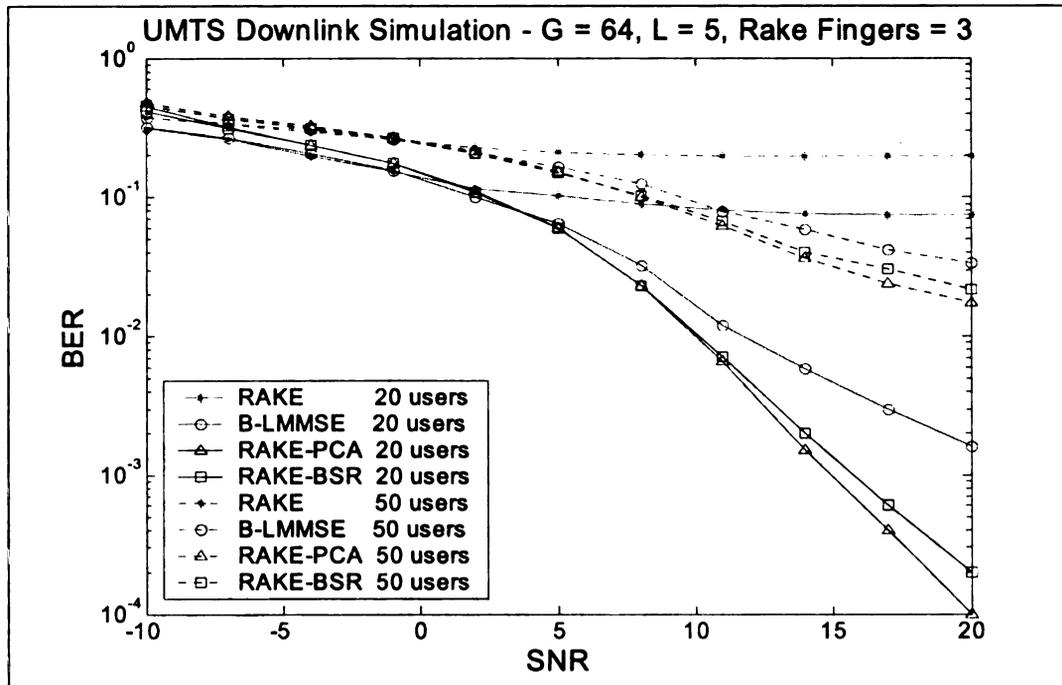
6.5.2 Simulation II: WCDMA (UMTS FDD) System

For the WCDMA Downlink simulation, 5 multipaths are assumed to exist similar to the previous simulation, all the user-specific codes are OVSF with a spreading gain G of 64. The long scrambling code has the frame-length of 38400 chips (10ms). A performance comparison of the proposed RAKE-BSR and RAKE-PCA algorithms with conventional RAKE, B-LMMSE under various network conditions is presented in Figures 6.7 and 6.8. In the first comparison (Fig. 6.7.a), the channel estimation is done for all the multipaths in the signal generation model. It is observed that the performance of the proposed algorithms RAKE-PCA and RAKE-BSR is very similar to B-LMMSE for lower SNR values. Under good SNR conditions, the performance of the B-LMMSE algorithm is better than the proposed algorithms.

However, note that for the B-LMMSE the auto-correlation matrix is estimated using 5000 symbols, while only 1000 instantaneous adaptations (for $K = 50$) are done for the proposed algorithms. In case more adaptations are done, the performance of proposed algorithms will approach the LMMSE limit, but even the performance attained with a limited number of iterations exhibits their effectiveness.



(a)



(b)

Figure 6.8. WCDMA Downlink System for $K=20, 50$ users (a) Performance Comparison with Perfect Channel Estimation, (b) Performance Comparison with Imperfect Channel Estimate

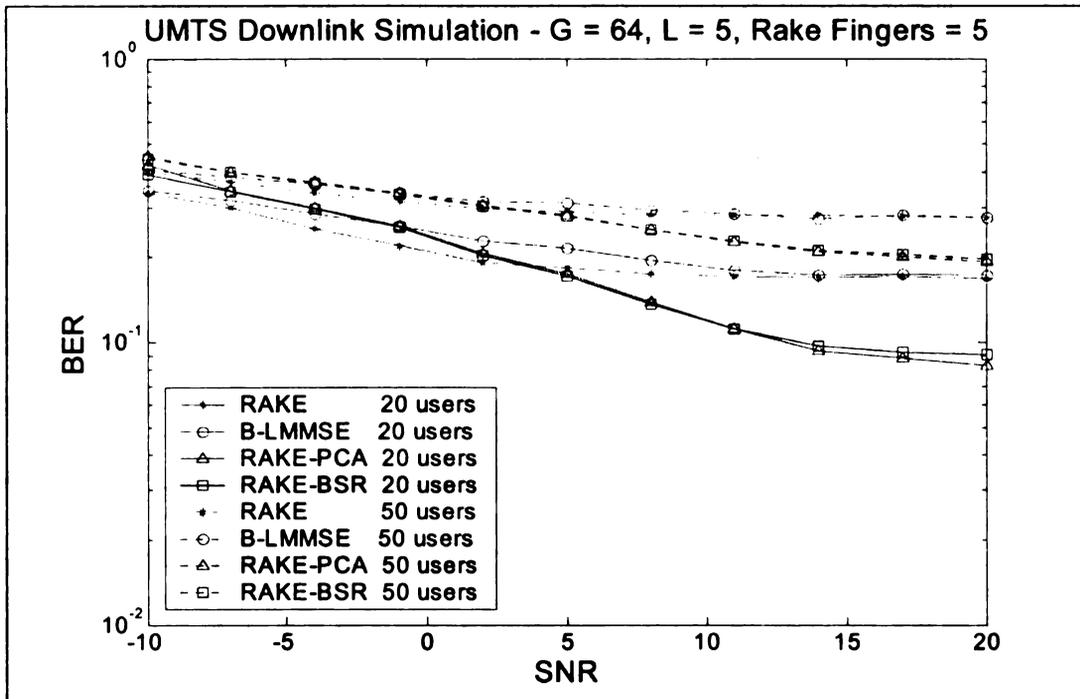


Figure 6.9. WCDMA Downlink System for K=20, 50 users: Performance Comparison with Imperfect Channel Estimate and Inter-Cellular Interference

In the second comparison (Figure 6.7.b), the channel estimate is restricted to the three dominant paths only, which correspond to the choice of delays 0, 1 and 2 in this case. In this case it is observed that the detection performance of the proposed RAKE-BSR and RAKE-PCA algorithms exceeds B-LMMSE at all SNRs. Comparing both RAKE-PCA and RAKE-BSR, it is observed that the performance of RAKE-PCA is approximately 1% better than RAKE-BSR. But RAKE-BSR has other advantages (as discussed in section 6.4.2.2) of stability and smaller energy of the computed filtering matrix \vec{W} . Another important observation in this case is that at very low SNRs, RAKE gives the best performance, but as the SNR improves the proposed algorithms can achieve very small BER. Therefore, a SNR based switching mechanism can be developed

to switch between the standard RAKE and the proposed algorithms, which just constitute an additional adaptive stage in the structure of the RAKE receiver.

In the third scenario (Figure 6.8), both the channel estimate is assumed to imprecise as in the previous comparison with only three dominant multipaths out of five estimated. In addition, the received signal is corrupted by extra-cellular signals of the neighboring cell BS. This is a realistic scenario in busy metropolitan areas, where there exist several dominant multipaths and the cell size is also kept relatively small to maximize the number of users per unit area. Inter-cellular interference is also critical when the UE/MS is on the cell boundary and undergoes a soft (or soft-softer) hand-over. The intercellular interference is unmodeled MAI and severely limits the performance of the detection algorithms. For the purpose of simulation this auxiliary BS interference is assumed to have half the energy of the primary BS. It is observed that the proposed RAKE-BSR/RAKE-PCA algorithms exhibit better immunity to this excessive unmodeled MAI and retain their qualitative performance advantage over the conventional algorithms.

6.6 Remarks

Three adaptive info-theoretic algorithms for CDMA systems have been proposed. Firstly a purely blind multi-user detection (BMUD) algorithm for the case of DS-CDMA systems, or WCDMA systems using short codes has been proposed. This algorithm gives better performance relative to all other user detection algorithms. This performance advantage further dominates conventional techniques under practical constraints of accuracy in on-line channel estimation and unmodeled environment MAI effects. Two other info-theoretic extensions have also been proposed on top of the standard RAKE detector namely RAKE-Blind Source Recovery (RAKE-BSR) and RAKE-Principal

Component Analysis (RAKE-PCA). These detection schemes add an adaptive information theoretic stage, based on higher-order statistics for RAKE-BSR and second-order statistics for RAKE-PCA, to the standard RAKE detector. This makes the resultant hybrid detector more robust to imperfections in channel estimates; unmodeled MAI and other slowly varying channel effects etc. Of these proposed algorithms, RAKE-BSR is more immune to synchronization errors between UE/MS and BS, possesses faster convergence and is better capable to cater for various dynamic and time-varying channel effects as compared to RAKE-PCA. However, for the presented WCDMA results RAKE-PCA demonstrates a slight performance advantage. This is due to the fact that the simulated channel imperfections are only linear, the channel synchronization is perfect and the underlying code structure is orthogonal.

Chapter 7

CONCLUSIONS

This dissertation is dedicated to the stochastic blind adaptive nonlinear signal processing problem of Blind Source Recovery (BSR). Blind Source Recovery represents unsupervised estimation of original source signals directly from the measurement space, with or without precise MIMO environment identification. A discussion on the BSR problem definition, applications, choice of the state space representation and practical issues is included in Chapter 1. In chapter 2, a focused survey of the major theoretical development in the field of BSR as related to this work has been provided. The natural gradient and its application to the static and dynamic BSR adaptation has also been reviewed.

Chapter 3 is the crux of this work. Using state space as a compact network representation, the BSR optimization framework has been described for both non-linear and linear structures. The natural gradient learning has been incorporated to develop equivariant linear state space update laws for the cases of minimum phase and non-minimum phase environments. In the case of minimum phase environments, separate adaptation algorithms have been developed for feedforward and feedback demixing

networks. It is demonstrated, via simulations, that the derived BSR algorithms exhibit robust convergence irrespective of the adopted network structure and various parameter initializations. Further, it is observed that based on the initialization of parameters the overall network may converge to different but equivalent state space representations. Since BSR denotes blind recovery, state space BSR algorithms have been extended using parametric source distributions. A couple of formations are introduced in chapter 4 with the capability to estimate the types of the underlying source distributions and eventually control the shape of the nonlinear score function in all the BSR update laws. These complementary adaptation algorithms add the capability to estimate the underlying distribution structure of original sources to the BSR framework. This results in extensive algorithms that can be applied in an autonomous fashion to mixtures of multiple source distributions. This is an important development as typically the nature of corruption in the measurement space is unknown. It is demonstrated via simulations that the proposed adaptive score functions can indeed efficiently capture the stochastic nature of actual sources and cause proficient convergence of BSR algorithm for mixtures of several different types of sources.

Although it is theoretically tractable to have the dimension of mixture space equal to the number of original sources, but in a blind setting the actual number of sources is unknown. This deficiency in the a priori knowledge or the structure of a recovery problem may ensue an undercomplete or an overcomplete BSR problem. The transformation of an undercomplete BSR mixture into a complete BSR mixtures has been discussed and related issues have been pointed out. For the overcomplete mixtures, there exists no solution unless the mixture can be projected onto a sparse temporal or spectral

analysis space. Algebraic ICA algorithm has been proposed to perform static mixing environment estimation for sources that exhibit sparseness in the time-domain. AICA approach has been combined with an L_1 -norm interior point linear programming approach to recover overcomplete mixtures of speech.

Chapter 6 discusses the application of BSR framework to the domain of CDMA wireless communication networks. An in-depth signal model has been utilized to develop a good understanding of the basic DS-CDMA implementations and the modern 3G WCDMA or UMTS FDD implementations. Three BSR based solutions are presented for the critical forward link multi-user detection in these networks. Extensive comparison of the proposed BSR user detection schemes with conventional downlink user detection formats indicate their effectiveness and robustness under conditions of congestion, channel estimation artifacts and unmodeled multiple access interference.

Quadratic Independence Measure (QIM) and Algebraic Matrix Distance Index (AMDI) are two new BSR performance measures that have been an outcome of this work and are presented in appendices D.2 and D.3.

The major highlights of this work include

- Using an info-theoretic, parametric, state space BSR optimization framework and efficient natural gradient learning, robust BSR update laws have been derived for various state space arrangements.
- Structural environment configurations such as the minimum phase and the non-minimum phase have been treated comprehensively. Linear feedforward and feedback structures for BSR have been proposed and implemented.

- The convergence of the natural gradient BSR algorithms has been investigated, via extensive simulations, for various demixing network structures and parameter initializations.
- Investigation into efficient parametric modeling of the source distributions, resulting in two new source density models proposed for sub- and super-gaussian sources, respectively.
- Adaptive estimation of underlying source distributions was achieved for linear convolutive BSR. This results in completely blind algorithms, which do not even require the knowledge of the type of underlying source distributions to be estimated. The comprehensive algorithms are also capable to demix / decorrelate mixtures of multiple source distributions.
- A new data space, cost-effective, algebraic independent component analysis algorithm, called Algebraic ICA (AICA) has been introduced.
- In conjunction with source inferencing algorithms for overcomplete mixtures, AICA has been applied to the realm of overcomplete blind source separation of speech with encouraging results.
- Three new BSR based multi-user detection algorithms have been proposed to the field of CDMA wireless communication networks. The proposed algorithms have been simulated for both generic DS-CDMA and modern 3G WCDMA (and CDMA-2000) communication standards. The BSR MUD algorithms demonstrate a favorable performance advantage to conventional user detection schemes for CDMA wireless communication networks.

Although, the dissertation concludes here, but there are several, yet to be tackled, open issues in BSR. Some of the possible extensions of this work are

- For the case of nonlinear environments, efficient BSR algorithms need to be derived and their performance needs to be characterized for various different types of nonlinearities.
- BSR performance evaluation of the Algebraic ICA algorithm needs to be done on more realistic sparse overcomplete data, e.g., electromedical data (i.e., EEG, MEG to name a few), astrophysical data etc.
- BSR formulations have been applied to the CDMA downlink wireless communication networks in this work. Currently the work is being extended to the CDMA uplink channel. There are several other applications in communication systems, wireless networks, acoustics, imaging, text, astronomy, finance etc., where application of BSR techniques can be very useful.

APPENDICES

Appendix A

List of Abbreviations and Acronyms

3G	3 rd Generation (wireless networks)
3GPP	3rd Generation Partnership Project
AICA	Algebraic Independent Component Analysis
AMDI	Algebraic Matrix-Distance Index
AR	Autoregressive filter
ARMA	Autoregressive Moving Average Filter
ARMAX	Autoregressive Moving Average Filter with Exogenous Variables
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BLER	Blocking Error Rate
B-LMMSE	Block Linear Minimum Mean Squared Error Detector
BLUE	Best Linear Unbiased Estimator
BMMR	Blind Mixing Matrix Recovery
BMUD	Blind Multi-user Detection
BP	Basis Pursuit
BPSK	Binary Phase Shift Keying
BS	Base Station
BSD	Blind Source Deconvolution
BSR	Blind Source Recovery
BSS	Blind Source Separation
cdf	Cumulative Density Function
CDMA	Code Division Multiple Access

CDMA2000	Wireless Standard Name (U.S. equivalent of WCDMA)
CPICH	Common Pilot Channel
CRC	Cyclic Redundancy Check
DM	Distortion Measure
DPCCCH	Dedicated Physical Control Channel
DPCH	Dedicated Physical Channel
DPDCH	Dedicated Physical Data Channel
DS-CDMA	Direct Sequence Code Division Multiple Access
EPP	Exploratory Projection Pursuit
FB	Feedback Network
FDD	Frequency Domain Duplex
FEC	Forwarded Error Correction Codes
FF	Feedforward Network
FIR	Finite Impulse Response
GGASF	Generalized Gaussian Adaptive Score Function
GICA or geo-ICA	Geometric Independent Component Analysis
GTF	Global Transfer Function
HASF	Hyperbolic Adaptive Score Function
ICA	Independent Component Analysis
IID or iid	Independent Identically Distributed
IIR	Infinite Impulse Response
IP-LP	Interior Point Linear Programming
IS-95	EIA Interim Standard 95 (also called CDMAone)
KL	Kullback Lieblar Divergence
LMMSE	Linear Minimum Mean Squared Error
LMMSE	Linear Minimum Mean Squared Error Detector
MA	Moving Average Filter
MAI	Multiple Access Interference
MCBD	Multichannel Blind Source Deconvolution
MF	Matched Filter
MIMO	Multiple Input Multiple Output
MISI	Multichannel Inter-symbol Interference

MLE	Maximum Likelihood Estimation
MS	Mobile Station
MSE	Mean Squared Error
MUD	Multi-user Detection
OSE	Overcomplete Source Estimation
OVSF	Orthogonal Variable Spreading Factor Codes (Variable Length Walsh Codes)
P/S	Parallel to Serial Converter
PCA	Principal Component Analysis
pdf	Probability Density Function
PDF	Probability Distribution Function
PN	Pseudo-noise
QIM	Quadratic Information Measure
QoS	Quality of Service
QPSK	Quaternary Phase Shift Keying
RAKE-BSR	RAKE with Blind Source Recovery Detector
RAKE-PCA	RAKE with Principle Component Analysis Detector
S/P	Serial to Parallel Converter
SF	Spreading Factor
SINRM	Maximum Signal to Interference plus Noise Ratio
SM	Separation Measure
SNR	Signal to Noise Ratio
SUD	Single User Detection
TB	Transport Block
TDD	Time Domain Duplex
TF	Transfer Function
TFCI	Transport Format Combination Indicator
TPC	Transmit Power Control
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
WCDMA	Wideband Code Division Multiple Access
ZF	Zero Forcing

Appendix B

Kullback Lieblar Divergence: A Performance Functional for Blind Source Recovery

Kullback Lieblar Divergence (or the relative entropy) [38, 69, 72, 100] between two probability laws $Q(X = x_i) = q_i$ and $P(X = x_i) = p_i$ is defined as

$$K_X = \int_{x \in X} p_i \ln \left(\frac{p_i}{q_i} \right) dx \quad (\text{B.47})$$

The relative entropic divergence of the mutual probability distribution function (pdf) of a random output vector \mathbf{y} (Kullback-Lieblar divergence) with respect to the product of marginal distribution functions is a measure of the independence of the marginal output variables. Kullback-Lieblar divergence is, therefore, an appropriate performance measure for blind source recovery based on the minimization of mutual information.

In the continuous case this relation is given by

$$L(y) = \int_{y \in Y} p_y(y) \ln \left| \frac{p_y(y)}{\prod_{i=1}^n p_{y_i}(y_i)} \right| dy \quad (\text{B.48})$$

where

$p_y(y)$ is the probability density function of the random output vector y

$p_{y_i}(y_i)$ is the probability density function of the i^{th} component of the output vector y

This functional $L(y)$ is a “distance” measure with the following properties

$$\text{i) } L(y) \geq 0 \quad (\text{B.49})$$

$$\text{ii) } L(y) = 0 \text{ iff } p_y(y) = \prod_{i=1}^n p_{y_i}(y_i) \quad (\text{B.50})$$

This measure provides an estimate of the degree of dependence among the various components of the recovered output signal vector and is appropriate to be used as the objective functional in the optimization framework of the BSR problem.

For the discrete case, the functional becomes

$$L(y) = \sum_{y \in Y} p_y(y) \ln \left| \frac{p_y(y)}{\prod_{i=1}^n p_{y_i}(y_i)} \right| \quad (\text{B.51})$$

The functional can be further simplified assuming the statistical properties of the output signals to be ergodic.

$$L(y(k)) = \sum_{y \in Y} p_y(y(k)) \ln \left| \frac{p_y(y(k))}{\prod_{i=1}^n p_{y_i}(y_i(k))} \right| \quad (\text{B.52})$$

Further simplification can be achieved using the assumption that as the algorithm approaches convergence the components of the output vector $y(k)$ will become statistically less dependent, ultimately approaching independence as close as possible. Therefore the above functional can be re-written in its simplified mutual information form using the entropy of signals, namely,

$$L(y(k)) = -H(y(k)) + \sum_{i=1}^n H_i(y_i(k)) \quad (\text{B.53})$$

where

$H(y(k))$: Entropy of the signal vector $y(k)$, given by

$$H(y) = -E \left[\ln |p_y(y)| \right] = \begin{cases} - \int_{y \in Y} p_y(y) \ln |p_y(y)| dy & \text{Continuous Case} \\ - \sum_{y \in Y} p_y(y) \ln |p_y(y)| & \text{Discrete Case} \end{cases} \quad (\text{B.54})$$

$H_i(y_i(k))$: Marginal entropy of a component signal $y_i(k)$.

Appendix C

State Space Network: Specialized Filtering Structures

One of the advantages of using state space model is that most signal processing filter structures form special cases where the constituent matrices of the dynamic state equation take specific forms. As an example, implementation for both the Infinite Impulse Response (IIR) and the Finite Impulse Response (FIR) filters using state space is discussed in this appendix [109, 110].

C.1. IIR Filtering

Consider the case where the matrices A and B are set in the canonical form I (or the controller form). In this case A and B are conveniently represented as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & \cdots & A_{1L} \\ I & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{C.1})$$

where

A : matrix of dimension $Lm \times Lm$.

A_{1j} : Block sub-matrix of dimension $m \times m$, may be simplified to a diagonal matrix

I : Identity matrix of dimension $m \times m$

0 : Zero matrix of dimension $m \times m$

B : matrix of dimension $Lm \times m$

The state matrix is given by

$$X_k = X(k) = \begin{bmatrix} X_1(k) \\ X_2(k) \\ \vdots \\ X_L(k) \end{bmatrix} \quad (\text{C.2})$$

where

$X(k)$: is a $Lm \times m$ dimensional state vector for the filter, and each

$X_j(k)$: is an $m \times m$ dimensional component state vector

For this model the state model reduces to the following set of equations representing an IIR filtering structure.

$$\begin{aligned} X_1(k+1) &= \sum_{j=1}^L A_{1j} X_j(k) + m(k) \\ X_2(k+1) &= X_1(k) \\ &\vdots \\ X_L(k+1) &= X_{L-1}(k) \end{aligned} \quad (\text{C.3})$$
$$y(k) = \sum_{j=1}^L C_j X_j(k) + Dm(k)$$

For an IIR filter represented in the state space canonical form I, there is no update law required for B , while for the matrix A , it is needed update only the first block row and

therefore the state equation update law (see section 2.1.2) reduces to the following block update rule

$$\Delta A_{1j} = -\eta \frac{\partial H^k}{\partial A_{1j}} = -\eta (f_{A_{1j}}^k)^T \lambda_{k+1} = -\eta \lambda_{k+1} X_j^T(k) \quad (C.4)$$

The special structure in matrix A also effects the update law for the co-state equations (propagating in future time) which reduces to

$$\begin{aligned} \lambda_1(k) &= \lambda_2(k+1) + C_1^T \frac{\partial L^k}{\partial y_k}(k) \\ \lambda_2(k) &= \lambda_3(k+1) + C_2^T \frac{\partial L^k}{\partial y_k}(k) \\ &\vdots \\ \lambda_L(k) &= C_L^T \frac{\partial L^k}{\partial y_k}(k) \end{aligned} \quad (C.5)$$

Solving specifically for time k and then using time shift for time $k+1$, we obtain the recursive form for the update as

$$\begin{aligned} \lambda_1(k) &= \sum_{j=1}^L C_j^T \frac{\partial L^k}{\partial y_k}(k+j-1) \\ \lambda_1(k+1) &= \sum_{j=1}^L C_j^T \frac{\partial L^k}{\partial y_k}(k+j) \end{aligned} \quad (C.6)$$

which is structurally similar to the natural gradient algorithm derived for blind source separation/ deconvolution by [155] and can be implemented in a similar fashion by using the usual time delayed version of the algorithm implemented by buffer storage memory.

Using (C.6), we can further simplify (C.4) for the update of the block sub-matrices in A as

$$\Delta A_{1j} = -\eta \sum_{j=1}^L C_j^T \frac{\partial L^k}{\partial y_k}(k+j) X_j^T \quad (C.7)$$

C.2. FIR Filtering

In this case, the first row of block sub-matrices A_{1j} in (C.1) is zero, therefore the above filtering model (C.3) reduces to the FIR filtering structure. In this case the state space model reduces to

$$\begin{aligned}
 X_1(k+1) &= m(k) \\
 X_2(k+1) &= X_1(k) \\
 &\vdots \\
 X_L(k+1) &= X_{L-1}(k) \\
 y(k) &= \sum_{j=1}^L C_j X_j(k) + Dm(k)
 \end{aligned} \tag{C.8}$$

where the state dynamic equations reduce merely to the delays of the mixture inputs i.e.

$$\begin{aligned}
 X_1(k) &= m(k-1) \\
 X_2(k) &= m(k-2) \\
 &\vdots \\
 X_L(k) &= m(k-L)
 \end{aligned} \tag{C.9}$$

In this case only matrices C and D need to be updated. Using the MIMO controller form, both matrices A and B contain only fixed block identity and zero sub-matrices and are conveniently absorbed in (C.9).

In case of non-minimum phase environments, the demixing network is typically constituted as a double-sided FIR filter [14, 129]. Discrete time finite length double sided FIR filter can be used to approximate the otherwise unstable theoretical demixing network due to the presence of poles outside the unit circle [22]. The implementation of these non-causal double sided filter is handled by using a time delayed version of the above algorithm with the corresponding increased requirement of buffered storage. The

primary advantage of these doubly finite filters being their ability to converge at “stable” solutions even for non-minimum phase environments [14, 129].

One possible method to deal with the problem of the filter approximation length and resulting delay is in the frequency domain implementations of the algorithm, where the converged solution represents a 1024 or higher tap double sided FIR filter in equivalent frequency domain. The result can be converted to the time domain equivalent by appropriately using the inverse FFT followed by chopping or windowing techniques to optimally contain maximum possible filter power spectral density (PSD) while minimizing algorithmic delay and hence the buffering memory requirement [73, 76]. However, recent studies have shown that the frequency domain techniques have an upper bound on the performance as the filters lengths become large [18, 90].

Appendix D

Performance Benchmarks for Blind Source Recovery

Appendix D.1. Conventional Performance Benchmarks

In this section we will provide an overview of several performance benchmarks that have been employed by various researchers for the problem of blind source recovery [113, 131].

D.1.1. Signal to Noise Ratio (SNR)

Signal to noise ratio is a well-known communication benchmark, which defines the ratio between the desired signal and the unwanted noise/distortion powers usually expressed in dBs.

$$SNR = 10 \log \frac{P_s}{P_n} \quad (D.1.10)$$

where P_s – represents the signal power, and

P_n – represents the noise and/or unwanted signal power

This benchmark is applicable in situations where the desired signal is known and is therefore not very suitable for BSR problems where no such reference signal exists.

D.1.2. Multichannel Intersymbol Interference (MISI)

MISI is an extension of the communication benchmark inter-symbol interference, which quantifies the smearing of a communication channel by an adjacent channel due to their overlapping temporal and/or spectral characteristics [52, 98]. MISI is useful as a dispersion measure as it is insensitive to the overall gain and mean group delay. In the case of BSR type problems, MISI is a measure to determine the “distance” between the global, i.e., the combined mixing and demixing network transfer functions and an identity impulse transfer function. This is a measure of the global diagonalization and permutation as achieved by the demixing network [9 and the references therein]

$$MISI_k = \sum_{i=1}^N \frac{\left| \sum_j \sum_p |G_{pij}| - \max_{p,j} |G_{pij}| \right|}{\max_{p,j} |G_{pij}|} + \sum_{j=1}^N \frac{\left| \sum_i \sum_p |G_{pij}| - \max_{p,i} |G_{pij}| \right|}{\max_{p,i} |G_{pij}|} \quad (D.1.11)$$

where for a linear case

$G(z) = H(z) * \bar{H}(z)$ – represents Global Transfer Function,

$\bar{H}(z) = [A_e, B_e, C_e, D_e]$ – Transfer Function of Environment

$H(z) = [A, B, C, D]$ – Transfer Function of Network

This is the performance benchmark of choice for most of the research on BSR. However, the main drawback of this benchmark is that it requires knowledge of the mixing

environment, which is available offline only in synthetic simulation environments and hence cannot be employed in practical blind scenarios.

Another variation of the above performance measure uses a squared version of the global transfer function matrix $G_{p_{ij}}$, i.e., the performance measure relies on L_2 norm instead of the L_1 norm used above. This technique has an additional inherent flaw that small error terms will be de-emphasized.

D.1.3. Maximum signal to interference plus noise ratio (SINRM)

This measure uses the maximum signal to interference plus noise ratio as a measure to determine separation. Interference in source k is constituted by all sources j such that $j \neq k$. Then SINR of source k at output i of separator W is defined as [32]

$$SINR_k(w_i) = \frac{\sigma_k^2 |w_i^H a_k|^2}{w_i^H R_{vk} w_i} \quad (D.1.12)$$

where

w_i - is the i^{th} column of separator matrix W

a_k - is the k^{th} column of mixing matrix A

σ_k^2 - is the input power of source k

R_{vk} - is the temporal mean (for nonstationary sources) of the correlation matrix of the noise plus interference for source k , defined by

$$R_{vk} = R_x - \sigma_k^2 a_k a_k^H \quad (D.1.13)$$

Again this measure requires knowledge of mixing transfer function, which is unknown in blind problems.

D.1.4. Mean Squared Error (MSE)

Mean squared error is probably the best-known adaptation measure. MSE uses the L_2 norm of the error as a measure to determine the convergence of an algorithm. In the absence of a desired reference to compute an explicit error measure as in the BSR case, this error may be computed as the difference between the current and the previous value of an adapted parameter or using a probabilistic score function. Although using MSE one can determine whether an algorithm has reached its steady state but there is no way to determine whether the algorithm has converged to the desired solution or a spurious local minimum, and as such is not suitable for quantification of BSR algorithm. For any parameter γ , MSE is given by

$$MSE(\gamma) = \begin{cases} |\gamma(k) - \gamma(k-1)|^2 \\ |\gamma(k) - \psi(\gamma(k))|^2 \end{cases} \quad (D.1.14)$$

Some other performance indices proposed in [113] for synthetic audio environments are given below

D.1.5. Distortion Measure

The distortion in the j^{th} separated output can be defined as

$$D_j = 10 \log \left(E \left\{ \left(m_{j,s_j} - \alpha_j y_j \right)^2 \right\} / E \left\{ \left(m_{j,s_j}, s_j \right)^2 \right\} \right) \quad (D.1.15)$$

where

$$\alpha_j = E \left\{ m_{j,s_j}^2 \right\} / E \left\{ y_j^2 \right\} - \text{is a scaling factor}$$

$E \{ \cdot \}$ - is the statistical expectation operator

y_j - is the corresponding output for the j^{th} source, and

m_{i,s_j} - is the contribution of the j^{th} source to the i^{th} microphone.

D.1.6. Separation Measure

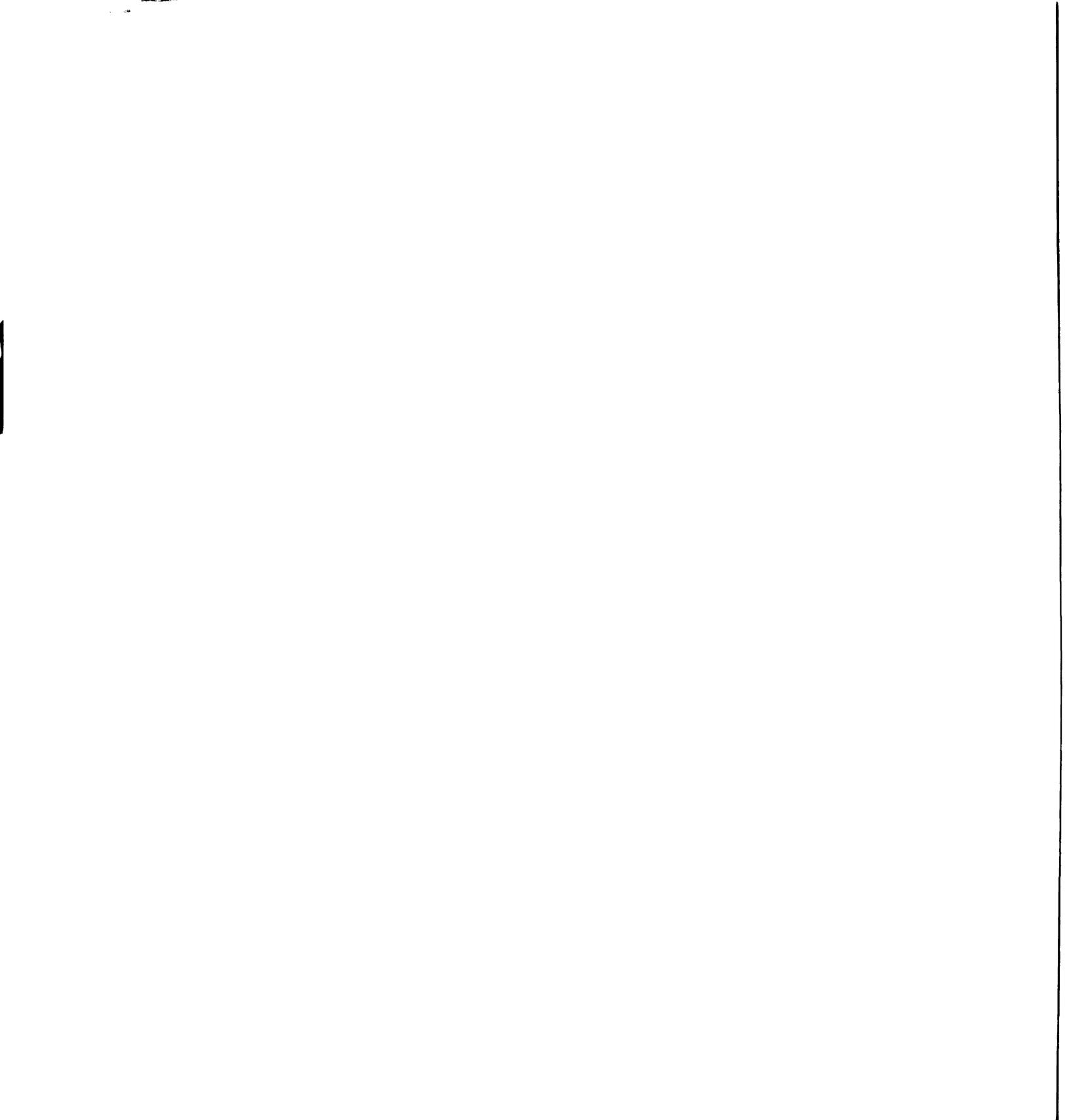
The quality of separation for the j^{th} separated output can be defined as

$$S_j = 10 \log \left(E \left\{ \left(y_{j,s_j} \right)^2 \right\} / E \left\{ \left(\sum_{i \neq j} y_{j,s_i} \right)^2 \right\} \right) \quad (\text{D.1.16})$$

where

y_{j,s_i} - represents the j^{th} output when only i^{th} source is active.

The last two benchmarks can also be represented in the frequency domain and are applicable for synthetic evaluations only, where the original source signals are known and their contribution to an observed mixture & output can be estimated.



Appendix D.2. Quadratic Information Measure (QIM)

Quadratic Information Measure (QIM) is a novel performance index to measure the statistical independence of data sequences. In this work, QIM has been proposed and applied it to the framework of blind source recovery (BSR) that includes blind source separation, deconvolution and equalization. This performance index is capable of measuring the mutual independence of data sequences directly from the data. This information theoretic independence measure is derived based on Renyi's quadratic entropy [100] estimated by a finite data length Parzen window [93] using a gaussian kernel. Simulation results are presented to validate the performance of the proposed benchmark and compare it with other established benchmarks.

D.2.1. Practical Issues in Estimating Independence of Output Sequences

Stochastic blind signal processing tasks such as blind source separation, deconvolution and equalization have been the focus of wide interest during the last two decades. This wide interest is due to several attractive and diverse application domains for these kinds of problems that include blind classification of topics in an internet chatroom, communication systems, audio and acoustics, finance and marketing, astronomy and physics, bio-medical, space and geo-exploratory applications. However, a daunting task remains to reliably classify the performance of a number of proposed algorithms especially in a fashion that makes them suitable for practical implementation. Most of the presented research on the topic relies on performance benchmarks that require knowledge

of either the original source signals or the mixing environment transfer function. Both these quantities are otherwise assumed to be unknown (no precise knowledge of the above mentioned quantities is actually what makes these tasks blind). Therefore, these performance benchmarks render themselves useful for synthetic simulations only. In a practical situation, while making observations or recording a sequence of available information, precise knowledge of the original signals or the environment transfer function is never available. Compounding this situation with the inherent indeterminacies of permutation and scaling in the blind signal recovery problem, the dilemma of having an unknown number of sources, the unknown order of filters required say for a deconvolution type of setup etc. severely limits the capability to determine the quality of signal separation algorithms in practical situations.

The use of a new benchmark has been proposed to determine mutual independence of a batch of signals for source recovery formulations. Unlike most other simulation performance benchmarks employed for BSR, this benchmark can be computed directly from the output data of the BSR network. It may also be applied directly to the observed mixture data in order to quantify the transformation effect of the BSR network. This novel benchmark is based on Renyi's Mutual Information, see [97, 150]. In the context of BSR, the use of this independence performance benchmark ensues the debate on the correspondence between statistical independence and source recovery. In this appendix, an overview of Renyi's quadratic entropy is first provided and it is discussed that how it can be utilized for estimation of signal independence. This is followed by a discussion on how to apply QIM to BSR type problems. At the end, simulated results are presented and observations are made from this study.

D.2.2. Independence Benchmark using Quadratic Mutual Information

A general benchmark for determination of statistical independence based on Renyi's quadratic mutual information has been proposed in this work [131]. Also discussed are salient features of this measure and use of Parzen Window method for its estimation. For a more detailed discussion on Renyi's quadratic mutual information, see [97, 100]. The framework was applied to the problem of blind source separation by Principe et. al. for the development of a generalized information theoretic learning framework; where the learning of the demixing structure can be derived directly from data. However, the resulting update laws are computationally very expensive. Instead, it is proposed to utilize a finite data-length quadratic entropic measure as a performance benchmark for achieving statistical independence.

D.2.2.1. Renyi's Entropy

Renyi's entropy definition is derived from his proposed theory of means [100] and is given by

$$H = \varphi^{-1} \left(\sum_{k=1}^N p_k \varphi(I(p_k)) \right) \quad (\text{D.2.17})$$

where,

$\varphi(\cdot)$ - is a continuous and strictly monotonic function subclass of Kolmogorov-Nagumo functions [69, 92]. To satisfy the constraints of an information measure

$$\varphi(x) = \begin{cases} x & \text{Shannon's Entropy} \\ 2^{(1-\alpha)x} & \text{Renyi's Entropy} \end{cases} \quad (\text{D.2.18})$$

$I(p_k)$ - any information measure, e.g., $I(p_k) = -\log(p_k)$ is Hartley's information measure (or Shannon's info-theoretic measure [38]).

Simplifying the above relation, (D.2.17)

$$H_{R_\alpha} = \frac{1}{1-\alpha} \log \left(\sum_{k=1}^N p_k^\alpha \right); \alpha > 0, \alpha \neq 1 \quad (\text{D.2.19})$$

Renyi's quadratic entropy is the special case for $\alpha = 2$, i.e.,

$$H_{R_2} = -\log \left(\sum_{k=1}^N p_k^2 \right) \quad (\text{D.2.20})$$

D.2.2.2. Parzen Window Estimator

For the purpose of estimating the squared probability density for the quadratic mutual information, a Parzen kernel estimator [93] is employed. The Parzen estimator takes the form of a convolution of an estimator kernel with the observations. Using this estimator, the pdf $p_y(z)$ of a random vector $y \in \mathfrak{R}^m$ is given by

$$\hat{p}_y(z) = \frac{1}{N} \sum_{i=1}^N \kappa(z - y_i) \quad (\text{D.2.21})$$

where

$y_i \in \mathfrak{R}^m$ is the i^{th} observation vector

$\kappa(\cdot)$ is a kernel function that satisfies the conditions for a pdf. There are several such well-known kernel functions, e.g., the gaussian kernel, the cauchy kernel or the uniform kernel [93]. For ease of implementation, the Gaussian kernel is chosen with covariance matrix Σ , i.e.,

$$\kappa(z) = G(z, \Sigma) = \frac{1}{(2\pi)^{M/2} |\Sigma|^{1/2}} \exp\left(-\frac{z^T \Sigma^{-1} z}{2}\right) \quad (\text{D.2.22})$$

This choice is motivated by an integral property of the Gaussian kernel, which results in an efficient and exact computation of Renyi's quadratic entropy, for the case of 2 gaussian kernels

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{k=1}^N G(z_k - y_i, \Sigma_1) G(z_k - y_j, \Sigma_2) \\ = \int_{-\infty}^{\infty} G(z - y_i, \Sigma_1) G(z - y_j, \Sigma_2) dz = G(y_i - y_j, \Sigma_1 + \Sigma_2) \end{aligned} \quad (\text{D.2.23})$$

where

$y_i \in \mathfrak{R}^m$ and $y_j \in \mathfrak{R}^m$ are two data vectors in the space

Σ_1 and Σ_2 are covariance matrices for corresponding Gaussian kernels.

D.2.3. Quadratic Independence Measure (QIM)

The proposed performance benchmark is based on the Cauchy-Schwartz Inequality. For L_2 norm, the inequality is

$$\|x\|^2 \|y\|^2 \geq (x^T y)^2 \quad (\text{D.2.24})$$

Using the above inequality, the Kullback-Lieblar divergence is "approximated" by taking the logarithm of both sides of (D.2.24) and rearranging

$$\log \frac{\|x\|^2 \|y\|^2}{(x^T y)^2} \geq 0 \quad (\text{D.2.25})$$

For two pdfs $f(x)$ and $g(x)$, the above inequality can be used as a divergence measure between the two pdfs. The relation is given by

$$P_{cs}(f, g) = \log \frac{\left(\int_{-\infty}^{\infty} f(x)^2 dx \right) \left(\int_{-\infty}^{\infty} g(x)^2 dx \right)}{\left(\int_{-\infty}^{\infty} f(x)g(x) dx \right)^2} \quad (\text{D.2.26})$$

with the following properties

- $0 \leq P_{cs}(\cdot) \leq 1$
- $P_{cs}(\cdot) = 0$ if and only if $f(x) = g(x)$
- $P_{cs}(f, g) = P_{cs}(g, f)$

expressing (D.2.26) in terms of quadratic entropy

$$P_{cs}(f, g) = 2H_{R_2}(\sqrt{f(x)g(x)}) - (H_{R_2}(f(x)) + H_{R_2}(g(x))) \quad (\text{D.2.27})$$

In order to apply this criterion for the measurement of independence in the source separation framework, let

$f(x)$ represent the joint probability density estimate of the signal vector using Parzen estimator, and

$g(x)$ represent the product of marginal probability density estimates using the Parzen estimator

For k observations of an output vector $Y \in \mathfrak{R}^N$, the above relation can be expressed as

$$P_{cs}(y) = -\log[M_{CI}(Y)] \quad (\text{D.2.28})$$

where

$M_{CI}(Y)$ -represents the normalized cross information measure

$$M_{CI}(Y) = \frac{M_{UCIP}(Y)^2}{M_{JIP}(Y) \prod_{i=1}^N M_{MIP}(Y_i)} \quad (D.2.29)$$

$M_{JIP}(Y)$ - represents the joint information measure

$$M_{JIP}(Y) = \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k \left(\prod_{l=1}^N G(Y_{il} - Y_{jl}, 2\sigma^2 I) \right) \quad (D.2.30)$$

$M_{PCIP}(Y_j, Y_l)$ - represents the partial marginal information measure

$$M_{PCIP}(Y_j, Y_l) = \frac{1}{k} \sum_{i=1}^k \left(G(Y_{jl} - Y_{il}, 2\sigma^2 I) \right), l = 1 \dots N \quad (D.2.31)$$

$M_{MIP}(Y_i)$ - represents the marginal information measure

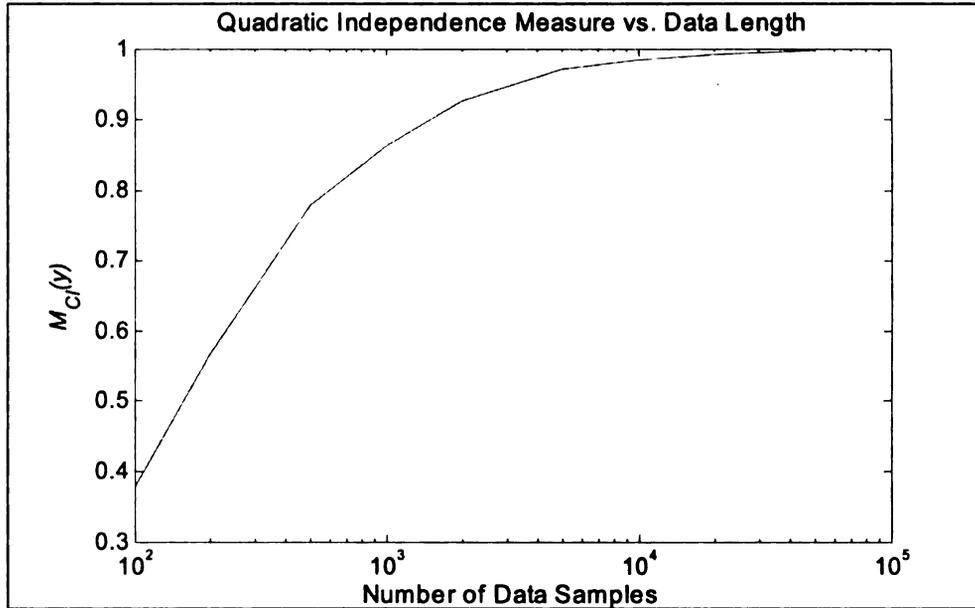
$$M_{MIP}(Y_i) = \frac{1}{k} \sum_{j=1}^k \left(M_{PCIP}(Y_j, Y_l) \right), l = 1 \dots N \quad (D.2.32)$$

$M_{UCIP}(Y)$ - represents unnormalized cross information measure

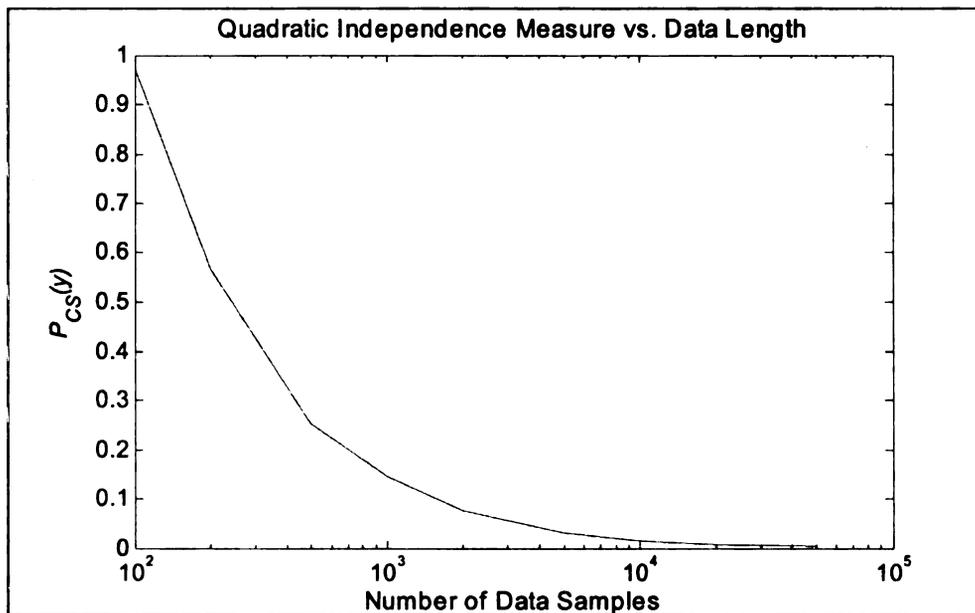
$$M_{UCIP}(Y) = \frac{1}{k} \sum_{j=1}^k \left(\prod_{l=1}^N M_{PCIP}(Y_j, Y_l) \right) \quad (D.2.33)$$

In case of linear instantaneous and convolutive mixtures, separability can be defined as a combination of output independence and diagonalization of the global transfer function [7, 28, 36, 37, 106, 110]. The diagonalization measure ensures that all sources are extracted and that the algorithm did not converge to a spurious solution in the data space, where estimated data sources are independent but different from the original sources. In a completely blind scenario, there is no known criterion to determine this diagonalization without explicit knowledge of the mixing transfer function. However, for a finite known number of network outputs practically one can verify that all the separated

sources are distinct and independent. This ensures separation of all non-gaussian source distributions, whereas for a gaussian source this ensures separation from other distributions only, while the reverberative effects might still be present.



(a)



(b)

Figure D.1. Effect of length of observation sequence on (a) Cross Information Measure,
(b) Quadratic Independence Measure (QIM)

The effectiveness of this performance measure depends on the length as well as the probability distributions of the constituents in an observation sequence. In Fig. D.1, a graph for QIM performance measure is being provided as a function of length of observation data. The data has three constituents that include speech, communication signals, and gaussian noise.

D.2.4. Simulations Results

Due to space limitation, only one simulation is being provided. The mixing environment is given by

$$\sum_{j=0}^{m-1} A_j m(k-i) = \sum_{i=0}^{n-1} B_i s(k-i) + v(k) \quad (D.2.34)$$

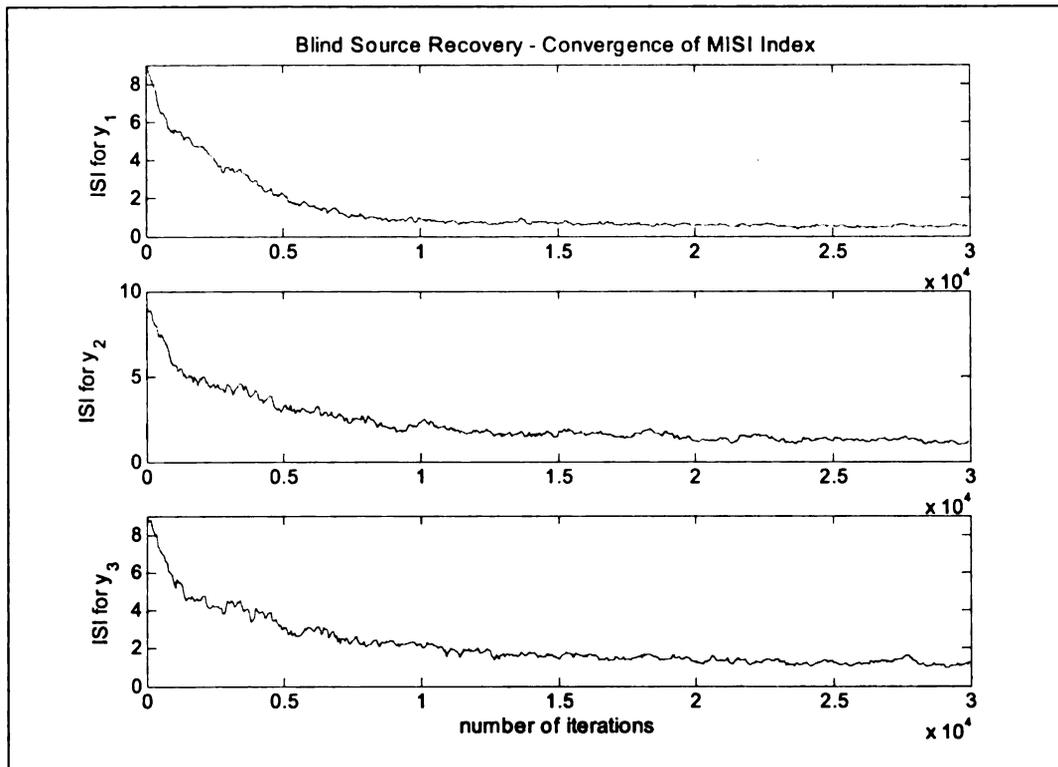
where

$$A_0 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, A_1 = \begin{bmatrix} 0.5 & 0.8 & -0.7 \\ 0.8 & 0.3 & -0.2 \\ -0.1 & -0.5 & 0.4 \end{bmatrix}, A_2 = \begin{bmatrix} 0.06 & 0.4 & -0.5 \\ 0.16 & -0.1 & -0.4 \\ -0.3 & -0.06 & 0.3 \end{bmatrix}$$

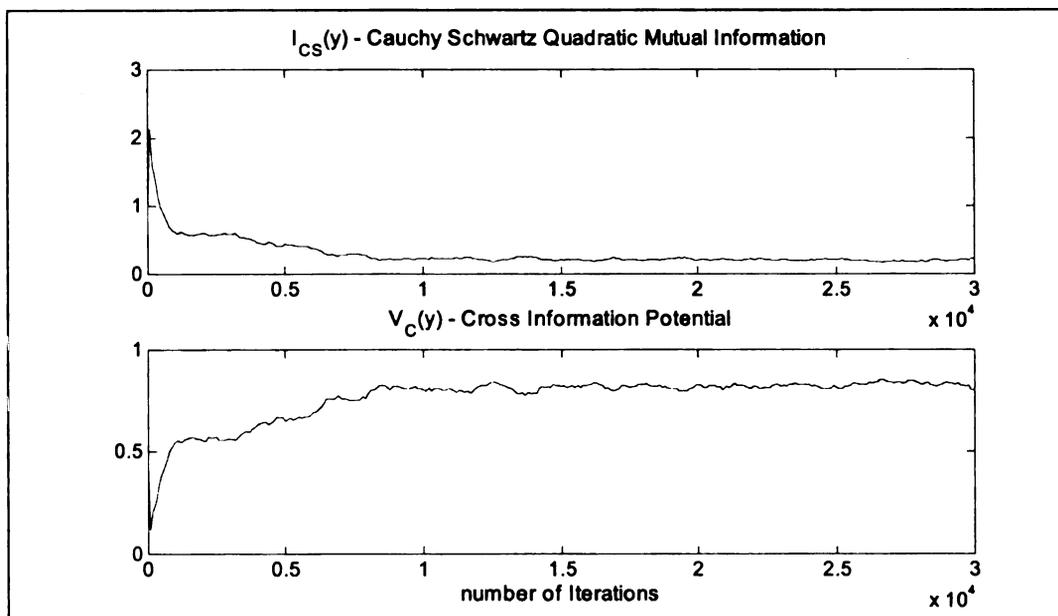
$$B_0 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.3 & 1 & 0.1 \\ 0.6 & -0.8 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} 0.5 & 0.5 & 0.6 \\ -0.3 & 0.2 & -0.3 \\ -0.2 & -0.43 & 0.6 \end{bmatrix}, B_2 = \begin{bmatrix} .125 & 0.06 & 0.2 \\ -0.1 & 0 & 0.4 \\ 0.08 & -0.13 & 0.3 \end{bmatrix}$$

$v(k)$ - Additive Gaussian noise

The feedforward separation results for each channel using MISI and QIM (computed with a batch of 1000 samples) are shown below



(a)



(b)

Figure D.2. (a) Convergence of MISI performance Index, (b) Corresponding Performance of QIM performance Index

D.2.5. Observations

The proposed quadratic independence benchmark [131] has been thoroughly investigated for BSR problems that include both minimum phase and non-minimum phase systems. Due to the normalized nature of QIM, it is suitable for performance comparison of demixing achieved by different algorithms. The primary inhibition is that computationally the benchmark has a quadratic relationship with the length of the data set. Therefore for practical situations, it is recommended to use either the QIM measure with a shorter data set or another computationally inexpensive measure such as MSE to determine the convergence of the algorithm. Once the algorithm has converged, a larger data chunk may be used to determine the achieved performance level using QIM.

Appendix D.3. Algebraic Matrix Distance Index

(AMDI)

Algebraic Matrix Distance Index (AMDI) [132-134, 137] is primarily a scaling, sign and permutation invariant numerical index to estimate the “similarity” between the column-bases of two matrices with conformal dimensions. These attributes make AMDI suitable for comparison of any two matrices of similar dimensions including results for algorithms such as AICA and geo-ICA. Unlike the multichannel intersymbol interference (MISI, also called Amari Performance Index), which measures the diagonalization of the global solution (if it can be calculated) in ICA or BSS. AMDI is used to compare the estimated mixing matrix with the synthetic or known mixing matrix to gauge the performance of the algorithms that compute the mixing matrix instead of the demixing matrix. Furthermore, for the case of overcomplete ICA, although no MISI can be calculated, yet AMDI can be used to verify the correctness of the estimated mixing matrix.

This Algebraic Matrix-Distance Index (AMDI) is a symmetric “matrix-distance” measure. Unlike geometric measure [115, 117], AMDI estimates the distance between two column-normalized matrices independent of possible sign and permutation ambiguities. The columns of the matrices are normalized by projecting them onto the

conformal unit hyper-sphere by dividing all the elements of each column by the norm of the corresponding column. The AMDI is given by

$$AMDI(W, H) = \frac{n - \sum_{rows} \max(|W^* H|)}{n} + \frac{n - \sum_{cols} \max(|W^* H|)}{n} \quad (D.3.1)$$

where

W, H – are two column-normalized matrices of similar dimensions.

The Algebraic Matrix Distance Index (AMDI) has the following properties:

- $0 \leq AMDI(W, H) \leq 2$; $AMDI(W, H) = 2$ only if both matrices are orthogonal to each other in the n -d space, also
- $AMDI(W, H) = AMDI(H, W)$

In order to better visualize the convergence properties of ICA, at times it is useful to compute the normalized AMDI instead. The normalization is done by the maximum AMDI during the estimation phase, which is typically AMDI at initialization.

Appendix E

Multi-Input Multi-Output (MIMO)

Canonical Form

This appendix describes the steps required to convert a MIMO filter definition, as used in all simulation examples to the canonical state space representation. For all the state space BSR simulations and derivations, the network structure is assumed to be in MIMO controller form [17, 103]. In this form, the matrices A and B have the structure

$$A = \begin{bmatrix} -Q_1 & -Q_2 & \cdots & -Q_{M-1} & -Q_M \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}, B = \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (\text{E.1})$$

$$C = [P_1 \quad P_2 \quad \cdots \quad P_{M-1} \quad P_M], D = [P_0] \quad (\text{E.2})$$

where the state space network/filter can be represented by

$$H(z) = [P(z)][Q(z)]^{-1} \quad (\text{E.3})$$

with

$$P(z) = \sum_{i=0}^n P_i z^{-i} \quad (\text{E.4})$$

$$Q(z) = \sum_{i=0}^n Q_i z^{-i} \text{ with } Q_0 = I_N \quad (\text{E.5})$$

The polynomial matrices $P(z)$ and $Q(z)$ of (E.3) are derived from a MIMO transfer as explained in the following section.

E.1. Transformations required for a MIMO Transfer Function

Consider a $p \times m$ multi-input multi-output transfer function of the form

$$H(z) = \begin{bmatrix} \frac{n_{11}(z)}{d_{11}(z)} & \frac{n_{12}(z)}{d_{12}(z)} & \dots & \frac{n_{1m}(z)}{d_{1m}(z)} \\ \frac{n_{21}(z)}{d_{21}(z)} & \frac{n_{22}(z)}{d_{22}(z)} & \dots & \frac{n_{2m}(z)}{d_{2m}(z)} \\ \vdots & \vdots & & \vdots \\ \frac{n_{p1}(z)}{d_{p1}(z)} & \frac{n_{p2}(z)}{d_{p2}(z)} & \dots & \frac{n_{pm}(z)}{d_{pm}(z)} \end{bmatrix} \quad (\text{E.6})$$

Defining, the monic least common denominator of each column i to be $l_i(z)$ of dimension say $1 \times n$, i.e.,

$$l_i(z) \triangleq LCM(d_{1i} \times d_{2i} \times \dots \times d_{pi}) \quad (\text{E.7})$$

We can represent each term of the transfer function as

$$\frac{n_{ij}(z)}{d_{ij}(z)} = \frac{n_{ij}(z)}{l_j(z) \left\{ \frac{d_{ij}(z)}{l_j(z)} \right\}} = \frac{n_{ij}(z) \left\{ \frac{l_j(z)}{d_{ij}(z)} \right\}}{l_j(z)} \quad (\text{E.8})$$

Now, defining the modified numerator polynomials of dimension $1 \times n$ as

$$\bar{n}_{ij}(z) \triangleq n_{ij}(z) \left\{ \frac{l_j(z)}{d_{ij}(z)} \right\} \quad (\text{E.9})$$

Using the above definitions, we can represent (E.6) as

$$H(z) = \begin{bmatrix} \frac{\bar{n}_{11}(z)}{l_1(z)} & \frac{\bar{n}_{12}(z)}{l_2(z)} & \cdots & \frac{\bar{n}_{1m}(z)}{l_m(z)} \\ \frac{\bar{n}_{21}(z)}{l_1(z)} & \frac{\bar{n}_{22}(z)}{l_2(z)} & \cdots & \frac{\bar{n}_{2m}(z)}{l_m(z)} \\ \vdots & \vdots & & \vdots \\ \frac{\bar{n}_{p1}(z)}{l_1(z)} & \frac{\bar{n}_{p2}(z)}{l_2(z)} & \cdots & \frac{\bar{n}_{pm}(z)}{l_m(z)} \end{bmatrix} \quad (\text{E.10})$$

or

$$H(z) = \begin{bmatrix} \bar{n}_{11}(z) & \bar{n}_{12}(z) & \cdots & \bar{n}_{1m}(z) \\ \bar{n}_{21}(z) & \bar{n}_{22}(z) & \cdots & \bar{n}_{2m}(z) \\ \vdots & \vdots & & \vdots \\ \bar{n}_{p1}(z) & \bar{n}_{p2}(z) & \cdots & \bar{n}_{pm}(z) \end{bmatrix} \begin{bmatrix} 1/l_1(z) & 0 & \cdots & 0 \\ 0 & 1/l_2(z) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/l_m(z) \end{bmatrix} \quad (\text{E.11})$$

therefore, by defining

$$N(z) \triangleq \begin{bmatrix} \bar{n}_{11}(z) & \bar{n}_{12}(z) & \cdots & \bar{n}_{1m}(z) \\ \bar{n}_{21}(z) & \bar{n}_{22}(z) & \cdots & \bar{n}_{2m}(z) \\ \vdots & \vdots & & \vdots \\ \bar{n}_{p1}(z) & \bar{n}_{p2}(z) & \cdots & \bar{n}_{pm}(z) \end{bmatrix} \quad (\text{E.12})$$

and

$$D(z) \triangleq \begin{bmatrix} l_1(z) & 0 & \cdots & 0 \\ 0 & l_2(z) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & l_m(z) \end{bmatrix} \quad (\text{E.13})$$

(E.11) can be expressed as

$$H(z) = N(z)D^{-1}(z) \quad (\text{E.14})$$

Note that (E.14) and (E.3) although similar are not equivalent. To express the above matrices $N(z)$ and $D(z)$ in (E.14) as $P(z)$ and $Q(z)$ of (E.3), we need to re-arrange the

polynomials in the matrices to be in sub-matrices containing co-efficients of descending power of z , i.e., for a causal $P(z)$, we have

$$P(z) \triangleq \left[\begin{array}{cccc} \bar{n}_{11,0} & \bar{n}_{12,0} & \cdots & \bar{n}_{1m,0} \\ \bar{n}_{21,0} & \bar{n}_{22,0} & \cdots & \bar{n}_{2m,0} \\ \vdots & \vdots & & \vdots \\ \bar{n}_{p1,0} & \bar{n}_{p2,0} & \cdots & \bar{n}_{pm,0} \end{array} \right] + \left[\begin{array}{cccc} \bar{n}_{11,1} & \bar{n}_{12,1} & \cdots & \bar{n}_{1m,1} \\ \bar{n}_{21,1} & \bar{n}_{22,1} & \cdots & \bar{n}_{2m,1} \\ \vdots & \vdots & & \vdots \\ \bar{n}_{p1,1} & \bar{n}_{p2,1} & \cdots & \bar{n}_{pm,1} \end{array} \right] z^{-1} + \dots + \left[\begin{array}{cccc} \bar{n}_{11,n} & \bar{n}_{12,n} & \cdots & \bar{n}_{1m,n} \\ \bar{n}_{21,n} & \bar{n}_{22,n} & \cdots & \bar{n}_{2m,n} \\ \vdots & \vdots & & \vdots \\ \bar{n}_{p1,n} & \bar{n}_{p2,n} & \cdots & \bar{n}_{pm,n} \end{array} \right] z^{-n} \quad (\text{E.15})$$

and

$$Q(z) \triangleq \left[\begin{array}{cccc} l_{11,0} & 0 & \cdots & 0 \\ 0 & l_{22,0} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & l_{pm,0} \end{array} \right]^{-1} \left[\begin{array}{cccc} l_{11,1} & 0 & \cdots & 0 \\ 0 & l_{22,1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & l_{pm,1} \end{array} \right] z^{-1} + \dots + \left[\begin{array}{cccc} l_{11,M} & 0 & \cdots & 0 \\ 0 & l_{22,M} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & l_{pm,M} \end{array} \right] z^{-M} \quad (\text{E.16})$$

where,

$$P(z) = \left[P_0 + P_1 z^{-1} + \dots + P_n z^{-n} \right] \quad (\text{E.17})$$

$$Q(z) = \left[Q_1 z^{-1} + \dots + Q_M z^{-M} \right] \text{ and } Q_0 \triangleq I \quad (\text{E.18})$$

$\bar{n}_{ij,k}$: represents the k^{th} filter tap for the ij^{th} modified numerator filter, and

$l_{ij,k}$: represents the k^{th} filter tap for the ij^{th} common column denominator filter

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] 3GPP, Multiplexing and Channel Coding (FDD), in TS 25.212, V3.3.0. 2002.
- [2] 3GPP, Physical channels and mapping of transport channels onto physical channels (FDD), in TS 25.211, V3.3.0. 2002.
- [3] 3GPP, Spreading and Modulation (FDD), in TS 25.213, V4.5.0. 2002.
- [4] 3GPP, UE Radio Transmission and Reception (FDD), in TR 25.101, V4.5.0. 2002.
- [5] Amari, S., Theory of Adaptive Pattern Classifiers. IEEE Trans. on Electronic Computers, 1967. EC-16(3): p. 299-307.
- [6] Amari, S., Differential-Geometrical Methods in Statistics Lecture Notes in Statistics 1985. 28.
- [7] Amari, S., Learning and Statistical Inference, in Hand-book of Brain Theory and Neural Networks, M.A. Arbib, Editor. 1995, MIT Press: Cambridge, MA. p. 522-526.
- [8] Amari, S. Neural Learning in Structured Parameter Spaces-Natural Riemannian gradient. in Neural Info. Processing Systems, NIPS-96. 1996: MIT Press.
- [9] Amari, S., Natural Gradient Works Efficiently in Learning. Neural Computation, 1998. 10: p. 251–276.
- [10] Amari, S., Natural Gradient Learning for Over- and Undercomplete Bases in ICA. Neural Computation, 1999. 11(8): p. 1875-1883.
- [11] Amari, S. and J.F. Cardoso, Blind Source Separation: Semi-Parametric Statistical Approach. IEEE Trans. on Signal Process, 1997. 45(11): p. 2692–2700.
- [12] Amari, S., T.-P. Chen, and A. Cichocki, Stability Analysis of Adaptive Blind Source Separation. Neural Networks, 1997. 10 (8): p. 1345–1352.
- [13] Amari, S., A. Cichocki, and H.H. Yang, A New Learning Algorithm for Blind Signal Separation, in Advances in Neural Information Processing Systems. 1996, MIT Press: Cambridge, MA. p. 757-763.
- [14] Amari, S., S.C. Douglas, A. Cichocki, and H.H. Yang, Multichannel Blind Deconvolution and Equalization using the Natural Gradient, in Proceeding of IEEE Workshop on Signal Processing, Adv. in Wireless Communications. 1997: Paris, France. p. 101–104.

- [15] Amari, S. and M. Kawanabe, *Information Geometry of Estimating Functions in Semiparametric Statistical Models*. Bernoulli, 1997. 3: p. 29-54.
- [16] Amari, S. and H. Nagaoka, *Methods of Information Geometry*. 1999: AMS and Oxford University Press.
- [17] Antsaklis, P.J. and A.N. Michel, *Linear Systems*. 1997: Electrical and Computer Engineering Series, McGraw-Hill.
- [18] Araki, S., S. Makino, R. Mukai, T. Nishikawa, and H. Saruwatari, *Fundamental Limitation of Frequency Domain Blind Source Separation for Convolved Mixture of Speech*, in 3rd Int'l Symposium on ICA and BSS, ICA-2001. 2001: San Diego, CA. p. 132-137.
- [19] Attkick, J.J., *Could Information Theory provide an Ecological Theory of Sensory Processing ?* *Network: Computation in Neural Systems*, 1992. 3(2): p. 213-251.
- [20] Barlow, H.B., *The coding of sensory messages*, in *Current Problems in Animal Behavior*, W.H. Thorpe and O.L. Zangwill, Editors. 1961, Cambridge University Press: MA. p. 331-361.
- [21] Bell, A.J. and T.J. Sejnowski, *An Information-Maximization Approach to Blind Separation and Blind Deconvolution*. *Neural Computation*, 1995. 7: p. 1129–1159.
- [22] Benveniste, A., M. Goursat, and G. Ruget, *Robust Identification of a Non-minimum Phase System: Blind Adjustment of a Linear equalizer in Data Communications*. *IEEE Trans. on Automatic Control*, 1980. 25: p. 385–399.
- [23] Bofill, P. and M. Zibulevsky, *Blind Separation of More Sources than Mixtures using the Sparsity of their Short Term Fourier Transform*, in *Proc. of Int'l Symposium on Independent Component Analysis and Blind Source Separation, ICA-2000*. 2000: Helsinki, Finland. p. 87-92.
- [24] Bofill, P. and M. Zibulevsky, *Underdetermined Blind Source Separation using Sparse Representation*. *Signal Processing*, 2001. 81(11): p. 2353-2362.
- [25] Broeck, C.V.d. and P. Reimann, *Unsupervised Learning by Examples: On-line versus Off-line*. *Physical Review Letters*, 1996. 76: p. 2181-2191.
- [26] Cao, J. and N. Murata, *A Stable and Robust ICA Algorithm Based on t-distribution and Generalized Gaussian Distribution Models*, in *Neural Networks for Signal Processing IX*. 1999. p. 283-292.
- [27] Cardoso, J.F., *Blind Signal Processing: Statistical Principles*. *Proceedings of the IEEE*, 1998. 90(8): p. 2009–20026.
- [28] Cardoso, J.F., *Entropic Contrasts for Source Separation: Geometry and Stability*, Chapter 4, in *Unsupervised Adaptive Filtering*. 1999, John Wiley & Sons.

- [29] Cardoso, J.F. and B. Laheld, Equivariant Adaptive Source Separation. IEEE Trans. on Signal Processing, 1996. 44: p. 3017–3030.
- [30] Charlier, G., Application, de la theorie des probabilités a l'astronomie, Part of the traite. 1931, Paris: Gauthier-Villars.
- [31] Chen, S.S., D.L. Donoho, and M.A. Saunders, Atomic Decomposition by Basis Pursuit, in Technical Report. 1996: Dept. of Statistics, Stanford University, CA
- [32] Chevalier, P., On the Performance of Higher Order Blind Source Separation Methods, in IEEE-ATHOS workshop on HOS. 1995: Begur Spain. p. 30-34.
- [33] Cichocki, A. and S. Amari, Adaptive Blind Signal and Image Processing. 2002: John Wiley and Sons.
- [34] Cichocki, A., S. Amari, and J. Cao, Neural Network Models for Blind Separation of Time Delayed and Convolved Signals. IECE Trans. Fundamentals, 1997. E82-A(9): p. 1595-1603.
- [35] Cichocki, A., S.I. Amari, and J. Cao. Blind Separation of Delayed and Convolved Sources with Self Adaptive Learning Rate. in Int'l Symposium on Nonlinear Theory and Applications, NOLTA-96. 1996. Kochi, Japan.
- [36] Cichocki, A., R. Unbehauen, and E. Rummert, Robust learning algorithm for blind separation of signals. Electronics Letters, 1994. 30(17): p. 1386-1387.
- [37] Common, P., Independent Component Analysis, a new concept ? Signal Processing, 1994. 36(3): p. 287-314.
- [38] Cover, T.M. and J.A. Thomas, Elements of Information Theory. 1991: Wiley Interscience Series.
- [39] Cristescu, R., T. Ristaniemi, J. Joutsensalo, and J. Karhunen, Blind Separation of Convolved Mixtures for CDMA Systems, in Proc. of the European Signal Processing Conference, EUSIPCO-2000. 2000: Tampere, Finland. p. 619-622.
- [40] Erten, G. and F.M. Salam, Real Time Separation of Audio Signals Using Digital Signal Processors, in Proceedings of the 40th Midwest Symposium on Circuits and Systems. 1998. p. 1237 -1240.
- [41] Erten, G. and F.M. Salam, Voice Output Extraction by Signal Separation, in Proceedings of the 1998 IEEE International Symposium on Circuits and Systems. 1998. p. 5–8.
- [42] Erten, G. and F.M. Salam, Voice Extraction by On-line Signal Separation and Recovery. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, 1999. 46(7): p. 915–922.

- [43] Gaeta, M. and J.-L. Lacounme, Source Sparation without a priori Knowledge: The Mximum Lkelihood Slution, in Proc. Eur. Signal Processing Conf. 1990. p. 621-624.
- [44] Gesbert, D., J. Sorelius, P. Stoica, and A. Paulraj, Blind Multi-User MMSE Detector for CDMA Signals in ISI channels. IEEE Communications Letters, 1999. 3(8): p. 233-235.
- [45] Gharbi, A.B.A. and F.M. Salam, Separation of Mixed Signals in Dynamic Environments: Formulation and some Implementation, in Proc. of 37th Midwest Symposium on Circuits and Systems. 1994: Lafayette, LA. p. 587-590.
- [46] Gharbi, A.B.A. and F.M. Salam, Implementation and Test Results of a Chip for the Separation of Mixed Signals. IEEE Transactions on Circuits and Systems, 1995. 42(11): p. 748–751.
- [47] Gharbi, A.B.A. and F.M. Salam, Algorithms for Blind Signal Separation and Recovery in Static and Dynamic Environments, in Proceedings of 1997 Int’l Symp. on Circuits and Systems. 1997: Hong Kong. p. 713–716.
- [48] Girolami, M., An Alternative Perspective on Adaptive Independent Component Analysis Algorithms, in Technical Report, Computing and Information Systems. 1998: Paisley University, Scotland.
- [49] Girolami, M., A.J. Bell, and T.J. Sejnowski, A Unifying Information Theoretic Framework for Independent Component Analysis. Int’l Journal on Mathematical and Computer Modeling 2000. 38: p. 1–21.
- [50] Girolami, M. and C. Fyfe, Extraction of Independent Signal Sources using a Deflationary Exploratory Projection Pursuit Network with Lateral Inhibition. IEE Proceedings on Vision, Image and Signal Processing, 1997. 14(5): p. 299-306.
- [51] Girolami, M. and C. Fyfe, Generalised Independent Component Analysis through Un-supervised Learning With Emergent Bussgang Properties, in Proc. International Conference on Neural Networks. 1997: Houston, TX. p. 1788-1791.
- [52] Haykin, S., Adaptive Filter Theory. 4 ed. 1996: Prentice-Hall.
- [53] Haykin, S., Neural Networks: A Comprehensive Foundation. 2 ed. 1998: Prentice Hall.
- [54] Haykin, S., Unsupervised Adaptive Filtering, ed. S. Haykin. Vol. I and II. 2000: John Wiley and Sons.
- [55] Heikkila, M.J., A Novel Blind Adaptive Algorithm for Channel Equalization in WCDMA Downlink, in 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC-01. 2001: San Diego, CA. p. A-41--A-45.

- [56] Herault, J. and C. Jutten, Space or Time Adaptive Signal Processing by Neural Network Models AIP Conference Proceedings, 1986. 151: p. 206-211.
- [57] Holma, H. and A. Toskala, WCDMA for UMTS Radio Access for Third Generation Mobile Communications. 2001: John Wiley & Sons, Inc.
- [58] Honig, M., U. Madhow, and S. Verdú, Blind Adaptive Multiuser Detection. IEEE Trans. on Information Theory, 1995. 41(4): p. 944-960.
- [59] Hooli, K., M. Latva-Aho, and M. Juntti, Performance Evaluation of Adaptive Chip-Level Channel Equalizers in WCDMA Downlink, in IEEE International Conference on Communications. 2001. p. 1974–1979.
- [60] Hyvarinen, A., J. Karhunen, and E. Oja, Independent Component Analysis. 2001: Wiley Inter-Science Series.
- [61] Hyvarinen, A. and E. Oja, A Fast Fixed-Point Algorithm for Independent Component Analysis. Neural Computation, 1997. 9: p. 1483–1492.
- [62] Joutsensalo, J. and T. Ristaniemi, Blind Multi-user Detection by Fast Fixed Point Algorithm without Prior Knowledge of Symbol Level Timing, in Proc. of the IEEE Signal Processing Workshop on Higher Order Statistics. 1999: Ceasarea, Israel. p. 305-308.
- [63] Jutten, C. and J. Herault, Blind separation of sources, an adaptive algorithm based on neuromimetic architecture. Signal Processing, 1991. 24(1): p. 1-31.
- [64] Karhunen, J., Neural Approaches to Independent Component Analysis and Source Separation, in Proc. of 4th European Symp. on Artificial Neural Networks (ESANN-96). 1996: Bruges, Belgium. p. 249-266.
- [65] Karhunen, J. and J. Joutsensalo, Representation and Separation of Signals Using Non-linear PCA Type Learning. Neural Networks, 1994. 7(1): p. 113-127.
- [66] Kärkkäinen, K.H.A. and P.A. Leppänen, The Influence of Initial-Phases of a PN Code Set on the Performance of an Asynchronous DS-CDMA System. Wireless Personal Communications, 2000. 3(3): p. 279-293.
- [67] Karvanen, J., E. J., and K. V., Adaptive Score Functions for Maximum Likelihood ICA. Journal of VLSI Signal Processing, 2002. 32: p. 83-92.
- [68] Karvanen, J. and V. Koivunen, Blind Separation Methods Based on Pearson system its Extensions. Signal Processing, 2002. 82(4): p. 663-673.
- [69] Kendall, M. and A. Stuart, The Advanced Theory of Statistics. Vol. I. 1977: Charles Griffin and Co. Ltd.

- [70] Kenney, J.F. and E.S. Keeping, The k-Statistics, in Mathematics of Statistics, V. Nostrand, Editor. 1962: Princeton, NJ. p. 99-100.
- [71] Khalil, H.K., Nonlinear Systems. 3 ed. 2002: Prentice Hall.
- [72] Kullback, S., Topics in Statistical Information Theory. 1987, New York: Springer-Verlag.
- [73] Lambert, R., Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures. 1996.
- [74] Lambert, R. and A.J. Bell, Blind separation of Multiple Speakers in a Multipath Environment, in Int'l Conference of Acoustics, Speech and Signal Processing, ICASSP-97. 1997: Munich, Germany. p. 423-426.
- [75] Lee, T.W., A.J. Bell, and R. Lambert, Blind separation of Delayed and Convolved Sources. Advances in Neural Information Processing Systems, 1997. 9: p. 758-764.
- [76] Lee, T.W., A.J. Bell, and R. Orglmeister, Blind Source Separation of Real-World Signals, in Proc. of IEEE Int'l Conference on Neural Networks ICNN-97. 1997: Houston, TX. p. 2129-2135.
- [77] Lee, T.W. and M. Girolami, Advances in Independent Component Analysis. 2000: Springer Verlag.
- [78] Lee, T.-W., M. Girolami, and T.J. Sejnowski, Independent Component Analysis using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources. Neural Computation, 1999. 11(2): p. 417-441.
- [79] Lee, T.W. and T.J. Sejnowski, Independent Component Analysis for Sub-Gaussian and Super-Gaussian Mixtures, in 4th Joint Symposium on Neural Computation. 1997, Institute for Neural Computation. p. 132-140.
- [80] Lewicki, M. and T.J. Sejnowski, Learning Overcomplete Representations. Neural Computation 2000. 12: p. 337-365.
- [81] Lewis, F.L. and V.L. Syrmos, Optimal Control. 2 ed. 1995: Wiley Interscience Series.
- [82] Li, H. and H.V. Poor, Iterative Channel Estimation and Multiuser Detection in Multipath CDMA Channels, in Proc. of IEEE Int'l Conference on Communications 2003: Anchorage, Alaska, USA. p. to appear.
- [83] Linsker, R., A local learning rule that enables information maximization for arbitrary input distributions. Neural Computations, 1997. 9(8): p. 1661-1665.
- [84] Mahram, H.H., H.E. Boll, and G. Alirezaei, Performance Evaluation of Advanced Receivers for WCDMA Downlink Detection, in Proc. of 5th Int'l Symposium on

Wireless Personal Multimedia Communications, WPMC-2002. 2002: Honolulu, Hawaii, USA.

[85] Mathis, H., T.P. Von-Hoff, and M. Joho, Blind Separation of Signals with Mixed Kurtosis Signs using Threshold Activation Functions. *IEEE Transactions on Neural Networks*, 2001. 12(3): p. 618-624.

[86] Michel, A.N. and D. Liu, *Qualitative Analysis and Synthesis of Recurrent Neural Networks*. 2002, New York: Marcel Dekker.

[87] Miller, J.H. and J.B. Thomas, Detectors for Discrete-Time Signals in Non-Gaussian noise. *IEEE Transactions on Information Theory*, 1972. IT-18(2): p. 241-250.

[88] Nadal, J.P. and N. Parga, Redundancy Reduction and Independent Component Analysis: Conditions on Cumulants and Adaptive approaches. *Neural Computation*, 1997. 9(7): p. 1421-1456.

[89] Nicholls, J., *From Neuron to Brain*. 4 ed. 2001: Sinauer Associates Incorporated.

[90] Nishikawa, T., H. Saruwatari, and K. Shikano, Blind Source Separation Based on Multi-Stage ICA Using Frequency-Domain ICA and Time-Domain ICA, in *Int'l conference on Fundamentals of Electronics, Communications and Computer Sciences, ICFS-2002*. 2002. p. 7-12.

[91] Opper, M., Online versus Offline Learning from Random Examples: General Results. *Physics Review Letters*, 1996. 77: p. 4671-4674.

[92] Papoulis, A., *Probability, Random Variables and Stochastic Processes*. 1984, New York: McGraw-Hill.

[93] Parzen, E., On the Estimation of a Probability Density Function and the Mode. *Annals Mathematical Statistics*, 1962. 33: p. 1065-1076.

[94] Pearlmutter, B. and L. Parra, A Context-Sensitive Generalization of Independent Component Analysis, in *1996 International Conference on Neural Information Processing*. 1996: Hong Kong.

[95] Pearson, K., *Contributions to the Mathematical Study of Evolution*. *Philosophical Transactions of the Royal Society A*, 1894. 185: p. 71-110.

[96] Prator, O., C. Unger, A. Zoch, and G.P. Fettweis, Performance of Adaptive Chip Equalization for the WCDMA Downlink in Fast Changing Environments, in *IEEE 7th Int'l Symposium on Spread Spectrum Technology and Applications, ISSSTA-02*. 2002: Prague, Czech Republic. p. 273-277.

[97] Principe, J., D. Xu, and J. Fisher, Information-Theoretic Learning, Chapter 7., in *Unsupervised Adaptive Filtering*, S. Haykin, Editor. 1999, John Wiley & Sons.

- [98] Proakis, J.G., *Digital Communications*. 4 ed. 2001: McGraw-Hill.
- [99] Puntonet, C.G. and A. Prieto, An Adaptive Geometrical Procedure for Blind Separation of Sources. *Neural Processing Letters*, 1995. 2.
- [100] Renyi, A., On measures of Entropy and Information. 1 ed. *Selected Papers of Alfred Renyi*. Vol. 2. 1976: Budapest Akademia Kiado.
- [101] Ristaniemi, T. and J. Joutsensalo, Advanced ICA-based Receivers for Block Fading DS-CDMA Channels. *Signal Processing*, 2002. 82(3): p. 417-431.
- [102] Roth, Z. and Y. Baram, Multi-dimensional Density Shaping by Sigmoids. *IEEE Transactions on Neural Networks*, 1996. 7(5): p. 1291-1298.
- [103] Rugh, W.J., *Linear System Theory*. 2 ed. 1996: Prentice Hall, NJ.
- [104] Rumelhart, D.E., G.E. Hinton, and R.J. Williams, Learning Internal Representations by Error Propagation, in *Parallel Distributed Processing*, S. Amari, Editor. 1986, MIT Press: Cambridge, MA. p. 318-362.
- [105] Sabala, I., A. Cichocki, and S. Amari. Relationships between Instantaneous Blind Source Separation and Multichannel Blind Deconvolution. in *Proc. Int. Joint Conference on Neural Networks*. 1998. Alaska USA.
- [106] Salam, F.M., An Adaptive Network for Blind Separation of Independent Signals, in *Proc. of International Symposium on Circuits and Systems*. 1993: Chicago, IL. p. 431-434.
- [107] Salam, F.M. and G. Erten, Blind Signal Separation and Recovery in Dynamic Environments, in *Proceedings of 3rd IEEE Workshop on Nonlinear Signal and Image Processing 1997: Mackinac Island, MI*.
- [108] Salam, F.M. and G. Erten, Exact Entropy Series Representation for Blind Source Separation, in *Proceedings of IEEE SMC Conference*. 1999. p. 553-558.
- [109] Salam, F.M. and G. Erten, The State Space Framework for Blind Dynamic Signal Extraction and Recovery, in *Proceedings of the IEEE International Symposium on Circuits and Systems*. 1999. p. 66-69.
- [110] Salam, F.M., G. Erten, and K. Waheed, Blind Source Recovery: Algorithms for Static and Dynamic Environments, in *Proc. of INNS-IEEE International Joint Conference on Neural Networks*. 2001. p. 902-907.
- [111] Salam, F.M., A.B.A. Gharbi, and G. Erten, Formulation and Algorithms for Blind Signal Recovery, in *Proceedings of the 40th Midwest Symposium on Circuits and Systems*. 1998. p. 1233-1236.

- [112] Salam, F.M. and K. Waheed, State-Space Feedforward and Feedback Structures for Blind Source Recovery, in Proc. of 3rd International Conference on Independent Component Analysis and Blind Signal Separation. 2001: San Diego, CA. p. 248–253.
- [113] Schobben, D., K. Torkkola, and P. Smaragdis, Evaluation of Signal Separation Methods, in Proc. of 1st Intl. workshop on ICA and BSS, ICA-99. 1999: Aussois, France. p. 261-266.
- [114] Taleb, A. and C. Jutten, Sources Separation in Post-Nonlinear Mixtures. IEEE Transactions on Signal Processing, 1999. 10(47): p. 2807–2820.
- [115] Theis, F.J., A. Jung, E.W. Lang, and C.G. Puntonet, A Theoretical Model for Linear Geometric ICA, in Proc. of 3rd Int'l Symposium on Independent Component Analysis and Blind Source Separation, ICA-2001. 2001. p. 349-354.
- [116] Theis, F.J. and E.W. Lang, Formalization of Two-Step Approach to Overcomplete BSS; , in IASTED Int'l Conf. on Signal & Image Proc. SIP-2002. 2002: Hawaii, USA.
- [117] Theis, F.J. and E.W. Lang, Geometric Overcomplete ICA, in Proc. of 10th European Symp. on Artificial Neural Networks, ESANN 2002. 2002: Bruges, Belgium. p. 217-223, .
- [118] Torkkola, K., Blind Separation of Convolved Sources based on Information Maximization, in IEEE Workshop on Neural Networks for Signal Processing. 1996: Kyoto, Japan. p. 423–432.
- [119] Tsybkin, Y.Z., Foundation of the Theory of Learning Systems. 1973, New York: Academic Press.
- [120] Verdu, S. Minimum Probability of Error for Asynchronous Gaussian Multiple-Access Channels. in IEEE Transactions on Information Theory. 1986.
- [121] Verdu, S., Multiuser Detection. 1998: Cambridge University Press.
- [122] Waheed, K., K. Desai, and F.M. Salem, Adaptive RAKE-Blind Source Recovery Algorithms for 3GPP UMTS/WCDMA Downlink Receivers, in 2003 IEEE International Conference on Robotics, Intelligent Systems & Signal processing, RISSP-2003. 2003: Changsha, Hunan, China. p. to appear.
- [123] Waheed, K., K. Desai, and F.M. Salem, Blind Multi User Detection in DS-CDMA Systems using the Natural Gradient based Symbol Recovery Structures, in 4th International Symposium on Independent Component Analysis and Blind Source Separation, ICA-2003. 2003: Nara, Japan. p. 727-732.
- [124] Waheed, K., K. Desai, and F.M. Salem, Natural Gradient based Blind Multi User Detection in QPSK DS-CDMA Systems, in 2003 IEEE-INNS Joint International Conference on Neural Networks, IJCNN-2003. 2003: Portland OR. p. to appear.

- [125] Waheed, K., K. Desai, and F.M. Salem, Blind Info-Theoretic Multi-User Detection Algorithms for DS-CDMA and WCDMA Downlink Systems. IEEE Transactions on Neural Networks, submitted 2003: p. to appear.
- [126] Waheed, K. and F.M. Salam, Blind Source Recovery: Some Implementation and Performance Issues, in Proceedings of the 44th IEEE Midwest Symposium on Circuits and Systems. 2001: Dayton, OH. p. 694–697.
- [127] Waheed, K. and F.M. Salam, Blind Source Recovery using an Adaptive Generalized Gaussian Score Function, in 45th IEEE Int'l Midwest Symposium on Circuits and Systems. 2002: Tulsa, OK p. 418-421.
- [128] Waheed, K. and F.M. Salam, Cascaded Structures for Blind Source Recovery, in 45th IEEE Int'l Midwest Symposium on Circuits and Systems. 2002: Tulsa, OK. p. 656-659.
- [129] Waheed, K. and F.M. Salam, State Space Blind Source Recovery of Non-minimum Phase Environments, in 45th IEEE International Midwest Symposium on Circuits and Systems. 2002: Tulsa, OK. p. 422-425.
- [130] Waheed, K. and F.M. Salam, State-Space Blind Source Recovery for Mixtures of Multiple Source Distributions, in Proc. of IEEE Int'l Symposium on Circuits and Systems. 2002: Scottsdale, Arizona. p. 197–200.
- [131] Waheed, K. and F.M. Salam, A Data-Derived Quadratic Independence Measure for Adaptive Blind Source Recovery in Practical Applications, in 45th IEEE Int'l Midwest Symp. on Circuits and Systems. 2002 Tulsa, OK. p. 473-476.
- [132] Waheed, K. and F.M. Salem, Algebraic Independent Component Analysis, in 2003 IEEE International Conference on Robotics, Intelligent Systems & Signal processing, RISSP-2003. 2003: Changsha, Hunan, China. p. to appear.
- [133] Waheed, K. and F.M. Salem, Algebraic Independent Component Analysis: An approach for Separation of Overcomplete Speech Mixtures, in 2003 IEEE-INNS Joint International Conference on Neural Networks, IJCNN-2003. 2003: Portland OR. p. to appear.
- [134] Waheed, K. and F.M. Salem, Algebraic Overcomplete Independent Component Analysis, in 4th International Symposium on Independent Component Analysis and Blind Source Separation, ICA-2003. 2003: Nara, Japan. p. 1077-1082.
- [135] Waheed, K. and F.M. Salem, Blind Source Recovery: A State Space Formulation, in Contemporary Issues in System Stability and Control with Applications, D. Liu and P.J. Antsaklis, Editors. 2003, Birkhauser Boston, MA. p. 171-196.
- [136] Waheed, K. and F.M. Salem, New Hyperbolic Source Density Models for Blind Source Recovery Score Functions, in IEEE International Symposium on Circuits and Systems. 2003: Bangkok, Thailand. p. to appear.

- [137] Waheed, K. and F.M. Salem, Algebraic Independent Component Analysis and its Application to Overcomplete Blind Source Separation. Neural Processing Letters, submitted 2003: p. to appear.
- [138] Waheed, K. and F.M. Salem, Blind Source Recovery for Multiple Source Distributions using Adaptive Score Functions. Neural Processing Letters, submitted 2003: p. to appear.
- [139] Waheed, K. and F.M. Salem, Blind Source Recovery: A Framework in the State Space. Journal of Machine Learning Research, submitted 2003: p. to appear.
- [140] Waheed, K. and F.M. Salem, State Space Feedforward and Feedback Structures for Blind Source Recovery. Neural Processing Letters, submitted 2003: p. to appear.
- [141] Wang, X. and H.V. Poor, Blind Equalization and Multiuser Detection in Dispersive CDMA Channels. IEEE Transactions on Communications, 1998. 46(1): p. 91-103.
- [142] Wang, X. and H.V. Poor, Subspace Methods for Blind Joint Channel Estimation and Multiuser Detection in CDMA Systems. Wireless Networks, 2000. 6(1): p. 59-71.
- [143] Website, 3rd Generation Partnership Project Website. <http://www.3gpp.org>. 2003.
- [144] Website, Blind Source Recovery. <http://www.egr.msu.edu/bsr>. 2003.
- [145] Website, CDMA Group website. <http://www.cdg.org/index.asp>. 2003.
- [146] Werner, S. and J. Lilleberg, Downlink Channel Decorrelation in CDMA Systems with long Codes, in IEEE Vehicular Technology Conference, VTC-99 1999: Houston, Texas, USA. p. 1614 –1617.
- [147] Widrow, B., A Statistical Theory of Adaptation. 1963: Oxford Pergamon Press.
- [148] Woods, J.W. and H. Stark, Probability and Random Processes with Applications to Signal Processing. 3 ed. 2001: Prentice Hall.
- [149] Wu, Y., Q. Zhang, W. Zhu, and S.-Y. Kung, Spreading Code Assignment in an Ad-hoc DS-CDMA wireless network, in International Conference on Communications, ICC-2002. 2002: New York, NY. p. 3066 –3070.
- [150] Xu, D., Energy, Entropy and Information Potential in Neurocomputing, in Ph.D. Thesis 1998, Department of Electrical Engineering, University of Florida: Florida.
- [151] Yang, H.H., S. Amari, and A. Cichocki, Information Back-Propagation for Blind Separation of Sources from Non-linear Mixtures, in Proc. of Int'l Conference on Neural Networks ICNN-97. 1997: Houston, TX. p. 2141–2146.

- [152] Zhang, J.-G., A.B. Sharmaand, and W.C. Kwong, Optical CDMA-based ATM Switches Supporting Equal-Bit-Rate and Variable-Bit-Rate Communication Services, in IEEE Global Telecommunications Conference, GLOBECOM -98. 1998. p. 3221 -3226.
- [153] Zhang, L.Q., S. Amari, and A. Cichocki, Natural Gradient Approach to Blind Separation of Over- and Under-complete Mixtures, in First International Workshop on Independent Component Analysis and signal Separation (ICA99). 1999: Aussois, France. p. 455-460.
- [154] Zhang, L.Q., S. Amari, and A. Cichocki, Semiparametric Approach to Multichannel Blind Deconvolution of Non-minimum Phase Systems. Advances in Neural Information Processing Systems 2000. 12: p. 363-369.
- [155] Zhang, L.Q. and A. Cichocki, Blind Deconvolution of Dynamical Systems: A State Space Approach. Journal of Signal Processing, 2000. 4(2): p. 111-130.
- [156] Zhang, L.Q., A. Cichocki, and S. Amari, Multichannel Blind Deconvolution of Nonminimum Phase Systems using Information Backpropagation, in Proc. of 5th Int'l Conf. on Neural Information Processing. 1999: Perth, Australia. p. 210-216.
- [157] Zhang, L.Q., A. Cichocki, and S. Amari, Kalman Filter and State-Space Approach to Multichannel Blind Deconvolution, in IEEE Workshop on Neural Networks for Signal Processing. 2000: Sydney, Australia. p. 425-434.
- [158] Zibulevsky, M. and B.A. Pearlmutter, Blind Source Separation by Sparse Decomposition in a Signal Dictionary. Neural Computation, 2001. 13(4): p. 863-882.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02470 8632